SCIENTIFIC AMERICAN

BREACHING WHALE

$2.50

March 1985

Pure Intelligence

The 1985 Mitsubishi Galant is perhaps the most ingeniously engineered luxury sedan in the world.

The beauty of its brilliance begins with its clean, aerodynamic lines.

A 2.4-liter computerized, fuel-injected engine provides smooth, responsive power. An electronically controlled, four-speed automatic overdrive transmission manages it with precision and ease.

Its electronic power steering improves parking and road-feel.

An available ECS™ electronically controlled suspension system monitors load and speed conditions, and automatically adjusts itself.

From the ergonomic, six-way driver's seat everything is at your fingertips. You command a host of assist items with an electronic control module at your disposal.

Mitsubishi Galant

Call 1-800-447-4700 for your nearest Mitsubishi Motors Dealer.

Pure Indulgence.

The ETACS III™ system directs several timed convenience functions by means of a microprocessor.

An optional stereo even offers you steering wheel-mounted switches.

The rear cabin is first class, too. With sumptuous reclining seats. A center console. Two-way headrests. And adjustable thigh supports.

The list of opulent features and advanced electronics is seemingly endless. But the new Galant has one clever innovation you will probably appreciate more than all the others.

A totally rational price tag.

The ingenious new Galant Sedan is indeed a luxurious example of high technology that is pure intelligence, pure indulgence.

And pure Mitsubishi.

MITSUBISHI MOTORS

Takes you where you want to be.™

Established 1845 **SCIENTIFIC AMERICAN** March 1985 Volume 252 Number 3

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $24 PER YEAR IN THE U.S. AND ITS POSSESSIONS; $33 PER YEAR IN ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.

SCIENTIFIC AMERICAN

CORRESPONDENCE

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

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

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

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 5969, New York, NY 10017. The date of the last issue on your subscription is shown in the upper right-hand corner of each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, on a mailing label of a recent issue) as well as the new one.

Name

New Address

Street

City

State and ZIP

Old Address

Street

City

State and ZIP

THE COVER

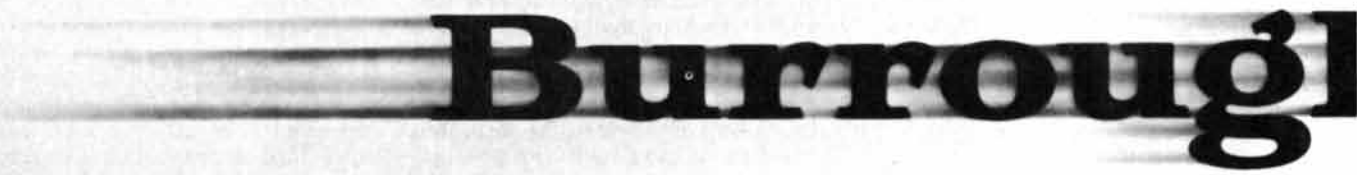
The painting on the cover portrays a humpback whale (*Megaptera novaeangliae*) leaping, or, in the language of both 18-century whalers and modern investigators of whale behavior, breaching. The breach is a purposeful maneuver. Its association with social activities suggests that it is a form of communication (see "Why Whales Leap," by Hal Whitehead, page 84). The whale on the cover is making a "true breach," the commonest form, in which the animal emerges from the water on its side, twists in the air and lands on its back. The other form of breach is the "belly flop," in which the whale emerges from the water with its back uppermost and lands on its belly. Clustered on the throat grooves of the whale in the painting are several species of parasite, including whale lice and acorn barnacles, which in turn carry goose barnacles. A humpback may carry half a ton of parasites.

THE ILLUSTRATIONS

Cover painting by Peter Ott

Page	Source	Page	Source
14–20	Edward Bell	55	Janet Finan and William Fore
38	AP/Wide World Photos	56–57	Ilil Arbel
40	Andrew Christie	58	Janet Finan and William Fore
41	Earth Satellite Corporation	59–60	Ilil Arbel
42	Jet Propulsion Laboratory, California Institute of Technology	75–80	Ian Worpole
43–44	Andrew Christie	85	Lewis Trusty, Animals, Animals
47	Larry W. Esposito, University of Colorado at Boulder	86–93	Patricia J. Wynne
48	Eric M. Eliason, U.S. Geological Survey (*top*); Jet Propulsion Laboratory, California Institute of Technology (*bottom*)	95	West Air Photography
49	Donald B. Campbell, Arecibo Observatory	96	Andrew Tomko
50	Peter G. Ford and Gordon H. Pettengill, Massachusetts Institute of Technology	97–100	Alan D. Iselin
51–52	Ilil Arbel	103	J. Narayan, Oak Ridge National Laboratory
53	V. L. Barsukov, V. I. Vernadsky Institute of Geochemistry and Analytical Chemistry	104–109	Gabor Kiss
		110	Gabor Kiss (*left*), Sandia National Laboratories (*right*)
		112	Gabor Kiss
		115–118	Jerome Kuhl
		122	C. W. Schwartz, Animals, Animals
		124–126	Michael Goodman

In a major independent survey, one computer company consistently had the fastest service response time.



Survey of user satisfaction with customer service—
International Data Corporation, July 1984

LETTERS

To the Editors:

In his article "The Spanish Ship of the Line" [SCIENTIFIC AMERICAN, December, 1984] John D. Harbron shows a talent for bringing life to history. I believe, however, that the diagram of the *Santísima Trinidad* on pages 118 and 119 errs in depicting the ship's figurehead as a royal lion. On page 160 of Geoffrey Bennett's excellent book *The Battle of Trafalgar* (Naval Institute Press, 1977) Bennett quotes a Midshipman Badcock's description of the *Trinidad* as she drew near the English line: "Her appearance was imposing; her head splendidly ornamented with a colossal group of figures, painted white, representing the Holy Trinity, from which she took her name."

Badcock also described the exterior paint of the *Trinidad* as "four distinct lines of red, with white ribbon between them." This is in contrast to the painting in the article, which may illustrate an earlier color scheme.

In recounting the virtues of the *navíos,* Mr. Harbron neglects to mention the ineptitude of their crews. Spain had certainly been eclipsed as a naval power in the two centuries since the Armada; it is known that, much as Nelson considered French seamen inferior to his own tars, the French admiral Villeneuve thought of Spanish sailors as peasants. This may have been the aristocratic view, but the clumsy handling of the *Trinidad* at Trafalgar is a matter of history. After a preliminary tussle with H.M.S. *Victory,* she allowed H.M.S. *Neptune* to cross her stern. This is the classic maneuver known as "crossing the T," which brings the firepower of the full broadside to bear on the enemy's weakest point. The raking fire of the smaller (98-gun) *Neptune* succeeded in bringing down all three of the mighty Spaniard's masts in about 35 minutes. Once dismasted, even a magnificent weapon such as the *Santísima Trinidad* was little more than an armed barge; the best shipbuilders in Havana could not help her.

DONALD A. HOLLWAY

Dallastown, Pa.

To the Editors:

I enjoyed reading Mr. Hollway's letter and agree with him that the Spanish crews at Trafalgar did not meet the same high standards as seamen that their ships met as machines of war. The Spanish commander-in-chief, Admiral Don Federico Carlos de Gravina, knew that his crews were not ready to fight the English. He apparently said so during an acrimonious conference of senior Spanish and French naval officers at Cadiz on October 11, 1805, 10 days before the battle. Many of Gravina's men were still recovering from severe epidemics of malaria that had swept southern Spain from 1802 to 1804, including port cities from which many of his sailors came.

As for the *Santísima Trinidad*'s colors, none of the eyewitnesses at Trafalgar concur. According to Carola Oman's 1947 biography of Nelson, the ship's colors were "vermilion and white." A Lieutenant Halloran in H.M.S. *Britannia* was sure she "was of a rich lake color."

JOHN D. HARBRON

Islington, Ontario

To the Editors:

"Bells of Note" ["Science and the Citizen," SCIENTIFIC AMERICAN, December, 1984] suggests that 2,400-year-old Chinese bells were accurately tuned by an absolute standard in which reference frequencies are governed by powers of 2 such as 256 hertz. Frequencies are not absolute quantities but are measured with respect to arbitrarily defined units of time such as the second. In the absence of evidence that the Chinese of that era not only had the ability to make precise frequency measurements but also had actually adopted the second (1/86,400 of the Mean Solar Day) as a unit of time, it seems doubtful that they deliberately tuned a bell to 256 hertz.

KENNETH J. KAHN

Lawrence, N.Y.

The Editors respond:

An assumption went unstated in the "Science and the Citizen" item. If the makers of the Zeng chimes knowingly tuned their bells to frequencies that correspond to powers of 2 when they are measured in hertz (cycles per second), China in that era must have counted time in seconds. Evidence for the supposition is circumstantial. It is thought that when the bells were cast, China already divided the year into $365\frac{1}{4}$ equal days, each with 24 equal hours. Water clocks of the period are known to have been calibrated in quarter-hour increments. The case for finer subdivisions rests mainly on extensive correspondences between the cultures of ancient China and Babylon, where divisions by 60 were common, and on the apparent importance of the number 60 in ancient China.

The case for an ancient Chinese second also rests on the chime set itself—on the close correspondence between powers of 2 and the frequencies of the bells that sound octaves of C, expressed in cycles per second. The measured frequencies of the bells are 64.8, 128.6, 256.4, 514.1 and 1,024.6 hertz.

It is hard to conceive that the ancient bellmakers could count even as few as 64 vibrations per second. The beat heard when two bells are sounded at the same time, however, has a frequency that equals the greatest common factor in their frequencies. Thus bell pairs with frequencies of 64 and 72 hertz and 72 and 80 hertz would both have beat rates of eight cycles per second, perhaps a countable number. From it a bell tuner could conclude that the highest and lowest bells differ in frequency by 16 cycles per second. If by means of comparisons with the pitches sounded by strings of different lengths the tuner determined the ratio of the highest and the lowest frequency to be 5 : 4, he could correctly deduce the actual frequencies to be 64 and 80 hertz and the frequency of the intermediate bell to be 72.

The assumption that the makers of the Zeng bells meted out time in seconds and determined vibration rates by a method such as the one outlined above is one way to account for what is otherwise a remarkable coincidence.

To the Editors:

Leon M. Lederman pleads eloquently for support of fundamental science ["The Value of Fundamental Science," SCIENTIFIC AMERICAN, November, 1984]. Less noble human instincts, however, such as greed for power, still pay much higher social dividends than intellectual curiosity does in many parts of our planet.

The wielders of power fear high-quality science not only because their pseudoscientific gospel is usually on a collision course with scientific facts but also because they too are in the market for brilliant minds. And so potential Einsteins often continue to emulate Napoleon.

If Spaceship Earth is in a state of mutiny, it is not because of science but rather because of the absence of science from many societies. Can scientists help? Perhaps, but only by crossing barriers of local interest and acting in unison. And time is running out.

BORAN A. LEONTIĆ

University of Zagreb
Yugoslavia

Stolichnaya
The Vodka

For gift delivery anywhere* call 800-528-6148 (Arizona 602-957-4923).

*Except where prohibited/Major credit cards accepted/Stolichnaya Vodka 80 and 100 proof, distilled from grain. Imported by Monsieur Henri Wines, Ltd. White Plains, NY 10604

Moving mosaic: **Car-phones are suddenly making news everywhere. The telephone has been unchained from its wired connection by a breakthrough technology called cellular radio. Working like a giant mosaic of many**

transmission zones or cells—joined to form a network—it's made mobile phones as easy to own and use as the phones in your office or home. Through Contel Cellular Inc., Continental Telecom is a pioneer in this latest example of everyday communication on the move. From telephony to satellites. Architects of telecommunication.

FOR A FREE BROCHURE ON ALL OF CONTEL'S CAPABILITIES WRITE TO:
CONTEL, DEPT. 5012, 245 PERIMETER CENTER PARKWAY, ATLANTA GA 30346 © 1984 CONTINENTAL TELECOM INC.

CONTEL®

REVOLUTIONARIES.

They were revolutionaries from the beginning—created to be the best six passenger cars you could buy. And for 1985, they still are. Aries 2-door, 4-door, and wagon offer the lowest prices, highest standard mileage rating, longest warranty, and best overall value*. All the great qualities that have helped Aries develop a tradition of solid resale.

But while the practicality under the sheetmetal has remained intact, *in* the sheetmetal, a lot has changed. And the changes are simply beautiful. Aries now adds a sophisticated new look to a proven tradition of over 15 billion front-wheel-drive miles.

Aries LE gives you over 40 standard features. Some 73 improvements and innovations—both standard and optional. And, with offerings like the Premium Equipment Discount Package, savings of $225 when you order an AM/FM radio, rear window defroster, power steering, automatic speed control, tilt steering wheel, and other popular options.**

Ask any of over 450,000 American families who've owned Aries what they think. Then ask your Dodge dealer about buying or leasing* a new Aries. It's traditional K-car value in a beautiful new style that makes Aries 2-door, 4-door, and wagon true revolution-aries again for 1985.

Dodge
DIVISION OF CHRYSLER CORPORATION

AN AMERICAN REVOLUTION

*1985 6-passenger model comparisons: lowest base and comparably equipped sticker prices; highest standard mileage—26 EPA city est. mpg, use to compare, actual mpg will vary with options, driving habits and conditions, and vehicle's condition, CA est. lower; longest warranty—5 years/50,000 miles, whichever comes first. Limited warranty on powertrain and outer body rust-through, excludes leases, deductible applies. Ask for details. **Aries LE savings based on sticker prices of package items if purchased separately. Ask for details.

BUCKLE UP FOR SAFETY.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

MARCH, 1935: "Larger than any like construction project ever undertaken by man, the San Francisco–Oakland Bay bridge is being speeded to its completion. When at last the 45,000,000 persons and 5,000,000 automobiles transported each year by ferries between San Francisco and Alameda counties begin to move across the structure, they will enjoy the distinction of riding over the world's longest and also highest above-water bridge, supported by piers sunk deeper than any heretofore poured under water. The bridge is being constructed by the California Toll Bridge Authority as a state enterprise at a cost of 77,200,000 dollars."

"Industry is making huge outlays to promote the art of packaging. Skillful packaging has been found to quicken the sale of products to a phenomenal degree; it opens the way to substantial economies in the shipment of goods, and it is instrumental in reducing losses incurred by deterioration and spoilage of consumer articles. Packaging once meant little more than wrapping products because they had to be wrapped to be moved, or at best to carry the trade-mark of brand quality; today that is a very small though fundamental part of it. The cardboard, paper and string stage is outgrown, and major credit for it must go to Cellophane, for it was this transparent material that made the consuming public 'package conscious' in a new sense."

"The New World, discovered and colonized by Europe 500 years before Columbus, was lost again because of rickets, modern archaeological excavations in Greenland suggest. At Herjolfsnes, on the lonely Greenland coast, several skeletons of Viking women, disinterred and studied by Prof. F. C. C. Hansen of Copenhagen, exhibit severe pelvic deformations. The abnormalities are due to osteomalacia (rickets in its severe form), according to Dr. J. Preston Maxwell, British physician and professor of gynecology at the Union Medical College at Peiping, China. Did the Nordics perhaps refuse to take the native cod-liver oil that has enabled their despised Eskimo neighbors to survive to the present day?"

"The cap of the Washington monument was examined with great interest by scientists when repairs to the famous landmark made it possible to inspect the aluminum pyramid. After 50 years' service as a lightning rod, 550 feet above the ground, the metal that cost 12 dollars a pound when the monument was built showed no signs of deterioration. In fact, the engraved inscriptions on the sides of the pyramid were perfectly legible. Today aluminum sells for 21 cents a pound, but when this cap was made, the metal was such a curiosity that the cap was exhibited in the window of Tiffany's and people shook their heads dubiously at the new-fangled metal and doubted whether it would stand up under the severe conditions of weathering to which it would be exposed."

"Sixty percent of Eastern Air Line's flying is now done at night. Two daily round-trip, eight-hour passenger, mail and express schedules are flown between New York and Miami, together with a flight of eight hours and 50 minutes between Chicago and Miami."

MARCH, 1885: "The Aeronautical Society of Great Britain proposes to take a practical step toward the attainment of the end for which it was established, by holding an Aeronautical Exhibition. The objects for exhibition will be: (1) Models of designs for the accomplishment of aerial navigation by mechanical means only. (2) Models of designs for the accomplishment of aerial navigation partly by buoyancy and partly by mechanical means. (3) Models constructed to elucidate either of the first two objects, which are capable of flight and carrying their own motive power. (4) Machines constructed upon a scale calculated to carry a weight equal to that of a man. (5) Light motors. (6) Balloons, navigable or otherwise. (7) Balloon material and appliances for propulsion or otherwise. (8) Kites or other aerial appliances of that character. (9) Objects of interest connected with aeronautics."

"Washington was laid out mathematically. It was mapped on a grand plan; and strange to say, it has finally been realized. Its success results wholly from external causes. Left to itself, Washington would have sunk long ago in its primeval mud, and future generations would hardly have known that such a city had once stood there. It has no elements of independent commercial prosperity. A single act of Congress, trundling the capital to some other part of the Union, could sweep Washington into nonentity. It is a political and social center. It is the home of the government. This, and this only, has made it what it now is—one of the finest cities in the Union."

"BY THOMAS A. EDISON: The efforts made with a view to long-distance telephoning have already proved quite satisfactory in a commercial way, and promise excellent results. Conversation has been conducted between Cleveland and New York, and is now daily carried on between New York and Boston to a limited extent. The great difficulty in long-distance telephoning is the loss of the current by static induction on the earth and wires in close proximity. If a single wire could be placed sufficiently high as to amply clear all the mountain tops, one could whisper around the world with perfect ease. One thing, however, is now certain: that the time is close at hand when the telephone will be perfectly successful in an unbroken circuit for a distance of at least 300 miles. It is probable that by means of repeating stations communication can be had over all parts of the United States."

"BY THOMAS A. EDISON: Two years' experience proves beyond a doubt that the electric light for household purposes can be produced and sold in competition with gas. It is immaterial whether the electrical energy is used for light or for other purposes. It is so easy of control, the apparatus required so inexpensive, that it can be used as a motor power for purposes innumerable. The function of electricity as a motor for household purposes will be hardly less useful than its value in illumination. Electricity as a motive power will not be confined to household or factory purposes. Experiments have perfectly proved the practicability of the electric locomotive, and indicate that it will be largely adopted in the future in place of the steam locomotive. Various experiments have been made with a view to the electric propulsion of carriages, cabs, drays, etc. The drawback has been that the power has been obtained from secondary or storage batteries, the depreciation in which is so rapid, and the weight of the receptacle so great, that until some radical improvements are made in connection with the storage of electricity, or the production of the same directly from coal, we cannot hope to see the subtle fluid used as a means of propelling street conveyances."

THE AUTHORS

DAVID HAFEMEISTER, JOSEPH J. ROMM and KOSTA TSIPIS ("The Verification of Compliance with Arms-Control Agreements") are respectively professor of physics at the California Polytechnic State University, a Ph.D. candidate in physics at the Massachusetts Institute of Technology and associate director of M.I.T.'s Program in Science and Technology for International Security. Hafemeister holds a B.S. from Northwestern University in mechanical engineering and a Ph.D. from the University of Illinois at Urbana-Champaign in physics. He was at the Los Alamos Scientific Laboratory from 1964 to 1966, when he moved to Carnegie-Mellon University. In 1969 he joined the faculty of the California Polytechnic State University, and he served as a special assistant in the U.S. Senate from 1975 to 1977 and in the Department of State from 1977 to 1979. Hafemeister was a visiting scientist at M.I.T. in 1983 and 1984, when most of the work discussed in the article was done. Romm, who was graduated from M.I.T. with a B.S. in physics in 1982, works with Tsipis in the Program in Science and Technology and is a visiting researcher at the Scripps Institution of Oceanography, doing his thesis work on acoustical ocean tomography. Tsipis, a native of Greece, came to the U.S. in 1954 to study electrical engineering and physics. His B.S. and M.S. in physics are from Rutgers University and his Ph.D., which he got from Columbia University, is also in physics. Tsipis joined the M.I.T. physics department in 1966.

RONALD G. PRINN ("The Volcanoes and Clouds of Venus") is a professor in the department of earth, atmospheric and planetary sciences at the Massachusetts Institute of Technology. He holds a B.Sc. in chemistry and pure and applied mathematics and an M.Sc. in chemistry from the University of Auckland in New Zealand and an Sc.D. in chemistry from M.I.T. Since 1971 he has held various positions on the M.I.T. faculty; in 1981 he was visiting associate professor in the division of geological and planetary sciences of the California Institute of Technology. He has been a fellow of the American Geophysical Union since 1981.

CARLO M. CROCE and GEORGE KLEIN ("Chromosome Translocations and Human Cancer") are respectively institute professor at the Wistar Institute of Anatomy and Biology in Philadelphia and professor of tumor biology at the Karolinska Institute in Stockholm. Croce, who was born in Milan, got his M.D. from the University of Rome in 1969 while working with Giuseppe Andres in the Istituto di Clinica Medica there. Croce is also Wistar Professor of Human Genetics at the University of Pennsylvania School of Medicine and associate director of the Wistar Institute. Klein was born in Budapest and received his M.D. degree in 1951 from the Karolinska Institute, where he is now head of the department of tumor biology. A member of the Royal Swedish Academy of Sciences, he is a foreign member of numerous other honorary societies and is editor of *Advances in Cancer Research* and *Advances in Viral Oncology*.

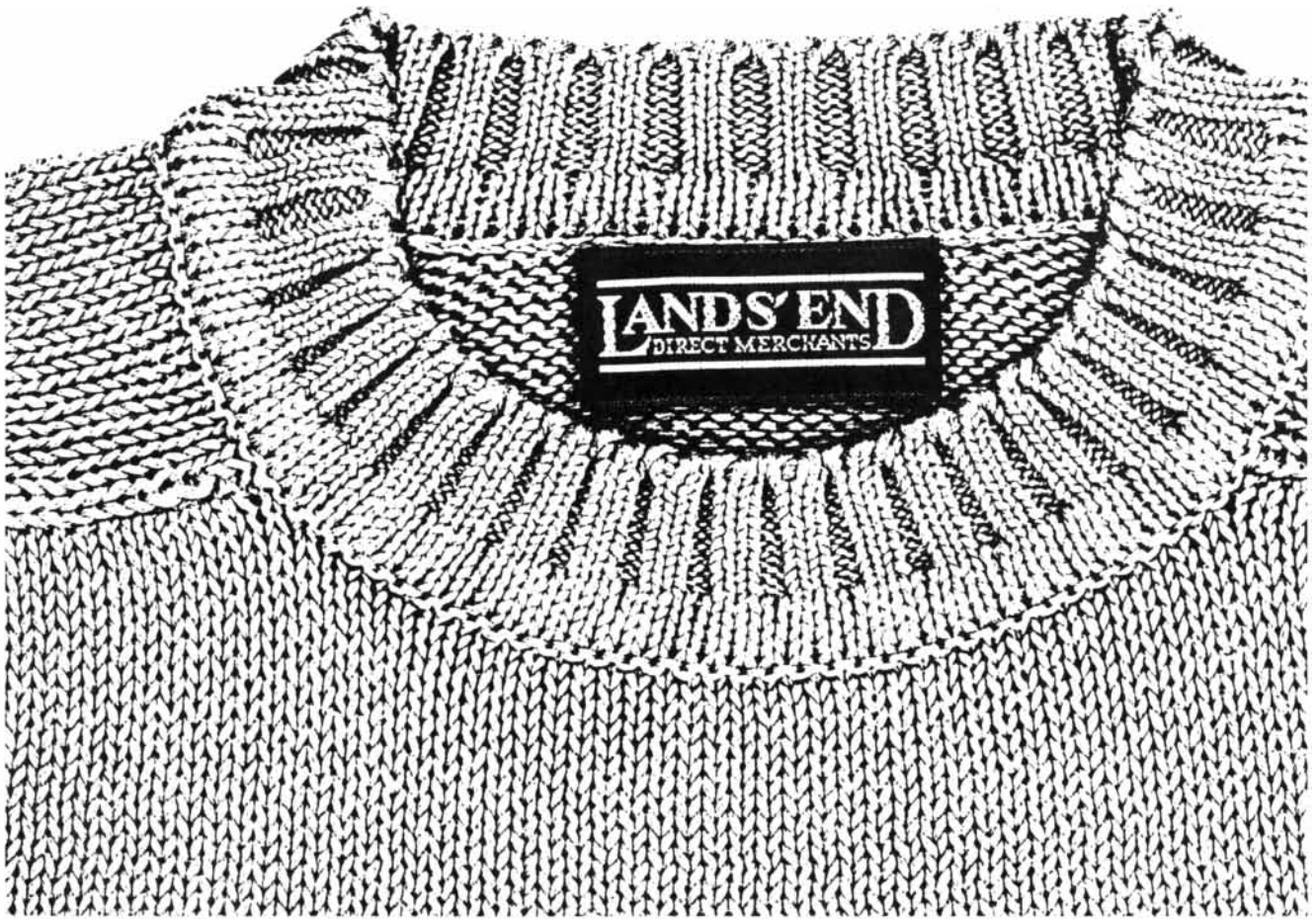
DANIEL Z. FREEDMAN and PETER VAN NIEUWENHUIZEN ("The Hidden Dimensions of Spacetime") are respectively professor of applied mathematics at the Massachusetts Institute of Technology and professor in the Institute for Theoretical Physics at the State University of New York at Stony Brook. Freedman did his undergraduate work at Wesleyan University and earned a doctorate in physics in 1964 from the University of Wisconsin at Madison. In 1968 he joined the faculty at Stony Brook, and he moved to M.I.T. in 1980. Van Nieuwenhuizen got doctorates in mathematics and physics at the University of Utrecht in 1969 and went to Stony Brook in 1975. At Stony Brook, Freedman and van Nieuwenhuizen collaborated (with Sergio Ferrara) on research that led to the discovery of supergravity in 1976. Since then they have been actively engaged in research on supergravity and related subjects.

HAL WHITEHEAD ("Why Whales Leap") is assistant adjunct professor at the Newfoundland Institute for Cold Ocean Science, where he has been since 1981. Raised in England as a member of a sailing family, he first saw whales in 1974, when, after receiving his B.A. in pure mathematics at the University of Cambridge, he sailed a small boat from Connecticut along the coast of Nova Scotia to Newfoundland. After some time spent as a research assistant to Roger Payne, who was then at the New York Zoological Society, he returned to Cambridge. There he earned a Diploma in Mathematical Statistics in 1977 for a dissertation titled "A Statistical Analysis of the Development of Play Behaviour in Kittens." In 1981 he got a Ph.D. in zoology, also from Cambridge, for work on the behavior and ecology of humpback whales.

R. J. MERCER ("A Neolithic Fortress and Funeral Center") is reader in archaeology and acting head of the department of archaeology at the University of Edinburgh. His M.A. in archaeology was granted by the University of Edinburgh in 1967. He joined the faculty in 1974 after five years as an inspector of ancient monuments in the British Department of the Environment, where his chief duties were preserving and recording ancient monuments in southern and southwestern England. During the past 10 years Mercer has developed a major project concerned with recording prehistoric, medieval and industrial remains in Scotland and southwestern England.

S. THOMAS PICRAUX and PAUL S. PEERCY ("Ion Implantation of Surfaces") are both at the Sandia National Laboratories. Picraux is supervisor of the ion-solid interactions research division and Peercy is manager of the ion-implantation and radiation-physics research department. Picraux received a B.S. in electrical engineering at the University of Missouri in 1965 and a Ph.D. in engineering science and physics from the California Institute of Technology, having spent a year studying physics as a Fulbright fellow at the University of Cambridge. He did research at the Chalk River Nuclear Laboratories in 1967 and 1968 and joined the staff of Sandia in 1969. Picraux is a fellow of the American Physical Society. Peercy got his doctorate from the University of Wisconsin at Madison in 1966 and then joined the staff of Bell Laboratories in Murray Hill, N.J. He went to Sandia in 1968. Peercy is program chairman of the Materials Research Society.

ERIC BLOCK ("The Chemistry of Garlic and Onions") is professor of chemistry at the State University of New York at Albany. He was graduated from Queens College of the City University of New York with a B.S. in 1962 and received his Ph.D. in chemistry from Harvard University in 1967. He was at the University of Missouri from 1967 to 1981 and then moved to Albany. Currently a fellow of the John Simon Guggenheim Foundation, he is a consultant to the Société Nationale Elf-Aquitaine, the French oil company. Block's major area of research interest is the organic chemistry of sulfur.



Cotton. The most versatile and comfortable fibre in the world. Naturally.

(And no one offers you more all-cotton items than Lands' End)

Can you imagine life on this planet without cotton? It feels so good and fresh and airy. And without the slightest problem it can look lustrous and silky just as easily as it looks fresh and crisp.

Here at Lands' End we hold cotton in the highest esteem. We've learned to appreciate how versatile it is. How well it takes and holds colors true to the spectrum. And how beautifully it responds to fine design.

Giving Nature a hand with raw cotton.

You don't set out to improve on Nature lightly. So, at Lands' End, when we plan a new cotton garment, we start with the best cotton there is—100% cotton, combed and, where appropriate, mercerized. Mercerizing involves treating the yarn or the fabric to give it a softer hand—a softer feel to the touch, that is. This treatment lets the fabric retain richer dye colors and put up fierce resistance to shrinkage.

This Spring, once again, cotton is king.

You'll note in our catalogs that here and there we feature items made of other fabrics. But far and away, it is cotton which reigns as king. Whether it's our cotton twills, cotton shirtings, or cotton in a variety of sophisticated knits, Lands' End offers cotton in so many products reciting them almost adds up to a chant.

You'll find cotton jeans, knit shirts, and a range of cotton ties. And cotton shirts, cotton sweaters, cotton trousers, cotton skirts. And don't miss our sleek trim cotton turtlenecks, and our rugged cotton shorts, stretch or otherwise. Cotton sweats are ideal for you who're into Spring training. And cotton socks and dresses need no explaining. But how often do you think of cotton in luggage? Yet it's a mainstay fibre throughout our many Lands' Ends-made soft luggage lines.

Yes, we have cotton by the catalog-full.

We say it advisedly because it's true. *No one offers more all-cotton items than Lands' End.* At prices throughout which constitute value. To check us out order our latest catalog today. By mail, via the coupon below. By phone, on our toll-free 800 number(800-356-4444), attended 24 hours a day.

And don't forget, 100% cotton stays new looking whether you hand wash it, have it dry cleaned, or beat it against a rock. We guarantee it the way we guarantee every other item we offer. Unconditionally.
As in GUARANTEED. PERIOD.

Please send free catalog.
Lands' End Dept. Q-27
Dodgeville, WI 53595

Name _____

Address _____

City _____

State _____ Zip _____

Or call Toll-free:
800-356-4444

COMPUTER RECREATIONS

A Core War bestiary of viruses, worms and other threats to computer memories

by A. K. Dewdney

When the column about Core War appeared last May, it had not occurred to me how serious a topic I was raising. My descriptions of machine-language programs, moving about in memory and trying to destroy each other, struck a resonant chord. According to many readers, whose stories I shall tell, there are abundant examples of worms, viruses and other software creatures living in every conceivable computing environment. Some of the possibilities are so horrifying that I hesitate to set them down at all.

The French spy thriller *Softwar: La Guerre Douce* (English translation to be published by Holt, Rinehart & Winston) provides a geopolitical fantasy of this type. Authors Thierry Breton and Denis Beneich spin a chilling yarn about the purchase by the Soviet Union of an American supercomputer. Instead of blocking the sale, American authorities, displaying studied reluctance, agree to the transaction. The computer has been secretly programmed with a "software bomb." Ostensibly bought to help with weather forecasting over the vast territory of the Soviet Union, the machine, or rather its software, contains a hidden trigger; as soon as the U.S. National Weather Service reports a certain temperature at St. Thomas in the Virgin Islands, the program proceeds to subvert and destroy every piece of software it can find in the Soviet network. To the extent that such scenarios represent real possibilities, I am tempted to say, "If we must have war, by all means let it be soft." On the other hand, the possibility of an accident due to the intimate connection between military software and weapons-control systems gives me pause.

INSTRUCTION	MNEMONIC	CODE	ARGUMENTS		EXPLANATION
Move	MOV	1	A	B	Move contents of address *A* to address *B*.
Add	ADD	2	A	B	Add contents of address *A* to address *B*.
Subtract	SUB	3	A	B	Subtract contents of address *A* from address *B*.
Jump	JMP	4	A		Transfer control to address *A*.
Jump if zero	JMZ	5	A	B	Transfer control to address *A* if contents of address *B* are zero.
Jump if greater	JMG	6	A	B	Transfer control to address *A* if contents of *B* are greater than zero.
Decrement: jump if zero	DJZ	7	A	B	Subtract 1 from contents of address *B* and transfer control to address *A* if contents of address *B* are then zero.
Compare	CMP	8	A	B	Compare contents of addresses *A* and *B*; if they are unequal, skip the next instruction.
Split	SPL	9	A		Split execution into next instruction and the instruction at *A*.
Data statement	DAT	0		B	A nonexecutable statement; *B* is the data value

The list of instructions for Core War programs

Before I describe the experiences of various readers with hostile programs it would be worthwhile to summarize Core War for those who missed the May 1984 column:

Two players write one program each in a low-level lanquage called REDCODE. The programs are placed in a vast, circular arena called Core. In reality Core is simply an array of several thousand locations whose last address is contiguous to the first. Each battle-program instruction occupies one location in Core. A Memory Array Redcode Simulator executive program (MARS for short) runs the battle programs by alternately executing one instruction of each, in the manner of a simple time-sharing system: the two programs attack each other and seek in turn to avoid damage or to repair it. A simple mode of attack can be executed by means of MOV instructions. For example,

MOV #0 1000

causes the number 0 to be placed in the location whose address lies 1,000 locations beyond this instruction. The previous contents of that location are thereby erased. If the 0 were placed on top of an enemy instruction, it too would be wiped out and the program would no longer be executable. The enemy would lose the game.

Since no computer, whether personal or mainframe, comes equipped with REDCODE and a convenient battle array, such features must be simulated. Guidelines for writing a simulation program were and still are available from the offices of *Scientific American* at a cost of $2 to cover postage and handling. Please address your request to Core War, *Scientific American,* 415 Madison Avenue, New York, N.Y. 10017. Last year several hundred readers obtained the guidelines and a large percentage of them wrote Core War game programs.

Inspired by a June 1959 *Scientific American* article on self-reproducing mechanisms by L. S. Penrose, Frederick G. Stahl of Chesterfield, Mo., created a miniature linear universe in which humble creatures lived, moved and (after a fashion) lived out their destinies. Stahl writes:

"Like Core War, I set aside a closed, linear segment of main memory in which a creature was simulated by modified machine language. The machine was an IBM Type 650 with drum memory. The creature was programmed to crawl through its universe eating food (nonzero words) and creating a duplicate of itself when enough

The Spirit of America

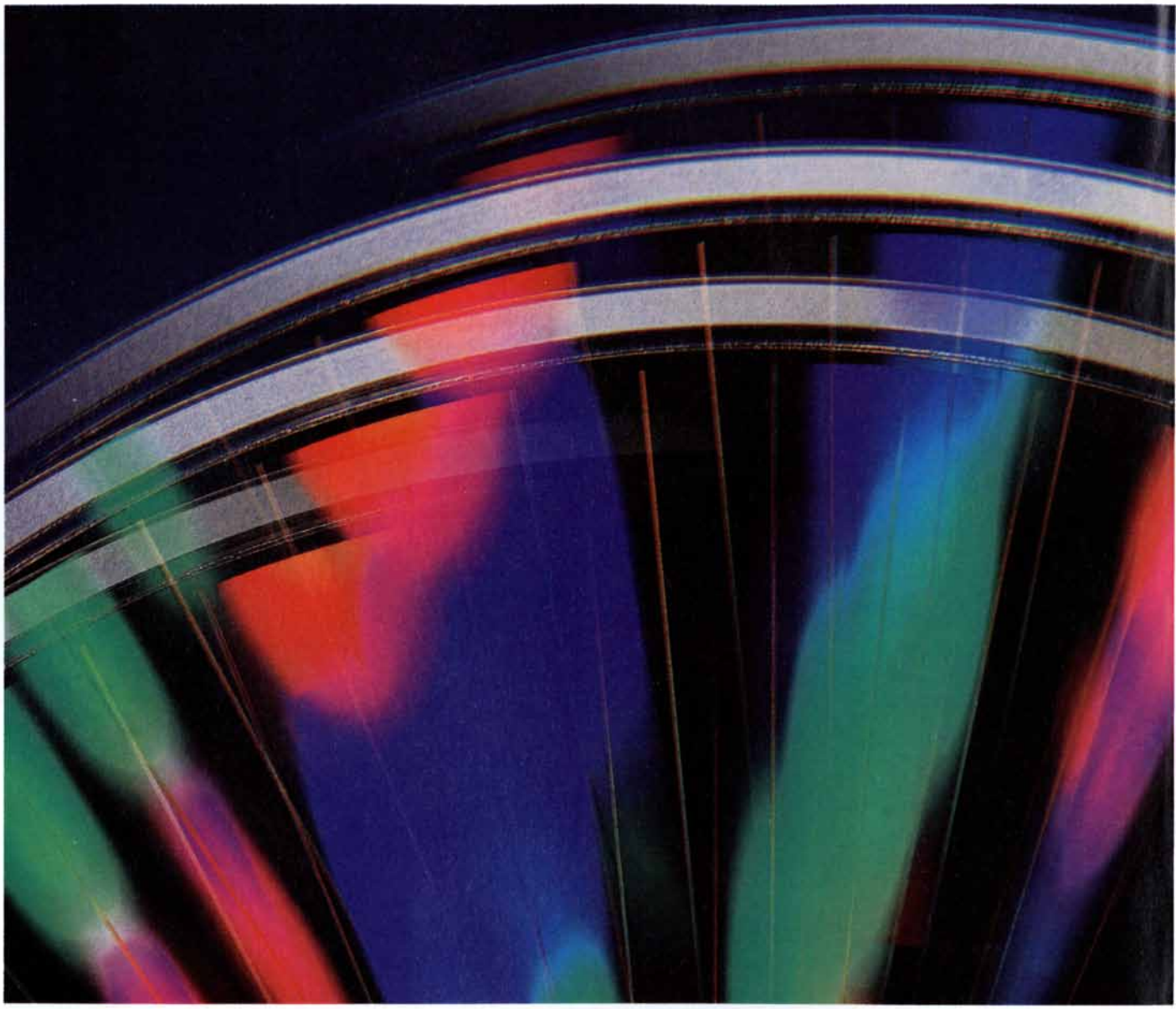
Mabry Mill, Virginia by Clyde H. Smith

Where the woodland farmer flourished, the miller was not far behind. Independent and enterprising, he signalled the coming of trade and prosperity. And looking to the future, he relaxed at day's end with America's native whiskey: Kentucky Bourbon. Old Grand-Dad still makes that Bourbon much as we did 100 years ago. It's the spirit of America.

For a 19" x 26" print of Mabry Mill, send a check or money order for $4.95 to Spirit of America offer, P.O. Box 183V, Carle Place, N.Y. 11514.

Old Grand-Dad

Kentucky Straight Bourbon Whiskey. 86 Proof. Old Grand-Dad Distillery Co., Frankfort, KY 40601 © 1984 National Distillers



A single 12" optical disk can store all the text appearing on 40,000 letter-size, typewritten pages—amazing memory capacity resulting from Hitachi technology.

MEMORY

Since the first electronic brain began "thinking" almost four decades ago, Hitachi has been steadily advancing the art of machine memories. From paper tape and punch cards. To magnetic tape and memory drums. To the very latest optical technologies for high-density information storage.

Technology that never forgets

Today, the results of Hitachi research are in use all around you. Floppy disks that can hold hundreds of pages' worth of text for ready reference via business or personal computers. Magnetic "bubble memories" which allow industrial robots to memorize dozens of job routines. Semiconductor memory devices for data processing, satellite communications systems, office automation equipment, and even home audio/video components.

Our engineers are using microelectronic skills to allow much more information to be packed into smaller and smaller units. They have found ways to etch micron-wide circuits on 1/2 cm^2 silicon chips to accommodate over one million bits of data. They have refined the magnetic storage properties of Hitachi hard disks to almost flawless levels.

In fact, we are constantly coming up with innovations and new applications. Quite recently: A mirror-like optical disk no larger than an LP record, which permits laser inscription of pictures and sounds as well as computer-encoded text and numbers for laser scanning of any bit of information in just a quarter of a second.

These examples demonstrate a few of the ways in which Hitachi is improving upon basic technology. Then using it to create practical tools that meet your needs...and those of professionals in banking, education, research, and virtually every other field you can name.

The best of worlds is yet to come

Our vision of the future includes whole libraries of information carried in memory packets even smaller than a standard textbook. Visual memory banks that allow designers to create new works of art from stored images. Automotive microcomputers that recall routes, driving conditions and destination data. And much, much more.

We'd like you to share in the benefits of our scientific research, covering the next generation of microchips, sensors and other electronic devices. For improved business efficiency. For a higher quality of life. Two goals we've pursued for 75 years as part of our commitment to a better world through electronics.

WE BELIEVE MEMORY DEVICES EXPAND THE POTENTIALS OF THE HUMAN MIND

HITACHI

BMW TRANSFORMS THE AUTOMOBILE INTO A HIGHER FORM OF INTELLIGENCE.

It seems that the same automakers who recently were extolling the comforts of their upholstery have now discovered another part of your anatomy.

Today, the emphasis is on heady gadgetry, such as those dashboards whose displays exhibit more in the way of entertainment value than informational value.

And no automobile stands more staunchly opposed to this trend than the BMW 528e.

BRAINPOWER VS. HORSEPOWER.

The BMW 528e is a luxury sedan as dedicated to furthering the science of internal combustion as it is to advancing the art of driving.

Thus, it is equipped with a microprocessor-based engine management system that controls fuel injection and ignition based on intelligence gathered from electronic sensors located throughout the engine.

The 528e's 'Eta' engine is another tribute to inspired engineering. A low-revving, high-torque design that aims its impressive output at low to medium speeds. So it doesn't simply perform exuberantly on the highway, but also "gives its best in the 'daily use' RPM range..." (Motor Trend). And achieves a remarkable level of efficiency—20 mpg, 25 highway*—in the process.

HIGHER EDUCATION FOR DRIVERS.

Still, all the 528e's electronic intelligence is not devoted to engine management. An equal amount is directed toward informing the car's most important moving part. Namely, the driver.

For example, there's the 528e's new Onboard Computer, which provides vital trip data. And the Service Interval Indicator, which recommends oil changes and inspections based on actual driving habits, not actuarial mileage tables.

Intelligence is also abundantly manifest in the ergonomically sound layout of the 528e's instruments and controls. And in its steering system, which automatically adjusts its level of power assist to provide ease of handling without ever anesthetizing your sense of the road.

In fact, even the 528e's body itself is programmed. With 'crush zones' at the front and rear to absorb and dissipate impact in the event of an accident.

Of course, a car such as this is not for everyone. But rather, the BMW 528e is for those individuals able to distinguish between a car that embraces superficiality, and one that is, instead, the embodiment of intelligence.

A higher form of intelligence which you are invited to put to the test at your nearest BMW dealer.

THE ULTIMATE DRIVING MACHINE.

*EPA-estimated fuel efficiency figures for comparison only. Your actual mileage may vary depending on speed, weather and trip length. Actual highway mileage will most likely be lower. The BMW trademark and logo are registered. European Delivery can be arranged through your authorized U.S. BMW dealer.

food was accumulated. Like Core War, I had an executive program which kept track of who was alive and allocated execution time among the living creatures. I called it the 'Left Hand of God.'" Stahl goes on to discuss his program's ability to reproduce. He also describes an interesting mutation mechanism: a program being copied might experience a small number of random changes in its code. However, Stahl reports, "I abandoned this line of work after one production run in which a sterile mutant ate and killed the only fertile creature in the universe. It was apparent that extraordinarily large memories and long computer runs would be needed to achieve any interesting results."

A similar story concerns a game called Animal in which a program tries to determine what animal a human is thinking of by playing a form of Twenty Questions. David D. Clark of the Massachusetts Institute of Technology Laboratory for Computer Science writes that the employees of a certain company devotedly played Animal. While it resembles neither a battle program nor even Stahl's simple creatures, Animal achieved the ability to reproduce itself in the corridors of core through the efforts of a programmer to enhance a key feature of the game: when the program guesses incorrectly what animal the human has in mind, it asks the human to suggest a question it might ask to improve its future performance. This feature, Clark continues, led the programmer to invent a certain trick for making sure that everyone always had the same version of Animal.

"On a very early computer system, which lacked any shared directory structure, but also lacked any protection tools, a programmer invented a very novel way of making the game available to several users. A version of the game existed in one user's directory. Whenever he played the game, the program made a copy of itself into another directory. If that directory had previously contained a copy of the game, then the old version was overwritten, which made the behavior of the game change unexpectedly to the player. If that directory had previously had no version of Animal, the game had been offered to yet another user."

Clark recalls that Animal was such a popular game that eventually every directory in the company system contained a copy. "Furthermore, as employees of the company were transferred to other divisions... they took Animal as well, and thus it spread from machine to machine within the company." The situation would never have become serious had it not been for the fact that all those copies of this otherwise innocuous game began to clog the disk memory. Only when someone devised a more "virulent" version of the game was the situation brought under control. When the new version of Animal was played, it copied itself into other directories not once but twice. Given enough time, it was thought, this program would eventually overwrite all the old versions of Animal. After a year had passed, a certain date triggered each copy of the new Animal program. "Instead of replicating itself twice whenever it was invoked, it now played one final game, wished the user 'goodbye' and then deleted itself. And thus Animal was purged from the system."

```
1   IF PEEK (104) = 134 GOTO 10
2   POKE 104, 134: POKE 134 * 256,0
3   PRINT CHR$(4) "RUN APPLE WORM"
10  HOME : POKE - 16302,0: POKE - 16304,0: POKE 1023,160
20  FOR I = 0 TO 94: READ D: POKE 1024 + I, D: NEXT I
30  POKE - 16368,0
40  IF PEEK ( - 16384)<128 GOTO 40
50  CALL 1024
100 DATA 160,225,200,185,255,3,153,127,4,192,95,208,245,
    160,18,190,76,4,24,189,128,4,105,128,157,128,4,189,129,
    4,105,0,157,129,4,192,13,208,18,238,23,4,173,23,4
200 DATA 141,151,4,206,31,4,173,31,4,141,159,4,136,208,211,
    173,167,4,72,173,176,4,141,167,4,104,141,176,4,76,128,
    4,7,20,25,28,33,46,55,61,65,68,72,75,4,16,40,43,49,52
```

A worm that inhabits Apples

Ruth Lewart of Holmdel, N.J., once created a monster (of sorts) without even writing a program. Working on her company's time-sharing system, she was preparing a demonstration version of a teaching program when she decided to make a backup copy on another time-sharing system. When the original system began to seem sluggish, she "switched to the backup system, which was very responsive—for all of three minutes, by which time there was no response and utter chaos on the screen of my graphics terminal. It was not possible for any user to log on or to log off from the system. The conclusion was inescapable—my program was somehow at fault! Despite my panic, I suddenly realized that I had specified an ampersand as the terminal's field separator character. But the ampersand was also the character used by the computer system to spawn a background process! The first time the computer read from the screen, it must have intercepted the ampersands meant for the terminal, and spawned a number of processes, which in turn each spawned more processes, ad infinitum." A frantic long-distance call informed a system administrator of the source of the disease and the mainframe computer was then shut down and restarted. Needless to say, Lewart changed the ampersand to a less dangerous character. Her program "has been humming happily ever since."

Even though Core War programs are not spawned in this manner, additional copies can enhance their survival. Several readers suggested three copies of a program be made so that the copy currently executing could use the other two copies to determine whether any of its instructions were wrong. The executing program could even replace a faulty instruction with a viable one. A similar idea lies behind Scavenger, a program designed to protect mass-storage files from error when backup copies are made on magnetic tape. Arthur Hudson, who lives in Newton, Mass. (and works for yet another unnamed company), writes: "Anyone who used much magnetic tape found himself beset by an alien force called the Law of Joint Probability." Hudson goes on to cite various errors connected with the handling of tapes and shows that, although each kind of error has a relatively small chance of happening, the probability of at least one of them occurring is uncomfortably large. He continues:

"Fear not, Scavenger is with you: If you place a mass-storage file in its care, it will copy the file on three magnetic tapes without bothering you with housekeeping details. Even if the computer crashes logically (as it did several times per day), the run backlog usually will not be destroyed; when the computer comes back up, whatever Scavenger worms are in the backlog will run in their turn. Each tape is writ-

ten by a separate run scheduled from a master runstream."

Owners of Apple computers should beware a mean little program called Apple Worm, created by Jim Hauser and William R. Buckley of California Polytechnic State University at San Luis Obispo. Written for the Apple II in 6502 assembly language, this species of worm replicates itself on a merry little journey through the host Apple. Initially one loads a special BASIC program [*see illustration on preceding page*] that in turn loads the worm into low memory (the part with low addresses). The BASIC program, on the other hand, occupies high memory.

"Because the Worm is loaded into one of the graphics areas of the machine, you can watch the Worm as it begins its headlong (actually, tail-long) dash into high memory.... After the Worm leaves the graphics window... you can wait until the Worm erases all of high memory (including the BASIC program) and crashes into the system ROMS."

Hauser and Buckley plan to publish a collection of worms in the not too distant future. They have designed a Worm Operating System and have even written a video game with Worm as one of its main characters.

Another software threat has been propounded by Roberto Cerruti and Marco Morocutti of Brescia, Italy. Inspired by the translation of the column on Core War in the Italian edition of *Scientific American, Le Scienze,* the two sought a way of infecting the Apple II computer, not with a worm but with a virus. Reports Cerruti:

"Marco thought to write a program capable of passing from one computer to the other, like a virus, and 'infecting' in this way other Apples. But we were not able to conceive it until I realized that the program had to 'infect' floppy disks, and use the computers only like a media from a disk to the other. So our virus began to take shape.

"As you know every Apple diskette contains a copy of the Disk Operating System, which is bootstrapped by the computer at power on. The virus was an alteration in this DOS, which at every write operation checked his presence on the disk and, if not, would modify in the same way the DOS on the disk, thus copying itself on every diskette put in the drive after the first power up. We thought that installing such a DOS on some disks used in the biggest computer shop in our city, Brescia, would cause an epidemic to spread in the city.

"But was it a real epidemic, of such unharming viruses? No, our virus should be malignant! So we decided that after 16 self-reproduction cycles, counted on the disk itself, the program should decide to re-initialize the disk immediately after bootstrap. Now the awful evil of our idea was clear, and we decided neither to carry it out, nor to speak to anybody about our idea."

That was kind of Cerruti and Morocutti. In a personal computer the disk operating system is the ultimate arbiter of the fate of programs, data and all else. In the scheme just described the infected disk operating system erases the disk whence it came and can therefore never be loaded again except from a new disk. The diseased DOS could even cause an irritating message to be displayed periodically:

IS YOUR DISK
SLIPPING?
It's time you got
DOS DOCTOR
available on disk at a
computer store near you

The viral infection just described has already happened on a small scale. Richard Skrenta, Jr., a high school student in Pittsburgh, wrote such a program. Instead of wiping disks or displaying advertisements, this form of infection caused subtle errors to appear throughout the operating system.

"All of this seems pretty juvenile now," writes Skrenta, but "Oh woe to me! I have never been able to get rid of my electronic plague. It infested all of my disks, and all of my friends' disks. It even managed to get onto my math teacher's graphing disks." Skrenta devised a program to destroy the virus, but it was never as effective as the virus itself had been.

There is a good problem implicit here and I would be both unimaginative and irresponsible for not posing it: In one page or less describe DOS DOCTOR, a program on disk that somehow stamps out such electronic epidemics. Many disks used by a personal computer contain copies of its DOS. When started up, the computer obtains its copy of the DOS from the disk. This DOS will still be in charge when other disks, also containing copies of the DOS, are run. If it is infected, the DOS currently in charge may alter the other copies of the DOS or even replace them with copies of itself. But how to counteract such virulence?

In the initial version of Core War the main challenge was to enable battle program A to protect itself from stray hits generated by battle program B. If such protection could be more or less guaranteed, then evolution of the game was to proceed to the next level, where programs would have been forced to seek each other out and develop concentrated attacks.

In an effort to guarantee such protection, I suggested in the May column the instruction

PCT *A*

where *A* is the relative address (either direct or indirect) of an instruction that is to be protected. A single attempt to change the contents of that address would be prevented by MARS, the game's supervisory system. The next attempt, however, would succeed. It seemed to me that by employing a simple loop, any battle program could protect all its own instructions from stray bombs long enough to be able to launch an undistracted probe for the other program. The illustration on this page displays such a self-protecting program in schematic form. The protection loop consists of six instructions, four of them executed at each cycle through the loop. Thus a battle program of n instructions (including the loop) would require nearly $4 \times n$ executions to have complete protection from a single hit. This salutary shielding is hardly proof against a dwarf program that hurls two shots at each location.

There is another use of this instruction, unforeseen in the earlier Core War article. Stephen Peters of Timaru, New Zealand, and Mark A. Durham of Winston-Salem, N.C., independently thought of using PCT offensively. A program called TRAP-DWARF lays down a barrage of zeros in the usual

```
DAT                0          /pointer to address being protected
ADD    #1          —1         /increment protection address
PCT    @—2                    /protect the address
CMP    #102        —3         /if all 102 instructions protected...
JMP    2                      /...then leave loop
JMP    —4                     /otherwise re-enter it.
                .
                .
                .
BODY OF MAIN BATTLE PROGRAM
                .
                .
                .
```

This loop protects combatants from stray bombs

way but then protects each deposit against overwriting. This means that an unwary enemy program may fall into one of these traps in the course of writing itself into a new area. The instruction meant for the location occupied by a protected zero would of course have no effect on that location. Later, when the new program's execution reaches that address, it dies because 0 is not an executable instruction. PCT may be worthy of inclusion in some future version of Core War but I shall shelve it for now in the interest of simplicity, the game designer's touchstone.

Other reader ideas varied from the two-dimensional Core array suggested by Robert Norton of Madison, Wis., to the range-limitation rule suggested by William J. Mitchell of the mathematics department at Pennsylvania State University. Norton's idea is largely self-explanatory but Mitchell's suggestion requires some elaboration. Allow each battle program to alter the contents of any location up to a distance of some fixed number of addresses. Such a rule automatically keeps DWARF from doing any damage outside this neighborhood. The rule has many other effects as well, including a strong emphasis on movement. How else can a battle program get within range of an opponent? The rule has much merit and I hope that some of the many readers with a Core War system of their own will give it the further exploration it deserves.

Norton also suggests that each side in a Core War battle be allowed more than one execution. The same idea occurred to many other readers. Indeed, I have decided to adopt the suggestion. Core War now assumes a previously lacking wide-open character.

The change is made by adding the following instruction, called a split, to the official Core War list [*see illustration on page 14*].

SPL *A*

When execution reaches this point, it splits into two parts, namely the instruction following SPL and the one *A* addresses away. Because this immediately allows each Core War player to have several programs running at once, it is necessary to define how MARS will allocate such executions. Two possibilities exist.

To illustrate them suppose one player has programs A_1, A_2 and A_3, whereas the other player has programs B_1 and B_2. One alternative is to have all the first player's programs run, followed by those of the second player. The order of execution would thus be A_1, A_2, A_3 and then B_1 and B_2. The cycle would

repeat indefinitely. The second alternative is to have the programs of the two players alternate. In this case the sequence would be A_1, B_1, A_2, B_2, A_3, B_1 and so on. The two schemes are quite different in effect. The first scheme puts great emphasis on unlimited proliferation and seems thereby to limit the role of intelligence in the game. The second scheme, however, implies that the more programs either player has running, the less often each will be executed. A law of diminishing returns seems appropriate in this context and I have therefore adopted the second scheme. The purpose of the game, in any event, is to bring all enemy programs to a halt.

The new instruction is rife with creative possibilities. As an illustration of the humblest issue possible, there is a battle program called IMP GUN:

```
SPL 2
JMP −1
MOV 0 1
```

Consider what happens when execution first arrives at the top of this program. The instruction SPL 2 means there will be two executions allotted to this program later: both JMP −1 and MOV 0 1 will be performed. The first instruction causes the program to recycle and the second sets an IMP in motion. The IMP will move down, of course, since the target of the MOV command will always be the next address, as indicated by the (positive) 1. The IMP is thus spawned on each program cycle and an endless stream of IMP's run pattering through Core bent on the destruction or subversion of hostile programs. At first glance it may seem that no defense is possible against such an army of IMP's, but in fact one is. Enter IMP PIT, an even simpler program set in motion by an SPL command in some larger assembly of instructions wishing to protect its upper flank:

```
MOV #0  −1
JMP −1
```

At each execution, IMP PIT places a zero just above itself in the hope of zapping an oncoming IMP. Here the execution-allocation rule is critical. If IMP GUN belongs to A and IMP PIT belongs to B, then A needs *n* turns to execute *n* IMP's; only one IMP can arrive at the location just above the IMP PIT. Other things being equal, B has to execute IMP PIT only once to terminate an arriving IMP.

In the expanded Core War game, one imagines each side generating and deploying small armies of programs individually shaped to detect, attack, protect and even repair. Many subtleties such as the one suggested by John McLean of Washington, D.C., await further investigation. McLean imagines a specialized trap program that places JMP commands at various addresses throughout the Core array in the hope of landing a JMP command inside an enemy program. Each JMP so placed would transfer execution of the enemy program to the trap program, causing it to go over to the enemy, so to speak.

One major problem in need of resolution emerges from the melee of battle programs. What is to prevent a battle program for one side from attacking its colleagues? A recognition system appears to be necessary.

Among the many readers who have constructed Core War systems three deserve special mention: Chan Godfrey of Wilton, Conn., Graeme R. McRae of Monmouth Junction, N.J., and Mike Rosing of Littleton, Colo., have taken special care in defining and documenting their projects. I should particularly like to make Rosing's documents available to readers, but there is another and better idea that includes this possibility and solves other communication problems as well. If any reader with a Core War system already running will volunteer to act as the director of a Core War network, then documentation of various systems, rule suggestions, interesting programs and battles can be communicated to all participating Core War users. One volunteer will be chosen as director; the remaining volunteers might help with potential functions such as a newsletter, rules committee and so on, according to interest. In a future article I shall give the name and address of the network director.

Reports from readers who played with the ecology of the planet Wa-Tor ["Computer Recreations," December, 1984] are still pouring in, and so it will be impossible to discuss more than a few of the many interesting experiences described. Generally speaking the selection of the right parameters produced robust fluctuations in the populations of sharks and fish. Some readers, anxious to make Wa-Tor more like the earth, added special features to their programs. The game does invite complication and this is welcome. The major disadvantage of introducing a variant system, however, is that (other things being equal) comparisons with the standard system become dangerous.

Initial system-builders were Jean H. Anderson of Lauderdale, Minn., Stephen R. Berggren of Satellite Beach, Fla., Milton Boyd of Amherst, N.H., J. Connett of Minneapolis, Minn., Edgar F. Couda of Park Ridge, Ill., Don C. Hopkins of Champaign, Ill., Jim Lemon of El Segundo, Calif., Fredric Stevens of Davis, Calif., and Kenneth D. Wright of Grayling, Mich.

Among the questions these and other readers dealt with was that of measuring duration of survival. Clearly there is no problem for eternal populations, but it would be useful to have a measure of less-than-eternal scenarios. Measurement by chronons, as Stevens points out, can be misleading when extended life spans and breeding times are chosen for sharks. Measurement by cycles also has problems: what is a cycle? Stevens makes the amusing observation that if one's sharks and fish survive enough repetitions of the basic random-number cycle, an earlier configuration will repeat itself in concert with the cycle and eternal life is thereafter guaranteed.

A number of readers including David Emanuel of Oak Brook, Ill., Richard G. Fizell of Fort Washington, Md., and John S. Lew of the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y., described modern theories that are helpful in the analysis of Wa-Tor. I have yet to hear the final word on whether stochastic matrices will enable us to derive specific survival probabilities from arbitrary parameter combinations. It is interesting to note, however, that the Lotka-Volterra equations have (since their formulation in 1931) been further elaborated to consider diffusion as a factor affecting both predator and prey. Diffusion forces the smoothly varying solutions of the Lotka-Volterra equations into more complicated shapes. A historical note from Lew revealed that Alfred J. Lotka was an American mathematician who, a decade earlier, had formulated much the same equation as Volterra did.

Boyd exploited a phase diagram to analyze shark/fish population dynamics. At each time t plot the current numbers x of fish and y of sharks as the coordinates of a single point. As time advances and populations cycle, the point describes an erratic orbit about a fixed eye, or center. Boyd used the technique to study the effect of ocean dimensions on survival. He writes that "for the more rectangular worlds, the orbits lost their eyes, the trajectories became more jittery, and eventually became random walks." Square oceans are evidently preferable.

Among the innovations introduced by readers were a shark life-force, mutations, dual fish populations and plankton. I neglected to mention in my December column that the fish of Wa-Tor graze on omnipresent and omni-

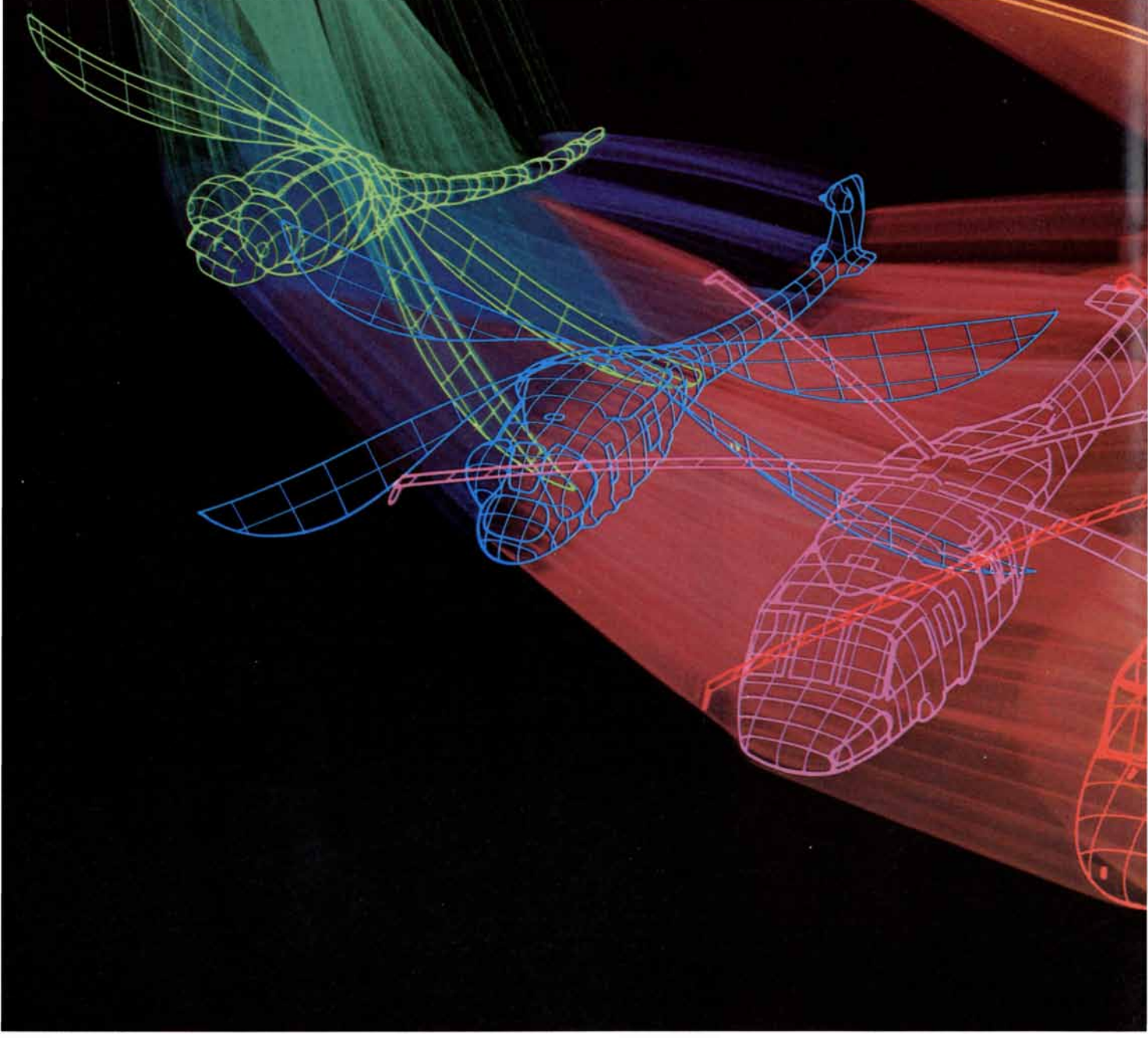
abundant oceanic plankton. Lemon made this feature explicit by placing plankton at every point not occupied by a shark or a fish. Plankton breed into otherwise empty spots and have the same relation to fish that fish have to sharks. Eternal populations exist here as well.

Couda's sharks gain or lose points of life-force depending on how well they eat. They can thus survive much longer without food than the primitive sharks of the standard Wa-Tor can. Couda sent plots (as did many of the other Wa-Tor programmers) that are remarkably similar to the Hudson's Bay Company data.

Connett uses two species of fish. One is the standard Wa-Tor variety; the other always breeds into any empty point to the south or east. Because of its mobile tendency, the second species often outlasts the first. Rudy Iwasko of Sacramento, Calif., proposed that sharks and fish be given characteristics of size, speed and agility. These were to be under genetic control. Berggren wrote his system, called EVOLVE, two years ago. It resembles WATOR except that it lets the animals evolve according to environmental pressures. In this way, Berggren reasoned, populations would arrive at an equilibrium favoring long-term survival.

No one succeeded in solving the toroidal pursuit problem. I shall now reveal one half of the solution so as not to deny readers the pleasure of finding the other half. Remember that at each turn the fish moves and then the two sharks move. As in Wa-Tor, standing still is not allowed. Imagine four rays emanating from the lone fish. Each ray follows a diagonal and twists around the torus, sooner or later rejoining itself. Once both sharks occupy a pair of opposite rays, it does not matter which way the fish moves; one shark pursues at a constant distance and the other shark closes in. The fish is doomed. I leave it to readers to discover how sharks hunt the rays, so to speak.

A new candidate for the five-state busy beaver was discovered on December 21, 1984, by George Uhing of Bronx, N.Y. Uhing's Turing machine starts on a blank tape and prints 1,915 1's before halting. The result was independently confirmed by Allen H. Brady of the University of Nevada and by Raphael M. Robinson of the University of California at Berkeley. Described by Brady as "astounding," Uhing's machine seems to justify the skepticism both mathematicians had expressed that Uwe Schult's machine (reported in "Computer Recreations," August, 1984) was the five-state busy beaver. It produced only 501 1's.



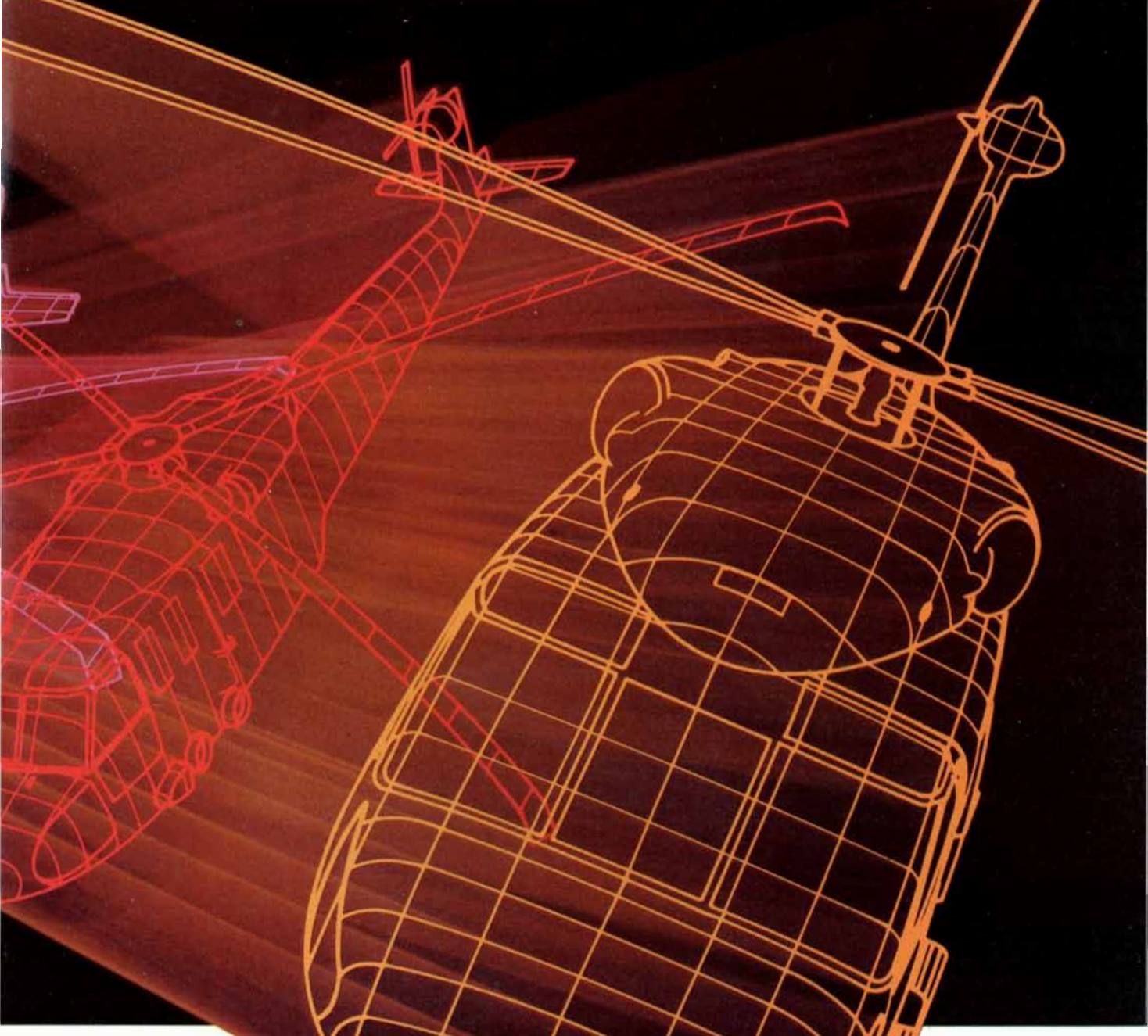
Artificial Intelligence at Sikorsky Aircraft

Exploring the evolution of vertical flight—at the speed of light.

The dragonfly perfected the fine art of vertical flight about 250 million years ago. Nature's organic design system gave us a perfect example of form following function.

Today, United Technologies Sikorsky Aircraft is building an "organic" design system for advanced vertical flight aircraft. And it's an information challenge that staggers the imagination.

Says Joe Piteo, Sikorsky Chief of CAD/CAM: "Our goal is to gather in one central system *everything* our engineers know. We want to extract and store the very process of their decision-making. Once we've added this 'creative logic component,' our system will

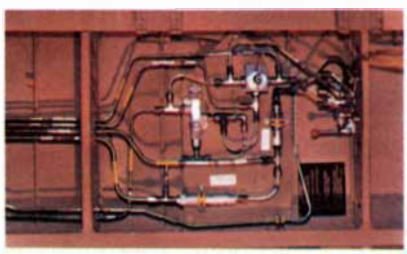


make a quantum leap from being merely a repository for data, to being an *information* base that is truly 'intelligent.' "

Using CAD/CAM software and two high-speed IBM 3081 computer systems, the Sikorsky team has tackled the design and manufacture of fluid lines. More than 2,000 tube drawings and their related design and manufacturing logic have been fed into the data system.

"Down the line," Michael Adami, project coordinator, says, "as more and more parameters are included in the software, the system will be able to 'see' more and more of the overall design. Ultimately, for example, the aircraft's tubing could be considered as a structural element and, as such, used in achieving the ideal blend of function, weight and strength."

"Our goal," says Piteo, "is to weld together these islands of knowledge and the automation of industrial design into a powerfully productive 'organism' which understands the relationship of its parts and works toward a common end."



With the help of an "intelligent" system, fluid line systems like these can be designed in minutes instead of weeks.

For more information on IBM's programs for engineers and scientists write IBM Engineering and Scientific Marketing, 1133 Westchester Ave., White Plains, NY 10604.

IBM®

'man·ij

Manage

- *Handle or direct with a degree of skill; e.g., apply high technology to global markets.*
- *Achieve; e.g., increased productivity through strategic investments supporting the work of 106,000 engineers, scientists and skilled employees.*
- *Succeed in accomplishing, as in record sales, record earnings, record return on equity.*

We're not out to change the world. Just to supply the technology it takes to make it better.

But technology alone doesn't bring 20.3% Return on Equity or nine straight years of increased earnings.

That takes management. Ours provides a culture that stimulates and directs the growth of high technology in four diverse business areas. To learn more about us, write: Rockwell International, Department 815S-62, 600 Grant Street, Pittsburgh, PA 15219.

Rockwell International

...where science gets down to business

Aerospace/Electronics
Automotive/General Industries

BOOKS

The size of life, aquanauts, a tale of old bones, U.S. atoms, Tokyo bagels, software

by Philip Morrison

ON SIZE AND LIFE, by Thomas A. McMahon and John Tyler Bonner. Scientific American Books, Inc. ($29.95). SCALING: WHY IS ANIMAL SIZE SO IMPORTANT? by Knut Schmidt-Nielsen. Cambridge University Press ($29.95; paperbound, $9.95). "The cells in an animal's body move through the world together not like a flock of birds but like the crew members and paying passengers on a steamship. We can visualize a liver cell standing at the deck rail... thinking that one of its kind would last alone out there for only a few minutes; it is too specialized to make it on its own." Thus McMahon and Bonner state an important theme shared by both volumes: As members of a community cells are obliged to obey common rules. Something tells each liver cell how fast to metabolize, even how long to live. There is one other major principle. Although cells are about the same size in all animals from water flea to whale, a grand order governs energy metabolism in warm-blooded animals of every size, and probably in other living creatures as well.

That grand order is expressed in the classical mouse-elephant curve, Max Kleiber's law. The measured points march along the fitted line, obedient to the relationship that sets the metabolic energy generation in every unit of body mass of mammal or bird. Power per gram goes smoothly down from skittery mouse to placid cow, and well beyond in both directions, inversely as the fourth root of body mass.

Birth provides dramatic confirmation of the principle. The infant all at once changes from a disciplined if fast-growing organ of the mother's body to a free-living six-pound primate. Within 36 hours the infant's metabolic rate doubles, coming to fit the mouse-elephant curve, not for the mother's weight but for the baby's. How can this happen?

No baby, of course, is in form a scale model of its mother. Although the heart weight simply scales, brain and kidney are proportionately larger for infants. The decrease in specific activity among larger mammals is not to be accounted for by a decrease in the relative size of their active organs.

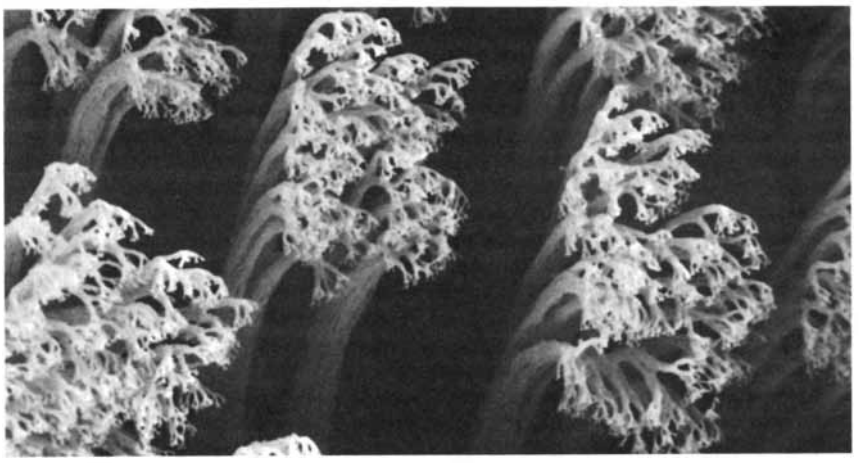
The investigation is technically difficult; the early work on tissue-culture metabolism in the 1920's was grossly in error. The constancy of the results the early physiologists reported is now seen to be due to their inability to provide conditions adequate for high respiration rates in tissue. Nowadays it appears that the various tissues of the mammalian body do obey the constraints set by the scale of the whole organism from which they are taken. The total metabolic rate is reasonably well represented by summing up the rates measured for the key tissues in vitro: mouse tissues in culture respire faster than rat tissues, which consume more oxygen per gram than dog tissues do, roughly as we expect. Look to a more molecular level. Although the hemoglobin concentration in mammals is independent of scale, the concentration of key respiratory enzymes such as cytochrome oxidase falls with body weight just as the specific metabolic rate does. For rats, rabbits, sheep and cattle the census of mitochondria in liver cells per gram of body weight again follows the ruling power law, and a recent study of muscle mitochondria showed the same overall scaling in a dozen species. The total body population of mitochondria—tiny, intricate, bacterialike organelles wherein resides the wheel of linked enzymes responsible for oxidative metabolism—rises with overall metabolic rate about as expected, systematically falling below proportionality. The liver cell at that metaphorical deck rail is not the same cell for mouse and man. Mouse cells hold more mitochondria, symbionts bound by ancient contract to supply biochemical fuel in exchange for shelter; theirs is the greater need.

The best account of a large-scale design rationale behind the mouse-elephant curve is the work of Thomas McMahon, one of the authors of the first book. Both texts present it well. It is an unexpected consequence of the idea that the mechanical loads on the supporting skeleton of mammals are limited by failure under bending. This yields neatly confirmed predictions for bone length and diameter in hoofed animals (and in the proportions of trees!). Plausible extension to concordance between the muscular system and the scaling of the bones it works on predicts the now classical metabolic result. Hints that similar rules apply even to animals of very different structure suggest the topic is not yet closed.

McMahon is a physicist-engineer; it is presumably his experience that is responsible for the delightful excursions of the McMahon-Bonner volume well beyond the natural history of size. Nails, bicycle-track design and aircraft engines are discussed; their relative simplicity forms a path to understanding the subtler, patient engineering of nature. The mouse-elephant curve finds its counterpart in a plot of brake horsepower against engine mass for internal-combustion engines. Anchoring the mouse end of the curve is the not-so-mousy Webra Speedy, a model-aircraft engine weighing about a quarter of a pound. A 100-ton ship's diesel stands at the upper limit. The fit is good, although imperfect; the principles invoked are simple. The most instructive result for engines, whose varieties of fuel, ignition and cylinder number seem to belie any unity, is a slow decline of r.p.m. with engine mass, just right to cancel the observed rise in stroke size with weight. The two together imply the constancy of peak piston speed. The inference is that engine designers are aware of peak strain energy per unit of volume of metal and seek to reach it under a broadly similar factor of safety.

The straight-line plots on log-log graph fields that spot the pages of these books are not precise. Rather, they are unifying and instructive. As Professor Schmidt-Neilsen reminds us, even the deviations seem not to be mere noise; "they appear to be secondary signals that convey a specific message," whether about the value of many engines on a large airplane or about the extra metabolic power appropriate for marine mammals, who must stay warm in so powerful a cooling medium as icy seawater.

Locomotion is a topic emphasized in both books. Organisms fly, swim, run, walk and jump; some climb on smooth vertical walls, others stride nimbly across the surface-tension film of water. Under analysis most of these means yield striking regularities. Here the physics is somewhat more demanding. One simple and illuminating result is the fernlike structure in



Branched endings of the tufts on a gecko's toe pad, enlarged 2,350 diameters

the feet of the elegant wall-climbing gecko. So large an animal—much bigger than the fly on the wall—needs a remarkable length of adhesive interface between limb and smooth surface. There it is, cunningly folded within gecko toe pads at submicron scale, hidden from the eye but not from the scanning electron beam. Even simpler is the account of the oxygen demand of animals running uphill and down at different speeds. A mouse breathes at a constant rate whether it runs up or down a 15-degree grade. Chimpanzees double their oxygen use during such a climb. Thus can squirrels clamber effortlessly up and down tree trunks as human climbers cannot. (The result is another corollary of the mouse-elephant curve.)

The McMahon and Bonner volume, with its artful metaphors, its dozens of color photographs and its wide overview of the physical and biological issues of size, is the most interesting and the freshest among the excellent series of introductory books to which it belongs. It came as a real gift, a text for beginners and adepts alike on a fascinating topic in a field that lacks recently published texts. (A fine monograph by Robert Henry Peters on the larger significance of such biological order was noticed here in September, 1984.)

Then the Schmidt-Nielsen work appeared. It has narrower scope; it offers few photographs but many graphs; it is centrally and deeply biological. Its concise arguments, engaging observations and overall physiological wisdom are all one would expect from an author whose own series of informing books is a long-admired mainstay of quantitative biology.

McMahon and Bonner treat the Reynolds number in detail, to establish its physical meaning, whereas Schmidt-Nielsen describes it mainly as a handy index of incipient turbulence. On the other hand, the joint authors do not much attend to biochemistry or cell biology, whereas Schmidt-Nielsen offers an excellent account of mitochondria and does not neglect enzymes and their substrates. A direct comparison would be close to odious; these books are a pair of aces unexpectedly dealt to the general reader in physical and biological science alike.

LIVING AND WORKING IN THE SEA, by James W. Miller and Ian G. Koblick. Illustrated by William Boggess. Van Nostrand Reinhold Company ($32.95). About 800 men and women in all the world are aquanauts, defined in this devoted and knowing chronicle as people who have lived on the sea floor, open to its ambient pressure, for at least 24 continuous hours. (The far more numerous submariners voyaging month after month safe inside strong pressure hulls, breathing air at surface density, need not apply!) A certain *tristesse* lies here between the lines; the narrative details the rise and fall of this adventurous form of engineering exploration, which proved too costly for the underwater amateurs and scientists who can best use it, yet not efficient enough for well-heeled commercial enterprises of undersea construction and maintenance.

The practical development began ashore in the pressure chambers of the U.S. Navy in about 1957. By 1964 both Edwin A. Link and Jacques-Yves Cousteau had joined the effort with admirable effectiveness; a couple of dozen divers, including Madame Simone Cousteau, had by then lived and worked on the ocean floor, as deep as 400 feet, for as long as a month. The diver's ancient dream of more time on the bottom had been fulfilled after 5,000 years. Sea-floor habitats, spheres and cylinders of many sizes, single or clustered, sought the bottom in continental seas from the high Arctic to the Sea of Japan; a couple even occupied lake floors. Some habitats were meticulously engineered, some were hastily improvised from plastic tents, railroad tank cars and even a cement mixer. The idea caught on: there appeared the enthusiasts, television coverage, intriguing names for projects, a race for records at depth, "all-girl teams" of aquanauts, as the media put it. Even landlocked Czechoslovakia did not ignore the art; clubs of brave sports divers there organized several habitats. Sixty-five sea-floor shelters in all were emplaced by 17 countries. Within a decade this vivid activity was all but over; just one habitat, Hydrolab, is still in regular scientific use at a depth of 50 feet in the U.S. Virgin Islands. The majority of the world's aquanauts have lived and worked in that single lab.

Life at high pressure has its risks, which are now well known. Too much oxygen is toxic; the remedy is to supply breathing gas whose total oxygen density is close to that of normal air. At high pressure, supply a proportionately smaller fraction of oxygen. Nitrogen at high pressure induces effects on the central nervous system, the narcosis of the deep, affecting judgment and even leading to stupor. Instead dilute the oxygen mainly with helium. Its effects, which include a timbre of speech reminiscent of Donald Duck, are very small.

Given the right chemistry, the chief risk is not high ambient pressure but rather pressure change. If decompression is swift, the gas released from solution forms tiny bubbles in all the fluids of the body; easily fatal nerve-tissue damage and circulation embolisms result. Slow decompression is therefore obligatory. At about 100 feet decompression takes at least a full shift, and a diver has no time left to work below. The way out became clear in about 1942. After a sufficiently long stay at high pressure the diver is saturated. No more gas will dissolve. Let her work and sleep therein; there is no need for daily decompression, and she can put in day after full day at high pressure without danger. Indeed, we know of no reason why a properly sheltered and supported diver could not live at high pressure for a year, decompressing only for vacation.

The sea-floor habitats were the first exploitation of saturation diving. The divers live well enough in their habitat; they sleep there, they cook and rest. (From time to time they have been known to catch lobster to enhance the rations dropped from the tender above.) The surface tender supplies food, water, information, electrical

power, breathing gas for excursions out of the habitat and emergency aid. Bad weather at sea routinely threatens the surface link, and decompression to surface pressure takes time. To skimp on that time is fatal; the partnership has become a danger.

The expansion in undersea construction demanded in the past decades by the offshore oil and gas industry has brought a large increase in saturation diving. But nowadays the pressure jump is made right on shipboard or platform. The divers live and sleep under pressure, all right, but there on the surface, in a deck decompression chamber. They are shipmates with their life-support system; no one is underseas at cable end in bad weather or at night. Weather and gear permitting, they daily report to work below, at a pressure close to their chosen off-duty pressure in the deck chamber. They travel to the depths in some modern form of diving bell at a similar pressure. The installation is expensive but safe; unlike the sea-floor habitat, it needs no fragile links.

Even the days of these surface-based saturation divers seem numbered. At depths of 300 feet or more robot vehicles, their television cameras and remote electronic grips and tentacles becoming more effective every year, will soon win out. Little one- and two-person submersibles, their pressure hulls filled with normal breathing air, are also much in use. Soon the deep-diving saturation systems will become as scarce as sea-floor habitats are today. For modest depths the human divers may retain a role, with their decompression times occupying at most a couple of days.

In an exposition that is mainly of historical interest, the engineering and architecture of the sea-floor habitat are nicely set out. There are unique problems: the Donald Duck voice of people breathing helium is a source of communications trouble. The electronic remedy of frequency shifting is available. Helium is chilling to the divers because of its high conductivity, and it leaks easily into television tubes and similar sealed electronics. Humidity is unhealthily high over the seawater lapping at the entrance hatch. Air conditioning is much needed. Ports and windows offend the pressure-hull designers, but the divers love them. There are safe designs, executed in both glass and plastics. A fine eight-foot transparent acrylic sphere was used as a winter habitat at a depth of 40 feet in Resolute Bay under three feet of sea ice. It must have been interesting to watch the Canadian divers, floodlit, walking in their inflatable suits upside down on the ice ceiling.

Many such personal accounts by aquanauts are found here. Early in one program wild sea lions came to call on the habitat. Earthquake shock was experienced in several habitats.

There are tips on many points of operation and design. For cables through the pressure hull, be careful with the O-rings. Divers find water beds agreeable. Glue sniffing, an accident of rubber-suit repair, is more dangerous at high pressure than it is at low. Parrots, frogs and rabbits all make satisfactory pets. Once fed, the local fish can become frequent guests; one big grouper inhaled an entire tin of sardines.

The submerged habitat will return, maintain these authors, two Florida veterans of sea-floor life. Archaeologists, biologists, students and romantics can benefit from dwelling on the bottom, instead of merely making day trips from a confined equal-pressure world. There is a plan under way to place a new habitat for the use of University of Southern California science students 1,200 feet off Catalina Island, at a depth of 50 feet in a bed of giant kelp. All its surface support will come by way of umbilicals from shore. (One hopes they have no plan to pump nutritive slurry into the habitat for food, as one Soviet project did.) With the darkling kelp forest and its high-flying otters in mind, every reader will assent to the authors' remark that the sea-bed resort hotel for diving tourists is "an idea whose time has come." Catalina may be the pioneer.

ASCENT TO CIVILIZATION: THE ARCHAEOLOGY OF EARLY MAN, by John Gowlett. Alfred A. Knopf, Inc. (paperbound, $14.95). In a deep, hidden gully on the plains of eastern Colorado the spade of the archaeologist has disclosed that 9,000 years ago hunters stampeded a couple of hundred terrified bison to their death. In the echoing sea cave below the cliffs of La Cotte de St. Brelade on the island of Jersey other excavators have found heaps of large skulls and shoulder blades, some carefully stacked. Hunters there had made their kills just as had those on the distant Great Plains. Here their game was not bison but mammoth and rhino, unknown in European memory. No small island can harbor many such beasts; plainly the hunt was active long ago, while the seas were withdrawn and Jersey was part of the mainland. The stratigraphy suggests a time a quarter of a million years ago; the Old World tells a story of humanity much longer than any we can read in the New.

Look at another newly studied site, Chesowanja in the Kenyan Rift Valley. There the signs of hominids occupy in space a shallow hollow a square kilometer or so in area; the traces are found at a time depth of more than 1.5 million years. Among small tools unearthed by the thousands from the earliest beds the most exciting artifact is a set of lumps of fired clay. Magnetic measurements of baking temperature are consistent with a campfire; there is a pattern of stones suggesting the hearth. This may be the oldest sign of human control of fire. A few clear pages of text accompany a spread showing reconstructions of the area in color perspectives at five times during its entire Pleistocene history. There are aerial photographs of similar landscapes in the Rift as they look today, and a few personal close-ups, one of an archaeologist intently at work on a half-buried bone with paintbrush and spiked tool (the furrowed brow is that of Bwana Bernard Kanunga, a Kenyan colleague of the author's).

All of this would be meritorious in a study of a single site, particularly the perspectives. But in fact the book is an overall survey of the entire span of hominid archaeology: from the first signs, footprints in the volcanic ash, through the long span of *Homo erectus* to the coming of *H. sapiens,* and finally the spread over all the world by means of new skills, art and domestication. The last tenth of the book offers a capsule account of metals, writing, wheels and ships, and closes with a photograph of the suspension bridge across the Humber. The subtitle states the content well; the rest of the title seems more a publisher's contrivance.

Two pages treat the origins of fire. Caves shelter the signs of hearths, but on the open sites of the deep past it is only by good luck that signs of fire can be made out. After all, "neither lightning, nor lava flow, can build a hearth arrangement." Animals do use natural fires; hawks and cheetahs station themselves to catch prey fleeing the flames. Perhaps it was not the Titan who first brought fire to humankind, but slow episodes of gradual learning. We know of no peoples who did not use fire, even if some did not kindle it. Those early hominids who made stone tools knew well how to work with wood, the material around which fire making is focused. The task of fire building is no more challenging intellectually than knapping blades out of flint. We may have to face the conclusion, perhaps still not certain, that the scavenging primates a couple of million years back were fire users "to all intents and purposes... already human in their basic character."

Dr. Gowlett argues on the skeptical side of the trendy theories of the day. The change from the earlier hominid

Models of everyday life: figurines of bakers and brewers from an Egyptian tomb

forms to the later seems to him gradual and prolonged. The widespread hand-ax culture may yet disclose an early role of Asia, although so far Africa is the main site of our first ancestry. The close length-and-breadth correlation in hand-ax shape at early sites, a million years ago, shows the presence of mental templates for their form. The copying of the form might have been purely visual, but it is plausible that language was already there. It is mental modelmaking, not forms of social aggregation, that marks humankind. Making tools, fire or sentences demands abilities to act according to some internal model. "Social co-operation is the ultimate basis for our communication, but it does not explain its refined intricacy. The simplest general explanation is that ... natural selection has constantly favoured the ability to anticipate events and actions." We are at an early stage of understanding our past.

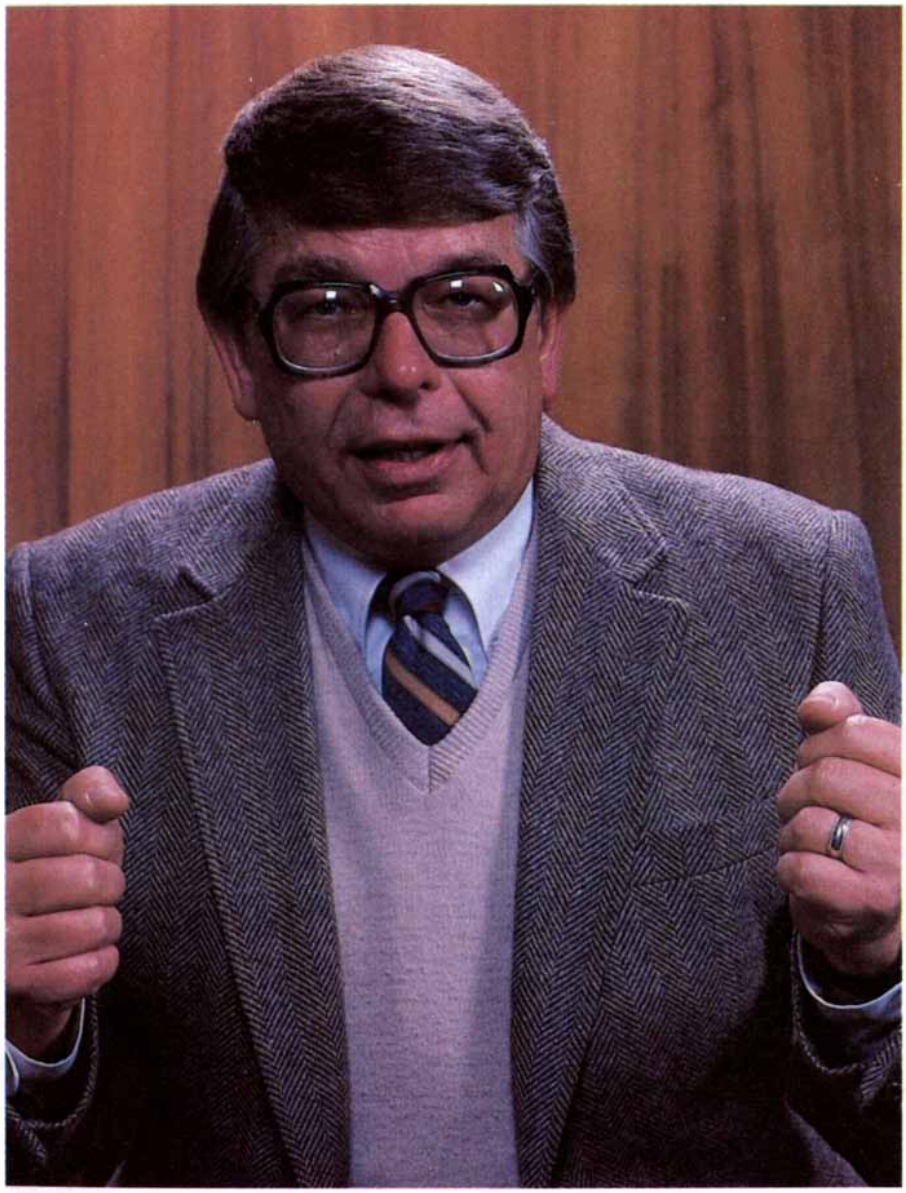
The text ends in hope. "It is very difficult to say that the human course has been retrograde....the people of the past ... not only survived, but steadily added to the cultural heritage.... If precedent counts for anything, then the long perspective suggests that we shall continue to get by."

The volume is organized into some 92-page spreads, each half text and half image. The pictures have been collected or made anew in one of those expert London establishments devoted to showy but comprehensive treatments. The staccato organization and the visual emphasis have in fact been given altogether first-rate intellectual value by the author, as the examples above should show. Dr. Gowlett is a young Oxford archaeologist who has dug lately and extensively in the Rift; his familiarity with new methods, new sites and new conclusions, and above all his sensitivity to the nature of scientific evidence, put this inexpensive book at the top of general studies of the kind. It is exciting to the eye, readable and honest.

THE AMERICAN ATOM: A DOCUMENTARY HISTORY OF NUCLEAR POLICIES FROM THE DISCOVERY OF FISSION TO THE PRESENT, 1939–1984, Robert C. Williams and Philip L. Cantelon, editors. University of Pennsylvania Press ($30; paperbound, $8.95). Surprising riches in a little room, this compilation of documents by two historians presents the main issues through well-excerpted primary texts. The direct evidence of their times, these materials range from laws and treaties to the comments and directives of presidents and physicists, journalists and bishops. There are chronological tables of tests and of weapons accidents, but no visual images at all.

The generous editors offer more than their title claims. The oldest selection is the work of the prescient H. G. Wells, who told the readers of his novel in the year 1914 about the highways thronged with "shimmering shapes of silvered steel" run by the cheap atomic engines his fantasy foresaw for the 1950's. But in his vision there burst over those happy travelers before 1960 "the unquenchable crimson conflagrations of the atomic bombs," lumps of pure Carolinum, storing "in a handbag an amount of latent energy sufficient to wreck half a city.... The power of destruction ... was now the only power left in the world." It is fiction, of course, but it is given deserved primacy here by the real letter that went from Leo Szilard to a friend along with these few pages of the Wells novel 50 years ago. Szilard applied for patents on possible nuclear chain reactions in June, 1934, before fission but after neutrons, expressing his doubt in the memorandum printed here that coal mining or oil production could survive

WHAT SHOULD YOU WORKING "PARTNERSHIP"



"At Ford Aerospace, HP automatic test sets cut test time by 90%, saving $3 million a year.

"We've used HP instruments and computers for 10 years. In that time, we've quadrupled the circuits on our communication satellites while cutting test time by 90%. By automating with HP test sets we save up to $3 million a year."

Ford Aerospace designs and manufactures commercial communication satellites. In Palo Alto, Ca., John Merizon, Manager of Systems Test & Operations says: "HP's engineering staff helped us assemble HP instruments and computers to build 4 universal satellite test systems.

"Because satellite testing is a 24 hours a day, 7 days a week operation, downtime is irretrievable and translates directly into schedule slips which could cost us up to $30,000 a day. HP saves us the time and money of routine service calls by training our staff to calibrate and maintain HP equipment in-house. Because our HP systems deliver 99% uptime, we've cut down on repeat testing and virtually eliminated costly delays.

"HP is available for each launch with spare parts and a support team to back us up. Based on the flexibility and ruggedness of HP instruments and HP's consistent high performance, we plan to expand our test program with additional HP equipment."

EXPECT FROM A WITH HEWLETT-PACKARD?

'At Mount Carmel East Hospital, an HP Patient Care Management system delivers 99% uptime and improves department productivity.

"Our HP system enables us to streamline unit operations, improve productivity and reduce operating costs."

Mount Carmel East Hospital, in Columbus, Ohio, is a member of the Holy Cross Health System. The hospital uses HP's Patient Care Management System—an HP 3000 computer-based information system.

Judy Willis, Vice President of Nursing says: "The HP system links clinical and administrative data into one network. By customizing the system to meet each department's needs, we've improved our response time. On-line order processing facilitates fast, accurate billing. And by eliminating delays due to lost or inaccurate data, we've reduced operating costs—a savings we can pass on to our patients.

"The modularity of the HP 3000 allows us to accommodate future growth. And with HP's commitment to R&D, we've not only invested in a computer but in our future. Our HP system delivers 99% uptime. With that type of reliability, we believe we have the best price/performance system available today."

Results with assurance.

HEWLETT PACKARD

0002508

after a few years of successful nuclear development.

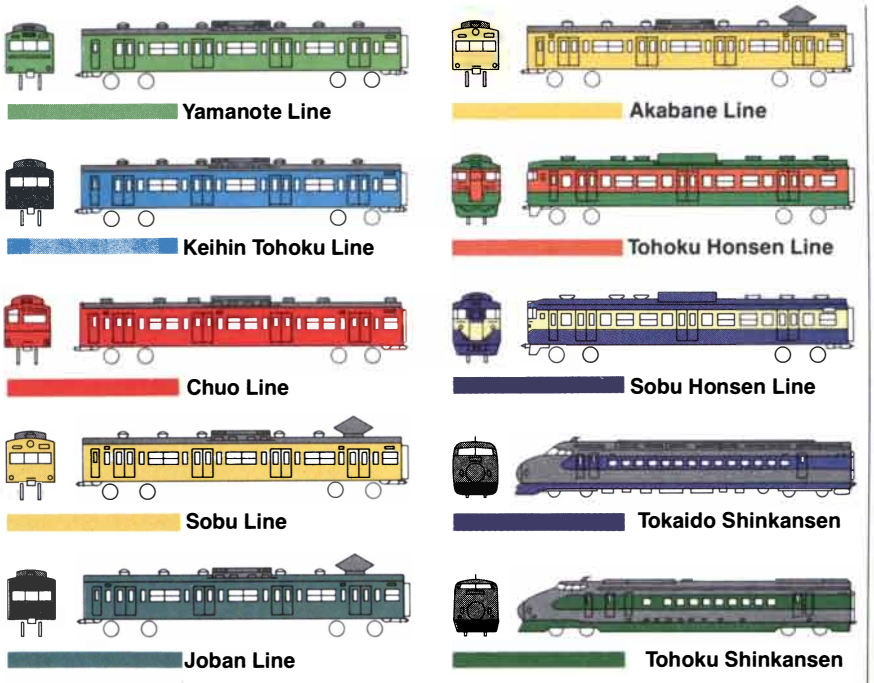
The latest documents cited appeared at the end of 1982; they include a long-secret account by Hans Bethe of the beginnings of thermonuclear weapons, the pastoral letter of the Catholic bishops of the U.S. on war and peace in the nuclear age, and a speech of President Reagan's on the MX, "the right missile at the right time." In between lies the history of our atomic times in 70-odd texts: the Manhattan project, the H-bomb decision, the matter of J. Robert Oppenheimer, the theory and practice of the arms race, the Enrico Fermi fast breeder and Three Mile Island.

Every such collection, even one so useful and representative as this brief anthology, induces knowing readers to conceive additions. A table of strategic-weapons inventories of the mid-1980's and a concise listing of nuclear power plants worldwide, both widely available, would add valuable concreteness to this policy-centered collection. The single terse policy statement one most misses was sent to Secretary of War Henry L. Stimson just a month after the Trinity test, over the signature of J. R. Oppenheimer "for the Panel" (with Compton, Fermi and Lawrence): "We believe that the safety of this nation—as opposed to its ability to inflict damage on an enemy power—cannot lie wholly or even primarily in its scientific and technical prowess. It can be based only on making future wars impossible. It is our unanimous... recommendation to you that ... all steps be taken, all necessary international arrangements be made, to this one end."

Field Guides

Here are two recent displays of the data, reasoned summaries of what is out there in some well-defined subset of the world.

TOKYO ACCESS, by Richard Saul Wurman. Distributed in the U.S. by Random House, Inc. (paperbound, $11.95). Bright as a flag, this bilingual text is printed in five colors to code shops, gardens, nightlife and food, museums and hotels. A tall-format pocket book, it lays out all public—but not private—Tokyo with many maps and drawings, district by district, door by door in a few thousand entries. It spans the city as well as any visitor might want, from the 1891 quartz benchmark that sets the zero of elevation for all Japan, "even that of Mt. Fuji," to the electrical bazaar of Akihabara, where they sell chips, components, sing-along audio wares and appliances, in hundreds of shops large and



Orient expresses and locals: color-coded subways and trains from Japan

small. There are a fine page on soba (buckwheat noodles), drawings of the bridges of the city, a summary of its history, a map of its industrial surroundings, the location of the Noh theaters and the Sword Museum, a capsule account of Tokyo Disneyland (the biggest of all of them), identification of the best restaurants for horsemeat and for bagels, and a report on the "silent, *en-masse* procession" in navy blue suits commuting in and out of Tokyo Station that the author sees as the "backbone of *Japan-the-economic-power*."

There are many Japanese contributors, even including poets. One writes: "Tokyo is a giant spoon . . . an indiscriminate mixture of foods—delicious, evil-tasting, sweet, spicy, fragrant or reeking . . . everything is gathered here . . . dip in and ladle out forever . . . and never reach the bottom." Tokyoites and *gaijin* alike all carry maps there; this is surely a fine one. Perhaps it is even better for stay-at-homes who will merely read of what has become of Edo, high and low. The editor offers half a dozen other guides in the same style to U.S. cities south, east and west.

WHOLE EARTH SOFTWARE CATALOG, Stewart Brand, editor-in-chief. Quantum Press/Doubleday (paperbound, $17.50). The "access to tools" of a decade or two ago, those catalogues and books on inflatables, obstetrics and Swedish chain saws seem far away. But this entry into the dazzling new marketplace of "empowering . . . mind tools," the hardware and software of personal computing, is carried out with the same critical and comparative insight and the same energetic mix of social iconoclasm and philosophical hopefulness that distinguished the chief editor and his crew in more turned-on times. Eleven domains of software and hardware are broken out, from play through accounting to miscellaneous applications. The team—it is a larger one now—recommends more than 350 programs, as well as computers and books. The work remains timely, candid, explicit and informed. Well illustrated, the book is a design and typographic success. Although it is no longer a uniquely catholic compilation, its quality remains unequaled within that little universe confined to doped silicon.

Would-be buyers will manifestly be repaid in full for using it. There is much of value beyond the market too. Stewart Brand remains the candid aphorist: "I think the notion of 'letter-quality' printers is about as deep as 'wood-quality' station wagons. Most . . . are expensive, gawdawful noisy, and huge, and they can't even do graphics." There is solid matter of general interest everywhere among the ratings. "CitiCorp in Manhattan sends its data locally by laser beams and microwave, coaxial cables, and fiberoptic lines running down the IRT subway line . . . since 1918, the bank has had a pneumatic tube system in the IRT." No system today carries as much data for the bank faster than those big old cylinders blown uptown at 75 m.p.h. stuffed with a load of floppy disks.

	MERKUR XR4Ti	BMW 318i*	SAAB 900 TURBO	AUDI 4000S QUATTRO	VOLVO TURBO
COUNTRY OF MANUFACTURE	West Germany	West Germany	Sweden	West Germany	Sweden
LAYOUT	Front engine/ rear drive	Front engine/ rear drive	Front engine/ front drive	Front engine/ all wheel drive	Front engine/ rear drive
ENGINE	2.3L OHC 4-cylinder Fuel-injected Turbocharged	1.8L OHC 4-cylinder Fuel-injected	2.0L DOHC 4-cyl. 16 Valve Fuel-injected Turbocharged Intercooled	2.2L OHC 5-cylinder Fuel-injected	2.1L OHC 4-cylinder Fuel-injected Turbocharged Intercooled
TRANSMISSION	5-speed manual 3-speed auto	5-speed manual 3-speed auto	5-speed manual 3-speed auto	5-speed manual auto N/A	4-speed manual + Overdrive 4-speed auto
HORSEPOWER (SAE net @ rpm)	175/5000	101/5800	160/5500	115/5500	162/5100
TORQUE (ft-lbs @ rpm)	200/3000	103/4500	188/3000	126/3000	175/3900
SUSPENSION, front & rear	MacPherson strut/ Independent semi-trailing arms	MacPherson strut/ Independent semi-trailing arms	Transverse A-Arms/ Beam axle, trailing arms, Panhard Rod	MacPherson strut/ Independent coil/shock absorber struts	MacPherson strut/ four link, live axle, Panhard Rod
STEERING	Rack-and-pinion	Rack-and-pinion	Rack-and-pinion	Rack-and-pinion	Rack-and-pinion
STEERING TURNS, lock-to-lock	2.8	3.9	3.7	3.1	3.5
BRAKES, front & rear, in.	10.2 Disc/ 10.0 Drum	10.2 Disc/ 9.0 Drum	11.0 Disc/ 10.6 Disc	10.1 Disc/ 9.6 Disc	10.3 Disc/ 11.0 Disc
TIRES	195/60HR-14	195/60HR-14	195/60VR-15	195/60HR-14	195/60R-15
WHEELBASE, in.	102.7	101.2	99.1	99.4	104.3
VEHICLE LENGTH, in.	178.4	176.8	186.6	176.6	188.8
CURB WEIGHT, lbs.	2920	2392	2885	2824	3045
WEIGHT DISTRIBUTION, front & rear, %	53/47	55/45	61/39	59/41	54/46
SEATING CAPACITY	5	5	5	5	5
FUEL CAPACITY, gals.	15.0	14.5	16.6	18.5	15.8
AERODYNAMIC DRAG, coefficient	0.33	0.38	N/A	0.42	N/A
1985 EPA FUEL ECONOMY** City/Highway, mpg	19/24	N/A	19/25	18/22	17/22
0-60 MPH Acceleration, sec.	7.8	11.4	8.7 Est.	9.5	N/A†

*1984 Model Year Data **EPA estimates. Actual mileage will vary with maintenance, options, driving conditions and driving habits. †Volvo Turbo 0-55 MPH Acceleration is 6.9, 0-60 MPH N/A
SOURCE: 1984/85 Manufacturers' Data

NOT TO COMPETE WITH THE LIKES

BUT

That's a very ambitious goal. But then, this is a very ambitious automobile.

The Merkur (pronounced mare-koor) XR4Ti is a German sports sedan designed and engineered specifically for those who appreciate the exhilaration of a finely tuned road machine.

Its shape is unusual because it achieves a level of aerodynamic efficiency that is unusual. In fact, it has a drag coefficient of only 0.33—the best of any car in its class.

The XR4Ti's 4-cylinder, 2.3-liter overhead cam engine uses a Garrett AiResearch® "blow-through" turbocharger in conjunction with electronic fuel injection to develop a healthy 175 horsepower.* The "blow-through" design delivers excellent throttle response by introducing the fuel downstream of the turbo compressor. The result is 0-60 mph acceleration in only 7.8 seconds.

Responsiveness is further enhanced by the XR4Ti's exceptionally quick steering. Only 2.8 turns lock-to-lock. Variable-ratio, power-assisted rack-and-pinion, of course.

And because all roads are neither smooth nor straight, the XR4Ti has a fully independent suspension with nitrogen-pressurized

MERKUR

MERKUR XR4Ti IS HERE.
ITS OBJECTIVE IS
OF BMW, SAAB, VOLVO AND AUDI.
FRANKLY, TO SURPASS THEM.

struts and shocks. And for added stability during hard cornering, there are stabilizer bars front and rear.

The XR4Ti's dedication to the art of driving is apparent in the functional design of the interior. The instrument panel wraps around the driver, providing ready access to all vital controls. The instrumentation is of analog design to allow for easy reading with just a quick glance, day or night.

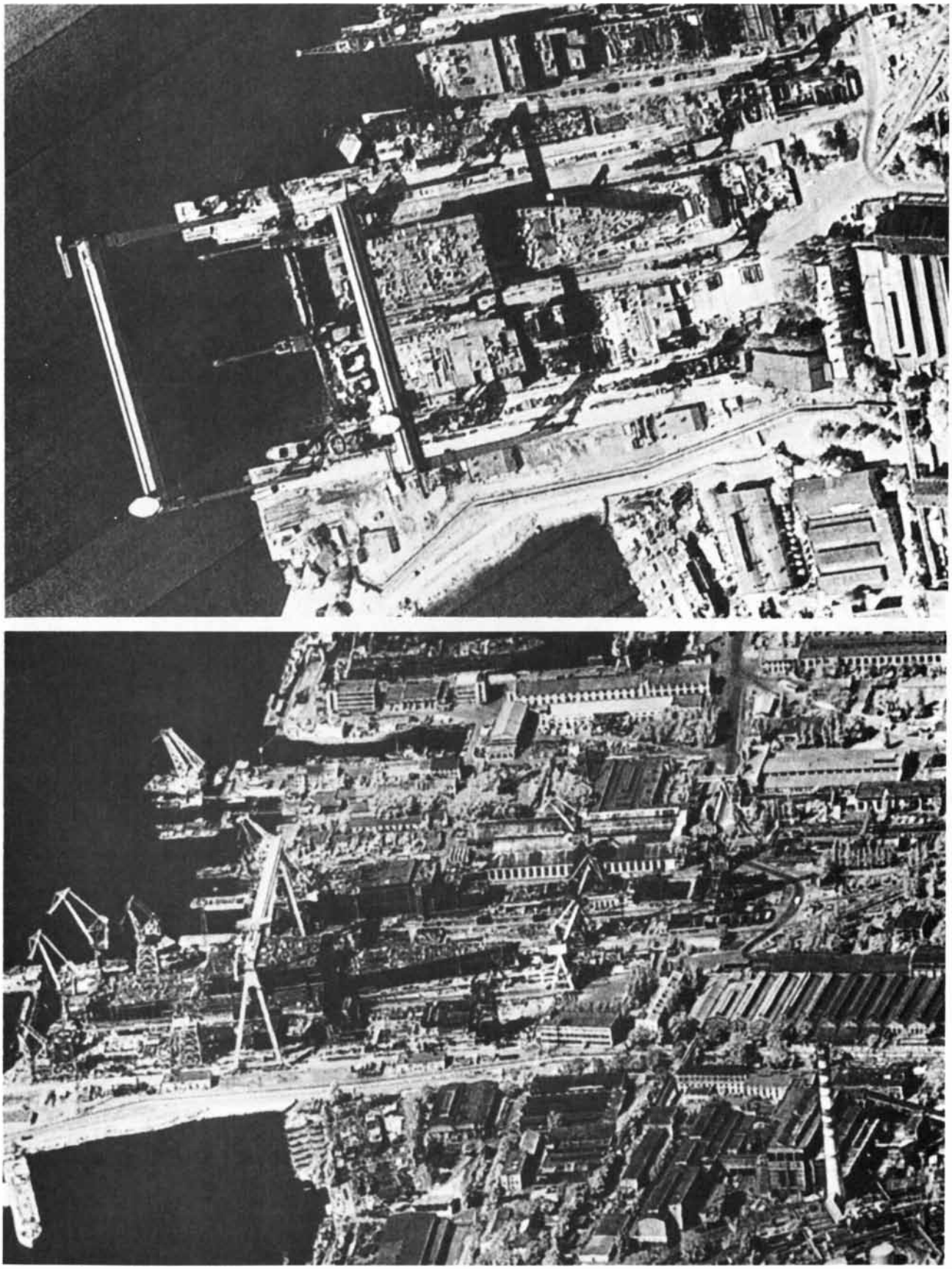
Its comprehensive list of standard equipment includes 5-speed manual transmission, Pirelli P6 tires, cast-aluminum alloy wheels, left- and right-side rear seat 3-point safety harnesses, air conditioning, heated electric outside mirrors, and Grundig® AM/FM stereo cassette.

The XR4Ti is the first of an entire new line of Merkur automobiles coming to America. It is assembled in Germany by the famous coach builder Karmann and available only at selected Lincoln-Mercury Dealers. Visit one bearing the Merkur sign and experience the exhilaration that is XR4Ti. *Based on SAE Standard J1349

Ford

IMPORTED FROM GERMANY
EXCLUSIVELY FOR LINCOLN-MERCURY

MERKUR XR4Ti



COMPUTER-ENHANCED SATELLITE PHOTOGRAPHS reveal the Soviet navy's first nuclear-powered aircraft carrier under construction at a shipyard on the Black Sea. The 75,000-ton warship, which is expected to be named the *Kremlin* when it becomes operational in 1994, is being built in two sections; the 264-meter-long bow section and the 73-meter-long stern section can be seen in adjacent slips under the giant gantry crane. The photographs were first published last year in *Jane's Defence Weekly*.

Established 1845 SCIENTIFIC AMERICAN March 1985 Volume 252 Number 3

The Verification of Compliance with Arms-Control Agreements

Military activities in the U.S.S.R. can be unilaterally monitored by the U.S. with the aid of a wide spectrum of remote-sensing technologies, including high-resolution satellite photography

by David Hafemeister, Joseph J. Romm and Kosta Tsipis

The recent decision by the U.S. and the U.S.S.R. to resume bilateral arms-control negotiations was heralded by both governments with public statements emphasizing the need to halt and even reverse the growth of their nuclear arsenals and related weapon systems. Given the present climate of mistrust between the two nations, it is imperative that compliance with the terms of any treaty that might emerge from the new negotiations be unilaterally verifiable by each side. Verification requires that each nation have a reliable and objective means of monitoring the military activities of the other. Accordingly both nations have over the years developed a completely new family of systems for gathering such information from a distance. The systems are collectively referred to as national technical means of verification.

The issue of verification is quite separate not only from the legal question of whether an observed activity constitutes a treaty violation but also from the political question of what to do about a violation once it has been detected. Nevertheless, verification is more than just a technical issue. In the U.S., for instance, verification has been a focal point of the recurrent political debate over the merits of various existing and proposed arms-control treaties. Opponents of any particular treaty tend to argue that the U.S. cannot verify Soviet compliance well enough to maintain national security under the terms of the treaty, whereas advocates of the treaty tend to argue that the U.S. can.

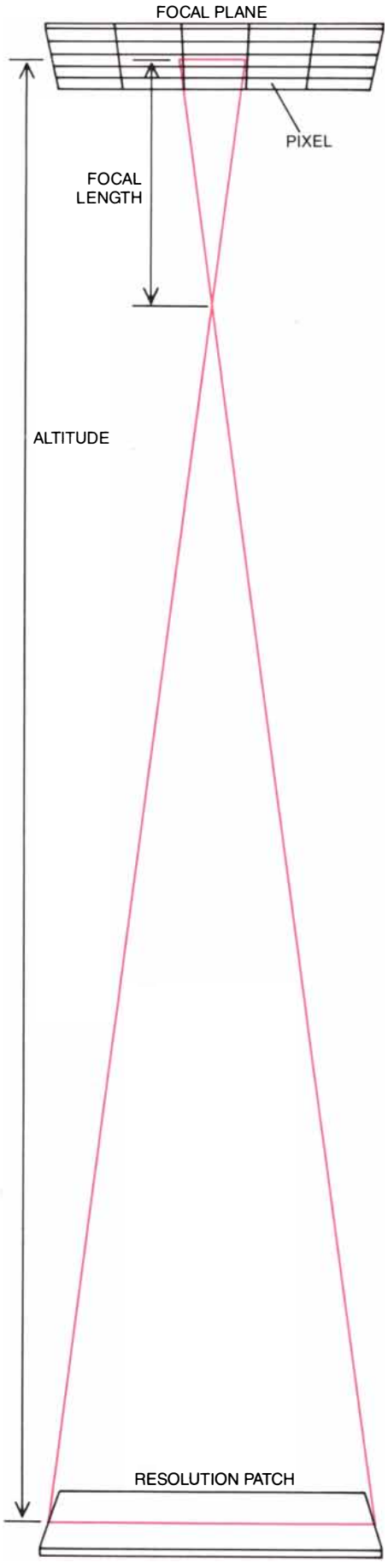
In general the question of whether or not a treaty is verifiable by the U.S. can be reduced to two narrower questions. First, at what level of clandestine Soviet activity would U.S. security be jeopardized? Second, is the U.S. system of verification capable of detecting that level of activity?

What are the technologies needed to observe the development, testing, production and deployment of a weapon system? The earliest detectable signs of development might be communications between officials and scientists working on the weapon; some of these signals could be intercepted by monitoring radio waves. Alternatively, a spy might observe the weapon or its plans directly. According to a statement made in 1979 by William J. Perry, then Undersecretary of Defense for Research and Engineering, "we monitor the activity at the [Soviet missile] design bureaus well enough that we have been able to predict every ICBM [intercontinental ballistic missile] before it ever began its tests.... We have always in the past been able to detect an ICBM's existence before it ever went into tests."

Once a weapon arrives at the stage of testing outside the laboratory, the full spectrum of acoustic and electromagnetic energy can be exploited to gain information about it. Visible-light detectors such as those on photoreconnaissance satellites can "see" the weapon directly; infrared detectors on satellites can sense the heat emitted by a rocket exhaust plume during tests; radar can track a weapon in the air; sonar can track one in the water, and seismometers can detect and assess an underground nuclear test.

After a weapon has been fully tested it moves into production. Production involves large-scale transportation to and from manufacturing plants, an activity that can be monitored with the aid of visible light or infrared radiation. Finally, the operational deployment and training of personnel with the new weapon can also be monitored. The main difference in observing these processes is that in general testing and production must be detected while they are happening, whereas deployment involves activities that continue for years and can be detected at any time after they begin.

Determining whether the U.S.S.R. is violating the terms of a treaty, however, does not require observing everything having to do with the laborious process that culminates with the weapon's deployment. On the contrary, only parts of the activities involved, if they became known in sufficient detail, would be enough to identify their purpose and betray the violation of the agreement. This process of detection and identification is facilitated in a number of ways by the natural synergy among the various means of gathering information from inside the U.S.S.R. Information gathered by one monitoring system (say the electronic intercep-



tion of messages) can tell another system (such as a photoreconnaissance satellite) where to look and what to look for. Furthermore, pieces of information gathered by one system, inconclusive by themselves, can reveal the nature of an ongoing activity when they are interpolated and correlated with other pieces gathered by other monitoring systems.

Putting the pieces of the puzzle together is as important as gathering them. There is, however, a measure of ambiguity and uncertainty inherent in this task. Some detected activities might not be observed in enough detail to establish that they represent violations of the provisions of some treaty. Other activities might be observed in great detail, but the relevant treaty provisions may be too ambiguous to determine whether or not those activities are violations. The efforts of the technical community have been devoted to reducing these uncertainties by developing surveillance technologies that produce highly detailed images and records of intercepted signals that are as free of noise as possible.

Ongoing activities and events that may produce visible permanent changes in the physical environment are most reliably detected by means of satellite photoreconnaissance, that is, the periodic photographing of scenes inside the U.S.S.R. by optical systems carried on a satellite in a comparatively low orbit. By photographing a scene repeatedly under the same conditions, one can see whether there have been changes in it. Given enough detail, it is possible to recognize the activity that caused the change. A special kind of camera collects light reflected by the scene and forms an image that is a pattern of light and dark (or colored) dots on a photosensitive recording surface. Each dot is called a pixel, or picture element. The size of the pixels, the focal length of the camera and the height of the satellite determine the system's resolution, which is the size of the smallest object on the ground that the system can distinguish [*see illustration on this page*]. The finer the detail available in a picture, the easier it is to detect changes from one picture to the next, and the smaller are the changes that can be detected. The increase in resolution enhances the information content of the pictures. Enhanced information content augments the probability that the activity responsible for the observed changes can also be identified.

Experience has shown that the presence of an object in a scene can be detected if the object is at least as big as the size of the resolution patch. If it is eight times bigger, it can be recognized (as, say, a tank or a truck), and if it is 12 times bigger, it can be identified (as, say, an older Soviet T-62 tank or a newer T-72). The size of the resolution patch of U.S. reconnaissance satellites is said to be smaller than 10 centimeters. Accordingly one can surmise that such systems can detect an object on the ground of approximately the same length, and that they can completely identify an object less than 1.5 meters across. The identification of objects and activities becomes even easier if pictures formed by reflected radiation of different wavelengths are available. By combining images of a scene made in infrared and ultraviolet radiation as well as visible light, an analyst can deduce additional information about an object or can detect camouflaged ones.

Once an image of the ground scene is formed by the camera, the image must be recorded and transmitted to ground-based facilities for analysis and interpretation. The recording can be done either on photographic film or on a two-dimensional array of photosensitive electro-optical detectors called charge-coupled devices, or CCD's. A detector of this type transforms the amount of light it receives in a short, fixed period of time into a proportional amount of electric charge. Thus the pattern of light creates an electrical replica of itself on the array of detectors. This pattern is converted in turn into a sequence of numbers, which is then transmitted to a receiver on the ground. Equipment in the ground station transforms the electrical replica of the light pattern recorded by the array into a picture. The process is then repeated and a new image is recorded.

In systems with a large focal length electro-optical arrays can achieve pixel sizes comparable with the best photographic film. In addition they offer a

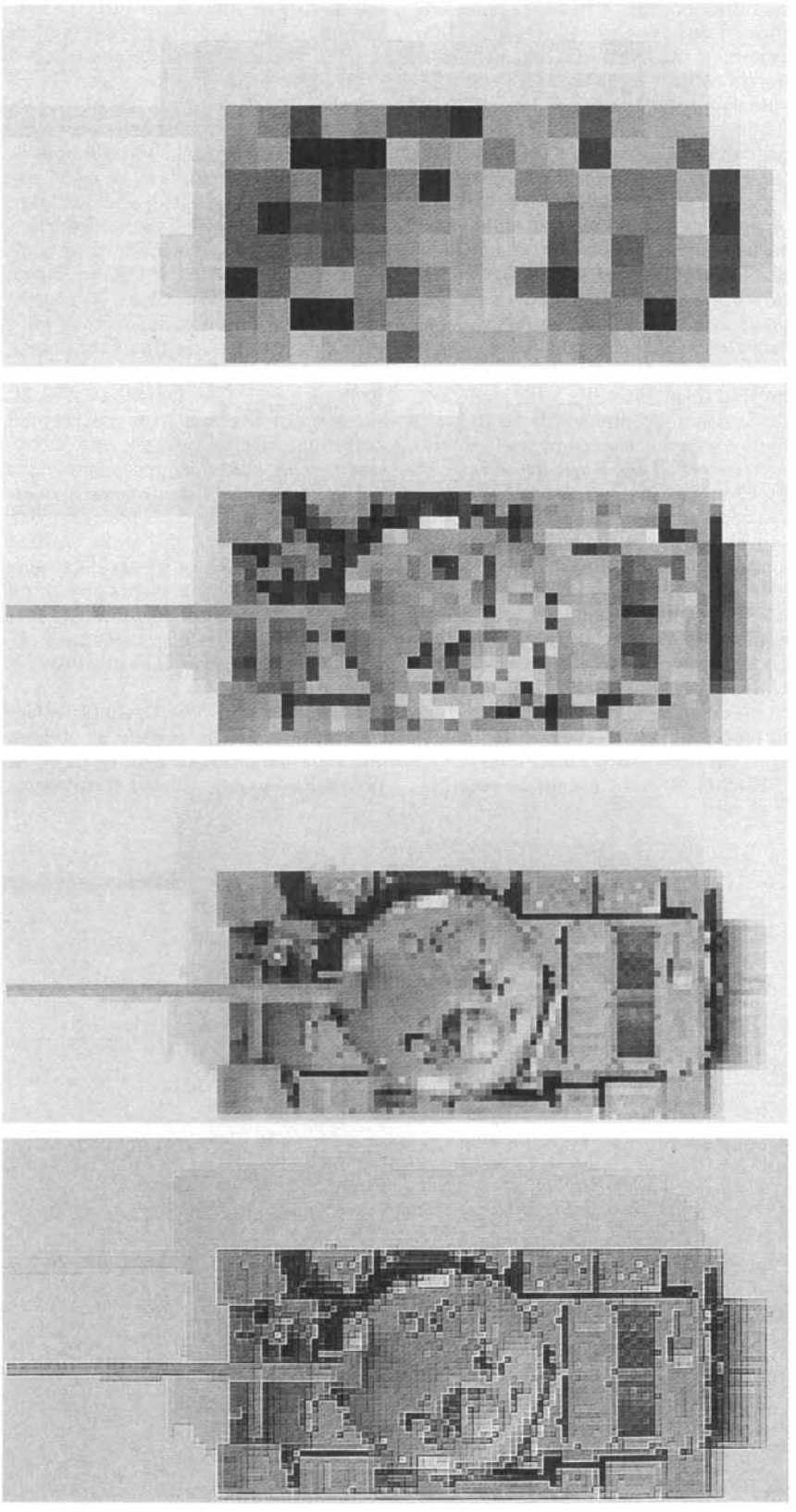
GROUND RESOLUTION of a photoreconnaissance satellite is defined with reference to the idealized diagram at left. The diameter of a single resolution patch of the ground scene (designated r) is given by the formula $r = (h/f)d$, in which h is the height of the satellite, f is the focal length of the optical system and d is the diameter of a pixel, or picture element, in the recording medium. (All units are generally computed in centimeters.) The size of a pixel is determined either by the graininess of the photographic film or by the size of an individual cell in the charge-coupled device (CCD) used at the focal plane of the optical system to record the image. The resolution patch of the current generation of U.S. photoreconnaissance satellites is said to be roughly 10 centimeters.

number of advantages over film. Images recorded on film can be transmitted in either of two ways. The film can be processed on board the satellite in a manner similar to the way a Polaroid camera develops its film; in that case the image is subsequently transmitted by a system similar to a television camera. Alternatively the film can be ejected from the satellite in special capsules, which reenter the atmosphere and, with the aid of a parachute, slowly descend until they are snared in midair by specially equipped aircraft. Both of these methods cause delays in the reception of the pictures and limit the useful life of the satellite, since the cameras eventually run out of film. The photosensitive arrays installed on U.S. satellites can, with the aid of a relay satellite, transmit images of the scenes they fly over in real time, as they are being observed. The pictures recreated on the ground have a much larger dynamic range than those obtained by means of photographic film; as a result they are not as sensitive to large changes in the intensity of illumination.

After the image of a scene has been received by a ground station it is possible to enhance its optical quality with the aid of fast computers and specialized processing devices. The technique is known as digital image processing. For example, the blurring of lines and shapes on a picture caused by turbulence, uncompensated motion, overexposure or poor contrast can be removed by artificially altering the gray areas of a picture and exaggerating the dark and light ones. The image processing enables one to identify characteristic patterns (such as the appearance of a missile silo observed from above) in pictures of the ground, or to complete a scene somewhat obscured by clouds.

The information-gathering capabilities of photoreconnaissance satellites are limited by the fact that a visible-light system cannot function at night, and by the fact that visible, ultraviolet and infrared radiation cannot penetrate through cloud cover. Thus bad weather and the polar night may prevent U.S. photoreconnaissance satellites from observing large parts of the U.S.S.R. for prolonged periods. This problem has been dealt with by turning to another part of the electromagnetic spectrum: radio waves.

Because the waves emitted by a radar system have a longer wavelength than visible light waves do, they remain largely unaffected by cloud cover and rain. Hence they can "illuminate" their target regardless of the weather. Since radars generate their own reflect-



EFFECT OF INCREASING RESOLUTION is simulated in this series of images of a plastic model of a Soviet tank. A photograph of the model was optically scanned to produce a digital record on magnetic tape. The information on the tape was then processed in such a way as to reproduce the image at three different pixel sizes, corresponding roughly to a resolution at which the object can be detected as a trucklike shape (*top*), a resolution at which the object can be recognized as a tank (*second from top*) and a resolution at which the tank can be identified as a Soviet T-62 (*second from bottom*). The enhanced image at the bottom was subjected to special processing steps to sharpen the edges and adjust the contrast.

ed radiation, they can function at any time of day. The resolution of a radar system is proportional to the wavelength of the radiation it emits and inversely proportional to the length of its antenna; thus radar images of the ground normally have a very low resolution. The resolution can be increased by making the satellite's fairly small antenna appear to the reflected radar waves as if it were a very long antenna. This is accomplished with a method known as synthetic-aperture radar (SAR). A SAR satellite combines elaborate electronic processing of waves reflected by the target with the relative motion of the satellite with respect to the surface of the earth to achieve high-resolution radar images [see "Radar Images of the Earth from Space," by Charles Elachi; SCIENTIFIC AMERICAN, December, 1982]. The resolution of a SAR image is not as high as that of a comparable picture made with visible light; there is a practical limit to how large the "virtual" antenna can be made to appear to radio waves. It is good enough, however, to enable the U.S. to continue monitoring many activities in areas of the U.S.S.R. that are obscured by cloud cover or are in daylight for only a few hours a day.

Radars do not have to be satellite-based to provide useful information. Ground-, ship- and plane-based radars are the primary means by which the U.S. obtains information about Soviet missiles while they are being tested. This coverage is a good example of the synergy created by the variety of U.S. monitoring capabilities. The test firing of Soviet ballistic missiles, which takes place at the Plesetsk and Tyuratam missile-launching facilities in central Asia, is preceded by many activities. For example, early warning of a forthcoming test could come from a photoreconnaissance satellite that photographed the missile being prepared on its launch pad. The signal to turn on all sensors could come from intercepted communications or from the early-warning satellites in geosynchronous orbits above the Equator; such satellites "stare" at the U.S.S.R. and can detect the infrared radiation emitted by the fiery plume of a rocket exhaust as soon as the missile clears any cloud cover over the launching site. At that point the detection of radio waves becomes the primary means of gathering information.

Large, ground-based radars follow the motion of the missile as it goes through the various stages of its boost phase. Radar can measure continuously and with great precision the velocity of the missile and hence its acceleration by measuring the Doppler shift of the reflected wave. When the launching site of a missile is too deep inside Soviet territory to be observed by conventional ground-based radars, the U.S. utilizes "over the horizon" radars. These devices project beams that the ionosphere, acting like a mirror, reflects into the interior of the U.S.S.R. A radar of this type also provides useful information about the velocity of a missile.

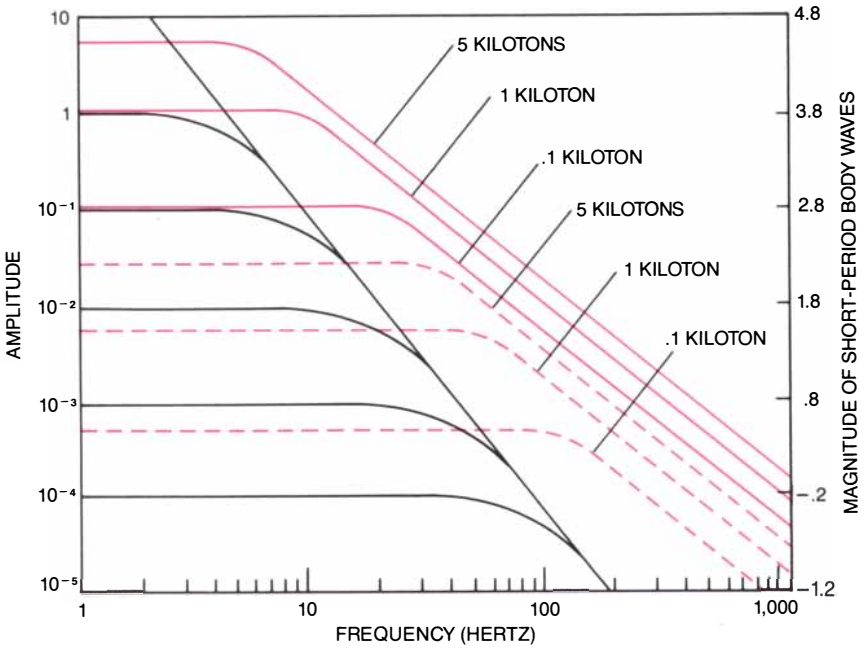
Radio receivers both on the ground and on satellites in high orbit intercept the stream of messages the missile sends back to the ground during testing that describe the performance and conditions of the various parts of the missile. The messages, known collectively as telemetry, include information on how much fuel is being burned, what commands the guidance system is sending to the rocket, and what temperature and pressure different parts of the missile are experiencing. All these signals, which are needed by the missile's engineers to check whether it is functioning as designed, are also intercepted by U.S. monitoring facilities. By correlating such te-

RADAR IMAGE of a large eddy in the Gulf of Mexico reveals a ship's wake (*bottom*). Variations in surface roughness associated with the eddy and the wake were recorded by the synthetic-aperture radar on board the *Seasat* satellite. The processing of the image was done by workers at the Jet Propulsion Laboratory of the California Institute of Technology. The main advantage of satellite-borne radar systems for military surveillance is that they are able to function at any time of the day or night and in any kind of weather.

lemetry data with information about the motion of the missile gathered by ground-based radars, U.S. intelligence officers know not only that the U.S.S.R. has tested a missile but also what many of its performance characteristics are. Some telemetry signals from such tests have been encoded by the U.S.S.R. in order to make them incomprehensible to U.S. analysts, a practice that has been protested by the U.S. on the grounds that the unratified SALT II treaty stipulates the only signals that can be encrypted are those not directly relevant to the verification of the treaty [see "Science and the Citizen," February]. Even without access to unencrypted telemetry, however, the U.S. can detect the Soviet missile test itself with high confidence and can also determine such important performance characteristics of a ballistic missile as the number of reentry vehicles it is carrying.

The final phase of a ballistic-missile test is the return of its reentry vehicles to the atmosphere and their impact on the ground. Since atmospheric forces affect the flight paths of the reentry vehicles, information about their performance in this terminal phase provides an important measure of the missile's accuracy and is crucial for the U.S.S.R. to have. Consequently, during Soviet missile tests, the reentry vehicles are aimed at instrumented impact areas either on the Kamchatka Peninsula or in the central Pacific Ocean. The U.S. has developed a range of sensing systems that monitor the return of the Soviet reentry vehicles to these areas. Very precise phased-array radars (which can detect an object the size of a basketball from thousands of kilometers away and can simultaneously track hundreds of such objects), radio receivers, infrared and visible-light telescopes equipped with cameras and rapid-scan spectrometers record in great detail the various forms of radiant energy emitted and reflected by the reentry vehicle. Some of these instruments are based on Shemya Island at the tip of the Aleutian Islands chain and at Kwajalein atoll in the Pacific; others are carried by ships and planes and even by small sounding rockets launched in anticipation of the arrival of the reentry vehicle.

From all these data, correlated and interpreted by intelligence analysts, the U.S. can estimate with great confidence many of the characteristics of the tested Soviet weapon. (For example, how big is its booster? How long is its range? How heavy are the reentry vehicles?) Other characteristics are much more difficult to determine with confidence. Accuracy, for instance, can only be determined statistically



ABILITY TO DISCRIMINATE underground nuclear explosions from earthquakes now extends down to explosive yields on the order of one kiloton or even less, as shown in this graph, based on the work of Jack F. Evernden of the U.S. Geological Survey. The black curves represent earthquakes. The solid color curves correspond to nuclear explosions in soft rock and the broken color curves to "decoupled" nuclear explosions in mined cavities.

from the flight path and impact point of the reentry vehicles. If the U.S. does not know at exactly what point the Soviets were aiming their reentry vehicles, then considerable uncertainty is introduced into the U.S. estimation of their accuracy. Determining the reliability of Soviet ICBM's and reentry vehicles is also statistically uncertain. Some characteristics are extremely difficult to establish. The potential yield of the weapon that will be placed on the reentry vehicle can be only crudely approximated from the vehicle's weight. Nevertheless, the unilateral technical means of monitoring the testing of Soviet ballistic missiles available to the U.S. can be relied on with confidence not only to detect such tests but also to provide the U.S. with a great deal of information about the performance of the tested weapon.

The nuclear warheads that missiles carry must be developed by means of underground testing. This state of affairs is the consequence of the Limited Test Ban Treaty, signed by the U.S., the U.K. and the U.S.S.R. in 1963, which banned such tests in the atmosphere. The detection technology needed to monitor underground tests consists of seismometers that detect the acoustic waves generated by an explosion, recording equipment and computers that analyze the data. Atmospheric testing, if resumed, would be easily monitored by special satellites and by an array of other techniques, including many discussed above.

Underground explosions generate elastic waves that propagate both on the surface and through the crust of the earth over very long distances. The magnitude and point of origin of such waves can be determined with the help of sensitive seismometer arrays. The seismometers are permanently "on," and so they detect and record all earth tremors, including those caused by an explosion.

A persistent problem in the seismic detection of underground nuclear test explosions has been the fact that such explosions yield elastic waves roughly similar to those of many other events, such as naturally occurring earthquakes; the waves are recorded the same way by the arrays of seismometers. Thus the risk existed that a nuclear explosion could be mistaken for an earthquake, and vice versa. Subsequent research effectively eliminated this difficulty for all but the smallest nuclear explosions [see "The Verification of a Comprehensive Nuclear Test Ban," by Lynn R. Sykes and Jack F. Evernden; SCIENTIFIC AMERICAN, October, 1982]. Yet two conceivable sources of evasion remained. The U.S.S.R. might arrange to test a small nuclear explosive during an earthquake. The seismic waves from the earthquake might obscure the signal generated by the detonation. It also seemed possible for a time that the en-

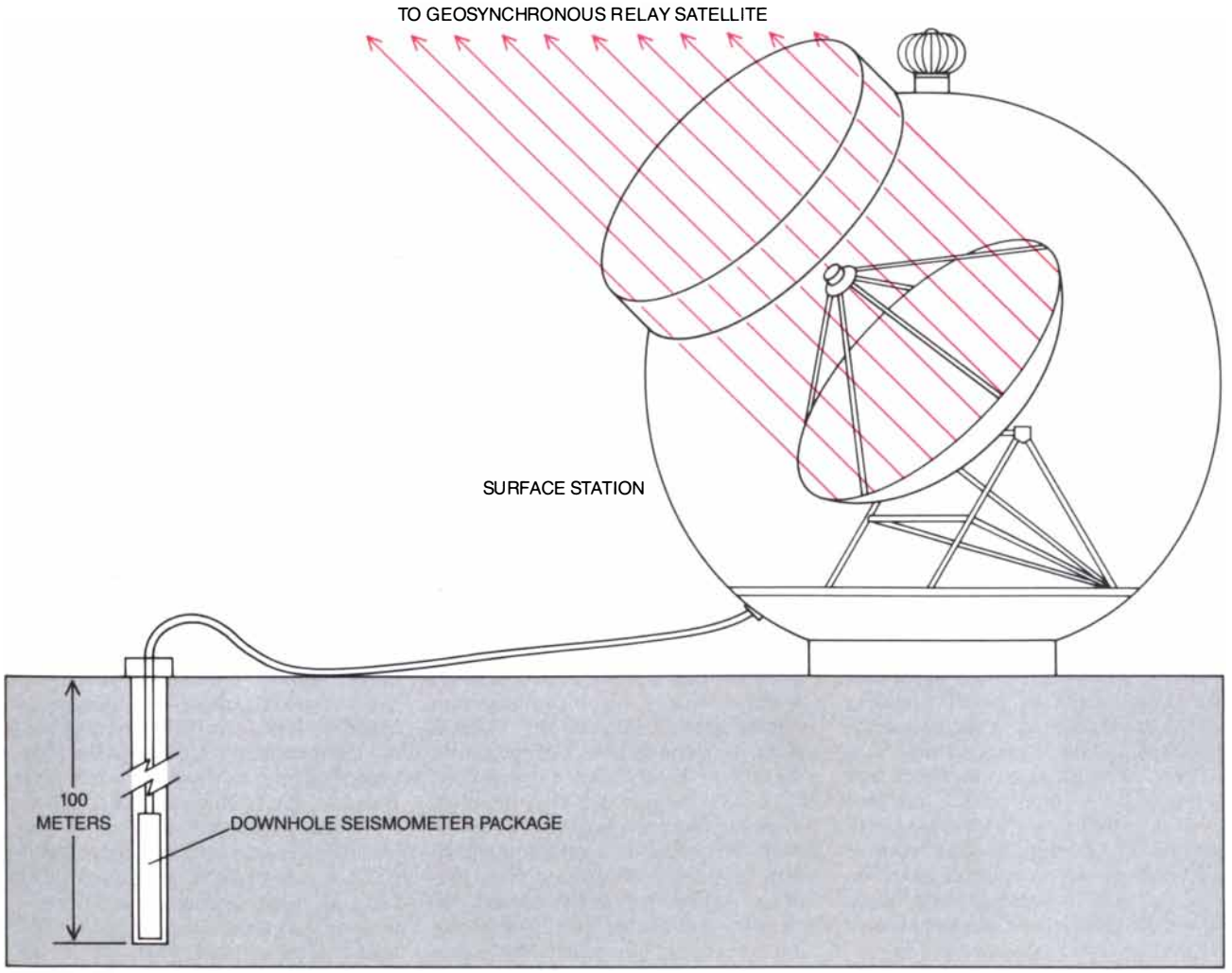
ergy from a low-yield device could be effectively muffled by detonating the weapon in a large underground cavity, making the blast undetectable outside the U.S.S.R. It has recently been reported, however, that a team of geophysicists at the U.S. Geological Survey's office in Menlo Park, Calif., has developed a means of unambiguously detecting a nuclear explosion as small as one kiloton or even smaller, given improved monitoring equipment.

The new detection technique takes advantage of the fact that an earthquake is a widespread event; as a consequence the acoustic energy it emits is concentrated at long wavelengths. A nuclear explosion, on the other hand, is a pointlike event that releases a huge amount of energy abruptly into the environment. Both the limited extent and the suddenness of such an event force it to release its energy in much shorter waves, which have higher frequencies. As a result the waves from a nuclear explosion can be detected and recognized by their frequency characteristics even if they are accompanied by those of an earthquake.

For example, waves from an earthquake can obscure the signal from a one-kiloton decoupled explosion up to frequencies of 10 hertz or so [*see illustration on preceding page*]. The earthquake waves, however, contain no detectable energy in frequencies above about 30 hertz, whereas waves from an explosion contain detectable energy up to frequencies of several hundred hertz. Thus seismometers tuned to detect only high-frequency waves will not "hear" the earthquakes; the seismometers will not be confused by the constant noise of the earth but will easily detect the explosion-generated waves.

Because high-frequency acoustic waves do not travel very far through the earth's crust, their detection would require placing seismometers inside the U.S.S.R. This potentially delicate political problem has been solved by the development of an unmanned seismic station. Workers at the Sandia National Laboratories have devised, tested and are now operating five unmanned seismic stations in the U.S. and Canada for practice and for demonstration purposes [*see illustration below*]. The stations communicate constantly by satellite relay with a central control facility in the U.S., making them virtually tamperproof. The U.S.S.R. accepted a proposal for the emplacement of such stations on its soil during negotiations for a comprehensive test-ban treaty, which were interrupted by the U.S. in 1980 after the Soviet invasion of Afghanistan. Consequently there appears to be no technical obstacle to monitoring with great confidence underground nuclear tests releasing as little as one kiloton or less of energy.

The three types of national technical means of monitoring Soviet activities discussed above—those that create



UNMANNED SEISMIC STATION developed by workers at the Sandia National Laboratories could be helpful in verifying compliance with a comprehensive nuclear-test-ban treaty. A system of five such stations is currently being operated in the U.S. and Canada.

images, those that detect electromagnetic waves emitted or reflected by missiles during tests and those that listen for acoustic waves—exemplify the sophistication of U.S. monitoring capabilities. Indeed, the level of detail the U.S. is able to detect suggests the kinds of treaty provision the U.S. can and cannot verify. The point is significant: the guiding principle for agreeing to any treaty provision must be that the U.S. is able to have confidence it can detect violations that could threaten its national security. Whether the U.S. can detect such violations depends both on what it considers threatening to its national security and on the physical properties of the weapon system or the military activity limited by the treaty.

For example, given unmanned U.S. seismic stations inside the U.S.S.R., what kind of treaty provisions limiting underground nuclear tests can U.S. seismic detectors verify compliance with? We have already shown that with such stations the U.S. can detect underground nuclear tests with an energy yield as small as a kiloton. Does this capability enable the U.S. to enter a treaty banning all underground nuclear tests confident that any Soviet violations that could harm U.S. security would be detected? The U.S.S.R. could attempt to test significantly below the U.S. detection level, conducting a clandestine test of, say, 100 tons. The Soviets have already conducted hundreds of much larger tests, however; what is more, a 100-ton test would be 1,000 times smaller than their smallest strategic nuclear weapon and 250,000 times smaller than their largest. They would not be likely to derive any useful information from such a low-level test, even if it were a test of a new type of weapon. Therefore it is improbable that a test significantly below the U.S. detection level of one kiloton could threaten U.S. security. Conversely, any test that might give them useful information would certainly be detected. Hence the comparison between what the U.S. must confidently detect and what the U.S. can confidently detect in the case of underground nuclear tests leads to the conclusion that the U.S. could verify a comprehensive test ban well enough to maintain national security.

Similar assessments can be made with regard to the ability of the U.S. to monitor Soviet missile tests. For example, is this ability adequate to monitor an agreement limiting improvements in accuracy in ICBM's? Unilaterally determining missile accuracy by observing ballistic-missile tests is statistically uncertain at best, and so accuracy improvements achieved in defiance of a limitation would be difficult to monitor. On the other hand, if the U.S. decides it is important to constrain Soviet accuracy, there is another option: banning all ballistic-missile tests. Verifying that a ballistic-missile test has taken place can be done with high confidence. Indeed, the probability that the U.S. would detect a single Soviet flight test is certainly higher than 90 percent, given the variety of means the U.S. has to detect these tests. At least 20 test flights are required to determine with confidence the accuracy of a new missile. Therefore, even if the U.S. can be confident that it can detect the test of a Soviet ballistic missile only 90 percent of the time, its chances of not detecting one of 20 Soviet tests is only one in 100 billion billion (10^{20}). In short, a treaty banning ballistic-missile testing altogether could be confidently verified. In addition, since the U.S. can determine the number of reentry vehicles a ballistic missile is capable of delivering, another verifiable treaty would be one that would restrict flight testing to missiles designed to carry a single warhead.

When it comes to monitoring the production and deployment of Soviet strategic weapons there is little doubt that U.S. photoreconnaissance satellites can confidently detect and count such large strategic-weapon delivery systems as ICBM's, submarine-launched ballistic missiles (SLBM's), submarines and bombers. The key question is: With what accuracy must such weapon systems be monitored in order to ensure against erosion of U.S. national security by violations of treaties limiting or forbidding their production or deployment?

Keeping activities that violate a treaty banning an entire weapon system secret is particularly difficult. The large-scale activities involved in developing, testing, producing and deploying significant numbers of any large weapon are easily observed. Indeed, any effort to conceal such activities would be handicapped by the fact that the Soviets do not know exactly what the U.S. can detect. Therefore they do not know what to attempt to hide, and they do not know with what thoroughness they must try to conceal the details of visible changes that the proscribed production of a weapon would cause.

Confidence in verifying numerical limits on a weapon system would depend on the number of weapons permitted. In general it is hard to have confidence in the verification of agreements that involve small numbers of weapons. A complete ban of a missile system would be easier to verify than an agreement that allows, say, 100 such missiles on each side. If the Soviets were supposed to have none, as soon as the U.S. monitoring systems detected one it would become apparent that the Soviets had violated the agreement. If the agreement were to allow 100, however, it would be quite difficult to know whether they had 100 or 120. Arms-control agreements that completely forbid a weapon system, a practice or an activity are much easier to monitor and verify for compliance than agreements that allow small numbers of weapons.

The U.S.S.R., however, has dozens of submarines and bombers, some 1,000 SLBM's and about 1,400 ICBM's. The U.S. is often said to be able to verify these numbers to within 10 percent or better (a credible claim in the light of the previous discussion of U.S. verification capabilities). This estimate implies that the U.S. can verify compliance with numerical limits accurate to the order of a few submarines and bombers and perhaps 100 SLBM's or ICBM's. Given the large numbers of weapons both nations now have, such verification accuracy seems more than adequate to maintain U.S. security under a treaty that limited the production of strategic weapons.

One special case is the problem presented by cruise missiles: small pilotless drones several meters long. The main complicating factor here is that the same type of cruise missile can be fitted to carry either a nuclear warhead or a conventional one. Hence differentiating cruise missiles carrying nuclear warheads from those carrying conventional warheads is an extremely difficult task. It is likely that any verifiable treaty dealing with cruise missiles must cover the total number of cruise missiles allowed, without distinguishing between nuclear and conventional warheads.

We have discussed a variety of powerful remote-sensing technologies but only some of their verification applications. (For instance, we have not described how these technologies could serve to verify a treaty dealing with antisatellite weapons.) Moreover, the actual intelligence capabilities of the U.S. are much greater than we have described here. Not only are there too many verification technologies to discuss in such an article, but also (and more significant) the nature of the intelligence-gathering process is such that many of the methods and sources by which the U.S. gains knowledge of Soviet actions are classified. Even our necessarily limited discussion, however, suggests that the rich remote-sensing capability of the U.S. can adequately verify a wide range of treaties.

The Volcanoes and Clouds of Venus

Radar maps of Venus and chemical analysis of its atmosphere and crust imply the existence of active volcanoes. The sulfur gases they release form a global cover of sulfuric acid clouds

by Ronald G. Prinn

To a planetary scientist one of the earth's most intriguing features is its geologic activity. Heat flowing from the interior of the planet drives processes that constantly reshape the surface. A hallmark of this activity is volcanism: the emergence of hot rock and gases through cracks in the crust. Volcanic eruptions have also been observed on Jupiter's moon Io, but on no other body in the solar system. Yet Venus, the earth's nearest planetary neighbor, is similar to the earth in many ways. It is about the same size, has about the same mass and was formed in the same region of the condensing solar nebula. The two planets might therefore be expected to have undergone a similar evolution. Are there active volcanoes on Venus?

For many years answers to that question could be no more than speculative, because a thick, permanent layer of clouds thwarted efforts to study the Venusian surface. In the past half decade the situation has changed. Although the clouds block visible light, they are transparent to radio waves and microwaves, and the surface has been mapped comprehensively by a radar instrument on board *Pioneer Venus,* which has been orbiting the planet since 1978. The maps, along with more recent high-resolution radar images from two Russian spacecraft and from radio telescopes on the earth, have revealed volcanolike structures.

Furthermore, the very permanence of the opaque clouds is strong, albeit indirect, evidence of active volcanism. Occupying the altitudes between 50 and 70 kilometers, the clouds are composed of concentrated sulfuric acid and an ultraviolet-absorbing material that is probably elemental sulfur. In recent years several entry probes have survived the corrosive attack of these substances and the intense heat at the surface (460 degrees Celsius) long enough to measure the composition of the planet's atmosphere and crust. As a result it is now possible to decipher the interaction of the two and to understand the complex cycle of photochemical and thermochemical reactions that transform sulfur gases into cloud particles. The data imply that sulfur gases are continually injected into the atmosphere, by a mechanism that could only be volcanic. Indeed, the detection by the *Pioneer Venus* orbiter of dramatic changes in the abundance of sulfur dioxide in the atmosphere above the cloud tops suggests that Venus has been shaken by massive eruptions within the past decade.

The prima facie evidence for the existence of volcanoes on Venus, whether active or extinct, comes from the radar studies. Beginning in the 1960's radio telescopes at Arecibo in Puerto Rico and at Goldstone in California were aimed at Venus. They provided the first reliable estimate of its radius (6,052 kilometers, compared with the earth's 6,378 kilometers) and spin period (243 days), as well as the first images of major surface features such as the massive northern continent Ishtar Terra. In 1977 R. Stephen Saunders and Michael C. Malin of the Jet Propulsion Laboratory suggested that Theia Mons, in the region known as Beta Regio, is a large shield volcano: a roughly symmetrical construct formed by chronic, nonexplosive eruptions of hot lava that flows over large distances before solidifying. At 700 kilometers in diameter Theia Mons is much larger than the Hawaiian shield volcanoes but smaller than Olympus Mons on Mars, which appears to be a giant example of this kind of mountain.

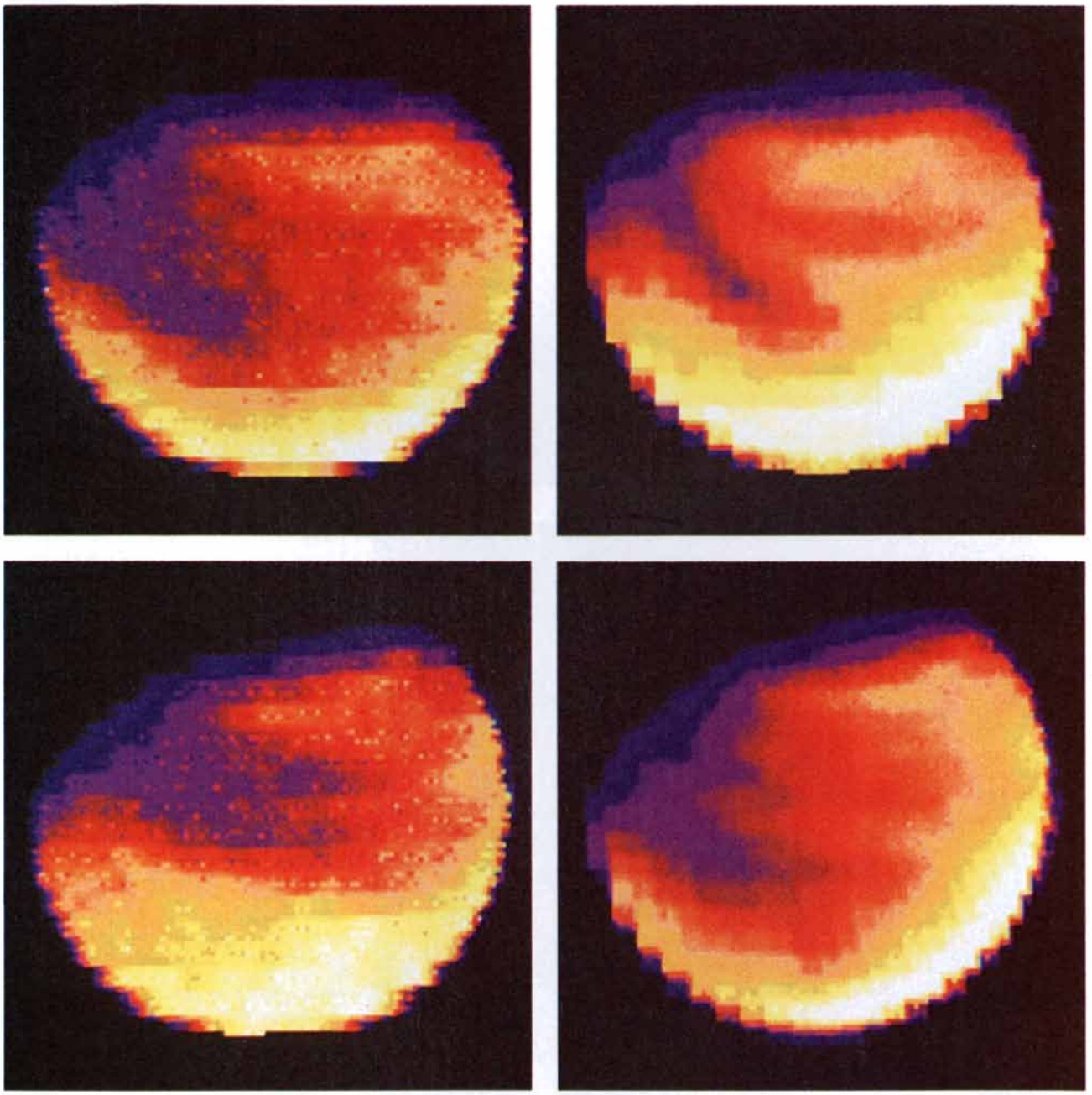
When *Pioneer Venus* entered Venusian orbit in December, 1978, it became possible to generate radar images of higher spatial resolution and also to map the surface relief of the planet. Based on these data Harold Masursky of the U.S. Geological Survey, Gordon H. Pettengill of the Massachusetts Institute of Technology and their colleagues argued in 1980 that the whole of Beta Regio, including the two mountains Theia Mons and Rhea Mons (both more than 4.5 kilometers high), is an enormous volcanic structure formed by the slow piling up of lava. Later George E. McGill and his colleagues at the University of Massachusetts at Amherst suggested the volcanoes were actually more modest lava structures on top of an elongated dome of uplifted crust.

The latter theory implies regional rather than localized upwelling of magma under Beta Regio. It is supported by the observation, from perturbations in the orbit of *Pioneer Venus,* that gravity in the region is stronger than average for Venus. The upwelling may even be strong enough to drive some horizontal motion of the crust. In 1983 Donald B. Campbell and his coworkers at the Arecibo Observatory made an image of Beta Regio with a resolution about 10 times greater than *Pioneer Venus* can achieve; the image shows a major linear rift, flanked by volcanic structures, cutting through the region. In any case it is generally agreed that volcanic activity of some kind has occurred at Beta Regio. The brightness of the radar images suggests the activity may have been geologically recent: brightness is an indication of a rough surface, a surface that is relatively young and unweathered.

Beta Regio is not the only region where volcanolike structures have been observed. Isolated peaks and pinnacles detected by *Pioneer Venus* in Atla Regio, west of Beta near the equator, may be individual volcanoes, according to James W. Head of Brown University and other workers. The Russian spacecraft *Venera 15* and *Venera 16,* which have been orbiting Venus since October, 1983, have transmit-

ted images of previously unobserved great circular features. As much as several hundred kilometers in diameter and relatively low in elevation, they mark the surface of Ishtar Terra and other areas. A. T. Basilevsky, V. L. Barsukov and their colleagues at the V. I. Vernadsky Institute of Geochemistry and Analytical Chemistry in Moscow have interpreted these features as huge volcanic domes that have collapsed, leaving folds of crust around their periphery. (A nonvolcanic explanation of the circular features would be that they are simply ancient, weathered meteorite craters.)

The radars on board the Venera orbiters yield images with a horizontal resolution of from one kilometer to two kilometers, whereas the resolution of the *Pioneer Venus* instrument is at best 30 kilometers. Unfortunately the Russian missions were planned to image only about a third of the planet, and their coverage has excluded most of Beta Regio. The next major improvement in knowledge of surface structures is likely to come from the *Venus Radar Mapper,* which the National Aeronautics and Space Administration

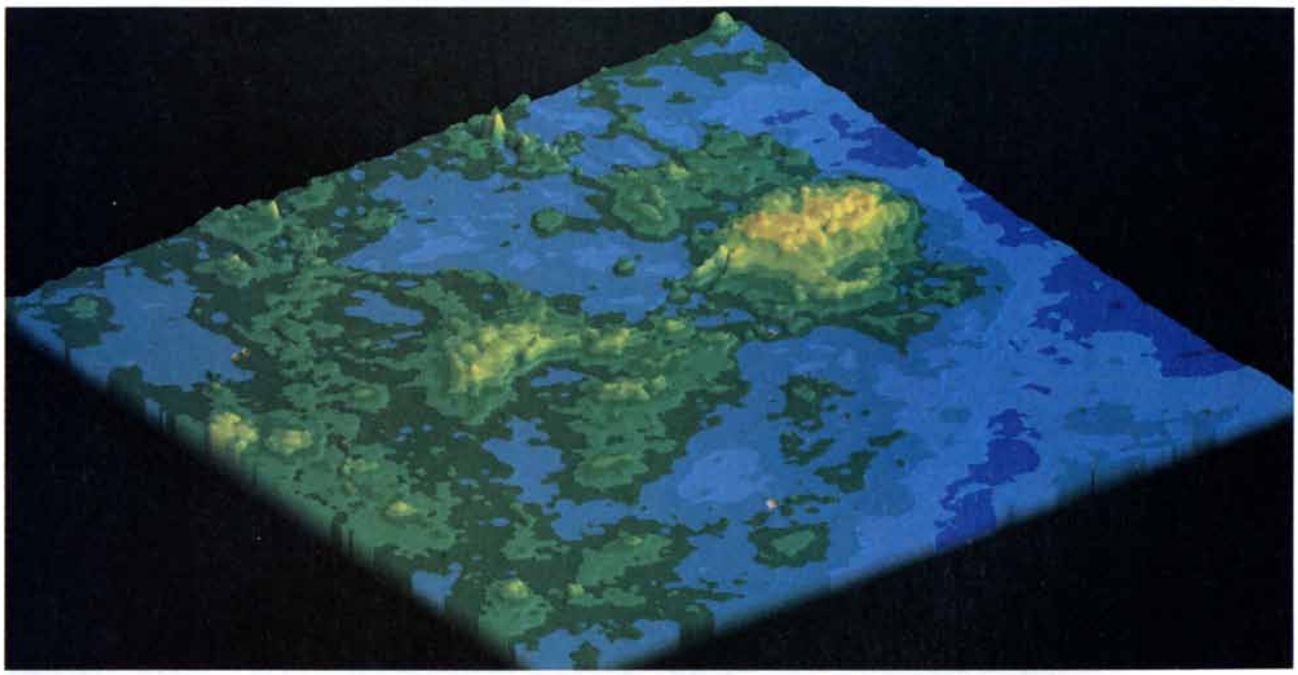


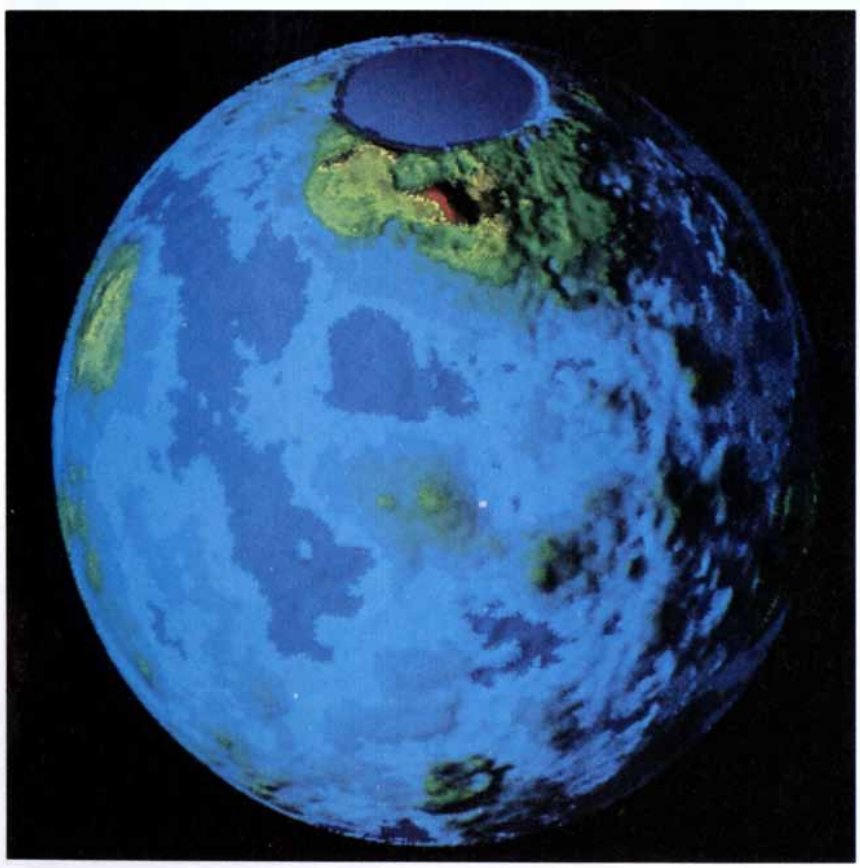
SULFUR DIOXIDE shows up as dark blue streaks in images of the Venus cloud tops based on measurements by the *Pioneer Venus* ultraviolet spectrometer. The images represent radiation at a wavelength of 207 nanometers, which sulfur dioxide absorbs strongly but which the clouds reflect. The data were gathered over a five-day period in August, 1984, as the planet's atmosphere, which rotates much faster than the surface, completed one rotation under the spacecraft. The first (*upper left*) and last (*bottom right*) images show the same part of the atmosphere. From measurements such as these Larry W. Esposito of the University of Colorado at Boulder has determined that the average abundance of sulfur dioxide above the clouds has declined by more than 90 percent since the spacecraft entered Venusian orbit in late 1978. At that time the abundance was unexpectedly high. The pattern suggests a massive volcanic eruption injected sulfur dioxide into the upper atmosphere before the orbiter arrived; the gas has since been forming sulfuric acid.

has scheduled for launching in 1988 and which will cover the entire planet at a resolution of .2 kilometer.

Are any of the volcanic structures discovered so far active? The question cannot be answered from the radar images alone. To answer it without being able to observe eruptions directly requires knowledge of the composition of the Venusian crust and atmosphere.

The first tantalizing bit of information on the chemistry of the crust came in 1975 from the *Venera 8, 9* and *10* landers. Equipped with gamma-ray detectors, these probes measured the abundances of radioactive potassium, uranium and thorium in the crust. Although the variability among landing sites was significant, the radioactive elements seem to be present on Venus in amounts comparable to the levels





TOPOGRAPHY OF VENUS has been mapped by a radar altimeter on board the *Pioneer Venus* orbiter. As the spacecraft orbits the planet the altimeter measures the distance to the surface; subtracting the distance to the planet's center of mass, which is known from precise tracking of the orbit, yields the radius of the planet at each point. On the maps the highest altitudes are yellow and red and the lowest are dark blue. Most of the Venusian surface is remarkably flat. Whereas 35 percent of the earth's surface is continental and 65 percent is ocean floor, Venus has only two continentlike features, and they account for less than 5 percent of the surface. One of these, Ishtar Terra, is visible near the north pole on the global map (*left*). (The pole itself is blank because the spacecraft's orbit does not take it over the poles.) The red area in Ishtar is Maxwell, the highest mountain on Venus, which reaches more than 11 kilometers above the average elevation and is thus comparable to Mount Everest. Maxwell may be volcanic in origin, but it is probably not active. The green-yellow area southwest of Ishtar, on the left edge of the globe, is Beta Regio, a likely region of current volcanic activity. In the enlarged perspective view (*above*) it is the broad highland with a cluster of pinnacles. The two highest peaks in Beta, Theia Mons and Rhea Mons, are more than 4.5 kilometers high. To the south of Beta (*at left in picture*) lies Phoebe Regio, which also has volcanolike structures. The enlarged view was made by the U.S. Geological Survey; the global image was made at the Jet Propulsion Laboratory.

found in terrestrial continental surface rocks. Since radioactive decay is the principal source of the earth's internal heat, the Venera data suggest that the amount of heat generated inside Venus is roughly comparable to that generated in the earth.

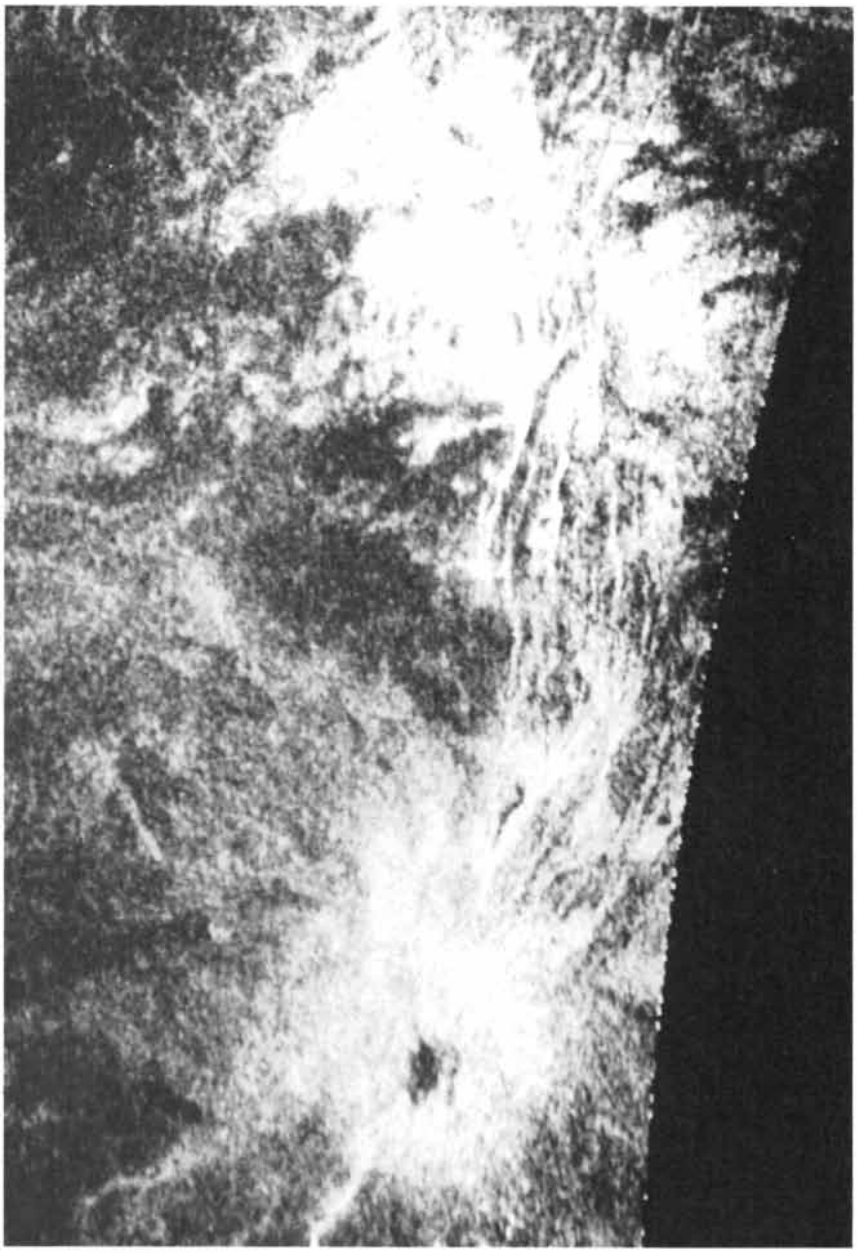
Somehow the heat of radioactive decay must escape from the planet's interior. The most efficient path is some form of volcanism, which in its general sense refers to any convective flow of hot material to the surface. On the earth well over half of the heat loss occurs at midocean ridges; there the plates that make up the lithosphere (the planet's rigid outer layer) diverge and lava wells up from the fluid asthenosphere to form new crust. A second form of volcanism is the island-arc volcanoes associated with oceanic trenches, where two plates collide and one is subducted into the asthenosphere. As Raymond E. Arvidson of Washington University and other workers have pointed out, however, the radar maps of Venus show no evidence of a global network of ridges and trenches. It may be, as Don L. Anderson of the California Institute of Technology has suggested, that the Venusian lithosphere is thinner than the earth's, in addition to being hotter and less dense, and that it therefore is too buoyant to be subducted into the fluid interior of the planet.

In the absence of clear evidence for plate tectonics, William M. Kaula and Lynn M. Muradian of the University of California at Los Angeles have concluded that two mechanisms of heat loss are possible on Venus: the eruption of isolated "hot spot" volcanoes like those of Hawaii, which are not associated with plate boundaries, and heat transfer by conduction at thin, "domed" areas of uplifted crust. The pinnacled, domed and rifted structure of Beta Regio suggests both types of heat loss have taken place there. In general it seems unlikely that conduction, a very slow process in rock, could be the sole mode of cooling. The need for an adequate cooling mechanism is a strong argument for active volcanism on Venus, but it is not conclusive.

Volcanism has another function: the discharge of gases from inside the planet. The discharge from terrestrial volcanoes includes carbon dioxide, nitrogen, water vapor, sulfur gases and the noble-gas isotopes argon 40 and helium 4, which are products of radioactive decay. All these gases have been detected in the Venusian atmosphere by the *Pioneer Venus* and *Venera 11, 12, 13* and *14* entry probes.

Indeed, on both Venus and the earth volcanic "outgassing" of the planet's interior early in its history is believed to have been the primary source of the atmosphere. The two atmospheres are vastly different, however: that of Venus is 96 percent carbon dioxide, whereas the earth's is 78 percent nitrogen. The difference arises in part from the presence on the earth of oceans, which have drawn most of the carbon dioxide out of the atmosphere and stored it as carbonates in the crust. If Venus ever had oceans, they have long since boiled off, and the hydrogen has escaped into space.

In comparing the two atmospheres the earth's crustal carbonates must therefore be included in its carbon dioxide inventory; the fact that Venus' atmosphere is 90 times as massive must also be taken into account. The total amount of carbon dioxide and ni-



RADAR IMAGE OF BETA REGIO suggests that Theia Mons and Rhea Mons are shield volcanoes flanking a linear rift in the crust. Made at 12.6-centimeter wavelength with the 300-meter radio telescope of the Arecibo Observatory, the image represents variations in surface roughness: bright areas are relatively rough and unweathered, dark areas are relatively smooth. Theia is the circular feature at the bottom of the image; its brightness is attributed to geologically recent lava flows. The linear features running north to Rhea are thought to be a rift, in part because they parallel a canyon discovered in topographical studies. The pattern suggests regional upwelling has lifted the crust in Beta, producing the rift and the volcanoes. The image has a horizontal resolution of approximately two kilometers.

trogen on Venus turns out to be about 30 percent less than on the earth. The concentration of argon 40 is roughly one-third of the terrestrial value, while the abundance in the crust of the argon source, radioactive potassium 40, is about the same on Venus as it is on the earth. The evidence suggests the two planets have experienced at least comparable amounts of outgassing and thus comparable levels of volcanism over the course of their history. Nevertheless, the presence of carbon dioxide, nitrogen and argon 40 does not show that outgassing has occurred on Venus recently; these gases are relatively stable and can survive in the atmosphere over geologic periods. For firmer evidence of recent volcanic activity one must turn to the clouds and to the sulfur gases that produce them.

Even on Venus sulfur is a minor atmospheric constituent—the gaseous progenitors of the clouds make up .02 percent and the cloud particles themselves a mere .00002 percent of the atmosphere—yet the clouds have a remarkable effect on the planet's climate. They reflect nearly four-fifths of the total incident sunlight, particularly at red and yellow wavelengths. As a result Venus absorbs much less solar energy than the earth in spite of being closer to the sun. Of the energy not reflected back into space, two-thirds is deposited in the clouds, which are absorptive at ultraviolet and near-infrared wavelengths; only one-third reaches the lower atmosphere and the surface. (Without the clouds the surface temperature would be much higher still.) On the earth the opposite is true: two-thirds of the solar input is absorbed at the surface.

The evidence that the dominant constituent of the clouds is concentrated sulfuric acid—75 percent by mass—is indirect but convincing. The visible and infrared reflection spectrum of the cloud particles is close to that of sulfuric acid. Studies of the polarization of sunlight reflected off the clouds indicate that the particles are spherical, which implies they are liquid droplets, and that they have a high refractive index of 1.44. These results exclude most candidates other than sulfuric acid. In particular they exclude water, which has a refractive index of 1.33 and which vaporizes at the temperatures prevalent in the lower region of the clouds. Furthermore, the concentration of both sulfur dioxide and water vapor in the clouds decreases rapidly with increasing altitude, suggesting the two substances are undergoing chemical reactions that ultimately produce sulfuric acid (H_2SO_4).

Sulfuric acid is a familiar component of the earth's atmosphere. It is present in dilute form in acid rain and also in concentrated form (as on Venus) in a very thin layer of stratospheric clouds. The acid is produced from sulfur dioxide as well as from hydrogen sulfide (H_2S), dimethyl sulfide ($(CH_3)_2S$) and carbonyl sulfide (OCS). The primary source of sulfur dioxide is the burning of fossil fuels, whereas the three reduced gases are mainly metabolic by-products of various sulfur bacteria. But sulfur dioxide and two of the reduced sulfur gases (hydrogen sulfide and carbonyl sulfide) are also common effluents from terrestrial volcanoes, along with hydrogen chloride and hydrogen fluoride.

All these volcanic effluents have been detected in the Venusian atmosphere. Since they are highly reactive and therefore short-lived, and since Venus seems to be lifeless, one might suppose their very presence to be evidence of current volcanic activity. In the late 1960's, however, John S. Lewis, now at the University of Arizona, suggested these gases are "cooked" out of surface rocks by the intense heat. The gases are added to the atmosphere, Lewis argued, at the same rate at which they are removed by reactions with the surface, so that there is no net flow of a compound into or out of the atmosphere. In other words, the sulfur gases, the hydrogen chloride and the hydrogen fluoride in the atmosphere are in chemical equilibrium with minerals in the crust.

According to Lewis' model, the mixing ratio, or atmospheric concentration, of molecular oxygen is determined by a reversible equilibrium reaction. In this process iron oxide in mineral form (FeO) and the calcium sulfate mineral anhydrite ($CaSO_4$) react with carbon dioxide to produce the calcium carbonate mineral calcite ($CaCO_3$), the iron sulfide mineral pyrite (FeS_2) and oxygen. The predicted oxygen mixing ratio is extremely small, and it determines the oxidation state of the sulfur gases. As a result the dominant sulfur gas would be the reduced gas carbonyl sulfide, at a concentration of about 600 parts per million. Hydrogen sulfide, also a reduced compound, would have a mixing ratio of about 130 p.p.m.; that of sulfur dioxide would be just 16 p.p.m.

Since Lewis advanced his equilibrium model new data, which I have been involved in interpreting, have shown the model's predictions to be wrong, at least as far as the sulfur gases are concerned. The mixing ratio of sulfur di-

REFLECTIVITY of the Venusian surface to 17-centimeter radio waves varies considerably with location, indicating that surface composition varies also. The map extends from 40 degrees north latitude to 20 degrees south and from 270 to 340 degrees east longitude. Red indicates the highest reflectivity, blue the lowest. The red area at the upper left is the central part of Beta Regio; Theia Mons has one of the highest reflectivities observed so far. A likely explanation is that its surface consists of volcanic rocks with substantial inclusions of pyrite, a highly conductive, sulfur-containing mineral. Unfortunately none of the Venera landing sites (*black dots*) were in regions of high reflectivity. The map was prepared by Peter G. Ford and Gordon H. Pettengill of the Massachusetts Institute of Technology.

oxide has been measured at 150 p.p.m., nearly 10 times its equilibrium value. The most recent measurements, made in 1982 by gas chromatographs on board the *Venera 13* and *Venera 14* probes, place the total abundance of carbonyl sulfide and hydrogen sulfide at about 150 p.p.m., much less than their equilibrium total of 730 p.p.m. The earlier *Venera 11, Venera 12* and *Pioneer Venus* probes found even lower levels of carbonyl sulfide and hydrogen sulfide and thus a greater relative importance for sulfur dioxide.

The equilibrium model cannot account for the high levels of sulfur dioxide. It is now clear that the abundances of sulfur gases in the Venusian atmosphere are not determined by reversible reactions with the crust that proceed with equal speed in both directions. Rather, sulfur is injected into the atmosphere by one reaction, undergoes a sequence of transformations to form sulfuric acid and is then removed from the atmosphere, eons later, by a different reaction with the crust.

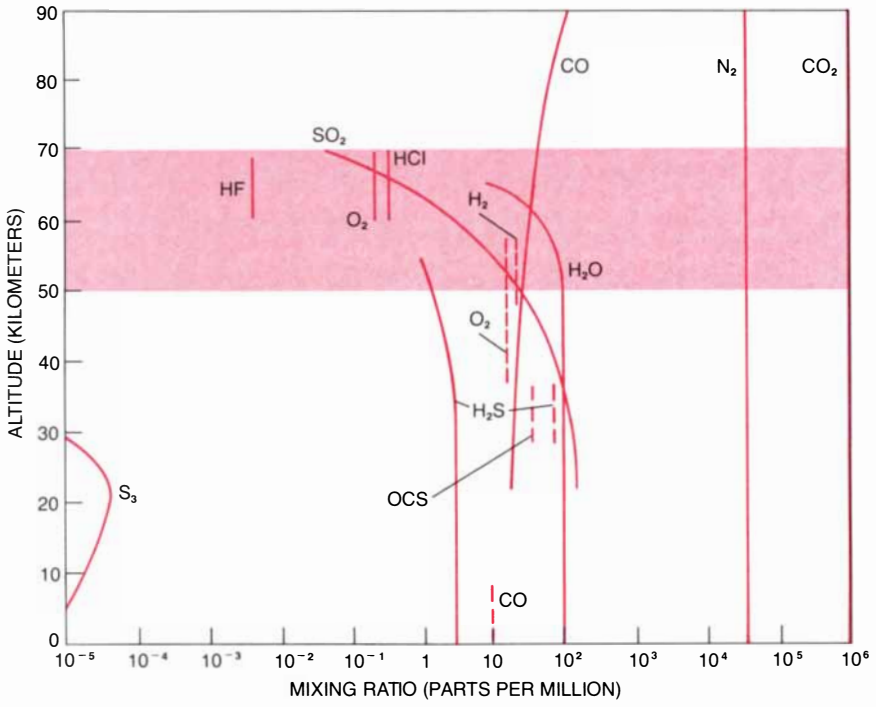
The sulfur cycle on Venus represents a balance between two opposing chemical regimes: a photochemical regime and a thermochemical one. Photochemical reactions increase the oxidation state of sulfur and produce the sulfuric acid clouds. The oxygen comes from the decomposition of carbon dioxide into carbon monoxide and oxygen by ultraviolet light in and above the clouds. Near the surface, high temperatures and densities drive thermochemical reactions that result in a net reduction, or deoxidation, of sulfur. These reactions regenerate the gaseous precursors of the clouds.

The cycle can be broken down into three parts: "fast" and "slow" atmospheric subcycles and a geologic subcycle. The fast subcycle begins in the middle and upper atmosphere (above the cloud base at 50 kilometers) with the oxidation, under the influence of ultraviolet light, of sulfur dioxide to sulfuric acid. The oxidation reactions are catalyzed by chlorine and hydrogen compounds derived from the photodissociation of hydrogen chloride. Sulfuric acid then sinks into the hot lower atmosphere below the clouds and evaporates. The resulting sulfur trioxide reacts thermochemically with carbon monoxide to regenerate carbon dioxide and sulfur dioxide. A sulfur molecule completes a loop through the fast subcycle in about a year.

In the slow subcycle sulfuric acid and elemental sulfur are formed in and above the cloud region through photochemical oxidation of carbonyl sulfide and hydrogen sulfide; in the lower atmosphere the two sulfide gases are oxi-

ELEMENT	VENERA 13	VENERA 14
MAGNESIUM (MgO)	11.4 ± 6.2	8.1 ± 3.3
ALUMINUM (Al_2O_3)	15.8 ± 3.0	17.9 ± 2.6
SILICON (SiO_2)	45.1 ± 3.0	48.7 ± 3.6
POTASSIUM (K_2O)	4.0 ± 0.6	0.2 ± 0.1
CALCIUM (CaO)	7.1 ± 1.0	10.3 ± 1.2
TITANIUM (TiO_2)	1.6 ± 0.5	1.3 ± 0.4
MANGANESE (MnO)	0.2 ± 0.1	0.2 ± 0.1
IRON (FeO)	9.3 ± 2.2	8.8 ± 1.8
SULFUR (SO_3)	1.6 ± 1.0	0.9 ± 0.8
OTHERS	3.9	3.6

SURFACE COMPOSITION of Venus was determined at two sites by X-ray-fluorescence instruments on board *Venera 13* and *Venera 14*. In calculating the relative importance of the elements (as a percentage of total weight) each element was assumed to be present as its oxide. The concentrations are similar to those found in certain volcanic basalts on the earth; an exception is sulfur, which is more abundant on Venus. Sulfur is much less abundant on Venus than calcium, indicating that most of the calcium is in oxide form in silicates or carbonates rather than in sulfates. Calcium oxide must be removing sulfur dioxide from the atmosphere; the two react spontaneously when the sulfur dioxide level is above equilibrium.



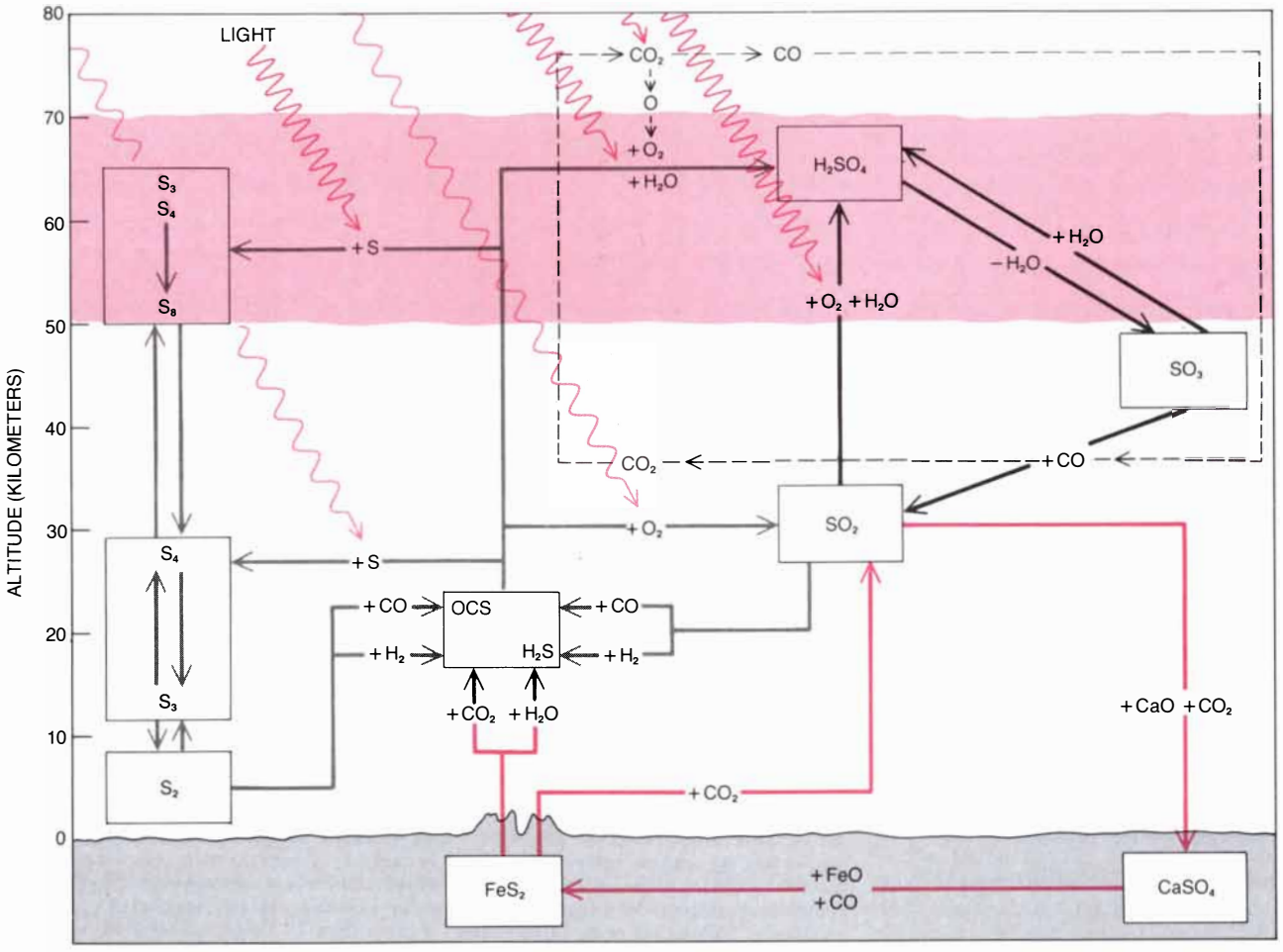
ATMOSPHERIC COMPOSITION of Venus has been analyzed with earth-based telescopes and by various entry probes, most recently by *Venera 13* and *Venera 14* (*broken lines*). The mixing ratios, or concentrations, of some important compounds are shown as a function of altitude on a logarithmic scale; for example, carbon dioxide (CO_2), which constitutes 96 percent of the atmosphere, is more than 10 times as abundant as nitrogen (N_2) and roughly 10,000 times as abundant as sulfur dioxide (SO_2), whose mixing ratio is about 150 parts per million. The termination of lines indicates the absence of measurements beyond a particular altitude range and not the absence of the compounds themselves. The recent Venera missions found higher levels of hydrogen sulfide (H_2S) and of carbonyl sulfide (OCS) than had been detected previously. The values for hydrogen (H_2) and oxygen (O_2) are upper limits; these compounds have not actually been detected. Sulfur dioxide and water vapor show parallel declines in abundance in the cloud region, indicating that they are reacting to form sulfuric acid cloud particles. The Venusian atmosphere also contains detectable amounts of hydrochloric acid (HCl) and hydrofluoric acid (HF).

dized to sulfur and perhaps to sulfur dioxide under the influence of near-ultraviolet light, which is transmitted by the clouds. The formation of elemental sulfur is an important feature of the slow subcycle: it accounts for the clouds' absorption of radiation in the ultraviolet range. Sulfuric acid, sulfur dioxide and elemental sulfur are ultimately reduced by molecular hydrogen and carbon monoxide to re-form hydrogen sulfide and carbonyl sulfide in the hot lower atmosphere. The slow subcycle probably takes about a decade to complete, because sulfur becomes trapped in the fast subcycle for some years before it is reduced.

The total mixing ratio of sulfur compounds (and thus the thickness of the sulfuric acid clouds) is determined by the interaction of atmosphere and crust in the geologic subcycle. Enough information is now available to understand the interaction. The most important new data come from the X-ray-fluorescence instruments on board the *Venera 13* and *Venera 14* landing craft. Using these data, Yu. A. Surkov and his colleagues at the Vernadsky Institute recently determined the elemental composition of the crust for the first time. (When an element is exposed to X rays from such an instrument, it fluoresces in the X-ray range with a characteristic spectrum.) The Venusian surface rocks contain a relatively large amount of sulfur, but otherwise their elemental abundances are similar to those found in certain terrestrial basalts, a type of volcanic rock.

The geologic subcycle begins with the production of sulfur gases, particularly carbonyl sulfide and hydrogen sulfide, by thermochemical reactions of the sulfur-containing mineral pyrite with carbon dioxide, water and carbon monoxide. Once in the atmosphere, the gases are pumped many times through the fast and slow atmospheric subcycles. The photochemical links between the two result in a net conversion of the sulfide gases into sulfur dioxide, and so sulfur dioxide accumulates in the atmosphere beyond its thermochemical equilibrium level. Sulfur dioxide is removed from the atmosphere and stored as calcium sulfate in the crust through a reaction with carbon dioxide and calcium oxide. After hundreds or even millions of years the calcium sulfate is buried; it then reacts with iron oxide to regenerate pyrite and complete the subcycle.

Surkov and his colleagues have found abundant calcium in the Venusian surface rocks. The concentration of sulfur in the rocks is much lower, indicating that most of the calcium is in oxide form in silicates and carbonates rather than in sulfates. Since calcium oxide is available and since the atmospheric mixing ratio of sulfur di-



SULFUR CYCLE responsible for producing the clouds of Venus consists of three subcycles. The geologic subcycle (*colored lines*) begins when volcanic pyrite (FeS_2) reacts thermochemically, either underground or at the surface of lava flows, to form sulfur dioxide and the sulfide gases H_2S and OCS. In the slow (*gray lines*) and fast (*black lines*) atmospheric subcycles, photochemical reactions, driven by ultraviolet light within the clouds and near-ultraviolet light below them, oxidize the sulfur gases to form the cloud particles: sulfuric acid (H_2SO_4) and various forms of elemental sulfur (S). The oxygen comes from the photodissociation of carbon dioxide. In the lower atmosphere evaporation and thermochemical reduction by carbon monoxide (CO) and hydrogen destroy the cloud material and regenerate the gaseous cloud precursors. Photooxidation leads to a net conversion of the sulfide gases into sulfur dioxide, which reacts with calcium oxide (CaO) at the surface to produce calcium sulfate ($CaSO_4$), thereby returning sulfur to the crust.

VENUSIAN PLAIN was photographed in March, 1982, by the Soviet landing craft *Venera 14*. The average temperature at the surface of Venus is 460 degrees Celsius, the atmospheric pressure is 100 times that of the earth and the atmosphere is highly corrosive. Nevertheless, the instruments on *Venera 13* and *Venera 14* survived for several hours and transmitted a series of photographs.

oxide exceeds its equilibrium level, the two substances must be reacting spontaneously to form calcium sulfate. Hence there is no doubt that sulfur dioxide is currently being drained from the atmosphere by the calcium oxide "sink." Its presence at 10 times the equilibrium mixing ratio can be explained only by a geologically very recent injection of sulfur gases.

The hypothesized sulfur source, pyrite, is a common secondary mineral in terrestrial volcanic rocks. The reactions that produce sulfur gases might occur at the surface of pyrite-rich lava flows. Or they might take place below the surface, with the sulfur gases then being ejected by volcanic eruptions. Either way it is a volcanic mechanism that ensures a steady flow of sulfur gases into the atmosphere.

Pettengill and Peter G. Ford of M.I.T. have recently found indirect evidence for vast areas of exposed pyrite in Beta Regio, where the radar maps had already revealed volcanic structures. Observations carried out by the *Pioneer Venus* orbiter showed that the reflectivity of the Venusian surface to radio waves of 17-centimeter wavelength varies widely, from a low of 3 percent to a high of 40 percent. One highly reflective area is Theia Mons in Beta Regio; Pettengill and Ford argue that the composition of this mountain must be distinctly different from that of the poorly reflecting lowlands surrounding it. High reflectivity implies high electrical conductivity, and remarkably few rock types have the conductivity necessary to explain the reflectivity of Theia Mons. Rock with substantial inclusions of highly conductive pyrite is the most likely candidate. Compared with Beta Regio, the areas where *Venera 13* and *Venera 14* set down show very low reflectivity; this suggests that the sulfur detected by these probes was largely in the form of poorly conducting calcium sulfate rather than pyrite.

The pyrite on the surface in Beta Regio could be the product of recent lava flows, but it could also have been exposed by the weathering of older deposits. Because the clouds prevent the direct observation of active lava flows or of plumes of dust and gas, it is not possible to "prove" that volcanic eruptions are taking place on Venus today, as the spectacular *Voyager* photographs did for Io. On the earth, however, massive eruptions—such as that of El Chichón in 1982—send large amounts of haze particles and sulfur gases into the upper atmosphere, where they remain for months or even years. If massive eruptions are occurring on Venus, they might be expected to leave an observable signature above the clouds.

Apparently they do. In one of the more surprising discoveries of the *Pioneer Venus* mission, Larry W. Esposito of the University of Colorado at Boulder reported last year that the orbiter's ultraviolet spectrometer recorded a 90 percent drop between 1978 and 1983 in the levels of sulfur dioxide and sulfuric acid haze particles above the clouds. In 1978 the levels had greatly exceeded the values workers had expected on the basis of observations over the previous 15 years from the earth; there may have been a similar anomalously high abundance of haze particles in the late 1950's. Esposito has proposed that in both the late 1950's and the late 1970's energetic volcanic eruptions injected sulfur dioxide directly into the upper atmosphere. The sulfur dioxide is then converted into sulfuric acid, which sinks back into the lower atmosphere. Gradually the high abundance of sulfur dioxide and of haze particles above the clouds returns to normal.

It is not necessary to assume that volcanic plumes can rise undiluted some 70 kilometers to the cloud tops—in my view an unlikely event. Even if the hot volcanic gas itself did not rise above the clouds, the convective energy of a massive eruption would propagate upward as intense gravity waves, gaining in amplitude and ultimately breaking in the cloud region much as oceanic waves break on a beach. The concentration of sulfur dioxide is 500 times greater at the cloud base than at the cloud tops, and so turbulent stirring caused by breaking gravity waves could explain the episodic increase in the sulfur dioxide level in the upper atmosphere. In any case the explanation requires large volcanic eruptions.

Another observation from *Pioneer Venus* may indicate just where eruptions are occurring. Frederick L. Scarf of TRW Inc. and Christopher T. Russell of U.C.L.A. reported recently that an antenna on the spacecraft has picked up low-frequency radio bursts thought to be emitted by strokes of lightning. The bursts are clustered conspicuously above several surface features: Beta Regio, Atla Regio and Phoebe Regio, which lies south of Beta. The clustering is difficult to explain if the lightning is attributed to random convection in the atmosphere. All three regions, however, have been identified by their topography as possibly being of volcanic origin, and on the earth lightning discharges have often been seen in the plumes of erupting volcanoes. The evidence suggests that similar discharges occur on Venus.

None of the arguments I have presented is entirely convincing on its own, but together they make a very persuasive case: there are active volcanoes on Venus, and this volcanism is a key link in the chemical cycle that produces the clouds. As yet there is no clear evidence of plate tectonics, and so the volcanoes of Venus are probably isolated hot spots rather like Mauna Loa in Hawaii. The levels of volcanic activity on the two planets seem to be roughly comparable; some investigators believe eruptions are even more frequent on Venus. At the very least it is now safe to conclude, after decades of speculation, that the cloud-covered planet, like the earth, is still evolving, still geologically alive.

Chromosome Translocations and Human Cancer

Chromosomes in a cell of the immune system sometimes "trade" segments of DNA. This process can activate cancer-causing genes by placing them near genetic sequences that enhance their activity

by Carlo M. Croce and George Klein

Every human cell contains oncogenes, genes that have the potential to cause cancer. These genes apparently carry out normal functions until a malignant change takes place. What is it that changes the oncogene from a normal part of the cell's genetic machinery into a source of cancerous, or neoplastic, transformation?

In the past decade several different mechanisms by which an oncogene may be activated have been discovered [see "A Molecular Basis of Cancer," by Robert A. Weinberg; SCIENTIFIC AMERICAN, November, 1983]. Sometimes an oncogene is activated by a "point mutation"; a small segment of the gene is altered by radiation or a chemical carcinogen. Another way an oncogene can be activated is through "amplification," in which the oncogene is somehow replicated many times so that several active copies of it are present in the same cell. When this happens, the gene may be expressed at an inappropriately high level; in other words, the cell may make too much of the protein encoded by the gene. Even a protein that is necessary for the proper functioning of the cell may have cancerous effects when it is produced in large quantities [see "The Proteins of Oncogenes," by Tony Hunter; SCIENTIFIC AMERICAN, August, 1984].

An oncogene could also be activated by incorporation into a retrovirus (a virus whose genetic material is made of RNA rather than DNA). When a retrovirus infects an animal cell, it can pick up from the cell an unactivated oncogene, which becomes part of the genetic complement of the retrovirus and its progeny. This process sometimes activates the oncogene, and so subsequent infection by that strain of retrovirus can induce a neoplastic transformation in a different cell. At present it is not clear what role these mechanisms of activation play in the development of human tumors, since few human tumors carry an oncogene activated in any of these ways.

Our work and the work of other investigators has shown there is yet another mechanism that can activate an oncogene. This mechanism operates in some cancers of cells in the immune system called *B* cells. The primary function of a *B* cell is to produce antibodies (immunoglobulins), the molecules that recognize and bind to the "nonself," or foreign, molecules called antigens. The genes that encode the production of antibodies must be expressed at a high level for the *B* cell to fulfill its function properly. Genetic sequences within the genes that encode antibody production increase the activity of such genes in *B* cells. If a rearrangement of the *B* cell's chromosomes (rodlike strands of DNA that contain the cell's genes) somehow juxtaposed such a sequence with an oncogene, then expression of the oncogene would be enhanced. Malignant transformation would seemingly become a primary part of that cell's function.

We have found that such rearrangements do indeed occur in Burkitt's lymphoma, an extremely fast-growing malignancy of the immune system. The rearrangements result from reciprocal translocations between two of the *B* cell's chromosomes: a segment of each chromosome breaks off and moves to the end of the other chromosome [*see illustration on page 56*]. In most of these translocations an oncogene moves into a position near one of the sequences that enhances antibody production; less often the oncogene remains in place and the enhancing sequence shifts.

We came on this mechanism in the course of research begun in the late 1970's to identify the chromosomes that contain the genes responsible for antibody production. Once we had mapped the locations of these genes, we noticed they lay on precisely the same chromosomes that were already known to be translocated in Burkitt's lymphoma cells. Our later research showed that the two segments of genetic material that translocate between chromosomes contain, respectively, an oncogene and a gene encoding part of the antibody molecule.

In order to discover which chromosomes contain the genetic information that codes for antibody production, one of us (Croce) and his associates employed an experimental technique involving hybrids between mouse and human somatic cells (body cells rather than eggs or sperm).

Hybrid cells are produced by mixing mouse and human *B* cells in a medium containing a chemical or viral fusion factor, which joins some mouse and human cells. The resulting hybrids contain both human and mouse chromosomes. Each hybrid cell of this type may lose some human chromosomes during cell division (although it retains its entire complement of mouse chromosomes), and so as the cells divide and multiply successive generations have fewer human chromosomes; after many generations each of the hybrid cells will have only a few human chromosomes.

To determine which chromosomes hold the genes coding for parts of the antibody molecule, we examined several panels of such cells. Any cell producing a part of the molecule must have contained one of the necessary chromosomes. We were able to determine which chromosomes hold the genes for any particular part of the antibody molecule by noting which human chromosomes were always

present in cells producing that part but absent in any cells not producing it.

The antibody molecule is made up of four protein chains; the chains link in two identical pairs to form a Y shape [*see illustration on page 57*]. The longer chain in each pair is called the heavy chain, and the shorter one is called the light chain. Each chain consists of two characteristic regions: the "variable" region and the "constant" region. The variable region recognizes and binds to antigens; the constant region specifies the task to be performed by the antibody (called its effector function) after it has encountered and bound an antigen. There are many different types of variable region, because antibodies are highly selective: each binds to only one specific antigen. In contrast, only two types of constant region occur in light chains (called kappa and lambda respectively) and only 10 types of constant region occur in heavy chains. Thus antibodies to different antigens may well perform the same task. Each mature *B* cell can produce only one type of antibody, and the cell's chromosomes contain DNA coding for the variable and constant regions specific to that antibody [see "The Genetics of Antibody Diversity," by Philip Leder; SCIENTIFIC AMERICAN, May, 1982].

In 1979 one of us (Croce) and his associates found that hybrid cells containing human chromosome 14 were the only ones to produce human heavy chains. Apparently the genes coding for production of heavy chains lie on chromosome 14. Using similar experimental techniques Jan Erikson, Joanne Martinis and one of us (Croce) found in 1981 that chromosome 22 encodes the light chains bearing the lambda constant region. In 1982 O. Wesley McBride and his associates at the National Cancer Institute and Terence H. Rabbitts and his collaborators at the Medical Research Council's Laboratory of Molecular Biology in

CHROMOSOMES from a hybrid of mouse and human cells include human (*pale*) and mouse (*dark*) chromosomes. One of the human chromosomes (*arrow*) has undergone a translocation: a segment of one end has broken off and has been replaced by a segment from another chromosome. Translocations can activate oncogenes (genes that cause cancer) by placing them near enhancers, genetic sequences that increase activity of certain other genes on the same chromosome. Because hybrid cells contain some, but not all, of the human genetic complement, they can be used to determine the chromosomes encoding a human product: any human product produced by a cell containing just one human chromosome must be produced by that chromosome. The authors used such hybrid cells to identify the chromosomes containing certain oncogenes and to study the effects of various translocations on the regulation of these oncogenes.

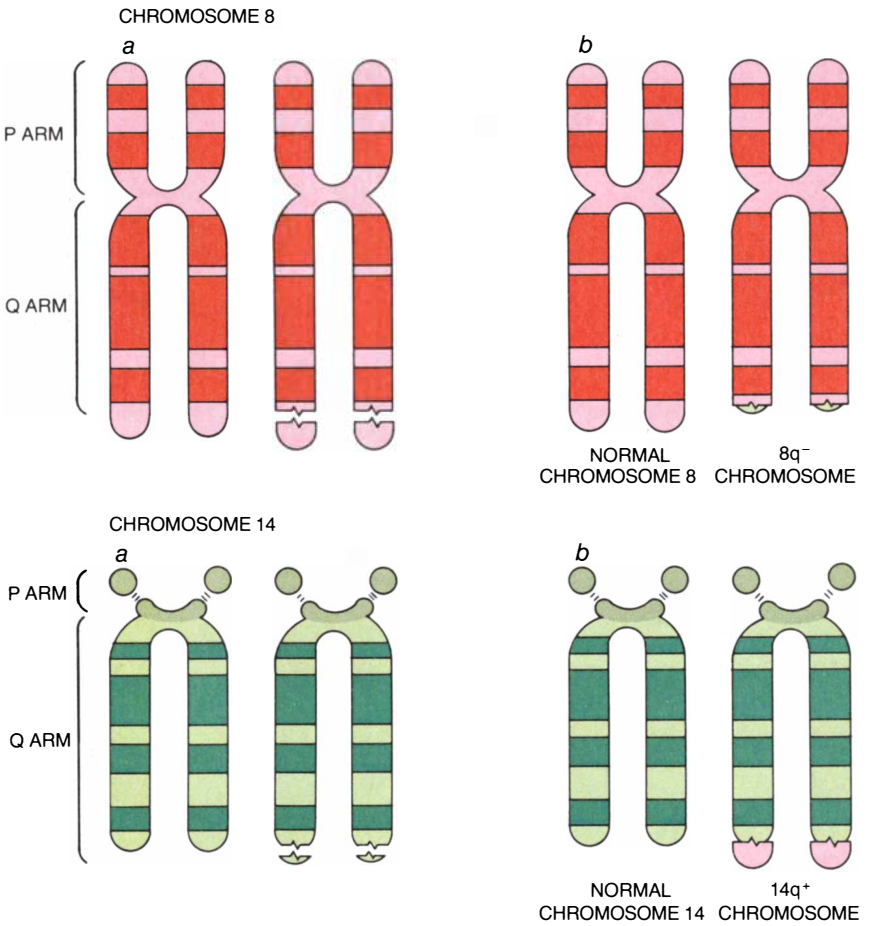
Cambridge found that chromosome 2 encodes the light chains carrying the kappa constant region.

These results fitted neatly with work that had been done a decade earlier on chromosome translocations. In 1972 George Manolov and Yanka Manolova, working at the University of Lund in Sweden, found an irregularity in the chromosomes of many cells that were affected by malignant Burkitt's lymphoma: one of the chromosomes in the 14th pair (a human somatic cell has 23 distinct pairs of chromosomes) was elongated. Noting that one section of the chromosome, called the *q* arm, was abnormally long, the Manolovs called the unusual chromosome $14q^+$.

Subsequently Lore Zech, in collaboration with one of us (Klein) at the Karolinska Institute, found that the $14q^+$ chromosome is the result of a reciprocal translocation: it forms when an end segment of a chromosome in the eighth pair breaks off and joins chromosome 14. A segment of chromosome 14 makes the opposite transition and joins the end of chromosome 8. The rearranged eighth chromosome is called $8q^-$, because it has a foreshortened *q* arm. More recently other workers found that two other chromosome translocations may occur in Burkitt's lymphoma cells. Both involve chromosome 8; in one kind of translocation (occurring in approximately 16 percent of the cases of Burkitt's lymphoma) the reciprocal shift is between chromosomes 8 and 22. In approximately 9 percent of the cases the translocation involves chromosomes 2 and 8. Three of the chromosomes affected by these translocations, chromosomes 2, 14 and 22, are involved in the production of antibodies.

The association between Burkitt's lymphoma and antibody production soon proved to be closer. One of us (Croce) and Erikson at the Wistar Institute of Anatomy and Biology and Janet Finan and Peter C. Nowell of the University of Pennsylvania School of Medicine found that the point at which chromosome 14 breaks during the translocation with chromosome 8 is situated precisely within the section of chromosome 14 that encodes the immunoglobulin heavy chain. For these experiments we once again used hybrids between mouse and human cells; in this case we used cells from the mouse immune system that had been transformed by a type of cancer called a plasmacytoma. Each hybrid cell contained, in addition to the mouse genetic complement, at least one chromosome from a human Burkitt's lymphoma cell.

As we had expected, hybrid cells with the normal chromosome 14 (the chromosome of the 14th pair that had not been affected by the translocation) possessed genes for antibody production; those with the normal chromosome 8 did not [*see illustration on page 59*]. On the other hand, hybrids with a chromosome 14 that had been involved in a translocation (the $14q^+$ chromosome) contained the genes for the constant regions of heavy chains but not for the variable regions. Chromosome 8 that had taken part in the translocation contained the genes for the variable regions. These results are evidence that chromosome 14 breaks between the genes coding for the variable and constant regions of the heavy chain, and that the genes coding for the variable region move to chromosome 8. The heavy-chain locus (the part of chromosome 14 that encodes the heavy chain) is thus directly involved in one of the translocations that is characteristic of Burkitt's lymphoma.



RECIPROCAL TRANSLOCATION between chromosomes 8 (*red*) and 14 (*green*) causes most instances of Burkitt's lymphoma, a malignancy of *B* cells in the human immune system. A segment from the end of chromosome 8 breaks off (*a*) and moves to chromosome 14 (*b*). The reverse translocation moves a segment from chromosome 14 to chromosome 8. Such reciprocal translocations place an oncogene from chromosome 8 near a gene on chromosome 14 that usually codes for production of part of the antibody molecule. A mechanism that enhances production of antibodies in normal *B* cells then activates the oncogene.

At this point it was clear the mechanism of Burkitt's lymphoma was somehow related to the genes that code for the production of antibodies; our next clues to the nature of the relation came from studies of oncogenes. Because Burkitt's lymphoma affects the *B* cells, we were particularly interested in an oncogene designated *c-myc*, a human oncogene closely related to the oncogene *v-myc*, which causes a *B*-cell lymphoma in chickens that have been infected with avian myelocytomatosis virus.

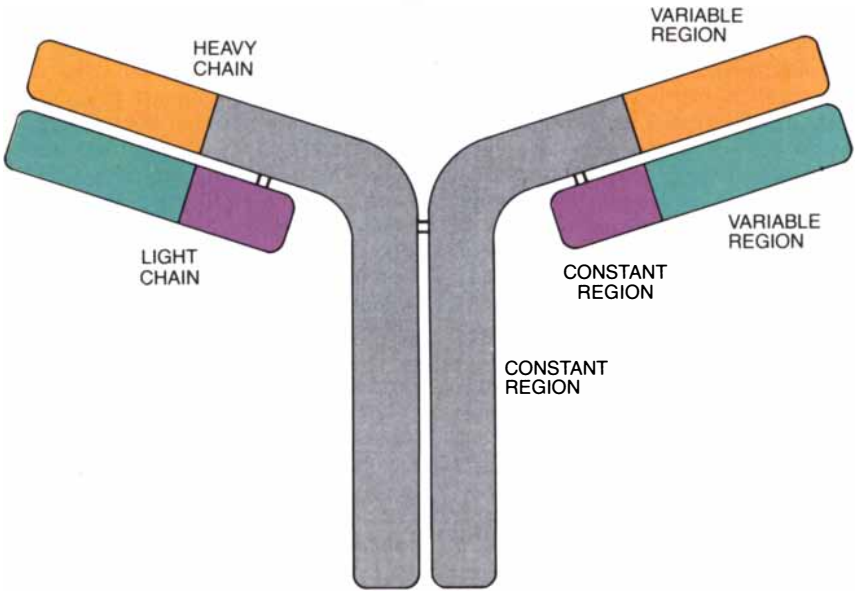
In collaboration with Riccardo Dalla-Favera and Robert C. Gallo of the National Cancer Institute, we used the close relation between human and avian *myc* genes to construct a probe that would identify hybrid cells containing the human *c-myc* oncogene. Our probe was a radioactively labeled segment of human DNA whose genetic sequence was very similar to that of the *v-myc* oncogene.

To tell whether a hybrid cell contained the human *c-myc* gene, we used an enzyme to cut the cell's genetic material into small segments; we then separated the segments by size, using a process called gel electrophoresis. Next we exposed the DNA to a solution containing radioactive probe DNA. Both the cellular DNA and the probe had been "denatured"; that is, in each case the two complementary strands of DNA that make up that molecule's double-helix shape had been separated. Because the *c-myc* probe and the human *c-myc* gene are nearly identical, the labeled strands of probe *c-myc* "hybridized" with the cellular *c-myc;* that is, single strands of probe DNA joined to complementary strands of cellular DNA. When we washed away the solution of probe DNA, any probe that had hybridized was left behind. After this washing process any cell whose genetic material had hybridized with the radioactive probe could be discerned by a specific radioactive band that appeared on the filter paper holding the sorted DNA.

We used the probe to examine a panel of mouse-human hybrid cells in order to determine the chromosomal location of the human *c-myc* gene. We first examined hybrid cells with normal human chromosomes and found that human chromosome 8 was present in all cells containing the human *c-myc* oncogene and absent in those without it; we concluded that the *c-myc* oncogene lies on chromosome 8.

Next we examined hybrid cells that contained the translocated chromosomes 8 and 14, derived from unions between mouse cells and human Burkitt's lymphoma cells. We observed that the *c-myc* oncogene resides in the small segment of chromosome 8 that consistently translocates to chromosome 14 in Burkitt's lymphoma cells containing the translocation between chromosomes 8 and 14. This result indicated that translocations involving the *c-myc* oncogene play a fundamental role in development of Burkitt's lymphoma.

Interestingly, similar specific chromosomal translocations have also been observed in mouse plasmacytomas by Shinsuke Ohno, Francis Wiener and Jack Spira, working at the Karolinska Institute with one of us (Klein) in collaboration with Michael Potter and his associates at the National Cancer Institute. Their study found that malignant antibody-producing cells of mice carried a characteristic translocation between chromosome 15 and the mouse chromosomes that have either the heavy-chain genes or the genes for the light-chain kappa region (mouse chromosomes 12 and 6 respectively). These results suggested that immunoglobulin genes have a role in mouse plasmacytomas.



ANTIBODY MOLECULE consists of two identical pairs of protein chains joined to form a split Y shape. In each pair there are a heavy chain and a light chain, and each chain has a variable region and a constant region. Most cases of Burkitt's lymphoma are caused by a translocation of an oncogene to the genetic locus encoding the heavy chain. Other cases are caused by translocations involving genes for the light-chain constant regions.

Later experiments done by one of us (Croce), Dalla-Favera and Gallo, in collaboration with Stuart A. Aaronson of the National Cancer Institute and Philip Leder of the Harvard Medical School, showed that the *c-myc* oncogene translocated to chromosome 14 in human Burkitt's lymphoma may be arranged in several ways.

The oncogene consists of three exons (DNA segments that are transcribed to make messenger RNA, or mRNA, during the process of expression as proteins) broken up by two introns (segments consisting of DNA that is not transcribed into mRNA and is thus not expressed as protein). The structure of the oncogene was analyzed by Dalla-Favera and Gallo and their associates and by Rosemary Watt, Giovanni Rovera and one of us (Croce) at the Wistar Institute. In some Burkitt's lymphoma translocations the breaking point on chromosome 8 is "upstream" of the entire *c-myc* oncogene and all three exons of the gene are translocated to chromosome 14; in other cases, however, the breaking point is "downstream" of the first exon and only the second and third exons are translocated [*see illustration on page 60*]. In this case the oncogene is attached "head to head" with one of the heavy-chain genes on chromosome 14; that is, the translocated segment of DNA from chromosome 8 runs in a direction opposite to that of the DNA from chromosome 14.

Other experiments have shown that a similar rearrangement occurs in translocations underlying mouse plasmacytomas. These studies were first carried out by Michael D. Cole and his associates at the St. Louis University Medical Center, and later by Leder, by one of us (Croce) in collaboration with Kenneth B. Marcu of the State University of New York at Stony Brook, and by Jerry Adams and Suzanne Cory of the Walter and Eliza Hall Institute of Medical Research in Melbourne. In mouse plasmacytomas the *c-myc* oncogene is rearranged head to head with an immunoglobulin heavy-chain gene. It is not yet clear, however, whether the oncogene is translocated to the heavy-chain locus or whether it remains on mouse chromosome 15 while the heavy-chain locus is translocated near it.

In spite of the several possible chromosome rearrangements in cells of human Burkitt's lymphoma, we found that the protein produced by the *c-myc* gene was qualitatively the same. Specifically we found that the first *c-myc* exon does not code for a protein; protein synthesis begins at the second exon. Therefore it was not the rearrangements of the *c-myc* gene during translocation that had activated its oncogenic qualities; the cancerous effect of translocation is not due to some alteration within the gene.

If the *c-myc* product is the same in

normal cells and in Burkitt's lymphoma cells, what is the oncogenic consequence of the chromosome translocation in Burkitt's lymphoma? Perhaps the translocation somehow causes the *c-myc* gene product, small quantities of which may be necessary for the cell's function, to be expressed at abnormally high levels.

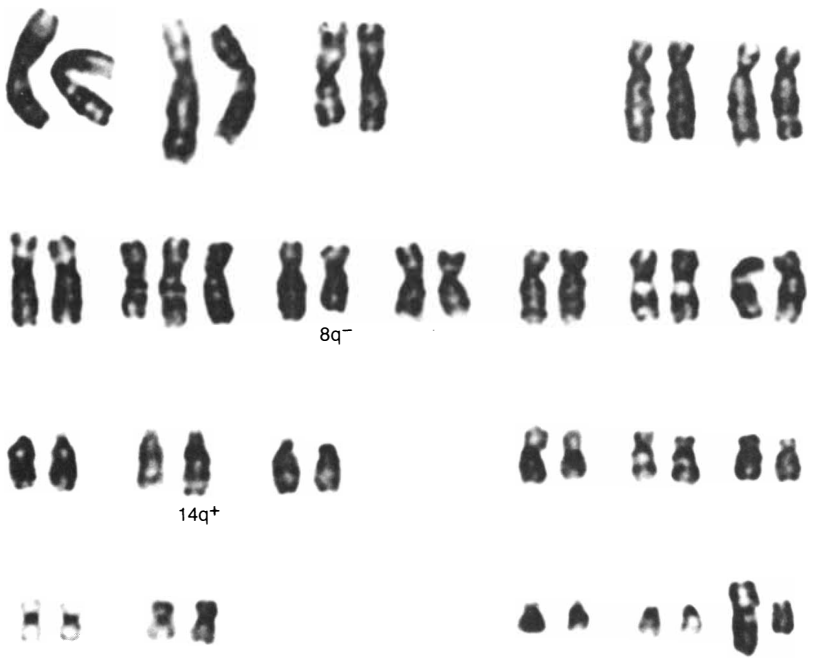
In other words, it is possible the translocation enables the *c-myc* gene to evade the mechanisms that normally control its expression. If this is the case, there should be a difference between the levels of expression of the translocated gene and the normal *c-myc* oncogene in the same cell. Cells with the $14q^+$ chromosome should have elevated levels of *c-myc* mRNA, the genetic material that represents an intermediate step between the presence of a gene on a chromosome and its expression as a protein. Cells with normal chromosome 8 should have low levels of *c-myc* mRNA.

With that possibility in mind Kazuko Nishikura and our associates at the Wistar Institute undertook further experiments with hybrids of human cells and mouse plasmacytomas. Using a method that enabled us to distinguish human *c-myc* mRNA transcripts from mouse *c-myc* mRNA, we found that the *c-myc* gene on the $14q^+$ chromosome is expressed at high levels; the *c-myc* gene on normal chromosome 8 is relatively silent in the same kind of plasmacytoma cell.

In parallel experiments we introduced a *c-myc* gene on normal chromosome 8 into mouse plasmacytoma cells. The gene had come from noncancerous human *B* cells. We found that this gene, which had been expressed (albeit at low levels) in normal human *B* cells, was shut off completely in mouse plasmacytoma cells. Adams and Cory found in other studies that the untranslocated mouse *c-myc* gene in mouse plasmacytoma cells is not expressed. Thus whereas the normal (untranslocated) *c-myc* gene is repressed in the background of a mouse plasmacytoma cell, a *c-myc* oncogene that is translocated to the heavy-chain locus on chromosome 14 somehow escapes the mechanisms that normally control transcription.

We also examined *c-myc* mRNA transcripts from Burkitt's lymphoma cells carrying a variant translocation. In these cells the first exon of the *c-myc* gene remains on chromosome 8 and the other two exons are rearranged head to head with the genes on chromosome 14. (Each of these cells had a normal eighth chromosome in addition to the translocated chromosome.) In these cases it is relatively easy to tell the difference between mRNA of the translocated gene and that of the gene from normal chromosome 8: the translocated gene has been partially rearranged and hence its mRNA transcripts will be different. In such cells Abbas ar-Rushdi and our other collaborators observed high levels of mRNA from the translocated *c-myc* gene but not from the normal *c-myc* gene.

These results indicate the *c-myc* oncogene becomes deregulated as a result of proximity to genes that code for antibodies. This conclusion is strengthened by observations concerning two translocations occurring in Burkitt's lymphoma that do not involve chromosome 14. One of these translocations is between chromosome 8 and chromosome 22, the chromosome that contains genes encoding light chains of the lambda type. The other is between chromosome 8 and chromosome 2, the chromosome that contains genes encoding kappa light chains. In both of these translocations, as one of us (Croce) showed, in collaboration with Nowell and with Gilbert Lenoir of the International Agency for Research on Cancer in Lyons, the *c-myc* gene remains on chromosome 8, where it is joined by a sequence that encodes antibody production (in one case by the lambda light-chain locus and in the other by the kappa locus). Either translocation can activate the oncogene by making it unresponsive to the mechanisms that normally control expression. Apparently the *c-myc* oncogene does not have to move in order for it to be expressed at elevated levels.



TRANSLOCATED CHROMOSOMES of a Burkitt's lymphoma cell differ in length from their untranslocated counterparts. In this cell one of the chromosomes in the eighth pair has undergone a translocation with a chromosome from the 14th pair. The translocated chromosome 8 has been shortened and the translocated chromosome 14 lengthened.

What is responsible for the deregulation of the *c-myc* oncogene taking part in these translocations? An experimental observation suggests an answer: the translocated *c-myc* oncogene of Burkitt's lymphoma is repressed in hybrid cells that are based on mouse fibroblasts (cells of connective tissue), whereas it is expressed at high levels in hybrids based on plasmacytoma cells (malignant antibody-producing cells). It appears the translocation has an oncogenic effect only in a cell that produces antibodies, one in which the chromosomal regions that are necessary for antibody production are particularly active.

Such chromosomal regions contain a type of genetic sequence called an enhancer. Enhancers are sequences of DNA that seem to increase the levels of transcription of certain other genes on the same chromosome; they are a recent discovery and little is known about how they function. Workers in the laboratories of Kathryn L. Calame of the University of California at Los Angeles, of Walter Shaffner of the University of Zurich and of Susumu

Tonegawa of the Massachusetts Institute of Technology have found there are enhancing sequences within the segment of DNA that codes for one type of immunoglobulin heavy-chain constant region. Recent studies at the Wistar Institute suggest the presence of additional enhancers in the heavy-chain locus. In addition David Baltimore and his associates at M.I.T. have found enhancing sequences within the chromosomal regions encoding the light-chain kappa constant region.

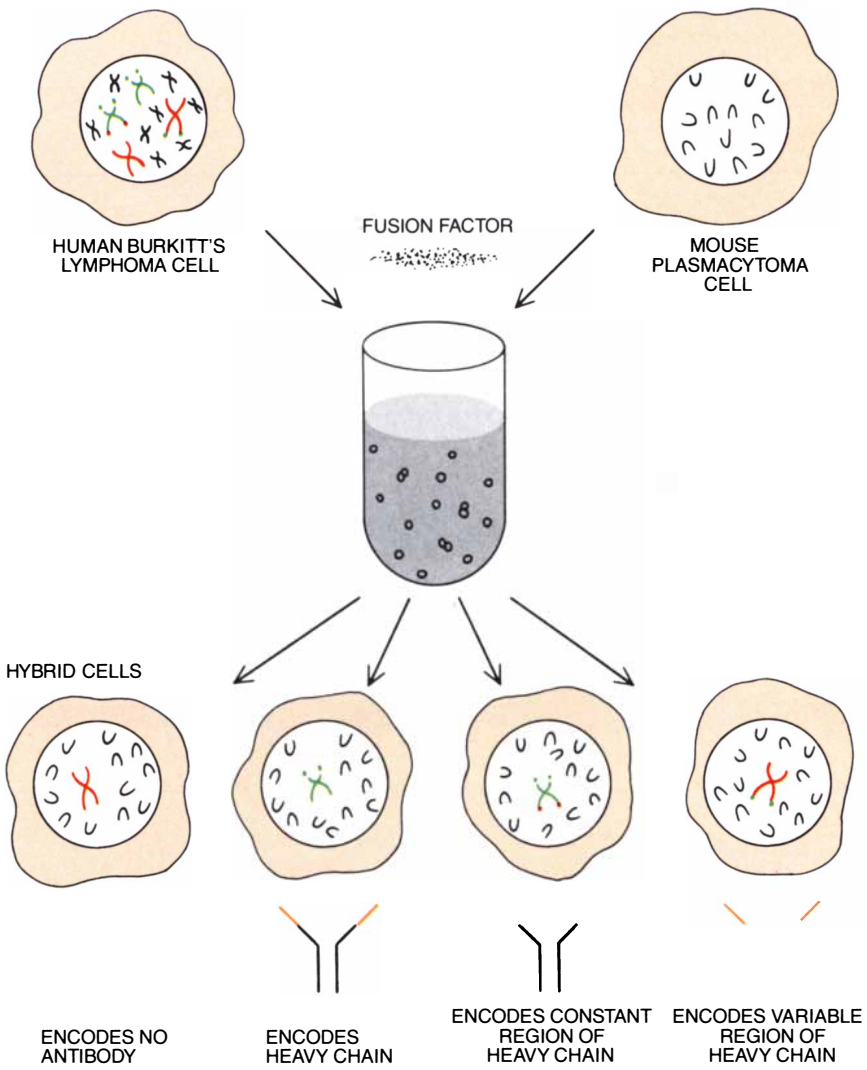
These findings suggest a possible mechanism for Burkitt's lymphoma: chromosome translocations within a *B* cell place a *c-myc* oncogene in juxtaposition to enhancers; the enhancers are capable of activating transcription over considerable distances. The *c-myc* gene is then expressed in the same way that immunoglobulin genes are expressed in a normal *B* cell. In a sense the expression of the *c-myc* oncogene becomes a part of the cell's specialized function.

Recent results indicate this mechanism of oncogenesis may be responsible for many other malignancies involving cells of the human immune system. Jorge J. Yunis of the University of Minnesota Medical School has developed a new method of banding, or staining, chromosomes that makes possible a highly precise detection of specific chromosomal changes in malignant cells. His applications of this technique suggest definite chromosomal changes mark the majority of *B*-cell malignancies. Translocations between chromosome 14 and segments of either chromosome 11 or chromosome 18 are common in *B*-cell lymphomas in adults, in human chronic *B*-cell leukemias and in multiple myeloma. This observation, combined with our knowledge of the role of the immunoglobulin heavy-chain locus on chromosome 14 in Burkitt's lymphoma, indicates human oncogenes may lie on chromosomes 11 and 18; this conjecture is supported by work done by Yoshihide Tsujimoto at the Wistar Institute in collaboration with Yunis and Nowell. We have found that the breaking points in these translocations were consistently clustered in short segments of chromosome 11 or 18; in addition the breaking points always lie in front of the segment of chromosome 14 that encodes the constant region of the heavy chain. We have proposed the designations *bcl-1* and *bcl-2* (*B*-cell lymphoma/leukemia 1 and 2) for the two putative oncogenes on chromosomes 11 and 18.

Observations made in the study of Burkitt's lymphoma open two new major areas for investigation. First, there is the matter of enhancers. What are the precise sequences of DNA that make up an enhancer, and how does the enhancer increase the level of transcription of certain genes? The other area opened for investigation concerns the *c-myc* oncogene. What is the function of the *c-myc* gene in a normal cell, and why should the expression of *c-myc* at elevated levels cause malignancy?

In addition to these new areas of research, our work suggests new experimental approaches to the study of *B*-cell neoplasms. Many of these malignancies involve the translocation of an unknown oncogene to the heavy-chain locus on chromosome 14. The heavy-chain locus is relatively well known, and there are nucleic acid probes that make it possible for an investigator to study segments of DNA close to it. Since translocations tend to bring the oncogene into close proximity with the heavy-chain locus, such probes will provide investigators with the tools to identify, isolate and characterize genes connected with the majority of human *B*-cell cancers. In this way knowledge of the genetics of antibody production would yield knowledge of the genetic structure of newly isolated oncogenes.

Work with chromosome transloca-

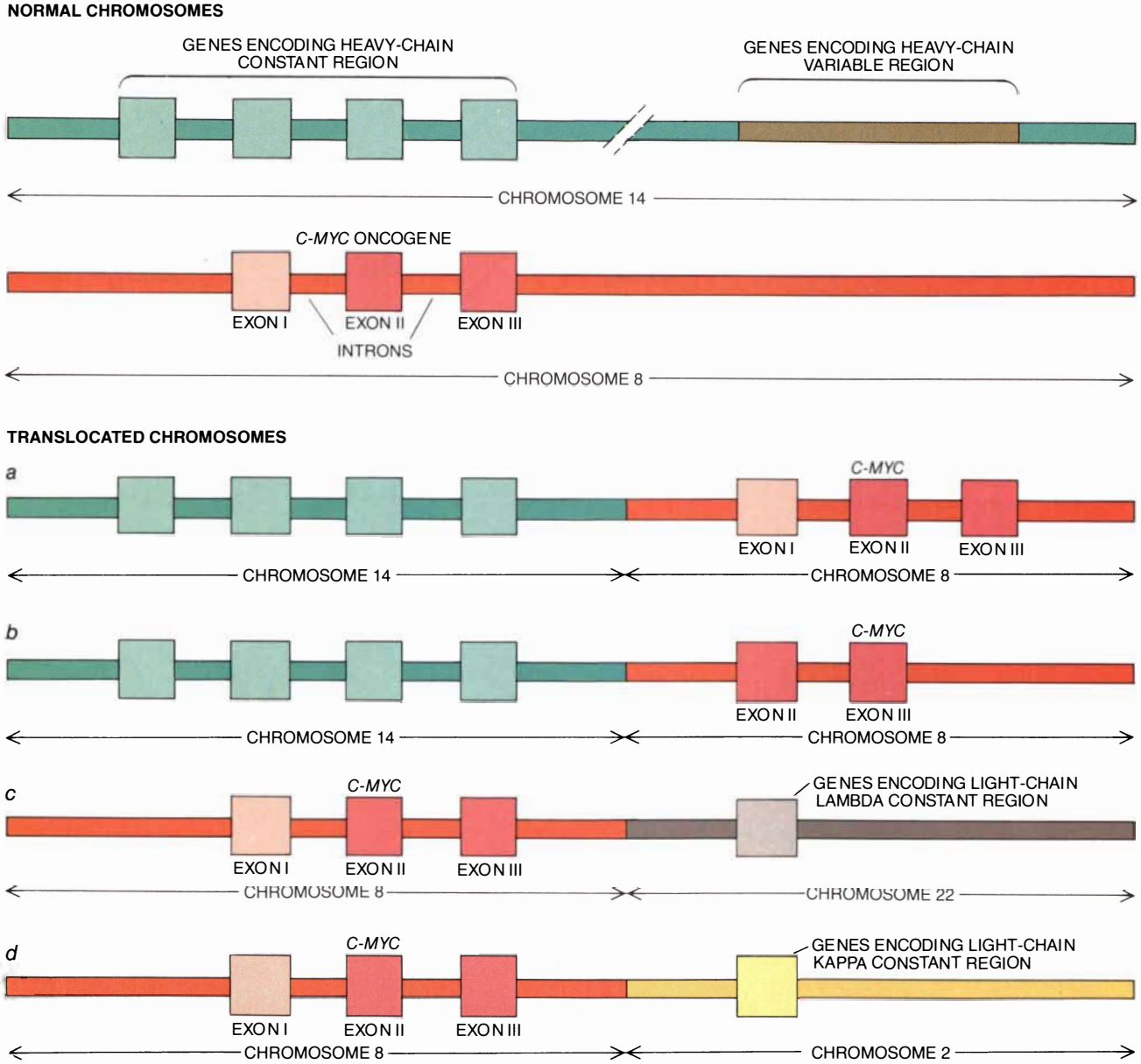


BREAKING POINT on translocating chromosome 14 lies within the region that encodes the antibody heavy chain, as this experiment showed. A human Burkitt's lymphoma cell, containing both normal and translocated chromosomes 8 and 14, was fused with a mouse plasmacytoma cell (a cancerous cell of the mouse immune system). Each hybrid cell retained one of the chromosomes from the human cell. Hybrid cells containing normal chromosome 8 (*red*) produced no antibody. Cells with normal chromosome 14 (*green*) produced the antibody heavy chain. The chromosome 14 involved in the translocation contained genes for only the constant regions of the heavy chains, and the involved chromosome 8 contained genes for the heavy-chain variable region. Apparently in the process of translocation chromosome 14 breaks directly between loci encoding constant and variable regions.

tions may also lead to new methods of diagnosing and characterizing cancers of the immune system. The chromosomal breaking points, for example, cluster within short segments of DNA in *B*-cell malignancies that exhibit the translocations between chromosomes 11 and 14 or between chromosomes 14 and 18. It should therefore be possible to develop DNA probes that are specific to these small segments. A tissue sample could then be taken from the affected area of a patient, and the DNA probes could be used to determine precisely which kind of chromosomal rearrangement is responsible for the malignancy.

Very recent results indicate that the lessons learned from *B*-cell malignancies may also be applicable to malignancies of *T* cells, the other major constituent of the immune system. One of us (Croce), in collaboration with Rovera and with Mark M. Davis of Stanford University, has found that the gene for the alpha chain of the *T*-cell receptor lies within the region of chromosome 14 that is involved in some translocations characteristic of certain cancers affecting *T* cells.

Our studies of the mechanism underlying Burkitt's lymphoma thus have implications beyond this one disease. The translocations in Burkitt's lymphoma seem to provide a model for the majority of human *B*-cell cancers (and perhaps for *T*-cell cancers as well). In addition knowledge of the translocation mechanism will provide powerful experimental tools not only for the study of other cancers but also for the study of the mechanisms that control genetic expression during the normal development and function of the human immune system.



VARIETY OF TRANSLOCATIONS can cause Burkitt's lymphoma. In the commonest case (*a*) all three exons (sequences of DNA that may encode proteins) of the *c-myc* oncogene move from chromosome 8 to a section of chromosome 14 that is adjacent to the genes encoding the constant region of the antibody heavy chain. Alternatively (*b*), chromosome 8 can break at the first intron (segment of "nonsense" DNA, which is not transcribed into mRNA) of the oncogene, in which case only two exons move to chromosome 14. In other possible translocations the *c-myc* oncogene remains on chromosome 8 while genes encoding the constant region of the antibody light chain join it. In one such case (*c*) genes that code for constant regions of the "lambda" type are translocated from chromosome 22; genes from chromosome 2, which encode "kappa" constant regions, can also take part in such a translocation (*d*).

A FEW GOOD REASONS TO BUY THE NEW CANON PERSONAL COMPUTER.

The Canon Personal Computer runs all popular IBM PC compatible software.

If you're in the market for a smart, fast, modestly priced computer that's compatible with the IBM PC, look no further. Canon has just created the ideal one.

The Canon Personal Computer can run hundreds of programs designed for the IBM PC. No modification is required. (Even the layout of the keyboard is compatible with the IBM PC.)

And with its 8086 16-bit microprocessor, memory board and display adapter, the Canon Personal Computer can process complex data in record time.

It expands as you expand.

As your needs change, so can our Personal Computer. It offers you five expansion slots. A printer interface, asynchronous communication port and mini floppy disk controller are built in. So only one slot is taken up, with a display adapter. That leaves four expansion slots to use for optional hardware, including IBM PC compatible boards. And with Canon's memory expansion board, you can increase the standard 256K bytes to an enormous 512K bytes.

You also have the choice of either a monochrome or color graphic display unit; each with pedestals for easy adjustment. And, when you're ready, Canon offers a full line of quality printers designed to meet any need.

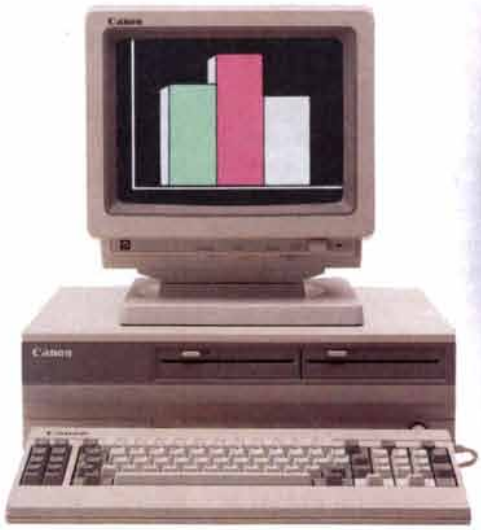
Our Personal Computer comes with Canon quality as a standard feature.

Perhaps the best feature our new Personal Computer has to offer can be summed up in a single word—Canon. Our worldwide reputation for quality and dependability has been earned in everything from sophisticated cameras to hard-working office machines.

So when you buy the Canon Personal Computer, you know you're not only getting a great computer, you're getting a great company to stand behind it.

How much does all of this cost? A lot less than you'd expect.

For more information, see your local computer dealer or call 1-800-323-1717, x325. (In Illinois, 1-800-942-8881, x325.)



Canon®

Canon U.S.A., Inc., One Canon Plaza, Lake Success, NY 11042.

SCIENCE AND THE CITIZEN

Patching Genes

Investigators will soon be ready to test recombinant-DNA therapies for correcting genetic disorders. Such techniques would be applied only to certain somatic cells (body cells, as opposed to eggs or sperm); they would not create inheritable alterations in the patient's DNA or affect the human gene pool. Nonetheless, the prospect of human gene therapy has raised some worries outside the scientific community.

As part of an effort to alleviate public concern, a working group within the Recombinant DNA Advisory Committee (RAC) of the National Institutes of Health has published a draft of a document titled "Points to Consider in the Design and Submission of Human Somatic-Cell Gene Therapy Protocols."

Among the diseases that might be amenable to gene therapy are disorders of the immune system that occur when the gene for a particular enzyme is deficient. To treat such a disease some of the patient's bone-marrow cells might be removed and infected with a virus into which a functioning gene has been spliced. Once such a gene has been incorporated into the marrow cells' genetic material, the cells would be reimplanted.

The draft document lists questions that will be considered by RAC when deciding whether or not to accept a proposal to test such therapy in human subjects. The list is a long one and includes many specific questions about the nature and objectives of proposed experiments. There are also quite detailed questions about the experimental methods to be followed, the theoretical and practical bases of such methods, and the safety procedures to be employed.

In addition the questions raise various ethical issues. "A major concern of the group," says panelist Robert F. Murray of Howard University, "is the protection of the subject from injury, abuse or exploitation." The panel considers informed consent an important issue. As with any new therapy, it will be difficult for both patient and physician to weigh the positive and negative aspects of receiving treatment.

Many of the panelists believe it is most important to ensure the privacy of subjects and their families. The working group met during the fall of 1984, when much publicity was being given to "Baby Fae," a human infant who received a baboon's heart, and to William J. Schroeder, the second person to receive an artificial-heart implant. Panelists are also concerned that institutions might exploit the new techniques for publicity.

Committee members stress that the draft document consists of questions to be considered, not strict guidelines that must be followed; they hope to use the questions as a basis for the evolution of a formal review procedure. The review process for the first proposals will be open to the public.

Several members of the working group that produced the document have said it is an attempt at "deliberate overkill"; the committee members do not in general think somatic-cell gene therapy presents novel or exotic hazards. Their objective is to create a review procedure that is "beyond reproach." Samuel Gorovitz of the University of Maryland, a member of the group, comments: "In my mind the most important single question involves devising a process that will engender justified public confidence in the process of review.... Even public fears that are irrational and unjust are real fears and must be responded to."

Ear to the Ground

The male frogs whose calls fill warm nights may have a quieter means of communication as well. Edwin R. Lewis of the University of California at Berkeley and Peter M. Narins of the University of California at Los Angeles report in *Science* that one species of frog appears to communicate not only vocally but also by means of vibrations in the soil.

Frogs are known to be acutely sensitive to substrate vibrations; a gentle footfall can quiet a calling male many meters away. In earlier research Lewis and his co-workers determined that the sensitivity lies in a structure of the ear known as the saccule, which probably functions in frogs, says Lewis, "as an inertial motion sensor." The saccule of frogs contains a large calciferous mass; Lewis believes that as soil vibrations are transmitted through a frog's body the mass is displaced less than surrounding tissue, setting up strains that are registered by nerve fibers.

The seismic sensitivity of frogs helps them to detect danger. Now Lewis and Narins have found it may also mediate communication. They write that white-lipped frogs, a species native to Puerto Rico, generate and respond to seismic signals of their own, a behavior never before reported in vertebrates.

The investigators made simultaneous recordings of sounds and substrate vibrations in the frogs' habitat. They found that when a male frog calls, it generates a seismic signal. As males chirp they crouch low, pressing their vocal sac to the soil. The vibration may result as the sac, expanding with each chirp, thumps the ground.

The thump is not merely an artifact of the call, however; it appears to act as a signal to other males. When the authors mimicked the thump by tapping the soil with a fingertip or a rubber mallet, frogs several meters away responded with the chuckling sounds that are believed to characterize interactions among white-lipped males.

What advantage do the animals derive from this second channel of communication? The vibrations propagate through moist ground at about 100 meters per second, a third the speed of sound. Lewis and Narins propose that the delay between a chirp and the accompanying seismic signal might indicate to other males the distance of the signaling male, thereby enabling the frogs to space their territorial claims.

New Order

An international team of metallurgists has reported the discovery of a new form of solid matter that is intermediate between a crystal and a glass. An alloy of aluminum and manganese, it may be the first real example of a quasicrystal, a hitherto unobserved arrangement of atoms proposed independently on theoretical grounds.

A crystal is a highly ordered solid structure in which the constituent atoms, ions or molecules are arranged in a characteristic pattern that is repeated periodically in three dimensions. Because the constituents tend to line up along certain axes, every crystal also has a characteristic long-range orientational, or directional, order, which manifests itself in a particular set of symmetries. A glass, in contrast, is solidified disorder: its constituents are distributed in a way that exhibits neither periodic nor orientational order. The newly discovered structure is distinguished by the fact that it has no periodic order and yet has a long-range orientational order.

The first experimental evidence of the existence of the new form of matter was found last year at the National Bureau of Standards by Dan Shechtman, a visiting Israeli investigator. Shechtman observed that when he bombard-

We're often asked why a Leica produces such superior photographs.

The answers may surprise you.

It is generally conceded, by professionals and serious amateurs alike, that Leica® cameras produce superior photographs. Photographs more vibrant, more alive, more three-dimensional than other photographs.

So it is tempting to say that this is simply because Leica produces superior cameras.

Which is true, of course. But hardly a complete explanation.

Because it is only when you fully understand the depth of our commitment to perfection that you can truly appreciate our extraordinary cameras.

Ten tons of glass.

Leica lenses are famous. The world's finest. So it won't surprise you to learn that we ourselves make the optical glass for them.

What will surprise you is how much glass it takes. And how few lenses we end up with.

Every year, our scientists use over ten tons of raw glass. Mixing and melting special formulations in pure platinum crucibles. Refining glass to ultimate optical standards. Discarding what others would consider acceptable. Scrutinizing every finished lens individually. Making sure it combines high color saturation with delicate color differentiation and unexcelled sharpness. Making sure it is perfect.

Ten tons of glass to make a few thousand lenses. No wonder no other manufacturer dares try to imitate us.

Never plastic.

And the mechanical components of our lenses are made just as exactingly. Critical moving parts are brass and aluminum, never plastic. Lens helixes are lapped to ten-micron tolerances, then polished by hand with jeweler's rouge. Thus eliminating the need for heavy greases used by other manufacturers. Greases which can affect focusing performance.

Body talk.

Which leads us to another surprise. Leica camera bodies are not designed and built just to enhance Leica lens performance. They're designed and built to enhance yours.

Again, take focusing. Focusing depends on the photographer's ability to see the image. So we've built the brightest viewfinder possible. Using silver instead of aluminum coatings on our prism. And seventeen dichroic coatings on our primary SLR mirror. To reflect maximum undistorted light.

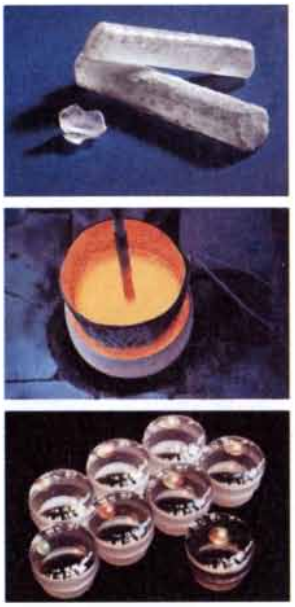
And for creative freedom, we've put an SLR exposure control at your fingertips. So you can switch instantly from manual or aperture priority with spot metering to program, shutter, or aperture priority with full scene integration.

The art beyond the science.

There are hundreds more examples we could give of Leitz's technical accomplishments. Of the specific superiority of Leica design and manufacture.

But at the very highest reaches of any discipline, there are areas beyond mere rational explanation.

And at Leitz,® a camera is more than just metal and glass and tests and formulations. It is pride and integrity and an almost fanatical dedication to perfection. It is skills and secrets passed on



from father to son. From generation to generation.

At Leitz, camera making is an art beyond a science.

But that shouldn't surprise you.

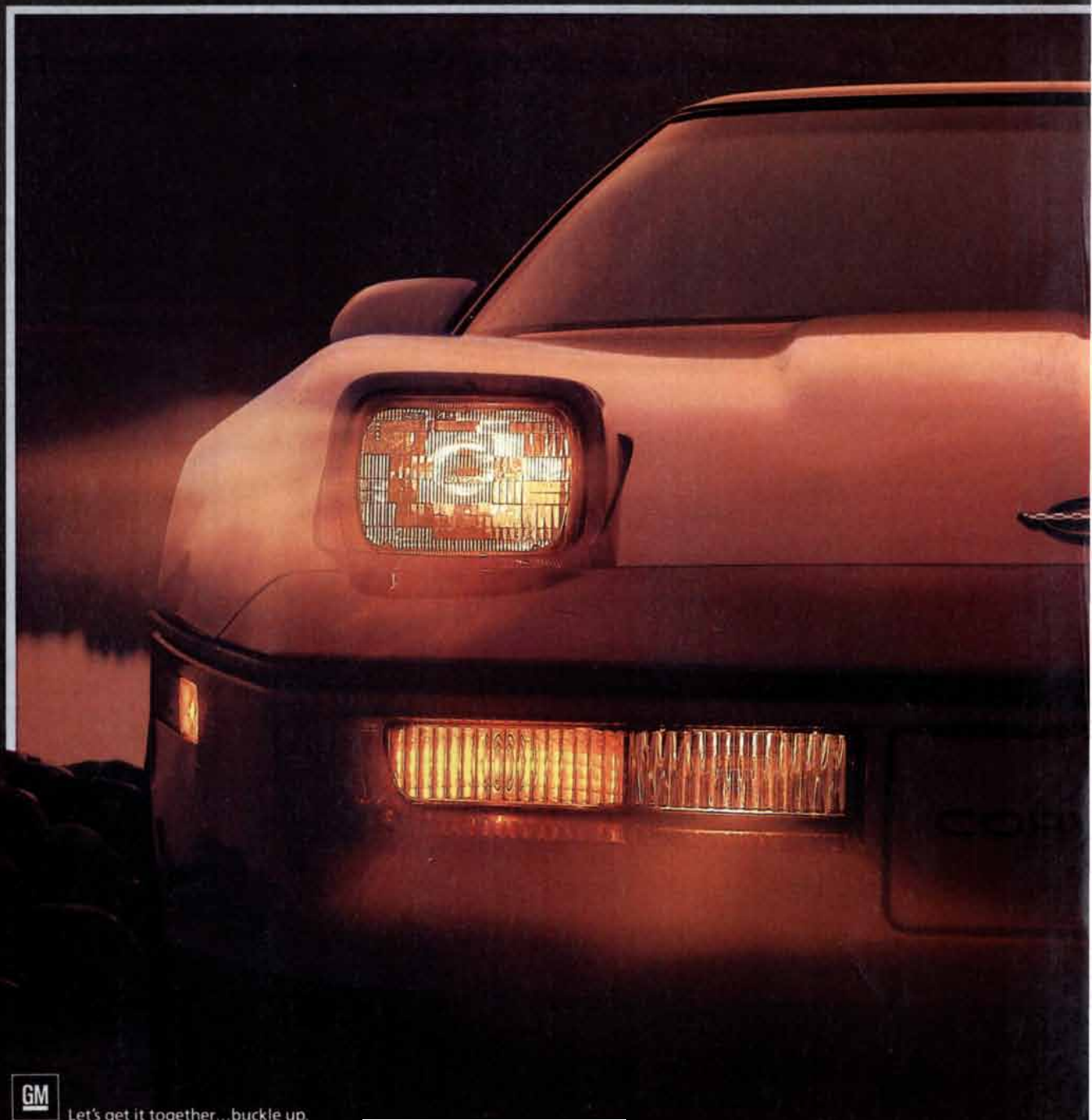
After all, so is photography.

For further information about Leica cameras call 1-800-223-0514 or write E. Leitz, Inc. 24 Link Drive Rockleigh, NJ 07647.

Leitz means precision. Worldwide.

CORVETTE

THE AMERICAN THAT WENT EYE TO EYE WITH EUROPE... AND EUROPE BLINKED.



GM

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

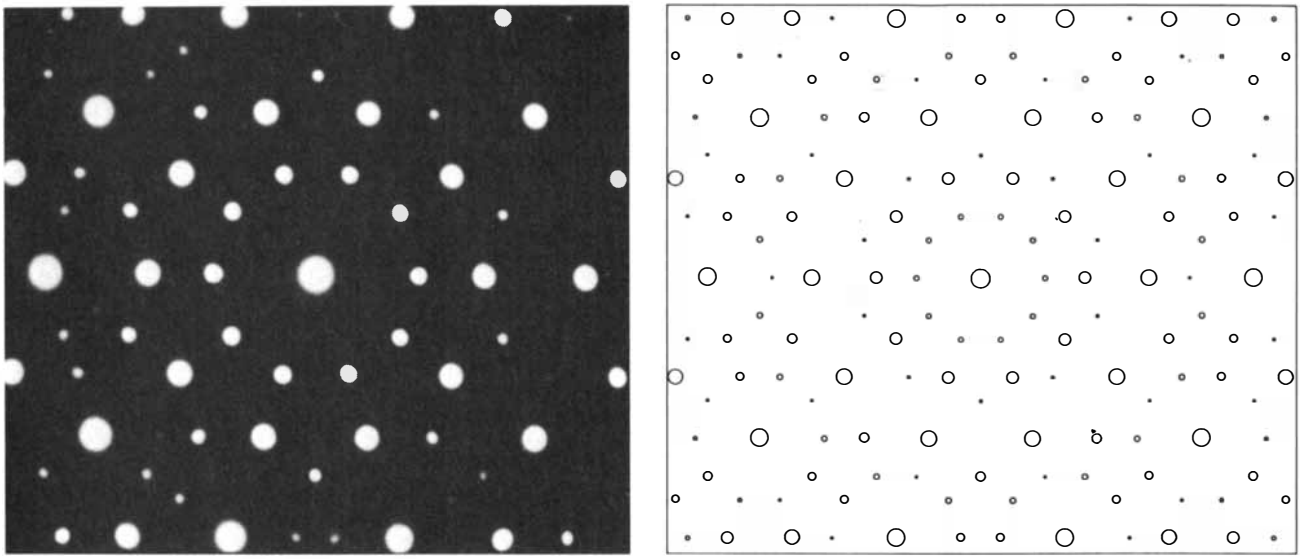
Exotic Europeans have traditionally dominated the high-performance sports car class. But in recent United States Automobile Club performance trials, the exotic American was the overall winner. The 1985 Corvette. So please join Ferrari, Porsche, Lotus and Lamborghini in a hearty welcome to a new world-class champion. The 1985 Corvette.

USAC Competitive Rank

	Corvette	Lamborghini Countach	Porsche 944	Ferrari 308 GTSi	Lotus Esprit Turbo	Porsche 928S
Total Points	21	18	14	11	11	9
Acceleration 0-60 (sec.)	4 (6.00)	6 (5.33)	1 (7.95)	3 (6.43)	5 (5.95)	2 (6.66)
Braking 60-0 (ft.)	6 (129.2)	3 (135.7)	4 (135.2)	2 (143.1)	1 (144.7)	5 (135.1)
Slalom (sec.)	6 (6.13)	3 (6.38)	5 (6.33)	4 (6.36)	2 (6.40)	1 (6.62)
Lateral Acceleration (g's)	5 (.91)	6 (.92)	4 (.86)	2 (.83)	3 (.85)	1 (.82)
Price as Tested	$26,703	$103,700	$26,121	$60,370	$50,384	$49,495

Scoring based on an Olympic system in which first place is awarded 6 points for each event. USAC certified tests, January 1985. All cars listed were latest models available for sale in the U.S. at time of testing and were equipped with various high-performance options. Corvette's Manufacturer's Suggested Retail Base Price is $24,891 including dealer prep. Tax, license, destination charges and optional equipment additional.

TODAY'S CHEVROLET



Evidence of nonperiodic order: actual electron-diffraction pattern **(left)**, *computed pattern* **(right)**

ed minute samples of a metallic solid composed of 86 percent aluminum and 14 percent manganese with an electron beam, the diffracted components of the beam produced sharp spots on a photographic film, similar to those obtained from a single crystal. Examination of the electron-diffraction patterns, however, revealed a peculiar symmetry inconsistent with any conceivable crystalline structure. Indeed, the appearance of a number of fivefold axes of symmetry in the diffraction patterns indicated an icosahedral configuration of the metal atoms that cannot be repeated periodically.

According to a report in *Physical Review Letters* by Shectman and three of his colleagues, Ilan Blech, Denis Gratias and John W. Cahn, the icosahedral phase forms during rapid cooling of the melt by a nucleation and growth mechanism that is characteristic of a direct transition from the liquid phase to the solid one.

As it happens, the possibility of an icosahedral phase of solid matter is implicit in the concept of quasicrystals, put forward recently by two theoretical physicists at the University of Pennsylvania, Paul J. Steinhardt and Dov Levine. According to Steinhardt and Levine, a quasicrystal, like a crystal, would have long-range orientational order; instead of being periodic, however, a quasicrystal would have a special atomic arrangement the two theorists describe as quasiperiodic. The three-dimensional quasicrystalline model, they note, is related to certain nonrepetitive, plane-filling arrangements of two-dimensional shapes, called Penrose tiles after their inventor, the British mathematical physicist Roger Penrose.

Writing in a subsequent issue of *Physical Review Letters,* Steinhardt and Levine comment on the striking correspondence between a computed electron-diffraction pattern of their hypothetical icosahedral quasicrystal and the actual electron-diffraction pattern found by Shechtman and his co-workers. "It is apparent," the theorists say, "that the atomic arrangements in the alloy must be closely related to the arrangement of lattice points in the quasicrystal." Referring to the recent experimental finding, they remark that "if real quasicrystalline materials exist . . . , they are sure to possess a wealth of remarkable new structural and electronic properties."

Healing Hormone

Within a few years the irreversible paralysis that often follows severe spinal injuries may be prevented in some patients by a drug administered at the scene of the accident. The drug, a synthetic analogue of the naturally occurring substance thyrotropin-releasing hormone (TRH), would be effective if the spinal cord has been crushed but not if it has been severed. There is no indication that TRH promotes the healing of old spinal injuries.

The prospect of a drug to prevent paralysis has been raised by work done by Alan I. Faden of the University of California at San Francisco Medical Center. Faden and his colleagues have shown that both natural TRH and a slightly modified synthetic analogue can prevent paralysis in laboratory animals if given within hours after a spinal injury.

Natural TRH is a peptide chain, three amino acid units long, that is secreted by the hypothalamus. The main physiological function of the hormone is to stimulate the pituitary gland to release thyroid-stimulating hormone (TSH). TSH in turn triggers the release of hormones from the thyroid.

It has been known for about a decade that much of the damage following a spinal trauma results from the body's response to injury rather than from the injury itself. The crushing of the spinal cord can lead to severe local damage of the gray matter. The surrounding white matter, on the other hand, may be more or less intact.

The gray matter contains horizontally oriented bodies of nerve cells and nerve fibers that carry commands to and from peripheral systems. The white matter consists of long vertical fibers that transmit messages to and from the brain.

Paralysis usually results from damage to the white matter. Hence if the white matter is intact, the patient will not immediately be paralyzed. Changes occurring in response to injury, however, can spread the damage to the white matter over a period of hours and lead to paralysis. Such changes can include the reduction of the flow of blood to the spinal cord.

The period of minutes or hours before the white matter is damaged constitutes a "therapeutic window" during which paralysis can be prevented, Faden said. His experiments show that TRH operates in the "window" by interrupting the normal response to trauma. The mechanism by which the peptide works is not known in detail. It is known, however, that TRH inhibits the activity of the endorphins, which are natural opiates that have a role in the complex postinjury response.

According to Faden, the optimum "therapeutic window" may last for no more than a few hours. Victims of spinal-cord injury typically reach specialized tertiary-care centers four hours after the accident, he notes. Hence any drug intended to prevent irreversible paralysis must be administered at the scene of the accident and must be potent enough to last until the patient reaches a specialized-care facility.

Natural TRH would probably not be an ideal substance for use in emergency care. The natural hormone is metabolized within five minutes and therefore to be effective it must be administered by continuous intravenous infusion. Faden has recently tested a synthetic TRH that is manufactured by the German company Grünenthal GMBH.

The synthetic analogue has been shown to be effective for at least several hours. Moreover, it is somewhat more potent than the natural hormone. Thus much of the clinical interest in TRH now centers on the artificial substance. Grünenthal is beginning trials of the synthetic analogue in Europe. In the U.S. Faden and his colleagues have recently completed tests of the safety of natural TRH in human beings and intend to begin trials of its clinical effectiveness soon.

Faded Genes

The evolutionary history of a species can be read in part from its DNA. Segments of the genetic material, extracted and purified from animal tissue, are compared with DNA from another species. The extent to which the two segments carry the same sequence of base pairs—constituents of the double strand of DNA—is a measure of how long ago the two species began to diverge. Workers in the laboratory of Allan C. Wilson at the University of California at Berkeley report in *Nature* that this method can be applied even when there is no living tissue from which to extract the DNA: they have managed to determine the sequence of two small segments of DNA from the quagga, a species of zebra that has been extinct since 1883.

The investigators, Russell Higuchi and his associates, did their analysis on a small, preserved sample of muscle taken from a quagga that died 140 years ago. The DNA Higuchi and his co-workers were able to extract came not from the nuclei of quagga cells but from mitochondria: constituents of a cell that contain their own DNA and are responsible for much of the cell's metabolic activity. An animal inherits its mitochondria, complete with DNA, directly from its female parent. Since each animal's mitochondrial DNA is derived from only one parent rather than from both, mitochondrial DNA is not subject to random reshuffling from one generation to another; consequently changes in mitochondrial DNA may act as a kind of "evolutionary clock."

Higuchi and his collaborators determined the genetic sequence of two segments of quagga mitochondrial DNA, 112 and 117 base pairs long. When they compared the sequences with corresponding ones in the mitochondrial DNA of a mountain zebra, they found the segments from the two species differed by a total of 12 base pairs; two of the differences would have caused the quagga DNA to encode amino acids different from those encoded by the mountain-zebra DNA. The quagga and the mountain zebra may have shared a recent ancestor.

Preliminary results with DNA from other zebras suggest that another species, Burchell's zebra, may be a still closer relative of the quagga. The Berkeley workers were trying to determine whether the quagga and the mountain zebra are more closely related to the true horse than they are to other zebras (Burchell's zebra and Grevy's zebra), as proposed by Debra K. Bennett of the University of Kansas. Higuchi believes his research is disproving Bennett's hypothesis.

Bacteriological Warfare

In the continuing controversy over alleged misuses of antibiotics one fact is generally assumed to have been established: bacterial resistance to antibiotics is increasing. That assumption was strongly challenged in a paper published last October in the *Journal of Clinical Microbiology.* Victor Lorian and Barbara A. Atkinson of the Bronx-Lebanon Hospital Center in New York concluded that overall bacterial resistance to antibiotics in the U.S. "remained virtually unchanged" during a recent 12-year period.

Their conclusion was based on statistical analysis of data for the years from 1971 through 1982, most of which had been collected (from between 150 and 329 hospitals, depending on the year) by a national monitoring service. The data, covering some 10 million bacterial samples isolated from both inpatients and outpatients, showed what percentage of the isolates of each bacterial species were susceptible to the effect of antibiotics in common use at the time.

Lorian and Atkinson chose to study eight species, which they picked on the basis of their frequency of isolation and ability to cause fatal infection. For each species they plotted the percentage of isolates found annually to be susceptible to a number of antibiotics. In the case of a few species a curve for susceptibility to a given antibiotic rises slightly and for a few species a curve drops significantly. In general, however, the curves are quite flat: the susceptibility of most organisms to most antibiotics appears not to have varied by more than a few percentage points over the 12-year period.

Commenting on the controversial practice of adding antibiotics to ani-

The quagga has been extinct since 1883; investigators have cloned some of its genes

Introducing connections for

With the new IBM Personal Computer AT/370, you can work with hundreds of VM programs and with mainframe data right at your desk. And only IBM offers a personal computer with VM capabilities.

Now that's a Smart Desk with very powerful connections.

With the PC AT/370, you have the power of 64-bit System/370 architecture and the ability to address up to eight megabytes of virtual memory. Plus access to FORTRAN 77, Pascal and other scientific languages.

The PC AT/370 lets you solve problems independently, only drawing on the host resources as needed. You can even record and analyze the results of an experiment or send them to the host for processing.

And you can enjoy the benefits of faster response time and greater data privacy—without having to compete for host computer time.

What's more, the PC AT/370 has a 20 or 40 megabyte storage capacity and the capability of running thousands of existing IBM PC programs. So this powerful connection is also a very versatile one.

One further point: If you already have a PC AT or XT, it can easily be upgraded to a PC AT/370 or XT/370 respectively. Ask your IBM marketing representative for the details.

And if you want to receive literature on how the PC AT/370 can turn your desk into a Smart Desk, call 1 800 IBM-2468, Ext. 428, or send in this coupon.

IBM®

3-85

IBM
DRM, Dept. KG/428
400 Parson's Pond Drive
Franklin Lakes, NJ 07417

☐ Please send me more information on the new IBM PC AT/370.
☐ Please have an IBM marketing representative contact me.

Name_____
Title_____
Company_____
Address_____
City_____ State_____ Zip_____
Phone_____

The Smart Desk from IBM

very powerful high-level thinkers.

mal feed, Lorian says: "I believe our data invalidate speculation that this practice could increase bacterial resistance in patients. I am afraid those concerned are trying to solve a problem that does not exist."

Lorian's paper and his conclusions are now being challenged by several investigators in medicine and microbiology. Their criticism is focused on the aggregated nature of the basic data. The susceptibility of most isolates of most species may indeed change only very slowly, some critics say; what really matters, however, is the increased resistance of particular species in specific places at certain times.

James M. Hughes of the U.S. Public Health Service's Centers for Disease Control cites his ongoing study of nosocomial infections (those developed by patients during a stay in a hospital). It shows that in certain species the level of resistance is higher in large teaching hospitals (which deal with many serious infections that need to be treated aggressively) than it is in smaller ones; the difference may be obscured when data from all hospitals are lumped together. Other critics maintain that the failure to distinguish between inpatients and outpatients obscures the high level of resistance in organisms causing nosocomial infections.

The data do not reveal which of the millions of isolates caused serious infections and which did not. Nor, Stuart B. Levy of the Tufts University School of Medicine points out, is any account taken of what may be most significant: the sudden appearance, in a single organism, of multiple resistance to several antibiotics. Levy also alleges that tracking susceptibility rather than resistance can in itself give a false picture. For example, a seemingly small decrease from 96 to 92 percent in susceptibility corresponds to a significant doubling of resistance, from 4 to 8 percent.

Minkowski's Jet

A thin, focused beam of ionized gas streaming from the core of the elliptical radio galaxy NGC 541 is triggering "a vigorous burst of star formation" in a nearby region of gas, according to Wil van Breugel of the University of California at Berkeley and his colleagues. The workers report their observations in a paper to be published in *The Astrophysical Journal*.

Stars form when clouds of gas and dust collapse. To start the process some external force, say the gravitational effect of a nearby galaxy, generally must provide a "push." Van Breugel's group has now shown that the impact of a cosmic jet, of which the plasma beam emanating from NGC 541 is an example, can also provide the needed force.

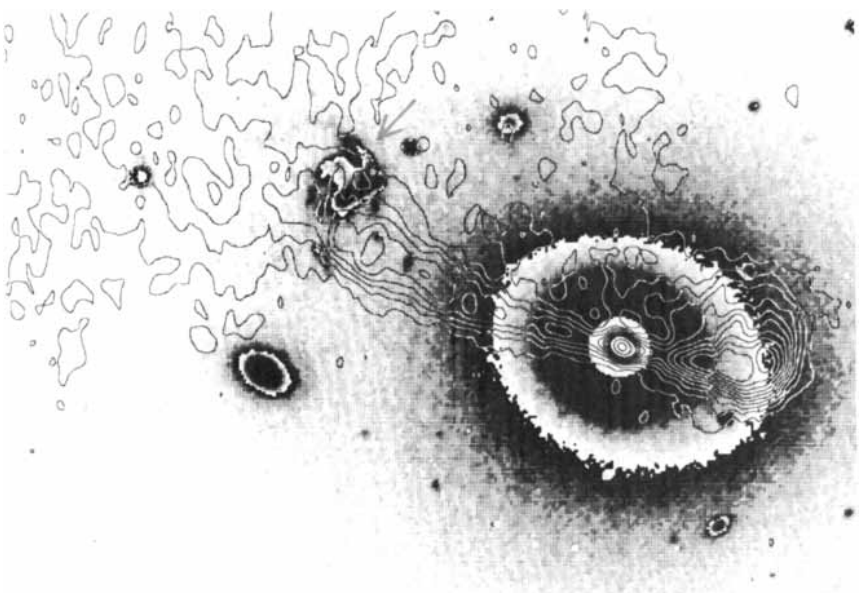
Van Breugel and his colleagues made their discovery in the course of their research on jets. As hallmarks of the violent activity at the center of many galaxies these structures have attracted considerable interest. They emit radio waves but are usually invisible at optical wavelengths. In the past few years, however, workers have found several regions of optical emission that are associated with jets. The hope has been that the investigation of these structures would reveal what is going on within the jets.

According to current theory, the optical emissions appear when a radio jet strikes a region of relatively dense gas. Clouds of gas are accelerated, heated and ionized in the turbulent boundary layers of the jet. As the gas is carried along it cools and radiates at discrete lines in the optical spectrum.

The "emission-line region" associated with the jet from NGC 541 is a spectacular one: a peculiar, bright blue structure called Minkowski's Object. Observations made by van Breugel's group with the Very Large Array radio interferometer in New Mexico and with several optical telescopes show that the radio and optical structures are indeed interacting. Minkowski's Object lies at the northeastern edge of the jet, and its optical emissions are most intense where it makes contact with the jet's boundary layers. "Downstream" from the contact point the velocity of the luminous gas increases, suggesting it has been accelerated by the fast-moving plasma; the jet's radio contours, in turn, are deflected and more diffuse.

Yet the radiation from Minkowski's Object differs from that of other emission-line regions. Its spectrum resembles the spectra of giant hydrogen clouds ionized by radiation from hot, young stars; specifically it resembles the spectrum of the galaxy NGC 7714, which is known to be undergoing an intense burst of star formation. The bright blue light emanating from Minkowski's Object is also characteristic of young stars.

Presumably the jet, in entraining the clouds of gas that make up the object, precipitates their collapse. Radiation from the new stars then ionizes the gas and gives rise to optical emissions. Van Breugel and his colleagues estimate that the total mass of young stars in Minkowski's Object is about 12 million times the mass of the sun—enough to qualify the region as a new galaxy, albeit an unusual one.



Superposed radio and optical images of NGC 541 jet and Minkowski's Object **(arrow)**

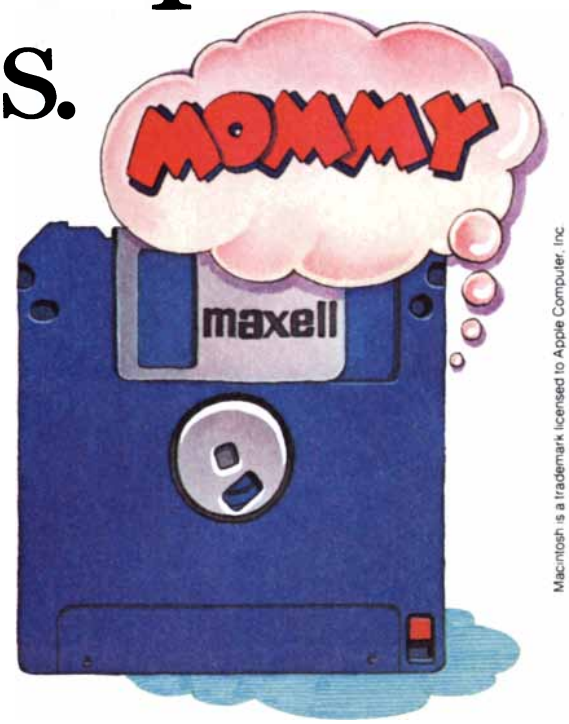
Social Diseases

Both retailers and physicians need to keep up with recreational fads—the retailer in order to nurture the current craze with appropriate clothing and accessories, the physician in order to recognize and treat the highly specific ailments to which almost any novel human activity gives rise. Take break dancing. In recent issues of medical journals a number of letters to the editor report more or less serious injuries sustained by young male practitioners.

In *The New England Journal of Medicine* physicians at two New York hospitals describe three cases of spinal injury, one of them a severe compression

Maxell Gold.

The 3½″ microdisk that gives Macintosh™ more to chew on and helps the HP® Touchscreen PC tap your deepest memories.



Whether you're in search of primeval wisdom, or polishing up financial projections, Maxell makes the Gold Standard microdisk for your computer. In fact, there's a Gold Standard for virtually every computer made. Maxell floppys are industry leaders in error-free performance. Each comes with a lifetime warranty. And each microdisk is perfectly compatible with your 3½″ drive.

maxell®
IT'S WORTH IT.

Maxell Corporation of America, 60 Oxford Drive, Moonachie, N.J. 07074

Macintosh is a trademark licensed to Apple Computer, Inc.

fracture and dislocation of the neck.

In a single issue of *The Journal of the American Medical Association* four letters give warning of break dancing's traumatic potential. Emergency-room physicians at a New York hospital report a neck dislocation sustained by a 15-year-old. It seemed the result of "a hyperflexion-rotation injury, caused by the weight of the body and shearing forces placed on the cervical spine" when he spun on his head.

Within a six-week period two teenaged break dancers came to a New Orleans hospital's emergency room complaining of acute pain in the scrotum. Thorough examination revealed they had suffered testicular torsion: twisting of the spermatic cord, which incorporates blood vessels and nerves along with the sperm-carrying vas deferens. Both patients required surgery to remove a damaged testicle.

In San Diego physicians who treated two patients within a week for painful swelling of the upper back checked with some colleagues on the prevalence of break-dancing injuries. Five of the 12 physicians they questioned told of recent cases: another spinal bruise, fractures of the forearm and of the collarbone, a neck dislocation, a torn knee ligament and severe sprains of the ankle and thumb.

A rather different injury is reported by a doctor in a New York City suburb who saw two cases of alopecia, or patchy baldness, in 17-year-old boys. Constant spinning on the top of the head "had eroded the hair shafts down to the scalp." The same correspondent notes a case of alopecia with a different recreational etiology. A young woman with hair loss in a strip across the top of the head was not a break dancer but a jogger with walkmania. Her earphone band was too wide and tight.

Joggers expose themselves to other risks, one of which may not be well known: they can be attacked by a bird of prey. In a letter to *The New England Journal* physicians in a community hospital in northwestern Switzerland tell of treating 12 male joggers over a two-year period for scalp scratches and lacerations. "The birds attacked by diving from behind and continuing to dive as long as the joggers were in motion." In five cases the assailant was identified as a European buzzard (*Buteo buteo*); almost all the attacks took place during the buzzard's breeding season. "Nature," the Swiss doctors comment, "may not allow intrusion without revenge."

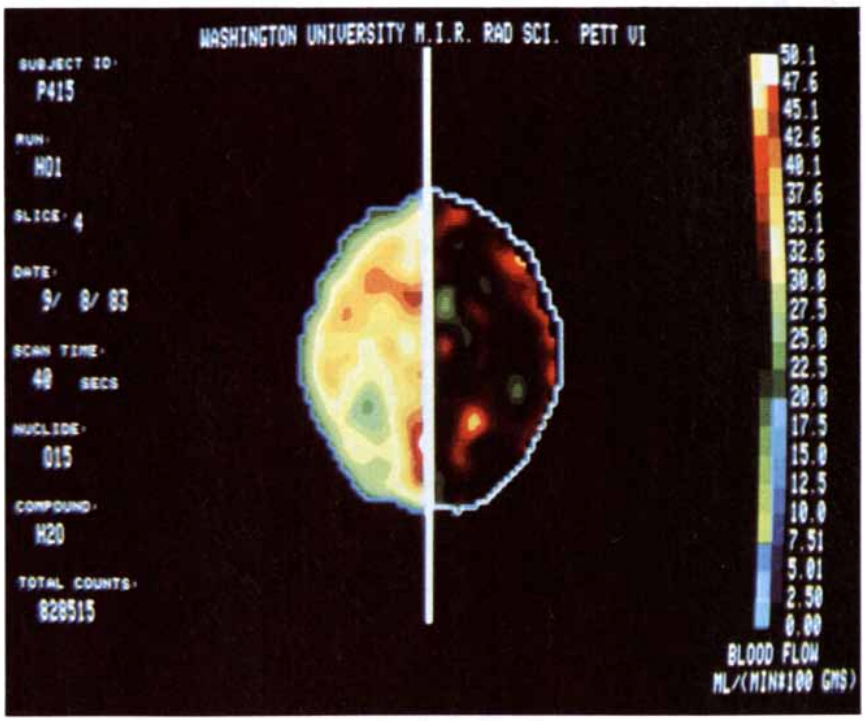
Why Panic?

Panic disorder, also called anxiety neurosis, is marked by recurrent attacks of fear in the absence of anything frightening. It afflicts from 2 to 5 percent of the general population (females twice as commonly as males) and from 10 to 14 percent of the patients seen by cardiologists. The victim feels severe anxiety, often accompanied by light-headedness, numbness and tingling in the hands or feet, a choking sensation and chest discomfort. The heart races or pounds. Severe trembling and sweating are also likely. The affliction is often regarded and treated as a psychiatric disorder or as a consequence of hyperventilation. Now a group at the Washington University School of Medicine has found what appears to be a physiological abnormality that correlates strongly with the illness. It is an unequal flow of blood to the parahippocampal gyrus, one of the regions of the brain believed to mediate panic, anxiety and vigilance.

The Washington University group consists of Eric M. Reiman, F. Kevin Butler and Eli Robins of the department of psychiatry and Marcus E. Raichle and Peter Herscovitch of the department of neurology and neurological surgery and the Mallinckrodt Institute of Radiology. The investigators took as their starting point the long-known fact that sodium lactate (a chemical related to the lactate, or lactic acid, that builds up in the blood after exercise) can precipitate the symptoms of panic attack when it is infused into the bloodstream of people susceptible to the disorder.

They gave volunteers a radioactive marker so that an image of the brain could be obtained with a positron emission tomography (PET) scanner. As Raichle puts it, the brain scans "showed a marked difference in blood flow between the right and left sides of a specific part of the brains of patients in whom lactate was known to provoke a panic attack." People without a history of panic disorder do not show the abnormality.

Describing the work in *Nature,* the Washington University investigators said they believe it is "the first study to identify a discrete brain abnormality in patients with this severe form of anxiety." Raichle notes that "the difference in blood flow between hemispheres probably correlates with differences in metabolic rates. Although you might suppose that this difference is a long way from an emotional response, the activity of nerve cells is closely tied to brain blood flow because the brain stores no oxygen." A promising site for further research, he said, is the blood-brain barrier: a network of membrane-enclosed capillaries in the brain that prevents many of the substances in the blood from reaching cells in the brain. "It's possible that the blood-brain barrier in persons with panic disorder is defective, allowing lactate to reach brain cells. In addition we will be measuring oxygen utilization, which is the link between blood flow and the activity of nerve cells."



Unequal flow of blood in the brain of a victim of panic disorder

What good is a car that could make it to the next century, if you can't get past the first payment.

The new Volkswagen Jetta offers the longevity European road cars are famous for. At a price mere mortals can afford. It's sporty, roomy and as reliable as a VW. We tested it through 3.7 million miles of hell and high water. To prove its durability, we back it with our new 2-year Unlimited-mileage Protection Plan*.

Jetta. Because a car can't get you to the next city, let alone the next century, if you can't afford the car.

The new Jetta $7,995.** **It's not a car. It's a Volkswagen.**

*Protection Plan: 2-year unlimited-mileage, limited warranty on entire car except tires; 3-year unlimited-mileage, limited warranty on corrosion perforation. See U.S. dealer for details. **Mfr's sugg. retail pricing, excluding tax, title, dealer prep and trans. $7,775 for 2-door model. $7,995 for 4-door model as shown. (Metallic paint, alloy wheels are extra cost options.) **For details call 1-800-85-VOLKS.**

Seatbelts save lives.

The Hidden Dimensions of Spacetime

Spacetime, usually thought of as four-dimensional, may have as many as seven extra dimensions. Eleven-dimensional structures now under study might give a unified account of the four basic forces of nature

by Daniel Z. Freedman and Peter van Nieuwenhuizen

On May 29, 1919, the shadow of a total eclipse of the sun swept across the Atlantic from western Africa to northern Brazil. Expeditions mounted by the British government at the instigation of Sir Arthur Stanley Eddington were ready to observe the stars near the darkened disk of the sun. One of Eddington's main objectives was to test a new theory of gravity put forward by Einstein four years earlier. The theory is best known by the name general relativity. In it Einstein advanced the startling intellectual claim that the geometry of the universe is determined by the matter and energy within it. More precisely, according to general relativity, space and time are knit into a four-dimensional mathematical structure called spacetime. The force of gravity is understood as an effect of the so-called intrinsic curvature of spacetime.

The observers of the eclipse set out to test one of the predicted observational effects of Einstein's curved spacetime. According to general relativity, the path of light from stars in the vicinity of the sun would be curved by the sun's gravitational pull. When the disk of the sun moved near a star, the star would appear to be shifted from its usual position in the sky. A test of the theory had to await a solar eclipse because only then could stars be seen near the sun. The observations of the eclipse made Einstein world-famous. The stars were shifted by just the predicted amounts, and so the success of Einstein's geometric approach to gravity was dramatically confirmed.

Although general relativity is concerned with geometry in only four dimensions, Einstein's imaginative work opened the door to even bolder applications of his basic idea. In the same year the concept of a four-dimensional universe was being confirmed by observational astronomy, Theodor Franz Éduard Kaluza, a virtually unknown privatdocent, or junior scholar, at the University of Königsberg in what is now the city of Kaliningrad in the U.S.S.R., dispatched a paper to Einstein. In the paper Kaluza proposed that the four dimensions of spacetime be supplemented with a fifth, spatial dimension.

Kaluza's aim in introducing a fifth dimension was to give a unified account of all the known forces of nature. There were only two fundamental forces recognized at the time: gravitation, which was described by general relativity, and electromagnetism, which was described by the theory of James Clerk Maxwell and others. The two forces appear to be profoundly different. For example, all particles are subject to gravity but only charged particles are subject to electromagnetism. In 1914 Gunnar Nordström of Helsingfors (now Helsinki) University had tried to give a unified description of the two seemingly distinct forces by showing that both arise from a five-dimensional form of electromagnetism. Nordström's approach had to be abandoned because it could not explain the bending of light near the sun. Kaluza showed that the two forces can both emerge from a five-dimensional version of general relativity.

In the past decade many physicists have taken a renewed interest in Kaluza's geometric program for unifying the forces of nature. For the current program geometric structures in even more than five dimensions must be considered because four forces are now known instead of two. The two additional forces are the strong nuclear force, which binds protons and neutrons together within the atomic nucleus, and the weak nuclear force, which is responsible for certain kinds of radioactive decay. Moreover, it is now recognized that quantum-mechanical effects cannot be omitted from any scheme for unification. One of the most exciting recent developments of the program is a version of a theory called supergravity. Although there are, within supergravity, several possibilities for the number of dimensions of spacetime, the theory is most elegant mathematically when it is formulated in 11 dimensions.

Why does one need 11 dimensions? The number arises from a curious mathematical coincidence. Theories of supergravity can be formulated in any number of spacetime dimensions up to 11, but in 12 or more dimensions the theory seems to break down. On the other hand, seven is the smallest number of hidden dimensions needed to accommodate the three nongravitational forces into a theory like Kaluza's. Taken together with the four dimensions of ordinary spacetime, the seven hidden dimensions would lead to an 11-dimensional universe. It is remarkable that the mathematical requirements for supergravity coincide with the physical constraint imposed by the description of the forces.

General Relativity

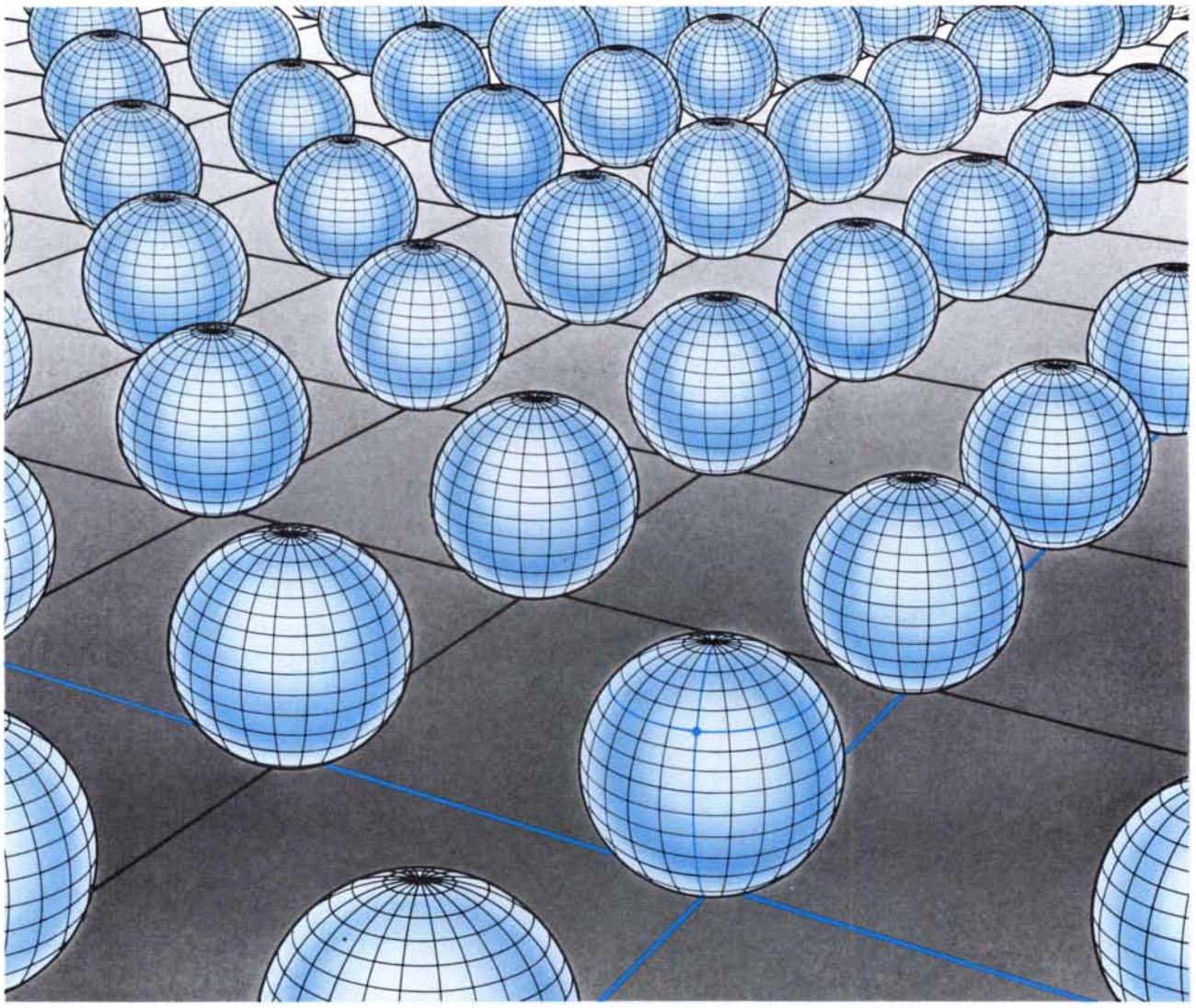
Einstein's general theory of relativity is the crowning achievement of classical physics. In its essence supergravity, like every other theory that is based on Kaluza's geometric ideas for unifying the forces of nature, is an extension of the ideas of general relativity. Einstein proposed general relativity after nine years of searching for a theory of gravitation. The theory sought was to be in agreement with his theory of special relativity and with the experimental observation, known since Galileo, that all bodies follow the same trajectory in a gravitational field. Einstein reasoned that since the trajectory of a freely falling body does not depend on its mass or on its internal composition, its motion under gravity must be related to the properties of spacetime itself. He then showed how to interpret the force as a manifestation of a property of spacetime called its curvature.

To better appreciate this point of view, imagine the curved surface of a sphere. The surface is two-dimensional because one must give two coordinates, such as the latitude and the longitude, in order to specify a point. The shortest path that connects two points on the sphere and lies entirely on the surface is the shorter arc of the great circle that passes through the points. This basic geometric fact is constantly applied in determining the most efficient air routes over the earth. One can also imagine a rippled surface that is more complicated than the sphere, but again there is a path of shortest distance on the surface that connects any two points. Such a path is called a geodesic, from the Greek words meaning division of the earth.

In general relativity spacetime is a four-dimensional analogue of a rippled surface. It is four-dimensional because four coordinates must be given to specify a point. A point in spacetime can be a physical event, such as a collision between two particles. The event is specified by where and when it happens: by its three spatial coordinates and its time. A geodesic in spacetime is the analogue of a geodesic on a surface. It is a path in spacetime between two events that is singled out by the geometry of spacetime. According to general relativity, any particle acted on solely by the force of gravity follows a geodesic in spacetime. In this way general relativity explains Galileo's observation that all freely falling bodies follow a common trajectory.

Kaluza's Unified Theory

Because Kaluza's description of unified forces adopted the approach taken in general relativity, he sent his article to Einstein for a recommendation. At the time an article could be published only if it had been endorsed by a well-known physicist, and from his position as privatdocent Kaluza had little influence and only the money from small fees paid by the students attending his lectures. Einstein, who had also begun



SEVEN HIDDEN DIMENSIONS of the universe, which have been proposed in a theory that seeks to unify the forces of nature, can be pictured as a small, compact structure such as a sphere that is associated with every point in space and every moment in time. In Einstein's general theory of relativity space and time are combined into a four-dimensional structure called spacetime. Astronomical observation shows that on a large scale spacetime has a nearly flat, or Euclidean, geometry. The plane in the illustration represents the geometry of ordinary spacetime; coordinate lines along one axis represent space and coordinate lines along the second, perpendicular axis represent time. The spheres shown at the intersections of the coordinate lines represent the hidden, curled-up dimensions postulated in the new theory. The illustration can only suggest the appearance of the postulated structure. The spheres should be imagined as tangent to the plane at every point. Moreover, the spheres and the plane actually give rise to only four dimensions instead of 11. The four dimensions are the four coordinates whose values must be given in order to specify a point (*color*).

his career as a privatdocent, was immediately fascinated by the article, but in a series of letters to Kaluza he suggested that certain problems of the theory be studied further before publication. Two and a half years later Einstein changed his mind, and he sent Kaluza a postcard in which he proposed to recommend publication. The article appeared in the journal *Sitzungsberichte der Berliner Akademie* in 1921, with the title "On the Problem of Unification in Physics."

The search for a unified description of all seemingly unrelated physical phenomena has always been an important theme of scientific investigation. As we have mentioned, in Kaluza's theory the ordinary forces of gravity and electromagnetism arise from a five-dimensional version of general relativity. In order to account for the fact that five dimensions are not observed, Kaluza simply assumed that quantities such as curvature do not depend on the fifth coordinate. Particles follow geodesics in five dimensions, but these paths appear in four dimensions as the paths of particles subject to the combined forces of gravity and electromagnetism.

From the present perspective the most apparent shortcoming of Kaluza's theory is that gravity and electromagnetism are not the only fundamental forces in nature. The strong nuclear force and the weak nuclear force had not been recognized in 1919 because they act only over a short range comparable with the diameter of the nucleus. Accelerators that can probe dynamic processes at such short distances had not yet been constructed.

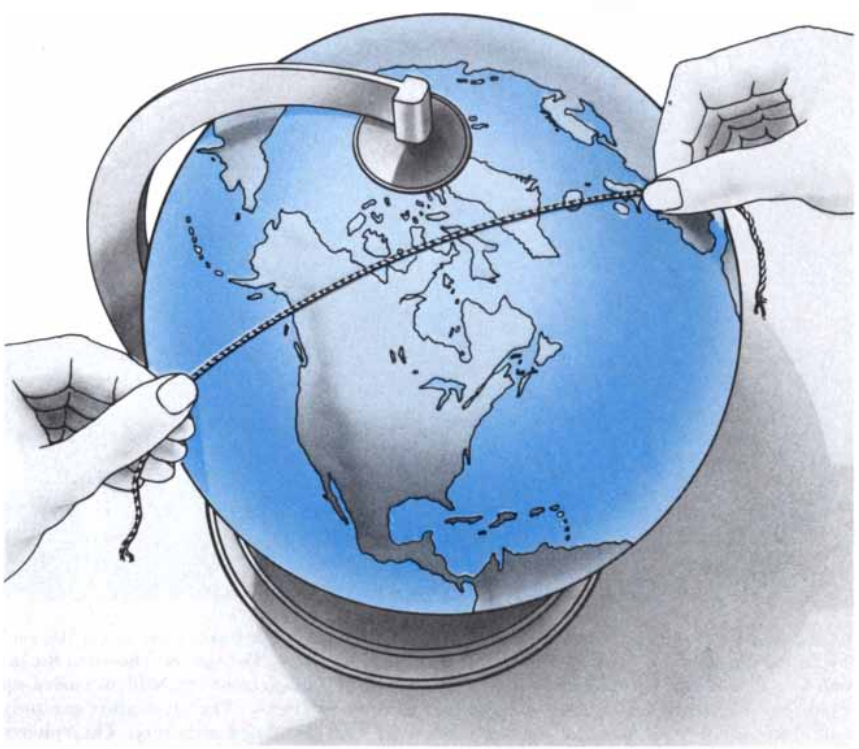
At the time of publication of Kaluza's paper, however, there was a more prominent shortcoming of the theory: it failed to address itself to an important set of phenomena that are now understood as quantum-mechanical effects. Kaluza was well aware of this deficiency. In the conclusion to his paper he wrote: "Every [classical, or deterministic and mechanistic, theory] that claims universal validity is threatened by the sphinx of modern physics, quantum theory." Nevertheless, in Kaluza's theory as well as in Einstein's general theory of relativity a classical view of the world is taken for granted.

According to the classical view, all physical objects including the smallest elementary particles act like bullets subject to one or more of the fundamental forces. For large-scale phenomena the classical view works quite well, but it fails totally to account for processes on the atomic scale. By 1919 many of the failures of classical explanations for atomic and subatomic processes had become apparent.

Historically the most important failure of classical physics was its account of the atomic spectrum. Experiments show that atoms emit light in discrete spectral lines, corresponding to a set of frequencies, or colors, that are a characteristic of the emitting atom. According to the classical theory, however, an atom should emit light of all frequencies because in that theory electrons orbiting the atom must spiral in continuously toward the nucleus. Moreover, in the classical picture the spiral path of the electron would quickly lead to the collapse of the atom, and so matter as we know it could not exist.

The resolution of this conundrum and other difficulties led to the development of quantum mechanics. In quantum mechanics the strict determinism of the classical theory is abandoned. The spiral paths of electrons around the nucleus are replaced by wave patterns in spacetime. The intensity of a wave pattern determines the probability of finding an electron at a particular point.

Wave patterns that are stationary correspond to long-lived states of motion of the electron, and each state of motion has a characteristic energy. Light is emitted at discrete frequencies, which correspond to discrete spectral lines, when the electron makes a sudden jump from one state to another. The state of motion corresponding to the lowest allowed energy is stable, and so atoms do not collapse in the quantum theory as they would according to classical physics. The wave patterns of the electrons arise as solutions to a differential equation formulated by Erwin Schrödinger, in which both time and the three spatial coordinates are treated as variables.



GEODESIC ON A SPHERE is the shortest distance between two points that can be measured on its surface. On a globe a geodesic is the shorter arc of the great circle that passes through the two points. It can be determined by holding a taut string between the points.

The Fifth Dimension

In 1926, still very early in the quantum era, the Swedish physicist Oskar Klein set out to determine whether or not quantum mechanics is compatible with Kaluza's five-dimensional theory. Klein wrote down a version of Schrödinger's equation having five variables instead of four. He showed that the solutions of the equation can be interpreted as waves moving in gravitational and electromagnetic fields of ordinary, four-dimensional spacetime. In quantum mechanics the waves can also be interpreted as particles. The name Kaluza-Klein theory is now given to all theories, within a quantum-mechanical framework, that attempt to unify the fundamental forces of nature in a spacetime having more than four dimensions.

In the original papers of Kaluza and Klein it is not clear whether the fifth dimension is to be understood as physically real or merely as a mathematical artifice necessary to obtain gravity and

electromagnetism in a uniform manner. The introduction of quantum mechanics, however, suggests plausible answers to several important questions about the physical reality of an extra dimension. In what way could the new dimension be physically real? Why has such a fundamental aspect of the universe gone undetected up to now? How could one discover the extra dimension experimentally?

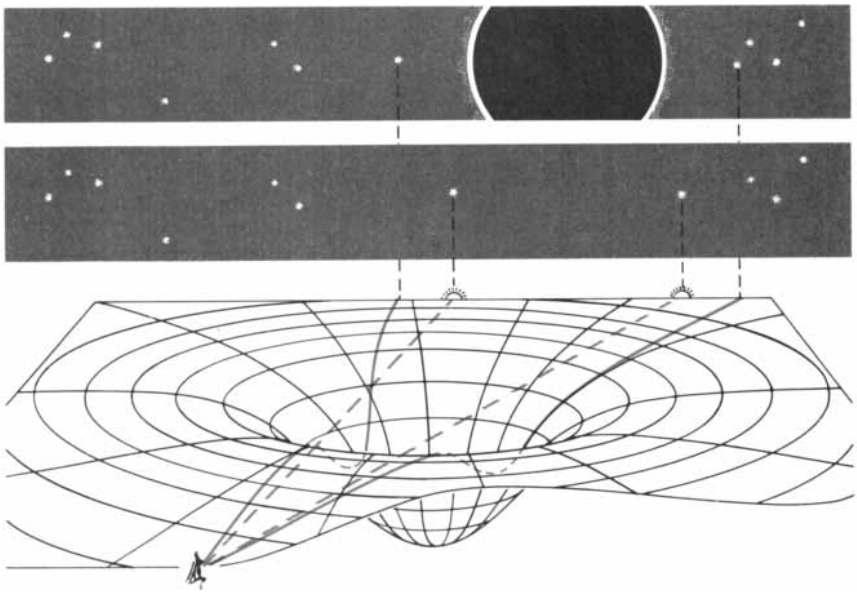
To begin to answer these questions, imagine a line of indefinite length to which a small circle is associated at every point. If the circle is actually constructed at every point along the line, the resulting structure is a cylinder of indefinite length. The one-dimensional line and the one-dimensional circle can be said to generate the two-dimensional cylinder.

In a similar way one can generate a four-dimensional structure from the two-dimensional plane and the two-dimensional sphere. The new structure can be understood as a plane on which a sphere is erected at every point. It is four-dimensional because specifying a point in the plane and specifying a point on the sphere both require two coordinates [*see illustration on page 75*].

The line and the plane in the previous two examples represent the nearly flat geometry of the four-dimensional spacetime in which we live. The circle and the spherical surface represent the extra dimension or dimensions of a higher-dimensional spacetime. A five-dimensional spacetime can be understood as the structure generated by a circle and ordinary four-dimensional spacetime; one possible structure of a six-dimensional spacetime is generated by ordinary spacetime and the surface of a sphere. In these structures there is a circle or a sphere associated with every point in space and every instant of time.

One can now explain how the fifth dimension of spacetime might be physically real in Kaluza's theory but still have gone undetected up to now. A fundamental concept of quantum mechanics is Werner Heisenberg's principle of uncertainty. Any particle can be thought of as a packet of waves spread over a certain region of space. According to the uncertainty principle, the minimum size of the region depends on the energy of the particle: the greater the energy of the particle, the smaller the minimum size of the region.

To detect a small spatial structure one must use a microscope. A microscope is essentially an instrument that illuminates a structure with photons of light, electrons or beams of some other particle. The resolution of the microscope is the minimum size of the region that can be illuminated, and so because of the uncertainty principle the resolution depends on the energy of the particles in the illuminating beam. Particles of progressively higher energies are needed to see structures of progressively smaller size.



BENDING OF STARLIGHT passing close to the sun is an effect predicted by Einstein's general theory of relativity. According to the theory, the geometric structure of spacetime in the vicinity of the sun is curved by the sun's mass in a way suggested by the curved set of coordinate axes shown in the diagram. Light must follow a geodesic in spacetime, and so the lines of sight to the stars are curved when the stars are near the disk of the sun in the sky (*color*). When the stars are viewed during a solar eclipse, they appear to be shifted away from the sun. The broken black lines indicate lines of sight when the sun is not nearby.

Suppose the fifth dimension were curled up into an exceedingly small circle. To detect the circle the energy of the particles illuminating it would have to be sufficiently high. Particles with too little energy would be effectively spread uniformly over the circle, and so it could not be detected. The most powerful accelerators today produce particles energetic enough to resolve structures 10^{-16} centimeter across. If the circle in the fifth dimension is smaller than 10^{-16} centimeter, it could not yet have been resolved.

Massive Particles

There is a more indirect way the presence of a fifth spatial dimension could be inferred. Just as stationary wave patterns in the atom correspond to long-lived states of motion of the orbiting electrons, so stationary waves on the circle of the fifth dimension correspond to particles that could be observed in the laboratory. The stationary wave patterns must fit evenly into the circumference of the circle. Hence the wave either must have a constant amplitude, or an integral number of oscillations must cover the entire circle: one, two or three oscillations and so forth [*see illustration on page 80*].

The mass of each observable particle depends on its wavelength, which is the circumference of the circle divided by the number of oscillations the wave makes around the circle. The shorter the wavelength, the greater the energy of the wave and the higher the mass of the associated particle. In Kaluza's theory the particles with the lowest mass are the ones whose associated wavelength is infinite; in other words, the amplitude of the wave in the fifth dimension is constant. Such particles have zero mass.

The first massive particle in the theory is the one whose wavelength is equal to the circumference of the circle. Its mass is therefore inversely proportional to the circumference. The mass of the second heavy particle is two times the mass of the first, because it corresponds to the wavelength that fits exactly two times into the circumference of the circle. Similarly, the other allowed patterns of stationary waves on the circle generate a series of particles whose masses are integral multiples of the mass of the first heavy particle.

There is an argument given by Klein that leads to an estimate of the mass of the first heavy particle. Since Kaluza's theory attempts to unify the forces of gravity and electromagnetism, the first heavy particle also has an electric charge that is inversely proportional to the circumference of the circle. On the other hand, the charge of all observed elementary particles is an integral multiple of the charge of the electron. If

one assumes the first heavy particle carries that charge, its mass can be calculated. The answer is astonishingly large: the mass is 10^{16} times the mass of the proton, which is heavier than 10,000 bacteria. There is no way that present or prospective accelerators could generate such particles, but they may have been produced in the big bang. Most of them would have decayed by now, but some of them may still be detectable.

Since the massive particles in Kaluza's theory are so heavy, the only particle in the theory that could correspond to the particles observed today is the particle of zero mass. It is now known, although it was not appreciated at the time the theory was formulated, that more subtle quantum-mechanical effects can lead to a finite, nonzero mass for the particle predicted by the theory. Thus the massless particle in Kaluza's theory and other massless particles in generalizations of the theory can account, at least in principle, for the particles that are observed.

The circumference of the circle in the fifth dimension that could give rise to the massive particles predicted by the theory is correspondingly small: about 10^{-30} centimeter. To resolve a structure of such small size with an instrument based on current technology would require an accelerator several light-years across.

After Klein's investigations and subsequent work by Einstein and by Wolfgang Pauli, there was little progress made on Kaluza's basic idea of unification until the late 1970's. Indeed, until then much of the modern work on the unification of forces was based on a strategy that does not call for a higher-dimensional spacetime. The strategy can be traced to a different proposal for the unification of gravity and electromagnetism made by the German mathematician Hermann Weyl in 1918. The central idea of Weyl's theory was that the description of a force is not altered by any changes in the length scales of the rulers or the time scales of the clocks that are carried as measuring instruments to various points in spacetime. This principle is called gauge invariance, after the gauges, or measuring instruments, to which Weyl referred. Such a theory is called a gauge-field theory, or, for short, a gauge theory.

Electroweak Unification

Weyl's own theory did not give a physically correct account of gravity, and it has largely been abandoned. Nevertheless, the principle of gauge invariance has become the pivot of modern theories of elementary particles. In 1954 C. N. Yang of the State University of New York at Stony Brook and Robert L. Mills of Ohio State University developed a class of gauge theories known as non-Abelian gauge theories. The theories make an important generalization of Maxwell's theory of electromagnetism, and the mathematical theory of symmetry groups assumes a central role. In group theory one studies operations, such as the rotations and mirror reflections of a solid object, that leave the object unchanged in appearance. For example, the appearance of a sphere is not changed by any rigid rotation about its center, and the group that expresses this symmetry mathematically is called $SU(2)$.

Many theoretical physicists have studied non-Abelian gauge theories. In 1967 Steven Weinberg, now at the University of Texas at Austin, Abdus Salam of the International Centre for Theoretical Physics in Trieste and John C. Ward, now at Macquarie University in New South Wales, applied several important contributions by Peter Higgs of the University of Edinburgh, Sheldon Lee Glashow of Harvard University and others to show that a non-Abelian gauge theory could unify the electromagnetic force and the weak nuclear force. Some predictions of the theory, which is called the electroweak theory, were confirmed experimentally in the early 1970's, but the most spectacular evidence was found in 1983 at CERN, the European laboratory for particle physics. In that year three particles, the W^+, W^- and Z^0 vector bosons, were discovered having exactly the masses predicted by the electroweak theory.

The success of the electroweak theory led theoretical physicists to propose another non-Abelian gauge theory called quantum chromodynamics, which can describe the strong nuclear force. In this theory the proton and neutron are made up of more fundamental particles called quarks. The strong force arises from interactions of the quarks with eight vector bosons known as gluons. Quantum chromodynamics also appears to be confirmed experimentally.

Although the electroweak theory and quantum chromodynamics are rather different gauge theories, the three forces they describe can be further unified by bringing both components into a single non-Abelian gauge theory based on a larger mathematical symmetry group. Such theories are called grand unified theories. Their predictions have not yet been experimentally confirmed, but the ideas are so attractive that many physicists think some version of the grand unified theories will give the correct, unified account of the strong, weak and electromagnetic forces.

What is missing from the grand unified theories is the force of gravity. Hence it is natural to ask whether the

April 21, 1919

The idea that the electric field quantities are mutilated...has also frequently and persistently haunted me. The idea, however, that this can be achieved through a five-dimensional cylinder-world has never occurred to me and would seem to be altogether new. I like your idea at first sight very much....

If no apparently fatal and irreparable objection occurs to me when reading your detailed exposition, I will be pleased to present your paper on the subject to the academy here.

April 28, 1919

I have read through your paper and find it really interesting. Nowhere, so far, can I see an impossibility. On the other hand, I have to admit that the arguments brought forward so far do not appear convincing enough. I would like to suggest considering the following (perhaps before you publish your paper, even though I do not like to permit myself to counsel you in this matter).

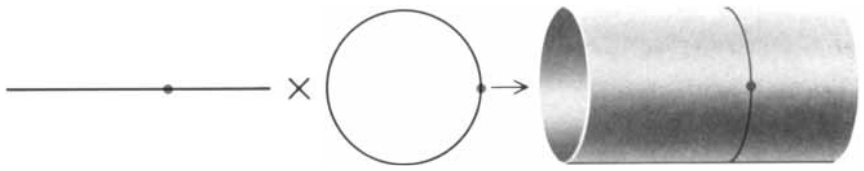
According to your basic idea, it should be assumed that geodesic lines which are oblique to the sections...should give the trajectories of electrically charged particles under the simultaneous action of the gravitational and electric field. *If you could show that this occurs with the accuracy guaranteed by our empirical knowledge,* I would be as good as convinced of your theory.

October 14, 1921

I am having second thoughts about having restrained you from publishing your idea on a unification of gravitation and electricity two years ago. Your approach seems in any case to have more to it than the one by H. [Hermann] Weyl. If you wish I shall present your paper to the academy after all, provided you send it to me.

LETTERS FROM EINSTEIN to Theodor Franz Éduard Kaluza show the evolution of Einstein's response to Kaluza's ideas. The dates show it was more than two years before Einstein recommended the publication of Kaluza's paper, which finally appeared in 1921.

grand unified theories can be joined together with gravitation as a Kaluza-Klein theory in higher dimensions. Kaluza's original theory required five dimensions because it included only one vector boson, namely the photon associated with the force of electromagnetism. The weak nuclear force requires the three recently discovered vector bosons, the strong nuclear force requires the eight gluons, and grand unification requires between 10 and 500 more vector bosons. The exact number of additional vector bosons depends on which version of the grand unified theory is adopted.



KALUZA'S THEORY regards the fifth dimension as a circle associated with every point in ordinary spacetime. If one dimension of ordinary spacetime is represented as a line, an analogue of the five-dimensional structure proposed by Kaluza can be visualized. It is a line to which a circle is associated at every point, or in other words a cylinder. A circular cross section of the cylinder represents the structure of empty, five-dimensional spacetime.

Modern Kaluza-Klein Theories

Although there is not a one-to-one correspondence between the number of vector bosons needed and the number of dimensions, it is roughly correct to say that more vector bosons require more spacetime dimensions. Thus the inclusion of strong and weak forces in the Kaluza-Klein framework would require a spacetime of even more than five dimensions. The extra dimensions could be physically real and yet unobserved provided they curl up into a higher-dimensional "surface" analogous to the circle in Kaluza's theory or to the surface of a sphere.

The recent attempts to include the strong and weak forces in a Kaluza-Klein theory began with the work of Bryce S. DeWitt of the University of Texas at Austin, Y. M. Cho of Seoul National University, Peter G. O. Freund and Mark A. Rubin of the University of Chicago, Eugene Cremmer, Bernard Julia and the late Joel Scherk of the University of Paris and John H. Schwarz of the California Institute of Technology.

The first problem for modern Kaluza-Klein theories is the number of extra dimensions that must be included. Because there is not yet a consensus about which version of the grand unified theories is correct, the number of vector bosons is also undecided. Hence the number of extra dimensions required in a Kaluza-Klein theory is both uncertain and arbitrary.

A second problem is accounting for the observed elementary particles. In quantum theories such as the non-Abelian gauge theories there are two classes of elementary particles, namely the bosons and the fermions. We have already mentioned the bosons, which are the carriers of the fundamental forces. For example, in the quantum-mechanical view the force of gravity is caused by the continuous exchange of bosons called gravitons between two massive bodies. The result of the exchange is seen in the laboratory as an attraction between the two bodies. There is no difficulty in deriving the bosons from a Kaluza-Klein theory. The higher-dimensional gravitational field can readily lead to bosons in the four-dimensional world.

The fermions, which constitute the second class of elementary particles, assume a completely different role in physics. Unlike the bosons, which transmit forces, the fermions make up all the bulk matter in the universe. The electron, the neutron, the proton and the neutrino are fermions. Indeed, the quarks that make up the neutron and the proton are fermions as well.

How can the fermions be accounted for in a Kaluza-Klein theory? They cannot be derived from a bosonic gravitational field. The only way to obtain them is to add one or more fermionic fields to the higher-dimensional theory. The fields would then lead to the fermions observed in four dimensions. The number of fermionic fields included in the theory is arbitrary because there is no theoretical principle on which the number is based.

Supergravity

There are many interesting studies of Kaluza-Klein theories in an arbitrary number of dimensions for which the fermionic fields are added "by hand." The arbitrariness, however, detracts from the simplicity of Kaluza's original idea. It is desirable to have a theory in which the number of fermionic fields and the number of dimensions are given naturally by the structure of the theory.

Supergravity is such a theory. In the first place it is an extension of general relativity in which bosons and fermions are treated on an equal footing. The bosonic graviton, for example, has a fermionic partner called the gravitino. In Einstein's version of general relativity one can add or drop the fermions at will, but in supergravity there is a fermionic partner for each boson. Thus in supergravity the fermions needed to describe the structure of matter are present in the theory from the beginning.

The number of dimensions in supergravity is also constrained. As we pointed out above, theories of supergravity probably do not work in dimensions higher than 11. Beyond that number the mathematical requirements for a partnership of the bosonic and the fermionic fields cannot be met. Moreover, Edward Witten of Princeton University has shown that at least seven hidden dimensions must be added to the four dimensions of spacetime in order to incorporate the strong, weak and electromagnetic forces into a Kaluza-Klein framework. There is a third feature of 11-dimensional supergravity that is circumstantial but highly attractive to theorists. Although in all dimensions lower than 11 there are several versions of supergravity that are mathematically distinct, in 11 dimensions the theory is unique.

The minimal ingredients of a Kaluza-Klein theory include the gravitational field, which gives rise to the bosons, and a fermionic field, which accounts for the fermions of our world. There must also be at least one bosonic field in addition to the gravitational field, which acts as the source that drives the compactification, or the curling, of the extra, hidden dimensions. Remarkably, the 11-dimensional version of supergravity includes precisely these three ingredients.

Even more surprising to the theorist is the fact that the extra bosonic field naturally leads to only two kinds of compactification. In one kind seven of the 11 dimensions curl up into a small, hidden structure: such a compactification would explain why the number of dimensions readily observable in the world is four. The alternative is that only four dimensions curl up, and this scenario would lead to a seven-dimensional world. Future physicists may be able to discover why the four-dimensional world is preferred.

To develop a Kaluza-Klein theory based on supergravity in 11 dimensions the physicist must first solve the equations of supergravity. Many solutions give rise to a spacetime structure generated by a four-dimensional spacetime and a small, closed, seven-

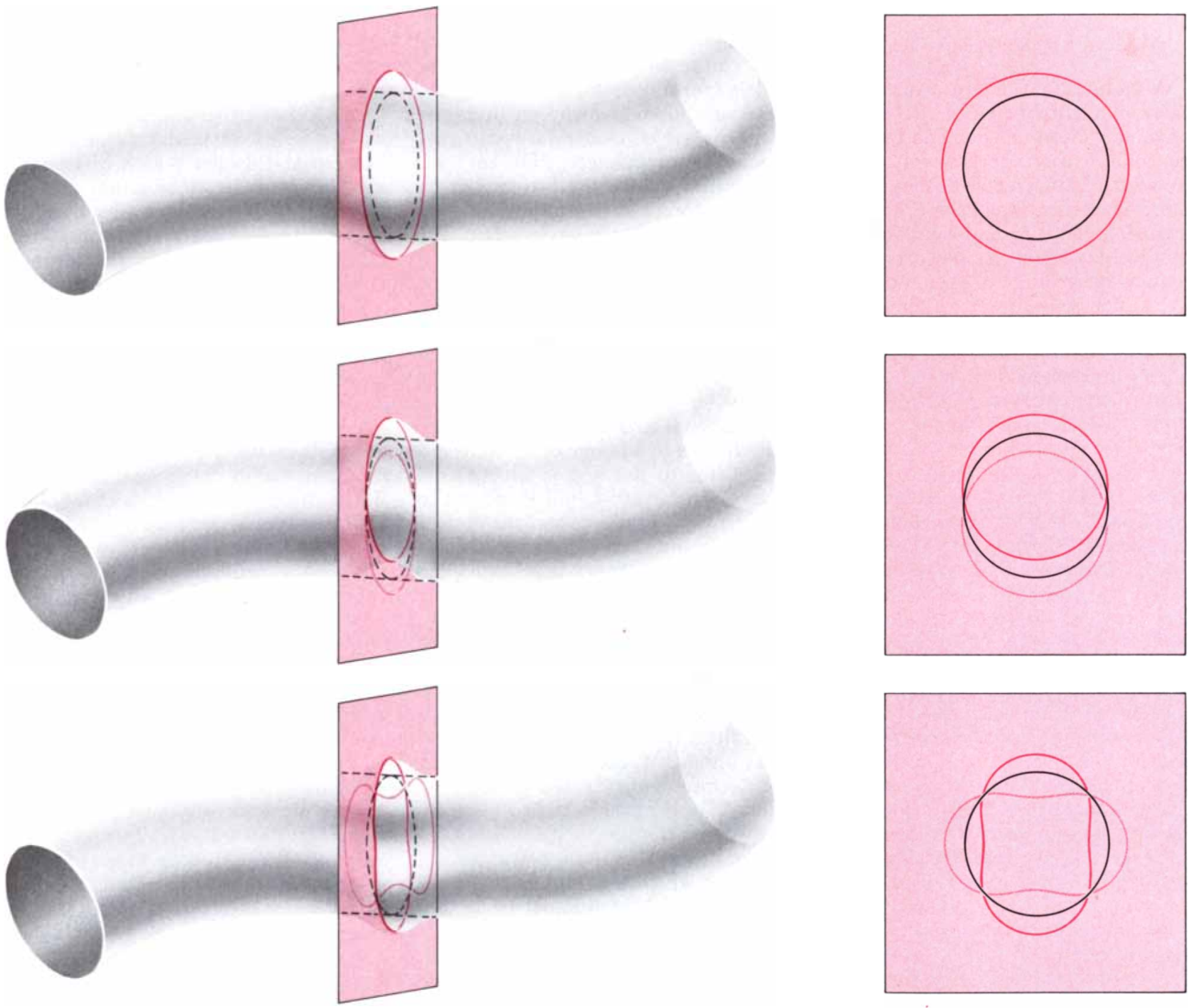
dimensional surface. One then studies the symmetry group of each surface corresponding to a solution of the equations, and the symmetry group determines the non-Abelian gauge theory that is to be unified with gravity. Different closed surfaces have different symmetry groups, and each one determines a different grand unified theory of the nongravitational forces.

The final step in the development of a Kaluza-Klein theory is to analyze the complicated stationary wave patterns allowed by the closed surfaces. These patterns determine the masses and other properties of the particles predicted by the theory in ordinary, four-dimensional spacetime. Each of the seven-dimensional surfaces that appear as a solution of the equations of supergravity must be analyzed in this way.

Theoretical Results

Most of the investigative effort has been devoted to two cases. In the first case the curled-up dimensions form the simplest and most symmetrical structure possible in seven dimensions, namely the seven-dimensional analogue of the sphere. Much of the work on the seven-dimensional sphere has been done by Michael J. Duff and Christopher N. Pope of the Imperial College of Science and Technology in London, François Englert of the Free University of Brussels, Bernard de Wit of the State University of Utrecht and Hermann Nicolai of CERN.

The second case is a set of surfaces having the symmetry group needed for the strong, weak and electromagnetic



PARTICLES may be associated with the curled-up, circular fifth dimension in Kaluza's theory. According to quantum mechanics, every particle can also be understood as a wave. If some integral multiple of the wavelength fits exactly around the circumference of the circle in the fifth dimension, the particle corresponding to the wavelength should be able to exist in ordinary, four-dimensional spacetime. The first kind of wave that fits onto the circle is a wave of constant amplitude around the entire circle. If ordinary, curved spacetime is represented as a curved line, the higher-dimensional spacetime generated by the curved line and the circular fifth dimension is a curved cylinder. The wave of constant amplitude appears as a bulge in the cylinder (*top*); its cross section is shown at the top right. The particle corresponding to this wave has no mass in Kaluza's theory. The second kind of wave oscillates once around the circle. It too appears as a bulge and its cross section is in effect a cosine wave plotted around the circle as if the circle were the horizontal axis of a graph (*middle*). The closed figure formed by the wave is shown as a dark-color curve. The figure precesses, or rotates, around the circle, and one of its subsequent orientations is shown as a light-color curve. Kaluza's theory predicts the particle associated with this wave is 10^{16} times the mass of the proton. The third and higher kinds of wave divide the circle into two, three or more equal parts. The third kind of wave is a cosine wave whose wavelength fits exactly two times around the circle (*bottom*); it too precesses as shown. The particle associated with the third wave is twice as heavy as the particle associated with the second wave.

forces. These surfaces have been studied by Witten, by Leonardo Castellani, Ricardo D'Auria and Pietro Fré of the University of Turin and by others.

Unfortunately the detailed results of the studies do not predict a four-dimensional world that resembles the one we know. There are three main problems. The first is called the chirality problem because it concerns the handedness of the fermions predicted by the theory. The chirality of a fermion is determined by the sense of its quantum-mechanical spin with respect to the direction of its motion. Every 11-dimensional structure studied so far predicts an equal number of left- and right-handed neutrinos. The neutrinos observed in nature, however, are always left-handed; there seem to be no right-handed neutrinos.

The second problem is called the cosmological problem, and it concerns the curvature predicted for ordinary four-dimensional spacetime. If one makes the reasonable assumption that the seven extra dimensions form a compact structure so small that it has not yet been observed, the remaining four dimensions of spacetime become highly curved as well. Astronomical observations suggest, on the contrary, that the curvature of the universe over a large scale is zero or close to zero. In Kaluza-Klein theories not based on supergravity the problem can be avoided. One can add a constant to the equations called the cosmological constant, which has the effect of canceling the curvature of four-dimensional spacetime even when the other seven dimensions are highly compact. Such freedom to adjust the underlying equations is not present in 11-dimensional supergravity.

The third problem for 11-dimensional supergravity is called the quantum problem, but there is hope its resolution could eliminate the first two problems as well. The theories underlying the Kaluza-Klein program are based on quantum-mechanical equations, and such equations lead to infinite quantities having no obvious physical interpretation. The infinite quantities present a general difficulty for virtually all quantum theories of gravity. To avoid them the theorist has been forced to make approximations that neglect some of the quantum effects. Eventually one might hope either to demonstrate that the infinities result from the approximation procedure rather than from the theory itself, or to find a special theory in which the infinities are absent.

In the past several months some theoretical physicists have become excited by the prospect that the problem of infinite quantities and possibly the other problems we have mentioned can be resolved by a type of theory called a superstring theory [see "Dual-Resonance Models of Elementary Particles," by John H. Schwarz; SCIENTIFIC AMERICAN, February, 1975]. Superstring theories have some of the attractive properties of supergravity. To be consistent mathematically they must be constructed in 10-dimensional spacetime, and in 10 dimensions there are very few theories possible. It has been known for some time that infinite quantities are absent in superstring theory at the first level of approximation of the quantum effects. Some physicists now believe the infinite quantities are absent at all levels of approximation.

Superstring Theory

In a string theory particles are associated with the vibrational motions of a one-dimensional string in a higher-dimensional space. The major difference between a string theory and a field theory such as supergravity is in the way one must count the number of particles predicted by the two theories. If the seven extra dimensions of higher-dimensional supergravity were not curled up into a closed surface, 11-dimensional supergravity without compactification would predict a finite number of particles. An infinite number of particles arise in supergravity only because of compactification. For example, in Kaluza's five-dimensional theory there is an infinite series of particles because there is an infinite series of stationary wave patterns that fit onto the circular fifth dimension. In superstring theory, on the other hand, there are infinitely many particles even if there is no compactification of the extra dimensions. The infinitely many particles in superstring theory correspond to the infinitely many wave patterns that can persist on the string.

Most of the particles that arise in superstring theory have an extremely large mass: more than 10^{19} times the mass of the proton. Nevertheless, the theory also predicts about 1,000 massless particles. Until recently the mutual interactions of these particles were thought to be equivalent to the interactions described by a 10-dimensional version of supergravity, and there were two reasons that version of supergravity was not under intensive study. First, there seemed to be no solutions to the equations of the theory in which six dimensions curl up and leave a four-dimensional spacetime with reasonable properties. Second, the equations themselves become inconsistent when they are interpreted at the quantum level. The 10-dimensional version of supergravity, and consequently the mutual interactions of the massless particles described by the superstring theory, did not seem to be relevant for the Kaluza-Klein program.

Recently Michael Green of Queen Mary College in London and Schwarz have demonstrated that the interactions of the massless particles in superstring theory differ slightly from their interactions in the 10-dimensional version of supergravity. The effects are subtle ones, caused by the infinitely many heavy particles that are present in the superstring theory but not in supergravity without compactification. When the effects of the heavy particles are included, one obtains consistent equations at the quantum level.

This recent success has stimulated renewed and vigorous effort on the compactification of the six extra dimensions in superstring theory. In many ways the problem is even more difficult than it is in 11-dimensional supergravity. The properties of the six-dimensional surfaces required in superstring theory are more complex mathematically than the properties of, say, the seven-dimensional sphere. Nevertheless, there is much incentive to solve the problem, and there is some indication that the two other major problems in supergravity, namely the chirality problem and the cosmological problem, do not arise in the superstring theory.

Future Development

There is often a long time between the development of elegant theoretical ideas and the precise formulation of predictions that can be tested experimentally. It took 13 years, for example, to find the correct way of applying non-Abelian gauge theories to the unification of the fundamental forces. The current lack of clear indications that the ideas of supergravity and Kaluza-Klein theory are experimentally correct does not necessarily indicate the ideas are wrong. They may simply require further theoretical work.

There is also a relation between the development of ideas in fundamental physics and new mathematical concepts. For example, it was possible to bring supergravity to its present level of sophistication because the mathematical theory of anticommuting numbers was ready to be applied. It may be that a deeper understanding of the role of space and time in the quantum theory will require the development and incorporation of additional mathematical ideas. The current interest in higher-dimensional theories of gravity may be only a first step toward that understanding.

INTRODUCING THE 16-VALVE, INTERCOOLED SAAB TURBO.

IN ANY OTHER CAR, THIS MUCH POWER WOULD BE IRRESPONSIBLE.

Remember muscle cars? "Getting rubber"? "Slam shifting"? "Blowing someone's doors off"?

Screech. Graunch. Whoosh. The golden songs of a golden era. Tributes to the prodigality of power in the hands of impetuous youth. Redolent of nostalgia. And like many things nostalgic, out of place in the present.

Now that gas is back, so is muscle. Only today, you have a choice. You can once again buy a car that uses it frivolously.

Or you can buy the Saab 900 Turbo.

The new 16-valve, intercooled Saab Turbo engine is fuel-efficient, like the first and second generation Saab turbocharged engines that preceded it. While being considerably more powerful at the same time.

Having four valves per cylinder (with dual-overhead camshafts) helps the new Saab Turbo engine inhale and exhale easier. It enables the spark plug to be located at the center of the hemispherical combustion chamber.

In strictly practical terms, this means combustion is more complete and efficient. So compression can be raised without raising the octane of the fuel needed to achieve it.

Increased compression means better throttle response. With no penalty in fuel economy. Both mean greatly increased driving enjoyment.

Intercooling lowers the temperature of the air mass compressed by the turbocharger. Which means boost can be increased (it has) and, along with it, power (it has too).

The intercooling and 16-valve head combine to produce the kind of mid-range torque you may remember from the halcyon days of the muscle car.

*1985 SAAB PRICE LIST**

900 3-door	*$11,850*
900 4-door	*$12,170*
900S 3-door	*$15,040*
900S 4-door	*$15,510*
Turbo 3-door	*$18,150*
Turbo 4-door	*$18,620*

Automatic transmission $400 additional.

But, unlike the power of yesteryear, the power of the Saab Turbo has been harnessed to be of maximum practical advantage in a real world.

How to put power to its best use.

Making a car go faster used to entail nothing more complex than squeezing into it the largest engine and carburetors that would fit.

With steering that was often, to be charitable, a bit vague. Tiny drum brakes. A chassis that was nothing if not flexible.

And a body that had little crashworthiness going for it other than sheer mass.

Saab thinks other things should precede power before it can be put to responsible uses.

First, you should build a body that is not only rigid, but safe.

Under that, you should put the best shock absorbers available (gas filled) so the power can be transmitted to the road no matter what kind of road it's transmitted to.

The brakes should be more than equal to the power they're meant to hold in check. Four-wheel disc brakes.

These brakes should ride within wheels and tires capable of resolute traction. Light-alloy, 15″ wheels with low profile, high-performance radials.

Steering should be precise, channeling power towards its proper course.

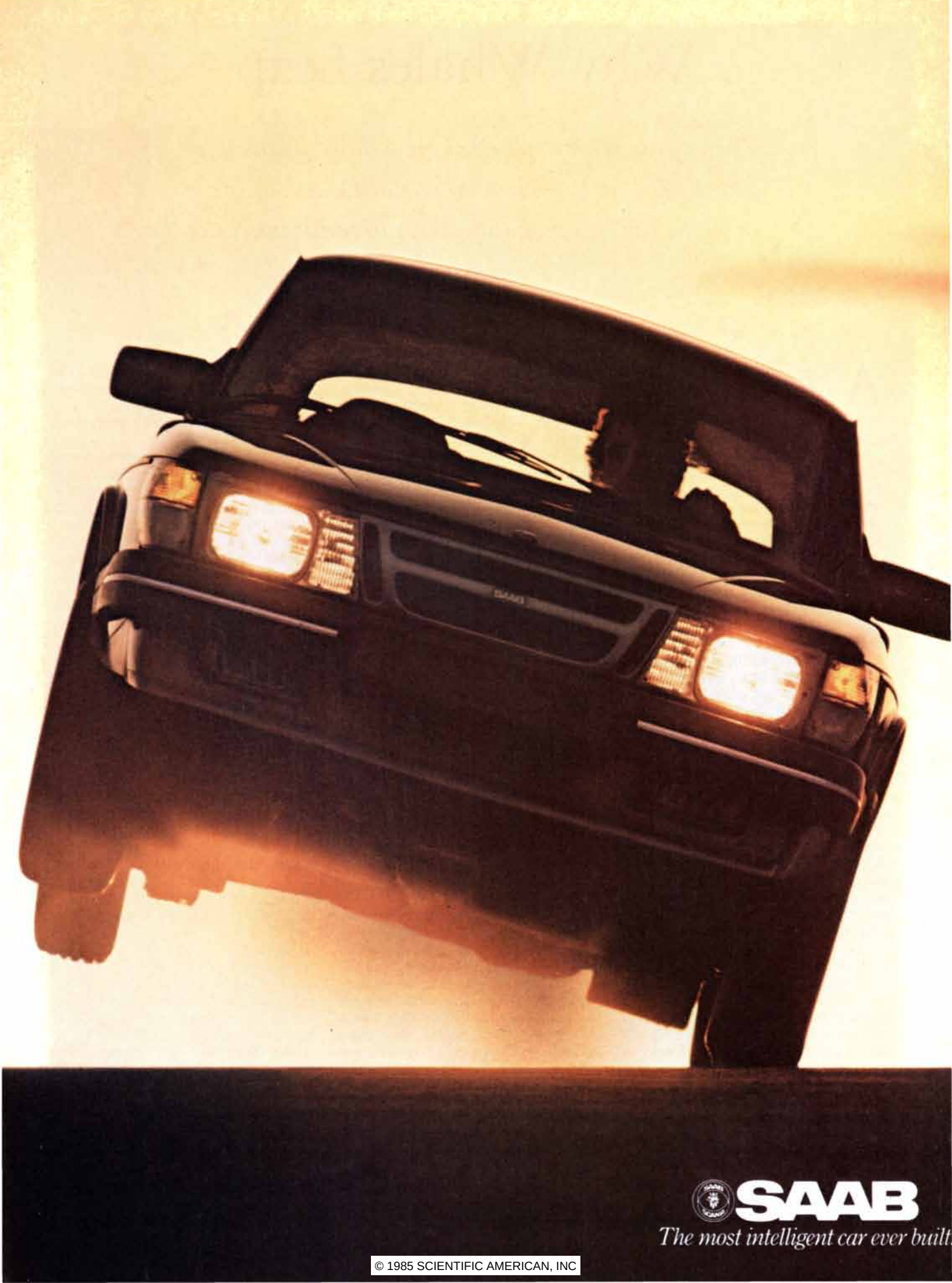
And power should be transmitted through the front wheels, where it can do the most practical good.

All these things make a Saab the perfect environment for an engine like the intercooled, 16-valve Turbo engine being introduced in 1985.

Of course, if you look for them, you will find more powerful cars on the market.

But none that use their power so wisely or as well.

**Manufacturer's suggested retail prices. Not including taxes, license, freight, dealer charges or options. There are a limited number of turbo models available with Saab's Exclusive Appointments group, which includes: leather upholstery, fog lights, and electric sunroof, at additional cost.*



Why Whales Leap

The action, which is called breaching, seems to be purposeful. It is associated with the social aspects of whale life and probably serves in communication

by Hal Whitehead

A whale's leap from the water is almost certainly the most powerful single action performed by any animal. It is called breaching, a term that whalers of the 18th and 19th centuries gave this dramatic activity and that present-day investigators of the phenomenon have retained. Considering the great bulk and weight the whale must lift in breaching, one wonders why the animal does it.

A breach provides the only opportunity most human observers have to see an entire whale, and it has inspired a wide variety of impressions. Thus J. N. Reynolds, recounting for readers of *The Knickerbocker* in 1839 the adventures of whalers in the Pacific, wrote: "Occasionally, a huge, shapeless body would flounce out of its proper element, and fall back with a heavy splash; the effort forming about as ludicrous a caricature of agility, as would the attempt of some over-fed alderman to execute the Highland fling." To Herman Melville the breach was sublime. "Rising with his utmost velocity from the furthest depths," Melville wrote in *Moby Dick,* "the Sperm Whale thus booms his entire bulk into the pure element of air, and piling up a mountain of dazzling foam, shows his place to the distance of seven miles and more. In those moments, the torn, enraged waves he shakes off, seem his mane."

The whalers of earlier centuries, searching for their quarry in slow sailing vessels, had many opportunities to observe the whales they were trying to catch. For years the anecdotes told by such men formed the basis of what was known of breaching and other kinds of whale behavior. Among the explanations of breaching they proposed, somewhat anthropomorphically, were feeding, stretching, amusement, being chased by swordfish and an "act of defiance," which was presumably directed at the whalers.

In the past few years scientific observations of whales in the open ocean have begun to yield useful quantitative data on many aspects of their behavior, including the breach. Roger Payne of the U.S. World Wildlife Fund and his associates have contributed many insights through their long study of southern right whales (*Eubalaena australis*) off the Valdés Peninsula in Argentina. Other important studies include work with gray whales (*Eschrichtius robustus*) off Baja California by Kenneth S. Norris of the University of California at Santa Cruz and a number of other investigators and observations of humpback whales (*Megaptera novaeangliae*) off Hawaii by James D. Darling of the University of California at Santa Cruz, Peter Tyack of the Woods Hole Oceanographic Institution and others. My own work has also been mainly with humpbacks that are in the western North Atlantic off Newfoundland during the summer and on Silver Bank in the West Indies during the winter.

Such long-term observations are crucial to an understanding of breaching because the phenomenon is generally rare. Most whales are seldom seen to breach. Hence it usually takes many years to witness even a moderate number of breaches. In this respect the research on Silver Bank was particularly important. Humpback whales from the western North Atlantic congregate there during the winter months for mating and calving. They reach a density approximating one whale per square kilometer. Many of them breach: during our transects of some 200 kilometers across the bank for the purpose of estimating the size of the population we saw breaching in about 20 percent of the pods (usually containing from one whale to four whales) we sighted.

A leap by a humpback entails the lifting of as much biomass as would be accounted for by 485 people weighing an average of 68 kilograms (150 pounds) each. The largest humpbacks reach lengths approximating 15 meters (49 feet) and weigh 33 metric tons (72,765 pounds).

The breaches of the humpback and of other whales known to breach range from a full leap clear of the water to a leisurely surge in which only half of the body emerges. In more than a fourth of the breaches by humpbacks at least 70 percent of the animal comes out of the water, but it is rare for the entire whale to be seen above the surface. Humpbacks breach at all angles up to 70 degrees with respect to the surface of the sea.

Payne has observed the breaching process while watching southern right whales from cliffs or small airplanes. The whale swims horizontally until it has developed enough speed. Then it tilts its head upward and raises its flukes, or tail. These actions convert the horizontal momentum into vertical momentum and the whale emerges from the water. Because of the horizontal approach, a whale can breach in water that is only a few meters deep.

Whales perform other actions that superficially resemble breaching. One of them is lunging. In this maneuver the whale thrusts no more than 40 percent of its body through the surface. A lunge can be executed horizontally, vertically or at any angle between those extremes. The whale can be oriented so that its dorsal surface or ventral surface is uppermost or so that it is lying on its side. Whales often are seen closing their jaws while lunging, sometimes ingesting a mouthful of plankton or small fish. Lunging is therefore usually considered to be associated with feeding. Humpbacks, however, can be seen to lunge as they try to outmaneuver one another in large groups, for example when from two to 10 males compete for access to a female among them. Lunging, then, happens when a whale breaks the surface as an unin-

tentional result of an underwater maneuver. A breach, on the other hand, seems to be purposeful.

Another activity in which certain aquatic animals intentionally jump above the surface is porpoising. The animal makes a series of horizontal leaps while traveling fast. Robert W. Blake of the University of British Columbia has calculated that by making such leaps a small whale or a dolphin minimizes frictional drag. He has also shown that large whales would not benefit in this way by porpoising, and indeed I have never seen humpbacks do it.

Breaches fall naturally into two types, which I call belly flops and true breaches. In the belly flop the whale remains dorsal side up throughout the breach and lands on its belly. In a true breach the animal emerges from the water on its side, twists with flailing flippers and lands on its back. Humpbacks make true breaches about 80 percent of the time; about 20 percent of the time they belly flop.

A belly-flopping whale is more likely to be seen to blow, or exhale, than a whale making a true breach. Payne has suggested that belly flopping might be as painful for whales as it is for people. It does, however, leave the blowhole clear of the water for a longer time than the true breach does. It thus might be the choice when the whale wants to breathe during a breach.

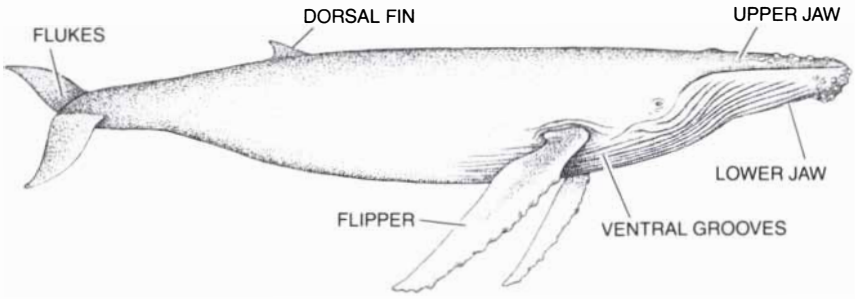
Breaches are often executed in sequence. A particular whale may breach every 40 seconds or so for a few minutes. Among pods of humpback whales in the western North Atlantic the mean sequence length was 9.4 breaches. (The mean includes occasions when a sequence consisted of only a single breach.) Usually all the breaches appeared to be made by a single animal. One sequence we observed on Silver Bank consisted of 130 breaches in 75 minutes, probably all by the same animal.

Within a sequence the tendency is for belly flop to follow belly flop, true breach to follow true breach. Among both humpbacks and right whales a breaching animal tends to lift successively less of its body out of the water as a sequence progresses. As one might expect in the circumstances, the whale seems to be getting tired.

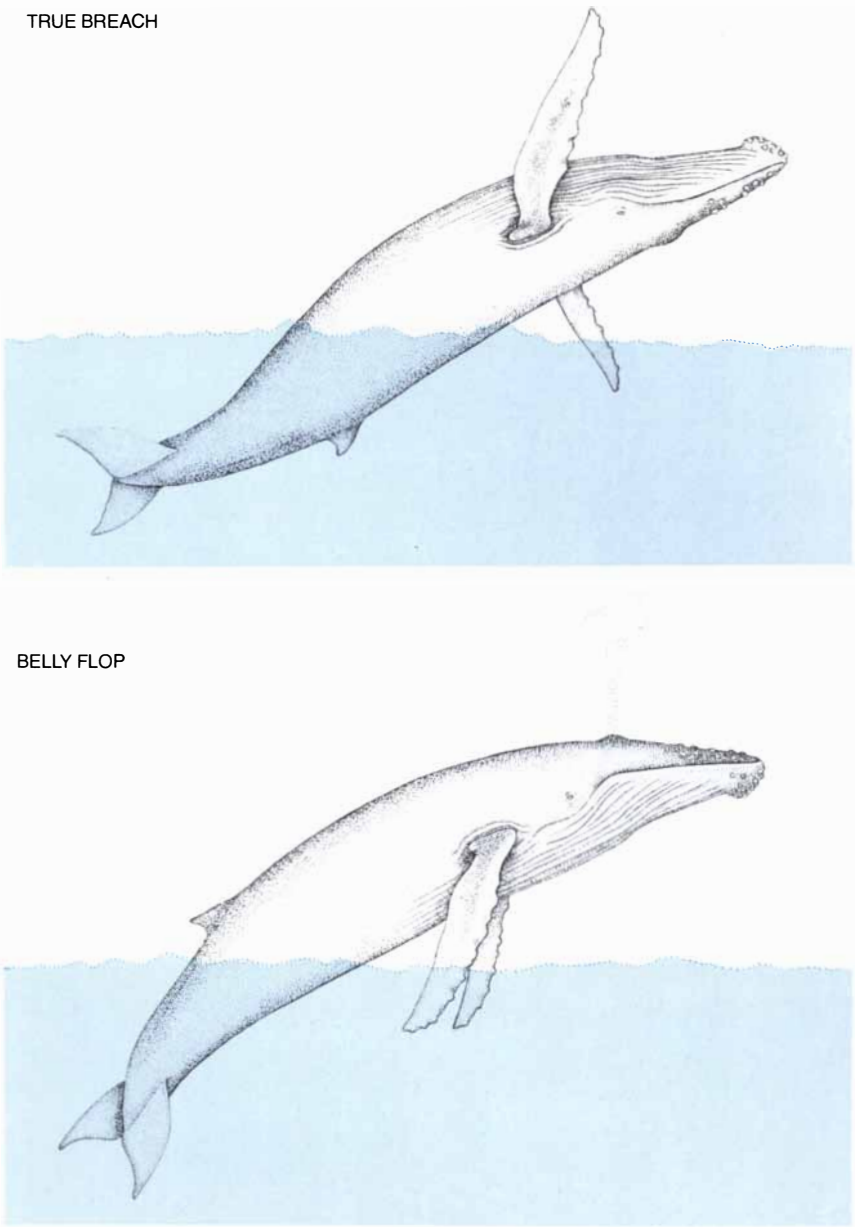
How much energy is a whale consuming as it makes a breach, and how much power is it developing as it leaves the surface? Using measurements from photographs of breaching whales, I have simulated the breaching process on a small computer. In a full breach, in which most of the animal leaves the surface of the water at an angle of about 35 degrees, a 12-meter adult humpback breaks the surface at about 15 knots (17 miles per hour). Because that is almost the maximum speed the animal can attain, a full breach represents the extreme use of a humpback's propulsive power.

The energy necessary to make such a

BREACHING WHALE was photographed in the Pacific Ocean near Hawaii. The whale, a humpback (*Megaptera novaeangliae*), was executing a true breach, in which the animal emerges from the water on its side, twists in the air and lands on its back. The other type of breach, done much less frequently, is a belly flop. Most of the breaches observed (approximately 80 percent) are true breaches.



HUMPBACK WHALE was the subject of most of the author's observations of breaching. The humpbacks he and his colleagues studied spend the summer off Newfoundland and the winter on Silver Bank in the West Indies. Social interactions among the humpbacks are probably more important in the winter, the season when the whales mate and give birth.



TRUE BREACH AND BELLY FLOP begin as the whale emerges from the water at any angle up to 70 degrees with respect to the surface. The animal lands on its back or belly respectively, as is depicted here. Often in a belly flop the whale blows, or exhales, suggesting that the belly flop (done infrequently) might be the choice when the animal wants to breathe. The belly flop keeps the blowhole clear of the water for a longer time than the true breach.

breach is roughly 2,500 kilocalories. The whale's resting metabolic rate is some 300,000 kilocalories per day. Hence the energy consumed in a breach is a little less than a hundredth of the animal's minimum daily caloric requirement. The energy translates into about 2.6 kilograms (5.7 pounds) of the capelin fish (*Mallotus villosus*), a major item in the humpback's diet. (Humpbacks often catch capelin in 100-kilogram gulps.) One breach is therefore not a particularly significant event in the daily energy budget of a whale. A sequence of 20 breaches or more, however, consumes a good deal of energy. It is not surprising that successive breaches are weaker.

It is less easy to say why a whale breaches. Studying the behavior of large whales has been likened to astronomy. The observer glimpses his subjects, often at long range; he cannot do experiments, and he must continually try to infer from data that are usually inadequate. Under conditions of this kind one way to investigate the function of an activity is to examine its context.

I have spent several hundred hours in small sailboats following groups of humpbacks through their daily routines. This work, together with the observations by Payne and others, is yielding a fairly clear picture of the circumstances in which whales breach. It is not giving rise to a set of firm rules about breaching; such certainty is usually not possible in studies of the behavior of advanced animals. The best one can do is to put forward statistically significant tendencies. What they suggest is that breaching is mainly associated with social interaction among whales, perhaps in communication and (among young whales) play.

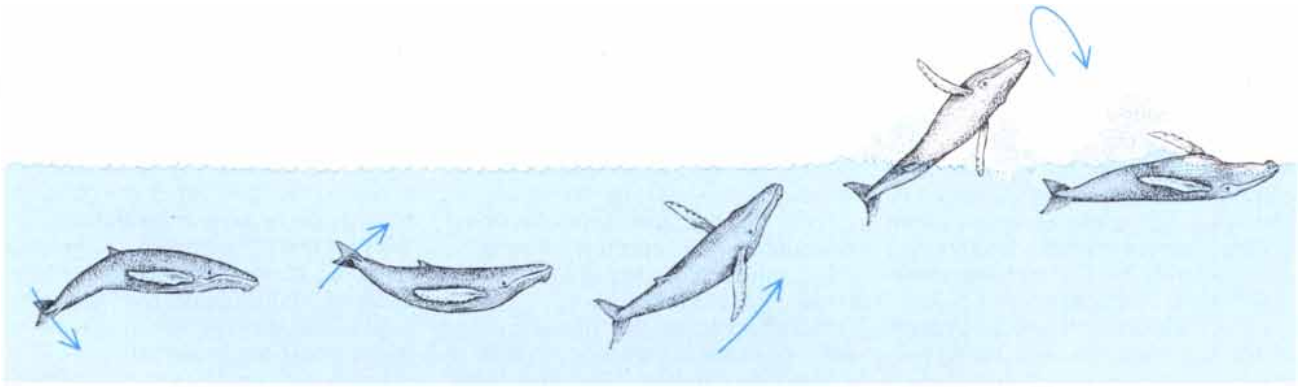
Whales often breach when a pod containing two or more humpbacks splits into two groups or when two pods (sometimes consisting of single whales) merge. A breach also often takes place within 15 minutes of a lobtail: a thrash of the whale's flukes onto the surface of the water. It may similarly be associated with flippering (a flipper lifted above the water and slapped down) and other demonstrations. Christopher W. Clark of Rockefeller University and Payne have observed similar patterns among southern right whales.

It is notable and apparently contradictory that humpbacks breach less in summer, even though groups split and merge more often then than they do in winter. Mating and calving take place in winter, however, and such social interactions are probably more important than the summer ones. Hence

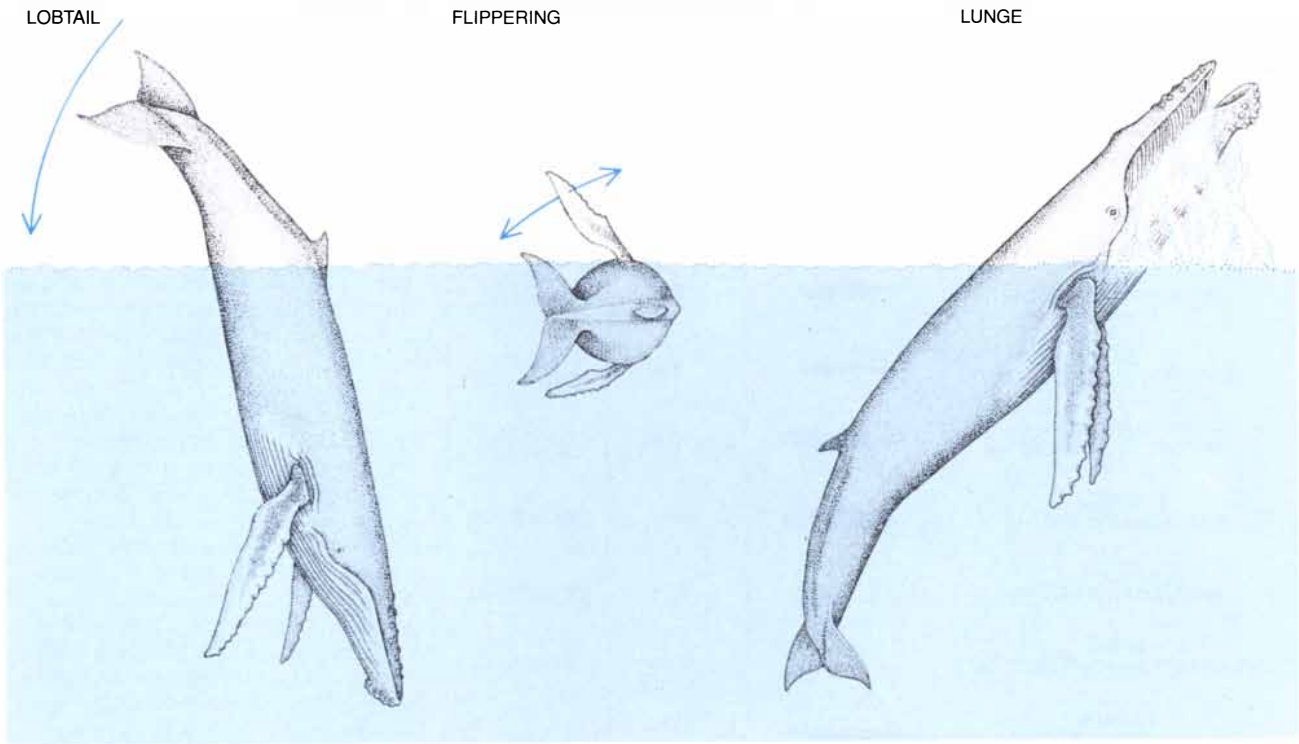
breaching rates are correlated not only with the number of social interactions taking place but also with their importance in the life of the whales.

An additional correlation between breaching and social activity is seen when one looks at the rates of breaching among different whale species. In examining this question I drew up a table that includes the ratio of mean mass to length cubed as an indication of rotundity. The more rotund species would seem to be less likely to breach because of unfavorable hydrodynamics. It is surprising, then, that observations show that they do it more frequently.

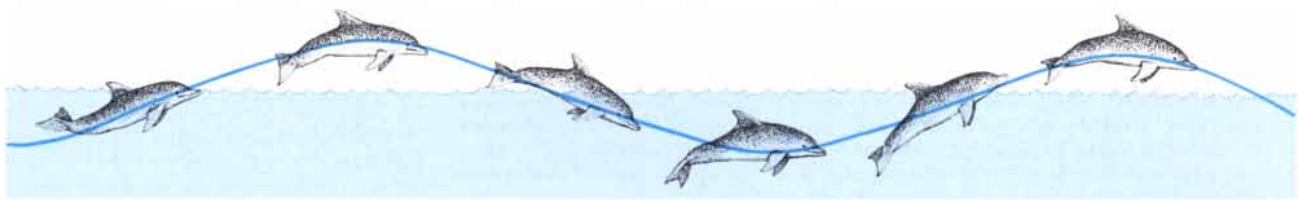
Right whales, gray whales and humpbacks—the three best-studied rotund species—congregate in winter on traditional breeding grounds. They sel-



EXECUTION OF BREACH begins when the whale, swimming more or less parallel to the surface, builds up speed. It raises its flukes and tilts its head upward, changing the horizontal momentum to vertical momentum. This whale is beginning a true breach.



ACTIVITIES RELATED TO BREACHING include lobtailing, in which the whale slaps its flukes on the surface, and flippering, in which a flipper is slapped on the surface. They seem to be purposeful. Lunging, which brings only part of the body out of the water, is evidently an unintentional result of an underwater maneuver. A breach often is done within 15 minutes of lobtailing or flippering.



PORPOISING consists of a series of horizontal leaps made when the animal is moving at high speed. For a small whale or a dolphin the action minimizes frictional drag from the water. Porpoising would not be efficient for large whales, and they seem not to do it.

dom feed there, subsisting instead on the energy stored in their thick layers of blubber. Social interactions are frequent and sometimes vigorous on those breeding grounds, and it is there that most breaching is seen.

In contrast the blue whale (*Balaenoptera musculus*), the finback (*B. physalus*) and the sei (*B. borealis*)—all slim—do not seem to frequent particular breeding grounds but remain dispersed during the winter months. This strategy probably reduces their net expenditure of energy, so that they do not need thick blubber layers. They may employ loud low-frequency sound or perhaps a monogamous social system to obtain access to mates. In any event they probably have rather few close-range social interactions.

Little is known of the social systems of the bowhead (*Balaena mysticetus*), Bryde's whale (*Balaenoptera edeni*) or the minke (*B. acutorostrata*), but the general impression among close observers is that among these baleen whales the more social species have the higher breaching rates. The sperm whale (*Physeter catodon*), a toothed whale that breaches frequently, has a particularly complex social system.

What other clues emerge from investigations of the context of breaching? One unexpected finding, obtained in several independent studies, is that whales breach oftener as the wind speed rises. It is not an abrupt increase during gales, when a whale might be trying to get a breath of air unaccompanied by spray, but rather a gradual increase over quite moderate wind speeds. Payne has speculated that whales might be using the breach as a means of communicating by sound (from the slap of reentry) when noise from wind and waves obscures their normal vocalizations.

Payne made another discovery that led him to think breaches might have a signaling function. He found that among southern right whales breaching begets breaching. In other words, the likelihood that an individual whale would breach increased when whales nearby were breaching. Intrigued by this finding, I carried out a rudimentary form of the procedure known as spectral analysis on some of our transects on Silver Bank. The results suggest that breaching humpbacks form clusters about 10 kilometers in diameter. A whale was more likely to be breaching if it was within 10 kilometers of other breaching whales. At the height of the season such a cluster might have 100 humpbacks, of which from 10 to 15 might be breaching. Under good conditions they might be expected to hear the sound of a breach over a distance of a few kilometers. Thus the findings lend tentative support to Payne's hypothesis that breaching has a signaling function. If other whales see or hear a breach, information has been conveyed. The message would at least be that a whale has breached.

Is a breach an efficient way of conveying any other message? It makes a spectacular sight and a loud noise for observers on the surface, but most of the other whales are below the surface at the time of a breach. Even in the clearest water the limit of underwater vision is about 50 meters. Under favorable conditions, however, sound can travel quite far in seawater. The question therefore becomes whether a whale can generate louder sounds, in at least a few frequency ranges, by breaching than it can by vocalizing. Little information is available on the strength of the underwater sound produced by a breach, and no information is available on whether whales try to maximize their output of sound during a breach.

A breach might also be a display intended as an act of aggression, as a challenge, as a show of strength or as a maneuver in courtship. Aggression seems to be a fairly unlikely motive. A whale is a smooth and well-cushioned animal, and so it is difficult to see how a breach could do significant harm unless the victim was much smaller than the breacher. I know of a case where a humpback landed on a boat during a series of breaches off Newfoundland, but the incident seems to have been an accident rather than a display of aggression by the whale. During many months at sea in small boats, I have never felt that any of the thousands of breaches we observed was aggressively directed toward us. Moreover, the whale can probably display aggression more effectively by administering a blow with its flukes.

A whale making a full breach is exhibiting its maximum power to any whale within sight or earshot. Hence the breach might be useful as a courtship display, a challenge or a show of strength. A female might choose a

SPECIES	RATIO OF MASS TO LENGTH CUBED	BREACHING RATE
HUMPBACK *(MEGAPTERA NOVAEANGLIAE)*	10.6	VERY FREQUENT
RIGHT *(EUBALAENA AUSTRALIS)*	16.2	OFTEN BREACHES
GRAY *(ESCHRICHTIUS ROBUSTUS)*	14.3	OFTEN BREACHES
SPERM (MALE/FEMALE) *(PHYSETER CATODON)*	10.7/19.1	OFTEN BREACHES
BOWHEAD *(BALAENA MYSTICETUS)*	26.7	OCCASIONAL
BRYDE'S *(BALAENOPTERA EDENI)*	6.1	OCCASIONAL
MINKE *(BALAENOPTERA ACUTOROSTRATA)*	12.3	UNUSUAL
FINBACK *(BALAENOPTERA PHYSALUS)*	4	RARE
BLUE *(BALAENOPTERA MUSCULUS)*	6.3	ALMOST NEVER
SEI *(BALAENOPTERA BOREALIS)*	3.6	ALMOST NEVER

BREACHING AND ROTUNDITY seem to be positively linked: the rounder the whale is, the more likely it is to breach. Rotundity is indicated by the ratio of the whale's mean mass to the cube of its length. In general the slimmer whales do the least breaching, although the hydrodynamics of breaching would seem to favor them. Apparently the correlation arises from the fact that the rotund species frequently engage in the kind of social activity often accompanied by breaching, particularly when they congregate in the winter on traditional breeding grounds. They do not eat much in winter, subsisting mainly on their reserves of blubber. The slimmer whales are much less social and probably feed consistently all year.

BREAKTHROUGH: TAKE THE HEADACHES OUT OF HIGH ALTITUDES.

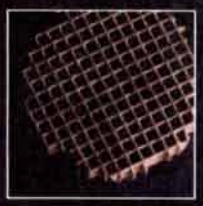
Planes going a long distance fly at high altitudes to conserve fuel. But at those altitudes, some oxygen atoms cluster together in threes to form ozone that can give passengers headaches and respiratory problems.

Our engineers reasoned that it is easier to pull the ozone apart than to build heavy filters to keep it out. A concept was developed using catalytic screens in the air ducts to attract the ozone for an instant as it passes, allowing the atoms to reform as oxygen and literally turning a noxious gas into a breath of fresh air.

We're creating breakthroughs not only in airline travel, but in space manufacturing, information systems, and helicopter design and production.

We're McDonnell Douglas.

Material used is Deoxo® catalyst substrate ozone abater manufactured by Englehard Corporation Specialty Chemicals Division.



MCDONNELL DOUGLAS

Almost priceless.

Hewlett-Packard presents the first 32-bit UNIX™ system under $5,000.

It's the one computer for both the technical and administrative sides of your job.

Introducing the Integral Personal Computer.

Don't let the small size fool you. This is definitely a full-fledged member of the HP 9000 family of engineering workstations.

With a 16/32-bit MC68000. A graphics co-processor. 800KB of standard memory expandable to 5.8MB.

And a multitasking operating system based on UNIX System III. It's a UNIX system that's easy to use. Its powerful but friendly user interface means even novices can now tap the power of this ideal operating system for technical tasks. There are no cryptic commands to learn. And there's an optional mouse.

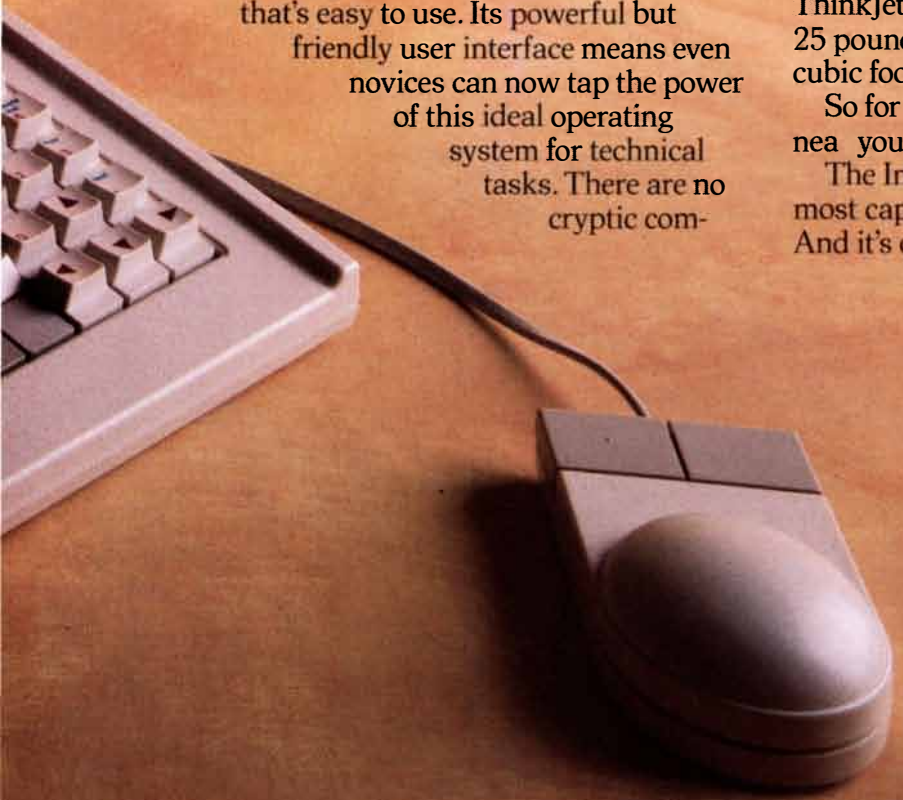
It's a personal computer without the limitations. Like other PCs, it runs a variety of popular spreadsheet, graphics, database and word processing software. But unlike other PCs, it has the multitasking operating system, powerful microprocessor and programming flexibility for your specialized applications. Just what you'd expect from Hewlett-Packard.

Developing custom solutions on the Integral PC is easy, with optional programming tools such as HP Technical BASIC, HP-UX "C" and Real-Time Extensions.

And the Integral PC goes where you need it. The complete system—including our built-in ThinkJet graphics printer—packs into a single 25 pound package that takes up less than a cubic foot.

So for the name of the dealer or HP sales office nea you, call 1-800-FOR-HPPC.

The Integral Personal Computer. Perhaps the most capable personal computer you can buy. And it's only $4,995*



HEWLETT PACKARD

PG02505

UNIX is a trademark of AT&T Bell Laboratories.
Mouse is optional. *$4,995 is suggested U.S. list price; prices elsewhere may vary.

Give people the tools they need, and there is no limit to what they achieve.

Announcing a major breakthrough in word processing technology.

Now, state-of-the-art comes easy.

While many businesses need a full-featured word processing program, few people in business have the time to learn one.

So we studied how people worked, learned and thought. We also drew from our experience with over 1,250,000 WordStar owners. And we created a totally new word processing program that does it all.

Easily.

With new WordStar 2000, the keys you press are the keys you'd expect to press, so they're easier to remember ("c" for copy, "p" for print, etc.). Or, you may use your computer's "function" keys. If you ever remove something by mistake, "undo" puts it back.

With WordStar 2000, an office temp with only moderate computer experience can be up and word processing in a matter of minutes. As for the complete novice, MicroPro's exclusive on-screen "tutor-in-your-computer™" makes learning fast and fun.

Never before have so many advanced word processing functions been so easy to learn and use.

With WordStar 2000, you press fewer keys to get more things done.

Simple keystrokes give you "windows," to work on several documents simultaneously. Press another few keys for WordStar 2000's built-in five function math. Or its built-in spelling corrector. Get WordStar 2000's special Plus package and you can also manage your own mailing lists. And even use electronic mail.

WordStar 2000 is the easy word processing program you'll never outgrow. Of course, the original WordStar and WordStar for PCjr will continue to provide substantial and proven word processing capability for those with more modest budgets.

See your local MicroPro dealer today, or call (800) 227-6703 [in California (800) 632-7979] for the dealer nearest you.

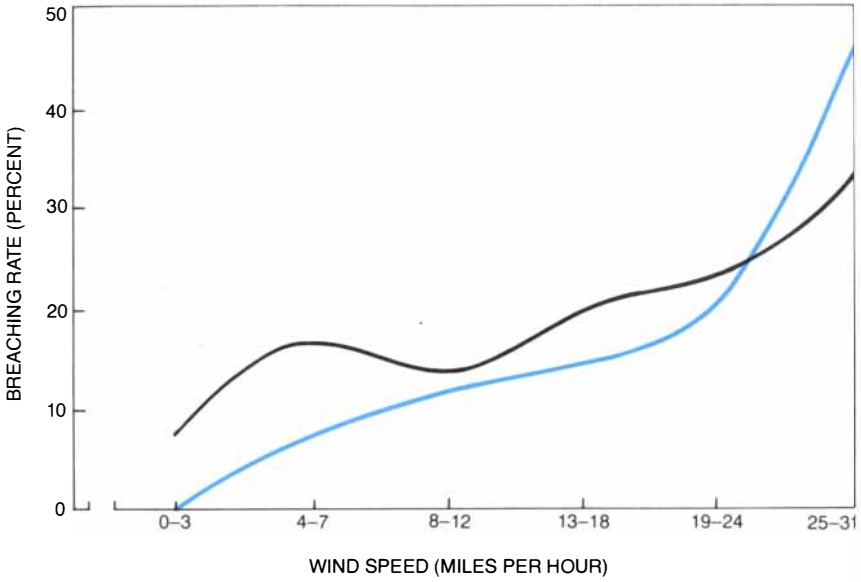
So you can remove all limits from what you can achieve.

Now available for IBM PC®/AT®/XT® and compatibles with 256K RAM. Upgrade from WordStar to WordStar 2000 through your dealer, or call MicroPro Customer Update (800) 227-5609, 9am-3pm P.S.T.

WORDSTAR® 2000

Now there are no limits.™

MicroPro.



BREACHING AND WIND SPEED also show a positive correlation. The correlation is charted for data derived by the author for humpbacks in 1978 (*black*) and 1980 (*color*). Roger Payne of the U.S. World Wildlife Fund believes breaching increases as the wind rises because it aids communication when the noise of the wave obscures normal vocalizations.

mating partner at least partially on the basis of the strength of his breach or his ability to keep up a strong output of power or sound during a sequence of breaches. Such a male would be demonstrating strength and stamina and so perhaps (indirectly) genetic fitness.

Similar correlations might make a breach useful as a challenge or a show of strength directed at other males competing for access to a particular female. Breaches by right whales and humpbacks are often seen when males are engaged in such a competition.

One must also consider the somewhat blurry concept of play. People watching an animal perform an action with no immediately obvious function tend to call it play. As a result the concept has become a catchall category for otherwise inexplicable behavior, in which breaching has often been included. Recently play has had serious attention from a number of biologists and students of animal behavior, and it is now generally regarded as a valid (but hard to define) behavioral category. If breaching is an important activity for whales, and if the way in which it is done influences its effectiveness, there are good selective reasons why calves and possibly adults might "play" at it.

Breaching has most of the characteristics of other activities that animal behaviorists call play: it is common in social contexts, it is often done by young animals and in many instances it has no obvious function. Some investigators have speculated that a purpose of play in other young animals is to aid the development of musculature; breaching might serve this role in young whales.

The most spectacular breaches are made by the youngest whales. Right-whale, gray-whale and humpback calves begin breaching when they are only a few weeks old. The breaches are often vigorous and may run on in long sequences. On Silver Bank calves breached oftener than adults. Indeed, it would be rather exceptional for adult animals to engage regularly in such a vigorous activity as a form of play. Hence it seems unlikely that play is the main function of breaching among adult whales.

The findings I have reported and the hypotheses I have discussed do not indicate any single clear function for breaching. The evidence suggests the activity has several functions. Although there are strong correlations with sociality and breaches have characteristics that would make them effective as signals of physical prowess, no evidence conclusively supports either hypothesis.

My subjective evaluation is that breaching often serves to accentuate other visual or acoustic communication. It is a kind of physical exclamation point. Just as people raise their voice, gesticulate with their hands or jump up and down to emphasize a communication, so the whale breaches. And, like eavesdroppers, human observers usually miss the message, noting only its salient features.

A Neolithic Fortress and Funeral Center

Excavations at Hambledon Hill in southwestern England reveal that in about 3600 B.C. elaborate funeral rites began to be carried out there. As the funeral center decayed, a huge fortress was built

by R. J. Mercer

Hambledon Hill is a landmark of striking proportions that sits astride the valley carved by the Stour River through the chalklands of southwestern England. A Neolithic herdsman who looked up to the hilltop in about 3400 B.C. would have seen an impressive sight. Crowning Hambledon Hill was a huge defensive enclosure with three concentric ramparts. The inner rampart, the most formidable of the three, was supported by 10,000 oak beams as thick as telephone poles. In the ditch around the ramparts human skulls placed at intervals added an eerie note to the appearance of the fortifications.

The Neolithic complex at Hambledon Hill had not always been intended for defense. The impressive fort was the final step in a process of modification and expansion at the site that began in about 3600 B.C. and may have lasted for several hundred years. From an archaeological point of view what is even more arresting than the fortifications is the possibility that in its early phases Hambledon Hill served as a setting for elaborate funeral rituals. It appears that when a member of one of the communities near the hill died, the body was exposed to the elements in a special area. In some instances the body may have been accompanied by offerings of precious objects. After the flesh had fallen off the bones, some corpses may have been selected for burial in another place with further ceremonies.

The energetic building program at Hambledon Hill was the outcome of important social developments in the Neolithic period, which began in Britain in about 4000 B.C. with the transition from hunting and gathering to agriculture. The first agricultural communities probably did not have a stable enough economic base to build permanent settlements; few traces of the earliest agricultural communities have been found. Within a few hundred years, however, the economic stability provided by agriculture enabled some members of the community to turn to activities other than farming or tending their herds. The fortifications and funeral rituals at Hambledon Hill were products of the energies liberated from subsistence activities by the development of agriculture. Hambledon Hill is one of the largest Neolithic sites excavated in Europe so far and it is one of the few known Neolithic sites where funeral rituals were highly developed. As such the site is helping to provide a novel picture of life during the New Stone Age in Britain.

How has the picture of life and ritual at Hambledon Hill been assembled? The Hambledon Hill Excavation and Field Survey Project, which lasted from 1974 to 1982, was an emergency effort to rescue Stone Age artifacts from the encroachment of modern agriculture. It had long been known that there were Neolithic artifacts on Hambledon Hill. As early as 1913 Heywood Sumner, a British architect and antiquarian, had detected the remains and drawn up one of the first site plans of a Neolithic enclosure to be published in Britain.

A brief exploratory survey in 1959 by Desmond Bonney of the Royal Commission for Historical Monuments showed that other Neolithic earthworks existed on the hill and suggested the complex might repay full-scale excavation. In the early 1960's plowing of the sheep pasture on the site was begun and continued plowing would have destroyed the evidence needed to understand the huge monument. Therefore in 1974 an excavation was undertaken that ultimately covered 60,000 square meters.

The field project revealed the remains of several related Neolithic structures on the hill [*see illustration on page 97*]. Most of the artifacts connected with funeral rites were found in a large enclosure (the main causewayed enclosure) at the center of the hill. Two long barrows, low mounds that may have had a ceremonial function, faced each other across the main enclosure, one to the south and one to the north. A smaller enclosure (the Stepleton enclosure) occupied the tip of the southeastern spur of the hill. The Stepleton enclosure was probably a domestic area for at least part of the time it was used in the Neolithic period.

The defensive fortifications that circled the hilltop were strongest on the southern and western slopes of the hill, between the Stepleton enclosure and the main enclosure. It is possible that there was a third enclosure in a strate-

HAMBLEDON HILL is a great outcrop of chalk that dominates an area of rich pastureland. A large Stone Age necropolis and later a fortress once crowned the hill. Little of the Neolithic monuments can be seen in the photograph. (The striking earthworks on the northern spur of the hill [*bottom*] are from a later Iron Age fort built over part of the Neolithic fortifications.) Most of the remains relating to Neolithic funeral rites were found at the center of the hill and on the Stepleton spur (*top left*). The Neolithic fortifications protected the entire hilltop; they were particularly strong on the southern and western slopes (*upper right*).

gic position for defense on the northern spur of the hill, but it lies under a much later hill fort from the Iron Age and has not yet been excavated. Although the defensive structures and the funeral monuments lie close together, they were probably not strictly contemporaneous; unraveling the chronology of the phases of construction at the site was one of the greatest challenges of the field project.

In addition to Hambledon Hill's strategic location, several economic factors undoubtedly influenced the choice of the imposing hilltop as a site for building. The hill dominates an area well supplied with two resources that were critical in the early Neolithic period: grazing land and flint. The rich pastureland at the foot of the hill includes the Vale of Blackmore to the west and the skirts of the chalk uplands of Cranborne Chase to the east. The chalklands are rich in flint and the hunting weapons and axes of prehistoric hunters abound there.

As a fertile region where prehistoric hunters were already concentrated, the area around Hambledon Hill was a natural site for the early development of agriculture. The process of development may have been spurred by the arrival of immigrants from the European continent with new techniques and materials, but this part of the Neolithic period in Britain is not well understood. In any event, by about 4000 B.C. the transition to an agricultural economy was under way.

In the transition from hunting-and-gathering groups to stable agricultural communities a new type of architectural structure—the walled enclosure—had a significant role. Walled enclosures served as means of controlling resources, delimiting areas where specialized activities (such as toolmaking) were carried out and defending a community from attack. About 60 Neolithic enclosures are known in southern England, ranging in size from one hectare to 70 hectares. (A hectare is about 2.5 acres.) A ditch generally ran around the outer circumference of the wall of the enclosure. The overall design varied from a simple enclosure with one ditch to sites with up to five concentric rings of ditches.

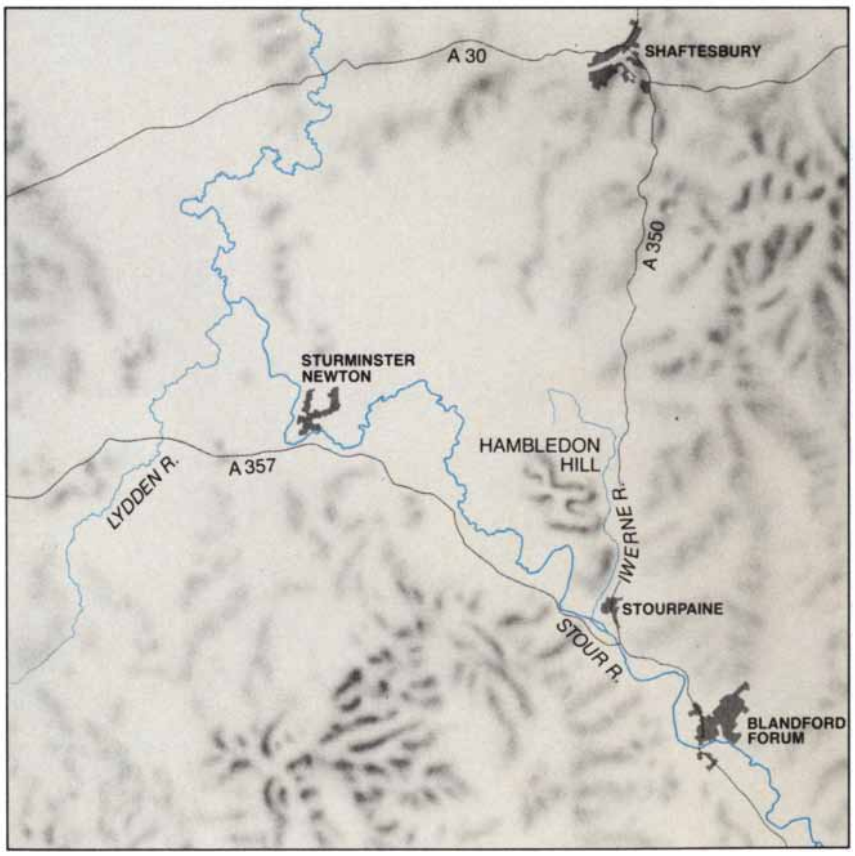
Regardless of the complexity of their design, almost all Neolithic enclosures shared one construction feature: the ditches around the enclosure were not continuous. Instead the ditches were interrupted at irregular intervals by causeways perpendicular to the long axis of the circular depression. The presence of the causeways suggests that the ditches were not conceived as defensive barriers in their own right but as quarries for the construction of the internal bank or rampart. In some instances the bank survives, although it is generally much reduced by erosion. The construction material and style of the bank probably depended on the available resources and the function of the enclosure.

At Hambledon Hill the main causewayed enclosure, set in the center of the hill, formed one of the primary focuses of the field work. Since so little was known about Hambledon Hill in 1974 when the Excavation and Field Survey Project began, the preliminary digging strategy was a simple one. About 20 percent of the interior of the main enclosure was excavated to determine what the enclosure had been used for in Neolithic times.

Roughly the same proportion of the enclosing ditch was also excavated. The work in the ditch provided additional information about the function of the site. The excavation of the ditch enabled the field-work team to set up a chronology of the phases of use of this part of the Hambledon Hill complex. The chronology was based on radiocarbon dating and a careful reconstruction of the strata that accumulated in the ditch after it was first dug.

The initial digging revealed that the subsoil of Hambledon Hill had been extensively damaged by erosion and agriculture. It soon became apparent that the hill had been plowed during the late Stone Age, the Bronze Age and the Iron Age as well as in Roman and medieval times. The millenniums of plowing had removed between 70 centimeters and one meter of the topsoil.

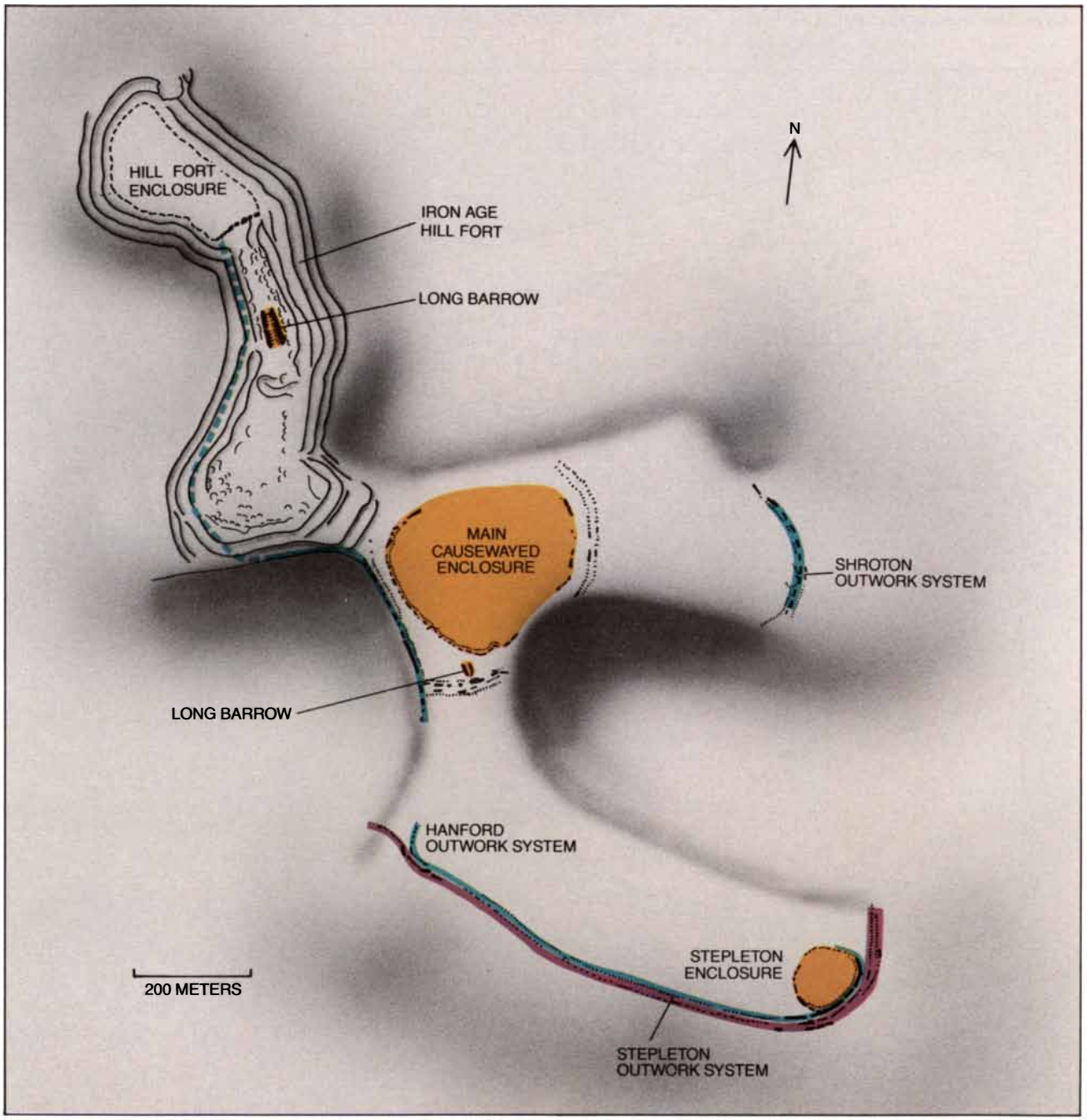
The removal of so much soil posed a major archaeological difficulty because with the soil were torn away



STRATEGIC SETTING of Hambledon Hill made it a natural place for an early Neolithic community to build. The hill commands the corridor cut by the Stour River through the chalklands of southwestern England. It dominates the rich grazing lands in the Vale of Blackmore to the west. In about 4000 B.C., at the beginning of the Neolithic period, population may have been concentrated in the area, drawn by the availability of two key resources: pastureland and flint for making stone tools. The Neolithic fortress on the hilltop would have been visible for miles around to the members of the communities living in the lowlands at the foot of the hill.

most of the postholes from the foundations of buildings and other wood structures. Hence little can be said about the dwellings and other structures that might have stood on Hambledon Hill. That is a considerable loss. Many of the pits in the main enclosure, however, were deep enough for their bottom parts to have survived the plowing, and these pits yielded much fascinating detail about ritual practices in Neolithic times.

It was clear that in many instances the pits had been dug and left open to allow a sediment of naturally eroded chalk to be deposited in the bottom. Only when this had happened were carefully chosen objects deposited in the hole. The assemblages of objects, which include pottery, stone axes and red-deer antler, appear to have been ritual offerings that may have accompanied corpses laid out in the enclosure.

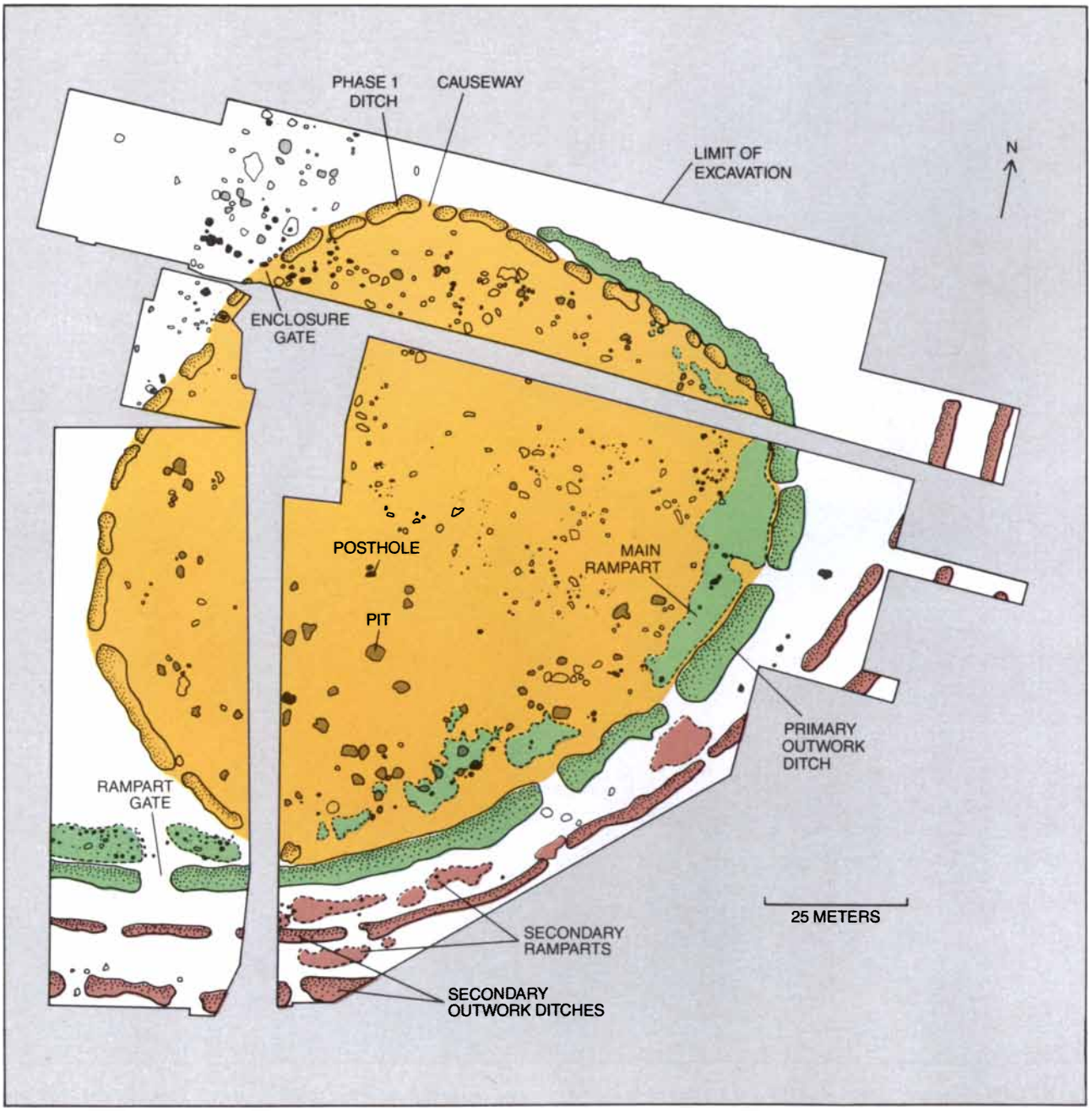


SITE PLAN shows how Hambledon Hill was developed. In its early phase (*orange*) the site was a center for funeral rituals. The main causewayed enclosure at the center of the hill was where the bodies were first laid out. After the flesh had fallen off the bones some bodies may have been buried in a pair of long barrows positioned north and south of the main enclosure. The Stepleton enclosure may have been a dwelling place for a small, privileged group of people who conducted the funeral rituals. As the funeral center fell into disuse the Neolithic community transformed the site into a fortress. A rampart was built on the southern and western slopes and the Shroton spur (*green*). A third enclosure, on the northern spur under the Iron Age fort, may have been the command center of the fortress. Later two additional ramparts (*purple*) reinforced the main one along the southern slope, which is gentle and hence vulnerable.

The objects in these assemblages undoubtedly had a high prestige value for the members of the Neolithic community. Fragments of pottery recovered from the pits are apparently the remnants of complete vessels. Compared with shards from other parts of the site, the pottery includes a high proportion of wares imported from distant regions such as Cornwall and Devon. Some of the vessels are quite large and were made with a skill that prompted imitations (only partially successful) in local materials.

Among other imported items found in the pits are stone axes. Analysis of the stone shows that the axes came from Cornwall, southern Wales and Borrowdale in Cumberland. The value of these funeral deposits to the Neolithic inhabitants is suggested by the fact the axes were desirable enough to justify importation over what were then vast distances: 400 kilometers in the case of Borrowdale. Even more striking was the retrieval of two axes, one of nephrite and one of jadeite, that do not come from Britain at all but probably come from Brittany or perhaps even farther afield.

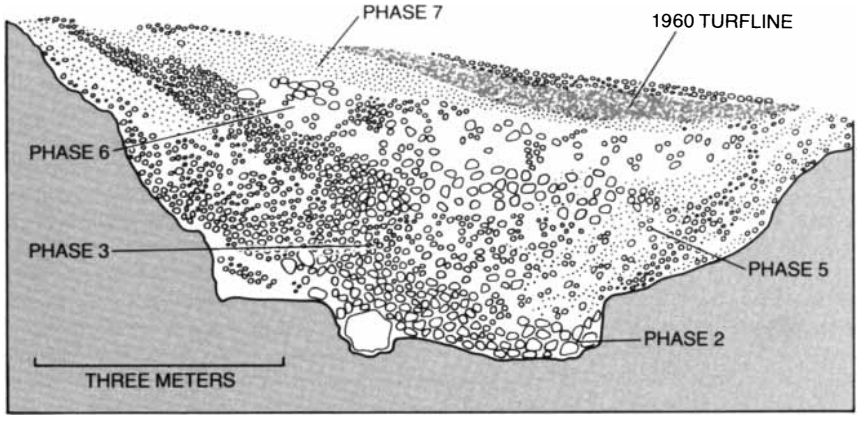
The items found in the main enclosure thus carry the suggestion of ritual practices. The excavation of the ditch around the main enclosure has helped to confirm that rituals were carried out and has also provided information



STEPLETON ENCLOSURE was probably a domestic area (*orange*), as indicated by the refuse found there. The digging of the Phase 1 ditch furnished material for a bank of which little remains. The gate at the upper left was the entrance to the enclosure. Postholes within the enclosure may have been foundations for dwellings; pits contained the remains of feasting and flintworking. After the original bank collapsed, fortifications were built over part of the enclosure. The primary outwork ditch yielded material for the main rampart (*green*), which was entered by the gate at lower left. Secondary ditches yielded material for two smaller ramparts (*purple*).

about what kind of rituals they were. The ditch was originally dug to furnish material for the rampart around the enclosure. The remains of the rampart are flimsy, to say the least. What evidence there is suggests that the rampart was a timber-framed case into which a mass of chalk from the ditch was packed to produce an impressive but ultimately unstable barrier.

After the rampart was erected a pastelike silt accumulated at the bottom of the ditch as material washed down the sides of the depression. At many points on the circumference of the ditch the accumulated silt appears to have been carefully cut out of the bottom of the ditch. The paste that had been gathered in this way was probably employed to maintain and repair the rampart.



CROSS SECTION through the ditch around the main enclosure shows strata that helped to establish the chronology of Neolithic events at the site. Phase 1 was the cutting of the ditch itself, hence no Phase 1 remains appear in the ditch. The chalk removed from the ditch was used to build the wall of the enclosure. Phase 2 deposits included human bone and what appear to be ritual offerings of red-deer antler, stone implements and pottery. The Phase 3 layer consists of the wall of the enclosure, which collapsed into the ditch as the main enclosure ceased to be used. The Phase 5 layer reflects the ritual recutting of the ditch and the deposit of additional offerings. In Phase 6 a flint cairn was built over the original ditch. Phase 7 is a post-Neolithic accumulation of cultivation soil. (Phase 4 deposits, which resemble those of Phase 5, were not present at the point where this cross section was made.)

The extraction of the primary silt by the builders deprived us of archaeological information about the initial use of the enclosure. After the removal of the silt, however, an assemblage of objects that appear to have a ritual character were placed on the ditch floor. The deposits, which were probably offerings of some kind and which might originally have been in leather bags, include human bone, animal bone, flint tools and pottery.

In addition to these items, objects more clearly connected with the disposal of human bodies were placed in the ditch. A series of skulls were laid right side up at irregular intervals on the ditch floor. A considerable quantity of broken human bone is also interspersed throughout the bottom layers of the ditch. Among the mass of bone two child burials survive intact under neatly constructed flint cairns.

Perhaps more significant are the trunk and thighs of a youth of about 15 whose body had clearly been lying exposed in the ditch in a relatively advanced state of decay. When the body had begun to fall apart, parts of it may have been dragged to the bottom of the ditch by dogs or other predators that gnawed the bones extensively.

The suggestion that bodies were intentionally exposed in or around the main enclosure at Hambledon Hill could help to answer two significant questions about Neolithic funeral practices. The excavation of long barrows at other sites has shown the barrows were the final resting place for skeletons or parts of skeletons that had been allowed to weather elsewhere. Where had the bones in the long barrows been exposed to the elements?

Furthermore, there is a puzzling imbalance in the bones retrieved from the barrows. In general the skeletons or partial skeletons include the remains of relatively few women and children; preference was given to the remains of adult males. The absence of children's bones is particularly striking in view of the very high mortality rate that undoubtedly prevailed among infants and children in prehistoric times. What happened to the bones of women and children?

A reasonable hypothesis suggested by our findings is that the bodies of the community members were laid out in exposure centers such as the main enclosure at Hambledon Hill. After the flesh had fallen off the bones, a few skeletons may have been selected for burial in long barrows. If the hypothesis is correct, then exposure centers such as the Hambledon Hill main enclosure formed part of a ritual process that included the long barrows.

Some support for this hypothesis comes from the fact that 60 percent of the great quantity of bone found at the Hambledon Hill main enclosure is from young children. Moreover, it would appear that bone from male corpses and from female corpses is present in roughly equal proportions. Thus all members of the community or of a subgroup of the community are represented. To link the main enclosure more directly to the barrows would require evidence from the long barrows on the site.

Such direct evidence is not yet available from Hambledon Hill. The southern barrow, 20 meters long, was bulldozed during pasture improvement in the 1960's. The bulldozer obliterated the mound of the barrow, destroying much of the evidence that would be needed to test the hypothesis. Similarities between the artifacts found in the barrow ditch and the artifacts found in the ditch of the main enclosure suggest the two monuments were linked in the minds of the builders. The loss of the barrow's mound, however, makes it impossible to link the corpses exposed in the main enclosure with those buried in the barrow.

The mound of the northern barrow, which is 66 meters long, is intact, but the barrow has not yet been excavated. The reason is that the northern barrow lies in a protected position within the later Iron Age hill fort. Since the Hambledon Hill project was in a sense a rescue effort, work was focused on the most vulnerable monuments and the excavation of the protected northern barrow was not undertaken. Although the material found in the main enclosure at Hambledon Hill is suggestive, a final answer to the question of the connection between the main enclosure and the long barrows will require further work.

The final major component of the hilltop funeral complex was the Stepleton enclosure. Much smaller than the main enclosure, the Stepleton enclosure was recognized as a Neolithic structure by means of field walking, ground survey and aerial photography. It is probable that in the early phases of building at Hambledon Hill the Stepleton enclosure was a small, simple structure with an entrance facing uphill toward the main enclosure.

Postholes found in the Stepleton

enclosure show that buildings stood there. The evidence is too sketchy to reconstruct the buildings but material found near the enclosure suggests the structures were dwellings. For example, the oldest deposits in the ditch around the enclosure contain little human bone (and no skulls). The deposits do, however, include large quantities of debris from the working of flint and red-deer antler, which were typical Neolithic domestic pursuits.

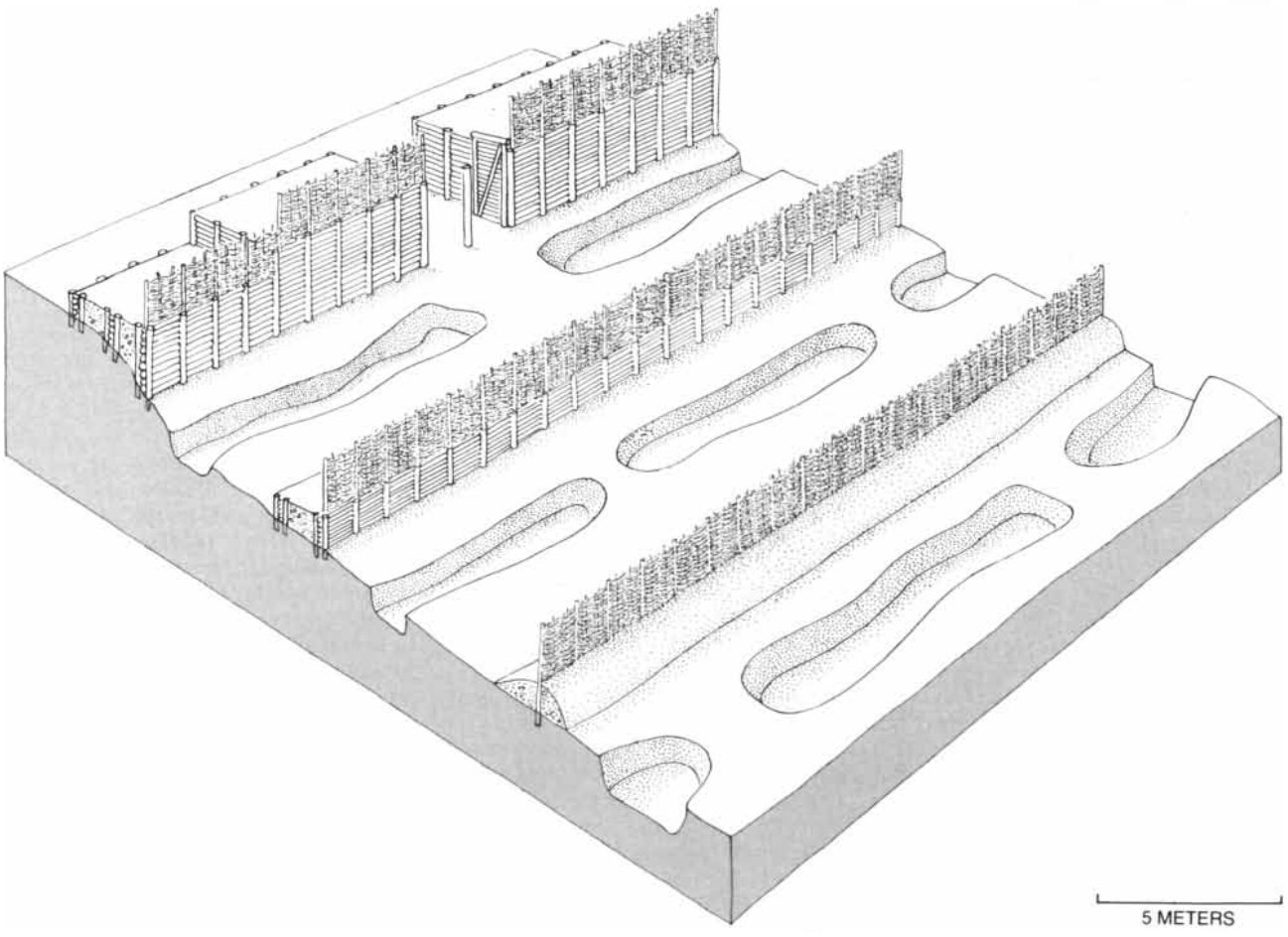
The enclosure was more than just a place of work: animal bones found in and around the Stepleton enclosure indicate that a good deal of feasting went on there. The state of the bones shows that much meat was wasted and there is little evidence that bones were smashed to yield humble soups or stews. Whoever lived at Stepleton liked the choicest kind of Neolithic food—roast meat on the bone—and did not hesitate to throw away the less desirable cuts. The pattern of consumption detected in and around the enclosure suggests that a small and relatively privileged group may have lived at Stepleton. It is tempting to speculate that this group presided over the elaborate funeral rituals that may have linked the main enclosure with the twin long barrows.

While Hambledon Hill was still at its height as a funeral center a period of modification began that probably lasted for 200 or 300 years. The cumulative effect of the modifications was to turn Hambledon Hill into a great fortress. The transformation was a gradual one, and acts of veneration of the dead were apparently carried out for a long time in the main enclosure.

The first defensive modification was probably undertaken after the main enclosure, the Stepleton enclosure and the barrows had fallen into some disrepair. A major earthwork was constructed along the southern part of the hill. The earthwork, a causewayed ditch backed by a timber-framed rampart, may have enclosed all the vulnerable points of approach to the entire summit of the hill, an area of about 60 hectares.

Little is known of the first system of earthworks enclosing the site because they were destroyed soon after being built during the reconstruction and strengthening of the defenses. The ditch segments were deepened and an enlarged timbered rampart was built over the remains of its earlier counterpart. The second outwork also protected the entire 60-hectare hilltop.

Like its predecessor, the enlarged rampart had a timber-framed "box" type of construction. Vertical supports in the form of oak beams as thick as telephone poles were placed about a meter apart along the front and back faces of the wall. To provide stability the upright supports were probably braced by horizontal beams passing through the rampart itself. Overall some 10,000 huge oak beams may have been employed in constructing



THREE RAMPARTS flanked the southern side of Hambledon Hill when the hilltop was at its height as a fortress. The main, inner rampart was a mass of chalk packed into a "box" frame. The frame was built of 10,000 oak beams with the diameter of telephone poles. At each entry point two gates swung shut onto a huge central oak post. The screening above the ramparts consisted of interwoven saplings. The second rampart also had an oak frame; the third rampart was simply a chalk mound. The ditches were interrupted by banks, or causeways, suggesting that they were not defensive barriers but merely served to provide construction material for the ramparts.

the timber frame. A project of such magnitude must have made great demands on the capacity of an early agricultural community to deploy its labor resources.

While the hilltop was being transformed into a fortress the funeral center apparently began to be neglected. At some point the ramparts of the main enclosure and the Stepleton enclosure collapsed into their respective ditches and the smaller of the long barrows tumbled into its ditch.

In spite of their dilapidation the main enclosure and the barrows apparently retained some of the original ritual function. Examination of the strata in the ditch of the main enclosure showed that after the rampart around the enclosure collapsed, pits were dug in the chalk rubble that filled the ditch. The pits, filled with ash, pottery, human bone and animal bone, were in some instances deep enough to penetrate to the bottom of the original ditch.

Later a narrow slot was cut around the circumference of the old ditch. In some parts of the ditch the slot was cut and recut as many as four times. Some sections of the new slot were filled with chalk rubble, others were left open to silt up and still others were filled with rich deposits of animal bone, pottery and flint tools. The digging of the pits and the cutting of the slot may have been acts of veneration linked to the burials that were once performed in the enclosure.

The final stage of ritual use of the main enclosure was the construction of a linear flint cairn over the original ditch. Excavation of the ditch around the smaller of the long barrows shows that the same sequence of ritual acts prevailed there: a slot was carefully cut in the rubble-filled ditch and later a flint cairn was built on the site of the old ditch. The parallelism suggests the two monuments were still linked in the minds of the builders as part of a funeral complex, albeit a diminished one. Near the Stepleton enclosure, which appears to have remained a domestic area, no such signs of commemoration were found.

While the character of the monuments inside the rampart slowly changed, the strengthening of the defensive arrangements continued. After the main outwork was built the construction of two additional ramparts reinforced the vulnerable southern slope of the hill. The outermost wall consisted of earth unsupported by timber; the middle rampart had a timber "box" like that of the inner wall.

Community members entered the hilltop fortress through three great timber-lined gateways. One entrance was next to the Stepleton enclosure; the second was on the Hanford spur of the hill between the Stepleton enclosure and the main enclosure; the third was on the eastern spur of the hill. Each entry point consisted of two gates flanking a large central post. A roadway 2.5 meters wide passed through the entryway, which tapered slightly toward the interior of the rampart and was lined with massive oak posts.

When the outer ramparts were finished, Hambledon Hill must have been an impressive site. It would have been visible for miles from the flat pastureland to the west in the Vale of Blackmore, where the community's herds grazed. The southern and western sides of the hill were rimmed by a timber-framed rampart 2,500 meters long. On the steep western flank a terrace had been excavated to provide a stable base for the rampart. On the southern flank a multiditch outwork system 1,200 meters long provided an imposing defensive obstacle.

The command center of the massive and grisly fortifications may have been the enclosure on the northern spur of the hill. This enclosure, which is almost totally obscured by the superposed Iron Age hill fort, was detected during the excavation by a combination of aerial photography and assiduous field walking on the part of Roger Palmer, the project surveyor. The 4.5-hectare enclosure, which is set in an excellent position for defense, has not yet been excavated.

The completion of the three rings of ramparts made Hambledon Hill a huge defensive structure. Yet the fortress was far from impregnable. Indeed, the archaeological record makes it clear that the site was abandoned after an attack that resulted in the burning of a large segment of the timbered outwork. The segment, some 200 meters long, is on the southeastern spur of the hill.

The fire can hardly have been accidental, since the entire timber structure was on fire, with the oak posts burning right down into their sockets. To achieve this effect the rampart would almost certainly have to have been torched. As the outwork burned, the timbers and then the rubble core of the rampart collapsed into the ditch. Much of the rubble is itself scorched, testifying to the intensity of the blaze.

Both attackers and defenders probably took losses in the struggle around the flaming outwork. The intact skeletons of two young males were found in the rubble and the condition of the skeletons makes it clear that they were quickly buried. One of the two was probably carrying an infant, which he crushed under him when he fell. The bearer of the infant was apparently killed by a finely worked leaf-shaped arrowhead that penetrated the thoracic cavity from the back.

Another young male, abandoned dead on the lip of the outwork ditch, was not covered by the falling debris. The condition of the bones indicates that his body was soon discovered by predators of all kinds. A fourth skeleton, found in the upper fill of the Stepleton enclosure ditch, may have been dragged there and dismembered by dogs or wolves.

Some of the victims of the attack were interred with more ceremony. On the north side of the Stepleton enclosure are two deliberate burials that can be linked to the conflagration. One body, that of a young male, was carefully laid in a pit that was then filled with heavily scorched chalk; the only known source of such chalk is the rampart destroyed in the attack.

The evidence for a violent end to the hilltop complex is unambiguous, but the circumstances of the attack are not clear. In about 3300 B.C., during the mid-Neolithic, there may have been a period of social upheaval brought on by economic or environmental factors. Several Neolithic sites are known to have been abandoned at this time, some of them littered with leaf-shaped arrowheads like the one found at Hambledon Hill.

Such general questions remain for the moment a matter of speculation. More particular questions, including the strategy of the final attack on Hambledon Hill, also lack answers. It seems likely that the central objective of the attack was the fortified enclosure on the northern spur of the hill that lies under the Iron Age hill fort. The skeletons and scorching found so far might represent the remains of a preliminary skirmish.

To tell whether that is the case will require excavation of the buried northern hill fort. Fortunately the northern spur of Hambledon Hill has recently been acquired by the British Nature Conservancy Council and will therefore remain protected from plowing and development. The excavation of the northern enclosure would undoubtedly yield further clues to the Neolithic history of Hambledon Hill and the story of its demise. Nonetheless, the work that has already been done at Hambledon Hill is giving an intriguing picture of the energy and imagination of the descendants of Britain's first agriculturists.

Ion Implantation of Surfaces

When foreign ions are embedded in the surface of a material, changes in structure and composition lead to novel surface properties. The technique is increasingly valuable both in industry and in fundamental research

by S. Thomas Picraux and Paul S. Peercy

It is the surface of a material that confronts the environment. There an object experiences corrosion and wear, is exposed to light and interacts with external electric and magnetic fields. As mechanical and electronic parts are miniaturized, surface-to-volume ratios grow ever larger, placing a greater premium on surface characteristics. Yet the electrical, mechanical, optical or chemical properties wanted on the surface of an object often conflict with requirements for its bulk properties, such as low material cost, high strength or ease of fabrication.

A technique known as ion implantation offers unprecedented freedom and control in tailoring surfaces with traits that are independent of bulk properties. Reduced to a beam of ions and electrostatically accelerated into the surface of a material, a controllable quantity of almost any element can be mingled with the host material. The foreign ions combine with atoms of the host material to form a layer perhaps one ten-millionth of a meter thick at a predetermined depth in the surface region. Unique structures and compositions can result: metals that are immiscible as liquids can be alloyed, and one material can be introduced into another at a concentration that could never be achieved at the high temperatures ordinarily needed to mix them. Surface properties are transformed accordingly, often in unexpected and favorable ways.

Ion implantation originated in the early 1960's, as avenues of research from otherwise disparate areas of science converged. Physicists were directing beams of energetic particles onto matter in investigations of atomic collisions and of the structure of nuclei. Nuclear chemists were trying to determine the changes in crystal structure that occur in solids when radioactive nuclei emit energetic particles in the course of their decay. Meanwhile workers in semiconductor electronics were searching for new ways of doping silicon—incorporating atoms of a foreign element into its crystal structure in order to change its electrical properties. Miniaturization and the effort to build integrated circuits that united a multitude of transistors, diodes, resistors and capacitors on a single chip demanded a more precise means of doping than the two techniques then in use: exposing silicon to the dopant in gaseous form or coating the silicon with a film of the dopant. Both techniques relied on the diffusion of dopant atoms into the silicon at high temperature, resulting in a concentration profile that changed gradually with depth.

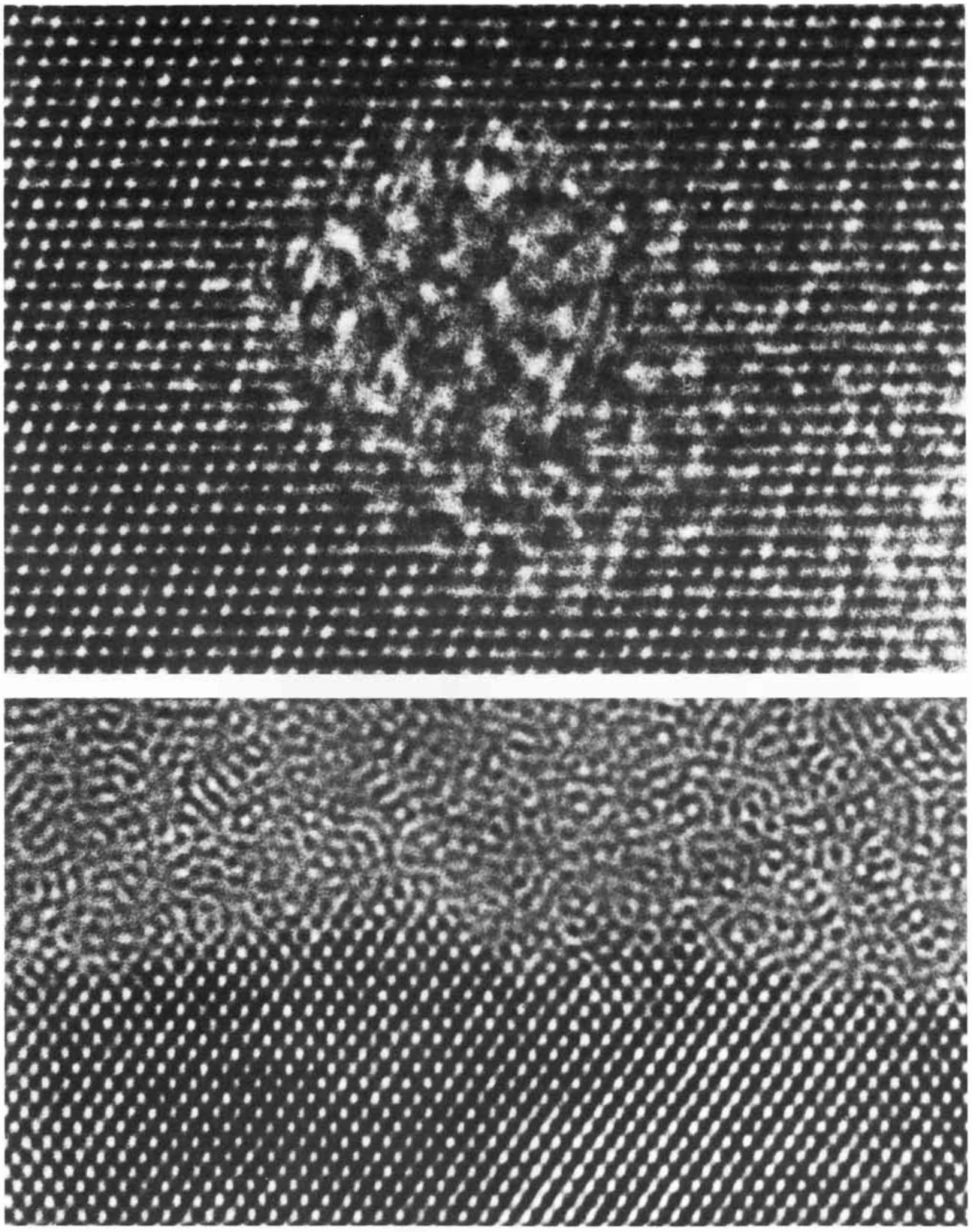
Because the electrical operations of integrated circuits take place in very thin layers, a method was needed that offered precise control over the number and depth of the dopant atoms. Using beams of energetic ions like those generated by physicists and chemists for fundamental research, semiconductor workers found they could dope silicon with the necessary precision. The process revolutionized the microelectronics industry, making possible the development and manufacture of low-power, high-speed semiconductor devices that today are familiar in simple hand calculators and sophisticated computers. We and other investigators have been exploring the possibilities for ion-beam alteration of other materials: metals and insulating materials such as ceramics and glass. We are finding that the technique has promise both as a commercial process and as a tool for materials science.

Although the goal is different, the apparatus used in the ion-beam treatment of materials differs little from that used to probe nuclear interactions. The ions originate at one end of an accelerator in a chamber in which electrons boil from a heated filament and are accelerated in electric fields. If the element to be ionized is available in gaseous form, as is elemental argon or carbon in the form of carbon dioxide, it is introduced directly into the chamber. There collisions with the high-velocity electrons strip away the element's own electrons, ionizing the gas to create a plasma. Volatile solids such as tin can be heated in an oven opening off the plasma chamber in order to vaporize them; the energetic electrons then ionize the vapor.

For nonvolatile solids the process is more complex. A plate of the material is placed in the chamber and a gas such as argon is ionized. The plate is given a negative charge; it therefore attracts the positive gas ions, which collide with the surface of the material, dislodging some atoms directly and heating the surface enough to vaporize other atoms. Electron collisions ionize the freed atoms.

An electric field then extracts the plasma from the chamber. Most uses require an ion beam of fairly high purity. To that end a magnet encircling the tube deflects the desired ions by a precise angle. Lighter ions are deflected further, out of the path of the beam. Heavier ions also stray from the beam; they are deflected by smaller angles than the wanted ions. A series of electric fields focus and accelerate the purified beam to its final energy, sending it into the implantation chamber, where it strikes the target material. Because the beam is narrow, typically only a few centimeters in diameter, it must be swept across the surface of the target to implant the material evenly. In many cases electrodes flanking the entrance to the implantation chamber sweep the beam from side to side; in other arrangements the target is moved instead.

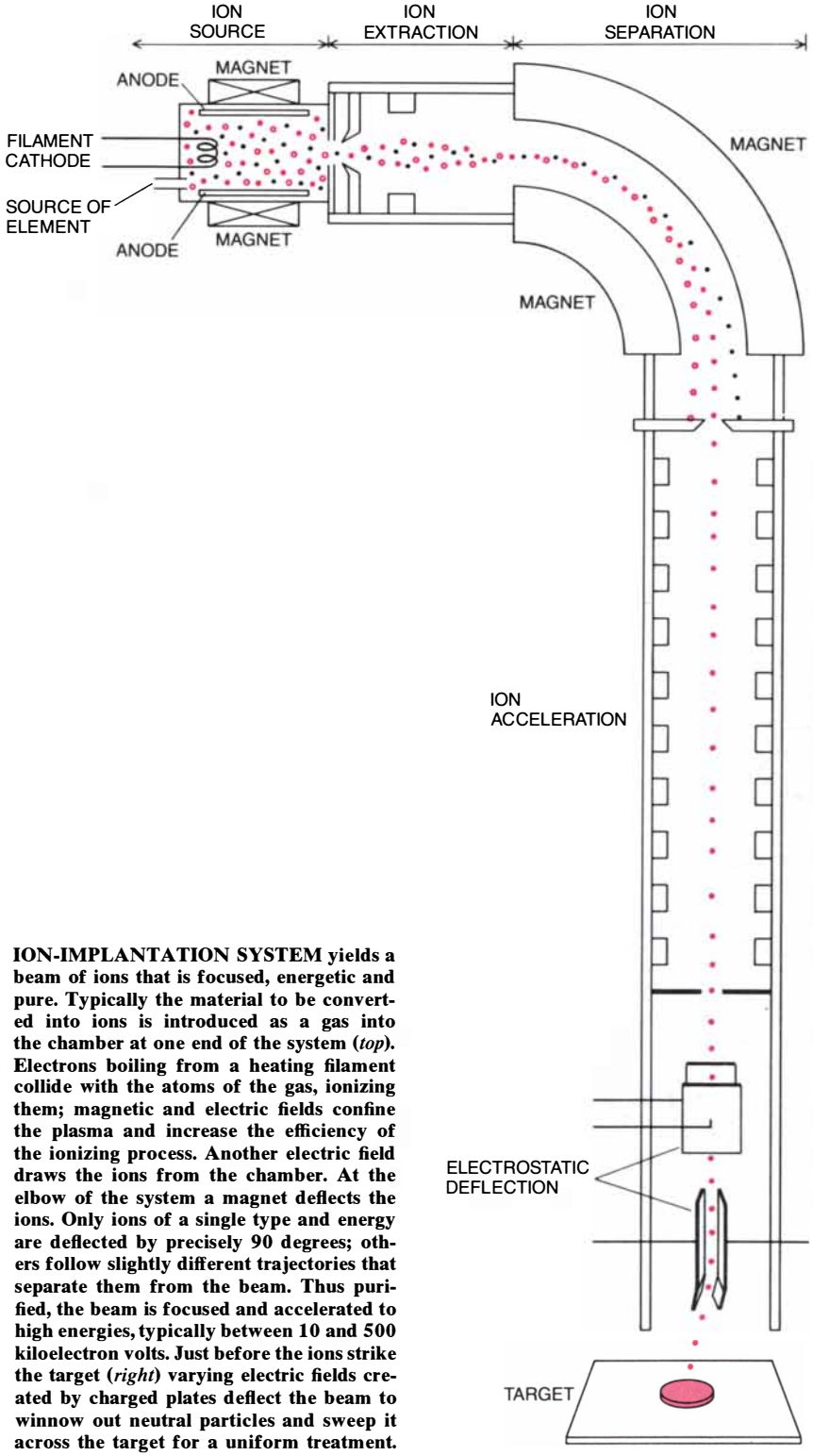
As the ions enter the target material they lose energy through collisions with electrons and nuclei and quickly come to rest. The most probable penetration depth, referred to as the projected range, can be calculated from



TRANSFORMED CRYSTAL STRUCTURES are one effect of ion implantation. These transmission electron micrographs of silicon, made by J. Narayan of the Oak Ridge National Laboratory, show the crystal lattice in cross section. Each bright spot corresponds to a chain of atoms in the crystal lattice. The fuzziness near the center of the top image shows the loss of crystal regularity that resulted when an ion of bismuth, accelerated to an energy of 100 kiloelectron volts, entered the lattice. If such bombardment is continued, disorder spreads throughout the implanted surface layer of the crystal, transforming it into an amorphous state. The bottom image shows the sharp boundary between an amorphous region and the undisturbed crystal structure below it. The mottled texture in the amorphous silicon is an artifact of the imaging process and does not correspond to the atomic structure of the ion-implanted layer.

the type and energy of the ions and the characteristics of the material; for typical ion-beam energies of from 10 to 500 kiloelectron volts, projected ranges vary between 10 and 1,000 nanometers (billionths of a meter), depending on the type of ion and host material. Individual ions, of course, do not come to rest precisely at the projected range. Instead their depths fall in a distribution that can be represented as a bell curve centered on the projected range. The width of the curve, characterized by the standard deviation of the depths around the projected range, is known as the range spread.

It is possible to anticipate, and therefore to control, not only the depth distribution of the implanted atoms but also the change in composition they produce in the host material. The calculation takes into account the ion flux of the beam, the number of atoms per unit of volume in the target and the exposure time. Suppose the composition of a 100-nanometer-thick surface region containing 5×10^{22} atoms per cubic centimeter is to be altered by 10 parts in a million. At an ion-beam flux of 10^{14} ions per square centimeter per second an exposure of a twentieth of a second per unit of area would suffice. A 20 percent change in composition would require about 20 minutes of exposure for every unit of area. To achieve a uniform distribution of implanted atoms throughout the thickness of the treated region the energy of the ion beam can be varied so that the individual distribution curves add in a nearly constant distribution profile.



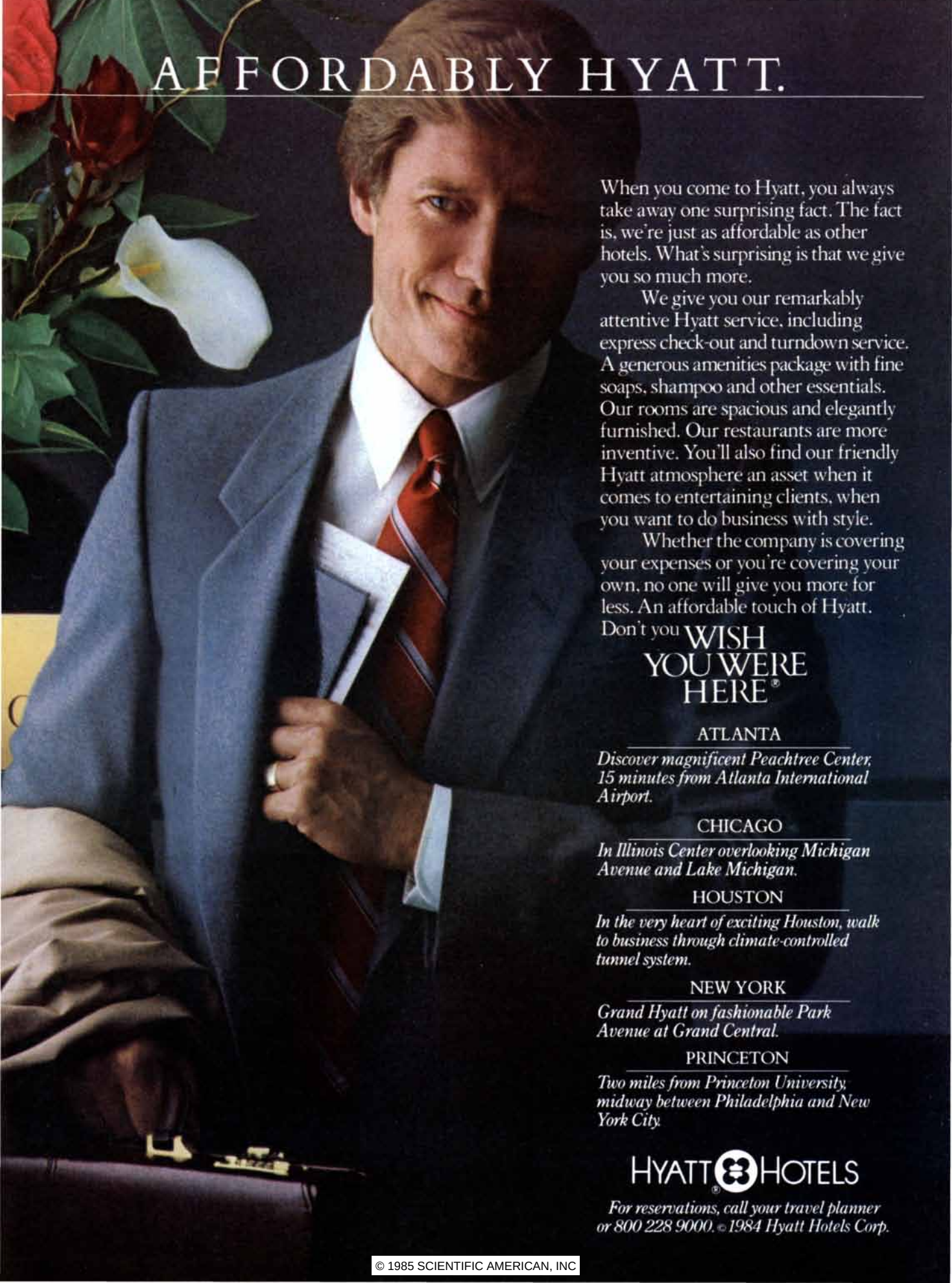
ION-IMPLANTATION SYSTEM yields a beam of ions that is focused, energetic and pure. Typically the material to be converted into ions is introduced as a gas into the chamber at one end of the system (*top*). Electrons boiling from a heating filament collide with the atoms of the gas, ionizing them; magnetic and electric fields confine the plasma and increase the efficiency of the ionizing process. Another electric field draws the ions from the chamber. At the elbow of the system a magnet deflects the ions. Only ions of a single type and energy are deflected by precisely 90 degrees; others follow slightly different trajectories that separate them from the beam. Thus purified, the beam is focused and accelerated to high energies, typically between 10 and 500 kiloelectron volts. Just before the ions strike the target (*right*) varying electric fields created by charged plates deflect the beam to winnow out neutral particles and sweep it across the target for a uniform treatment.

The violent intrusion of foreign ions into the lattice of a crystalline material damages it in a variety of ways. Host atoms are knocked out of their lattice positions, leaving lattice vacancies; displaced host atoms and implanted ions can lodge between the host's atomic planes as interstitials. Such point defects can accumulate to form clusters of vacancies or of interstitials; large numbers of defects can congregate at a single atomic plane in a flat assemblage known as a dislocation loop. With continued damage these loops can grow and intersect to form networks of dislocation lines.

The disorder in the crystal lattice is known as displacement damage or radiation damage; the damage left in the track of a single ion is called a collision cascade. When both the ion and the host atoms have high atomic numbers and heavy nuclei, the damage mounts quickly; it accumulates more slowly when the material and the ions have lighter nuclei. As a general rule one atom is displaced for each atomic plane an ion penetrates. Ion treatment sufficient to change the composition of a layer by 1 percent will generally displace every atom within the implanted depth several times.

In a few cases displacement damage yields useful changes in electrical or mechanical properties. More often radiation damage and its effects are unwelcome. In semiconductors dopant atoms that come to rest between atomic planes as interstitials cannot serve their proper role as electron donors or acceptors, whereas vacancies interfere with the activity of the semiconductor by capturing electrons or positive charges. Achieving optimal properties requires that the dopant ions be incorporated into the crystal lattice and the vacancies be removed.

Many point defects are mobile and



AFFORDABLY HYATT.

When you come to Hyatt, you always take away one surprising fact. The fact is, we're just as affordable as other hotels. What's surprising is that we give you so much more.

We give you our remarkably attentive Hyatt service, including express check-out and turndown service. A generous amenities package with fine soaps, shampoo and other essentials. Our rooms are spacious and elegantly furnished. Our restaurants are more inventive. You'll also find our friendly Hyatt atmosphere an asset when it comes to entertaining clients, when you want to do business with style.

Whether the company is covering your expenses or you're covering your own, no one will give you more for less. An affordable touch of Hyatt.

Don't you WISH YOU WERE HERE®

ATLANTA

Discover magnificent Peachtree Center, 15 minutes from Atlanta International Airport.

CHICAGO

In Illinois Center overlooking Michigan Avenue and Lake Michigan.

HOUSTON

In the very heart of exciting Houston, walk to business through climate-controlled tunnel system.

NEW YORK

Grand Hyatt on fashionable Park Avenue at Grand Central.

PRINCETON

Two miles from Princeton University, midway between Philadelphia and New York City.

HYATT® HOTELS

For reservations, call your travel planner or 800 228 9000. © 1984 Hyatt Hotels Corp.

SCIENCE/SCOPE

Communications satellites are helping Indonesia unite its 150 million inhabitants, according to a study conducted by the East-West Center in Honolulu. Indonesia's population, spread across a vast archipelago of over 13,600 islands, speaks more than 250 languages. Satellites have helped bridge the gap between the urban minority and the rural majority through improved telephone and television service. Since Indonesia inaugurated Asia's first and the world's third domestic communications satellite system in 1975, the number who can speak the national language, Bahasa Indonesia, has risen from 69% to 91%. Educational TV programs broadcast via satellite to rural areas have also prompted improved farming methods, including the use of new varieties of rice. Indonesia's Palapa satellite system was designed and built by Hughes Aircraft Company.

A klystron amplifier tube designed to operate 10 years—three times the current design life—is improving reliability and reducing life-cycle costs in Hughes' AML terrestrial microwave signal distribution systems. Key to its longevity is a coating layer of osmium ruthenium alloy for the tube's cathode, which allows the amplifying electrons to be emitted at lower temperatures. The resulting tenfold reduction in evaporation of barium from the cathode extends operating life. The technique, an outgrowth of work done for satellite communications amplifiers, is finding application with cable television companies.

An RF-excited waveguide carbon-dioxide laser has been introduced for use in medicine and industry. The compact 20-watt laser, designated the Series 3900, is the first RF-excited laser to be commercially available from Hughes. It features stable output and a clean mode for excellent control of the beam's cutting edge. "Hardseal" construction and advanced metal-ceramic processing techniques insure long life. The tube has a hardseal gas valve to permit factory refill. The laser is available in three configurations—air-cooled, water-cooled, and a bare tube that measures 1.75 inches in diameter, about the size of a conventional helium-neon laser.

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

Hughes is seeking experienced engineers and scientists to further develop advanced spacecraft systems and components for commmunications satellites—successors to the 20 that will have been launched from the space shuttle by 1986. Openings are in the fields of: software, computers, and data processing systems; components and component survivability; microwave communications; space electronics; control electronics; spacecraft design and integration; engineering mechanics; propulsion and electrical power; guidance and control; spacecraft manufacturing; and systems test and development. Send your resume to Dan Frownfelter, Hughes Space & Communications Group, Dept. S2, S4/A300, P.O. Box 92919, Los Angeles, CA 90009. Equal opportunity employer. U.S. citizenship required.

For more information write to: P.O. Box 11205, Dept. 66-9, Marina del Rey, CA 90295

HUGHES
AIRCRAFT COMPANY

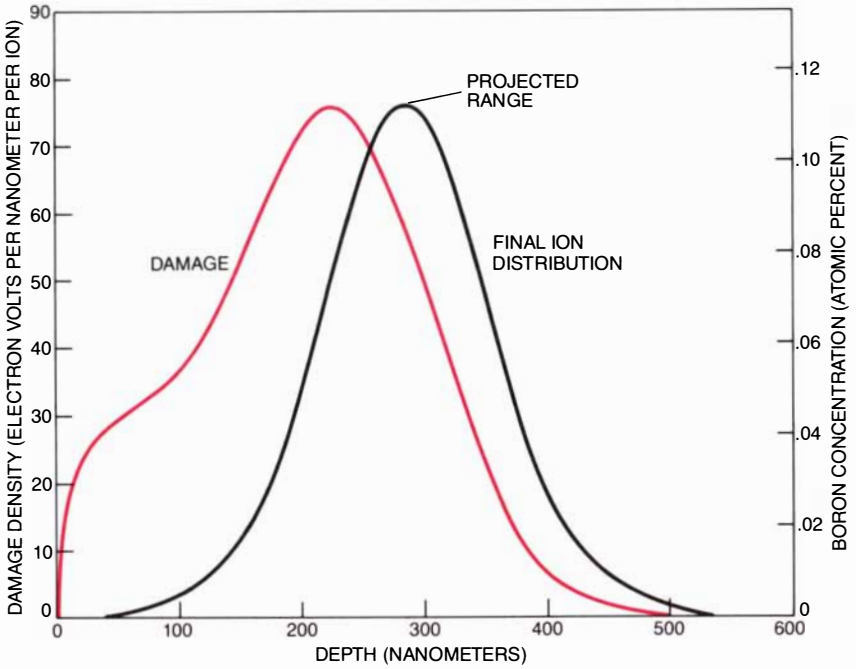
short-lived at room temperature. Interstitials fill vacancies or migrate to the surface. Defect clusters and loops, however, are usually stable and must be removed through annealing: heating the material to foster the rearrangement of atoms into the energetically more favorable configuration of a well-ordered crystal lattice.

One or several steps of ion implantation, followed by annealing to incorporate the dopant atoms into the silicon crystal's lattice, have become part of the standard manufacturing process for microelectronic circuits. The electrical basis of such devices is the junction between semiconductor regions that differ in their electrical properties. One region contains dopant atoms that readily release electrons, fostering electrical conductivity; its activity is known as *n*-type, or negative, conduction. The adjacent region contains a different dopant whose atoms readily accept electrons from the host material, leaving positively charged holes within the host material's sea of electrons. The holes are mobile, giving rise to *p*-type, or positive, conduction. The conjunction of regions of *n*- and *p*-type electrical activity creates an electric field, vital to the operation of semiconductor circuits.

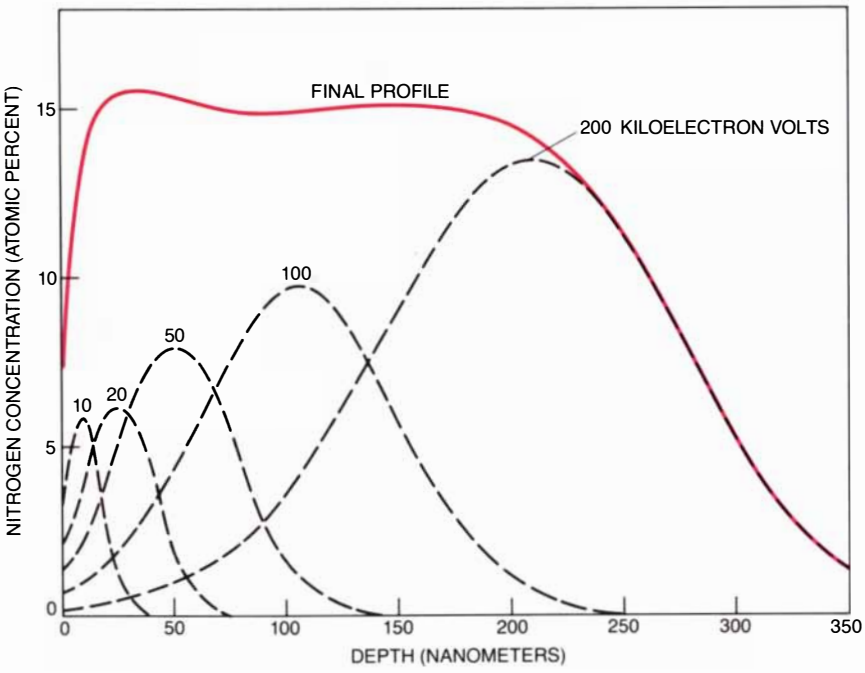
Ion beams supply the dopants; in silicon, phosphorus or arsenic is usually used for *n*-type activity and boron for *p*-type. Before each round of ion treatment the silicon is coated with a layer of a masking material that is sensitive to light or to electrons. Electron-beam lithography or photolithography is used to transfer the circuit pattern to the masking. Chemicals strip away the exposed areas, leaving the rest of the surface shielded from the ions. After implantation the chip is heated to between 600 and 1,000 degrees Celsius to anneal it.

With this basic process well established, workers are now exploring ion-beam technology for use with new generations of semiconductor devices. The manufacture of the Very Large Scale Integrated circuits (VLSI) currently being developed may dispense with the masking step, for example. Laying out the dense, fine-line circuit features needed for large-scale integration demands resolution close to the limits of electron beams and visible light. Sharply focused ion beams offer much higher resolution; ion beams could in principle be used directly, without intervening steps of masking, to dope features less than a micrometer across, whose electrical activity might be controlled by as few as 100 dopant atoms.

Another new technology, that of



DEPTH DISTRIBUTION of ions implanted at a uniform energy takes the form of a normal curve (*black curve*). The depth at which the curve peaks, known as the projected range, varies with the energy of the beam and the type of ions and host material. The collisions that stop incident ions also displace atoms, damaging the crystal structure of the host material. The collisions occur at depths shallower than those at which the ions finally lodge. The colored curve shows the distribution of the damage, expressed as electron volts of energy absorbed per incident ion by atomic collisions in each nanometer of material thickness. The curves are for silicon implanted with a total of 10^{15} boron ions per square centimeter of surface; the beam used in the implantation process had an energy of 100 kiloelectron volts.

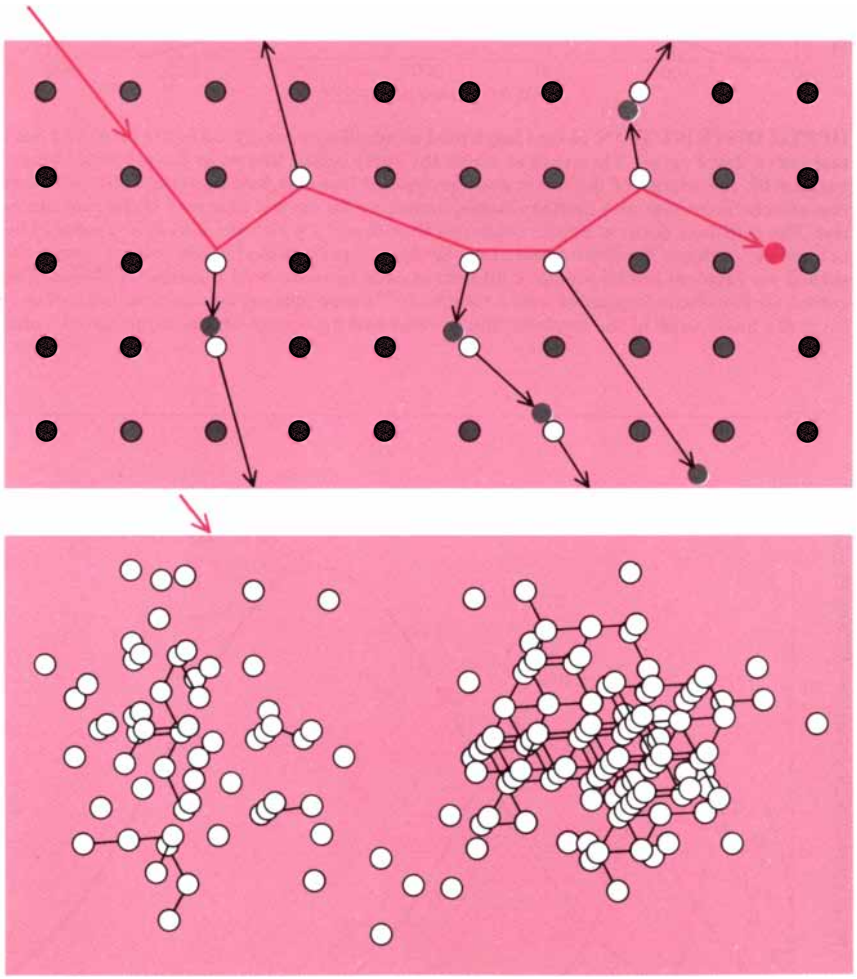


UNIFORM CONCENTRATION OF IONS in a surface layer can be achieved by varying both the energy of the implanting beam and the number of ions delivered at each beam energy. Each broken-line curve represents the depth distribution of nitrogen ions implanted in iron at a single beam energy. In successive treatments both the beam energy and the number of ions implanted were increased. The solid curve, which represents the sum of the individual concentration profiles, shows that the stepwise treatment yielded a concentration that was uniform from just below the surface to a depth of about 200 nanometers.

compound semiconductors, will call for more complicated annealing processes. Such semiconductors promise faster circuits: they consist not of a single element but of a compound, such as gallium arsenide, through which electrons and positive holes move more quickly. Within the crystal structure the two or more host elements make up interlocking sublattices, and each element must be returned to its respective sublattice following the atomic displacements resulting from ion implantation. The annealing can be complicated further when one of the elements is volatile, as arsenic is. During extended annealing some of its atoms may escape, upsetting the chemical proportions of the semiconductor and altering its electrical properties. Promising techniques for overcoming the difficulty include the use of capping layers to retain the more volatile element, rapid, high-temperature annealing using optical pulses and annealing in an atmosphere containing an overpressure of the volatile element.

The possibility of using ion beams not only to treat a semiconductor for a specific type of electrical activity but also to produce a new semiconductor medium is emerging. For example, the microminiaturization accompanying the development of VLSI requires that electrical activity be confined to exceedingly thin layers of semiconductor and that electrical interference between adjacent elements in an integrated circuit be minimized. One way to do so is to create circuits in a medium known as silicon-on-insulator (SOI), in which a layer of silicon perhaps half a micrometer thick rests on an insulating substrate. SOI circuits would need less power than present circuit designs, work faster and be less sensitive to radiation (a consideration important in space and defense applications). They are also more resistant than other miniaturized circuits to latch-up, a disruptive process in which current flows between neighboring transistors.

The difficulty of growing sufficiently thin single-crystal silicon layers on an insulator has slowed adoption of SOI technology. Ion implantation circumvents the problem. A layer of silicon dioxide, an insulator, can be created when oxygen ions are implanted in a silicon crystal under appropriate conditions. Careful regulation of the energy of the implanting beam and the temperature of the silicon makes it possible to form the insulating region not at the surface of the silicon crystal but half a micrometer below it. Microcircuit elements can then be fabricated in the resulting film of electrically isolated, unaltered silicon.



DAMAGE to the host material's crystal structure takes two basic forms, as is shown schematically at the top. Atoms of the host material are knocked out of place by each energetic ion, leaving vacancies (*circles*), and the displaced atoms and the implanted ions often lodge between the atomic planes of the crystal as interstitials. When both the ion and the host substance have heavy nuclei, a single incident ion can leave extensive damage. The bottom image shows the distribution of vacancies that resulted when a single molybdenum ion penetrated the surface of a tungsten needle; the lines connect vacancies on adjacent lattice sites. The arrow indicates the ion's approximate point of entry. The three-dimensional distribution was mapped by Ching-Yeu Wei and David N. Seidman of Cornell University. They used a technique known as field evaporation to expose successive atomic layers on the needle surface; they imaged the structure of each exposed layer using field ion microscopy.

In semiconductors ion implantation is used almost exclusively to alter electrical properties. In metals changes in mechanical and chemical properties promise to be of first importance. The changes are all the more valuable because they do not come at the cost of altered dimensions or bulk properties, nor do the altered surfaces suffer the problems of adhesion that often plague surface coatings.

An ion-implanted surface can transform a metal, increasing its hardness and resistance to wear, oxidation or corrosion, lowering its coefficient of friction and modifying a range of other properties, including magnetic, superconductive and optical behavior. Whereas dopant concentrations of a few parts per million bring about *n*- and *p*-type electrical activity in semiconductors, significant changes in the surface properties of a metal usually require the concentration of implanted ions to reach several atomic percent. Altered properties across a useful area of metal therefore require long exposure times or an accelerator capable of delivering a large flux of ions.

Of the surface properties over which ion beams offer control, wear has received the most attention. It is a logical choice, since most of the important processes that control wear take place in the outermost micrometer of material—the region ion implantation affects directly. Moreover, many traditional treatments known to improve surface hardness and wear resistance also involve the introduction of foreign elements into the near-surface re-

gion of the material. One such treatment is the nitriding of steel, commonly used for bearing surfaces, valve parts and cylinder linings in engines and for other parts whose use requires a hard surface. In nitriding the object is heated to about 500 degrees C. in an atmosphere that is rich in nitrogen, usually in the form of ammonia. Nitrogen diffuses into the surface, where it reacts with the iron to form nitride precipitates—small crystals of nitride compounds within the crystal structure of the steel.

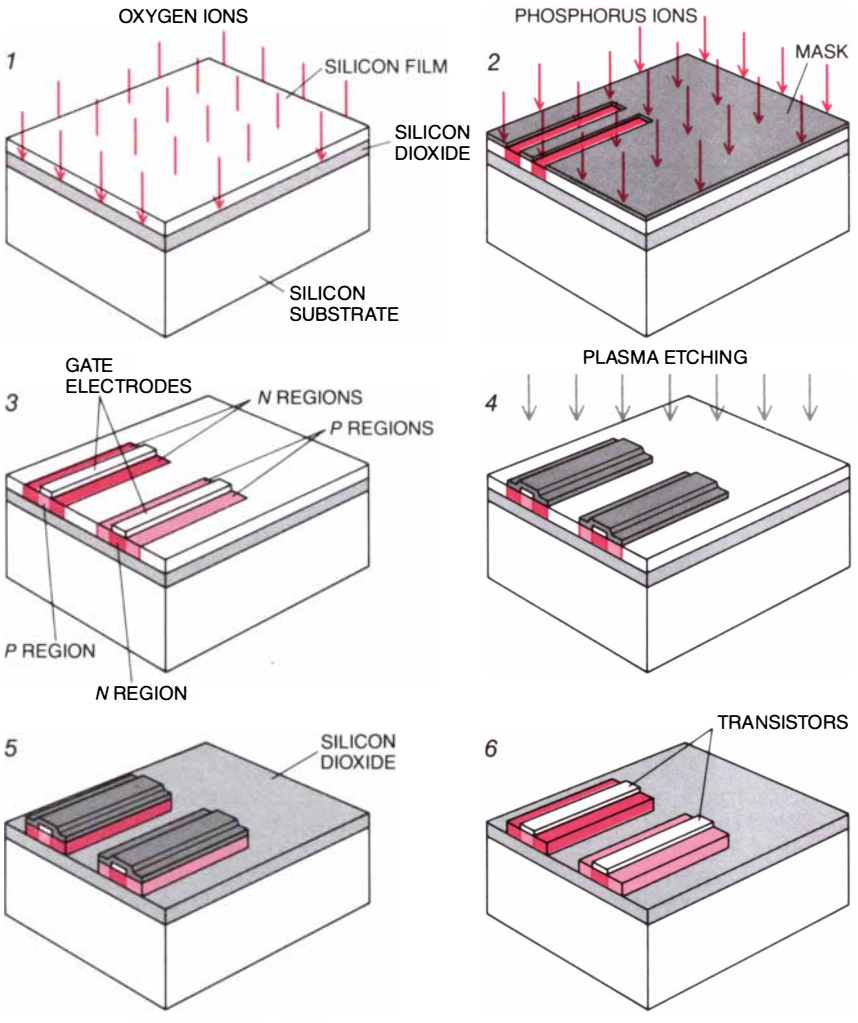
Geoffrey Dearnaley and his associates at Harwell in England pioneered the introduction of nitrogen by means of ion beams rather than diffusion, and other groups elsewhere have extended their work. Nitrogen ions are easier than most to supply in large numbers with an ion accelerator, because nitrogen is available as a gas and is easily ionized. As a result reasonable treatment times produce the 20 to 40 atomic percent of nitrogen needed to improve surface hardness and wear resistance. The implanted nitrogen ions either are incorporated singly into the crystal lattice, in which case they are said to be in solid solution, or take the form of nitride compounds, dispersed as fine precipitates.

Tests in which an implanted steel disk is rotated under a steel pin that makes contact with the surface at an adjustable load show that, in relation to untreated steel, wear drops by a factor of as much as 10. The degree of improvement varies, however, with the elemental composition of the steel, its prior heat treatment and its crystal structure; very hard bearing steels, for example, show no improvement at all. Moreover, in some steels combined nitriding and implantation can give even more striking improvements.

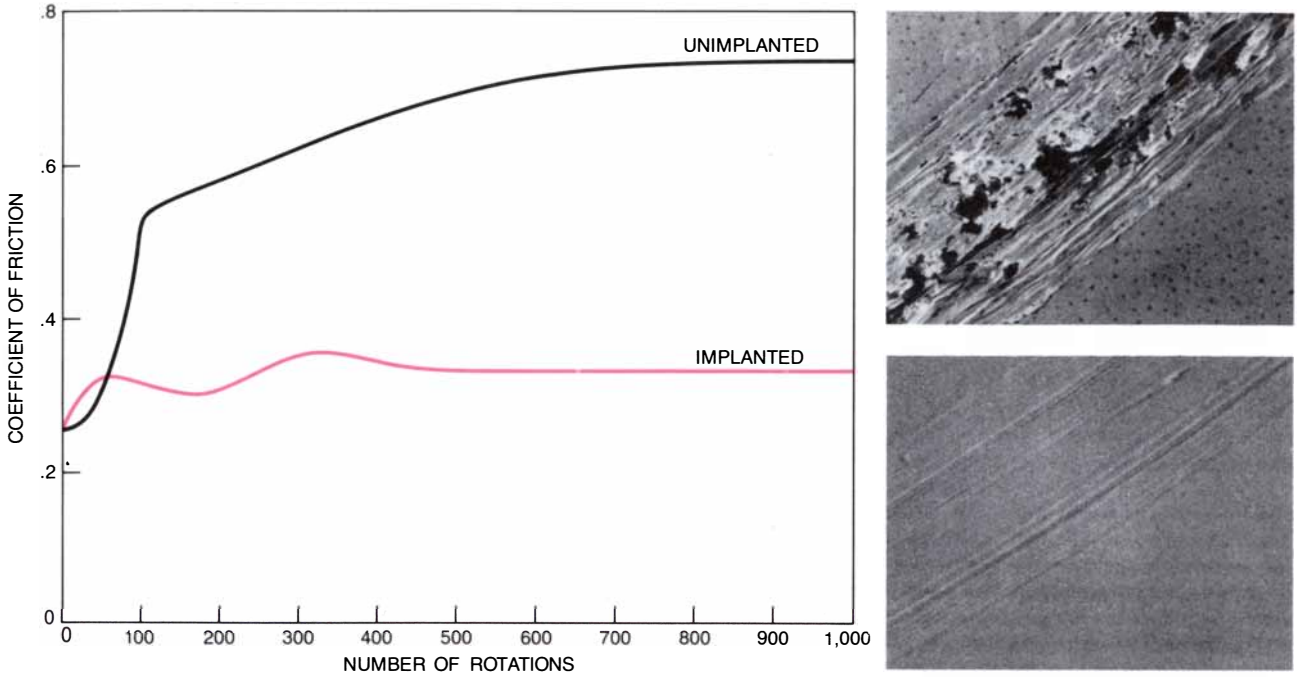
On an atomic level the increased wear resistance is not understood fully. The introduction of foreign atoms and crystal lattices interrupts the regularity of the steel's crystal structure. The disruption may increase the hardness of the steel by helping it to resist dislocation flow, in which surface stresses cause individual atomic layers to slip past one another atom by atom. Rate of wear reflects not only the hardness of a surface, however, but also specific mechanisms of wear, in many cases poorly understood, such as abrasion and adhesion.

In other metals nitrogen implantation is believed to change the dominant mechanism of wear. Implanted nitrogen slows the wear of layers of chromium plating by factors of up to 100. Nitrogen implantation probably causes the layer of chromium to expand, thereby closing cracks, thought to be the weak points at which wear begins. Dramatic improvements also occur in a material known as Ti-6Al-4V, a titanium alloy that is used in aerospace structures and contains 6 percent aluminum and 4 percent vanadium. In experiments unlubricated, nitrogen-implanted surfaces wore 1,000 times more slowly than untreated surfaces; friction was also reduced, to one-third its normal level. The implanted nitrogen promotes the formation of a surface oxide layer that seems to act as a lubricant. The lubrication slows the process of micromachining that occurs as two surfaces rub together.

Other instances of increased wear resistance reflect a change in the phase of a material. When workers at the Sandia National Laboratories and the Naval Research Laboratory treated stainless steel with beams of carbon and titanium ions in order to produce a surface alloy containing 20 percent of each element, the steel's surface region lost its crystalline order and was transformed into an amorphous, or glassy, state—a phase of steel that cannot be formed by other means. A drop of 50



SILICON-ON-INSULATOR CIRCUIT is made using beams of ions in an exploratory process. In the first step a beam of oxygen ions bombards a crystal of silicon; the ions come to rest below the surface and combine with the silicon to form a layer of silicon dioxide, an insulator (*1*). After the silicon is annealed to repair the displacement damage left in the crystal by the passage of the ions, the circuits are fabricated in the thin surface film of electrically isolated silicon. The silicon is coated with a photosensitive masking to which a pattern is transferred by means of photolithography. The regions of masking exposed to light are etched away, uncovering appropriate areas of the silicon for ion-beam implantation of electrical dopants such as boron and phosporus to produce *n*- and *p*-type electrical activity (*2*). Repeated steps of masking and ion treatment and the deposition of gate electrodes define circuit elements, in this case two transistors (*3*). New masking is applied to protect the transistors, and the surrounding unaltered silicon is etched away (*4*). The transistors, each of them about 500 nanometers thick and a few micrometers wide, are left as islands of silicon on the insulating silicon dioxide (*5*). After the masking is stripped away (*6*) a complete circuit is formed by interconnecting a large number of transistors made in the same way.



REDUCED FRICTION can follow ion implantation. Graphs show the coefficient of friction (the force needed to slide a surface across another surface divided by the force with which the surfaces are pressed together) for two stainless-steel plates that were rotated under a stainless-steel pin without lubrication; the study was done by David M. Follstaedt, Larry E. Pope and their colleagues at the Sandia National Laboratories. One of the plates (*black curve*) was untreated; the other one (*colored curve*) had been bombarded with carbon and titanium ions, transforming the steel surface from crystalline order into an amorphous, or glassy, state. After about 100 rotations the coefficient of friction of the implanted surface stabilized at less than half the value for the untreated surface. Lowered friction slowed wear on the implanted steel. Magnified 500 times in a scanning electron micrograph (*bottom*), the implanted surface appears smooth after 1,000 rotations, in contrast to the rough aspect of the untreated steel in a similar electron micrograph (*top*).

percent in the steel's coefficient of friction resulted, and its wear rate slowed by a factor of about 10.

These and other implantation processes for reducing friction and wear in metals are already attracting commercial attention. Interest focuses on ion treatment for critical components such as bearings in jet engines and in gyroscopes and for precision tools for industrial processes, in which the cost of a part or of stopping production to replace it makes longevity crucial. Workers in the U.S., Britain and elsewhere have found that implantation of nitrogen ions lengthens the lifetime of metal-forming tools made of steel and of cobalt-cemented tungsten carbide from two to five times. Precision dies for plastics face abrasive wear from the fibers and mineral particles used as fillers in the plastics; they respond even more dramatically to ion treatment, showing tenfold increases in longevity.

Titanium-alloy hip-joint prostheses, whose longevity is crucial, may also benefit from implantation. In an environment of body fluids, the motion of the metal ball joint in the polyethylene socket can degrade the joint severely. Investigators have found that implanted nitrogen can slow the wear of the alloy when it is in contact with polyethylene by factors of up to 1,000.

Parts whose roles expose them to wear often face corrosion as well; plastic-forming dies, for example, are both abraded and attacked by residual chemicals, and the degradation of hip prostheses results from corrosion by body fluids as well as from wear. Electrochemists have found that the implantation of ions can bolster a material against both corrosion and wear. When 5×10^{16} atoms of platinum per square centimeter are implanted in stainless steel, a surface layer of alloy forms that can protect the steel from attack by a 20 percent solution of sulfuric acid during 80 days of immersion. Although platinum is a precious metal, the treated layer is so thin that the cost of the implanted platinum is only a few hundredths of a cent per square centimeter.

The surface modification of metals by ion implantation appears to be on the threshold of widespread commercialization, both in its conventional form and as a modified technique known as ion-beam mixing. In this process a film of material is deposited as a vapor; simultaneously or subsequently the surface is exposed to a beam of ions. Collision cascades mix the deposited material with the host. Small numbers of ions suffice to break down the boundary between the film and the substrate, bonding the coating tightly; larger numbers of ions can mix the constituents thoroughly, forming an alloy. Because the process usually requires fewer ions than are necessary to create a layer of alloy through implantation alone, it is more economical than conventional implantation.

Ion-beam modification is only beginning to fulfill its promise as a commercial surface treatment for metals, but it has already shown its worth as a tool for research. Ion beams can be used to simulate physical and chemical stresses, allowing laboratory observation of the response of metals. In studies of the mechanisms of corrosion workers have used the technique to introduce single varieties of ions into metal surfaces in controlled numbers in order to determine the roles individual chemical species play in corrosion.

Ion beams have also served to simulate the radiation and heavy neutron bombardment found in the interior of breeder reactors. In the 1970's investigators discovered that the neutrons caused metals used in experimental breeder cores to swell, increasing in volume by several percent and interfering with close mechanical tolerances. It appeared that the neutrons left dense displacement cascades. At reactor operating temperatures the vacancies accumulated to form small voids

Chivas Brothers introduces Royal Paisley Scotch.

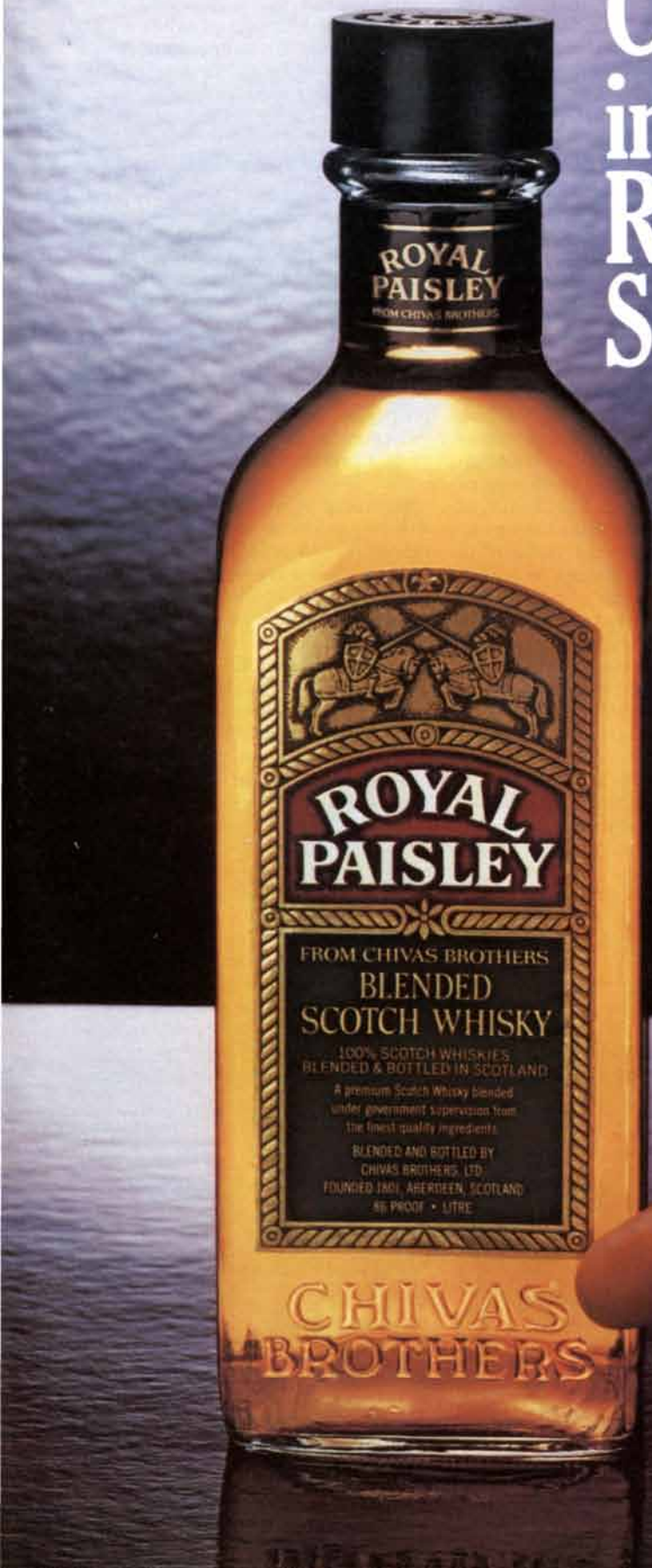
Scotch drinkers have come to know that when you see the words "Chivas Brothers" on the outside of a bottle, you're sure to find exceptional Scotch on the inside.

And Royal Paisley is no exception.

Made from Scotland's most prized whiskies, Royal Paisley is remarkably rich and incomparably mellow. Yet it's priced at about $11 a bottle.*

Best of all, Royal Paisley offers you something only a few of the world's finest Scotches possess.

The name Chivas Brothers.



The taste is certified right on the bottle.

*750 ML size. Price varies slightly by locale.

throughout the metal. The search for alloys resistant to such swelling required a means of testing them in the laboratory. Ion bombardment, with its capacity for producing the same kind of damage far more quickly than the neutrons themselves, was well suited to the purpose.

Ion beams have aided similarly in the evaluation of materials for the inner walls of the plasma devices used in fusion research. Exposure to hot hydrogen plasma erodes most metals: the energetic ions collide with surface atoms and transfer enough momentum to dislodge them completely. This process, known as sputtering, can cause the container wall to recede and can contaminate the plasma. Ion beams provided a way to re-create the process in the laboratory so that resistant alloys could be identified.

Because it can permeate a surface with almost any concentration of another material, ion implantation has proved a powerful tool for investigating the interactions of substances in alloys. Accurate measurements of the rate at which foreign elements diffuse through a metal, for example, can ordinarily be done only at high temperatures. Because of the precise control it offers over the depth at which the ions come to rest, however, ion implantation can produce greater concentrations and steeper concentration gradients than conventional alloying. As a result measurable diffusion, evident in a change in the depth distribution of the implanted atoms, occurs at temperatures closer to those the metals face in everyday uses.

The measurement technique relies on the fact that for a given implantation energy the implanted atoms' depth distribution can be measured with high precision. Implantation of the foreign ions is the first step; the depth profile of the implanted atoms is then determined by means of a technique called ion-beam scattering. A high-energy beam of helium ions is trained on the metal. A small proportion of the ions will collide with nuclei and rebound, at energies characteristic of the mass of the nuclei they struck. Since the helium ions lose energy at a known rate as they travel through the metal, the depth at which the collisions took place can be determined. The depth profile of the implanted atoms is constructed from the number of scattered helium ions and their energy.

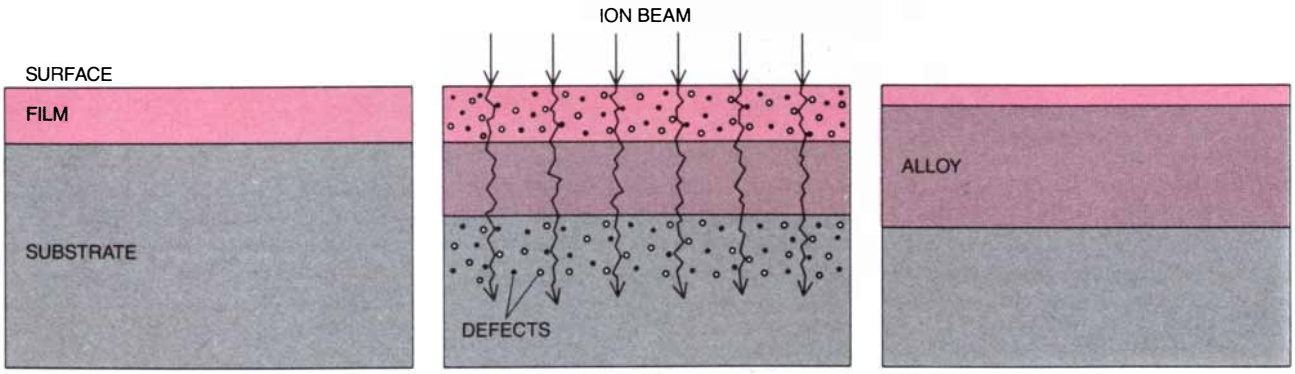
The metal is then heated to a specified temperature for a preset length of time. Once it has cooled it is again subjected to ion-beam scattering, yielding a new depth profile. The diffusion that took place when the metal was heated will be reflected in a broadening of the distribution curve. The degree of broadening and the time and temperature in which it occurred give the diffusion rate with far greater accuracy than other means offer.

The extreme sensitivity of the technique makes it possible to measure long-term diffusion processes within a reasonable length of time. Temper embrittlement typifies such processes; it occurs in steel parts at moderate temperatures when unwanted elements such as antimony, tin and phosphorus accumulate at boundaries between the distinct crystalline grains of the metal, thereby weakening it. Ion implantation allows the migration and trapping of foreign elements to be observed and characterized in the laboratory at the temperatures to which metals are exposed in actual use.

Samuel M. Myers, Jr., and his colleagues at Sandia have demonstrated that ion implantation can also aid in the study of solid solubility, another metallurgical phenomenon. The technique enables investigators to introduce precise numbers of solute atoms into a metal at low temperature. When the concentration of the foreign substance exceeds its solubility, the atoms that can no longer be accommodated within the host metal's crystal lattice begin to form precipitates. The determination of the depth profile of the implanted atoms by ion-beam scattering, combined with the detection of precipitates through electron microscopy, allows the solid solubility of the implanted substance to be measured at a relatively low temperature. Such information is invaluable in developing new alloys, since an alloying element can have very different effects on metal properties depending on whether it is incorporated in the form of a solution or of a precipitate.

As emerging technologies call for new materials and place novel demands on old ones, new uses for ion implantation will emerge. Ion implantation is now showing potential as a means of tailoring the surfaces of insulating materials such as glass, ceramics and polymers. Among the specific technologies in which it may play a role are those of optical waveguides, magnetic-bubble memories, ferroelectric ceramics, high-temperature ceramics for engines, and polymers with conductive surfaces.

In waveguides made out of silica glass, for example, the implantation of nitrogen ions causes radiation damage and forms silicon nitride precipitates, both of which help to confine light within the waveguide by changing the refractive index of surface layers. Hydrogen ions, implanted in garnets used for magnetic-bubble memories, cause displacement damage that changes the material's innate magnetic properties. The implanted regions serve to confine



ION-BEAM MIXING reduces the number of ions needed to form a surface layer of alloy, thereby lowering the cost. The substance to be alloyed with the host metal is applied as a surface film (*left*) by conventional means, often by deposition as a vapor. Accelerated ions such as nitrogen supply the energy needed to mix the substances (*center*). The mixing occurs as incoming ions collide with atoms of the film and the substrate, displacing them. The excitation of atoms in the implanted region and the defects left in the crystal structure by the bombardment also cause atoms to migrate. Atoms shift to fill vacancies in the lattice (*circles*) and to accommodate displaced atoms (*dots*) within the crystal structure. Extensive bombardment can thicken the mixed region until it approaches the surface (*right*).

the bubbles to particular propagation tracks in high-density memories.

Ion implantation also enhances the properties of ferroelectric ceramics, photosensitive ceramics that record images when simultaneously exposed to intense near-ultraviolet light and to an electric field. Radiation damage and the chemical alteration resulting from bombardment with ions of aluminum and argon lead to changes in electrical behavior that increase the material's photosensitivity by a factor of more than 10,000. The treatments also shift the sensitivity into the visible region of the spectrum. Implanted ceramics can record photographic images in sunshine or white light within reasonable exposure times. In other ceramics implantation alters mechanical properties. For all their hardness, substances such as titanium diboride and alumina are quite brittle, a trait that has limited their use. Workers at the Oak Ridge National Laboratory have now shown that the implantation of certain ions can toughen the surface regions of such ceramics.

Perhaps the most recent additions to the list of materials in which ion implantation has been explored are polymers. Most polymers are insulators, but when they are implanted, their conductivity can increase by as much as 14 orders of magnitude, to levels comparable to that of graphite. The increase in conductivity is thought to reflect changes in the molecular structure of the polymer and a sharp increase in its carbon content, which occurs as bonds linking the carbon to hydrogen and other volatile elements are broken, freeing the volatiles. A variety of polymers can be rendered conductive through implantation. The technique may make it possible to use a polymer for the electrical connectors that link elements in a microelectronic circuit. A polymer made conductive by implantation could also replace metal in shielding designed to protect sensitive instruments from electromagnetic interference.

It is too early to say which of ion implantation's many new applications will reach commercial fruition; one materials scientist referred to some ion-beam technologies as "solutions in search of a problem." Given the power of the technique, its role in future industry, however difficult to predict, is certain to be noteworthy. Even when it is not used to modify surfaces directly in commercial processes, ion implantation will affect the available range of materials and their uses indirectly, as ion-beam studies in the laboratory yield insights into the fundamental properties of materials.

The Chemistry of Garlic and Onions

A number of curious sulfur compounds underlie the odor of garlic and the crying brought on by slicing an onion. The compounds also account for medical properties long ascribed to garlic and onions

by Eric Block

The world has always been divided into two camps: those who love garlic and onions and those who detest them. The first camp would include the Egyptian pharaohs who were entombed with clay and wood carvings of garlic and onions to ensure that meals in the afterlife would be well seasoned. It would include the Jews who wandered for 40 years in the Sinai wilderness, fondly remembering "the fish which we did eat in Egypt so freely, and the pumpkins and melons, and the leeks, onions and garlic." It would include Sydney Smith, the 19th-century essayist, whose "Recipe for Salad" includes this couplet: "Let onion atoms lurk within the bowl, / And, scarce-suspected, animate the whole."

The camp of the garlic and onion haters would include the Egyptian priests who, according to Plutarch, "kept themselves clear of the onion.... It is suitable neither for fasting nor festival, because in the one case it causes thirst, and in the other tears for those who partake it." The camp would include the ancient Greeks, who considered the odor of garlic and onions vulgar and prohibited garlic and onion eaters from worshiping at the Temple of Cybele. It would include Bottom, who in *A Midsummer Night's Dream* instructs his troupe of actors to "eat no onions nor garlic, for we are to utter sweet breath."

Chemists must be included among the garlic and onion lovers. For them the reasons are professional: chemists have long been attracted to substances that have strong odors, sharp tastes and marked physiological effects. Investigations made by chemists over more than a century establish that cutting an onion or a garlic bulb releases a number of low-molecular-weight organic molecules that incorporate sulfur atoms in bonding forms rarely encountered in nature. The molecules are highly reactive: they change spontaneously into other organic sulfur compounds, which take part in further transformations. Moreover, the molecules display a remarkable range of biological effects. The lacrimatory, or tear-inducing, quality of an onion is only one example. Certain extracts of garlic and onions are antibacterial and antifungal. Other extracts are antithrombotic, that is, they inhibit blood platelets from forming thrombi (aggregations of themselves and the protein fibrin). In short, they act to keep blood from clotting.

Garlic and onions are members of the lily family. Their botanical names are *Allium sativum* and *Allium cepa* ("allium" may derive from the Celtic word *all,* which means pungent). Both are among the oldest of all cultivated plants: their origins, most likely in central Asia, predate written history. For thousands of years they have been a part of folk medicine. The Codex Ebers, an Egyptian medical papyrus dating to about 1550 B.C., gives more than 800 therapeutic formulas, of which 22 mention garlic as an effective remedy for a variety of ailments including heart problems, headache, bites, worms and tumors.

The Egyptians were not alone. Aristotle, Hippocrates and Aristophanes recommended garlic for its medicinal effects. The Roman naturalist Pliny the Elder cited numerous therapeutic uses for both garlic and onions. Dioscorides, chief physician to the Roman army in the first century A.D., prescribed garlic as a vermifuge, or expeller of intestinal worms. During the first Olympic games in Greece garlic is said to have been ingested by athletes as a stimulant.

In India garlic has served as an antiseptic lotion for washing wounds and ulcers. In China onion tea has long been recommended for fever, headache, cholera and dysentery. Folk medicine is often intertwined with legend, as in the case of Four Thieves' Vinegar. In 1721, the story goes, four condemned criminals were recruited to bury the dead during a terrible plague in Marseilles. The gravediggers proved to be immune to the disease. Their secret was a concoction they drank consisting of macerated garlic in wine, which immediately became famous as *vinaigre des quatre voleurs.* It is still available in France today.

Along with these folk prescriptions come more contemporary endorsements of garlic and onions. On one line of evidence they emerge as mild antibiotics. In 1858 Louis Pasteur reported that garlic is antibacterial. More recently Albert Schweitzer is said to have made use of garlic in Africa for the treatment of amoebic dysentery. In the two world wars garlic was used as an antiseptic in the prevention of gangrene. Laboratory investigations show that garlic juice diluted to one part in 125,000 inhibits the growth of bacteria of the genera *Staphylococcus, Streptococcus, Vibrio* (including *V. cholerae*) and *Bacillus* (including *B. typhosus, B. dysenteriae* and *B. enteritidis*). Moreover, garlic juice exhibits a broad spectrum of activity against zoopathogenic fungi and many strains of yeast, including some that cause vaginitis.

On another line of evidence garlic and onions emerge as antithrombotics. Again the evidence is both old and new. In France horses suffering from blood clots in the legs were fed garlic and onions. More recently, in 1979, G. S. Sainani and his colleagues at the B. J. Medical College of the University of Poona in India published the results of an epidemiological study of three populations that consumed differing amounts of garlic and onions. The subjects were vegetarians in the Jain community of India who ate garlic and onions in liberal amounts (at least 50 grams of garlic and 600 grams of onions per week), in smaller amounts (no more than 10 grams of garlic and 200 grams of on-

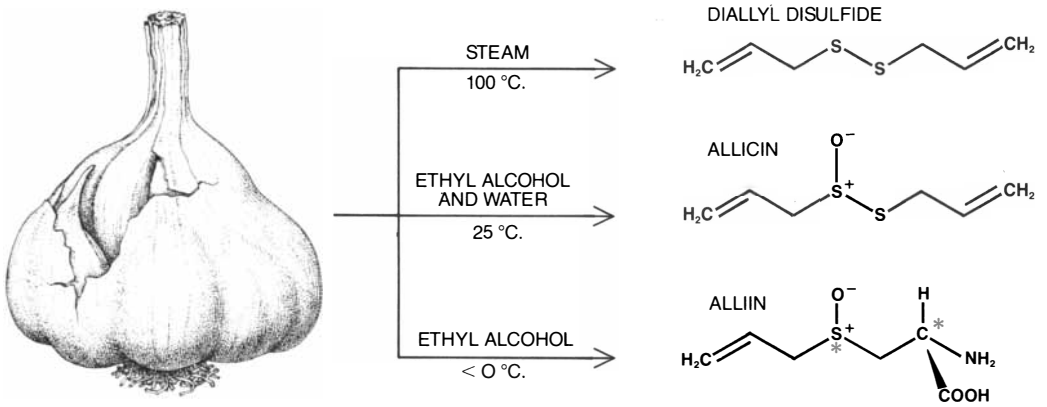
ions per week) or never in their lives.

The group of garlic and onion abstainers proved to have the shortest blood-coagulation time. In that group, moreover, the blood-plasma level of the protein fibrinogen proved to be the highest. (Part of the process of blood clotting is the conversion of fibrinogen into fibrin.) Studies done throughout the 1970's had already suggested that oils extracted from garlic and onions inhibit the aggregation of platelets. The folklore concerning garlic and onions seemed to be gaining some credence.

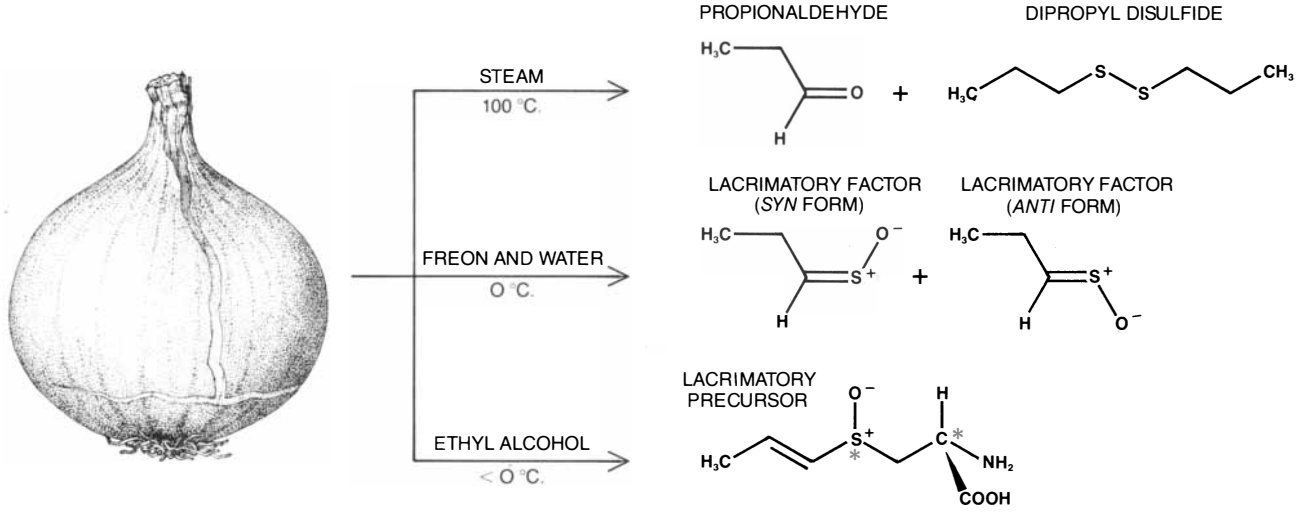
How do garlic and onions produce their effects? The answers must be sought on the molecular level, among the substances in garlic and onions. One of the earliest chemical studies was made in 1844 by the German chemist Theodor Wertheim. His subject was garlic. Wertheim attributed garlic's appeal "mainly to the presence of a sulfur-containing, liquid body, the so-called garlic oil. All that is known about the material is limited to some meager facts about the pure product, which is obtained by steam distillation of bulbs of *Allium sativum*. Since sulfur bonding has been little investigated so far, a study of this material promises to supply useful results for science."

Wertheim employed steam distillation. He put garlic in boiling water; the steam rising from the vessel included small amounts of garlic oil. Distillation of the oil yielded some strong-smelling volatile substances. Wertheim proposed the name allyl (from *Allium*) for the hydrocarbon group in the oil, and schwefelallyl ("allylsulfur" in English) for the volatiles. "Allyl" is still used today: it refers to groups of structure $CH_2{=}CHCH_2$, or in chemical shorthand C_3H_5. Numerous compounds that include an allyl group have a pungent odor.

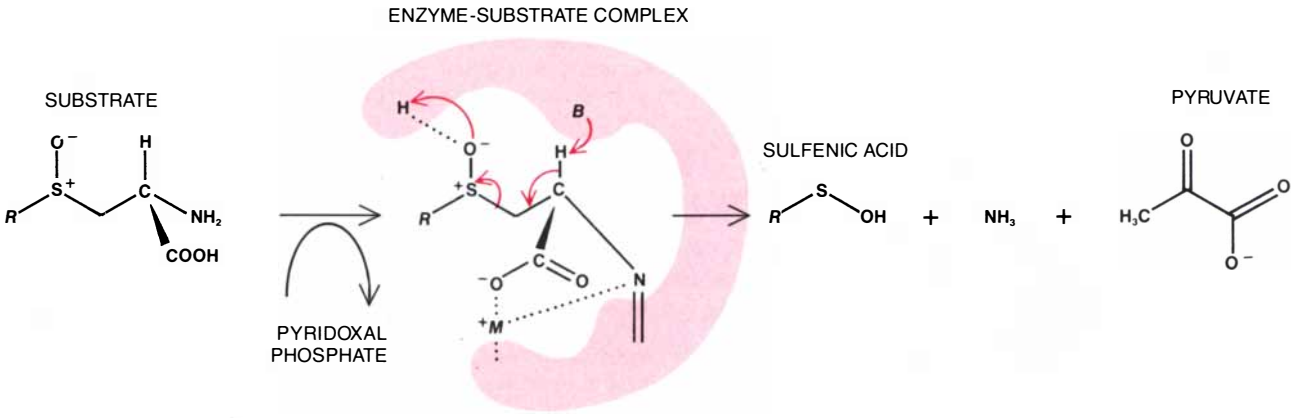
In 1892 another German investigator, the chemist F. W. Semmler, applied steam distillation to cloves of garlic, producing one or two grams of an evil-smelling oil per kilogram of garlic. In turn the oil yielded diallyl disulfide ($C_6H_{10}S_2$, or more precisely $CH_2{=}CHCH_2SSCH_2CH{=}CH_2$), accompanied by lesser amounts of diallyl trisulfide and diallyl tetrasulfide [*see upper illustration below*]. The steam distillation of 5,000 kilograms of onions yielded a rather different oil. It incorporates propionaldehyde



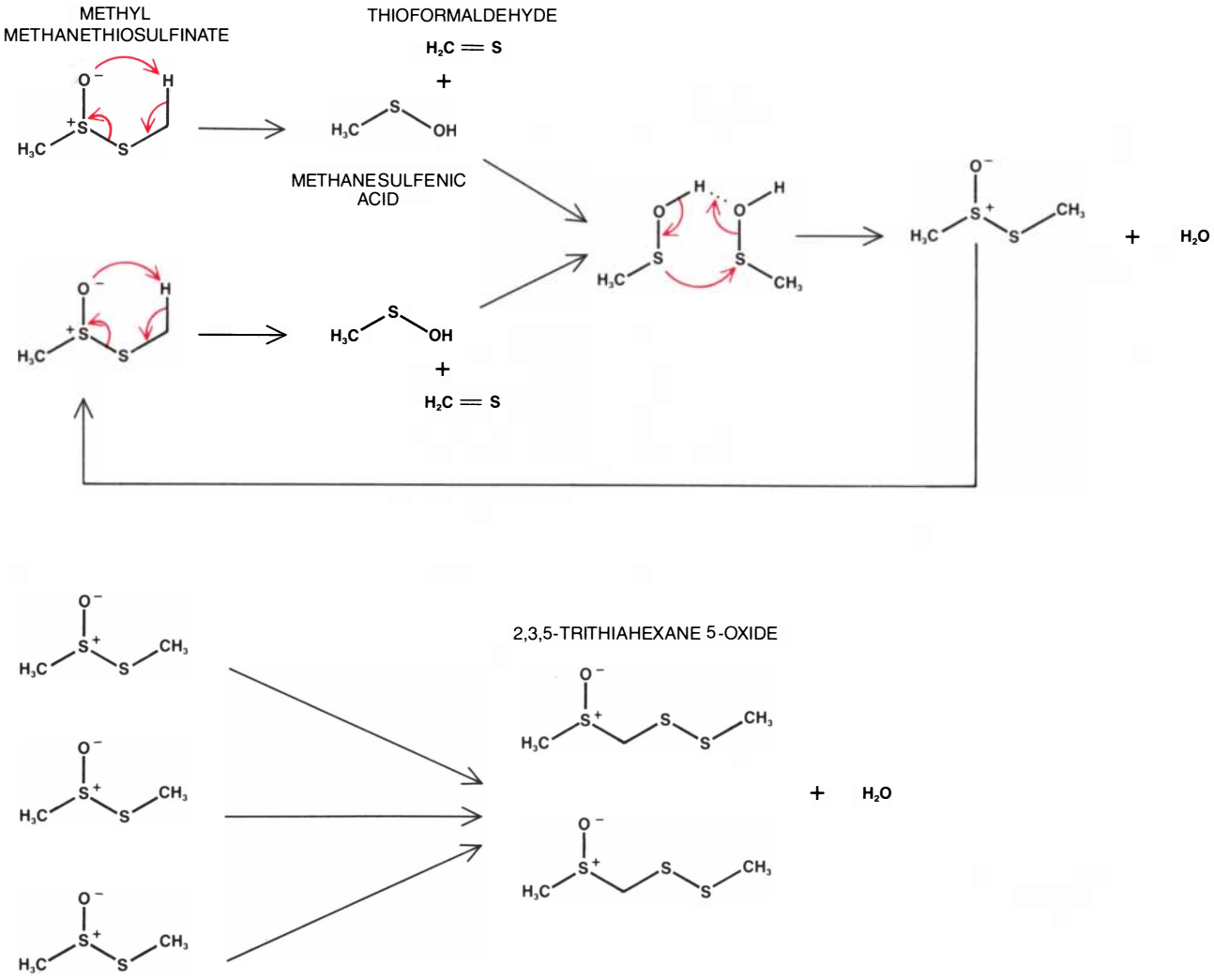
SULFUR COMPOUNDS EXTRACTED FROM GARLIC depend on the conditions of the extraction. The harshest technique is steam distillation, that is, the boiling of garlic followed by the extraction of compounds from condensed steam; it yields diallyl disulfide (*top*). A gentler technique employs as a solvent ethyl alcohol at room temperature; it yields the oxide of diallyl disulfide, called allicin (*middle*), which is the source of the odor of garlic. A still gentler technique employs pure ethyl alcohol at a subzero temperature; it yields alliin, a molecule with optical isomerism, or mirror-image chemical forms, at sulfur and carbon atoms (*asterisks*). (Four forms are possible, but only one is found in garlic.) An enzyme converts alliin into allicin. For each molecule only the carbon skeleton is shown.



SULFUR COMPOUNDS EXTRACTED FROM ONIONS also depend on the conditions of the extraction. Steam distillation yields propionaldehyde and dipropyl disulfide (*top*). The solvent Freon (mixed with water at zero degrees Celsius) yields the lacrimatory factor (*middle*): the substance that makes people cry when slicing an onion. The factor has two forms, designated *syn* and *anti;* the *syn* form predominates. Pure ethyl alcohol at a subzero temperature yields the lacrimatory precursor (*bottom*). It is a structural isomer of alliin; that is, the two differ only in structure. In onions an enzyme converts the lacrimatory precursor into the lacrimatory factor.



ALLINASE ENZYME catalyzes the conversion of several sulfur compounds in garlic and onions. Most notably, in garlic it acts on alliin; in onions it acts on the lacrimatory precursor. Here the catalysis is shown schematically for a generalized substrate. (For example, if *R* is an allyl group, C_3H_5, the substrate is alliin.) A cofactor, pyridoxal phosphate, acts on the substrate so that it forms a complex with the enzyme; the binding includes an electrostatic interaction of the substrate and a metal ion (M^+). A basic group, *B*, on the enzyme then removes a proton, or hydrogen ion, from the substrate, triggering the collapse of the substrate and the release of a sulfenic acid, RSOH, along with ammonia and pyruvate. A chemical reaction is in essence a transformation of the chemical bonds in molecules, which can be symbolized by the movement of pairs of electrons; arrows (*color*) show the likeliest movements of the pairs.



DECOMPOSITION PATHWAYS for methyl methanethiosulfinate, a homologue, or simplified version, of allicin, helped to elucidate the pathways for allicin itself. Along one path (*top*) methyl methanethiosulfinate decomposes into methanesulfenic acid and thioformaldehyde; then two molecules of methanesulfenic acid combine to regenerate a molecule of methyl methanethiosulfinate. Along another path (*bottom*) three molecules of methyl methanethiosulfinate self-condense to produce 2,3,5-trithiahexane 5-oxide.

(C_2H_5CHO) as well as numerous sulfur compounds, including dipropyl disulfide ($C_6H_{12}S_2$).

The next key discovery in the chemistry of garlic and onions was made in 1944 by Chester J. Cavallito and his colleagues at the Sterling-Winthrop Chemical Company in Rensselaer, N.Y. They established that methods less vigorous than steam distillation yield rather different substances. Cavallito applied ethyl alcohol to four kilograms of garlic at room temperature and eventually produced six grams of an oil whose formula was $C_6H_{10}S_2O$. The oil proved to be both antibacterial and antifungal. Against *B. typhosus* it was more potent than penicillin or sulfaguanidine; otherwise it was less potent than penicillin.

Chemically, Cavallito's oil is the oxide of diallyl disulfide, the principal substance Semmler had isolated by steam distillation half a century earlier. Its exact chemical formula is $CH_2=CHCH_2S(O)SCH_2CH=CH_2$, which makes its name forbidding. The name is allyl 2-propenethiosulfinate. It should be said that the nomenclature of chemistry is cumbersome but precise in its characterization of a molecule. Each part of a name conveys the structure of a section of the molecule's carbon skeleton or signals the interruption of the skeleton by atoms such as sulfur. In allyl 2-propenethiosulfinate the 2 specifies that a carbon-carbon double bond (=) begins on the second carbon atom from the point of attachment of sulfur. Brackets surrounding an atom or group of atoms signify that the atom or atomic group is off the main chain of the molecule.

In any case, Cavallito called his discovery allicin. It is a chemically unstable, colorless liquid that accounts for the odor of garlic—much more so than the diallyl sulfides do. Allicin is the subject of two U.S. patents in Cavallito's name, but its clinical use as an antibacterial agent was abandoned after a brief trial because of the substance's odor.

Although allicin is responsible for the smell of garlic, a garlic bulb exhibits little or no odor until it is cut or crushed. In 1948 Arthur Stoll and Ewald Seebeck of the Sandoz Company in Basel showed why. Allicin develops in garlic when an enzyme initiates its formation from an odorless precursor molecule, which Stoll and Seebeck identified as (+)-S–allyl–L-cysteine sulfoxide, or $CH_2=CHCH_2S(O)CH_2CH(NH_2)COOH$. (The + and the L signify a particular spatial arrangement for the sulfur atom and for the carbon atom attached to nitrogen.) Evidently the cutting or crushing of garlic enables the enzyme, called allinase, to come in contact with the precursor of allicin.

Stoll and Seebeck named the precursor alliin; it accounts for about .24 percent of the weight of a typical garlic bulb. It can be formed by attaching an allyl group and an oxygen atom to the sulfur atom in the amino acid cysteine. Alliin can also be taken from garlic, but the chemical conditions of the extraction must be mild. Subsequent crystallization affords extremely fine, colorless, odorless needles.

Alliin is a remarkable molecule. In particular it was the first natural substance found to display the property called optical isomerism due to mirror-image forms at sulfur as well as at carbon. Optical isomerism arises when a molecule has mirror-image forms and nature favors one form over the other. A solution of the substance then proves capable of rotating a beam of polarized light. In alliin, mirror-image configurations are possible at both sulfur and carbon. Under the influence of allinase alliin decomposes to 2-propenesulfenic acid [*see top illustration on opposite page*]. The enzyme acts preferentially on the isomer of alliin designated (+), the form that produces a clockwise rotation of a beam of polarized light. In turn 2-propenesulfenic acid dimerizes, or pairs with a second molecule of 2-propenesulfenic acid, to give allicin.

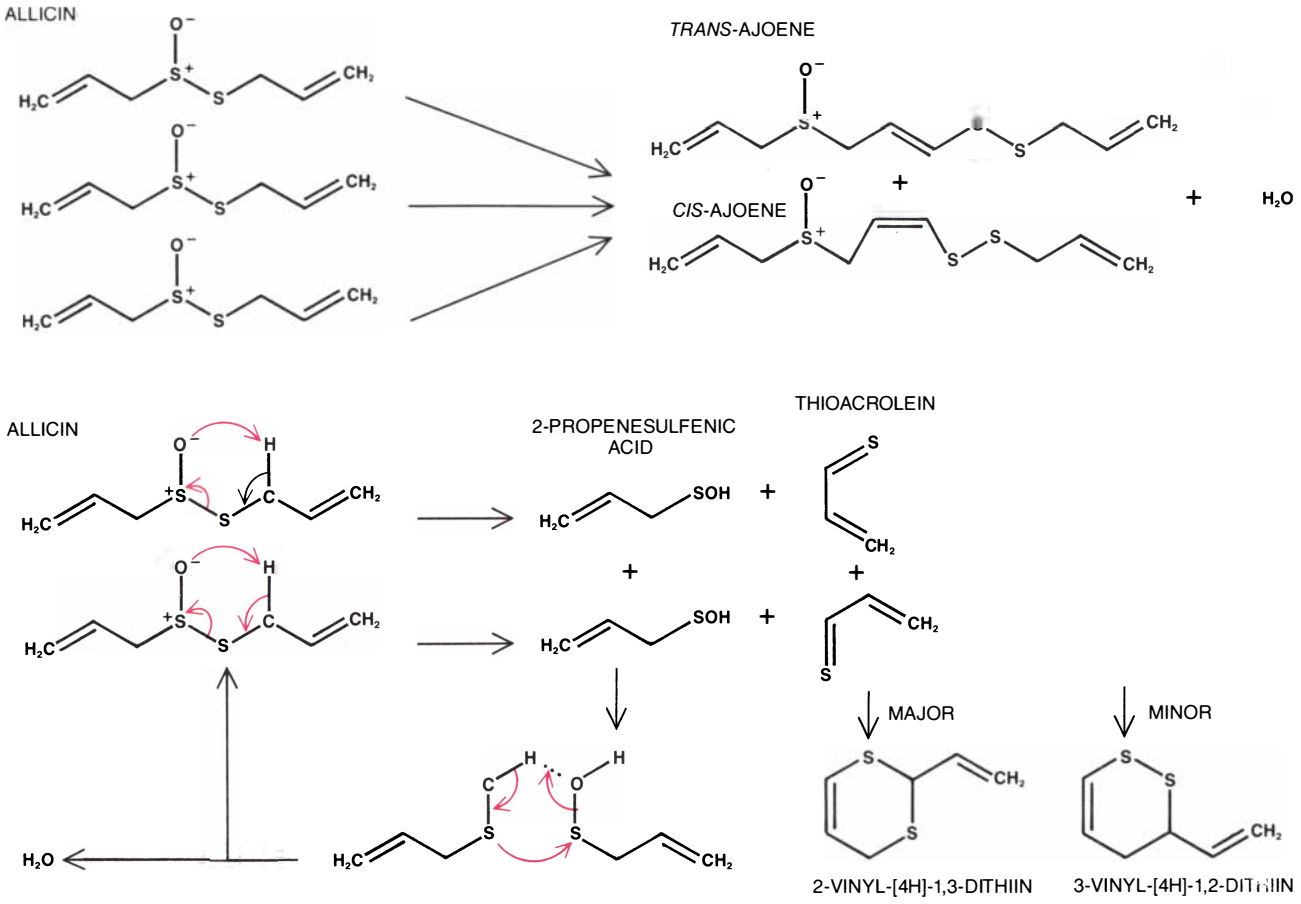
While the chemistry of garlic was being investigated, that of the onion was also emerging. In 1961 the Finnish biochemist Artturi Virtanen (who received the 1945 Nobel prize in chemistry for his work in animal husbandry) showed that onions contain *trans*-(+)-S–(1-propenyl)–L-cysteine sulfoxide, a positional isomer of alliin [*see bottom illustration on page 115*]. That is, its chemical content is identical with that of alliin; only its structure differs. (Specifically it differs in the position of a double bond, which, as the 1 in its name indicates, is attached directly to sulfur.) *Trans*-(+)-S–(1-propenyl)–L-cysteine sulfoxide is the lacrimatory precursor, or LP: the onion's allinase enzyme converts it into the lacrimatory factor, or LF, the substance that makes people cry when they slice an onion.

The chemical formula of the lacrimatory factor is C_3H_6SO, which fits more than 50 different chemical structures. Virtanen postulated that the actual structure is $CH_3CH=CHS(O)H$ and not the alternative possibility $CH_3CH=CHS–O–H$, in which the oxygen atom is on the main chain of the molecule's carbon skeleton. (These compounds are both named *trans*-1-propenesulfenic acid.) In the meantime, however, W. F. Wilkins, a graduate student at Cornell University, proposed a structure of $C_2H_5CH=SO$, which corresponds to the name propanethial S-oxide. A decade later, in 1971, M. H. Brodnitz and J. V. Pascale of the International Flavors and Fragrances Company in Union Beach, N.J., confirmed Wilkins' hypothesis.

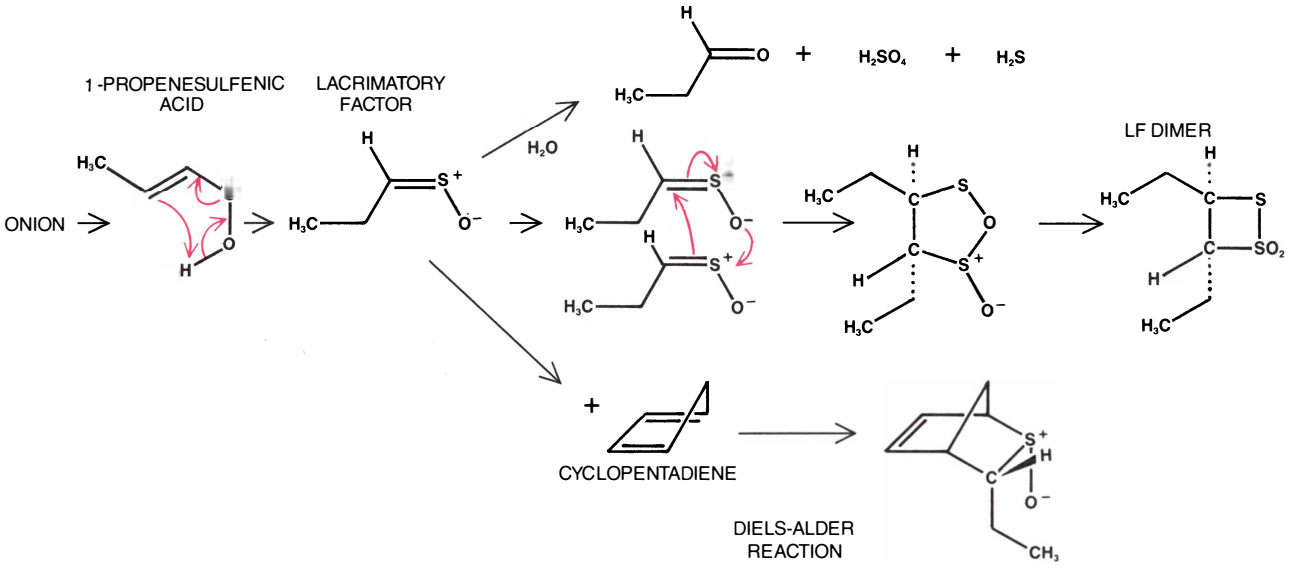
In garlic, then, an allinase enzyme converts alliin into allicin, the odoriferous constituent of garlic. In onions an allinase enzyme converts the lacrimatory precursor into the lacrimatory factor. These are not the only actions of the enzymes. In both garlic and onions the enzymes act on a number of molecules; in chemical parlance they have a number of substrates. All of them are sulfur-containing substances synthesized in garlic and onions by chemical sequences that begin with the sulfur-containing amino acid cysteine. From them the allinase enzymes produce several sulfenic acids, RSOH, where R signifies a radical: either allyl ($CH_2=CHCH_2$), 1-propenyl ($CH_3-CH=CH$), methyl (CH_3) or propyl (C_3H_7). The by-products of the reactions are pyruvate ($CH_3C(O)CO_2^-$) and ammonia (NH_3).

Recent work indicates that the reactions require the participation of an additional substance, or cofactor, called pyridoxal phosphate. Apparently the cofactor and the substrate interact, so that the substrate is converted into an activated form. A basic group in the enzyme (that is, a proton-capturing group) then initiates the release of sulfenic acid. For their part the sulfenic acids are highly unstable and undergo further reactions spontaneously.

My own work in the chemistry of garlic and onions began in 1971 as a further exploration of the properties of allicin. My colleagues and I at the University of Missouri at St. Louis initiated a study of the chemical transformation of methyl methanethiosulfinate, $CH_3S(O)SCH_3$. The compound is the simplest homologue of allicin: on the one hand, it includes a chemical group (S(O)H) that is central to the chemistry of allicin; on the other hand, its carbon skeleton is simpler than that of allicin. With John O'Connor I discovered two unusual processes [*see bottom illustration on opposite page*]. In one process the decomposition of methyl methanethiosulfinate gives methanesulfenic acid, CH_3SOH, and thioformaldehyde, $CH_2=S$. In turn two molecules of methanesulfenic acid, which is remarkably reactive, combine (with the loss of a molecule of water) so that a molecule of methyl methanethiosulfinate is re-



DECOMPOSITION OF ALLICIN proceeds on several pathways. In one of them (*top*) three molecules of allicin combine, producing two molecules of a substance called ajoene. The mechanism was suggested by studies of methyl methanethiosulfinate. For its part ajoene proves to be antithrombotic: it is at least as potent as aspirin in preventing the aggregation of blood platelets and thus in keeping blood from clotting. Ajoene has two forms, designated *trans* and *cis*. The latter is slightly more potent. Along another path (*bottom*) allicin self-decomposes, giving 2-propenesulfenic acid and thioacrolein. Both are highly reactive. The self-condensation of two molecules of 2-propenesulfenic acid regenerates a molecule of allicin; the self-condensation of two molecules of thioacrolein yields two types of cyclic compound by a chemical process termed a Diels-Alder reaction. Both cyclic compounds are mildly antithrombotic.



PATHWAYS FOR LACRIMATORY FACTOR are complex. In the immediate forerunner of the factor, 1-propenesulfenic acid, the SOH group is attached to a double bond. The proximity favors an internal transfer of hydrogen (*colored arrows*), and hence formation of the factor. The factor is highly reactive. It can undergo hydrolysis (*top*), giving propionaldehyde, sulfuric acid and hydrogen sulfide. It can combine with itself (*middle*), producing a curious four-atom ring. It can be locked into a bicyclic structure (*bottom*).

stored. In the second process methyl methanethiosulfinate self-condenses, yielding 2,3,5-trithiahexane 5-oxide, $CH_3S(O)CH_2SSCH_3$.

Twelve years later our work proved important in elucidating the structure and mode of formation of the garlic antithrombotic factor. Mahendra K. Jain and Roger W. Crecely of the University of Delaware, working with Rafael Apitz-Castro and Maria R. Cruz of the Venezuelan Institute of Scientific Investigations in Caracas, produced several garlic extracts that were particularly active in preventing the aggregation of blood platelets. The most active extract had the chemical formula $C_9H_{14}S_3O$. In close collaboration with our colleagues from Delaware and Venezuela, Saleem Ahmad and I, working at the State University of New York at Albany, established the structure of the compound as $CH_2{=}CH{-}CH_2S(O)CH_2CH = CHSSCH_2CH = CH_2$, or 4,5,9-trithiadodeca-1,6,11-triene 9-oxide. We named the compound ajoene, after *ajo* (pronounced aho), Spanish for garlic.

My earlier work on the self-condensation of methyl methanethiosulfinate suggested that ajoene might form by self-condensation from allicin. We verified the proposal by simply heating allicin with a mixture of water and an organic solvent such as acetone [*see top illustration on opposite page*]. Experiments have since indicated that as an antithrombotic agent ajoene is at least as potent as aspirin. Studies by the Delaware and Venezuela groups, in collaboration with James Catalfamo of the New York State Department of Health in Albany, suggest that ajoene acts by inhibiting fibrinogen receptors on platelets. More specifically, there may be an interaction of the hydrocarbon, sulfur-oxygen and sulfur-sulfur groups of ajoene with chemically complementary groups on the surface of platelets, which could otherwise bind to fibrinogen. Further experiments now under way should establish the utility, if any, of ajoene as a drug.

A second aspect of the chemistry of the allicin-homologue methyl methanethiosulfinate also proved to be instructive. I have noted above that the decomposition of methyl methanethiosulfinate yields thioformaldehyde ($CH_2{=}S$). The same type of process seems to take place for allicin. In particular the decomposition of allicin yields thioacrolein, $CH_2{=}CHCH{=}S$. It is a highly reactive compound, sapphire blue in color. Hans Bock of the University of Frankfort has shown that thioacrolein dimerizes to produce two cyclic compounds. We have found them in garlic, in the ratio Bock would predict. The dimerization proceeds by a Diels-Alder reaction, in which a four-atom unit of one molecule and a two-atom unit of another molecule combine to form a six-atom ring. Diels-Alder reactions are among the most important in organic chemistry.

Concerning onions a problem persisted: the lacrimatory factor had been identified as $C_2H_5CH{=}SO$, or propanethial S-oxide, but that molecule has two isomers. In an effort to learn which one an onion contains, Larry Revelle, Robert E. Penn and Ali Bazzi did studies in my laboratory. To extract the lacrimatory factor they minced frozen onions, applied a solvent (Freon), produced a residue by removing the solvent at -78 degrees Celsius and distilled the residue in vacuum at -20 degrees. Two independent molecular spectroscopic techniques (microwave spectroscopy and nuclear magnetic resonance spectroscopy) then established the structure of the factor. It is largely *syn*-propanethial S-oxide; the *anti* isomer is present only in trace amounts. In the *syn* form the ethyl group (C_2H_5) at one end of the molecule's carbon chain is near the oxygen atom at the other end of the chain.

Two experiments in my laboratory cleared up a further problem. In garlic the allinase enzymes produce sulfenic acids; in onions they produce propanethial S-oxides, which are distinct from sulfenic acids. Specifically, propanethial S-oxides belong to the class of compounds called sulfines. Quite different chemical processes would have to be occurring if the sulfines were forming directly, and that conclusion is unsettling because an enzyme usually displays one mode of action, not several.

In the first experiment Penn established by means of spectroscopy that the structure of methanesulfenic acid (the simplest sulfenic acid) is $CH_3S{-}O{-}H$ rather than $CH_3S(O)H$. Then in the second experiment Penn and I found that when *trans*-1-propenesulfenic acid (the lacrimatory factor, according to Virtanen) is prepared by chemical means, it rapidly rearranges to *syn*-propanethial S-oxide. If we assume (based on the first experiment) that *trans*-1-propenesulfenic acid has the structure $CH_3CH{=}CHS{-}O{-}H$ rather than $CH_3CH{=}CHS(O)H$, the means of rearrangement can be deduced: it takes the form of an internal transfer of hydrogen [*see bottom illustration on opposite page*]. We conclude that the initial stage in the formation of the lacrimatory factor in onions does produce a sulfenic acid (1-propenesulfenic acid), which then changes rapidly into the true lacrimatory factor: *syn*-propanethial S-oxide.

The lacrimatory factor in turn is highly reactive. In the laboratory it can undergo hydrolysis, giving (among other things) sulfuric acid. It can dimerize, giving an LF dimer whose curious structure incorporates a four-atom ring. As Alan Wall and I have found, it can undergo a Diels-Alder reaction with cyclopentadiene, a highly reactive cyclic molecule containing a four-atom, so-called diene unit. The reaction locks the *syn* structure of the lacrimatory factor into a rigid molecular framework of two linked rings.

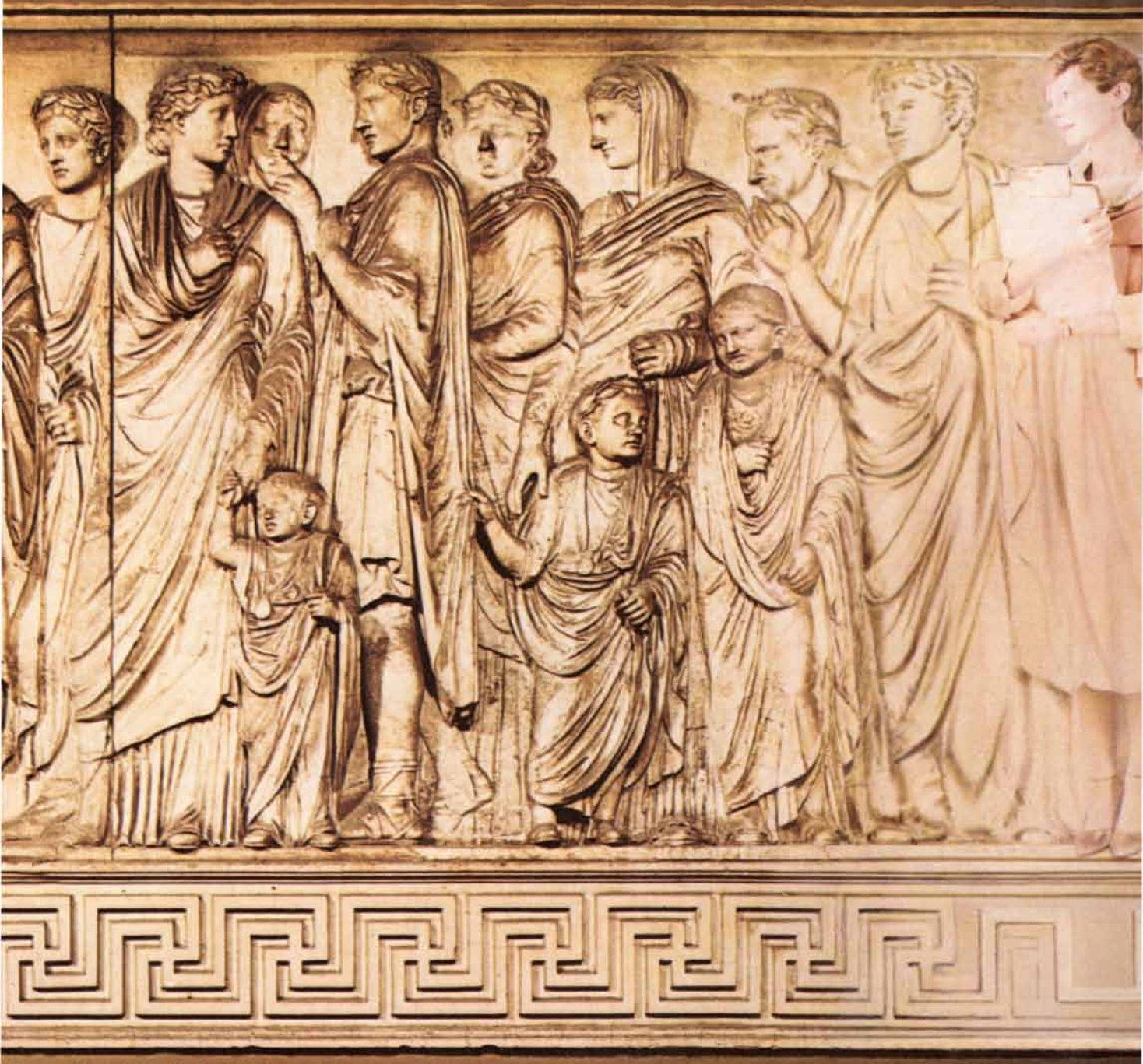
The chemical properties of the lacrimatory factor account for the efficacy of the methods employed in the kitchen to lessen the discomfort of dealing with onions. Chilling an onion reduces the volatility of the lacrimatory factor. Peeling an onion under running water washes the factor away; the factor is water-soluble.

Why did nature incorporate into garlic and onions the chemical apparatus to manufacture allicin and the lacrimatory factor? Since allicin is antifungal as well as antibiotic, it could offer the garlic plant protection against the bulb decay induced by fungi. And since the lacrimatory factor is irritating and repugnant to certain animals, it too has survival value.

There remains a question concerning the antithrombotic property of certain molecules in garlic. My colleagues and I have never detected ajoene or the cyclic antithrombotic compounds in dehydrated garlic powder, nor have we found them in pills, oils, extracts or other proprietary garlic preparations. The probable explanation is that the manufacture of most such products begins with the steam distillation of garlic. For now the beneficial effects attributed to garlic are best obtained from fresh garlic. Of course, self-medication with concoctions prepared from garlic and onions must not be substituted for proper medical diagnosis and treatment. Common sense will suggest the most important reason; the nose may suggest another. Garlic and onions provide a lingering reminder of their ingestion because sulfur compounds introduced into the bloodstream find their way into exhaled air and perspiration. Garlic's virtues and faults are perhaps best summarized by Sir John Harrington in "The Englishman's Doctor," written in 1609:

> Garlic then have power to save
> from death
> Bear with it though it maketh
> unsavory breath,
> And scorn not garlic like some
> that think
> It only maketh men wink
> and drink and stink.

The Ara Pacis commemorated an era of unity and prosperity unique in the history of the world.



Frieze on the south side of the Ara Pacis, Rome. 13 B.C.

MANUFACTURING PLANTS: ANN ARBOR, MI • ATLANTA, GA • CONCORD, NH • CREEDMOOR, NC • DALLAS, TX • MARLTON, NJ • MINNEAPOLIS, MN

Northern Telecom introduces a new era of technological unity and prosperity: the OPEN World.

The Age of Augustus. A time when peace, a universal language, and international trade brought such prosperity that the Romans built a monument to it: *Ara Pacis*, the altar of peace.

Today Northern Telecom is helping to create a new era of prosperity with a rational approach to information management we call the OPEN World.* It unites old and new technology from many suppliers—computers, telephones, and office equipment—into a single integrated information network.

Everyone has access to the information they need. Everything works together.

Northern Telecom is the world leader in digital telecommunications, the new universal language that makes it all possible. But the OPEN World is more than advanced technology. It's a corporate commitment to provide systems that evolve instead of age; that are easy to use; that are cost-effective.

The OPEN World opens up a whole new world of opportunities.

nt northern telecom

For further information write: Northern Telecom Inc., Public Relations Dept., 259 Cumberland Bend, Nashville, TN 37228. Or call (615) 256-5900, Ext. 4264.

**Trademark of Northern Telecom Limited*

MORRISVILLE, NC • MORTON GROVE, IL • NASHVILLE, TN • RALEIGH, NC • SAN DIEGO, CA • SANTA CLARA, CA • WEST PALM BEACH, FL

THE AMATEUR SCIENTIST

A field formula for calculating the speed and flight efficiency of a soaring bird

by Jearl Walker

Many birds are able to soar (fly without flapping their wings) by circling in rising air. A close observer of such a bird can learn much about the aerodynamics of the flight and much about how different species of birds accomplish the task of flying. Paul MacCready of AeroVironment Incorporated in Monrovia, Calif., has developed a technique by which an amateur can calculate a bird's coefficient of lift, which is a measure of how efficiently the bird flies. The bird's speed and radius of circling can also be calculated.

MacCready is known for his construction of human-powered aircraft. His earliest observations of soaring birds inspired the construction of the *Gossamer Condor,* which in 1977 won the Kremer Competition by being the first man-powered aircraft to fly a figure-eight course. (The aircraft is now on display at the National Air and Space Museum.) The feasibility of the flight arose from MacCready's calculations on bird soaring, specifically his calculation of the output of power that is required.

MacCready began his observations and calculations on bird soaring in 1976 by watching turkey vultures (*Cathartes aura*) as they glided smoothly in circles. With the help of his children he measured the time a bird took to complete a circle and also the bird's angle of bank (the angle between the horizontal and the plane of the bird's wings). Holding a protractor at arm's length, MacCready measured the bank angle as a bird flew toward him and as it flew away. The average of the two measurements and the time of circling were recorded.

MacCready also observed the soaring of black vultures (*Coragyps atratus*) and of the best soarers of all—frigate birds (*Fregata magnificens*). He again monitored frigate birds while he was on vacation in 1980 at La Paz in Mexico. In 1982 he videotaped the flights with a zoom-lens television camera. When the tape was reviewed, he measured the bank angle on the monitor's screen. A clock incorporated into the tape player recorded the time.

The results in 1980 and 1982 were surprisingly different. In 1980 the average bank angle was 23 degrees and the average time for a full circle was 9.1 seconds. In 1982 the average angle was 39 degrees and the average time was 9.4 seconds.

MacCready noted that the flights were at a much higher altitude in the second set of observations. The higher flights made measurements of the an-

A turkey vulture soaring in a thermal

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

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

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

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

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

FUJI.

Nobody gives you better performance.

 to Film U.S.A., Inc., Magnetic Products Div., 350 Fifth Avenue, NY, NY 10118

gles less accurate because of perspective. To correct for this problem he reduced the average bank angle to 36 degrees for the 1982 data. Still the puzzle remained. Why did the birds soar at a steeper angle in the higher flights? I shall return to this question after I introduce MacCready's calculations on flight efficiency.

A gliding bird descends slowly in relation to the air through which it flies. If the air is rising faster than the bird descends, the bird gains altitude in relation to the ground. A circling bird usually maneuvers to stay within the rising air. The lift on the bird's wing is a force from the air that is perpendicular to a line connecting the wing tips. It is also perpendicular to the flight path. If the bird is flying level, the lift is entirely vertical and counters the bird's weight. When the bird banks, the lift is tilted from the vertical by the angle of bank and acquires a horizontal component. Only the vertical component of the lift is available to counter the bird's weight. If the component is too small, the bird sinks in relation to the rising air. If the component is too large, the bird rises.

The horizontal component of the lift provides a centripetal force maintaining the circular motion. The size of the component depends on the strength of the lift and the bank angle.

MacCready derived a mathematical formulation whereby one can compute the speed of the bird once the bank angle and the time of circling are measured. His formula relies on three relations: (1) A bird soaring in a circle has matched the vertical component of lift to its weight. (2) The horizontal component of the lift generates circular motion and thus is equal to the bird's mass multiplied by the centripetal acceleration, which can be calculated as the square of the speed divided by the radius of the circle. (3) The speed of an object in circular motion is equal to the circumference of the circle divided by the time required to complete a circle.

Armed with these three relations, MacCready found that the speed of the bird is proportional to the circling time multiplied by the tangent of the angle of banking. Once the speed is known the radius can be computed from the relation between the speed and the circumference.

The full formulation is displayed in the top illustration on the opposite page. The relations are also plotted as a graph in the bottom illustration. The graph can be used in the field. First measure the bank angle of a bird soaring in a circle and the time it takes to complete a circle. Mark the time on the abscissa of the graph. Now move upward on the graph until you reach a straight line marked with the bank angle you measured. (You may have to interpolate between two such lines if the angle is not shown.) From that intersection move directly to the ordinate, from which you read the speed of the bird. For example, if a bird completes a circle in nine seconds at an angle of 20 degrees, its speed is about 5.1 meters per second.

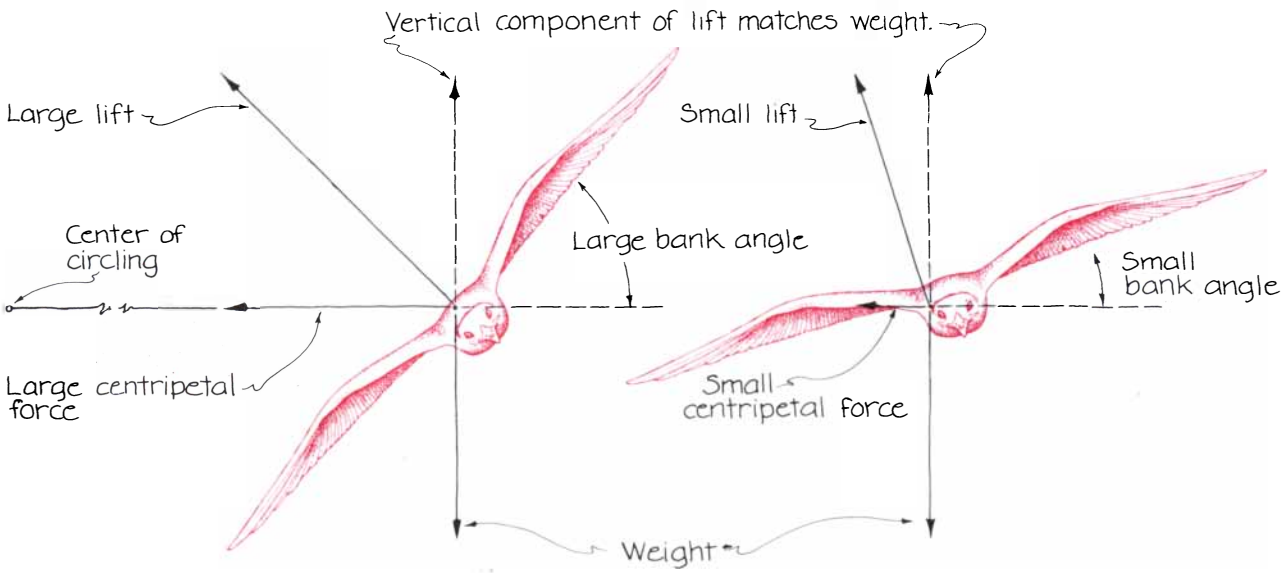
You can also read the radius of the circle from the graph. When you have moved directly upward from the point representing the circling time to the intersection with a line of bank angle, determine the radius from the nearest curved line representing radii. In my example the radius is about 7.3 meters.

MacCready calculated on the basis of his field observations in 1980 that the birds glided at an average speed of 5.9 meters per second in a circle with an average radius of 8.5 meters. The birds he observed in 1982 glided at 10.4 meters per second in a circle with a radius of 15.3 meters.

Another parameter is the bird's coefficient of lift, which is defined in terms of the loading of lift on the wings. The loading is the lift divided by the total surface area of the top surfaces of the wings. The coefficient of lift is the ratio of the lift loading to the excess pressure generated at the front of the bird by the forward movement. This excess pressure is equal to half of the density of the air multiplied by the square of the bird's speed.

The actual value for the coefficient of lift depends on the bird's angle of attack, the angle at which its wings meet the passing air. The largest useful value for the coefficient is obtained for the angle that results in stall, the instance in which the airflow over the top of the wing breaks away from the wing instead of moving smoothly to the rear edge. MacCready is interested in how large the coefficients can be for the soaring birds. Are they larger than the coefficients that aircraft built by human beings can achieve?

As it stands, the definition of the coefficient of lift is not useful for an amateur watching birds soar, because the lift is always unknown. Therefore MacCready has reexpressed the definition in terms of quantities an amateur



The components of lift

can obtain. He replaces lift with the bird's weight. (The ratio of weight to the area of the top surfaces of the wings is termed the wing loading.) The bird's speed is replaced with the speed it would have if it were gliding in a straight line with the same coefficient of lift. This straight-line speed is the actual speed multiplied by the square root of the cosine of the bank angle. The ratio of the straight-line speed to the actual speed can be represented as a graph [*see illustration on next page*]. The ratio is about 1 when the bank angle is small, but it can be appreciably smaller for larger angles.

To complete the calculation for the coefficient of lift one needs the ratio of weight to wing area (the wing loading) for the type of bird observed. (You can assume that the ratio is the same for all birds of the same type.) For the frigate birds observed by MacCready the ratio is about 35.1 newtons per square meter. (The newton is the metric unit of force. At one newton a body with a mass of one kilogram would be accelerated at a rate of one meter per second per second.) The density of air, which is also required in the calculation, depends on the air pressure and temperature. Assuming that the birds he observed as they soared were approximately at sea level and in air at a temperature of 15 degrees Celsius (59 degrees Fahrenheit), MacCready took the air density to be 1.225 kilograms per cubic meter.

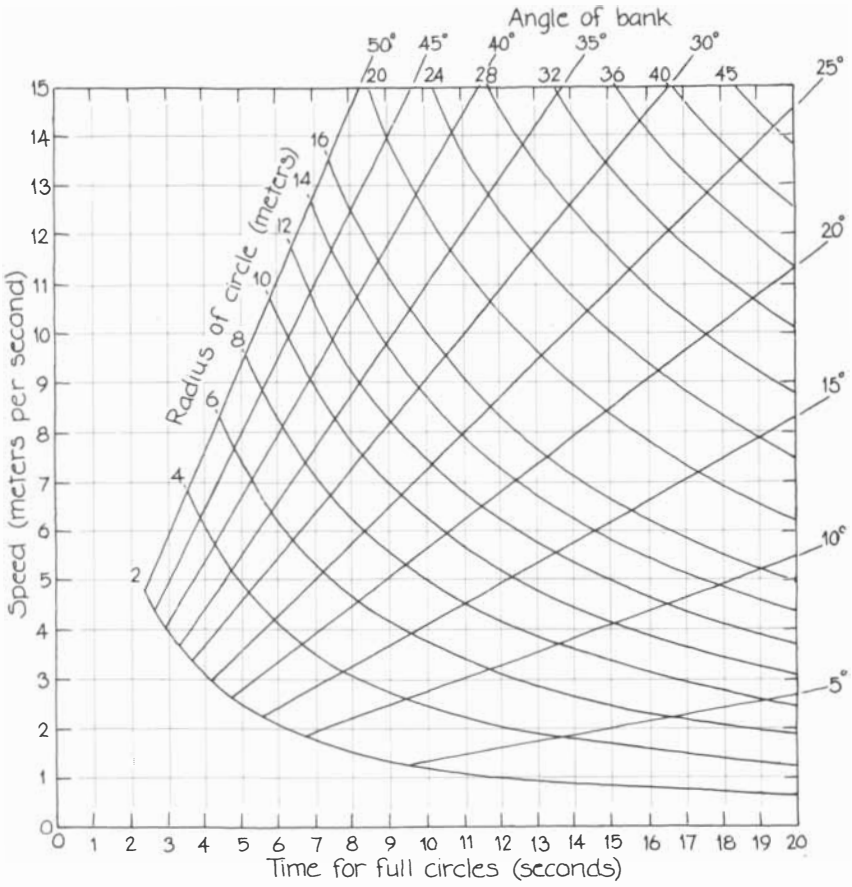
With these values MacCready calculated a coefficient of lift of 1.79 from the 1980 data and .65 from the 1982 data. The two results are different because the straight-line speeds (which depend on actual speed and bank angle) are different.

The puzzle about the different bank angles in the two sets of observations in Mexico reappears in the calculations of the coefficients. When MacCready made his first observations of birds soaring in 1976, he expected to find that all birds would soar with the same coefficient of lift. His computations of that year for the turkey vulture, the black vulture and the frigate bird met this expectation. The surprise in the observations of 1980 and 1982 was that the same type of bird does not always soar with the same coefficient of lift. Because the winds in 1980 were smooth and gentle and the winds in 1982 were turbulent and strong, MacCready concluded that the birds made adjustments to change the coefficients of lift depending on the meteorological conditions.

In the same way, the pilot of a sailplane flies faster when he is circling in a turbulent thermal (a rising region of

t — Time for complete circle	Matching of vertical forces: $L \cos\theta = mg$
L — Lift	Horizontal component of lift provides centripetal force: $L \sin\theta = \frac{mv^2}{r}$
g — Acceleration of gravity	Relation of speed and radius: $v = \frac{2\pi r}{t}$
w and mg — Weight of bird	Resulting expressions for speed and radius: $v = \frac{tg \tan\theta}{2\pi}$, $r = \frac{vt}{2\pi}$
v — Speed	Associated straight-line speed: $v_a = v\sqrt{\cos\theta}$
v_a — Associated straight-line speed	Coefficient of lift: $C_L = \frac{L/S}{dv^2/2} = \frac{W/S}{dv_a^2/2}$
θ — Angle of bank	
r — Radius of circle	
S — Surface area of wing top	
L/S — Loading of lift on wing	
W/S — Loading of weight on wing	
d — Air density	

Paul MacCready's equations



A field guide for determining the speed and circling radius of a soaring bird

warm air) than he does when he is in a gentle one. The higher speed makes the craft less vulnerable to stall if turbulence momentarily increases the angle of attack of the wing. It also gives the pilot greater control. In addition to flying faster the pilot banks the sailplane more sharply in a strong thermal.

Soaring birds make such adjustments more effectively because they are more experienced and have better sensors and more ways of adjusting their wings. The videotapes MacCready made in 1982 reveal that in turbulence the birds were constantly altering their bank angle and the shape of their wings, particularly their twist.

The wing loading on some hang gliders is in the same range as that on some birds, such as the black vulture. Therefore the pilot of such a glider could utilize a thermal almost as small as the ones birds fly in. The bird, however, still has the edge in controllability. Colin J. Pennycuick of the University of Bristol, an expert in avian biology, once remarked: "I suspect that even if you could simulate a bird's control system, you would need a cockpit so full of levers that it would take an octopus to fly it." Moreover, since birds have shorter wings that take up less of the radius of turning, they avoid some of the wing twist needed to keep the lift the same over both wings. That adjustment is harder with the larger wings of a hang glider.

A sailplane has a greater wing loading than hang gliders or birds and flies much faster. It is also quite efficient and sinks through the air more slowly. Nevertheless, it cannot utilize a thermal as small as the ones birds exploit.

Because the turkey vulture has a substantially smaller wing loading than the black vulture, it can soar in smaller and weaker thermals. It can also soar earlier in the day (when the thermals are weaker). The frigate bird, with its long, slender wings, is the best ocean soarer of all. It can even make use of some of the gentle convective cells that develop in light wind when the water is warmer than the air.

The lift coefficient for frigates in calm conditions is 1.8. That is surprisingly large for the aerodynamic conditions of bird soaring. The high value indicates the frigate can glide with a low speed that normally would make other birds or a hang glider stall. MacCready suggests further observations might result in a lower coefficient for the frigate. If they do not, one must conclude that the frigates have a better engineering design than model airplanes of similar size operating at similar speeds and that aerodynamicists have something to aim at.

Last month I described how a peculiar afterimage can be generated in a dark room when one observes a scene briefly illuminated with a bright flash of light. You will see few details of the scene because of the dazzle of the light, but if you keep your gaze steady when the room is again dark, a detailed afterimage soon appears; it resembles a snapshot of the scene. The afterimage can be so vivid that details such as printed words can be recognized.

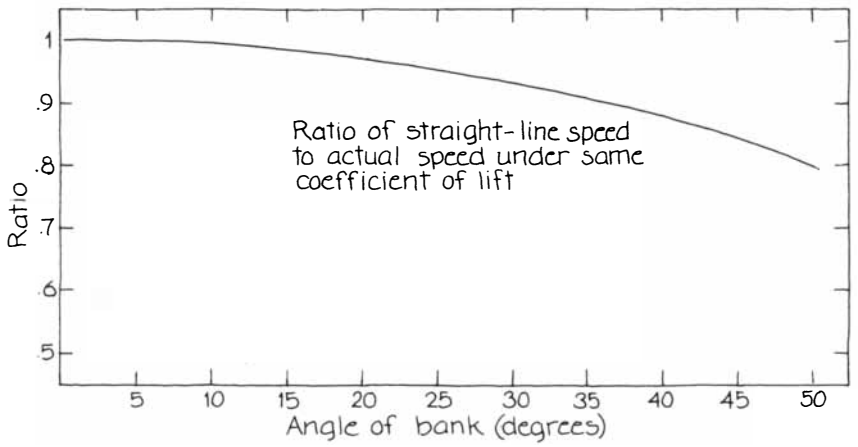
Part of the article was based on research by Edward H. Adelson of RCA's David Sarnoff Research Center. He offers another demonstration of this afterimage. An observer adapted to the dark views an apparently white card illuminated with a flash of light. In the afterimage he perceives the word RODS written on the card although there was no evidence of the word during the flash.

To set up this demonstration Adelson first writes RODS in thick block letters on the white card with a yellow highlighter pen. He then cuts a slit about a quarter of an inch wide in a red filter, covers the slit with a blue filter and attaches the composite to the flashgun. At the flash the gun sends out red light except from the slit, through which it sends blue light. When you set off the gun after you have adapted your eyes to darkness, you will see only a blank card bathed in red light. After a few tenths of a second the word RODS will appear.

Adelson explains the perception of the word in terms of the light absorbed by the yellow ink. In the red light the card is blank because the ink absorbs none of the light. In the blue light the card shows the word in dark gray because the ink absorbs blue. The cones of the retina are dominated by the bright red light, which masks the gray of the word formed in the dimmer blue light. The rods, insensitive to the red, record the presence of the word, but they are initially saturated by the blue light, and so the word is not perceived. As the rods become desaturated they eventually allow the perception of the word in the afterimage.

You may have to adjust the amount of blue from the flash. If there is too much blue, you will see the word during the flash of light and therefore will not be surprised to see it in the afterimage. If there is too little blue, you will not see the word in the afterimage.

You can also arrange the card so that the letters of the word are perceived one at a time. Hold the flashgun about six inches from the *S* and angle it so that the entire word is illuminated. The *R* is illuminated least by the flash. Rods receiving an image of the *R* are least saturated. Hence you perceive the *R* sooner than you do the other letters. The *O* is perceived next as the corresponding rods become desaturated. The *S* is the last letter to be perceived because the rods receiving its image were illuminated most and so require the greatest time to become desaturated.



A means of determining a bird's straight-line speed

Howard Brody of the University of Pennsylvania and Joseph P. Straley of the University of Kentucky have pointed out an error in my discussion of racquetball last September. I had written that in an elastic bounce of a ball from a floor the vertical velocity is reversed but unchanged in amount. Thus each subsequent bounce should be equally high.

As obvious as this conclusion was to me, I forgot it when I watched a highly elastic ball bounce across my kitchen floor. When I put spin on the ball, it seemed to advance in a pattern of alternating high and low bounces. What I saw was an illusion: the height of each bounce was in fact virtually the

same. Because of the spin, however, the ball came off the floor at different angles. If the horizontal velocity was low after a bounce, the ball rose at a steep angle. When the horizontal velocity was high, the angle was shallow. I misinterpreted the difference in angles as being a difference in height.

Brody also pointed out that in tennis the role of spin is more complex than it is in racquetball. "The high bounce of a topspin shot and the low bounce of a backspin shot result not from the interaction of the ball with the court surface but from the fact that the ball's vertical velocity is increased by topspin and decreased by backspin," he wrote. "This change is due to the Magnus effect." (The term refers to the unequal air pressure on opposite sides of the ball resulting from the interaction of the spinning surface with the passing air. Depending on the spin, the pressure difference can lift the ball or push it downward.) The effect is smaller on the racquetball because the ball is smoother.

Joseph W. Kennedy of Miami has written to me about several more unsolved puzzles that are associated with racquetball strokes. He delivers one of his strokes when he is near midcourt on the right-hand side. The ball hits the front wall, the rear of the left wall and the middle left of the back wall and then drops to the floor in front of the middle right of the back wall. An experienced opponent can anticipate this final position and return the ball.

Kennedy hits several of these shots and then makes a slight change, leading the stroke with his wrist. (He is right-handed.) Although the ball appears to hit the walls in the same places as before, it reaches the floor much closer to the back wall. The opponent is likely to be too far away from the wall to make a good return. Kennedy thinks he may be putting a spin on the ball by his wrist action.

Another peculiar (and rare) shot hits a front corner squarely at high speed. The ball drops to the floor and is unreturnable. Surely this results from the friction on the ball as it hits two walls simultaneously, but I do not know whether anyone has worked out the details of the collision.

A "kill shot" is directed horizontally and at high speed toward the bottom few inches of the front wall. The rebound is close to the floor and difficult to return. Yet sometimes, when the ball strikes the wall within two or three inches of the floor, it rises to a height of four or five inches on the rebound and is easier to return. Why does it rise? Perhaps it gains lift in the manner described by Brody.

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

COMPUTER RECREATIONS

SOFTWAR: LA GUERRE DOUCE. Thierry Breton and Denis Beneich. Éditions Robert Laffont, 1983.

CORE WAR GUIDELINES. D. G. Jones and A. K. Dewdney. Available from SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017; March, 1984.

WAR GAMES. Philip Elmer-DeWitt in *Time,* Vol. 124, No. 16, page 109; October 15, 1984.

THE VERIFICATION OF COMPLIANCE WITH ARMS-CONTROL AGREEMENTS

VERIFICATION AND SALT: THE CHALLENGE OF STRATEGIC DECEPTION. Edited by William C. Potter. Westview Press, 1980.

THE KEYHOLE SATELLITE PROGRAM. Jeffrey Richelson in *The Journal of Strategic Studies,* Vol. 7, No. 2, pages 121–153; June, 1984.

ARMS CONTROL VERIFICATION. Edited by Kosta Tsipis and Penny Janeway. Pergamon-Brassey's International Defense Publishers, in press.

THE VOLCANOES AND CLOUDS OF VENUS

THE SURFACE OF VENUS. Gordon H. Pettengill, Donald B. Campbell and Harold Masursky in *Scientific American,* Vol. 243, No. 2, pages 54–65; August, 1980.

VENUS. Edited by D. Hunten, L. Colin, T. Donahue and V. Moroz. University of Arizona Press, 1983.

PLANETS AND THEIR ATMOSPHERES: ORIGIN AND EVOLUTION. J. Lewis and R. Prinn. Academic Press, 1984.

CHROMOSOME TRANSLOCATIONS AND HUMAN CANCER

TRANSLOCATION OF IMMUNOGLOBULIN V_H GENES IN BURKITT LYMPHOMA. Jan Erikson, Janet Finan, Peter C. Nowell and Carlo M. Croce in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 79, No. 18, pages 5611–5615; September, 1982.

HUMAN C-*MYC ONC* GENE IS LOCATED ON THE REGION OF CHROMOSOME 8 THAT IS TRANSLOCATED IN BURKITT LYMPHOMA CELLS. Riccardo Dalla-Favera, Marco Bregni, Jan Erikson, David Patterson, Robert C. Gallo and Carlo M. Croce in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 79, No. 24, pages 7824–7827; December, 1982.

TRANSLOCATION OF THE C-*MYC* GENE INTO THE IMMUNOGLOBULIN HEAVY CHAIN LOCUS IN HUMAN BURKITT LYMPHOMA AND MURINE PLASMACYTOMA CELLS. R. Taub, I. Kirsch, C. Morton, G. Lenoir, D. Swan, S. Tronick, S. Aaronson and P. Leder in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 79, No. 24, pages 7837–7841; December, 1982.

SPECIFIC CHROMOSOMAL TRANSLOCATIONS AND THE GENESIS OF B-CELL-DERIVED TUMORS IN MICE AND MEN. George Klein in *Cell,* Vol. 32, No. 2, pages 311–315; February, 1983.

THE HIDDEN DIMENSIONS OF SPACETIME

SUPERGRAVITY AND THE UNIFICATION OF THE LAWS OF PHYSICS. Daniel Z. Freedman and Peter van Nieuwenhuizen in *Scientific American,* Vol. 238, No. 2, pages 126–143; February, 1978.

SUPERGRAVITY THEORY IN 11 DIMENSIONS. E. Cremmer, B. Julia and J. Scherk in *Physics Letters,* Vol. 96B, No. 4, pages 409–412; June 19, 1978.

DYNAMICS OF DIMENSIONAL REDUCTION. Peter G. O. Freund and Mark A. Rubin in *Physics Letters,* Vol. 97B, No. 2, pages 233–235; December 1, 1980.

SEARCH FOR A REALISTIC KALUZA-KLEIN THEORY. Edward Witten in *Nuclear Physics B,* Vol. 186, No. 3, pages 412–428; August 10, 1981.

WHY WHALES LEAP

ACOUSTIC COMMUNICATION AND BEHAVIOR OF THE SOUTHERN RIGHT WHALE (*EUBALAENA AUSTRALIS*). Christopher W. Clark in *Communication and Behavior of Whales,* edited by Roger Payne. Westview Press, Inc., 1983.

ENERGETICS OF LEAPING IN DOLPHINS AND OTHER AQUATIC ANIMALS. R. W. Blake in *Journal of the Marine Biological Association of the United Kingdom,* Vol. 63, No. 1, pages 61–70; 1983.

STRUCTURE AND STABILITY OF HUMPBACK WHALE GROUPS OFF NEWFOUNDLAND. Hal Whitehead in *Canadian Journal of Zoology/Journal Canadien de Zoologie,* Vol. 61, No. 6, pages 1391–1397; June, 1983.

A NEOLITHIC FORTRESS AND FUNERAL CENTER

CAUSEWAYED ENCLOSURES. I. F. Smith in *Economy and Settlement in Neolithic and Early Bronze Age Britain and Europe,* edited by D. D. A. Simpson. Leicester University Press, 1971.

INTERRUPTED DITCH ENCLOSURES IN BRITAIN: THE USE OF AERIAL PHOTOGRAPHY FOR COMPARATIVE STUDIES. R. Palmer in *Proceedings of the Prehistoric Society,* Vol. 42, pages 161–186; 1976.

CAUSEWAYED CAMPS AND EARLY NEOLITHIC ECONOMIES IN CENTRAL SOUTHERN ENGLAND. Graeme Barker and Derrick Webley in *Proceedings of the Prehistoric Society,* Vol. 44, pages 161–186; 1978.

ION IMPLANTATION OF SURFACES

ION IMPLANTATION IN SEMICONDUCTORS. James W. Mayer, Lennart Eriksson and John A. Davies. Academic Press, 1970.

ION IMPLANTATION. Frederick F. Morehead, Jr., and Billy L. Crowder in *Scientific American,* Vol. 228, No. 4, pages 64–71; April, 1973.

TAILORED SURFACE MODIFICATION BY ION IMPLANTATION AND LASER TREATMENT. S. T. Picraux and L. E. Pope in *Science,* Vol. 226, No. 4675, pages 615–622; November 9, 1984.

THE CHEMISTRY OF GARLIC AND ONIONS

THE BOOK OF GARLIC. Lloyd J. Harris. Holt, Rinehart and Winston, 1975.

THE LACRYMATORY FACTOR OF THE ONION: AN NMR STUDY. Eric Block, Larry K. Revelle and Ali A. Bazzi in *Tetrahedron Letters,* Vol. 21, No. 14, pages 1277–1280; 1980.

(E,Z)-AJOENE: A POTENT ANTITHROMBOTIC AGENT FROM GARLIC. Eric Block, Saleem Ahmad, Mahendra K. Jain, Roger Crecely, Rafael Apitz-Castro and Maria R. Cruz in *Journal of the American Chemical Society,* in press.

THE AMATEUR SCIENTIST

SOARING BIRD AERODYNAMICS—CLUES FOR HANG GLIDING. Paul B. MacCready, Jr., in *Ground Skimmer,* No. 45, pages 17–19; October, 1976.

SOARING AERODYNAMICS OF FRIGATE BIRDS. Paul B. MacCready, Jr., in *Soaring,* Vol. 48, No. 7, pages 20–22; July, 1984.

Our competition leaves you but one choice: Jeep Wagoneer.

Compare: Jeep Wagoneer Limited has luxurious leather bucket seats that'll fit you like a glove. The Buick Estate Wagon doesn't. Wagoneer has a larger standard engine and higher mileage* than Volvo Turbo Wagon. More cargo space than the Dodge Caravan.** Four doors and room for five. Plus something no other luxury wagon offers: the convenience and security of "shift-on-the-fly" from two-wheel to four-wheel drive with the simple flick of a switch.

Jeep Wagoneer. It has practically everything—except serious competition.

Only in a Jeep Wagoneer

*Wagoneer Limited 18 EPA EST MPG, 21 HWY EST. Use these figures for comparison. Your results may differ. California figures lower. **With optional seven-passenger seating in Dodge Caravan and five-passenger seating in Wagoneer. AVAI[illegible]G JEEP DEALERS. SAFETY BELTS SAVE LIVES.