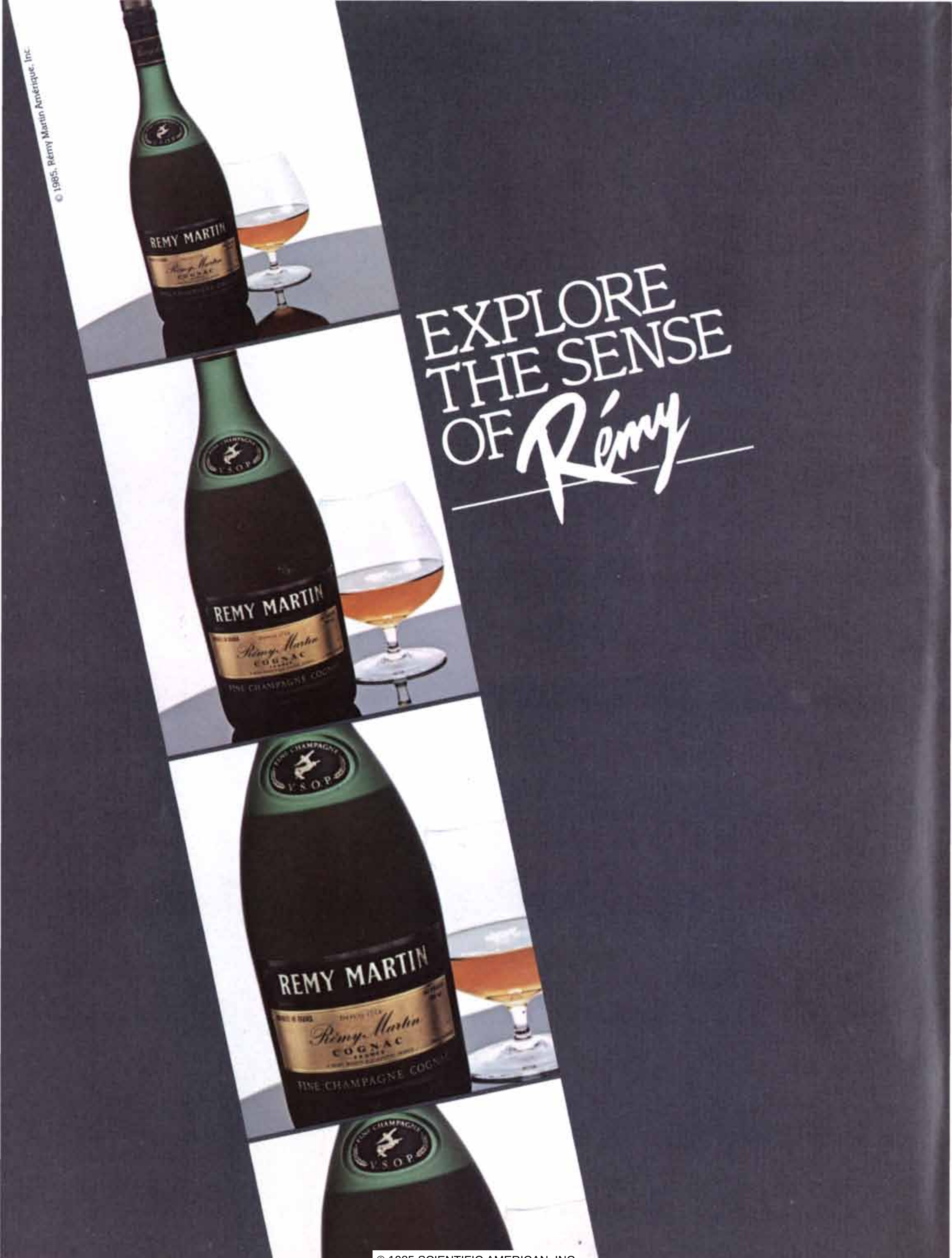
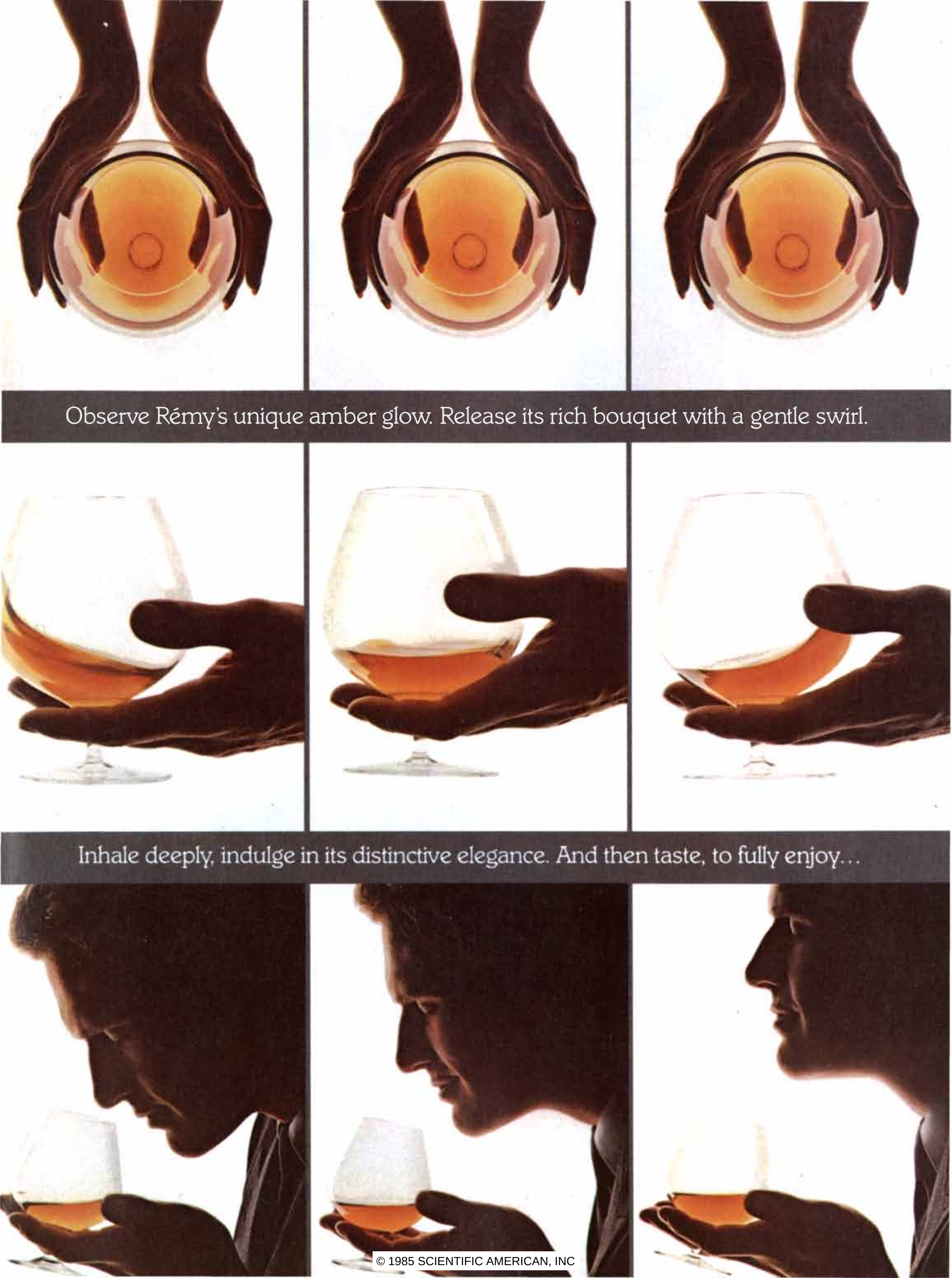
SCIENTIFIC AMERICAN

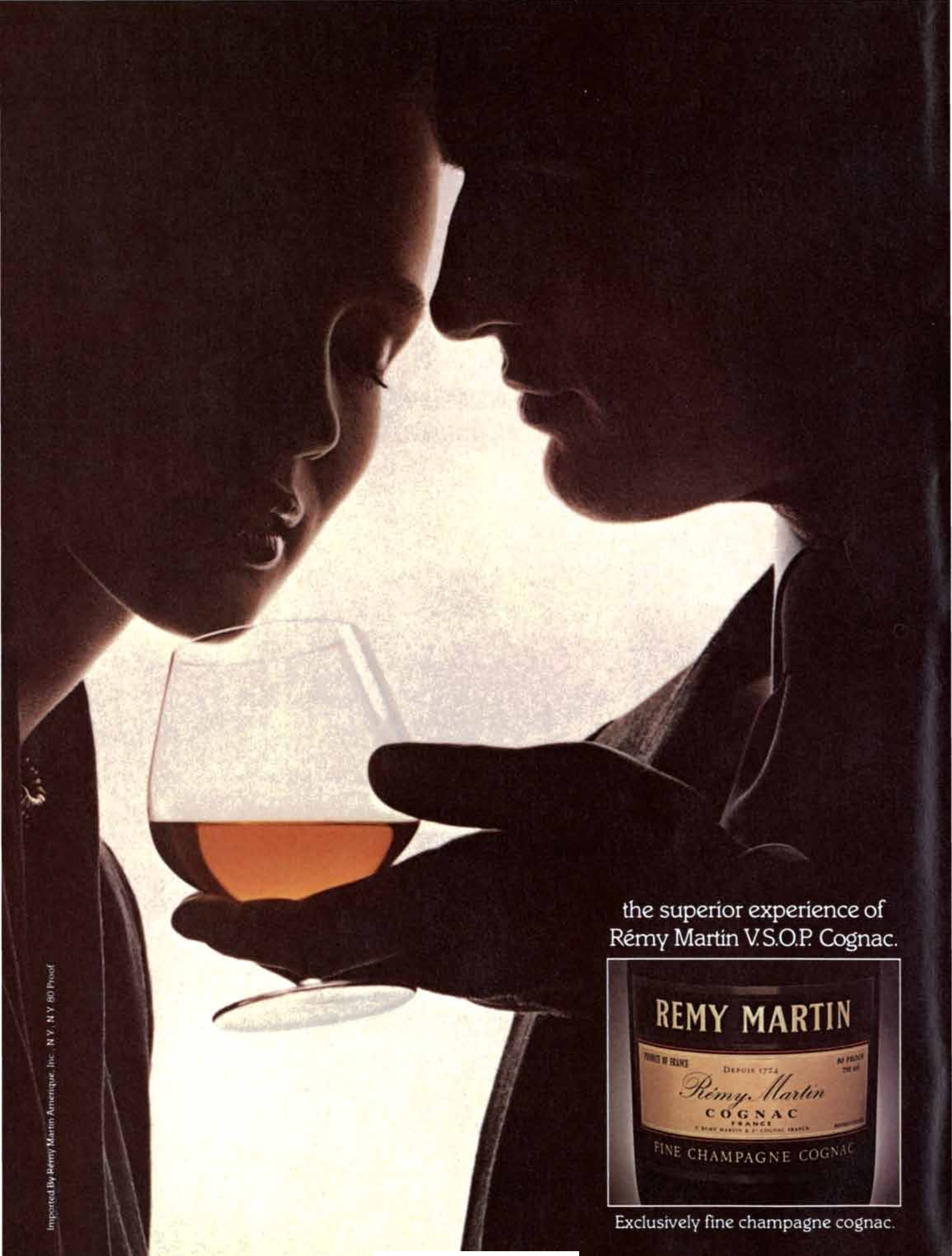
HUMAN-POWERED AIRCRAFT

$2.50

November 1985


© 1985, Rémy Martin Amérique, Inc.
REMY MARTIN
Rémy Martin
COGNAC
FINE CHAMPAGNE COGNAC
V.S.O.P.
EXPLORE
THE SENSE
OF Rémy


Observe Rémy's unique amber glow. Release its rich bouquet with a gentle swirl.
Inhale deeply, indulge in its distinctive elegance. And then taste, to fully enjoy…


the superior experience of
Rémy Martin V.S.O.P. Cognac.
REMY MARTIN
DEPUIS 1724
Rémy Martin
COGNAC
FRANCE
FINE CHAMPAGNE COGNAC
Exclusively fine champagne cognac.
Imported By Rémy Martin Amerique, Inc. N.Y., N.Y. 80 Proof

Established 1845 **SCIENTIFIC AMERICAN** November 1985 Volume 253 Number 5

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. COPYRIGHT © 1985 BY SCIENTIFIC AMERICAN, INC. ALL RIGHTS RESERVED. PRINTED IN THE U.S.A. NO PART OF THIS ISSUE MAY BE REPRODUCED BY ANY MECHANICAL, PHOTOGRAPHIC OR ELECTRONIC PROCESS, OR IN THE FORM OF A PHONOGRAPHIC RECORDING, NOR MAY IT BE STORED IN A RETRIEVAL SYSTEM, TRANSMITTED OR OTHERWISE COPIED FOR PUBLIC OR PRIVATE USE WITHOUT WRITTEN PERMISSION OF THE PUBLISHER. 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.

This transverse image of the brain was generated by a GYROSCAN magnetic resonance imaging system. It shows hydrocephalic ventricles and periventricular plaques characteristic of multiple sclerosis. The GYROSCAN is currently classified as an investigational device.

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.

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN for the flat fee of $1.25 per copy of each article or any part thereof. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970. Specify CCC Reference Number ISSN 0036–8733/84. $1.25 + 0.00.

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 shows the human-powered aircraft *Monarch B* in flight. *Monarch B* was designed and built at the Massachusetts Institute of Technology (see "Human-powered Flight," by Mark Drela and John S. Langford, page 144). Generally the pilot powers the aircraft by pedaling a chain drive that turns the propeller. At .5 horsepower the aircraft attains speeds of 19 to 23 miles per hour. To provide the additional horsepower needed for climbing, the pilot starts an electric motor powered by a battery that he has charged before flight by pedaling a generator. The pilot's hands grasp a control stick by means of which he can cause the craft to climb, descend or turn. *Monarch B* has a wingspan of 61 feet and is 40 feet long. At cruising speed the propeller makes about 210 revolutions per minute.

THE ILLUSTRATIONS

Cover painting by Ted Lodigensky

Page	Source
13	SCIENTIFIC AMERICAN
21–27	Ilil Arbel
52–58	Ilil Arbel
61	Ian Worpole
62–64	Andrew Christie
65	Trevor C. Weekes, Fred Lawrence Whipple Observatory (*top*); Andrew Christie (*bottom*)
66	Rick F. Harnden, Jr., Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory
67	Trevor C. Weekes, Fred Lawrence Whipple Observatory, and John C. Geary, Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory
68–69	Andrew Christie
71	David Bacon, University of Alberta
72–75	George V. Kelvin, Science Graphics
76	Feroze N. Ghadially, University of Saskatchewan
77–78	George V. Kelvin, Science Graphics
117	EROS Data Center
118–119	Benita L. Murchey and David L. Jones, U.S. Geological Survey
120	Anita G. Harris, U.S. Geological Survey
121–125	Hank Iken, Walken Graphics
127	Enid Kotschnig
128–134	Tom Prentiss
138	Warren W. Burggren and Alan W. Pinder, University of Massachusetts at Amherst
139–142	Tom Prentiss
144–145	Steven Finberg
146–150	Hank Iken, Walken Graphics
153	James Kilkelly
154–159	Jerome Kuhl
161	James Kilkelly
162	Andrew Tomko (*top*), James Kilkelly (*bottom*)
164	Patricia J. Wynne
165	Andrew Tomko
166–168	Patricia J. Wynne
171–172	Mount Wilson and Las Campanas Observatories
173–177	Michael Goodman

LETTERS

To the Editors:

Figures from the 1985 *Statistical Abstract of the United States* and from U.S. budget documents fail to confirm C. Arden Miller's argument ["Infant Mortality in the U.S.," SCIENTIFIC AMERICAN, July] that funding for Federal programs relevant to infant mortality (IM) has been decreasing. Figures for Aid to Families with Dependent Children have shown small increases each year and will continue to rise through 1989 under Administration proposals. The budget for food stamps rose from 1980 through 1983, decreased 9 percent in 1984, decreased between 1 and 2 percent in 1985, and is scheduled to rise between 4 and 10 percent each year through 1989. The Child Nutrition Program held steady from 1982 through 1985 before a 10 percent drop scheduled for 1986. The budget for Women-Infant-Children, which Miller shows as rising by 4 percent from 1982 through 1985, actually rose 88 percent in that period. Federal Medicaid funding rose 32 percent from 1982 through 1985 and will continue to rise through 1989. State Medicaid has been rising almost as fast. Alleging that decreases in funding through 1985 (future) caused a slowdown in improvement of IM in 1983 and 1984 (past) is inappropriate.

Even if Miller's allegations about program funding are granted, $28 billion in cuts would be responsible for the difference between the observed 1984 IM rate of 10.6 per 1,000 live births and the rate of 10.2, expected if the 5 percent annual decline had continued. This translates into about 1,500 lives, achieved at a cost of $1.8 million each. The assumption that identifiable programs can inexpensively improve IM rates is unwarranted.

The variables important for IM (race, age, prenatal care, weight gains, smoking, drug use, education, marital status and income) make it clear that cultural changes must occur if IM rates are to continue to improve. In any case, closer analysis of state health programs in states where IM rates have suffered setbacks (21 states in 1983–84, including Alabama, Colorado, Georgia, Hawaii, Nebraska, New Hampshire, Utah and Vermont) might yield more useful results.

Miller's citation of China as a positive example is questionable because of the difficulty of collecting data on China, the lack of any independent verification of IM gains and the possibility that hidden female infanticide lowers documented IM. Why Cuba is cited now that we have the analyses reported by Nick Eberstadt in the *Wall Street Journal* (December 10, 1984; page 26) showing that Cuba has systematically falsified IM data, I cannot imagine. Cuban IM may still exceed 40 per 1,000, as it did 20 years ago.

RAYMOND R. WHITE

Palo Alto, Calif.

To the Editors:

The political bias displayed in Miller's report should have no place in *Scientific American*. After listing numerous factors involved in infant mortality, most of which he concedes have not been studied statistically or in detail, he cites the reduction in Federal programs as the cause for the slowed rate of decline. He might just as well have claimed that the rise in births out of wedlock, shown in one of his charts, has leveled off as a result of the reduction in Federal funds.

W. BRYSON SCOTT

Lovingston, Va.

To the Editors:

The writers raise four issues, which these comments address:

1. Government spending. Dr. White is correct in pointing out that Federal spending for programs important for infant survival, after declining in 1982, has increased in recent years. These increases have come about because of the growing proportion of the population in poverty. Even as total welfare expenditures have increased, however, the eligibilities and benefits for the programs have been cut back. This means that a diminishing proportion of people in need are actually served by the programs, and those people who are served receive fewer benefits. When measured against 1981 levels of service, Women-Infant-Children is the only program that increased (by about 4.4 percent); the programs that decreased include Aid to Families with Dependent Children (AFDC), food stamps, child nutrition and maternal and child health.

The Congressional Budget Office reported in testimony on April 28, 1983: "In the last three years, benefits have declined significantly relative to the number of potentially eligible families, and in 1982 alone spending levels fell by about 5 percent in real terms. Two offsetting factors have affected outlay levels in this period. On the one hand, the number of low-income families has increased considerably since 1979, causing both eligibility and applications for benefits to rise. On the other hand, major cuts in these programs would have reduced outlays on them substantially had the recession not increased the number of beneficiaries. Even so, between 1981 and 1982 nominal expenditures for AFDC and food stamps for families with children each dropped by about $200 million."

White's reckoning of the cost of preventing one infant death fails to take into account the fact that the infant mortality rate serves as a surrogate measure for morbidity among infants who survive. Cost-benefit studies have generally favored programs that improve services and supports to pregnant women.

2. Cultural changes. Again, White is correct in pointing out that cultural changes affecting, for example, the age of childbearing, marital status and substance abuse would have an important impact on infant survival. Until our society knows how to bring about such changes while still preserving freedoms that most of us cherish, I believe the best we can do is to implement supports and services that help to mitigate the ill effects of demographic and behavioral trends. Coercive societies approach these problems with less trepidation.

3. Vital-statistics data. The reliability of vital-statistics data is always suspect. The 1984 UNICEF report gives an infant mortality rate for Cuba of 19 in 1981 and 70 in 1960. In an excellent review of vital statistics for 1983 (*Pediatrics,* Vol. 74, No. 6, pages 981–990; December, 1984) Myron E. Wegman reports that the rate for different regions varies between 10.8 and 21.4. The destruction of some female infants in some regions of China is deplorable but can hardly account for China's impressive record for overall infant survival.

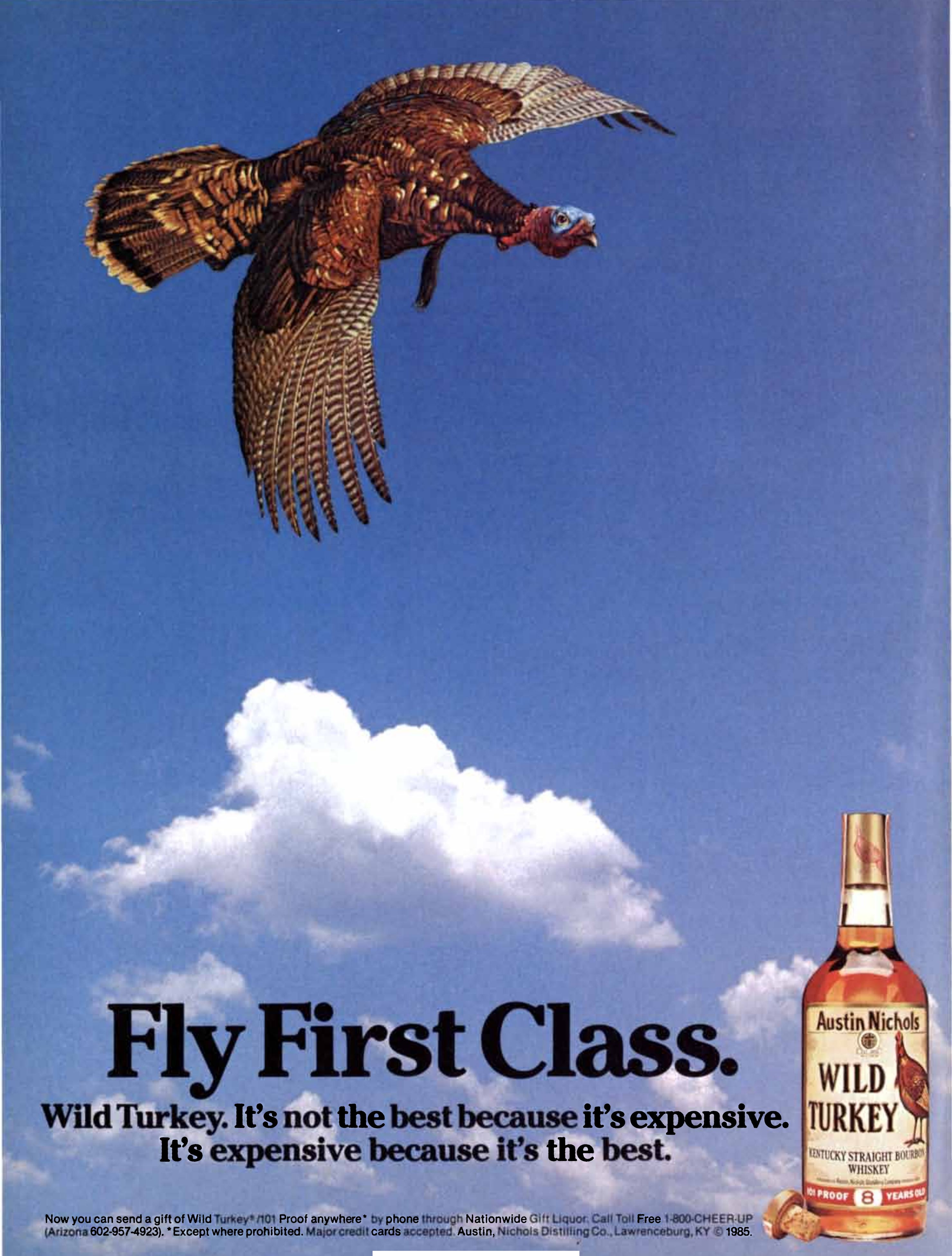
4. Political bias. Mr. Scott is correct in pointing out that my article is political, rather than a pure exercise in statistical inference, and that it is biased. Public policy in this country derives from negotiation among competing interest groups, each promoting different social values. Statistical inference, while producing data that may elevate the quality of negotiations, will not by itself lead to policy that is better-understood or better-formulated. A consideration of social values must enter the analysis. My bias is that the most vulnerable people in our society need protections only government can assure. I am also biased in believing a national interest is served by promoting the health and vigor of our people.

C. ARDEN MILLER

PURE ADRENALIN.

If you're ready for a rush, so is Dodge Daytona Turbo Z. At the tender age of two it had already mastered such tasks as leaping from zero to fifty in only 5.39 seconds. A very small number that translates into a rather big fact: Our Turbo Z beat Chevy's Z28.* Proof that beefy V-8s aren't the be-all and end-all to going in a straight line fast. With a bit of finesse you can fuel-inject and turbocharge half as many cylinders to do the job. But don't take our word for it, consult the street folk who write for *Car and Driver, Motor Trend, Road & Track*, et al. We're rather proud to say that they're rather impressed.

Of course, speed isn't the only rush to be had from the Turbo Z. The whole point of this machine's performance is *total* performance. And that means when the straight line turns into a curve, you've got front-wheel drive, nitrogen-charged shocks,


Fly First Class.
Wild Turkey. It's not the best because it's expensive.
It's expensive because it's the best.
Austin Nichols
WILD
TURKEY
KENTUCKY STRAIGHT BOURBON
WHISKEY
101 PROOF
8
YEARS OLD
Now you can send a gift of Wild Turkey®/101 Proof anywhere* by phone through Nationwide Gift Liquor. Call Toll Free 1-800-CHEER-UP (Arizona 602-957-4923). *Except where prohibited. Major credit cards accepted. Austin, Nichols Distilling Co., Lawrenceburg, KY © 1985.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

NOVEMBER, 1935: "'Visual hearing' is receiving considerable attention at Ohio State University. 'Lip-reading' (a misnomer, since the movements and expressions of the whole face and other parts of the body enter into visual hearing) has long been used by the hard-of-hearing, but its use in supplementing minor defects of hearing is a fairly recent development. For students with hearing loss the university provides a course of training whereby their latent visual-hearing ability can be strengthened considerably. A motion-picture method for accomplishing this result has been in successful use for the past five years."

"One of the main reasons for the alarming decline of wildlife in the United States is the lack of organization on the part of sportsmen. There are millions of individuals interested in fish and game, yet there is no articulate voice to speak for them. Now, however, there appears a light in the cloudy picture of conservation. Jay N. ('Ding') Darling, well known for his two years of work as the chief of the Bureau of Biological Survey, appears as one of the guiding spirits of the American Wildlife Institute. The function of this organization is to influence and guide the inarticulate sentiment for the preservation of wildlife."

"The scheduled operations of American air transport lines are showing remarkable growth. The figures on passenger-miles for the first six months of three years are so striking as scarcely to need comment: 1933—73,288,579; 1934—88,955,113; 1935—160,013,357. Each month this year shows an increase in activity over the preceding month and over the corresponding month of 1934."

"In the first six months of this year 851 airplanes were produced in the United States, a 14 percent increase over the corresponding period in 1934. These new planes include 517 for domestic civil use, 173 for military purposes and 161 for export."

"That especially far-reaching developments are about to be made in the fields of mining, power-plant equipment, television, aviation and automobiles is indicated by a barometric reading of the annual volume for the fiftieth year of the *Engineering Index*. It records engineering advancement in all its branches in all parts of the world for the past year."

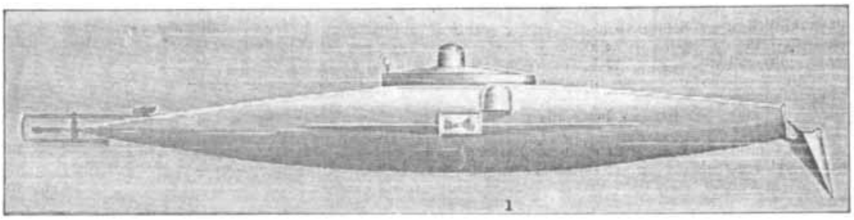
SCIENTIFIC AMERICAN

NOVEMBER, 1885: "The National Academy of Sciences met at Albany, N.Y., on Nov. 10, and enjoyed a large and distinguished attendance. During the four days of the session a substantial number of papers were presented; many of them were of more than ordinary interest. Prof. Young gave a *résumé* of the history of that erratic star in Andromeda and quoted Monck's hypothesis accounting for its luminosity on the ground that it may be a dark star passing through the nebula. Prof. Pickering's paper 'Stellar Photography' attracted marked attention. He cited a recent victory in gaining the impression of stars so distant or so minute as to be beyond the discovery of the most powerful telescope yet constructed. Major Powell's description of the stone ruins on the Colorado and the Rio Grande pointed to the conclusion that the arid regions now so characteristic of the interior of the continent were once fertile and well watered. Prof. Graham Bell made a report on hereditary deafness."

"The new submarine boat which has been built at Stockholm upon the plans of Mr. Nordenfelt has had an interesting and successful trial. Ever since the American civil war, naval engineers have been striving to solve the problem of submarine navigation, but until now with very little success. Mr. Nordenfelt's invention, however, appears to fulfill the numerous requirements for overcoming the dangers and difficulties of maintaining, driving and directing a boat beneath the water. The boat is built of steel and is cigar-shaped, with a glass conning tower in the center, from which the commander can keep a lookout. There are three engines, one to work the screw in the stern, which propels the vessel, and two to work the propellers on each side, which compel the boat to sink and maintain her at a certain depth. The motive power is steam. When the boat sinks, the fires have to be sealed, and reserve steam is used, which is kept at high pressure in two tanks. With this the boat has been driven for five hours at a speed of three miles an hour. Her speed on the surface is eight knots. The crew numbers three. During their submarine existence they have to subsist on the amount of air they take with them in the hull. Four men have subsisted for six hours without any inconvenience. The boat is 64 feet long and the central diameter is 9 feet."

"We are apt to regard the rain solely as a product of distillation and, as such, very pure. Yet a great number of industrial processes, domestic combustion of coal, natural changes in vegetable and animal matter, terrestrial disturbances such as tornadoes and volcanic eruptions, vital exhalations, etc., are discharged into the atmosphere and, whether by solution or mechanical contact, descend to the surface of the earth in the rain. The acid precipitation around alkali and sulphuric acid works is well known; the acid character of rains collected near and in cities and the remarkable ammoniacal strength of some local rainfalls have been fully discussed."

"A few months since, we gave an account of the experiments which were being made by the Central Pacific Railroad Company with petroleum as fuel on some of their steamboats. Since that time they have been able to determine more in detail concerning the results. On the freight steamer *Thoroughfare* they saved $7,000 in the cost of fuel in the five months they were using oil as compared with the five months of the same season last year, when they were burning coal. Besides saving 44 per cent in actual fuel they got rid of four firemen, which makes an additional saving of $240 per month. The oil costs $1.70 per forty-gallon barrel, or about four cents per gallon. Coal costs about $7 per ton."



Nordenfelt's submarine

The uses

Summary:

Even the smoothest voice is discontinuous, especially in conversation. Data communications has bursts of message and periods of silence, too. Even TV has some "bursty" traits. GTE scientists are isolating silences and inserting other messages into them. This permits voice and data to coexist on the same channel at the same apparent time. The development stems from parallel research in microelectronics, silence detection, speech, voice compression and signal processing.

Without basic change, or vast growth, telephone networks will be unable to cope with the anticipated traffic of the 1990's. The proliferation of personal computers and data terminals has already placed a strain on switching and transmission facilities. It has also placed demands on networks that are much different from the original voice-communications concept, in which average time of connection was three minutes.

Today, far shorter and far longer connections abound, more subscriber lines are in demand, and there are growing needs for enhanced services and faster switching.

Out of research dating from 1979, GTE has developed a switching system that promises not only to triple present transmission capacity but also to process calls 20 times faster. The system is called Burst Switching.

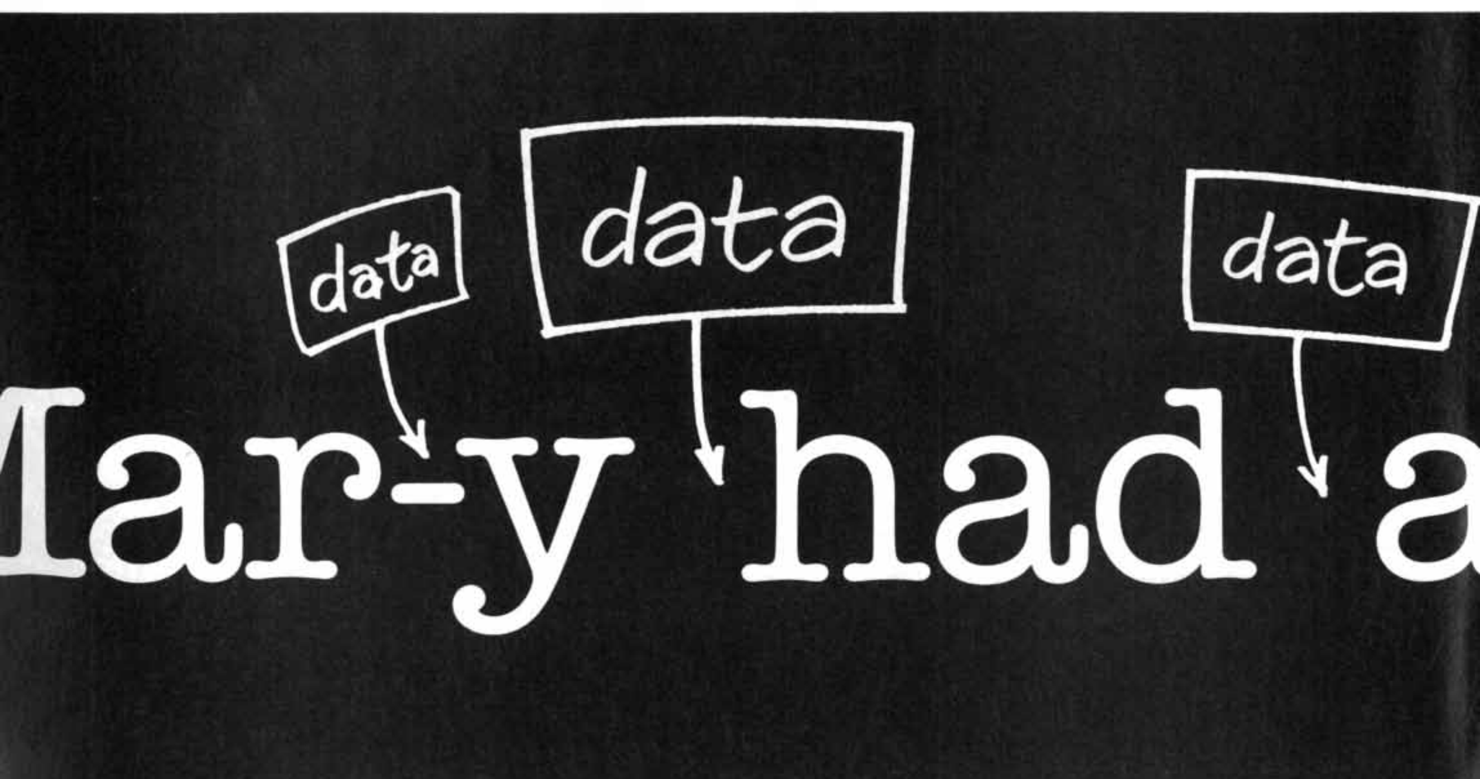
The nature of speech.

Our world is full of holes. Matter is mostly empty space. Conversation is mostly silence. But, even though speech is 2/3 silence interspersed with bursts of sound from 0.1 to 1.5 seconds long, if that speech goes over a telephone line, the line is locked up for the duration.

But, with Burst Switching, we can shoehorn other messages into the silences, automatically easing the pressure on transmission facilities. Theoretically, in fact, we triple transmission capacity.

VHSIC.

Through Very High-Speed Integrated Circuits (in which we are currently researching devices with submicron feature size), we are able to make and break telephone connections at increasingly high speeds. Voice lines need be dedicated only for the very brief duration of voice bursts. At other times, channels are available for other voice messages, or for data streams which are also "bursty" in nature. In addition, video, because of its built-in redundancy, can be considered to have bursts, too.



# of silence.

Message compression.

The capacity needed to transmit speech can be made even smaller if the information that must be sent to make it recognizable can be minimized. Our scientists have reduced the 64 kb/s signals to 16 kb/s while retaining high quality.

Thus, transmission-capacity requirement is reduced by a factor of four.

We are working, as well, on techniques for compressing video signals from 90 Mb/s to 64 kb/s. This will have special relevance for such activities as video conferencing.

So transmission capability grows and switching becomes faster—and we can now envision future telephone systems able to carry billions of simultaneous calls.

The box at the right lists some of the pertinent papers GTE people have published on Burst Switching and related subjects. For any of these, you are invited to write GTE Marketing Services Center, Department TPIIE, 70 Empire Drive, West Seneca, NY 14224.

Burst Switching experimental model.

Pertinent Papers.

Burst Switching–An Introduction, IEEE Communications Magazine, November 1983.

New Switching Concept Integrates Voice and Data Bursts, PROFILE, September 1983.

A PCM Frame Switching Concept Leading to Burst Switching Network Architecture, IEEE Communications Magazine, September 1983.

Application of the Burst Switching Technology to the Defense Communications System, Proceedings 1983 IEEE Military Communications Conference, MILCOM '83, Washington, D.C.

Performance Evaluation of a Distributed Burst-Switched Communications System, Proceedings Second Annual Phoenix Conference on Computers and Communications, March 1983.

A Complementary Speech Detection Algorithm, Proceedings of GLOBECOM '83, November 1983.

GTE

data data data

lit-tle lam

In Burst Switching, the roughly 65% silence in speech can be filled with data streams and other messages, effectively tripling transmission capacity.

THE AUTHORS

JOHN CAIRNS ("The Treatment of Diseases and the War against Cancer") is professor of microbiology at the School of Public Health of Harvard University. After getting his medical degree from the University of Oxford he moved to Australia, where he studied the multiplication of viruses and the structure of DNA molecules. He then came to the U.S. and, until 1968, served as director of the Cold Spring Harbor Laboratory of Quantitative Biology. In 1973 he returned to England to write *Cancer: Science and Society* (W. H. Freeman and Company, 1970). At Harvard, Cairns divides his professional attention between study of problems in public health and research on the origin of mutations.

P. KEVIN MACKEOWN and TREVOR C. WEEKES ("Cosmic Rays from Cygnus X-3") have collaborated on research projects in physics since their undergraduate days. MacKeown is professor of physics at the University of Hong Kong. He has a B.Sc. (1962) from University College in Dublin and a Ph.D. (1965) from the University of Durham. He held research positions at the Tata Institute for Fundamental Research in Bombay, the University of Maryland at College Park and Louisiana State University before moving to Hong Kong in 1970. In his work MacKeown has literally gone to extremes: cosmic-ray stations at 17,500 feet above sea level in the Andes and neutrino experiments 8,000 feet down in a gold mine. Weekes is an astrophysicist at the Smithsonian Institution's Fred Lawrence Whipple Observatory in Arizona. Like his coauthor he did undergraduate work at University College, where he also received a Ph.D. in 1966. He then went to Arizona to work at the Smithsonian Astrophysical Observatory as a National Research Council postdoctoral fellow. In 1967 he joined the staff of the Whipple Observatory. From 1969 to 1976 he was the resident director of the observatory. Weekes was granted a D.Sc. by the National University of Ireland in 1978.

ERNESTO CARAFOLI and JOHN T. PENNISTON ("The Calcium Signal") are professors of biochemistry respectively at the Swiss Federal Institute of Technology and at the Mayo Medical School. Carafoli was born and educated in Italy. He earned his medical degree in 1957 from the University of Modena and then worked at the university's Institute of General Pathology. While he was associated with the institute he spent some years at Johns Hopkins University as a postdoctoral fellow. In 1973 he took his current position. Carafoli spent much of this year as Fogarty Scholar-in-Residence at the National Institutes of Health (NIH); he will return to the NIH for part of next year. Penniston received a Ph.D. from Harvard University in 1962 for work on nucleic acids. After a year of teaching chemistry at Pomona College he moved to the Institute for Enzyme Research of the University of Wisconsin at Madison, where his research focus shifted to cell membranes. From 1971 to 1976 he was associate professor of chemistry at the University of North Carolina at Chapel Hill. He then accepted a position as consultant in molecular medicine at the Mayo Clinic and associate professor of biochemistry at the Mayo Medical School. Penniston was made professor and consultant in cell biology in 1979.

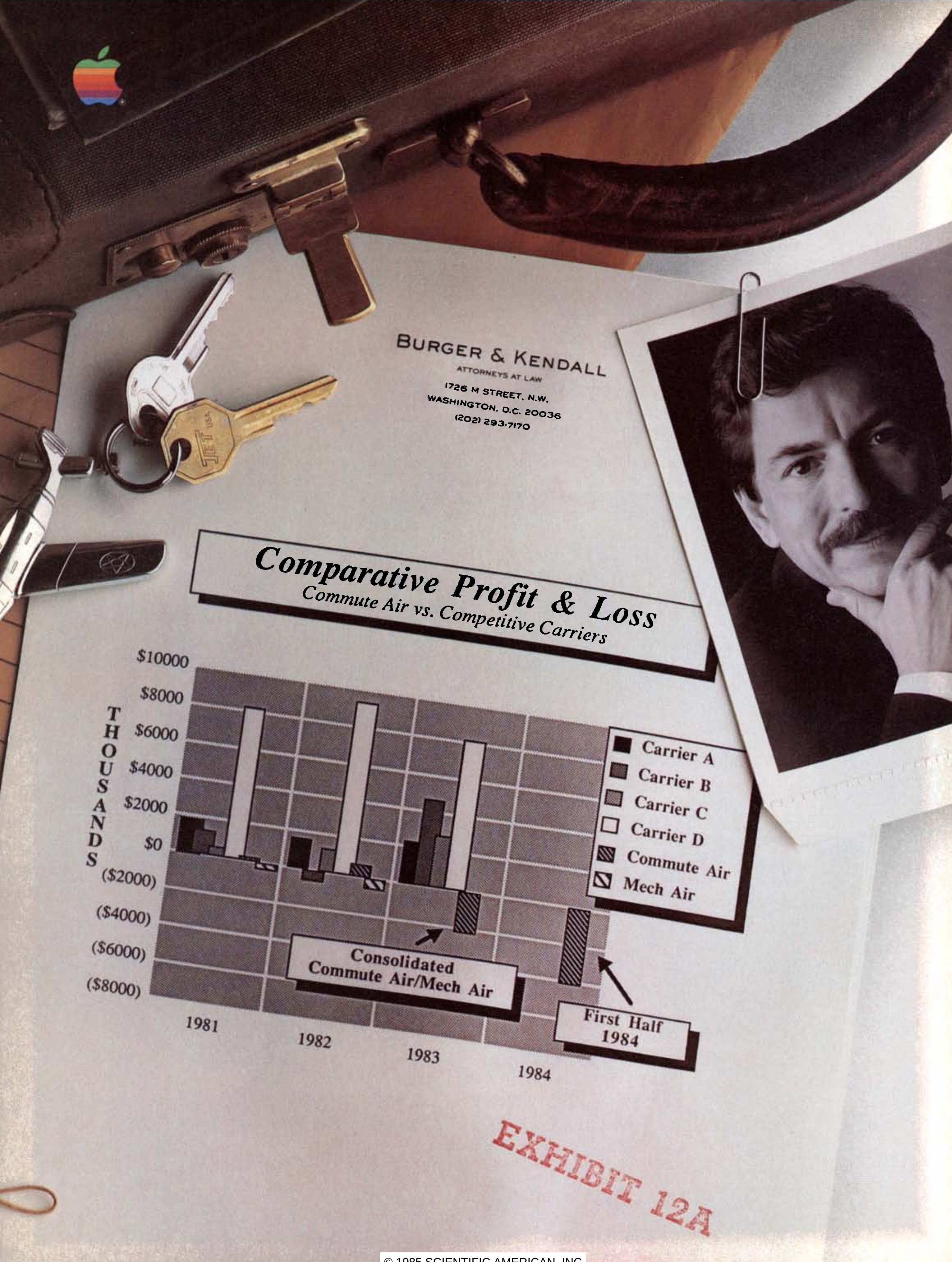
DAVID G. HOWELL ("Terranes") is a research geologist for the U.S. Geological Survey. He was graduated from Colgate University in 1962 and did graduate studies at the University of California at Santa Barbara, where he got a master's degree in 1969. He served three years in the U.S. Army before earning his Ph.D. from Santa Barbara in 1974. He then joined the Geological Survey. In 1980 he took a concurrent post as consulting professor at Stanford University.

MARTIN E. FEDER and WARREN W. BURGGREN ("Skin Breathing in Vertebrates") are respectively assistant professor in the department of anatomy at the University of Chicago and associate professor in the department of zoology at the University of Massachusetts at Amherst. Feder was graduated from Cornell University in 1973 and received a Ph.D. from the University of California at Berkeley in 1977. He spent two years as a postdoctoral fellow at the University of Chicago and a year as visiting research professor at Silliman University in the Philippines before taking his present position. Burggren did his undergraduate work at the University of Calgary and his graduate work at the University of East Anglia. After two years as a postdoctoral fellow at the University of British Columbia he joined the faculty of the University of Massachusetts. In 1982 Burggren was made associate professor.

MARK DRELA and JOHN S. LANGFORD ("Human-powered Flight") are doctoral candidates in the department of aeronautics and astronautics at the Massachusetts Institute of Technology and have collaborated on two human-powered aircraft projects. Drela got his bachelor's and master's degrees from M.I.T. Drela's academic specialty is computational fluid dynamics. Langford's undergraduate and graduate degrees are also from M.I.T.: a B.S. (1979) and an M.S. (1984) in aeronautics and astronautics, as well as an M.S. (1983) in political science and public policy. He is currently employed by the Institute for Defense Analyses in Alexandria, Va.; his job meshes with his research interests, which center on areas where aeronautics and public policy meet. Drela and Langford are now working on the *Daedalus* project, a feasibility study of a human-powered flight between Crete and the Greek mainland.

ROBERT McIVOR ("Smart Cards") is a manufacturing manager at Motorola, Inc., in Austin, Tex. He was educated at Kirkaldy Technical College and the University of Strathclyde in his native Scotland. He joined Motorola in 1970, and after rising to the position of operations manager in Scotland he came to the U.S. His responsibilities in his current position include all operations and production control for microprocessor manufacturing, as well as the management of projects for major customers. McIvor plays a leading role in coordinating and managing Motorola's participation in the development of smart cards, and he belongs to the national committee that is setting uniform standards for the devices.

JEAN S. AIGNER ("Early Arctic Settlements in North America") is professor of anthropology at the University of Alaska in Fairbanks. She began her undergraduate studies at the University of California at Los Angeles, but the enjoyable experience of an archaeological field trip to the Aleutian Islands prompted her to transfer to the expedition's sponsoring institution, the University of Wisconsin at Madison. There she completed her B.A., M.A. and Ph.D. degrees. While she was at Madison she developed interests in early circumpolar settlements and ancient Chinese archaeology. In 1969 she joined the faculty at the University of Connecticut. In 1977, following a sabbatical in the Federal Republic of Germany and Czechoslovakia, she moved to Alaska. In 1980 and 1983 Aigner did archaeological research in the People's Republic of China.


BURGER & KENDALL
ATTORNEYS AT LAW
1726 M STREET, N.W.
WASHINGTON, D.C. 20036
(202) 293-7170
Comparative Profit & Loss
Commute Air vs. Competitive Carriers
THOUSANDS
$10000
$8000
$6000
$4000
$2000
$0
($2000)
($4000)
($6000)
($8000)
Carrier A
Carrier B
Carrier C
Carrier D
Commute Air
Mech Air
Consolidated
Commute Air/Mech Air
First Half
1984
1981
1982
1983
1984
EXHIBIT 12A

"If you can't make your point with a Macintosh, you may not have a point to make."

On August 10, 1984, in a federal building in Philadelphia, an attorney named Jim Burger and a computer named Macintosh™ won their very first arbitration.

Hands down.

We like to refer to the case as: *Jim & Macintosh vs. The Surprised Lawyer & His Art Department.*

Multiplan. Chart. Word.

It involved an airline/labor dispute. And enough paperwork to cover the entire floor of Jim's office in the weeks before the hearing.

There were binders full of industry comparatives on profit and loss. Copies of wage agreements. Wage disagreements. Testimony. Research notes scrawled on legal pads.

It was too complicated a job to recruit help from legal assistants. Too urgent to wait for the usual secretarial face-lift.

There was only one source Jim could turn to for help: his Macintosh.

In a succession of 16-hour days, using powerful, easy to use software like Microsoft's Multiplan, Chart and Word, MacDraw™ and ThinkTank 512, Jim turned an undeniable mess into an indisputable case.

In part, by presenting the arbitrator with a 50-page document containing exhibits like the one exhibited here.

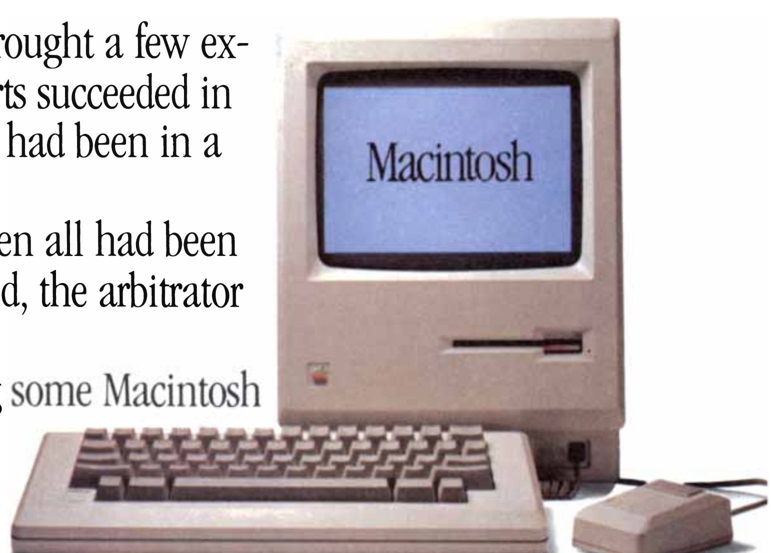
"It gave me a tremendous amount of confidence having that document under my arm," Jim remembers.

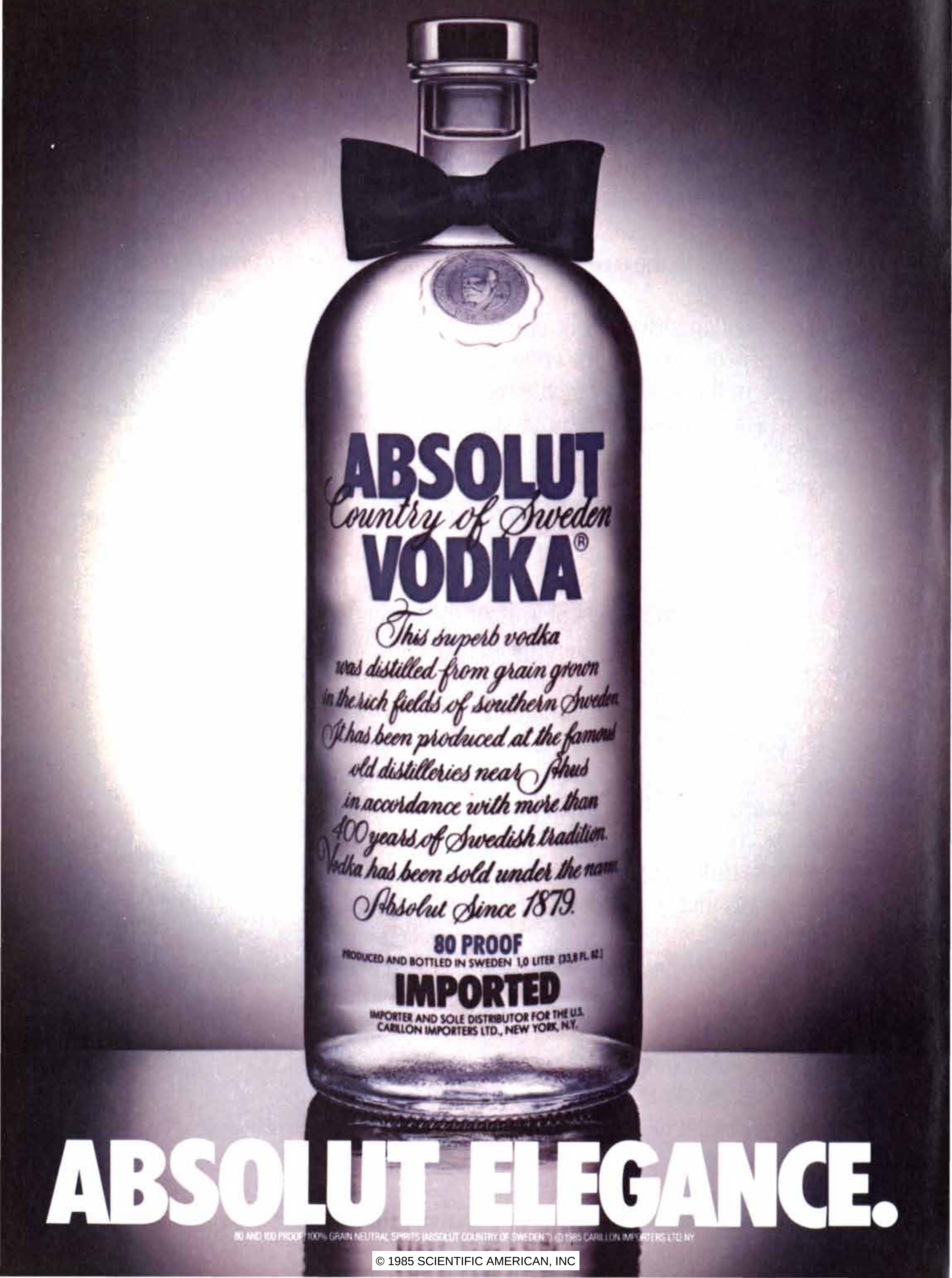
Of course, the opposing lawyer had brought a few exhibits of his own. Unfortunately, all his charts succeeded in demonstrating was that his art department had been in a hurry to do them.

To make a long arbitration short, when all had been said and done and objected to and overruled, the arbitrator decided in favor of Jim's client.

Jim's client decided in favor of buying some Macintosh computers of his very own.

And we decided it all made another great case for Macintosh.




ABSOLUT
Country of Sweden
VODKA®
This superb vodka was distilled from grain grown in the rich fields of southern Sweden. It has been produced at the famous old distilleries near Åhus in accordance with more than 400 years of Swedish tradition. Vodka has been sold under the name Absolut since 1879.
80 PROOF
PRODUCED AND BOTTLED IN SWEDEN 1,0 LITER (33,8 FL. OZ.)
IMPORTED
IMPORTER AND SOLE DISTRIBUTOR FOR THE U.S. CARILLON IMPORTERS LTD., NEW YORK, N.Y.
ABSOLUT ELEGANCE.
80 AND 100 PROOF/100% GRAIN NEUTRAL SPIRITS (ABSOLUT COUNTRY OF SWEDEN™) © 1985 CARILLON IMPORTERS LTD NY

COMPUTER RECREATIONS

Exploring the field of genetic algorithms in a primordial computer sea full of flibs

by A. K. Dewdney

Imagine an abstract sea inhabited by abstract organisms called finite living blobs, or flibs. Each flib is equipped with the simplest decision-making apparatus possible. This is the biological equivalent of what computer scientists call a finite automaton. Each flib also contains a single chromosome consisting of a string of symbols that encodes the automaton. The flibs inhabit a primordial, digital soup in constant flux. These changes must be predicted accurately by the flib if it is to survive.

In the primordial soup I recently set simmering in my computer, flibs that predicted poorly died out. The best predictors left progeny that sometimes improved on ancestral performance. Eventually a line of perfect predictors evolved.

Flibs and their evolutionary tendencies illustrate nicely a form of programming known as the genetic algorithm. Pioneered by John H. Holland of the University of Michigan in the 1960's, the technique is sometimes able to solve difficult problems by evolving a sequence of approximate solutions. New solutions are produced by mating the best of the old solutions with one another. Before long a new solution that is superior to its parents appears and joins the list of preferred breeders. Genetic algorithms have been applied with some success to pattern recognition, classifier systems, pipeline operation, symbolic layout and a small number of other problems. In my computer soup the technique yielded superior flibs. Was this success due to the general efficacy of the genetic-algorithm method or to the simplicity of the predictive task facing the flibs? The question is hard to answer. It can be pondered and the underlying phenomenon can be reproduced by any interested reader who has a computer within reach.

A finite automaton has a finite number of states; an input signal causes it to change automatically from one state to another. The kind of automaton used in a flib also generates signals. Incoming and outgoing signals are represented within the automaton by symbols. When a signal is received, the automaton changes state and emits a second signal.

A state-transition table is useful for representing the process. For example, a finite automaton that is capable of assuming three states, *A, B* and *C*, and that can handle afferent and efferent 0's and 1's fits nicely into a 3-by-4 table. For each state the automaton might find itself in, and for each symbol it might receive, there are two entries. The first entry gives the corresponding output symbol; the second entry gives the state that the automaton next assumes:

	0		1	
A	1	*B*	1	*C*
B	0	*C*	0	*B*
C	1	*A*	0	*A*

The automaton represented by this table might well find itself in state *C* at some time. If the automaton receives a 1, the table tells us the automaton will generate a 0 and enter state *A*.

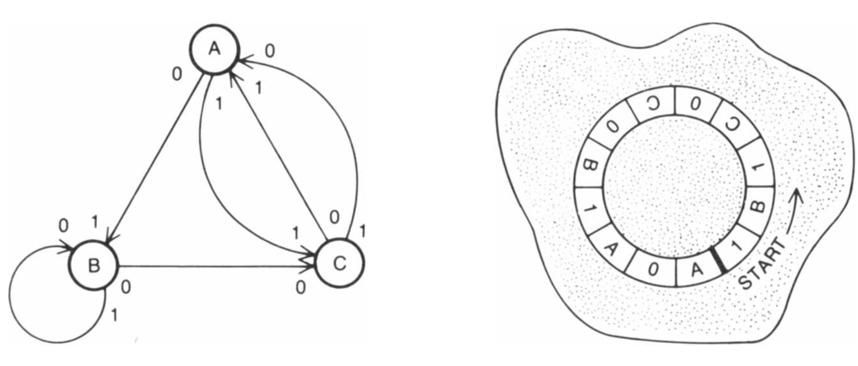
Another representation, easier for humans to read, is the state-transition diagram, in which circles represent states and arrows represent transitions. If an automaton goes from one state to another when it receives a specific symbol, an arrow should be drawn from one state circle to another. The arrow should be labeled both with the input symbol that caused the transition and with the resulting output symbol [*see illustration below*].

A finite automaton always begins its operations in a specific state, called the initial state. At each tick of an imaginary clock a new symbol arrives, a new symbol leaves, and a new state is entered. The automata used in my flibs all send and receive the same two symbols, 0 and 1.

How is one to interpret the behavior of a flib if so little is known about the creature's biology? Therein lies the joy of abstraction. The symbols received by the automaton are merely sensory messages from the environment. In corresponding fashion, an output symbol can be viewed as a response by the organism to the environment's most recent condition.

The concept of a flib is so flexible that input and output can represent a great variety of specific biological phenomena. For example, an input signal could represent a chemical or temperature gradient. The corresponding output symbol could be a command to an effector that controls cilia, or a spore-forming mechanism. A task of great importance to a creature wishing to evolve to some minimally acceptable level (say that of a university professor) is to predict the environment. To a flib the environment is a seemingly unending sequence of 0's and 1's. Insofar as symbols received indicate significant events, there is clearly some advantage in the ability of a flib to predict the next symbol, particularly if under some more specific interpretation of flib functioning the flib's survival were enhanced.

Most flibs are rather poor at predicting their environment in this sense. For example, the flib described by the



A state-transition diagram **(left)** *and a corresponding flib with its chromosome* **(right)** state-transition table given above responds to the environmental sequence

0111000010110...

with the outputs

1000011001000...

At each stage of its operation the flib's output is its prediction of the next symbol to arrive from the environment. To find the number of correct predictions shift the output sequence one symbol to the right and compare it bit by bit with the input sequence. Count the number of matching symbols. In this case the flib predicted correctly only six of the 12 incoming symbols, a score that is no better than might result from random guessing.

One can easily demand too much from a finite automaton. Indeed, it is unfair to ask a flib to predict any nonperiodic environment. Readers might like to ponder this point for a moment. Why must a perfectly predicted sequence of input symbols consist of the same basic string endlessly repeated? For example, the 3-state flib that failed the prediction test just set for it succeeds brilliantly on the following environmental sequence:

010011010011010011...

Here the environment marches to the beat of a simple repetition, 010011.

There are several dozen 3-state flibs, but only a few of them can predict this sequence perfectly. Among flibs that have more than three states perfect predictors for a given environmental sequence are rare and become more so as the number of states increases. Predictability depends heavily on the period of the sequence: no n-state flib will ever be able to predict the sequence that results from continued repetition if the basic string of symbols is too long. There is evidently a relation between the number of states a flib can have and the largest period in a sequence that it predicts perfectly. Readers might enjoy discovering the relation for themselves. What is the longest period an n-state flib can predict?

A flib is more than a finite automaton trying to predict its environment; it has a chromosome. Flibs periodically breed (by some unknown method). An examination of the chromosome in its relation to a flib's finite automaton shows how the inherited genes determine the behavior of the offspring. Start with the state-transition table and strip away the rows, one at a time, from top to bottom. Join the rows together end to end and then join the beginning of the string to its end. The result is a circular chromosome.

Before the final joining operation, the chromosome of our 3-state exemplar appears as a string of 12 genes:

1B1C0C0B1A0A

Strictly speaking, the symbols in this string are alleles. An allele is a specific form of a gene that appears at a given locus. As such, a gene can be specified either by its name or by its locus. Thus the seventh symbol from the left controls a flib's output symbol when it is in state *B* and a 1 is received from the environment. The locus here is 7.

I recently set up a primordial soup containing 10 4-state flibs in my personal computer. Before 1,000 of the time units I call chronons had passed none of the original flibs was alive. All had been replaced by superior predictors. The display screen showed the highest and lowest scores attained in the current population. The lowest score fluctuated a good deal; the highest score crept slowly upward [*see illustration at left*]. Just when I was beginning to give up hope that a perfect predictor would evolve, one suddenly appeared, whereupon the highest score jumped to 100.

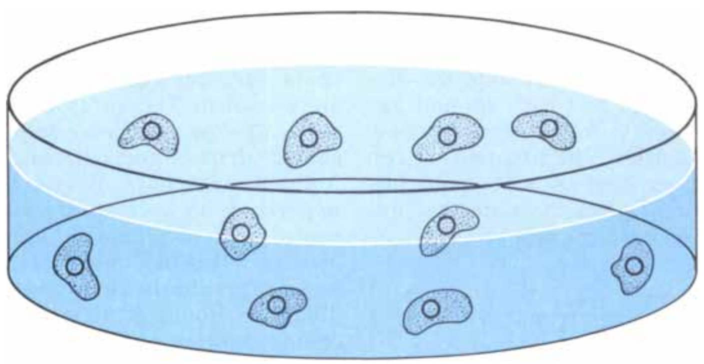
All of this raises the question of just how flibs evolve in my computer soup. Periodically a cosmic ray zips through the broth and strikes a random chromosome at a random locus; the result is that a specific gene is changed from one allele to another. For example, in the following 4-state flib chromosome the gene at locus 3 controls the output symbol for the transition from state *A*, when the creature receives a 1:

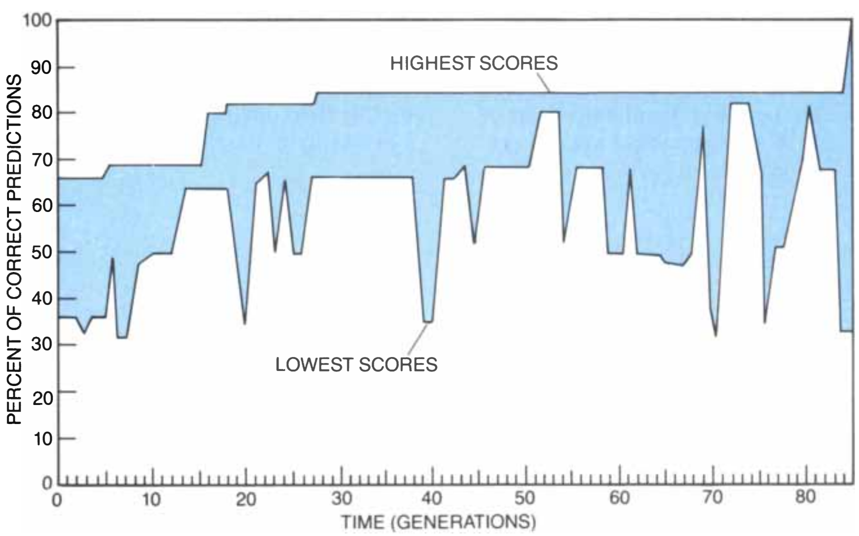
0D1C0D0B1A0C1B1A

A cosmic ray striking this gene changes the chromosome slightly:

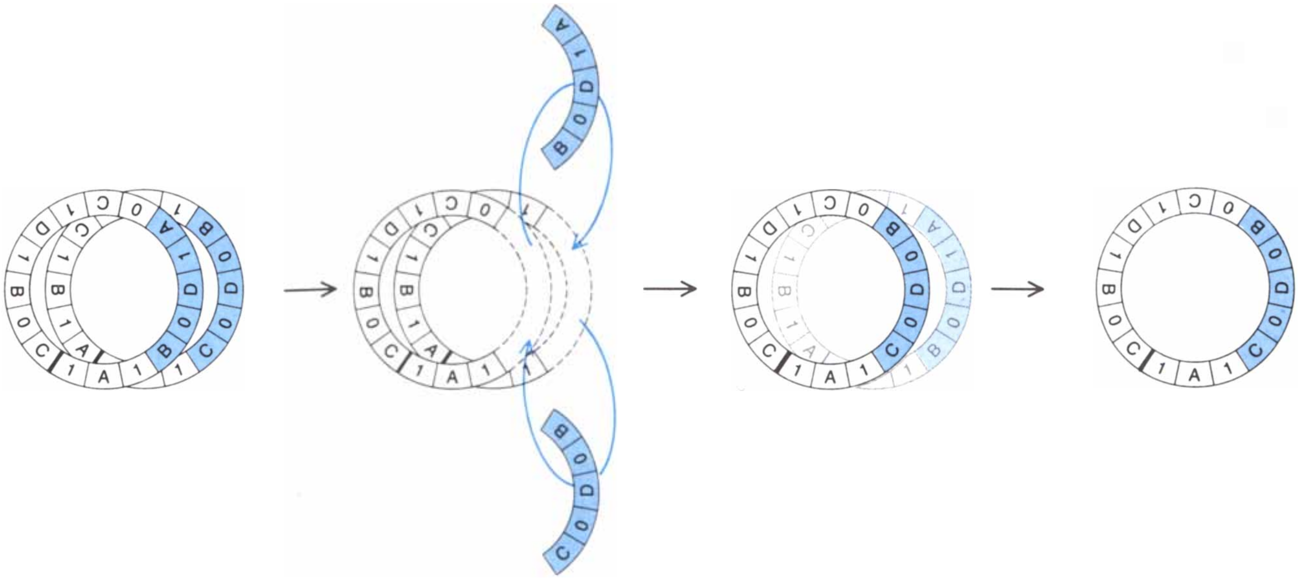
0D0C0D0B1A0C1B1A

Mating is the other source of variation in the flib gene pool. During the mating season the highest-scoring flib shuffles genes with a randomly selected flib. The offspring bears a composite chromosome. One part comes from the superior parent, the other from





A soup of 10 flibs **(top)** *evolves a perfect predictor* **(bottom)**



Crossover of two flib chromosomes, and the resulting chromosome of the offspring **(right)**

the winner of the mating lottery. The composition resembles a phenomenon called crossover that takes place in real chromosomes. In flib chromosomes crossover can be illustrated by combining the first (unaltered) chromosome listed above with another:

```
1A1B0D1A0C1D1B0C
  ↑       ↑
```

Arrows indicate randomly selected crossover points. The offspring's chromosome is identical with that of the second parent as far as the first crossover point. Between points it is identical with the first parent's chromosome. After the second point it is again identical with the chromosome of the second parent [*see illustration above*]:

1A1C0D0B0C1D1B0C

Before actually writing and testing the primordial program, I was somewhat skeptical of the value of crossover breeding. I was surprised to find, however, that if the first parent is a reasonably good predictor, the offspring tends to be as well.

Readers may judge the issue themselves by writing a program called AUTOSOUP. Listed, the program does not extend much beyond a single page. It consists of four modules embedded in a loop. A limit that defines the top score should be set. As long as the top score is less than the limit the program should continue to run through the four modules.

In the first module the 10 flibs are scored on a sequence of 100 environmental symbols. The second module identifies the flibs with the highest and lowest scores that result. In the third module the top-scoring flib is bred with a randomly selected mate. The offspring of this union replaces the bottom-scoring flib. In the fourth and last module a cosmic ray arrives, strikes a random flib and causes a mutation. Just before the program invokes the third (breeding) module a random number is selected. If the number falls below a certain threshold, the program will skip around the breeding module and execute the mutation module immediately. The threshold can be set to any level. Certain settings, however, are better than others; if the breeding module is executed too often, the small population quickly becomes dominated by the genes of the top-scoring flib. The gene pool loses diversity and evolution slows to a painful crawl if not to a downright standstill. Evolution slows down, in any event, as the scores get higher. The top-scoring flib remains on the scene for a lengthening period because it becomes increasingly unlikely that flibs superior to it will evolve.

Four arrays are useful in AUTOSOUP. They are called *chrom, state, score* and *e. Chrom* is a two-dimensional array of 10 flibs and 16 genes. *Chrom*(i,j) refers to the jth gene of the ith flib. *State* and *score* contain the current state and score of the 10 flibs. The fourth array, *e,* contains the basic string used to generate environmental symbols. This array is received from the keyboard at the beginning of the program.

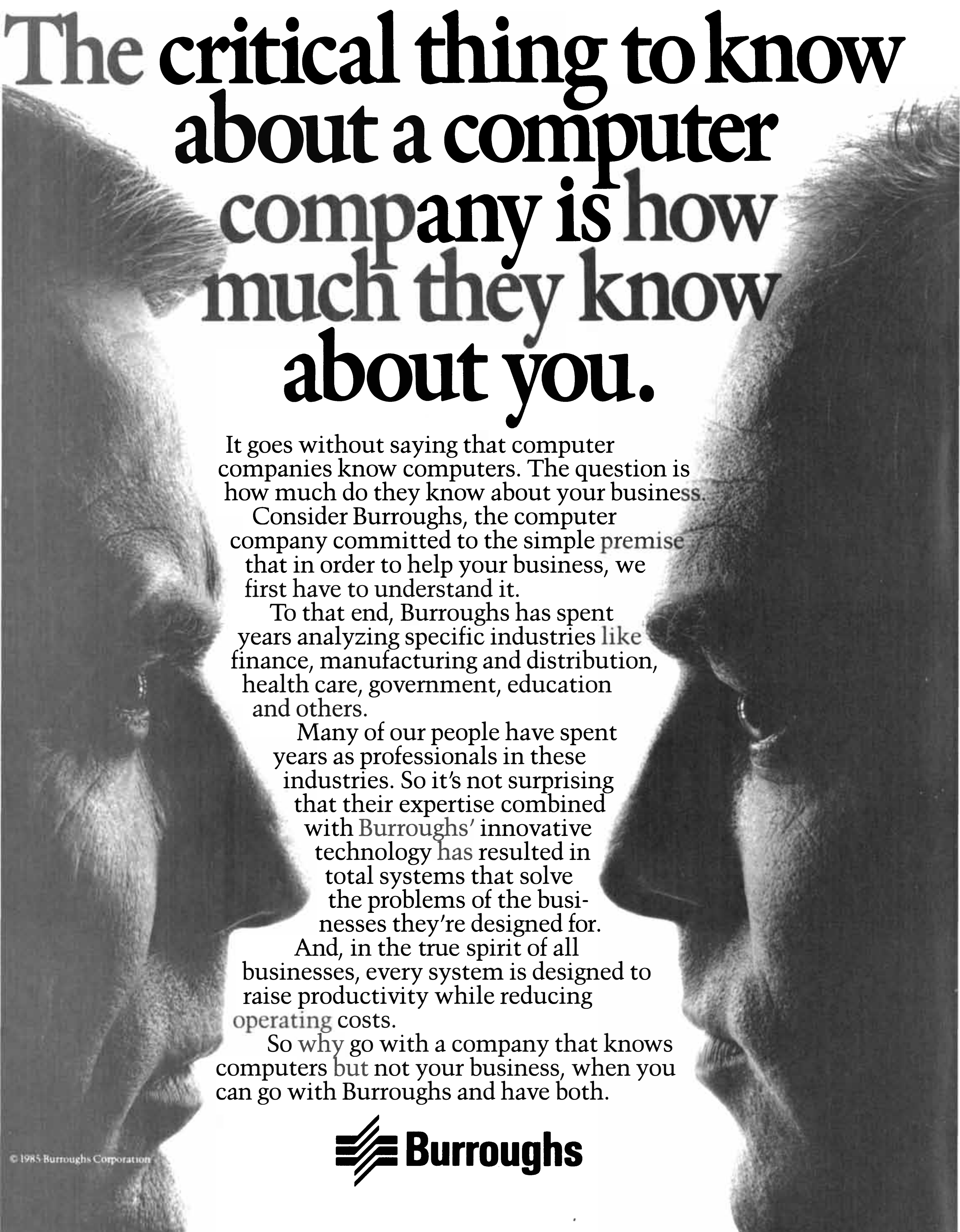
Flibs are evaluated by means of a double loop. The outer loop generates 100 environmental symbols and the inner loop increases the score of each flib if it correctly predicts the next symbol. One can test 4-state flibs adequately on environments of period 6, a challenge of intermediate difficulty. Perfect predictors may require a day's run to evolve in a period-8 environment, whereas period-4 environments are almost no challenge at all. Two tricks are useful in this module. The first trick retrieves the next environmental symbol from the outer-loop index i by computing i modulo 6; the result is the remainder of i on division by 6. The number can be used to index the array *e.* As i runs from 1 to 100 the computed index runs through the array repeatedly, producing the proper sequence of environmental symbols. Given the index of the current symbol, the next symbol is easy to compute and look up. This symbol is compared with the prediction made by each flib in turn.

The second trick enables the program to find the flib's next state quickly and determine its output merely by scanning the chromosome. Instead of representing the four states by *A, B, C* and *D,* the numbers 0, 1, 2 and 3 are used as entries in the array *state.* If the environmental symbol is called *symb,* the output of the ith flib can be found by first using a simple formula:

$$l = 4 \times state(i) + 2 \times symb\,.$$

Then *chrom*(i,l) should be identified. The locus l on the ith flib's chromosome yields its output when the creature is in state i and is receiving input *symb.* The next state occupies the locus $l + 1$.

The module that determines the top and bottom flibs uses an exercise common in elementary programming courses: given an array of n numbers, write a program that finds the largest number. The solution involves setting a variable called *top* to 0 and then scan-


The critical thing to know about a computer company is how much they know about you.
It goes without saying that computer companies know computers. The question is how much do they know about your business.
Consider Burroughs, the computer company committed to the simple premise that in order to help your business, we first have to understand it.
To that end, Burroughs has spent years analyzing specific industries like finance, manufacturing and distribution, health care, government, education and others.
Many of our people have spent years as professionals in these industries. So it's not surprising that their expertise combined with Burroughs' innovative technology has resulted in total systems that solve the problems of the businesses they're designed for.
And, in the true spirit of all businesses, every system is designed to raise productivity while reducing operating costs.
So why go with a company that knows computers but not your business, when you can go with Burroughs and have both.
Burroughs
© 1985 Burroughs Corporation

ning the array within a simple loop. Each entry is then compared with *top*. If it is larger than *top, top* should be replaced by that entry. The index should be saved as well. The same procedure can be inverted and used to find the bottom score. This time a variable, *bot,* should be set at 100 and replaced by entries that are smaller.

The third module breeds the top-scoring flib with a randomly selected member of the population. The only difficulty in writing this segment arises in the selection of the two crossover points. I think it is easiest to select two random integers c_1 and c_2 from the range 1 through 16. If c_1 is greater than c_2, their values should be exchanged. With just a little finesse readers will discover how three loops that range from 1 to c_1, c_1 to c_2 and c_2 to 16 supply the machinery to move elements of *chrom* on the breeding rows into the destination row, which is occupied by the doomed flib with the lowest score.

In the fourth module a random flib index and a random *locus* should be selected. The parity of *locus* determines whether a state gene or an output gene is to mutate. If the value is 0, then 1 modulo 2 should be added to the number already stored there. This flips the bit, so to speak. If the *locus* value modulo 2 is 1, then 1 modulo 4 should be added to the array entry. This changes the state stored there.

Have I cheated? Surely a systematic change of state from 0 to 1 to 2 to 3 and back again is hardly a random effect. My answer is that it is random enough: the number of states is small enough so that one cannot expect the final outcome of the program to be much different from the outcome when more randomly selected states prevail. Indeed, I also cheated in a mild way by choosing c_1 and c_2 so carelessly: the method guarantees an advantage for certain substrings in relation to others. Again, I think differences between AUTOSOUP and a statistically corrected crossover selection procedure would be slight. Either way there is so much juggling of genes and cracking of chromosomes that the top flib is hard put to recognize its own grandchildren.

The only parts of AUTOSOUP as yet unspecified are its beginning and its end. The flibs initially occupying the soup should be selected randomly. For each gene in each flib an integer should be selected from the appropriate range and assigned to that gene. Finally, when a flib first exceeds the limit set in the outer loop, AUTOSOUP should print it.

Readers embarking on this genetic adventure are warned that there is much exploring to do. Perhaps some explorers will become addicted. Questions to be answered concern the presence of evolution and its speed. When an environment period is too long for a 4-state perfect predictor to evolve, how fit do the flibs get? How do changes in the length of the period affect the amount of time it takes a perfect predictor to evolve? Nothing in the AUTOSOUP description prevents extending the program to 5- and 6-state flibs. One can even alter the program to explore nonperiodic environments or ones that occasionally change their basic string.

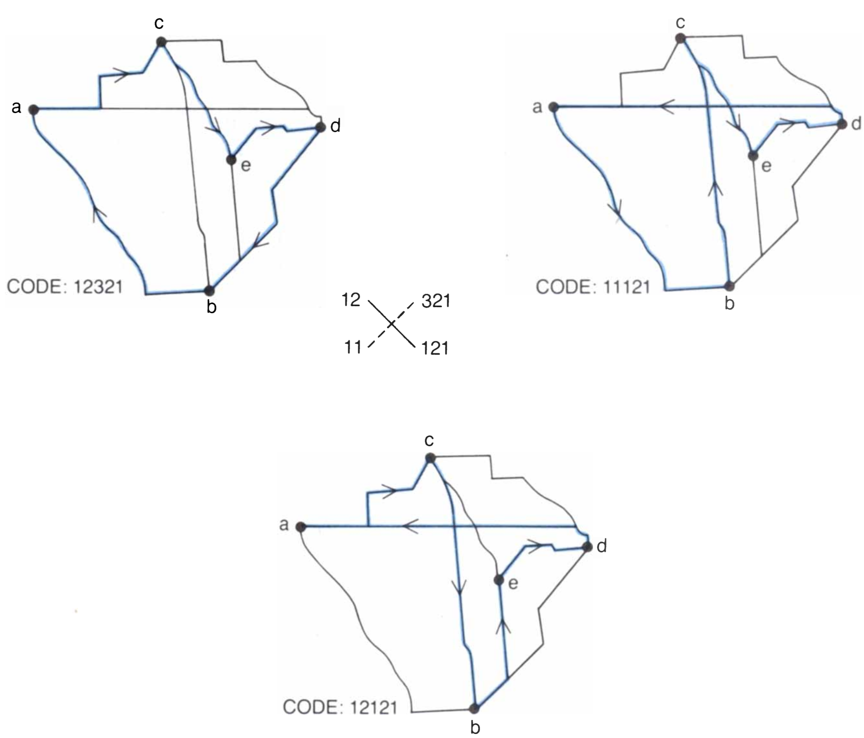
Automaton soup was inspired by a book that appeared in the early 1960's. Titled *Artificial Intelligence through Simulated Evolution,* the book describes a series of experiments in the evolution of automata by Lawrence J. Fogel, Alvin J. Owens and Michael J. Walsh. Automata were asked to predict periodic sequences and were allowed to evolve much the same way as our flibs. No breeding or crossover was allowed in this austere study, however.

It was Holland who suggested that I add the crossover facility to the automaton soup. As noted above, Holland is the acknowledged father of the genetic algorithm. Practitioners of the discipline, growing steadily in number, met at their first large-scale, funded conference, held at Carnegie-Mellon University. They discussed a wide range of theory and applications. A problem explored in several papers serves as an interesting introduction to the subject of genetic programming.

Called the traveling salesman problem, it poses the following challenge: Given a map of *n* cities connected by a network of roads, find the shortest tour of the *n* cities. Such a tour can then be used by a salesman or saleswoman to minimize travel expenses. In spite of my inclusion of salespeople of both sexes, the foregoing description has a 1940's ring to it. But more modern methods of travel and actual costs are easily incorporated without changing the mathematical skeleton implicit in the statement.

It is entirely possible that a tour of minimum length can be made to evolve just as perfect flib predictors evolved from lesser flibs. Each tour should be encoded in a chromosome. The shortest tours should be bred in the hope of obtaining yet shorter offspring. Crossover yields the chromosomes of the progeny.

It is a beguiling task to choose a good representation for a tour. For example, if one simply uses a list of the cities in the order visited, the offspring may not even be a tour. To get around this difficulty the authors of one paper, John Grefenstette, Rajeev Gopal, Brian Rosmaita and Dirk Van Gucht of Vanderbilt University, propose an ingenious chromosome. The representation for a five-city tour such as *a, c, e, d, b* turns out to be 12321. To obtain such a numerical string reference is made to some standard order for the cities, say *a, b, c, d, e.* Given a tour such as *a, c, e, d, b,* systematically remove cities from



Two parent salesman tours **(top)**, *and an offspring* **(bottom)** *resulting from genetic crossover*

A THOROUGH EXPLANATION OF THE 16-VALVE SAAB 900 ENGINE.

This is for all those people who were not far-sighted enough to take Auto Shop as an elective in high school.

Where do we start? Well, let's begin with the familiar: the car you're presently driving.

The overwhelming odds are that for each cylinder your car has, it has two valves. One to let the fuel in and run the engine; the other to let the gases out after they've been burned.

Hence the rule: Make the valves bigger and you increase what's called the *volumetric* efficiency. (In plainer English, the larger the valve, the easier it is for gas to come in and exhaust to go out.)

Unfortunately, the bigger you make the valves, the farther away you move the spark plugs from the center of the combustion chamber. And that starts to wreck what's called the *thermal* efficiency.

The problem: How to increase one efficiency (the volumetric) without decreasing the other (the thermal).

Engineers had been solving that problem for years in their competition and rally cars. Instead of just making the valves bigger, they simply doubled the number of valves.

Each cylinder, then, gets two valves to ingest the fuel and two valves to throw off the exhaust.

Getting back to your car. If it has a four-cylinder engine, it probably has eight valves.

1986 SAAB 900S	
Power output:	*125 HP/92 kw @ 5500 rpm*
Max. torque:	*123 ft. lbs./166 NM @ 3000 rpm*
Compression ratio:	*10.1:1*
Fuel injection:	*Bosch LH electronic*
Ignition:	*Bosch electronic with knock detector*

1986 SAAB 900 TURBO	
Power output:	*160 HP/118 kw @ 5500 rpm*
Max. torque:	*188 ft. lbs./255 NM @ 3000 rpm*
Compression ratio:	*9.0:1*
Fuel injection:	*Bosch LH electronic*
Ignition:	*Bosch electronic, Hall effect*

Whereas a Saab 900S or Turbo (the descendants of rally and competition cars) has sixteen.

Back in the days of cheap gas and free glassware, none of this engineering cleverness was needed. (Well, it *was* needed; it just wasn't called for.)

All you had to do then was worry about performance (V-8 engines and the like) and throw efficiency to the wind.

Today, Saab throws nothing to the wind.

Engine weight: On a football field, 180-pound cornerbacks are faster than 290-pound defensive tackles. They have less weight to carry around. So, too, the four-cylinder, two-liter Saab engine block. Everything is attuned to keeping that weight down: The cylinder head, camshaft cover, and intake manifold are built from lightweight aluminum alloys instead of clunkier materials.

Fuel injection: The fuel is not burned mindlessly; the computer-like Bosch fuel injection always maintains the most efficient mixture of gas and air.

Even the stroke the piston makes is only this long:

3.07 inches. So hours and hours and hours and hours of high cruising speed don't create unnecessary wear and tear on the engine.

Saab offers both turbocharged and 16-valve naturally aspirated engines.

In the legendary Saab Turbo. And the not-yet-legendary new Saab 900S.

They prove that there are more ways than one to achieve what is more appropriate to poets than to engineers: the picture of driving exhilaration you see on the right.

SAAB

The most intelligent cars ever built.

Saabs range in price from $12,285 for the 900 3-door 5-speed to $18,695 for the 900 3-door 5-speed, 16-valve Turbo. 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.

A MORE THOROUGH EXPLANATION.

Expected.

It's come up before.
It will come up again.

And that's what's wonderful about it. Some of the best things in life are predictable, repetitive, expected. Thank heaven they are.

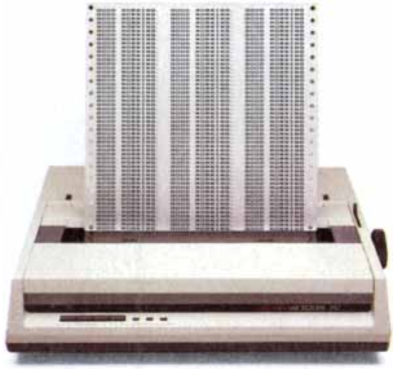
Even in business, it's nice to have some things you can count on. Things that work so dependably, so expectedly, they're completely forgettable.

Well, let's hope not completely forgettable.

When it comes to "expected" performance in printers for personal computers, the name is Okidata. Our record for reliability—for dependable, day-in, day-out performance—has set the industry standard.

Or, to put it more expectedly, we think it's the best.

Okidata printers work expectedly with IBM®, Apple®, and all other personal computers.



OKIDATA®
an OKI AMERICA company

We put business on paper.

the standard list in the order of the given tour: remove *a,* then *c, e* and so on. As each city is removed from the special list, note its position just before removal: *a* is first, *c* is second, *e* is third, *d* is second and, finally, *b* is first. Hence the chromosome 12321 emerges. Interestingly, when two such chromosomes are crossed over, the result is always a tour [*see illustration on page 27*]. With such a representation tours can now be bred, so to speak, for fitness.

Most practitioners of the art of genetic algorithms readily admit that the traveling salesman problem is one of their greatest challenges. Although experiments with the representation just described were not very encouraging, there are other genetic algorithms that perform better on the problem.

Still, no genetic algorithm has ever been able to conquer the traveling salesman problem in any well-accepted sense. This is undoubtedly owing to the difficulty of the problem itself. Because it is what theorists call NP-complete, it may be doomed to eternal insolubility in the practical sense.

Since its appearance in this column last August the MANDELZOOM program has been incarnated in hundreds of homes, schools and workplaces. Although the program has apparently awed adults, intrigued teenagers and frightened a few small children, to my surprise the mail on iteration diagrams, the secondary topic, nearly equals that on the Mandelbrot set.

Many readers have lost themselves in the colored intricacies of the Mandelbrot set by zooming ever deeper. Some readers who are determined to have their own color images but who lack color-display equipment have been ordering pictures from John H. Hubbard of Cornell University. I have been told by Heinz-Otto Peitgen, the Mandelbrot explorer whose images graced the August pages, that those images too are for sale; they are included with dozens of others in a color catalogue from Design-Büro-Weisser, Lothringer Strasse 23, D-2800 Bremen 1, West Germany. Hubbard and Peitgen together learned the art of image generation from Mandelbrot.

For readers whose equipment is limited to black and white, I should have thought of shades of gray. Such pictures can be nearly as inspiring as their colored counterparts. The best gray images were produced by David W. Brooks, who works with equipment at Prime Computer, Inc., in Framingham, Mass., to compute and plot his pictures. In his fabulous and delicate riots of halftones each shade of gray is rendered by tiny black squares of a certain size; the squares are made by a laser printer. Brooks has been searching for the tiny filaments that are thought to connect miniature Mandelbrots to the main set. So far they have not appeared at any magnification used by Brooks. Mandelbrot has advised him that they are probably infinitesimal.

Those with less sophisticated equipment can still work with shades of gray on a black-and-white monitor. John B. Halleck of Salt Lake City varies the density of points per pixel to indicate different shades.

Another approach depends on black and white contours. Yekta Gursel of Cambridge, Mass., has generated views of the Mandelbrot set that rival the ones Brooks generates. Gursel replaces a discrete spectrum of colors with alternating bands of black and white. Gary J. Shannon of Grants Pass, Ore., suggested the same technique and Victor Andersen of Santa Clara, Calif., took it to an extreme. He suggested changing from black to white (or the converse) whenever the *count* variable changes from one pixel to its neighbor.

Two other explorations are worth mentioning. James A. Thigpenn IV of Pearland, Tex., uses height instead of color. The Mandelbrot set becomes an immense plateau seen from an angle, with a complicated arrangement of spiky hills approaching the plateau in various places. Richard J. Palmaccio of Fort Lauderdale dispenses with the set altogether. His interest is in tracking individual complex numbers in the course of iteration. Their choreography near the boundary can result in spiral ballets or circular jigs.

The function $z^2 + c$ gives rise to the Mandelbrot set. Naturally other functions are possible, but they produce other sets. For example, Bruce Ikenaga of Case Western Reserve University has been exploring what appears to be a cubic cactus. The function $z^3 + (c - 1)\ z - c$ produces a prickly and uncomfortable-looking set (at least in stark black and white) surrounded by mysterious miniature spiral galaxies.

There are mysteries in iteration diagrams as well: when the integers modulo *n* are squared, each number migrates to another, in effect. The iteration diagram appears when each number is replaced by a point and each migration is replaced by an arrow. I raised several questions about such diagrams. How many components do they have? Readers sent diagrams documenting their explorations for various values of *n*.

The largest diagrams were completed by Rosalind B. Marimont of Silver Spring, Md. She examined the integers modulo 1000 and reported four pairs of components in the resulting iteration diagram. Each component sported a single attractor, as usual, and the largest attractors had 20 numbers. As a mathematician Marimont is allowed to conjecture that the integers modulo 10^k will produce $k + 1$ pairs of components and that the largest attractors will have $4 \times 5^{k-2}$ numbers.

Stephen Eberhart of Reseda, Calif., investigated the case where n is a Fermat prime (a prime number of the form $2^{2^k} + 1$). Here the number 0 forms an attractor by itself and the remaining numbers all lie in one single, grand tree. A number-theorist friend affirms that this will always be the case for Fermat primes and that the tree is binary: each internal point has two incoming arrows.

Iteration diagrams, like numbers, can be multiplied. If n is the product of two relatively prime numbers, say p and q, the iteration diagram for the integers modulo n is the product respectively of the diagrams for p and q. This interesting observation was made by Stephen C. Locke of Florida Atlantic University. Locke has also described a fascinating relation between the nth iteration diagram and a diagram of a seemingly different kind, one in which the numbers, instead of being squared, are merely doubled. When n is a prime, the latter diagram for the integers modulo $n - 1$ is the same as our nth iteration diagram, except for a single isolated number forming an attractor by itself. Much the same observation was made in number-theory terms by Noam Elkies of New York City.

A powerful tool for analyzing the (squared) iteration diagrams was developed by Frank Palmer of Chicago. Apparently all the trees attached to a given attractor are isomorphic. This means essentially that they have precisely the same form.

Finally, Bruce R. Gilson of Silver Spring, Md., and Molly W. Williams of Kalamazoo, Mich., examined a quite different generalization of the numbers from 0 through 99. These may be regarded as numbers to different bases. As numbers to the base 3, for example, one would count 00, 01, 02, 10, 11, 12, 20, 21, 22 before arriving once more at 00. Such numbers also produce iteration diagrams that look like those arising for integers modulo n. Gilson proved the diagrams always have paired components when n is even but not a multiple of 4.

There was an error in the iteration diagram presented in the August column for the integers from 0 through 99. Two arrows were missing from two components and the direction of one attractor was reversed.

Here are 147 reasons to buy at Elek-Tek, not to mention the fastest delivery anywhere.

Save 30% to 43%

off Manufacturer Suggested Retail prices on

EPSON • Okidata • Star • TOSHIBA • COMREX • OLYMPIA

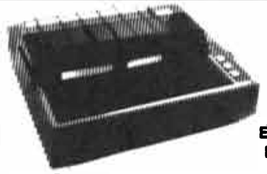
HEWLETT-PACKARD

73. **HP-71B Computer** .... $ 399
74. **HP-82400A** Series 70 card reader 125
75. **HP-82401A** HP-IL interface 95
76. **HP-82402A** 4K memory module 60
77. **HP-82700A** 8K memory module 150
78. **HP-82441A** Assembly/Forth 120
79. **HP-82480A** HP-71 math pac 75
80. **HP-82482A** HP-71 finance pac 60
81. **HP-82483A** Surveying Pac 120
82. **HP-82484A** Curve Fitting 75
83. **HP-82485A** Text Editor 60
84. **HP-82488A** Data Com. Pac 120

OMEGA

BERNOULLI BOX *(Modular)*

1. 10 meg ½ height Drive for IBM-PC/XT/AT & compatibles 1675
2. 20 meg ½ height Drive for IBM-PC/XT/AT & compatibles 2335
3. Non-Bootable Interface Card ... 104
4. Bootable Interface Card ... 234
5. 10 meg cartridges for above (3 pak special) 125



EPSON FX 85

EPSON®

15. LX 80 .... $ 215
16. LX 90 220
17. RX 100+ 300
18. FX 85 340
19. FX 185 475
20. LQ 1500 parallel .... CALL
21. DX10 Daisy Wheel 10CPS .. 230
22. DX20 Daisy Wheel 20CPS ... CALL

TOSHIBA

45. Toshiba P341 .... CALL
46. Toshiba P351 .... CALL
47. Toshiba P1340 430

OKIDATA

48. Oki 182 .... CALL
49. Oki 192/193 .... CALL
50. Oki 84 .... CALL

star

51. SG 10 .... $ 235
52. SD 10/15 .... CALL
53. SR 10/15 .... CALL

HP-ThinkJet

HP-2225 $375

85. HP-41CV .... $ 168
86. HP-41CX 245
87. Optical Wand 95
88. Card Reader 145
89. Printer 283
90. Quad R.A.M. (for HP41C) .. 60
91. Ext. Memory Module 60
92. Ext. Function Module 60
93. Time Module 60
94. HP-IL Loop Module 95
95. Digital Cassette Drive 400
96. Printer/Plotter (HP-IL) 335
97. **HP-9114A** Disk Drive 600

HARD DRIVES

Dysan CORPORATION

Internal Hard Disk Subsystems

6. **10 MEGABYTE** 500
7. **20 MEGABYTE** 600

Each system includes:

- Micro Science Drive
- Western Digital Controller
- 2 cables
- Documentation
- Dysan 1yr limited warranty
- For use in IBM PC, XT or most compatibles

PRINTER BUFFERS

8. **DDP-16** 64K Par./Par. 70

QUADRAM

9. **QRMP8** 8K Parallel/Parallel 125
10. **QRMPS8** 8K Parallel/Serial 139
11. **QRMSP8** 8K Serial/Parallel 139
12. **QRMSS8** 8K Serial/Serial 139

POWER PROTECTORS

13. **PowerMite PG600-S9** 6 outlet Surge Suppressor ... 30
14. **ISOBAR 8** 8 outlet Surge Suppressor ... 50

PRODUCTS FOR IBM-PC

23. **Amdek 310A** Amber Monitor .... $ 150
24. **Generic Multi** Multifunction Board, 64K 129
25. **Generic Multi 384K** Multifunction Board, 384K 169
26. **AST Six Pak +** Multifunction Board, 64K 225
27. **AST Six Pak + (loaded)** Multifunction Board, 384K 279
28. **AST Megaplus II** Multifunction Board, 64K 270
29. **Quadram Quadboard** Multifunction Board, 0K 195
30. **Quadram Quadboard** Multi. Board, 64K/384K .. 210/267
31. **Orchid Tech.** PC Turbo 186 570
32. **Paradise** Modular Graphics Card 290
33. **Hercules** Monochrome Card 329
34. **Hercules Color** Color Graphic Card 170
35. **Novation 490605-1** 2400BPS inc. Mite Software .. 620
36. **Novation 490603** ½ Card Modem 2400BPS No software 425
37. **Novation 490603-1** As above inc. MS-DOS Software . 490
38. **Hayes 1200B** Internal modem w/software ... 359
39. **AT&T 4000** 300/1200 Ext. Modem 335
40. **Hayes 1200** External modem 399
41. **Hayes 2400** External modem 599
42. **US Robotics Courier 2400** Ext. 2400B Smart Modem ... 460
43. **US Robotics Telpac** Telecomm. Software 75
44. **TEAC FD55B** ½ ht. DSDD Disk Drive 90

IBM is a trademark of International Business Machines Corp.

PRODUCTS FOR APPLE

54. **Toshiba HND 04DT** ½ Height DSDD Disk Drive . $ 70
55. **U—Print IIc-16K** Par Interface external 85
56. **Grappler PLUS** Graphics Interface 80
57. **Buffered Grappler** 16K Buffer (Expand to 64K) .. 150
58. **Kensington System Saver** Fan and Surge Suppr. 69
59. **Super Cooling Fan** Fan and Surge Suppr. 29
60. **Gamma** Disk Controller Card 48
61. **Gamma II** SS Disk Drive for II, IIe, IIplus ... 125
62. **Gamma IIc** SS Disk Drive for IIc 125
63. **Rana ELITE 2** DSSD Disk Drive 335
64. **Rana ELITE 3** DSDD Disk Drive 420
65. **ERAM 80** 80 col; 64K RAM for IIe ... 115
66. **Novation Apple Cat II** 300 Baud Modem 200
67. **Novation 212** 300/1200 Baud Modem .. 390
68. **Nov 49059112** Mod. & Soft. for MacIntosh . CALL
69. **Hayes Micromodem IIe** 300 Baud Modem 145
70. **Hayes Smartmodem IIc** 300 Baud Modem for IIc .. 240
71. **Wico 501030** Analog Joystick 36
72. **Amdek Monitors** ..... CALL
73. **Zenith Monitors** ...... CALL

SLIMLINE Shirtpocket Styled Power Packed Programmable LCD PROBLEM SOLVERS

98. **HP-11C** Scientific 58
99. **HP-15C** Scientific 90
100. **HP-12C** Financial 90
101. **HP-16C** Programmer 90

hp HEWLETT PACKARD — PERSONAL COMPUTATION DEALER

HUGE SAVINGS ON DISKETTES

	Dysan	maxell	3M	SONY	wabash	MEMOREX	BASF
3½″ SSDD	23.00	25.00	25.00	25.00	20.00	—	—
DSDD	29.00	37.00	33.00	37.00	24.00	—	—
5¼″ SSDD	17.00	14.00	14.00	—	11.50	11.50	11.00
DSDD	21.00	18.00	18.00	—	12.50	14.00	12.00
SSDD96TPI	24.00	24.00	24.00	—	—	—	—
DSDD96TPI	30.00	29.00	29.00	—	—	—	—
5¼″ DSDDHD (For IBM AT)	38.00	39.00	35.00	—	24.00	—	24.00
8″ SSDD**	22.00	29.00	25.00	—	19.00	—	—
8″ DSDD**	26.00	32.00	29.00	—	20.00	—	—

**Unformatted

Call for Quantity pricing for 10 boxes or more.

3M DATA CARTRIDGES

142. DC100 .... $ 14.00
143. DC300A 18.00
144. DC300XL 21.00
145. DC300XL/P 22.00
146. DC600A 23.50
147. DC1000 .... 15.00

Call for Quantity pricing for 10 cartridges or more

CALL TOLL FREE 800-621-1269 EXCEPT Illinois, Alaska

ELEK-TEK, inc. 6557 N. Lincoln Ave., Chicago, IL 60645 (312) 631-7800 (312) 677-7660

Corp. Accts. Invited. Min. Ord. $15.00. Visa or MasterCard by Mail or Phone. Mail Cashier's Check, Mon. Ord., Personal Check (2 wks. to clear) Add $4.00 1st item. (AK, HI, P.R., Canada add $10.00 first item) $1.00 ea. add'l shpg. & handl. Shipments to IL address add 7% tax. Prices subj. to change. WRITE for free catalog. RETURN POLICY: Defectives Only: Most products replaced within 30 days of purchase with identical merchandise only. Computer and large peripherals replaced only when defective on arrival (within 3 work days of delivery). Other problems covered by mfr. warranty. ALL ELEK-TEK MERCHANDISE IS BRAND NEW, FIRST QUALITY AND COMPLETE. Delivery subject to availability. DUNS #09-718-0517

BOOKS

A glowing coma of books presages the apparition of Halley's faithful comet

by Philip Morrison

HALLEY'S COMET: A BIBLIOGRAPHY, compiled by Ruth S. Freitag. Library of Congress, from Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 ($26). FIRE AND ICE: A HISTORY OF COMETS IN ART, by Roberta J. M. Olson. Published for the National Air and Space Museum by Walker and Company, N.Y. ($24.95). MANKIND'S COMET, by Guy Ottewell and Fred Schaff. Astronomical Workshop, Furman University, Greenville, S.C. 29613 (paperbound, $20). COMET, by Carl Sagan and Ann Druyan. Random House ($24.95). THE COMET BOOK: A GUIDE FOR THE RETURN OF HALLEY'S COMET, by Robert D. Chapman and John C. Brandt. Jones and Bartlett Publishers, Inc. (paperbound, $14.95). HALLEY'S COMET, by Francis Reddy. AstroMedia Corp. Distributed by Tide-Mark Press, P.O. Box 813, Hartford, Conn. 06142 (spiralbound, $9.95). Welcome to Comet Halley. Just about now it has passed northward through the plane of the earth's orbit to spend half a year nearby after a lifetime spent far on the southern side. It is this constancy we mark, the Newtonian fidelity of motion, not the fickle appearance of that night plume, mostly dust motes dancing in the distance, at best an uncertain, faint and much touted spectacle.

Suppose a heavy late-winter fall of snow were to blanket the entire East from Cincinnati to Boston. Were it magically to free itself from the pull of the earth before packing too solidly, such a blanket might be rolled very lightly into one gigantic, fluffy snowball four miles in diameter. (Its own minute self-gravity presses no more heavily than the weight of a down comforter.) Assume that it had been prepared for launching at a certain spot in the cold dim distances of planetary space, far southward of the flattened region the planets travel. Let the snowball be thrown not quite straight toward the sun, at a determinate if rather leisurely speed, so that in 1948 it begins a return voyage. The ball is spinning slowly on its axis a couple of rotations per day. It progresses counter to the direction of revolution that all the planets share. By this Christmas your snowball, falling faster and faster, will have only slightly overshot the focal sun, to arch back past it again under the strong attraction of that ringmaster, which was once so far away and is now nearby. The earth orbits in that inner neighborhood; we shall pass close to the moving snowball twice, first in early December, 1985, on its way in, and again in mid-April, 1986, on its way back out. In April we shall approach the comet to a distance of about 30 million miles, about a third of our nearly constant radius from the sun. (In 1910 earth and comet passed each other within half that distance, and the earth traveled through the end of the gaseous tail.)

That fanciful orbiting snowball is a fairly realistic model of Halley's comet (in the jargon, Comet P/Halley, the *P* standing for periodic). Periodic it is of old, requiring about a human lifetime for each full orbital passage; our calculations can date with precision 45 rhythmic returns to our neighborhood, their beat slightly syncopated under the pull of Saturn or Jupiter as those planets occasionally pass near the comet's path. The earliest visit for which there is probable witness was in 240 B.C., recorded in the youth of the book-burning tyrant of Ch'in, the ruler with that famous funeral escort of terra-cotta warriors. The annals of the time record a "broom star" during the years before he unified much of China to establish the imperial pattern that endured until 1910, a year that saw another return of P/Halley.

No hair, no comet. The very word comes from the Greek for hair. We speak of the tail of the comet, but the locution is a poor one. A comet's hair streams always outward from the sun, whether the comet is falling in or moving away. Two winds fly out from the sun, photons and protons, and two kinds of tail fly with those winds. The orbital motion has little effect. The irregular surface of dirtied snow sublimes to vapor and dust whenever that snowball, the cometary nucleus, ventures close enough to the sun to feel the heat. Then a gassy halo grows, from which a streamer of plasma emerges to glow blue in the proton wind, knotted and kinked by magnetic fields borne out from the sun. An accompanying diffuse stream of dust—no such celestial snowball can be really pure and clean—may feather out gracefully too under the pressure of radiation, a long plume of dust motes lit by golden sunlight. A comet wastes its substance to provide the glory of the tails. But the matter of the showy tails is extremely dilute: the snow P/Halley will lose during this apparition amounts to only as much of its model snow blanket as might cover Long Island.

There is a third significant emanation from every bright comet. We might call it the comet's tale. Human residents on the earth eye these celestial tourists by no means quietly. Comets inhabit myth worldwide; books on comets abound in every printed tongue; images multiply in all the mediums of art. The comet is for human culture less a real outsize snowball than it is form, portent and metaphor. Among the tidal bulge of three dozen books that have risen this year out of American publishing to greet P/Halley (the standard of quality held so high that most of them are competent volumes) we have chosen to describe a diverse and appealing subset.

Two works bring meticulous scholarship to their subject. The first of these is a thick annotated bibliography that lists 3,235 references to this comet, the latest of which run through the middle of 1984. They have been culled by Ruth S. Freitag from those five or 10 gigapages of print held in the Library of Congress. The learned compiler makes no claim that her listing is comprehensive, yet few if any workers can have cast a wider net. English-language texts dominate an assemblage in which the languages of all Europe are also plentiful, and tongues farther eastward are not ignored. The profound and the absurd both occur; daily cartoons jostle thick tomes of dry calculations. Surely Puck would have found enough here to support his estimate of us mortals.

Consider only a few citations. In 1682 young Edmund Halley made notes and calculations about the September comet. His notebook page was published by Arthur Eddington in 1910 (it is reproduced in the Freitag book). Two decades later Halley came to realize that the 1682 comet shared an unusual orbit with bright comets that had been seen earlier, in 1607

and in 1531. The comet, one comet now, was his in a new way, and it justly still bears his name.

Meanwhile at Marburg in Hesse in 1682 a hen laid a marvelous egg marked with a starry design; it was duly recorded by one Johann G. Brand (*Hessisches Cometen-Ey*..., with engravings). Marvelous eggs turn out to be just as periodic as comets. On May 17, 1910, P/Halley present again in the sky and on every tongue, Mme Haydée Bouyard reported that one of her hens had produced an egg "with a distinct image of a comet on the shell." Watch for the comet egg of 1985–86! The phenomenon is not at all a trivial correlation; it is causal. At all times there are eggs that bear marks; we can accept the 1910 report from France. A marked egg is of interest only in comet season, preadapted to fit the news, in a paradigm not very different from the working of the immune system or even of organic evolution itself.

One can with some effort follow the rise of the special fear of mass asphyxiation that marked the 1910 return. The zealous popularizer Camille Flammarion forecast that the earth would pass through the comet's tail in May. The spectroscopists had by January already identified the bands of cyanogen (CN) in the tail. The reporters did the rest. They knew all about cyanide. Public reassurances were many, but further alarms, even from prophetic astronomers, were not few. The consequences were pills and panic—real, anticipated and feigned. In Cardiff they noticed strong odors "like old boots burning" during the very night of tail passage: the comet's gases. Or perhaps only old boots? In China the end of the ancien régime was close at hand. There comets had long announced the mandate of heaven; collapse and not cyanogen was the issue.

A comet accentuates the horror of William Blake's **The Ghost of a Flea**

Roberta Olson is the art historian at Wheaton College who recognized in 1978 that it was Halley's comet Giotto portrayed in a famous painting [see "Giotto's Portrait of Halley's Comet," by Roberta J. M. Olson; SCIENTIFIC AMERICAN, May, 1979]. In a superb fresco at the Scrovegni (Arena) Chapel in Padua, a cometary star of Bethelehem glows brightly above the Wise Men who have come to adore the Virgin and Child. The comet's apparition in 1301 had been awesome and much remarked in northern Italy. The artist painted the fresco between 1303 and 1306, the intricately executed image an "advanced artistic statement" without precedent in his or any other's work. "Comets are nature's icons," Olson says at the beginning of her striking chronicle of comets in Western art.

The book comes at the time of the exhibition Olson has arranged for the National Air and Space Museum during the months P/Halley is near. The text is helpful in the interpretation of the 120 images she has collected, from Roman coins and seals up to the present day. Artists are at work for this apparition too. They will have to achieve extraordinary success to match two images here that trapped a reviewer's eye. One was William Blake's painted panel of *The Ghost of a Flea,* the powerful human figure with scaly skin and red "eagre tongue" standing for all such men "as were by nature bloodthirsty to excess." The baleful comet streams huge across the night sky behind that wicked soul. The image speaks loud to our grim times. It is a relief to then come on a Miró autumn landscape, golden comet in a playful orbital dance across an orange sky, above a crimson-headed hare that resembles a toy dragon.

Three books, all well written and well produced, offer up-to-date summaries of the science of comets and of Halley's comet. They differ in style and scope in interesting ways. Guy Ottewell and Fred Schaff are two gifted and assiduous amateurs of astronomy. Their volume, handsomely composed, all in black and white, centers squarely on P/Halley. Intended for the cultivated nonscientific reader, it has a fascinating section giving an account of each of the comet's apparitions.

The "grand sequence" ranges from the Shang dynasty through Octavius, Pliny, Attila, Norman William and English Harald, past Giotto, Apian, Kepler, Halley, down to 1910, when the comet served as material for Mark Twain and H. G. Wells. The witty chronicle gradually turns from the tale of emperors and their follies to the


In the field of data communications, we get the bugs out.
Bugs. Bugs. And more bugs.
All too often, that's what you get with high-tech data products and services.
But nobody wants to be pestered by problems with computer networks that are supposed to solve problems. And keep data moving. We know. We had to solve the same problems ourselves.
In fact, we had to build our own computer systems because what we needed didn't exist.
Today, we're experts at integrating data systems into a business.
Because, like you, we needed computers to process purchase orders, and handle billing. Computer systems we designed even analyze and correct traffic problems. Often before they happen.
Today we have one of the largest collections of mainframe computers in the world. As well as over 875 mini-computers and 1,025 micro-computers. And they all work together to help us serve over 10,000,000 customers daily.
Which is why, when you come to us for data communications systems and services, you know we can make them do what you want them to do.
So to avoid bugs in today's high-tech jungle of communications, call us.
NYNEX is the parent company of New York Telephone and New England Telephone, plus other companies that offer business communications equipment, cellular services and directory publishing.
For more information, call 1 800 441-1919.
In Utah, call 1 800 453-9000.
It pays to know us.

growth of modern astronomy, a strobe light on human history. Then the authors render the orbit of each of 48 passages (a couple are still to come) both in a geocentric perspective (that is, against the stars) and in a three-dimensional Copernican perspective.

This leads to a typology of encounters, no simple matter but artfully spun out of examples. During this encounter our view is not an excellent one, although it is not wholly mediocre. The brightest meeting we know was that of April, A.D. 837. The comet then passed only three or four million miles away from the earth. The recitation is a work of geometric passion, beautifully executed; the meaning of the earth's axial precession and perturbations in comet motion are made real by their long-term presence. Another page neatly examines the proper pronunciation of the name Halley. The cautious verdict is: Unsure, but on balance likely to rhyme with valley. There is much more, a store of devotion and wit in the amateur's finest style.

The volume by Carl Sagan and Ann Druyan is the joint work of a veteran professional planetary scientist and a writer of talent. One must begin by observing that Professor Sagan has become deservedly so widely known in American life that his publishers can confidently invest in a volume of unusual largesse. There are more than 300 illustrations, the majority in color; a score or more of these are vivid paintings of astronomical scenes never perceived but arguably and carefully imaged by knowing artists such as Jon Lomberg, Don Davis and Anne Norcia. A comet plunges into the fiery solar surface, another impacts the ancient ocean, a shower of comets confronts a dinosaur: these are striking constructions, although for some readers they are tinged with the uneasiness of undeclared approximations.

The text shares that imaginative quality. Ten chapters treat the nature of comets, recounting the history of our knowledge; there is a splendid chapter on Halley himself, and an unusual bestiary of bizarre cometary forms and visits. Seven more chapters deal with the origin and fate of comets, including the issue of their relation to the widespread extinctions that punctuate the fossil record.

The last few chapters take up the future of comets. They begin with the future near at hand, describing the satellites that will probe right into the comet. Twin Soviet spacecraft have caromed off Venus toward the comet. Among their instruments is a clever dust analyzer designed and built in the U.S. The analyzer will fly to mutual benefit but not because the governments so agreed; such cooperation has lapsed by American intent. It happens that Roald Sagdeev of Moscow and John Simpson of Chicago are friends and colleagues close enough to get around such obstacles. Simpson, under official eyes, anxious to give no scandal by impermissible "technology transfer," did not embody the state of the art in his design. The Soviet engineers were puzzled that the American instrument was the only one their probe carried that lacked any microprocessor. All the Soviet instruments used them. "Simpson smiled."

There is also a shadowy prophecy about comets. Perhaps we shall mine them for their water content, or board them to establish space colonies. These ably written extrapolations draw in even a skeptical reader. The focus in Sagan and Druyan is not on Comet Halley; it is on generalization, on beginnings and endings, on comparisons and metaphor that can lend content to numbers otherwise not significant to readers with little training. A special quality is added by their wide-ranging view of culture; the quoted poems and sayings often come from sources much fresher than the established classics of European worthies. We are given a wider look at the human record, and it is a pleasure.

Theories are nicely handled, particularly in the realm of the solar system. A number of excursions into the possible origin of myth and symbol—Why are comets so fearful?—are interesting even when they are not immediately compelling. It is not easy to accept the origin of the swastika found in so many cultures as the relic of a multijet comet spinning nearby a few millenniums back. Anyone who likes to make marks might well invent the cross after a dozen marks, and the swastika after a dozen crosses. Independent invention is easy to believe for so symmetrical a symbol—one, moreover, carrying two opposing meanings and a dual form.

Altogether this is a book as full of new ideas as it is packed with worthwhile images. It gives a first-rate summary of the received material in addition, particularly the newly seen possibilities of comets: objects that may have repeatedly wreaked destruction on life on the earth, just as they may have given that life in the first place.

Robert D. Chapman and John C. Brandt are two professional cometary scientists at NASA's Goddard Space Flight Center. They have written plainly, in a style more narrative than textbooklike, a compact but quite full book on comets. One section goes into some detail on P/Halley's orbit shown from a geocentric standpoint and includes sky maps for finding our visitor. Diagrams and spectra abound; there is a brief section of color plates that reproduce good comet photographs, as well as the Bayeux tapestry and the Giotto fresco. A listing of the identified atomic and ionic species in comets along with formulas is about as technical as the material becomes. The book is a bargain both in dollars and in the quickness with which it satisfies curiosity. Treatment of the human background is brief; the lives of Newton and Halley are covered in two or three pages each, rather than in a chapter. The space probes now on the way are adroitly—if succinctly—handled.

Francis Reddy's attractive spiral-bound book has large, colorful pages. The brief and easy text is divided about equally between the nature and the history of Comet Halley. The book offers pictures of the Bayeux tapestry, Giotto's fresco and a 3D cutout model of the orbit. There are also some 20 celestial maps and other aids to finding our visitor, drawn for skies north and south, winter and spring. These are about the most attractive and the easiest for beginners to use among many published charts. Here the aim is to make Halley's comet accessible for the newcomer to sky-watching. There is one major task for those who want to see the object: first find your way to a dark sky, beyond any lighted city. Unless from your vantage you can clearly see the Milky Way, travel. Otherwise television will have to suffice.

We close with a few words about those who have studied comets. Two seniors among contemporary astronomy come first. One is Jan Oort of Leyden. He worked out the dynamical arguments, persuasive if unverified by observation, showing that comets are stored by the trillion, slowly circling in a vast, invisible circumsolar sphere so far away that the sun they obey appears to be only the brightest among a number of similar stars. Gravitational nudges from passing stars continually disturb a few comets so that they fall inward, to become trapped and soon waste away in our radiant neighborhood. Fred Whipple of Harvard has convinced us that the comets are physically the dust-laden snowballs discussed above, mountain-size fragments in planetary space, spinning, outgassing, even splitting as they fly by. (His own new book on comets will appear in time for the Halley visit.)

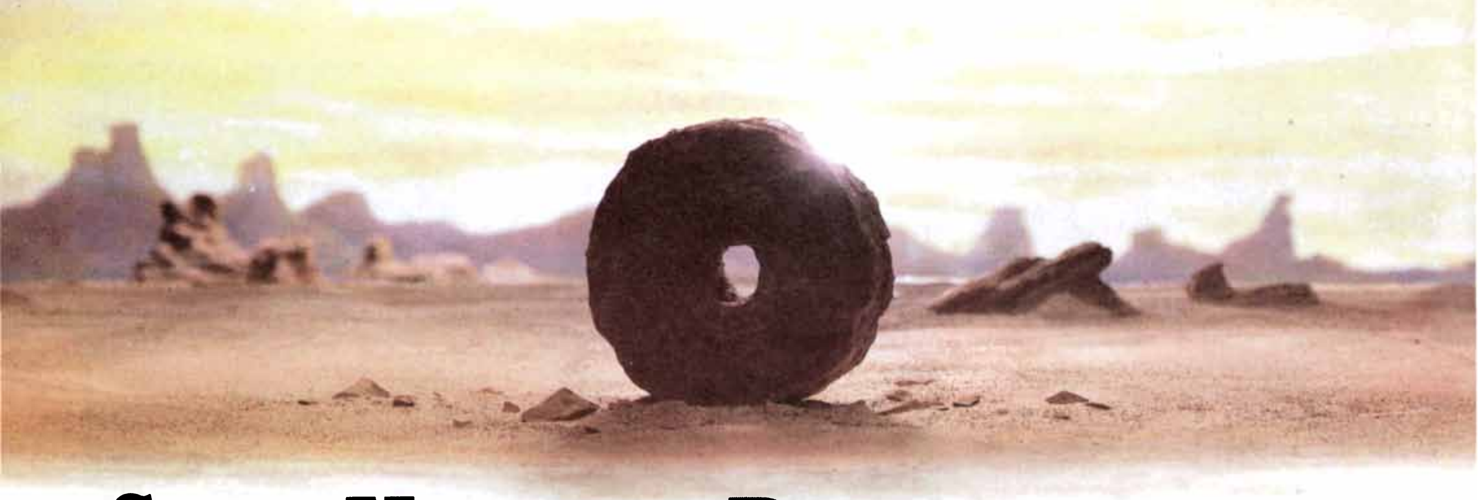
Both our modern cometeers share with Edmund Halley, in whose name P/Halley now makes its round, the admiration and friendship of their astronomical generation. Halley the man orbits forever in the shadow of the unmatched Newton, but he was a gifted, original, versatile and productive scientist, and a human being as adventuresome, generous, loving and sweet as Newton was retiring, cold, solitary and austere.

It is a fine year to think of Halley; it is in his spirit to hope that by 2061 when his comet next returns to perihelion we shall have detected the Oort cloud and sampled and analyzed many a snowy comet confected according to Whipple's recipe. Perhaps we humans shall even have learned how to survive as knowing watchers for a good many more passages of Comet P/Halley.

FIRE, by John W. Lyons. Scientific American Books, Inc. (available to members of the Scientific American Library, $24.95). Our fascination with flame is manifested on the jacket of this original work for the general reader, which reproduces the portrait by Georges de la Tour of a notorious young woman lost in reflection and regret before a tranquilly burning lamp. Such a simple flame, from oil lamp or candle, has long stood as metaphor for life itself. It is evidently a steady-state phenomenon, responsive to small disturbances but resistant to them, not something inert and stone-hard. Patently it feeds on its fuel, to release heat and wastes. In a limited sense it is even self-reproducing; a candle flame will kindle another just like itself if it makes contact with the right substrate, an unlit candle.

It was Michael Faraday who first used the familiar candle flame as vehicle for a classical essay introducing the methods and results of science. Dr. John W. Lyons knows a good thing when it is glowing; his first full chapter follows Faraday's wonderful lead, in up-to-date terms. We see the interplay of the two flows, heat and mass, learn something of the chemistry of wax combustion (by no means a closed chapter in science) and form some understanding of the physiology of this proto-organism.

The source of the flame is that tiny pool of molten wax. Thence come the vapors diffusing upward to decompose, first to simpler reactive radicals that emit the faint blue spectral light, then down to the carbon atoms that condense into particles of black soot. The opaque soot is hot enough to glow bright yellow with thermal radiation. The tiny carbon particles oxidize in turn in the fresh air, plentiful at the upper end of the flame, to enter the final gaseous products of combustion, water vapor and carbon dioxide convected upward beyond the luminous region. Of course, the air is a necessary if dilute combustion partner. Eighty percent of it is the inert gas nitrogen, merely a sink for heat. A candle light-



SOME HISTORIC BREAKTHROUGHS DON'T TAKE AS MUCH EXPLAINING AS COMPUSERVE.

But then, some historic breakthroughs could only take you from the cave to the tar pits and back again.

CompuServe, on the other hand, makes a considerably more civilized contribution to your life.

It turns that marvel of the 20th century, the personal computer, into something useful.

Unlike most personal computer products you read about, CompuServe is an information service. It isn't software. It isn't hardware. And you don't even have to know a thing about programming to use it. You subscribe to CompuServe—and 24 hours a day, 7 days a week, it puts a universe of information, entertainment and communications right at your fingertips.

A few of the hundreds of things you can do with CompuServe.

COMMUNICATE

EasyPlex™ Electronic Mail lets even beginners compose, edit, send and file messages the first time they get online. It puts friends, relatives and business associates—anywhere in the country —in constant, convenient touch.

CB Simulator features 72 channels for "talking" with thousands of other enthusiastic subscribers throughout the country and Canada. The chatter is frequently hilarious, the "handles" unforgettable, and the friendships hard and fast.

More than 100 Forums welcome your participation in "discussions" on all sorts of topics. There are Forums for computer owners, gourmet cooks, veterinarians, pilots, golfers, musicians, you name it! Also, Electronic Conferencing lets businesses put heads together without anyone having to leave the shop.

Bulletin Boards let you "post" messages where thousands will see them. You can use our National Bulletin Board or the specialized Bulletin Boards found in just about every Forum.

HAVE FUN

Our full range of games includes "You Guessed It!", the first online TV-style game show you play for real prizes; and MegaWars III, offering the ultimate in interactive excitement. And there are board, parlor, sports and educational games to play alone or against other subscribers throughout the country.

Movie Reviews keep that big night at the movies from being a five star mistake.

SHOP

THE ELECTRONIC MALL™ gives you convenient, 24-hour-a-day, 7-day-a-week shopping for name brand goods and services at discount prices from nationally known stores and businesses.

SAVE ON TRIPS

Travelshopper℠ lets you scan flight availabilities (on virtually any airline — worldwide), find airfare bargains and order tickets right on your computer.

Worldwide Exchange sets you up with the perfect yacht, condo, villa, or whatever it takes to make your next vacation *a vacation*.

A to Z Travel/News Service provides the latest travel news plus complete information on over 20,000 hotels worldwide.

MAKE PHI BETA KAPPA

Grolier's Academic American Encyclopedia's Electronic Edition delivers a complete set of encyclopedias right to your living room just in time for today's homework. It's continuously updated... and doesn't take an inch of extra shelf space.

The College Board, operated by the College Entrance Examination Board, gives tips on preparing for the SAT, choosing a college and getting financial aid.

KEEP HEALTHY

Healthnet will never replace a real, live doctor—but it is an excellent and readily available source of health and medical information for the public.

Human Sexuality gives the civilization that put a man on the moon an intelligent alternative to the daily "Advice to the Lovelorn" columns. Hundreds turn to it for real answers.

BE INFORMED

All the latest news is at your fingertips. Sources include the AP news wire (covering all 50 states plus national news), the Washington Post, USA TODAY Update, specialized business and trade publications and more. You can find out instantly what Congress did yesterday; who finally won the game; and what's happening back in Oskaloosa with the touch of a button. And our electronic clipping service lets you tell us what to watch for. We'll electronically find, clip and file news for you...to read whenever you'd like.

INVEST WISELY

Comprehensive investment help just might tell you more about the stock you're looking at than the company's Chairman of the Board knows. (Don't know who he is? Chances are, we can fill you in on that, too.) CompuServe gives you complete statistics on over 10,000 NYSE, AMEX and OTC securities. Historic trading statistics on over 50,000 stocks, bonds, funds, issues and options. Five years of daily commodity quotes. Standard & Poor's. Value Line. And more than a dozen other investment tools.

Site II facilitates business decisions by providing you with demographic and sales potential information by state, county and zip code for the entire country.

National and Canadian business wires provide continuously updated news and press releases on hundreds of companies worldwide.

GET SPECIALIZED INFORMATION

Pilots get personalized flight plans, weather briefings, weather and radar maps, newsletters, etc.

Entrepreneurs use CompuServe too for complete step-by-step guidelines on how to incorporate the IBMs of tomorrow.

Lawyers, doctors, engineers, military veterans and businessmen of all types use similar specialized CompuServe resources pertinent to their unique needs.

And now for the pleasant surprise.

Although CompuServe makes the most of any computer, it's a remarkable value. With CompuServe, you get low start-up costs, low usage charges and local phone-call access in most major metropolitan areas.

Here's exactly how to use CompuServe.

First, relax.

There are no advanced computer skills required.

In fact, if you know how to buy breakfast, you already have the know-how you'll need to access any subject in our system. That's because it's "menu-driven," so beginners can simply read the menus (lists of options) that appear on their screens and then type in their selections.

Experts can skip the menus and just type in "GO" followed by the abbreviation for whatever topic they're after.

In case you ever get lost or confused, just type in "H" for help, and we'll immediately cut in with instructions that should save the day.

Besides, you can either ask questions online through our Feedback service or phone our Customer Service Department.

How to subscribe.

To access CompuServe, you'll need a CompuServe Subscription Kit, a computer, a modem to connect your computer to your phone, and in some cases, easy-to-use communications software. (Check the information that comes with your modem.)

With your Subscription Kit, you'll receive:

- a $25 usage credit.
- a complete hardcover Users Guide.
- your own exclusive user ID number and preliminary password.
- a subscription to CompuServe's monthly magazine, *Online Today.*

Call **800-848-8199** (in Ohio, 614-457-0802) to order your Subscription Kit or to receive more information. Or mail this coupon.

Kits are also available in computer stores, electronic equipment outlets and household catalogs. You can also subscribe with materials you'll find packed right in with many computers and modems sold today.

☐ Please send me additional information.
☐ Please send me a CompuServe Subscription Kit.
☐ I am enclosing my check for $39.95, plus $2.50 handling. *(Add sales tax if delivered in Ohio.)*
Please make check payable to CompuServe Information Services, Inc.
☐ Charge this to my VISA/MasterCard
#__________
Expiration Date __________
Signature __________
Name __________
Address __________
City __________
State __________ Zip __________

MAIL TO:

CompuServe®

Customer Service Ordering Dept.
P.O. Box L-477
Columbus, Ohio 43260 PRI—3311

An H&R Block Company

EasyPlex and ELECTRONIC MALL are trademarks of CompuServe, Incorporated. Travelshopper is a service mark of TWA.

ed in pure oxygen is a lively piece of fireworks.

The capillary wick is essential to candle flame. It draws fluid wax upward into the midst of the radiant flux of blue light that surrounds the wick. There the absorbed energy can both vaporize that fluid and initiate flame reactions. The hotter, brighter yellow flame region stands too high above the candle to sustain the flame on its own; only a few percent of its isotropic energy flux can strike and hence warm the candle top. Without a wick a small disk of candle wax or an oil lamp will not sustain combustion: not enough energy is fed back from the radiant flame.

Scale matters; a large enough pool of oil will indeed support sustained flaming, since the region of bright flame a few centimeters above the surface sees down to the wide sustaining surface over nearly half of the directions in space. Now the feedback is strong enough; a sufficiently broad pool of oil will burn freely without a wick. Candle design, however, is still no science but a subtle empirical art.

The next chapters take a similar look, first at wood fires on the hearth and in stoves, then at industrial furnaces, at the cylinders of automobile engines, at the big gas turbines and finally at the six-million-kilowatt flame of a Saturn C-5 rocket. Wood burns by induced thermal decomposition to combustible gas; the burning surface is both self-heated and heated by neighboring surfaces. The balance between rates of air supply and net heat loss by radiation and to the flowing air is critical. An open fireplace is only about 6 percent efficient; 10 times as much warmed room air as is needed for the fire floods up the chimney as waste. If the ambient air that enters the room is cold enough, the crackling fire can become overall not a source of but a sink for room heat.

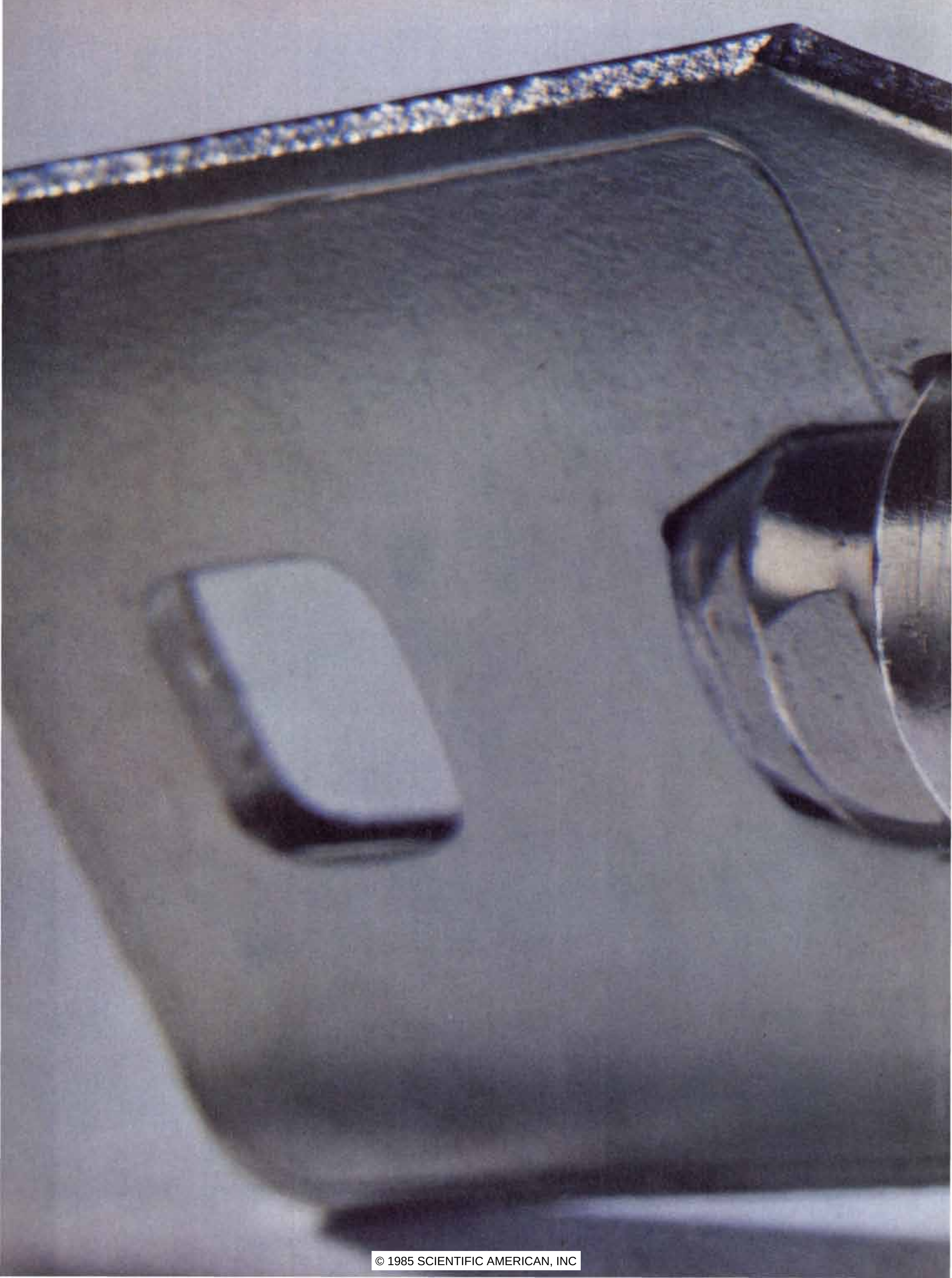
The latter half of the book centers on Dr. Lyons' professional work. He has for a long time been responsible for large-scale engineering research on the prevention and control of unwanted fires. Obviously the dauntingly complicated geometry and fuel mix of a burning house or hotel are taxing to a field of engineering not mature enough to design even a candle from scratch. Expensive and intricate experiments at full scale, model studies done at quarter scale and quite elaborate computer simulation have nonetheless enhanced understanding of the process.

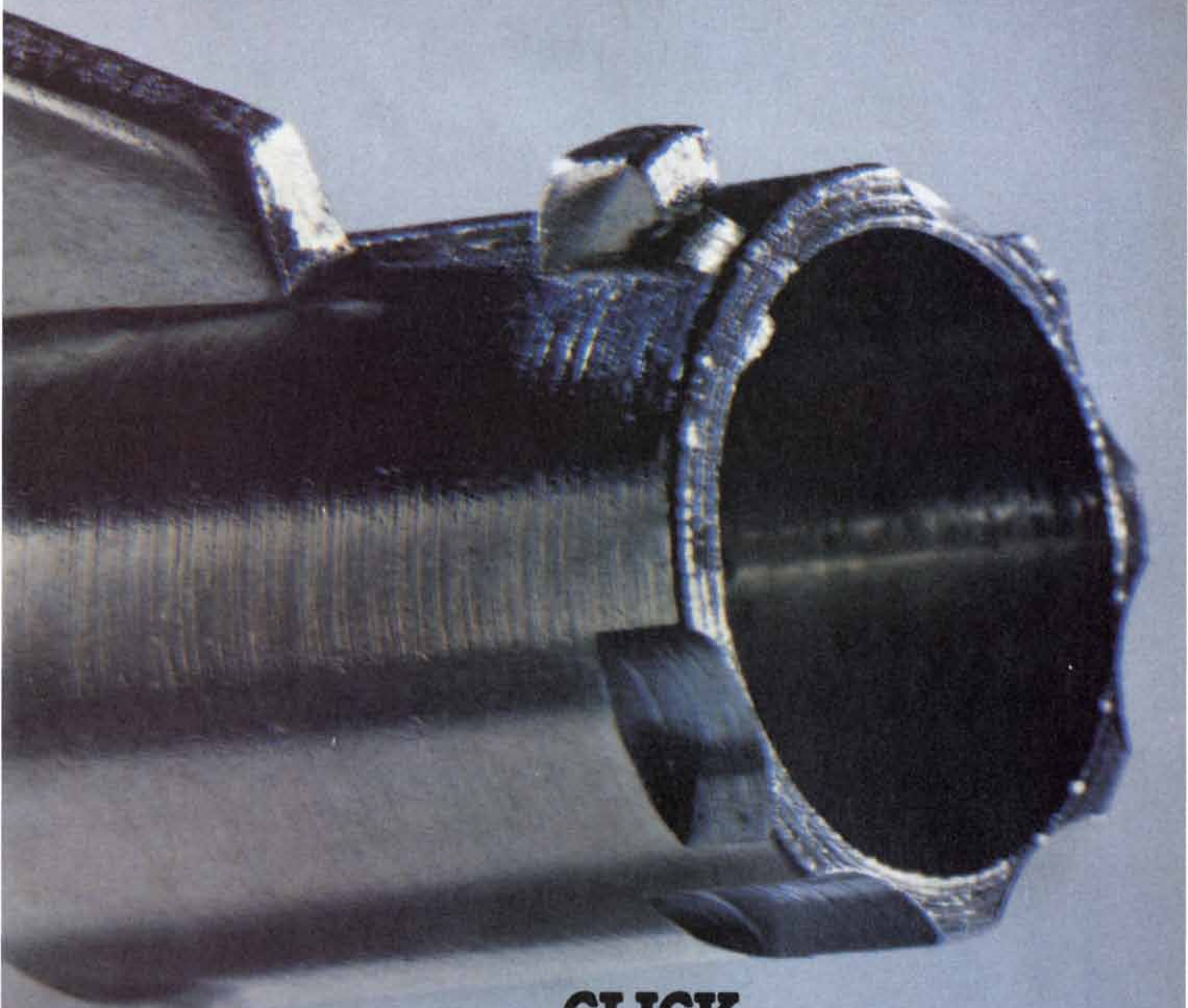
Like other growth phenomena the room fire surprises by its nonlinearity. It grows slowly at first; escape seems feasible. Then it rapidly transforms to full combustion at all the fuel surfaces within the room, including furniture and wood walls and floors. A frightening sequence in color shows this flashover in an instrumented experimental room. The flames first rose two minutes after a burning cigarette fell into the upholstery, so to say. Two minutes later a thick layer of hot gas occupied the space at the ceiling. At six minutes that hot layer became deep enough to pour out at the top of the open door and enter the corridor. A minute and five seconds later every surface in the room burst into flames. There can be no more growth in burning area; the process proceeds rather steadily but at high rates, controlled by the mechanisms of burning at depth. All at once escape becomes doubtful.

Diverse materials and their forms interact. Flow patterns are beyond simple rules. Yet computer models can now predict with an accuracy of about 15 percent the time to flashover for a much studied type of room. The software to handle real cases is at hand; often the simulation takes a few thousand iterations, with half a million equations solved in each example, coupled volume element by volume element. Supercomputers do the work efficiently. One sequence of photographs shows an impressive graphic simulation of the spread of a room fire (although it is not captioned explicitly enough for easy understanding). It is time to put this gain into practice as part of computerized urban fire codes of a new, more interactive, more quantitative and more expert kind.

Volatile gases emanating from incomplete room combustion can spread rapidly through big fireproof buildings. The gases flow through corridors along partially burning ceilings. Survivors often report that suddenly "the fire was everywhere." Most likely it spread not as flame but as hot unburned gases flowing from enclosures where the oxygen had been depleted, to ignite suddenly as soon as it entered the fresh air of new open spaces. One is reminded of the urban fire storm that can result from simultaneous fires ignited over a large section of a city by a host of small incendiary bombs, as in Dresden and Tokyo, or by the fearful radiant fireball of a nuclear weapon.

Today more than half of the victims of fires die not from flame but from inhalation of the toxic gases themselves, mostly carbon monoxide, but laced with hydrogen cyanide, hydrogen chloride and irritant aldehydes. Ducted ventilation schemes for pressurizing the fire stairs and exhausting air from the floor that is burning, all under computer control, might be a valuable strategy at least for large buildings of many stories. The Federal Building in Seattle is a pioneer in this





line. The good idea of reducing the sources of toxic gases by controlling the nature and amount of the ubiquitous polymer foams and films offers limited hope; wood itself turns out to be bad enough, although we have only sparse data and not much from direct bioassay. Big old wood houses that burn kill many poor children in inner cities. Mobile homes newly built under Federal regulations are safer.

We are beginning to acquire some ability to control what was once left to fate. What we do not yet control is the structure of society, the subtle system that allocates risk and resources by its action and inaction alike. The engineer-author, an energetic analyst and strategist, concludes somberly on rather a note of resignation: "Perhaps there will be a resurrection of public interest, in another decade or after a particularly catastrophic fire." Automobile accidents, after all, are worse, and we know that war is potentially far more threatening.

THE NATURAL HISTORY OF THE PRIMATES, by J. R. Napier and P. H. Napier. The MIT Press ($19.95). The British primatologists John and Prue Napier have during two decades prepared both popular books and stringently technical volumes within their irresistible and active discipline. Now they have compiled another handsomely illustrated brief survey. Half of the book is a compact, full profile of these our kin, primates figured and described genus by genus the world around. The rest presents a clear and current account of the classification, origin, structure and behavior of the most distinctive order of mammals. The last chapter reviews the current understanding of human evolution, the jigsaw puzzle of our own origin that we crave to assemble out of so fragmentary a set of fossil pieces.

On their title page the Napiers repeat a drawing of a fond pair of dusky titis, small monkeys of the South American rain forest, their long tails entwined in a droll and endearing intimacy, a ritual "of social rather than sexual significance." Titi couples live as a nuclear family group; the two parents share alike the care of their single offspring almost from birth. Primates dwell in social units of quite varied nature. For 50 years it was held that sexual attraction bound primate groups, since the females were thought to be in continuous estrus throughout the year, in contrast to the short seasonal receptivity of most mammalian females. True for some important forms, such as humans, this does not hold for

Male, female and infant proboscis monkeys in the Twycross Zoo many of the primates observed nowadays, who nonetheless live in coherent groups. Primate social adaptation has complex roots in advantages we can so far only surmise.

The little vervet monkey of Kenya sounds four distinct calls of alarm coded for the predators leopard, eagle, snake and baboon. On hearing the eagle alarm the vervets look upward and run for cover. The snake alarm induces them to scan the ground nearby. The leopard signal makes them run up a nearby tree. (One timid vervet infant touchingly sounded the eagle alarm for a falling leaf.) Far less than language, this modest communication system is on the way there.

The primate genus of the wily macaque monkeys is remarkable. The macaques have adapted to the challenge of ground living, although they still resemble in form and locomotor style the quadrupedal African monkeys constrained to climbing in trees. Their 16 species have learned to live almost anywhere and to eat almost anything, from tree bark to crabs. As "Barbary apes" their presence sustains imperial Gibraltar, we are told, while other macaque species abound across all North Africa, around tanks and temples anywhere eastward from the Indus, thence overseas to Indonesian rain forests and the light snows of a Honshu winter.

The precision grip that daily serves us so well is not lacking in baboons or macaques, although no hands are as deftly prehensile as our own. The characteristic striding heel-and-toe human gait is marked in ashy footprints at Laetoli. Those early walkers more than three million years ago had cranial cavities still apelike. The celebrated studies of our chimpanzee cousins in the field have shown they are sometimes able to cooperate.

The primate talent is for specialized and subtly apt behavior realized by visible anatomical structures that remain plastic and unspecialized. This statement points to neuroelectronics as the locus of selection. The Napiers are skeptical that any single behavioral threshold was key to the patent for humans. Hands, toolmaking, language, culture—these and more are shared to a degree by primates still far from human. That the genus *Homo* has evolved so rapidly probably shows at work the selective cultural feedback loop that grew along with the successes of cultural behavior. It is the swift rise in the combinatorial richness of our behavior that has made us human, not this or that particular clever trait. Prove all things, counseled St. Paul, and—here we primates modify his preachment—hold, but not too fast, that which is good.

Comparis
been m

Honda has a way of bringing out the best in everyone. It's called the Accord.

Whenever we introduce a new Accord, everyone else returns to the drawing board. To try their best. Now we've done it again.

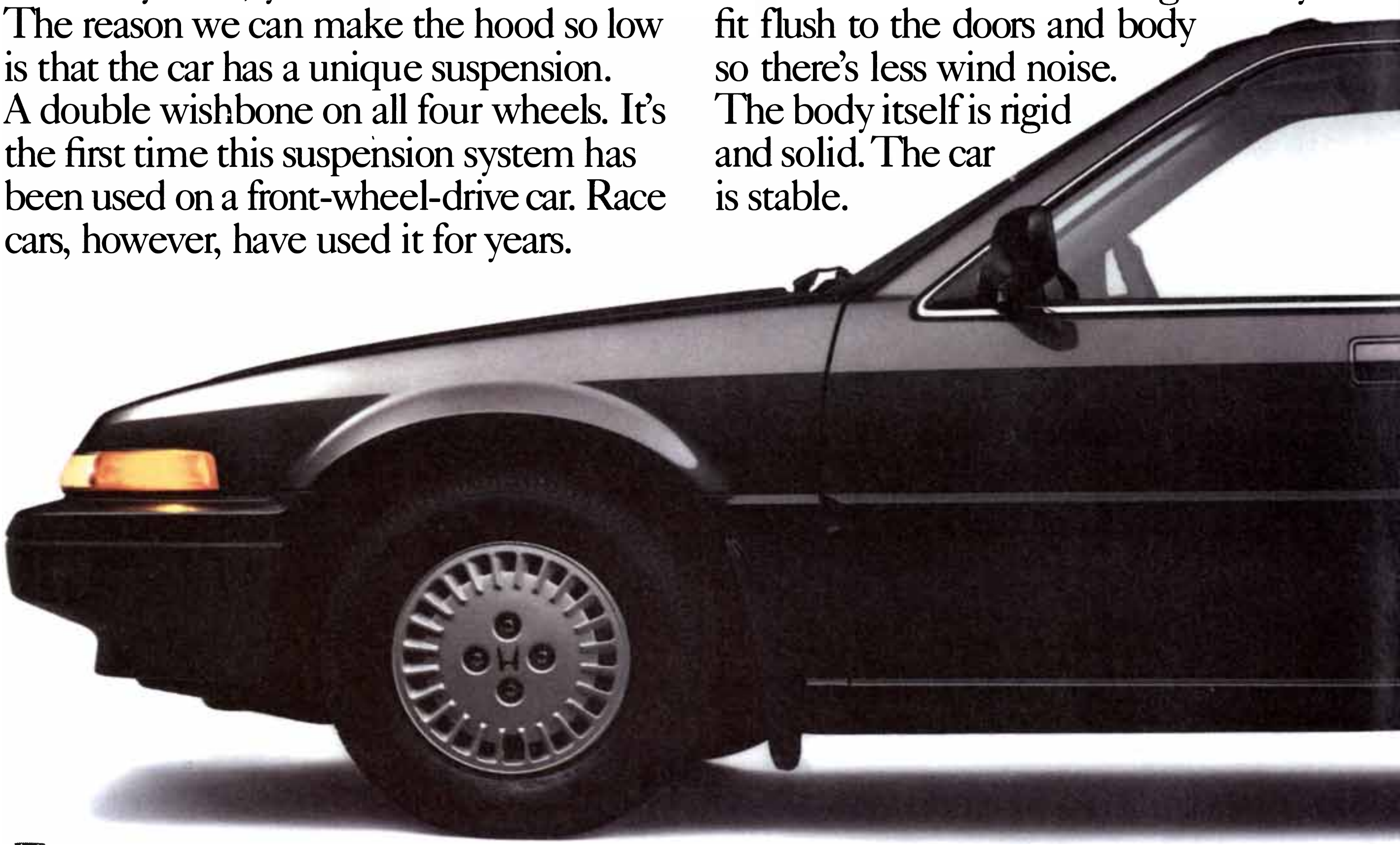
This is the new Honda Accord LX 4-Door Sedan. There's nothing else like it.

The hood is lower for aerodynamic efficiency. Also, you can see better over it. The reason we can make the hood so low is that the car has a unique suspension. A double wishbone on all four wheels. It's the first time this suspension system has been used on a front-wheel-drive car. Race cars, however, have used it for years.

The new suspension does more than give the car a beautiful shape. It gives it a lovely ride, too. The handling is outstanding. We promise you'll be impressed.

You'll also be very comfortable. There is more headroom even though the roofline is lower. The car is also wider.

You'll also see through a lot more glass because the windows are larger. They all fit flush to the doors and body so there's less wind noise. The body itself is rigid and solid. The car is stable.



ons haven't
ade yet.

Conveniences have been well considered. This luxurious car comes with air conditioning. A new electronic stereo system with a cassette player. There are power windows and door locks. A power radio antenna. Dual power remote mirrors.

You might call the engine more of a powerhouse. Its size has been increased to two liters. And its mounts have been redesigned to absorb vibrations. It's smooth.

Variable-assist power steering takes the rough work out of parking. Cruise control makes long trips even more relaxing. And the trunk is substantially larger so you can carry more luggage.

While these extras might be available as optional equipment on other cars, they come standard on this car. We invite your comparison. HONDA

The New Accord

Established 1845 SCIENTIFIC AMERICAN November 1985 Volume 253 Number 5

The Treatment of Diseases and the War against Cancer

To judge the progress in the war against cancer one must understand the role of trial and error in the evolution of medicine and know something about the natural history of the commoner forms of cancer

by John Cairns

Cancer is a disease that will touch most of us, directly or indirectly. One in three Americans gets cancer at some time in his or her life, and one in five dies of it. This article measures the progress that has been made in the treatment of cancer. Because it is a rather controversial subject, I describe how such judgments are made and what sources of information are available. Before discussing cancer, it is useful to consider briefly the problem of developing a treatment for any disease.

The molecular biology of all living creatures shows an underlying unity, even though the particular form of each species reflects chance events in its evolutionary history. We can be confident that any newly discovered animal will prove to have the same kinds of informational macromolecules and the same genetic code as other forms of life, but even if we were given a complete description of the animal's habitat, we could not predict exactly what the animal would look like because that would depend on the history of its ancestors. Still less could we predict the animal's diseases. For example, who could have guessed that *Homo sapiens* would share with the humble guinea pig the unenviable distinction of being incapable of synthesizing ascorbic acid or share with armadillos a susceptibility to the bacterium that causes leprosy, or that intestinal cancer usually occurs in the large intestine of humans and in the small intestine of sheep?

If we cannot predict the existence and salient characteristics of each disease, we certainly have no basis for deciding a priori how diseases should best be treated. It would seem to be a short step, therefore, to conclude that the treatment of diseases ought to be based on the results of some rational system of trial and error. Yet this is a rather new idea in the annals of medicine. For example, a famous early 19th-century comparison of the fate of a group of patients with pneumonia, who were bled at various stages of their disease, showed that bloodletting did not affect either the average duration or the fatality of their disease. Most were ill for two to three weeks, and about 25 percent of them died. The author of the study did not go so far as to suggest that these patients might actually have done better if they had been left alone, but still he was attacked for daring to think that patients could be compared with one another. As one of his critics put it, "By invoking the inflexibility of arithmetic in order to escape the encroachments of the imagination, one commits an outrage upon good sense." (In the absence of proper clinical trials the issue remained in doubt, and it was not until well into the 20th century that bloodletting fell completely out of favor.)

With diseases such as pneumonia or cancer, which are sometimes fatal and sometimes not, there may be no way of determining whether any particular patient's survival was predestined or should be attributed to the treatment. And so it becomes necessary to compare the response of groups of patients rather than the response of one or two individuals. Some 100 different kinds of human cancer are recognized. Each has its characteristic behavior that includes average age of onset, rate of growth and tendency to spread and to be lethal. Each must therefore be considered as if it were a separate disease. Furthermore, most cancer patients are already well into middle or old age, so that some way must be found to correct for other, "competing" causes of death. After all, even the most successful treatments would not be expected to protect a 90-year-old patient from all forms of mortality. Like the 19th-century doctor who studied pneumonia, the work of the modern clinician starts therefore with an investigation of what happens to patients with each type of cancer and how their life expectancy compares with that of people who do not have cancer.

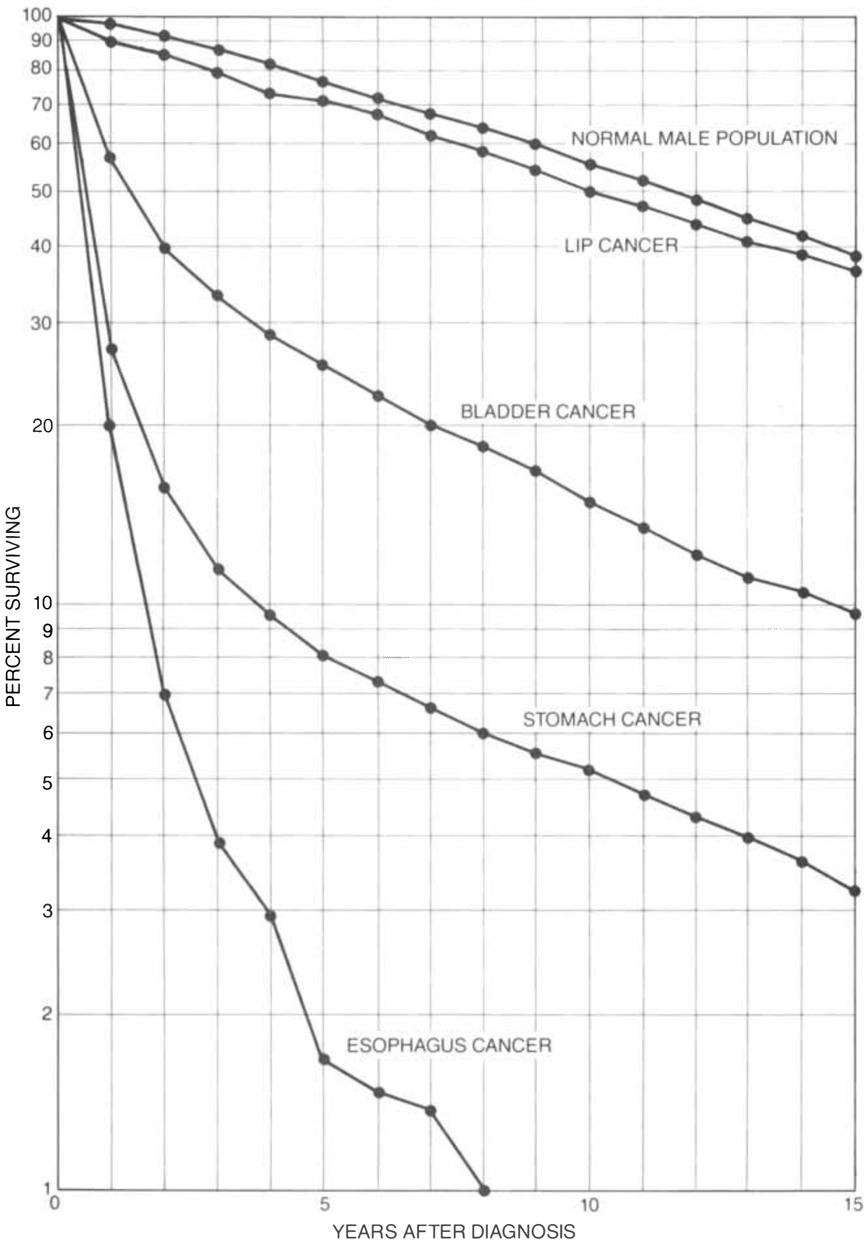
Cancer arises when some cell in the body starts to multiply without restraint and produces a family of descendants that invade the surrounding tissues. Such invasion may be followed by metastasis, or spread to distant sites, by way of the lymphatics and the bloodstream. This process of metastasis is the main reason for the lethality of cancer, because it puts the disease beyond the reach of surgery and local irradiation. Some cancers, for reasons not known, are incapable of metastasis (for example, most forms of skin cancer). These cancers are easily dealt with, unless of course the process of local invasion is itself lethal (as it can be, particulary in certain forms of brain cancer). At the other extreme, the normal cells of the bone marrow and lymphatic system are already programmed to move around the body, and so it is not surprising that cancers arising among these cells (the leukemias and lymphomas) are likely to be disseminated throughout the body from the start. Most forms of cancer lie between these extremes.

Although it has recently become possible to study directly the changes in gene structure and function found in certain cancer cells [see "A Molecular Basis of Cancer," by Robert A. Weinberg; SCIENTIFIC AMERICAN, November, 1983], we still know little about what controls the multiplication and territorial restraints of most cells and tissues. For the time being, therefore, our knowledge of the behavior and prognosis of each type of cancer remains largely empirical. Over the past century pathologists have built up a classification of human cancers according to the origin and category of cell involved, and they have further subdivided each type of cancer according to the appearance of the cells and their general growth pattern. This classification is important because the different cancers behave in very different ways. Some tend to be rapidly fatal and some are not; most kinds occur predominantly in old age but a few occur only in children; some are common and some are rare; some are common in rich nations and rare in poor ones, and for others the reverse is true.

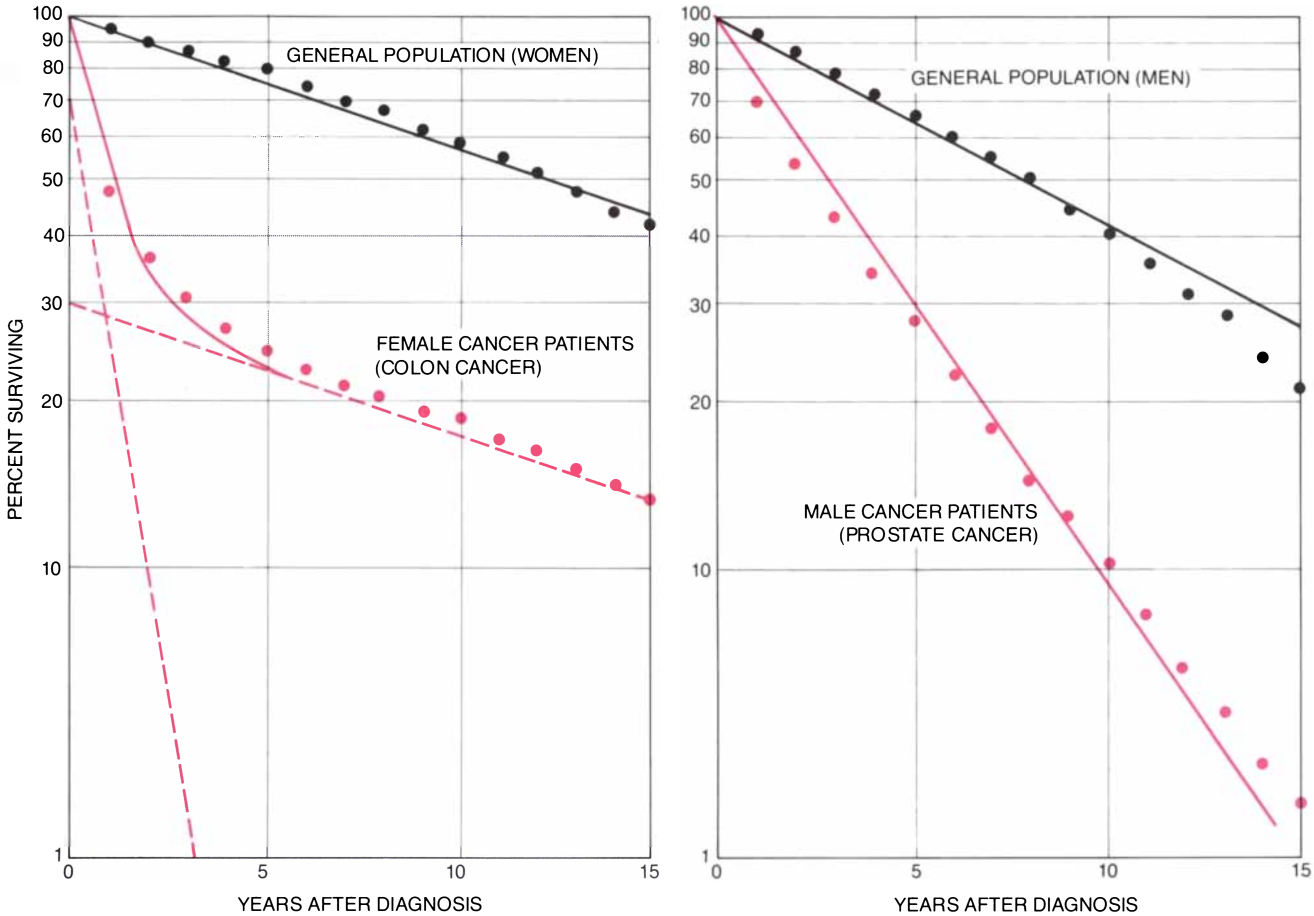
Since World War II cancer registries have been set up in several states and nations to record the changing trends in cancer incidence and mortality. The records collected by some of these registries yield a rather precise picture of the natural history of cancer, and that is the necessary starting point for any discussion of treatment. The Cancer Registry of Norway has published its findings in book form and is therefore a particulary accessible source. Norway has a population of about 3.5 million, and the registry has followed the fate of the 200,000 cancer cases diagnosed there since 1953. As a sample of these statistics, the illustration at the left shows the pattern of survival of men with four kinds of cancer, chosen to demonstrate how differently various cancers behave. At some sites, such as the lip, cancer is associated with a negligible decrease in life expectancy; at other sites, such as the esophagus, it is nearly always rapidly fatal.

A group of patients can be considered cured of their cancers if they die at about the same rate as the general population, which they would if, thanks to their treatment, they had been returned to the common pool. The essential step, therefore, in determining how often a particular type of cancer can be cured or controlled is to estimate from statistics what fraction of cancer patients die at the same age as they would have died if they had not had the cancer (that is, what fraction are dying from causes unrelated to their cancer). The survival rate of Norwegian women who have colon cancer, for instance, has been compared with the survival rate of the general population of women characterized by the same age distribution. Most of the patients die rather soon after diagnosis, but a sizable minority, about 30 percent, die at the same rate as the general population (that is, behave as if they have been cured). That is what we would expect if some of the patients die of their cancer and some do not.

The Norwegian registry lists 40 major sites of cancer for males and 43 for females. Similar calculations can be made for each site. By summing the estimates I calculate that about 25 per-



SURVIVAL OF NORWEGIAN MEN who had various types of cancer is compared with the survival of similarly aged males in that nation's population as a whole. In the years after diagnosis lip cancer caused a negligible decrease in life expectancy, whereas cancer of the esophagus was nearly always rapidly fatal. The survival rates for patients who had cancers of the bladder and the stomach fell somewhere between those extremes. The statistics cover the years from 1953 through 1964 and come from the Cancer Registry of Norway.



CANCER PATIENTS are considered cured if they die at about the same rate as the general population. The graph at the top left compares the survival rate of Norwegian women who had colon cancer (*colored dots*) to the survival rate of the general population of women who had the same age distribution (*black dots*). The solid gray line shows that the general population had a half-life of about 13 years. (A half-life is the time required for half of a given population to die.) The solid colored line is the summation of the two broken lines and shows that the women with colon cancer can be considered to fall into one of two categories: 70 percent had a half-life of eight months and 30 percent had a half-life of 13 years. In other words, a subgroup consisting of 30 percent of the women died at the same rate as the general population. Some cancers, on the other hand, show no such subgroup and in this sense should perhaps be considered incurable by present methods. Cancer of the prostate is one example; the graph at the top right compares the survival rate of Norwegian men who had prostate cancer (*colored dots*) to the survival rate of similarly aged men in the general population (*black dots*). The solid gray line shows that the general population had a half-life of eight years, the solid colored line that the cancer patients had a half-life of three years. About a third of these cancer patients suffered no loss of life span as a result of their disease.

cent of all male cancer patients and about 40 percent of all female cancer patients died of causes unrelated to their cancer. In other words, about a third of all Norwegian cancer patients suffered no loss of life span as the result of their disease.

In these statistics from the 1950's and 1960's we are looking at the results of treatment by surgery, occasionally backed up by X-irradiation when the primary tumor was inaccessible to surgery. It is the picture of what used to happen before the advent of screening programs, chemotherapy and numerous clinical trials. The major ancillary aids to surgery, such as blood transfusion, antibiotics and improved forms of anesthesia, had been developed and disseminated by the early 1950's, so that the deciding issue, for nearly every patient, had by then become the extent of spread of the cancer at the time of surgery. Once a cancer had metastasized to sites beyond the reach of surgery or radiotherapy, the patient's fate was almost entirely determined by the rate of growth and further spread of the residual cancerous cells. Only through the intervention of some other cause of death would the patient be spared from death by cancer.

The importance of the extent of spread at the time of operation is shown in the illustration on the next page, which gives the survival of women with colon cancer according to the state of their disease. Plainly the patients' chances were least good if their cancer had already spread when it was first diagnosed. There are, however, two possible explanations for this effect. The first and more obvious interpretation is that the crucial determinant is the time of diagnosis; according to this view, the cancers that have already metastasized when they are first seen have simply been left too long, but they could have been diagnosed earlier while they were still localized.

The less obvious explanation is that we are seeing here not so much a variation in time of diagnosis as a variation in ability to spread and produce metastases. In other words, the localized cancer might have been destined to stay localized for many years after it had become detectable, whereas the cancer that had metastasized might consist of cells so apt to spread that the cancer would already have produced metastases when it was very small and still undetectable. If the first explanation is the correct one, earlier diagnosis could bring great benefit; if the second is correct, there might be negligible benefit. The actual benefits of earlier diagnosis must therefore be determined by properly controlled trials, and this must be done for each kind of cancer since what is true for cancer of the breast may not be true for lung cancer, and so on.

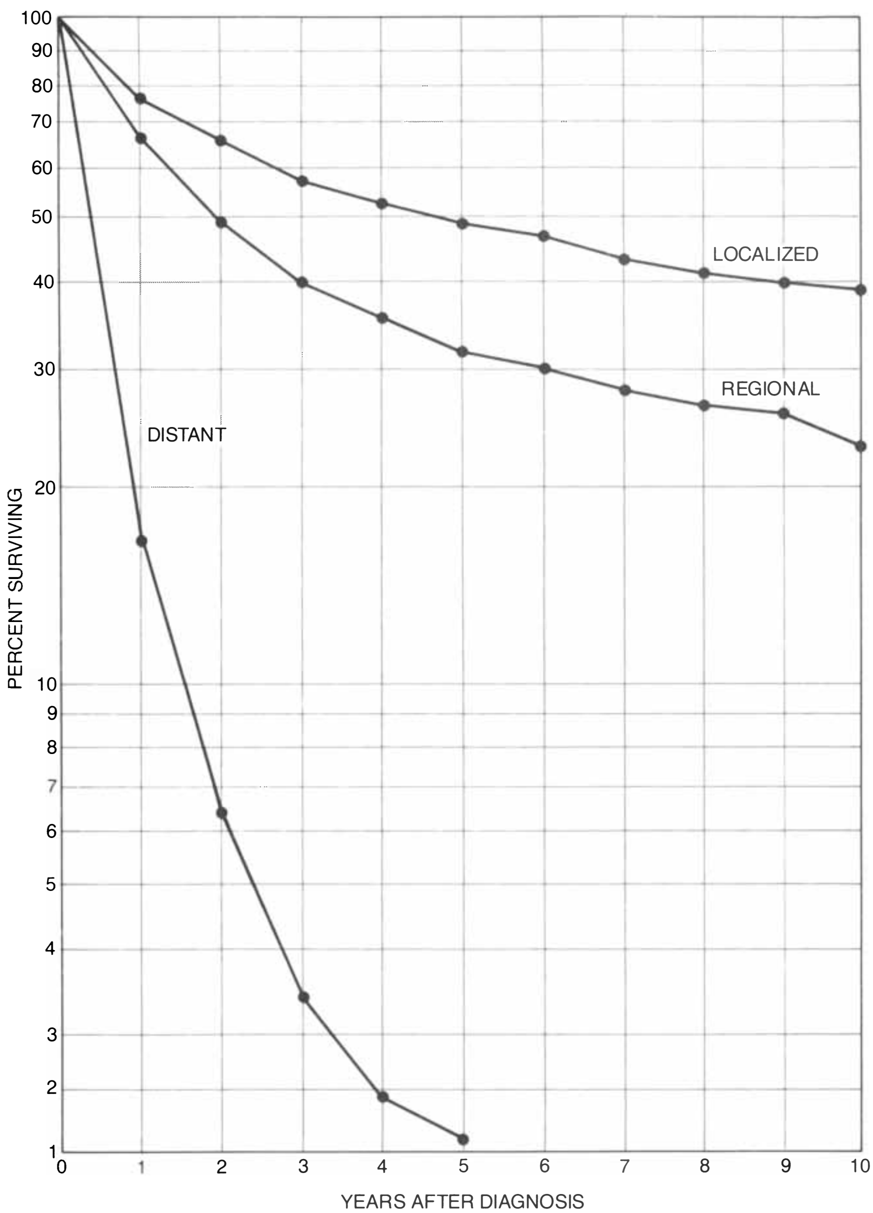
A famous example of a trial to measure the effect of early diagnosis was the study of breast cancer started 20 years ago in New York and funded by the National Cancer Institute. The study followed 62,000 women aged 40 to 64 who were covered by the Health Insurance Plan of Greater New York. The women were separated into several categories according to age, family size and income. Each category was then evenly and randomly divided into two groups. Those in one group, the "test" group (consisting of 31,000 women), were offered a free annual check by physical examination and X-ray mammography for early evidence of breast cancer. The remaining 31,000 women, the "control" group, were given no special encouragement to be examined and were left undisturbed. The annual checks of the test group were continued for four years, and the study has monitored the subsequent fate of both groups since then.

The trial was therefore designed to answer the following practical question: Does the act of offering such free annual examinations bring any measurable benefit (that is, is it possible to intercept breast cancers before they have spread, and will a high enough proportion of women agree to be examined)? The answers were promising indeed. Two-thirds of the women in the test group had at least one examination, and in the nine years of follow-up the test group as a whole suffered significantly fewer deaths from breast cancer (91 deaths compared with 128 in the control group).

The study also showed how important it is to have randomly selected control groups. Overall, the two groups showed about the same incidence of breast cancer and the same general mortality from causes other than cancer—as they should, since the groups were chosen at random. The remarkable finding was that in the test group the 10,200 women who refused to be examined had a higher general mortality, but a slightly lower incidence of and mortality from breast cancer, than the 20,800 women who agreed to be examined.

The explanation lies in the fact that the separation of the test group into the examined and the unexamined was the result of self-selection. The women who refused to be checked were less interested in their health and proved to have a lower level of education than the women who agreed to be checked. Because breast cancer is commoner among the well-educated than it is among the indigent and less-educated, these differences between the two self-selected categories of women are actually not surprising. There is therefore an important lesson to be learned here. If the comparison had been simply between the women who were examined and those who refused to be, the study could have reached the ludicrous conclusion that annual examination for the early signs of breast cancer lowers general mortality and slightly raises mortality from the very disease the exercise is designed to prevent. But because a randomly selected control was included, the study produced a reliable estimate of the actual benefits of screening for breast cancer.

To summarize the results, the New York Health Insurance Plan's study (and a similar trial, recently reported from Sweden) suggests that about a fourth of the total mortality from breast cancer (that is, a fourth of some 35,000 deaths a year) might be prevented if all women in the U.S. over the age of 50 were offered a free examination every one to three years. The



EXTENT OF METASTASIS, or spread, of a cancer has a large impact on the chance of long-term survival after diagnosis. The data shown here are from a study of Norwegian women between the ages of 55 and 74 who were diagnosed as having cancer of the colon. In some of the patients (*top curve*) the cancer was localized in the gut wall, and two-thirds of these women appear to have suffered no loss of life expectancy. When the cancer was seen to have spread to the local lymph glands (*middle curve*), only a third of the patients had a normal life expectancy. When the cancer was seen to have undergone metastasis to distant sites (*bottom curve*), the prospects for long-term survival were most unfavorable.

			DEATHS FROM OTHER CAUSES IN FIRST FIVE YEARS		BREAST CANCERS DETECTED IN FIRST FIVE YEARS		DEATHS FROM BREAST CANCER IN FIRST NINE YEARS	
		NUMBER OF WOMEN IN EACH GROUP	NUMBER	DEATHS PER 1,000 WOMEN	NUMBER	INCIDENCE PER 1,000 WOMEN	NUMBER	DEATHS PER 1,000 WOMEN
TEST GROUP	EXAMINED	20,800	421	20	225	11	63	3.0
	REFUSED	10,200	429	42	74	7	28	2.8
	TOTAL	31,000	850	27	299	10	91	2.9
CONTROL GROUP	TOTAL	31,000	877	28	285	9	128	4.1

SCREENING FOR EARLY DIAGNOSIS of breast cancer can sometimes intercept the disease before it has spread, as a study funded by the National Cancer Institute indicates. The study followed 62,000 women between the ages of 40 and 64 who were covered by the Health Insurance Plan of Greater New York. The women were divided randomly into two groups: a test group and a control group. Those in the test group were offered a free annual check by physical examination and X-ray mammography for early evidence of breast cancer. About two-thirds of the women in that group chose to have at least one examination during a four-year period. Those in the control group were given no special encouragement to be examined and were left undisturbed. After nine years of follow-up the test group suffered significantly fewer deaths from breast cancer (91 deaths) than the control group (128 deaths).

expense of such a program (more than $100 million per year) has stopped it from being widely adopted, but at least the estimates are there, ready to be brought into any calculation of priorities, costs and benefits.

The other main screening procedure currently in use is the Pap smear, or Papanicolaou test, for the early diagnosis of precancerous changes in the cervix. Overall, on a worldwide basis, carcinoma of the cervix is the commonest lethal cancer of women. Like most other forms of cancer, but unlike breast cancer, it is much commoner among the poor and the less-educated than it is among the educated. In the U.S., for instance, the rates in the highest and the lowest social classes differ about fivefold.

The Pap smear was invented in the late 1920's by Aurel Babès in Bucharest and developed by George N. Papanicolaou of the Cornell University Medical College. The test involves the microscopic examination of cells scraped from the surface of the cervix, at the entrance to the uterus. In the years following World War II the Pap smear came into use as a way of detecting the early stages in the development of cervical cancer. By now it has been applied one or more times to at least 75 percent of the adult women in the U.S. No attempt was made to set up a properly controlled trial like the breast-cancer trial, and for this reason the benefits of the test have not been precisely established. Indeed, the issue is complicated.

The mortality from cervical cancer in the U.S. has been steadily declining probably since the 1930's, presumably because during this period the average levels of hygiene, affluence and education have gone up. Since the Pap smear was introduced in the U.S. at a time when cervical cancer was already on the decline, overall national mortality figures cannot be used as evidence for its success. Nor is it fair simply to compare the mortality among women who have been tested with the mortality among those who have not; in the absence of strong pressure the better-educated will be more likely to agree to testing than the less-educated, and so they would be expected to have a lower mortality from cervical cancer even if the test were of no benefit.

The case for the effectiveness of the Pap smear rests on two separate observations. First, whenever it is possible to compare otherwise similar populations of women who were offered testing programs that started at different times (for example, the women in the different provinces of Canada or the different Scandinavian countries), it does seem that the decline in mortality from cervical cancer invariably accelerated at the time testing became widespread. Second, women who are found to be in the early stages of cervical cancer but who do not then return to the clinic for treatment subsequently experience a much higher mortality from cervical cancer than the women who do return for treatment. Neither of these arguments is absolutely ironclad, but together they are strong enough to make it no longer ethical to conduct a proper trial of the Pap smear.

At various times attempts have been made to establish programs for the early diagnosis of other kinds of cancer. Screening programs for cancer of the skin and mouth have proved effective. Unfortunately few of the sites of the common lethal cancers are as accessible as the breast, the cervix, the skin and the mouth, and there is little sign so far that any of the other programs will prove worth applying on a large scale. In principle, screening programs can succeed only for those cancers characterized by an early precancerous state that is detectable or for those cancers that commonly go through a prolonged stage in which they are detectable but still have not spread beyond the reach of surgery.

Regrettably, not all types of cancers fall into either of these categories. Some years ago, for example, a large-scale trial of the early diagnosis of lung cancer indicated that no great benefit comes from having the disease detected by chest X rays before it has produced any symptoms; it seems that by the time most lung cancers are detectable they have spread too far to be treated. For each type of screening program it is therefore important to set up, early on, a proper trial on a randomly selected group of subjects, because there is no intuitive way of knowing which cancers can be intercepted by early diagnosis.

To summarize, screening programs for earlier diagnosis sometimes bring benefits and sometimes do not. Aside from questions of efficacy, however, it seems unlikely that any country—even one as rich as the U.S.—will ever be able to afford to test the majority of its population annually for the earliest signs of each of the major cancers.

It remains a depressing truth that fewer than 50 percent of cancer patients can be cured by surgery. A tremendous effort has therefore gone into discovering adjuvant forms of treatment that can be given following surgery. The three widely employed techniques of adjuvant therapy are hormonal. treatment, X-irradiation and chemotherapy.

Hormonal treatment is the obvious form of adjuvant therapy to apply to cancers that arise in hormone-responsive tissues such as the breast and the prostate gland. Beginning in the 1890's the ovaries were removed from women who had spreading breast cancer in the hope that the consequent drop in circulating estrogens would slow the growth of the cancer cells. To achieve the same effect without removing the ovaries certain structural analogues of estrogen such as tamoxifen are now employed, which work by blocking the estrogen receptors of the cancer cells. Similarly, the growth of prostate cancer can often be slowed or inhibited by removing the testes or by giving the patient estrogens. Although not all breast and prostate cancers respond to hormonal control, one advantage of this form of adjuvant therapy is that its side effects are usually minor.

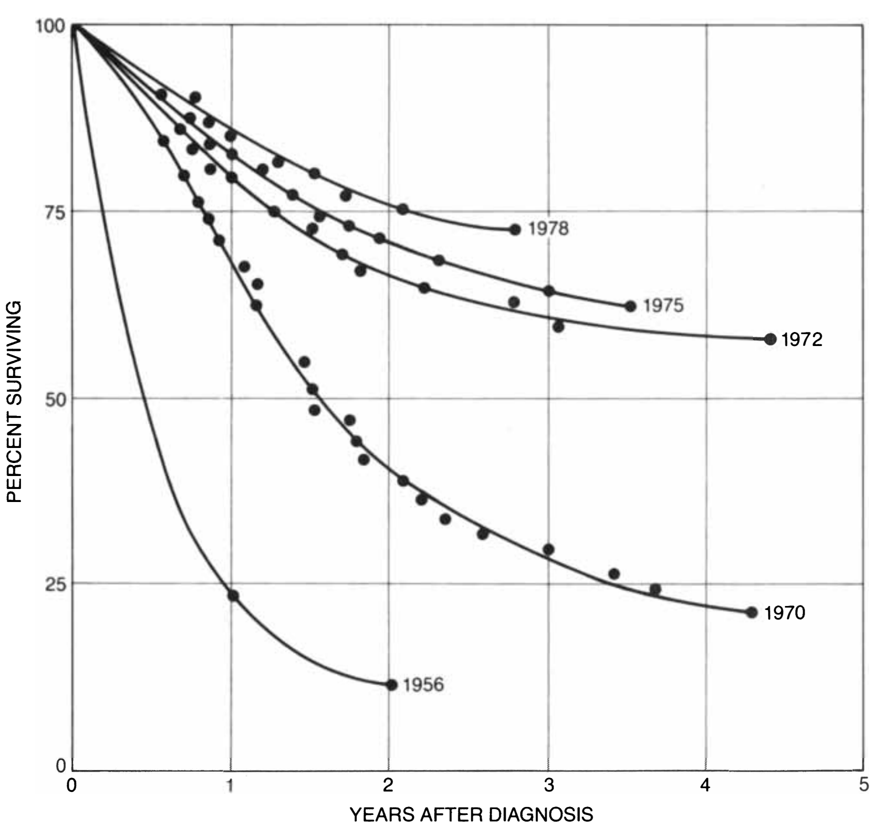
Soon after Wilhelm Roentgen discovered X rays in 1895, investigators found that the radiation could damage human tissues. As a result X rays were soon being tested as a means of treating breast cancers that had undergone local recurrence after surgery. X-irradiation is now one of the mainstays in cancer therapy. But there are limits to its use. Excessive whole-body irradiation damages the immune system, the bone marrow and the lining of the intestines. It is the damage to these tissues that is the basis of radiation sickness. (We now interpret most of the effects of radiation in terms of damage to the genetic material, DNA, and this will presumably tend to be greatest in the tissues undergoing the fastest cell division because the more often a cell divides, the less time it has available for the repair of any damage to its DNA.) The treatment of cancer by X rays therefore depends on the relative sensitivity of the cancer compared with the normal tissues that surround it and on whether the radiation can be concentrated on the cancer.

How much radiation can be tolerated by each part of the body is now accurately known. Furthermore, it is feasible, with the high-energy X rays available today, to concentrate radiation into any target organ quite precisely. It has therefore become possible to treat the more sensitive cancers such as Hodgkin's disease, cancer of the cervix and one kind of testicular cancer without producing an unacceptable level of radiation sickness. The majority of cancers, however, cannot be cured by radiation because the dose of X rays required to kill all the cancer cells would also kill the patient.

The next major contribution to the treatment of cancer, cytotoxic chemotherapy, originated with the observation that one of the long-term toxic effects of the mustard gases used in World War I was damage to the bone marrow. (Incidentally, just like X rays, these highly reactive, toxic chemicals proved capable of damaging DNA.) Not long after World War II, when the mutagenicity of mustard gases ceased to count as a military secret, trials were undertaken to test the efficacy of such radiomimetic chemicals (chemicals that produce effects similar to those of radiation) in treating cancer. The early results were encouraging, and in the next 20 years many more chemicals were added to the list of drugs used for chemotherapy.

At present a large number of chemotherapeutic agents are employed, in one combination or another, for the treatment of cancer. Some are prepared synthetically (for example cyclophosphamide, certain nitrosoureas and more recently certain organic metal compounds such as *cis*-platinum). Others are natural toxins (for example plant alkaloids such as vincristine and fungal toxins such as the actinomycins). Nearly all these reagents bind to DNA and cause damage that the cell cannot repair properly. This seems to be the basis of their toxicity, for both the cancer cells and the normal tissues of the body. The other group of chemicals currently in use are certain antimetabolites that block the synthesis of DNA or its precursors (for example fluorouracil, cytosine arabinoside and methotrexate).

The first efforts at chemotherapy concentrated on childhood leukemias and lymphomas for two reasons. First, these cancers, because they are dispersed from the outset, were almost inevitably fatal. Second, the patients, because they were young, had much more to gain from a cure than old people. Although it proved quite easy to achieve a temporary remission with chemotherapy, in nearly every instance the cancer eventually returned,



CHEMOTHERAPY has greatly improved the prospects of survival for children who have leukemia. Only about 10 percent of the children diagnosed in 1956 as having leukemia were still alive two years after diagnosis; by 1978 the number of two-year survivors was roughly 70 percent. The data are from Denman Hammond of the Children's Cancer Study Group.

and it subsequently resisted further chemotherapy. This was not really surprising because even the smallest detectable cancer consists of at least a billion cells; any population of cells as large as that can be expected to contain some variants that are better at growing in the face of selection pressure. As a result two principles have become established that—rightly or wrongly—have determined the subsequent development of most kinds of chemotherapy. If all the cells in a cancer are to be destroyed, it may be necessary not only to use each chemical agent at the highest tolerable level but also to use different agents simultaneously.

With suitable combinations of chemicals it is now possible to cure many kinds of childhood cancer that would otherwise be rapidly fatal. For instance, most children with leukemia can now apparently be cured; to be more precise, a minority relapse and die during or soon after the end of their course of chemotherapy, but the majority enter what at the very least is a prolonged period of relapse-free survival. The hope now is that these survivors will prove to have a normal life expectancy. Similar results have also been recorded for other childhood cancers. The best measure of these successes is to be found in the national mortality statistics. In the early 1950's in the U.S. about 1,900 children under the age of five died of cancer each year. The rate is now down to about 700 per year, suggesting that two-thirds of all children with cancer are now being cured of it.

The reduction in the annual mortality of older children and young adults has been less spectacular, with the following notable exceptions. Hodgkin's disease used to be inevitably fatal, but now most patients can be cured. This represents a saving of some 1,000 lives in the U.S. each year for all age groups combined. About 35 percent of testicular cancers were fatal before chemotherapy, but now roughly a third of these deaths can be prevented, a saving of some 300 lives a year. Finally, choriocarcinoma, a rare cancer of the placenta that occurs about once per 40,000 pregnancies in the U.S., can now be cured by chemotherapy, a saving of perhaps 20 to 30 lives a year. Overall, however, the gains have been limited. The latest figures for the U.S. show about 7,000 deaths per year from cancer under the age of 30, compared with the 10,000 we would have expected if the death rate had remained unchanged since the 1950's.

It is important to note that so far in this discussion there is no discrepancy between the total number of cures estimated from the apparent cure rates for certain cancers and the actual change in the inventory of deaths recorded at the national level. Each year about 3,000 patients under the age of 30 are being cured by chemotherapy who otherwise would have died.

Only 2 percent of the patients who die of cancer are under 30, however. For the vast majority of cancers, which arise in older patients, the results of chemotherapy are much more controversial. The figures on mortality, assembled and published by the statisticians at the National Cancer Institute, show several major changes in the past 25 years. Deaths from lung cancer are on the rise, particularly in women, as the delayed result of the increase in cigarette smoking. Deaths from cervical cancer are going down, thanks in part to the Pap smear. The death rate for stomach cancer continues its unexplained downward trend, which started in the 1930's, and many less common cancers are drifting slightly in one direction or another. Apart from the success with Hodgkin's disease, childhood leukemia and a few other cancers, it is not possible to detect any sudden change in death rates for any of the major cancers that could be credited to chemotherapy. For the old and middle-aged, therefore, the picture is more one of stability than it is one of change.

Those who organize cancer centers and supervise the many clinical trials of chemotherapy look for ways to circumvent these relentless statistics. Sometimes they explain away the unchanging statistics for mortality by pointing out that the national statistics are inevitably a few years behind the times and therefore do not reflect the most recent advances in treatment. Although this point is absolutely correct, it has been made repeatedly in the past 10 years but has never been vindicated by national statistics when these eventually became available. For the most part, however, the organizers disregard the figures for mortality and simply point out that the fraction of patients who are still alive five years after diagnosis has been steadily increasing for nearly every kind of cancer. They attribute this increase in five-year survival to steady improvements in methods of treatment.

Before we conclude that the statistics for mortality are unreliable we should, however, consider whether there could be any systematic error in the national figures for survival rates. As we have seen, it is possible to translate the results of clinical trials into numbers of lives saved on a national scale when considering cancers such as childhood leukemia, where the diagnosis is not in doubt and the outcome depends entirely on whether the treatment is successful or not. But for cancers that are not invariably fatal the calculation is fraught with difficulty because it turns out that we have no certain way of estimating how many lives are waiting to be saved.

To take one rather extreme example, a fourth of all U.S. males over the age of 70 who have died from other causes can be shown on routine postmortem examination to have small cancers of the prostate. We know from incidence data, however, that fewer than 10 percent of those cancers were destined to produce symptoms and still fewer would have proved fatal. Therefore any campaign to detect and treat prostate cancers while they were still small and before they had produced any symptoms is certain to include many "cancers" that would not have been detected except for the campaign. Even if the campaign saved no lives, the inclusion of these additional, nonfatal "cancer" cases would inevitably increase the proportion of "patients" who survived.

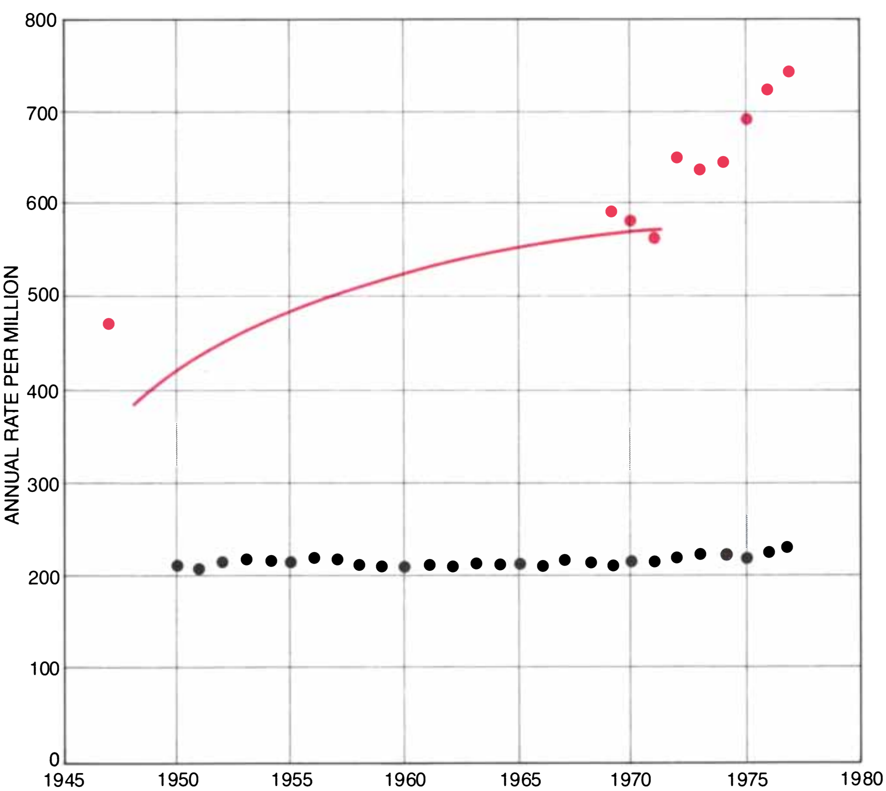
Something like this appears to have happened in the U.S. over the past 30 years. Although clinical trials have failed to demonstrate any major advances in the treatment of cancer of the prostate since the introduction of hormone therapy in the 1940's, the five-year relative survival of cases is reported to have gone up from 43 to 63 percent. The national statistics, however, show that it is the reported incidence of new cases that has changed, going from 400 per million men per year in the late 1940's to about 700 per million per year in the late 1970's; the death rate has remained steady at about 210 deaths per million per year. The survival rate has therefore increased not because fewer men are dying from prostate cancer but because more men are being classified as having prostate cancer.

Similar artifacts probably affect the survival rates for many other types of cancer, particularly cancer of the breast. It has therefore become a principle, at least among many cancer epidemiologists, that the comparison of the survival of patients in different eras is not in general an acceptable measure of therapeutic success (unless, as in the case of childhood leukemia, it is clear that there has been no change in the definition of what constitutes disease). For most forms of cancer, therefore, physicians are forced back once again to the "inflexible arithmetic" of clinical trials. Groups of patients must be separated at random and given the various rival treatments. The average survival rates of the subgroups will then show which treatments do good and which do harm.

The best-studied case concerns the use of various adjuvant therapies following surgery for breast cancer. Recently the results from a large number of trials have been brought together and summarized. Altogether the trials covered about 5,000 women who were treated with various toxic chemotherapeutic agents. A similar number of women, selected at random, received no extra treatment after surgery. The patients have now been followed for between one and 10 years. During this period the treated group as a whole has had about 25 percent fewer deaths than the control group; for women under the age of 50 the reduction has been by about a third. Whether these patients have really been cured by their treatment will not be known until the two groups have been followed for many years, but even if there has been only a postponement of death, this may have been worthwhile. In any event, the chemotherapy of breast cancer can offer real benefits, although so far they are rather modest and not to be compared with the results for certain childhood cancers.

Armed with just these facts, it would be tempting to conclude that chemotherapy should be given to every woman with breast cancer; after all, one-third fewer deaths among women under 55 could conceivably be translated into saving 2,000 or 3,000 lives each year in the U.S. alone. The actual story, unfortunately, is more complicated than these statistics suggest. A six- or 12-month course of chemotherapy not only is a very unpleasant experience but also has its own intrinsic mortality. Furthermore, many of the drugs used are known to be carcinogenic, and one of the long-term effects of chemotherapy is that somewhere between 5 and 10 percent of the surviving patients die of leukemia in the first 10 years after treatment. These may seem like relatively minor hazards for a patient who has an advanced and rapidly growing cancer, but they would be serious considerations for a woman who has a small and apparently localized cancer of the breast. Her chance of dying of her cancer within five years is only about 10 percent even if she receives no additional treatment after surgery.

Still other considerations come into the decision. The same set of trials also included a large study of treatment with the estrogen-inhibitor tamoxifen. This too produced a significant reduction in the number of deaths, although perhaps not as large as the reduction brought about by chemotherapy. (It is interesting that tamoxifen seemed to be most effective in women over 50, whereas cytotoxic chemotherapy seemed to be most effective in younger patients.) Since tamoxifen generally produces only minor side effects, its widespread use could be advocated more readily than the widespread use of cytotoxic drugs. The issue now is to decide whether cytotoxic chemotherapy has anything to offer patients with breast cancer that cannot be achieved with tamoxifen. Judging from the available trials, the likeliest answer is that many young patients can be benefited, but for patients over 50 the combination of chemotherapy and tamoxifen does not seem to produce better results than tamoxifen alone.



INCREASE IN SURVIVAL of cancer patients occurs if the definition of what constitutes a cancer is widened to include conditions that are not destined to be fatal. This appears to be the explanation for the reported increase in survival of men with cancer of the prostate. Between the late 1940's and the late 1970's the mortality rate in the U.S. (*black dots*) has remained steady at about 210 deaths per million men per year. In the same period, however, the number of reported cases has risen from about 400 per million per year to about 700. (The colored dots show the incidence recorded in various national surveys and the colored line shows the increase recorded in Connecticut.) The increase in incidence is presumably the reason the five-year survival of men with cancer of the prostate has increased from 43 to 63 percent over 30 years although trials have not shown any major advance in treatment.

The role of chemotherapy in the treatment of the other major cancers of adults is much less well documented. Various trials have suggested that in adults some cancers such as ovarian cancer do sometimes respond to chemotherapy. In addition chemotherapy and local irradiation can be used to shrink cancers that arise in inaccessible sites such as certain regions of the head and neck. Overall, however, in terms of duration of survival the results have been more often negative than positive. One recent report, for example, described a trial of chemotherapy in the treatment of colon cancer. More than 600 patients who had received standard surgery were randomly allocated to various forms of adjuvant therapy. Roughly half of the patients received cytotoxic chemotherapy (fluorouracil and an alkylating agent), but their survival proved to be indistinguishable from the survival of the controls, who received no additional treatment. Of the roughly 190 patients who had chemotherapy and did not die of their cancer in the six-year trial period, one patient died from the immediate consequences of the treatment and seven died of leukemia.

In spite of these rather sobering findings several cytotoxic drugs are now commonly employed. The Connecticut Cancer Registry, for instance, reports that about a fourth of all cancer patients are recorded as having some form of chemotherapy during their initial stay in the hospital. The National Cancer Institute estimates that more than 200,000 patients receive chemotherapy in the U.S. each year. For a dangerous and technologically exacting form of treatment these are disturbing figures, particularly since the benefit for most categories of patients has yet to be established. Furthermore, the number of patients who are being cured can hardly amount to more than a few percent of those who are treated.

All told, adjuvant treatments now avert a few thousand (perhaps 2 or 3 percent) of the 400,000 deaths from cancer that occur each year in the U.S. Even without the invention of any additional drugs this figure might conceivably be pushed up to 5 percent; an extra 1 percent, for example, could possibly come from the treatment of breast cancer with cytotoxic agents. These are very real gains and a fitting memorial to the many thousands of patients who took part in the early trials of chemotherapy.

The fortitude and altruism of these patients have not, however, been matched by any comparable sense of responsibility on the part of those who determine national policies. By the 1960's cigarette smoking had been established as the major cause of lung cancer. Since then, however, few nations have made much of an effort to contain the further expansion of the tobacco industry. Unfortunately there are huge financial incentives for nations to sit back and do nothing. The cigarette is a readily taxable commodity; in the U.S. it provides the Federal and the state governments with about $6 billion a year. More important (at least for the British government, and perhaps also in the eyes of the U.S. Government), smoking cuts down the bill for old-age benefits because it reduces life span.

At the price of a slight increase in costs for health care the current smokers in the U.S. will on the average each have saved the U.S. Government about $35,000 in Social Security payments simply because they will on the average die sooner than nonsmokers; most of the deaths occur after retirement and are not from cancer but from cardiovascular disease and chronic lung disease, the incidence of which is also raised by smoking. The loss of life span represents a total saving of some $10 billion a year over the next half century or so.

Some countries have banned all tobacco advertising, and this has had an almost instant effect on tobacco sales. The failure of the U.S. Government to take such a step far outweighs all the advances made in the treatment of cancer since the advent of modern surgery. From 1953 on lung cancer has been the commonest fatal cancer in American males, and about now it is expected to surpass breast cancer and become the commonest fatal cancer in females. The waste of life is truly astonishing. Thanks to the cigarette, the U.S. now suffers a completely unnecessary additional 100,000 deaths per year from lung cancer. These numbers dwarf the 5,000 to 10,000 lives that are being saved by chemotherapy. So far the war against cancer is being lost because (to stay with the metaphor of war) we continue to tolerate the presence of a fifth column in our midst.

The conquest of the commonest of all lethal cancers depends, therefore, on the will power of governments and not on the skill of physicians or the ingenuity of scientists. Fortunately the affluent and better-educated are now smoking less than they used to. Because they tend to set the pace, the trend may eventually spread to the population as a whole.

For the other major cancers the issues are less clear-cut. In order of descending numerical importance these are cancer of the large intestine, breast, prostate and pancreas. Because each cancer is common in some countries and rare in others, each must be driven by external causes that are prevalent in some parts of the world and rare or absent in others. Even within the U.S. it is possible to find certain groups of people for whom the death rate from cancer is only about half the average national rate. Surely this proves that most forms of cancer are preventable.

This is not a very startling conclusion. None of the important causes of death has been primarily controlled by treatment. The death rates from malaria, cholera, typhus, tuberculosis, scurvy, pellagra and the other scourges of the past have dwindled in the U.S. mainly because humankind has learned how to prevent these diseases, not simply because they can be treated. Indeed, even cardiovascular mortality (the commonest of all forms of death in developed countries) has begun to decline in the U.S., suggesting that some change in circumstances or lifestyle is tending to prevent its occurrence. And so there are many grounds for believing that when any major disease is tackled on a national scale, the chief effort should be to prevent its occurrence. To put most of the effort into treatment is to deny all precedent.

Cancers of the cervix and of the liver are usually due to a viral infection, and each should be preventable by immunization. This could save 14,000 lives per year in the U.S. and perhaps as many as 500,000 in the world as a whole. The causes of most of the other important cancers are not yet known well enough for anyone to predict how or when they will be prevented. But eventually they will be, because they do have causes that await discovery.

The prospects for great advances in the treatment of cancer are not as obvious. The available cytotoxic drugs are not particularly discriminating in their action, being toxic for any rapidly dividing cell. Indeed, it is at first sight surprising that chemotherapy should ever be successful. It is important to remember, however, that the cancers most effectively treated by chemotherapy fall into rather special classes. First, there are the cancers that arise in cells left over from the process of embryogenesis (the cancers special to infants), in cells of the germ line (certain testicular and ovarian cancers) and in fetal cells trapped in the mother (choriocarcinoma of the placenta). These cancers have in common the unusual feature that they are cells in an alien environment. It is quite possible that the body has mechanisms for destroying such leftovers, particularly if chemotherapy has somewhat reduced their numbers.

The only other cancer readily curable by chemotherapy is Hodgkin's disease. This is a most unusual cancer because it is made up of a mixture of several types of cell and until recently was actually classified as some form of chronic infection. In short, the extreme peculiarity of the list of cancers known to be cured by cytotoxic drugs suggests that the present forms of chemotherapy may be seldom if ever sufficiently specific to kill every cell in a cancer and yet spare the normal tissues of the patient. Whether any of the common cancers can be cured by chemotherapy has yet to be established.

What then can be said about the long-term prospects? No one knows what new forms of chemotherapy may be invented, or when they will be invented. While such discoveries are awaited, more effort should be directed to certain proved forms of screening and much more effort to prevention. It seems bad cost-accounting for the Federal Government to subsidize chemotherapy of the common cancers of adults and not to subsidize the screening of women for breast cancer. Worse, it is surely an act of folly to pour hundreds of millions of dollars every year into giving a growing number of patients chemotherapy while doing virtually nothing to protect the population from cigarettes.

Cosmic Rays from Cygnus X-3

After decades of fruitless search astronomers have found a source of high-energy particles and gamma rays bombarding the earth. The object is thought to be a binary star at the edge of the galaxy

by P. Kevin MacKeown and Trevor C. Weekes

Few questions have perplexed astrophysicists for as long, or challenged their imagination as intensely, as that of the origin of cosmic radiation. The properties of the radiation have been revealed gradually over the decades since 1912, when its existence was first demonstrated by the Austrian physicist Victor F. Hess. It consists primarily of atomic nuclei stripped of electrons and moving at close to the speed of light. The nuclei are so energetic that the power generated in our galaxy in cosmic rays is thought to be far greater than the output radiated in, say, X rays or radio waves. Yet whereas many of the most exciting advances in astronomy have come from the detailed analysis of X-ray and radio sources, until recently the source of cosmic rays was largely a matter of speculation. They seem to come from everywhere, raining down on the earth from all directions at a uniform rate.

Now at last a major source has been found. The object is called Cygnus X-3 because it is the third-brightest X-ray emitter in the constellation Cygnus; it was first observed by X-ray astronomers in the late 1960's. It gained notoriety in 1972 when it underwent a violent outburst in which the intensity of its radio emissions increased a thousandfold. Recently it has also been shown to be a source of high-energy gamma rays. It is the gamma rays that have identified Cygnus X-3 as a source of cosmic radiation: they can have been produced only by charged particles moving at close to the speed of light.

There is still no consensus on how the particles are accelerated to such high speeds; several models of the process have been proposed. Most workers agree that Cygnus X-3 is a binary star system lying at least 37,000 light-years away, roughly at the edge of the galaxy. In visible light it cannot be seen even with the largest telescopes, because dust clouds in the spiral arms of the galaxy obscure it from view. The emissions observed at other wavelengths, however, indicate that it is intrinsically one of the two or three most luminous objects in the Milky Way.

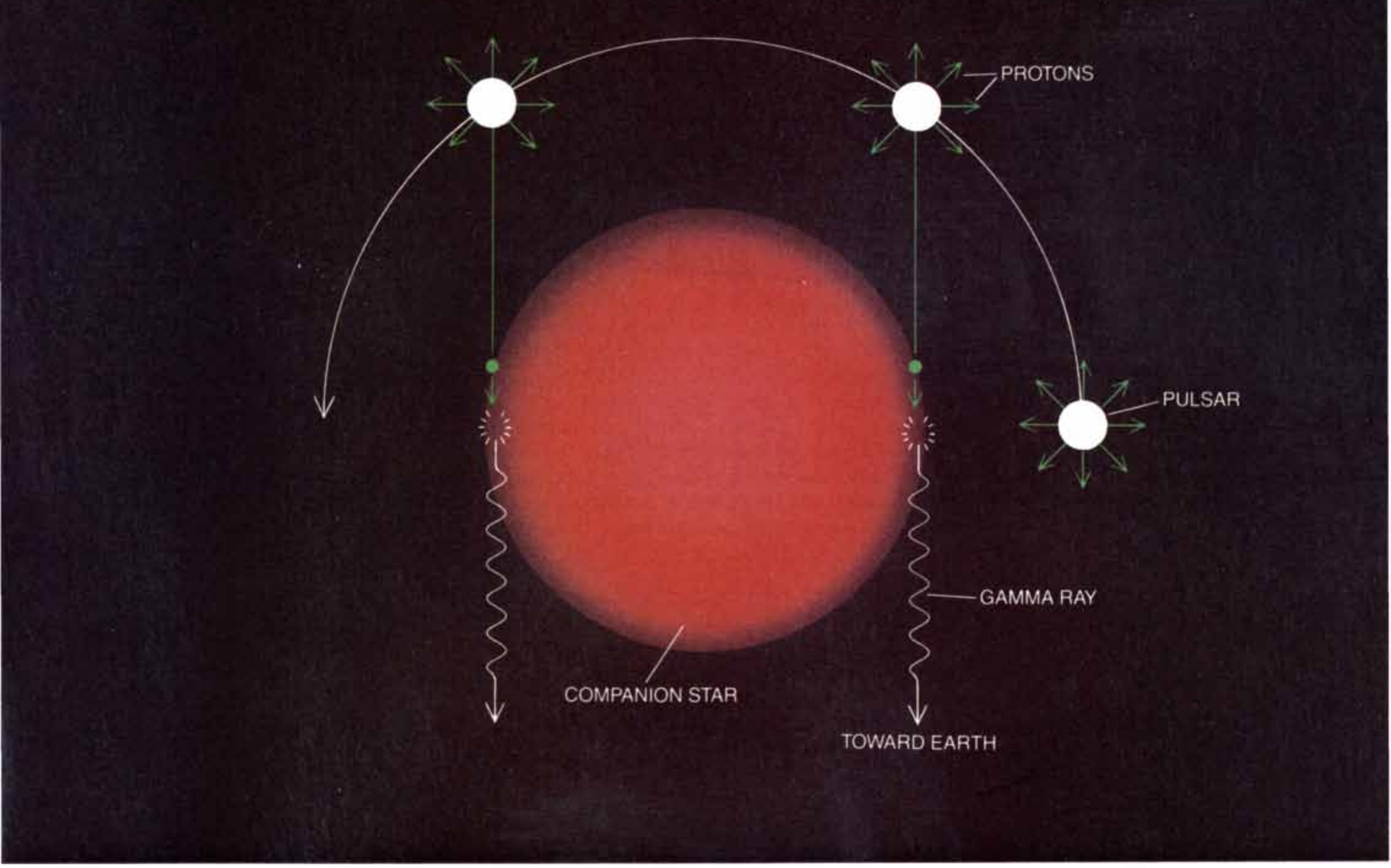
Any source of cosmic radiation would have to be a powerful particle accelerator to account for the tremendous energy of the individual cosmic rays. The most energetic of them have energies greater than 10^{20} electron volts. (That is 100 million times more than the energy expected to be imparted to particles in the proposed Superconducting Supercollider, which would be the most powerful manmade particle accelerator.) The other end of the cosmic-ray energy spectrum is defined somewhat arbitrarily: any quantum greater than 10^{8} electron volts arriving from space is considered a cosmic ray. The definition encompasses not only particles but also gamma-ray photons, which are quanta of electromagnetic radiation.

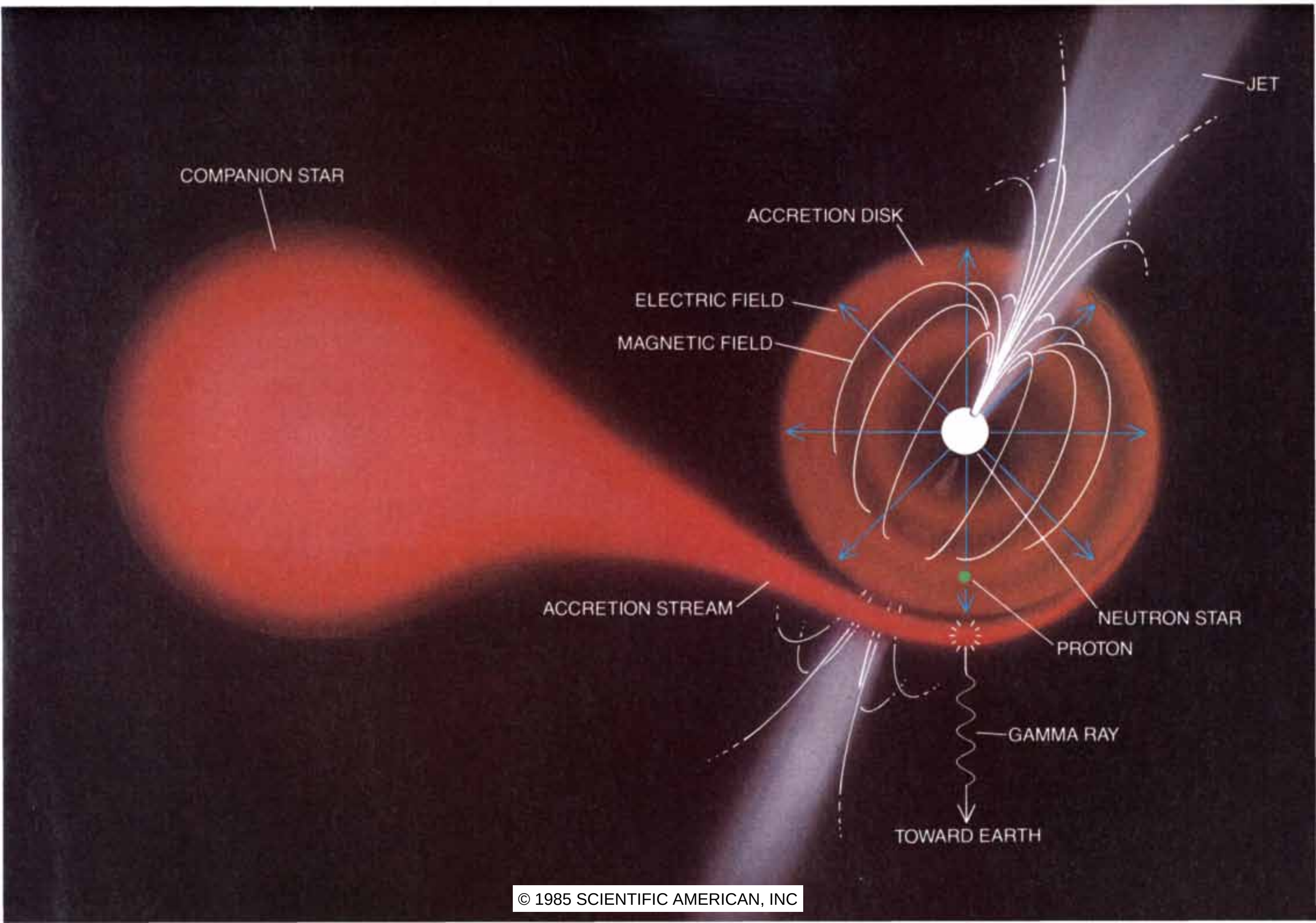
Fewer than .1 percent of all cosmic rays are gamma rays, however; the rest are particles. Electrons account for about 1 percent of the total flux, and a few positrons (anti-electrons) and antiprotons are also present, but the vast majority of cosmic rays are atomic nuclei. Some 92 percent of them are hydrogen ions, which are simply individual protons stripped of their electrons. Helium, the second-lightest element, accounts for about 6 percent of the flux. The other common elements are present in much smaller amounts.

The most important fact concerning the chemical composition of cosmic radiation is that it closely resembles, over most of the observed energy range, the composition of the galaxy as a whole. The point may seem trivial, but actually it eliminates several hypotheses about the origin of cosmic rays. For example, they cannot all be relics of the big bang, which produced hydrogen almost exclusively. Nor can they all come from old, highly evolved stars, because then a much higher fraction of them would be expected to be heavier nuclei. (The elements heavier than helium are thought to have been synthesized inside stars and during supernova explosions by means of fusion reactions.) Cosmic rays either must come from an object of typical composition or must originate in a large number of quite different objects that together produce an average mixture of elements.

Other observations exclude the possibility that the cosmic radiation bombarding the earth today could be the result of a single local, catastrophic event, such as a solar outburst or the explosion of a nearby supernova. First, an examination of the ionization trails

MODELS OF CYGNUS X-3 assume that it is a binary star system in which one component is a dense neutron star. In one model the neutron star is a rapidly spinning pulsar (*top*). The pulsar accelerates protons to cosmic-ray energies and ejects them in all directions. Some protons strike gas nuclei in the outer layer of the companion star, generating high-energy gamma rays that continue along the proton trajectory. The gamma rays are detected on the earth during two phases of the pulsar's orbit. In the accretion model (*bottom*) the neutron star's gravitational field draws matter off the companion star, and its magnetic field induces an intense electric field in the rotating disk of accreted matter. Protons are accelerated to cosmic-ray energies along the electric field lines; some collide with gas nuclei in the accretion stream and produce gamma rays. High-energy gamma rays from Cygnus X-3 have identified it as a source of cosmic radiation at least 37,000 light-years away.


PROTONS
PULSAR
GAMMA RAY
COMPANION STAR
TOWARD EARTH


JET
COMPANION STAR
ACCRETION DISK
ELECTRIC FIELD
MAGNETIC FIELD
ACCRETION STREAM
NEUTRON STAR
PROTON
GAMMA RAY
TOWARD EARTH

left in meteorites by charged cosmic-ray particles suggests that the density of the rays in the solar system has remained fairly constant over the past billion years. The ionization trails are visible as defects in a meteorite's crystal structure; the number of them, along with the age of the meteorite (which is known from radioisotope dating), yields the average rate at which the meteorite has been struck by cosmic rays.

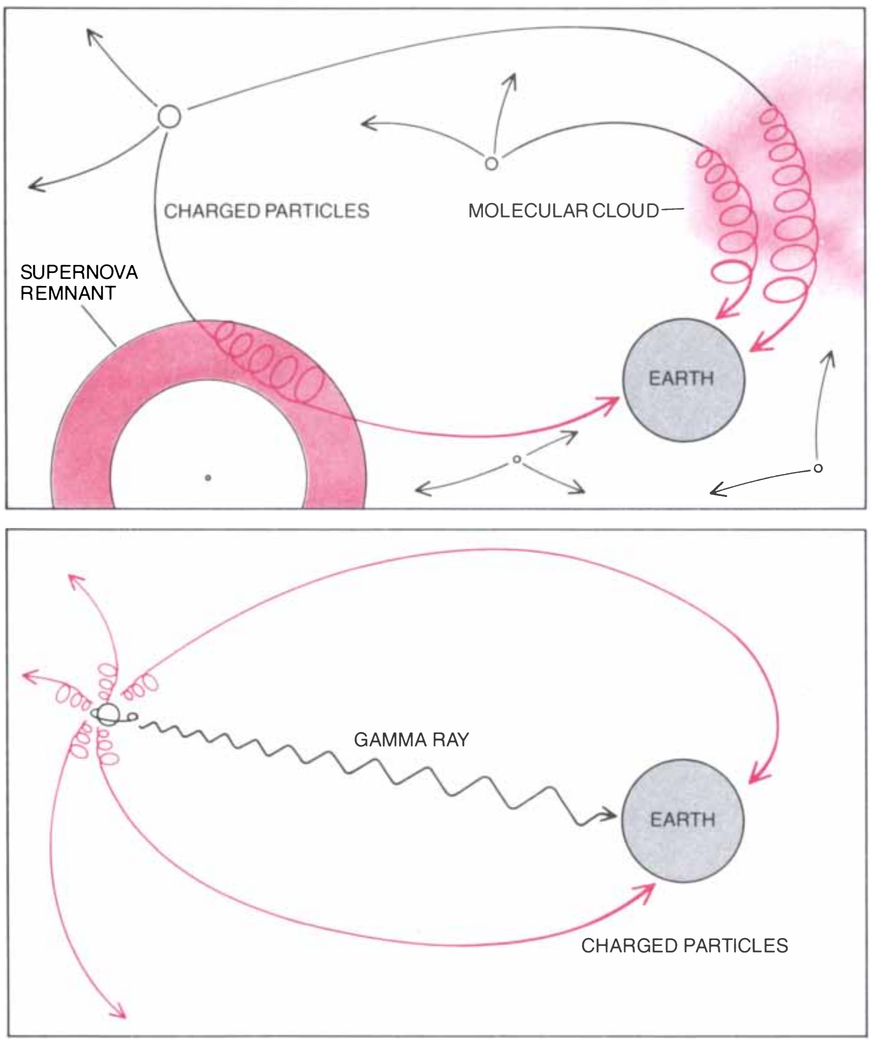
Second, the cosmic-ray density in the solar system appears to be typical of the entire galaxy. The evidence for this conclusion is indirect: it comes from the observation throughout the galaxy of synchrotron emissions at radio wavelengths. Synchrotron radiation, which has a characteristic polarization, is generated primarily by relativistic electrons moving in a magnetic field. A relativistic electron in space is a cosmic ray; by assuming that electrons and other cosmic-ray particles are present in synchrotron-emission regions in the same proportions observed near the earth, one can arrive at a rough estimate of the overall cosmic-ray density in distant parts of the galaxy.

If one knows how long high-energy particles travel in space before colliding with interstellar gas nuclei, one can even calculate the rate at which the galaxy must be producing cosmic rays in order to maintain the observed density. Estimates of the cosmic-ray lifetime are based on a detailed examination of the composition of the radiation. Although the composition is broadly similar to that of the galaxy as a whole, the light elements lithium, beryllium and boron are present at enhanced levels. These elements are not expected to be produced by fusion reactions in stars. They almost certainly result from the fragmentation of heavy elements that have collided with stationary gas atoms in interstellar space. From their relative abundance in cosmic radiation and from an estimate of the density of interstellar gas, one can calculate how long the average cosmic ray has been in space: about 20 million years.

This figure, together with the observed energies, leads to an estimate of the power output of the galaxy in cosmic radiation. Obviously any such calculation contains large uncertainties, but the correct value is thought to lie in the range of 10^{40} to 10^{41} ergs per second. The galaxy generates less energy in the form of cosmic rays than it radiates as visible light (10^{44} ergs per second) but more than it radiates as radio waves (10^{39} ergs per second) and X rays (2×10^{39} ergs per second) combined. Whatever the source of cosmic rays may be, it is clear they are produced at a prodigious rate.

Historically there have been two major schools of thought on the origin of cosmic radiation. The first view, originally put forward by the physicist Enrico Fermi, holds that cosmic-ray nuclei are injected into space with comparatively low energies and then are accelerated to high velocities through collisions with clouds of magnetized gas, shock fronts from supernova explosions or other large, energetic features of the interstellar medium. Low-energy nuclei are known to be ejected by ordinary stars; according to the Fermi model, such nuclei might acquire cosmic-ray energies step by step, in a long random walk through space. Thirty years ago this "distributed acceleration" model was widely favored because it avoided the need for postulating exotic and as yet undiscovered objects capable of imparting relativistic velocities to particles all at once.

In recent years the Fermi model has fallen out of favor. Detailed calculations have shown that the energies of nuclei are likely to increase enough on their journey through interstellar space to account for the observed cosmic-ray energies. Since Fermi first proposed the model, moreover, observations of the exotic and the extremely energetic have become commonplace in astronomy. No one in 1950 could have foreseen the discovery of, say,



ORIGIN OF COSMIC RAYS has been explained in two ways. According to the Fermi model (*top*), cosmic rays are produced in interstellar space. Low-energy particles emitted by ordinary stars are accelerated to cosmic-ray velocities when they collide with moving clouds of magnetized gas or with expanding shock fronts from supernova explosions. An alternative view is that cosmic rays come from a small number of exotic objects capable of accelerating particles to high energies (*bottom*). The charged nuclei forming most of the radiation are deflected by the galactic magnetic field, and so their directions do not indicate their source. In contrast, gamma rays are electrically neutral and travel straight from the source.

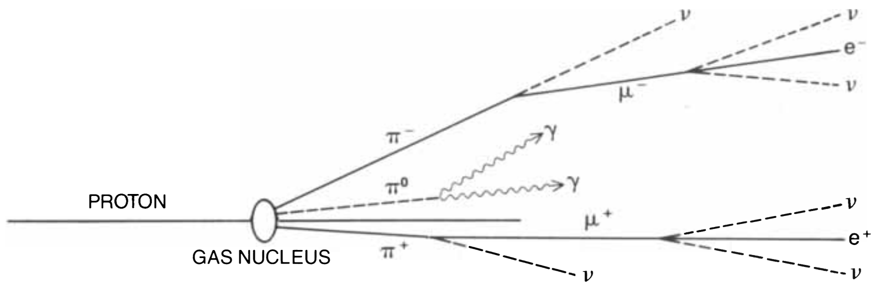
pulsars: dense, rapidly spinning neutron stars with magnetic fields trillions of times as intense as the magnetic field of the earth.

Today many investigators adhere to a second school of thought: the notion that cosmic rays receive all their energy from discrete sources, with interstellar space acting only as a diffusive medium. Many types of source have been proposed, including pulsars, supernova explosions and T Tauri stars (young, variable stars that sometimes undergo rapid increases in brightness). Some of these objects are numerous enough to account for the total energy contained in cosmic radiation. They seem, however, to be incapable of producing the entire energy spectrum.

A number of workers have therefore argued that cosmic rays, or at least the most energetic ones, must come from outside our galaxy. Synchrotron emission from other galaxies shows that they too contain cosmic rays. A relatively small number of galaxies have violently active cores, some of which expel particles in relativistic jets; quasars are thought to be extreme examples of such "active galaxies." An active galactic core might conceivably accelerate particles to energies that are beyond the reach of an individual star or a binary in our own galaxy. On the other hand, this does not imply, as some workers have suggested, that *all* cosmic rays must originate outside the Milky Way.

One might think the sources of cosmic rays could be identified easily enough by making a few simple observations with a cosmic-ray telescope. Such is not the case. Cosmic-ray nuclei come from all directions at almost exactly the same rate—not because the sources lie everywhere but because the particle trajectories are scrambled by the galactic magnetic field. Since the nuclei are electrically charged, they do not travel in straight lines through interstellar space; instead they are constrained to spiral along the lines of the magnetic field, about which little is known other than that it is disordered and chaotic. When the particles arrive at the earth after 20 million years of wandering in the galaxy, their paths are hopelessly tangled.

In order to pinpoint a source one must study components of the radiation that are relatively unaffected by magnetic fields. The obvious candidates are electrically neutral particles. Neutrons can be excluded from consideration, because they would decay to charged particles (protons and electrons) long before they could reach the earth from a cosmic-ray source.



GAMMA RAYS FROM CYGNUS X-3 are thought to be produced by the collision of high-energy protons from the neutron star with gas nuclei from the companion star. Some of the kinetic energy of the proton is converted into unstable particles called pions. Each charged pion produces a neutrino and a muon, which decays into two more neutrinos and an electron or a positron. The neutral pion decays into two highly energetic gamma rays.

Neutrinos are more promising; they would certainly be generated in the nuclear reactions occurring at a cosmic-ray source, and their ability to penetrate obstacles without being absorbed is well known. Unfortunately the same property makes them very difficult to detect. As yet there are no telescopes capable of detecting high-energy neutrinos from cosmic sources, although such devices are planned.

The approach taken by us and by other workers—one that has now borne fruit with the confirmed detection of cosmic radiation from Cygnus X-3—is to observe gamma-ray photons, which like neutrons and neutrinos are electrically neutral. An object that accelerates particles to cosmic-ray velocities will almost inevitably produce gamma rays as well; conversely, high-energy gamma rays are emitted only by relativistic particles. Gamma rays are therefore an effective probe of the source of cosmic radiation even though they account for less than .1 percent of the total flux. At most wavelengths, moreover, they interact very little with the thin interstellar gas, so that they arrive in the solar system virtually unattenuated.

Cosmic gamma rays do not, however, reach the surface of the earth. Before they have penetrated more than a fortieth of the dense atmosphere they interact with gas nuclei. Hence they can be detected directly only by instruments carried above the atmosphere. Furthermore, satellite instruments are limited in size, and so they cannot detect fluxes smaller than one photon per square meter per month. The flux of the gamma rays whose origin is most puzzling, those with energies greater than 10^{12} electron volts, is below that limit. They cannot be studied directly.

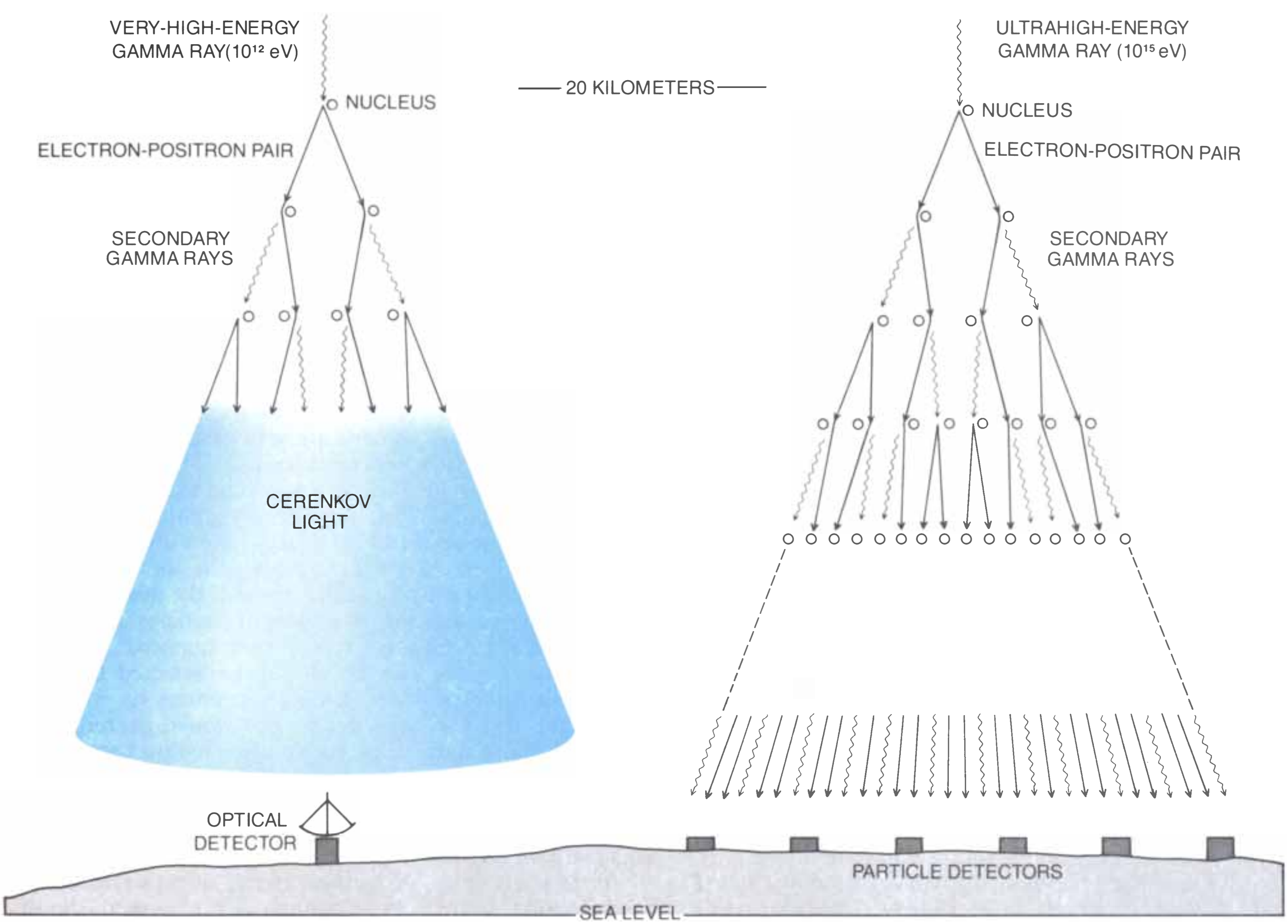
Fortunately these gamma rays can be studied indirectly from the ground because they initiate observable showers of secondary particles and photons in the atmosphere. An air shower (also called an electromagnetic cascade) begins when a cosmic gamma ray interacts with the electric field of a gas nucleus in the upper atmosphere, typically at an altitude of 20 kilometers. The energy of the gamma ray is converted into matter: usually into an electron and a positron, each of which carries roughly half of the energy of the gamma ray. After a short distance each of the two fast-moving particles is itself deflected in the electric field of a gas nucleus, and some of its energy is radiated as a gamma-ray photon in the process known as bremsstrahlung (a German word meaning braking radiation). The secondary gamma rays then produce more electron-positron pairs. In this way the energy of the primary gamma ray is shared among a geometrically increasing number of particles and photons. The cascade peters out when the energies of the individual quanta are so low that other absorptive processes compete with bremsstrahlung and pair production.

If the primary gamma ray is an ultrahigh-energy photon (10^{14} electron volts or more), the air shower may reach the ground. By then it consists of thousands of particles and gamma rays spread over hundreds of square meters. It can be detected by an array of Geiger counters (in which the charged particles ionize gas molecules, allowing a current to flow between two electrodes) or scintillation counters (in which flashes of light triggered by the particles in a suitable medium are registered by photomultiplier tubes). When the arrival direction of the primary gamma ray is vertical, the secondary particles strike all the detectors in an array simultaneously. If the primary gamma ray strikes the atmosphere at an angle, its arrival direction can be deduced from small differences, typically less than one ten-millionth of a second, in the arrival times of shower particles at different detectors. The energy of the primary gamma ray can be estimated from the number of particles striking the detectors.

The air shower precipitated by a very-high-energy gamma ray (between 10^{11} and 10^{14} electron volts) does not reach the earth's surface. Before the shower particles are absorbed in the atmosphere, however, they trigger flashes of visible bluish light that can be detected from the ground. The light, called Cerenkov radiation, is generated whenever a particle moves through a medium at a speed exceeding the speed of light in the medium; it is actually emitted by electrons in the medium itself, which are disturbed by the passage of the particle. Cerenkov radiation is an electromagnetic analogue of the bow wave generated by a power boat moving faster than the speed of wave propagation in water. It is beamed along the particle trajectory and reaches the ground as a disk of light. Because the light is in the visible range, it is attenuated only slightly by the atmosphere.

On moonless nights flashes of Cerenkov light can be detected with a simple, inexpensive system consisting of a mirror that focuses the light onto a phototube. More elaborate devices, including the one in use at the Fred Lawrence Whipple Observatory in Arizona, consist of many mirrors assembled into a dish that focuses light onto an array of phototubes. Such devices not only detect Cerenkov light but also map its intensity. The arrival direction of the cosmic gamma ray, which corresponds roughly to the axis of the Cerenkov light disk, can then be determined to within less than a quarter of a degree of arc. In contrast, the resolution of current particle-detector arrays is only about one degree of arc. Because particle detectors can be used 24 hours a day, however, and not just on dark nights, they remain the principal method for observing cosmic gamma rays with energies of 10^{15} electron volts or more.

Both types of device suffer from a serious weakness, in that they do not allow one to say of a particular air shower that it was initiated by a gamma ray. The problem is that charged cosmic-ray particles interacting with atmospheric nuclei set off showers similar to those caused by gamma rays. Yet as a result of the influence of the galactic magnetic field, the flux of charged cosmic rays is essentially equal in all directions. A source of cosmic gamma rays may therefore reveal itself as a directional anisotropy: an excess of air showers coming from a particular direction. The anisotropy would necessarily be minute, because charged cosmic-ray particles account for about a thousand times as many air showers as gamma rays do. That is one of the reasons it has required years of painstaking gathering and analysis



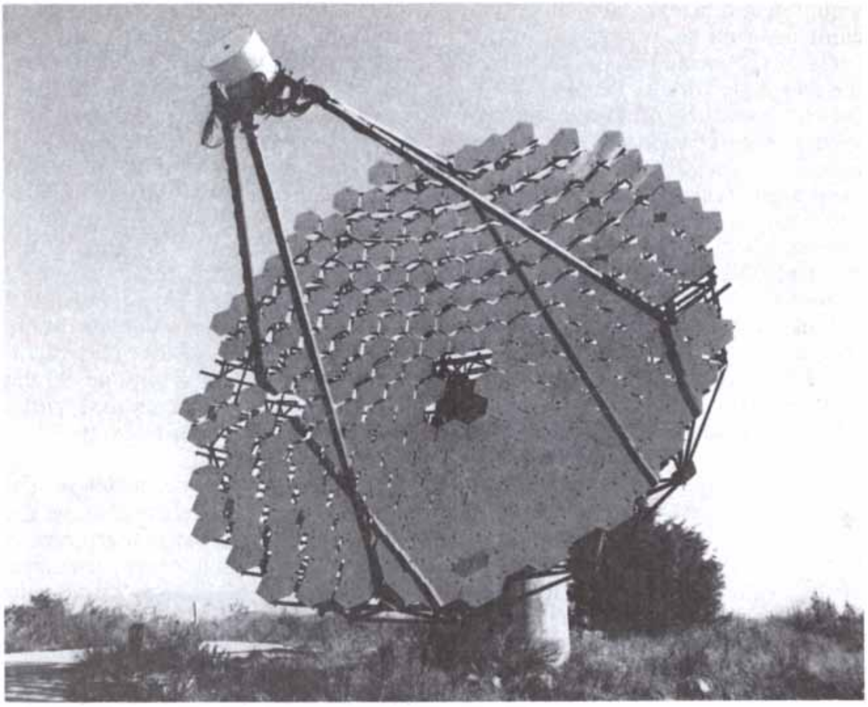
AIR SHOWERS triggered by cosmic gamma rays make it possible to determine the arrival direction of the gamma rays. A gamma ray does not usually reach the earth's surface; instead it interacts with an atmospheric nucleus at an altitude of about 20 kilometers. The energy of the gamma ray is converted into an electron and a positron. Each particle emits a secondary gamma ray when it is deflected in the electric field of another nucleus. The secondary ray then gives rise to another electron-positron pair, and the process continues until the energy of the primary ray is dissipated. An air shower triggered by an ultrahigh-energy gamma ray may reach the ground, where the particles can be observed with Geiger counters or other detectors. The air shower from a very-high-energy gamma rays is absorbed in the atmosphere, but some fast-moving secondary particles generate visible Cerenkov light that is detectable on the ground.

of data to identify a source of cosmic rays.

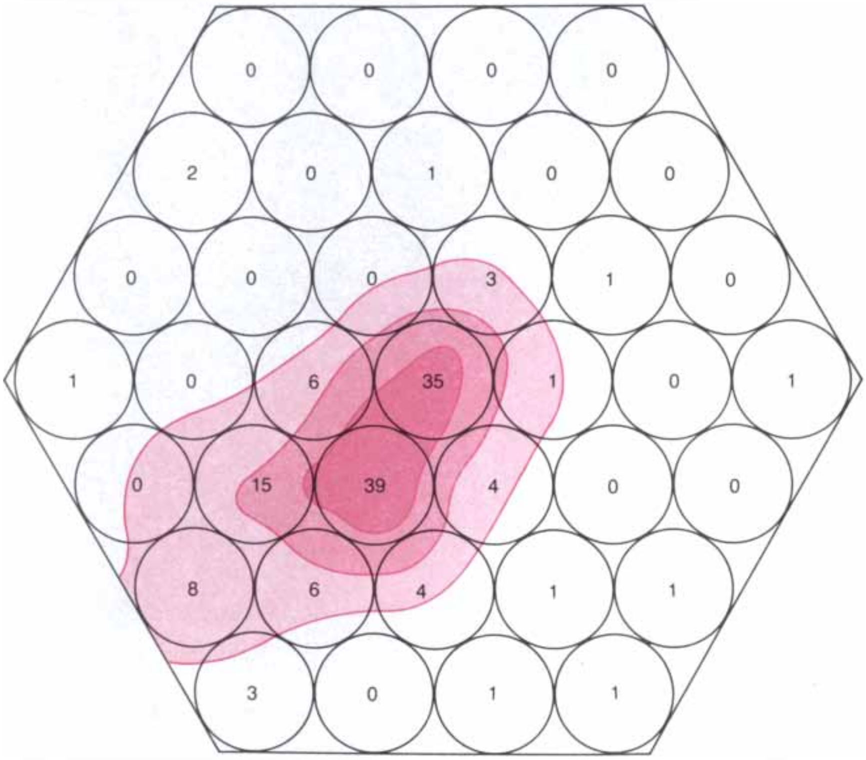
The story of Cygnus X-3 begins well before the recent recognition of its importance as a source of cosmic rays. It was discovered as an X-ray source in 1966 by Riccardo Giacconi, Paul Gorenstein, Herbert Gursky and J. R. Waters, who were working at American Science and Engineering, Inc., in Cambridge, Mass. With the help of a rocket-borne telescope (required because X rays too are absorbed by the atmosphere) the workers identified three X-ray sources in Cygnus. Although the emissions from Cygnus X-3 were considerably weaker than those from the other two sources, they did hold a clue to its energetic nature. The emissions were deficient in low-energy X rays, which are preferentially absorbed by interstellar gas. The deficiency suggested that Cygnus X-3 is much more distant and hence that it must be intrinsically brighter than the other two sources.

When the National Aeronautics and Space Administration launched its first satellite X-ray telescope (*Uhuru*) in 1970, many X-ray signals, including the one from Cygnus X-3, were found to vary with time. The Cygnus X-3 signal varied cyclically with a period of 4.8 hours. The observation suggested that the object consisted of two stars locked in a binary orbit; the periodic drop in the intensity of the emissions was attributable to the passage of one star in front of the other. Equally important, at a time when the position of X-ray sources could be specified only crudely, the periodicity of Cygnus X-3 provided a way to determine whether it radiated at other wavelengths. Infrared studies of the Cygnus area subsequently revealed a point source with the distinctive 4.8-hour period. The more precise positional fix yielded by the infrared studies in turn enabled workers to identify Cygnus X-3 as a variable (although not a periodic) radio source.

Then on September 2, 1972, a fortuitous discovery ended the relative obscurity of Cygnus X-3 by showing just how dramatically variable its radio emissions could be. Early in the evening, while waiting for his main target to rise above the horizon, Philip C. Gregory, working at the Algonquin Radio Observatory in Ontario, decided to pass the time by pointing his radio telescope at Cygnus X-3. Ordinarily a weak radio source, it was on that night one of the brightest objects in the sky, radiating at a thousand times its normal intensity. An equivalent phenomenon for an optical astronomer



CERENKOV-LIGHT TELESCOPE used by the authors and their collaborators at the Fred Lawrence Whipple Observatory in Arizona is known as the 10-meter reflector. It consists of 248 mirrors that focus light onto a hexagonal array of 37 phototubes. On dark nights the instrument can easily detect the flashes of Cerenkov light set off by air showers.



CERENKOV-LIGHT MAP of an air shower was made with the 10-meter reflector at the Whipple Observatory. The circles represent the phototubes. The numbers are proportional to the intensity of the light recorded by each phototube; the colored lines are intensity contours. From such a map the axis of the air shower and hence the arrival direction of the cosmic ray can be calculated. In this case the axis of the shower is parallel to the axis of the detector and displaced from the detector toward the lower left. (The map records only a part of the edge of the conical shower, which may be hundreds of meters in diameter.) Cosmic gamma rays from Cygnus X-3 are found as an excess of showers from that direction.

would be a dim star that suddenly became as bright as Jupiter.

Gregory immediately telephoned his colleague Robert M. Hjellming at the National Radio Astronomy Observatory in Green Bank, W.Va. Hjellming confirmed the observation. That same night, fearing the outburst might end before it could be adequately documented, the two workers called as many other astronomers as they could. Within a few days news of the Cygnus X-3 outburst had spread to nearly every observatory in the world. Regular observing programs were interrupted, and telescopes of all types—radio, infrared, optical, X-ray and gamma-ray—were aimed at Cygnus. Never before had there been such a concentration of effort on a single object.

The complex radio signal proved to be particularly informative. For one thing, the signal contained absorption lines attributable to specific clouds of interstellar hydrogen whose approximate locations in the galaxy are known. (Hydrogen absorbs radiation at a wavelength of 21 centimeters; because the interstellar clouds are moving at different velocities, each cloud has a characteristic absorption line that is Doppler-shifted from the 21-centimeter line by a different amount.) The absorption lines established a lower limit on the distance of Cygnus X-3: it must be more distant than the farthest hydrogen cloud, which is believed to lie about 37,000 light-years away, near the edge of the galaxy. From the minimum distance and the observed X-ray luminosity of Cygnus X-3 one can calculate its intrinsic X-ray luminosity. The calculation confirms that it is one of the brightest sources in the galaxy, with a power output of at least 2×10^{37} ergs per second.

Furthermore, the spectrum of the radio emissions observed during the 1972 outburst indicated that Cygnus X-3 was ejecting high-energy particles. Several days after Gregory's discovery the radio emissions reached a peak of intensity. They then began to wane in a distinctive way: the emissions at the longest wavelengths, which correspond to photons of the lowest energies, persisted for the longest time. The pattern is characteristic of synchrotron radiation from relativistic electrons injected into a weak magnetic field. As the electrons spiral around the field lines they emit photons, initially at short wavelengths but eventually, after much of their energy has been radiated, only at longer ones. The decline in intensity of the radio emissions from Cygnus X-3 was interrupted by new bursts of radio activity, including a major flare-up about a month after the initial outburst; apparently the ejection of energetic particles was an ongoing process.

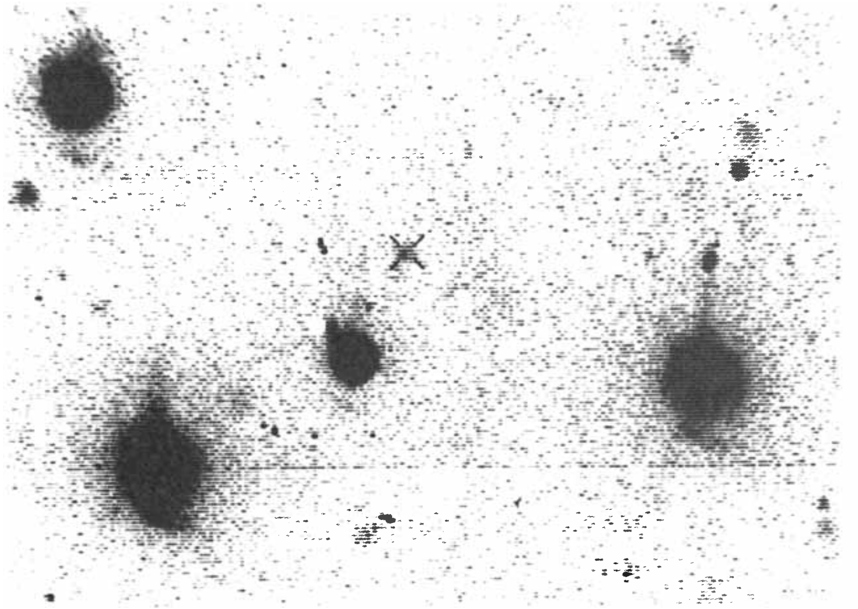
Indeed, 10 years later, in 1982, another giant outburst was recorded. Robert Geldzahler and his colleagues at the Naval Research Laboratory took advantage of the event and of the newly completed Very Large Array radio interferometer in New Mexico to make the first maps that showed Cygnus X-3 to be anything other than a point source. The radio-emission region was elliptical, and as the outburst progressed the ellipse gradually became more elongated. Geldzahler and his co-workers interpreted the elongation as evidence that Cygnus X-3 was emitting jets of particles at velocities of about a third the speed of light. Radio jets emanate from the cores of a number of active galaxies, but only one other source in our galaxy has been observed to emit them: the strange object called SS 433.

X-RAY IMAGE OF CYGNUS X-3 was prepared by Rick F. Harnden, Jr., of the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory from data gathered by the orbiting Einstein Observatory. Cygnus X-3, which was discovered as an X-ray emitter, is the largest and brightest source on the image; it is one of the brightest sources in the galaxy. The other sources are all young, massive stars.

A source of high-energy particles would be expected to emit gamma rays; hence it is not surprising that gamma rays from Cygnus X-3 were first detected during the 1972 radio outburst. Using a balloon-borne instrument, A. M. Galper and his colleagues at the Moscow Engineering Physical Institute observed a strong flux of low-energy gamma rays from the direction of Cygnus. The energy of the gamma rays was about 4×10^7 electron volts. Later in 1972 the NASA satellite *SAS-2* recorded gamma radiation with an energy of 10^8 electron volts. In analyzing the signal Richard C. Lamb, Carl E. Fichtel, Robert C. Hartmann, Donald A. Kniffen and David J. Thompson of NASA's Goddard Space Flight Center found a 4.8-hour periodicity that clearly identified the signal as coming from Cygnus X-3.

Nevertheless, it was still not obvious that Cygnus X-3 was very pertinent to the problem of the origin of cosmic rays. Gamma rays with energies of 10^8 electron volts are at the low end of the cosmic-ray spectrum; conceivably they might be generated by processes not associated with high-energy particles. To be considered a significant source of cosmic rays an object would have to emit gamma rays of at least 10^{12} electron volts.

The first such gamma rays from Cygnus X-3 were observed by Arnold A. Stepanian and his colleagues at the Crimean Astrophysical Observatory. Since 1970 they had searched for galactic sources of very-high-energy gamma rays by detecting Cerenkov light from air showers; their equipment consisted of four army-surplus searchlight mirrors with phototubes at the foci. On hearing news of the Cygnus X-3 outburst in September, 1972, the Crimean workers aimed their mirrors toward Cygnus. In 11 nights of observation they detected a strong flux of 10^{12} electron-volt gamma rays. The signal exhibited the characteristic 4.8-hour periodicity, although the peak was sharper than the peaks at other wavelengths and also came at a different phase of the cycle.



NEAR-INFRARED IMAGE OF CYGNUS X-3 (*bars*) was made by one of the authors (Weekes) and John C. Geary of the Center for Astrophysics using a charge-coupled device attached to the Multiple Mirror Telescope, on Mount Hopkins in Arizona. The visible light from Cygnus X-3 is absorbed by intervening dust clouds in the galaxy, but its infrared emission suggests that it is intrinsically fairly bright in the visible range as well.

At this point Cygnus X-3 should have attracted the attention of cosmic-ray theorists, but it did not. The results of the Crimean group were largely ignored, perhaps in part because they were not published in the issue of *Nature* in which most of the papers on the outburst appeared. The fact that the prevailing theoretical models of X-ray sources did not provide for the emission of very-high-energy gamma rays may also have contributed to the neglect. In defense of the theorists one must acknowledge that observations of air showers had produced some false alarms in the past.

The Crimean workers nonetheless continued their observations. In every observing season between 1972 and 1979 they measured the signal from Cygnus X-3; the amplitude declined and the shape of the light curve (the time variation of the emissions) changed, but the signal was always there. Finally, in 1980, the Crimean results were independently confirmed. Sean Danaher, David Fegan and Neil A. Porter of University College in Dublin, collaborating with one of us (Weekes) at the Whipple Observatory, detected the emissions from Cygnus X-3 with a small Cerenkov telescope. In the following two years observations by two other groups established beyond a doubt that Cygnus X-3 is a source of 10^{12} electron-volt gamma rays and thus of high-energy particles.

Since then data gathered by particle-detector arrays have shown that Cygnus X-3 also emits ultrahigh-energy gamma rays. Between 1976 and 1980 workers at the University of Kiel in West Germany kept such an array in continuous operation. Their primary purpose was not to search for a cosmic-ray source; instead it was to study high-energy interactions in air showers generated by cosmic rays in the energy range from 10^{15} to 10^{16} electron volts. In order to determine the distribution of the secondary particles, the workers had to measure the arrival direction of cosmic rays to within about one degree of arc. As a result their data base was also well suited for identifying directional anisotropies in the cosmic-ray flux.

In 1983 Wilhelm Stamm and Manfred Samorski undertook the laborious task of searching through five years of data for evidence of gamma-ray emissions from Cygnus X-3. They found a clear excess of air showers originating in the correct general direction, but the margin of error in the determination of arrival directions was too large to allow a definite statement that Cygnus X-3 was the source of the signal. The conclusive evidence lay in the arrival times of the showers. When Stamm and Samorski analyzed the times for the showers originating in the direction of Cygnus X-3, the familiar 4.8-hour periodicity emerged.

Soon after these results were reported they were confirmed by a group at the University of Leeds. The angular resolution of the array operated by the Leeds investigators from 1978 to 1982 was not as good as that of the Kiel experiment, but it was good enough to detect a strong signal from Cygnus X-3. In addition the Leeds group determined that Cygnus X-3 shows no detectable emissions at energies above 10^{16} electron volts.

The object is nonetheless extraordinarily luminous at ultrahigh energies. Its observed flux in the energy range from 10^{15} to 10^{16} electron volts is 6×10^{36} ergs per second, which is on the order of a thousand times greater than the output of the sun at all wavelengths. Moreover, gamma rays of energies near 10^{15} electron volts are preferentially absorbed in interstellar space; they are thought to be annihilated in collisions with the low-energy microwave photons that pervade the universe as a result of the big bang. Hence the flux actually emitted by Cygnus X-3 must be larger still. Assuming Cygnus X-3 lies at the minimum distance of 37,000 light-years, its output in ultrahigh-energy gamma rays must be about three times more than the observed flux, or roughly 2×10^{37} ergs per second.

Furthermore, gamma rays are only a small part of the total flux of cosmic radiation from Cygnus X-3. The various theoretical models of the object agree that the gamma rays are probably emitted in high-energy interactions of charged nuclei, mainly protons. In such interactions only some 3 percent of the released energy would go into gamma rays; the rest would be carried away by electrons, positrons and neutrinos. If Cygnus X-3 is radiating 2×10^{37} ergs per second in ultrahigh-energy gamma rays, it must be ejecting at least 30 times as much energy, or

6×10^{38} ergs per second, in the form of charged cosmic-ray particles.

The estimate is conservative; because it is based on the gamma rays observed arriving at the earth, it does not take into account the likelihood that Cygnus X-3 emits cosmic rays in all directions. Since the galactic output in cosmic radiation is only between 10^{40} and 10^{41} ergs per second, a small number of sources like Cygnus X-3—say a few dozen—would explain most of the observed flux of cosmic rays. Only the cosmic rays with energies above Cygnus X-3's apparent limit of 10^{16} electron volts would remain to be explained. They may well come from other galaxies.

What makes Cygnus X-3 such a powerful particle accelerator? Because of its periodic emissions, most workers agree it must be a binary star system. X-ray binaries are not uncommon in the galaxy. They are thought to consist of a dense, collapsed star orbiting a massive normal star. If the orbit is close enough, the intense gravitational field of the dense star can pull matter off the surface of its companion. As the gaseous matter falls toward the dense star it is accelerated to tremendous velocities; collisions between molecules heat the gas to tens of millions of degrees Celsius, and some of the thermal energy is radiated as X rays.

In the case of Cygnus X-3 the energy needed to accelerate particles to cosmic-ray velocities must also come from the dense star, which is probably a neutron star. A neutron star is more massive than the sun but is only about 10 kilometers in diameter; it is the collapsed core of a massive star whose outer shell has been blown off in a supernova explosion. The collapse compresses the star's magnetic field, increasing its intensity by as much as a trillion times. The star's rotation rate also increases, because its angular momentum is conserved during the collapse. In the ensuing millenniums the spinning star may gradually slow down.

In a model of Cygnus X-3 proposed by Thomas Vestrand of the University of New Hampshire and David Eichler of the University of Maryland at College Park the neutron star is still spinning rapidly: it is a pulsar. The energy imparted to the cosmic rays comes from the rotational kinetic energy lost by the pulsar as it slows down. Pulsars emit synchrotron radiation, and so they must be capable of accelerating electrons, at least, to relativistic energies. Although the details of the process are murky, the spinning magnetic field would induce an electric field that might then accelerate charged particles—protons as well as electrons—from the surface of the pulsar.

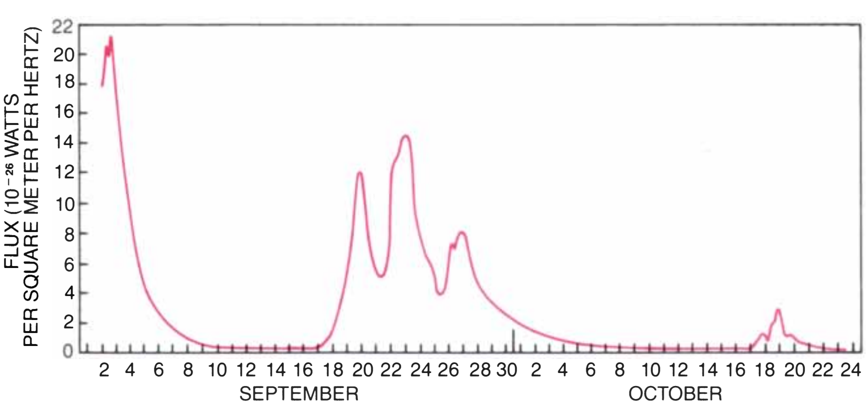
According to the model advanced by Vestrand and Eichler, the Cygnus X-3 pulsar ejects high-energy protons in all directions. Most of the protons escape into space but some enter the gaseous outer layer of the companion star. When a proton strikes a gas nucleus, it produces three kinds of pion, of which one kind decays into two gamma rays. The gamma rays travel close to the original trajectory of the proton, and the model predicts they will be beamed toward the earth at just two points on the pulsar's orbit around the companion star. This would explain the twin-peak light curve recorded in some observations of Cygnus X-3 at energies of 10^{12} electron volts. In addition A. Michael Hillas of Leeds has shown that the pulsar model can account for the observed intensity of the gamma rays emitted by Cygnus X-3 at all energies from 10^{12} to 10^{16} electron volts—provided the pulsar accelerates protons to an energy of 10^{17} electron volts.

That energy is 100 times greater than what is ordinarily considered the upper limit of a pulsar's acceleration capability. The Vestrand-Eichler model requires pulsars to be much more efficient at accelerating particles than they are generally thought to be. It also assumes the pulsar in Cygnus X-3 is spinning extremely rapidly, between 100 and 1,000 times a second. So far only two pulsars have been observed with a rotation rate in that range. The discovery of a fast pulsar in Cygnus X-3 would obviously lend strong support to the pulsar model.

A second model, proposed by Ganesar Chanmugam of Louisiana State University and Kenneth Brecher of Boston University, does not require the neutron star to be spinning rapidly. Instead the energy converted into cosmic rays derives from the process that is also responsible for the X-ray emission: the accretion of matter by the neutron star from its massive companion. As the accretion disk rotates about the neutron star, the star's magnetic field induces an electric field in the plane of the disk. Protons are accelerated radially outward along the electric field lines. Some of the protons penetrate the accretion stream, where they collide with gas nuclei and generate gamma rays. This model too has its drawbacks. For instance, the electric field has to be intense, and it is not clear whether such a field can be sustained in the gaseous disk.

The Chanmugam-Brecher model would explain why high-energy gamma-ray emissions seem to have been detected from two other X-ray binaries (Vela X-1 and Hercules X-1) that do not contain fast pulsars. Still, decision between the two models of cosmic-ray production—and a final understanding of Cygnus X-3—is probably not imminent. Much has still to be learned, and some of it will doubtless be surprising.

For example, the cosmic radiation from Cygnus X-3 may teach investigators something new about particle physics. Although the excess of air showers coming from the direction of Cygnus X-3 is attributed to high-energy gamma radiation, no one has ever positively identified gamma rays as the cause of the excess; the conclusion is reached by eliminating neutrons and neutrinos, the only other known candidates. But are all the candidates known? Is the physics of interactions at 10^{15} electron volts, an energy far be-



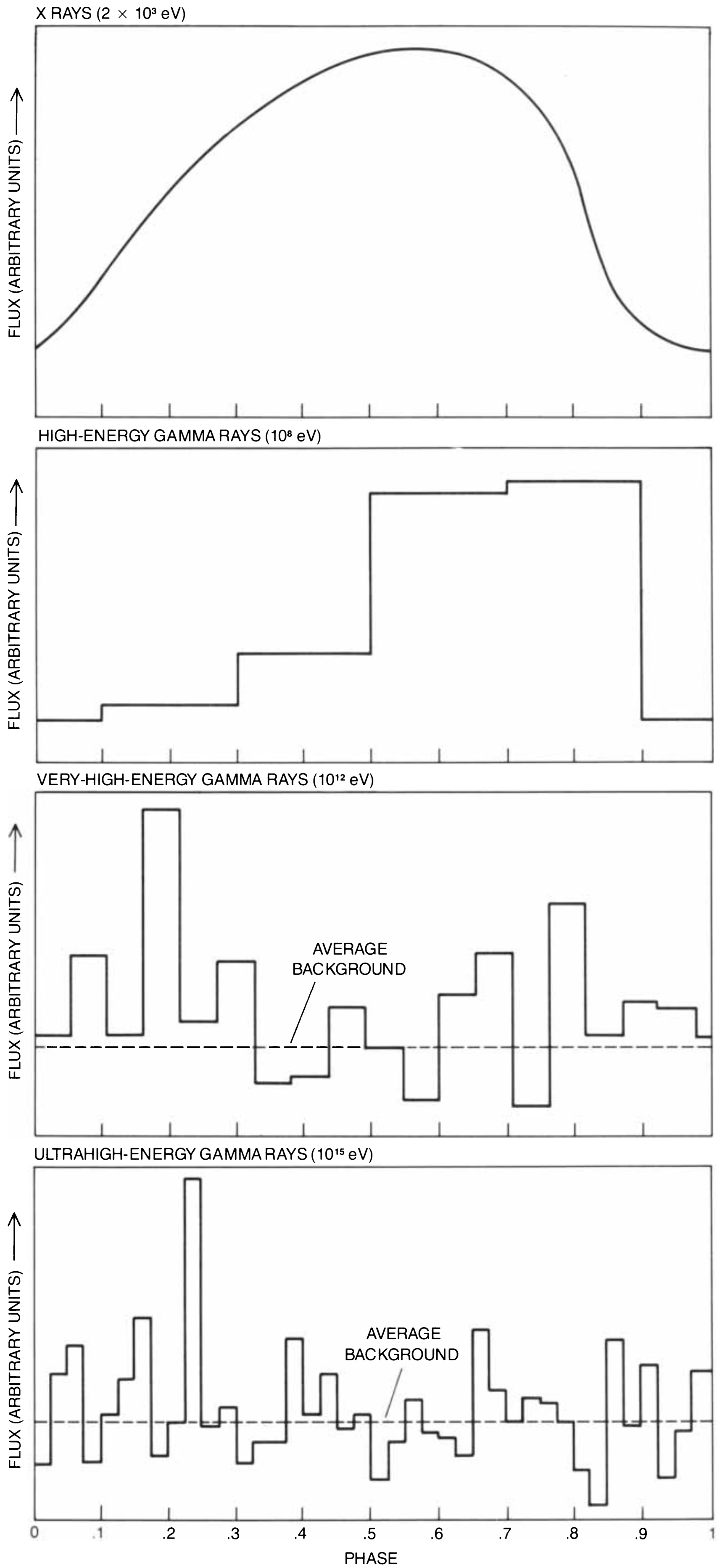
RADIO OUTBURST from Cygnus X-3 in September, 1972, showed that it was capable of accelerating particles to high energies. The outburst was discovered shortly before it peaked, at which point the radio flux from Cygnus X-3 was a thousand times more intense than usual. Here the outburst at a frequency of 8,085 megahertz is plotted. At lower frequencies (and lower energies) it lasted longer, a pattern characteristic of synchrotron radiation from relativistic electrons. The peaks suggest Cygnus X-3 was ejecting new bursts of particles.

yond the reach of manmade accelerators, really understood? Or might the flux of cosmic radiation from Cygnus X-3 include a new type of particle?

Recent reports from two proton-decay experiments, one in the Soudan mine in Minnesota and one in the tunnel under Mont Blanc, are at least suggestive of something new. Both experiments are designed to detect the decay products of the proton, which may include muons (short-lived, negatively charged particles). As a result the apparatus also records the arrival directions of muons generated in cosmic-ray air showers. From these data one can glean the arrival directions of the cosmic rays themselves.

After analyzing a year's worth of data the Soudan group has reported an excess of cosmic rays from the direction of Cygnus X-3. The signal bears the characteristic 4.8-hour periodicity. A similar periodicity is seen in data from the Mont Blanc experiment. Air showers triggered by gamma rays, however, are expected to contain very few muons and should not produce a measurable signal in these detectors. The results thus suggest that some unknown particle emitted by Cygnus X-3 may be causing air showers. Alternatively, and more conservatively, the interactions of high-energy gamma rays in the atmosphere may not be understood; perhaps they can produce a lot of muons after all.

Such uncertainties should not obscure our central point: For the first time a significant source of highly energetic cosmic rays has been identified. After decades of rather unconstrained speculation, astrophysicists now have a prototype of a cosmic-ray accelerator, even if they do not fully understand how it works. In the coming years more sensitive gamma-ray telescopes may detect new cosmic-ray sources that are simply weaker versions of Cygnus X-3. The end of a long search seems close at hand.



PERIODICITY of gamma-ray emissions from Cygnus X-3 matches the 4.8-hour periodicity observed at X-ray wavelengths. The periodicity is attributed to the orbit of the binary star. The gamma-ray flux curves are averages made by superposing emissions from successive 4.8-hour time intervals. If the signal varied randomly, the superposition would yield a flat curve. Peaks in the superposition imply that the peaks in the signal recur at the same phase of successive time intervals; thus the signal is periodic. The phase difference between the 10^8-electron-volt emission and the emission at higher energies suggests that the signals are from different regions of the binary star system.

The Calcium Signal

The calcium ion controls processes ranging from muscle contraction to cell division. An array of proteins that are specialized for binding calcium regulate its concentration in the cell and mediate its effects

by Ernesto Carafoli and John T. Penniston

A common trigger precipitates biological events as diverse as the contraction of a muscle and the secretion of a hormone. The trigger is a minute flux of calcium ions. Calcium is one of the body's "second messengers": it relays electrical and chemical messages that arrive at a cell's surface membrane to the biochemical machinery within the cell. To control cellular processes effectively, calcium itself must be regulated. Thus cells have evolved an elaborate system of proteins that interact with the calcium ion, governing the transmission and reception of the intracellular message. Knowledge of these intricacies may lead to greater clinical control over intracellular calcium, a possibility that has broad implications for the treatment of disease.

Several other ions besides calcium are common in the biological environment; they include the doubly charged magnesium ion and the singly charged sodium, potassium and chlorine ions. Why, in the course of evolution, was calcium chosen as a second messenger? For a substance to act as an intracellular messenger a target protein, usually an enzyme, must be able to bind it tightly and with high specificity. The binding alters the conformation of the enzyme molecule, thereby changing its state of activity. To alter the number of affected molecules the messenger must undergo large swings in concentration. A tenfold increase in concentration may be needed to change the state of an enzyme in a cell from "off" to "on."

A simple ion such as calcium cannot be created or destroyed easily. Instead its free concentration in the cytosol (the watery medium in which the intracellular organelles are suspended) must be regulated by compounds that alternately bind the ion, removing it from solution, and release it so that it can carry its message. To be able to distinguish the messenger ion from other ions in the cell and hold it tightly, such binding compounds must be structurally complex. A cellular function that demands a structurally complex substance is almost always served by a protein.

Calcium is far better suited to tight and specific binding than other common ions are. Ions of potassium and chlorine are relatively large in radius. They do not fit into compact binding sites on proteins. The sodium ion has a smaller radius, about the same as that of the calcium ion, but because it bears only a single charge, sodium is bound relatively loosely in the complexes it forms with proteins. The very large polyatomic ions common in the biological milieu, such as phosphate and bicarbonate, are even less able to form tight complexes.

This process of elimination leaves the magnesium and calcium ions. Both of them are small ions with a double charge, capable of binding tightly to other compounds. What explains the evolutionary preference for calcium? Robert J. P. Williams of the University of Oxford has examined the binding chemistry of calcium and magnesium. In complexing with a protein both ions readily bind to six electron donors, usually oxygen atoms, in an octahedral arrangement. One donor occupies each vertex of the octahedron and adjacent bonds lie at right angles. The calcium ion, which is larger than magnesium and has a more complex electronic structure, can also form bonds with a total of seven or eight electron donors.

Because of its small size, the magnesium ion attempts to draw the oxygens of a protein with which it forms a complex into a very tight and regular configuration. A protein is not sufficiently flexible, however, to form a regular cavity that is compact enough to match the small dimensions of the magnesium ion. Instead of establishing a full set of six bonds with the protein, a magnesium ion will also link with water molecules. The substitutions greatly weaken the strength of the binding, because fewer bonds between the ion and the protein must be broken to free the magnesium. Calcium requires a less drastic change in conformation of the binding protein; hence the ion can form a full six bonds with the protein.

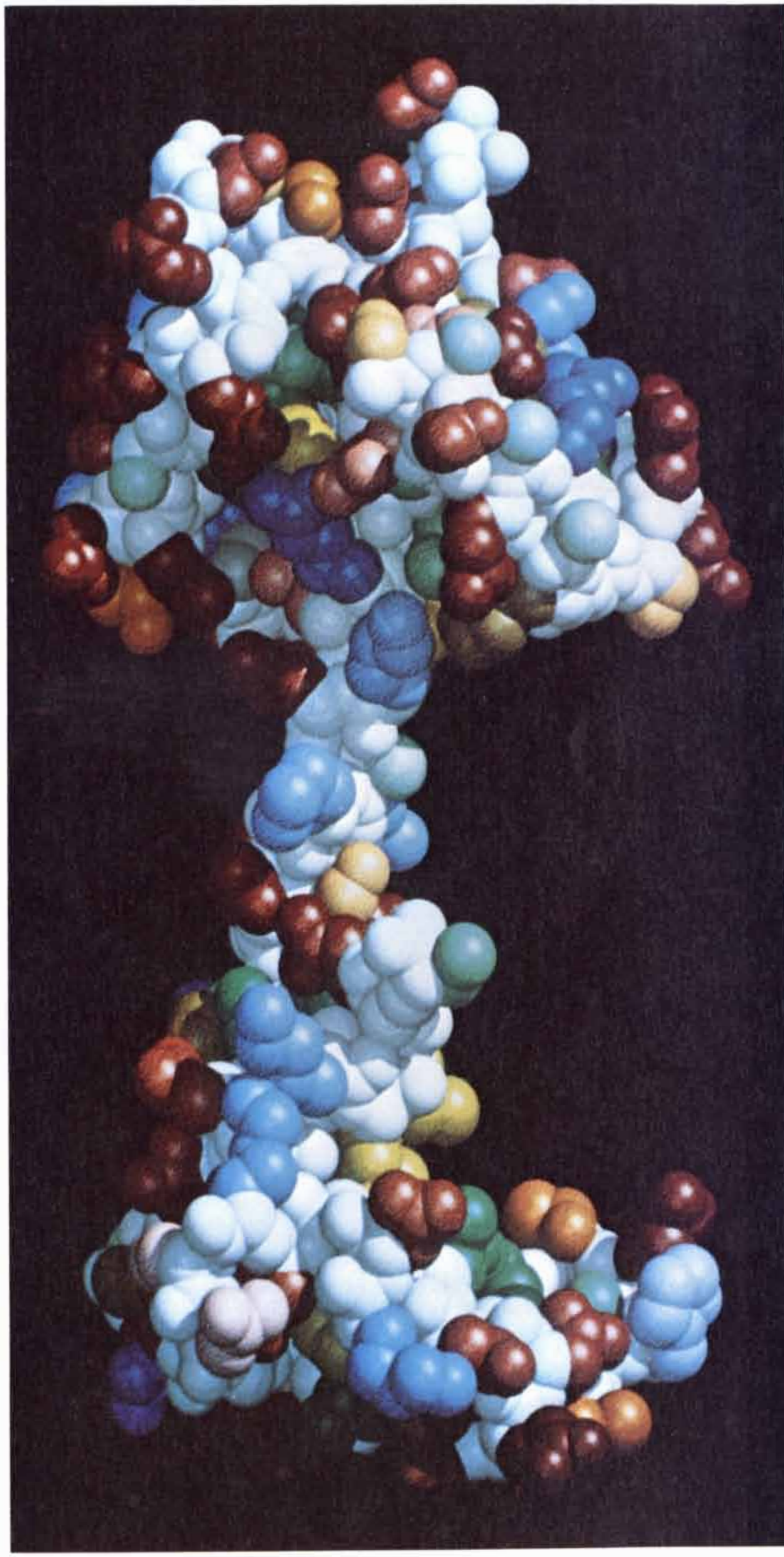
Calcium binding not only is tighter than that of magnesium but also is more specific. Because of its larger radius and variable number of bonds, calcium can adapt to irregularly shaped binding sites. Thus a protein can bind calcium very tightly while excluding magnesium, which is present in the cytosol in a concentration a thousand times as high. In contrast, a protein that will bind magnesium will often accommodate calcium equally well or better. Both tightness and specificity of binding are essential to service as an intracellular messenger; only calcium meets both requirements.

The proteins that bind calcium fall into two classes: proteins incorporated in the membranes of the cell, where they govern the passage of calcium into and out of the cytoplasm and the interior of intracellular organelles, and soluble proteins in the cytoplasm and in the organelles themselves. The proteins of the second class play some role in controlling the concentration of calcium in the cell, but the amount of calcium a soluble protein can bind is limited by the number of protein molecules. Membrane-bound proteins can regulate calcium concentration more efficiently because they act as transporters: a single protein molecule can bind a calcium ion on one side of the membrane, move it to the other side, release it and repeat the process indefinitely. The major role of the soluble proteins in the cytosol and the organelles is instead to mediate the signal's intracellular effects.

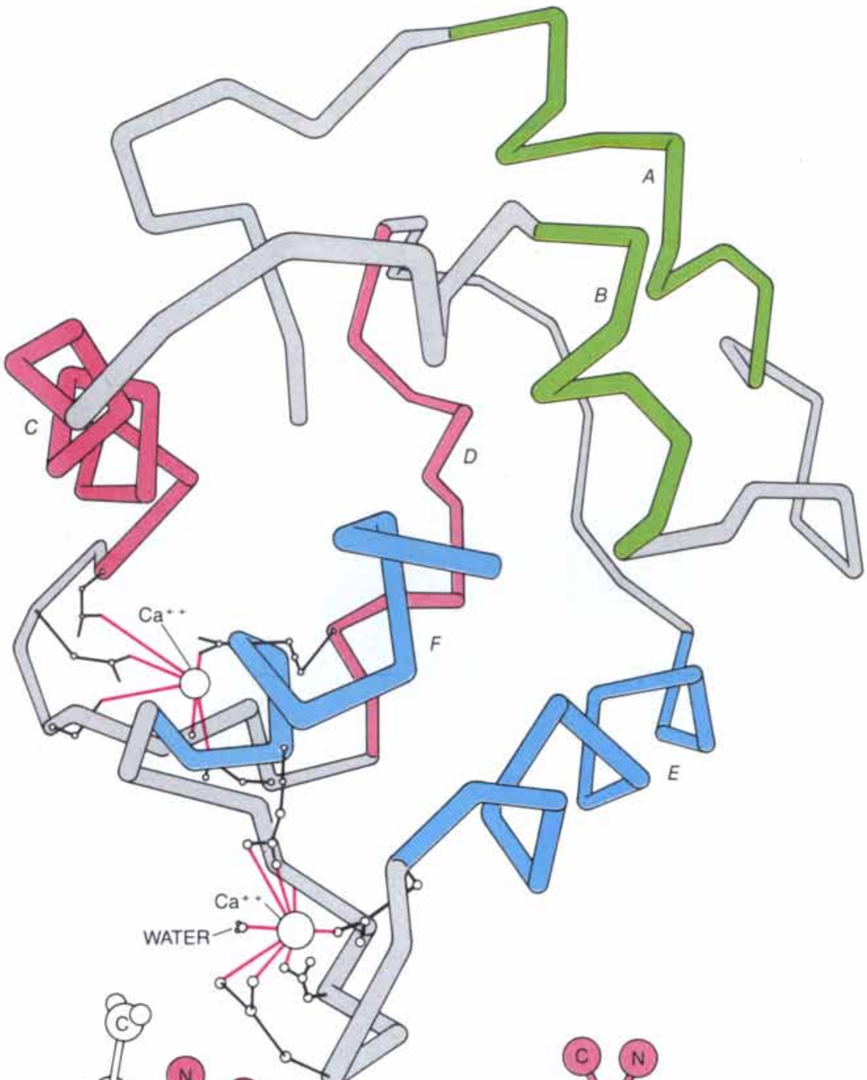
In some cases a cytosolic calcium-binding protein itself is responsible for the effects of the calcium signal. Usually the calcium-binding protein acts as an intermediary, activating other enzymes. Such a protein was first identified in 1967 by Setsuro Ebashi and his co-workers at the University of Tokyo. They described a protein that enables calcium to trigger contraction in skeletal muscle and named it troponin. Subsequent studies showed that troponin is a complex of three distinct subunits, of which one, troponin *C*, is the calcium receptor. The protein complex is an integral part of muscle fiber. When the muscle cell is at rest, a part of the troponin molecule acts as an inhibitor, preventing myosin, the contractile protein of muscle, from linking with actin, the filamentous protein. When the *C* subunit binds calcium ions, the shape of the troponin molecule changes and the inhibitor part swings aside, allowing myosin to bind with actin so that the muscle tissue can contract.

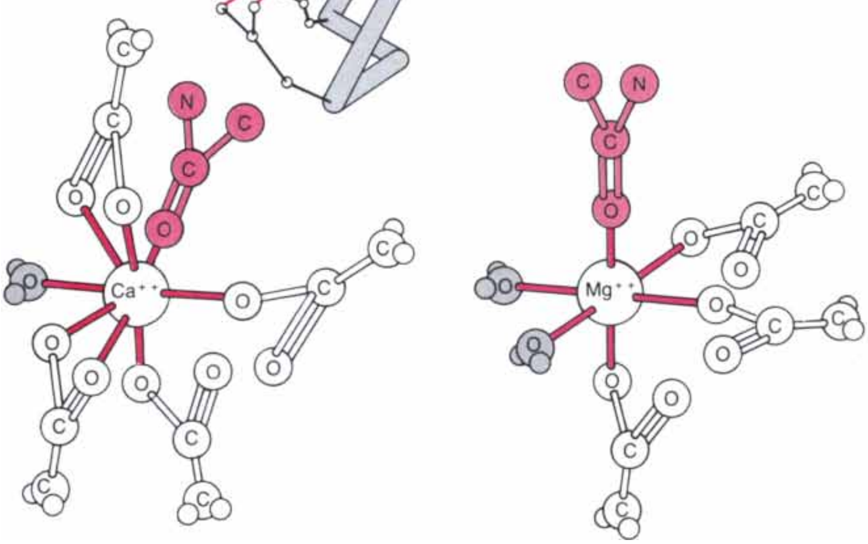
Study of a second cytoplasmic protein, parvalbumin, which is abundant in the skeletal muscles of fishes, reptiles and amphibians, helped to establish how proteins bind calcium. Parvalbumin's function is not certain, but Robert H. Kretsinger and his associates at the University of Virginia used X-ray crystallography to determine its three-dimensional structure. The parvalbumin molecule contains three similar regions, each of them consisting of a calcium-binding loop 10 to 12 amino acids long flanked by two helical sequences about 10 amino acids long. Within each calcium-binding loop between six and eight oxygen atoms on the side chains of the protein and on the protein's peptide backbone donate electrons to bind the ion. One loop is inactive because two of the amino acids needed to bind calcium have been replaced by other amino acids.

Until very recent crystallographic



CALCIUM-BINDING PROTEINS are essential to the element's role as an intracellular messenger. Some of the proteins that interact with calcium control its concentration in the cell, thereby producing signals; others receive the signals. Shown here is troponin *C*, a protein of skeletal and heart muscle. When calcium ions bind to the molecule, the protein's shape changes, triggering the physiological events that lead to muscle contraction. In this computer-generated image the spheres correspond to individual atoms and the colored regions represent side chains having distinct chemical characteristics. Osnat Herzberg and Michael N. G. James of the University of Alberta based the image on data from X-ray analysis of the protein.





PARVALBUMIN, a protein found in fishes, reptiles and amphibians, is typical of the proteins that bind calcium tightly and with high specificity. The molecule (*top*) has six helical regions (*color*) interspersed with three calcium-binding domains; a pair of helixes and the domain they flank constitute a "hand." The hand including helixes *A* and *B* is nonfunctional; bound calcium ions are shown in the other two hands, which encompass helixes *C* and *D* and helixes *E* and *F*. The details of calcium binding in the *EF* hand are shown (*bottom left*). The ion forms bonds with six oxygen atoms on the side chains of the protein and one on the protein's backbone (*color*). An eighth bond links the calcium with the oxygen in a water molecule (*gray*). The bond lengths and the angles between adjacent bonds vary. A hypothetical attempt to fit a magnesium ion, calcium's main competitor for binding sites on protein, into the hand (*bottom right*) shows why it has a much higher affinity for calcium. The magnesium ion's electronic structure and its smaller size demand that its six bonds define a compact octahedron. Since parvalbumin is not sufficiently flexible, the ion must substitute bonds with two water molecules for bonds with the protein, weakening the binding.

studies revealed the structures of other calcium-binding proteins only their amino acid sequences were known. Comparison with the amino acid sequence of parvalbumin made it possible, however, to anticipate the crystallographic results. In such comparisons the region of the protein containing helixes *E* and *F* and the calcium-binding loop they flank (together known as the *EF* hand) served as a model in the search for the calcium-binding sites on other proteins.

Such studies have shown that other calcium-binding proteins generally resemble parvalbumin. In particular they have more than one calcium-binding domain. It is likely that the shape of the protein changes depending on the number of occupied binding sites. Each of the resulting range of forms could interact with a different target protein or have a different effect on the same target.

It is thought that calmodulin, a third important calcium-binding protein, may function in this way. Calmodulin has four binding sites; recent crystallographic studies have shown that all resemble the *EF* site of parvalbumin. A calmodulin molecule with only one or two occupied sites may have different properties from one in which calcium occupies all four sites. Such variation in calmodulin's properties could help to explain the vast number of proteins with which it interacts. It appears to be the commonest translator of the intracellular calcium message, and it is present in all kinds of mammalian cells.

Calmodulin acts on enzymes that catalyze the making or breaking of bonds between phosphorus and other atoms common in biochemicals. Among such enzymes are the protein kinases, which can phosphorylate (add phosphate groups to) other proteins. Phosphorylation is often the switch that turns an enzyme on or off. Hence calmodulin regulates numerous cellular functions. Recent research has implicated it in the release of hormones from endocrine glands. It also affects cell shape and division, probably by acting on the microtubules, slender tubules made up of the protein tubulin that form an intracellular framework. Calmodulin can, for example, trigger the breakdown of microtubules when other regulatory proteins are present. Since cancer cells differ from normal ones primarily in their shape and rate of division, calcium-activated calmodulin may play a role in the biochemistry of the disease.

In skeletal muscle calmodulin's regulatory function is well established. A rise in intracellular calcium triggers an instantaneous response that is mediated by troponin: the contraction of the muscle. It also results, however, in a longer-term metabolic change mediated by calmodulin. Calcium-activated calmodulin mobilizes a protein kinase that phosphorylates a second enzyme. Thus activated, the second enzyme catalyzes the breakdown of glycogen, a storage form of glucose. The cell then metabolizes the glucose to provide energy for muscular work.

Calcium binding by protein is crucial to the generation of the intracellular calcium signal as well as to its reception. The calcium-binding proteins that reside in the plasma membrane of the cell and in the membranes of two kinds of organelle, the endoplasmic reticulum and the mitochondria, are the main regulators of the free calcium level in the cytosol and the organelles. By binding the ion and moving it in and out of the cytosol the proteins control the swings in concentration that constitute the calcium message. They also keep the normal concentration within the cell almost infinitesimally low. In a typical mammalian cell the concentration of free calcium is about 0.1 micromolar, or four millionths of a gram per liter—about a ten-thousandth the concentration in the blood plasma. (The molarity of a solution is the number of grams of solute per liter, divided by the solute's atomic or molecular weight.)

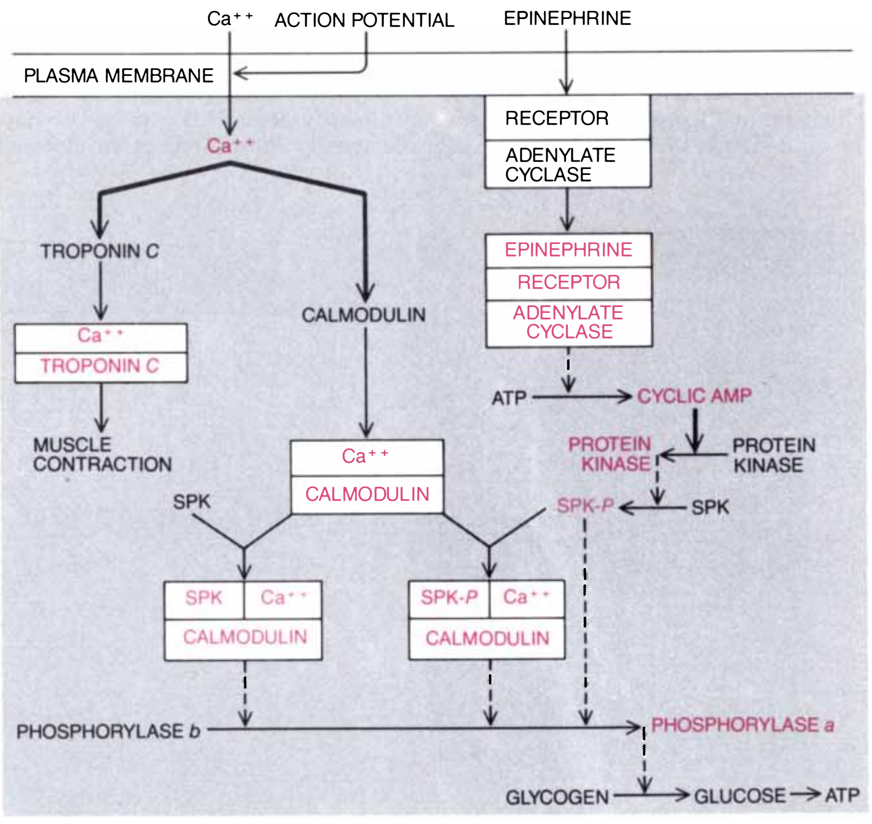
In 1976 Annemarie Weber of the University of Pennsylvania pointed out that low intracellular calcium is a necessary condition for the phosphate-driven metabolism that is characteristic of higher organisms. The energy-rich fuel for most cellular processes is adenosine triphosphate (ATP). Its breakdown releases inorganic phosphate. If the intracellular concentration of calcium were high, the phosphate and the calcium would combine to form a precipitate of hydroxyapatite crystals—the same stony substance that is found in bone. Ultimately calcification would doom the cell.

A low calcium concentration also makes the use of the ion as an intracellular messenger energetically inexpensive. The movement of calcium ions across membranes requires energy, usually supplied by ATP. If the resting level of calcium in the cell were high, a large number of ions would need to be transported into the cytoplasm to raise the concentration by the factor of 10 that is ordinarily needed to activate an enzyme; afterward the excess calcium would have to be expelled from the cell. The normal very low calcium level means that relatively few ions need to be moved, with a relatively small expenditure of energy, to regulate an enzyme. In contrast, the energetic cost of regulation by the other important intracellular messenger, cyclic adenosine monophosphate (cyclic AMP), is high: it must be synthesized and broken down each time it carries a message, and both steps require a significant investment of energy.

The membrane-bound proteins that control intracellular calcium make up at least seven distinct transporting systems, encompassing four different biochemical mechanisms. They are found both in the plasma membrane that envelops the cell as a whole and in the membrane of organelles: the endoplasmic reticulum (found in most cells), the related sarcoplasmic reticulum (found in muscle cells) and mitochondria. The transporters also vary in function: some admit calcium to the cytosol in response to an external stimulus, generating the calcium message, whereas others remove the calcium and maintain the low intracellular concentration that makes the messenger system efficient.

Among the transporters that export calcium from the cytoplasm another distinction holds. Some of the proteins bind and transport the ions only when calcium has reached a concentration of one micromolar or so, well above the resting level, but they typically can transport large numbers of ions. These low-affinity transporters presumably serve to remove calcium from damaged or stimulated cells. Other transporters bind calcium at much lower concentrations but have a lower transport capacity. These high-affinity transporters fine-tune the level of calcium in an unstimulated cell.

The plasma membrane of the cell contains two transporters responsible for expelling calcium from the cell.



CALCIUM IN SKELETAL MUSCLE has a dual role. Mobilized when a nerve impulse triggers an action potential at the plasma membrane, calcium ions diffuse into the cell (*heavy arrows*) and interact with two calcium-binding proteins, troponin *C* and calmodulin, converting them from an inactive (*black*) into an active form (*color*). Troponin *C* causes the muscle to contract. Calmodulin combines with inactive synthase-phosphorylase kinase (SPK) to produce a partially active form (*light color*). SPK in turn catalyzes the conversion of phosphorylase *b* into its active form, phosphorylase *a*. (Broken arrows indicate catalytic processes.) Phosphorylase *a* catalyzes the breakdown of glycogen into glucose, which the cell metabolizes to produce ATP; ATP supplies the energy required for muscular contraction. Calcium ions act in synergy with cyclic AMP, another intracellular messenger. When a hormone such as epinephrine binds to a cell-surface receptor, adenylate cyclase catalyzes the synthesis of cyclic AMP. Cyclic AMP diffuses into the cell and triggers enzymatic steps that lead to the phosphorylation of SPK. Phosphorylated SPK (SPK-*P*) is a partially active form; when it combines with calcium-carrying calmodulin, it becomes fully active.

The high-affinity system is an ATPase: an enzyme that splits ATP to derive energy. The energy enables it to move calcium out of the cell up the steep concentration gradient that prevails across the plasma membrane. The transporter contains, in a single molecule, all the machinery needed to transport calcium: a part that splits ATP and a part that binds and transports the ion. The pump was first discovered, by H. J. Schatzmann of the University of Bern, in red blood cells. Since then we and many other workers have shown that such a pump is present in virtually every kind of mammalian tissue.

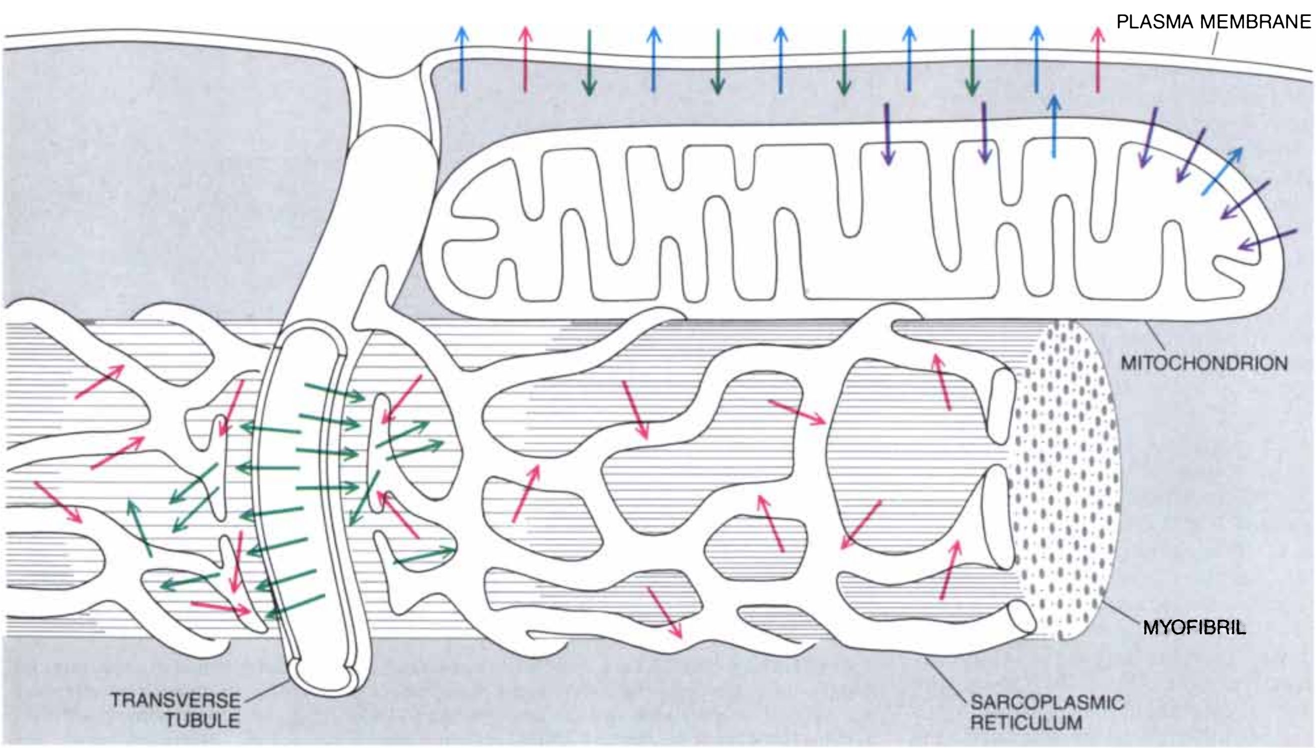
One of the mechanisms by which this calcium-pumping ATPase is regulated enabled us, working with Verena Niggli of the Swiss Federal Institute of Technology in Zurich, to purify the protein from red-cell membranes. When intracellular calcium links with calmodulin, the calmodulin activates the calcium pump by binding to it. The pump then lowers the intracellular calcium concentration. Although a single red-cell membrane contains only about 5,000 calcium-pump molecules, which amount to only a thousandth of the total membrane protein, the specificity and tightness with which calmodulin binds to the transporter molecules made it possible to isolate them from the membrane.

We linked calmodulin to a solid support and exposed it to calcium and an extract from human red-cell membrane. The ATPase bound to the immobilized calmodulin and became temporarily insoluble. We washed the insoluble pump protein using a calcium-containing solution, which removed other proteins. Rinsing with another solution, this one free of calcium, released the purified protein by removing calcium from the calmodulin, which then freed the ATPase. We incorporated the purified pump into artificial membranes and showed that it was still capable of splitting ATP and pumping calcium.

Niggli examined the properties of the purified pump and showed that as it ejects calcium from the cell, the pump imports protons (hydrogen ions), probably in a ratio of two protons for each calcium ion. Since a calcium ion carries a double positive charge, the exchange makes the pump electrically neutral. It is therefore not affected by the difference in electric charge that prevails across the plasma membrane of many cells.

Our study of the purified ATPase showed it may also be activated by substances other than calmodulin. One is an acidic lipid in the cell membrane called phosphatidylinositol bisphosphate, which decreases in abundance when a cell is stimulated by certain hormones. Since the same hormones also cause an influx of calcium, a decrease in the amount of the lipid stimulating the pump following hormonal stimulation may leave the cell with a higher free-calcium level than is normal. Because of its high calcium level, such a cell would be easier to stimulate a second time. It would in effect remember its stimulation history, an important feature for certain nerve cells.

It also seems likely that in the living cell the splitting of the ATPase by proteolytic enzymes (enzymes that break down proteins) may act as a regulatory mechanism. The pump protein in the plasma membrane of red blood cells is one of the largest single-chain proteins known. A reduction of 50,000 in its molecular weight of 138,000 does not prevent it from operating; indeed, removal of the extra protein activates the purified protein just as calmodulin



PUMPS, EXCHANGERS AND CHANNELS, seven in all, control the level of calcium in the cytosol (the fluid of the cytoplasm) of a heart-muscle cell. Calcium enters the cell through calcium channels (*green*) in the plasma membrane; calcium channels are abundant in the transverse tubule (an invagination of the plasma membrane) and are concentrated where it approaches the sarcoplasmic reticulum. An influx of calcium from the transverse tubule causes the sarcoplasmic reticulum to release more calcium, probably through similar membrane channels concentrated in the ends of the reticulum. The calcium triggers contraction of the muscle myofibril. A slow, continuous release of calcium from mitochondria occurs by means of a sodium-calcium exchange system (*blue*). Calcium is expelled from the cell by an enzyme known as a calcium-pumping ATPase (*red*) and by a sodium-calcium exchanger. An ATPase in the membrane of the reticulum returns calcium to the interior of that organelle. In the inner mitochondrial membrane a uniporter (*purple*), so called because it moves calcium without compensatory transport of other ions, also removes the ion from the cytosol.

or acidic lipid does. The apparently superfluous protein presumably has a regulatory role or takes part in other, unknown functions of the molecule.

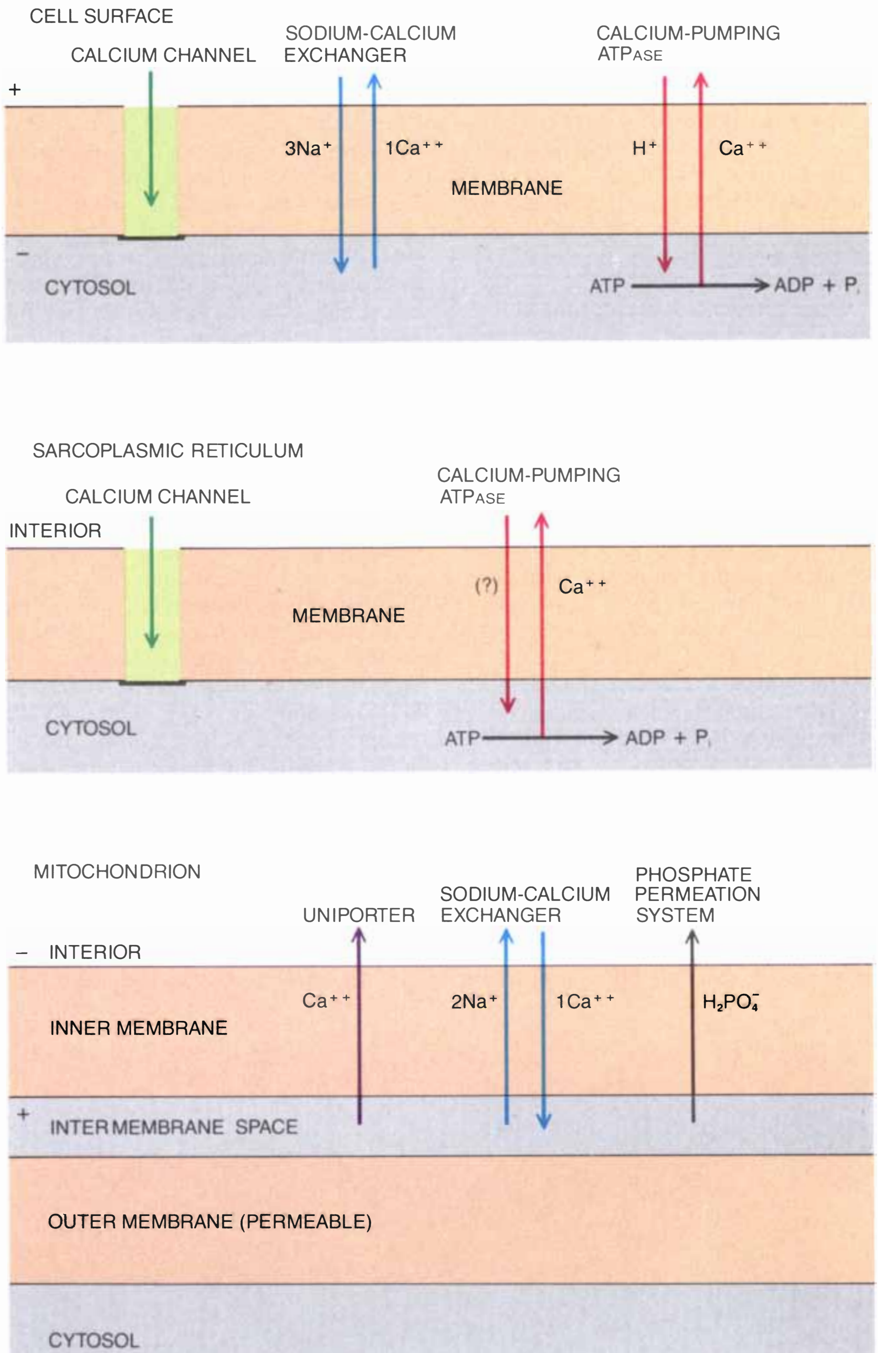
A fourth regulatory mechanism has been shown to operate in heart-muscle plasma membrane. It is the phosphorylation of the pump protein by a protein kinase, which is activated by cyclic AMP. The relative importance in living cells of the four means of activating the plasma-membrane calcium pump is not known, but the number of mechanisms testifies to the importance of the pump in cell physiology.

The plasma membrane ATPase, a high-affinity pump, responds to minute increases in intracellular calcium. More drastic swings in concentration activate the other calcium-removal system of the plasma membrane, the so-called sodium-calcium exchanger. This transporter protein is particularly abundant in excitable cells: nerve and muscle cells, which repeatedly experience rapid influxes of calcium in response to electrical impulses. The sodium-calcium exchanger, which was first identified in squid heart muscle and nerves, does not generate the energy for calcium transport independently, unlike the calcium-pumping ATPase.

Instead the exchanger derives some of the energy needed to drive calcium out of the cell from the concentration gradient of sodium. Like calcium, sodium is more abundant in the fluid surrounding the cell than it is in the cytoplasm. The ratio of sodium's concentration outside the cell to its concentration in the cell is about 10 to one, whereas the ratio for calcium is greater than 1,000 to one. By itself the small sodium gradient contains barely enough energy to force calcium outward against that ion's steep concentration gradient.

It has been established, however, that three singly charged sodium ions move into the cell for every doubly charged calcium ion that is exported. The charge imbalance enables the transport process to tap the energy embodied in the membrane potential: the gradient of voltage across the membrane of an excitable cell. In the resting state of the cell the inside of the membrane has a negative charge of 90 millivolts. In effect the transport process allows charge to flow down the gradient, because the net effect of one cycle of the exchanger is to move one positive charge into the cell. Thus both chemical and electrical gradients drive the exchanger.

The capacity of the exchanger is high: in a single heart cell the protein can remove as many as 3×10^9 calci-



DETAILS OF CALCIUM TRANSPORT across a muscle cell's three calcium-transporting membranes vary. The molecular structures of the transporter proteins are obscure but their modes of operation are known. Calcium channels in the plasma membrane (*top*) open when the voltage across the membrane changes; they allow calcium ions to flow into the cell down the steep gradient of concentration that prevails across the plasma membrane. The sodium-calcium exchanger exports calcium from the cell, depending in part on the concentration gradient of sodium to drive calcium ions; in addition, because it generates a charge imbalance (importing three singly charged sodium ions for each doubly charged calcium it expels), the exchanger is powered by the voltage gradient across the membrane. The calcium-pumping ATPase also expels calcium, deriving energy for the process by splitting ATP into adenosine diphosphate (ADP) and inorganic phosphate (Pi). The membrane of the sarcoplasmic reticulum (*middle*) is thought to release calcium through channels similar to those of the plasma membrane; the release is driven by a concentration gradient. A calcium-pumping ATPase accepts energy from ATP to drive calcium back into the reticulum. Calcium release through the inner mitochondrial membrane (*bottom*) takes place through a sodium-calcium exchanger, driven by the opposing gradients of sodium and calcium. Mitochondria import calcium by means of a uniporter, which allows the positively charged calcium ions to migrate toward the negatively charged interior of the organelle. A third transporter in the inner mitochondrial membrane imports phosphate ions, which react with calcium in the mitochondria to form hydroxyapatite, the same substance that mineralizes bone.

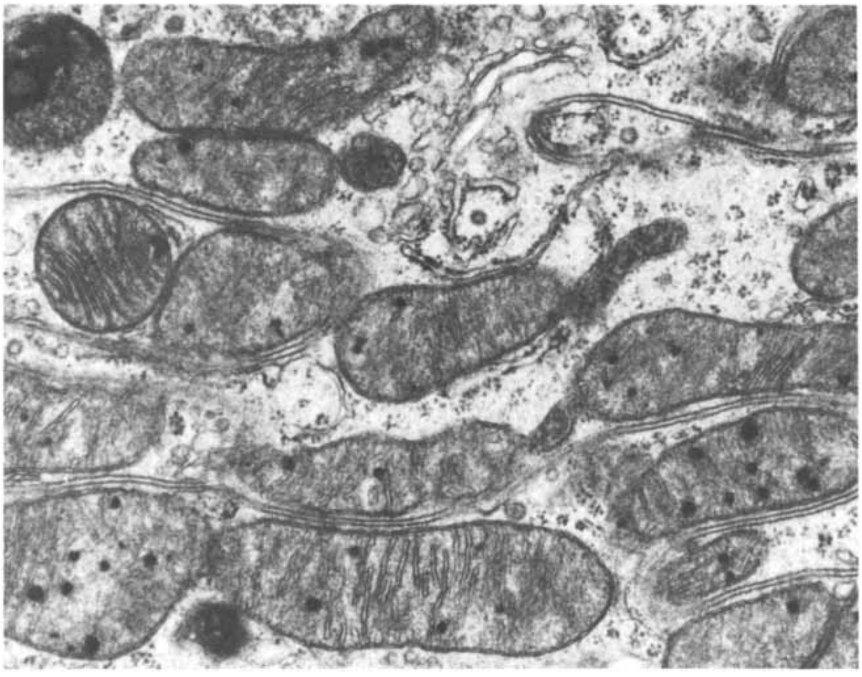
um ions per second from the cytosol. It does so only when the intracellular concentration is quite high; the exchange reaches half its maximum velocity at a calcium concentration of at least one to five micromolar, more than 10 times the resting level. Such findings confirm that the exchanger serves mainly to remove large quantities of calcium from excitable cells following stimulation.

How does calcium enter the cell? In both excitable and nonexcitable cells the ions enter through calcium channels: molecular pores in the membrane that allow large numbers of ions to move down the concentration gradient and into the cell. In excitable cells calcium channels open in response to the action potential: a reversal in the normal voltage gradient across the membrane. An action potential is triggered when a messenger molecule (usually a neurotransmitter) binds to a receptor at the surface of the cell, causing an initial depolarization, or change in the voltage gradient. Voltage-sensitive sodium channels in the membrane then open, admitting positively charged sodium ions into the cell. The influx causes the charge inside the membrane to change from negative to positive.

Calcium channels, which normally are closed, are sensitive to the membrane potential. They begin to open as the potential rises above −30 millivolts; at a potential of +30 millivolts about 70 percent of a cell's calcium channels are open. The action potential peaks at about +40 millivolts. At that point other ionic events (including an outflow of potassium ions) return the membrane potential to −90 millivolts, and the influx of calcium ends. To explain the sensitivity of calcium channels to voltage it is widely assumed that one of their protein components, of which there are probably three, acts as a voltage sensor. Such a protein would have an electrical polarity, which would enable it to reorient itself in the mouth of the channel as the electrical field changed.

The voltage-sensitive calcium channels remain open for only about a millisecond, during which they admit about 3,000 calcium ions. Even when the membrane potential is at its highest, the calcium channels in a cell's membrane are continually opening and closing, with 70 percent of them open at a given instant. That percentage can increase significantly under the influence of cyclic AMP. Cyclic AMP, which is made at the plasma membrane when hormones bind to cell-surface receptors, activates enzymes that phosphorylate the calcium channels. Phosphorylation greatly increases the probability that a channel will open.

Since nonexcitable cells do not experience action potentials, their calcium channels cannot be controlled by voltage. In such cells calcium channels may stay open permanently, or they may be regulated by cyclic AMP, which could gradually open and close them. In either case the number of channels or their calcium conductance would need to be quite small to prevent large influxes of calcium. Nevertheless, the calcium channels of nonexcitable cells share several important properties with the channels of excitable cells. Calcium-channel blockers, drugs widely used in heart disease because of their ability to modulate heartbeat, inhibit both kinds of channel, as do ions such as cobalt, which inactivate the channels by binding tightly to their calcium receptors.



DEPOSITS OF CALCIUM in the form of hydroxyapatite are enlarged 30,000 diameters in this electron micrograph of mitochondria in a mouse kidney cell. Mitochondria absorb calcium from the cytosol and store it as hydroxyapatite when its concentration rises far above the normal level. In this case the high intracellular level of the ion was achieved by treating a mouse with parathyroid hormone, which reduces calcium excretion from the kidneys. Feroze N. Ghadially of the University of Saskatchewan provided the micrograph.

In principle mechanisms that transport calcium in and out of the cell through the plasma membrane could suffice to control intracellular calcium. In cells such as those of muscle, however, in which calcium triggers a quick and transient response, the release and removal of the ion must take place very rapidly. A system of release and uptake situated in the cell interior, in an organelle, suits the purpose better because it is closer to the calcium's site of action. Moreover, the traffic of calcium within the cell may require less energy than transport between the cytoplasm and the extracellular space: the concentration gradients across the membranes of organelles may be less steep than the gradient across the plasma membrane.

In heart-muscle cells investigators have measured the area of the calcium-transporting membranes and combined the results with information on the activity of the transporter proteins in each membrane. The calculations have shown that in heart muscle intracellular calcium movements, rather than transport across the plasma membrane, account for most of the calcium that is injected into or removed from the cytosol.

Two of the mechanisms engaged in intracellular calcium movement reside in the endoplasmic reticulum and the sarcoplasmic reticulum. Both kinds of reticulum consist of a network of flattened tubes that encloses a separate space within the cell; one function of the space is to sequester calcium from the cytosol. In tissues other than muscle the endoplasmic reticulum also has other functions, such as the synthesis of proteins and the detoxification of foreign compounds. In muscle cells the major function of the sarcoplasmic reticulum is to govern the short-term fluctuations in free calcium that cause the contraction of muscle fibers.

It now seems likely that the sarcoplasmic reticulum releases calcium through channels analogous to those of the plasma membrane. It appears that the opening of the channels in the reticulum follows an initial surge of calcium into the cytoplasm through plasma-membrane channels opened by an action potential. Calcium channels are particularly concentrated in a feature of the cell membrane known as the transverse tubule, which extends into the muscle fiber near an enlargement of the sarcoplasmic reticulum. Thus the initial influx occurs very near the site of later intracellular release.

It is not known how the influx of extracellular calcium leads to calcium release from the sarcoplasmic reticulum of muscle cells. In other cells recent studies have shown that a breakdown product of phosphatidylinositol bisphosphate, the same membrane lipid that activates the calcium-pumping ATPase in the plasma membrane, triggers a flow of calcium from the endoplasmic reticulum. Hormones that bind to cell-surface receptors cause splitting of the lipid. One fragment, diacylglycerol, presumably remains part of the membrane. There, in combination with another factor, it binds calcium and activates certain protein kinases that have regulatory functions. The other breakdown product, inositol trisphosphate, dissolves in the cytoplasm and causes the calcium release. Workers are currently investigating whether such a mechanism also operates in muscle, and what its relation might be to the initial current of calcium that is a prerequisite of contraction.

The calcium that triggers muscle contraction is pumped back into the sarcoplasmic reticulum by a calcium-pumping ATPase. Like the ATPase of the plasma membrane, the transporter responds quickly to small increases in cytosolic calcium concentration, but it differs from the membrane protein in many details of its structure and regulation. The transporter protein is extremely abundant in muscle cells. In rabbit skeletal muscle, for example, up to 90 percent of the sarcoplasmic reticulum consists of the protein.

MEMBRANE (CALCIUM-REMOVAL SYSTEM)	PERCENT OF TOTAL CALCIUM-TRANSPORTING MEMBRANE	CONCENTRATION (MICROMOLAR) .1	1	10
		PERCENT OF TOTAL CALCIUM UPTAKE		
CELL SURFACE	.8			
(CALCIUM-PUMPING ATPASE)		3.6	.2	.07
(SODIUM-CALCIUM EXCHANGER)		27	3	1.9
SARCOPLASMIC RETICULUM (CALCIUM-PUMPING ATPASE)	12.1	69	90	47
MITOCHONDRION (UNIPORTER)	87	0	6.4	51

COMPARISON OF AREAS AND ACTIVITIES of calcium-transporting membranes in a heart-muscle cell shows that calcium transport into organelles, rather than export from the cell, is the primary means by which this cell controls the ion's concentration in the cytosol. (In other cell types export is more important.) At concentrations of .1 and one micromolar (typical respectively of a resting and a contracting cell) the sarcoplasmic reticulum accounts for most calcium removal. When the concentration reaches 10 micromolar, a level found only in an injured or diseased cell, the mitochondria become dominant.

Whereas the sarcoplasmic reticulum functions as a swift and sensitive modulator of intracellular calcium, the other organelle that controls calcium, the mitochondrion, acts as a long-term, large-scale regulator. Mitochondria are the site of cellular respiration, part of the process that liberates energy from foodstuffs. That they are also able to accumulate and release calcium was discovered in 1961 by Frank D. Vasington and Jerome V. Murphy of Johns Hopkins University. It is now generally agreed that mitochondria are unable to effect rapid, minute (submicromolar) changes in calcium concentration. Instead mitochondria are crucial in preventing large fluctuations of cellular calcium above and below the range of concentration that is normal.

Mitochondria are enclosed by two layers of membrane. The outer membrane is permeable to ions, and so its role in calcium transport is passive. Calcium uptake across the inner membrane is powered by the mitochondrial membrane potential of about 180 millivolts, in which the charge inside the membrane is negative. The potential enables a putative transporter protein embedded in the membrane to move the positively charged calcium ions inward, toward the region of negative charge. The transporter acts as a uniporter, moving calcium without compensatory movements of other ions.

In isolated mitochondria maintained under conditions similar to those in the living cell the uptake of calcium reaches half the maximum rate only when the concentration of the ion in the cytosol rises to 25 micromolar, a value well above the normal intracellular level. Hence the mitochondrial calcium uniporter is of low affinity. Its transport capacity, however, is high: the convoluted inner mitochondrial membrane represents about 90 percent of the total calcium-transporting membrane in most of the cells of higher organisms.

Mitochondria have a storage capacity for calcium that matches their uptake ability. Separate transporters in the inner membrane move phosphate ions from the cytosol into the mitochondria. There the phosphate and calcium form masses of hydroxyapatite. In this way the mitochondria can accumulate large amounts of calcium. Normal cells usually do not retain much calcium in mitochondria. Injury to a cell, however, generally reduces the ability of the plasma membrane to exclude calcium. Injured cells take on large amounts of calcium from the extracellular medium; mitochondria absorb the excess.

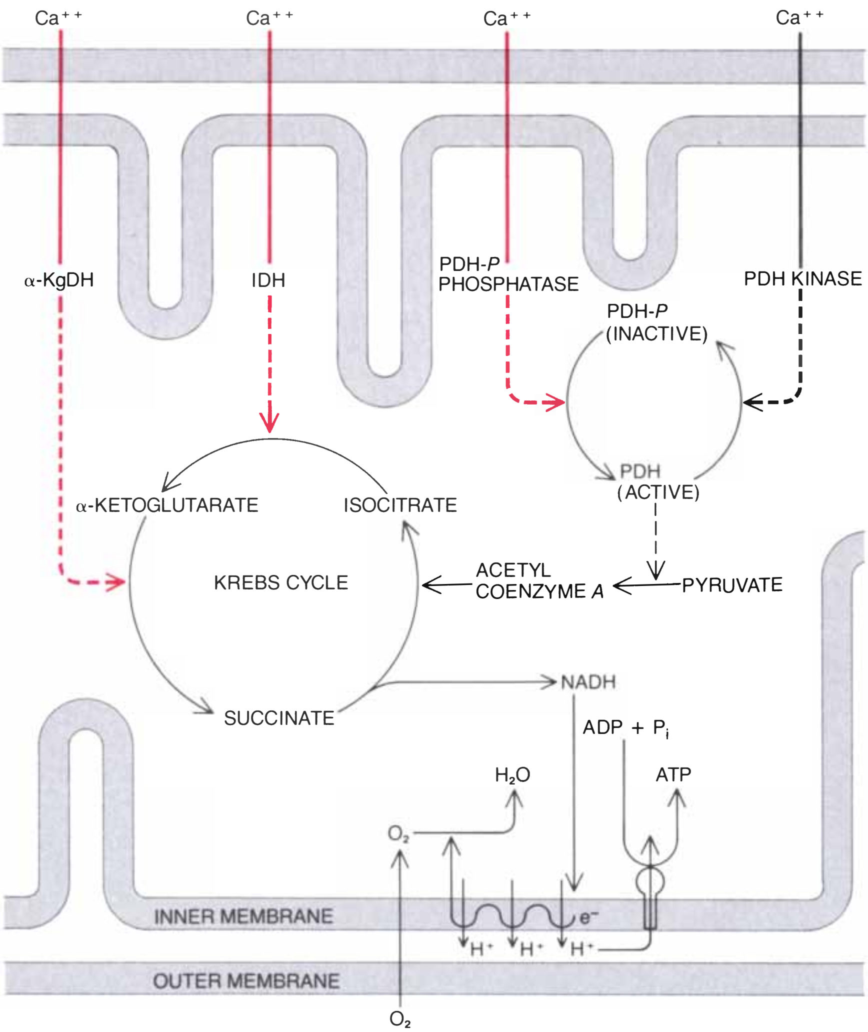
The buffering ability of mitochondria is finite; sufficiently large amounts of intracellular calcium eventually overcome their storage capacity. Calcium, normally a vital carrier of signals, then becomes an ionic assassin of the cell. The reactions calcium ordinarily modulates proceed continuously and uncontrollably, and the excess ions activate other reactions that do not occur in a normal cell.

In less cataclysmic situations, during which the mitochondria are not swamped, they buy time for the cell by absorbing the excess calcium; after the calcium storm has abated they release the ions at a rate that does not disturb cell metabolism. The release takes place through a transporter that, like the exchanger of the plasma membrane, catalyzes the exchange of sodium and calcium. In contrast to the plasma-membrane exchanger, which creates a charge imbalance by bringing in three sodium ions for every calcium it exports from the cell, the mitochondrial exchanger is electrically neutral: it transports sodium and calcium ions in a ratio of two to one. Neutral charge enables it to release calcium from the mitochondrion slowly in spite of the high membrane potential, which would tend to accelerate an electrically unbalanced exchange. In some types of mitochondria, such as those in liver cells, the sodium-calcium exchanger protein is scarce or absent and the mechanism of calcium release is uncertain.

In heart-muscle cells both the uniporter and the sodium-calcium exchanger seem to operate continuously, at a low level, cycling calcium into and out of the mitochondria. The opposing transport processes constitute a futile cycle. The cycle wastes some energy by dissipating the mitochondrial membrane potential, which is generated by respiration. But the ceaseless operation of the transporters ensures that the mitochondria are prepared to defend the cell against a calcium surge.

The calcium stored in mitochondria is not inert; it also has a metabolic role. Richard M. Denton of the University of Bristol recently showed that when mitochondrial free calcium reaches a concentration above about one micromolar, it affects the activity of at least four enzymes in the organelle. The enzymes catalyze steps in the production of substances that carry chemical energy derived from the oxidation of foodstuffs to the respiratory chain. The respiratory chain, a series of electrochemical processes taking place across the inner mitochondrial membrane, transfers the energy to the cellular energy carrier, ATP. The effect of calcium is to speed up the metabolic processes that supply the respiratory chain thereby increasing the production of ATP.

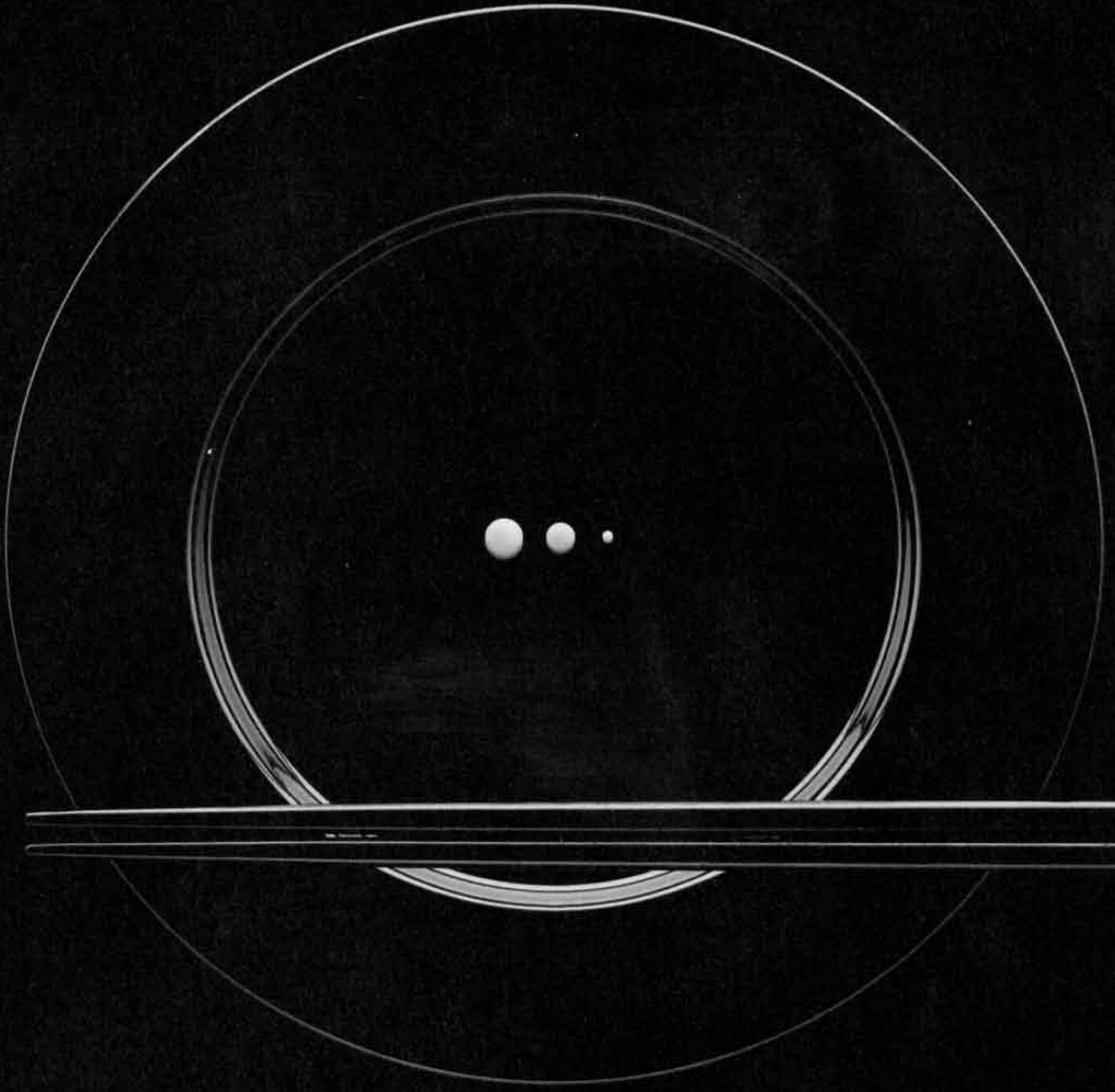
This chain of events may serve as a feedback mechanism that helps the cell to defend itself against excessive levels of the ion. When the cell is challenged by very high intracellular calcium, it needs to operate all its calcium-removal mechanisms. All of them require energy generated by the respiratory chain. The two calcium-transporting ATPases use ATP directly. The plasma-membrane sodium-calcium exchanger taps the energy stored in the membrane potential, which is maintained by a sodium-potassium pump that is also an ATPase. The mitochondrial uniporter exploits the mitochondrial membrane potential, which is generated by respiration. Through its metabolic effects an initial rise in calcium in the mitochondria ensures that energy for continued calcium transport will be available.



CALCIUM QUICKENS RESPIRATION, the process by which mitochondria synthesize ATP. Calcium stimulates (*colored arrow*) an enzyme that activates pyruvate dehydrogenase (PDH); it inhibits (*heavy black arrow*) the enzyme that inactivates PDH. The increased supply of active PDH speeds the conversion of pyruvate (a breakdown product of glucose) into acetyl coenzyme *A*, which is the raw material for the series of reactions known as the Krebs cycle. Increased acetyl coenzyme *A* speeds up the Krebs cycle, as do other effects of calcium: stimulation of two enzymes that catalyze steps in the cycle. The net result is that more of the energy-carrying compound NADH is supplied to the respiratory chain on the inner mitochondrial membrane. There oxidation releases the energy, powering ATP synthesis.

The mechanisms that regulate calcium in the cell do not operate in isolation. Networks rather than simple pathways typify cell physiology, and intracellular signaling systems are no exception. The modification by cyclic AMP of calcium channels and the calcium-pumping ATPase in the plasma membrane is one example of the interactions among messenger systems. Calcium regulation is also intertwined with that of other ions. Fluxes of sodium and potassium, by initiating and ending action potentials in excitable cells, control the opening and closing of calcium channels. In turn the intracellular calcium that builds up following a series of action potentials can open potassium channels in the plasma membrane, changing its electrical character and thereby making the membrane more resistant to subsequent depolarization.

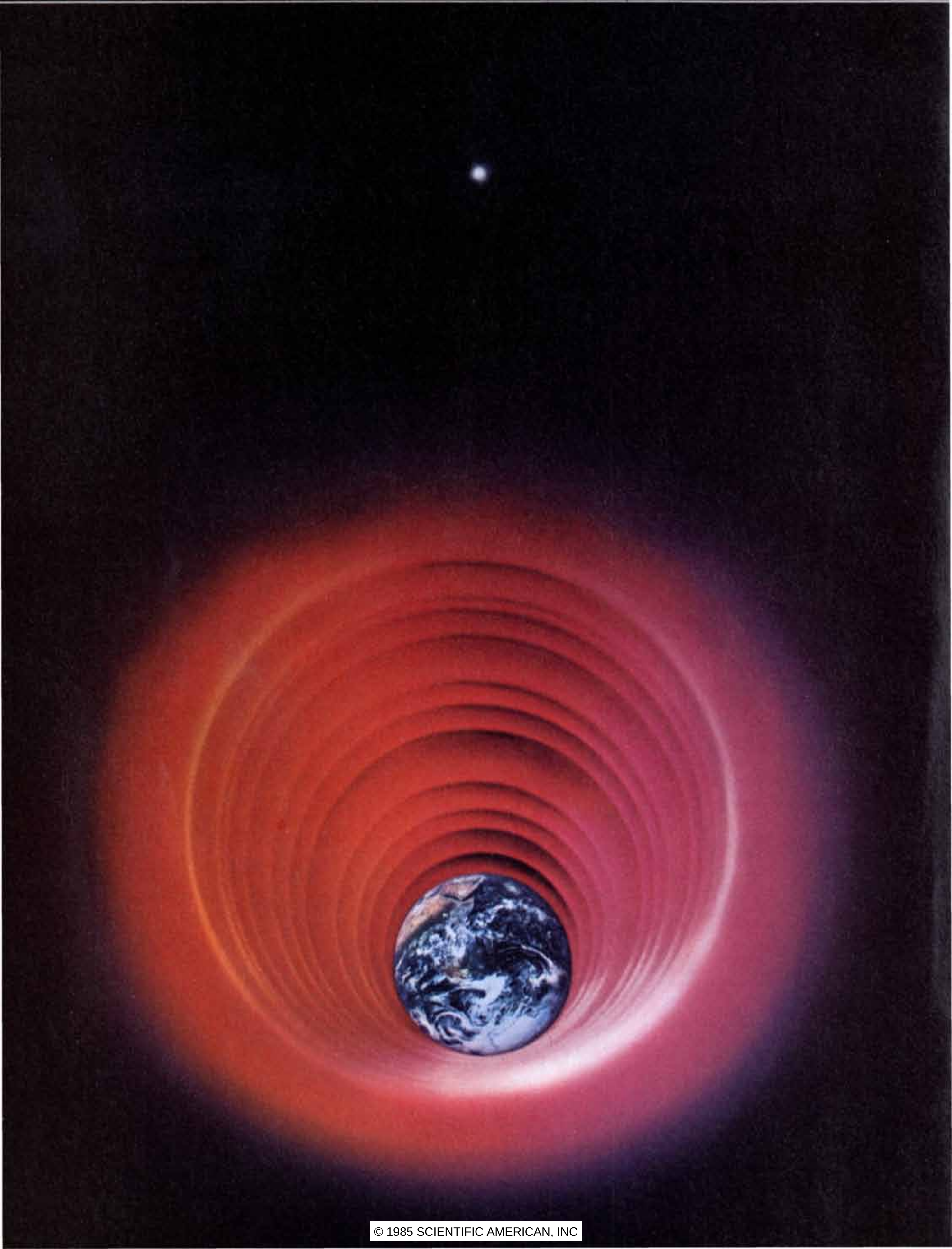
The intricacies of calcium regulation are only now being unraveled. But one of the oldest of all drugs, digitalis, exemplifies the clinical importance of the interactions. In the 18th century William Withering, an English physician, described the drug's ability to strengthen the heartbeat. It is now known that digitalis and the related substances in use today raise the level of intracellular calcium in the heart. They do not act directly on the calcium transporters; instead they affect the sodium-potassium pump in the cell membrane. The drugs inhibit the pump, raising the level of sodium in the cell. Higher intracellular sodium reduces the sodium gradient across the plasma membrane but increases it across the mitochondrial membrane. Thus the plasma-membrane sodium-calcium exchanger ejects little or no calcium while the mitochondrial exchanger injects calcium into the cytosol at an increased rate. The result is higher intracellular calcium and, for many heart patients, a longer and more secure life.



JAPANESE TECHNOLOGY TODAY.

East and West meet with the traditional *hashi* resting on *yōzara*, the traditional Western plate used by many Japanese. The setting displays three dielectric ceramic filters used in high-frequency transmissions.

Developments such as these are examples of the new technologies and new directions research and development will take through the 20th century which are part of this comprehensive report from inside Japan.



How the Matsushita Earth Sensor harnesses infrared rays from Earth to guide a satellite 22,000 miles in space.

Component technology is changing the way we see, feel and hear the universe around us. From tiny infrared sensors that steer a satellite through space. To cameras that take pictures with a microchip instead of film. In the microworld of components, Matsushita (pronounced Mot-SOOSH-ta) Electric is making great strides.

Putting a satellite in space is one thing. Keeping it in precise orbit is another. Matsushita sensors detect faint infrared rays that emanate from carbon dioxide in the Earth's atmosphere. Then convert these rays into electrical signals that keep the satellite in a precise orbit.

COMPONENTS THAT SAVE YOUR LIFE.

Closer to home, Matsushita sensors can also steer workers away from danger in nuclear power plants, research institutes, and hospitals. About the size of a credit card, our Thermoluminescent Dosimeter Radiation Detection System has a sensor that automatically measures radiation around workers. It goes off before they do.

Other Matsushita sensors let you breathe easy. The sensor in our Air Pollution Monitor Telemetry System measures pollutants in the environment.

COMPONENTS THAT SEE IN THE DARK.

Your industrial security picture will always be bright. Our patented Extended-Red Newvicon® tubes allow our Lunar-Lite cameras to see in pitch black, as little as 30/1,000,000ths of a footcandle. Perfect for areas too impractical to light.

Because Newvicon tube cameras see what other cameras can't, they're used by broadcasters and consumers. The unique Newvicon tube shoots in light as low as one birthday candle with excellent picture quality. This has helped make *Panasonic* the leading brand of consumer video cameras.

Sometimes no tubes is good news. Like our camera that uses a single-chip image sensor instead. It works where conventional cameras won't, even in extreme heat and pounding vibration. It takes a beating and keeps on shooting.

Then there's our "filmless" color still camera. Images are turned into electronic signals. Stored on a 1.85″ floppy disc. Played back on a TV monitor. And retrieved as hard copies through a color printer. Imagine the possibilities for information storage. For medicine. For education. For industry.

COMPONENTS THAT SAVE A MEMORY.

A power failure of a mere millisecond can wipe out a million-dollar computer program. Matsushita can prevent this with the first rechargeable lithium battery power supply. Rechargeable up to 1,000 times. And 50% more voltage than NiCad batteries.

Matsushita was also first to develop the world's smallest pin-type primary lithium battery. How small? So small that our Bone Conduction Hearing Aid fits into the stem of an eyeglass frame. No wonder Matsushita is one of the world's largest battery manufacturers. Producing over 1,500 types.

"Look after the components and the products will look after themselves," said Konosuke Matsushita, our company founder, more than 60 years ago. That's why Matsushita has one of the largest research and development budgets of any Japanese corporation. Expenditures of $820 million in 1984. Up 15% from the preceding year. With over 18,000 Matsushita employees developing new technologies and new products in 39 research and development organizations.

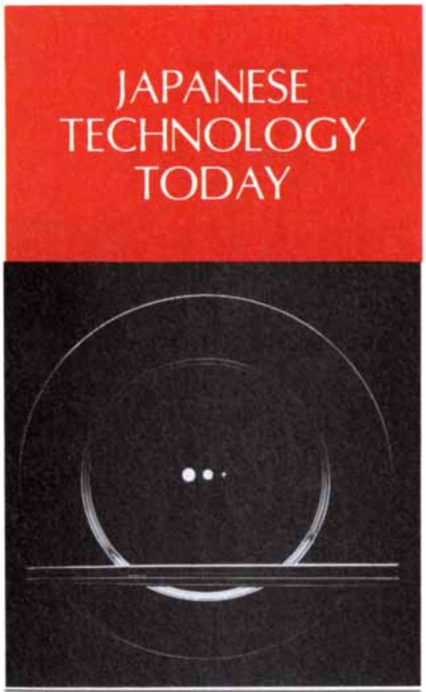
This commitment to R&D has resulted in more than 14,000 innovative products. Sold worldwide under the *Panasonic, Technics, Quasar* and *National* brand names. Down on Earth or in outer space, Matsushita is in orbit for the future.

EXPO'85

Visit our pavilion at the 1985 International Exposition, Tsukuba, Japan. No IV-E-0002

MATSUSHITA ELECTRIC

PANASONIC TECHNICS QUASAR NATIONAL



THE AUTHORS

The 1985 edition of "Japanese Technology Today" was written by Gene Adrian Gregory and Akio Etori.

Mr. Gregory is Professor of International Business in the Department of Comparative Culture at Sophia University. He is a former associate editor of *U.S. News & World Report,* a contributor to *Far East Economic Review* and author of *The Japanese Challenge in Europe*.

Mr. Etori is an award-winning and respected writer on science and technology, having covered these areas for over 20 years. He is managing editor of *Saiensu,* the Japanese-language edition of *Scientific American* and Japan's leading science publication.

THE COVER

Photographer Lynn St. John dramatically highlights three dielectric ceramic filters developed by Matsushita Electric. Combining crystal titanium dioxide with a ceramic compound made of barium-oxide-titanium-dioxide-lanthanum dioxide, these dielectric ceramic filters achieve excellent temperature and time stability along with a reduction in size by almost 90 percent compared to conventional components. Such properties are important in high-frequency communications including cellular mobile phones, telemetry and satellite transmissions. These components and additional information are available in the U.S. from Panasonic Industrial Company, ECD, Dept. Ceramics.

Art direction and supervision of the text and illustrations were provided by Ted Bates Advertising/New York.

Tsukuba Expo '85 marked a new epoch for Japan. Just as the Meiji Restoration launched Japan on its course of modernization, and the postwar reconstruction served to shift Japanese resources and energies to building a productive civil infrastructure, this third stage of Japan's modern era will necessitate charting a new course.

Having caught up with other advanced countries in applied research and development, the Council for Science and Technology (CST) noted in its November-1984 recommendation to the government, Japan no longer has any model to follow into the future. It is, therefore, absolutely imperative for Japan to take the initiative in pushing forward to fashion the future. Particularly, the Council noted in introducing the guidelines for that crucial task, Japan must stake its existence upon the intellectual creativity of its people, the sole resource of the nation. Even more than other nations blessed with more abundant and diversified resources, Japan's future is linked to the further development of science and technology. The challenge is therefore clear.

In a period of relative stagnation of the world economy, advanced Western nations are confronted with severe constraints on their growth, largely due to an excessive allocation of resources to military pursuits which no longer assure the advance of civilian science and technology and tend to dissipate vital economic forces. Japan must contribute not only to global economic revitalization, but also to the forward thrust of all science and technology. To achieve these objectives and to assure their widest possible benefits, it is essential that they be pursued cooperatively with not only the leading industrial countries, but also the developing countries of the Third World.

THE NEW TECHNOLOGIES

Research will take few new directions. Rather, fresh ideas will come to the fore, based upon a continuity of work already under way. Unexpected insights may emerge to spur new levels of exploration through the combination of new technologies, but basically the way ahead is clearly pointed toward reaping the full benefits of microelectronics, genetic engineering, new materials, quantum physics and astronautics already fully in view.

If the direction of future development is discernible in R&D programs already well advanced, the pace of change is indicated in the momentum they have gathered in their formative stages. Over the years, the momentum has increased as successive new waves of technology have joined in. As a result, in the four decades since the end of World War II, the pace of innovation in Japanese industry has accelerated steadily, overtaking and surpassing the most advanced level of technology in the West.

The missing characters in this drama of rapid technological change, it has been commonly observed, were the scientists. University laboratories were under-equipped and starving for funds. And government research institutes worked quietly in modest facilities, spread out randomly over metropolitan Tokyo.

But, as if the urgent need for enhanced emphasis on basic research were foreseen, by 1983 a new Science City emerged, full-blown, on the Tsukuba plain some 60 kilometers north of Tokyo. Where just a decade before half a dozen country towns were clustered amid pine forests and farmland, 51 governmental and private research institutes, plus two new universities, comprising more than 1,700 buildings, brought together more than 11,000 researchers in what is generally agreed to be one of the world's largest and best-equipped research centers. Belying its reputation for parsimony in expenditures on basic research facilities, the Japanese government invested more than $5.3 billion in the Tsukuba installations from 1970 to 1983, providing the necessary conditions and equipment for creativity in advanced science.

By early 1985, the foundations for a massive creative effort by Japanese industry were in place. In the past two years, there has been a feverish rush to establish laboratories for basic research. Large integrated firms such as Hitachi Ltd., NEC Corp., Toshiba Corp., Fujitsu Ltd., and Matsushita Electric, have all upped R&D generally, budgeting approximately 20 percent of their collective R&D outlays for fundamental research.

Moreover, the current boom in basic-research investment is not confined to electronics. Biotechnology and new materials are the other main fields of high technology targeted by the majority of the new basic-research laboratories as the most important for corporate survival in the 21st century. Pharmaceutical, chemical, and food companies such as Eisai Co., Mitsubishi Chemical Industries, Asahi Chemical Industry, Kyowa Hakko, and Ajinomoto, are devoting sharply increased resources to fundamental research in fields such as protein engineering for the production of artificial proteins and new drugs.

With this vastly increased flow of corporate funds into basic research, government spending will increase apace in selected fields of science and technology in the coming years. Together, private and governmental R&D outlays are expected to rise to 3.5 percent of national income, which is considerably higher than the current average of 2.8 percent for West-

JAPANESE TECHNOLOGY TODAY

ern countries. This amounts to a quantum increase over the 1.7-percent level of 1977, when the gross national income was much smaller than at present.

But domestic expenditures no longer reflect the total R&D effort, particularly for multinational corporations. A growing number of Japanese firms are also budgeting for basic research in their subsidiaries abroad. Kyocera, the Kyoto-based ceramics high-flyer, has constructed a research center in Washington state to be staffed by both Japanese and American researchers. Sumitomo Electric will establish a combined production-research facility for optical fibers in North Carolina. U.S.-based facilities are under way for R&D in both office equipment for Ricoh and automotive electronics for Nippon Denso. And a number of Japanese are funding research at major American and European universities.

In short, a whole new technological system is emerging in Japan to chart new directions into the 21st century. Like the Japanese manufacturing systems that developed during the 1970's and 1980's, this new synthesis of science and technology to accelerate the processes of creativity has its distinctive features.

Among them is a systematic approach to identify R&D priorities. Comprehensive surveys have been conducted by government agencies, private research institutes, banks, and other firms dealing in securities to assess the importance and the development of virtually all industrial technologies, present and future. Those which are most advanced in Japan are targeted for further R&D necessary to sustain the lead, and those which are at once lagging behind the U.S. and susceptible to more rapid development in Japan are identified for major efforts to bridge the gap (see Table 1).

Anticipating the 1984 recommendations of the CST, MITI designated 12 R&D projects in its vision for the 1980's, to develop basic technologies for industries expected to flourish in the 1990's. In each industry, the key technologies require 10 or more years' investment for development, which is too risky for private enterprise. The technologies selected as the focus of MITI assistance were all critical to the further development of microelectronics, new materials, and biotechnology (see Table 2 and Table 3).

On the face of things, nothing seems very special about these major clusters of technologies earmarked for major growth to become the backbone of new industries of the 1990's and the 21st century. They seem no different from those of other advanced countries. But the management of each of these clusters of technologies differs significantly from that in the U.S. and Europe. Furthermore, the synergism achieved in integrated industrial structures driven by the dynamic process of creative destruction and diversification into higher value-added production makes a distinctive pattern and direction for accelerated technological change.

Electronics technologies make up by far the most important cluster. Heavy emphasis is being given to computers, communications, and consumer electronics. Automotive and medical electronics, and office and factory automation, are also fields of intensive R&D. And these are all supported by massive investments in microelectronics and optical electronics.

But what is amazing about all of this is the following: As of 1984, of the total expenditures on electronics R&D, approximately 95 percent was funded by private industry, compared with about 50 percent private funding in the U.S. And, in spite of the fact that the total spending on research was considerably less than in the U.S., the Japanese electronics industry has taken the lead in an expanding range of technologies—across the board in consumer electronics, including high-definition television as well as other video and audio technologies; in optical communications and facsimile technology; in robotics and factory automation; and in most passive components and commodity semiconductor devices.

Second only to electronics, biotechnology has become a high-priority focus of the Japanese R&D effort. Nervous observers in the U.S. see the possibility of Japanese technology's moving into the forefront of a field that might replace chemistry in importance before the end of the decade. This could happen.

Japanese biotechnological research begins with solid foundations in highly sophisticated fermentation technology, which is deeply implanted in the Japanese experience. More important, it is precisely those firms with one to three centuries of experience in traditional fermentation technology that are leading the development of the new biotechnological system. By joining traditional and new biotechnologies in the same firms, Japanese research benefits from a rare synthesis of men, money, and production capabilities, which in turn enables these firms to bring new technologies to market more rapidly and at less cost than can the smaller-scale, venture-capital–backed businesses in the U.S.

Materials technologies form a third

TABLE 1. JAPAN'S COMPETITIVENESS IN HIGH TECHNOLOGIES

TECHNOLOGIES	JAPAN	U.S.	EUROPE
ELECTRONICS			
Industrial Robots	X		
CAD		X	
Computer Services		X	
Fixed Disk Drives		X	
Medical Electronics		X	
Sensors	X		
VLSIs	X	X	
Semiconductor-Prod. Equipment	X		
CATV (service and equipment)	X	X	
Communication Satellites		X	X
Videotex	X	X	X
NEW MATERIALS:			
Engineering Plastics		X	
High-Performance Polymers		X	X
Fine Ceramics (functional)	X		
Fine Ceramics (structural)		X	
Amorphous Alloys		X	
High-Purity Silicon	X	X	
Gallium Arsenide	X	X	
BIOTECHNOLOGY:			
Plant Factories		X	X
Biopharmaceuticals	X	X	
Biotech Equipment		X	

"X" indicates relative lead.
Source: Industrial Bank of Japan

Introducing a Nikon for total beginners who don't plan to stay that way.

This picture was taken by setting the N2000 in the programmed mode, focusing and pushing the shutter button. Simple.

This picture was taken in the manual mode, with a Nikon 15mm super-wide angle lens, between f 11 and f 16 at ¼ second. Exposure compensation of + ½ stop, flash fill and skylight filter. Not so simple.

If you're new to 35mm SLR photography, there's no better place to begin than with the new Nikon N2000.

It's auto-wind, auto-load, and can automatically set the film speed.

And with programmed exposure control, all you have to do is focus and shoot.

But you won't be a beginner forever. So the N2000 also has a high-speed programmed mode, aperture priority automatic and manual modes, TTL flash control, plus the ability to accept over 60 Nikon lenses.

So as you progress from beautifully simple pictures like the one on the left to complex beauties like the one on the right, one thing you won't have to change is your camera.

Nikon®
We take the world's greatest pictures.®

Nikon USA CLUB *TO RECEIVE FREE MEMBERSHIP IN THE NIKON USA CLUB, RETURN THE NIKON USA WARRANTY APPLICATION. © NIKON INC., 1985* cluster in the emerging system of research-intensive industries. In recent years, Japanese industry has focused resources on the development of a wide range of new ceramics, engineering plastics, amorphous metals, and carbon fibers. Chemical compound semiconductors, particularly gallium arsenide; the hydrogen-storing alloy essential for deriving energy from hydrogen in a practical way; the optical fiber which forms the sinews of information network systems; ceramics used in microelectronic device substrates and packaging; and engineering plastics used in video tape recorders: all have undergone rapid development in recent years to become the driving force of high-flying entrepreneurial firms such as Kyocera and to revitalize entrepreneurial companies epitomized by Sumitomo Electric, the copper-cable maker that has emerged as a world leader in optical-fiber technology.

Improving the utility of ceramics is clearly a major thrust of research. More than 300 companies, including major electronics, metals, and traditional industrial ceramics firms, are pushing forward the state of the art. Made from metal oxides that constitute 90 percent of the earth's crust and thus assure an almost inexhaustible supply of raw material, fine ceramics have many applications and are critically important to high-technology areas such as semiconductor production, development of new energy sources, optical electronics, biological mimetics, and aerospace manufacturing.

But in addition, fine ceramics hold promise for a wide variety of more down-to-earth applications in internal combustion engines, machinery, tools, and magnetic materials. Heat-resistant and structural ceramics are also being perfected for use in high-temperature gas turbines, nuclear reactors, and magneto-hydrodynamic (MHD) power generating systems.

In spite of delays, mainly due to quality-control problems, the fine-ceramics market in Japan had reached ¥630 billion in 1983, and is expected to grow to between ¥1.24 trillion and ¥1.41 trillion in 1990 and to more than ¥3 trillion in 2000. In the process of development, fine-ceramics technology will serve as an instrument of revitalization for the materials industry, resulting in a considerable reduction in petroleum consumption and a more sparing use of rare metals.

Two other clusters of technologies targeted for intensive basic R&D, alternative energy and aerospace, may not come to fruition as soon as microelectronics, biotechnology, and fine ceramics. But the pervasive application of the latter three technologies will set the stage for a radical restructuring of the energy business, based upon a widening range of alternative energy sources—from fuel cells and photovoltaics to nuclear fusion and MHD.

THE NEW SYNTHESIS

While microelectronics technologies are now well advanced in Japan, and the large, vertically integrated and highly diversified electronics/communications/computer manufacturers now have third- or fourth-generation laboratories that are excellently equipped and churning out new products at a dizzying pace, biotechnology and new materials are only now on the verge of commercial applications, or will be within a decade or so. Most important, however, is the realization that the full effect of these three basic technologies, along with new energy and aerospace, comes not from isolated specialization but from innovative juxtaposition or combination.

Japanese industry has proven particularly adept at managing this synthesis of

TABLE 2. BASIC TECHNOLOGIES FOR FUTURE INDUSTRIES: MITI'S PROJECTS FISCAL YEARS 1981–1990 (Costs in billions of yen)

THEME	TERM [YRS]	BUDGET FY '83	DEVEL. COST	RECIPIENT*
NEW MATERIALS:				
1. High-performance ceramics	10.00	0.84	13.00	TRG (Fine Cer.)
2. Synthetic membranes for new separation technology	10.00	0.52	10.00	TRG (High-Polymer Infra.)
3. High-performance plastics	10.00	0.33	5.00	do
4. Synthetic metals	10.00	0.30	6.00	do
5. Advanced alloys with controlled crystalline structures	8.00	0.56	8.00	R&D A. (Next Gen. Comp. Materials
6. Advanced composite materials	8.00	0.65	11.00	do
BIOTECHNOLOGY:				
7. Bioreactors	10.00	0.45	11.00	TRG (Biotech Devel.)
8. Large-scale cell cultivation	9.00	0.38	5.00	do
9. Utilizing recombinant DNA	10.00	0.36	10.00	do
FUTURE ELECTRON DEVICES:				
10. Superlattice devices	10.00	0.42	8.00	R&D A. (New Electron Device)
11. Three-dimensional ICs	10.00	0.73	9.00	do
12. Fortified ICs for extreme conditions	8.00	0.30	8.00	do
Total 12 themes		5.84	104.00	

TRG = Technology Research Group
R&D A. = R&D Association
*Participants are listed in Table 3
Source: MITI

JAPANESE TECHNOLOGY TODAY

high technologies. Optics has been combined with electronics to obtain an exciting and mind-boggling range of new products cutting across the boundaries of existing industries and giving birth to new ones. Mechanics has been blended with electronics in myriad applications to produce new families of robots, machine tools, and home appliances in exponential profusion. Photovoltaics using silicon wafers have been linked to the production of microelectronic devices and solar batteries used in pocket calculators and wristwatches. And the combination of microbiology and electronics is about to produce still other synthetic technologies, bioelectronics and bio-optical electronics, to probe the frontiers of the unknown.

Although this trend toward synthesis is evident in Western countries, their obsessive emphasis on specialization still tends to compartmentalize high technologies in smaller production units managed by highly qualified Ph.D.'s and financed by venture capitalists. In Japan, synthesis has become the predominant pattern. Vertically integrated, horizontally diversified firms and the Japanese capacity for team play combine to provide the necessary conditions for the most efficient management of synthesis. Not coincidentally, then, it is the Japanese who are adding to Western vocabularies zany new terms such as "mecha-tronics" and acronyms such as "C&C" to express the strategic marriage of computers and communications in a common industrial strategy.

As a result, previously clear industrial lines are becoming blurred, making it virtually impossible to distinguish the new ones as they emerge. The embryonic bio-industry in Japan, for example, is composed not of specialized, new biotechnology businesses, but rather of firms that are classified as sake brewers, soy-sauce makers, food and dairy processors, chemical companies, pharmaceutical houses, textile manufacturers, steel makers, or electrical-equipment producers. A company in the forefront of biotechnology research may also appear among the leaders in new-materials technology. Appliance makers with wide experience in mecha-tronics are emerging as leaders in the robotics race.

Combining technologies, driven mainly by the imperative of diversification to assure continued growth and sustained employment, has also enabled Japanese engineers to use their experience to design high-quality new products. Take copiers, for instance. When Japanese camera makers began making copiers, quite naturally engineers built them the same way they had made precision optical equipment. "We conceive copiers in the way we conceive cameras," Sam Kusumoto, president of Minolta's U.S. sales subsidiary, reflects. "Copiers have lenses, camera obscura, and papers instead of film. All sorts of mechanisms are assembled in a very limited space to form a compact unit. That is why Japanese copying machines are more elaborate than those of U.S. makers."

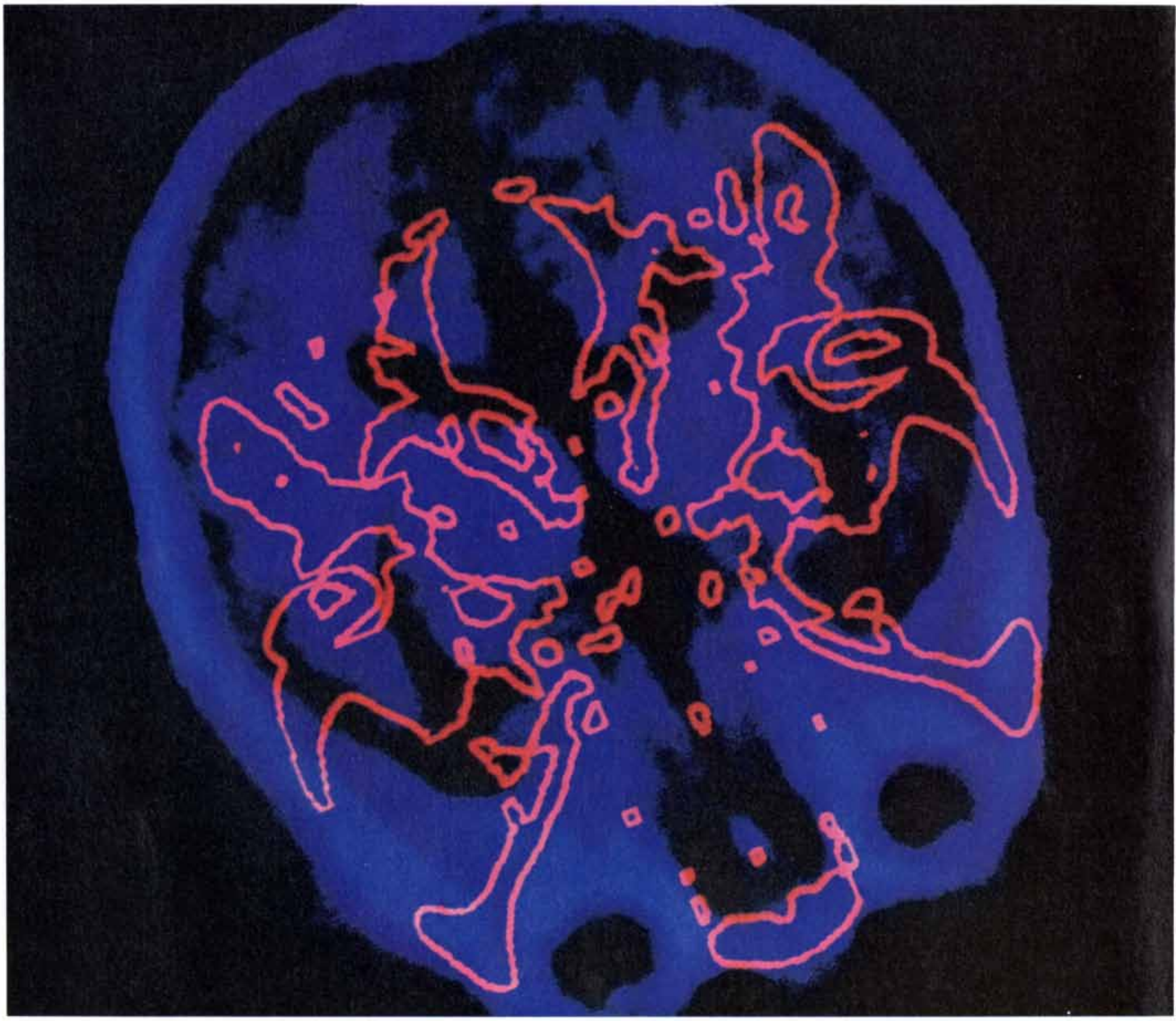
Now the experience of computerizing cameras is being applied to copiers. The idea is to develop a machine that can instantly make a copy of information stored in a computer, without converting that information into graphs or characters. Currently, computer data must first be projected on a screen before being copied. But the intelligent copier will skip that process. "With further progress," Kusumoto adds, "we will eventually have newspapers transmitted via data communications networks to computer terminals and produced on copiers. Copiers can also be attached to every TV set to reproduce in hard copy anything that appears on the screen that the viewer wants to keep. In the future, the copying machine will be one of the more intelligent of computer and communications terminals."

The logic of synthesis in the management of high technology is equally compelling. Not only is the combined impact of the various technologies far greater than the impact of each managed separately, dramatically broadening the range of products and accelerating the pace of innovation; but this broadened product range and accelerated innovation are also functions of multiple relational syntheses: those of government and industry, of companies in the same industry

TABLE 3. BASIC TECHNOLOGIES FOR FUTURE INDUSTRIES: PARTICIPANTS IN MITI'S PROJECTS

THEME	PARTICIPANTS
1. High-performance ceramics	Toshiba, Showa Denko, Denka, Kyocera, Shinagawa Hakurenga, Asahi Glass, Toyoda Koki, Kobe Steel, Nihon Tokushu Togyo, Toyota Motors, Inoue Japax, IHI, Kurosaki Yogho, Nihon Glass, Sumitomo Denko
2. Synthetic membranes for new separation technology	Asahi Chemical, Asahi Glass, Kuraray Sumitomo Denko, Toyobo, Daiseru, Teijin, Toray, Mitsubishi Kasai
3. High-performance plastics	Asahi Chemical, Sumitomo Chemical, Sumitomo Denko, Teijin, Toray
4. Synthetic metals	Asahi Chemical, Teijin, Toray, Mitsubishi Kasei, Mitsubishi Yuka
5. Advanced alloys with controlled crystalline structures	Hitachi, Kobe Steel, Daido Steel, Mitsubishi Metal, Hitachi Metal, Sumitomo Denko, IHI
6. Advanced composite materials	FABRICATION: Mitsubishi Heavy Industries, Fuji Heavy Industries, Toyota Motors, Toshiba Kikai, IHI, Mitsubishi Denki, Kawasaki Heavy Industries HIGH POLYMER: Teijin, Toray, Mitsubishi Kasei, Nihon Carbon
7. Bioreactors	Kao, Daiseru, Denka, Mitsui Chemical, Mitsubishi Kasei, Mitsubishi Gas
8. Large-scale cell cultivation	Asahi Chemical, Ajinomoto, Kyowa Hakko, Takeda Pharmaceutical, Toyo Jozo
9. Utilizing recombinant DNA	Sumitomo Chemical, Mitsui Toatsu, Mitsubishi Chemical
10. Superlattice devices	Fujitsu, Hitachi, Sumitomo Denko
11. Three-dimensional ICs	NEC, Oki, Toshiba, Mitsubishi Denki, Sanyo, Sharp, Matsushita
12. Fortified ICs for extreme conditions	Toshiba, Hitachi, Mitsubishi Denki

Source: MITI



A medical breakthrough: Hitachi computed tomography yields clear color pictures of both bones and organs, such as this cross-section of a human skull, by combining images obtained by using X-rays and superconductive magnets.

engaged in cooperative research projects, of customers and manufacturers, of the laboratory and factory floor, of different divisions within a company, and of management and labor within the enterprise. Each of these relationships, when imaginatively managed, serves to attain the optimal synergism by synthesizing complementary resources and actions.

To the amazement of foreign observers, in the much-touted very-large-scale-integration (VLSI) project of the late 1970's, Japanese semiconductor makers spent vast sums on the development of new production technologies which were then shared with smaller equipment makers at little or no cost to the latter. This continuing collaborative relationship, the synthesis of customer and equipment vendor, assures the rapid diffusion of new microelectronics interactive technology and accelerates progress in Japanese microelectronics technology as a whole.

But industrial equipment is not the only sector in which this synthesis of user and vendor is a determinant for manufacturing. Medical-equipment manufacture is in a class by itself.

Few medical-equipment makers have managed this critical synthesis of producer and user more effectively than has Olympus Optical Co., Ltd., in the development of fiberscope technology. Since 1950, when 30 gastroenterologists gathered at the first "gastrocamera" meeting in Tokyo, Olympus has systematically worked closely with distinguished members of the medical community the world over to perfect a wide range of fiberscope instruments, combining expertise in optics and electronics with the latest developments in medicine and related sciences. "Our working relationship with the medical community is most important," Olympus President Toshiro Shimoyama observes. "We learned this from our experience with microscope manufacture. We can design microscopes by ourselves, of course. But it is much more important that professors, doctors, and lab technicians who use these instruments participate in the design stage, to assure their usefulness. Nor can we leave this participation to chance. We organize our external communications and cooperation very carefully to assure professional user inputs to the design and production at all critical stages.

"In the development of our fiberscopes and related equipment, we have invited not only Japanese doctors to contribute to this process," Shimoyama notes, "but doctors from other countries as well. Many doctors have come from European countries, even Eastern countries—many from Russia, China—at our invitation, to study the use of the gastrocamera and fiberscope at research centers in Japan. And this process has continued for over 30 years.

"We do not confine this cooperation to the equipment itself," he explains further. "Our research laboratories also have close working relations with universities for the development of semiconductor devices needed for the equipment. Take our static-induction transistor. It was developed by Professor Nishizawa of Tohoku University, to our specifications. He invented the device and we make it. And we have similar working relationships with Johns Hopkins University and others throughout the world."

As a result of this mutual effort, today thousands of doctors around the globe are using Olympus fiberscope instruments. Few would consider any alternative. Olympus has succeeded in developing a unique technological sensitivity to doctors' needs and an ability to adapt its products to meet medical requirements. An 80-percent market share is a powerful testimonial to the effective synthesis of equipment maker and user.

For technology to respond appropriately to continuing feedback from users, it is essential that it be applications-oriented and that the entire process be conceived of as an incremental change. And the focal point of change is the factory floor, where ideas are transformed into useful products. This pervasive pragmatism, rather than any cultural constraints, explains the Japanese rejection of the "big bang" approach to invention and the apparent lack of interest in gargantuan leaps forward into the unknown.

Experience in the semiconductor industry confirms to managers and policymakers the wisdom of this approach. Japanese integrated-circuit (IC) makers have succeeded in moving ahead in the race to pack increasingly higher densities of elements and circuitry on a minuscule silicon chip, at least in part because they have approached the innovative process incrementally, building step-by-step on existing technology and know-how rather than attempting to stretch their creative powers to the limits at each stage in the process. As a result, new generations of integrated circuits have been produced efficiently and up to the necessary quality levels sooner than the competition abroad. Fujitsu America president Masaka Ogi points out, "in areas such as fiber optics or semiconductors, we can compete with or even surpass American companies." Epson America president Yasuhiro Tsubota agrees, "the strongest advantage of Japanese manufacturers is their production technology which enables them to produce high quality at low cost."

Step-by-step innovation draws deeply on past experience at all stages of the organization, from the laboratory to the point of sale, thus obtaining optimal advantage of the learning process.

The new emphasis on basic research does not represent a fundamental departure from this practice, but rather will mainly constitute its broadening and extension. This, as often as not, will entail injecting new technology into old products. Just as Mitsubishi Heavy Industry is designing ships by computers to be navigated by computers, Mitsubishi Motors is building advanced electronics and new materials into its future models of passenger cars; home computer terminals are being built into what once was a plain old telephone; facsimile devices are being added to television receivers; and all are being combined into integrated information systems.

Just as quality and efficiency are the business of everyone in an enterprise, so is new-product development. Here is the ultimate synthesis, from the laboratory to the consumer, with the factory floor as the vital link.

At Sumitomo Metals, Executive Vice President Hiroshi Kojima reflects, this synthesis is achieved through the *jishu kanri* (JK) system, which seeks to combine individual development and work satisfaction through quality control and creativity at all levels of the company. "These voluntary, small-group JK activities trace their origin to the American ZD [zero-defects] movement," Kojima points out. "But the objectives of this movement have been adapted to the Japanese way of people working together in small groups. Our excellent personnel themselves should, in my view, be taken as a sort of technology in the widest sense of that term."

This synthesis finds its most effective expression not in formal functional structure or profit-center organization units, but in project teams involving representatives of all units of the company concerned with the development of a particular product or technology. At Hitachi, this project-team approach to R&D management is formalized in a system called *tokubetsu kenkyu* (special research), or simply *tokken* for short, in which three or four project teams are formed each month, drawing their members from the appropriate divisions or subsidiaries of the group. Teams may work together for two, even five, years drawing on re-

INTRODUCING MINOLTA

MAXXUM™

AUTOFOCUS SYSTEM

ONLY THE HUMAN EYE FOCUSES FASTER.

Today, you can experience a quantum leap in 35mm SLR photography.

Minolta's new Maxxum is the easiest SLR ever; the only one with automatic focusing built right in.

Imagine! You don't have to fiddle with a focusing ring. With uncanny speed and amazing accuracy, Minolta Maxxum actually *sees* your subject and automatically snaps it into perfect focus for sharp pictures every time. All you do is compose and shoot.

Maxxum's built-in Motorized Film Control, with DX capability, loads, advances, rewinds the film and allows you to shoot up to two frames a second.

Maxxum's exciting array of Autofocus lenses includes five versatile zooms. Its exclusive Automatic Multi-Program Selection (AMPS) gives you the best possible program for the lens you choose. And it automatically changes programs as you zoom.

You can also shift into Aperture Priority, Shutter Priority or Creative Manual to call your own shots.

This speed, ease and flexibility springs from breakthrough technology protected by over 300 patents, including more than 30 World's Firsts in camera engineering. It's a system so sophisticated that, with Maxxum Flash units, it even focuses in the dark.

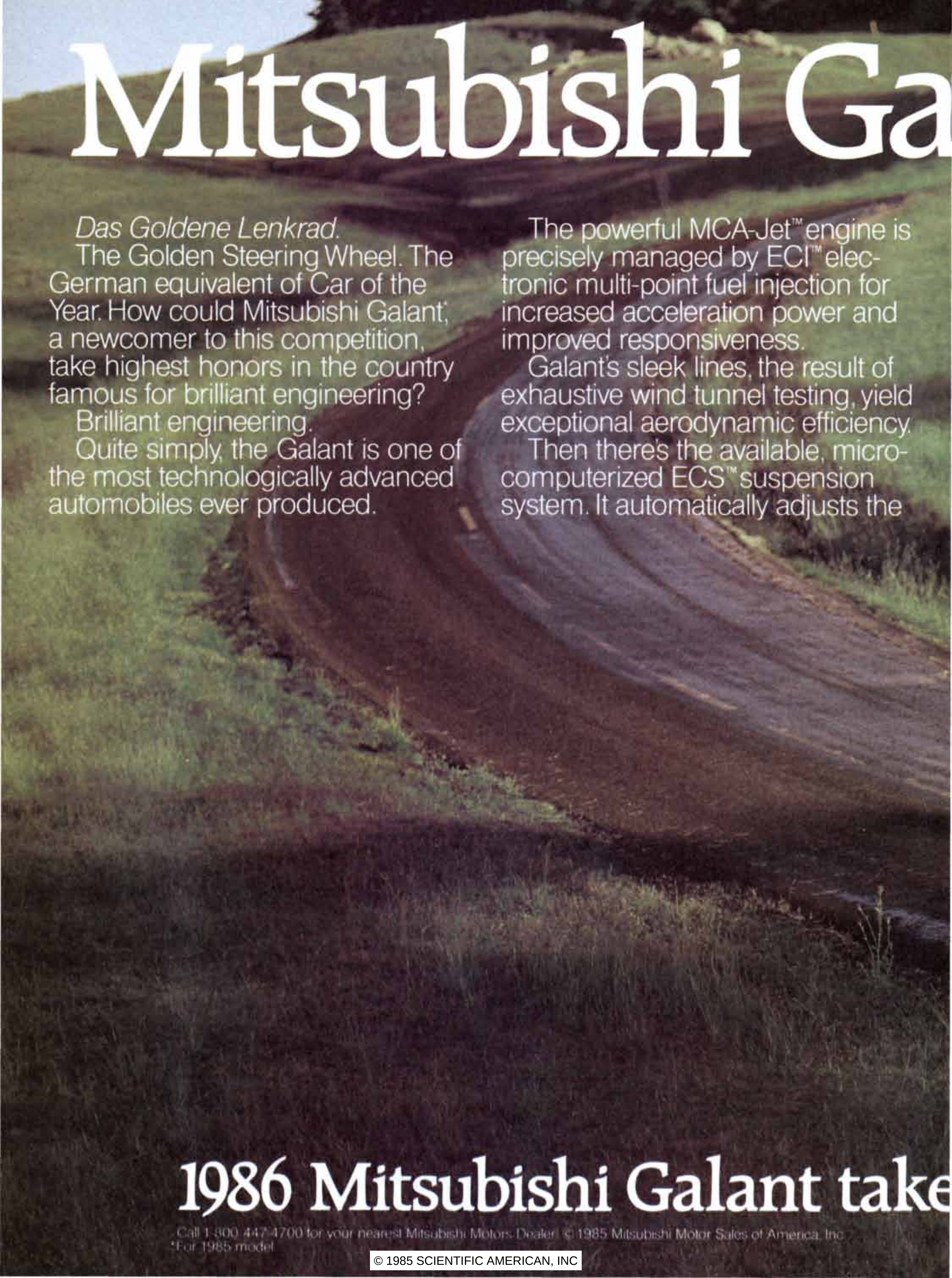
Experience Maxxum now. Get perfect pictures in the blink of an eye.

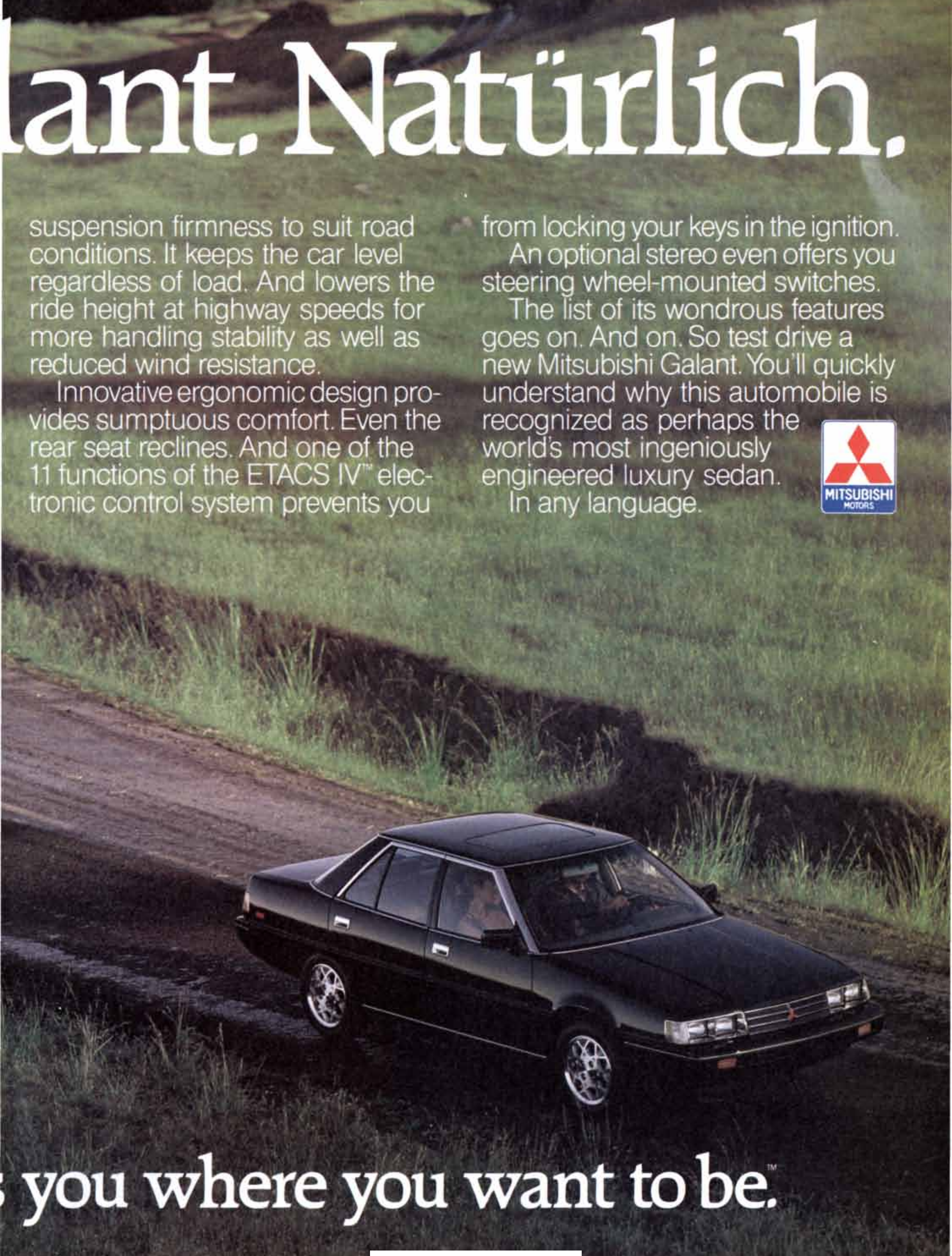
EUROPEAN CAMERA OF THE YEAR 1985

Be certain that the valuable Minolta U.S.A. 2-year camera/5-year lens limited warranty cards are packaged with your products. For more information, see your Minolta dealer or write: Minolta Corporation, Dept. MX-7, 101 Williams Drive, Ramsey, NJ 07446. In Canada: Minolta Canada, Inc., Ontario. © 1985 Minolta Corporation.

ONLY FROM THE MIND OF MINOLTA

MINOLTA


Mitsubishi Ga
Das Goldene Lenkrad.
The Golden Steering Wheel. The German equivalent of Car of the Year. How could Mitsubishi Galant,* a newcomer to this competition, take highest honors in the country famous for brilliant engineering?
Brilliant engineering.
Quite simply, the Galant is one of the most technologically advanced automobiles ever produced.
The powerful MCA-Jet™ engine is precisely managed by ECI™ electronic multi-point fuel injection for increased acceleration power and improved responsiveness.
Galant's sleek lines, the result of exhaustive wind tunnel testing, yield exceptional aerodynamic efficiency.
Then there's the available, micro-computerized ECS™ suspension system. It automatically adjusts the
1986 Mitsubishi Galant take
Call 1-800-447-4700 for your nearest Mitsubishi Motors Dealer. © 1985 Mitsubishi Motor Sales of America, Inc.
*For 1985 model


ant. Natürlich.
suspension firmness to suit road conditions. It keeps the car level regardless of load. And lowers the ride height at highway speeds for more handling stability as well as reduced wind resistance.
Innovative ergonomic design provides sumptuous comfort. Even the rear seat reclines. And one of the 11 functions of the ETACS IV™ electronic control system prevents you from locking your keys in the ignition.
An optional stereo even offers you steering wheel-mounted switches.
The list of its wondrous features goes on. And on. So test drive a new Mitsubishi Galant. You'll quickly understand why this automobile is recognized as perhaps the world's most ingeniously engineered luxury sedan.
In any language.
MITSUBISHI
MOTORS
you where you want to be.™

sources of the various organizational units represented to carry out their R&D work. Electron-beam lithography was a project of one such *tokken* team working in tandem with the five-company cooperative VLSI research association in the latter half of the 1970's. Synthesis of resources, expertise, and energy was thus obtained on two levels of interorganizational cooperation: first, interdepartmental within the group, and second, intercorporate between the leading makers in the semiconductor industry. Both of these were ultimately able to draw upon the basic research work done at MITI's Electronics Technical Laboratory in Tsukuba.

In industry, the effect of this *tokken* system is to enhance the entrepreneurial capacity of large-scale enterprises. The net effect is of singular import, giving to Japanese innovation a sustained renewing force and ability to carry through that is often absent in the U.S. and Western Europe. Since large-scale firms are prepared to invest in R&D and accept the risk of developing new technologies, there is little need to call on venture capital. New ventures are launched within established organizations.

"Venture capital in the U.S.," observes Yutaka Sasaguchi, president of Nikon's New York-based subsidiary, "is mainly concerned with making a fortune fast. Typically, venture capital wagers on an original design or concept and expands it rapidly into a very large firm, without time to build up a production base. And, to build market share fast, manufacturing is often located in countries like Singapore, Taiwan, or Mexico, where production costs are lower," Sasaguchi explains. "...But when they grow to compete with Japanese companies on an equal footing, they inevitably expose their weakness."

In sharp contrast, when the requirement for new approaches to microlithography in the VLSI era became evident in 1974, Nikon, a fine-optics company in search of opportunities for rational diversification, launched a long-term R&D undertaking to develop the required technology and production base. "In our case," Sasaguchi recalls, "the stepper technology was developed over a period of 10 years before we entered the market. It consists of many kinds of intricate techniques and incorporates more than a few original designs. In Japanese business and manufacturing, speaking of my own field of optical equipment, our development, processing, and marketing techniques are all proprietary know-how." Building on solid technological foundations and manufacturing experience leads not only to originality but also to a superb new production capability.

These advantages over the venture-capital approach are clearly manifest in both microelectronics and biotechnology. Although U.S. firms that use venture capital have been known for their ability to develop new technologies and bring their products to market, they often lack the necessary human and capital resources for sustained development, particularly in volatile markets or industries marked by radical, cyclical demand swings. Nor do specialized firms using venture capital have the opportunities for synthesis and instant in-house diffusion of technologies that vertically integrated and diversified firms have.

This capacity of major firms for investment in R&D has still another important consequence that distinguishes the Japanese approach to innovation. Since private industry is able and compelled, by market forces, to invest heavily in new-technology development, government is relieved of much of the burden. The emphasis on optimal synthesis in R&D management, then, is at once a cause and an effect of the predominant role of private enterprise.

As a result, in 1985 almost three fourths of the total ¥7.9 trillion spent on R&D will be in the private sector. Moreover, in high-technology industries such as electronics, that percentage is as high as 95 percent—a stark contrast to the 50-percent contribution of public funding for R&D in the U.S. electronics industry, where much of the military-related R&D has limited possibilities for synthesis within organizations or with other technological developments.

Nor is this pattern expected to change substantially with the increased emphasis on basic research. On the contrary, CST recommendations specifically call for governmental measures to extend preferential tax treatment to investments in basic research, and to credit guarantees to encourage private financial institutions to grant liberal loans for R&D funding. Other means of encouraging greater and more efficient use of capital markets for high-technology firms are also contemplated.

But money without men is not very useful in the promotion of new technologies. And here the Japanese approach to human-resource development has its distinctive features.

For well over a decade now, Japanese universities have been graduating more engineers per capita than any other advanced country. In fact, by 1980, more engineers graduated from Japanese universities than from U.S. universities. This means, since the population of the U.S. is twice as large as that of Japan, the number of new engineering graduates per capita was over 200 percent of the American level.

Equally important, Japanese industry is in a position to invest more in the continued education of engineers. Within three years after employment, new engineers, whose college training has been rendered obsolete by the rapid pace of technological change, begin a lifelong program of education within the major companies. Relying on in-house expertise, these training programs play a critical role in the management of technology synthesis.

The situation is not the same in basic sciences, however. Not only is the number of graduates in the various scientific fields relatively low, but neither the government, the universities, nor private firms are geared to the needs for continuing education of scientists. As a result, there is currently a dire shortage of qualified and experienced scientific researchers. This must be corrected if basic research is to develop, as called for by the CST recommendations.

Japanese universities are moving to increase postgraduate courses in both basic sciences and engineering, with support from government agencies and industry. And to involve these students in practical research during their postgraduate training, both government and industry are being actively encouraged to commission more of their research work to academic institutions. The government-industry-academia synthesis will, therefore, be an important feature of the new era of creativity.

Japanese policymakers understand full well that the new synthesis must be universal to be optimal. And they seem determined to pursue the logic of that synthesis to its ultimate realization.

Such a pursuit necessarily entails the globalization of structures: the optimization of technology on a global basis. Neither microelectronics nor biotechnology nor new-materials technology lends itself to development within restricted national markets. Economies of scale and learning can be achieved for basic technologies only on a global scale. High investment in R&D and capital-intensive products must be financed globally, in competitive marketing conditions. Customer service and market access often require both downstream assembly and production of components or other materials near the point of consumption. Effective management of R&D surpasses corporate and national capacities and often can be achieved only through synergy of re-

Are you overlooking a copier maker that's bigger than Xerox?

Look into the new Ricoh 6085.

Ricoh is a bigger maker of copiers than both Xerox® and IBM.® And we got to be so big by making a full line of copiers that win top marks for quality.

Like the new Ricoh 6085 console copier.

Even though it's new, the Ricoh 6085 is already number one in its class. It's the first mid-sized copier with so many features. For example, it gives you automatic duplexing, so you can make two-sided copies from one or two-sided originals. It also gives you a unique feature that lets you automatically combine two different images on a single copy. Plus automatic reduction and enlargement. And a really fast copy speed of 50 copies a minute.

But there's one thing about the new Ricoh 6085 that's not unique. It's made to the same high-quality standards all Ricoh copiers are.

So, look into Ricoh.

We're a copier maker that's too big to overlook.

800-222-0878. In NJ. 800-222-0879.

Ricoh Corporation, 5 Dedrick Place,
West Caldwell, N.J. 07006.

- ☐ *I'd like to look into the new Ricoh 6085.*
- ☐ *I'd like to look into the full line of Ricoh copiers.*

Name ________________

Company ________________

Address ________________

City ________ *State* ______ *Zip* ______

Phone ( )

RICOH®
We respond.

SA—11

sources that cut across national boundaries.

This new synthesis must be sought through global strategies. Thus, in 1984, the top nine Japanese electronics companies reportedly had 232 manufacturing facilities abroad, accounting for approximately 60 percent of the total number of overseas production units of all Japanese electronics companies. In biotechnology and new materials, however, globalization has taken quite a different form. Since Japanese firms abroad are all engaged in the development of new technologies, beginning from just about the same levels of technological development and capital-intensive R&D, there has been a tendency to spread risk and pursue synergies through cross-licensing, investments in joint research, or joint production and marketing ventures.

What has gone almost unnoticed is a change not merely in the geographical pattern of production and innovation but, even more fundamentally, in the entire approach to market building. Powerful forces have compelled Japanese firms to redefine efficiency in global terms which take into account secondary, not merely primary, effects, and which are consistent with a wide range of economic, social, and humanistic concerns.

When the Japanese electronics industry began developing its global strategies in the 1950's, it approached the task with remarkable adaptability. New production systems were developed to match new solid-state-electronics technology to the exigencies of market expansion. Rather than approach Southeast Asian markets with traditional mercantile strategies that sought maximization of short-term profits and a quick return on investments, Japanese electronics companies sought to externalize labor costs and benefits by a do-it-yourself approach of downstream production.

Matsushita Electric began this process of externalization over 25 years ago, with the establishment of its first overseas production unit in Taiwan to produce transistor radios and television receivers. Today, the leading Japanese consumer-electronics company has 47 manufacturing companies and 36 sales companies in 37 countries. "The policy of Matsushita Electric," says Dr. S. Kisaka, Executive Vice President of Matsushita Electric Industrial Co., "is to establish a company which will contribute to the country in which we operate. Our main aim is to be welcomed by the local people, not to make a lot of money there and bring it back to Japan. Our basic idea is to establish a company which will be run by local people for local people."

Overseas investments have grown apace with Matsushita's technological advance. In the 1960's and first half of the 1970's, investments were mainly in developing countries. With Matsushita's lead in color-television technology, production facilities were established in the U.S. and Western Europe. "But VTR [video-tape-recorder] technology has had the greatest influence on the company's internalization," according to Matsushita America Vice President Akira Yokoi. Matsushita's strong position in the American market has been sustained by domestic production backed up by a continued high level of Japanese R&D investment. The key, Yokoi believes, is Matsushita's long-range view. "This is the foundation of our technological strength," he asserts. "RCA and other U.S. companies developed some good VTR and videodisc technology, but since the development of these products did not progress as quickly as expected, R&D budgets were cut. We made many mistakes over the years, but we persevered. And the accumulated technology has now borne fruit. The important thing about R&D is continuity."

Significantly, the development of VTR technology at Matsushita began about the same time the first investments were made in overseas production. In a very real sense, continuity of R&D becomes the fulcrum of global strategies, and global strategies in turn assure the critical resources that make such continuity possible. The two are mutually reinforcing.

The Matsushita policy of value-added externalization through foreign direct investments is not the only way to optimize global returns from R&D investments, of course. Ricoh is the classic example of an innovative firm that has sustained a high level of R&D by cooperative development and original-equipment-manufacturer (OEM) supply. According to Yukihiro Wada, President of Ricoh America, under cooperative development agreements with AT&T, Digital Equipment Corp., and Hewlett-Packard, his company supplies printers, facsimile terminals, and copiers for sale under the buyer's brand. The U.S. partners provide the software and marketing, in most cases, and participate in the product-development process; Ricoh's main contribution is production of the hardware, employing technologies developed through heavy investment in R&D. From its inception as a creation of the renowned Institute of Chemical and Physical Research, Ricoh has sought to maximize market strength through innovation, primarily by sustaining a high level of research activity.

What we see, then, is a pattern cutting across many high-technology industries: increasing externalization through increasing investment in overseas production or technological cooperation. Ultimately, this sustains a high level of innovative R&D.

At what point and for what industries and in what countries will this externalization be extended to include R&D itself? Developing countries insist that R&D investments should be made in their countries by multinationals, including Japan, to assure the transfer of technology. In advanced countries, foreign R&D laboratories are viewed as a mixed blessing: They may be receptors rather than transmitters of technology.

Clearly, externalization of R&D must be rational and mutually beneficial. At present, only the largest Japanese electronics companies find internationalization of R&D necessary and practical. Hitachi, Matsushita, and Sony have laboratories in the U.S. Sony America president Kenji Tamiya points out, "it would be too troublesome and time-consuming if we tried to develop new products in Tokyo in response to requests from our American customers." But, as Dr. Yahiko Yamada of Hitachi America's Technology Planning Group candidly admits, "The main purpose of R&D in the United States is to gather information on the most advanced technologies developed in Silicon Valley and adapt them quickly." Observes JVC's general manager for corporate planning and development in New Jersey, Junichi Egawa, "the scale of our business is becoming larger every year, and we cannot keep abreast of developments unless we are in constant contact with engineers of leading companies in the United States."

Similar investment patterns are emerging in the automotive industry, but for different reasons. Research by Japanese automakers in the U.S., for example, aims mainly at adapting cars to the tastes and technical standards of the market. "We have established an institute in Los Angeles to design cars and work on other projects," Mitsubishi Motors Executive Vice President Shinji Seki explains. "We have also established a laboratory in the States to perform research concerning exhaust gas. But there is very little technology in the U.S. which we need to learn about...."

THE MICROELECTRONICS REVOLUTION

Of the three main technologies poised for major growth in the remainder of this

JAPANESE TECHNOLOGY TODAY

century and well into the next, microelectronics has made the most remarkable advances in Japan, providing a foretaste of what can be expected from the other two. The Japanese electronics industry has already become the world's most diversified producer of electronic equipment. A relatively recent entrant to the industrial scene, electronics manufacturing and related software now account for more than ¥22 trillion in revenue each year, which is five times what it was in 1975. And by the year 2000, output of electronic products in Japan is expected to reach ¥153 trillion. This will make it Japan's largest industrial sector, surpassing the total output of steel, chemicals, and cars. Electronic machinery, including equipment and components, will already surpass automobiles in size of production by the end of this decade.

With the completion of the world's first major cooperative research undertaking in microelectronics, from 1976 to 1979, the Japanese electronics industry developed the leading-edge VLSI technology in the field—not based on technology imported from abroad but rather from original research.

During this stage, Japanese R&D in electronics shifted emphasis from mainly incremental improvements of existing technologies to the development of original technologies.

The most significant change to come from this shift is a sharp increase in R&D allocations. The electronics industry as a whole invests a substantially higher share of total sales than all others: 4.9 percent, compared to 3.3 percent in the chemical industry, 2.7 percent in the automotive industry, 2.6 percent in the machinery industry, and 1.6 percent in the steel industry. According to a recent report, "Key Players in the Japanese Electronics Industry," by Dodwell Marketing Consultants, R&D expenditures by the Japanese electronics industry will surpass a trillion yen by the end of 1985. Seven of the top 11 firms, ranked by levels of spending, were leading electronics companies in as early as 1979. But since then the role of these firms has mounted sharply. Not only have sales climbed at higher annual rates than in most other sectors; but the share of R&D expenditures has steadily increased.

These fundamental changes are likely to be accelerated as the Japanese electronics industry leads the way into the next generation of ultrahigh integration, 4 megabits (4M) and beyond. In February last year, Toshiba and Fujitsu took the initiative in introducing 1M technology. By mid-1985, all major Japanese semiconductor makers had developed and begun sampling 1M DRAM devices. Since Japanese suppliers had garnered 90 percent of the world market for 256K devices, this shift to higher levels of integration was largely in their hands. And prospects are good that the timing of the shift to 4M will be determined in Tokyo. "I think the game is over as far as development of semiconductors is concerned," Mitsubishi Electric's Silicon Valley president, Kenji Kitamura, asserts.

Two developments, at present in full swing, suggest that Japanese electronics technology will continue to play a pivotal role. First, gallium arsenide, another new development in semiconductor technology, is already well advanced in Japan; second, the work currently under way on the Fifth-Generation Computer will provide further advances into the postsilicon era.

Diffusion of fifth-generation technology has, of course, already begun. Not only have the major achievements of the project been publicized worldwide; they are already being incorporated in some automatic translation devices, office automation equipment, and intelligent robots. Sakae Shimizu, senior managing director of Toshiba Corp.'s R&D, describes the results of knowledge engineering as a continuous flow in which software is becoming the main stream. Automatic translation systems, introduced in 1985, will become increasingly efficient with the development of new parallel systems with expanded dictionaries. Voice-recognition word processors, which already exist in prototype, will be perfected gradually as knowledge engineering progresses and becomes ready for general use in 1995. Similarly, Shimizu expects intelligent robots to be perfected by the mid-1990's. In the meantime, Toshiba is developing expert systems for engineering picture processing and power-station early-warning systems designed to assure safety and appropriate maintenance.

CONCLUSION

A new global industrial system is emerging with the Japanese electronics industry as one of the principal architects. Traditional business organizations, management systems, and political and economic systems must change in response to this new reality. The very purposes of corporate activity must be redefined, and new measures of performance developed, if the full benefits of the microelectronics revolution are to be reaped. Even the role of the nation-state, and the uses of new technologies for peace and war, are subject to review. Timeworn ideologies, mainly of 18th- and 19th-century vintage, are rendered obsolete. The microelectronics revolution is not the sole motive power in this transformation, of course, but at this critical juncture in the development of our civilization, it is the prime mover in the common technological and economic sphere. Along with the ascendancy of Japan in particular and East Asia in general, this all adds up to an epochal shift that fundamentally changes not only the world industrial system, but the way ordinary people live their daily lives as well.

This is the meaning of Tsukuba and Expo '85.

The publishers, authors and Ted Bates Advertising/New York acknowledge the important contributions by the members of the international business community whose messages appear on these pages:

Epson Corporation

Hitachi, Ltd.

Japan Airlines Co., Ltd.

Matsushita Electric Industrial Co., Ltd.

Minolta Camera Co., Ltd.

Mitsubishi Motors Corporation

Nikon, Inc.

Nissan Motor Co., Ltd.

Olympus Optical Co., Ltd.

Ricoh Company, Ltd.

Sumitomo Metal Industries, Ltd.

Toshiba Corporation

Toyota Motor Sales U.S.A., Inc.

Special acknowledgment is made to the offficers and staff of Japan Society for their contributions to "Japanese Technology Today." For more information about Japan Society write to 333 East 47th Street, New York, N.Y. 10017 or call (212) 832-1155.

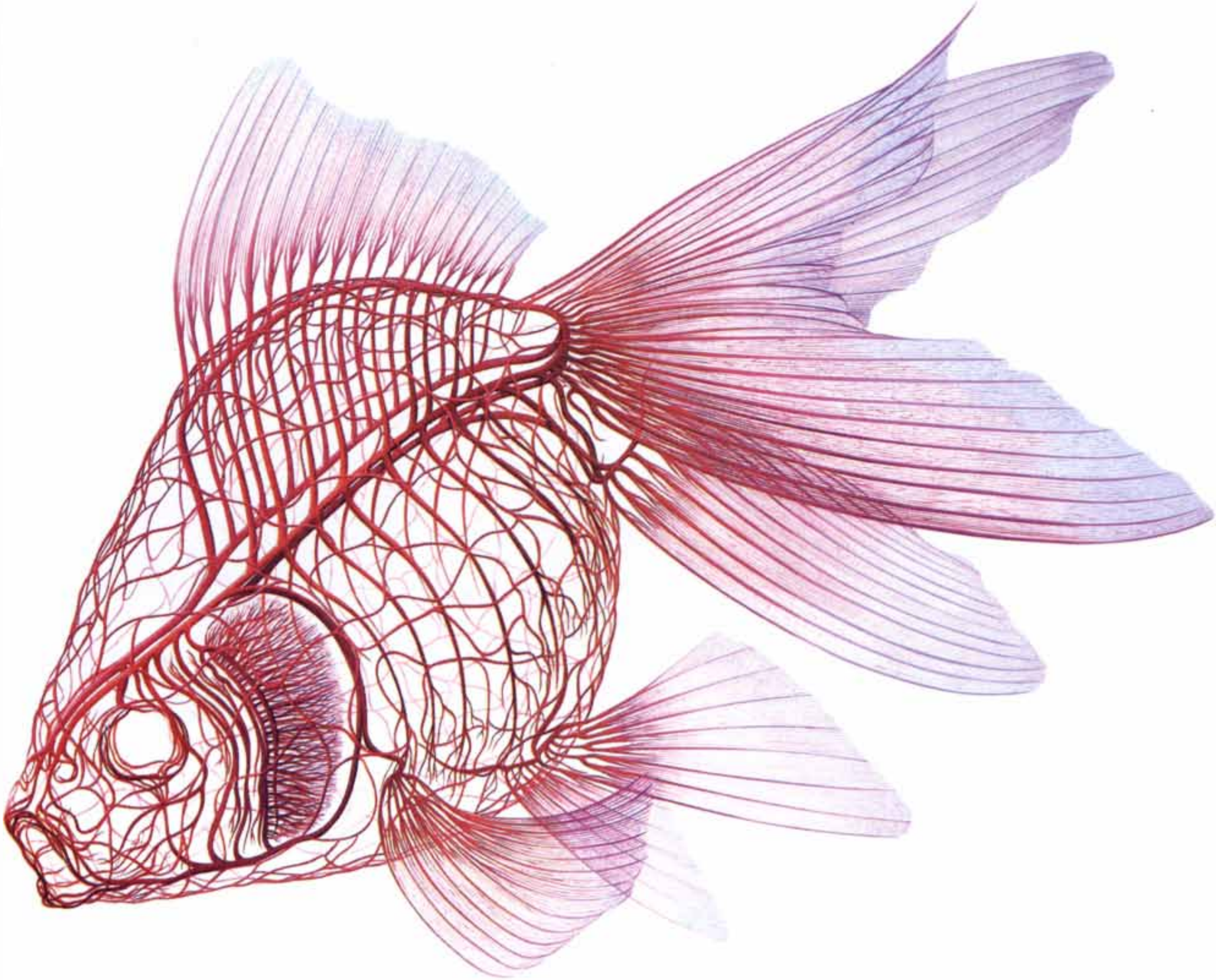
300 ZX GTP

This Nissan GTP ZX Turbo was designed to compete in the International Motor Sports Association (IMSA) GTP class of racing.

Inspired by the revolutionary Z-car engine, its mission is to prove technology on the track that will someday affect every Nissan we put on the road; and to take on the best and win.

When it comes to performance and technology... the Name is Nissan.

NISSAN



What you can't see makes the difference... no bones about it.

Toshiba takes out the unnecessary. We give you music without noise.

And we give your doctor a clear view of vascular systems. Without any bones. So it's easier to diagnose any problems. All with the latest digital technology.

Toshiba medical electronics is found in hospitals around the world. At the state of the art.

Because we're one of the largest electronics makers there is.

But there is something unique about Toshiba electronics. It works for people. In everything from medical equipment to video recorders to giant power generators.

We bring electronics to life. To make your life a little better. We're Toshiba.

DIGITAL FLUOROGRAPHY SYSTEM DIGIFORMER

In Touch with Tomorrow

TOSHIBA

Toshiba Medical Systems: Division of Toshiba America, Inc. 2441 Michelle Drive, Tustin, Calif. 92680 U.S.A. Tel:(714)730-5000 Telex:183-812

SCIENCE AND THE CITIZEN

Nuclear Famine

In the aftermath of a nuclear war it is now widely accepted that a dense, high cloud of smoke and dust would cover most of the Northern Hemisphere. In one or two weeks the normal circulation of the atmosphere at high altitudes would spread the cloud in thinner layers over the Southern Hemisphere as well. The devastating consequences for the temperature and sunlight at the surface of the earth have been summarized by the term nuclear winter. The term has been criticized on the ground that many post-nuclear-war scenarios might not lead to the deep and sustained cold that is characteristic of winter in the mid-northern latitudes. An international body of 300 scientists from more than 30 countries, including the U.S. and the U.S.S.R., has now reached consensus that such criticisms are largely academic. The stresses a strategic nuclear attack or exchange would impose on the world agricultural system and on other systems people might depend on for food would lead to massive starvation in combatant and noncombatant nations alike.

The conclusions of the scientists are set forth in a two-volume study, *Environmental Consequences of Nuclear War*. The study is the outcome of three years of investigation by the scientists under the coordination of the Scientific Committee on Problems of the Environment (SCOPE), a permanent committee of the International Council of Scientific Unions (ICSU). In October several of the scientists outlined their findings in hearings before the Senate Armed Services Committee.

According to Sir Frederick Warner, chairman of the Steering Committee for the study, the SCOPE scientists intentionally stopped short of considering the possible political and social disruptions that would result from a nuclear war. Russian investigators made it clear that their participation was contingent on limiting the discussions to physical, atmospheric, ecological and biological effects.

The physical and atmospheric studies are based primarily on three-dimensional computer simulations of unprecedented sophistication. A. Barrie Pittock of the Commonwealth Scientific and Industrial Research Organization and five other atmospheric scientists collaborated on the volume that summarizes the findings of the simulations. The ecological and agricultural predictions are derived from computer models, historical analogues, statistical analyses, expert judgment and laboratory tests of the responses of plants and animals to various kinds of stress. Mark A. Harwell of Cornell University integrated most of these findings for the SCOPE report. In spite of its self-imposed silence about the social responses to nuclear war, the report presents a case-by-case analysis of the stored food supplies in 15 countries, a physical constraint that no surviving society could ignore.

Earlier studies of the potential post-war environment have drawn attention to the effects of the dust and debris that would be drawn into the atmosphere by nuclear explosions and the complex hydrocarbons that would be injected by burning forests. According to the SCOPE study, however, the most serious threat to food supplies following a nuclear war would be suspended particles of soot, or elemental carbon. Soot particles strongly absorb energy at the wavelengths associated with the heat and light energy of the sun. Nevertheless, they readily transmit the long-wavelength energy reradiated by the earth into space.

Large quantities of soot are released by the burning of fossil fuels and materials derived from fossil fuels such as plastics, rubber, asphalt, roofing materials and chemicals. All such materials are concentrated in urban and industrial centers. Because many military targets are geographically isolated from such centers, a so-called counterforce nuclear strategy might therefore appear to eliminate most of the dangerous effects of soot in the atmosphere. The authors of the SCOPE report explicitly reject this view. They write: "Enough important military and strategic targets are located near or within cities so that... even relatively limited nuclear attacks directed at military-related targets could cause large fires and smoke production."

Even if early rainfall washed away about half of the smoke, a major nuclear exchange could leave about 30 million tons of black, sooty smoke in the upper atmosphere, circulating as high as 10 to 15 kilometers. If the war were begun during the northern spring or summer, the heat of the sun would carry the soot even higher and shift it toward the Equator. The warming of the upper layers of smoke could also stabilize the atmosphere and keep the air from mixing vertically, thereby extending the time the particles would stay aloft to several months or more.

The net effect would be a significant average cooling even over equatorial regions, and an average reduction of between 40 and 70 degrees Fahrenheit over the northern interior continental landmasses. The sunlight reaching the surface under large, patchy clouds of smoke could drop to less than 1 percent of normal, and in some areas the rainfall associated with the convective movement of the atmosphere would practically cease.

The most surprising conclusion of the SCOPE report is the vulnerability of agriculture even to much smaller disruptions. For example, even if the average cooling were much less than 40 degrees F., extreme fluctuations could develop about the average. A short-term frost at a critical point in the growing season can destroy a year's crop. Rice is particularly sensitive. If the transient temperatures were to reach even the low 50's F., the crop would be lost, although the rice plants would survive. An average drop of only about five degrees F. would eliminate a year's crop of cereal grains in Canada and the U.S.S.R., and an average decrease of between five and 10 degrees F. would virtually eliminate agricultural production in the Northern Hemisphere.

Rice production is also seriously threatened by the loss of convective rainfall. The SCOPE report predicts large crop losses from the cessation or displacement of the monsoon rains. In the countries of Africa, Asia and the Pacific such losses would be disastrous. Imports from other nations would presumably be cut off or at least seriously disrupted after a nuclear war, and so these countries would be left to fend for themselves. The authors of the report predict the deaths from the famine in India that would be caused indirectly by a nuclear war would exceed all the casualties caused by the direct effects of blast, fire and radiation in the U.S. and the U.S.S.R. combined. In Africa more people would die of the indirect effects of nuclear famine than would die of the direct effects of nuclear war in Europe.

The SCOPE scientists also investigated many secondary factors that could substantially reduce crop production. For example, global increases in ultraviolet radiation could result from the partial breakdown of the ozone layer in the upper atmosphere. Soil, air and water could be contaminated not only by radioactivity but also by the release of toxic chemicals into the environment. The loss of fertilizer, fuel and pesticides would lead to a breakdown of mechanized agriculture, even if crops could be made to grow. Other natural sources of food such as marine life would be adversely affected by the loss of sunlight, but even if they were not affected, they could not replace the agricultural losses.

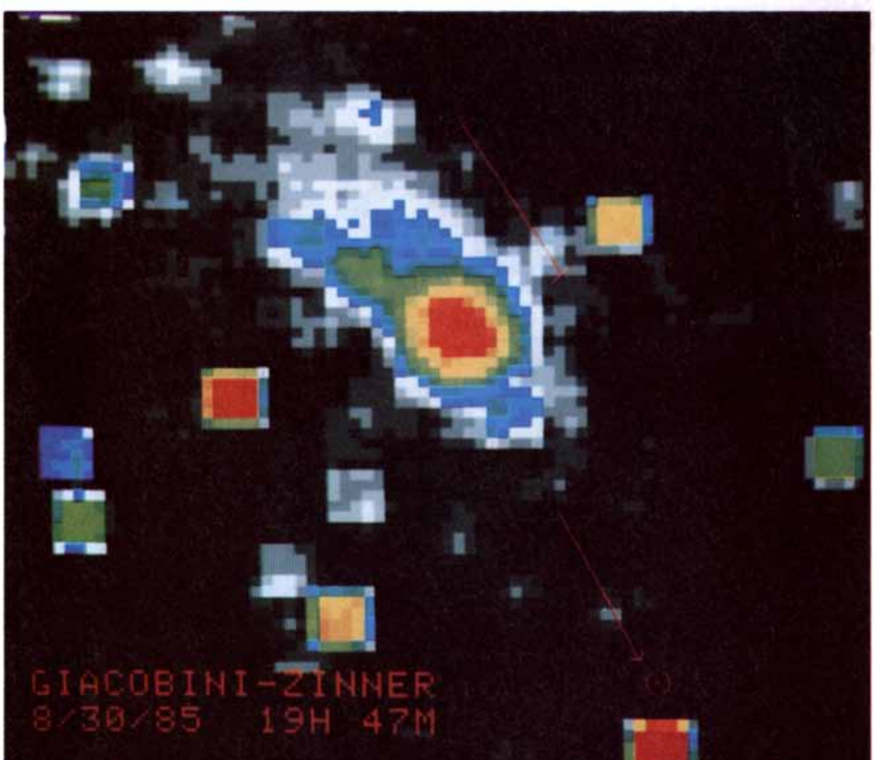
Furthermore, such environmental factors might well interact in ways that magnify their individual effects. The vulnerability of crops to disease and pests might be increased by radiation and air pollution. Reduced temperatures could depress the activity of insects in pollinating crops. The enforced low-calorie diet of the survivors could increase their susceptibility to disease.

In its study of the food stores that would remain after a nuclear war, the SCOPE report makes the optimistic assumption that within each country the distribution of the remaining food supplies would not be affected by the war. Even with this assumption the survivors of the war in the U.S. would have only about three years in which to reestablish agricultural systems or face starvation. In most other countries stored food would last only from three to six months.

The social responses to nuclear war that are avoided by the SCOPE study may still be addressed in other scientific forums. Can the likely effects of food hoarding and conflict over the scarce nutritional resources remaining to a postwar society be quantified? To what extent would the destruction of market, transportation and communications systems affect the perfect distribution of remaining food supplies that is assumed by the SCOPE report? How would imperfect allocation affect the distribution of necessary nutrients, such as vitamins and proteins? Although the SCOPE scientists advocate the study of such issues, wide scientific consensus about them would seem difficult to achieve.

On ICE

The sense of showmanship of the National Aeronautics and Space Administration (NASA) seems as keen as ever. Recently several hundred scientists and reporters were gathered in an auditorium at the agency's Goddard Space Flight Center in Greenbelt, Md., to listen as audio transmissions from the first spacecraft to fly through the tail of a comet were beamed over a distance of 45 million miles. NASA investigators later told the audience that although data from the mission of that spacecraft, the International Cometary Explorer (ICE), supported the prevailing model of comets, it also raised several interesting questions.



COMET GIACOBINI-ZINNER was recently the target of the International Cometary Explorer (ICE), a spacecraft of the National Aeronautics and Space Administration (NASA). The false-color image was obtained by another NASA satellite, the International Ultraviolet Explorer (IUE). Red corresponds to the brightest levels of light. The arrow points to the sun.

ICE was launched into space more than seven years ago. Initially known as the International Sun-Earth Explorer (ISEE) III, the probe was sent into orbit between the earth and the sun to gather data on the solar wind. ISEE III was then directed on a second mission to the side of the earth away from the sun, where it conducted the first deep survey of the earth's magnetic tail.

Planning for its third mission, the interception of the comet's tail, was begun in 1981. At that time it became clear that a tight budget and a steadfast commitment to the space shuttle would prevent the U.S. from launching a significant number of unmanned interplanetary spacecraft, including one to rendezvous with Halley's comet during the then distant year 1986. That engagement will be kept next March by a flotilla of European, Soviet and Japanese spacecraft. In an attempt to send a representative from the U.S., Robert W. Farquhar of Goddard proposed that the ISEE III could be directed to Halley's comet. When workers determined that the satellite's transmitting power was not sufficient to traverse the some 100 million miles that would separate the comet from the earth, Farquhar suggested an alternative target: comet Giacobini-Zinner. Like Halley's comet, Giacobini-Zinner exhibits periodic motion, although its period, $6\frac{1}{2}$ years, is much shorter than Halley's 76 years. Its point of closest approach, 45 million miles from the earth, is well within the satellite's transmission range.

To obtain the higher speed required for its third mission, the satellite was directed to the moon, from which it received a gravitational kick as it swung around that body. When the spacecraft left the earth-moon system on December 22, 1983, it was renamed the International Cometary Explorer. Nearly two years later ICE passed through the tail of comet Giacobini-Zinner, a scant 5,000 miles behind its nucleus, or densest region. (A comet's tail can be 500,000 miles long.) Traveling at a speed of 46,000 miles per hour, ICE remained in the tail for 20 minutes.

There ICE detected water molecules, carbon monoxide and dust particles. John C. Brandt, chief of the Laboratory for Astronomy and Solar Physics at Goddard, notes that the presence of such material "fits in nicely with our picture of comets as large dirty snowballs." The "dirty snowball" model, first proposed 35 years ago by Fred L. Whipple of the Smithsonian Astrophysical Observatory, holds that the nucleus of a comet consists of a mixture of ice, silicates (a class of minerals) and perhaps metals.

NASA investigators and others were

TANDY... Clearly Superior™

We edged out the competition with the 10-Mb Tandy® 1200 HD, a "mirror image" of IBM's PC XT for only $1999.

It uses virtually every software package made for the PC XT. It's expandable with the same hardware. The keyboard follows the same layout, with improved placement of the shift keys. So what's the difference between the Tandy 1200 HD and the IBM® PC XT?

The Tandy 1200 HD costs $1,896 *less* than the PC XT. And we include the printer adapter, an option on the XT.

So if you're using IBM PCs, expand with ten-megabyte hard disk Tandy 1200 HD's—at substantial savings.

And when it comes to service and support, Radio Shack is second to none.

In-store training is conducted at 431 Computer Centers nationwide. Low-cost delivery and set-up, new-user orientation, and on-site training are available in all major market areas. We also offer a telephone "hotline" as well as carry-in and on-site service.

So come in and find out how your business can lease a Tandy 1200 HD system for only $85 per month*. And get the edge on *your* competition.

Available at over 1200 Radio Shack Computer Centers and at participating Radio Shack stores and dealers.

Radio Shack®
COMPUTER CENTERS
A DIVISION OF TANDY CORPORATION

Send me a Computer Catalog.
Mail to: Radio Shack, Dept. 86-A-493
300 One Tandy Center, Fort Worth, Texas 76102

Name ______
Company ______
Address ______
City ______
State ______ Zip ______
Telephone ______

*Commercial lease for a Tandy 1200 HD with monitor and adapter, plus applicable use/sales tax. Tandy 1200 HD prices apply at Radio Shack Computer Centers and participating stores and dealers. IBM and the IBM logo are registered trademarks of International Business Machines Corp. Pricing and service availability as of 5/1/85.

surprised by the apparent lack of "bow shock." Astronomers have theorized that just as the bow of a moving ship creates a wave, so the interaction of a comet with the solar wind should create a sharply defined magnetic boundary around the comet. ICE detected no bow shock.

Although the NASA scientists were enthusiastic, they were disappointed that they would not participate in the main event. James M. Beggs, NASA administrator, commented: "I regret that the United States did not go to Halley's. We would have enjoyed the Halley encounter more if there had been an American spacecraft in that armada of spacecraft going to Halley's."

Fine Structure

Surrounding the thread of genetic material that gives a virus its identity is a capsid, or protein shell, that determines its biological properties. To understand how a virus interacts with its host one must know not only the capsid's composition but also its structure and topography.

Until recently three-dimensional structures had been determined only for a few plant viruses. Now groups from Purdue University and the Research Institute of Scripps Clinic have worked out the structure of two animal viruses at the level of individual atoms. Writing in *Nature,* the Purdue group, led by Michael G. Rossmann, describes the structure of a strain of rhinovirus, the agent of the common cold. The Scripps workers, under James M. Hogle, deciphered the structure of a poliovirus strain soon after the Purdue group's success; they describe their results in *Science.*

The cold and polio viruses belong to the picornaviruses. Members of the family are about 300 angstrom units (30 billionths of a meter) in diameter; they have a faceted shell that resembles an icosahedron (a 20-sided solid). The high symmetry of the picornaviruses makes it feasible to crystallize them and study their structure by X-ray crystallography.

When X rays pass through a crystal, they are scattered by electrons in the crystalline planes; the scattered waves interfere with one another, yielding a diffraction pattern. The pattern reveals the orderly arrangement of units in the crystal. Determining the structure of the units themselves requires calculations incorporating the relative amplitude (brightness) and phase of the waves that are scattered to different points in the diffraction pattern. The amplitude of the scattered radiation is easy to measure; finding the relative phase of the waves is difficult.

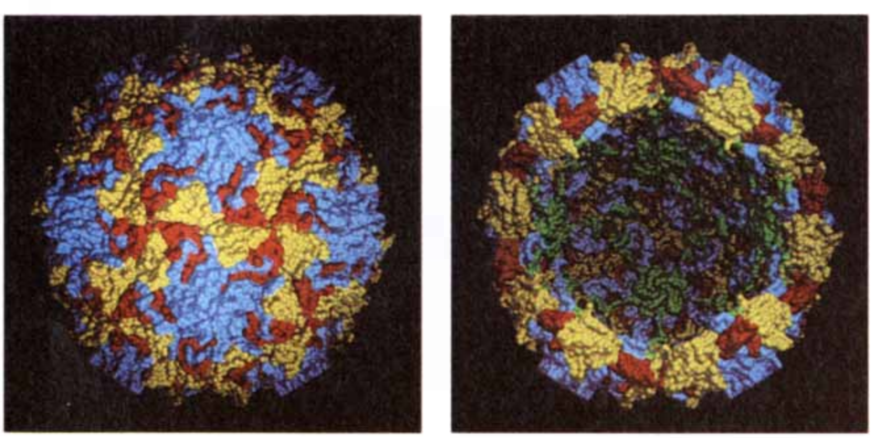
Both the Purdue and the Scripps groups owe their success to a method of determining phase that was recently pioneered by other workers. In X-ray crystallography of biological molecules the usual procedure is to soak the material in a solution containing a heavy element such as gold or platinum. Incorporated into the molecules at specific sites, the heavy atoms scatter X rays much more strongly than do the lighter atoms characteristic of organic compounds. Once the heavy atoms' positions within the crystal unit have been charted, they can serve as benchmarks for determining the phase of X rays scattered from other atoms.

Both groups determined phase with heavy atoms early in their work, but in refining their data to a resolution of about three angstroms (the scale of individual atoms) they turned to the new method. It is based on "noncrystallographic symmetry": in this case symmetries present in the icosahedral virus particles but not apparent in the crystals they form. The hidden symmetries mathematically constrained the possible phases of the scattered X rays; repeated calculation cycles, performed on vast amounts of raw data (a process for which the Purdue group employed a supercomputer), yielded maps of electron density in the viruses. From the maps the configuration of protein chains could be inferred.

Earlier X-ray results, combined with the known sequence of the RNA in certain picornaviruses, had indicated that the picornavirus capsid consists of 60 identical subunits, each made up of four proteins. The new data show how the four proteins intertwine to form the subunits in the cold and polio viruses and how the subunits are packed together in the viral shell.

The structural data also reveal the sites to which antibodies bind as they defend the body against each virus. Known mutations in the viral RNA yield strains of cold and polio viruses that are resistant to antibodies. The mutations change specific amino acids in the capsid proteins. By determining where the protein changes occur in the capsid, the workers could pinpoint the antibody binding sites.

In theory a map of the immunological geography of a virus could aid in the creation of new vaccines. In the case of the cold virus, however, the new structural knowledge makes the prospect of a conventional vaccine seem remote. A cold vaccine would have to stimulate the production of antibody against some protein sequence that is shared by a large fraction of the 89 or more rhinovirus strains. Since the majority of the strains bind to the same kind of cell, their cell-binding sites are probably identical. An antibody that binds to the site and thus blocks it could confer immunity to many of the viral strains. The Purdue workers report, regretfully, that the rhinovirus cell-binding sequence probably lies well protected from antibody at the bottom of a deep cleft that winds around the virus.



PROTEIN SHELL OF POLIOVIRUS is made up of four protein subunits, distinguished by color in these computer-generated images. Each subunit is repeated 60 times to form a highly symmetrical structure. On the surface of the virus (*left*) three different subunits are apparent. The fourth subunit (*green*) is visible on the hollow shell's inner surface, exposed in a cutaway view (*right*). The cutaway reveals the shell thickness and the size of the cavity, which contains the viral RNA. Arthur J. Olson of the Research Institute of Scripps Clinic made the images working with AMS and RAMS, programs developed by Michael L. Connolly.

Cachectin

Cachexia is a syndrome that frequently affects individuals who have cancer or a serious infection. Even if such patients are fed an adequate diet, they experience a continuous loss of weight. A major factor in the cause of the syndrome appears to be cachectin, a hormonelike protein


Stolichnaya
The Gift Vodka
genuine Russian vodka
STOLICHNAYA
IMPORTED FROM THE USSR
Stolichnaya vodka
COOL BEFORE DRINKING
RUSSIAN VODKA
For gift delivery anywhere* call 1-(800)-238-4373.
*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

Cash or charge? Credit cards have changed the world's way of shopping—whether it's buying a great vacation or a great new outfit; whether shopping close to home or far away. As the premier electronic processor of national credit card transactions, NaBANCO, a Continental Telecom subsidiary, speeds up your credit-card shopping. In just four seconds or less your card and credit status are OK'd, your purchase is authorized, often clear across the country, even around the world. On-line cash registers and a nationwide communication network make such processing of over 200 million transactions a year by NaBANCO a routine thing—and a part of the telecommunication revolution that you experience almost every day. From telephony to satellites. Architects of telecommunication.

FOR A FREE BROCHURE ON ALL OF CONTEL'S CAPABILITIES WRITE TO:
CONTEL, DEPT. 5010, 245 PERIMETER CENTER PARKWAY, ATLANTA GA 30346

CONTEL

made by macrophages, one of the major cell types in the immune system.

Work by investigators at Rockefeller University and the Stanford University School of Medicine has now illuminated the crucial role cachectin plays in cachexia. The work also shows cachectin may induce shock when bacterial endotoxins cause the factor's release. Finally, cachectin has emerged as the long-sought tumor-necrosis factor, a substance that is secreted in the presence of bacterial endotoxins and can promote the rapid death of cancerous tissue.

During an infection cachectin travels to fat cells and helps to mobilize their energy reserves: it releases fat that can be broken down and then metabolized by the body. In addition to releasing fat cachectin seems also to reduce the concentration of enzymes that are crucial to the production and storage of fat. Without the enzymes fat no longer accumulates. In patients with cachexia the process is apparently carried to an extreme. Fat stores are depleted by excessive mobilization coupled with inadequate intake of calories and a hypermetabolic state: food is burned too quickly and inefficiently.

Anthony Cerami and Bruce Beutler of Rockefeller, working with Frank M. Torti, Barbara Dieckmann and Gordon M. Ringold of Stanford, report in *Science* that they have determined how cachectin depresses enzyme levels. Apparently the factor severely curtails the amount of messenger RNA transcribed from the genes that encode enzymes required for fat production. In the absence of messenger RNA the enzymes are not made.

The workers made this discovery by measuring the effect of cachectin on the levels of the messenger RNA's that code for the fat-producing enzymes. To do so they added cachectin to a batch of rodent fat cells, isolated all the RNA and measured (by means of a radioactive DNA probe) the level of the RNA's mediating fat formation. As expected, cachectin was associated with reduced levels of those RNA's.

The workers next studied whether other signs of cachexia could be induced when fat cells were exposed to cachectin. They observed that cachectin caused the cells to lose their accumulated lipid droplets and blocked the production of more fat, two reactions known to occur during cachexia.

Cerami, Beutler and I. W. Milsark, also of Rockefeller, were aware of indications that cachectin plays a significant role in the shock related to infection by gram-negative bacteria, a group that includes *Salmonella* and *Shigella*. To explore the relation they injected mice with an antibody that binds to cachectin and neutralizes it. Then they injected the mice with a bacterial endotoxin known to cause shock. The passively immunized mice were protected. In mice that had not received the antibody, the endotoxin stimulated the release of cachectin and the mice quickly died of shock. These results suggest a potential therapy for shock based on neutralizing cachectin's activity.

Further work by members of the Rockefeller group and other investigators revealed that cachectin and tumor-necrosis factor (TNF), which had been isolated by Lloyd J. Old and his colleagues at the Memorial Sloan-Kettering Cancer Research Institute, are the same substance. The first hint emerged when Beutler sent John Mathison of the Scripps Clinic and Research Foundation a sample of mouse cachectin. Mathison put the substance through a screening panel that included an assay testing for activity against tumor cells. Mathison was surprised to find that the cachectin killed the cells as effectively as TNF does.

Beutler then compared the partial sequence of amino acids he and his colleagues had derived for mouse cachectin with the corresponding segment of human TNF, whose structure was completely known. The two sections shared 15 out of 20 amino acids, suggesting that the two proteins are homologous. The conclusion was confirmed when Lucie Fransen at Biogen, SA, determined the complete sequence of amino acids in mouse cachectin, a task Beutler and his colleagues also completed successfully.

How does cachectin cause shock and mediate the destruction of tumor cells? The two effects may be related to cachectin's ability to reduce messenger-RNA levels. In shock endotoxin may cause a release of cachectin; the factor may then depress RNA levels and thereby shut down vital cellular functions. In tumor tissue endotoxin may also cause the release of cachectin, which is to say of TNF. In this instance, however, the factor may specifically affect RNA's that mediate the biochemistry of tumor cells.

Scientists at several biotechnology companies are eagerly analyzing cachectin by snipping the molecule apart or synthesizing segments of it and introducing the pieces into cell cultures. One investigator at Biogen speculates that the part of the molecule responsible for antitumor activity is different from the part responsible for shock. Work has begun there on second-generation molecules: fragments of cachectin that might initiate antitumor activity without producing such side effects as shock. At the Asahi Pharmaceutical Company in Japan clinical applications of cachectin are also being evaluated.

Piggyback Vaccine

Almost 200 years ago Edward Jenner gave the world its first vaccine when he inoculated an eight-year-old boy with cowpox virus to prevent smallpox. Now the cowpox virus, called vaccinia, may again serve as the vehicle for a major advance in disease prevention. Enzo Paoletti of the New York State Department of Health and his colleagues report in *Science* that they have derived from the virus a genetically engineered polyvalent vaccine that has simultaneously immunized rabbits against hepatitis *B*, herpes simplex and influenza.

Vaccinia lends itself to becoming a polyvalent vaccine because it is a large virus whose DNA can accommodate the insertion of many foreign genes. Once they have been inserted into vaccinia DNA and placed (as part of the vaccinia virus) in live cells, these genes manufacture proteins, exploiting the biochemical machinery of the host. In particular they direct the production of antigens: proteins that stimulate the immune system of the vaccinated animal to produce antibodies neutralizing the herpes, hepatitis and influenza viruses. Since these viruses are not present in the vaccine, side effects they might provoke—such as allergic reaction or low-level infection—are unlikely. In this respect the polyvalent vaccine resembles monovalent synthetic vaccines that present only an antigen instead of a whole virus in killed or attenuated form. The only whole virus here is vaccinia, which rarely provokes adverse reactions.

The investigators produced the polyvalent vaccine by extending already successful recombinant-DNA techniques for creating vaccinia-based vaccines against single viruses. First the team isolated the foreign gene sequence for the influenza antigen and embedded it in a fragment of vaccinia DNA. The fragment was introduced into cultured cells infected with vaccinia virus. When the virus replicated, the DNA fragment recombined at its normal site in the vaccinia DNA, thereby forming a recombinant vaccinia-plus-influenza vaccine. The process was repeated to incorporate isolated gene sequences from the herpes and the hepatitis virus. The full polyvalent vaccine was purified by cloning and was grown in tissue culture.

Tests of the effectiveness of the vaccine were carried out in rabbits. The workers found that the vaccine raised detectable levels of antibodies to all

So no one has to ask you how you're doing lately.

The 1986 Grand Am LE Sedan is a new statement of Pontiac's road car philosophy. Its power rack and pinion steering, five-speed gearbox and responsive 2.5 liter engine will handle the asphalt. Its sharply styled appearance will handle the rest.

PONTIAC GRAND AM

WE BUILD EXCITEMENT

LET'S GET IT TOGETHER BUCKLE UP.

A "business partnership" with Hewlett-Packard

- helps ETC Corporation reduce chemical analysis costs by one-third and double productivity.

- helps Container Corporation of America improve response time to customers--and save $600,000 a year.

three of the introduced antigens. It also protected two test animals against infection by the target viruses (and by vaccinia).

Because the DNA of vaccinia can accept dozens of genetic insertions, polyvalent vaccinia recombinants can in theory be created against five, 10 or more strains of influenza or other multistrain viruses. Whether the immune system would be able to respond to so many challenges at once has yet to be determined. Paoletti thinks the most exciting application of vaccinia recombinants would be the development of "cassettes" carrying genes for antigens of many infectious agents in a given geographic area. One cassette could be administered in, say, South America, whereas a different cassette would serve in North America.

Within the next two years Paoletti hopes to begin human tests of vaccinia recombinants against single viruses—herpes and hepatitis—followed quickly by polyvalent vaccines.

REPLICA of the Lake Turkana *Homo erectus* skeleton is at the National Museums of Kenya. Some of the material (*yellow, green, light gray*) was added for the reconstruction.

Skeleton Key

A year has elapsed since a team of paleontologists unearthed a 1.6-million-year-old skeleton of *Homo erectus* that is certainly the most complete example and probably one of the oldest examples of this extinct species yet found. The specimen has already yielded much new information about the early ancestors of modern man.

The first fragment of the skeleton was discovered by Bw. Kamoya Kimeu at a site in Kenya near the western shore of Lake Turkana, where a number of human fossils have been found. The dispersed bones and fragments of much of the rest of the skeleton were excavated by other members of a team led by Frank Brown of the University of Utah, John Harris of the Los Angeles County Museum, Richard Leakey of the National Museums of Kenya and Alan Walker of Johns Hopkins University.

Homo erectus, a species that made primitive tools and may have used fire, is thought to be the direct ancestor of *Homo sapiens,* the living species of man. The first hominids, or humanlike creatures, appeared on the earth some four million years ago, and for the next two million years several species apparently existed in Africa. The relatively large-brained *H. erectus* emerged there about 1.6 million years B.P. (before the present), and then in Asia about a million years B.P. The species disappeared from both continents about 300,000 years B.P.

According to Eric Delson of the American Museum of Natural History, the Kenyan find is important because it consists of almost the entire skeleton, including most of the skull; it lacks only hands and feet and most of the arm bones. It is therefore the first fossil early hominid in which both brain and body size can be accurately measured on the same individual.

Brown and his colleagues report in *Nature* that bone structure and tooth development in the fossil are consistent with those of a modern 12-year-old male. In spite of the apparent age, the lengths of the long bones of this individual are close to the mean lengths for modern adult males. He was nonetheless likely to have been somewhat under 1.68 meters (5′5″) tall: because the *H. erectus* skull was relatively flat, he would have been shorter than a modern human whose bones are the same length. "If a member of *Homo erectus* got on your bus," Delson notes, "it would be obvious that there was something unusual about him."

Study of this fossil is providing insights into patterns of growth and individual differences in these earlier hominids. For instance, it is clear from the shape of the pelvis that this specimen was a male. Comparison of his facial pattern with that of other fossil skulls has helped to identify the sex of these individuals. The authors and others conclude that there were substantial differences between male and female skull shapes in early *H. erectus.*

Furthermore, the narrow pelvic structure seen in the fossil may indicate that this species gave birth to small, immature infants. This fact, associated with the high level of infant dependence and parental care that is typical of humans, suggests these behavioral traits may actually have been present 1.6 million years ago.

The new skeleton should help to resolve a significant evolutionary question about *H. erectus:* Are the African forms members of the same species as the Asian ones? If they are, the species may have changed more over its 1.5-million-year span than has been thought. The answer hinges on accurate attribution and dating of many scattered and often fragmentary fossils thought to be *H. erectus.*

The Right Chemistry

Predicting the outcome of complex chemical reactions has long challenged theoreticians. A team of investigators at the University of Georgia reports that the task may be made easier and years of potentially unfruitful research may be avoided by exploiting certain branches of mathematics known as graph theory, topology and group theory. As well as saving perhaps millions of dollars in the development of drugs, the new approaches could also have a large impact on cancer research and the control of environmental pollution.

If the approaches are successful, they would supplement the traditional tools of theoretical chemistry, quantum mechanics and statistical mechanics. Quantum mechanics provides a way of predicting the behavior of atoms, and statistical mechanics is an elegant way to approximate large systems of atoms and molecules. In spite of the great success of these techniques, they do have limitations. When the Schrödinger equation (the fundamental equation of quantum mechanics) is employed to model atoms that have large numbers of electrons, for instance, the calculations become exceedingly complex and cumbersome.

The Georgia investigators hope to find short cuts. Dennis H. Rouvray, a theoretical chemist and member of the Georgia group, notes, for instance, that constructs known as "topological indices" can be used to describe the structure of compounds in purely mathematical terms. Specifically, indices can express the bonding network of a compound: the neighborhood relation between atoms.

Indices take on a wide variety of forms including an algorithm, a matrix, a code or even a simple number. Rouvray states that "in recent years such indices have been shown time and again in all sorts of different contexts to be extremely valuable predictors of chemical behavior." They could, for instance, be employed to predict physical properties such as freezing and boiling points, densities, viscosities, surface tensions, magnetic properties, heat capacities, refractive indices and dielectric constants.

On the biological side topological indices could predict carcinogenic, mutagenic and teratogenic behavior, as well as biological degradability and toxicological behavior. They may also prove worthwhile for predicting octane number in fuels and designing new drugs.

Also working on the prediction of physical and statistical properties of matter are E. Rodney Canfield, a computer scientist, and M. Howard Lee, a theoretical physicist. Canfield is now developing the computer software in order to utilize both topology and group theory. R. Bruce King, leader of the project, is concentrating on the design of new substances, such as novel conducting materials, that have the desirable properties of both metallic and nonmetallic materials.

In spite of the promise of the approach, it is still in its infancy. King says that "as far as we know, this is the first time in this country that a major group has been assembled with our specific objectives in this field." The Georgia team has received a $1.2 million contract from the Office of Naval Research, with the possibility of further funding from the National Science Foundation and the Environmental Protection Agency.

Airtight Case

Air trapped in old hourglasses, brass buttons and pendulum bobs may help to show how much the concentration of carbon dioxide in the atmosphere has increased in recent centuries. Workers at the Los Alamos National Laboratory are exploring the possibility, scouring the world for objects that are hollow, airtight and datably old. They will open the objects in a vacuum, ascertain by gas chromatography the amount of carbon dioxide in the trapped air and compare the readings with measurements of modern air.

The object is to obtain data bearing on the possibility that the burning of fossil fuels is increasing atmospheric carbon dioxide levels enough to cause changes in global climate patterns. Carbon dioxide is transparent to sunlight at visible wavelengths but is opaque to certain infrared radiation that would otherwise be transmitted from the earth to space. It therefore tends to trap heat. If the result were to be a significant greenhouse effect, some areas of the globe would become warmer and wetter.

Among the objects the Los Alamos group has found or been given so far are a stuffed quail in a glass case, model ships in sealed bottles and a set of old bottles of chemicals with their original seals. It remains to be seen whether these and the other old-air objects have remained airtight and whether the air in them appears not to have been changed in composition by chemical reactions with the walls of the container.

OLD AIR will be sought by Los Alamos National Laboratory workers in metal buttons collected at the laboratory, in an hourglass from the Adler Planetarium in Chicago and in a bottle containing a ship model at the Maine Maritime Museum (photograph by Lou Jones).

Terranes

They are fault-bounded blocks of crust that accrete to the ancient cores of the continents. The process makes the continents increase in extent and reworks them into what amount to geologic collages

by David G. Howell

For more than a century geologists have been seeking to understand the engine that produces the great geologic features of the earth. The early investigators conceived of geosynclinal cycles: great crustal downbowings that fill with sediment, followed by upheavals that create young mountain ranges. That picture gave way to the theory of plate tectonics. In this view, put forward in the 1960's, the dominant directions of motion are horizontal: the brittle, outer layer of the earth consists of large crustal plates that are constantly shifting. Where plates move away from one another, rifts develop and new ocean basins form; where plates collide, chains of volcanoes erupt along lines parallel to the zone of the collision; where plates slide past each other, along faults such as the San Andreas fault in California, great earthquakes can occur.

It is becoming apparent, however, that a further revision is in order. The crustal patterns engendered by plate-tectonic activity are ephemeral: tectonic forces rework the original patterns, carving out fragments of crust, dispersing them and refashioning them into groupings of disparate crustal blocks. At the same time new crustal blocks arise from volcanic processes and get swept into the reworking. The crustal plates are turning out, therefore, to be strange patchwork mixtures of crustal fragments, geologic collages consisting of pieces known as terranes.

The concept of terranes emerged in the 1970's, when conflicts over land use in Alaska required that the U.S. Geological Survey send teams of geologists to survey mineral resources. What they found was startling. The elucidation of a geologic pattern in one part of the state would lead to a prediction of what the pattern should be a few tens of kilometers away. Yet the actual pattern would be markedly different: the rock would be the wrong age and have the wrong composition. In brief, the straightforward application of the plate-tectonic theory failed to account for the geology of Alaska. The entire state proved instead to be an agglomeration of crustal fragments. Alaska is the crustal flotsam and jetsam of the ancient, vanished ocean that preceded the Pacific. It is a collage of terranes dismembered and repositioned over the past 160 million years by the wanderings and collisions of crustal plates. Pieces are still arriving from the south.

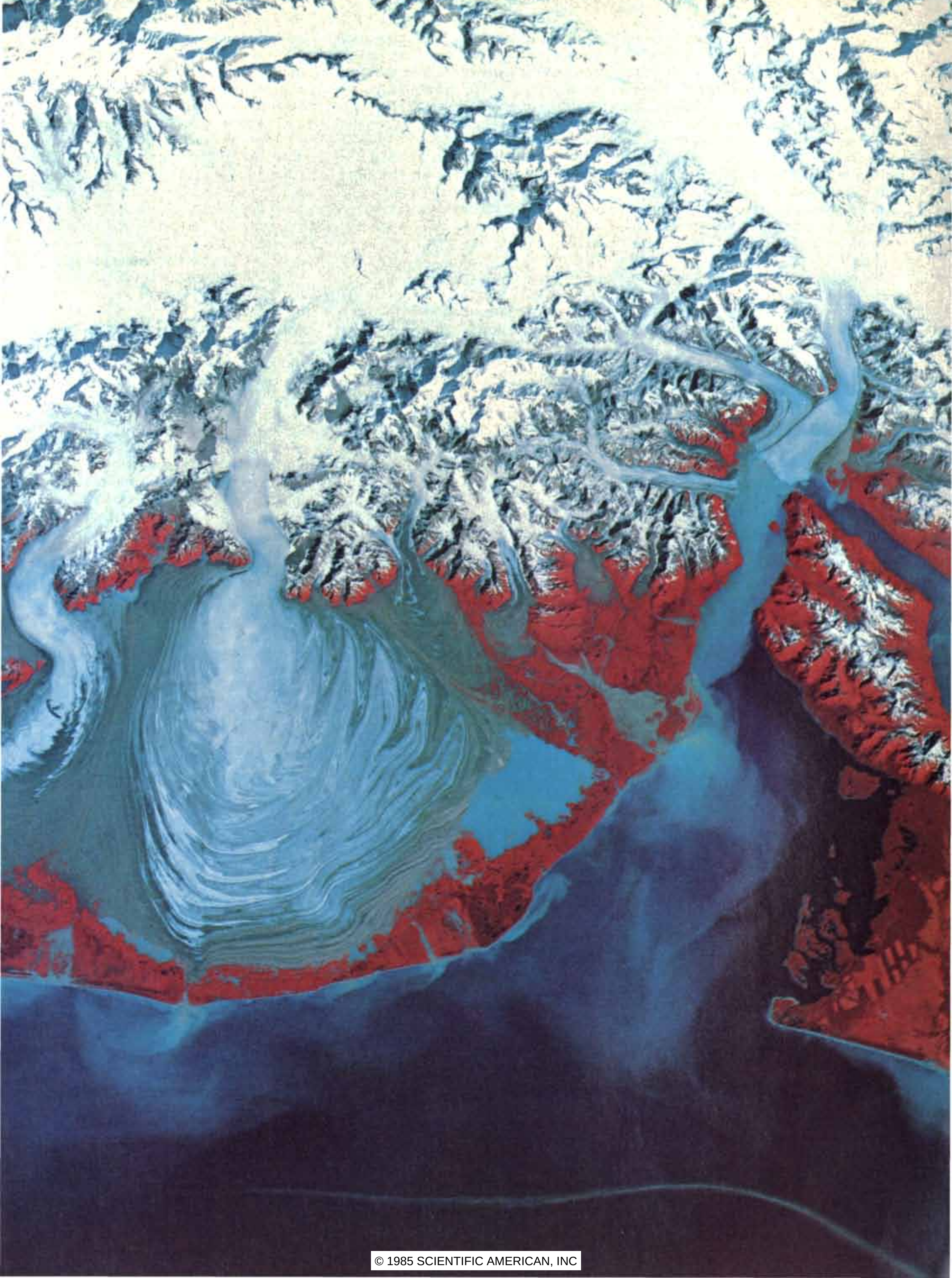
The Advent of Plate Tectonics

A brief review of the plate-tectonic theory helps to put terranes in perspective. After all, terranes represent only the newest aspect of the central effort in geology, an effort to grasp the vastness of time and comprehend the cumulative effects of slow movements in the earth. (Most tectonic processes plod along at rates equal to or less than the growth rate of a fingernail.) Fundamentally the crust of the earth has two domains: the oceanic crust, which is dense and homogeneous, and the continents, which are lighter and mineralogically heterogeneous. Terranes are the crustal fragments swept onto the ancient cores of the continents. In describing their histories one confronts the full consequences of the plate-tectonic theory, in which oceanic crust transports the continents as though it were a great, slow conveyor belt.

Oceans indeed are a central part of plate tectonics. Specifically, each ocean basin widens in the wake of diverging plates as magma, or molten rock, rises and solidifies to produce oceanic crust along a submarine ridge called an oceanic spreading center. Over geologic time the process may have slowed, and the network of spreading centers may well have decreased in extent. The present length of the network amounts to some 56,000 kilometers, and during the past two billion years the spreading velocities have probably averaged about five centimeters per year. The present-day Atlantic spreads at less than three centimeters per year; the most active parts of the East Pacific spread at about 16 centimeters per year. (These rates combine the velocities of the two plates moving in opposite directions from an oceanic spreading center.)

The multiplication of the two figures—the average spreading rate and the length of the spreading-center system—yields the estimate that new oceanic crust is formed today at a rate of 2.8 square kilometers per year. The area of the oceans is 310 million square kilometers. It follows that the oceans could have formed in a mere 110 million years. This is quite a realization. Before the theory of plate tec-

SIX TERRANES occupy the part of the coast of southeastern Alaska shown in this Landsat image. The coast is indented by Yakutat Bay; to its west the Malaspina Glacier descends from a group of peaks including Mount St. Elias and Mount Augusta. Under the ice and snow the rock represents parts of volcanic islands, parts of a displaced continental margin or the metamorphosed and remelted products of sediments within a sedimentary matrix. In each case the rock accreted to the ancient core of North America during the past 100 million years: the terranes are disparate fragments of crust swept together by the motions of the earth's great crustal plates. In southern Alaska the terranes are elongated bodies resulting from the slicing of the crust by faults to the south. Even today the faulting continues to spread terranes northward. A rotation has also taken place. Caught between the North America plate and the Pacific plate, the terranes of southern Alaska have turned clockwise.



tonics the oceans were taken to be the oldest parts of the earth, simply because they are the lowest parts of the surface. The idea was that old rock is colder, hence denser and lower, than young rock. The remarkable youth of the oceans is now well confirmed by the samples of oceanic crust amassed by the Deep Sea Drilling Project. In the earth today the age of the oceanic crust ranges from essentially zero along the submarine ridge crests that mark spreading centers to no more than 180 million years in the eastern Pacific, the part of an ocean floor most distant from a ridge. Over the past two billion years as many as 20 oceans may have been created and destroyed.

Oceanic crust rises at a ridge, moves across the width of an ocean basin and descends along a trench that marks what is called a subduction zone. The crust's surface, however, is not flat. On a broad scale the oceanic rock cools and compresses as it spreads away from the ridge crest where it was emplaced in the crust, and so the crust systematically sinks.

Moreover, mountains known as seamounts or oceanic plateaus are widely distributed across the ocean floor. Many are high enough to form islands. As I shall show, they can be swept up by plate-tectonic processes and thus become part of terranes. The primary constituent of a seamount is basalt, a dark volcanic rock that is rich in iron and magnesium and has a silica content of less than 50 percent. The rock ascends from "hot spots" under an oceanic plate. In fact, when the position of the hot spot is stable in the earth (that is, when the hot spot is stationary with respect to the core of the planet), long linear chains of volcanoes can form as the plate slides over the rising jet of magma. The Hawaiian Islands are part of such a chain. The rate of growth of the chain is rather rapid as geologic processes go, and yet the hot spot itself is seldom more than a kilometer in diameter. The annual global contribution of oceanic basaltic piles to the augmentation of the continental crust is estimated to be only about .2 cubic kilometer per year.

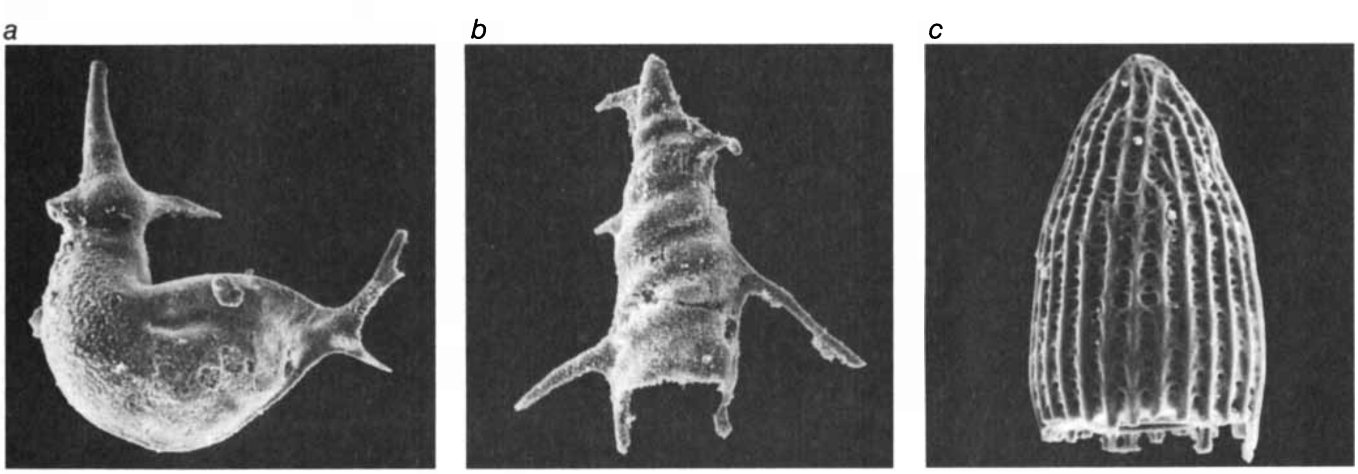
Subduction-Zone Volcanism

A greater contribution to the continental crust comes about when two plates collide. Along the zone of collision the denser plate descends, so that it encounters progressively greater temperature. The plate carries sediment and also water trapped in pores in the crustal rock. At a critical depth, usually between 100 and 150 kilometers, the water triggers a sequence of physical and chemical events, including the partial melting of rock, producing a magma that tends to be rich in volatile chemical elements, notably aluminum, potassium and sodium. The silica content ranges from 50 to 75 percent. The magma is stickier and more viscous than an oceanic, basaltic magma, and so it supports a buildup of pressure. The result is that the volcanism related to the earth's subduction zones tends to be explosive. Mount St. Helens and Krakatau are examples of the process, which has also formed numerous island arcs.

The plate-edge volcanoes are quite different from the intraplate islands and seamounts. For one thing the plate-edge volcanoes lie above great rising curtains of magma that parallel the trench marking the interface between two colliding plates. In all, the earth now has some 37,000 kilometers of plate-edge volcanic chains, and for each kilometer of such activity it is calculated that from 20 to 40 cubic kilometers of new siliceous material erupts in a million years. Worldwide, therefore, siliceous volcanic material joins the continental crust at a rate of roughly .75 to 1.5 cubic kilometers per year.

Clearly plate tectonics offers a rationale by which to examine terranes and classify additions of new material to the continental crust. In itself the oceanic crust is likely to contribute almost nothing, since the formation of oceanic crust at ridge-crest spreading centers is likely to be balanced by the loss of oceanic crust along subduction zones. The evidence is impressive. The assemblage of crustal rocks produced at spreading centers, called a MORB (midocean ridge basaltic) ophiolite and consisting of a characteristic sequence of three strata a total of six kilometers thick, is observed only rarely in folded mountain belts. (Mountain ranges almost always form by crustal folding.) Such belts are where MORB ophiolites would be found if oceanic crust augmented the continental masses.

Protruding above the oceanic crust, however, are the islands or island chains created by gentle, basaltic volcanism above hot spots and the island arcs created by the more explosive volcanism that parallels subduction zones. The ocean basins also include fragments of continental margins that



SKELETONS OF RADIOLARIANS, single-cell organisms that first appeared in the oceans some 500 million years ago, can now be freed from rocks such as flinty chert, which form from deep-sea sediment. The shape of the skeletons reveals the age of the rocks and so helps to establish the histories of terranes. The two skeletons at the left were extricated from mudstone of the Golconda terrane in north-central Nevada. One of them (*a*), of the genus *Pseudoalbaillella,* is some 290 million years old; it is enlarged 200 diameters. The other one (*b*), of the genus *Albaillella,* is 250 million years old; the magnification is 300 diameters. The two skeletons at the right were extricated from chert of the San Simeon terrane in southern California. One of them (*c*), of the genus *Archaeodictyomitra,* is be-

broke away from the continents when rifting occurred and new oceans opened. Moreover, on the ocean floor there are blankets of sediments currently totaling 170 cubic kilometers worldwide. The sediments represent river-borne continental debris, planktonic fossil remains and chemical precipitates from the sea. Some of the material on the ocean floor is subducted. Nevertheless, much of the sediment, the basaltic seamounts, the volcanic island arcs and the continental fragments is destined to be swept together: it becomes terranes, which increase the size of the continents.

The Nature of Terranes

In geology the word terrain simply designates the lay of the land. In contrast, the term terrane (the full name is tectonostratigraphic terrane) designates a crustal block, not necessarily of uniform composition, bounded by faults. It is a geologic entity whose history is distinct from the histories of adjoining terranes. Terranes come in many sizes and shapes, and they have varying degrees of compositional complexity. India, for example, is a single great terrane. Some of its individual rock formations have ages exceeding a billion years; nevertheless, over the past 100 million years India has acted as a single mass. (It was part of the margin of the great, but now shattered, megacontinent Gondwana and then broke away and drifted north to a collision with the southern margin of Asia.) Conversely, the terranes that did not originate as a fragment of some earlier continent generally embody a fairly simple history spanning less than 200 million years, the normal maximum survival time for an ocean floor. The composition of such terranes tends to resemble that of a modern oceanic island or plateau. Some terranes consist chiefly of consolidated pebbles, sand and silt; they represent sedimentary fans that accumulated in an ocean basin, commonly between colliding crustal fragments.

The geometry of a terrane is the product of its history of movements and tectonic interactions. Terranes born on an oceanic plate retain their shape until they collide and accrete. Then they are subjected to crustal movements that modify their shape. For example, the terranes of the Brooks Range in Alaska are great sheets stacked one on top of another. Elsewhere in the cordillera, or chain of mountain ranges, in western North America the terranes are elongated bodies. The elongation reflects the slicing of the crust by a network of northwest-trending faults including the San Andreas fault of California. In Asia the terranes tend to have retained the shapes they inherited from episodes of rifting; some smaller terranes, however, were caught in collisions between the larger ones and got distorted. The assemblage of terranes in China is being stretched and displaced in east-west directions as India continues to squeeze Asia from the south.

The precise history of the movement of an individual terrane is not always known. Indeed, it is only recently that paths have been documented for a few of the earth's terranes. Since by definition terranes are fault-bounded and distinct from their geologic surroundings, each of them must have moved a distance at least equal to its longest dimension. The actual distances vary greatly. Some basaltic seamounts now accreted to the margin of Oregon have moved a minimal distance, from a nearby offshore origin. Yet similar rock formations around San Francisco have come as far as 4,000 kilometers across the Pacific. At a rate of just 10 centimeters per year a wandering terrane could complete a circuit of the globe in just 400 million years. Little wonder that the continents are patchwork agglomerations of terranes.

Lines of Evidence

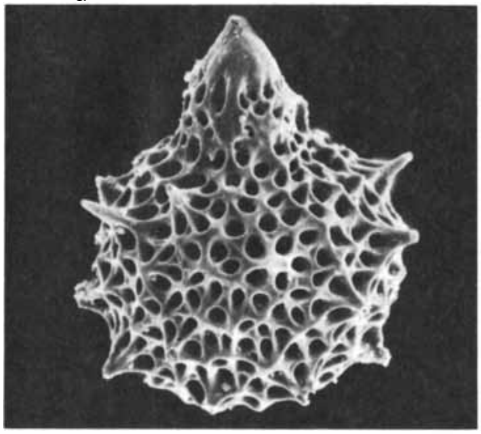
How can the history of a terrane be reconstructed? Fundamentally the origin of each rock unit composing a terrane gives insight into the evolutionary history of the terrane. The sedimentary rocks indicate depositional environments of the past: they suggest ancient river gravels, coral banks, deltaic sands, muds of a continental shelf or muds of an ocean abyss. The age of the sedimentary rocks is also important. One aid in determining age is the fossil record. Until recently, however, the record was incomplete: the fossil evidence came mostly from rocks deposited in ancient shallow marine environments. Only in the past decade has it become possible to determine the age of rocks representing deeper oceanic deposits—rocks that are essential to an understanding of the events that created many of the earth's mountain belts. The breakthrough involved radiolarians and conodonts.

Radiolarians are single-cell organisms that appeared in the oceans as early as the Cambrian period, some 500 million years ago. and were abundant until as recently as 160 million years ago. They occupied the upper levels of the ocean, but their skeleton consisted of silica, a substance of very low solubility at all levels of the sea. Accordingly abyssal muds are often rich in what is called radiolarian ooze: the accumulation of their skeletal silica. The rock called flinty chert, exploited for the making of arrowheads and knives by many peoples, is in fact the hardened ooze. Once the technique was perfected for extracting radiolarians from chert by dissolving the rock in a strong acid, thousands of chert assemblages could be dated.

Conodonts too are microscopic fossils, and like radiolarians they can now be extracted from rock by dissolving the rock in strong acids. Their biological identity was elusive. Now, however, it seems certain they are skeletal remains from the feeding apparatus of an extinct group of small, marine, wormlike animals that lived from 570 to 200 million years ago. Because conodonts are found in association with other fossil assemblages, it has been possible to establish a time scale of changing conodont morphology. In fact, for rocks more than 250 million years old the basis for radiolarian biostratigraphy (the determination of the age of a rock from the nature of the radiolarian fossils it holds) is the occasional joint occurrence of radiolarians and conodonts. By means of radiolarians and conodonts many rocks whose age was unknown a decade ago can now be dated. The results are often surprising. Sequences of old rocks are sometimes found to be lying on younger ones; hence great piles of strata seem to have been reshuffled. In some cases the stacking surfaces are parallel to the strata and the rocks show no obvious sign of repositioning.

Structural geology also has a part

d



tween 100 and 150 million years old and is enlarged 700 diameters. The other (*d*), perhaps of genus *Stichocapsa,* is of similar age and size. The scanning electron micrographs were provided by Benita L. Murchey and David L. Jones of the U.S. Geological Survey.

in the analysis of terranes. The reason is plain: the movement of rocks past (or over or under) one another imparts folds, crenelations, lineations and foliations, microscopic and macroscopic, all of which are a help in reconstructing the movement. Many of the folds one sees in the field, however, represent not the primary direction of motion but a later motion that consolidated terranes into tighter packages. The structural data are complemented by analysis of the composition of the rock. There seem to be no sure correlations between mineral assemblages or chemical composition and particular modes of origin, but there are generalizations that often hold true. For example, granite whose content of strontium is unusually rich in the isotope strontium 87 tends to result from the solidification of magma within an existing old continent, whereas granite poor in strontium 87 indicates an origin in an oceanic setting.

An important contribution to reconstructing the history of terranes derives from the study of remanent paleomagnetism: the alignment of minute magnetic particles in a rock, induced, at the time the rock formed, by the magnetic field of the earth. The investigations assume the earth's magnetic field is essentially that of a dipole, or bar magnet, coincident with the planet's axis of spin, so that the field lines at the Equator are horizontal (parallel to the surface of the earth) and the field lines at each pole are vertical (perpendicular to the surface). Between the Equator and the poles the field lines steepen. Most sedimentary and volcanic rock is deposited horizontally in sheets; hence the tilt of its remanent paleomagnetism can reveal the latitude at which the rock was formed. In addition the compass direction of the orientation of the magnetism can suggest that the rock has been rotated at some point in its geologic wanderings.

Collage Tectonics

Drawing on all these lines of evidence, I shall now take up, in terms of terranes, the rebuilding and reshaping of the continents, past and future. Some numbers are worth keeping in mind. The present global volume of continental crust is approximately 7.6 billion cubic kilometers; the oldest rock known is 3.8 billion years old. Dividing the first number by the second yields a simple linear estimate that the continents have grown at a worldwide rate of about two cubic kilometers per year—or roughly 65 cubic meters per second. The estimate may be too high; the growth processes in the hot, primitive earth may have been more rapid than the average. A host of crustal growth curves have been proposed; most of them hypothesize that between 70 and 80 percent of all crustal growth occurred more than two billion years ago. The final 20 to 30 percent of the current mass of the continents would then have accumulated over the past two billion years, at an average rate of from .7 to 1.1 cubic kilometers per year, a rate well within the rate of crustal contributions from modern volcanic arcs and oceanic seamounts.

The accumulation amounts to the plastering of terranes onto preexisting cratons, or continental nuclei, the oldest parts of the continental crust. The process can be followed in greatest detail through Phanerozoic time, the span of approximately 600 million years for which the fossil record of multicellular life is abundant. At the beginning of the span the existing continents (according to the paleomagnetic data) were isolated masses strung around the globe in the equatorial region. (The period from 700 to 500 million years ago was apparently an age of major continental breakup.) In the ensuing 350 million years the shifting of the continents resulted first in the agglomeration of two megacontinents, Gondwana and Laurasia, and then, 250 million years ago, in the union of the two to form the supercontinent Pangaea, a broadly crescent-shaped mass with a general north-south orientation. The old continental nuclei, augmented by terranes that had accumulated since the beginning of the Phanerozoic, then began to break up again some 200 million years ago along a new pattern of rifts resembling the modern 56,000 kilometers of globe-girdling oceanic spreading centers.

One can imagine, in the earth sever-

CONODONTS, unlike radiolarians, were part of a larger organism: they are skeletal remains, no more than a millimeter long, from the feeding apparatus of extinct marine worms that lived from 570 to 200 million years ago. Their shape and their surface ornamentation establish the age of the rock in which they are found; their color characterizes the maximum temperature attained by the rock, a matter of importance in exploration for hydrocarbons and certain minerals. (Between 50 and 550 degrees Celsius the organic matter in conodonts changes sequentially from pale yellow to brown, black, gray and white, and finally it loses all color; a temperature exceeding 150 degrees eliminates the possibility of liquid hydrocarbons.) The conodonts at the left, from the Brooks Range in northwestern Alaska, lived in shallow-water marine environments about 360 million years ago. Their color (pale yellow to light brown) indicates that the host rock never got beyond 90 degrees. The conodont at the center, from a small clast, or fragment, in a conglomerate outcropping along the Yukon River in eastern Alaska, lived at relatively shallow depths; the clast is rock that was deposited about 325 million years ago. The gray color indicates that the rock reached a temperature of at least 400 degrees. The conodonts at the right, from the Glacier Bay region of southeastern Alaska, lived in deep water approximately 230 million years ago. The blue-black color indicates a rock temperature of at least 300 degrees. The micrographs were provided by Anita G. Harris of the U.S. Geological Survey.

al hundred million years from now, the formation of a new supercontinent consisting of Asia and North and South America. The Pacific will have closed, following the subduction of the East Pacific spreading ridge. Meanwhile the Atlantic will have continued to widen. One can also predict that by then the colliding continents will have increased in size. The surface area of the continents ringing the Pacific today is 290 million square kilometers, of which the newly accreted terranes (that is, the post-Pangaea material) contribute approximately 25 million square kilometers, or 9 percent. If one assumes the crust has an average thickness of 20 kilometers, the crustal growth rate in the Pacific over the past 200 million years has been 2.5 cubic kilometers per year. The rate is somewhat misleading: the tectonic collages ringing the Pacific include some large terranes consisting of displaced continental crust that formed before Pangaea broke up. Examples include the eastern half of Mexico, the Brooks Range of Alaska, parts of northeastern Russia and most of the Malay Peninsula. Still, the preliminary investigations do seem to indicate that growth rates for the continental crust that rings the Pacific exceeded the global average rate of one cubic kilometer per year.

In this regard some recent studies of remanent paleomagnetism in limestone blocks of northern California are intriguing. The studies indicate that the limestone, which ranges in age from 85 to 100 million years, was deposited south of the Equator. Yet the age of sedimentary rock that now laps across the blocks suggests that the limestone (and associated basalts) became accreted to the margin of California no later than 38 million years ago. The figures require the plate carrying the limestone to have moved northward at a rate of between 15 and 30 centimeters per year, which is faster than the plates are moving now. The figures lead me to speculate that the growth rate of the continents may also vary. Perhaps the growth comes in cycles hundreds of millions of years in duration.

ONE VAST OCEAN, Panthalassa, dominated the surface of the earth some 250 million years ago, when essentially all the planet's continental crust was agglomerated into a single supercontinent, Pangaea. Since that time Pangaea has broken into the continents of today and the oceanic crust of the Panthalassan basin has been completely subducted (directed downward back into the earth's mantle). Its place is now taken by the ocean basins of today. The crustal flotsam and jetsam of Panthalassa (consisting of fragments of continental crust along with crust created by volcanic activity) yielded terranes that now augment the continents ringing the Pacific. Such terranes appear in orange in the next two illustrations.

The Importance of Sediments

Sediments play far more than a passive role in the growth or diminution of the continents. They too become part of terranes. For one thing, thick piles of sediments remain on the continents and accumulate along their rifted margins. Only about 30 percent of the sediments discharged from rivers make it beyond the continental margin and settle onto oceanic crust. In the second place, part of the sediments carried to subduction zones by the oceanic conveyor belt seems to get plastered onto the overriding crustal plate as an "accretionary prism" or becomes attached to the underside of the overriding plate. Then too, great piles of sediments are commonly caught between colliding crustal masses. An example is seen today in the Molucca Sea, where two island arcs are colliding. A sediment pile may be a terrane in itself; this is the case for the accretionary prisms in the region of Kodiak Island and the Gulf of Alaska. Alternatively, the pile may form the matrix in which terranes are embedded, as in the case of the Alaska Range.

In the world today the greatest single source of sediment is the towering landform resulting from the collision of India and Asia. There the Asian crust has overridden the Indian terrane, doubling the thickness of the crust and creating the Himalayan Mountains, and to their north the Tibetan plateau. Six great river systems, the Huang He, the Yangtze, the Irrawaddy, the Mekong, the Ganges-Brahmaputra and the Indus, drain the region, which amounts to only 4 percent of the world's land surface. Together they discharge into the oceans some 3.8 billion tons of sediment per year, or as much as 40 percent of all the sediment discharged by all the rivers of the earth.

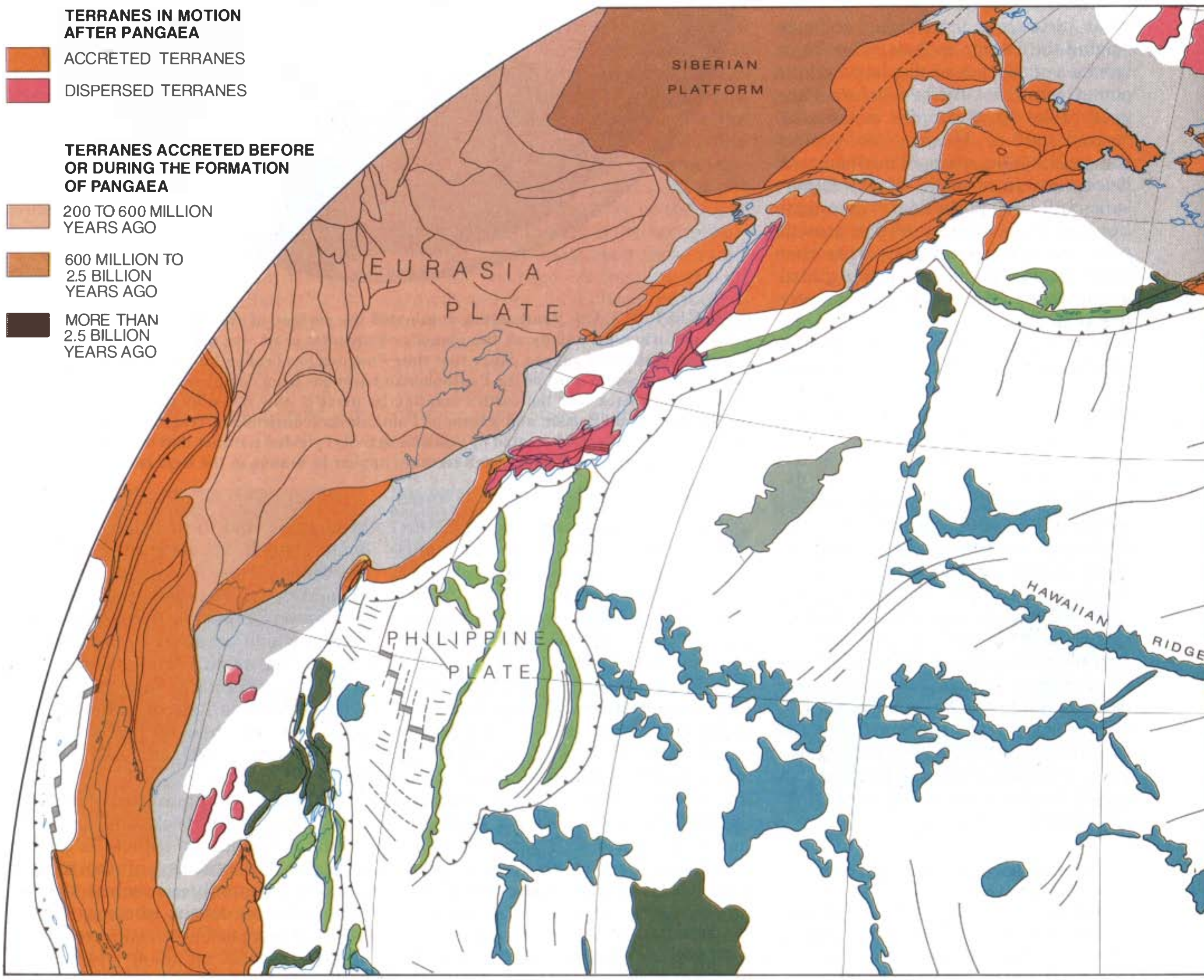
The sediment is composed of silt and clay, along with rock and mineral grains. A further constituent is water, owing to the porosity of the solids, which is commonly as much as 50 percent. If the net density of the sediment is two grams per cubic centimeter, the volume of the sediment discharged each year from Asia's rivers is 1.7 cubic kilometers; the volume worldwide is from 4.5 to 6.8 cubic kilometers. As the sediment compacts and becomes stone the porosity decreases to almost zero. The worldwide discharge therefore amounts to some 3.3 to 4.9 cubic kilometers of rock per year. (I have assumed a rock density of 2.75 grams per cubic centimeter, which is a bit greater than the density of common quartz.)

The long-term fate of most of the rock is unknown. Some of it could be subducted; some could be lifted between colliding continental massifs; some could be torn away from its place of accumulation (say in a river delta or a deep-sea deposit) and ultimately become accreted to a far-off continent. Still, the 3.3 to 4.9 cubic kilometers of rock that can form each year from sedimentation along continental margins and on the ocean floor far exceeds the .2 cubic kilometer contributed each year by basaltic volcanism and the .75 to 1.5 cubic kilometers contributed by explosive subduction-zone volcanism. Little wonder that sedimentary rock, or its metamorphosed equivalent, is a major component of mountain fold belts. Indeed, the estimates I have given suggest that as much as 75 percent of newly formed continental crust could consist of sediment and metamorphosed or melted products of sediment.

Terranes of the North Pacific

The making and shaping of continents out of terranes is best explored by examining specific regions of the world. My own interest is the Pacific, which I divide into quadrants. Each quadrant displays a distinctly different geologic pattern, reflecting contrasting histories of the accretion and dispersion of terranes.

In the northeastern quadrant, where the Pacific washes the western coastline of North America, terranes consisting chiefly of island arcs and other oceanic material have piled onto the coastline throughout the past 180 million years. The cordillera inland from the coast includes a network of northwest-trending strike-slip faults. Here the motion is primarily horizontal, along the plane of the earth's surface;

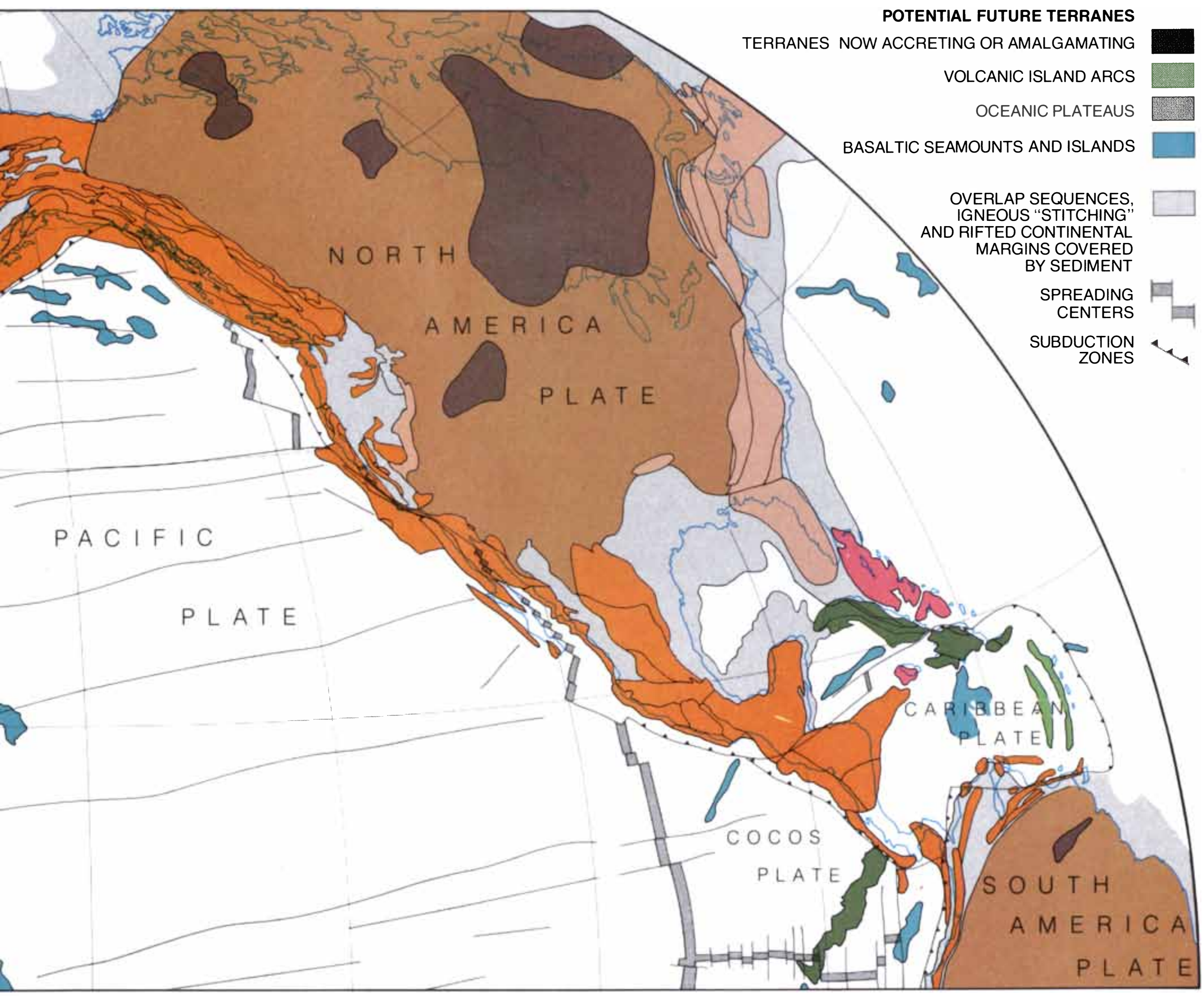


TERRANES OF THE NORTH PACIFIC dominate a map that shows somewhat more than a fourth of the surface of the earth. The spreading center of the Pacific is well to the east of the center of the Pacific; here part of it is visible south of Central America. The Pacific floor spreads from there, to be subducted along great trenches, which mark subduction zones along the margins of the Pacific. Colors indicate the ages at which terranes accreted. The first terranes (*brown*), older than 2.5 billion years, include the ancient cores of the continents. The youngest accretions (*orange*), which account for some 9 percent of the surface area of the continents ring-

thus the newly accreted terranes have been distended into a sequence of slivers. Wrangellia, a well-studied terrane that once lay at, or even south of, the Equator, is a good example. Wrangellia crashed into Oregon some 70 million years ago. Subsequent faulting has spread fragments of Wrangellia northward, leaving parts in eastern Oregon, on the Vancouver and Queen Charlotte islands and throughout the Wrangell Mountains of southern Alaska. Central and western Oregon consists of terranes swept in after Wrangellia was in place.

The Brooks Range of northern Alaska has a quite different story. There huge, thin sheets of strata representative of a continental margin have been thrust over one another, transforming a geography that once measured at least 500 by 1,000 kilometers into a crustal stack 500 by 300 kilometers. Analysis of clues to the directions of flow of the sediment-bearing fluids that contributed to the strata indicates that the entire stack must have moved to its current position, but from where is a matter of debate. According to one hypothesis, the stack emerged by a counterclockwise rotation from the islands of the Canadian Arctic. The Canadian basin of the Arctic Ocean would then represent the depression left by the pivoting landmass.

The northwestern quadrant of the Pacific includes Asia, Japan and the Philippines. Here the continental crust consists of old continental fragments, each of them flanked, or even surrounded, by belts of terranes that accreted during the Paleozoic era, from 600 to 250 million years ago. In effect the Siberian platform seems to have been a backstop on which a succession of terranes built outward. Along the southern border of the platform volcanic arcs and other crustal chunks piled up in the early Paleozoic, from 600 to 400 million years ago; they formed the Baikalean fold belt, the mountainous region between Mongolia and the Sea of Okhotsk. Then, from about 300 million years ago to about



ing the Pacific, represent crustal debris swept up by the subduction of Panthalassa. Modern oceanic plateaus, seamounts and island arcs represent possible future terranes. On the North American side of the Pacific the coast is formed by terranes consisting of oceanic debris, including island arcs, distended into slivers. Northern Alaska (the Brooks Range) represents terranes thrust one over another; the terranes consist of continental margin and oceanic crust. On the Asian side of the Pacific the continental crust consists of ancient terranes surrounded by belts of younger terranes. The ancient core of Asia was a backstop on which younger terranes accumulated.

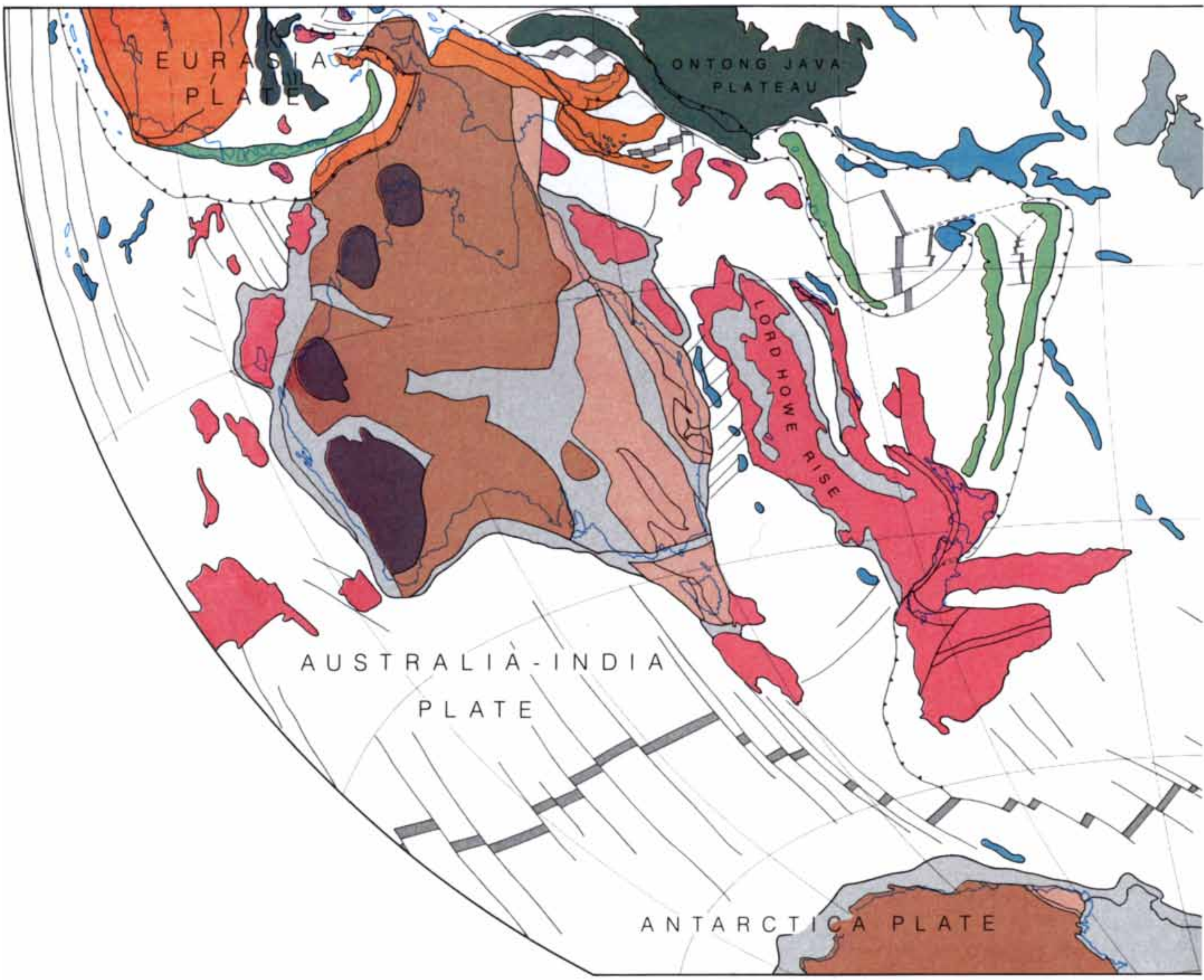
60 million years ago, when India came crashing in, a number of terranes (Tarim, Yangtze, the Sino-Korean massif, Indochina and finally India) came together to form Asia.

A similar episode of accretion is now in progress. The Ontong Java oceanic plateau is probably a rifted fragment of continental crust. Today it is mostly submerged. In size it is comparable to the Yangtze terrane of Asia. Accreted against the southern side of the Ontong Java plateau is part of the New Hebrides volcanic arc. Many other arcs lie nearby. If the Coral Sea closes up—an event that is hard to predict, since in some places the sea is opening, while in others it is closing—the Ontong Java plateau, wreathed with accreted volcanic arcs, is likely to form a major addition to Australia.

Terranes of the South Pacific

The southwestern quadrant of the Pacific, including Antarctica, Australia and New Zealand, is a region characterized tectonically by a radial dispersion: the landmasses result from the breakup of part of Gondwana that began between 120 and 100 million years ago, when a three-pronged rift system developed. One prong created the Tasman Sea; the other two prongs separated Antarctica from Australia and the Campbell Plateau of New Zealand. Fold belts in western Antarctica, eastern Australia and New Zealand suggest the earlier history of the region. Evidently episodes of accretion had built the continental crust outward from nuclei now positioned in eastern Antarctica and western Australia.

That leaves the southeastern quadrant of the Pacific, which includes the western shore of South America. It consists of contrasting regions. In the southern part of the quadrant new terranes are being carved in the Scotia Sea, which is cutting between the southern Andes and the Antarctic Peninsula, leaving small, fault-bounded crustal blocks such as South Georgia and the Orkney Islands. Northward, from southern Chile to Peru, the Andes extend in a nearly straight line for 3,000 kilometers. The region beckons

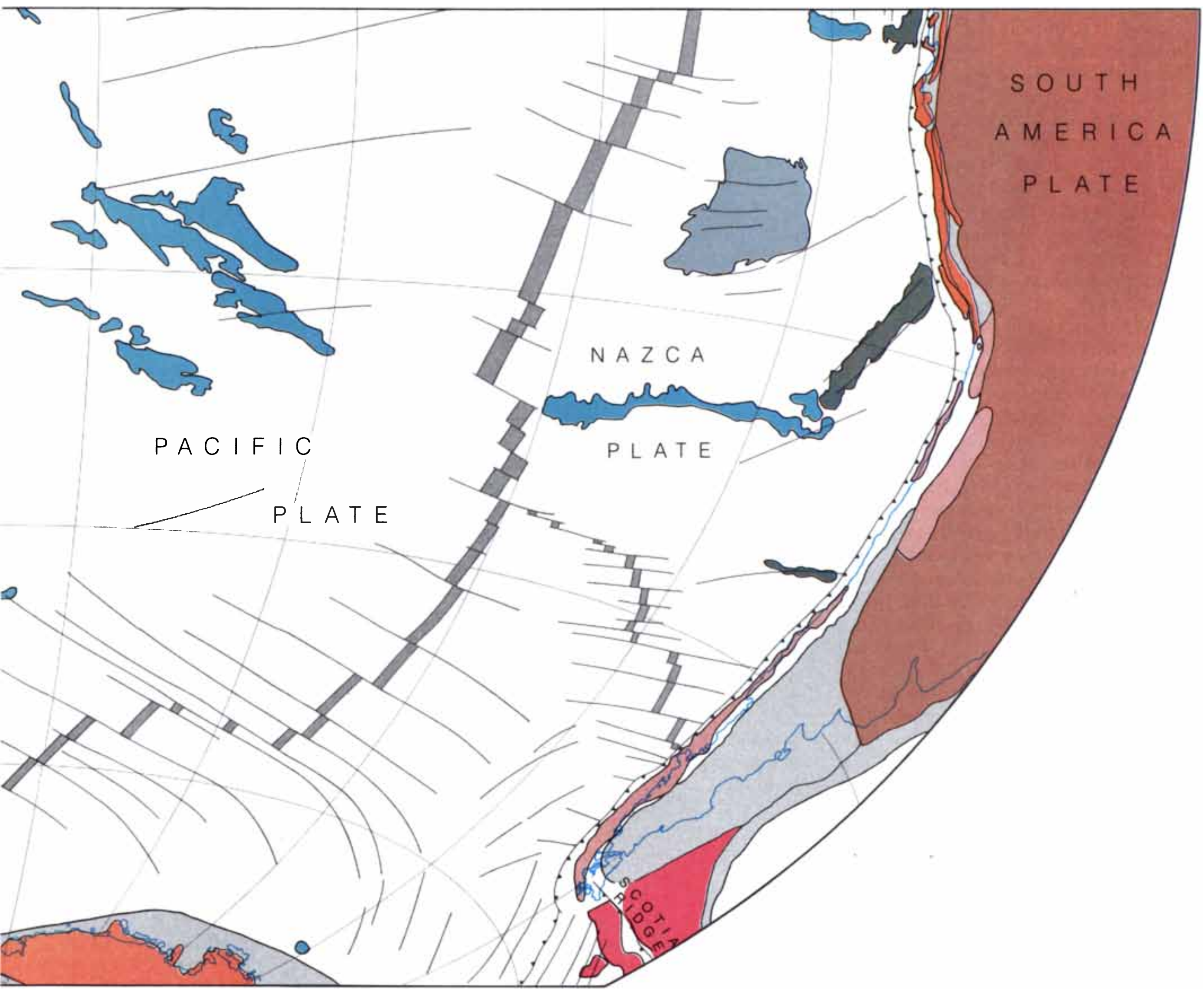


TERRANES OF THE SOUTH PACIFIC occupy a map that complements the preceding illustration and completes a display of the span of the Pacific. In the western part of the map Antarctica, Australia and New Zealand mark the remains of a three-pronged rift that began to open, from 120 to 100 million years ago, in the southern part of Pangaea, which formed from the southern megacontinent Gondwana. Northeast of Australia the accretion of terranes is now in progress. The Ontong Java oceanic plateau, a fragment of continental crust, has already been augmented, along its southern edge, by part of the New Hebrides island arc. Other arcs may

for more study, but from what is now known there appears to be little sign of terrane accretion, in spite of the past 200 million years of subduction. (Oceanic crust is being subducted under the continental crust of Chile and Peru.)

From central Peru to the Caribbean, western South America displays a patchwork fabric of accreted terranes. Here, as at the southern end of South America, crustal dispersion is active. For example, the petroleum-rich sediments under Lake Maracaibo in northwestern Venezuela fill part of a sag in the crust left when the Antilles island arc of the Caribbean plate slid eastward past the northern edge of South America. Like two bulldozers, the Scotia arc in the south and the Antilles arc in the north are advancing into the Atlantic, and along the flanks of the advance lies crustal wreckage, which presumably will serve as raw material for the building of future terranes.

The concept of terranes is becoming part of the plate-tectonic theory as the patterns and processes of continental growth are examined. Overall the budget for continental growth is dynamic. Volcanoes contribute from .75 to 1.5 cubic kilometers per year, and hot spots contribute .2 cubic kilometer per year. Meanwhile, owing to erosion, the continents lose as much as four cubic kilometers per year, but as much as three-fourths of the loss is recycled: the erosional sediment is lifted, folded, sometimes metamorphosed and sometimes melted, in the collisional, accretionary processes that raise mountains along the continental margins. The averaging of billions of years of geologic history probably masks pulses or even cycles of continental growth. In the aftermath of the breakup of Pangaea 200 million years ago an entire global-scale ocean, Panthalassa, has been consumed and the circum-Pacific continental crust has grown at a rate of as much as 2.5 kilometers per year, which seems to have exceeded the average rate of long-term growth. The events in the Pacific seem, therefore, to require complementary periods of tectonic tranquillity. Charting the detailed history of the earth in terms of terranes is the challenge now facing geology.



join it, and the entire assemblage may ultimately join Australia. In the eastern part of the map the Pacific plate is being subducted under the South America plate. South of the tip of South America new terranes are forming: the oceanic crust of the Scotia Sea is slicing eastward between South America and the Antarctic Peninsula, carving terranes along its margins before it gets subducted. The Pacific maps accompanying this article display results of investigations by David L. Jones, Erwin Scheibner of the University of Sydney, Zvi Ben-Avraham of Stanford University, Elizabeth R. Schermer of the Massachusetts Institute of Technology and the author.

Skin Breathing in Vertebrates

It can supplement or replace breathing through lungs or gills. Special adaptations of the skin and the circulatory system help to regulate the cutaneous exchange of oxygen and carbon dioxide

by Martin E. Feder and Warren W. Burggren

Perhaps because people breathe almost exclusively with their lungs, respiration is often thought of as taking place only in specialized organs: if not in the lungs, then in the gills of fishes and crustaceans, the tracheae of insects or the book lungs of spiders. Yet whenever a relatively thin membrane separates the respiratory medium (the air or water an animal breathes) from living cells or flowing blood, oxygen can enter the cells and the blood and carbon dioxide can leave them. The recent findings we shall discuss here suggest that the skin serves as an effective and highly regulated organ of gas exchange in many vertebrates.

Cutaneous gas exchange has long intrigued physiologists. Pioneering studies were made by August Krogh at the turn of the century in Denmark. Krogh obstructed the airflow to the lungs of frogs and observed that the skin could supply enough oxygen to the blood during the winter, when frogs are normally quiescent; during other seasons, however, lungs proved to be necessary. Krogh's research on blood circulation through capillaries earned him the 1920 Nobel prize in physiology or medicine.

Numerous experimental studies in the 1960's and 1970's examined the partitioning of gas exchange among an animal's respiratory organs: the skin, lungs and gills, if present. For example, by placing plastic masks over the faces of salamanders, Victor H. Hutchison and his students at the University of Rhode Island were able to determine what proportion of the animal's total oxygen is taken in through the skin and what proportion of the total carbon dioxide is eliminated through it. The results of these and many other such experiments provide a surprisingly long list of vertebrate skin breathers.

Probably the best-known of these skin breathers are the amphibians. Among amphibians it is not unusual for at least 30 percent of the total oxygen uptake and as much as 100 percent of the carbon dioxide elimination to take place across the skin. Frog larvae, for example, exchange approximately 60 percent of their respiratory gases through their skin even though they also have both gills and lungs.

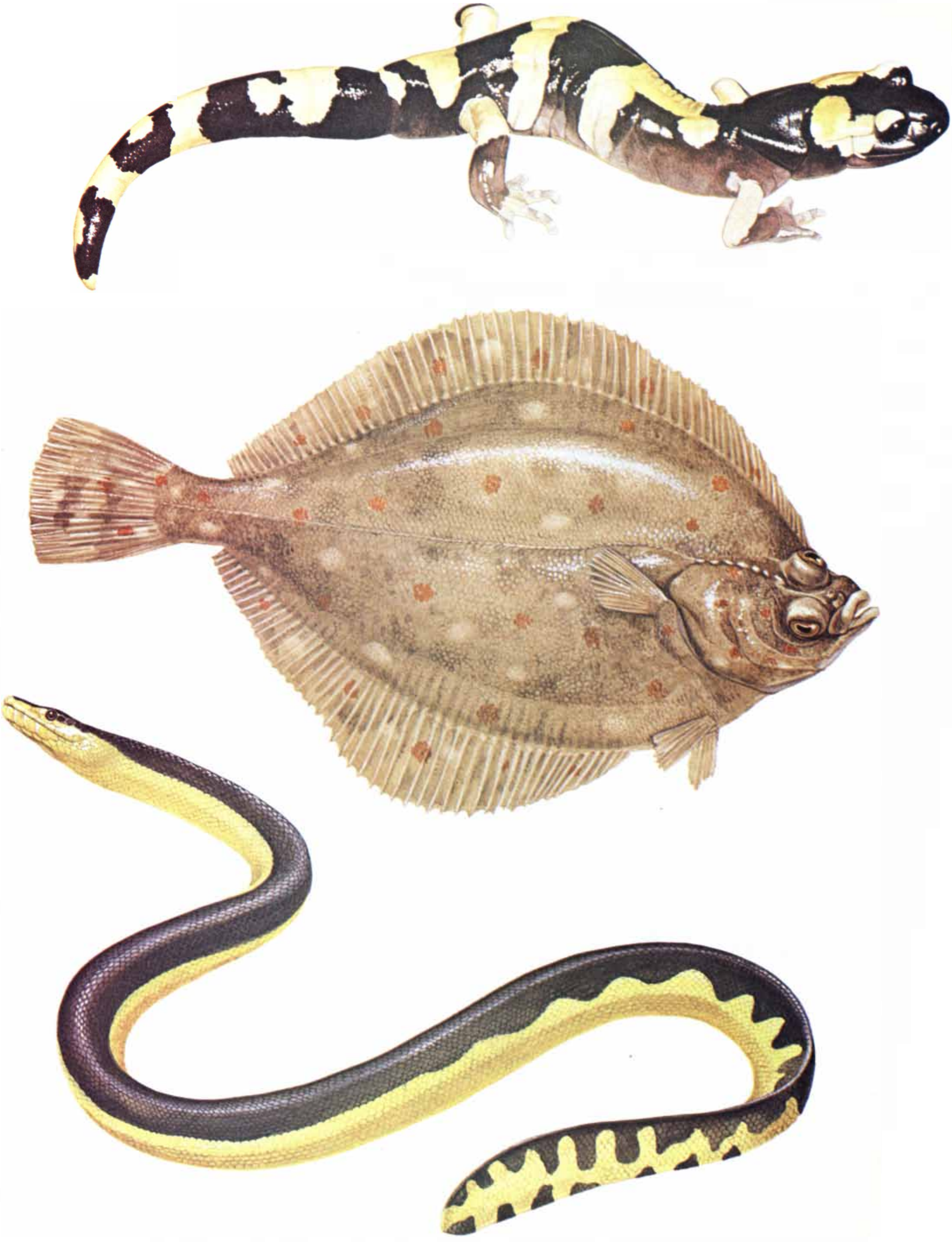
Amphibians that spend almost all their time on land rather than in water face potentially life-threatening difficulties because the same characteristics that make skin an effective membrane for gas exchange also facilitate water loss. Nevertheless, cutaneous gas exchange has been found in every species of terrestrial amphibian examined for this capacity. In fact, the skin is the sole respiratory organ in adult salamanders of the family Plethodontidae. Scores of these species inhabit terrestrial environments as diverse as the ground litter of New England woods and the canopy of tropical rain forests. Although some tropical plethodontids may attain a body mass of 150 grams and a body length of 24 centimeters, they accomplish all respiratory gas exchange between tissue and environment by way of their skin.

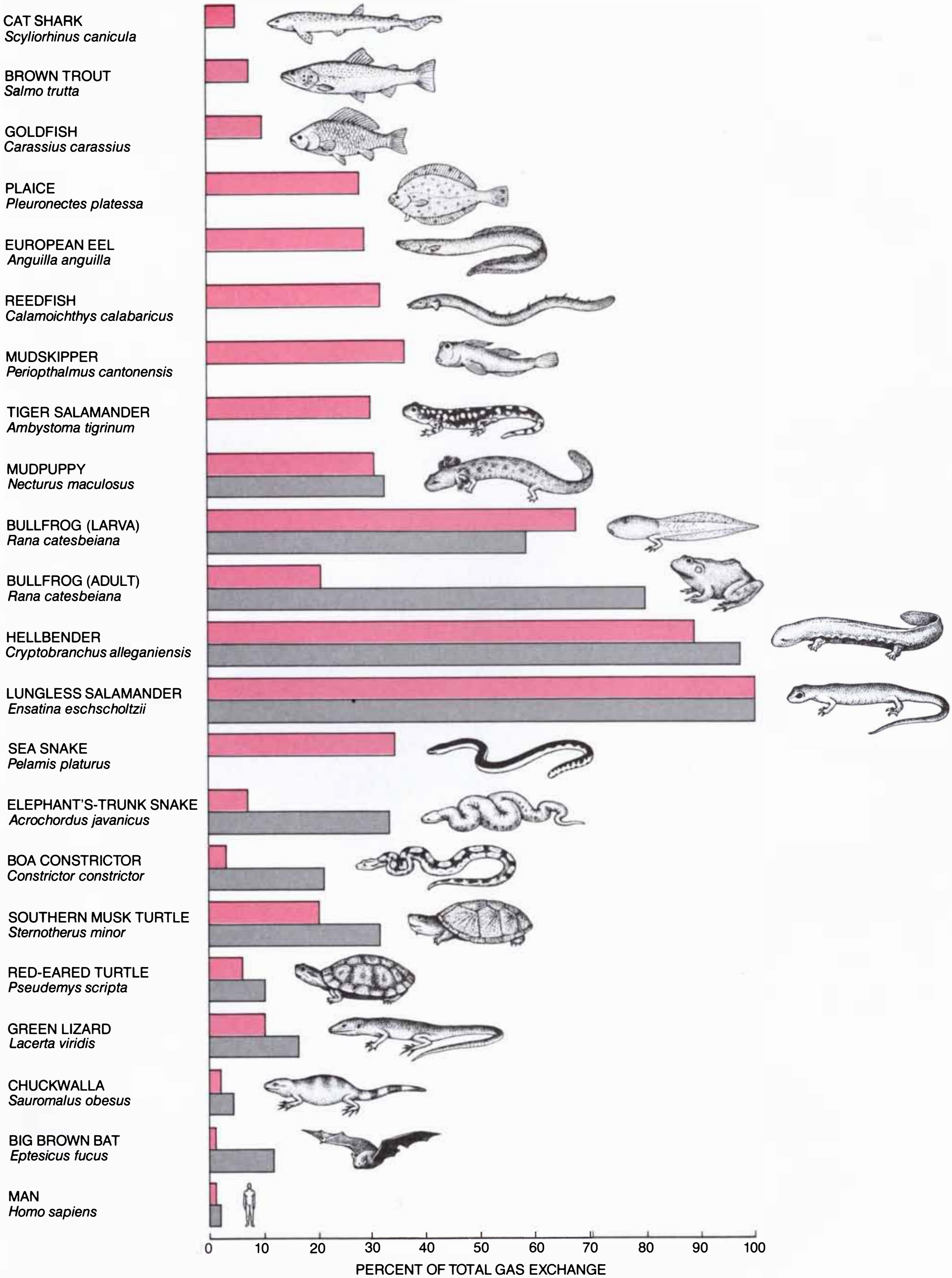
Skin breathing has been looked for in other vertebrates, from fishes to mammals. It has been demonstrated experimentally, for example, that amphibious (air breathing) fishes rely on their skin for as much as half of their gas exchange, particularly if they venture onto land: gills typically collapse in the air and become useless. The adequacy of the skin as an air-breathing organ seems to have escaped the attention of many biologists who assume that lungs were a prerequisite for the evolution of terrestrial animals.

Commoner types of fishes also rely on cutaneous gas exchange. Sharks, trout, cod and goldfish—to name a few—acquire between 5 and 30 percent of their total oxygen by way of the skin. Apparently neither the structure of the integument nor the shape of the body determines the extent to which a particular fish breathes through its skin. Both the flat, scaleless plaice and the elongated, heavily scaled reedfish derive about a third of their required oxygen cutaneously.

Reptiles have been shown to take advantage of cutaneous gas exchange as well. In spite of their thick shells, some freshwater turtles rely heavily or even totally on cutaneous gas exchange, particularly while passing the winter in ice-covered ponds. Snakes inhabiting both fresh and salt water typically exploit their skin as a respiratory organ. These reptiles often dive for long periods and supplement oxygen stores in the lungs and blood with oxygen acquired by way of the skin. Roger S. Seymour of the University of Adelaide in Australia has even suggested that some sea snakes use their skin to eliminate nitrogen during prolonged deep dives, thereby preventing decompression sickness (the formation of nitrogen bubbles in blood) as they surface. Many reptiles that live in arid environ-

THREE SKIN-BREATHING VERTEBRATES are depicted on the opposite page: a lungless salamander, a plaice and a sea snake. The salamander *Ensatina eschscholtzii* lives in the forests of California and Oregon; it is about five centimeters long when fully developed. The skin is the sole respiratory organ in adults of this species. The plaice *Pleuronectes platessa*, a commercially valuable flatfish of the North Atlantic, takes in 27 percent of its oxygen through the skin. The fish can grow to a weight of five kilograms and a length of 90 centimeters. The sea snake *Pelamis platurus* absorbs more than a third of its oxygen and excretes more than three-fourths of its metabolic carbon dioxide cutaneously. The snake reaches a length of about a meter; it ranges the tropical Pacific Ocean, between America and Africa.





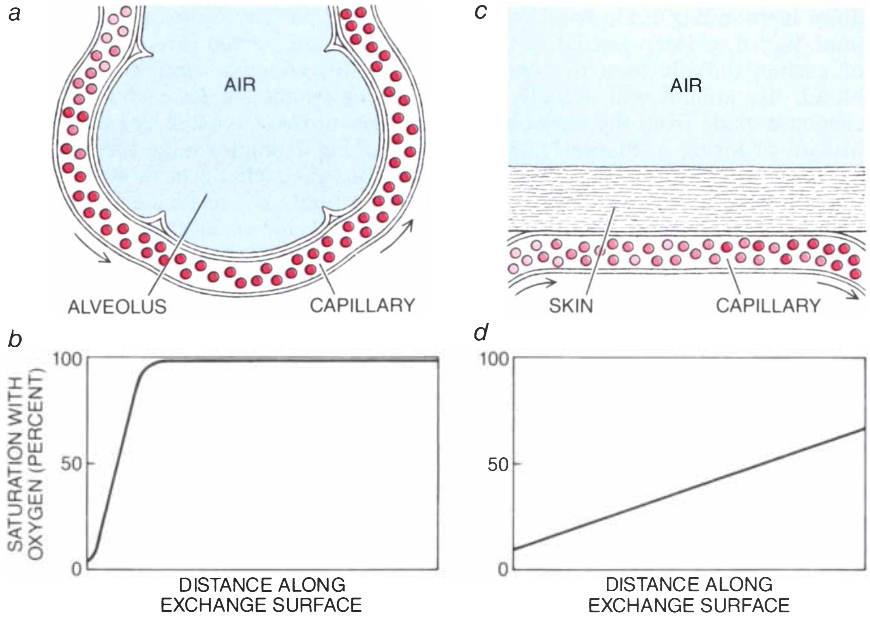
CUTANEOUS GAS EXCHANGE is widespread among vertebrates. Although it is most prominent in amphibians, this mode of respiration is also important in many other animal groups. Cutaneous excretion of carbon dioxide (*gray bars*) typically accounts for a larger fraction of total gas exchange than cutaneous oxygen uptake does (*colored bars*) in those cases where both gases were measured.

ments, where dehydration is a constant danger, have countered this problem by evolving thick scales, a shell or leathery skin. A few species, such as the chuckwalla lizard (*Sauromalus obesus*) of the southwestern deserts of the U.S., do exhibit some cutaneous gas exchange regardless of their thick, protective skin.

Although cutaneous respiration is often significant and even crucial in lower vertebrates, the skin is seldom an important avenue for gas exchange among higher vertebrates such as birds and mammals. Even if the furred or feathered integument of such creatures were as permeable as the skin of frogs, the higher vertebrates would need lungs to sustain their much greater metabolic rate. Lungs offer a larger and thinner surface for gas exchange than the skin can provide. Yet there are exceptions to this generalization. Clyde F. Herreid II and his colleagues at Duke University found that in bats as much as 12 percent of total carbon dioxide elimination may take place across the thin, well-vascularized wing membranes.

Young mammals and birds are often born or hatched with thin skin, richly supplied with blood and lacking fur or feathers. Cutaneous respiration may therefore be more important in the developing forms of such higher vertebrates than it is in the adults. There is probably significant cutaneous gas exchange in the mammalian embryo and fetus; gas exchange through the shell and the shell membrane is the major avenue of respiration available for the eggs of birds and indeed of all egg-laying vertebrates [see "How Bird Eggs Breathe," by Herman Rahn, Amos Ar and Charles V. Paganelli; SCIENTIFIC AMERICAN, February, 1979].

What about cutaneous gas exchange in people? Human skin certainly does not serve as a respiratory organ for the general benefit of the body tissues. Yet the skin, like any living tissue, consumes oxygen and generates carbon dioxide. All the respiratory gases consumed or produced by skin cells, as well as an undetermined additional quantity of carbon dioxide from capillary blood flowing through the skin, are exchanged directly between skin and air. In fact, the permeability of the skin is exploited medically when drugs are administered by applying patches to the skin.



CUTANEOUS OXYGEN UPTAKE is limited by the thickness of the skin's diffusion barrier: the distance between the respiratory medium (air or water) and the blood. Oxygen can diffuse quickly through the thin alveolar membrane of the mammalian lung (*a*); blood-cell hemoglobin is rapidly oxygenated (*dark red*) and blood in adjacent capillaries is completely saturated with oxygen (*b*). Because diffusion of oxygen through the thicker skin (*c*) is much slower, the blood in capillaries under the skin is never fully saturated with oxygen (*d*).

In spite of their number, studies of cutaneous gas exchange generally have discounted the prevalence of skin breathing among vertebrates by stressing its limitations. For example, investigators have tended to contrast the skin to the elegantly structured lungs and gills. They have emphasized the circumstances under which cutaneous gas exchange would be precluded and pointed to the thick skin or scales and the limited cutaneous surface area of most vertebrates as characteristics likely to hinder cutaneous gas exchange. The thrust of most of these contentions rests on the physical principles governing gas exchange.

Oxygen and carbon dioxide pass through the membrane of a gas exchanger by means of a physical process called diffusion. Diffusion is defined as the equilibrating flow of matter, through the random movement of molecules in a fluid or gas, from a region of high concentration to one of low concentration. Because the gas-exchange membranes of lungs and gills are typically very thin (less than a thousandth of a millimeter in the human lung), equilibration with the respiratory medium is achieved extremely rapidly. A more important limitation to gas exchange in lungs and gills is the speed with which the blood carries oxygen away from the exchange surface or delivers carbon dioxide to it. Rapid diffusion makes it possible for animals with lungs or gills to regulate gas exchange by simply increasing or decreasing blood flow through the respiratory organ.

In contrast, increasing the flow of blood to the skin has been thought to have little effect on total gas exchange in vertebrates. Johannes Piiper, Peter Scheid, Randall Gatz and Eugene Crawford of the Max Planck Institute for Experimental Medicine in Göttingen have shown that cutaneous gas exchange is diffusion-limited. That is, the diffusion of respiratory gases through the relatively thick skin of vertebrates is so slow that it is not able to transfer oxygen and carbon dioxide as rapidly as the gases are transported from or to the skin by the blood and the respiratory medium.

Other studies demonstrated another problem encountered by cutaneous gas exchange. Equilibration between the gas concentrations in the blood flowing through capillaries inside the skin and in the respiratory medium outside the skin is governed by Fick's law of diffusion: the rate of gas exchange is proportional to the difference between the partial pressure of the gas in the respiratory medium and its partial pressure in the blood. (Partial pressure, the pressure of a gas in a solution or a gas mixture, reflects both the concentration of the gas and its solubility.) A gas will therefore diffuse through the skin only if its local partial pressure on one side of the skin is greater than its partial pressure on the other side; the greater the difference is, the faster it will diffuse.

As studies by Donald C. Jackson and his colleagues at Brown University have shown, this relation may have fateful consequences for skin-breathing vertebrates. If the respiratory medium surrounding a skin-breathing animal has a greater partial pressure of carbon dioxide than the animal's blood, the animal will actually gain carbon dioxide from the environment instead of losing it. Similarly, if high temperature or activity increases carbon dioxide production, many skin-breathing amphibians cannot immediately excrete the excess; they must first wait for the internal carbon dioxide partial pressure to rise above the external partial pressure. Fick's law governs the diffusion of oxygen as well. Our own studies have shown that amphibians may actually lose oxygen when their blood contains more oxygen than the water surrounding them.

Human beings can excrete excess carbon dioxide by simultaneously increasing blood flow to the lungs and raising the respiration rate while maintaining a constant partial pressure of carbon dioxide in the blood. The internal gas partial pressures of skin-breathing animals, on the other hand, seem to be poorly controlled and subject to the ever changing balance between external gas concentrations and internal demands for gas exchange.

The fact remains that in spite of the daunting limitations imposed by the diffusion process, vertebrates do rely on cutaneous gas exchange to a significant degree. We recognized that this paradox could be resolved by considering the different mechanisms vertebrates have at their disposal to regulate cutaneous gas exchange. Some of these mechanisms seem obvious, although they had seldom been recognized as potentially important in this respect. Other suggested regulatory processes are subtler. The various regulatory mechanisms can be divided into two categories: mechanisms that are permanent morphological adaptations or responses to long-term changes (days, weeks or months) in the environment or in the animal's physiology, and mechanisms that are called on from minute to minute as an organism's immediate respiratory needs or the environment varies. Together these mechanisms support a remarkably effective capability for controlling cutaneous gas exchange.

One obvious way to regulate skin breathing depends on the fact that cutaneous gas exchange is limited in part by the total skin surface area. Consequently a change in surface area can increase or decrease the amount of gas exchanged through the skin. In spite of a rigid internal skeleton and a relatively fixed form, many vertebrates exhibit seasonal changes in surface area that augment cutaneous gas exchange precisely in this way.

For example, some male amphibians engage in elaborate and strenuous courtship rituals that may include repeated body movements lasting for hours. Associated with this behavior is a greatly increased requirement for oxygen uptake and carbon dioxide elimination. Apparently in response to the increased respiratory burden, parts of the skin in these amphibians gradually become enlarged or develop outgrowths during the courtship season. Such surfaces act as accessory gas-exchange organs.

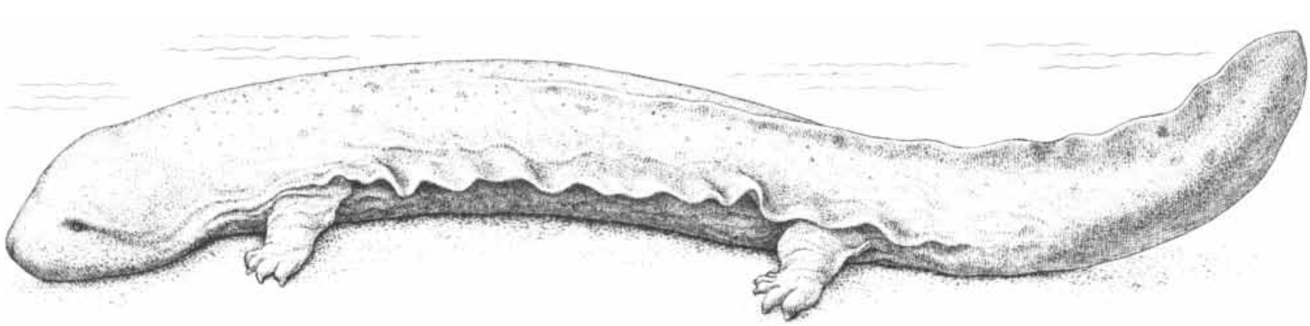
Changes in the skin structure to this end appear in the male hairy frog, *Astylosternus robustus.* Tiny dermal papillae that superficially resemble hair appear on its hindquarters. The enlarged tail fin and the dorsal crest developed by males of many species of newts during the courtship season also could contribute to the total amount of gas diffusing through the skin. The male skin enlargements typically regress when the breeding season ends. Moreover, such structures do not occur at all in the relatively quiescent females of these species.

In addition to these seasonal changes in cutaneous surface area amphibians have evolved a number of permanent morphological characteristics that promote skin breathing. Many amphibians have what seem to be disproportionately elongated bodies or tails. It has been argued that such shapes arose to equip these organisms with enough skin for adequate cutaneous gas exchange.

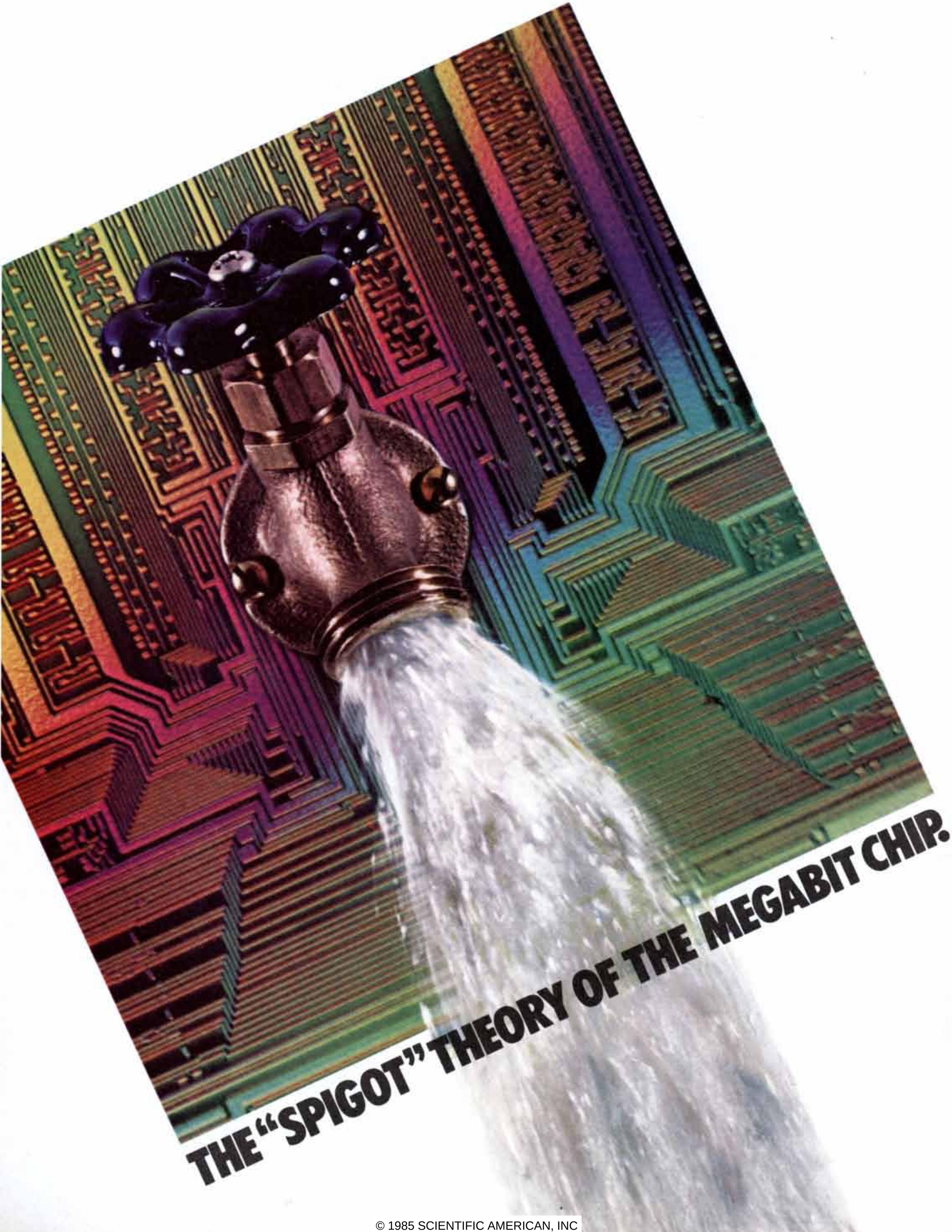
Other species of amphibians permanently bear numerous skin folds that also increase the surface area for cutaneous gas exchange. The most spectacular of these is the Lake Titicaca frog, *Telmatobius culeus,* which inhabits the lake for which it is named in the Andes between Peru and Bolivia. As Victor Hutchison and his colleagues at the University of Oklahoma discovered, the Lake Titicaca frog is so well adapted for skin respiration that it does not need to ventilate its lungs at all. One major explanation for this ability is the pendulous folds of skin that protrude from its hind limbs and trunk. Other species in the genus *Telmatobius* and in other genera of frogs have similar skin folds. In some types of amphibians such as the hellbender, *Cryptobranchus alleganiensis,* a large aquatic salamander, the skin folds are minute but numerous and are richly invested with capillaries.

A second adaptation enhancing the capability for cutaneous gas exchange is a reduction in the thickness of the skin to lessen its resistance to the diffusion of respiratory gases. In addition to being devoid of such physical barriers as hair, feathers or scales, amphibian skin is typically only between 10 and 50 micrometers (millionths of a meter) thick. Actually the significant morphological factor governing diffusion in this connection is not the total thickness of the skin but rather the thickness of the "diffusion barrier," the distance between the respiratory medium outside the skin and the blood flowing through the cutaneous capillaries. Hence any morphological change that reduces the distance between the bloodstream and the respiratory medium helps to increase gas exchange.

We demonstrated through a simple experiment that such changes are indeed possible within a fraction of an amphibian's lifetime. Frog larvae were reared in two enclosures, one containing well-aerated water and the other



HELLBENDER (*Cryptobranchus allegiensis*) is a large aquatic salamander found in swift-flowing streams of the eastern and central U.S. It reaches a length of about 70 centimeters. Although the hellbender has lungs, it breathes primarily through its very wrinkly skin.


THE "SPIGOT" THEORY OF THE MEGABIT CHIP.

There are megabit chips. And there are megabit chips. All of them can access more than 1 million bits of information. But not all of them can access these 1 million bits of information AT&T fast. And fast to AT&T is 20 million bits per second.

That means we can pour out 1,048,576 bits much faster than you can say "byte." And we can do it because only the AT&T megabit chip has a fast column access mode.

The "Trickle-Down" Theory Vs. The "Spigot" Theory

Most other megabit chips use a "page mode" to access their memory cells. When a chip is in this mode, its fastest data rate is about 10 million bits per second.

But the fast column cycle, developed by AT&T Bell Laboratories, pours out data twice as fast. And not only does data flow out faster, getting data out is easier with the fast column, because access timing requirements are less demanding. Valid data are always available during the entire access cycle.

And speed, ease of use and timing tolerance will make it easier for AT&T to design new and improved products and high-speed systems around the megabit chip.

The Megabit Special

Because this megabit chip is designed for manufacturability, you might want to know what, besides its fast column, makes it special. It's made using an advanced CMOS—Complementary Metal-Oxide Semiconductor—process that makes it possible to provide high performance at reduced power consumption.

As a matter of fact, it uses 1/8 the power per bit of most 256K DRAMs. The lower power requirement is a plus for any system using the megabit—allowing for lower operating costs and reduced cooling requirements.

The AT&T megabit chip is smaller than a dime, yet has more than 2 million elements on it. Among those 2 million-plus elements are more than 20,000 spare memory cells for safety's sake. If for some reason there's a bad cell in a chip, a computer-controlled test targets the cell—and its whole row or column—for replacement. A laser redundancy technique, pioneered by AT&T, then disconnects the offenders and automatically replaces the entire memory row or column from the spares.

Actual size

Cleanliness Is Next To Goodliness

A speck of dust on a megabit chip is like a boulder on a railroad track.

To keep our chip clean, it has to be made in a room that's C-L-E-A-N. And an AT&T class-10 clean room is one of the cleanest rooms in the world—where the air contains fewer than 10 particles of dust in a cubic foot of air. And the largest particle must be smaller than 1/150 the diameter of a human hair. That makes our room 10,000 times cleaner than a hospital operating room.

To keep our air clean, we have to keep it moving. Ours is constantly circulated and filtered by fans that could inflate the Goodyear blimp in only 30 seconds.

You Ain't Seen Nothing Yet

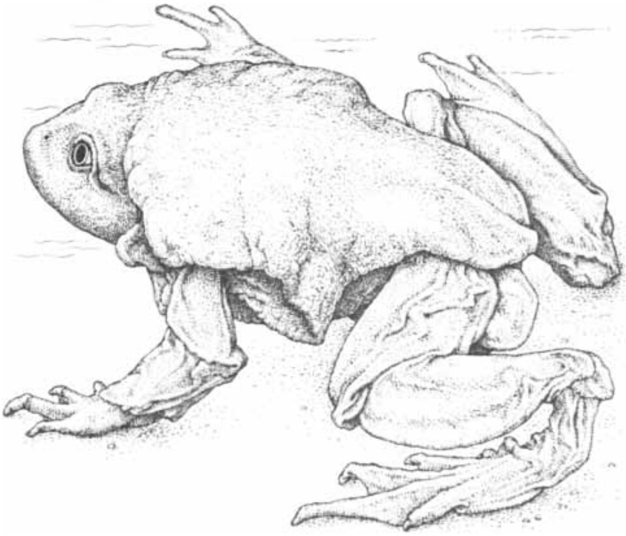
Looking toward the year 2000, today's technology predicts submicron design widths, with lines only 400 atoms wide. Chips containing 100 million components—50 times the current amount—could have tiny regions with capabilities that more than match the megabit chip.

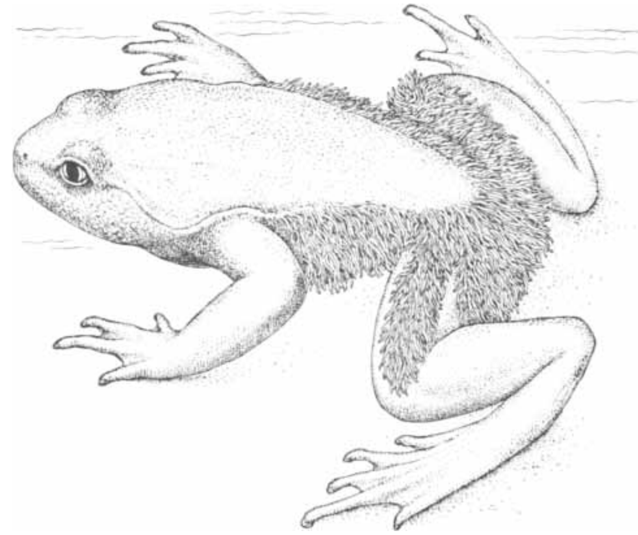
Now we're talking computer POWER, because these new chips with micro-parts could work at picospeeds (trillionths of a second). Current estimates suggest 10-picosecond transistors.

AT&T will not let the chips fall where they may. We'll put them where they'll do the most good—in the powerful and reliable digital systems creating the general-purpose information technology we call Universal Information Services. Our continuing flow of advanced technology is one reason why AT&T is the right choice.

AT&T

The right choice.





INCREASED SKIN SURFACE AREA is a prominent morphological adaptation in two skin-breathing frog species. Pendulous folds of skin have evolved on the Lake Titicaca frog *Telmatobius culeus* (*left*). These loose folds provide so much skin area for gas exchange that the Titicaca frog does not need to breathe through its lungs at all. The male hairy frog, *Astylosternus robustus* (*right*), sprouts dermal papillae on its hindquarters as an accessory gas-exchange organ to meet the respiratory demands of the mating season.

containing water with approximately half the oxygen concentration of the aerated water. After four weeks the diffusion barrier in larvae reared in the well-aerated water was 40 micrometers, a typical value. In contrast, the diffusion barrier averaged only 20 micrometers in animals reared in the oxygen-poor water. Furthermore, the cutaneous capillary network of the oxygen-deprived larvae was finer and denser. It appears these larvae underwent an acclimatory change that augmented their capacity for cutaneous gas exchange.

Even vertebrates with a relatively thick skin can carry on significant cutaneous gas exchange as long as evolution has led to advantageous placement of the cutaneous capillaries. In the scaly skin of some lizards, for example, the cutaneous capillaries either underlie the skin between the scales or run under the scale's hinges, where the scale is thinnest. In some snakes the cutaneous capillaries penetrate the scale itself. The dermal scales of fishes are generally covered by a layer of living tissue, an arrangement that places cutaneous capillaries above the diffusion-resistant scale and very close to the respiratory medium.

Adaptations such as skin folds, dermal papillae and generally thin skin, as well as the specialized circulation that may serve these structures, obviously develop slowly over days or weeks, if not time measured on an evolutionary scale. How can a vertebrate regulate cutaneous gas exchange instantaneously if its oxygen requirement suddenly increases (as it does during activity), or if it suddenly encounters a region of water that has a high carbon dioxide concentration?

In most of the vertebrates we have examined not all skin capillaries are constantly perfused with blood. Skin that is distant from underlying capillaries or that is underlain by nonperfused capillaries does not contribute to overall gas exchange; the functional surface area of the skin at any given time consists only of those skin regions that overlie perfused cutaneous capillaries.

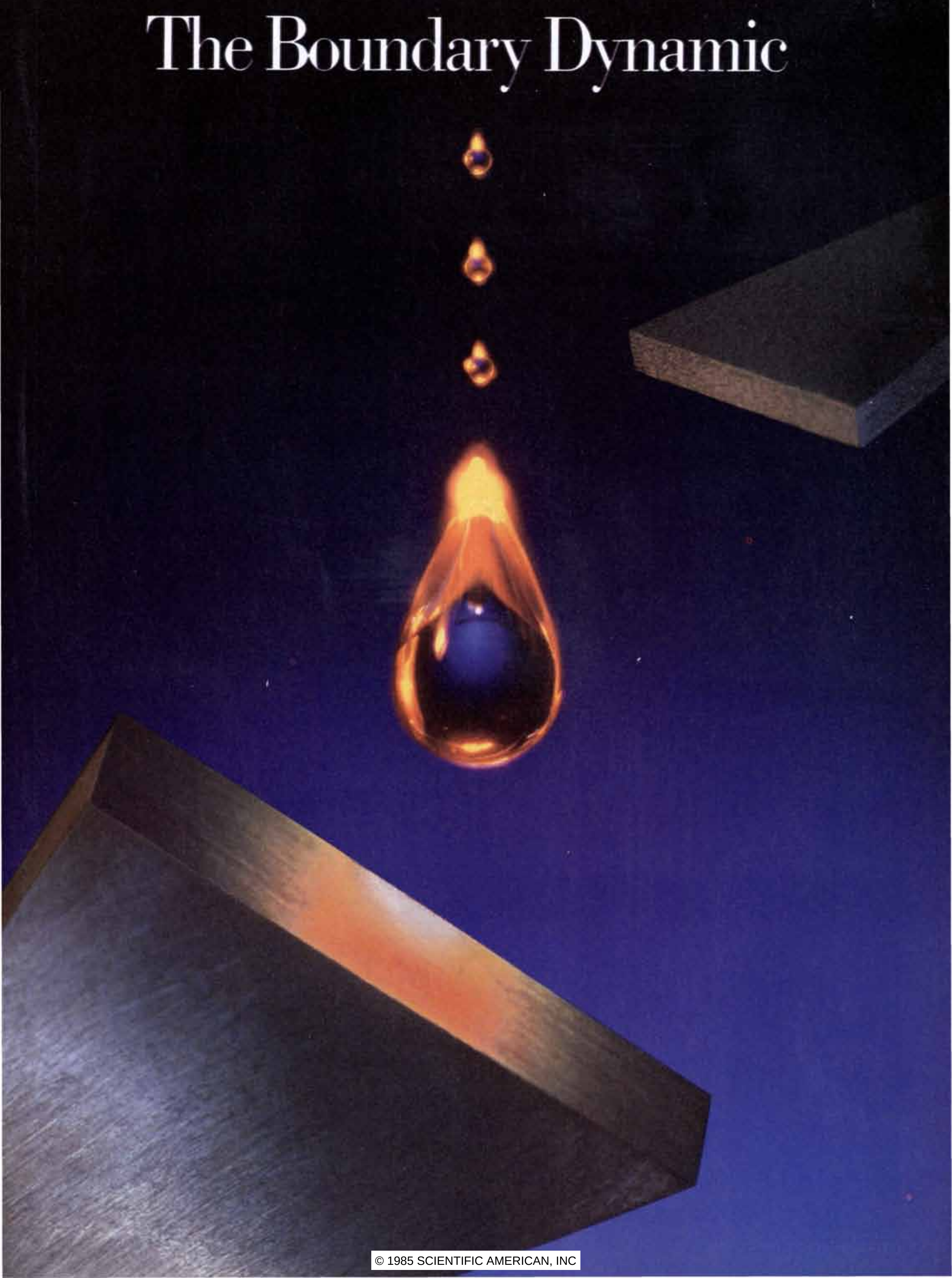
The initiation of blood flow through capillaries (what is called capillary recruitment) can occur in seconds. (Blushing is a classic, if nonrespiratory, example of this in humans.) Capillary recruitment in the skin of amphibians was documented many years ago by Piotr Poczopko and his colleagues in Poland. They found that in frogs breathing gas with an elevated carbon dioxide concentration, and in frogs prevented from using their lungs, the number of perfused skin capillaries increased by nearly a third.

Further experimental confirmation that capillary recruitment is linked to cutaneous gas exchange has emerged recently from our own studies of bullfrogs. Working at the University of Massachusetts at Amherst, we observed that when frogs submerged in water were exposed to air, the number of perfused skin capillaries rapidly fell by 60 percent; simultaneously carbon dioxide elimination by way of the skin fell by 44 percent. When the frogs were returned to water, capillaries were recruited and cutaneous carbon dioxide elimination resumed immediately.

More recent experiments conducted by Gary Malvin and Michael P. Hlastala of the University of Washington School of Medicine have also demonstrated that frogs control capillary blood flow in their skin. These investigations revealed that frogs can reduce gas loss through the skin by from 15 to 30 percent when they are exposed to an atmosphere devoid of oxygen, presumably by decreasing the amount of blood-perfused skin.

Just as important as morphological adaptations and capillary blood flow are physiological processes or behavioral responses that affect the partial pressures of the respiratory gases in the blood within the skin or in the respiratory medium outside the skin. The most important of these are processes that regulate the flow of either the respiratory medium or the blood along the diffusion barrier.

Ventilation, the flow of the respiratory medium into or past a gas-exchange organ, is clearly critical in lungs and internal gills. If ventilation stops, gas exchange in these internal organs declines precipitously because oxygen is quickly depleted (its partial pressure falls) and carbon dioxide is rapidly accumulated (its partial pressure increases) in the respiratory medium enclosed within the body. Ventilation in the context of cutaneous gas exchange, on the other hand, has often been considered unnecessary; after all, the skin of a vertebrate in air or in a large body of water is in constant


The Boundary Dynamic

The Boundary Dynamic

The performance of a polymeric adhesive depends on the properties and composition of its surface. Now a scientist at the General Motors Research Laboratories has developed and validated a theory that describes the coupled effects of diffusion and chemical reaction on the changing surfaces not only of adhesives, but of chemically reacting surfactant systems in general.

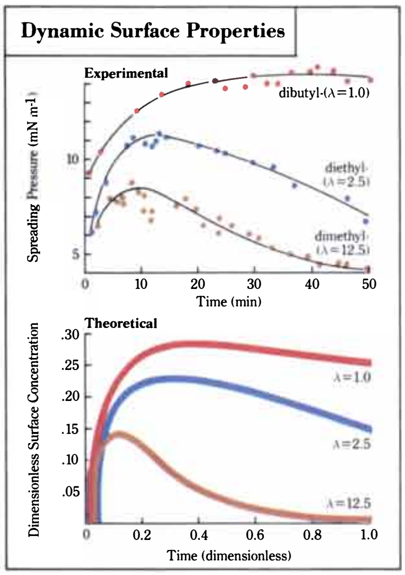


Figure 1: Experimental measurements of spreading pressure v. time for dialkylaminopropylamines with various Damköhler numbers (λ), and corresponding theoretical calculations of surface concentrations.

THE USE OF adhesives in the production of an automobile promises to make both the product and the process more efficient. Both weight and operations can be reduced. In practice, however, steel and other metallic surfaces are often contaminated by process lubricants. A durable bond depends on the ability of an adhesive to displace contaminants and to wet the substrate.

Assuring intimate contact between adhesive and substrate requires detailed knowledge of adhesive surface tension, since it is this property that controls displacement of contaminants and wetting. Up to now the surface tension of an adhesive has typically been assumed constant. In reality, though, surface-active components in the adhesive collect preferentially at the interface and also react, so that the surface composition varies with time, giving rise to dynamic surface tension. Variations can be large enough to significantly affect adhesive performance.

The understanding of time-dependent surface tension has been advanced by the work of Dr. Robert Foister, a scientist at the General Motors Research Laboratories. Investigation of dynamic surface properties of thermosetting adhesives led him to develop a general theory of adsorption kinetics in binary, chemically reacting surfactant systems. The significance of this theory is that it includes the coupled effects of surfactant diffusion and chemical reaction, making it possible for the first time to describe quantitatively the changing surfaces of such systems.

In a typical adhesive that polymerizes, or "cures," by chemical reaction (Figure 2), a surface-active curing agent (Solute 1) reacts with the host resin to form a second surface-active species (Solute 2) that is also reactive. Both solutes migrate to the surface, lowering the surface tension. Diffusion to the surface is driven by a potential energy gradient between the surface and the bulk, with the solute molecules experiencing a lower energy at the surface.

Dr. Foister derived appropriate transport equations to describe diffusion and chemical reaction in the bulk, in a subsurface region, and at the surface itself. The transport equations can be solved analytically if the chemical rate equations are assumed to be first order in the concentrations of reacting species, and if the subsurface and surface concentrations can be related to one another by a linear adsorption isotherm. For more complicated isotherms, a set of coupled, non-linear integral equations is generated.

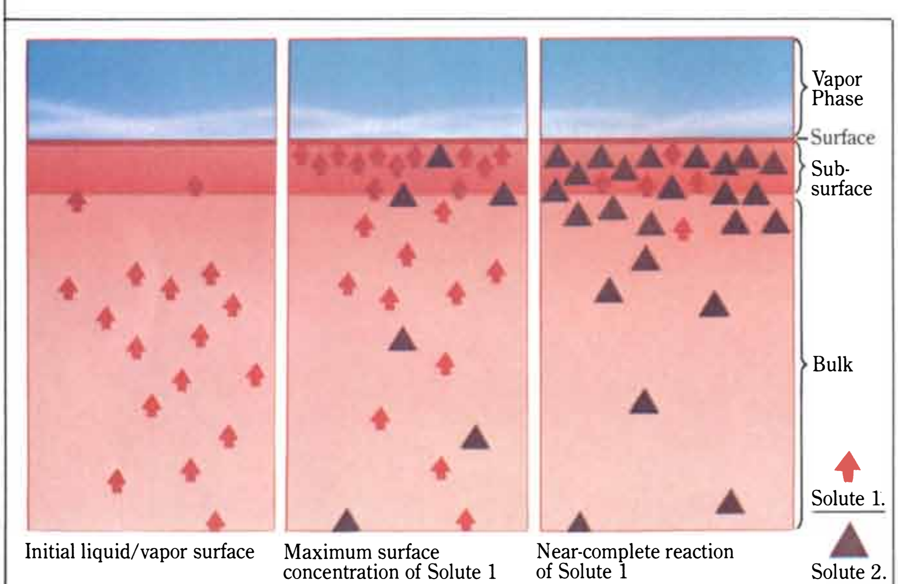


Figure 2: Evolution of an adhesive surface: Surface-active Solute 1 (red) reacts with host resin (pink tone) to form surface active Solute 2 (brown).

These must be solved numerically.

Analytical solution for the special case of the linear isotherm indicated that the change with time in surface concentration (and consequently in surface tension) is composed of two terms: first the diffusive flux of Solute 1 into the subsurface from the bulk, and second the depletion of this solute due to chemical reaction. Hence, the surface concentration of Solute 1 exhibits a maximum with time (Figure 2). This maximum in surface concentration corresponds to a minimum in surface tension.

MODIFYING the transport equations to include binary adsorption isotherms allowed for consideration of competitive adsorption of the two reacting and diffusing solutes. By solving these equations numerically and conducting dimensional analysis, Dr. Foister identified various dimensionless parameters as predictors of system behavior. The most important of these parameters was a dimensionless number (λ), of the Damköhler type, involving terms representative of reaction, diffusion, and adsorption.

$$\lambda = \frac{k\,(\Gamma_m\, a)^2}{4D}$$

Here k is the reaction rate constant of Solute 1, D its diffusivity, Γ_m its "surface capacity" (the maximum number of molecules absorbed per unit surface area), and a its "surface affinity" (a measure of its energy of adsorption). For an adhesive, lowering λ by reducing k (the reactivity of the curing agent), for example, would prolong the time to maximum, and would increase the value of the surface concentration at the maximum (see Figure 1, Theoretical). As a practical consequence, this would improve wetting by minimizing the surface tension.

In experiments using a series of dialkylaminopropylamine curing agents (dimethyl-, diethyl-, and dibutyl-) in a host epoxy resin matrix, good agreement has been demonstrated between theoretical predictions for surface concentration and the measured dynamic spreading pressure, which is the change in adhesive system surface tension due to the curing agent (Figure 1, Experimental).

"I expect," says Dr. Foister, "that the physical insights gained from this analysis can be applied to other reactive surfactant systems by using specifically tailored isotherms and chemical reaction schemes. Predicting surface behavior can certainly help us design better adhesives for specific applications, but it is also pertinent to the performance of anti-oxidants and anti-ozonants in synthetic rubber, for example. And applied to interfaces in biological systems, a suitably modified theory may prove valuable in understanding the phenomenon of enzyme activity."

General Motors

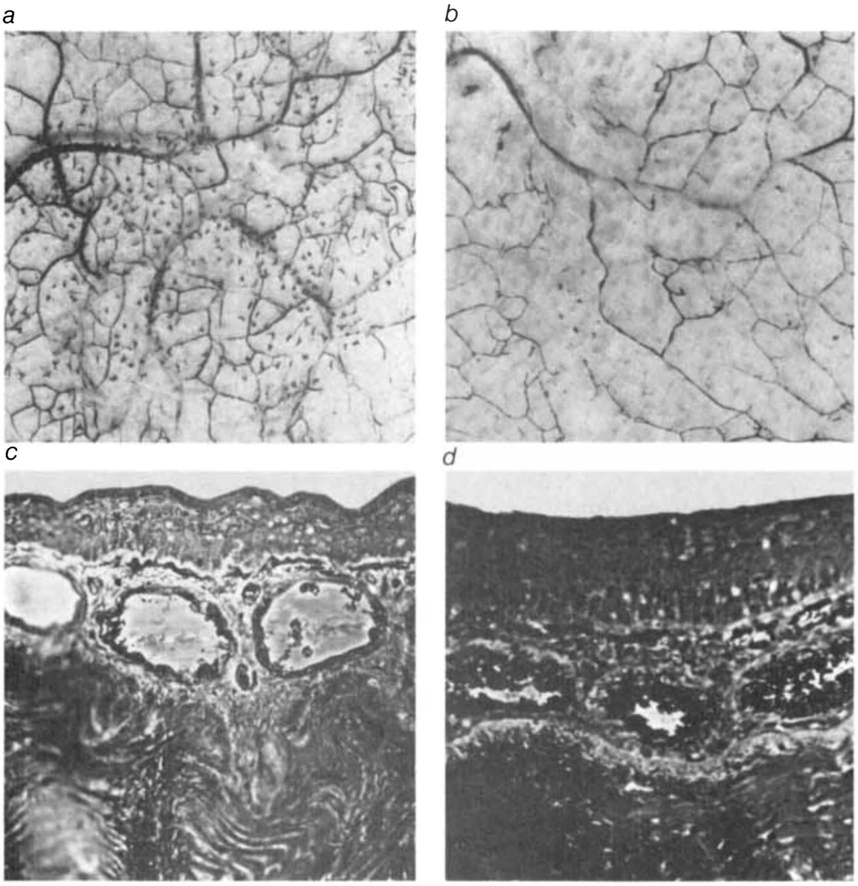
THE MAN BEHIND THE WORK

Dr. Foister is a Staff Research Scientist in the Polymers Department at the General Motors Research Laboratories.

Dr. Foister received his undergraduate degree from Guilford College, and holds a Ph.D. in Physical Chemistry from the University of North Carolina at Chapel Hill. His thesis dealt with the role of liquid inertia in the intrinsic viscosities of rod-like polymers.

He did post-doctoral work in Canada as a Fellow at McGill University in Montreal, and in the Applied Chemistry Division of the Pulp and Paper Research Institute of Canada, working on the microrheology of colloidal dispersions.

Dr. Foister joined General Motors in 1980. He is the leader of the Structural Adhesives Group in the GMR Polymers Department. His current research interests center on surface chemistry and adhesion.



MICROSCOPIC MORPHOLOGICAL RESPONSES that facilitate skin breathing under adverse conditions are shown in two pairs of photomicrographs. The capillary network in the skin of frog larvae becomes finer and denser when the larvae live in oxygen-poor water (*a*) than it does when they live in oxygen-rich water (*b*). The distance between capillaries and the surface of the skin in frog larvae can also vary depending on the oxygen content of the water: the capillaries of larvae reared in oxygen-poor water are closer to the skin surface (*c*) than the capillaries of larvae raised in water having a high oxygen concentration (*d*).

contact with an "infinite pool" of respiratory medium containing abundant stores of oxygen and only small amounts of carbon dioxide.

Yet the Lake Titicaca frog waves its large skin folds and the hellbender rocks its body. Moreover, both animals increase the frequency of these movements when the concentration of oxygen in the water decreases. Are these behaviors unrelated to respiration or might they serve to augment cutaneous gas exchange by ventilating the skin?

As we examined these peculiar behaviors we reflected on the physical considerations that influence the exchange of heat. Heat exchange is often severely limited in still air or water and is facilitated when the medium (or the heat emitter) is moving. If, for example, one sinks into a tub of hot water and remains still, one feels pain as heat enters the skin. As the heat leaves the layer of water surrounding the skin, the fluid forms a relatively cool boundary layer. Any rapid movement of the skin or the water will dissipate the boundary layer, allowing hot water again to come in contact with the skin and again to cause pain until the boundary layer is reestablished. Even in an infinite pool (or finite tub) of hot water, stagnation of the medium next to the skin can limit heat transfer.

We envision a similar relation for the exchange of respiratory gases, particularly in water. If both the animal and the water are stationary, oxygen diffusing out of water adjacent to the skin and into the bloodstream would create a diffusion boundary layer of low oxygen partial pressure. Because the rate of diffusion is proportional to the difference in oxygen partial pressures on each side of the diffusion barrier, cutaneous gas exchange would decrease. Movement of either the animal, the water or both could dissipate the boundary layer and thereby increase the diffusion of oxygen across the skin. We calculated that the diffusion boundary layer should offer a significant resistance to gas exchange at water flow velocities of four centimeters per second or less. Much lower velocities would suffice to dissipate the diffusion boundary layer in air.

To test our hypothesis we immobilized bullfrogs in wire-mesh envelopes to ensure that spontaneous body movements would not dissipate the boundary layer and that their cutaneous surface area would be constant and maximal. We then placed each animal, sandwiched in its envelope, in a leakproof chamber. The chamber was filled with water so that only the frog's nostrils were above the surface. We could measure the decline in oxygen concentration in the air and the water compartments and so calculate the respective pulmonary and cutaneous oxygen consumption. By actuating a stirrer at the bottom of the chamber we could ventilate the skin with the surrounding water.

The results support our hypothesis. When we stopped ventilating the skin, the cutaneous oxygen consumption (determined by measuring the oxygen concentration in the water) declined by about a third. This result is clearly in conflict with the notion that ventilation is unimportant in cutaneous gas exchange.

As we have stressed, the extent of cutaneous gas exchange is determined by many variables other than ventilation, including the skin's functional surface area and the oxygen partial pressure on the inside of the diffusion barrier. Might the decrease in cutaneous oxygen consumption associated with the stopping of ventilation in fact be due to one of these other factors?

Our previous experiments had impressed on us the importance of functional surface area and capillary recruitment. We therefore repeated the ventilation experiments with frogs in which we could observe the relative numbers of perfused and nonperfused capillaries. The amount of oxygen taken in through the skin could have been lessened by a reduction in the number of perfused capillaries in the frog's skin. We found the opposite to be the case: whenever stirring was stopped, the frogs recruited additional skin capillaries. Because capillary recruitment should increase cutaneous gas exchange, the decline in cutaneous oxygen consumption, observed whenever ventilation stopped, cannot be due to changes in functional surface area.

In a third experiment we measured the partial pressure of oxygen in blood carried by arteries leading to the skin. If the partial pressure of oxygen increased every time stirring stopped, this could explain the observed drop in the rate of oxygen diffusing from the water into the blood. Stirring had no effect on the blood's oxygen partial pressure, however. The results suggest that the formation of a diffusion boundary layer was indeed responsible for the decrease in cutaneous oxygen uptake, and that skin ventilation could serve to regulate gas exchange by dissipating the layer.

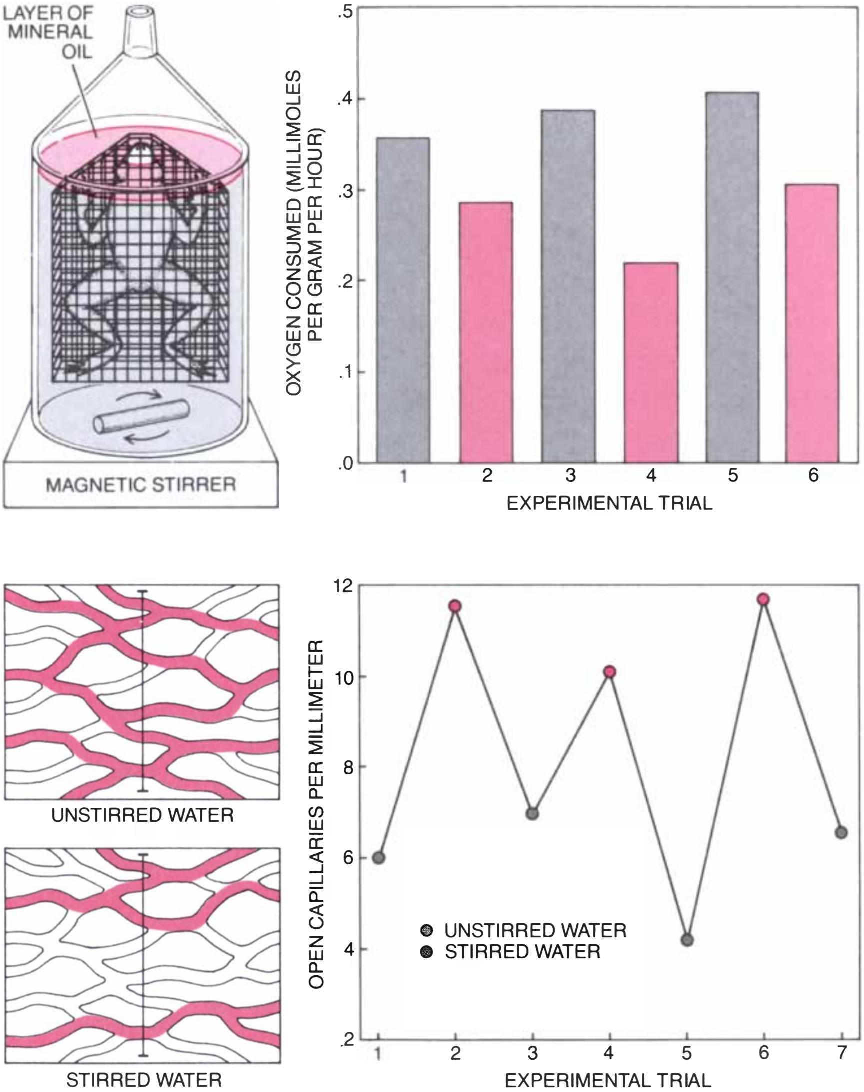
The partial pressures of oxygen and carbon dioxide in blood within cutaneous capillaries vary according to whether or not the blood is oxygenated. This provides another possible regulatory device. By controlling whether the blood flowing to the skin is primarily oxygenated or deoxygenated, an organism could presumably regulate the amount of oxygen and carbon dioxide diffusing through the skin.

Cutaneous gas exchange would be most effective if only deoxygenated blood flowed to the skin, in the same way that only deoxygenated blood flows to the lungs in mammals. The partial pressure differentials across the skin of both oxygen and carbon dioxide would be largest in this circumstance. The skin of most vertebrates, however, is no different from any other living tissue: it receives blood only from the major systemic arteries, which typically deliver oxygenated blood. Because the difference between the oxygen partial pressure of oxygenated blood and that of the respiratory medium is often rather small, oxygen uptake from the environment is normally limited by the very blood the skin cells need to live. Even under this seemingly major constraint the skin may nonetheless be important in carbon dioxide elimination because carbon dioxide levels in arterial blood may still be considerably greater than they are in the environment. This explains in part why carbon dioxide elimination typically exceeds oxygen consumption in cutaneous gas exchange among vertebrates.

Some vertebrates are indeed able to direct a fraction of their deoxygenated blood into the systemic arteries and thereby augment cutaneous gas exchange. Amphibians and reptiles, for example, have an incompletely divided heart that allows deoxygenated blood to flow to the skin without first traversing the lungs. Comparative anatomists have traditionally regarded this arrangement as a primitive and inefficient one. Kjell Johansen of the University of Aarhus in Denmark, Fred White of the Scripps Institution of Oceanography and others have suggested that the opposite is true. They argue that the heart structure of amphibians and reptiles is actually an important adaptation that allows these vertebrates to distribute blood where it would best promote gas exchange.

Amphibians, moreover, are unique in having cutaneous arteries, which direct transfer of deoxygenated blood to the skin. Blood can leave the single ventricle of an amphibian heart by either of two routes. One way is through the systemic arteries, which carry oxygenated blood directly to the brain, muscles, viscera and ultimately to the skin. The second route is through the pulmocutaneous arteries, which supply deoxygenated blood to the lungs by way of the pulmonary arteries and to the skin by way of the cutaneous arteries.

Graham Shelton and his colleagues at the University of East Anglia in England have shown that amphibians can channel deoxygenated blood into the pulmocutaneous arteries and thence selectively to either the lungs or the skin for gas exchange. The basis for such ability may lie in the structure of the heart and in the muscular sphincters that surround the pulmonary and cutaneous arteries after they diverge from the common pulmocutaneous



SKIN VENTILATION affects cutaneous gas exchange. A frog, immobilized in a wire-mesh envelope, was immersed in a water-filled chamber so that only its nostrils were above the surface (*top left*). The skin could be ventilated by actuating a magnetic stirrer at the bottom of the chamber. Cutaneous oxygen uptake could be monitored by measuring the decrease in concentration of oxygen in the chamber. (A surface layer of mineral oil prevented aeration of the water.) Cutaneous oxygen exchange (*top right*) was greater when the water was stirred (*gray bars*) than when the water was left still (*colored bars*). The reduction in oxygen consumption when the water was still was shown not to result from a decrease in capillary blood flow. A frog's leg was led through the side of the container so that the capillaries in the web of the foot could be seen under a microscope. The number of blood-filled capillaries that intersected a line in the microscope's eyepiece was counted while the water was stirred and while it was left still (*bottom left*). In the absence of stirring, and thus of skin ventilation, the number of blood-perfused capillaries actually increased (*bottom right*).

Where engine

The more you rely on sophisticated graphics, the more you need a Smart Desk equipped with an IBM 3270 Personal Computer/GX.

The 3270 PC/GX gives you easy access to host-based graphics and data. And, with IBM's new Graphics Editor and Picture Plotting software, you can work offline locally to draw your own charts, diagrams and free-form sketches—all with excellent resolution.

You can actually create or edit directly on the screen, using either a mouse or a tablet. What's more, you can select from a wide range of colors, lines and type styles to arrive at a design or presentation that works best for you.

If you want to make foils for a meeting or reproduce your design sketches, just zoom and scroll to position the graphics you want to reproduce. Then use IBM's Picture Plotting to plot the results. You also get a handy graphics index, so you can file and retrieve pictures at your convenience.

But there's more to the 3270 PC/GX than just interesting pictures. It can display up to four host graphics sessions, one PC session and two notepads concurrently. And the keyboard has a variety of ergonomic features including special function keys and a numeric keypad.

There's also a smaller model of the 3270 PC with advanced color graphics—the 3270 PC/G. Whether you choose the G or the GX, you're sure to find it a graphic example of the state of the art.

IBM 11-85
DRM, Dept. KW/428
400 Parson's Pond Drive
Franklin Lakes, NJ 07417

☐ Please send me literature on the IBM 3270 PC/G and GX.
☐ Please have an IBM marketing representative contact me.

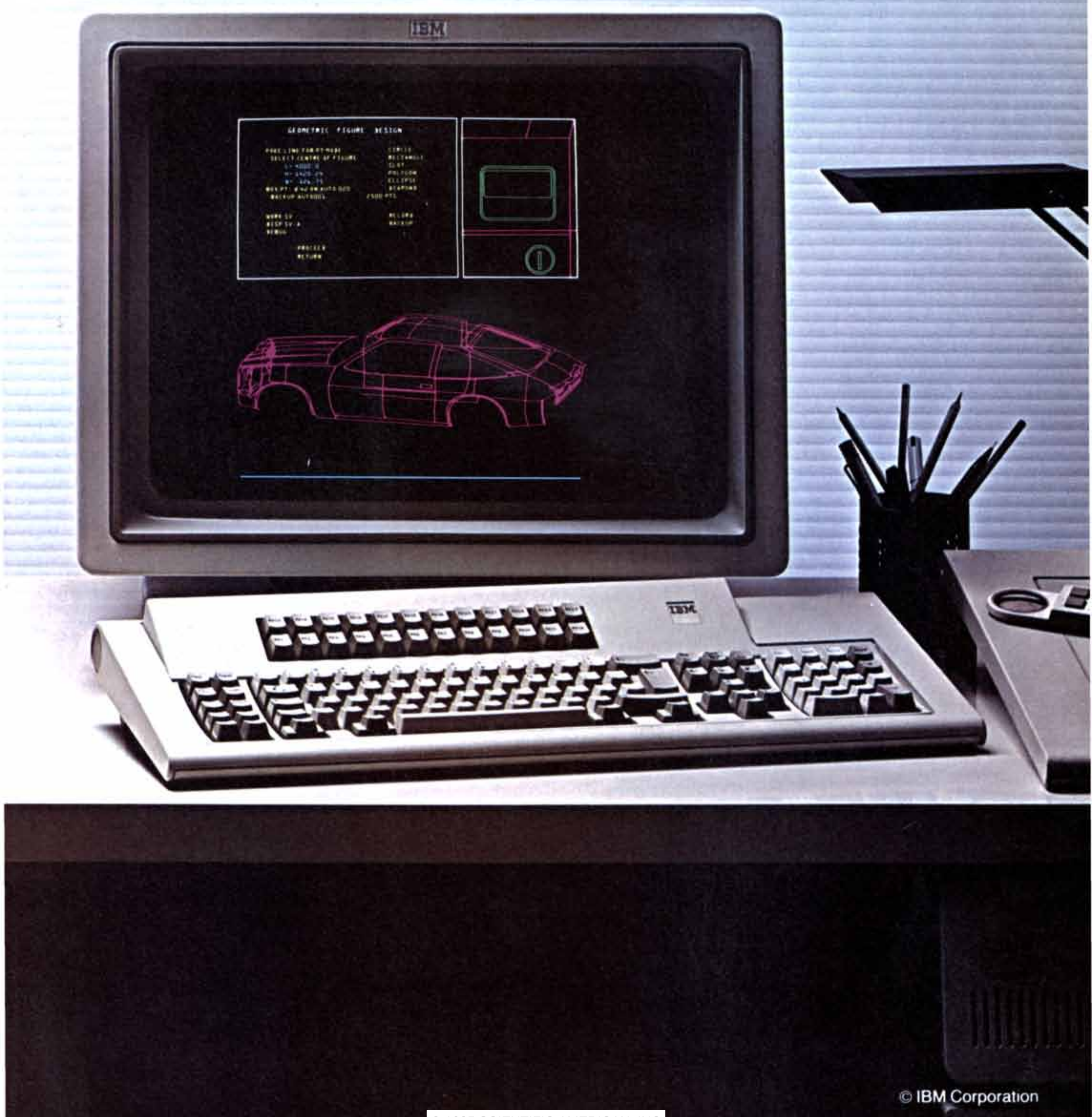
Name__________
Title__________
Company__________
Address__________
City_____ State_____ Zip_____
Phone__________

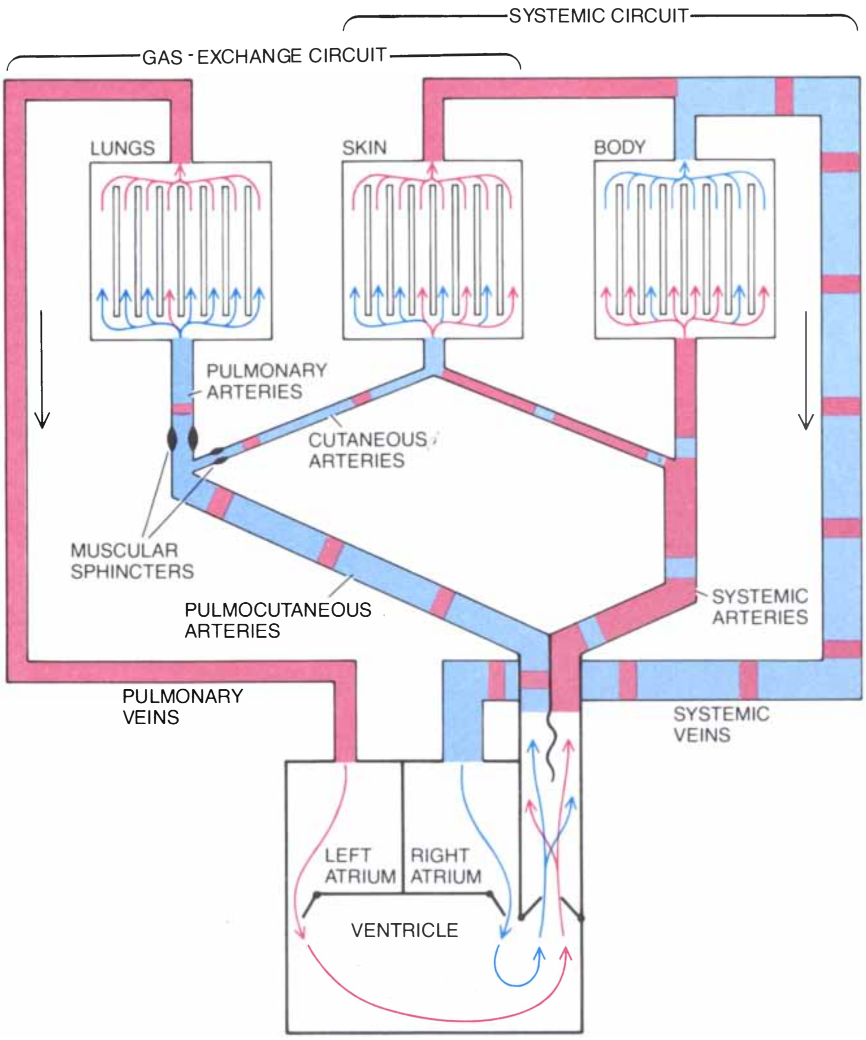
To receive literature, call 1 800 IBM-2468, Ext. 428, contact your IBM marketing representative or send in this coupon.

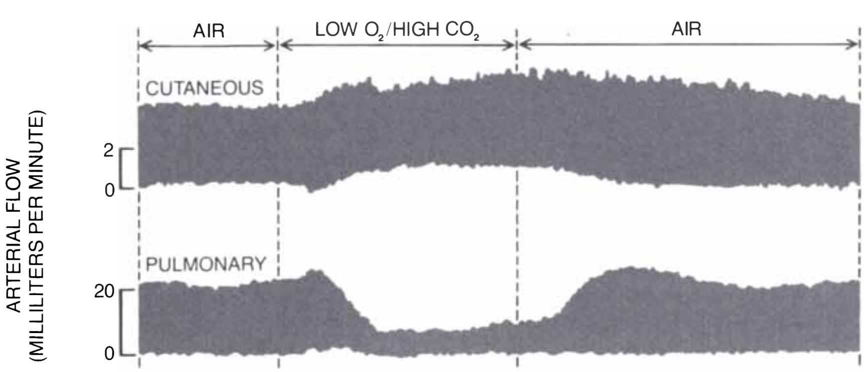
IBM

The Smart Desk from IBM

ers practice the art of decision making.







AMPHIBIAN CIRCULATORY SYSTEM, diagrammed here, allows deoxygenated blood (*blue*) to reach the skin and thereby promotes the diffusion of oxygen through the skin and into the blood. Most of the deoxygenated blood delivered to the skin comes from the heart by way of the cutaneous arteries (*top*); a small quantity is mixed into oxygenated blood (*red*) in the ventricle of the heart and is transported by the systemic arteries to the skin. (Some oxygenated blood also finds its way into the pulmocutaneous arteries.) In spite of the fact that a single ventricle pumps both oxygenated and deoxygenated blood into the arteries, a segregation of blood is maintained: primarily oxygenated blood is routed to the systemic circuit and primarily deoxygenated blood is routed to the gas-exchange circuit. Part of the primarily deoxygenated blood in the pulmocutaneous arteries can then be shunted to either the lungs or the skin, depending on which organ is better suited for respiration at the time. Electromagnetic cuffs measuring the blood flow within the cutaneous and pulmonary arteries recorded (*bottom*) the decrease in pulmonary arterial flow and the increase in cutaneous arterial flow when the lungs were flushed with gas rich in carbon dioxide and poor in oxygen. Muscular sphincters around the arteries may act as valves and so regulate the flow.

artery. The sphincters may act as valves that shunt blood flow one way or the other. In any case the exact distribution depends on the level of oxygen and carbon dioxide prevailing at these two respiratory organs.

Working with Nigel West of the University of Saskatchewan, we measured the distribution of pulmocutaneous blood between the skin and the lungs by implanting electromagnetic flow transducers around the pulmonary and cutaneous arteries of anesthetized toads. When we simulated the oxygen and carbon dioxide partial pressures typically found in the lungs of toads holding their breath, the toads diverted blood from the lungs to the skin. This response facilitates cutaneous gas exchange.

Such a reaction must also be relied on by frogs whenever pulmonary breathing is hindered, as it is when a frog is underwater. Recently Robert Boutilier, Mogens Glass and Norbert Heisler of the Max Planck Institute in Göttingen conducted a series of experiments similar to ours on intact, unanesthetized bullfrogs. They injected microscopic radioactive spheres into the bullfrogs' circulatory systems. The spheres, which were slightly larger than red blood cells, lodged in capillaries through which blood flowed. The distribution of blood flow between the lungs, skin and other body tissues could then be calculated from the radioactive emissions of the various body tissues exposed by dissection.

When frogs dived in oxygenated water after breathing gas mixtures low in oxygen, pulmocutaneous blood was preferentially distributed to the skin rather than to the lungs. Conversely, when frogs with air-filled lungs were submerged in water containing little oxygen, pulmocutaneous blood was distributed from the skin to the lungs. Clearly amphibians can regulate cutaneous blood flow both to optimize cutaneous gas exchange and to coordinate the respiratory activity of the skin with that of the lungs.

Although skin breathing may account for only a small component of total gas exchange in certain animals, in others it can play a major if not vital role. The sheer diversity of species that resort at least in part to cutaneous gas exchange should be sufficient to convince one that skin breathing is commonplace rather than exceptional in vertebrates. Through closer investigation cutaneous gas exchange has emerged as a well-regulated, energetically inexpensive process that can respond to immediate, prolonged and evolutionary changes in an animal's respiratory requirements.


The butterfly and the 747.
Near Los Angeles International Airport, where 747's reach for the sky, one of the smallest endangered species in the world reaches for its dinner.
The El Segundo Blue lives on wild buckwheat . . . on land that's part of an oil refinery. The people who work there protect the area and plant buckwheat year in, year out.
Now, for a creature little bigger than your fingernail, some of the danger of being endangered is past.
Do people really care that much for a tenth of a gram of beauty?
People Do.
Chevron
For more information write: People Do-B, P. O. Box 7753, San Francisco, Calif. 94120

Human-powered Flight

Human-powered aircraft exploit a little-understood flight regime. Aircraft that can negotiate it are fun to fly and may turn out to have uses in reconnaissance and planetary science

by Mark Drela and John S. Langford

During all the centuries in which people dreamed of human flight it was supposed the flier would provide the power, as a bird does. Yet only in the past 25 years—after the development of the propeller-driven airplane and the jet engine as well as the achievement of supersonic flight and space flight—has the human-powered aircraft come into its own. Its arrival is due to the development of a combination of crucial technologies: aerodynamic, propulsive and structural. Equally important was a somewhat earlier achievement: making the craft adequately controllable by a pilot for whom the task of generating a large amount of mechanical power is distracting. The craft have now reached a stage where some applications for the technology can be envisioned.

Human-powered aircraft probably would not have reached this stage without the stimulus supplied by a series of competitions sponsored by a few organizations and individuals. The first one, which took place in France between 1912 and 1922, was a project of the Peugeot company. It resulted in aircraft that were really only jumping bicycles: the operator pedaled hard to get up speed on the ground and then the winged craft would glide through the air for about 12 meters. Once the craft was airborne it had no means of propulsion.

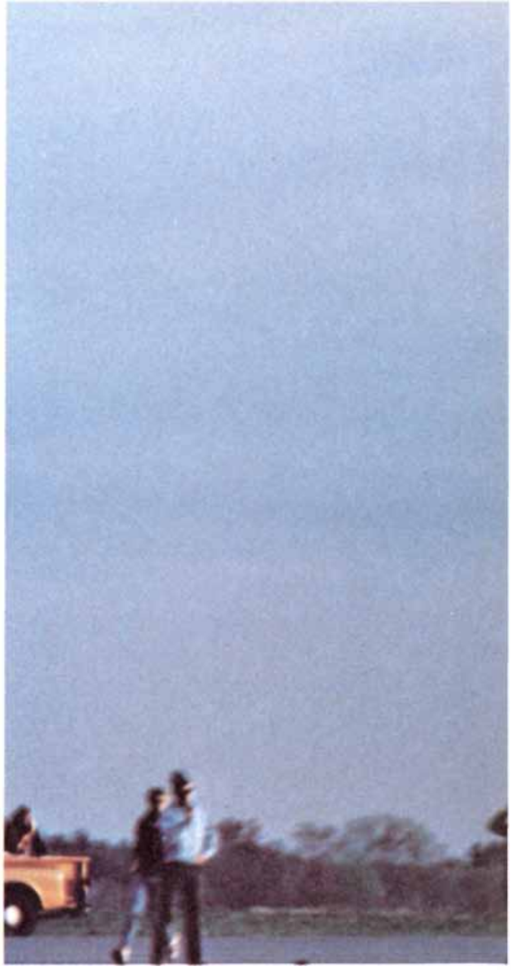
In 1935 the German aircraft *Mufli* went a step further: the pilot was able to drive a propeller after a catapult takeoff. Apparently the power requirement for level flight was more than the designers could achieve. The pilot could produce only enough power for an extended glide, the longest of which was 712 meters. *Mufli* competed for a prize of 5,000 marks offered by a group in Frankfurt for the first human-powered flight around two pylons 500 meters apart. Similar prizes were also offered in Italy and the U.S.S.R.; all went unclaimed.

The most famous competitions, and the ones that have led to genuine technological progress, have been sponsored by Henry Kremer, a British industrialist. In 1959 he offered a prize of £5,000 to the first entrant who could fly an aircraft around a one-mile, figure-eight course under human power alone. Eighteen years had passed and the prize money had increased tenfold when Bryan Allen of the U.S. successfully flew *Gossamer Condor* around such a course.

Kremer subsequently offered the largest prize in the history of aviation: £100,000 for the first human-powered flight across the English Channel. Again the winner was Allen, who pedaled *Gossamer Albatross* across the 21-mile strait between Folkestone and Cape Griz-Nez on June 12, 1979.

Both *Condor* and *Albatross* were large, fragile craft that became unmanageable in all but the gentlest breezes. Their success did not lead to widespread activity in human-powered flight. Four years after the Channel crossing Kremer responded by sponsoring another competition intended to make human-powered aircraft faster and thereby smaller and more practical. This time the goal was to achieve a relatively high speed around a triangular course of 1,500 meters. The prize was £20,000 to the first contestant who could complete the course in less than three minutes—a pace that implied a speed of about 32 kilometers (20 miles) per hour. Frank P. Scarabino of the U.S. won this prize in May, 1984, flying *Monarch,* a craft designed and built at the Massachusetts Institute of Technology. Prizes offered by the Royal Aeronautical Society are now available for flights that better the existing record by at least 5 percent; three such prizes have already been awarded.

Discounting stunts and turn-of-the-century winged bicycles, some 60 human-powered aircraft have been built. Most of them were inspired by the Kremer competitions. The designs can be grouped roughly into three generations, according to characteristic sets of aerodynamic and structural concepts [*see illustration on page 147*]. The



HUMAN-POWERED AIRCRAFT completes a speed trial at Hanscom Field in Massachusetts. The craft is *Monarch B,* designed and built at the Massachusetts Institute of Technology. The pilot is Frank P. Scarabino,

craft of the first generation were conceptually based on sailplanes (motorless gliders). They could make only straight-line flights, few of which exceeded one kilometer.

The second generation includes the first vehicles capable of sustained and controllable human-powered flight. *Gossamer Condor* is the best-known of them. It was built in California by a team led by Paul MacCready, Jr., and now is on display in the Smithsonian Institution's National Air and Space Museum. A lesser-known but somewhat more rugged craft is *Chrysalis,* a biplane built by students at M.I.T. in 1979. Second-generation craft have unusual configurations because the designers went beyond conventional ideas of what an airplane should look like.

Third-generation aircraft have been built for the speed competition and so are much smaller. Externally they look something like the first-generation craft, but they incorporate modern structural and aerodynamic technology and reflect the experience accumulated through the design and operation of the second-generation machines. Because of a provision in the rules of the speed competition, some of these machines also have an energy-reserve capability. Such a system enables the pilot to store his own energy in the craft for a short time before a flight (usually by pedaling a generator to charge batteries) so that he can draw on it during the flight. The machines are also considerably more sophisticated than their predecessors. For example, the pilot of the M.I.T. *Monarch* can tune the propeller electronically, thereby modifying the requirement for the speed of pedaling or the output of stored energy.

Although the specific tasks have differed in each Kremer competition, the designers of all three generations have faced a common problem: how to reduce the power required by the aircraft to the amount available from a human being. The second major problem of human-powered flight has been stability and control.

The power available from a human differs widely according to the person's age, training and motivation. A well-conditioned athlete can produce up to one kilowatt for short periods of time or a few hundred watts for several hours. Surprisingly in view of the many studies made, little conclusive evidence is available on such basic factors as whether it is better for the pilot to be vertically seated or recumbent. In the absence of sound physiological data the design decision is usually made for reasons of aerodynamics, structure or weight distribution.

The power required by an aircraft is the product of its aerodynamic resistance (drag) and its velocity. Low power can therefore be obtained by building a craft with low drag; flying slowly is also a way of reducing the power requirement.

who in 1984 won with this craft a prize of £20,000 offered by Henry Kremer of the U.K. for completing a 1,500-meter triangular course in less than three minutes. The rules of the Kremer competition and succeeding ones sponsored by the Royal Aeronautical Society stipulate that the craft must be at an altitude of at least two meters at the start and finish; the orange streamer being held by two members of the ground crew is two meters high. The crewman at the right also holds a stopwatch, as does an official judge (appointed by the Royal Aeronautical Society) standing in the near background. In winning the Kremer prize *Monarch B* averaged 21.5 miles per hour.

To achieve equilibrium in flight the lift (vertical force) produced by an aircraft's wings must equal the gross weight of the vehicle. Wing area is the most useful variable for the designer. In theory arbitrarily low flight speed and power requirements can be obtained with sufficiently large wings. In practice the wing area is limited by considerations of structural rigidity, weight, sensitivity to wind and the size of the buildings available for storing planes on the ground.

Drag at subsonic speeds has two components of comparable magnitude. One component arises from friction with the air. It is roughly proportional to the exposed area of the plane. The second component is an unavoidable consequence of generating lift and is known as induced drag. Friction drag can be reduced by decreasing the exposed area and employing aerodynamically efficient shapes. Induced drag can be diminished primarily by increasing the wingspan and by flying close to the ground. Both theoretical and practical factors limit the amount of reduction in each case.

Sailplanes have long represented the epitome of low-drag design, and so it was only natural that the first generation of human-powered aircraft resembled them. Designers essentially sought to reduce the weight of a sailplane by a factor of 10 while adding a propeller and allowing no compromise of aerodynamic principles. All the framework and bracing was internal. One can now see that the task was beyond the capability of the available structural technology. The resulting aircraft were heavy and small compared with the machines of the second generation. The low drag and relatively high flight speed created a power requirement that left the pilot little margin for maneuvering the vehicle.

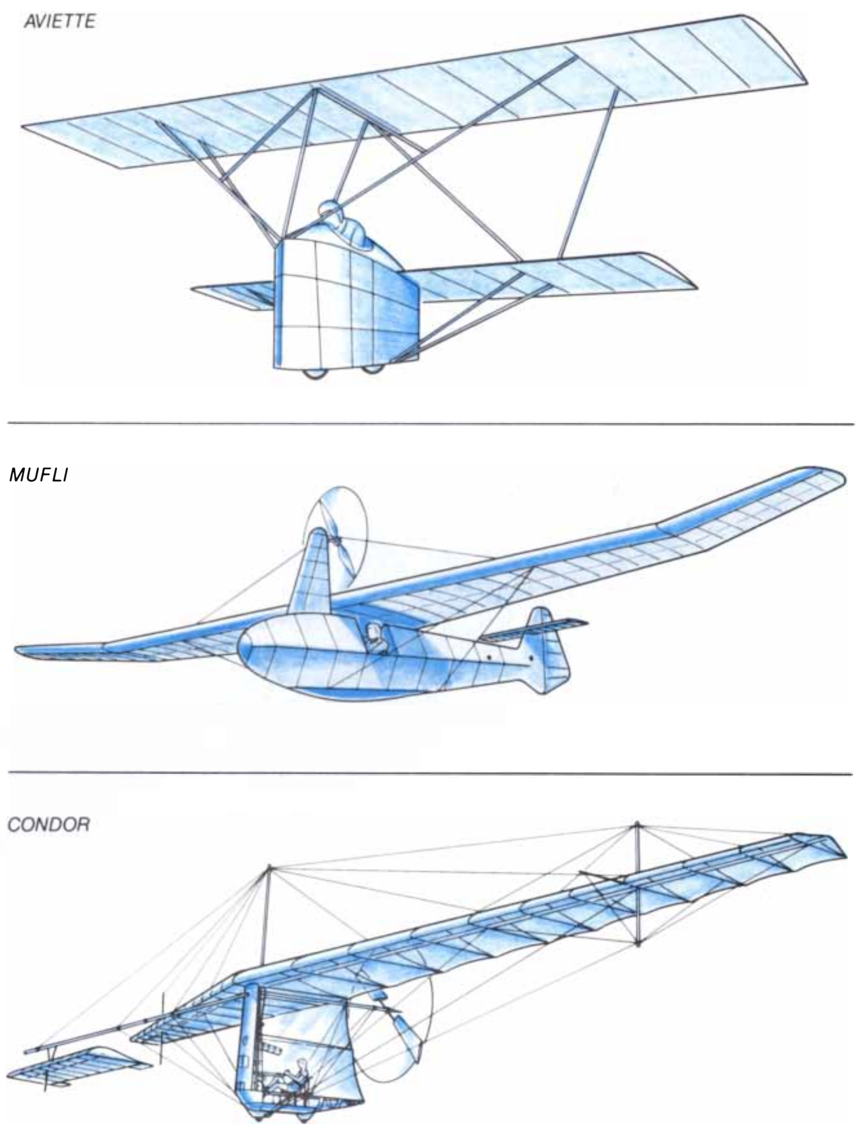
Second-generation craft embodied the low-speed approach to power reduction. The low-drag advantage of the sailplane was abandoned in favor of external bracing. The resulting increase in drag was offset by substantial increases in wing area and large reductions in weight. A low power requirement was achieved by the resulting low flight speed (approximately 16 kilometers per hour).

The machines of the second generation also incorporated the first workable solutions to the problem of stability and control. This achievement meant the designers had coped successfully with several effects that are normally not significant in conventional aircraft.

An example is acceleration. To accelerate an aircraft (when, say, it is making a banking turn) some of the surrounding air must also be accelerated. The craft is said to have an "apparent mass" in addition to its own mass. In conventional aircraft this additional component is negligible. In a human-powered aircraft it can be very important. As a result conventional control surfaces cannot generate the forces needed to deal adequately with the apparent mass, and so designers had to take a new approach.

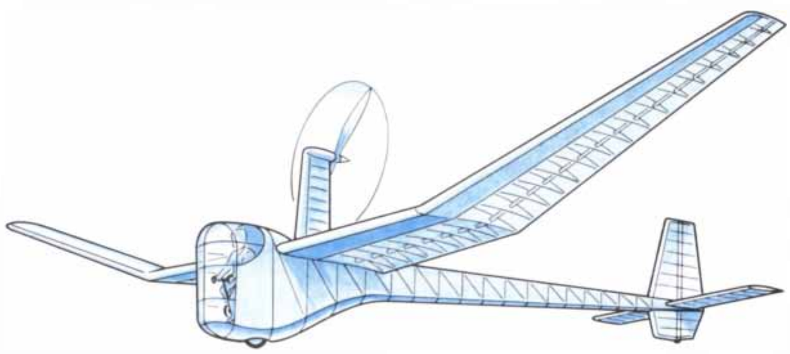
An aircraft must be controlled on three axes: yaw, pitch and roll [*see illustration on page 148*]. Usually a vertical rudder on the tail governs yaw, a horizontal elevator on the tail establishes pitch and horizontal ailerons on the wings determine roll. To begin a turn the pilot rolls the craft by means of the ailerons. This action tilts the lift vector of the wing, providing the side force needed for the turn. The rudder is then employed to "coordinate" the turn, keeping the nose pointed into the airstream. The ailerons control the roll rate, so that they are centered when the turn has been initiated and are used in the opposite direction to roll the aircraft out of the turn.

When ailerons are deflected, they impose on the wing a torque that tends to twist it along the axis of the span. The resulting change in the angle of



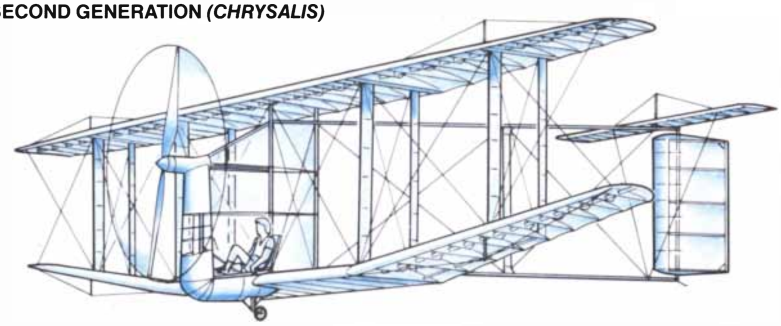
ADVANCING TECHNOLOGY of human-powered aircraft is marked by *Aviette, Mufli* and *Gossamer Condor. Aviette,* the winner of a competition sponsored by the Peugeot company between 1912 and 1922, was basically a jumping bicycle; it had no source of propulsion after it left the ground and so merely glided for a few meters. *Mufli,* a German craft operated in 1935, had a human-powered propeller, but the pilot could generate only enough power for an extended glide (712 meters was the longest one). *Condor* represents the class of human-powered craft that can fly indefinite distances and are fully controllable. In 1977 *Condor* won the first Kremer competition by completing a one-mile, figure-eight course.

FIRST GENERATION *(JUPITER)*



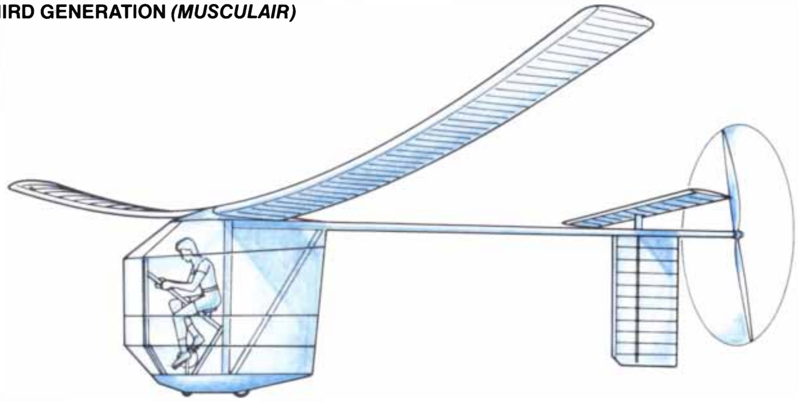
NAME	ORIGIN
MUFLI	GERMANY
PEDALIANTE	ITALY
SUMPAC	U.K.
PUFFIN I	U.K.
McAVOY	U.S.
VINE	S. AFRICA
MAYFLY	U.K.
PUFFIN II	U.K.
LINNET I	JAPAN
RELUCTANT PHOENIX	U.K.
LINNET II	JAPAN
MALLIGA	AUSTRIA
SM-OX	JAPAN
LINNET III	JAPAN
LINNET IV	JAPAN
MERCURY	U.K.
OTTAWA	CANADA
WRIGHT	U.K.
JUPITER	U.K.
TOUCAN I	U.K.
LIVERPUFFIN	U.K.
EGRET I	JAPAN
EGRET II	JAPAN
EGRET III	JAPAN
BURD I	U.S.
AVIETTE	FRANCE
EGRET IV	JAPAN
DEDAL III	POLAND
TOUCAN II	U.K.
STORK I	JAPAN
BURD II	U.S.
BLIESNER	U.S.
OLYMPIAN ZB-1	U.S.
ICARUS	U.S.
SKYCYCLE	U.S.
STORK II	JAPAN
NEWBURY MANFLIER	U.K.
PHOENIX	U.K.
PHILLIPS	U.K.

SECOND GENERATION *(CHRYSALIS)*



GOSSAMER CONDOR	U.S.
CHRYSALIS	U.S.
GOSSAMER ALBATROSS	U.S.
GOSSAMER PENGUIN	U.S.
MILAN '82	JAPAN

THIRD GENERATION *(MUSCULAIR)*



MONARCH	U.S.
HVS	W. GERMANY
BIONIC BAT	U.S.
PELARGOS	SWITZERLAND
MUSCULAIR	W. GERMANY
MONARCH B	U.S.
MAN-EAGLE	U.S.
SWIFT B	JAPAN

GROUPING OF CRAFT into three generations reflects major differences in the technology of human-powered flight. Aircraft of the first generation had internal wood trusswork; they were both heavy and fragile and could make only straight-line flights. Aircraft of the second generation had an aluminum-tube framework and external wires for bracing. These were the first fully controllable craft. Aircraft of the third generation are smaller and speedier. Modern materials such as graphite make cantilevered construction possible.

attack (and hence the lift) at each tip partially negates the effect of the aileron itself. For adequate control of roll the wing must therefore have enough rigidity to resist the twisting torque of the ailerons.

In the first two generations of human-powered aircraft the combination of large apparent mass and torsionally weak wings made ailerons ineffective. The problem was solved for *Gossamer Condor* by means of a canard: a control surface mounted on the fuselage to ride in front of the wing. On *Condor* the canard was tilted, producing a sideward force like that generated by a rudder, thereby achieving the desired yaw. The yawing motion produced a higher airspeed and a higher lift on the outside wingtip and a lower airspeed and lift on the inner one. The lift differential made the craft roll.

To keep the craft from banking too much a pilot flying *Condor* had to pull on the external bracing wires in order to twist the wings, much as the Wright brothers did on their *Flyer* of 1903. The maneuver increased the angle of attack (and hence the lift) on the inner wing and decreased it on the outer one. This action made sustained controlled turns possible.

Because human-powered aircraft of the third generation are smaller, their apparent-mass effects are smaller and the wing can be made considerably more rigid. Ailerons have proved practical for these machines.

We turn now to the three technological developments that have proved crucial to successful human-powered flight. They are high-lift airfoils, efficient propulsion systems and lightweight structures.

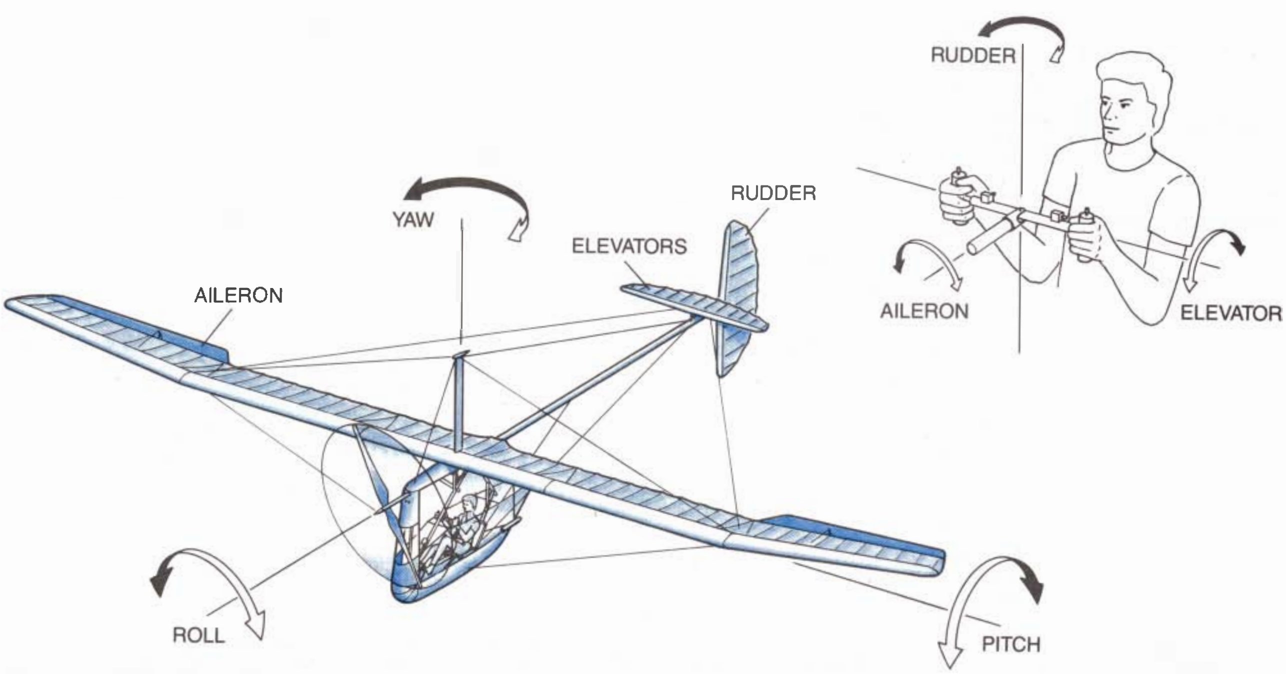
The main aerodynamic surface is the wing. Because it creates most of the drag, its cross-sectional shape (the airfoil) must be as efficient as possible. One measure of an airfoil's efficiency is the ratio of lift to drag ($L:D$). Another performance measure is the "power parameter," which is similar to $L:D$ but gives more emphasis to high lift. The higher the power parameter, the lower the power needed to sustain flight. Because low power is the primary concern in human-powered aircraft, a large power parameter is more important than a large $L:D$. To attain a high power parameter an airfoil must be capable of high lift but must not induce excessive drag.

The airfoil must also have a small pitching moment, meaning that it should tend to remain level along the axis of flight. A large pitching moment generates the same torques about the axis of the wingspan as ailerons do. (This is another reason the wings must be made torsionally stiff.) Higher stiffness invariably adds to the weight of the wing. In addition a high pitching moment destabilizes the aircraft and requires larger tail surfaces, which add weight and drag.

A factor that complicates the design of human-powered aircraft is that they operate in an unusual aerodynamic regime, normally the province of large birds and model airplanes. The regime is formally characterized by its relatively low Reynolds number, a dimensionless figure of merit that takes into account the speed, density and viscosity of the air together with the length of the body aligned with the flow. Typical aircraft operate at Reynolds numbers of between two million and 20 million; a vast store of information on that kind of flight has been built up since World War I. Human-powered aircraft operate at Reynolds numbers of less than one million, a poorly understood region of flight.

The low Reynolds number and the need for high lift, low drag and low pitching moment have required the designers of human-powered aircraft to adapt existing airfoils or to design new ones. The task is to tailor the distribution of pressures on the airfoil's surface. Loosely speaking, two types of airfoil could serve in human-powered aircraft: rear-loaded and front-loaded. The terms reflect the fact that the distribution of pressure on the top and bottom of the wing tends to be uneven, so that most of the load is carried either on the rear of the wing or on the front according to the choice made by the designer.

A typical rear-loaded airfoil offers a large lift-to-drag ratio that prevails through a fairly wide range of speeds and angles of attack. This type works well on sailplanes but not nearly as



THREE-AXIS CONTROL of a typical third-generation craft is achieved solely with the pilot's hands. (His legs are pedaling to provide the power for flight.) He controls roll by means of the ailerons, pitch by means of the elevator and yaw by means of the rndder.

well on human-powered aircraft. Its main disadvantage is its high pitching moment. This disadvantage and others, however, become less severe as the size of the aircraft decreases. The German *Musculair,* a successful third-generation craft, employed a rear-loaded airfoil.

Front-loaded airfoils offer high lift-to-drag ratios and also the highest power parameters, but only over a relatively narrow range of speed and angle of attack. Although those disadvantages make the airfoil unsuitable for sailplanes and most airplanes, it is ideal for human-powered aircraft, which perforce have a narrow speed range because of their limited power. Moreover, the airfoils have low pitching moments; hence a wing with a front-loaded airfoil can be built with a lower structural weight than can a wing with a rear-loaded airfoil.

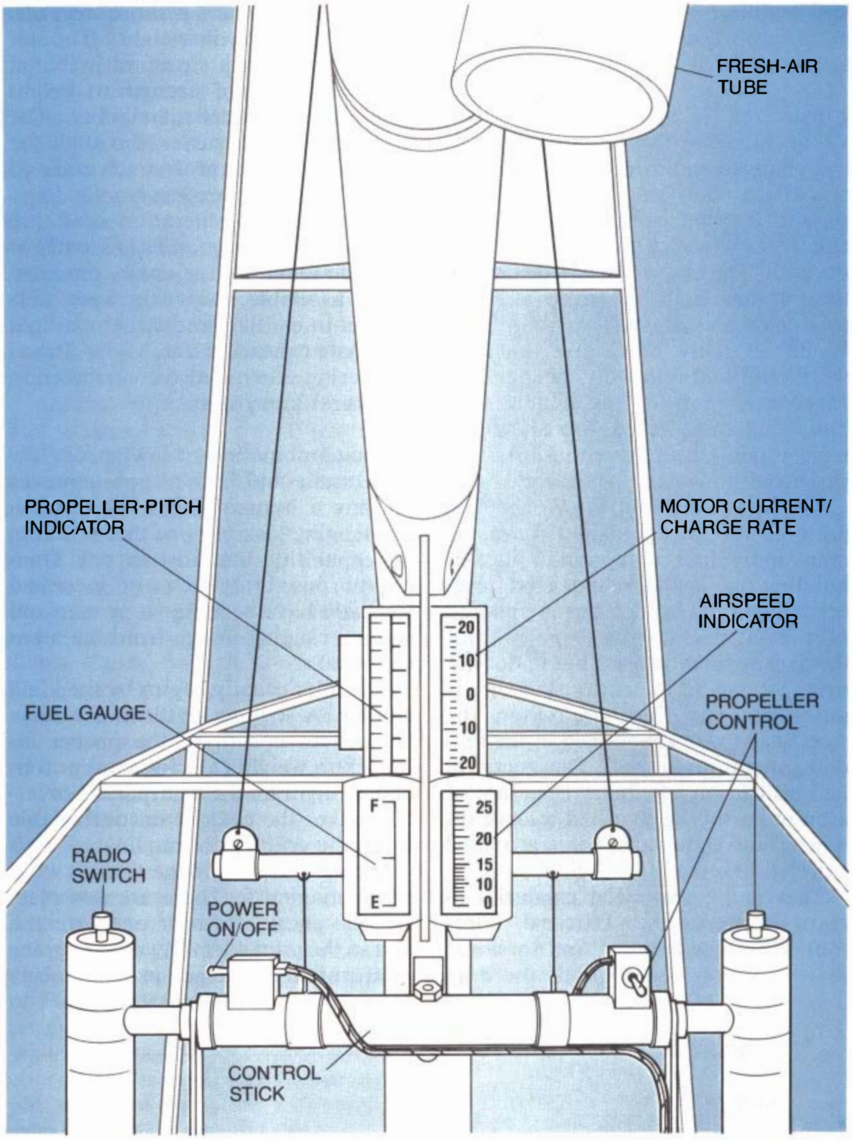
Propulsion is another area where high efficiency is important in human-powered aircraft. A propeller is by far the most effective means of transforming the mechanical power generated by the pilot's legs into a thrust sufficient to overcome the drag of the machine. One can conceive of other means of propulsion for human-powered aircraft, including flapping wings and jets of compressed air, but they have not yet been successful.

Any propulsive device (with the exception of a rocket) that generates thrust takes in air at flight speed and expels it to the rear at a higher speed as a jet. In the case of a propeller the jet is the slipstream: the air pushed aft of the propeller. A flapping wing pushes back an amorphous mass of air with each stroke.

In every case the jet carries kinetic energy that has been added by the propulsive device and that cannot be recovered; it is eventually dissipated as heat. As the jet velocity increases, the loss from wasted energy goes up faster than the gain in thrust. Thus efficiency dictates a device that takes in a large mass of air and adds to it only a small increment of velocity. This goal calls for a propeller that has a large diameter or for flapping wings that have a large span. (A compressed-air jet is inherently inefficient at the speeds of human-powered aircraft because of its high jet velocity.)

Although the flapping wing can in theory be made quite efficient, it has never been applied successfully to any aircraft powered by a human or carrying human passengers. To achieve high efficiency the wing must be twisted in one direction along the axis of its span on the downstroke and twisted in the other direction on the upstroke. Birds execute the maneuver quite well, but in a machine the combination of flapping and twisting creates severe mechanical and structural problems that get worse as the size of the craft increases. Hence a propeller is currently the only practical propulsive device for a human-powered aircraft.

The ideal of a large-diameter propeller faces certain constraints in a human-powered aircraft. A large propeller adds weight, which is something the designer is trying to avoid. Beyond a certain size the propeller tips are likely to strike the ground when the aircraft is taking off or landing. Thus the designer cannot achieve maximum efficiency by merely increasing the diameter of the propeller. Instead he must seek to reduce the efficiency-robbing kinetic energy in the slipstream of the propeller by careful attention to the distribution of the load on the blade. Air friction on the blades also influences the design of the propeller. An ample diameter and an optimum de-



CONTROL DEVICES of *Monarch B* are shown as the pilot sees them. The current/charge gauge records the rate at which the craft's battery is being charged or discharged; the charging takes place before the flight, when the pilot pedals a generator to store power. In flight he can draw on the stored power by means of the power switch, which actuates an electric motor. The fuel gauge records the amount of charge in the battery. The propeller-pitch indicator reflects the angle of the propeller blades, which the pilot can control by means of the propeller-control switch. By pushing the radio switch the pilot can talk to the ground crew; he can hear the radio at all times. *Monarch's* control stick bears a message of encouragement for the hardworking pilot: "You have great physical powers and an iron constitution."

sign enable a propeller on a human-powered aircraft to attain efficiencies approaching 90 percent.

Structural technology is the feature of human-powered aircraft that has changed the most since the vehicles of the first generation. In those early aircraft intricate trusswork made chiefly of wood provided form and strength. The truss is an efficient structure: it has high ratios of strength to weight and of stiffness to weight. Wood is easily obtainable, easy to work with and relatively inexpensive. Moreover, most of the people who build human-powered aircraft are or were model-airplane enthusiasts, experienced in working with wood.

On the other hand, the wood truss presents several drawbacks. It has so many individual pieces and joints that building one is a labor-intensive project. Mending a broken one is difficult. Moreover, if one truss member fails, the nearby members are put under unusual stress and the entire structure is put in jeopardy. For these reasons the wood truss was abandoned in the second-generation aircraft. Designers relied instead on a primary structure of aluminum tubing that had a large diameter and thin walls; wires provided external bracing.

The tubing was sized primarily to resist compression. External wires took all major bending and torsional loads. (At low flight speeds the drag created by the wires is more than offset by the saving in weight.) The advantage of such a structure is that it has a high ratio of strength to weight and provides excellent rigidity. The absence of wood trusses also made the second-generation craft much easier to repair than their predecessors.

Because third-generation craft are smaller and strong materials such as graphite and graphite-epoxy have become available, designers were able to turn to cantilevered structures that eliminate external wires. Mylar film as a covering skin has also contributed to structural improvement.

The combination of low speed, low altitude and limited power makes piloting a human-powered aircraft a challenging task but one that is within the capability of almost anyone. General-purpose craft such as *Condor* and *Chrysalis* have been flown by men and women ranging in age from the teens to the 60's.

The pilot usually begins by shedding clothes. A jogging outfit and a bicycle helmet constitute the proper attire: extra weight calls for extra power, and sunlight on the transparent covering makes the cockpit uncomfortable when the craft is not moving.

It is hard to get into the aircraft without damaging it. There are few places solid enough to bear one's weight, and so the pilot normally uses stepping platforms and is helped in by members of the ground crew. After running through a preflight checklist the pilot signals to the crew members holding the wingtips and begins to pedal.

The takeoff is surprisingly smooth. Most people flying for the first time are unaware that the craft is airborne until they hear the cheers of the ground crew. The cockpit is noisier than one would expect because of the whirring of the bicycle chain and the cyclic whooshing thump as the blades of the propeller pass the fairing.

In the air the pilot's main task is to concentrate on maintaining a steady attitude and airspeed. If the craft has been correctly trimmed for his weight, only small adjustments of the rudder are necessary. To climb the pilot pedals harder; to come down he or she reduces the pedaling rate.

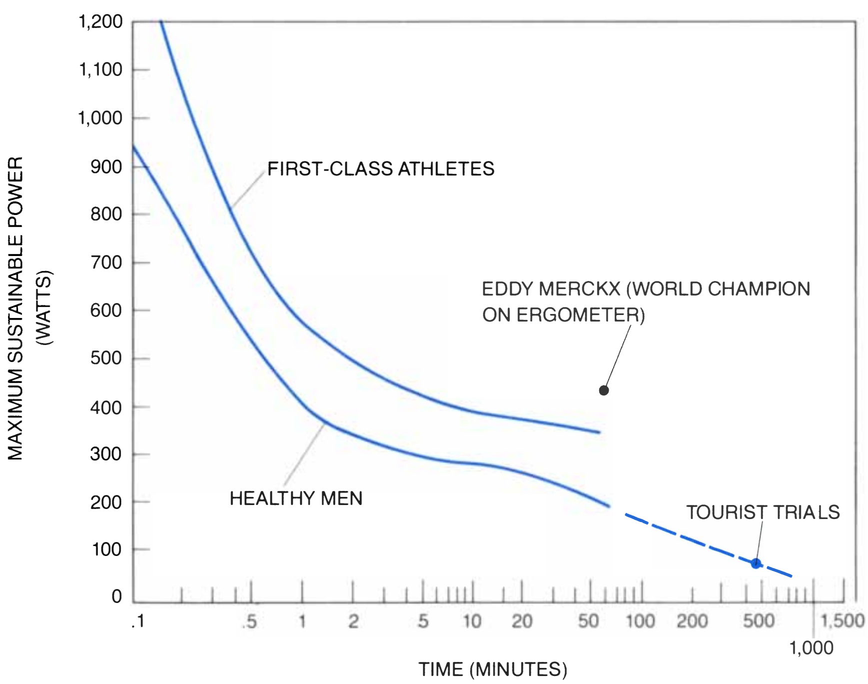
A human-powered aircraft reacts quite slowly to its controls; consequently inexperienced pilots are likely to overcontrol. Even more confusing is the tendency for the aircraft to respond differently on different axes. The pitch response is relatively fast, the roll response agonizingly slow. Making turns around a specific course, as is required in the speed competitions, calls for careful coordination and much practice.

The two primary dangers in flight are stalls and gusts of wind. The craft stalls when the airflow separates from the surface of the wing. The separation usually occurs because the pilot has let the craft's speed fall too low.

The inherently low flight speed makes gusts a special problem. Because the relation between wind speed and the speed of the aircraft is crucial, a gust of only five miles per hour is equivalent to one of 30 miles per hour or more on a small conventional airplane. Striking from the front, such a gust can overload and break the wing; from the rear it can cause a stall. A gust can also change both the flight path and the plane's attitude. Fortunately the low speed and the low altitude combine to make a human-powered aircraft fairly safe; crashes that demolish the airframe usually inflict only cuts and bruises on the pilot.

To land the pilot aligns the craft with the runway and reduces his rate of pedaling. The vehicle glides gently in and touches down softly.

Human-powered flight has been pursued mostly for its own sake, greatly spurred by the incentive of the various competitions. Nevertheless, the technologies that have evolved can be expected to have practical applications in at least three areas: human-powered flight itself, ultralight aircraft and a variety of reconnaissance and observational tasks.



HUMAN POWER varies according to the age, condition and motivation of the person. The range is indicated on this chart. The "tourist trials" line shows values deduced from cross-country bicycle races. One kilowatt (1,000 watts) is the equivalent of 1.3 horsepower. The data are derived from *Bicycling Science,* by Frank Rowland Whitt and David Gordon Wilson.

In human-powered flight the speed competition as it is now set up will continue until 11 more awards of £5,000 each have been won. The result will certainly be faster aircraft, perhaps attaining speeds of as much as 85 kilometers per hour. Such a speed will be difficult to achieve, however, even with highly efficient schemes for storing energy. It remains to be seen whether the proposed prizes will be sufficient to elicit the necessary investment of money and thought.

The Royal Aeronautical Society is examining the possibility of staging additional competitions. They would probably be aimed at making human-powered aircraft more practical and more rugged. At the other end of the spectrum it is now possible (in our view) to build a large, low-power craft that could turn the legend of Daedalus into reality, flying the 96 kilometers from Crete to the Greek mainland at about 22 kilometers per hour.

The relation between human-powered and ultralight aircraft (which are powered by small gasoline engines) is that the former, as they are increasingly designed for speed and utility, become more like the latter. Human-powered aircraft of the third generation cruise on about .5 horsepower and climb splendidly at two horsepower. The engines of today's ultralight aircraft produce from 30 to 50 horsepower. Improvements in the technology of human-powered aircraft can be expected to reduce this gap so that human-powered aircraft can perform some of the tasks now calling for ultralights and ultralights can function at lower power.

The final application has to do with high-altitude operations. High-altitude craft capable of prolonged flights are now being considered as unmanned platforms for reconnaissance, communication relays and sampling work in the stratosphere. A high-altitude craft operates at the low Reynolds numbers characteristic of human-powered aircraft. Hence the technologies developed to increase the structural strength and reduce the weight of human-powered craft will also benefit the high-altitude vehicles.

Eventually these technologies might find application in space. For example, the atmosphere of Mars, even though it is much less dense than the earth's atmosphere, could support winged flight at Reynolds numbers similar to those of human-powered aircraft. A winged, unmanned vehicle (an airborne analogue of *Lunar Rover*) would be an effective platform from which to examine the terrain and sample the atmosphere of Mars.

Smart Cards

Cards containing microcircuitry are more versatile and securer than conventional credit cards. A microelectronic chip must meet severe constraints to function in this unique environment

by Robert McIvor

Within a year it may be possible to carry much of the power of a personal computer in one small compartment of a billfold. It will reside in a device called a smart card: one or more microelectronic chips mounted in a piece of plastic the size of a credit card.

Smart cards that have slightly less computing power than a personal computer are already being produced and used in France in a number of applications. In September, Mastercard began issuing some 50,000 cards to American users in Washington, D.C., and Palm Springs, Fla., as part of a field trial. Eventually there may be many types of smart cards, each having a different level of sophistication. The most advanced of these might carry small liquid-crystal-display readouts and have the capability to encrypt and decrypt any dialogue with an external device; they will be powered by solar panels. The least sophisticated card might consist of little more than a simple processor and a small memory; it could serve, for example, as a "debit card" for long-distance calls on public telephones or as a complete personal medical history. It is not inconceivable that within a few years smart cards will be able to store digitized versions of their owners' signatures, retina prints or fingerprints; personalized smart cards could be used as highly secure keys, allowing access to telephone networks, corporate data banks and secure buildings.

In what is perhaps their most natural application, smart cards may replace credit cards and the cards that function in automatic banking machines. In a typical smart-card transaction, a cardholder puts the card in a "point-of-sale terminal": a special card-reading cash register. The terminal supplies the card with electric power and communicates with its microcircuitry through eight metal contacts on the card's surface. The cardholder is asked to enter a password on a keyboard, and the terminal verifies that the card is valid and the user is legitimate. When the sale is rung up, the amount of the transaction is stored in the card's memory, credited to the retailer's account and deducted from the cardholder's credit balance, which is also stored in the card's memory. The cardholder may be able to replenish the credit balance at an automatic banking machine.

The credit cards and automatic-banking-machine cards that could thus be replaced store information in embossed areas and magnetic stripes. Magnetic-stripe cards are entirely passive—they are merely a medium for storing information—and are vulnerable to many forms of fraud and abuse. The simplest form of abuse is overspending: since most purchases are not reported immediately to the company that has issued the card, a customer can spend much more than his or her credit limit by making a large number of relatively small transactions or one large unvalidated transaction.

Methods of fraud tend to be more sophisticated. For example, machines are available that copy information stored in the magnetic stripe of a credit card onto a dummy, or blank, card as they take a carbon-paper impression of the card's embossing. These devices closely resemble the machines that stores and restaurants use to create receipts and record purchases. Another type of fraud involves "personal identification numbers," or passwords. When a magnetic-stripe card is used in an automatic banking machine, the machine first asks the cardholder for a password. It then reads the correct password directly from the card and compares it with the given password. Hence at some point during the transaction the correct password must be brought into the banking machine's working memory. Any "hacker" who has some way to monitor that memory can learn the cardholder's password.

Smart cards have two essential properties that make them invulnerable to all these kinds of fraud and abuse. First, a smart card has a nonvolatile, programmable, read-only memory: a memory into which information can be placed after the card has been issued and that will remember any such information even when it is not connected to a power source. This memory can record the amount of each transaction and the total amount that has been spent, ensuring that the user does not spend beyond a set limit.

Second, each smart card contains its own central processing unit—essentially a small computer—which controls all the interactions between the memory and the various external devices that read the card and enter data onto it. The central processing unit and memory architecture can be constructed in such a way that certain parts of the card's memory are physically or logically inaccessible to anyone but the card's issuer: the central processing unit will not obey an instruction from anyone else to read or change those parts of memory.

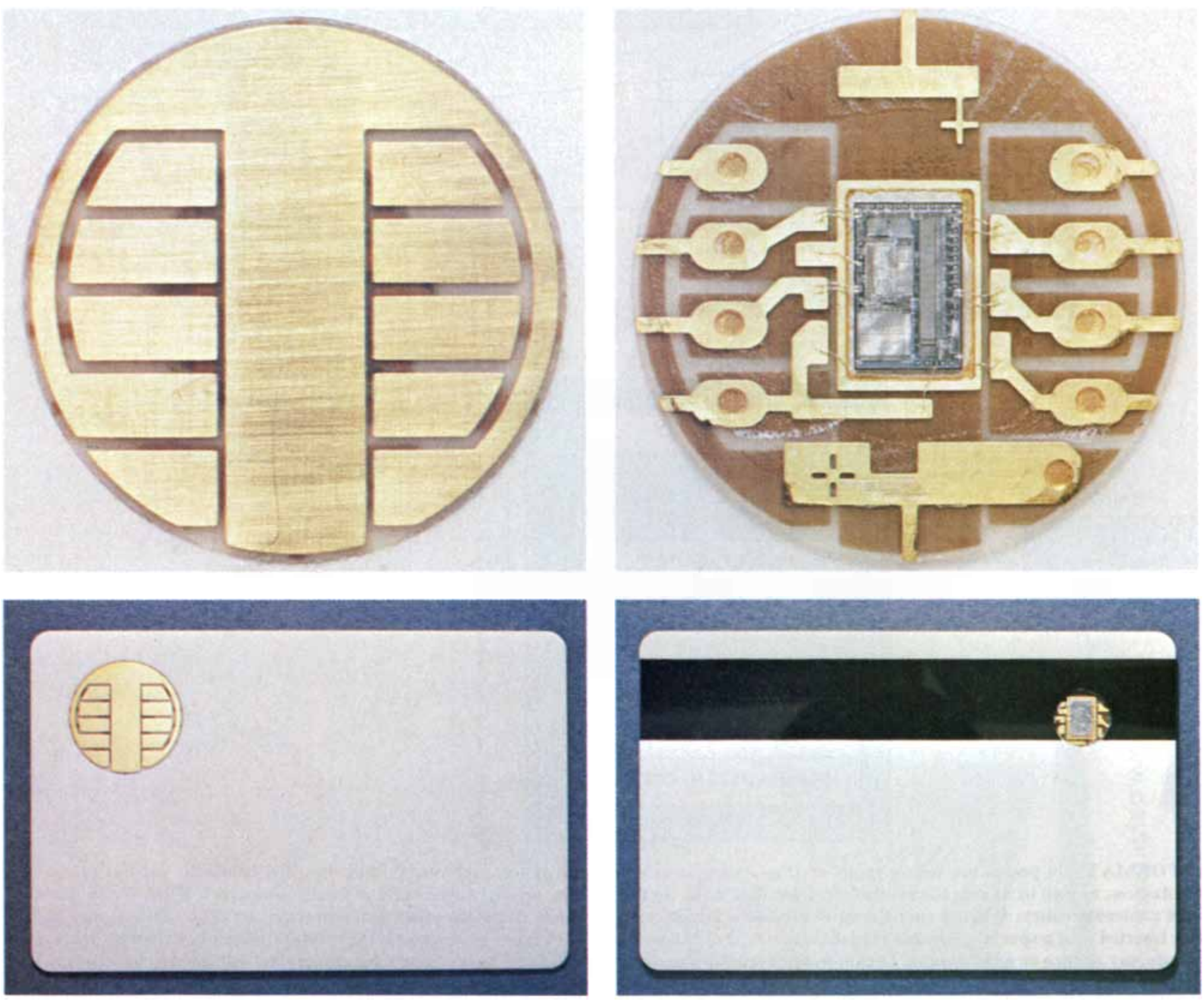
Through the central processing unit the card itself can examine any proposed password and compare it with the correct password, which is stored in a secret location in the card's memory. The card never has to reveal the correct password to any outside system. As a matter of fact, even the company that issues the card does not need to know the correct password. When the card is first issued, the cardholder can program the password directly onto it by means of a card-reading/writing machine. After the password has been entered and checked (the cardholder is asked to enter the password two or three times to ensure that it is not entered incorrectly), the card stores it in the memory's "secret zone."

In addition to the password the secret zone stores the cardholder's account balance, the card's serial number and a sequence of letters or numbers, chosen by the card's issuer, that can be used to determine whether the card is legitimate. Another zone of the programmable, read-only memory, called the open zone, might store the cardholder's name, address, telephone number and account number. The open zone can be read by any card-reading machine, but it cannot be altered; the card's central processing unit will not obey instructions to change any information in the open zone.

Whenever the card is used to make a purchase, such information as the amount of the purchase, the name and address of the store and the date is stored in a third zone of the memory, called the working zone. Information can be written there only under certain conditions (when the card is in a legitimate cash register, say) and can be read or written only with the cardholder's permission. A cardholder can buy a separate card-reading machine, which, when it is connected to a home computer, a television set or a printer, displays a complete record of all purchases made with the card.

In designing a smart card the engineer is presented with a particularly severe set of constraints. First of all, the existing technology—that of embossed, magnetic-stripe cards—limits the number of possible places the chip can occupy on the card. Because of the large number of devices already in use that read the embossing or magnetic stripe of a card, the locations of the stripe and the embossing cannot be changed. Thus the chip cannot be placed in the lower, embossed region of the card. On the other hand, a chip in the upper half of a card could interfere with the card's logo and, more important, might also interfere with the performance of the magnetic stripe.

Two locations, deemed to be the least unsuitable of those remaining, are now under consideration. If it is possible to install the chip so that it does not affect the magnetic stripe, the chip could be placed in the upper left-hand corner of the card. Alternatively, it could be placed slightly above the long axis of the card, near the card's left side. Unfortunately this is a region where mechanical stress is highest when the card is bent about the long axis. Part of the purpose of the

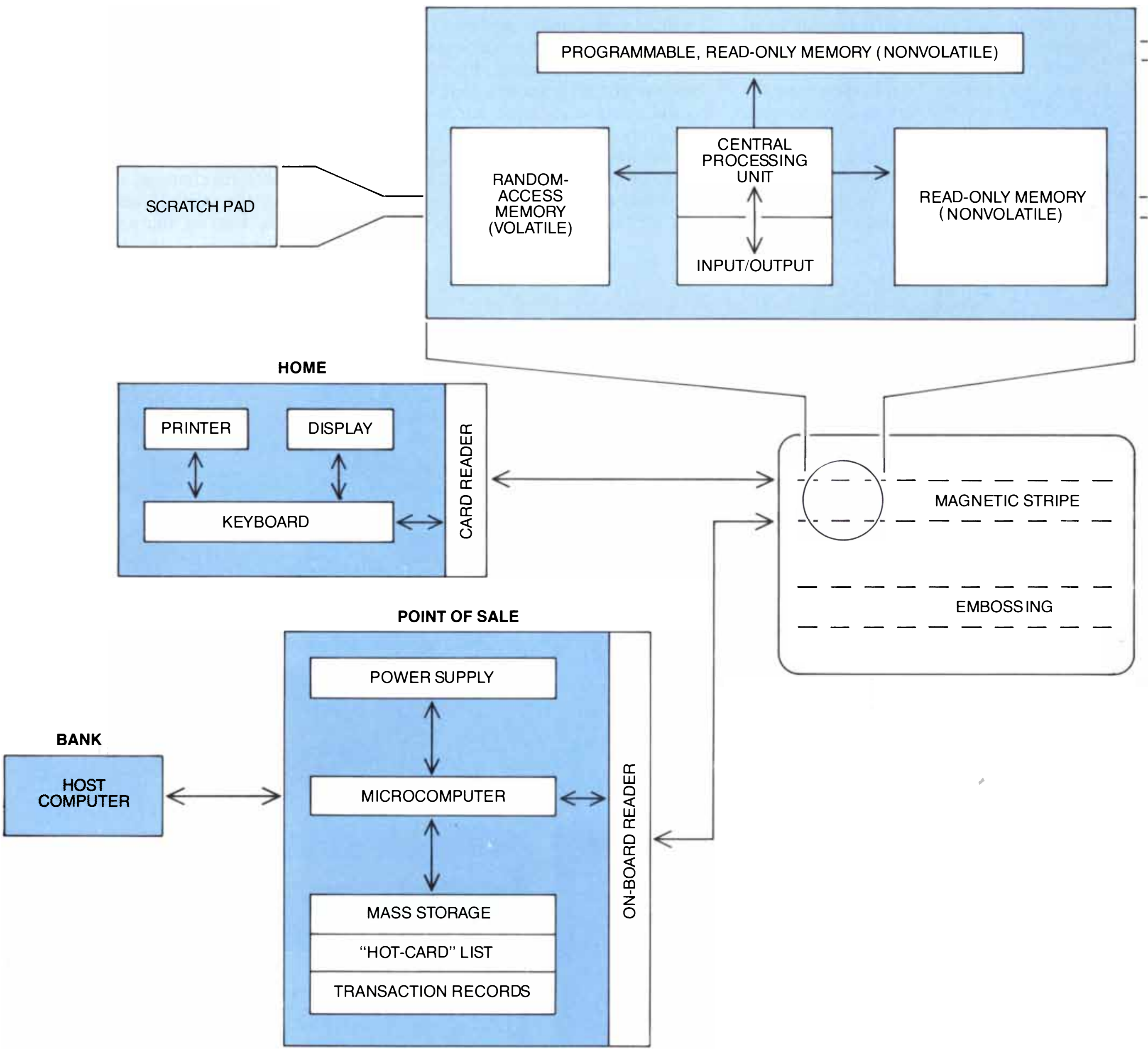


SMART CARDS, plastic cards containing microelectronic chips, can act as credit cards, keys or portable computers. For example, the chip (*top right*), which holds a microprocessor and several kinds of memory bank, can store and update a cardholder's credit-account balance and keep a complete record of all transactions made with the card. The chip communicates with external machinery through eight metal contacts (*top left*) on the front of the card (*bottom left*). In the photograph at the bottom right, part of a card's magnetic stripe has been removed to show one possible placement of the chip. The chip could also be placed just above the card's long axis.

Mastercard field trials, in which half of the cards have chips in one location and half have them in the other, is to determine whether the location of the chip has a significant effect on the card's reliability. The most acceptable location will be established as an international standard by the multilateral International Standards Organization.

There is another question that is decided largely by the environment in which smart-card microcircuitry must exist: Should all the processors and memories be integrated into one chip, or should each card carry two chips? In a two-chip card one chip would contain the central processing unit, a working memory and little else; the other would consist almost entirely of memory banks.

Chips that contain little more than a microprocessor and a working memory, and others that contain only memory, are readily available, and so little innovation would be needed to produce a card incorporating the functions on separate chips. It might be



INFORMATION passes between a smart card and various external devices, as well as among the several elements that make up the card's microcircuitry. When a smart card is used as a credit card, it is inserted into a special on-board reader (*bottom left*) at the point of sale, say a store or a restaurant. The on-board reader is connected to a microcomputer that has access to a memory containing a list of stolen cards and a complete record of the store's transactions. The point-of-sale machine periodically communicates with the card-issuing company's host computer (*far left*) both to report transactions that have occurred and to update the store's list of stolen cards. A special off-board reader (*middle left*) in the cardholder's home can provide the cardholder with a complete printed record of transactions. Within the card itself, the central processing unit, or CPU (essentially a small computer), is the nexus through which all information passes (*top center*). The CPU is connected to three types of memory. The read-only memory (*bottom right*) is established by the card's manufacturer and cannot be altered; it is nonvolatile (information stored in it remains even when the card is not connected to a power source). The read-only memory contains the card's operating system: the sequence of steps the central processing unit is to follow when it is given a command by an external device. Another memory, the random-access memory, is volatile: it cannot retain information after the card has been disconnected from a power source. It is used as a scratch pad (a place to store

much more expensive to design and fabricate a single chip containing all the necessary microcircuitry.

Nevertheless, there are two compelling reasons to put all the electronics on one chip. First, putting two chips in a setting that endures as much physical abuse (bending, torque, extreme temperatures and spilled liquids) as a credit card does would nearly double the chances of failure. Each of the two chips might fail just as readily as a single chip, and the failure of one would make the card useless. Second, if the memory is separated from the central processing unit, there must be an electrical connection between them. It might be possible to monitor the connection, thereby violating the security of the chip's memory.

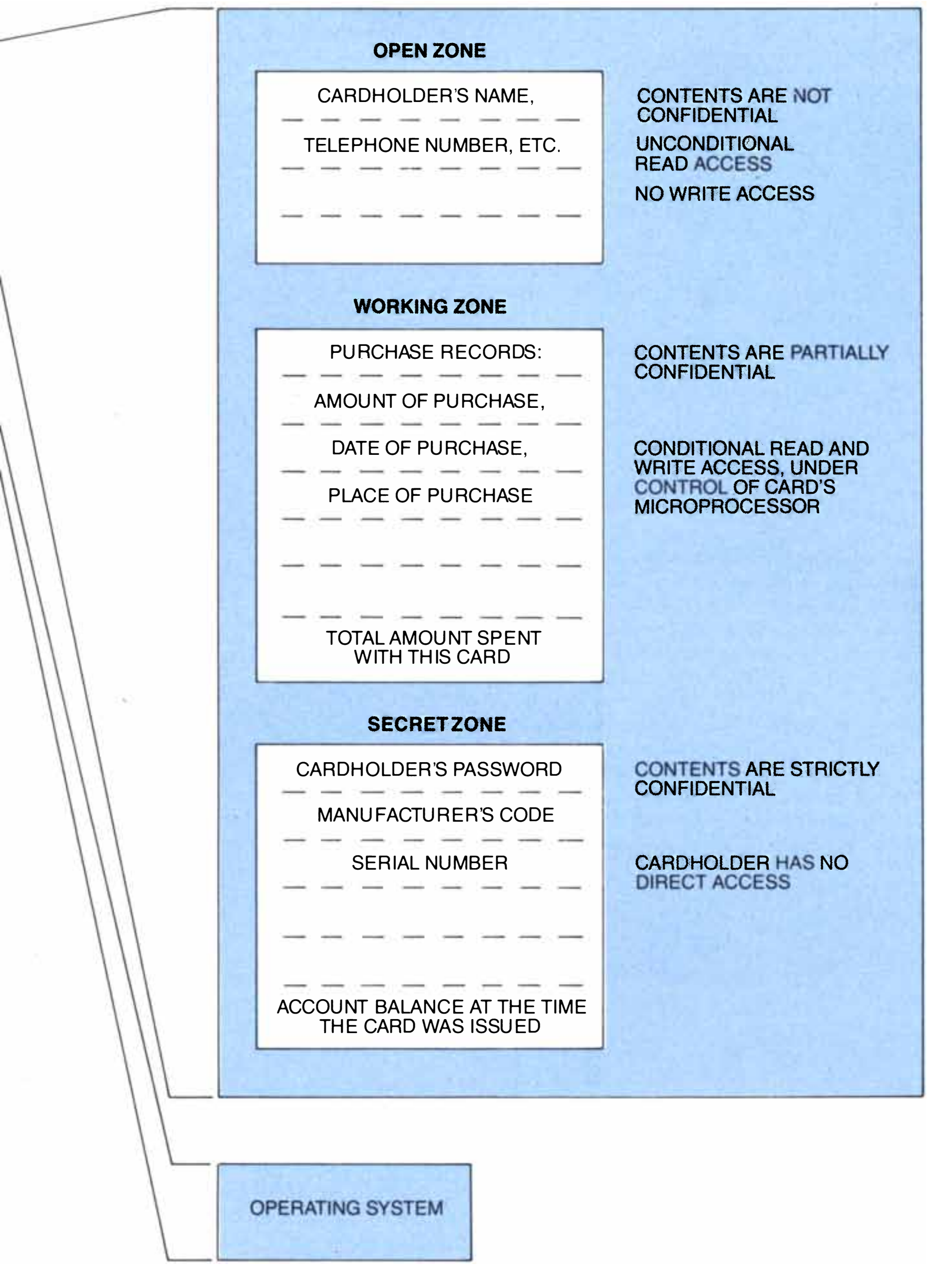
The unique environment in which the smart card's microcircuitry must operate places tight constraints not only on the overall design of the card but also on the design of the chip itself. For example, a chip that is to be embedded in a credit card must be thinner than the standard silicon wafer; chips designed for smart cards are typically .011 inch thick instead of the more standard .015 inch. Also, the area of the chip should be as small as possible: a chip that occupies a large area on the card will be subject to a greater amount of strain than a smaller chip when the plastic card is bent. It is therefore particularly important to select a type of microcircuitry that can be packed onto a chip very densely. The cost of producing the chip may also be an issue.

The chip should be able to operate well in an electronically "noisy" environment; for example, a debit card that is to be used in a gas pump must be able to function even when it is near other electronic devices, such as the electromechanical apparatus in nearby cars and in the pump itself. It is also important that the microcircuitry not consume large amounts of electric power. In many applications the card's central processing unit need not be particularly fast, however: most of the card's operation will involve dialogue with the user, and the response time of even the slowest microprocessors is sufficient to keep pace with the relative slowness of a human being punching a keyboard. (In some applications, on the other hand, such as high-security ones that require extensive encryption and decryption, high-speed operation could be important.)

Hence the design engineer has to choose circuitry that is small and can be packed densely on a chip. It must use as little power as possible, and it should be inexpensive to manufacture.

Perhaps the most important decision is to determine what kind of transistors to use. Transistors are the basic building blocks of microprocessors and memories; a chip may contain as many as 100,000 transistors. A transistor is essentially a switch: an input voltage applied to one element of the transistor creates or eliminates an electrical channel between two other elements. When there is a channel, current can flow through the device.

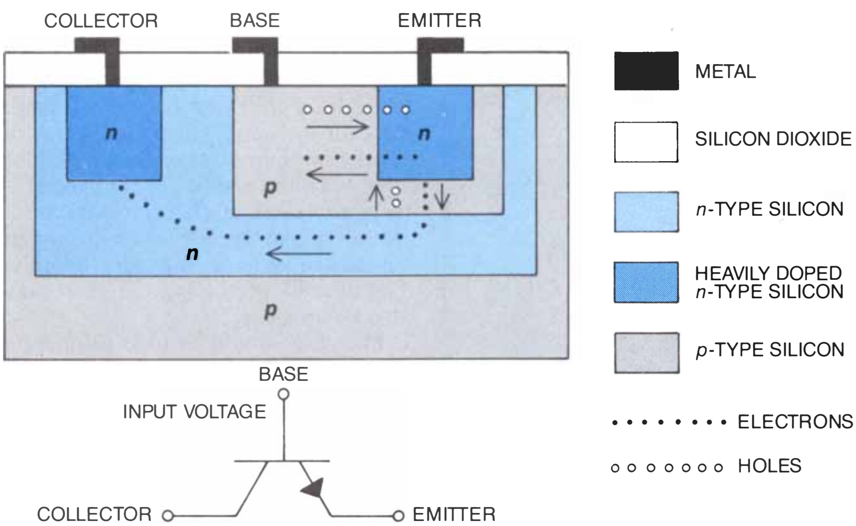
A transistor consists of several adjacent regions of a doped silicon crystal: a crystal to which impurity atoms have been added to change its electrical properties. If the added impurity atoms have more electrons in their va-



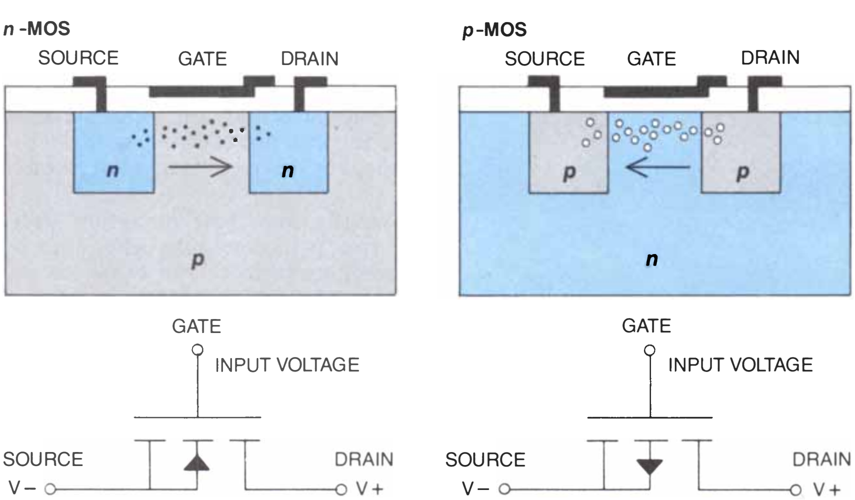
the intermediate results of calculations) by the central processing unit. The third memory, a programmable, read-only memory, gives the smart card much of its flexibility and utility. The programmable, read-only memory is divided into three zones (*right*). One of these, the secret zone, is accessible only to the company issuing the card: the card's CPU will not follow an instruction from anyone else to read data from the secret zone or to alter any information stored there. The secret zone contains such information as the cardholder's password and credit limit. Other information, such as the cardholder's name, is stored in an open zone. The open zone can be read by any card-reading device, but no information in it can be altered. The third zone, called the working zone, contains a complete record of purchases made with the card: the amount of the purchase, the place and date of purchase and the total spent on all purchases. The working memory can be read from and written to only under certain conditions (when the card is in a legitimate cash register, for example).

lence, or outer, shell than silicon atoms do, some electrons will not participate in the bonds that hold together the crystal lattice and so will be free to move about within the silicon. If an electric field is applied to the crystal, the free electrons will move, creating an electric current. Since the electrons that carry the current are negatively charged, silicon into which atoms that give up electrons have been introduced is said to be *n*-doped.

Conversely, impurity atoms can be added that have fewer valence electrons than silicon has. Then there will be a number of "holes," vacant sites that would normally be occupied by electrons, within the crystal. In a sense



***NPN* BIPOLAR TRANSISTOR acts as a switch: a voltage applied to an element called the base causes current to flow between two other elements, called the emitter and collector. The transistor is made of doped silicon: silicon to which impurity atoms have been added to change its electronic properties. In *n*-doped silicon there are extra electrons available to carry current. Conversely, in *p*-doped silicon there are "holes": vacant sites that could be occupied by electrons. Holes, acting as positively charged particles, can carry a current. In an *npn* bipolar transistor, current normally cannot flow from the emitter to the collector because electrons cannot penetrate the *p*-doped region of the base. When a positive voltage is applied to the base and the collector, however, some holes, repelled by a positive voltage, flow from the base into the emitter. In exchange, electrons flow from the emitter into the base. Most electrons flow through the base into the lightly *n*-doped surrounding region and then into the collector. In a wiring diagram (*bottom*) current is conventionally represented by arrows pointing opposite to the direction in which the negatively charged electrons flow.**



METAL-OXIDE-SEMICONDUCTOR (MOS) TRANSISTORS can be built in two ways. In an *n*-MOS transistor (*left*) current cannot flow through the *p*-doped region between the source and the drain unless a positive voltage is applied to a gate, which lies just above a thin layer of insulating silicon dioxide. The voltage attracts a layer of electrons, called an inversion layer, forming a conductive channel through which current can flow. Conversely, in a *p*-MOS transistor (*right*) a channel is formed by applying a negative voltage to the gate, attracting an inversion layer of holes. The holes flow from the drain to the source.

holes act as positively charged particles: when an electric field is applied to the crystal, holes move from atom to atom, producing an electric current. Because this current is made up of positively charged "particles," silicon with an excess of holes is said to be *p*-doped.

In one type of transistor, known as an *npn* bipolar transistor, a heavily *n*-doped region, called the collector, lies within a lightly *n*-doped region [*see top illustration at left*]. A *p*-doped region, called the base, also lies within the lightly *n*-doped region. Another heavily *n*-doped region, called the emitter, lies within the *p*-doped region and is completely surrounded by it.

As a rule electrons cannot flow from the emitter to the collector, even if a positive voltage (a voltage that attracts electrons) is applied to the collector; they cannot pass through the *p*-doped base. When a positive voltage is applied to the base, however, some holes pass from the base into the emitter and in exchange some electrons from the emitter are injected into the base. Many of these electrons pass through the base and into the lightly *n*-doped region in which the collector, base and emitter lie. From there they pass into the collector. In this way the positive voltage applied to the base acts as an input signal: it induces a current to flow between the emitter and the collector. The current is the transistor's output signal.

Bipolar transistors are relatively inexpensive to make, and they are fast switches (that is, they respond quickly when an input voltage is applied to the base). On the other hand, they consume a relatively high amount of power: current flows from both the base and the emitter whenever the transistor is "on" (whenever a positive voltage is applied to the base). Bipolar transistors are therefore not entirely suitable for many of the applications to which smart cards will be put.

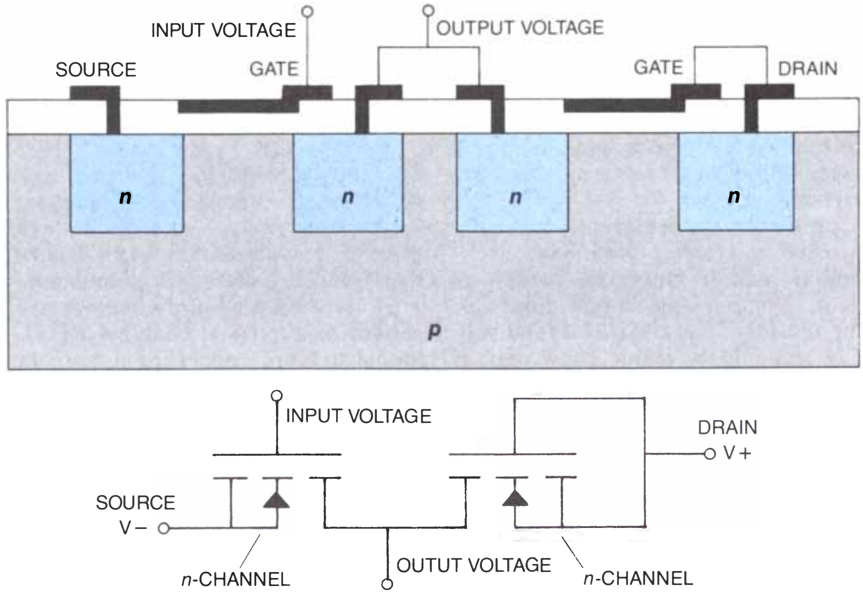
Much more promising than bipolar transistors are those built with metal-oxide-silicon, or MOS, technology. There are two types of MOS transistor, respectively known as *n*-MOS and *p*-MOS. An *n*-MOS transistor contains two islands of highly *n*-doped silicon called the source and the drain; both are embedded in a substrate of *p*-doped material [*see bottom illustration at left*]. A thin layer of silicon dioxide (an insulator) covers the silicon crystal, and a metal or silicon electrode, called the gate, is deposited on top of the region of silicon dioxide that is just above the *p*-doped silicon lying between the two *n*-doped regions.

As in a bipolar transistor, normally current cannot pass through the device; to pass from one *n*-doped region to the other it would have to traverse a *p*-doped region. When a positive voltage is applied to the gate, however, electrons from within the *p*-doped substrate are pulled into the region of *p*-doped silicon that lies just under the silicon dioxide layer between the source and the drain. These electrons form a layer that is called an inversion layer because it is made up of charge carriers that are opposite in sign to the holes that usually inhabit a *p*-doped semiconductor. The inversion layer acts as a conductive channel through which other electrons may flow from the source to the drain.

A *p*-MOS transistor operates on the same principle as an *n*-MOS transistor, but the source and drain are *p*-doped and lie in a substrate of *n*-doped material. In a *p*-MOS transistor current is carried by holes, rather than electrons, that pass from drain to source; a negative voltage, rather than a positive one, is needed to bring enough holes near the transistor's surface to form the channel between source and drain.

In an MOS circuit, as in a bipolar transistor, the voltage applied to the gate represents an input signal. Unlike the current of bipolar transistors, however, the current between the source and the drain is not usually used as the output signal. In most MOS technology the drain is electrically connected, either by a resistor or by a transistor whose channel is always conductive, to a line carrying a high voltage [*see illustration on this page*]. The source is connected to a line carrying a low voltage.

When the input voltage (the voltage applied to the transistor's gate) is low, the transistor is unable to carry a current and the drain has no direct electrical connection to the low-voltage line. The drain does, however, have a direct connection to a high-voltage line. It therefore assumes a high voltage. When the input voltage is high, the transistor can carry current; the drain is then connected to the low-voltage line and assumes a lower voltage. The voltage of the drain, which is low when the input voltage is high and high when the input voltage is low, is the output signal of the MOS device.



***N*-CHANNEL MOS SWITCHING DEVICE consists of two *n*-MOS transistors wired together. One transistor's gate and drain are wired together (*far right*); they are connected to a high-voltage line, and so current can always flow between the transistor's source and its drain. The other transistor's source (*far left*) is connected to a low-voltage line. Its drain is directly wired to the source of the first transistor; the voltage at this junction (*center*) represents the output. When the input voltage (*left*) is low, the output node is electrically isolated from the leftmost transistor's source: there is no layer of electrons to carry a current. The output node is connected, however, by means of the permanently open transistor, to the high-voltage line attached to the device's drain. The output node therefore assumes a high voltage. When the input voltage is high, the output node is no longer isolated from the low-voltage line attached to the device's source and therefore assumes a lower voltage.**

Such MOS circuits are reasonably fast, inexpensive to make and reliable. They can also be packed densely onto a chip. On the other hand, they are relatively sensitive to noise (electrical signals generated by neighboring devices), and although they require less power to operate than bipolar transistors, they still draw a lot of power. They may therefore be unacceptable in some applications of smart cards.

A variation on this kind of MOS technology, called complementary MOS, or CMOS, uses nearly an order of magnitude less power and is approximately half as sensitive to noise. It may prove to be the most appropriate technology for smart cards. A CMOS circuit is a single crystal containing both an *n*-MOS and a *p*-MOS transistor. The gates of the two transistors are wired together [*see illustration on next page*] and the voltage applied to them serves as the input signal. The responses of the two transistors to any input signal will be complementary: a signal that activates the *p*-MOS transistor will inactivate the *n*-MOS transistor and vice versa.

The drain of the *n*-MOS transistor is wired to the source of the *p*-MOS transistor, so that these elements are always at the same voltage. This voltage is the output signal. The source of the *n*-MOS transistor is connected to a line bearing a low voltage, and the drain of the *p*-MOS transistor is connected to a high-voltage line.

When the input voltage is positive, the channel of the *n*-MOS transistor is conductive and that of the *p*-MOS transistor is not. The drain of the *n*-MOS transistor and the source of the *p*-MOS transistor are therefore electrically connected to the low-voltage line and isolated from the high-voltage line. The output voltage hence assumes a low value. Conversely, when the input voltage is negative, current can flow through the *p*-MOS but not the *n*-MOS transistor. The drain of the *n*-MOS transistor and the source of the *p*-MOS transistor are then connected to the high-voltage line and isolated from the low voltage line. The output voltage hence assumes a high value.

Perhaps the most important feature of a CMOS device is that no current passes between the high-voltage line connected to the *p*-MOS transistor's drain and the low-voltage line connected to the *n*-MOS transistor's source except during the short periods when the input signal is switched between high and low voltage. Hence a CMOS device draws far less current than either a bipolar transistor or a pair of transistors that are both *n*-MOS or both *p*-MOS.

A CMOS device is also much less susceptible to electrical noise. To switch a standard *n*-MOS device, say from low output voltage to high output voltage, the input voltage need only be switched from some positive value, at which the conductive channel between source and drain is open, to zero voltage. (When the gate is set at zero voltage, no electrons are attracted to the surface of the device and there is no conductive channel.) To switch a CMOS device the input voltage must be switched from a positive value, which allows the *p*-MOS transistor to conduct, to a negative value, which allows the *n*-MOS transistor to conduct. The difference between the new input voltage and the old is twice as high for a CMOS device as it is for *n*-MOS circuitry, and so a spurious signal would have to be twice as intense to force the CMOS device into an error.

CMOS circuitry is currently more expensive to produce than *n*-MOS technology, and it cannot be packed as densely onto a chip. In part, however, the reason is that CMOS is a relatively new technology. Within a few years it will be possible to produce CMOS devices that are as small and inexpensive as *n*-MOS devices. Some prototypes of these devices, called high-performance CMOS, or HCMOS, devices already exist. The technology of an HCMOS device is essentially the same as that of a CMOS device, but the manufacturing equipment is more precise, allowing more components to be built in each unit of area. The separation between the source and the drain of a CMOS device is typically about three micrometers; that of an HCMOS device is less than one micrometer.

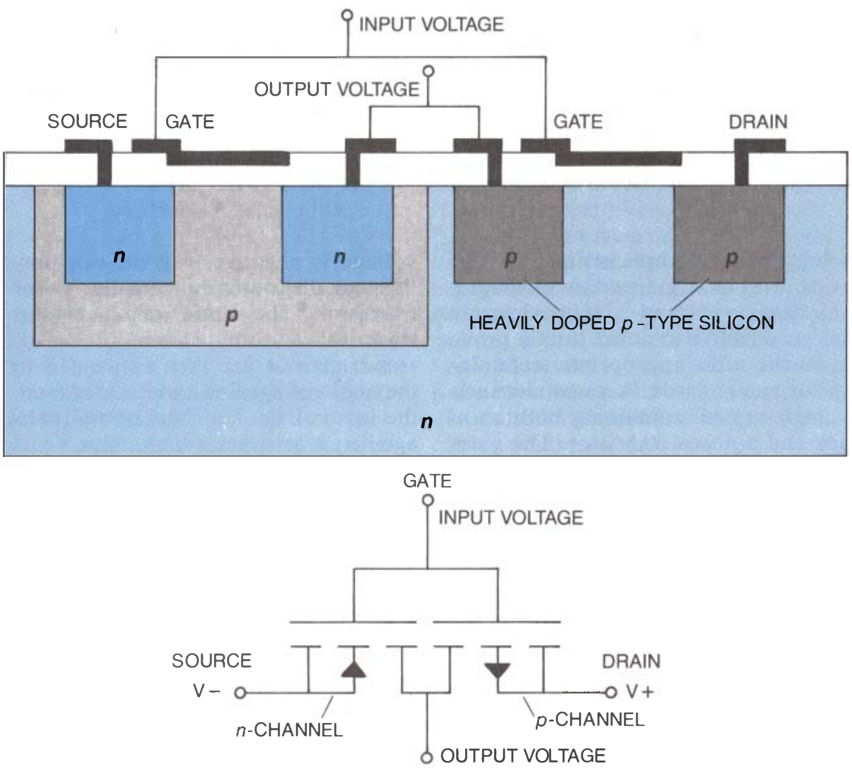
A final technological question concerns the smart card's memory. Every smart card will have at least three general kinds of memory. One of these, called a read-only memory, or ROM, is established at the factory and cannot be altered. It contains, for example, the card's operating system: the set of instructions the central processing unit (the part of the chip that performs logic operations) is to follow when it is given certain commands. Another memory, a volatile random-access memory, or RAM, is a very fast memory that is left blank by the manufacturer. Any part of the RAM can be altered by the card's central processing unit, but information in the RAM is erased whenever the card is disconnected from a power source. The RAM is used as a scratch pad: it stores intermediate results of calculations done by the central processing unit.

Most of the third kind of memory, the nonvolatile, programmable, read-only memory, is left blank by the manufacturer and can be altered by the central processing unit. Information entered into this memory is retained even after the power source is removed. It is this memory that houses the open, working and secret zones that help to ensure the security of information stored on a smart card.

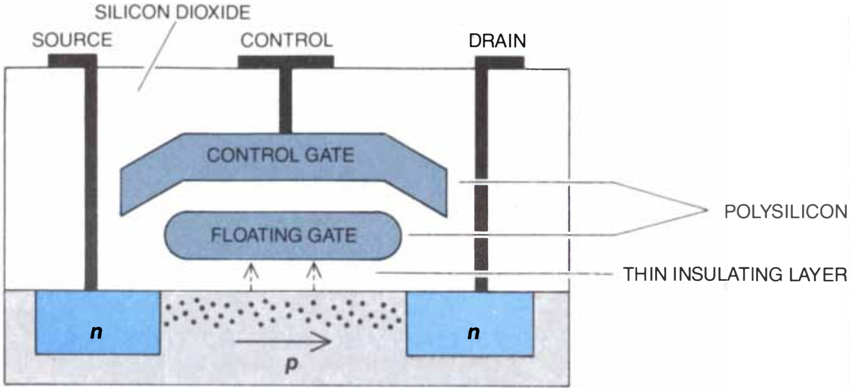
Until recently the only technology out of which this third kind of memory could be built was the so-called erasable, programmable, read-only memory, or EPROM. This is somewhat of a misnomer: information entered into an EPROM can be erased only by exposing it to ultraviolet light under controlled conditions. The cardholder must therefore return the card to the supplier in order to erase the memory in such a way that it can be reused. A card that relies on EPROM technology can therefore serve for only a limited period, after which no more memory space is available for storing such information as the date and amount of a purchase. Ideally a cardholder should be able to erase the memory of a previous year's transactions to make room for the next year's.

A novel type of memory device, known as the electronically erasable, programmable, read-only memory, or EEPROM, may soon make that possible. An EEPROM cell is built in a single crystal of insulating silicon dioxide that lies on a substrate of *p*-doped silicon. Just under the top surface of the crystal there is an island of polysilicon (a conductor) called the control gate [*see illustration on opposite page*]. The control gate is electrically connected to a control electrode that lies above the silicon dioxide crystal. Under the control gate is another island of polysilicon called the floating gate, which is electrically isolated from all the other components of the EEPROM cell. The floating gate is separated from the substrate by a thin layer of silicon dioxide. Within the substrate, on each side of the region that lies directly under the floating gate, are the source and drain, two islands of *n*-doped silicon.

The EEPROM cell relies on a quantum-mechanical property of electrons called tunneling. According to quantum-mechanical rules, when an electron is present in a conductive material that is separated from another conductive material by a thin insulating layer, there is a small probability that the electron will tunnel through the insula-



COMPLEMENTARY MOS (CMOS) SWITCHING DEVICE consumes much less power than a standard *n*-MOS or *p*-MOS switching device and may be the most appropriate technology for smart cards. In a CMOS device the gates of a *p*-MOS transistor and an *n*-MOS transistor are wired together; the voltage applied to them is the input signal. The actions of the two transistors will thus be complementary: any voltage that activates the *n*-MOS transistor will inactivate the *p*-MOS transistor, and vice versa. In addition the drain of the *n*-MOS transistor is connected to the source of the *p*-MOS transistor; the voltage at this node is the device's output signal. The source of the *n*-MOS transistor is connected to a low-voltage line, and the drain of the *p*-MOS transistor is connected to a high-voltage line. When the input voltage is positive, current can flow through the *n*-MOS transistor but not through the *p*-MOS transistor; the output node, which is thereby electrically connected to a low-voltage line and isolated from a high-voltage line, assumes a high voltage. Conversely, when the input voltage is negative, the output node is connected to a high-voltage line through the *p*-MOS transistor and isolated from a low-voltage line; it therefore assumes a high voltage. CMOS devices consume very little power because no current flows through the device except during the short period when the input voltage is changed from high to low or vice versa.



MEMORY CELL, part of an electrically erasable, programmable, read-only memory (EEPROM), retains information stored in it even after it has been disconnected from a power source. The presence or absence of electrons in an electrically isolated "floating gate" (*center*) signifies whether the cell is storing a 1 or a 0 binary bit. To draw electrons into the floating gate, a small positive voltage is applied to the drain and a large positive voltage is applied to the control gate. The layer of electrons thus attracted to the area under the floating gate forms a small current between the source and the drain. Some of these electrons, attracted by the positive voltage on the control gate, "tunnel" through a thin insulating layer of silicon dioxide to the floating gate, where they are trapped. To remove them, a large negative voltage is applied to the control gate while a small positive voltage is applied to the drain. Electrons then tunnel from the floating gate to the *p*-doped substrate, where they form a current into the drain. An EEPROM cell can be programmed and reprogrammed in this way many times.

tor and become trapped in the second conductor. In the case of an EEPROM, electrons tunnel between the *p*-doped silicon and the polysilicon floating gate. It is the presence or absence of extra electrons within the floating gate that signifies whether an EEPROM cell is storing the binary bit 1 or 0.

When there is no voltage applied to the source, drain and gate, there are no electrons available to tunnel. To produce them a small positive voltage is applied to the drain and a large positive voltage is applied to the control gate. The source, drain and control gate then behave like the source, drain and gate of an *n*-MOS transistor: a layer of electrons is attracted to the surface of the *p*-doped substrate, and a current flows through this layer from the source to the drain. Some of the electrons that make up the current tunnel through to the floating gate and are trapped there. When the voltages on the drain and the control gate are removed, the electrons remain in the floating gate and the EEPROM stores a logic-1 bit.

To remove the electrons from the floating gate a small positive voltage is applied to the drain while a large negative voltage is applied to the control gate. Electrons are then repelled by the control gate and tunnel back into the substrate, where they form a current between the source and the drain.

An EEPROM cannot be programmed and reprogrammed indefinitely. Every time electrons tunnel between the substrate and the floating gate a few are trapped within defects in the thin insulating layer of silicon dioxide. After about 10,000 cycles (that is, after the EEPROM has been reprogrammed from 0 to 1 and vice versa 10,000 times) enough electrons are caught to form leaky channels through which current can flow directly from the floating gate to the substrate. Nevertheless, a smart card equipped with an EEPROM memory will be sufficiently durable to serve for several years and will have enormous flexibility of function.

The smart card may effect major changes in the economic functioning of society. As keys, for example, smart cards may provide the degree of security necessary to make true computer networking a viable proposition; a complete data-bank-exchange and electronic-funds-transfer network can exist only if there is a way to ensure that no unauthorized users have access to the system. Smart cards may bring about this true union between telecommunications and computing.

Smart cards will also change the way simple commercial transactions are carried out. It has been quite some time since the system of barter, of trading one commercial item directly for another, was replaced by a system of trading items for standard units of wealth (say shells or pieces of gold). Eventually the units of wealth themselves were replaced by receipts: certificates that represented a certain amount of gold or silver. Those certificates have now been replaced by cash and checks. Not too far in the future paper money itself may be replaced by "units of purchasing power" stored electronically on smart cards.

Early Arctic Settlements in North America

Survival of the ancestors of the modern Eskimos and Aleuts hinged on the relation between culture and resources. One key to success appears to have been the capacity to exploit many sources of food

by Jean S. Aigner

The Arctic fringe of North America, stretching from the western coast of Alaska to Greenland, contains some of the least promising as well as some of the richest regions ever to be inhabited by human beings. Much of the northern land is tundra, a treeless plain whose surface is frozen for most of the year. Some parts of the long, northern coastline are blocked by ice for as many as nine months at a time. Yet maritime grasslands and wet forests characterize the south, where no ice forms on the ocean. When European explorers reached the Arctic in the 16th and 17th centuries, they found 90,000 hardy people whose way of life was well suited to their coastal habitats. These northernmost native Americans are the Eskimos and Aleuts, and their success in the Arctic rests on more than 10 millenniums of economic and cultural development.

Reconstructing that long prehistory by means of archaeological investigation is not an easy task in the vast Arctic region. Enough is known, however, for archaeologists and anthropologists to depict several crucial episodes in Arctic prehistory. This article takes up four of them. The first episode began about 9,000 years ago when a group of people, who may have been the ancestors of the modern Aleuts, settled on Anangula Island in the southern part of the Aleutian chain. The second took place 5,000 years later when people whose culture is called Pre-Dorset migrated into the eastern Arctic. The third, some 2,000 years ago, was the settlement in Alaska at Point Hope of the Ipiutak people, whose subsistence technology was evidently historically related to that of the Pre-Dorset people. The final episode began in about 985 B.P. (before the present), when the eastern Arctic was resettled by people of the Thule culture. By that time descendants of the Pre-Dorset were not so numerous and the adaptively superior Thule way of life became established across the region. Also, the Ipiutak culture was gone from Point Hope.

Three of these four episodes represent pioneering efforts. The village on Anangula was the first community known to have practiced the coastal way of life that is typical of Eskimo and Aleut culture. It is significant that the settlement was in the climatically more equable and rich region south of the Arctic zone. The small Pre-Dorset bands were the first human inhabitants of the eastern Arctic. The efficient whaling technology and innovative social organization of the Thule people provided much of the foundation of modern Inupiaq Eskimo culture. The fourth episode, the occupation of Point Hope by the Ipiutak people, was chosen not because it was revolutionary but because it demonstrates how rich Arctic culture could be in a favorable location. When the four episodes are combined, they begin to yield a picture of how human beings adapted to a highly variable habitat: the Arctic.

The native inhabitants whom the European explorers encountered in the Arctic made up a complex pattern of linguistic groups and modes of subsisting. Of a population totaling 90,000 at the time of European contact about 16,000 were Aleuts, who occupied the Aleutians and the adjoining part of the narrow finger of land called the Alaska Peninsula. The remaining 74,000 were Eskimos, who in turn were divided into two main linguistic groups. The many speakers of Alutiiq lived on the southern coast of Alaska. Even more numerous were related speakers of Central Yupik, who lived on the western coast of Alaska; their linguistic relatives who spoke Siberian Yupik inhabited islands in the Bering Strait and small enclaves on the Siberian coast. The other linguistic group, which was more widespread (although not more numerous), was made up of the speakers of Inupiaq, who occupied a territory stretching from northern Alaska all the way across northern Canada to Greenland.

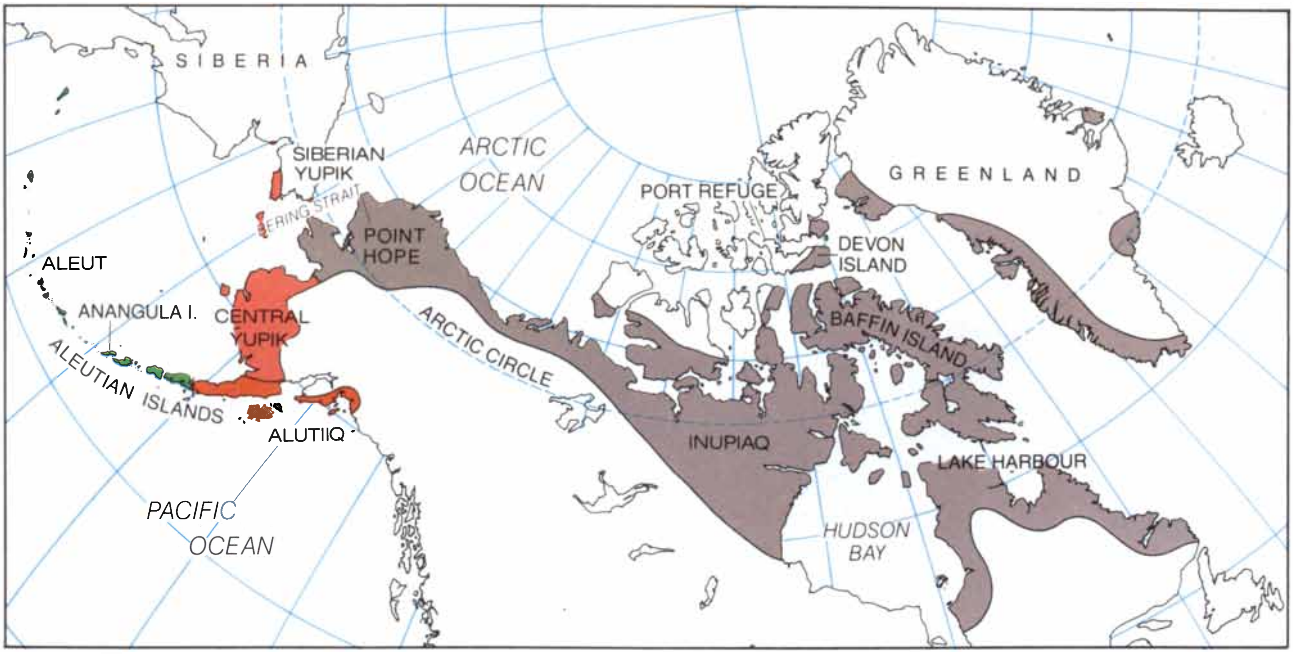
In spite of such language and cultural divisions all the Eskimo and Aleut groups shared an Asian origin. During the Wisconsin ice advance, which reached its peak between 18,000 and 16,000 years ago, so much water was frozen in the glaciers that the mean level of the world's seas fell by about 100 meters. One result was that eastern Siberia and Alaska merged into a single landmass called Beringia. The surface of Beringia probably consisted largely of tundra. Along the coast and perhaps across the frozen plain the ancestors of the Aleuts and Eskimos gradually expanded southeastward toward Alaska.

Among the people who came to Alaska by way of Beringia were the forebears of the groups that established the community at Anangula. They probably traversed the southern coast of Beringia. Although it is not easy to define the outlines of Beringia precisely, the southern coast may have corresponded to a line that extended southeast from Siberia toward the Aleutians. Thus a traverse of the coast would have terminated somewhere in the eastern Aleutian chain. During the height of the Wisconsin glaciation the Aleutians were under ice, along with much of southern Alaska. By about 11,000 B.P., however, the ice had melted and opened the islands for human habitation.

The journey along the coast of Beringia to the eastern Aleutian Islands undoubtedly took millenniums, but it was by no means time wasted. Indeed, the journey may have served as a laboratory for perfecting subsistence skills suited to Anangula. Unlike northern Alaska, the Aleutians have a sub-Arctic climate. The coastline is foggy, rainy and free of ice even in the winter. As the migrants moved south and east along the coast of Beringia they would have encountered increasingly warm and wet conditions. By the time they reached the Aleutians they were quite likely well prepared to survive in an ice-free coastal environment. The long learning experience was clearly valuable, because when my students and I excavated Anangula in 1970, we found the remains of a vital, robust village.

The prehistoric village site is a point of land extending from the tiny, treeless island of Anangula into Nikolski

BURIAL MASK was made about 1,500 years ago by people of the Ipiutak culture. Each of its parts was carved from a piece of walrus tusk. The parts may originally have been held together by sinew thongs. In modern Eskimo culture great power and ritual significance are attributed to masks, and the same may have been true in the Ipiutak culture. The mask came from a grave at Point Hope in northwestern Alaska. Ipiutak people are known to have been present there by 1600 B.P. (before the present). Several hundred years later they were gone, but they had probably contributed to aspects of modern Bering Sea Eskimo cultures and populations. At Point Hope the Ipiutak community was succeeded by groups of people who were probably the ancestors of the modern Inupiaq Eskimos.



ARCTIC REGIONS of North America are home to the modern Aleuts and Eskimos, who are divided into several major linguistic groups. The deepest division is between the Aleuts and the Eskimos. The Aleuts live in the Aleutian Islands and on the adjoining part of the mainland. Another linguistic division is between Inupiaq Eskimo and the group of Alutiiq, Central Yupik and Siberian Yupik Eskimo languages. Alutiiq speakers live on the southern coast of Alaska. Speakers of Central Yupik live on the western coast. Speakers of Siberian Yupik live on islands in the Bering Strait and in eastern Siberia. The most far-flung group is made up of the speakers of Inupiaq. Their territory, which is bounded on the west by northwestern Alaska and on the east by Greenland, includes Canada.

IPIUTAK GRAVE ORNAMENTS are among hundreds found at Point Hope; much of what is known about Ipiutak intellectual culture comes from findings made in graves. The objects shown here are made of carved walrus tusk. The cone-shaped objects at the upper left are artificial eyeballs with jet pupils. They were made to be inserted in the eye sockets of a skull. The object shaped like a figure eight at the upper right is part of an ornamental chain link; the small projection might represent a frog. The pointed object in the middle was lashed to some other object; it may represent the head of a bird. The head at the right of the three-link chain could be that of a walrus. These objects, along with the mask on the preceding page, are from the American Museum of Natural History.

Bay. Radiocarbon dating, supplemented by other evidence, shows that the site was occupied for several hundred years between 8750 and 8250 B.P. The village consisted of many small oval dwellings, each occupied by a single family. The people built their houses by digging an oval depression over which they raised a driftwood superstructure. The superstructure was covered by matting and a roof of living sod. Entry was through the roof. Artifacts found in and around the houses imply a division of labor between the sexes, much like the one that exists among Aleuts and Eskimos in the historic period (after contact with Europeans). Men and boys apparently spent much time on the house roof, manufacturing objects of stone, bone and wood and scanning the sea for sea mammals. Women and girls worked inside the house, making clothing, grinding pigments and tending the stone lamp in which they burned sea mammal fat to provide heat and light.

Even though men and women performed different tasks, each household carried out basically the same economic activities. Statistical analysis of the tools and by-products left by the manufacture of stone tools shows that there was little specialization among households; each one apparently made all the tools it needed. The size and contents of the homes imply that all families lived in about equal comfort. Furthermore, because the sub-Arctic maritime climate provides a wide variety of resources, food and other necessities were most likely available to all households. Rather than being integrated in a social hierarchy based on the ownership and distribution of scarce resources, the community at Anangula seems to have been largely egalitarian, perhaps because a living had to be made from the coastal zone and the sea demanded cooperation.

Indeed, the shoreline and sea were the only substantial sources of food on the Aleutian Islands. The treeless land supports few edible plants. Berries and greens may have provided a bit of dietary variety, but they could not have been the source of many calories. The species of grasses available on Anangula and other islands yield useful raw material for making containers, baby cradles and even socks, but they cannot be eaten by human beings. The sea and the beach yield a rich harvest. Resident Aleutian sea mammals, such as hair seals and sea lions, and migratory ones, such as fur seals and some whales, were the staples of the diet. Their flesh was supplemented by that of shorebirds and seabirds, shellfish, shallow-water fishes, deep-sea fishes and anadromous fishes such as salmon (a species that migrates annually from fresh water to salt water).

Considering how such prey were probably hunted, the size and permanence of Anangula take on particular significance. The modern-day Aleuts tend to hunt alone, in pairs or in small groups, but rarely in large, highly organized parties. The early residents of Anangula probably did the same. Such a hunting method does not require substantial social units. Continued survival, however, does require that a group achieve a certain minimum size. For example, two families acting as a unit would be extremely vulnerable to the death or maiming of their male hunters in a single accident. Hence a core group needed at least from six to eight families to ensure consistent success in the pattern of subsistence.

The number of houses and the distribution of other artifacts at Anangula suggest that roughly 75 people lived there, or twice the minimum number called for by the style of hunting. If a community of that size is to survive through recurrent periods of scarcity, it must have highly developed hunting techniques. Not only were the islanders' hunting techniques sound enough to support a large community but also they were reliable enough to support a permanent one. When game is scarce, one obvious strategy is to pick up the village and move. The people of Anangula, however, were such skilled hunters, and locally available resources were so abundant, that the villagers did not have to resort to such measures. Digging at later, nearby sites revealed a pattern with a permanent village like Anangula, a seasonal salmon-fishing camp and also temporary camps where pairs of hunters slept overnight. Although the main village on Anangula was apparently occupied continuously, people probably also followed a pattern using optional seasonal camping sites.

The vitality and permanence of Anangula are all the more impressive when one considers that it is the earliest coastal site in Alaska bearing a direct relation to ways of life in the modern period. Traditional Aleuts subsist by hunting sea mammals and fishing, a way of life similar to that of the prehistoric villagers on Anangula. Although not all scholars would agree with me, I think the community on Anangula was inhabited by the cultural ancestors of the modern Aleuts. A series of other, younger sites in the Aleutians appears to provide the intermediate links between Anangula and the Aleutian culture and people in the historic period.

The prehistory of the Eskimo people is somewhat more complex than that of the Aleuts, for several reasons. Not only are there many more Eskimos than there are Aleuts but also the Eskimo populations include groups with quite varied subsistence practices. In addition the antecedents of Eskimo culture do not form a single, clear line of descent as the antecedents of Aleut culture appear to do. Instead there are several false starts and new beginnings. One early example is Pre-Dorset culture. Although Pre-Dorset culture (really several cultural variants) flourished for a long time in the eastern Arctic, as did its descendant culture, Dorset, it did not continue in recognizable form in the historic period.

Where Pre-Dorset culture originated is a matter of disagreement among scholars. Some experts argue that it arose in Alaska; others say it originated in Siberia and was imported to Alaska by immigrants crossing the Bering Strait by boat. Archaeologists generally agree, however, that between 4300 and 4000 B.P. Pre-Dorset people expanded from northern Alaska toward the eastern Arctic, which until then was uninhabited. In their movement eastward they may have been abetted by a temporary warming phase in the Arctic climate. The relatively warmer conditions and the resulting changes in the patterns of coastal ice increased the population of sea mammals, the main source of food for Pre-Dorset communities.

Among the sea mammals the most important for Pre-Dorset hunters was the small ringed seal, which they hunted several times in the year. Ringed seals breed in the spring, and the Pre-Dorset people sought them as pups in their dens on the ice; they also hunted the adult seals as they basked in the late-spring sun or swam near the edge of the ice floe. Some groups probably kept dogs as hunting partners, perhaps to help find the pupping dens and possibly the winter breathing holes of seals as well. Some archaeologists suggest that Pre-Dorset groups focused on seals and that even in places where waterfowl and caribou are known to have been common they appear as relatively minor dietary elements. Fish and shellfish are often absent altogether. Animal bone remains are often rare in such sites and many seasons are almost entirely unrepresented. Therefore archaeologists disagree about the degree of focus on single items that actually characterize Pre-Dorset economies. Nevertheless, a current characterization of Pre-Dorset culture suggests it focused on one or two food sources (seals and caribou) and knew fewer hunting strategies than modern Eskimos do. Pre-Dorset descendants were at a competitive disadvantage when later (Thule) groups expanded into the eastern Arctic.

As the Pre-Dorset people moved east they colonized both the high Arctic (north of 75 degrees north latitude) and the low Arctic (south of 75 degrees). Two sites, one in each zone, give a sense of how they adapted to their environment. Port Refuge on Devon Island, not far from Greenland, lies north of 76 degrees and near the margin of human survival. The coastline is blocked by ice for at least nine months of the year. Yet local variations in air and water temperature and ocean currents produce open water areas called polynyas that are ice-free most of the year. Port Refuge Pre-Dorset groups positioned themselves on the coast near the polynya to hunt the sea mammals swimming through the open water. Excavations near Port Refuge by Robert McGhee of the National Museums of Canada offer a glimpse of their way of life.

Port Refuge was much too harsh a habitat to support a large, sedentary community such as the one on Anangula. The resident population was small and often moved in search of food. The Pre-Dorset people apparently had a variety of house types at sites they visited according to the season, length of stay and the availability of food. From summer into the fall the group sometimes visited a location known as the Cold Site, where the remains of 31 houses, tents and workshops for making tools have been found. Bits of food, mostly ringed seal, indicate that tents and workshops were occupied for a day or two at a time by families or small hunting parties. The houses were more substantial and were occupied by families for periods ranging from a week to three months.

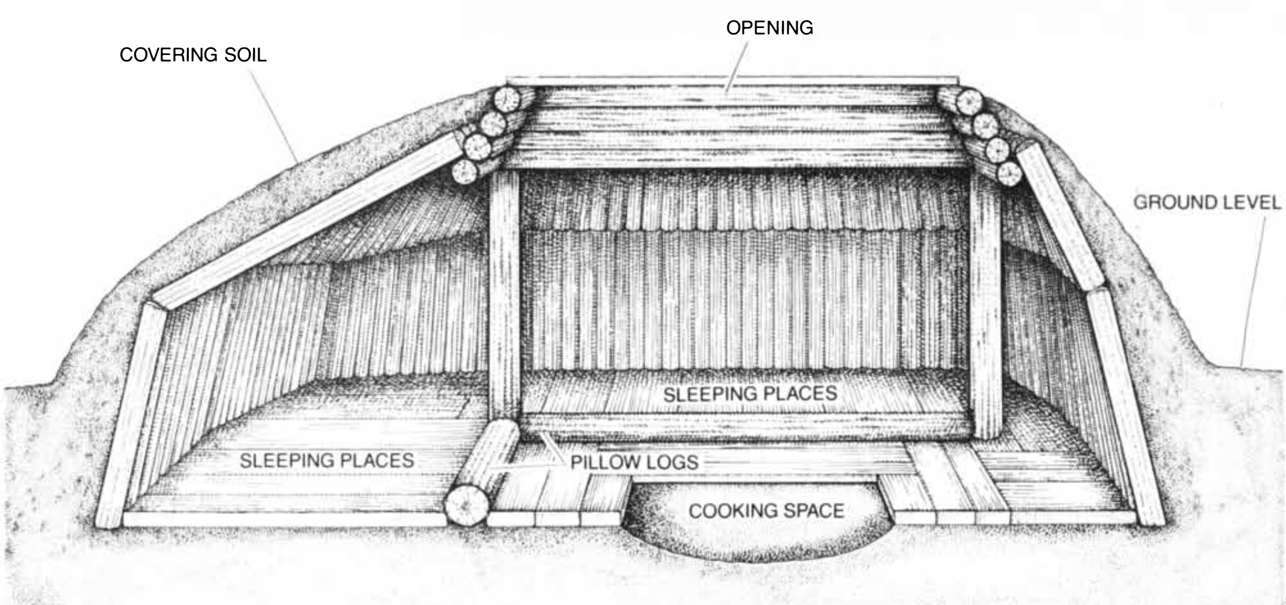
The typical Pre-Dorset house consisted of a shallow foundation with a driftwood frame. Skins were stretched over the frame and held in place at the bottom by gravel or stones. Large slabs of rock in parallel rows marked the central hearth and the work areas. The house generally faced the sea; the interior space was divided according to sex roles. On the left as one looked out to sea the woman sat sewing or tending the hearth. The man sat on the right and made stone or bone tools.

The Cold Site was not the only locale the group members visited in the summer. During the warm season Port Refuge families sometimes lived in houses or tents four kilometers away at the Upper Beaches site. In the spring some might camp five kilometers to the southeast at Cape Hornby, obtaining fresh water from snow melt as they hunted seals and did domestic chores. The absence of sites giving evidence of winter activities suggests that in the cold season the group camped on the sea ice in snowhouses. Pre-Dorset people are credited with this innovation.

The initial tenure of Pre-Dorset people in the high Arctic was brief. By about 3500 B.P. the warming trend that had accompanied their migration to the eastern Arctic was reversing. The cooling was sufficient to make the high-Arctic regions less attractive to people with Pre-Dorset styles of adaptation, and they withdrew. In the more hospitable low Arctic, however, they continued to live with considerable success. Archaeologists suggest that in the Lake Harbour region of Baffin Island from five to 10 Pre-Dorset bands with a total of about 150 members spread themselves along about 450 kilometers of coastline and survived there for many hundreds of years.

Work done by Moreau S. Maxwell of Michigan State University and his students suggests how the thinly dispersed groups might have lived. These ideas are subject to revision, of course, because Pre-Dorset sites with fauna are uncommon, and sites with bones are biased toward spring at the floe edge. Because of the local ice and tide conditions the resource base was modest but reliable. Ringed seals and the larger bearded seals were in the area year round. Walruses, beluga whales and even the larger bowhead whales were present during warm climatic intervals; fish were available. The land supported caribou and birds. From this variety, it is suggested, Lake Harbour Pre-Dorset people, and to a lesser extent descendant Dorset people, were inclined to focus on small seals and caribou. They were willing to shift their settlements as needed to take seal and caribou with the strategies and tools they had long known.

Some archaeologists believe that the

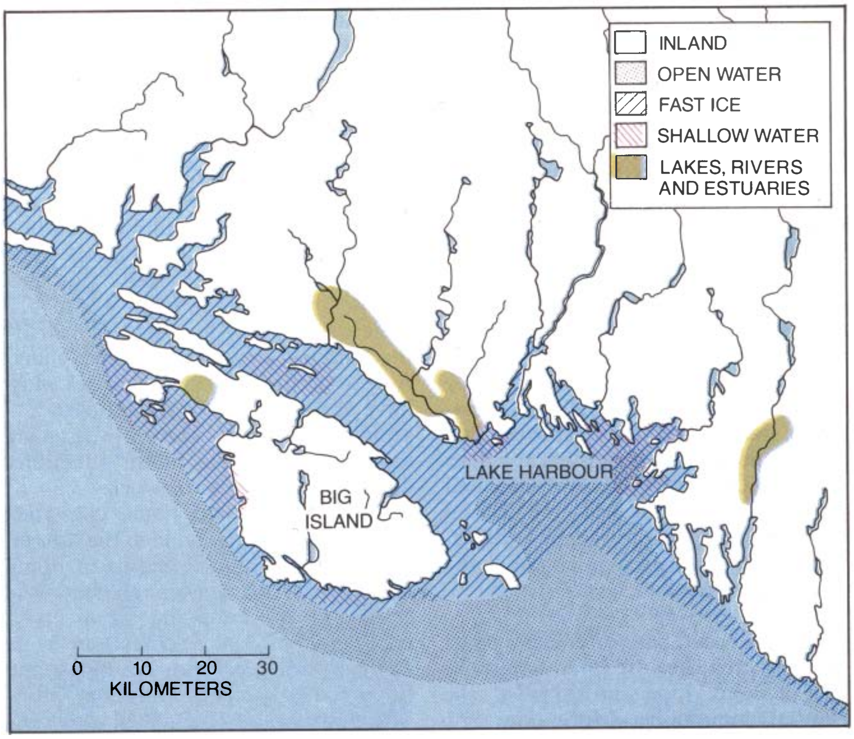


IPIUTAK HOUSE had a roughly square plan and may have been entered through a side passage in some cases. The house was set slightly below ground level. It had a log frame, log or perhaps plank flooring and walls and a covering of earth. Each house was probably occupied by one family, but examples suggest differences in family size and membership. Cooking was done over a fire in a central pit below floor level. The family members slept and worked on low benches around the periphery of the house. The reconstruction is based both on excavations at Point Hope and nearby Cape Krusenstern and on 19th-century sketches of modern Eskimo houses.

focused specialization of the Pre-Dorset signifies a culture that had both a limited technology and set of hunting strategies and a world view that strongly favored traditional ways of living. Such a traditionalist and rigid ideology might have served to discourage socioeconomic and technological innovation. Hypotheses concerning the intellectual culture of prehistoric peoples are slippery at best, but it is true that Pre-Dorset hunters lacked certain technologies that were of great benefit to later Arctic dwellers. For example, although they had dogs, we have no evidence for dog sleds. As a result the hunters may not have been able to travel quickly from a land settlement to the hunting ground during the spring sealing season and may have been forced to build snowhouses on the ice to be near the seals.

We also lack evidence for techniques that make open-water hunting of sea mammals highly efficient, such as winter sealing at breathing holes and those involving drag floats. Such methods have had a significant part in the adaptive success of modern northern Eskimos. Thus a way of conceptualizing differences between Pre-Dorset and historic Eskimos of the eastern Arctic is to place them along a continuum from a focused economy to one that is diffuse, and from a rigid world view to one that is flexible. Pre-Dorset culture focused on ringed seals and caribou. Descendant Dorset culture was more inclined to take some walruses, small whales and small game. But the techniques for taking these animals were more limited than the wider repertory used by historic Eskimos, who were very flexible, adapting as needed to both new settlement areas and economic strategies. The range of foods taken varied with time and place, as did methods of hunting and group organization.

The Pre-Dorset culture underwent enough changes for archaeologists to recognize a distinct culture known as Dorset. Although by 1485 B.P. Dorset sites were fewer than before and the total population appears to have been smaller across the eastern Arctic, we cannot say Dorset groups disappeared. There is enough evidence to tell us that some Dorset people were present in the eastern Arctic in 985 B.P. when Thule Eskimos moved into the area. Dorset culture as such disappeared soon thereafter in most areas, and in this sense it may be said to have failed. Nevertheless, without contact with a more dominant Eskimo culture such as Thule, Dorset might have continued in the region. That it did not suggests several scenarios. As a result of competition many groups may have been forced into marginal areas and have died out. Others may have become acculturated, taking on Thule techniques and strategies, and so have become invisible as Dorset to later archaeological inquiry.



LAKE HARBOUR REGION of Baffin Island was inhabited in succession by members of two prehistoric cultures: the Pre-Dorset (and descendant Dorset) and the Thule. The map shows the areas where the people of the Thule culture hunted their major types of prey. In winter, when the coast is icebound, seals were hunted at their breathing holes in the "fast," or solid, ice and also along the edge of the floe. Then and at other times sea mammals such as walruses and whales were hunted in open water. The shallow water and wetlands provided ducks and other waterfowl. The lakes, rivers and estuaries were a source of fish. Inland the Thule people hunted caribou and other game. As the map suggests, the Thule community was quite flexible in finding alternative sources of food. In contrast, the earlier Pre-Dorset community was focused on small seals and caribou as primary food sources. Whereas only 150 Pre-Dorset and Dorset people survived at Lake Harbour, about 250 Thule people were able to live there for more than 1,000 years. This map and the diagrams on the next two pages are based on the work of George Sabo of Michigan State University.

The final word has yet to be written about Pre-Dorset and Dorset in the eastern Arctic. Pre-Dorset people survived for many hundreds of years in some of the harshest habitats in the world and evolved into groups with Dorset culture. Dorset in turn had a long history before being submerged by the newly dominant Thule culture from which the modern Inupiaq people developed. While Pre-Dorset and Dorset regional groups occupied the eastern Arctic, Eskimos with distant historic connections evolved different cultures in the western Arctic. Furthermore, in areas of unusual richness those cultures were capable of great artistic achievement and social complexity on an initial technological base similar to that of the Pre-Dorset. An example of such achievement is provided by members of the Ipiutak culture in northwestern Alaska.

Remains found at Point Hope show that an Ipiutak community was present there by about 1600 B.P. Ipiutak culture was one form of Bering Sea Eskimo coastal culture that prevailed in northern Alaska at that time. The coastal culture appears to be related historically to the Arctic Small Tool Tradition, which is the name given by Alaskan prehistorians to archaeological remains that bear technological similarities to Pre-Dorset. The Ipiutak people of Point Hope can be thought of as the descendants of those who stayed behind when the forebears of the Pre-Dorset people began their migration to the eastern Arctic in about 4000 B.P.

For millenniums Eskimos of various cultures have been drawn to Point Hope because of the quantity and variety of food available there. The Ipiutak community possessed roughly the same hunting gear as their Pre-Dorset and Dorset counterparts to the east. Because their environment was rich and possibly because the community's hunting strategies differed from those of the Pre-Dorset, the diet of the Ipiutak was far more varied. A count of the animal bones from the Ipiutak site shows that ringed seals, walruses, bearded seals and caribou were taken in a ratio of about 4 : 2 : 1 : 1. Although bone counts cannot be translated into actual numbers of animals, we know that the bearded seal and the walrus are respectively four and eight times larger than the ringed seal, and they provided most of the meat and fat eaten in the village. Land species that were hunted nearby, possibly for their fur as much as for their meat, included polar bear, wolf, fox and even ground squirrel in addition to caribou. Migratory birds and fish supplemented the mammalian catch.

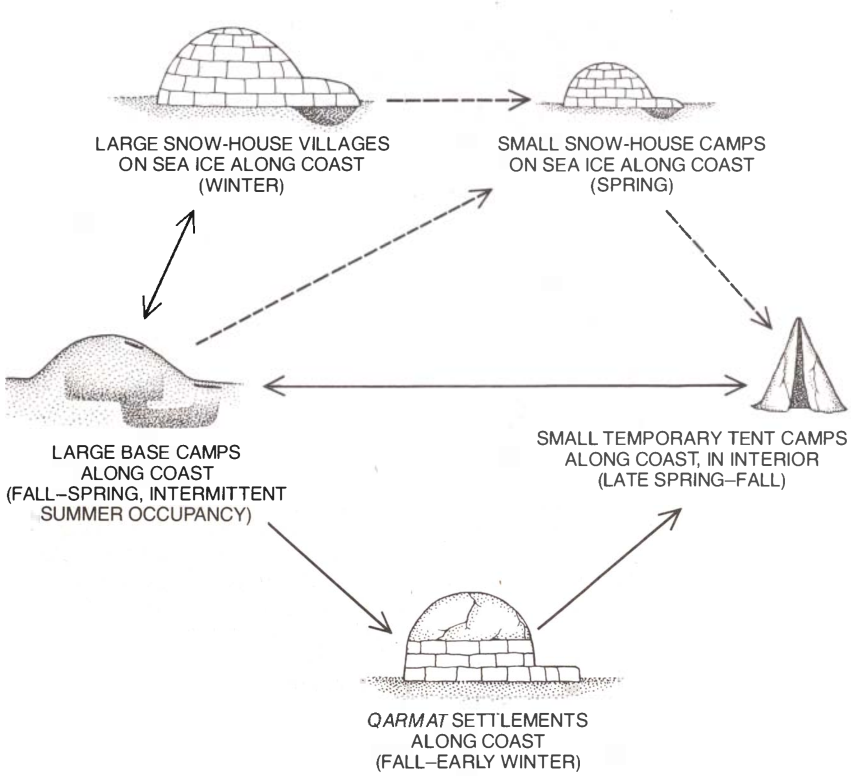
On this broad base of protein-yielding creatures the Ipiutak population of Point Hope constructed a substantial community. More than 575 houses were built on five low beach ridges that fringe the triangular point. Of course, not all of them were occupied at the same time. Based on density patterns typical of modern Eskimo villages I have estimated that at any given time between 75 and 100 people may have lived at Point Hope. Variation in house size suggests that within this community there may have been fairly large differences in wealth and status. About half of the square houses were from 3.5 to 4.5 meters on a side and another fourth were from 4.6 to 5.5 meters on a side. The rest were distributed among larger and smaller plans. A few of the largest structures may have corresponded to the modern Eskimo men's house, known in Inupiaq as *qargi,* that serves public functions and provides space for work.

The quality of the house construction and the goods found in the houses show that the Ipiutak people of Point Hope lived well. Perhaps the best evidence for the emerging social complexity of Point Hope comes not from objects associated with the living but from those accompanying the dead. The Ipiutak graves are in separate cemetery areas on the point. Since the 1930's, when Helge Larsen of the Danish National Museum and Froelich Rainey of the University of Alaska began the work, about 130 graves have been excavated and from them has come a trove of superbly carved ivory ornaments and ritual objects.

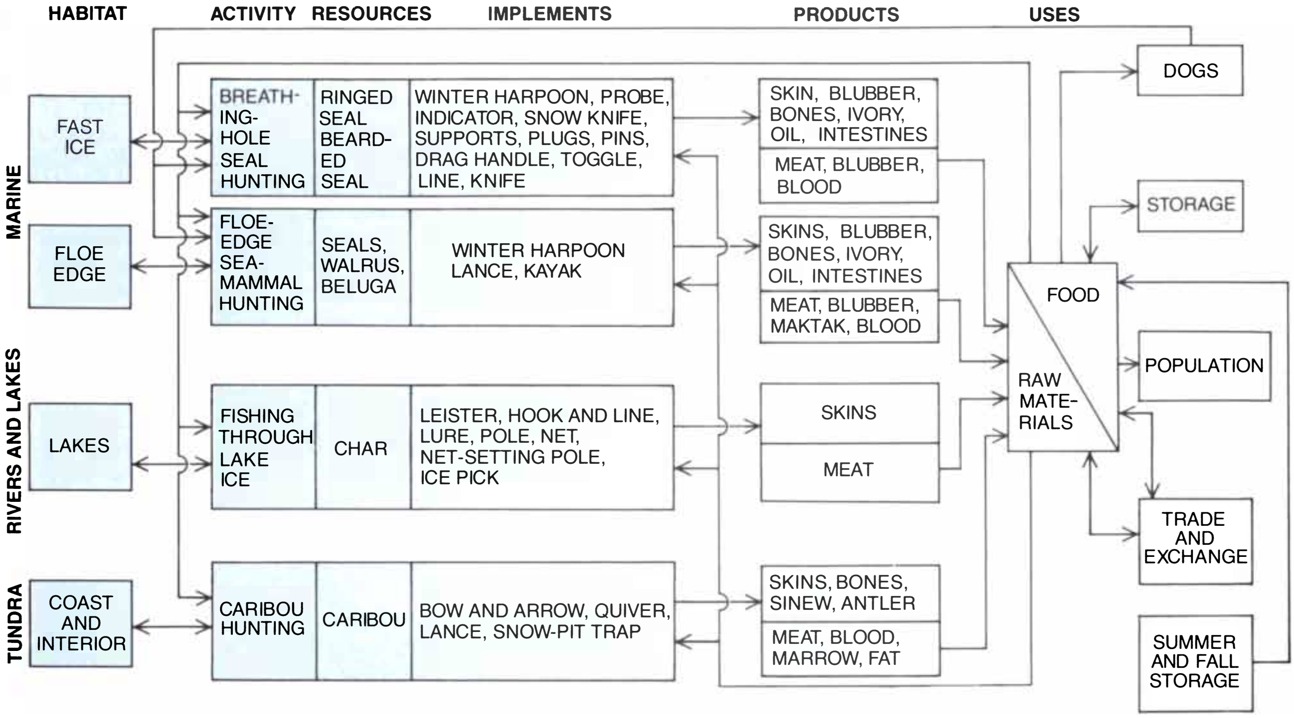
Of the Ipiutak graves at Point Hope, 59 are log tombs built into the ground. Within the tomb the body was generally laid on its back; the hands covered the pubic region and the head pointed west. Such a burial form is rare in the Eskimo world. About half of the 59 tombs contained artifacts, most of which were utilitarian objects such as those found in the houses. The remaining nonutilitarian artifacts include some striking finds. Several tombs held masks that consisted of multiple parts carved from walrus ivory. Some of the skulls were elaborately decorated with carved ivory. Objects resembling eyeballs, with pupils made of jet, were inserted into the eye sockets. Small carved birds were put into the nostrils. The mouth was hidden by an intricately carved ivory cover. Masks, which have considerable ceremonial significance among modern Eskimos, and skull decorations suggest a highly developed system of religious beliefs and ceremonies.

The 70 remaining Ipiutak graves were "midden burials" in which the body was laid out in a former living area on the earth or under a log frame. Some midden corpses were found simply laid out flat, but others had sometimes been dismembered for burial. Only in the midden burials have investigators found exotic ivory swivels and link chains beautifully carved from a single piece of walrus tusk or caribou antler. Together the midden graves and log tombs offer some tantalizing hints about the structure of the Ipiutak community. The quantity and quality of the carvings in some graves suggest a separate caste of shamans with a religious function. Objects found in the graves show that the community made distinctions between men's and women's work, in wealth and perhaps also in family standing and occupation. Such findings support the notion that the differences among community members in wealth and status were substantial.

The cultural richness of the Ipiutak community at Point Hope was no guarantee of permanence. Within a few hundred years of 1600 B.P., the date by which the community is known to have been established at Point Hope, it was gone. Perhaps the local Ipiutak people withdrew because of competition from other groups that possessed a more varied technology and a superior adaptive flexibility. Whatever the cause, the displacement



SEASONAL CYCLE of the Thule community at Lake Harbour was based on movements between several different types of houses. Houses in the permanent base camp had a foundation of flagstones and uprights made of whalebone. The frame was covered with sod. Large snowhouses with deep entries to trap the cold were built on the thick pack ice in winter. At times they provided shelter for several families. When the ice began to break up in spring, small snowhouses might be constructed for individual families. In the late spring and fall Lake Harbour Thule people moved through the interior and along the coast, camping in tents made of hide with wood frames. In the fall and early winter they occupied *qarmat* settlements along the coast. A *qarmat* is a tent structure with a foundation of ice blocks.



DYNAMIC ECONOMY of the Thule community at Lake Harbour is shown schematically in this diagram. The Thule economy was based largely on sea mammals: the ringed seal and the larger bearded seal, the walrus and the beluga whale. Thule hunters supplemented the sea-mammal kill with caribou, birds and fish. The varied base of prey species provided food for the Thule group, for their dogs and for storage. One fraction of the food eaten by human beings contributed to reproducing the population; another fraction provided nourishment for further hunting. The carcasses of dead prey also provided the raw materials for equipment, clothing and houses. The adaptive flexibility of the Thule enabled them to survive various climatic changes. The Thule are the progenitors of the adaptationally and culturally varied modern Inupiaq-speaking Eskimo populations who now inhabit much of the northern Arctic.

of the Ipiutak community from Point Hope presaged a much larger cultural episode that ultimately brought the innovative culture known as Thule to much of northern Alaska and the eastern Arctic.

The Thule culture's antecedents are found on the rich northeast Siberian and the Bering Strait islands. Here the ancient Eskimos developed an effective technology for hunting whales, walruses and bearded seals in open water. Among the most important of their tools were the throwing board and the drag float. The throwing board is a launcher that enables a hunter to impart considerable distance and momentum to a spear or harpoon. The drag float is an inflatable sealskin. When the float is attached to the line of a harpoon lodged in a sea mammal, it exhausts the whale and keeps it from diving. Ultimately some of the ancient Eskimos who possessed this technology moved from the offshore islands to the Alaskan mainland. These Thule ancestors replaced the Ipiutak community at Point Hope and expanded as far south as the place that is now Nome. In some areas they expanded at the expense of Eskimo groups that lacked the more versatile Siberian approach to technology and weaponry.

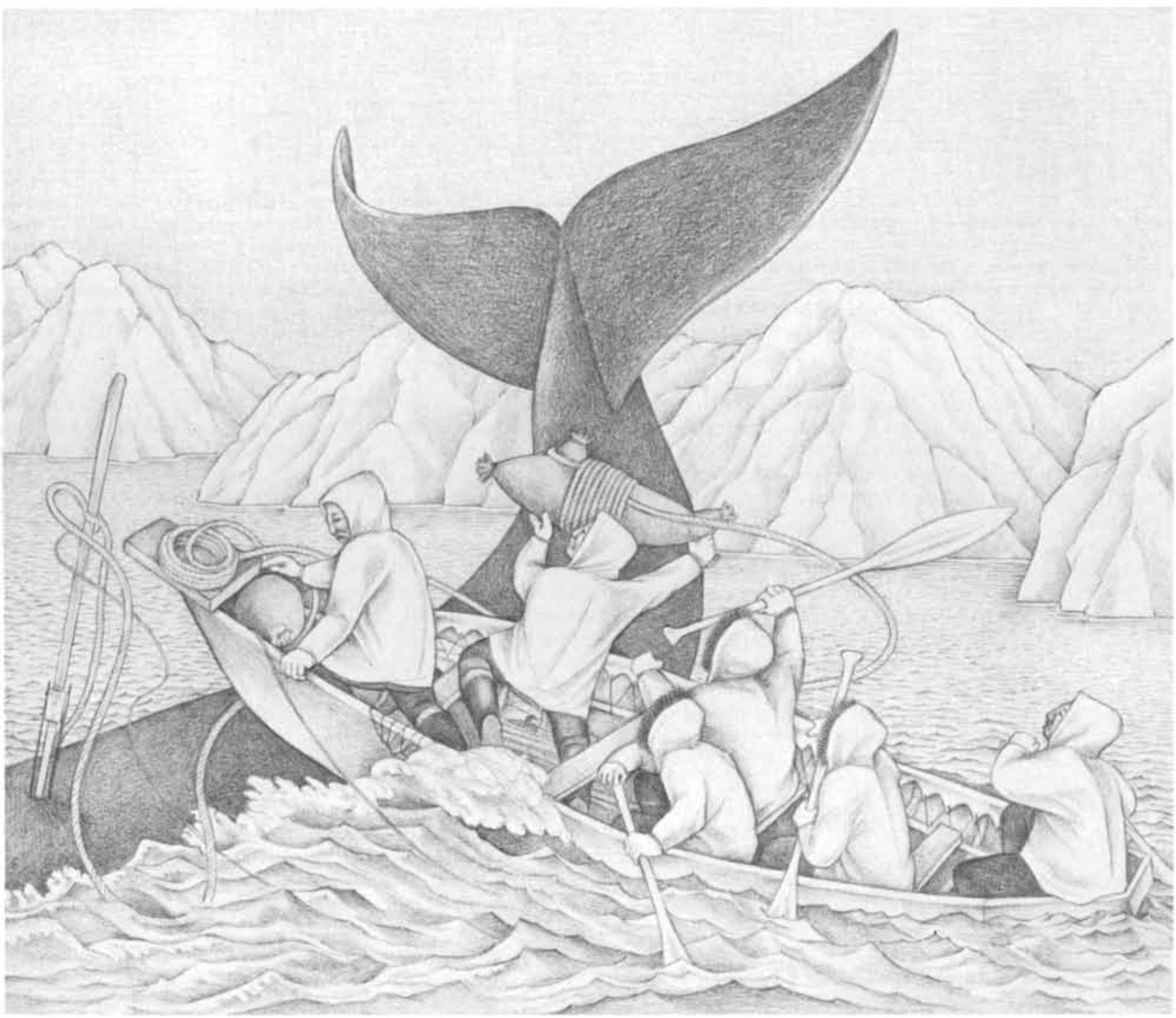
On the Alaskan mainland the ancient Eskimos' cultural development continued. Social forms were added that complemented the new hunting technology. Historic Eskimo whaling cultures suggest that at about this time groups began to organize around whaling captains, called *umialgit* in Inupiaq. A community might have one *umialik* or several depending on its size and prosperity. The *umialik* acquired the capital needed to build, outfit and maintain the whaling boat. He was also responsible for supporting his crew, which might include half a dozen men or more, and their families. Each *umialik* was associated with a *qargi*. In the *qargi* the crew repaired their whaling gear. The *qargi* was also the site of male ceremonial activities and of general ceremonies that included the families of crew members. The collective activities of the community put a heavy emphasis on sharing and on distributing the catch.

When effective methods of whaling were combined with a community structure based on the *umialik*, the result was the fully developed communal whaling culture practiced by Thule people. Thule culture did not remain confined to Alaska for long, although its richest expression remained there. Beginning about 1085 B.P. a warming trend in the Arctic climate extended the range of whales and the population size of walruses and bearded seals to the east. One proposed scenario has Thule groups in about 985 B.P. set out eastward, for reasons that are debated, with their highly efficient open-water hunting methods. Some of the places they settled were then uninhabited.

As the Thule people spread eastward they were aided not only by their highly productive whaling strategies but also by developed techniques for hunting seals in winter, use of dog sleds for transport and other innovations. For ambushing seals at their breathing holes through the ice, the hunter first found a breathing hole with the aid of his dogs. He cleared the snow from the ice above the opening, which he enlarged with an ice pick attached to the butt end of his harpoon shaft. A delicate bone-and-feather mechanism was placed over the hole to detect the seal's breath. The hunter waited patiently for the seal to come up for air, then speared it and pulled it through the ice with a line and handle designed for the purpose. He could sled his kill back home. With these simple, ingenious methods Thule groups were able to extend seal hunting through the winter. In favorable eastern locations Thule groups could store killed quarry by using ice for refrigeration. The surplus left the spring free for hunting whales and walruses.

Perhaps the greatest advantage conferred by Thule culture on its members, however, was not a specific technique but an overall adaptive flexibility. Drawing on a wide range of techniques and a world view that supported settlement moves and changing hunting strategies, Thule groups routinely took a wider range of animals than Pre-Dorset groups and used more locations. Such economic and social flexibility made it possible to maintain the equilibrium of the community under changing environmental conditions. Thule flexibility at several localities in the eastern Arctic has been studied by archaeologists and anthropologists. The most instructive of them is the Lake Harbour region of Baffin Island, where work done by George Sabo, a student of Moreau Maxwell's at Michigan State, demonstrates the superiority of the Thule adaptive strategy there. The region sustained a scattered Pre-Dorset and later Dorset populations of about 150 for several millenniums. Sabo's work shows that beginning not long before 985 B.P., and continuing for almost 1,000 years, the region supported a Thule population numbering 250.

When the Thule groups arrived at Lake Harbour, the warming trend was well under way and the sea mammal population was large and included bowhead whales. The early Thule bands locally practiced communal whaling, using the whale ribs and jaws to build the houses of their winter base camp. Social and ceremonial activities were concentrated in the winter settlement. Although the large bowhead whale was a staple, the Thule community had a varied supply of other game. Leaving aside the bowhead whale, their diet consisted of about 30 percent small seals (such as the ringed seal), 20 percent walrus, 20 percent caribou, 15 percent bearded seal, 10 percent polar bear and 5 percent beluga whale, a



COMMUNAL WHALING was an important factor in the success of Thule culture. The drawing shows a critical moment in the hunt for a bowhead whale. The whale has been harpooned for the first time; the harpoon is visible at the left. Attached to the harpoon line is a drag float (an inflated sealskin), which a crew member is about to throw overboard. Several floats were attached to the whale to exhaust it and prevent it from diving. The drag float was one of several crucial artifacts of Thule whaling technology. The technology was combined with a form of community organization well suited to whaling. The combination of advanced technology and flexible social organization made it possible for Thule culture to spread from its source near the Bering Strait across Canada to Greenland. Some Thule groups lived in places with whale populations; those groups launched whaling boats. Some lived in regions without whales; they concentrated on large sea mammals and other sources of food. The drawing is based on a painting by the Alaskan artist Howard Rock.

more broadly based diet than the diet of the Pre-Dorset people. In particular it includes the meat from several sea mammals in fairly well-balanced proportions.

By about 685 B.P. the warming phase was over. In the Lake Harbour area the increasing cold reduced the whale and caribou populations and increased the number of ringed seals. The Thule bands seem to have been equal to the climatic change. They adjusted to the absence of whales by taking more small seals, which increased to 40 percent of the nonwhale diet, and bearded seals, which increased to 20 percent. The new balance of prey species required changes in hunting style and even some in house design. To obtain enough food the Thule people more often split up into smaller groups, each traveling to a different hunting ground. In colder conditions the edge of the ice (and the accompanying winter food supply) was farther from land. The resourceful Thule groups abandoned the whalebone winter house and evidently replaced it with the snowhouse built on the ice. Snow structures housing several families, encouraging food sharing and enhancing feelings of group identity, developed about this time.

Under conditions of stress brought on by climatic change, people of the Thule culture responded with modifications of their way of life—in the locations and organization of settlements, in the tools and methods for subsistence and in the range and mixture of animals taken. The contrast to the more tightly defined patterns of Pre-Dorset culture is striking. The broader economic potential and the greater flexibility of Thule culture are undoubtedly reasons for the culture's persisting in such varied and robust form into the historic era. The lifeways of the Inupiaq-speaking Eskimos, scattered from northern Alaska to Greenland, are based largely on Thule culture. Yet the success of Thule culture is only one episode in the complex prehistory of human settlement in the Arctic. The village at Anangula, Pre-Dorset culture and the Ipiutak culture of Point Hope are equally instructive and historically relevant. Throughout all these episodes one theme recurs, namely that the relations between the natural environment and the settlers' culture determine the resources available to the community. The relations between nature and culture have a significant effect on human life in all regions of the world. Yet their significance seems particularly dramatic in what appears to us as that inhospitable yet obviously eminently habitable region, the Arctic.

THE AMATEUR SCIENTIST

How best to see Halley's comet while it is in view during the next few months

by Jearl Walker

Halley's comet is now passing among the inner planets and around the sun. The comet is appropriately named for the 17th-century English astronomer Edmund Halley; he was the first to discover that it periodically revisits the solar system. Employing laws of mechanics and procedures of observation previously outlined by his contemporary Isaac Newton, Halley found that comets recorded in 1531, 1607 and 1682 had similar orbits. Suspecting that all the sightings were of the same object, he concluded that the comet would return in 1758. It did, observed by an amateur astronomer on Christmas night. Actually recent work indicates that Halley's comet was seen as early as 240 B.C.; it has been observed approximately every 76 years since then. The present return is the fourth since Halley proposed his theory.

Halley's comet will be visible, albeit only faintly, to people living in the Northern Hemisphere, and so I shall discuss the observations that will be possible in those mid-latitudes. My information is derived from the International Halley Watch at the Jet Propulsion Laboratory of the California Institute of Technology and from books by Robert D. Chapman and John C. Brandt of the National Aeronautics and Space Administration's Goddard Space Flight Center [see "Books," page 34], and research by David W. Hughes of the University of Sheffield.

Because the comet will be faint, you will be able to see it only against a dark night sky. Because of the glare from artificial light scattered by the air, you will not be likely to observe it from residential areas. A bright moon will also make the comet invisible. A good test for an observation site is that the Milky Way should be visible. Allow at least 15 minutes for your eyes to adjust to the darkness.

You might try to see the comet with unaided eyes; binoculars will greatly improve the observation because they gather more light. Binoculars with a magnifying power of seven and a 50-millimeter objective lens should be best; binoculars with an objective diameter of 35 millimeters will also serve. At times the comet will occupy a large angle in your field of view, so that you may have to scan with the binoculars to see all of it. A telescope with a large magnification is not suitable: the part of the comet visible through a telescope is too small to contrast strongly enough with the background darkness. You might try a wide-angle telescope at its lowest magnification.

The illustrations on page 173 indicate where the comet's head will be on various nights in the next few months for an observer at 40 degrees north latitude (approximately the latitude of San Francisco, Denver, St. Louis and Washington, D.C.). For observers at higher latitudes the comet will be lower in the sky and will be above the horizon for shorter periods. Each illustration is a plot of the altitude and azimuth of the comet in degrees. Altitude is measured from the horizon; azimuth is measured clockwise from due north.

The small circles on the graphs indicate the position of the comet's head at hourly intervals on various nights. Positions before midnight are connected with a broken line, those after midnight with a solid line. The first circle on each path shows the date and hour (in local standard time on the 24-hour clock) when the comet's head will be at that location. For example, the top line in the upper illustration indicates that the comet's head will be at an altitude of about 38 degrees and an azimuth of about 108 degrees at 17:00 (5:00 P.M.) on December 1. Later that night the head will rise to a peak altitude of about 60 degrees when it is due south. Then it will descend until it sets in the western sky at about 3:00 A.M.

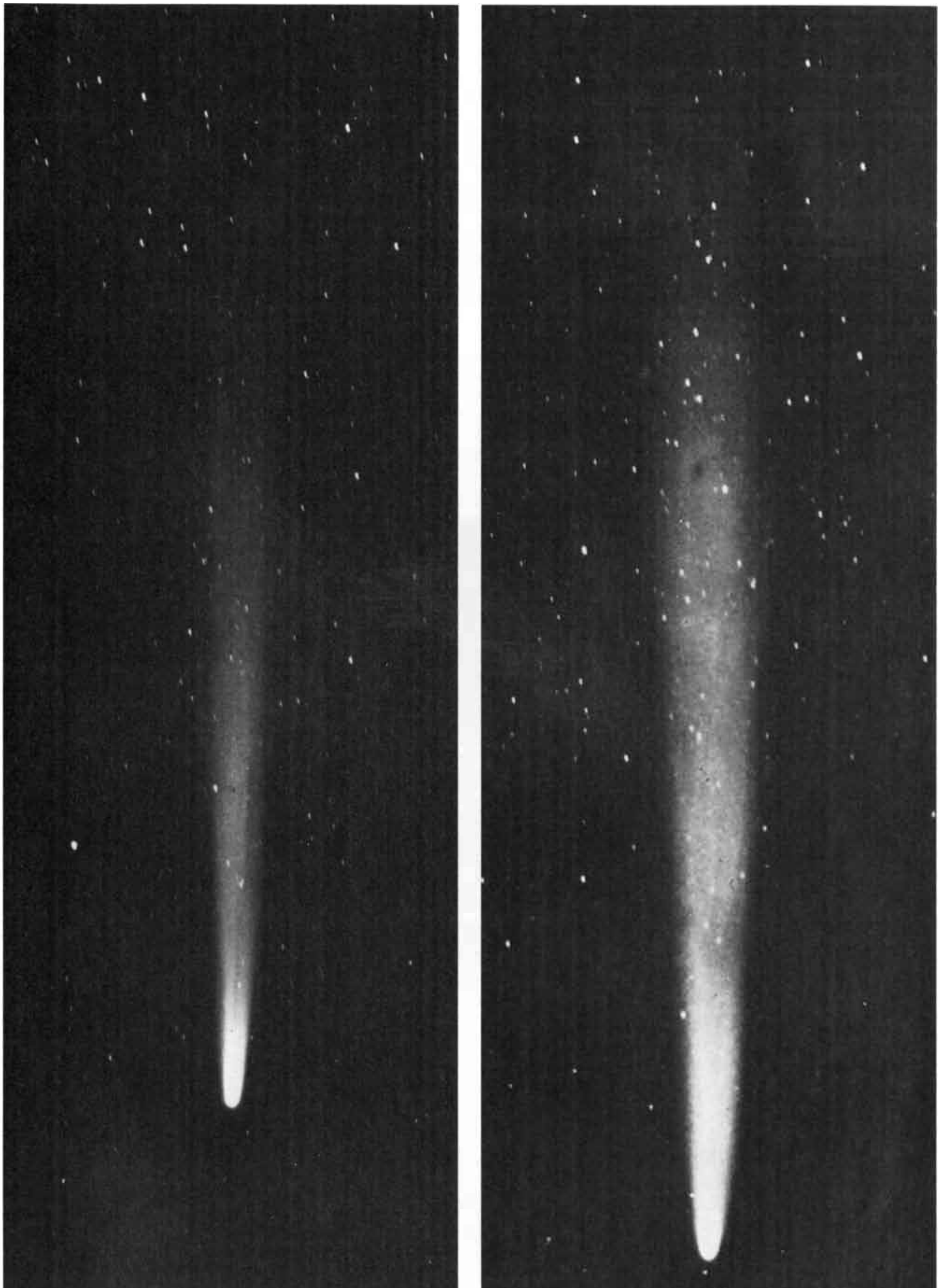
As the nights go by in December and January the altitude of the comet's head at the earliest nightly sighting will decrease. By January 24 the comet will be only briefly visible before it sets in the early evening. During the latter half of both December and January the moon will be bright while the comet is above the horizon and so the viewing will be poor or impossible. The moon also will make viewing difficult toward the end of February, the last week of March and the first and last days of April.

During February and March the observation periods will be short. Since the comet will be low in the sky, you must find an observation point that offers a low and clear horizon. The comet will also be dim during those months; because of its low position, light from it reaches you only after a long trip through the earth's atmosphere. In some locations the scattering of the light by air molecules and aerosols during the long passage may make the comet so dim that it is unobservable. Many experts believe the best opportunity to see the comet will come on about April 15, when the head will be bright and will lie somewhat above the horizon while the sky is dark.

The positions of comets and other celestial bodies are often given as angles measured in relation to the equatorial plane, an imaginary reference object that extends indefinitely outward from the earth's Equator. The plane intercepts an imaginary celestial sphere, the reference object for locating celestial bodies. The equatorial plane divides the northern and southern hemispheres of the celestial sphere. The sun crosses the equatorial plane twice each year, on the vernal and autumnal equinoxes.

In the top illustration on page 174 an observer is at the center of the equatorial plane, which is tilted from a horizontal plane extending through his horizon. (The angle of the tilt is related to the latitude of the observer: the higher the latitude, the smaller the angle.) The horizon is indicated by the symbols N, S, E and W. The field of stars seen by the observer lies on the celestial sphere, which appears to rotate about the north celestial pole because of the rotation of the earth.

Two coordinates—declination and right ascension—serve to locate an object on the celestial sphere. Declination is the angle between the object and the nearest point of the equatorial plane. Right ascension is the angle between that point on the plane and the position of the vernal equinox on the celestial sphere. By convention the angle for right ascension is measured in units of time; one hour is equivalent to 15 degrees. This coordinate system is useful because it rotates about the



Halley's comet on May 12 and 15, 1910, as photographed from Honolulu north celestial pole along with the entire celestial sphere. Hence the coordinates for a star or a comet remain essentially constant overnight. Moreover, the coordinates are independent of the observer's geographic location.

The bottom illustration on page 174 demonstrates the motion of Halley's comet along the celestial sphere for the next few months. Until December 23 the declination is positive, that is, the comet is "above" the equatorial plane. The declination becomes negative as the comet descends "below" the plane.

The motion of the comet can also be described in terms of three angles measured with respect to the ecliptic plane: the plane of the earth's orbit around the sun [*see bottom illustration on page 175*]. The orbital inclination of the comet is the angle through which the ecliptic plane must be rotated to coincide with the comet's orbital plane in such a way that the earth and the comet orbit the sun in the same direction. The orbital inclination of Halley's comet is 162 degrees.

The second angle of reference is the longitude of the ascending node: the point at which the comet passes upward through the ecliptic plane on its approach to the sun. The angle, 58 degrees for Halley's comet, lies between the direction of the vernal equinox and a line extending from the sun through the node.

The third angle has to do with the comet's perihelion (its closest point to the sun, which will be on February 9, 1986) and is called the argument of perihelion. It is the angle between two lines extending from the sun. One line passes through the perihelion, the other through the node. For Halley's comet this angle is about 112 degrees.

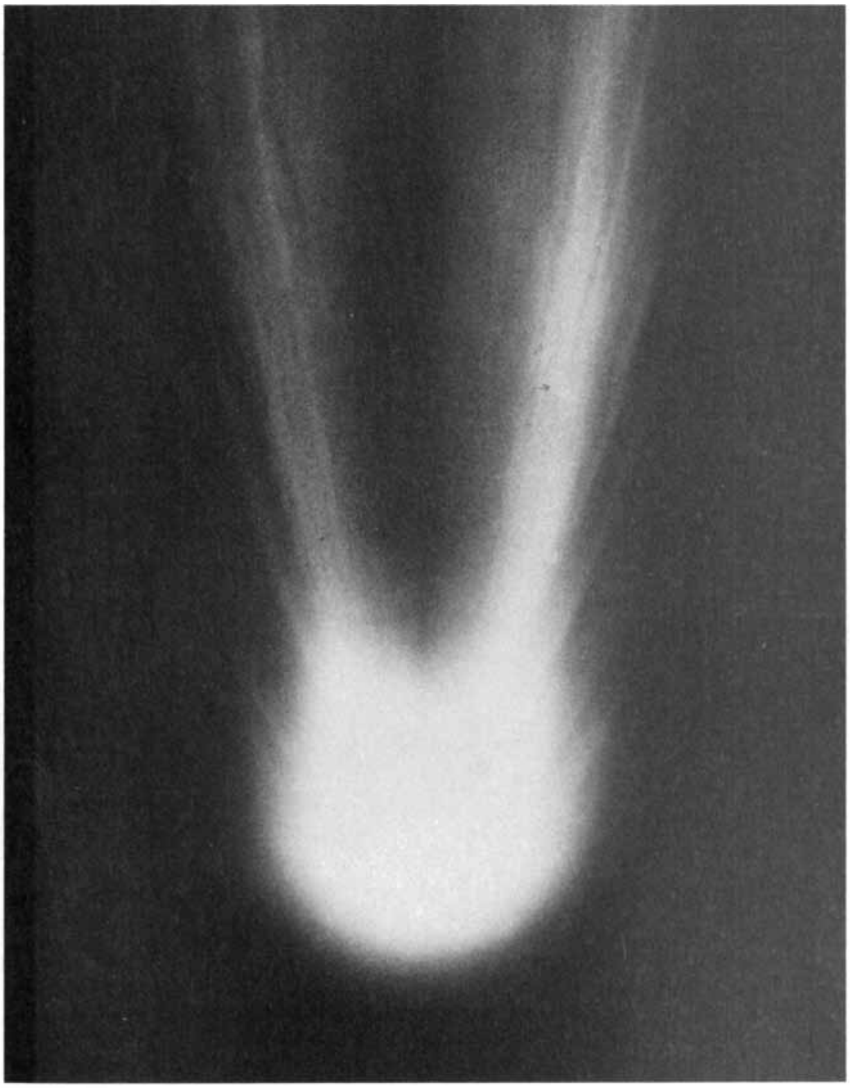
The illustration also includes approximate locations of the earth when it will be closest to the comet. Before perihelion the closest approach will be on November 27, 1985, when the separation will be about 93 million kilometers. After perihelion the closest approach will be about 48 million kilometers. It will take place on April 11, 1986. (In the comet's appearance in 1910 the separation was only 24 million kilometers, so that the comet looked much brighter than it will on its present visit.) When the comet reaches perihelion, it will be unobservable because it will be near the side of the sun that faces the earth and so one's view will be obliterated by the sun's glare.

The full orbit of the comet in the solar system is shown in the top illustration on page 176, along with its position in various years. One full orbit takes about 76 years, making Halley a short-period comet. Long-period comets have orbital periods of 200 years or more. The two types differ in other ways. The orbital planes of the short-period comets generally lie close to the ecliptic plane, and the comets usually orbit the sun in the same direction as the planets (counterclockwise as seen from above the ecliptic plane). The orbits of the long-period comets can incline at any angle, and many of them are clockwise.

Comets come, it is thought, from a giant cloud of the objects that orbit the sun at a distance of between 20,000 and 100,000 astronomical units. (One astronomical unit, 150 million kilometers, is the mean distance between the earth and the sun.) The cloud is named the Oort cloud after the Dutch astronomer Jan Oort, who proposed its existence in 1950.

At times the gravitational pull of a passing star draws comets out of the Oort cloud. Some go far afield and some enter the solar system. Many of the latter end up in large orbits with long periods. Others repeatedly pass close enough to Jupiter or another giant planet for the gravitational pull of the planet to shrink the orbit until the object becomes a short-period comet. Halley's comet is thought to be a member of this class.

The nucleus of a comet is a cohesive structure that ranges in width from one kilometer to several tens of kilometers. A comet may be irregular in shape rather than spherical; no observer has had a good look at a nucleus because the object is so small. Nevertheless, many of the characteristics can be inferred. About 25 percent of the mass is made up of dust particles similar in composition to carbonaceous-chondrite meteorites. The particles range in diameter from .1 micrometer to 10 micrometers. Larger dust particles and pebbles are also present. The other 75 percent of the mass consists of ice, mainly water ice and snow. Some comets may also have a rocky core. Pock-



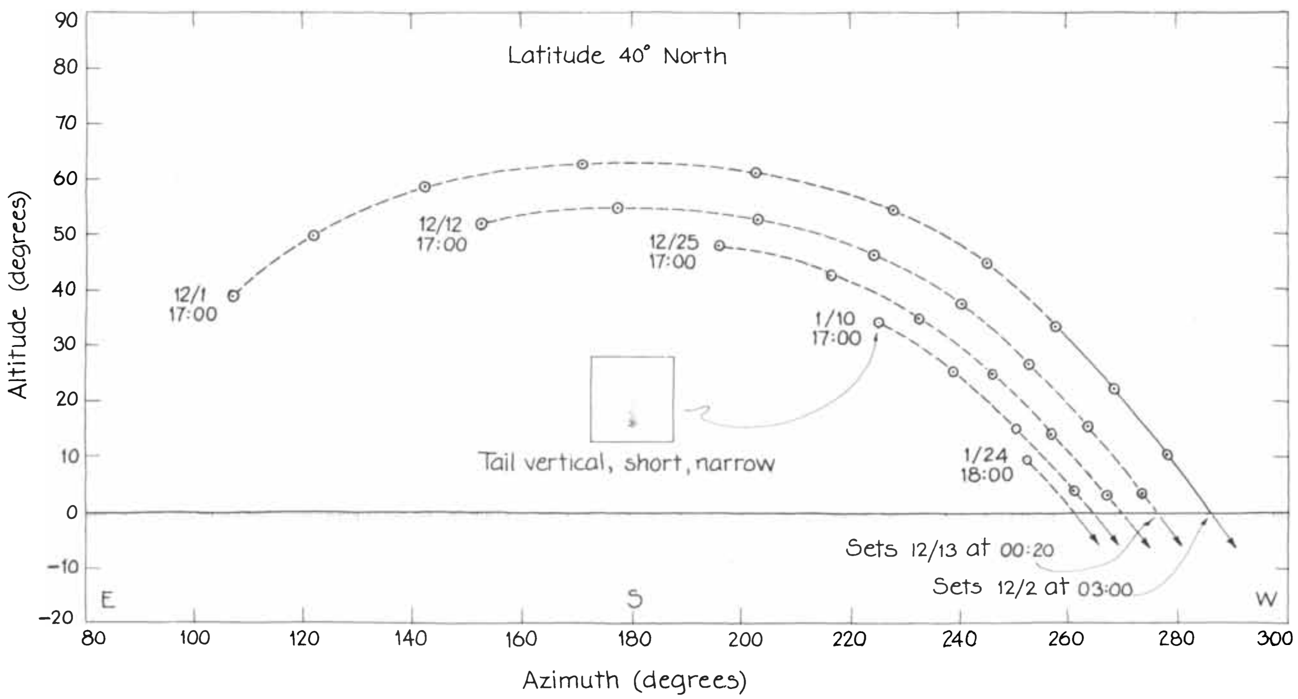
The head of the comet on May 8, 1910

ets in the loose packing of ice crystals and snow contain an assortment of such molecules as ammonia, methane and carbon dioxide.

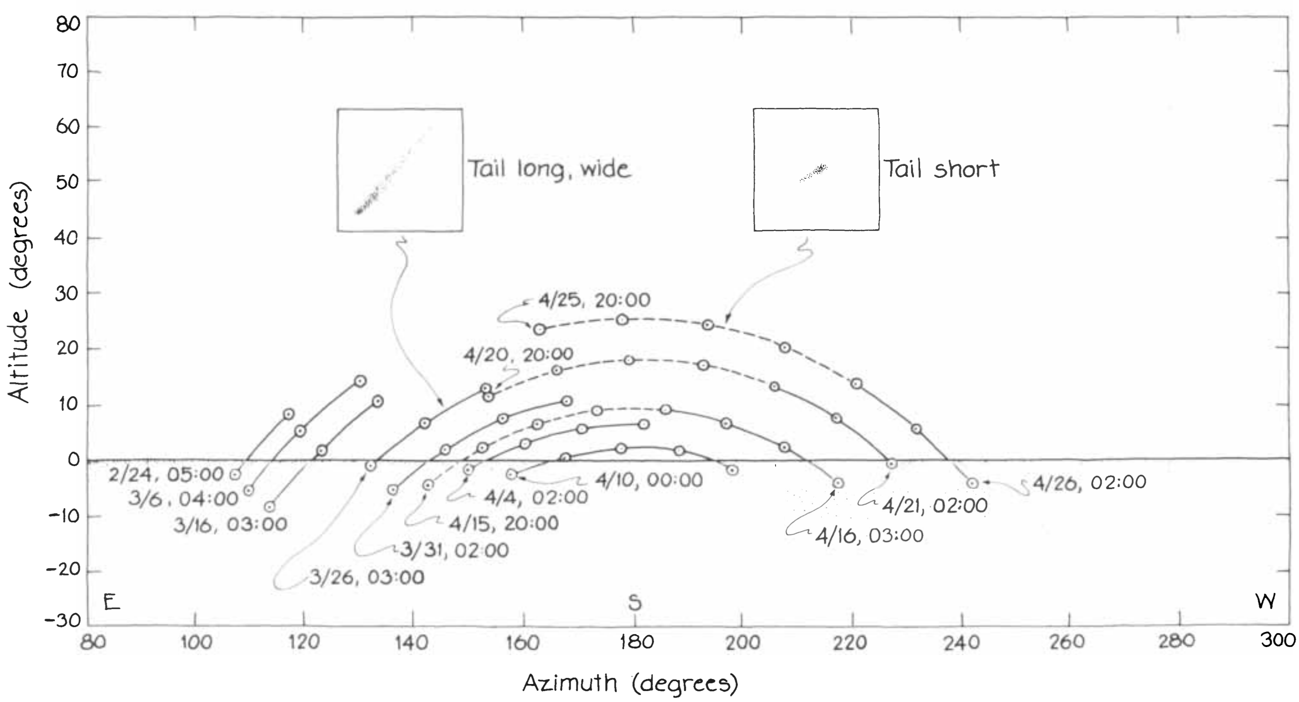
As a comet nears the sun the heat on the sunward surface transforms the ice directly into gas (a process called sublimation). Trapped molecules escape, and ultraviolet light from the sun dissociates and ionizes them to form simpler "daughter" molecules, atoms and ions. Spectroscopic examinations of comets reveal the presence of the cyanogen molecule (CN) when a comet is as much as three astronomical units from the sun.

As the comet moves closer to the sun spectroscopic records display more emission lines, showing that the number of daughter molecules is rising rapidly. They absorb energy of certain wavelengths and then emit energy as light either at the same wavelength (a process known as resonance fluorescence) or at a longer wavelength (fluorescence). In addition the released water molecules dissociate to form a huge hydrogen (H) cloud and a smaller hydroxyl (OH) cloud around the comet's nucleus. The diameter of the hydrogen cloud may be as much as .1 astronomical unit.

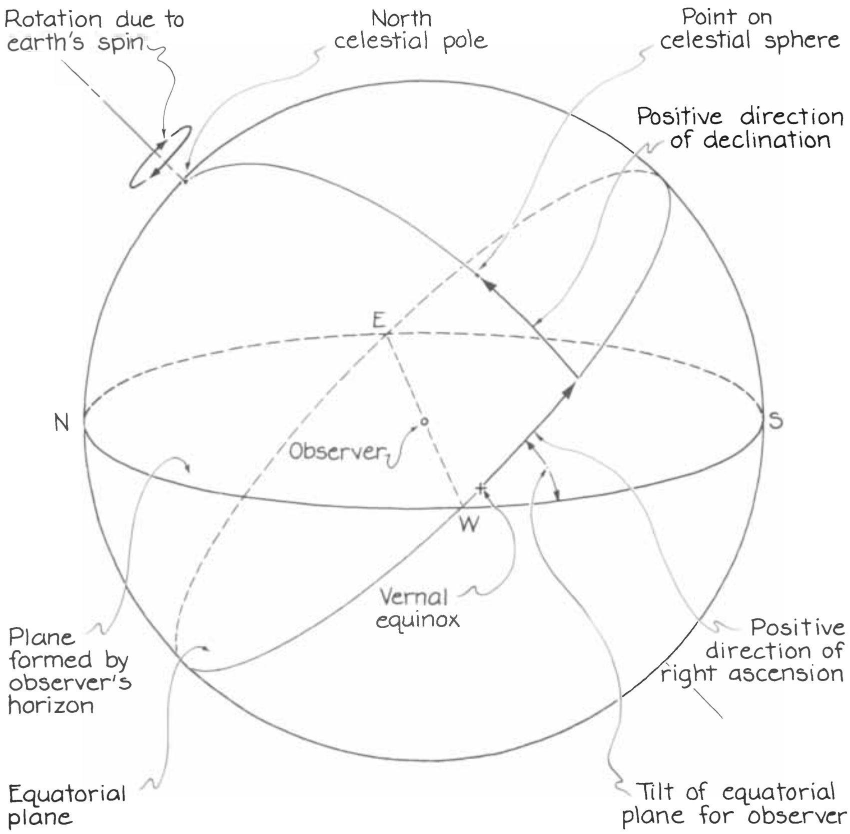
As the outer surface of the nucleus



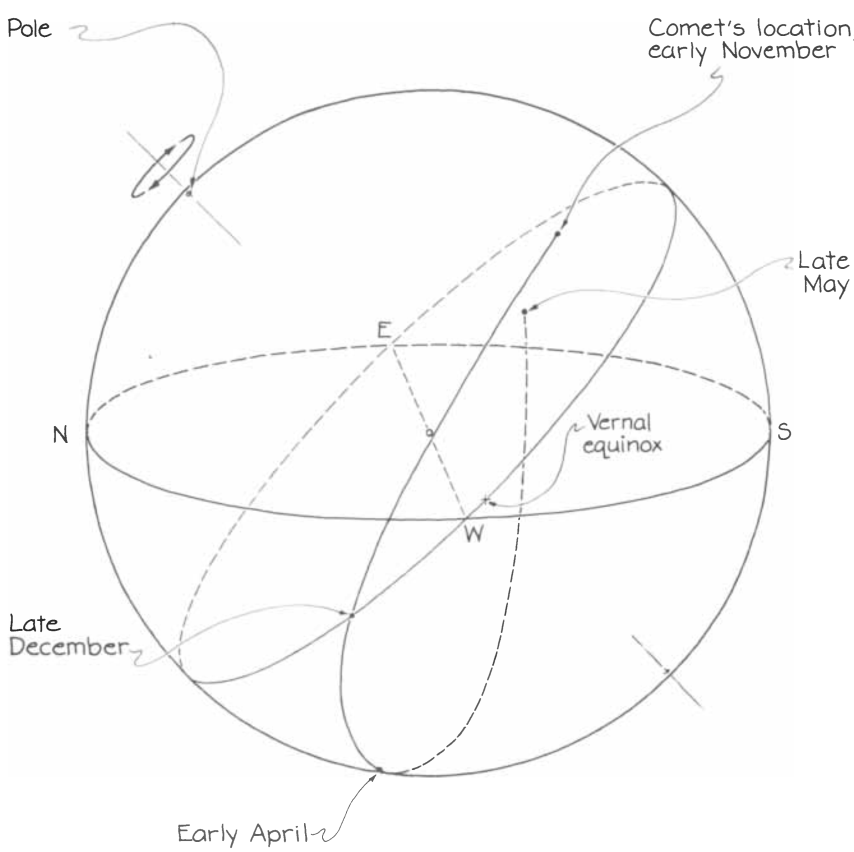
Positions of the comet in the Northern Hemisphere during December and January



Where the comet will be in February, March and April



The coordinate system based on the equatorial plane



The motion of Halley's comet on the celestial sphere

gets warm and loses ice it is transformed into a spongy dust layer that is from one centimeter to 10 centimeters deep. If the nucleus spins so that all the surface receives sunlight, the spongy layer forms over the entire nucleus. The layer insulates the deeper ice. The core of the nucleus probably remains at a temperature of approximately −150 degrees Celsius.

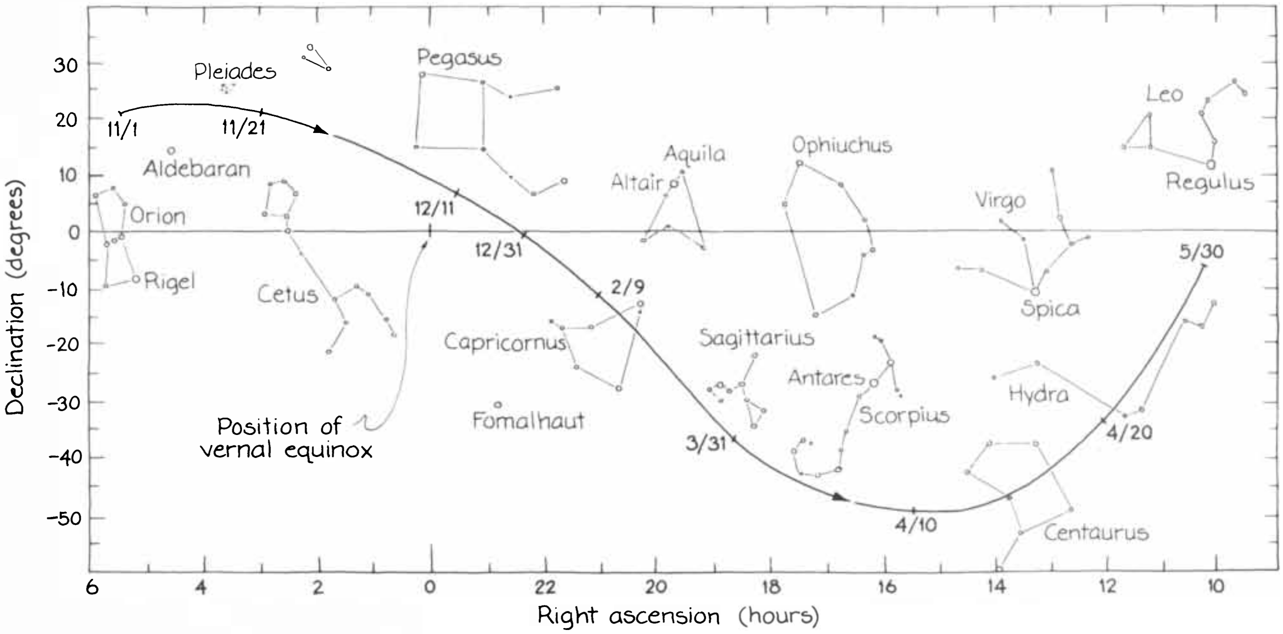
Dust particles on the surface gradually break off and are blown away by the pressure of gas molecules released by the sublimation of ice at the boundary of the spongy dust layer and the dusty ice. As the surface erodes, the boundary moves inward, maintaining the thickness of the layer.

The dust and neutral gas released by the nucleus form an elliptical, glowing cloud called the coma. It may have a diameter of as much as 100,000 kilometers. The coma increases in diameter until the comet is about 1.5 or two astronomical units from the sun. Thereafter the coma shrinks because the material expelled from the nucleus moves rapidly into two tails. At times the coma displays glowing streamers. They may result from uneven heating on the surface of the nucleus.

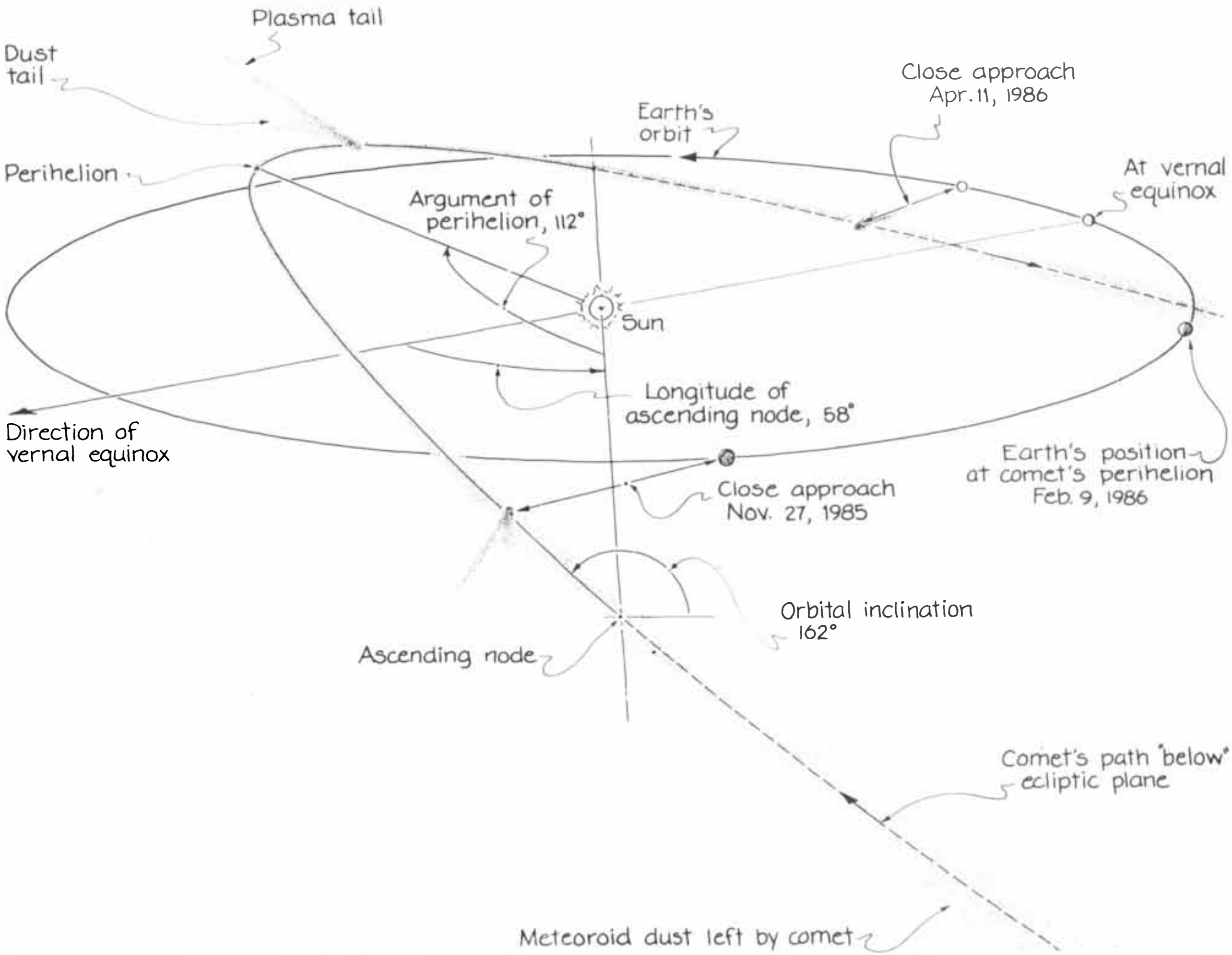
As the comet nears the sun it also begins to develop a plasma tail and a dust tail. The plasma tail is a straight stream of charged particles that points almost directly away from the sun. The dust tail forms a curve with its concave side facing toward the comet's previous locations. Both tails derive from the material expelled from the surface of the nucleus, and so they are more pronounced when the comet is near the sun.

The electrically charged atoms and molecules of the ejected material are forced into the plasma tail by the solar wind, which consists of a stream of protons and electrons moving outward from the sun. The collision between the solar wind and the ions emanating from the comet distort the interplanetary magnetic field so that the magnetic field lines wrap around the comet's head and extend along a radial line away from the sun. The cometary ions are forced to spiral around the radial lines. The spiraling ions constitute the plasma tail.

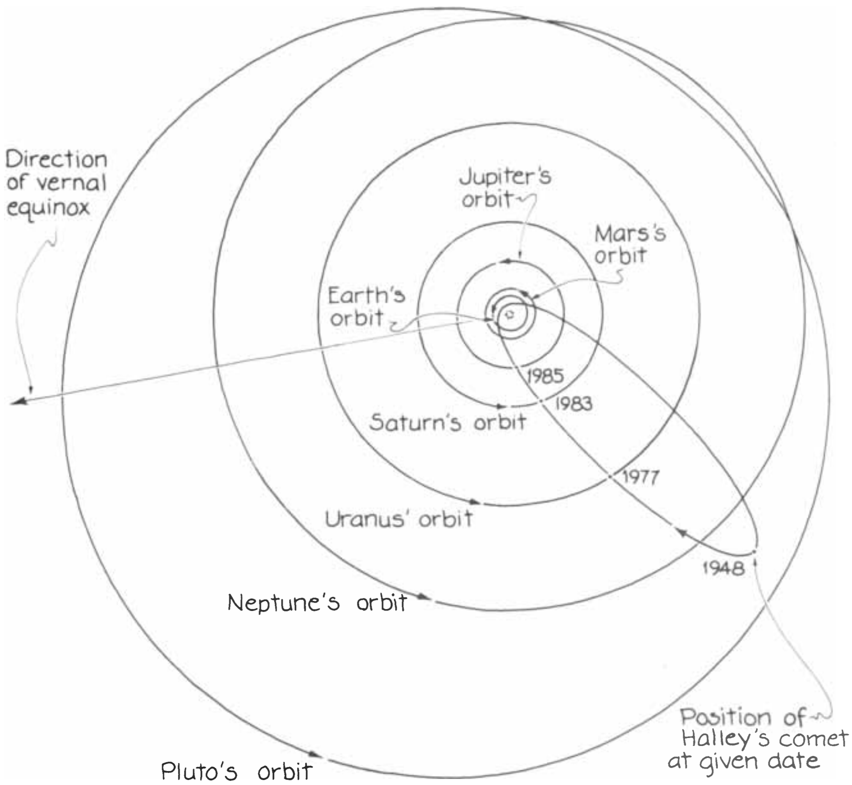
The tail is visible because the ions emit light. It is blue primarily because of the blue emissions from the positively ionized carbon monoxide molecule (CO^+). Often the tail develops kinks and waves due to irregularities in the ejections from the nucleus or in the solar wind. At times the tail separates from the nucleus, after which a new tail forms. The separations seem to occur when the comet crosses a boundary in the magnetic field surrounding



The comet among the constellations



Halley's orbit in relation to the orbital plane of the earth

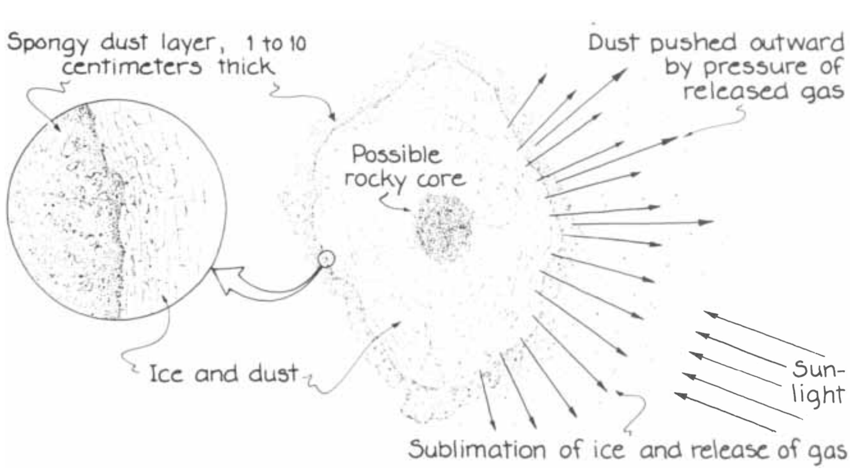


Halley among the planets

the sun. The arrangement of the field lines resembles a pinwheel consisting of spiraling sectors. In one sector the field lines extend toward the sun and in the adjacent sector they extend away from the sun. When a comet crosses the boundary between sectors, the change in the direction of the field lines releases the plasma tail. A new tail forms as the comet travels through the next sector.

The dust tail forms from the particles of dust released by the nucleus. These grains, each approximately a micrometer in diameter, scatter the sunlight with a peak intensity in the yellow portion of the spectrum. The tail is therefore yellow.

The tail develops because dust in the coma is pushed away from the sun by the pressure of the sunlight. (Light has momentum and pushes on a surface on which it shines.) Each grain is pulled radially inward by the sun's gravitational field and pushed radially outward by the pressure of the sunlight. Grains larger than a micrometer are dominated by the gravitational pull and end up orbiting the sun in a belt along the orbital path of the comet. The smaller grains are dominated by the light pressure and move away from the sun.



A model of a comet's nucleus

It is these smaller dust particles that form the dust tail [*see lower illustration on opposite page*]. I have assumed the particles are identical in size and so are pushed by the light at a uniform rate. Inasmuch as the grains are released with an initial momentum because the comet is moving, they move outward from the sun along a curved path. In the example given the particles released by the comet early in its orbit have moved outward to form the outer end of the tail. The particles released later in the orbit have not moved as far and so form the part of the tail closer to the nucleus. The composite tail gives rise to the illusion that dust particles ejected from the nucleus move along the tail.

In reality the particles composing the dust tail vary in size. The smaller ones are pushed outward from the sun faster than the larger ones, and so the tail is broader than it would be if the particles were of uniform size. Bursts of dust from the nucleus may further alter the apparent shape of the tail.

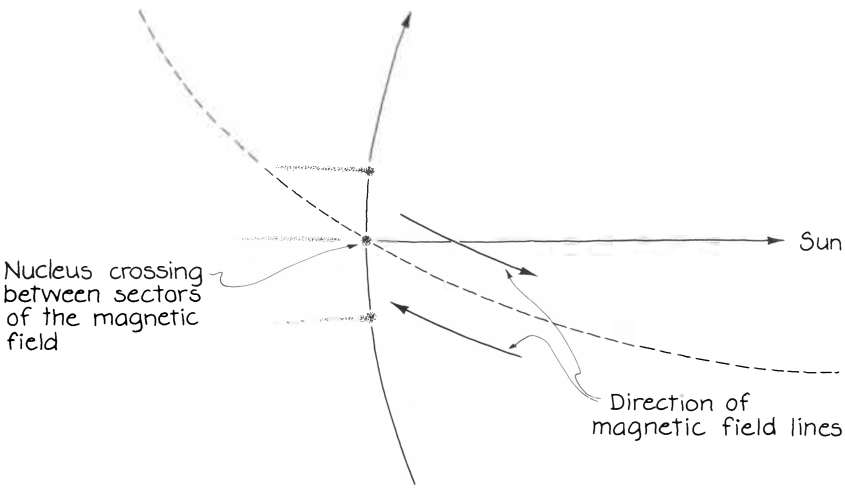
The dust tail normally becomes wider and longer after the comet has passed perihelion. By then the nucleus has swung around the sun, whereas the dust particles released previously have not. The apparent length of the tail depends on the observer's line of sight. If the tail is approximately along the line of sight, it occupies only a small angle in one's field of view. The best view comes when the line of sight is perpendicular to the tail.

At times a comet may have an antitail that seems to point toward the sun. The appearance is illusory. The antitail is actually an almost edge-on view of the larger dust particles strewn along the comet's orbit. They are visible if the observer is near the orbital plane of the comet and has a line of sight that virtually coincides with the path just taken by the comet. In these conditions the line of sight passes through enough dust to make the antitail visible.

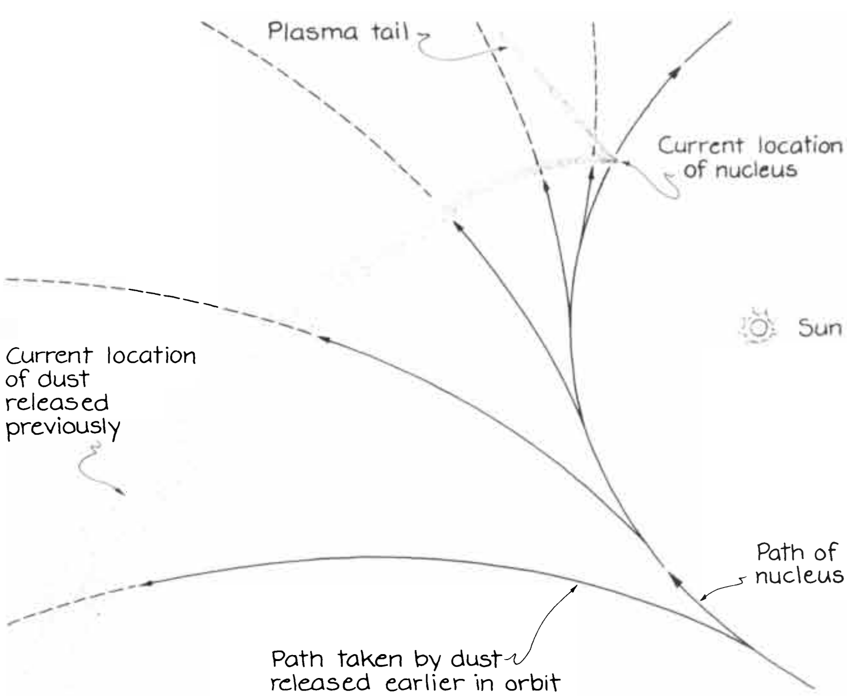
The dust tail of Halley's comet will probably be short, narrow and straight before the comet goes behind the sun. If viewing conditions are good, you might see it extending nearly vertically from the comet's head. Just before perihelion the tail will begin to broaden, but the glare of the sun will mask it from observation. As the comet emerges from the glare later in February, the tail will be wider, longer and more pronounced. It probably will be most visible early in April, when it will extend generally westward. After mid-April its width will decrease. The antitail for Halley's comet will form when the comet reappears in late February. It will be faint and short, however, and you will be able to see it only if conditions are ideal.

The larger dust particles released by a comet lie in a belt roughly along the comet's orbital path. These particles continue to orbit the sun. If the earth passes through the residual cometary dust, the particles ablate and burn as they enter the atmosphere, giving rise to meteor showers. The earth passes through the orbit of Halley's comet twice a year. The Eta Aquarid meteor shower in May and the Orionid shower in October both result from previous passages of Halley's comet around the sun.

The nucleus of Halley's comet is believed to have a mass of 2×10^{14} kilograms; the comet's mean diameter is about 10 kilometers. The proportions of dust and solid matter to water ice are average for comets. On its most recent pass through the inner solar system the comet ejected about 2×10^{11} kilograms of material, losing a surface layer about two meters thick. Even at this rate of loss of material Halley's comet is good for many more passes around the sun.



Forces on the comet's plasma tail



How the comet's dust tail forms

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

ARTIFICIAL INTELLIGENCE THROUGH SIMULATED EVOLUTION. Lawrence J. Fogel, Alvin J. Owens and Michael J. Walsh. John Wiley & Sons, Inc., 1966.

ADAPTATION IN NATURAL AND ARTIFICIAL SYSTEMS. John H. Holland. University of Michigan Press, 1975.

GENETIC ALGORITHMS FOR THE TRAVELING SALESMAN PROBLEM. John Grefenstette, Rajeev Gopal, Brian Rosmaita and Dirk Van Gucht in *Proceedings of an International Conference on Genetic Algorithms and Their Applications,* edited by John J. Grefenstette. The Robotics Institute, Carnegie-Mellon University, 1985.

THE TREATMENT OF DISEASES AND THE WAR AGAINST CANCER

A HISTORY OF CANCER CONTROL IN THE UNITED STATES, 1946–1971. Lester Breslow. Department of Health, Education, and Welfare Publication No. 78–1516, U.S. Government Printing Office, 1977.

CANCER MORTALITY IN THE UNITED STATES: 1950–1977. Frank W. McKay, Margot R. Hanson and Robert W. Miller. National Cancer Institute Monograph No. 59, U.S. Department of Health and Human Services, U.S. Government Printing Office, 1982.

MASS SCREENING FOR CONTROL OF BREAST CANCER. Philip Strax in *Cancer,* Vol. 53, No. 3, pages 665–670; February 1, 1984.

DECLINE IN U.S. CHILDHOOD CANCER MORTALITY, 1950 THROUGH 1980. Robert W. Miller and Frank W. McKay in *Journal of the American Medical Association,* Vol. 251, No. 12, pages 1567–1570; March 23/30, 1984.

COSMIC RAYS FROM CYGNUS X-3

AN ASTRONOMICAL PUZZLE CALLED CYGNUS X-3. R. M. Hjellming in *Science,* Vol. 182, No. 4117, pages 1089–1095; December 14, 1973.

GAMMA-RAY ASTRONOMY. Edited by H. S. W. Massey, R. D. Wills and A. W. Wolfendale. Royal Society of London, 1981.

HIGH ENERGY ASTROPHYSICS. Malcolm S. Longair. Cambridge University Press, 1981.

THE CALCIUM SIGNAL

CA^{++} UPTAKE BY RAT KIDNEY MITOCHONDRIA AND ITS DEPENDENCE ON RESPIRATION AND PHOSPHORYLATION. Frank D. Vasington and Jerome V. Murphy in *The Journal of Biological Chemistry,* Vol. 237, No. 8, pages 2670–2677; August, 1962.

ATP-DEPENDENT CA^{2+}-EXTRUSION FROM HUMAN RED CELLS. H. J. Schatzmann in *Experientia,* Vol. 22, No. 6, pages 364–365; July 15, 1966.

PURIFICATION AND PROPERTIES OF AN ADENOSINE TRIPHOSPHATASE FROM SARCOPLASMIC RETICULUM. David H. MacLennan in *The Journal of Biological Chemistry,* Vol. 245, No. 17, pages 4508–4518; September 10, 1970.

PURIFICATION OF THE (CA^{2+}-MG^{2+})-ATPASE FROM HUMAN ERYTHROCYTE MEMBRANES USING A CALMODULIN AFFINITY COLUMN. Verena Niggli, John T. Penniston and Ernesto Carafoli in *The Journal of Biological Chemistry,* Vol. 254, No. 20, pages 9955–9958; October 25, 1979.

TERRANES

WRANGELLIA: A DISPLACED TERRANE IN NORTHWESTERN NORTH AMERICA. N. J. Silbering and J. Hillhouse in *Canadian Journal of Earth Sciences,* Vol. 14, No. 11, pages 2565–2577; 1977.

CORDILLERAN SUSPECT TERRANES. Peter J. Coney, David L. Jones and James W. H. Monger in *Nature,* Vol. 288, No. 5789, pages 329–333; November 27, 1980.

SUSPECT TERRANES AND ACCRETIONARY HISTORY OF THE APPALACHIAN OROGEN. Harold Williams and Robert D. Hatcher, Jr., in *Geology,* Vol. 10, No. 10, pages 530–536; October, 1982.

THE GROWTH OF WESTERN NORTH AMERICA. David L. Jones, Allan Cox, Peter Coney and Myrl Beck in *Scientific American,* Vol. 247, No. 5, pages 70–84; November, 1982.

PHANEROZOIC ADDITION RATES TO THE CONTINENTAL CRUST AND CRUSTAL GROWTH. A. Reymer and G. Schubert in *Tectonics,* Vol. 3, No. 1, pages 63–77; February, 1984.

SKIN BREATHING IN VERTEBRATES

THE EVOLUTION OF AIR BREATHING IN VERTEBRATES. David J. Randall, Warren W. Burggren, Anthony P. Farrell and M. Stephen Haswell. Cambridge University Press, 1981.

A MODEL FOR EVALUATING DIFFUSION LIMITATION IN GAS-EXCHANGE ORGANS OF VERTEBRATES. J. Piiper in *A Companion to Animal Physiology,* edited by C. R. Taylor, K. Johansen and L. Bolis. Cambridge University Press, 1982.

CUTANEOUS GAS EXCHANGE IN VERTEBRATES: DESIGN, PATTERNS, CONTROL AND IMPLICATIONS. Martin E. Feder and Warren W. Burggren in *Biological Reviews,* Vol. 60, pages 1–45; 1985.

HUMAN-POWERED FLIGHT

FLUG DURCH MUSKELKRAFT. Hans-Georg Schulze and Willi Stiasny. Fritz Knapp, Frankfurt, 1936.

GOSSAMER ODYSSEY: THE TRIUMPH OF HUMAN-POWERED FLIGHT. Morton Grosser. Houghton Mifflin Company, 1981.

BICYCLING SCIENCE. Frank Rowland Whitt and David Gordon Wilson. The MIT Press, 1982.

SMART CARDS

SMART CREDIT CARDS: THE ANSWER TO CASHLESS SHOPPING. Stephen B. Weinstein in *IEEE Spectrum,* Vol. 21, No. 2, pages 43–49; February, 1984.

SMART CARDS—THE ULTIMATE PERSONAL COMPUTER. Jerome Svigals. Macmillan Publishing Company, 1985.

EARLY ARCTIC SETTLEMENTS IN NORTH AMERICA

THE HISTORY OF GREENLAND. Finn Gad. McGill–Queen's University Press, 1971.

ARCHAEOLOGY OF THE LAKE HARBOUR DISTRICT, BAFFIN ISLAND. Moreau S. Maxwell, Mercury Series Paper No. 6, Archaeological Survey of Canada, 1973.

THULE ESKIMO PREHISTORY ALONG NORTHWESTERN HUDSON BAY. Allen P. McCartney. Mercury Series Paper No. 70, Archaeological Survey of Canada, 1977.

EARLY MARITIME ADAPTATION IN THE ALEUTIAN ISLANDS. J. Aigner and T. Del Bene in *Peopling of the New World,* edited by Jonathan E. Ericson, R. E. Taylor and Rainer Berger. Ballena Press, 1982.

THE AMATEUR SCIENTIST

INTRODUCTION TO COMETS. John C. Brandt and Robert D. Chapman. Cambridge University Press, 1981.

COMETS. David W. Hughes in *Contemporary Physics,* Vol. 23, No. 3, pages 257–283; May/June, 1982.

ING FOR ENTERPRISING
ON THE TRADITION.

The Rolex Awards for Enterprise were inaugurated in 1976 with a single aim: "To encourage the spirit of enterprise in individuals throughout the world by acknowledging outstanding personal efforts or contributions made in selected categories of human endeavour." The qualities that the Awards set out to honour are the very ones that have long been shown by Rolex and by Rolex owners.

Now, Rolex takes pride in announcing a further 250,000 Swiss Francs and five specially inscribed gold Rolex chronometers, to be divided equally among five self-motivated individuals who have developed projects displaying the most outstanding spirit of enterprise.

AREAS OF ENTERPRISE.

Projects may be submitted in one of three categories: Applied Sciences and Invention; Exploration and Discovery; The Environment. Within these three broad categories, the Selection Committee will be looking for projects which display a spirit of enterprise together with the qualities of originality and likelihood of realisation.

The accompanying brief descriptions of winning projects from earlier years demonstrate the kind of enterprise that has impressed previous Selection Committees. The members of the current committee, listed here, are equally distinguished and will be applying the same criteria.

If your scheme falls outside the committee's area of expertise, Rolex will call on specialists around the world to advise on the special merits of a particular case.

A hardback book about the Awards will be published, entitled "Spirit of Enterprise – The 1987 Rolex Awards," richly illustrated in colour. It will contain details of some of the best and most fascinating, stimulating and challenging, ideas submitted by men and women throughout the world. The publicity given to projects by previous editions has often led to additional support from a wide range of sources.

HOW TO APPLY.

Your entry must be submitted in English on an official application form and reach the Secretariat before 31st March 1986.

To obtain an official application form for the 1987 Rolex Awards for Enterprise, together with the rules and conditions for entry, write to: The Secretariat, The Rolex Awards for Enterprise, P.O. Box 178, 1211 Geneva 26, Switzerland.

The Awards will be presented to the laureates in Geneva at the end of April 1987.

If you possess originality, imagination and initiative, you will not let this opportunity pass you by.

ROLEX
of Geneva


GM
A little out of the way place for dinner.
You know how it goes. You find a wonderful little spot with just the right ambience and the next thing you know, it gets discovered. Goodbye ambience.
Hello, Jimmy. S-15 Jimmy, by GMC Truck, lets you get away from it all. To anywhere you want to discover next. Jimmy offers civilized interior comforts and exterior looks stylish enough for any setting. With the four-wheel-drive model, that setting can be almost anywhere.
Even if it's more than a little out of the way.
What isn't out of the way is your nearest GMC Truck dealer. Just look in the Yellow Pages.
GMC
TRUCK
A truck you can live with
Let's get it together . . . buckle up.