SCIENTIFIC AMERICAN

ACOUSTICS OF VIOLINS

$2.00

October 1981

Now turn 35mm slides

Polaroid

Polaprinter Instant Slide Copier

Johnnie Walker
Black Label Scotch
YEARS 12 OLD

Success is often measured by how deeply you're in the Black.

12 YEAR OLD BLENDED SCOTCH WHISKY, 86.8 PROOF. BOTTLED IN SCOTLAND. IMPORTED BY SOMERSET IMPORTERS, LTD., N.Y. © 1981

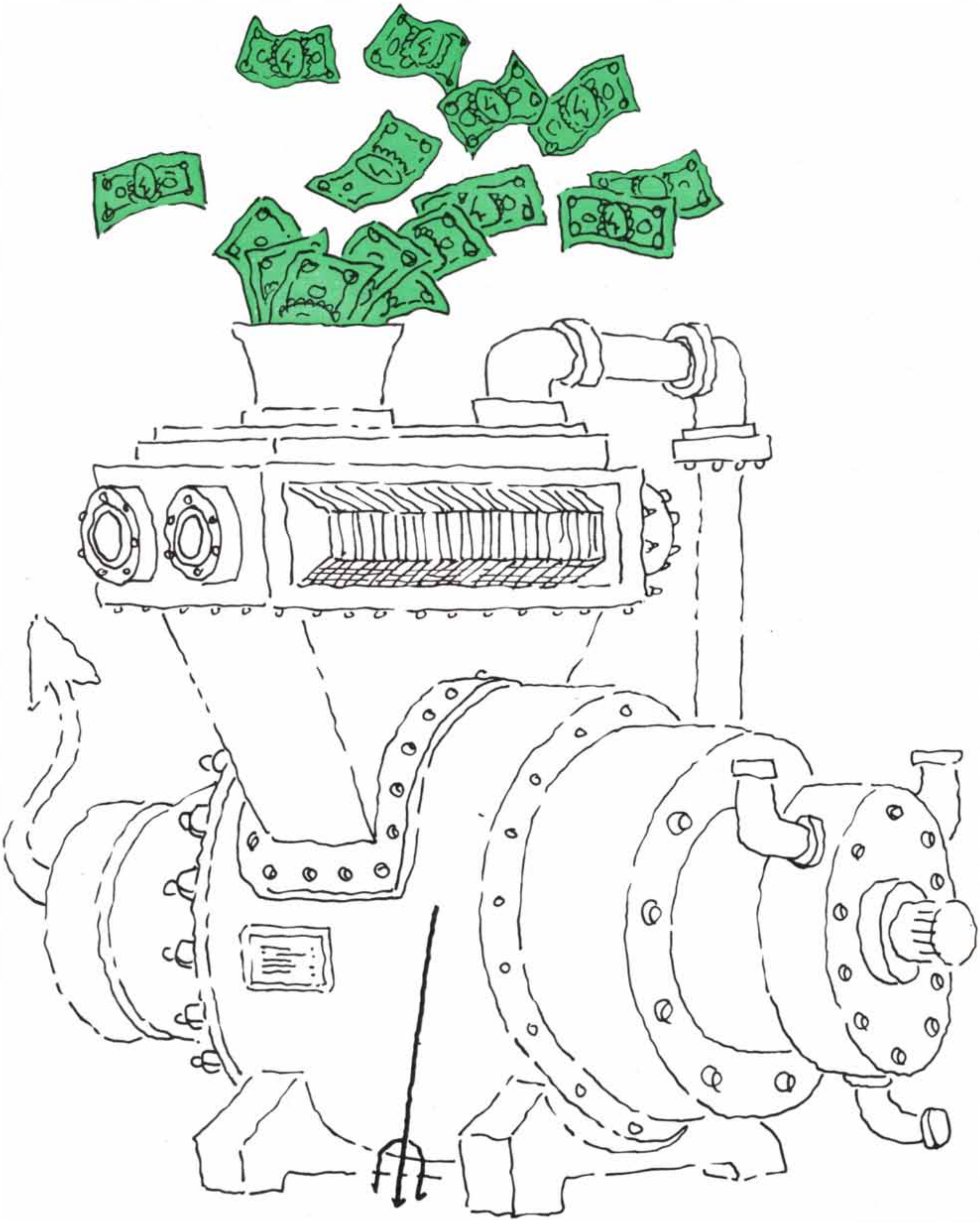
Established 1845 SCIENTIFIC AMERICAN October 1981 Volume 245 Number 4

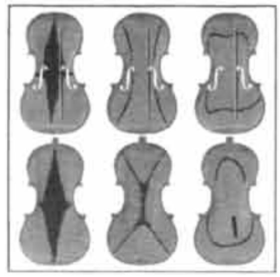
ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. COPYRIGHT © 1981 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: $21 PER YEAR, U.S., ITS POSSESSIONS AND CANADA; $27 PER YEAR, ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.

Imagine turning hot exhaust into cold cash.





THE COVER

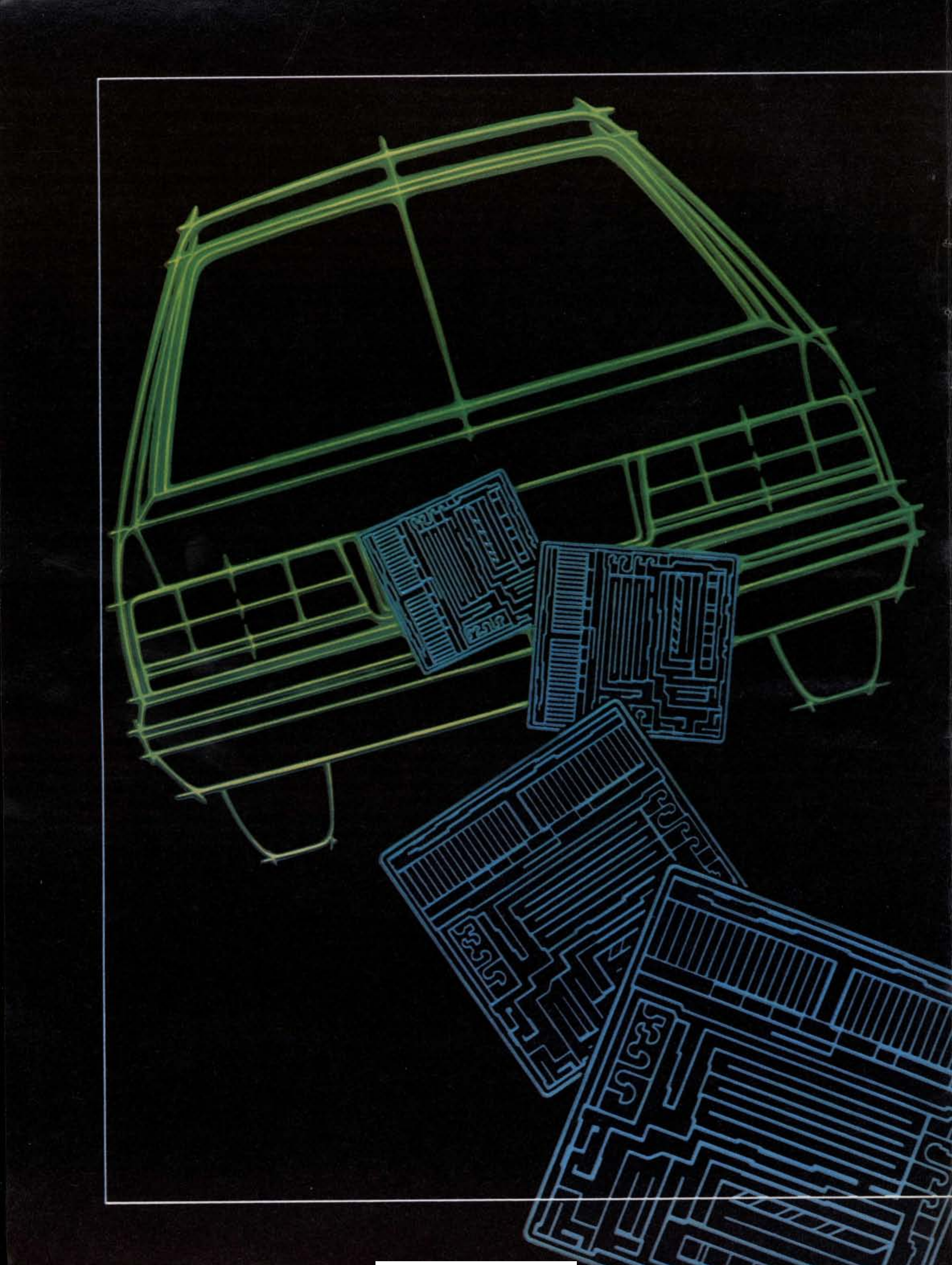
The painting on the cover shows tests of the vibrational properties of an unattached top and back plate of a violin (see "The Acoustics of Violin Plates," by Carleen Maley Hutchins, page 170). In the test the plate is mounted inside upward over a loudspeaker and particles of aluminum flake are sprinkled on it. The plate rests on four soft foam pads, each adjusted to support it at a nodal, or nonvibrating, point. The loudspeaker is centered under an antinode for the vibrational mode being tested. In response to a single-frequency sound from the loudspeaker the particles begin to bounce. Since the antinodal areas move vigorously, the particles are bounced quickly into the nodal areas, giving rise to a pattern that is characteristic of the particular mode. The modes are designated as mode 1, mode 2 and so on, beginning with the mode of lowest frequency. In the painting one sees from left to right the patterns of modes 1, 2 and 5 on each plate. The top plate is made from two pieces of spruce, the bottom one from two pieces of curly maple. The insides of the plates are left "in the white," that is, without sealer or varnish, and therefore display the natural color of the wood.

THE ILLUSTRATIONS

Cover painting by Marvin Mattelson

Page	Source
10	George V. Kelvin
26	© BEELDRECHT, Amsterdam/VAGA, New York, Collection Haags Gemeentemuseum
28–34	Ilil Arbel
55	Mayo Clinic
56	Jon Brenneis
57	Mayo Clinic (*top*), Jon Brenneis (*bottom*)
58–60	Jon Brenneis
62–63	Ilil Arbel
64	Andrew Tomko
65	K. G. McCracken
66	David I. Groves
67	Brian W. Logan
68–69	Andrew Tomko
70–72	David I. Groves
75	Allen Beechel
76	Allen Beechel (*top*), J. A. Tyson (*bottom*)
77–81	Allen Beechel
82	Allen Beechel (*top*); National Astronomy and Ionosphere Center, Cornell University (*bottom*)
149	Jorge Mancillas and Jacqueline F. McGinty (*top*), Nicholas Brecha and Harvey J. Karten (*bottom*)
152	Elena Battenberg and Floyd E. Bloom
153–167	Albert E. Miller
170	Daniel Bernstein
174	Alan D. Iselin
177	Carleen Maley Hutchins
178	Daniel Bernstein
179	Carleen Maley Hutchins
180	Carleen Maley Hutchins (*top*), Alan D. Iselin (*bottom*)
182–184	Daniel Bernstein
186	Alan D. Iselin
189	*Codex Atlanticus,* Giunti-Barbèra, Firenze, copyright 1973
190–196	Michael Goodman
198	André Wegener Sleeswyk (*top*), Michael Goodman (*bottom*)
200	Michael Goodman
203–213	Gabor Kiss
215	Earth Satellite Corporation
216	Andrew Christie
217	Brent S. Baxter
218–219	Brent S. Baxter and Oliver Faugeras
222–224	T. M. Cannon
225	Joel Trussell
227	R. F. Bonifield
228–235	Michael Goodman


SARONNO
The Taste of the Renaissance lives on
ILLVA
AMARETTO di SARONNO
ORIGINALE
A LIQUEUR PRODUCED BY
ILLVA SARONNO ITALY
56 PROOF
750 ml
LIQUEUR 56 PROOF SOLE U.S. DISTRIBUTOR, FOREIGN VINTAGES, INC., JERICHO, NY © 1981 A UNIQUE PRODUCT FROM ILLVA SARONNO, ITALY



LETTERS

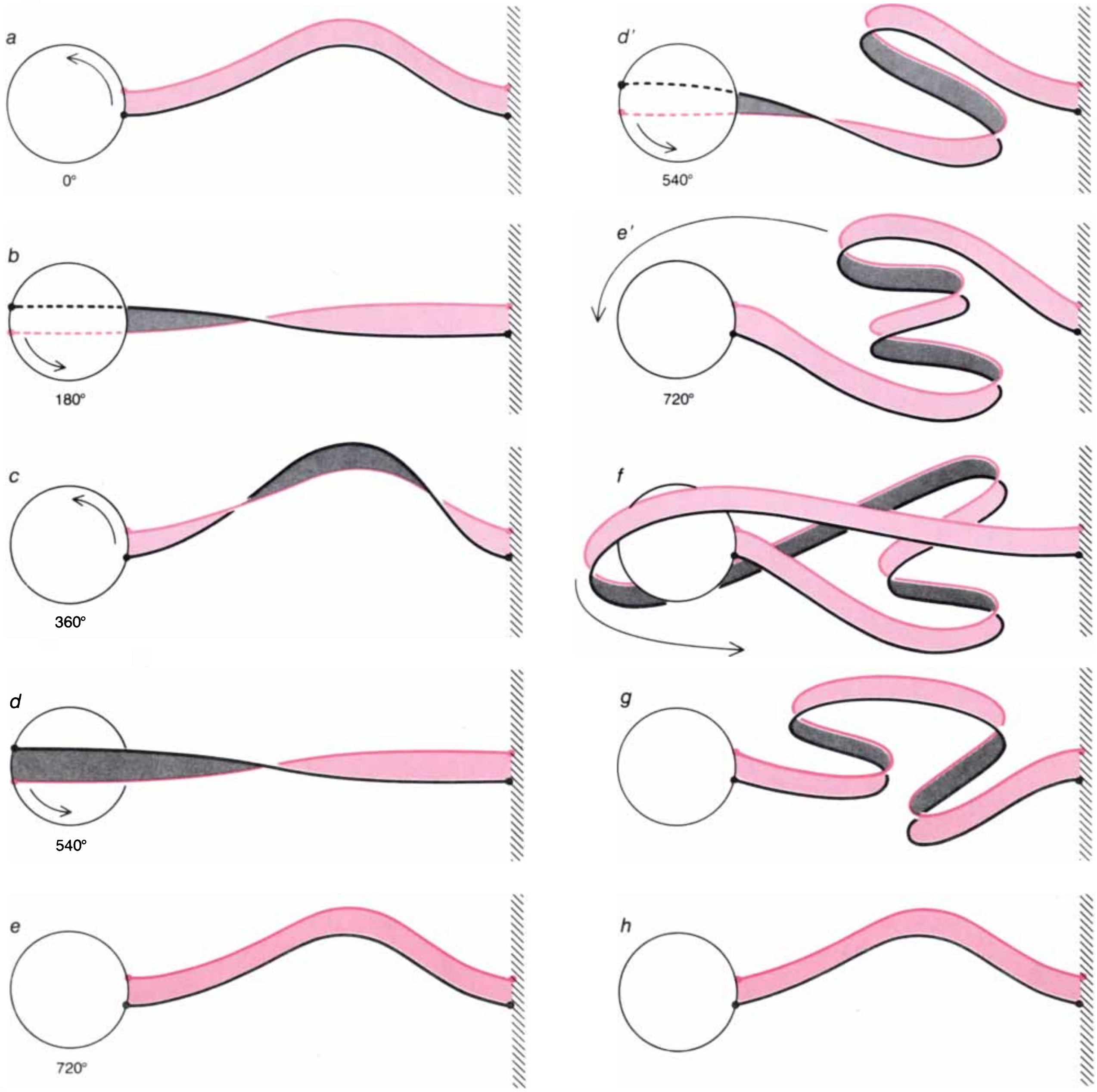
Sirs:

Bernstein and Phillips' very interesting article ["Fiber Bundles and Quantum Theory," by Herbert J. Bernstein and Anthony V. Phillips; SCIENTIFIC AMERICAN, July] starts out with a striking example of a remarkable rotation property of some objects, illustrated by a Philippine dancer doing the wine dance.... It is the fixed position of the dancer's feet with respect to the floor that makes the phenomenon so surprising. But the wine-dance pictures clarify the mystery.

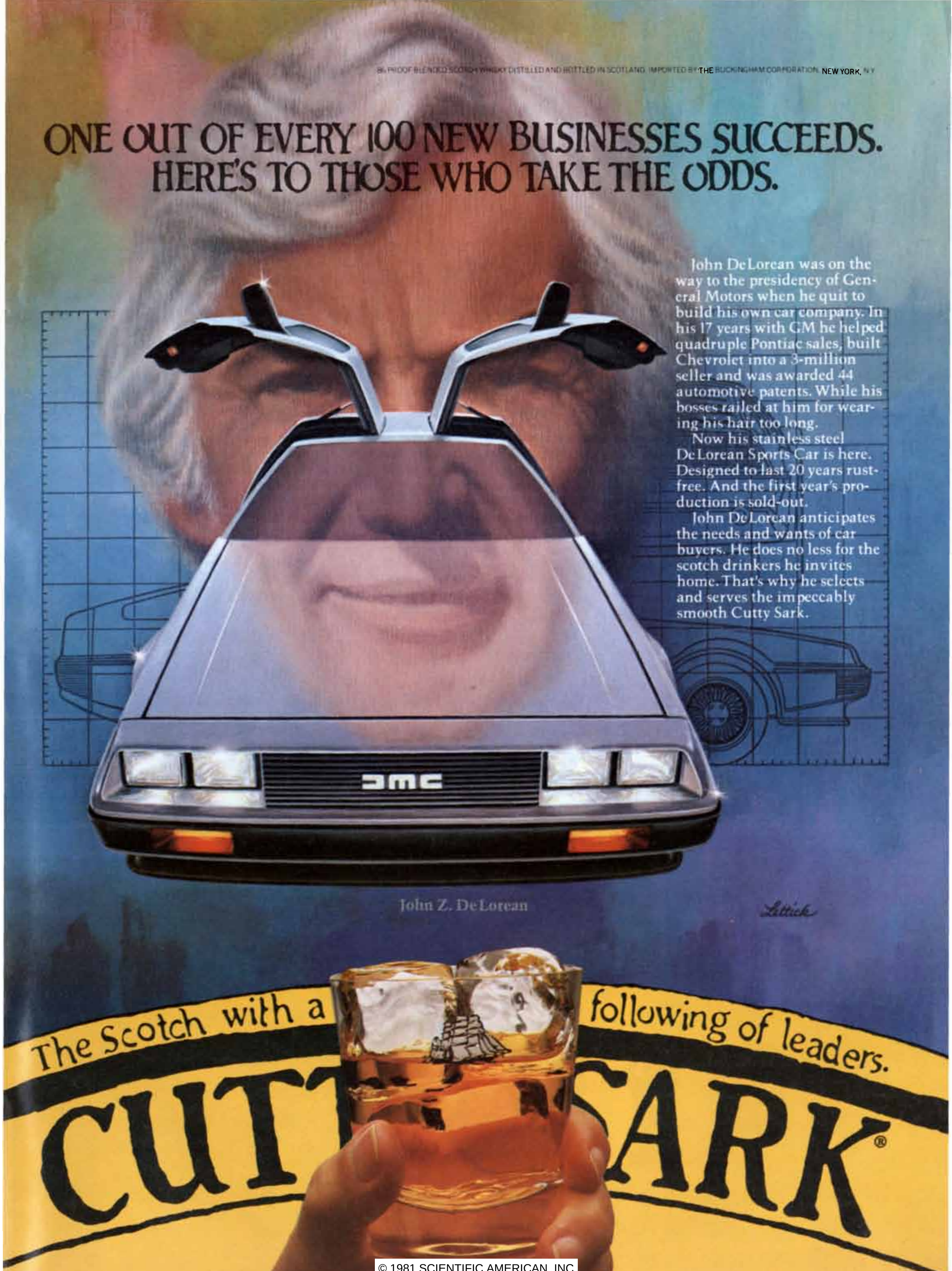
The surface of the wine is a key reference plane, and the dancer's hand is a fixed attachment to it. Consider the position of the dancer's elbow. At the start of the nine-frame series her elbow is above the wine surface. She turns her palm through 360 degrees (the position in frame 5), still keeping her elbow above the wine surface. But in frame 6 her elbow is even with the surface and in frame 7 it is well below the surface. In frame 8 her elbow is almost even with the wine surface again and in frame 9 it is back above the surface. The cycle is a key feature of mechanical examples of tangleable but self-disentangling rotation systems. (I mean a rotating object with a fixed but flexible attachment.)

Consider an idealized model consisting of a plate attached at its edge to a flat ribbon, which is in turn attached to a wall [*see illustration below*]. Begin to rotate the plate in its own plane, say counterclockwise when viewed from above. It must pass, say, over the ribbon at some point during a 360-degree rotation because it cannot pass through it. After the rotation the ribbon will have a 360-degree counterclockwise twist in it.

At this point the twist in the ribbon can be removed by at least two maneuvers each of which involves an additional 360-degree rotation of the plate in its own plane. (A third maneuver for removing the twist in the ribbon is just the "normal" untwisting one.) The secret of



***Two plate rotations need not twist a ribbon* (a–e). *If they do, the twist can be undone* (f–h)**


86 PROOF BLENDED SCOTCH WHISKY DISTILLED AND BOTTLED IN SCOTLAND. IMPORTED BY THE BUCKINGHAM CORPORATION, NEW YORK, N.Y.
ONE OUT OF EVERY 100 NEW BUSINESSES SUCCEEDS.
HERE'S TO THOSE WHO TAKE THE ODDS.
John DeLorean was on the way to the presidency of General Motors when he quit to build his own car company. In his 17 years with GM he helped quadruple Pontiac sales, built Chevrolet into a 3-million seller and was awarded 44 automotive patents. While his bosses railed at him for wearing his hair too long.
Now his stainless steel DeLorean Sports Car is here. Designed to last 20 years rust-free. And the first year's production is sold-out.
John DeLorean anticipates the needs and wants of car buyers. He does no less for the scotch drinkers he invites home. That's why he selects and serves the impeccably smooth Cutty Sark.
DMC
John Z. DeLorean
Lettick
The Scotch with a following of leaders.
CUTTY SARK®


WAVE FRONT

the first maneuver is that on the second 360-degree rotation the plate must pass under the ribbon. After the second 360-degree rotation the ribbon will again have described a 360-degree twist, but this time the twist is clockwise. The second twist cancels the first, so that after a 720-degree rotation of the plate the system is back to the start position. That is the wine-dance equivalent.

The second maneuver involves a different and less transparent operation. As before, the plate continues to rotate counterclockwise in its own plane, but now it again passes over the ribbon. As a result there are two counterclockwise twists in the ribbon. Both twists can be undone by passing the ribbon around the plate....

The examples indicate that the common notion of returning to the initial configuration after a 360-degree rotation actually corresponds to a special, degenerate situation where there are no external attachments to the rotating object. In the general case a 720-degree rotation is the one naturally needed for returning to the initial configuration, and the requirement seems to be imposed by the need for a symmetrical path of the rotating object with respect to the line of attachment. The rotating object must pass both above and below the line during its circuit....

ROSEMARIE SWANSON

College Station, Tex.

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

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

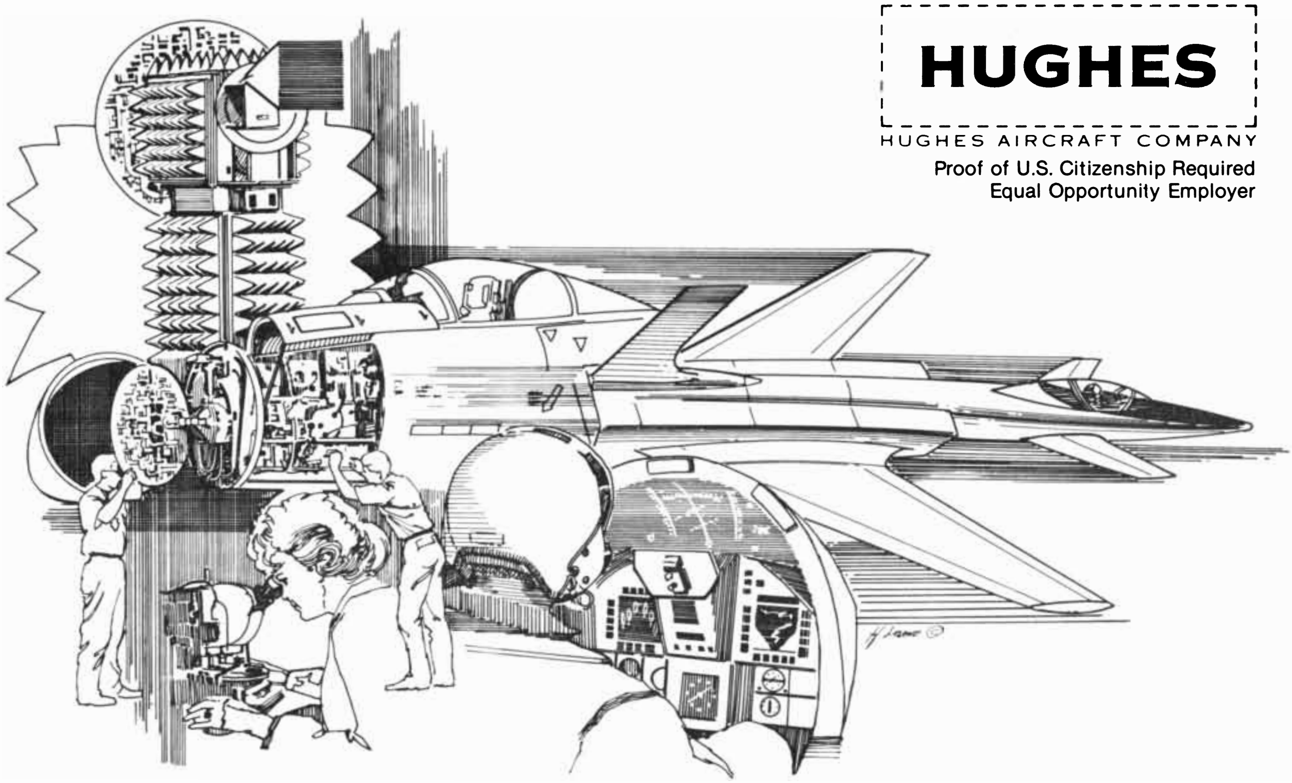
Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 60 cents.

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. For change of address, notify us at least four weeks in advance. Send both old and new addresses and enclose an address imprint from a recent issue. (Date of last issue on your subscription is shown at upper right-hand corner of each month's mailing label.)

Name__________

New Address__________

Old Address__________



50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

OCTOBER, 1931: "The sun and all other stars are tremendous heat engines, drawing energy from some internal source. It has long been known that no ordinary chemical process could produce more than a few millionths of the amount of energy required; and even the gravitational energy liberated by the gradual shrinkage of the star's mass accounts for only a few per cent. Perhaps the process is one of the building of atoms. Hydrogen, the lightest and simplest of atoms, consists of one proton and one electron. The heavier atoms have two or more electrons circulating outside a nucleus, which doubtless contains both protons and electrons locked together in some way not yet fully understood. The total number of electrons inside and outside the nucleus equals that of the protons in the nucleus, so that we might imagine the complex atom to be built up out of numerous atoms of hydrogen. But the weight of the heavier atom is in all cases less than that of the corresponding number of hydrogen atoms. This indicates that in the process of formation a large amount of energy would be liberated by the 'packing' of the constituents of the nucleus. No one, however, has even suggested from the side of atomic theory how any such atom-building process could actually occur. An important paper by R. d'E. Atkinson makes the first attempt. Starting with the recent wave theory of the constitution of atoms and their nuclei, he finds that at temperatures of 10,000,000 degrees or more a flying proton may sometimes hit the nucleus of a light atom in such a way that it goes in and sticks there, producing a new atomic nucleus of greater charge and weight. By repetitions of this process and by the capture of electrons by nuclei in a corresponding manner, heavier atoms would gradually be built up."

"The United States has a higher death rate from childbirth than any of the 20 other countries from which statistics are available, according to Howard W. Haggard, associate professor of physiology at Yale University. Available childbirth-mortality figures, he pointed out, range from 2.6 per 1,000 births to 6.6. Denmark and Italy lead the world with the lowest figure. The United States stands at the bottom of the list. 'In our country,' Dr. Haggard said, 'we have

A landmark 16-volume work... now available for the first time in 8 convenient volumes that contain every word of the original

From antiquity to the mid-twentieth century, here is the history and development of science throughout the world — presented in more than 5,000 vividly detailed biographies

More than 17 years in preparation, with over 1,500 contributors from 90 countries, including the Soviet Union, the *Dictionary of Scientific Biography* has received international acclaim.

More than a catalog of names, dates, and discoveries, the *Dictionary* covers an awesome body of knowledge clearly and precisely, helping the reader think through scientific problems, both in their historical context and in light of present-day knowledge.

- *Each article is written by an eminent historian of science; each completed by a carefully selected bibliography.*
- *A comprehensive index is included—a complete, separate volume, containing more than 75,000 topics and subtopics.*
- *The 8-volume set includes everything from the original 16 volumes—with no alterations in text or type-size.*

"May well be at this time the single most important vehicle for transmitting scholarship done by historians of science to the general world of learning."
—*Pfizer Award Committee, History of Science Society*

"A high point of historical scholarship and synthesis." —*American Scientist*

"A lasting contribution. —*Scientific American*

"The indispensable work for the scientist and the historian of science."
—*American Reference Books Annual*

READERS OF *SCIENTIFIC AMERICAN* CAN ENJOY THE PERSONAL USE OF THIS COMPLETE, PERMANENT SCIENCE LIBRARY—AT CONSIDERABLE SAVINGS.
Order the *Dictionary of Scientific Biography* through this exclusive offer, and these 8 handsome volumes can be yours at a saving of $100.00.

Awarded the 1980 Dartmouth Medal for Outstanding Reference Work of the Year

DICTIONARY OF SCIENTIFIC BIOGRAPHY

Edited under the auspices of the American Council of Learned Societies
Charles Coulston Gillispie, Editor in Chief

CHARLES SCRIBNER'S SONS
597 Fifth Avenue, New York, N.Y. 10017

ATT: Karen Day, Dept. SA
Please send me the 8-volume *Dictionary of Scientific Biography*, at $495.00* (shipping and handling included). I enclose check or money order for $495.00. **Or,** charge my
☐ Mastercharge ☐ Visa ☐ American Express
Acct. #__________ Exp. date______
*regular retail price $595.00
I understand that if I am not completely satisfied, I may return the set within 10 days for a full refund.

Name __________

Address __________

City______ State______ Zip______

☐ Please send me further information about the 8-volume DSB and about other Scribner science publications.

physicians specializing in obstetrics who are as skilled as any to be found elsewhere in the world. Our nurses are as well trained as those of any other country. We have many excellent maternity hospitals. But these means of saving lives and ensuring health have not been available to most of the mothers in this country.' "

"In a recent appeal to the scientists of the world Albert Einstein asked them to discontinue their research for the creation of new instruments of war. 'Those who believe that the danger of war is past,' he is quoted as saying, 'are living in a fool's paradise. We face today a militarism far more powerful and destructive than that which brought on the World War disaster.'"

"In July, 28 out of 57 delegations of nations signed at Geneva a convention for the limitation of narcotic drug manufacture. The drug traffic has increased at a most alarming rate during recent years, particularly in the United States. Drug addiction is not a vice; it is a disease. If the agreement to limit narcotic-drug manufacture is strictly adhered to, as is confidently expected, fewer lives will be ruined by the deadly powders."

SCIENTIFIC AMERICAN

OCTOBER, 1881: "The concluding meeting of the International Medical Congress was held on August 9 in St. James's Hall, London, where a general address was delivered by T. H. Huxley on the connection of the biological sciences with medicine. The conclusion of Professor Huxley's address was: 'The search for the explanation of diseased states in modified cell life; the discovery of the important part played by parasitic organisms in the etiology of disease; the elucidation of the action of medicaments by the methods and the data of experimental physiology—appear to me to be the greatest steps that have ever been made toward the establishment of medicine on a scientific basis. I need hardly say they could not have been made except for the advance of biology. There can be no question, then, as to the nature or the value of the connection between medicine and the biological sciences. There can be no doubt that the future of pathology and of therapeutics, and therefore that of practical medicine, depend upon the extent to which those who occupy themselves with these subjects are trained in the methods and impregnated with the fundamental truths of biology. I venture to suggest that the collective sagacity of this congress could occupy itself with no more important question than this: How is medical education to be arranged so that, without entangling the student in those details of the systematist that are valueless to him, he may be enabled to obtain a firm grasp of the great truths respecting animal and vegetable life, without which, notwithstanding all the progress of scientific medicine, he will still find himself an empiric?' "

"A device that did for cotton picking what Whitney's gin did for the work of freeing cotton lint from the seed would give an incalculable impetus to the extension of cotton culture. It is doubtful whether any phase of agricultural labor needs the aid of the inventor as badly, or promises as rich a return for a successful effort. Already a crop amounting in value to $300,000,000 or $400,000,000 is every year made difficult to secure, and subjected to serious hazard and no inconsiderable loss, through the lack of efficient harvesting machinery; and any rapid increase in the crop is prevented by the lack of laborers at the critical season. The problem is a complicated one, yet it cries aloud for a solution, and it promises a liberal reward to any who shall solve it wholly or in part. If the successful machine is reasonably simple and inexpensive, it will bring a speedy fortune to the inventor, prosperity to thousands of small planters, occupation for multitudes of mills and cheaper clothing for half the world."

"There is at present, in round numbers, 25,000,000 barrels of crude petroleum stored in iron tanks in the oil regions of Pennsylvania. It is an inland lake of oil that may be described as having reached its highest ebb, inasmuch as indications now point unmistakably to a falling off in the daily production of the wells and a consequent decline in the amount tanked. Not the least striking feature of the oil regions are the clusters of these enormous iron reservoirs, on hill and in valley, whose construction keeps actively employed great workshops and an army of men in Pittsburg, Titusville, Oil City and elsewhere. Of the 25,000,000 barrels fully one-half is owned by the United Pipe Lines (Standard Oil Company), the balance being owned by other pipe lines and by private parties. A single banking firm of New York owns half a million barrels stored in tanks and awaiting better prices."

"Ground was broken at the end of September at Portrush, Ireland, for an electric railway to the Giant's Causeway. It is the first electric railway in Britain. William Siemens is said to be a large contributor to the new enterprise. It is estimated that the expenses for haulage on a tramway such as this with horses would be 23 cents per mile and by steam about 15 cents, while it is supposed the working expenses by an electric motor will not reach two cents per mile."

ALL FEET ARE NOT CREATED EQUAL.

Those little footprints on your birth certificate aren't there for decoration. They were the best way you had of saying — I'm an individual; I'm unique.

And the moment you start treating your feet like they belong to someone else, they're going to let you know about it. Via blisters, shin splints, stress fractures or any number of other ailments.

No one has done more to get on an intimate basis with feet than Nike. We built one of the most sophisticated sports research labs in the world — the only one in the shoe industry — and staffed it with researchers in biomechanics, anatomy and exercise physiology.

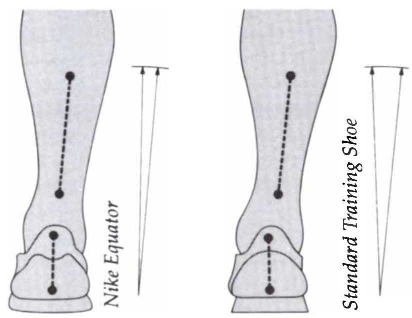
Feet, we've found, can be pretty articulate. But you have to know how to listen.

First, pay attention to their prints. There are three basic types, and you can spot yours the next time you step out of the shower.

According to our ongoing anatomy study, nearly four runners out of ten have something other than a normal arch.

Those with extremely low arches may take solace in the fact they share this trait with Henry Rono and Patti Catalano.

Unfortunately, some low arched feet overindulge. They're so flexible, they love to pronate. A little pronation is a good thing because it absorbs shock. Too much of a good thing, however, can lead to various knee and foot problems.



Two rear views of pronation, taken from high speed film. Although runner and speed are identical, angle between lower left leg and rearfoot is less in Nike Equator than standard training shoe.

To give them a bit of self-discipline, we designed the Equator. Through computer analysis of high speed film, our lab reports show this shoe reduces rearfoot motion up to five degrees, or slightly less than a hard orthotic.

The high arched foot has its own story to tell. And frequently, it's a shocker. If this foot is also rigid, as is often the case, it will do little to absorb impact.

That's why for the likes of Steve Ovett, Joan Benoit and Herb Lindsay, cushioning is everything.

We want shoes that do more than "feel" soft in the store. So we check out materials with dynamic load displacement tests. And run prototypes across force platforms to judge their shock attenuation.

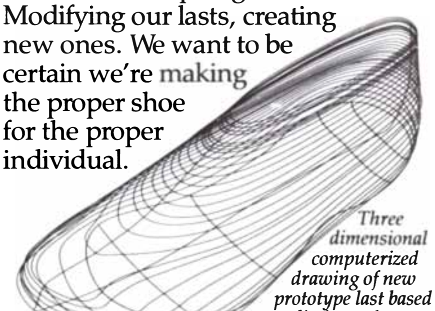
But don't think because you're blessed with a normal arch that you can give your feet just anything to wear.

After your next shower make two sets of footprints, one while sitting, one standing. If the second set is much flatter than the first, your feet are flexible. Look for a shoe with good motion control. If there's little difference, go for cushion.

Best of all, take your feet to an expert. A knowledgeable dealer can put you into the shoe your foot was created for. And if you have persistent health problems, don't mess around. See an orthopedist or podiatrist.

For our part, we're seeing runners. In the lab, at the schools, in meets and races all over the country.

We're compiling information. Modifying our lasts, creating new ones. We want to be certain we're making the proper shoe for the proper individual.



Three dimensional computerized drawing of new prototype last based on preliminary foot morphology study.

Of the thousands of runners we've seen so far, we've never met a foot we didn't like. Or couldn't help.

NIKE

Beaverton, Oregon

THE AUTHORS

LEONARD T. KURLAND and CRAIG A. MOLGAARD ("The Patient Record in Epidemiology") are members of the department of medical statistics and epidemiology at the Mayo Clinic. Kurland, who currently serves as chairman of the department, is also professor of epidemiology at the Mayo School of Medicine. A graduate of Johns Hopkins University, he got his M.D. from the University of Maryland in 1945. Following his internship he entered the U.S. Public Health Service, where he did an epidemiological study of multiple sclerosis. During this period he continued his academic work, getting two additional degrees, both in public health: an M.P.H. from the Harvard School of Public Health and a D.P.H. from Johns Hopkins. He went on to receive his clinical training in neurology at the Mayo Clinic, during which time he established the National Institutes of Health field station for neurological studies on Guam. From 1955 to 1964 Kurland was chief of the epidemiology branch of the National Institute of Neurologic Disease and Blindness. Molgaard, who is a research associate and instructor in the Mayo Clinic's department of medical statistics and epidemiology, is a graduate of Iowa State University. He did his graduate work at the University of California at Berkeley, obtaining his Ph.D. in anthropology in 1979.

DAVID I. GROVES, JOHN S. R. DUNLOP and ROGER BUICK ("An Early Habitat of Life") are at the University of Western Australia, where Groves is associate professor of geology and Dunlop and Buick are doctoral candidates in geology. Groves's degrees are from the University of Tasmania: a B.Sc. in 1963 and a Ph.D. in 1968. He worked for several years for the Geological Survey of Tasmania before joining the faculty at Western Australia in 1971. Since then, he writes, his main research interest has been in "the evolution and metallogeny of Archaean crust." Dunlop and Buick both got their B.Sc. degrees from Western Australia before entering the postgraduate geology program there.

JOEL M. WEISBERG, JOSEPH H. TAYLOR and LEE A. FOWLER ("Gravitational Waves from an Orbiting Pulsar") worked together on the subject of their article at the University of Massachusetts at Amherst. Since then Weisberg and Taylor have joined the faculty at Princeton University (respectively as assistant professor of physics and professor of physics) and Fowler has gone to work for Atmospheric and Environmental Research, Inc., in Cambridge, Mass. Weisberg has a B.S. from the Massachusetts Institute of Technology (1972) and a Ph.D. in physics from the University of Iowa (1978). In addition to his scientific work, he reports, "I lecture on the dangers of the arms race and work with local groups advocating arms control and disarmament." Taylor received his academic degrees from Haverford College (B.A. in physics, 1963) and Harvard University (Ph.D. in astronomy, 1968). An investigator of pulsars, he was the codiscoverer (with his student Russell A. Hulse) of the first pulsar known to be in an orbiting system. Fowler became interested in pulsars while he was an undergraduate at Cornell University. On getting his B.A. in physics from Cornell in 1973, he joined the pulsar-research group at the University of Massachusetts. He obtained his Ph.D. in astronomy in 1979 and then spent two years at the Max Planck Institute for Radioastronomy in Bonn before taking up his present job earlier this year.

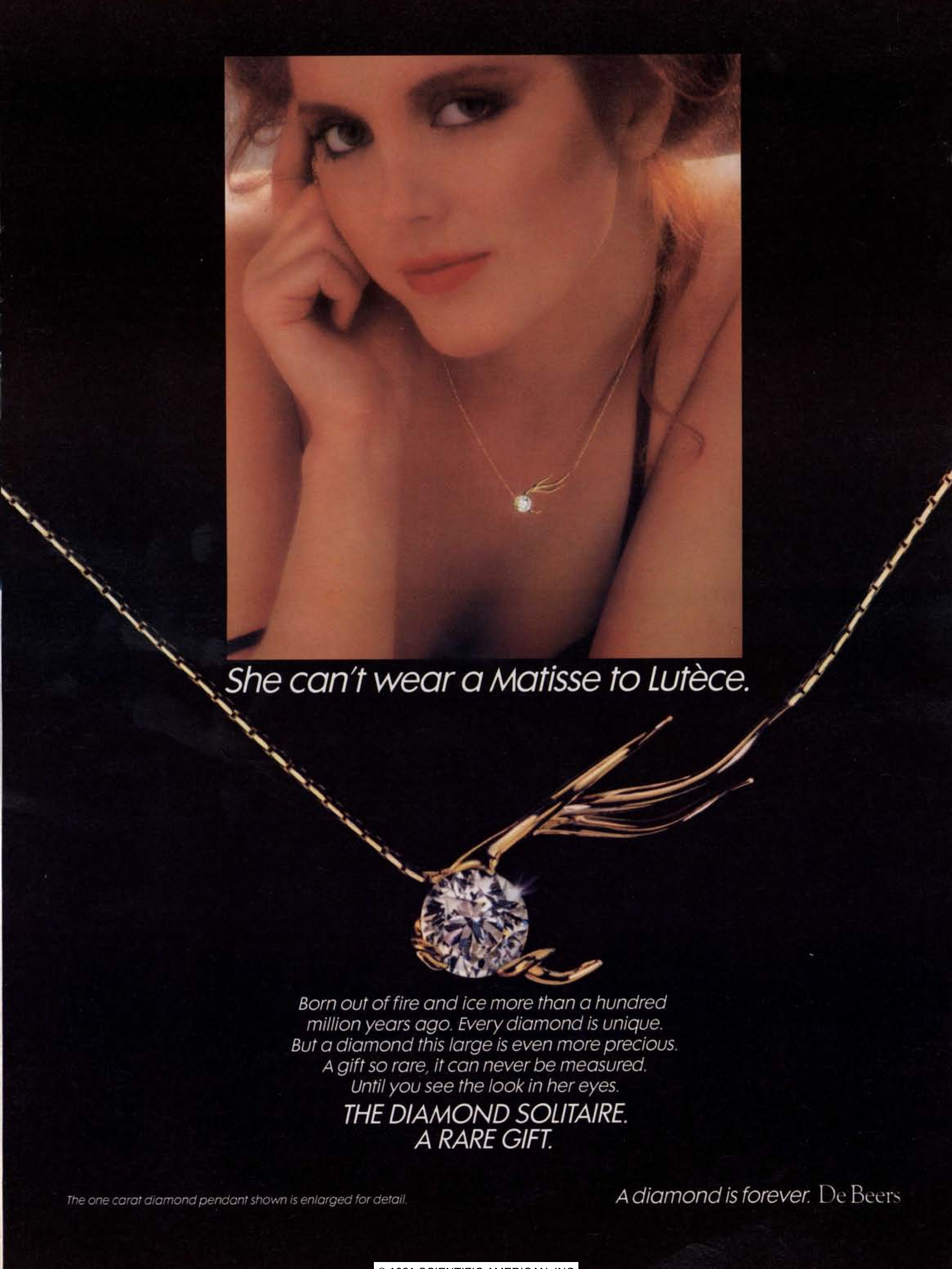
FLOYD E. BLOOM ("Neuropeptides") is director of the Arthur Vining Davis Center for Behavioral Neurobiology at the Salk Institute. He is a graduate of Southern Methodist University and the Washington University School of Medicine. Before being appointed to his current position in 1975 he was chief of the laboratory of neuropharmacology and acting director of the Division of Special Mental Health Research Programs of the National Institute of Mental Health at St. Elizabeth's Hospital in Washington. His interest in brain research, he says, "stemmed directly from my medical interest in how the brain monitors and controls blood pressure. The continued pursuit of that simple question has led me to develop and apply a variety of experimental approaches to the task of identifying where and how chemical transmitters communicate messages between nerve cells and how those signals are able to alter behavior."

CARLEEN MALEY HUTCHINS ("The Acoustics of Violin Plates") has been constructing stringed instruments and investigating their acoustics for more than three decades. To date, she reckons, she has made "150 violas, 25 violins, five cellos and some 40 instruments of the new violin family, ranging from a seven-foot contrabass violin to a tiny treble violin tuned an octave above the normal violin." For some 20 years she has served as permanent secretary of the Catgut Acoustical Society, an organization she helped to found (with Frederick A. Saunders of Harvard University) in 1963. In recent years Hutchins has been the recipient of a number of awards, including an honorary Doctor of Engineering degree from the Stevens Institute of Technology and the Silver Medal of the Acoustical Society of America.

ANDRÉ WEGENER SLEESWYK ("Vitruvius' Odometer") is professor of applied physics at the University of Groningen. He writes: "I was born in 1927 at Bindjei in Sumatra, at that time part of the Dutch overseas empire, and I spent most of my youth there, with intermissions in the Netherlands. From 1942 to 1945 I sojourned in Japanese internment camps, living under conditions I now recognize as approaching Third World poverty. After the war I was trained as a naval officer for the Royal Navy of the Netherlands. Later I studied mechanical engineering at the Delft Technological University and at the Massachusetts Institute of Technology. My Ph.D., in physics, is from the University of Amsterdam. After having worked in industry in Holland and in France I was appointed to the faculty at Groningen in 1964. I have taught physical metallurgy, thermodynamics and the history of technology, and I have published some 80 research papers in these fields."

R. I. G. HUGHES ("Quantum Logic") is assistant professor of philosophy at Yale University. A native of London, he was educated at Highgate School and the University of Cambridge. After receiving his B.A. in engineering at Cambridge in 1957, he worked in the professional theater for a time and then taught at schools in London and Canada for 13 years before starting the study of philosophy at the University of British Columbia in 1971. He got his Ph.D. from British Columbia in 1979 and has taught there, at the University of Toronto and at Princeton University; he joined the Yale faculty in 1980.

T. M. CANNON and B. R. HUNT ("Image Processing by Computer") began their collaboration while both were on the staff of the Los Alamos Scientific Laboratory. Cannon is currently spending a sabbatical year on leave from Los Alamos at the British Home Office, where he is working on the forensic applications of digital image deblurring. Hunt, who has a joint appointment as professor of electrical engineering and professor of optical sciences at the University of Arizona, is also on sabbatical leave; he is spending the year working at the University of Canterbury in New Zealand. He majored in aeronautical engineering as an undergraduate at Wichita State University and went on to obtain an M.Sc. in electrical engineering at Oklahoma State University in 1965 and a Ph.D. in systems engineering from the University of Arizona in 1967.


She can't wear a Matisse to Lutèce.
Born out of fire and ice more than a hundred
million years ago. Every diamond is unique.
But a diamond this large is even more precious.
A gift so rare, it can never be measured.
Until you see the look in her eyes.
THE DIAMOND SOLITAIRE.
A RARE GIFT.
The one carat diamond pendant shown is enlarged for detail.
A diamond is forever. De Beers

MATHEMATICAL GAMES

Euclid's parallel postulate and its modern offspring

by Martin Gardner

"Lines that are parallel
meet at Infinity!"
Euclid repeatedly,
heatedly,
urged.
Until he died,
and so reached that vicinity:
in it he
found that the damned things
diverged.

—PIET HEIN, *Grooks VI*

Euclid's *Elements* is dull, long-winded and does not make explicit the fact that two circles can intersect, that a circle has an outside and an inside, that triangles can be turned over and other assumptions essential to his system. By modern standards Bertrand Russell could call Euclid's fourth proposition a "tissue of nonsense" and declare it a scandal that the *Elements* was still used as a textbook.

On the other hand, Euclid's geometry was the first major effort to organize the subject as an axiomatic system, and it seems hardly fair to find fault with him for not anticipating all the repairs made when David Hilbert and others formalized the system. There is no more striking evidence of Euclid's genius than his realization that his notorious fifth postulate was not a theorem but an axiom that had to be accepted without proof.

Euclid's way of stating the postulate was rather cumbersome, and it was recognized early that it could be given the following simpler form: Through a point on a plane, not on a given straight line, only one line is parallel to the given line. Because this is not quite as intuitively obvious as Euclid's other axioms mathematicians tried for 2,000 years to remove the postulate by making it a theorem that could be established on the basis of Euclid's other axioms. Hundreds of proofs were attempted. Some eminent mathematicians thought they had succeeded, but it always turned out that somewhere in their proof an assumption had been made that either was equivalent to the parallel postulate or required the postulate.

For example, it is easy to prove the parallel postulate if you assume that the sum of the angles of every triangle equals two right angles. Unfortunately you cannot prove this assumption without using the parallel postulate. An early false proof, attributed to Thales of Miletus, rests on the existence of a rectangle, that is, a quadrilateral with four right angles. You cannot prove, however, that rectangles exist without using the parallel postulate! In the 17th century John Wallis, a renowned English mathematician, believed he had proved the postulate. Alas, he failed to realize that his assumption that two triangles can be similar but not congruent cannot be proved without the parallel postulate. Long lists can be made of other assumptions, all so intuitively obvious that they hardly seem worth asserting, and all equivalent to the parallel postulate in the sense that they do not hold unless the postulate holds.

In the early 19th century trying to prove the postulate became something of a mania. In Hungary, Farkas Bolyai spent much of his life at the task, and in his youth he discussed it often with his German friend Carl Friedrich Gauss. Farkas' son János became so obsessed by the problem that his father was moved to write in a letter: "For God's sake, I beseech you, give it up. Fear it no less than sensual passions because it too may take all your time and deprive you of your health, peace of mind and happiness in life."

János did not give it up, and soon he became persuaded not only that the postulate was independent of the other axioms but also that a consistent geometry could be created by assuming that through the point an infinity of lines were parallel to the given line. "Out of nothing I have created a new universe," he proudly wrote to his father in 1823.

Farkas at once urged his son to let him publish these sensational claims in an appendix to a book he was then completing. "If you have really succeeded, it is right that no time be lost in making it public, for two reasons: first, because ideas pass easily from one to another who can anticipate its publication; and secondly, there is some truth in this, that many things have an epoch in which they are found at the same time in several places, just as the violets appear on every side in spring. Also every scientific struggle is just a serious war, in which I cannot say when peace will arrive. Thus we ought to conquer when we are able, since the advantage is always to the first comer."

János' brief masterpiece did appear in his father's book, but as it happened the publication of the book was delayed until 1832. The Russian mathematician Nikolai Ivanovitch Lobachevski had beat him to it by disclosing details of the same strange geometry (later called by Felix Klein hyperbolic geometry) in a paper of 1829. What is worse, when Farkas sent the appendix to his old friend Gauss, the Prince of Mathematicians replied that if he praised the work, he would only be praising himself, inasmuch as he had worked it all out many years earlier but had published nothing. In other letters he gave his reason. He did not want to arouse an "outcry" among the "Boeotians," by which he meant his conservative colleagues. (In ancient Athens the Boeotians were considered unusually stupid.)

Crushed by Gauss's response, János even suspected that his father might have leaked his marvelous discovery to Gauss. When he later learned of Lobachevski's earlier paper, he lost interest in the topic and published nothing more. "The nature of real truth of course cannot but be one and the same in Marcos-Vasarhely as in Kamchatka and on the moon," he wrote, resigned to having published too late to win the honor for which he had so passionately hoped.

In some ways the story of the Italian Jesuit Giralamo Saccheri is even sadder than that of Bolyai. As early as 1733, in a Latin book called *Euclid Cleared of All Blemish,* Saccheri actually constructed both types of non-Euclidean geometry (we shall come to the second type below) without knowing it! Or so it seems. At any rate Saccheri refused to believe either geometry was consistent, but he came so close to accepting them that some historians think he pretended to disbelieve them just to get his book published. "To have claimed that a non-Euclidean system was as 'true' as Euclid's," writes Eric Temple Bell (in a chapter on Saccheri in *The Magic of Numbers*), "would have been a foolhardy invitation to repression and discipline. The Copernicus of Geometry therefore resorted to subterfuge. Taking a long chance, Saccheri denounced his own work, hoping by this pious betrayal to slip his heresy past the censors."

I cannot resist adding two anecdotes about the Bolyais. János was a cavalry officer (mathematics had always been

Baked Apple.

Last Thanksgiving, a designer from Lynn/Ohio Corporation took one of the company's Apple Personal Computers home for the holidays.

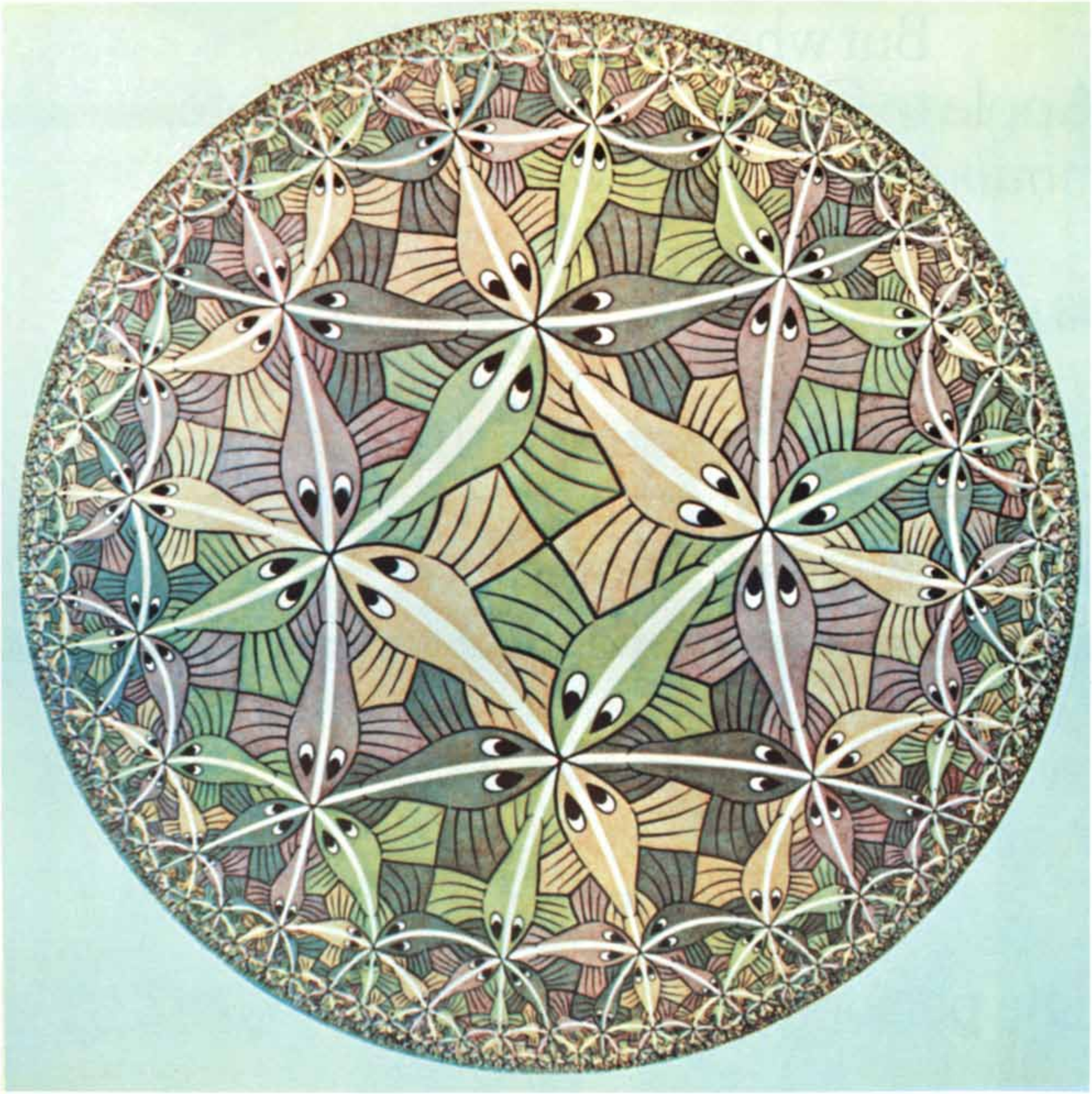
While he was out eating turkey, it got baked.

His cat, perhaps miffed at being left alone, knocked over a lamp which started

strictly a recreation) known for his swordsmanship, his skill on the violin and his hot temper. He is said to have once challenged 13 officers to duels, provided that after each victory he would be allowed to play to the loser a piece on his violin. The elder Bolyai is reported to have been buried at his own request under an apple tree, with no monument, to commemorate history's three most famous apples: the apple of Eve, the golden apple Paris gave Venus as a beauty-contest prize and the falling apple that inspired Isaac Newton.

Before the 19th century had ended it became clear that the parallel postulate not only was independent of the others but also that it could be altered in two opposite ways. If it was replaced (as Gauss, Bolyai and Lobachevski had proposed) by assuming an infinite number of "ultraparallel" lines through the point, the result would be a new geometry just as elegant and as "true" as Euclid's. All Euclid's other postulates remain valid; a "straight" line is still a geodesic, or shortest line. In this hyperbolic space all triangles have an angle sum less than 180 degrees, and the sum decreases as triangles get larger. All similar polygons are congruent. The circumference of any circle is greater than pi times the diameter. The measure of curvature of the hyperbolic plane is negative (in contrast to the zero curvature of the Euclidean plane) and everywhere the same. Like Euclidean geometry, hyperbolic geometry generalizes to three-space and all higher dimensions.

The second type of non-Euclidean geometry, which Klein named "elliptic," was later developed simultaneously by the German mathematician Georg Friedrich Bernhard Riemann and the Swiss mathematician Ludwig Schläfli. It replaces the parallel postulate with the assumption that through the point *no* line can be drawn parallel to the given line. In this geometry the angle sum of a triangle is always more than 180 degrees, and the circumference of a circle



M. C. Escher's **Circle Limit III** *is a tessellation based on a Euclidean model of the hyperbolic plane*

We believe that when you travel on business, there are certain things you deserve.

You deserve service that's not only fast, but friendly. From people who understand that warm attitudes are just as helpful to travelers as cool efficiencies.

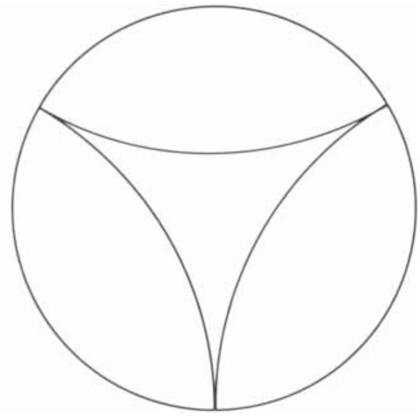
You deserve a personal welcome. A lot of respect. Extra help whenever you need it. And a commitment on our part to give you a level of attention that may surprise you, and that will certainly make renting a car from National one of the better experiences of your business trip.

Try us. Contact your travel consultant. Or call toll-free 800-328-4567. In Minnesota call 800-862-6064. In Alaska and Hawaii call 800-328-6321. In Canada call collect 612-830-2345. We're going to prove it: There *is* a difference between car rental companies, and it's called "National attention.SM"

We feature GM cars like this Buick Skylark.

You deserve National attention.

National Car Rental



Largest triangle in the hyperbolic plane

is always less than pi times the diameter. Every geodesic is finite and closed. The lines in every pair of geodesics cross.

To prove consistency for the two new geometries various Euclidean models of each geometry were found showing that if Euclidean geometry is consistent, so are the other two. Moreover, Euclidean geometry has been "arithmetized," proving that if arithmetic is consistent, so too is Euclid's geometry. We now know, thanks to Kurt Gödel, that the consistency of arithmetic is not provable in arithmetic, and although there are consistency proofs for arithmetic (such as the famous proof by Gerhard Gentzen in 1936), no such proof has yet been found that can be considered entirely constructive by an intuitionist [see "Constructive Mathematics," by Allan Calder; SCIENTIFIC AMERICAN, October, 1979]. God exists, someone once said, because mathematics is consistent, and the Devil exists because we are not able to prove it.

The various metaproofs of arithmetic's consistency, as Paul C. Rosenbloom has put it, may not have eliminated the Devil, but they have reduced the size of hell almost to zero. In any case no mathematician today expects arithmetic (therefore also Euclidean and non-Euclidean geometries) ever to produce a contradiction. Curiously, Lewis Carroll was one of the last mathematicians to doubt non-Euclidean geometry. "It is a strange paradox," the geometer H. S. M. Coxeter has written, "that he, whose Alice in Wonderland could alter her size by eating a little cake, was unable to accept the possibility that the area of a triangle could remain finite when its sides tend to infinity."

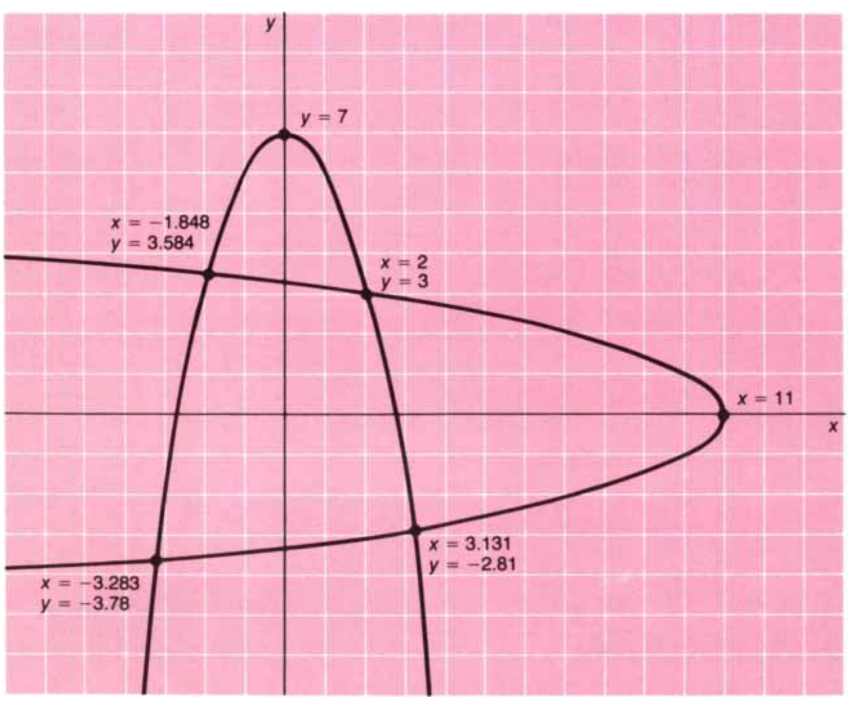
What Coxeter had in mind can be grasped by studying M. C. Escher's *Circle Limit III,* reproduced on page 26. This 1959 woodcut (one of Escher's rare works with several colors in one picture) is a tessellation based on a Euclidean model of the hyperbolic plane that was constructed by Henri Poincaré. In Poincaré's ingenious model every point on the Euclidean plane corresponds to a point inside (but not on) the circle's circumference. Beyond the circle there is, as Escher put it, "absolute nothingness."

Imagine that Flatlanders live on this model. As they move outward from the center their size seems to us to get progressively smaller, although they are unaware of any change because all their measuring instruments similarly get smaller. At the boundary their size would become zero, but they can never reach the boundary. If they proceed toward it with uniform velocity, their speed (to us) steadily decreases, although to them it seems constant. Thus their universe, which we see as being finite, is to them infinite. Hyperbolic light follows geodesics, but because its velocity is proportional to its distance from the boundary it takes paths that we see as circular arcs meeting the boundary at right angles.

In this hyperbolic world a triangle has a maximum finite area, as is shown in the top illustration on this page, although its three "straight" sides go to infinity in hyperbolic length and its three angles are zero. You must not think of Escher's mosaic as being laid out on a sphere. It is a circle enclosing an infinity of fish—Coxeter calls it a "miraculous draught"—that get progressively smaller as they near the circumference. In the hyperbolic plane, of which the picture is only a model, the fish are all identical in size and shape. It is important to remember that the creatures of a hyperbolic world would not change in shape as they moved about, light would not change in speed and the universe would be infinite in all directions.

The curved white lines in Escher's woodcut do not, as many people have supposed, model hyperbolic geodesics. The lines are called equidistant curves or hypercycles. Each line has a constant perpendicular distance (measured hyperbolically) from the hyperbolic straight line that joins the arc's ends. Note that along each white curve fish of the same color swim head to tail. If you consider all the points where four fins meet, these points are the vertexes of a regular tiling of the hyperbolic plane by equilateral triangles with angles of 45 degrees. The centers of the triangles are the points where three left fins meet and three mouths touch three tails. The 45-degree angles make it possible for eight triangles to surround each vertex, where in a Euclidean tiling by equilateral triangles only six triangles can surround each vertex.

Escher and Coxeter had corresponded from the time they met in 1954, and Escher's interest in tilings of the hyperbolic plane had been aroused by the illustrations in a 1957 paper on crystal symmetry that Coxeter had written and sent to him. In a lovely article titled "The Non-Euclidean Symmetry of Escher's Picture 'Circle Limit III'" (in the journal *Leonardo,* Vol. 12, pages 19–25; 1979) Coxeter shows that each white arc meets the boundary at an angle of almost 80 degrees. (The precise value is $2^{7/4} + 2^{5/4}$ arc-seconds.) Coxeter considers *Circle Limit III* the most mathematically sophisticated of all Escher's pictures. It even anticipated a discovery



Answer to the 7–11 problem


IMPORTED
Seagram's
V.O.
CANADIAN WHISKY
Taste!
Whether you mix it in a whisky sour, or pour it on the rocks, you'll see
why more V.O. is bought than any other imported distilled spirit in America.
More than any Scotch, Canadian, Rum, Gin, Vodka or Tequila.
VO
Enjoy our quality in moderation.
CANADIAN WHISKY. A BLEND OF CANADA'S FINEST WHISKIES.
6 YEARS OLD. 86.8 PROOF. SEAGRAM DISTILLERS CO., N.Y.C.

Coxeter did not make until five years after the woodcut was finished!

Elliptic geometry is roughly modeled by the surface of a sphere. Here Euclidean straight lines become great circles. Clearly no two can be parallel, and it is easy to see that triangles formed by arcs of great circles must have angles that add up to more than two right angles. The hyperbolic plane is similarly modeled by the saddle-shaped surface of a pseudosphere, generated by rotating a tractrix about its asymptote.

It is a misuse of the word "crank" to apply it to mathematicians who erred in thinking before the independence of the parallel postulate was established that they had proved the postulate. The same cannot be said of those amateurs of later decades who could not understand the proofs of the postulate's independence or who were too egotistical to try. Augustus De Morgan, in his classic compendium of eccentric mathematics, *A Budget of Paradoxes,* introduces us to Britain's most indefatigable 19th-century parallel-postulate prover, General Perronet Thompson. Thompson kept issuing revisions of his many proofs (one was based on the equiangular spiral), and although De Morgan did his best to dissuade him from his futile efforts, he was unsuccessful. Thompson also wanted to replace the tempered scale of the piano with an octave of 40 notes.

The funniest of the American parallel-postulate provers was the Very Reverend Jeremiah Joseph Callahan, president of Duquesne University in Pittsburgh. In 1931, when Father Callahan announced he had trisected the angle, *Time* gave the story sober treatment and ran his photograph. The following year Callahan published his major work, *Euclid or Einstein: A Proof of the Parallel Theory and a Critique of Metageometry* (Devon-Adair, 1932), a 310-page treatise in which he ascended to heights of *argumentem ad hominem.* Einstein is "fuddled," he "has not a logical mind," he is in a "mental fog," he is a "careless thinker." "His thought staggers, and reels, and stumbles, and falls, like a blind man rushing into unknown territory." "Sometimes one feels like laughing," Callahan wrote, "and sometimes one feels a little irritated.... But there is no use expecting Einstein to reason."

What Callahan found so irritating was Einstein's adoption of a generalized non-Euclidean geometry, formulated by Riemann, in which the curvature of physical space varies from point to point depending on the influence of matter. One of the great revolutions brought about by relativity theory was the discovery that an enormous overall simplification of physics is obtained by assuming physical space to have this kind of non-Euclidean structure.

It is now commonplace (how astonished, and I think delighted, Kant would have been by the notion!) to recognize that all geometric systems are equally "true" in the abstract but that the structure of physical space must be determined empirically. Gauss himself thought of triangulating three mountain peaks to see if their angles added up to two right angles. It is said he actually made such a test, with inconclusive results. Although experiments can prove physical space is non-Euclidean, it is a curious fact that there is no way to prove it is Euclidean! Zero curvature is a limiting case, midway between elliptic and hyperbolic curvatures. Since all measurement is subject to error, the deviation from zero could always be too slight for detection.

Poincaré held the opinion that if optical experiments seemed to show physical space was non-Euclidean, it would be best to preserve the simpler Euclidean geometry of space and assume that light rays do not follow geodesics. Many mathematicians and physicists, including Russell, agreed with Poincaré until relativity theory changed their mind. Alfred North Whitehead was among the few whose mind was never changed. He even wrote a book on relativity, now forgotten, in which he argued for preserving a Euclidean universe (or at least one of constant curvature) and modifying the physical laws as necessary. (For a discussion of Whitehead's controversy with Einstein, see Robert M. Palter's *Whitehead's Philosophy of Science,* University of Chicago Press, 1960.)

Physicists are no longer disturbed by the notion that physical space has a generalized non-Euclidean structure. Callahan was not merely disturbed; he was also convinced that all non-Euclidean geometries are self-contradictory. Einstein, poor fellow, did not know how easy it is to prove the parallel postulate. If you are curious about how Callahan did it, and about his elementary error, see D. R. Ward's paper in *The Mathematical Gazette* (Vol. 17, pages 101–104; May, 1933).

Like their cousins who trisect the angle, square the circle and find simple proofs of Fermat's last theorem, the parallel-postulate provers are a determined breed. A recent example is William L. Fischer of Munich, who in 1959 published a 100-page *Critique of Non-Euclidean Geometry.* Ian Stewart exposes its errors in the British journal *Manifold* (No. 12, pages 14–21; Summer, 1972). Stewart quotes from a letter in which Fischer accuses establishment mathematicians of suppressing his great work and orthodox journals of refusing to review it: "The university library at Cambridge refused even to put my booklet on file.... I had to write to the vice-chancellor to overcome this boycott."

There are, of course, no sharp criteria for distinguishing crank mathematics from good mathematics, but then nei-

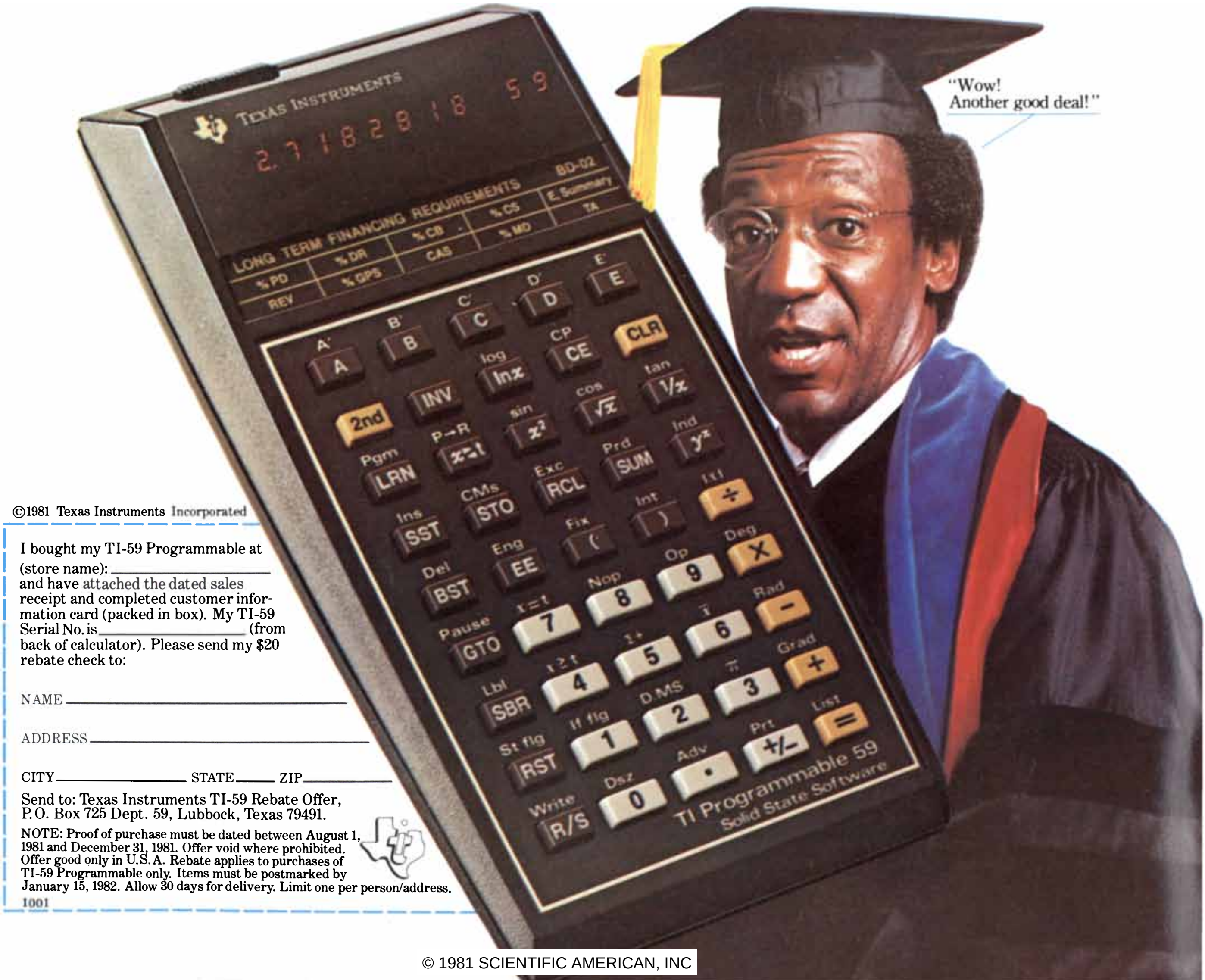
Get a $20 rebate on the TI-59 Programmable.

Even without the $20 rebate, the TI-59 is special —it's our most powerful programmable, and we've never offered it at a lower price.

The TI-59 gives you up to 960 program steps, or up to 100 memories, plus magnetic card read/write capability. You can also slip in one of TI's Solid State Software™ modules and successfully attack complex engineering, business, statistical and scientific problems. And by adding the optional PC-100C printer, you can record your calculations. But if that's not enough, any TI-59 owner can join our Professional Program Exchange for access to over 2500 additional programs.

So if you like the idea of having real programmable power, take us up on the rebate offer. Buy a TI-59 now, and fill out the coupon below. The offer ends December 31, 1981.

TEXAS INSTRUMENTS
INCORPORATED



I bought my TI-59 Programmable at (store name): ____________ and have attached the dated sales receipt and completed customer information card (packed in box). My TI-59 Serial No. is ____________ (from back of calculator). Please send my $20 rebate check to:

NAME ____________

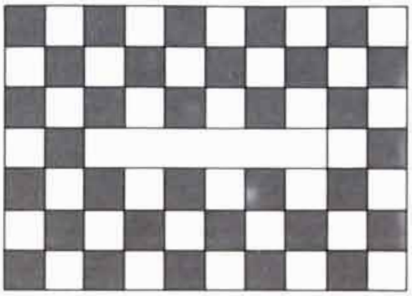
ADDRESS ____________

CITY ____________ STATE ______ ZIP ______

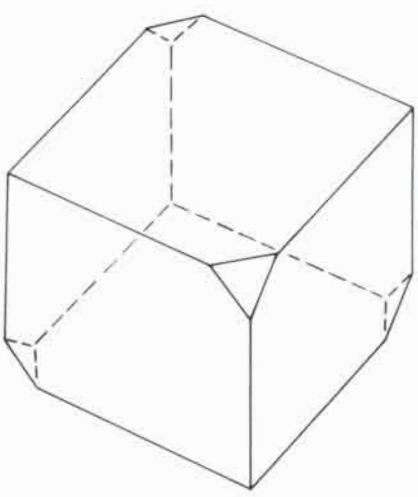
Send to: Texas Instruments TI-59 Rebate Offer, P.O. Box 725 Dept. 59, Lubbock, Texas 79491.

NOTE: Proof of purchase must be dated between August 1, 1981 and December 31, 1981. Offer void where prohibited. Offer good only in U.S.A. Rebate applies to purchases of TI-59 Programmable only. Items must be postmarked by January 15, 1982. Allow 30 days for delivery. Limit one per person/address.

1001



New chessboard-dissection puzzle



Truncated cube that forces six colors

ther are there sharp criteria for distinguishing day from night, life from nonlife and where the ocean ends and the shore begins. Without words for parts of continuums we could not think or talk at all. If you, dear reader, have a way to prove the parallel postulate, don't tell me about it!

The first problem in my August column asked for the use of a parabola to provide a quick proof of the number of solutions for the pair of equations: $x^2 + y = 7$ and $x + y^2 = 11$. The equations graph as two crossed parabolas, as is shown in the bottom illustration on page 28. Since the parabolas intersect at just four points, there are just four solutions. Only one ($x = 2$, $y = 3$) is in integers. Even if your graphing is imprecise, you can prove that the other three solutions are not in whole numbers by testing the numbers indicated by the lattice crossings nearest the intersections. The actual numbers, all of them irrational, have the approximate values given in the illustration.

The second problem, from Ronald L. Graham, asked how an infinite number of identical disks with diameters of less than 1, say 1/10, could be placed in the plane so that no two points in the disks are an integral distance apart. One solution is to place them with their centers on a parabola of the form $y = x^2$, with their centers at points (1,1), (3,9), (9,81),..., (3^k,3^{2k}),.... I lack space for Graham's unpublished proof, and so I reluctantly leave it as an exercise for interested readers.

One of the short problems in my April column called for the cutting of a patchwork quilt of size 9 by 12, with a 1-by-8 hole, into two parts that would make an undamaged quilt of size 10 by 10. Eric Stott wrote to say that if the damaged quilt is colored like a chessboard, the coloring is destroyed when the square is formed. In how few pieces, he asked, can the original be cut to make a square that preserves the coloring?

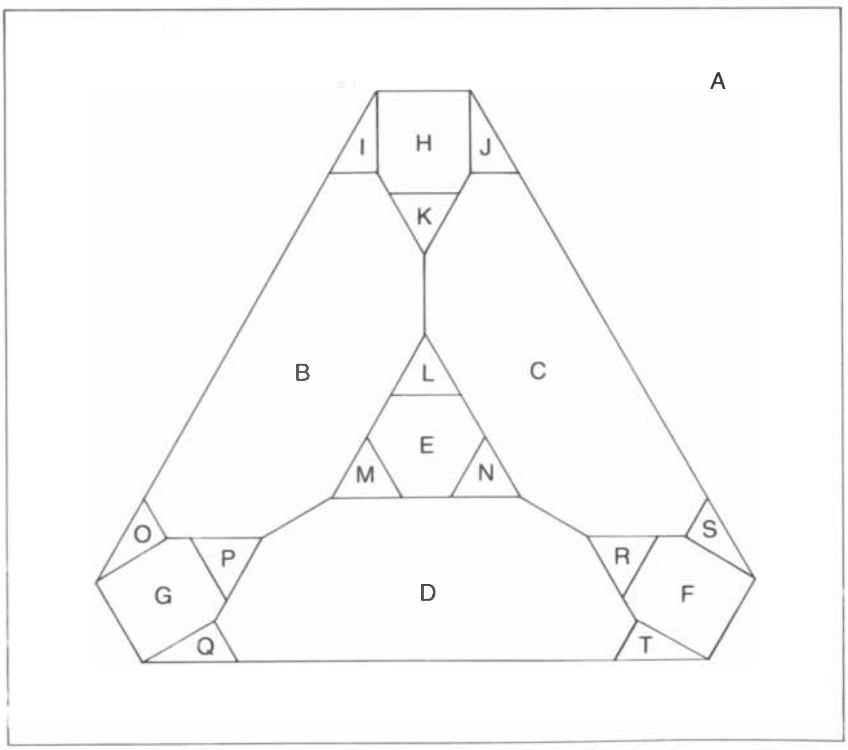
It proved to be an interesting question. There are hundreds of solutions with four pieces, but I finally found one (which I suspect is basically unique) that has just three. It can be generalized to all squares with even sides and (with slight variations) to all odd-order squares. If n is the square's side and is greater than 2, we assume the rectangle to be $(n - 1) \times (n + 2)$, with a hole of $1 \times (n - 2)$ parallel to the rectangle's long side and as centered as possible.

I give the new problem here with n equal to 8. Can you cut the mutilated rectangle shown in the upper illustration on the preceding page along lattice lines into three pieces that will form a perfect chessboard? You are allowed to rotate pieces and to turn them over. In my next column I shall give the solution.

Steven J. Brams's map-coloring-game problem, also in my April column, led Robert High, a mathematician with Informatics, Inc., of New York City, to some surprising results. Call the first player Min (who tries to minimize colors) and the second player Max (who tries to maximize colors). I had given a map of six regions on which Max can force five colors. High found that a projection of the cube gives a simpler six-region map for forcing five colors. Max merely places at each turn a new color on the face opposite Min's last play.

The map forcing six colors, which I gave in June, was a projection of the skeleton of a dodecahedron. High found that if four corners of a cube are replaced by triangular faces, as in the lower illustration on the preceding page, a projection gives a map of only 10 regions on which Max can force six colors. The strategy is less elegant than the one on the dodecahedral 12-region map and is a bit too involved to give here.

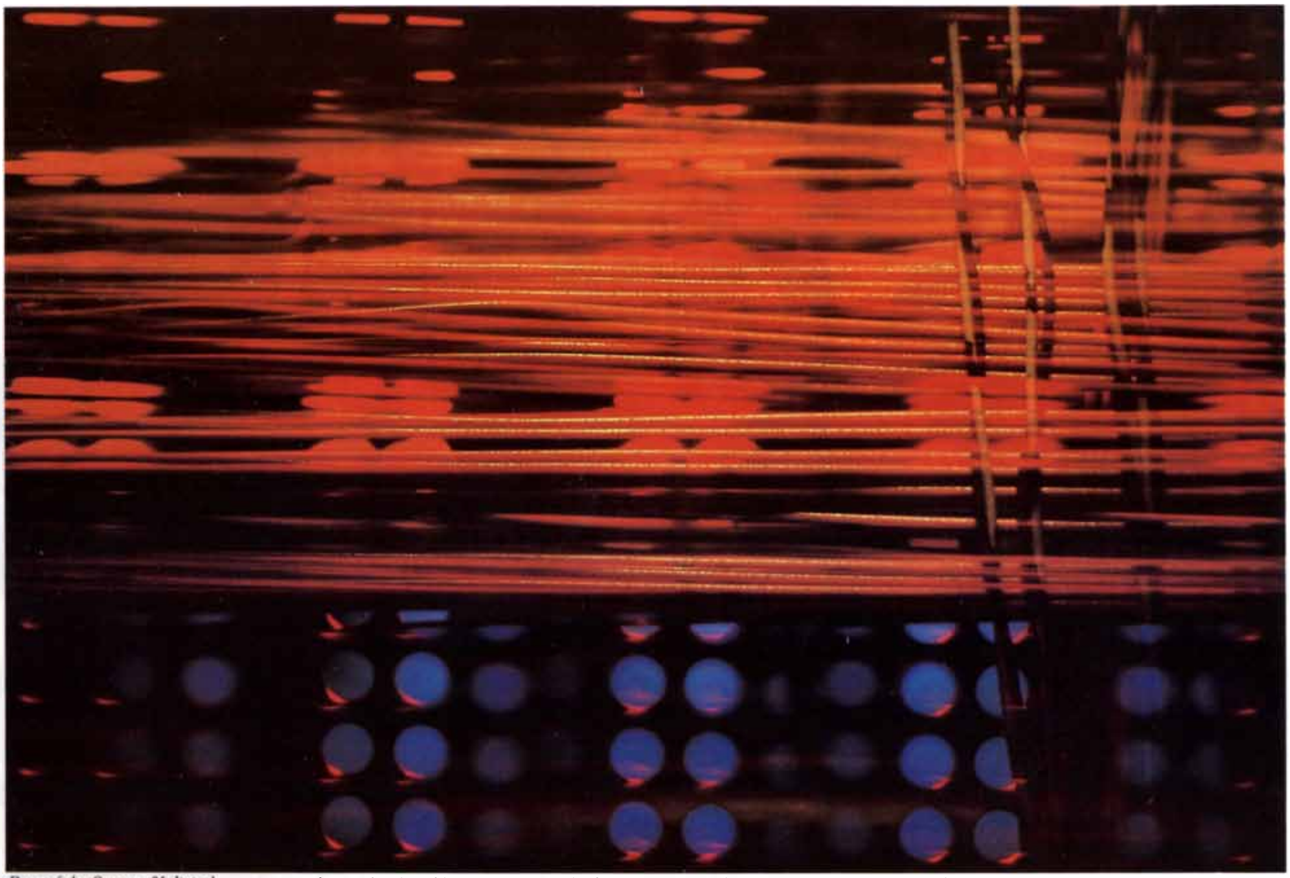
The biggest surprise was High's discovery of a 20-region map on which Max can force seven colors! Imagine each corner of a tetrahedron replaced by a triangular face, then each of the 12 new corners replaced by a triangular face. The illustration below shows a planar projection of the resulting polyhedron's skeleton. Here the strategy is more complex than the one for High's 10-region map, but he sent a game tree that proves the case. High conjectures that no planar map allows Max to force more than seven colors, but this remains unproved. It is even possible that there is no minimum upper bound.



Twenty-region map that forces seven colors

European Teleinformatics:

CREATING INTEGRATED SERVICES DIGITAL NETWORKS



Part of the System X digital switching exchange developed by GEC, ITT subsidiary STC, and Plessey together with British Telecom.

A special report on European Teleinformatics prepared by Joel Stratte-McClure

Recent technological developments and novel product applications in telecommunications, electronic data processing (EDP) and related high technology industries have merged numerous disciplines to produce a global communications revolution. The decreasing price and increasing performance of microprocessors and high speed integrated circuits (ICs) have created a myriad of new devices and approaches for collecting, storing, displaying, manipulating and moving information. The expanding fusion of the various disciplines will ultimately touch and transform every man-machine and machine-machine interface – including the 500 million telephone lines, over one million telexes and more than two million data terminals connected to public switched telephone networks (PSTN) throughout the world.

The integration of various converging disciplines is occurring so quickly that even nomenclature is proving difficult. The French call the overall development *télématique*, the combination of the French words for telecommunications and data processing. The International Telegraph and Telephone Consultative Committee (CCITT), a branch of the 154 nation Geneva-based International Telecommunication Union responsible for end-to-end standardization of international networks, places the new non-voice services under the umbrella of teleinformatics. Individual governments and companies, all formulating strategies for the technical, commercial and consumer future of communications, have their own sobriquets: compunications, advance communications service, information processing, new information technologies, office automation, the office revolution, office of the future, information management, the interactive information network, the information century, the informative marketplace, integrated information processing systems and teleprocessing – to randomly name some common appellations.

Teleinformatics basically incorporates telephones, television, facsimile and computer terminals, and other media into an harmonious network which permits data and other forms of information to be transferred speedily using advanced digital transmission and switching techniques. During the next few years a diverse range of business and consumer products and services will transform the telephone line and other links into multifunctional tools which will serve a multitude of communication requirements.

The technical cross-fertilization will ultimately affect global, national, community, office and home environments as an increasing number of clerical chores are handled electronically. Efficient digital public and private switching exchanges with stored program control (SPC) are replacing electromechanical systems; digital transmission systems using pulse code modulation (PCM) are permitting the rapid transport of voice, data, image and text; communicating word processors are replacing typewriters; low-cost intelligent facsimile machines are replacing copiers; inexpensive electronic storage systems are replacing bulk filing cabinets; and multi-purpose work stations are replacing the traditional office desk. The result will be considerable savings in time, space, trans-

They're part of the digital generation

The digital generation is upon us. And for youngsters just like these and millions worldwide, growing up in the 80's will mean becoming a part of that generation. This generation will discover all the advanced services that will benefit the home and office linked to integrated digital networks.

At Alcatel, we are pioneers in digital communications. We were the first company in the world to introduce digital technology to telephone subscribers, and that was achieved some ten years ago.

Today, one out of every two digital telephone central offices operating worldwide, is produced by CIT-Alcatel.

We feel proud to be an integral part of the digital generation we've helped create, and proud of the generation we strive to serve.

CIT-ALCATEL
33, rue Emeriau
75015 Paris - FRANCE
Tél. : (1) 571.10.10
Télex : 250927 F

port costs and energy as well as a flurry of debates about the political, social, financial and technical ramifications of numerous new products and systems reaching the office and consumer marketplace.

The prospective market is understandably large in many areas of teleinformatics. Arthur D. Little Inc. says the annual world capital expenditure for telecommunications in 1980 was $74.8 billion and projects that amount will climb to $161.8 billion in 1990. Sales of office automation equipment are expected to reach $90 billion by the year 2000 with tremendous growth in distributed terminals. Mackintosh Consultants, a British market research company, says the European IC market will quadruple by 1990. The market for fiber optics, personal computers and other teleinformatic products is still relatively untapped.

American and Japanese manufacturers maintain a commercial and industrial lead in numerous areas of teleinformatics, particularly in the fields of EDP and microelectronic components. The U.S. has benefitted from a vast home market with a high telephone density open to numerous innovations, while the Japanese have formed a common industrial strategy enabling them to increase penetration of international markets.

In fact, much of the world market for telecommunications is a captive one, and Philips of the Netherlands estimates that only 13-15 per cent is open to true competition. While North America, Asia, the Middle East and South America are important markets, a substantial portion of the technical future will involve Europe which contains one-third of the world's telecommunications market, 26 per cent of the EDP market and 19 per cent of the IC market.

European companies are active in all sectors of teleinformatics and are battling to increase market penetration outside their own territories. Domestically they are instituting national telecommunications, components and data processing programs to modernize their networks and prepare for the technical and commercial future of teleinformatics. Indeed, they believe they will lead the way in some areas.

Says Donald Ham, director of marketing and vice president of ITT Europe in Brussels, "Europeans have been ahead of the Americans and Japanese in innovative telecommunications systems such as videotex, teletex and other new product areas because of their ability to introduce systems in domestic networks."

EUROPE'S STRENGTHS AND WEAKNESSES

Europe's strengths are paradoxically its weaknesses. The government-controlled Post, Telephone and Telegraph (PTT) administrations provide determined and quickly implemented national programs, nationwide product specifications and standards, protection for domestic manufacturers, the much needed infusion of capital and other attributes.

"The PTT in France has been much more innovative than industry in defining the future of *télématique*," says Hervé Nora, head of new products and services for the Director General of Telecommunications.

"PTTs can introduce network systems and have the financial clout to do it quickly," adds Rolf-Dieter Leister at AEG-Telefunken in Frankfurt. "They provide a captive market and long-term direction for industry, but they also have their negative aspects."

These negative factors stymie the creation of a pan-European competitive market with standards that would enable Europeans to compete on a continental and worldwide scale with the resulting increase in product volume and establishment of product standards – which would assist Europeans in penetrating international markets at competitive prices.

Throughout Europe the emerging trend is to harmonize national products to European and international standards while lobbying for the concept of Europewide open markets. Numerous organizations are actively working towards these goals including the European Economic Community (EEC) in Brussels and the 26 member Conference of European Postal and Telecommunications Administrations (CEPT).

COLLABORATION OR COMPETITION

EEC initiatives have led to the creation of a pan-European data communications network, Euronet, and there is also a collaborative effort in the area of satellites through the European Space Agency. But outside these sectors Europeans have a substantial way to go before the establishment of a harmonized and deregulated European market which would be similar in size to that of the U.S. Individual countries and companies are lobbying at the CCITT for international standards based on *their* products but there is rarely a European effort to lobby for European standards.

"There's no question that in numerous areas of teleinformatics the Europeans, because of their advanced national programs, are setting the pace," says Eckart Hummel, a CCITT senior councillor in Geneva, "but the French, British or Germans rarely want the same thing because they have to protect themselves."

Nor have Europeans managed to collaborate on a commercial level. The microelectronics industry is an appropriate example. Europe currently imports 68 per cent of its IC requirements and during the past few years most European countries and/or companies launched all-encompassing components programs to establish technical and commercial independence. Many of these national programs have involved licensing agreements, joint ventures or acquisitions of U.S. firms – anything but a pan-European effort.

In fact, the EEC seems to be making little progress – largely for political, social and chauvinistic reasons – in convincing Europeans to cooperate.

"The history and tradition of Europe work against an overnight solution to the problem of pan-European cooperation," says a source at the EEC in Brussels. "The only way Europeans will ever get together is when it is financially beneficial and there is absolutely no way to proceed alone."

JOINT EFFORTS

Despite this situation there remains a great deal of discussion about future collaborative projects in Europe and considerable activity on a national, pan-European and international level as companies form alliances to consolidate their position in teleinformatics.

On a national level, for example, LM Ericsson has purchased computer manufacturer Data Saab. Three British manufacturers – GEC, ITT subsidiary STC, and Plessey – joined forces with British Telecom to develop the System X digital switching exchange. Telettra and Italtel in Italy have signed a memorandum of understanding to expand and market the Proteo switching exchange, while Olivetti is taking an interest in SGS-ATES. The French Matra group is rapidly expanding its interests, recently purchasing a large publishing and distribution concern.

On a pan-European level, LM Ericsson and Philips, with Bell Canada, are creating the $5 billion Saudi Arabian telephone network. Thomson-CSF and Siemens are collaborating on the $1.8 billion telecommunications project in Egypt. The French industrial group Saint Gobain-Pont-à-Mousson, which purchased stakes in Cii-Honeywell Bull and Olivetti, is currently prospecting for a telecommunications company. Philips has taken a 25 per cent interest in Grundig in West Germany.

On an international level, there are numerous agreements as European companies attempt to expand their market to the U.S. and other countries. To name a few, CIT-Alcatel has taken a 25 per cent share in Semi Process Industries; Thomson-CSF has marketing agreements with Continental Telephone, Xerox and 3M in the U.S.; Philips acquired Signetics; LM Ericsson has an agreement with Annaconda and Siemens has working arrangements with a number of foreign firms including Intel, Fujitsu, Corning Glass Works and Superior Cable. Most large companies have also purchased a string of U.S. concerns.

Meanwhile multinationals – especially ITT, IBM and Philips – are relying on their affiliates throughout Europe to expand their commercial and technical activities in various European countries.

"European companies must be open to joint ventures which will enable them to increase their markets," says Michele Principe, managing director of Italy's STET. "In Europe we must define a strategical line of cooperation which will result in numerous cross-licensing agreements between companies."

SUBSCRIBING TO DIGITAL SWITCHING

One area in which Europeans have assumed a substantial lead is in the application of time division multiplexing (TDM) switching and PCM transmission. These techniques became feasible when relatively low cost computering techniques were introduced to perform data processing chores – including call processing, circuit control, fault finding and administrative services. For example, microprocessors service subscriber lines and generate signals for making connections; signal processors at central exchanges control processing of calls; and computers at management centers perform overall operations supervision and maintenance functions. The combination of these digital techniques is the integrated digital network (IDN).

The European trend towards digital switching is an evolutionary one which will ultimately take advantage of digital techniques which boost productivity due to the increased use of integrated components that are decreasing in price; provide greater flexibility in engineering and operation while decreasing instalment costs; improve the quality of subscriber service due to non-blocked switching networks; allow a modular design and overall use of plug-in cards which will result in reduced installation time; require less surface space; have SPC features which allow more flexibility and reduced maintenance; and have higher quality speech channels.

More importantly, TDM switching and PCM transmission provide networks which are flexible enough to permit the connection of multiple terminals and the possibility of dialogue from terminal to terminal without costly interfaces.

Most large European companies have been or will soon begin manufacturing competitive digital systems for European and foreign markets; and the battle to sell these exchanges is fierce with rich rewards.

THE INTEGRATED SERVICES DIGITAL NETWORK

An important aspect of European teleinformatics today is the introduction of integrated services digital networks (ISDNs), networks capable of switching and transmitting voice, data and image with bidirectional 64 kilo bits per second (kbit/s) transmission rates using the public switching facilities which can access large bandwidth transmission systems, including satellites with TDM access techniques and fiber optic cables.

A number of voice, data and visual services will be progressively offered to the public when an ISDN is available. For example, it will permit a range of terminal types, from a simple digital telephone to a sophisticated data terminal, and a customer will have a variety of terminals either sharing a single line or connected as extensions to a private branch exchange (PBX) which has digital access to the local exchange.

The potential usefulness of the ISDN, explains Graham Oliver of British Telecom Research Laboratories, will depend on: the provision of adequate interworking capabilities with other networks so customers can communicate with all existing network users; the spread of the new service over a wide geographical area so customers can be rapidly connected to the evolving ISDN; and the emergence of international standards so international communications can take place using common protocols.

European companies believe the ISDN is an extremely important and integral aspect of the evolution of telecommunications. Individual companies and countries are advocating international standards for ISDN and the CEPT is working to standardize basic access while studying a variety of issues including: guidelines for handling strategy for multi-service digital traffic; network architecture; the definition of possible services; identification of access features; transmission guidelines; interface guidelines and signaling principles.

DIGITAL TRANSMISSION

The growth of TDM exchanges and increased data transmission, which currently accounts for only three per cent of total traffic, has resulted in the demand for PCM digital transmission systems which will coexist with analog networks and soon become the norm in telecommunications. Consequently transmission systems are being introduced employing hierarchial operational channels (2,8,34,140 and 560 mbit/s digital paths) to transmit over coaxial cable, microwave, radio relay, optical fibers and satellites.

Orders for digital equipment are increasing throughout Europe due to the continuing decrease in costs of electronic components and integrated circuits which make digital transmission an economical solution. Digital transmission techniques, in combination with TDM switching, have a number of cost-saving benefits: they reduce the required amount of coding and multiplexing equipment, more circuits can be established on existing cables than by employing analog techniques and they can be used for a variety of transmission services.

"The future for everyone in Europe in the PCM area," says CIT Alcatel's technical director Jacques Eldin, "is to increase circuit capacity which, of course, decreases the price per line."

OPTICAL FIBERS

Fiber optical cables – which accelerate the introduction of digital transmission techniques – are going to play a formidable role in Europe.

The advantages of fiber optical over traditional copper cables are many and include: large bandwidth with low loss; reduced size and weight; immunity to electromagnetic interference; electrical isolation; a better utilization of cable ducts; resistance to corrosion and high capacity. For example, two fiber strands can carry in excess of 10,000 telephone conversations as against 48 on copper cables. On-off audio, video and digital pulse signals are sent through the fibers, about the thickness of a human hair, by minute flashing pulses of light generated by a laser or light emitting diode. The signal is sent in the same manner whether it represents telephone, television, data or other transmission.

Ongoing projects in Europe are numerous and some recent developments include: the planned wiring of 1500 homes in Biarritz, France; an agreement between BICC, the British cable and construction group, and world leader Corning Glass Works to manufacture 100,000 km of fiber per year in the United Kingdom; another plant in Britain being built by STC, which is testing an underwater link in Loch Fyne in Scotland, with a capacity of 25,000 km a year; British Telecom's installation of 3,500 km of fiber in Britain's telephone network by the end of 1982, as the first phase of a program to go over to fiber completely for new trunk systems from the mid 1980s; the wiring of 24 homes in Berlin by AEG-Telefunken; a second link to Western Union in New York being supplied by CIT-Alcatel; sale of fiber optical cables from Philips for the Saudi Arabia network; and the development of a family of optical fiber turnkey systems by numerous companies including Thomson-CSF, GEC, Siemens, LM Ericsson and Telettra. In addition, firms throughout Europe are becoming specialists in producing the optoelectronic components required for transmitting and receiving the light signals as well as repeaters required for the cable.

SATELLITE COMMUNICATIONS

Europeans, largely through cooperative projects conducted under the auspices of the European Space Agency (ESA), are planning to launch numerous satellites for new applications in large bandwidths – including video teleconferences, high speed facsimile and rapid transfer of data between computers. Communications satellites include the European Communications Satellite (ECS) for television relay and L-Sat, a regional multi-role satellite with direct telephone, television, telex and data transmission facilities. There will be five European Communications Satellites and three maritime communications versions (MARECS) which are expected to handle a substantial proportion of European business traffic until the 1990s under the operation of Eutelsat, the 17 nation assembly of European PTTs.

Individual countries including Italy, Switzerland, France and Germany plan to launch domestic broadcast satellites. The Germans and French have formed the Eurosatellite group to create a direct broadcasting network manufactured by four partners: aerospace company MBB and AEG-Telefunken in Germany, Aerospatiale and Thomson-CSF in France. The French will also launch Telecom 1 in 1983, a satellite which will provide an 'intracompany' link with high speed digital wide band communications between separate branches of a subscriber's organization. Private satellites will be launched during the next few years primarily for television viewers but also for related services.

FRANCE: the télématique program

While there are no pan-European teleinformatics, telecommunications or data processing programs, there are a vast array of new products, services, and networks being initiated on national levels.

The French have a most ambitious national program – *télématique* – in which the government has allocated the expenditure of $60,000 million from 1975-85 to modernize the telecommunications network – increasing the number of subscriber lines from 6.2 million in 1974 to 28 million in 1987 – and to introduce a unified national policy which calls for the integration of numerous compatible products and services.

The *télématique* program, using terminals and improved transmission techniques, permits the simple connection of every telephone subscriber to virtually any information source. In other words, an ISDN. The first step in the program involves an electronic telephone directory which will replace the current paper telephone directories. 250,000 subscribers in the Ille and Vilaine region of western France will begin experimenting with the low-cost terminals this autumn and during the decade the units will be progressively introduced throughout

France with the predicted abolition of printed directories by 1995. The rationale behind the electronic directory is a reduction in costs for printed directories but it is clear that an electronic directory, compatible with other services, in every home will have a revolutionary impact.

Although it remains to be seen how quickly the electronic directory will be installed in France, there is little question that it has crushed some crucial barriers in teleinformatics.

"The electronic directory solves the problem of getting a terminal to the consumer," says Henri Sulzer, corporate development manager of Matra's *télématique* division.

Another interesting *télématique* experiment began in Velizy near Paris in March when households began receiving the French videotex service on a test basis. 2,500 terminals will be online this autumn. Known as *Télétel* (derived from the word telephone and television) the French videotex is connected directly to information stored in data bases through the telephone network and creates an interactive dialogue using an alphanumeric keyboard, compatible with the electronic directory. It enables professional and private users to consult a wide range of data bases or conduct multiple transactions. A teletext service, the direct one-way broadcast of static pictures, is also underway.

"There is tremendous potential for French products abroad," says Roy Bright, managing director of Intelmatique, a government-backed company promoting *télématique* products. "The five new products which industry is exporting include *Télétel*; the electronic directory; audiographic teleconfering for intergroup communications with visual support, telewriting, which transfers manuscript and graphic information from the writing pad to the television screen to display terminal in real time; and the smart card, a plastic card whose embedded microcomputer can be used for transactions, including point of sale and home purchase."

BRITAIN: the Prestel future

The British, who will have System X as the core of their ISDN, will have the system installed in 30 towns by the mid-1980s, and ISDN will be extended to cover the whole country by 1992. In addition, the government is deregulating some sectors of the telecommunications industry in an attempt to liberalize the market.

"British companies had an easy time under the umbrella of a government monopoly and were not competitive enough in the export field," says Jonathan Solomon at the Department of Industry (DoI) in London. "Deregulation should change things considerably."

Many items of telecommunications equipment are already supplied competitively including modems at more than 2,400 bit/s, telephone answering/recording machines, videotex terminals and facsimile machines, while PABXs of more than 100 extensions capacity have long been supplied direct by industry subject to type approval by British Telecom. Simple telephones and slower speed modems will join this list early in the deregulation program. The last area will be PABXs.

"We're opening up the market so business users can obtain better products from a wider choice," says the DoI's Solomon, "and we hope our European neighbors will follow suit."

Britain is firmly committed to the development of ISDN and is introducing such teleinformatic services as teletex, facsimiles, high speed facsimiles and other products.

"The British ISDN will provide a comprehensive, economic and flexible range of voice, data and visual communication services," says John Alvey at British Telecom.

Britain's primary contribution to teleinformatics to date is the development and introduction into public service of teletext and videotex. The Prestel videotex service has been operating for two years and is the largest information retrieval network in the world.

Prestel operating experience has taught British Telecom a number of important lessons. The first was that there is a long term need for videotex pages containing high quality photographic inserts. This is hardly surprising since newspapers and magazines use pictures frequently for editorial and advertising purposes.

The second important lesson is that the cost of videotex information to the customer is heavily influenced by the cost of entering information into the system. Consequently, a technique providing fully automatic picture generation from photographs will be more viable than one depending on manual composition of stylized graphics.

Many countries are now trying out videotex systems and Prestel-based, or Prestel-compatible, systems predominate. Full-scale public trials of such systems are under way in Holland (Viditel), West Germany (Bildschirmtext) and Hong Kong (HK Viewdata), all using the software and hardware developed in Britain for Prestel. The technology has also been sold to telephone administrations in Italy, Austria and Switzerland while fully Prestel-compatible systems are being brought into service in Finland (Telset), Denmark (Teldata), Spain (Videotext), Sweden (Datavision) and Norway (Teledata). Other videotex trials include Green Thumb (Kentucky), Captains (Japan), Viewtron (Florida) and Vidon (Calgary).

WEST GERMANY: pioneering teletex

What the British are doing for videotex, the Germans are doing for teletex. The Deutsche Bundespost, which operates one of the finest albeit most costly telecommunications systems in Europe, is stressing uniform construction of its telecommunications network with a universal service availability. It promotes a large choice of terminal equipment and is conducting tests with videotex, fiber optics, and other new services.

Siemens, for example, is pioneering teletex which is a synthesis of the typewriter and teleprinter and is the first step towards full-scale electronic mail. Transmission is at 2,400 bit/s and a page of 1500 characters can be sent in ten seconds – 20-30 times faster than telex – and received on telexes, electronic typewriters, word processors or other compatible terminals. It can transmit letters in upper and lower cases as well as a variety of symbols. In addition, Siemens is currently investigating textfax with experimental work stations which interactively perform the integrated processing and communication of text and images.

Teletex systems will be introduced in Sweden, Austria, France, Denmark, Belgium, Italy, the Netherlands, the UK and Norway within the next five years.

"We believe that products like teletex and textfax will immediately find a market," says Siemens senior vice president in charge of R&D, Karl Heinz Beckhurts. "They will be integrated into multi-service work stations and will be a significant step toward the introduction of ISDN."

Other German companies are also active in teleinformatics. AEG-Telefunken and its affiliate Telefonbau und Normalzeit have developed a digital telephone set which allows subscribers to make telephone calls while simultaneously using the set for videotex and other digital data based services. Another AEG-Telefunken subsidiary, Olympia Werke, has developed a word processing system for both Chinese and Roman characters and is gearing up to supply additional ranges of electronic office equipment.

Some companies, like Nixdorf, will concentrate their efforts on defined market segments. Nixdorf's strategy concentrates on the end users, particularly branch offices and small businesses, and its strength is in the area of decentralized data processing. Current development includes an integrated terminal capable of handling teletex and a new, more powerful, work station. The company will also be putting its PCM-TDM PBX system on the market next year.

ITALY: gearing up

Italy is just beginning to catch up with its European counterparts in terms of modernizing its communications network, the delay due to a number of frustrating political and financial problems. Earlier this year, however, the government decided to increase the network by one million subscriber lines per year. In addition, there are the beginnings of a rationalization of the industry.

"We must get Italian companies working together to streamline our R&D and products," says Michele Principe, managing director of STET, the government-owned telecommunications holding company. STET controls a number of Italian companies including the utility company SIP, the main equipment manufacturers Italtel, component manufacturer SGS-ATES and the teleinformatics company Selenia. An agreement between STET and Fiat subsidiary Telettra will allow joint coordination and the evolutionary development and export of the Proteo digital switching exchange while an agreement between Olivetti and SGS-ATES will enable a collaborative components policy.

"In the end Italy will be able to provide a complete systems and products range for teleinformatics," says Guido Vannucchi, director of public telecommunications at Telettra. "The industrial overhaul will allow us to bring our respective strengths to the industry's future."

In addition, some Italian companies are strongly promoting European ventures which will produce pan-European and international agreements. "The Italian plan for telecommunications has been so vague in the past that we are actually following the development going on in major PTTs like Germany's and France's," says Luigi Mercurio, managing director of Olteco, Olivetti's telecommunications arm. "We must cooperate with other European countries and build our work stations, video displays and data networks into a larger overall concept. Future industrial growth will require healthy, competitive, highly flexible company structures capable of stimulating innovations."

SWEDEN: evolutionary steps

European companies which lack captive domestic markets are obviously forced to concentrate their efforts on exports. Sweden, which sets the pace for most of Europe's Nordic countries, has been especially successful in this pursuit and its leading telecommunications company LM Ericsson, which has a successful AXE line of semi-digital exchanges and is actively promoting ISDN, is quickly moving into the teleinformatics field. Although half of its sales are still digital exchanges and other switching equipment, the company policy is to move into sales of peripherals attached to the system. Ericsson first established an information systems division to explore the market for data communication and office products and last year purchased 90 per cent of Data Saab – the country's computer company which has been successful in selling its line of Alfascope terminals to IBM and Univac mainframe users and its banking terminals to Citibank. The remaining ten per cent of Data Saab is held by the Televerket, Sweden's telecommunications board.

"The fact that we have a stake in Data Saab indicates our interest in promoting a host of new services in Sweden including teletex and videotex," says Lars Ackzell of the Swedish Telecommunications Administration. "A wide market of new services will also be a good basis for implementing a working ISDN at an early stage."

LM Ericsson has also established a joint venture with Annaconda to provide a marketing base for its PBX and other office oriented equipment and is looking for additional acquisitions in the future, particularly a software company.

"The Swedish market is too small to allow us to develop expertise in every area," says Ericsson's Lars-Olaf Noren. "Takeovers and joint agreements must be a key part of our strategy."

THE NETHERLANDS: digital optical recording

One of the largest European-owned multinationals is Philips in the Netherlands, which has a group of highly autonomous subsidiaries throughout Europe and other parts of the world capable of providing substantial talent in most areas of teleinformatics.

"It's difficult to find a company as diversified as we are," says Gerrit Jeeloff. "We have a worldwide presence which should allow us to maximize our potential in data processing, telecommunications and the related fields."

Although Philips is considerably hindered by Europe's closed markets the company is using its position to establish its own philosophy.

"Because we are based in a small country we must solve all the problems," says Jeeloff. "Consequently we have an all-encompassing ISDN network philosophy. And we are spreading out – purchasing Signetics Corp in the U.S., a share of Grundig in Germany and a Canadian word processor firm."

Philips expects to use its position in the consumer market as a selling vehicle for its teleinformatics products. It produced one of the forerunners of consumer video discs, for example, and has now developed a digital optical recording system – a necessary element for the automated office of the future.

AMERICAN INTERESTS IN EUROPE

The U.S. involvement in European teleinformatics affects every area from components to mainframe computers. American semiconductor companies – including Fairchild Camera and Instruments, Harris, Intel, Motorola, National and Texas Instruments – all have relationships of one kind or another with European firms. Word processing companies, measuring equipment firms, terminal manufacturers, office equipment suppliers, and teleinformatics companies are also very active in Europe.

The two major American concerns in Europe – IBM and ITT, which respectively employ over 100,000 and 180,000 persons – consider their numerous European companies truly European and are creating a wide variety of products for domestic markets in Europe.

ITT, whose presence in Europe began with the purchase of several well-established telecom manufacturing units in 1925, insists its affiliates take an active participation in European domestic programs. STC in the U.K. helped develop System X, Standard Elektrik Lorenz in Germany was awarded a development contract for videotex centers, CGCT in France is attempting to supply System 12 to the French network.

"We respond to opportunities for new products as seen in each individual European country through our local companies," says ITT's Donald Ham. "Our Network 2000 approach to ISDN, including System 12 and telematic products, largely results from our European activities. We expect that much of this technology is applicable to the U.S. market."

IBM works closely with PTTs to obtain approval for attaching its products to telecommunications networks and is supplying a variety of equipment to PTTs, including computers for directory assistance and itemized billing. In addition, the company has participated in CCITT study groups since 1961 and during the 1981-84 period will urge the definition of simple and stable interfaces between ISDN and terminals, so users can obtain a menu of basic transmission services.

GOING FURTHER IN EVERY DIRECTION

European companies plan to maintain relatively high R&D expenditures in all areas of teleinformatics directed towards the establishment of ISDNs as discussed in this report. They will continue seeking ways on national, and perhaps pan-European, levels to increase their technical independence from the Americans, especially in the area of components.

"Independence in microelectronics is vital to our *télématique* development," says Michel Camus, head of the French government's new Norbert Segard Microelectronics Center in Grenoble. "We will continue our efforts to rival American manufacturers."

In fiber optics, laboratories will be looking at higher and wider bandwidths over longer wavelengths permitting better utilization of fiber capacity. Software will continue to be a preoccupation for all phases of teleinformatics experience. Satellite usage will be increased dramatically and storage systems will be refined.

The world's knowledge will soon be accessible from one's desk and from enhanced television sets in the living room," suggests ITT's Jim Van Horn. "Throughout teleinformatics there will be an exciting compression of size and time and the expansion of function and bandwidth." Improvements will be made in new displays and the keyboard will ultimately be replaced by synthetic voice processing and new devices which convert handwriting to electronic signals.

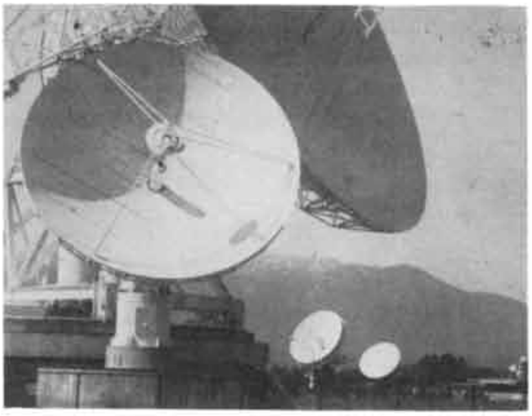
In addition, as predicted by authors including Aldous Huxley and George Orwell, there will be increased integration of information systems using the five senses and perhaps three dimensional imagery. Future information systems will be completely integrated.

The R&D activities at Siemens reflect the future with concentration in nine main areas: energy technology; information and communications systems; diagnostic and image-producing techniques; materials science and technology; microelectronics; electronic and mechanical fabrication processes; automation technology; quality assurance; and the optimization of manufacturing operations.

Much of the future will be spent discussing the future and dealing with a variety of questions. "Are customers going to be innovative enough to accept all these new systems?" asks Philip's Gerrit Jeelof. "Will tradition or innovation win the teleinformatics battle in Europe?" wonders Klaus Luft at Nixdorf. "Will Americans be the first to put new systems to work?" "What will happen when teleinformatics seriously competes with established services, like print journalism?" inquires Eckart Hummel at the CCITT. "The PBX will ultimately be a computer which happens to have a telephone line attached to it," says Thomson's Michel Carpentier. "But will we have a consolidated system or a number of interconnected networks? What will happen to packet and circuit switching with the advent of ISDN?" "We must seriously begin evaluating the impact of new services on the consumer," says Maurice Bernard, head of the French National Center of Telecommunications Research (CNET). "In many cases the user has been ignored and it is definitely the user who will decide the future of teleinformatics."

"It remains to be seen what impact all the technical advances in teleinformatics will have on unemployment, individual liberty and other areas," Nobel Peace Prize winner Sean MacBride said during a recent teleinformatics forum in Paris. "Are we moving too fast to remain in control?"

THE STET GROUP'S FOREIGN ACTIVITIES IN THE FIELDS OF TELECOMMUNICATIONS AND ELECTRONICS



In the last few years the STET Group's electronics and telecommunications sales have been intensified on the international markets. The STET Group controls several manufacturing companies which operate according to an integrated group strategy and represent the core of the Italian telecommunications and electronics industry.

Both the domestic and foreign activities of the Group are supported by the **CSELT** Research Centre, by the Institute of Telecommunications and Electronics for post-graduate training and specialization and by **CONSULTEL.**

Study, design, consultancy and technical assistance both in Italy and abroad in the fields of telecommunications and electronics are the main activities of **CONSULTEL.**

In the last years the results achieved in terms of acquisition of orders, especially for electronic equipment, are highly significant and demonstrate the mounting technical competitivity of the Group Companies' products in spite of notable international competition and consolidated positions of leadership on various markets.

The electronic manufacturing companies of the Group totalled foreign sales of 339 billion lire in financial year 1980.

The manufacturing companies in the Group have agencies abroad in 72 foreign countries covering the markets of Europe, South America, Africa, the Middle and the Far East.

ITALTEL, the largest manufacturing company of the Group, has been awarded a 12 million U.S. dollars contract for supplying Argentina with Multiplex FDM.

A further notable result has been achieved by **ITALTEL** in Brazil where in cooperation with Telebras is developing a Brazilian telephone switching system (called Tropico system) on the basis of the Italian Proteo System.

As regards the radar sector in which **SELENIA,** another manufacturing company of the Group, operates the Company has supplied air traffic control radars and other radar systems like RAN 12 L to many countries all over the world. Lately ten of these systems were supplied to Spain for a value of 15 million U.S. dollars.

In the international electronic component markets STET Group operates through **SGS-ATES** whose sales abroad reach about 75% of the total.

As regards the sector of plant installation, **SIRTI** is carrying out a policy of expansion in markets of particular interest. As a part of this policy, Sirti has obtained, as a member of an Italian consortium, a 520 million U.S. dollars contract in Libya for the supply and installation of a 6,000 km-long coaxial cable telecommunications system; the Company, whose share of the contract in worth 250 million dollars, has been assigned the design, installation and maintenance of the system.

STS, besides already-acquired contracts referring to delivery of equipment for Uruguay and Angola, has recently supplied a ground-station to Somalia in the framework of a cooperation agreement signed by **ITALCABLE** with the local P.T. Ministry.

Last February, the Company has awarded a contract for supplying Portugal with two ground-stations for telecommunications by satellite, which will be linked to the Intelsat world-wide system. **STS** will complete both systems with radio-links for connecting the ground-stations with switching centers and TV stations. In the typographic field the advanced technology introduced in the last few years, as well as the precious know-how achieved by **ILTE** on the international markets, have enabled the Company to offer a highly specialised and most efficient service: both in printing telephone directories and similar products and in gravure and off-set printing processes.

In 1980 **ILTE's** international sales have reached about 30% of the total amount.

EXPLORE NEW HORIZONS WITH TELEMATIQUE

Telematique – this multi-product programme being developed in France is the most advanced and cohesive project of its kind today.

A brief summary of the individual services now being implemented includes:
Teletel Videotex, Electronic Directory System, Smart Card Technology, Telewriting, Antiope Teletext, Low Cost Terminals, Mass Facsimile, Audiographics, Teleconferencing.

With a carefully coordinated R&D programme ensuring technical compatibility and common standards, this family of products maximises the key users benefits of: Low Cost and Ease of Use. This technology and know-how is now available internationally through INTELMATIQUE - the promotional arm of the French Telecommunications Administration.
If your organization is interested in these new vistas created by the convergence of computers and telecommunications, contact:

Intelmatique
FRANCE

THE ORGANIZATION FOR THE INTERNATIONAL PROMOTION OF TELEMATIQUE
98, rue de Sèvres 75007 PARIS. Tel.: (33) 1 306.16.36, Telex 203185 TELEMAT. F.

QUATRE H

BOOKS

The universe of Martin Gardner, astronomers of old Mexico and a romance of engineering

by Philip Morrison

MATHEMATICAL CIRCUS, Vintage Books, 1981 ($4.95); THE INCREDIBLE DR. MATRIX, Charles Scribner's Sons, 1977 ($8.95); THE ANNOTATED ALICE: ALICE'S ADVENTURES IN WONDERLAND AND THROUGH THE LOOKING GLASS, by Lewis Carroll, with an introduction and notes by Martin Gardner, New American Library, 1960 ($3.95); THE AMBIDEXTROUS UNIVERSE: MIRROR ASYMMETRY AND TIME-REVERSED WORLDS, second edition, Charles Scribner's Sons, 1979 ($4.95); FADS AND FALLACIES IN THE NAME OF SCIENCE, Dover Publications, Inc., 1957 ($4); SCIENCE: GOOD, BAD AND BOGUS, Prometheus Books, 1981 ($18.95); CONFESSIONS OF A PSYCHIC, by Uriah Fuller, Karl Fulves, P.O. Box 433, Teaneck, N.J. 07666, 1980 ($3.75); FURTHER CONFESSIONS OF A PSYCHIC, by Uriah Fuller, Karl Fulves, 1980 ($6); THE FLIGHT OF PETER FROMM, William Kaufmann, Inc., 1973 ($5.95). To praise Martin Gardner by these books: So long ago that the redoubtable Richard Feynman of Cal Tech was still a Princeton graduate student, Feynman and a few others developed the remarkable recursively folding paper structures called flexagons. In 1956 a professional free-lance editor and writer, a man with not one college course to elaborate his high school flair for mathematics but with a lively and trained philosophical mind, undertook to elucidate flexagons, then known only through the journals of recreational mathematics, for the readers of this magazine. The piece ran in December, 1956, to wide interest. The editors of *Scientific American* perceived a public thirst for informal mathematics.

Here the fibers of connection wind tightly; the wonderful four-volume *The World of Mathematics* compiled by the first author of this book-review column, James R. Newman, had in 1956 begun its unexpected rise to the runaway best-sellerdom it so patently deserved. The editors popped the question; Gardner agreed that recreational math could support a monthly column. "I rushed around New York and bought all the major references I could find...and started a library." Today Gardner has a couple of rooms full of books and papers, "all the books that come out on recreational math," a file of the issues of some 10 journals and a fruitful worldwide correspondence with knowing readers, many of whom send in new ideas. His column has appeared some 300 times; it has become the focus of a recognizable subculture.

This year Martin Gardner turns 65. Douglas Hofstadter, a young, wide-ranging mathematical mind as familiar with language as it is with computers, is the continuator of Gardner's column, and Gardner has moved from the Hudson River shore to the quieter hills of North Carolina. There he will continue to write in his own rhythm within his best-loved philosophical vein.

Most of those monthly columns have been collected into nine volumes issued over the years by various publishers; all but one are still in print. (The magazine columns are all indexed in the valuable 1979 cumulative-index volume of *Scientific American* from the start through June, 1978; no one has yet prepared an overall index to the collections in books.) Here we review only the latest book of the series; consult *Books in Print* for the rest.

It seems fitting to list a few of the columns that have left a permanent mark on this reviewer, however idiosyncratic the choices may be. (Remember that responses by letter, often more complete in the later book version than in the magazine, offer a rich trove.) To begin with word games, there was a fine response about pangrams, those sentences that seek economically to use all 26 letters. The best pangram (September, 1964) is ascribed to the famous theorist of communications, the mathematician Claude Shannon: Squdgy fez, blank jimp crwth vox. It is formally perfect, meaningful although it requires an explanatory context. The most erudite reader response (November, 1970, and *Mathematical Circus*) is surely that of George L. Hart III, who offered a classical palindrome in Sanskrit, a poem of 32 syllables called *sarvatobhadra,* "perfect in every direction."

Then there was the game of "life" (October, 1970, and February, 1971), a set of recursive rules that generate moving plane patterns whose asymptotic limits are remarkable. It must have consumed in its period of epidemic spread enough computer time to show up in the G.N.P. The studies of tiling, or tessellations of the plane, provide a swift path into serious math for people with no mathematical training and no facility with numbers. (These appeared in April, 1961; July, August and December, 1975, and January, 1977.)

The magnificent April Fools' spoof in 1975 gave an illustrated set of ingenious hoaxes. It drew several thousand responses, mostly from people who had found the mistake within their own specialty but were pleased to read of surprising breakthroughs against those other stodgy truisms. Nearly a thousand physicists wrote to defend special relativity against the well-known paradox displayed in the column. When an Australian paper some years later printed the fact that two Illinois mathematicians had proved the four-color-map theorem, one loyal reader protested that this proof must be wrong because Gardner had published a map that was a counterexample! More skeptical readers had colored the "counterexample" and sent it back to Gardner.

The most telling of the responses were to a hilarious column about Irving Joshua Matrix and his daughter Iva, which appeared in June, 1974. The two were then running a wonderful pyramidal factory producing pyramids with magical properties (here a long, authentic citation of the claims in the marvel literature was given) at Pyramid Lake in Nevada. In spite of everything in the pages, including a pudgy Paiute super named One-Tooth Ree, so many readers drove to Nevada to visit the site that even those who wrote later to complain that no trace of the establishment could be found were numerous. Even sadder was the invitation immediately extended to Gardner by a reputable New York publisher to put out a quick book on pyramid miracles. When that enthusiast learned the whole thing was a hoax, he was not dismayed. His offer stood; write the first book at once and then put out a second in refutation! (The account appears in *The Incredible Dr. Matrix.*)

Gardner's own favorites are the more philosophical columns; memorable ones include those on Newcomb's paradox of decision theory (July, 1973, and March, 1974), and two tours de force on nothing and on everything (February, 1975, and May, 1976). In August, 1977, the column gave an account of the unbreakable public-key cryptography system growing at M.I.T.; this helped to raise a controversy over censorship between the research community and the National Security Agency that has only lately calmed.

The most widely known of Gardner's books is his delightful *Annotated Al-*

ice. A big volume, with the indispensable original drawings by John Tenniel, it presents long marginal notes in witty and learned explication and amplification. Although there is "something preposterous about an annotated *Alice,*" as Gardner himself concedes in the first line of his introduction, the book is in fact both useful and endearing. To "Jabberwocky," for example, there are 25 notes, including the texts of the clever translations *Le Jaseroque* and *Der Jammerwoch.* Even more to the point, there is an early Carroll gloss on the difficult words, written years before *Alice* and fuller than the interpretation given by Humpty Dumpty. The success of Gardner's *Alice* induced several similar annotated classics, including a 1967 *Casey at the Bat,* now out of print. Gardner is predictably a devotee and a savant of Oziana and L. Frank Baum; so far he has published only introductions to various editions of Oz works.

Gardner's college training, like his general predilection, was directed toward philosophy, particularly the philosophy of science and of knowledge. Gardner has written four or five books expounding science for general readers. The new version of *The Ambidextrous Universe* is a special pleasure. Its reach includes the look of flipped photographs and paintings, a delightful palindromic (by words) short-short-short story, the Korean flag, the violation of parity and the intimate history of its discovery, the symmetries of magnetism, entropy, Dr. Edward Anti-Teller and a good deal more. Fresh, clear and sensible, it is a small classic of popular science; the current edition includes a few new pieces on difficult questions of time invariance.

Gardner's first general book, a little timeworn but still in print, is *Fads and Fallacies in the Name of Science.* Here a new theme first entered the extraordinary output of this more than mathematician and philosopher. He was a boyhood magician, and he has ripened into a keen critic of magic and a prolific designer of magical effects, with a wide publication in those esoteric circles. He never performs, not even as a public speaker; the closest he ever came to a direct appearance as a conjurer was in Christmas demonstrations of magic at Marshall Field's in Chicago before World War II. A direct union between his philosophical concern with the nature of evidence and his own behind-the-scenes experience with magic turned him quite naturally into an assiduous and effective critic of pseudoscience.

The early book began this theme, and Gardner's very latest (*Science: Good, Bad and Bogus*) is a collection of nearly 40 book reviews and magazine pieces mostly on the passing and disparate parade of unreliable wonders, such as the psychic surgeons, the learned Dr. Velikovsky, those many adepts who manage over the decades to see through blindfolds, and some academic self-deceivers such as John Hasted and John G. Taylor. Professor Hasted studies, among other things, young people who can, if they are not watched, somehow pass distorted paper clips into a sealed glass globe. Well, not quite sealed; you do need to leave a small hole, or curiously the paraphysical effect does not work! It is all in print, incredible at two levels. Here is our Gardner as an acute and responsible if polemical discussant; responses are by no means suppressed.

The two booklets by "Uriah Fuller," relatively expensive (as the publications of the magicians' world are), lead any serious reader close to the methods of the pseudopsychic conjurers. Those fancy Dans are audacious and subtle by turns. The key is not a special technique, although there are plenty of clever ones. The real base of the effects is the adaptive stream of intellectual and physical misdirection and deceptive simulation. The magician-psychic will exploit any small opportunity, will vary method, sequence and intent, will make use of any interruption of attention and any chance hint to attain the imprecise goal, which itself changes. Control over such behavior lies only in mastery of the details, through stratagems as clever and knowing as the tricks they seek to expose. It is no art for mere credulous professors.

One small example: psychologists have learned that a breath can move a light object that is otherwise quite untouched. A control is a glass jar placed over the object that is presumably to be moved by mental forces. Does this also eliminate the use of an unseen thread? No; on a hard tabletop every jar edge is likely to leave at some point enough clearance to freely pass a nylon fiber of microscopic diameter. Somehow an adroit magician can operate an obedient dancing handkerchief "on a well-lit stage before thousands" and it is just an unexplained trick, perhaps done with thread. But "a crude demonstration by Nina, in her dimly lit Russian apartment," supports volumes of pseudolearned commentary, darkly threatening the repeal of Newton's laws.

Finally, this engaging but toughminded man has a productive hand at fiction. Gardner will soon publish a book of poetry for children; for eight postwar years he prepared a story and a poem each month for the children's magazine *Humpty Dumpty,* with a frequent trick or stunt that on being performed physically damaged the page it was printed on, an amiable delight of the author. There are plenty of his short stories, mostly humorous ones, in *Esquire* and other magazines of the period.

Gardner's one full-fledged novel, with a complex genesis, is now out in paperback. It is a genuine *Bildungsroman;* the young protagonist—and maybe the older man whose novel it purports to be—shares Gardner's life. Youth gave him a profoundly fundamentalist religious background. It was undergraduate study of philosophy at the University of Chicago that carried him beyond the passionate simplicities of a Tulsa upbringing. Exactly this happens to Peter Fromm, the name echoing Sören Kierkegaard, and the book also supplies a vivid journal of life on a destroyer escort in the Atlantic, Gardner's own wartime experience. The core of this intellectual novel is a comprehensive examination of the most important trends in Protestant theology of the past few decades, particularly the subtle Christology of Karl Barth. Fromm (and surely Gardner) leaves the tub-thumping certainties of his youth to enter a domain of doubts and searchings, from which Gardner will certainly comment in the fruitful years ahead.

(The above remarks owe much to the kindness of two mathematician-admirers of Gardner, Peter Renz of W. H. Freeman and Company and Dana Richards of Indiana University at Indianapolis, who made available to this reviewer both published and private material from their treasury, and to *The Two-Year College Mathematics Journal* at Menlo College. To praise Martin Gardner buy these books!)

SKYWATCHERS OF ANCIENT MEXICO, by Anthony F. Aveni. University of Texas Press ($25). Latitude determines. On Stonehenge's Salisbury Plain a cold, clear midwinter day offers some seven hours of pale sunshine. In midsummer it is quite the other way around: the night is short and the day glows throughout two shifts. There people have good reason to notice the solstices, and they have done so demonstrably over four millenniums. At Copán in Honduras day and night are always sensibly equal, give or take a half hour or so. A solstice is only one detail of the simple solar path. A couple of days are nonetheless remarkable. Then the noon sun stands quite overhead; the shadow of a man or of a post disappears entirely. For Temperate Zone latitudes the cardinal points are by definition symmetrical: east and west equinoxes split the solstitial north-south arc in half. It is not surprising, however, that among the Chorti Mayas today "the rise-set directions of the sun on the days of zenith passage are actually regarded as east and west, replacing two of the conventional cardinal points," which lie a few degrees south.

This cheerfully readable and well-illustrated book, part reference text, part informal reportage of research results from manuscript and the field, is a long, helpful gloss on the study of archaeoastronomy, in the context of the sharp distinction between the skies of the Tropics and of the Temperate Zone. That dis-

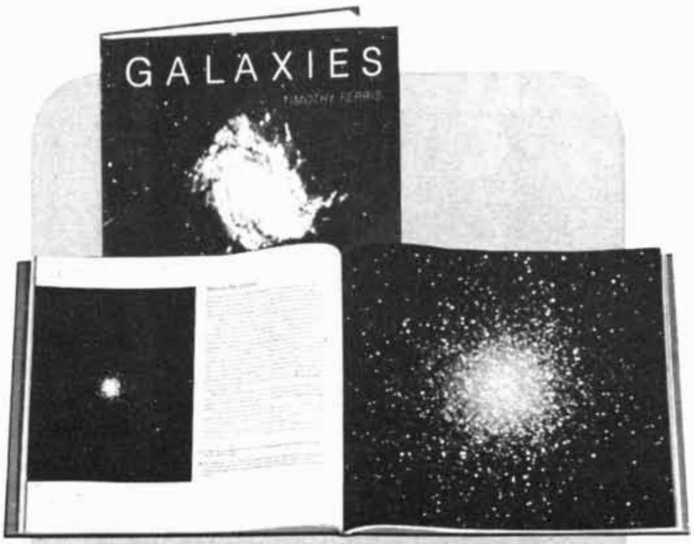
Explore the universe...

take any one of these 4 special selections

for only $2.95 (values to $75.00!)

as a member of Astronomy Book Club.

You simply agree to buy 3 more books—
at handsome discounts—within the next 12 months.



An authoritative reference and an adventure...
GALAXIES. Timothy Ferris.

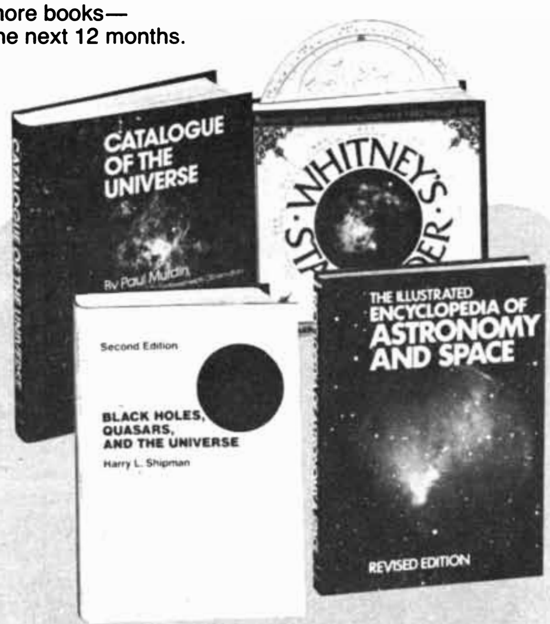
GALAXIES combines an outstanding portfolio of the best photographs to date of the most significant galaxies with a vivid text which details each galaxy and surveys the most recent findings and speculations.

The 39 color and 106 black and white photographs each possess a deep, luminous quality, perfectly suiting its subject. Also features 20 line drawings by noted illustrator Sarah Landry and a huge 13" × 56" gatefold photographic mosaic of the Milky Way. With an extensive bibliography, glossary and index.

"... a very good candidate for the most beautiful book in the world."—Isaac Asimov.

"Worthy of being in the library of anyone who loves the sky and nature."—Sky and Telescope.

Publisher's Price **$75.00**



Basic Astronomy Library

Explore the fantastic discoveries being made about the universe with this outstanding selection of four critically acclaimed astronomy books.

CATALOGUE OF THE UNIVERSE. Murdin et al. Three-volumes-in-one: an atlas, encyclopedia and book of descriptive astronomy. Surveys over 220 astronomical objects. More than 300 photographs.

THE ILLUSTRATED ENCYCLOPEDIA OF ASTRONOMY AND SPACE. Revised Edition. Ian Ridpath. Over 1,000 entries and 150 photographs and illustrations.

WHITNEY'S STAR FINDER. Newly-revised edition of this classic field guide to the heavens. Enables you to identify every major star in the Northern and Southern hemispheres. Comes with a convenient Star-Finder-Locater-Wheel.

BLACK HOLES, QUASARS, AND THE UNIVERSE. Second Edition. Harry Shipman. 344-page report on all the exciting advances being made at the frontiers of space.

Publisher's Price **$61.80**

tinction is not newly found by students of the relations between celestial patterns and human culture. At the turn of the century Zelia Nuttall, a pioneer archaeologist-ethnographer of the Peabody Museum of American Archaeology and Ethnology, called attention to its importance. Indeed, she did her best later to revive all over Mexico the old observance of solar zenith passage, as a contribution to the 1920's renaissance of the indigenous culture. On the right day in every village—and a table followed—"place a post in a vertical position" and tell "the children to attend dressed in white and with a crown of flowers on their heads." It is only for a decade or so, however, that the new wave of astroarchaeology has subsumed the point. The old chroniclers, those few European-trained scholars whose texts bring us the last details of the living cultures of Mesoamerica at the time of the Spanish Conquest, were shaky about it. Not being astronomers, they tended to impose the words of the European art on the subtly differing observances they still witnessed or heard about during the last generations of Maya, Aztec or Inca ceremony.

The sky is well known to us now, even in its changes. The long chapter here on astronomy with the unaided eye makes that quite clear even to the beginner. Striking diagrams (done by P. Dunham) and detailed text represent the best single introduction to the motions of the luminaries, catholic enough to bring out the importance of tropical location as well as of epoch. The book contains helpful tables (one could use still more) and even a couple of program listings for a small calculator. The finer points are not neglected: horizon elevation, extinction of stars near the horizon line and others. The entire piece reflects the experience of the author, an Amherst physicist turned astroarchaeologist, well known for his own field work and for the many students he has led to study the still unread documents of the minds of the past, documents in stone as on bark paper.

The most novel portion of the volume addresses the form and orientation of many important sites of Mesoamerica, south to Cuzco. The great city grid of Teotihuacán in the Valley of Mexico, "both grandiose and precise," has been studied by Aveni and others with care. Not only is the carefully mapped city with its pyramids oriented well off the cardinal directions, but even within the grid there is a fascinating departure from rectangularity. It is not error; the angles are uniform, and what appear to be surveyor's bench marks, a set of distinctive pecked-cross petroglyphs, manifest careful alignment by unknown builders. The orientation may mark the setting of the Pleiades and the direction of the sunset on the day of the zenith passage of the sun, as seen not at the site itself but far south in the latitudes of Copán. That indirection may determine "Teotihuacán north." It is striking that some 60 centers of the region share the clockwise shift of the city axis.

Even more remarkable, the pecked-cross glyphs are widespread; one pair is found far from the city sites, near Alta Vista, a small site hundreds of miles north in the state of Zacatecas, "within 10 kilometers of the present Tropic of Cancer." We are far from full understanding, but the facts do not allow us to ignore the possibility that the priest-astronomers of long ago had sought out the place where the sun turns. There is a good deal of similar material about other famous sites, including the strange round or arrowlike "observatories" that are found at Monte Albán and Chichén Itzá. Answers are fewer than puzzles, but ethnographic remnants and statistical examination seem to exclude the skeptical assertion that we are merely chasing coincidences unnoticed by the ancient builders. They observed and celebrated in their structures as in their codexes cycles of time and significant directions, more fully than we know now.

Another long chapter is quite distinct in its approach, although it is obviously relevant. It is a summary of the Maya calendar documents as they are now interpreted. The fundamentally linguistic subject is not easy, even in this careful sketch. One is certainly disposed to accept the notion that the 260-day count is a union of the properties of the factors of 260 with the fact that about 260 days separate the zenithal passages of the sun at the latitude of Copán; the zero date of the "long count" occurs on the very day of zenithal passage there on one of the best-accepted historical correlations of the Maya and Julian calendars. That the complex cycles of bright Venus also in some degree fit the 260-day framework looks like a later discovery. The 52-year cycle powerfully observed throughout the region is pretty surely the result of the smallest common multiple: $52 \times 365 = 73 \times 260$.

The title of the book is perhaps misleading. Its locus is all Mesoamerica, well south of the borders of modern or ancient Mexico. Indeed, it discusses the solstitial tendencies of the sky watchers of Arizona and Wyoming, and it considers the Andean Tropics. The zenithal sun was surely important in the Inca world; there is some sign that even the invisible passage of the sun through the nadir, the antizenithal point, did not escape notice. Perhaps that is no more surprising than the survival around Cuzco of dark zoomorphic constellations, seen in the gaps and rifts marking the bright star clouds of the southern Milky Way. The fox and the partridge (the Coalsack to modern watchers) are only two of the figures drawn here (from forthcoming

publications of Gary Urton). The mind of past cultures is rich and strange, like our own; its acute response to the complex signals of nature binds us all strongly in a linkage that foreshadows and transcends the canons of our strict and powerful science. We begin, however, to uncover the ancient roots of order.

THE SOUL OF A NEW MACHINE, by Tracy Kidder. Little, Brown and Company ($13.95). The digital sun never sets on imperial IBM. That power's most typical province is its mainframe trade in large, expensive computers for banks, manufacturers and Government agencies. Those users all require attentive service and repay persuasive attentions. IBM makes most of its chips for itself, but those intricate artifacts, for example microprocessors, are produced in large quantities for the market by the famous companies of Silicon Valley and Texas, not to mention Japan. Chips are now a decisive raw material for all the others who manufacture computers.

Although the neatness of division within the industry is speedily being blurred, there is still a recognizable third group: the minicomputer companies. Their complex cabinets sell like limousines, singly or in fleets, not by the handful like the chips. Their customers are not in urgent need of service contracts and systems recipes: the minis are bought by scientific and technical laboratories for data handling and instrument control, and in even larger numbers by companies that build them into more or less specialized complexes, aimed at a variety of functions, civil and military. These are all savvy customers, who prize performance, price and quick delivery above service or advice.

History has set the minicomputer in the marches of Boston, just as chips owe their allegiance to the Bay of San Francisco. There north of the famous Route 128 lie the main suburban plants of the Digital Equipment Corporation. DEC is a minigiant now, about a sixth the size of IBM, a kingdom with its own subtle culture, generally regarded as enlightened and quietly powerful. An intimate narrative of the conception and birth of a new minicomputer, this book dwells in a north-of-Boston plant of DEC's chief rival, a company counting its forces in the thousands, as DEC counts them by the tens of thousands and IBM by the hundreds of thousands. Data General was formed a dozen years back by secession from DEC of the leader of the design team that developed DEC's first big minicomputer success. Data General affects to be a kind of Prussia among those rival company-states, a band of indecorous upstarts in a continent of gentlemen. These attributes are as much deliberately cultivated images as they are reflections of reality. Growth is DG's grail, and grow it did along a smooth, steep curve for a decade. Its Eclipses and Novas were a steady success in the marketplace, showy, discounted, aggressively sold and highly profitable.

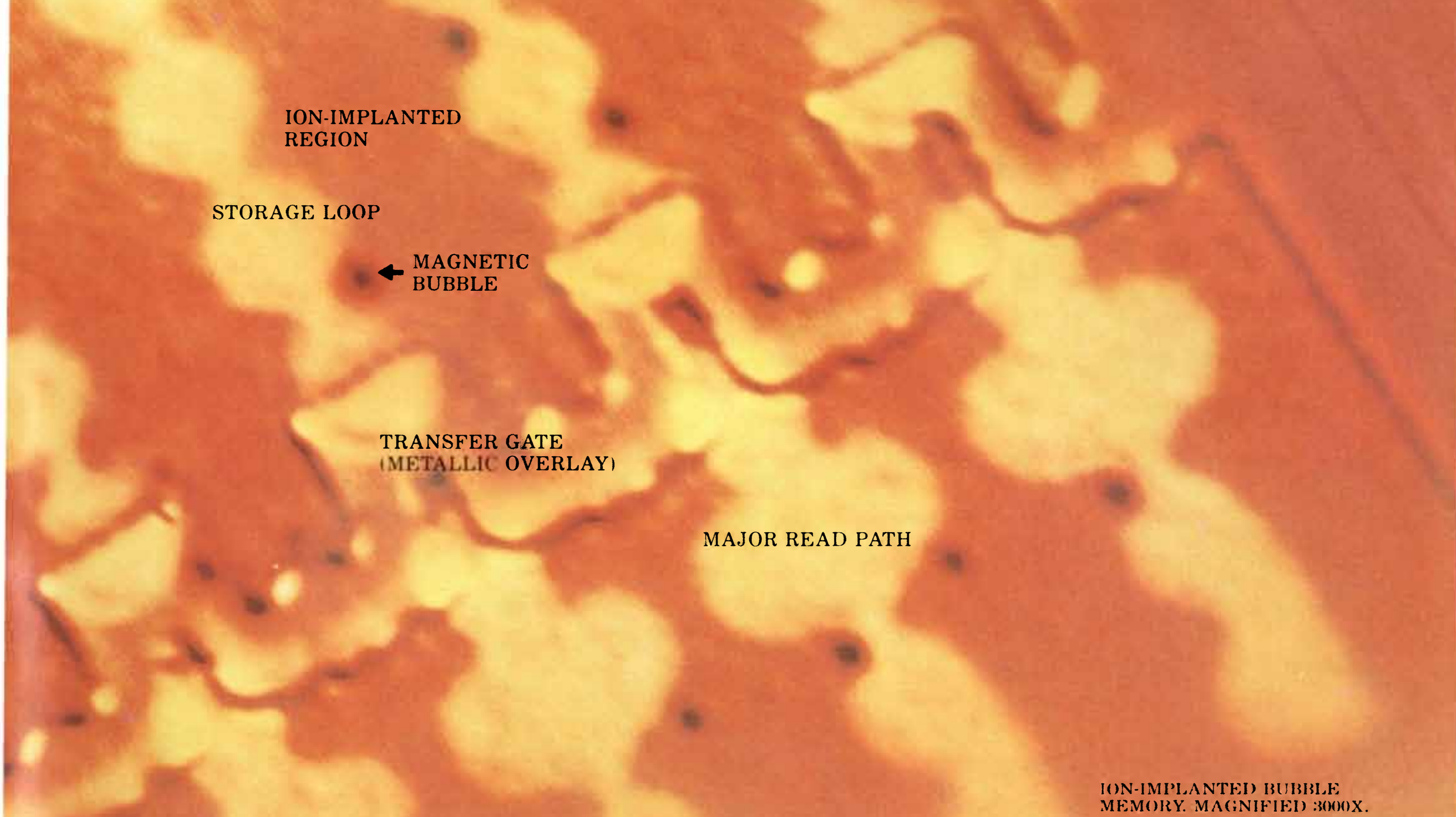
Late in 1978 DEC produced the first of a new line, its VAX machines. They can be called superminis; their principal architectural novelty was the 32-bit word, replacing the 16-bit standard. That change in the packet of symbols the machine can distribute at maximum speed increased its potential directory from 65,000 logical addresses to four billion. It has seemed plain from 1976 on that the future would belong to such machines, made possible by new chips.

Tracy Kidder keeps us in a brightly lighted half basement of Data General's Westborough Plant 14A/B from late 1978 to the fall of 1980. There we watch—as he did—and come to know and to admire the 30-odd young engineers—half of them brand-new recruits—who undertook to build a 32-bit Eclipse to rival VAX. It is guerrilla engineering. Hear two new voices: "'At IBM we wouldn't have gotten on a project this good. They don't hand out projects like this to rookies.' 'They don't hand out projects like this to rookies anywhere but Data General.'" Indeed, even within the company quite another laboratory 1,000 miles away had the 32-bit assignment seriously under way. The basement scheme was for DG something of insurance, something of a challenge to one responsive and brilliant project engineer, the catalyst of this "summer romance," the most original of the people we meet, and something of an irresistible temptation to DG's laconic chief upstairs, himself a gambler, dreamer and star designer.

The place is crowded with half-open little steel cubicles, the tops of their walls below eye level. Each holds a team member, some scruffy manuals and catalogues, a computer terminal on a desk and sometimes a drafting table. A couple of partitioned rooms full of cables and test gear would turn into circuit boards the lengthy logical games of the designers. Out of all of this, out of the determination and insight and hard, long work of the team, there was created in a year enough microcode to "fit into a volume eight inches thick; diagnostic programs amounting to thousands of lines of code; over 200,000 lines of system software; several hundred pages of flow charts; about 240 pages of schematics; hundreds and hundreds of engineering changes from the debugging; twenty hours of videotape to describe the new machine; and now a couple of functioning computers in blue-and-white cases." Birth, however, is not all. Life depends on the suppliers of chips (could they keep up a reliable flow?), the manufacturing department, the famous sales force and the unpredictable boom market. Indeed, in June of this year Data General marked the production of its 100,000th computer.

Two gifts inform this unusual book. One is a sensitive and sharp eye for personality and style; in the mode of the new reportage, the approach is almost novelistic. Apart from the fact that there is not enough time for the full flow and change of character, the reader can come to know as individuals quite a few of the Eclipse Group, their hopes and fears, their jokes and zany games, their science-fiction habit. The second is a pithy insight into just what a computer is, far from formal or complete but extremely helpful to the tyro. The staircase of computer languages, down from half-human high-level language to assembly language to microcode, is made quite clear if not detailed. "Much of the engineering of computers takes place in silence.... Alsing favored the porch and staring out at trees.... 'Writing microcode is like nothing else in my life.... The empty yellow pad sits there in front of me.... Finally, it starts to come.... You have a hundred L-shaped blocks to build a building. You take all the pieces, put them together, pull them apart.'" Each instruction is a precise list of 75 0's and 1's, the colorless words of a specific language, "like early Old English, in which there was no word for fighting and a poet who wished to convey the idea of battle had to describe one." There is much more: a quick key to computer architecture, materials and the crafts of construction.

Finally the book is, perhaps artlessly, an exercise in the economic and social critique of Thorstein Veblen. Once the newborn machines are complete they must be debugged. This search for errors embedded in profound complexity is demanding of men and machines. To be sure, the task is mediated by nurse computers and cleverly made diagnostic programs, together with the occasional shaking of some suspect circuit board. The company sent in technicians to share the agony on double shift. One young engineer found the stub of a technician's paycheck in the wastebasket. "These were first-rate technicians, and being technicians, they got paid for overtime. Engineers were professionals; they did not." The take-home for the technician was twice that of the engineer. The man and his boss burned the stub, so that "the troops wouldn't see it.... Holberger laughed. He said once again: 'I don't work for money.'" *Homo faber*. The regional sales manager has the last word, in a pep talk to his salesmen, once the Eagle (as the excited engineers but not the marketers knew their new machine) had flown. "'What motivates people?' he asked. He answered his own question, saying 'Ego and the money to buy things that they and their families want.'" The engineers and the price system still play different games.



ION-IMPLANTED BUBBLE MEMORY. MAGNIFIED 3000X.

Magnetic bubbles: Packing them in by the millions.

The more magnetic bubbles we can put on a chip, the more attractive this technology becomes as a mass memory. This means an inexpensive, reliable, and nonvolatile alternative for storing large amounts of data in telecommunications and information management systems.

We recently stored 8,798,406 bits in a single experimental chip. This indicates the potential for storing the equivalent of a 600-page novel in an area the size of a postage stamp. We did this by using simple, easily resolved patterns to guide the bubbles. To create these patterns, we implant high-energy ions directly into the magnetic bubble material through a mask. The result: smaller patterns, which permit smaller bubbles, and thus more bubbles per chip than in commercial bubble memories.

We're also studying how to manipulate bubbles by means of two wafer-thin conducting sheets instead of a pair of external coils. With this technique, we'll be able to cut the overall size of bubble devices by a third, move the bubbles 10 times faster, and also put far more bubbles on a chip than in today's commercial devices.

A group of Bell Labs scientists and engineers working on magnetic bubble technology holds over 150 patents in the field. About half of them have been awarded to Andrew H. Bobeck, a co-holder of the basic patent for the 1966 invention. Among our inventions:

- Magnetic bubble concepts and devices
- Garnet materials for bubble devices
- Method of growing epitaxial garnet films from a supercooled solution
- Basic technology for device manufacture
- Ion implant method of propagating bubbles
- Dual-conductor sheet method of propagating bubbles

In the Bell System, bubble memories are already at work in equipment that provides recorded voice announcements and in systems that administer and test digital networks. Eventually, the memories could be used in electronic switching systems, and in advanced home and business telephones.

Our goal is to make this technology even more economical and versatile for storing data electronically. Ultimately—through our partnership with AT&T, Western Electric and the Bell telephone companies—this translates into better service to Bell System customers.

For information about employment, write:

Bell Laboratories, Room HL-3J-238, 600 Mountain Avenue, Murray Hill, N.J. 07974

Please include a resume.

An equal opportunity employer.

Bell Laboratories

FORD LTD

The automatic overdrive comes standard. The highway mileage comes easy.

For 1982, Ford LTD remains a member of a special class of automobile. Full-sized. Full of room and quiet and comfort for the drivers who want and need it.

LTD offers up-to-the-minute technology that keeps this six-passenger car contemporary and efficient. For 1982, a 4.2 liter V-8 engine* and Ford's Auto-

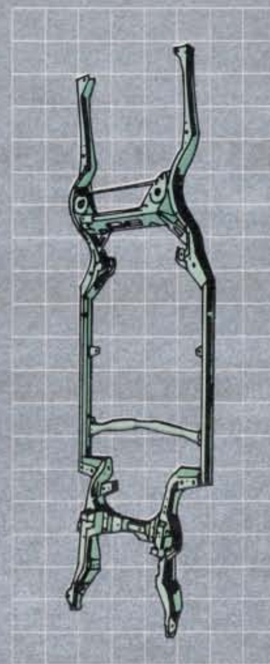
matic Overdrive transmission are standard. At about 40 MPH the Automatic Overdrive engages, cutting engine RPMs by about one third, reducing wear on engine parts and helping improve highway fuel economy.

26 EST HWY **18** EPA EST MPG

For comparison. Your mileage may differ depending upon speed, distance and weather. Actual highway mileage will be lower. Calif. and wagon ratings lower.

The quiet miles.

You needn't give up a quiet, comfortable ride to get this kind of highway mileage either. LTD features extensive sound insulation and a system of rubber body mountings tuned to minimize objectionable noise, vibration and harshness.



LTD's frame of steel.

*4.2 liter engine not available in California.

LOOK OUT HERE COMES

Full-sized comfort.

LTD is still built with body-on-frame construction. This frame of steel forms a solid base for a fine automobile. Of course, your comfort is of prime interest, too.

LTD's Interior Luxury Group, available on the LTD Crown Victoria, features split bench seats, with both sides reclinable, in a choice of super-soft vinyl or crushed velour upholstery.

Country Squire wagon. Luxuriously practical.

With the rear seat down, LTD Country Squire wagon offers almost 90 cubic feet of cargo room and the same sort of luxurious touches found in LTD's top-of-the-line Crown Victoria.

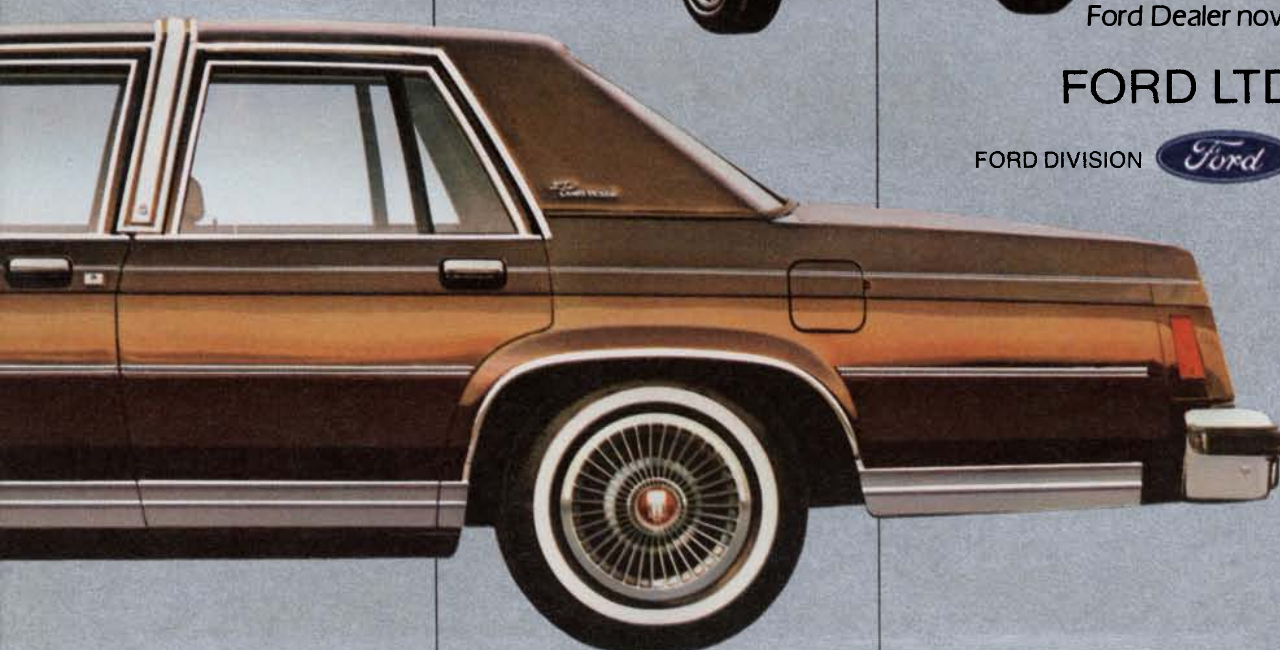
Tripminder® computer.

The Tripminder® computer option, new to LTD, displays time, elapsed time, trip average speed, current fuel economy, trip average fuel economy, total fuel used and more, all at the touch of a finger.

LTD is still the quiet, comfortable automobile so many Americans want. It's a car with a proud tradition. Whether you buy or lease, see LTD at your Ford Dealer now.

FORD LTD

FORD DIVISION



WORLD FORD

Established 1845 SCIENTIFIC AMERICAN October 1981 Volume 245 Number 4

The Patient Record in Epidemiology

The most accurate estimates of the rate of many diseases come from studies based on a huge collection of patient records at the Mayo Clinic. The records are indexed for retrieval by diagnosis

by Leonard T. Kurland and Craig A. Molgaard

Record keeping is a part of medical practice whose full utility is not readily appreciated. Maintaining a current record of all diagnoses, medications, laboratory procedures and other information for each patient helps to ensure the continuity of medical care and thereby contributes to the welfare of the individual patient; it may seem unlikely, however, that the patient record itself could contribute much to the advance of medical knowledge. Actually a patient record compiled as a matter of routine can serve a larger purpose. In particular a historical collection of the records of many patients is an invaluable resource for advances in clinical care and for epidemiology: the investigation of patterns of health and disease in a population. One such collection that has amply proved its worth is a large archive of medical records at the Mayo Clinic in Rochester, Minn. The records are indexed by diagnosis and by demographic characteristics of the patients. Information from the archive has been employed in determining the incidence of a wide range of diseases, including multiple sclerosis, parkinsonism, stroke, leukemia and almost every other form of cancer.

An incident in December, 1976, illustrates the capabilities of the archive. A national program of immunization against the "swine flu" strain of influenza was suspended when Public Health Service physicians noted among those who had been inoculated what seemed to be a relatively large number of cases of Guillain-Barré syndrome, a rare neuromuscular disorder. In order to define how many cases constitute a relatively large number it was necessary to know the base-line incidence of the disease: the incidence in the population as a whole in the absence of any disturbing factors. That information was supplied by a 1973 study of the rate of Guillain-Barré syndrome in the population of Olmsted County, Minn.; the study had been based on records drawn from the Mayo Clinic archive. Epidemiologists determined from further studies that recipients of the swine-flu vaccine were developing Guillain-Barré syndrome at a rate several times the usual one, and the program of immunization was not resumed.

A feature that distinguishes the Mayo Clinic archive as a source of data for epidemiology is its comprehensiveness. For almost a century the clinic has provided primary medical care for much of the population of Rochester and the surrounding areas of Olmsted County. Furthermore, since 1966 the diagnoses made by almost all the physicians in the county, including those who do not practice at the clinic, have been consolidated in a centralized file maintained at the clinic; hence with few exceptions every diagnosis made by a physician in the county is now indexed in the archive. Many records give a complete history of medical care from birth to death, and often to autopsy.

Another important property of the archive is ease of access to the patient records. Although the methods of storing and retrieving information have been changed several times, the organizing principle of the indexing system has remained constant since 1907. Any record can be retrieved by the patient's identification number or by the diagnosis, and so it is possible to identify essentially all cases of a disease diagnosed in Olmsted County in the past 75 years.

The organization of the Mayo Clinic record system is so well matched to the needs of epidemiology that one might think it was devised explicitly for such purposes. That is not the case; like other systems of medical record keeping, the one at the Mayo Clinic was established as an adjunct to patient care. The subsequent development of the archive is intertwined with improvements in diagnosis, surgery and data processing and with the expansion of the clinic itself.

The forerunner of the Mayo Clinic was a partnership of William W. Mayo and his two sons, William J. and Charles H., who were all surgeons. The elder William Mayo was a British immigrant who in the 1860's had settled in Rochester, then a small prairie town. He had helped to found St. Mary's Hospital, which was then the only hospital in Rochester, and the Mayos had arranged to practice surgery there. The growth of surgical expertise, however, made it increasingly difficult for an individual to master the entire field, and in about 1903 the practice was transformed from a partnership into a cooperative group. As new areas of specialization emerged physicians were added to the group. The idea of group practice was then an innovation in American medicine; it remains the basis of the Mayo Clinic, which is now the largest integrated group practice in the U.S.

While the Mayo Clinic remained small an informal system of record keeping sufficed. From 1885 until 1907 patient records were kept by individual physicians in leather-bound ledgers. Four or five case histories were usually entered on each page. The state of the art of diagnosis was such that entries were generally brief. The attending physician noted the date and the patient's name, age, place of residence and complaint; later he added the principal findings of the case and its outcome. Each laboratory and each surgeon maintained a similar set of ledgers. To trace the course of a patient's illness it was necessary to examine the ledgers of the attending physician, the laboratory and the surgeon.

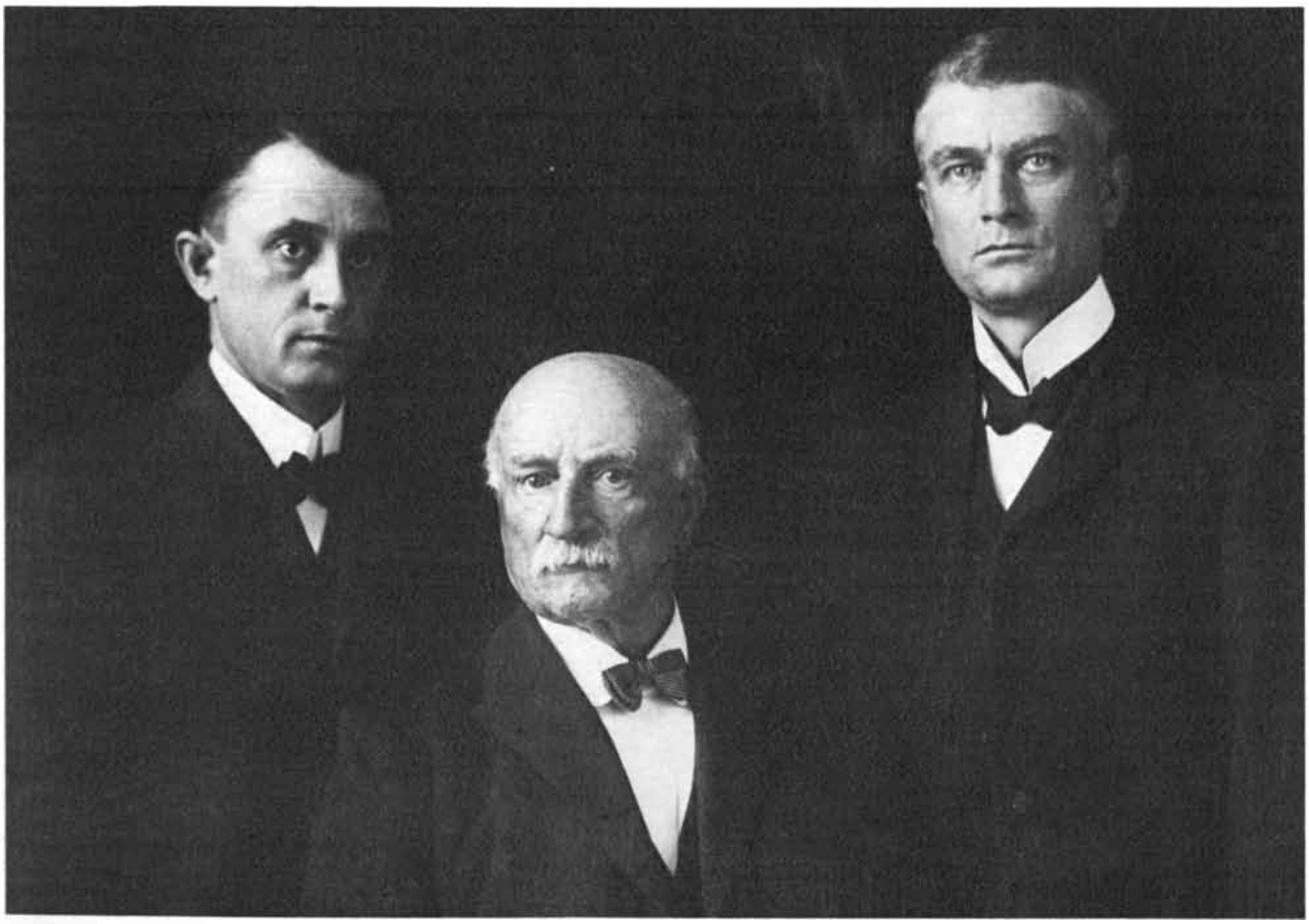
Because the clinic was small and diagnostic techniques were limited the system worked effectively. The Mayos employed the ledgers as a basis for clinical research, describing extensive series of surgical cases and reporting the results of new surgical techniques. In 1906, for example, Charles Mayo published an account of "Surgical Treatment of Goiter: Based upon Three Hundred Personal Operations." (Mayo had pioneered the modern surgical approach to the treatment of goiter.)

In the 23 years during which records were kept in ledgers the subtlety of differentiation that could be attained through diagnosis increased significantly. Before 1850 few diseases had been described in sufficient detail for them to be clearly distinguished. In the next 50 years new anesthetic agents and aseptic techniques greatly improved the effectiveness of surgery, but diagnosis remained a medical stepchild, serving primarily for the selection of patients who would benefit from surgery. Soon after 1900, however, new laboratory methods, practical X-ray devices and endoscopic techniques for direct visualization of internal organs were added to the physician's means of distinguishing one disease from another. As diagnosis became both more complex and more exact it was established as an independent procedure required in all forms of medical care.

These developments, together with an increasing patient load, made it cumbersome to trace a case history through the ledgers. In addition the possession of the ledgers by individuals implied an attachment to them that was at odds with the spirit of the group practice and also inhibited free access to the records for teaching or research.

Soon after 1900 the Mayos asked Henry S. Plummer, a young clinic associate, to reorganize the records. By 1907 Plummer had devised a system whereby a patient's records were all kept in one file. Each patient was assigned a unit number, which identified the various components of the record. The record included notes made by each physician who examined the patient; new record forms were introduced as new specialties were developed at the clinic. The results of laboratory tests were transferred in chronological order to forms designed for that information. Correspondence with the patient was also kept in the file, as were birth and death records for local residents.

Plummer's system ensured that all medical information pertaining to an individual patient could be found in a single dossier. Although this greatly enhanced the clinical care of the patient, it soon became evident that some means of access to groups of dossiers was needed for research and teaching. To meet the need Plummer created two simple indexes, one index organized by diagnosis and the other by surgical procedure.



WILLIAM W. MAYO (*middle*) and his sons William J. (*right*) and Charles H. were all surgeons. Their medical partnership in the small prairie town of Rochester, Minn., was the forerunner of the Mayo Clinic. The Mayos had arranged to practice surgery at St. Mary's Hospital in Rochester, which William W. Mayo had helped to found in the 1880's. Some 20 years later, as a result of a growing number of patients and the expansion of surgical expertise, the partnership was transformed into a group practice; new physicians were added as surgical specialties proliferated. Group practice remains the basis of the Mayo Clinic, which is the largest integrated group practice in the U.S.

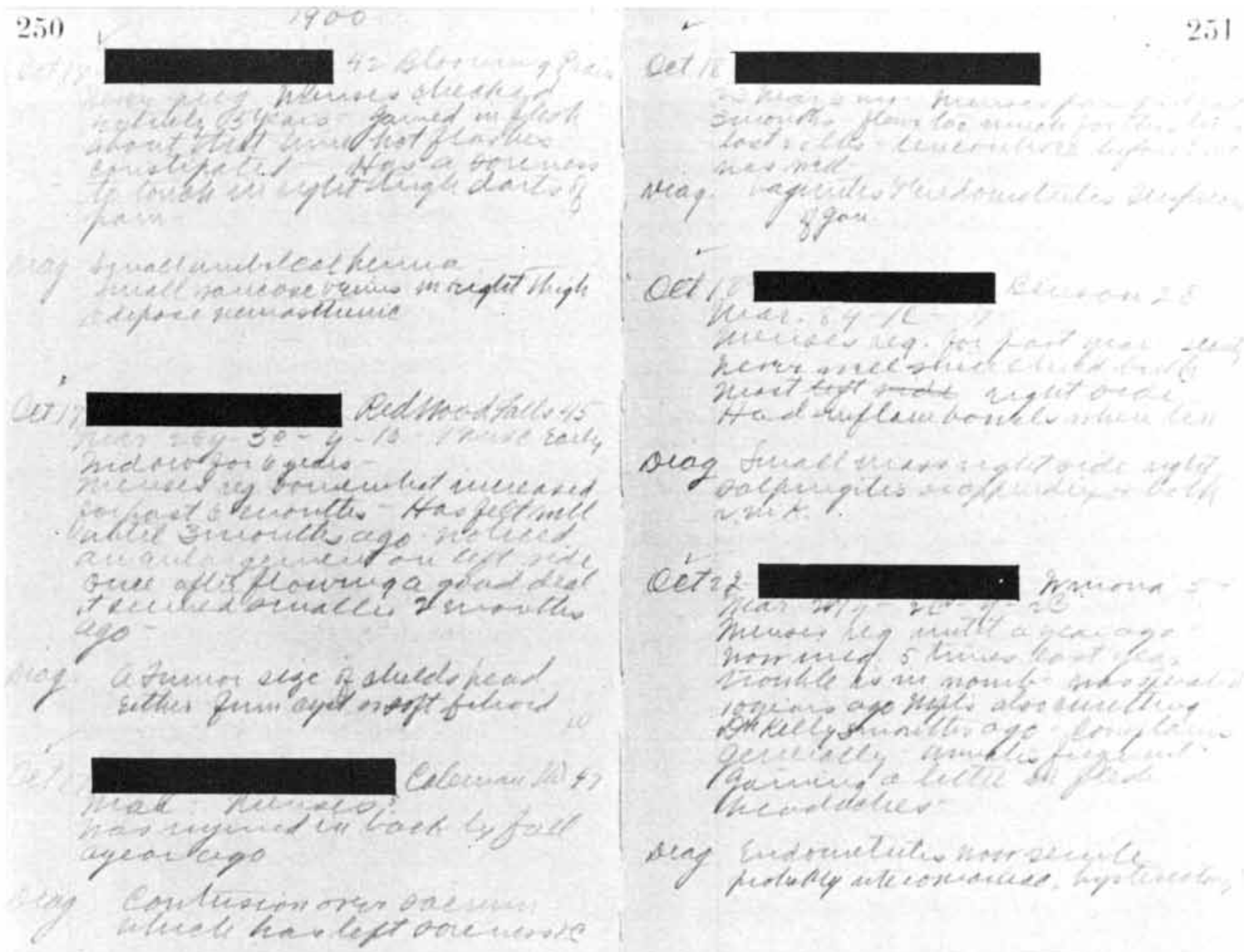
In both indexes the major headings were organs and organ systems; listed alphabetically under these headings were diseases or surgical procedures. Entries were carefully made by Plummer's secretary, Mabel Root, on five-by-seven-inch cards.

The dossier system and the Plummer-Root indexes represented a great advance over the ledger system, both for patient care and for clinical research. Plummer persuaded the Mayos to establish as clinic policy that the archive should serve as an institutional resource, so that all records would be available for teaching and research regardless of which physician had treated a particular patient. Individual records were easily retrieved for patient care by the number assigned to each patient; the indexes made it possible to locate records of groups of patients who had similar diseases or who had undergone similar surgical procedures. In the early decades of this century the system facilitated an increasing volume of clinical and surgical publication from the Mayo Clinic.

It was Plummer himself who made the first use of the diagnostic archive for an investigation of the rate of a disease in the local population. In 1931 his study of the frequency of diagnosis of goiter among Olmsted County residents was published. The archive's main function, however, remained the organization of records for patient care; the effectiveness of Plummer's system for the purpose is suggested by the fact that the diagnostic and surgical indexes remained practicable well into the 1930's, in spite of enormous changes in diagnostic technique.

The chief drawback of the Plummer-Root system, which became progressively more troublesome, was that diseases were listed under organs. Diseases that can affect many organs (such as cancer) were dispersed throughout the index. The lack of cross-indexing of diagnoses was another drawback, and one that frustrated attempts to study the relations among diseases. By the 1930's processes rather than structures had come to play the major role in the analysis of disease. Diagnostic terminology had also burgeoned, a development represented by the publication of a *Standard Nomenclature of Diseases and Operations* (SNDO) in 1935 by the American Medical Association. Ultimately the increased sophistication of diagnostic technique and the growth of the patient load overburdened the Plummer-Root system as they had the ledger system 25 years earlier.

Joseph Berkson of the clinic's department of physiology was asked to reorganize the file. Berkson had been trained at the Johns Hopkins School of Hygiene and Public Health and was proficient in methods of statistics and



PATIENT RECORDS in the early period of the Mayo Clinic were kept in leather-bound ledgers. Each physician, surgeon and laboratory maintained a set of ledgers. These are the notes of gynecological consultations for several days in October, 1900, of William J. Mayo. He recorded six diagnoses on these pages, making brief biographical notes and recording the symptoms and the diagnosis. The middle entry on the left-hand page includes the diagnosis of an ovarian tumor the size "of [a] child's head." The ledgers were the precursors of the current diagnostic archive and served both for patient care and for research. William and Charles Mayo consulted the ledgers in preparing accounts of long series of surgical cases. As the number of clinic patients grew, however, it became increasingly time-consuming to trace a patient's history through the series of ledgers. (The names of individual patients have been covered to preserve their privacy.)

data handling. In 1932 he introduced an index for prenatal and obstetrical care, and in 1935 he set up two more indexes, one for surgical procedures and one for diagnoses. An independent man, Berkson had little use for the classification system implicit in the American Medical Association's standard nomenclature. He also decided not to adopt another system that had become available, the *International Classification of Causes of Death* (later called the *International Classification of Diseases,* or ICD).

In place of these Berkson devised his own diagnostic index, although it did share features with the published classification systems. Like the standard nomenclature of the American Medical Association, Berkson's index had as headings both organs and diseases, reflecting increased interest in disease processes. In each disease category he created a classification labeled "Except as Above," much like the "Other and Unspecified" categories of the ICD. Berkson's index made it possible to classify more than 20,000 diseases and sites in the body, easily accommodating the current state of diagnostic technique. The numerical codes in the indexes were entered on Hollerith punch cards, which were then the most efficient medium for mechanical data processing. Berkson's diagnostic and surgical classifications and his introduction of punch cards represent the largest single advance in the automation of the indexes.

Still another feature of Berkson's diagnostic index was the inclusion of an item indicating Rochester residency on the punch cards. Since almost all care for significant disease in the area was then being provided by the Mayo Clinic, it became possible to identify quickly and by mechanical means all local patients who had had a given disorder. The preconditions for effective population-based epidemiological research had clearly been met. The primary uses of the archive, however, remained patient care and clinical research; it was not until 1948 that the archive was exploited for research into population-based disease rates. In that year Berkson and members of the neurology department concluded a pioneering epidemiological study of multiple sclerosis in Rochester.

Berkson's system worked brilliantly and is still used for retrieving all records dating from 1935 through 1974. His foresight in providing space for new diagnostic entities and in adding the catchall category "Except as Above" allowed the system to remain useful for almost 40 years. By the 1970's, however, the catchall category was being overrun by the proliferation of new diagnoses. Searching the "Except as Above" file had become an expensive and time-consuming process. To solve this problem one of us (Kurland), who had originally been attracted to the Mayo Clinic by the

HENRY S. PLUMMER, a young associate of the Mayos, reorganized the clinic's patient records in about 1907. In Plummer's system for the first time all records for an individual patient were kept in a single dossier, including the physician's notes, the surgeon's report and laboratory test results. To organize the files Plummer devised an index based on the patient's diagnosis and one based on surgical procedures. He also persuaded the Mayos to establish as clinic policy the ready availability of patient records for teaching or research, regardless of which physician had attended the patient. These innovations remain central to the modern diagnostic archive.

INDEX CARD in the diagnostic file devised by Plummer lists cases of a single disorder. The cards are filed under the heading of the organ of the body affected by the disorder; organs are arranged alphabetically. The card shown lists cases of exophthalmic goiter diagnosed in 1928; it was filed under "Thyroid Gland." (Exophthalmic goiter is an enlargement of the thyroid gland named for one of its symptoms: protrusion of the eyeballs.) Plummer's secretary, Mabel Root, entered in each square the date, the patient's age and sex, whether the goiter was recurrent and the patient's identifying number. The top entry in the column on the far left is for a 50-year-old man who attended the clinic on June 25 with a recurrent goiter. The card was one of those that supplied data for the first population-based study of disease rates done at the Mayo Clinic: Plummer's 1931 study of the incidence of goiter in the surrounding area.

possibility of doing population-based research of the kind represented by Berkson's study of multiple sclerosis, initiated the third major reorganization of the diagnostic index in 70 years.

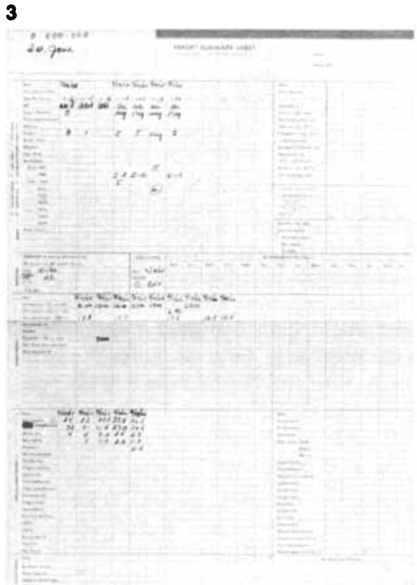
In the decades since Berkson developed his system the sophistication of diagnosis has grown at a rapid pace. By the 1970's several standard classifications of disease existed. At the Mayo Clinic the reorganization of the diagnostic index was based on the *Hospital Adaptation of the International Classification of Diseases, Adapted, Second Edition,* or H-ICDA-2, which was published in 1975. The clinic's needs required that the classification be modified to increase its specificity. The H-ICDA-2 system incorporated many improvements over its parent volume, the *International Classification of Diseases, 8th Revision,* published by the World Health Organization. Nevertheless, many categories consisted of groups of diseases rather than single entities. To overcome this difficulty additional digits were added to the code to identify individual diseases and their synonyms under each rubric of the classification.

The operation of the current diagnostic index requires the cooperative effort of many groups and individuals, each having a specific function with respect to the records. Plummer's numbering system is still employed, and each new patient at a Mayo facility is assigned a chronological seven-digit registration number. If the patient's primary physician seeks a consultation with a specialist, the consultant's findings are recorded on forms designed for that specialty, inserted in the patient record and forwarded to the primary physician. When a patient is dismissed, the diagnoses made during the current registration at the clinic are recorded on the master sheet in the dossier. The dossier is then sent to the completion desk in the medical-records department, where the forms are arranged in a prescribed order, the

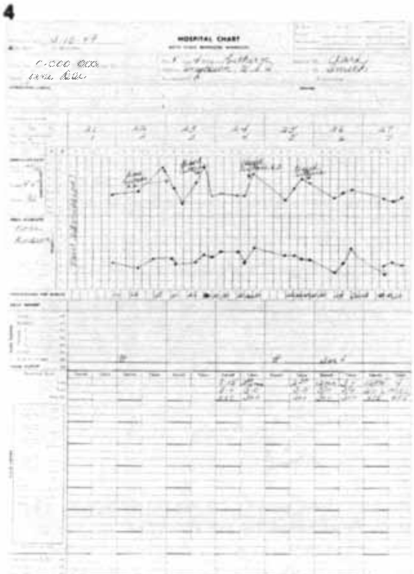
1

2

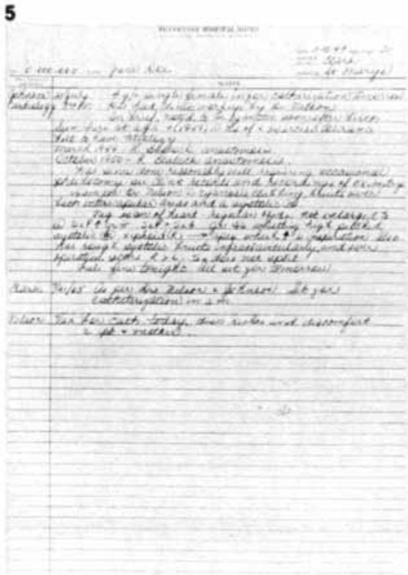
3



4



5



6


SURGICAL RECORD
MAYO CLINIC — ROCHESTER, MINNESOTA

Age 4 yrs. Sex F Section W-6 Date 3-16-49

Name Jane Doe Address Anytown, U.S.A.

Dr's Address Smalltown, U.S.A.

Operation advised by, Medical Consultant W-6 Surgical Consultant Jones

Tetralogy of Fallot. Right sided aorta.

Operation Indicated Blalock-Taussig procedure. In house.

8(243)

Date op. 3-17-49 Col. I-1 Jones Brown Green

Tetralogy of Fallot.

Anastomosis of end of left subclavian artery to side of left pulmonary artery (resection portion of fifth rib). Institution of temporary closed intercostal drainage.

1 #24 mushroom catheter.

Add. cond. to index:
Detail: Skin antiseptic: ether and mertniolate.

A posterolateral incision was made on the left, and a long segment of the fifth rib was resected. The pleura [illegible]. The lung was retracted. The pulmonary artery [illegible] dissected out. The left subclavian artery was [illegible] to sacrifice the first four branches of this [illegible] length. The subclavian artery was then ligated and divided as [illegible] as possible. The proximal end was rotated over, and an [illegible] anastomosis of the subclavian artery to the side of the [illegible] artery was performed using one row of everting silk sutures. [illegible] clamps were removed, there was no bleeding, and a definite thrill could be felt at the site of the anastomosis. The operative site was washed with saline. One catheter was inserted for temporary [illegible] drainage. The wound was closed in layers after 140,000 units of penicillin had been inserted.

STANDARDIZED FORMS make up much of the patient record in the modern Mayo Clinic system. The master sheet (*1*) holds biographical information, diagnoses and a list of surgical procedures; on the reverse is a summary of the patient's clinical history. Two forms present laboratory results; one form (*2*) is a verbal summary and the other (*3*) is a compilation of the numerical results of diagnostic tests. Hospital charts (*4*) trace the patient's course before and after surgery is performed. The attending physician's notes have their own form (*5*), as do descriptions of surgical procedures (*6*). The dossiers are filed according to patient numbers, which are assigned consecutively. When a dossier is called for, it is inserted in a plastic envelope and delivered by pneumatic tube or dumbwaiter to the current attending physician.

data are consolidated and the results of laboratory tests are transcribed onto summary sheets.

The completed dossier is sent to the medical-indexing unit, where nosologists, or specialists in diagnostic classification, assign the numerical code for each diagnosis. The code is entered directly into a computer file by means of an on-line terminal. The system allows a final check on the accuracy and uniformity of the coded data as the data are entered. When the nosologist enters the diagnostic code number, its translation appears on the screen. This technique has proved effective: periodic sampling indicates the error rate is less than 1.5 percent. The dossier is ultimately returned to the central record-storage area, which occupies two floors of a clinic building given over primarily to record storage and data handling.

The whereabouts of some 50,000 patient records in use at any given time are monitored by means of a manual location file, which is soon to be replaced by a computerized system. These active records are available almost immediately. Past records are also readily retrieved from the permanent file among records of 3.5 million patients, representing 10 million clinic visits and 25 million consultations.

The organization of the index of surgical procedures has undergone little revision since 1935, although there has been a large increase in the number of surgical procedures and in their variety. Hollerith punch cards are still employed for initial tabulation; cases for study are selected after a manual review of the records. The information coded includes not only a detailed listing of surgical procedures but also the diagnosis based on the findings of the surgeon and on the reports of the clinical pathologist.

A separate collection of pathology results from both surgery and autopsies is organized along the lines of the Plummer-Root file. All autopsies in Olmsted County are done at the Mayo Clinic and the records are stored there. For many years the autopsy rate for the Rochester population has been remarkably high: often as high as 70 percent and now about 60 percent. The availability of autopsy confirmation for clinical diagnoses adds greatly to the reliability of the information in the archive.

With the cooperative effort of physicians, nosologists and data processors the archive is now growing at a rate of about 885,000 medical diagnoses and 50,000 surgical procedures per year. About 20 percent of the diagnoses pertain to local residents. Although the clinic is famous as a referral center, it has always provided much of the primary care and almost all the specialty care in the Rochester area. The clinic draws patients from the 95,000 residents of Olmsted County (including 61,000 in Rochester) and from the area beyond in southeastern Minnesota. The population of the region is quite homogeneous, being mostly of northern European origin. The local economy is based mainly on farming and light industry. Rochester is about 75 miles from Minneapolis–St. Paul, the nearest large urban area.

The isolation of the region has contributed to the concentration of medical services at the Mayo Clinic. Nevertheless, there have always been a few independent practitioners of medicine in Olmsted County. In the early 1950's a second group practice, the Olmsted Medical and Surgical Group, was begun in Rochester, and in 1955 the Olmsted Community Hospital was opened, staffed by physicians of the Olmsted Group and by a few independent practitioners. In 1966 a grant from the National Institutes of Health made it possible to consolidate diagnoses and surgical procedures from these sources in the Mayo Clinic system. The information is indexed according to the clinic classifications, and a computerized file for local residents is maintained at the clinic. This final step has made the archive a virtually complete record of all visits to physicians in a medium-size city and its surrounding area. The consolidation of diagnostic information in 1966 thus laid the groundwork for an important series of studies of the base-line incidence of disease.

One of the first of the population-based studies indicated that research based on the diagnostic archive might yield a more accurate picture of disease than studies utilizing other kinds of record systems. Surveys in several European cities had suggested that multiple sclerosis had been gradually increasing in frequency for decades. The surveys measured the total number of cases of the disease in successive years; such surveys of prevalence include both old and new cases. To determine whether the rate of multiple sclerosis had been increasing, however, what was needed was a measure of incidence, or the number of new cases each year; the incidence of disease is not easily derived from a survey of prevalence.

By searching both the Berkson and the Plummer-Root indexes it was possible to identify every diagnosed case of multiple sclerosis in Rochester between 1905 and 1964. Because the dates of diagnosis were known it was comparatively simple to determine the incidence rate, which showed no significant change over the 60-year period.

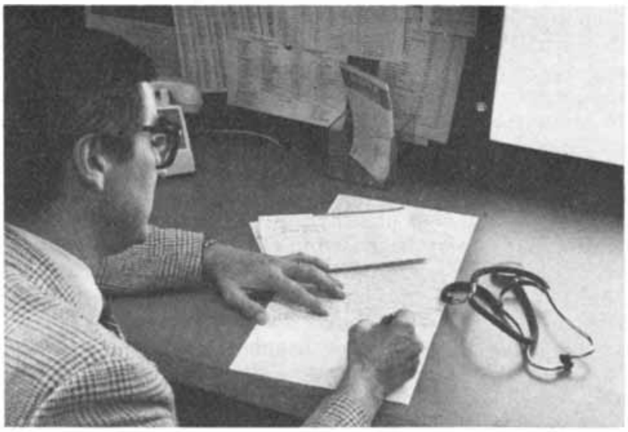
The Rochester findings differed from those of other studies in several additional ways. The length of time patients survived after diagnosis in Rochester was much greater than that in hospital studies of multiple sclerosis, where the average survival time was from 12 to 15 years; in Rochester the patients' average survival time was about 35 years. In addition the overall rate of diagnosed multiple sclerosis in Rochester was higher than that found in other studies. These differences reflect the fact that, with excellent medical care in an isolated area, a higher proportion of mild cases are diagnosed in Olmsted County than are diagnosed in other places where only hospitalized cases can be ascertained. The longer survival time probably reflects this diagnosis of a greater number of mild cases, which have a better prognosis, as well as earlier diagnosis of cases in general.

The completeness of the records at the Mayo Clinic allows epidemiological research based on the diagnostic archive to throw light not only on the patterns of disease but also on theories about the cause of disease. In a 1963 study David C. Poskanzer and R. S. Schwab of the Harvard Medical School noted a dramatic increase in the mean age of admission for parkinsonism at Massachusetts General Hospital; the age had increased from about 30 years in 1920 to roughly 60 years in 1960. Poskanzer and Schwab hypothesized that a single, simultaneous exposure to some disease-causing agent was responsible for all the cases.

In 1918 a worldwide influenza epidemic took some 500,000 lives in the U.S. Some of those who survived contracted encephalitis in the aftermath of an influenza infection; some encephalitis victims also subsequently developed parkinsonism. Poskanzer and Schwab proposed that there had been many cases of subclinical encephalitis in 1918, leading to parkinsonism after a long and variable latency period. They suggested that as those who had lived through the influenza epidemic aged together, the average age of onset for parkinsonism also rose. On the basis of this hypothesis Poskanzer and Schwab predicted a sharp drop in the incidence of parkinsonism in the 1970's as the number of those who had lived through the influenza epidemic dwindled. They also hypothesized that most cases of parkinsonism, whether or not they were preceded by symptoms of encephalitis, were caused by exposure to a virus.

Information from the diagnostic archive provided a test of these hypotheses. All cases of parkinsonism diagnosed in Rochester from 1935 through 1974 were identified. Autopsy reports were available for more than half of the cases, providing further confirmation of the diagnosis. In contrast to the increase noted at Massachusetts General Hospital, the mean age at diagnosis for the Rochester parkinsonism patients remained between 62 and 66 years throughout the 40-year period. The Rochester study thereby undermined the theory of causation proposed by

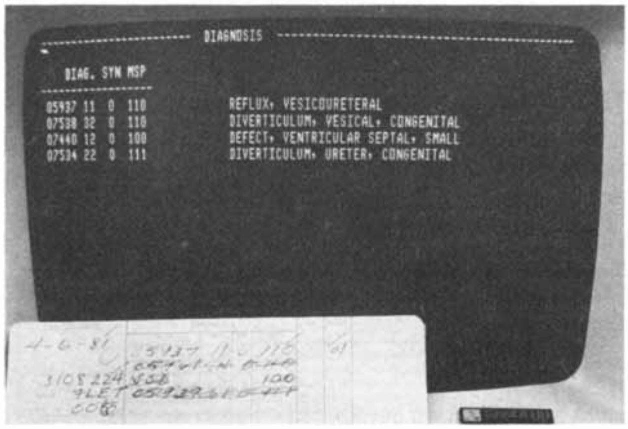
1



2

3

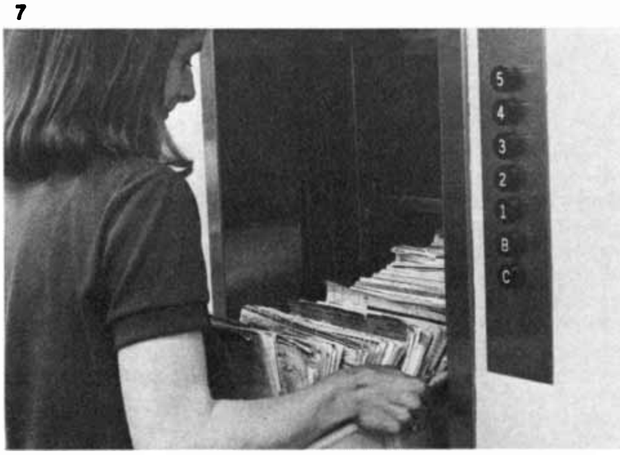
4


DIAGNOSIS
DIAG. SYN MSP
05937 11 0 110
07538 32 0 110
07440 12 0 100
07534 22 0 111
REFLUX, VESICOURETERAL
DIVERTICULUM, VESICAL, CONGENITAL
DEFECT, VENTRICULAR SEPTAL, SMALL
DIVERTICULUM, URETER, CONGENITAL

5

6

7


5
4
3
2
1
B
C

8

Poskanzer and Schwab; if there had been no steady increase in the age of onset, there is little reason to think a single exposure was responsible for all cases of parkinsonism. Furthermore, the Rochester records showed no decrease in incidence as the proportion of the population born before 1918 was reduced, as would have been predicted on the basis of Poskanzer and Schwab's hypothesis.

The differences between the results of the population-based study in Olmsted County and the Massachusetts General Hospital study highlight a bias that can arise in studies based on hospital records. The population of hospital patients reflects the selective patterns of referral from other medical facilities. Indeed, when the mean age of all patients with parkinsonism seen at the Mayo Clinic in the period from 1935 through 1964 was examined (as opposed to the age of parkinsonism patients from the local area), a rise similar to but smaller than the one noted by Poskanzer and Schwab was observed. The mean age at diagnosis also increased for referral patients with two other diseases that affect the nervous system: herpes zoster and stroke.

The reason for these increases is that a growing proportion of the elderly in the population now come to specialists for care. The Mayo Clinic study suggested that it was this factor rather than any change in the age distribution of those with parkinsonism that led to the observations at Massachusetts General Hospital. Older patients, particularly those with moderate symptoms, who might formerly not have sought treatment, are now more likely to consult a physician. The unchanging age of onset for many chronic diseases among the Rochester population can probably be attributed to the ready availability of excellent medical care that older people have enjoyed there for many years. The Rochester study of parkinsonism, along with studies of stroke, brain tumors, rheumatoid arthritis and other diseases, has helped to dispel the idea that the incidence of these diseases reaches a peak in the sixth decade of life and then declines. The studies have shown that the rates of many chronic diseases rise continuously throughout life.

The Rochester studies of base-line incidence thus call attention to some of the biases present in most collections of hospital records. The diagnostic archive is also useful for investigations that touch more directly on the factors responsible for disease. In the 1960's studies based on U.S. mortality statistics showed an increase in the death rate from leukemia from two per 100,000 people per year in the 1930's to about seven per 100,000 in the 1960's. It was suggested that increased exposure to ionizing radiation (from diagnostic tests or environmental radiation) or to chemical pollutants might have been responsible for the increase. The annual incidence of leukemia in Rochester, however, based on all cases diagnosed between 1935 and 1964, was found to have been steady at about eight per 100,000. The reason the national death rate was lower than the Rochester incidence rate is presumably that many cases of leukemia remained undiagnosed in areas of the country with inadequate medical services, particularly during the economic depression of the 1930's.

In 1978 the Rochester leukemia study was extended to include all cases diagnosed through 1974. No increase in incidence among women or young men was found. There had, however, been a rise in the incidence of leukemia among men 50 years old and older. Almost all the increase could be attributed to acute myelogenous leukemia, a form of the disease in which environmental factors are thought to be important. With the aid of the diagnostic index an investigation was undertaken to identify the factors that might have been responsible for the increase.

The case-control study is the epidemiological method most commonly employed to determine what factors increase the risk of contracting a disease. In such a study a group of people who have the disease are matched with a group of control subjects who do not. An attempt is made to match cases and controls in such a way that they are similar in most important demographic respects, including age, race, sex and sometimes socioeconomic status. If a certain kind of exposure (to ionizing radiation, say) appears more often in the medical histories of the cases than it does among the controls, the exposure is presumed to entail an increased risk of contracting the disease. The ratio of the proportion of cases who have been exposed to the proportion of controls who have been exposed is taken as an estimate of the increase in risk associated with the exposure.

Case-control studies are often done by selecting a series of cases from the records of one service at a hospital and a group of controls from other services at the same hospital. Investigations of this kind are informative and relatively easy to carry out; they have provided much of the current base of knowledge in epidemiology. They are, however, replete with methodological complexities. The selection of an appropriate group of controls is by no means straightforward; several biases may be introduced in the process. Control subjects (particularly those from other services at the same hospital) often have characteristics related to their being sick that systematically affect their chances of having had the exposure under investigation. To overcome such difficulties it is almost always necessary to exclude certain groups of control subjects; the criteria for exclusion are rarely unambiguous. Moreover, the selection of different groups of controls generally gives rise to quite different estimates of risk. Case-control studies also cannot provide an estimate of the base-line incidence or prevalence of a disease, only an estimate of the relative levels of risk among exposed and unexposed patients.

Case-control studies in Rochester based on the diagnostic archive avoid some of these shortcomings. In the first place the records of the clinic and the additional sources in the archive are of sufficient accuracy and completeness that they can provide base-line-incidence rates (as in the leukemia study). After this has been accomplished the record system can also supply a group of controls who have basic demographic backgrounds similar to those of the cases. The controls are particularly suitable for purposes of comparison because, rather than having a serious disorder, which complicates a study of this kind, they have often consulted the Mayo Clinic or one of the other local medical facilities for a minor ailment or a routine procedure such as a checkup, and are in good health.

In any single year about 80 percent of the community will have been seen at the clinic or at one of the other health-care facilities in the community and en-

CURRENT OPERATION of the diagnostic archive requires the close cooperation of many departments to index 885,000 diagnoses and 50,000 surgical procedures each year and to retrieve dossiers for hundreds of epidemiological studies. Eight phases of indexing and retrieval are shown. (*1*) A physician enters notes into the patient record. (*2*) Laboratory results are transcribed onto summary sheets; test results are recorded in the laboratory on yellow slips, then collated and summarized on standard forms. (*3*) The code for a diagnosis is looked up in the diagnostic index and recorded on a small card. (*4*) The code is checked and then entered into a computer file by means of a visual-display terminal. (Shown on the screen is the list of diagnoses for a single patient.) (*5*) The patient record is stored among 3.5 million dossiers in an area that occupies two floors of a clinic building. (*6*) A dossier is identified for retrieval at the request of the attending physician by means of a location file manned by eight clerks who monitor about 50,000 current records. (*7*) For an epidemiological study records are brought from the central storage area to the medical-statistics unit by a dumbwaiter. (*8*) Information is abstracted from patient records for an epidemiological study. The investigator provides a form listing the information needed from the dossier; trained abstractors record the data from each dossier.

02

I. Infective and parasitic diseases
II. Neoplasms
III. Endocrine, nutritional and metabolic diseases
IV. Diseases of the blood and blood-forming organs
V. Mental disorders
VI. Diseases of the nervous system and sense organs
VII. Diseases of the circulatory system
VIII. Diseases of the respiratory system
IX. Diseases of the digestive system
X. Diseases of the genitourinary system
XI. Delivery and complications of pregnancy, childbirth and the puerperium
XII. Disease of the skin and subcutaneous tissue
XIII. Disease of the musculoskeletal system and connective tissue
XIV. Congenital anomalies
XV. Certain diseases peculiar to newborn infants
XVI. Signs, symptoms and ill-defined conditions
XVII Injuries and adverse effects

027

240–246 Diseases of the thyroid gland
250–258 Diseases of other endocrine glands
260–269 Avitaminoses and other nutritional deficiencies
270–279 Other metabolic diseases

CLASSIFICATION OF A DISEASE in the current diagnostic index is based on a hierarchy of taxonomic groups. Each disease is given a unique eight-digit code number, which represents its niche in the classification system. The code combines the four-digit *International Classification of Diseases* (ICD) system with further classifications devised at the clinic. Each disorder has a place in five successively finer classes, beginning with one of the 17 major ICD groups, which together include all diseases, and ending with a small category made up of closely related disorders. Each group contributes one digit or more to the code. The taxonomy shown represents the classification

tered into the comprehensive index. In any three-year period the proportion exceeds 95 percent for most age groups. Matching of cases and controls by age, sex and city or rural residence is readily accomplished; perhaps the major feat is selection of the control whose registration number among controls is closest to that of the case. This ensures a similar interval of residence and medical coverage and increases the likelihood that the possible risk factors will have been recorded for both the case and the control.

A case-control study was undertaken to examine some of the factors that might be related to the development of leukemia, with the particular aim of explaining the recent increase in leukemia among men over 50. Data in the clinic records indicated the extent of diagnostic and therapeutic radiation among cases prior to the onset of symptoms of leukemia. Similar information was obtained for matched controls without leukemia. Such exposure was not found to be any greater among the cases than it was among the controls. A similar analysis was made to see whether health workers who might have increased exposure to radiation and farmers who probably have increased exposure to pesticides were in excess among those with leukemia. No such association was found in a recently completed study, although further investigations of larger cohorts are still recommended.

The diagnostic archive has been shown to be valuable in descriptive studies of the incidence and prognosis of disease and in case-control studies aimed at identifying risk factors. The optimum method in epidemiology, however, is neither of these but a prospective study, in which several large groups of subjects are followed for a long period. The groups may differ in important respects, such as exposure to suspected pathogens; they are monitored to determine whether exposure leads to a higher incidence of disease.

A well-designed prospective study is considered the most reliable means of showing an association between a suspected risk factor and subsequent disease because it eliminates many of the problems of the case-control study and approximates the experimental model of the physical sciences. A prospective study of women clients of the Family Planning Association in Oxford, England, has provided some of the most reliable data on the long-term health effects of oral contraceptives. A prospective study of the residents of Framingham, Mass., is yielding information about risk factors in the development of cardiovascular disease. The typical prospective study, however, is expensive and complicated to administer; moreover, biases are sometimes introduced when patients move away from the area and are lost to observation.

Epidemiologists have long been aware that some of the difficulties of the prospective design might be circumvented if the subjects could be traced not in life but in the medical records of a large population. Few sources of medical records provide sufficiently complete coverage of an area or sufficiently thorough medical histories for such a "historical cohort study." Some health-maintenance organizations, such as the Kaiser-Permanente Medical Care Program in the Western states, and the Group Health Cooperative in Seattle, now keep records that would be adequate. For most of the organizations, however, the records go back only a few years; in addition the health-maintenance organizations often draw patients from a relatively narrow socioeconomic stratum.

In many instances the Mayo Clinic diagnostic archive does satisfy the requirements of a historical cohort study. The archive makes it possible to identify a cohort exposed to a suspected risk factor, to efficiently monitor the cohort for some outcome (usually a disease process) and to estimate the incidence of the outcome in the general population (of Rochester) as a basis for determining whether the rate in the cohort is excessive. Several such studies have been done at the Mayo Clinic. One of the most important focused on the long-term consequences of the use of the synthetic estrogen hormone diethylstilbestrol (DES). In the 1940's and 1950's DES was widely prescribed during pregnancy to prevent possible miscarriage. In 1971 a case-control study done in Boston reported an elevated rate of vaginal clear-cell cancer, which is generally a rare condition, in young women who had been exposed to DES in utero.

The 1971 study, with its case-control design, could not estimate the baseline incidence of vaginal clear-cell cancer in the population at large or among those exposed to DES; it did demonstrate a strong association between in utero exposure to DES and this rare cancer. What was needed was a prospective study of those who had been exposed during gestation. The necessary information was found in the Mayo Clinic archive.

In 1943 Berkson had added to the files of the prenatal-care index a code for the use of estrogens during pregnancy. An examination of the records of all pregnancies in the Rochester area between 1943 and 1959 showed that DES and other estrogens had been administered in about 7 percent of them. At the time this was the only available information on the proportion of infants who had been exposed to DES. On this basis it has been estimated that from one to two million daughters and an equal number of sons had been exposed, along with their mothers. Medical histories and later examinations of several hundred of the exposed daughters in Rochester disclosed no cases of vaginal clear-cell cancer; the absence of cases, however, was

0273

- 270 Congenital disorders of amino acid metabolism
- 271 Congenital disorders of carbohydrate metabolism
- 272 Congenital disorders of lipid metabolism
- 273 Other and unspecified congenital disorders of metabolism
- 274 Gout
- 275 Plasma protein abnormalities
- 276 Amyloidosis
- 277 Obesity not specified as of endocrine origin
- 278 Other hyperalimentation
- 279 Other and unspecified metabolic diseases

02737

- 273.0 Cystic fibrosis
- 273.1 Disorders of porphyrin metabolism
- 273.2 Hemochromatosis
- 273.3 Hepatolenticular degeneration
- 273.4 Other disorders involving metabolism of minerals
- 273.5 Defective bilirubin excretion
- 273.6 Disorders of corticosteroid metabolism
- 273.7 Mucopolysaccharide disorders
- 273.8 Other disorders of renal transport
- 273.9 Other and unspecified

02737 12 0

02737	12	0	Hurler syndrome
02737	12	1	Lipochondrodystrophy
02737	12	2	Multiplex dysostosis
02737	13	0	Gargoylism
02737	14	0	Hunter syndrome
02737	21	0	Osteochondral dystrophy
02737	21	1	Morquios disease
02737	21	2	Osteochondrodystrophy
02737	21	3	Morquio-type kyphosis
02737	21	4	Mucopolysaccharidosis

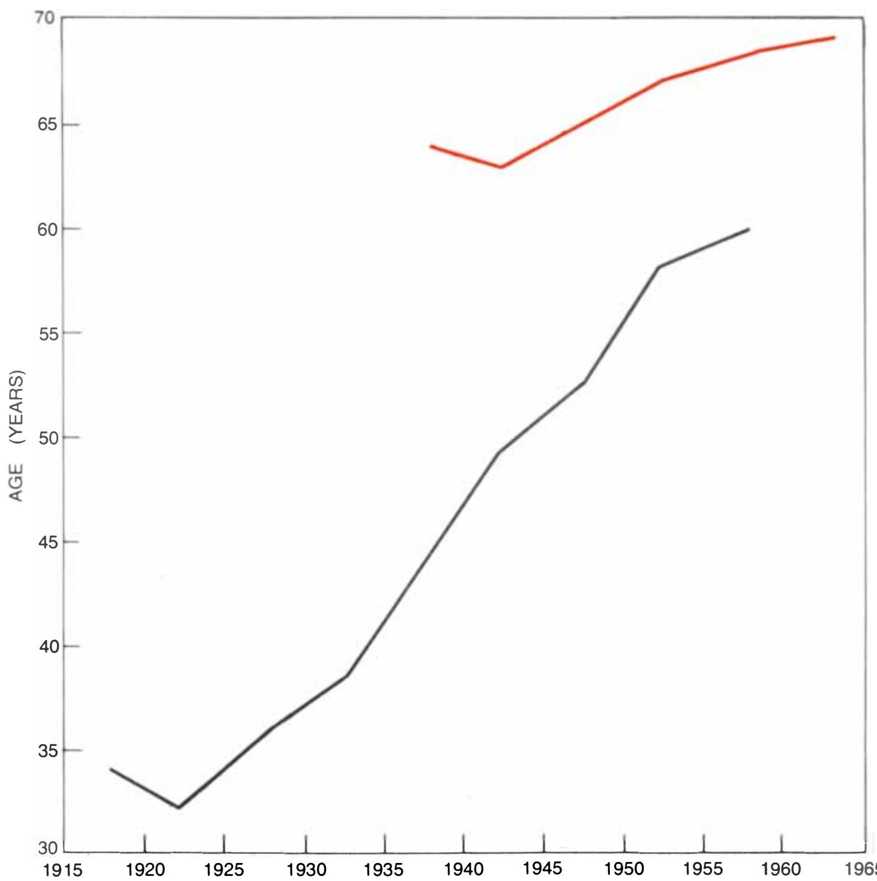
of Hurler syndrome, a genetic disorder of mucopolysaccharide metabolism that gives rise to heart disease and mental retardation and usually leads to death before age 10. The code for the syndrome is 02737 12 0. The initial 0 is an internal clinic code used for checking the accuracy of the data. The ICD code for the class of diseases that includes Hurler syndrome is 273.7, where the first three digits designate a broad group of disorders and the fourth digit specifies a subgroup. The clinic adds the three final digits: 12 identifies the disease and 0 picks out one name for it among synonyms. Thus 12 0 and 12 1 (lipochondrodystrophy) identify different names for one disorder.

still compatible with an incidence rate of the cancer among the exposed as high as nine per 1,000 by age 19. In view of the fear that a high proportion of the daughters exposed to DES might develop the cancer, the results of the Mayo Clinic report were reassuring.

Even the sample represented by the diagnostic archive, however, was too small to determine the incidence with certainty. For this reason a national registry of cases was established and prospective studies of 5,000 women from many parts of the country were begun in several medical centers in 1974. The results of the national studies indicate that the risk of contracting the vaginal cancer by age 20 is likely to be in the range from 1 to 10 per 10,000 among the exposed daughters. The studies have also shown that the daughters have an increased risk of miscarriage, and that some 35 percent of them have benign, or noncancerous, vaginal lesions.

The history of the DES project suggests the limits of the Mayo Clinic diagnostic archive as a resource for epidemiology. For many descriptive and analytical studies the archive is a better source of data than prevalence surveys, hospital records, vital statistics or the records of most health-maintenance organizations. The value of the archive is limited, however, by the size of the local population. As in the instance of DES, the archive is too small to provide definitive information about some rare but potentially important conditions. Furthermore, the homogeneity of the population limits the extent to which results based on the diagnostic archive can be generalized.

To overcome these limits it will be necessary to apply the Mayo Clinic model in other areas. If a diagnostic archive, including indexes based on clinical diagnosis and surgical procedure and with readily available supporting material such as autopsy and laboratory reports, were instituted in other medium-size cities capable of being covered by a single record-linkage system, the combined records would provide epidemiological knowledge that is now beyond our grasp. Such archives could provide a check on the results of the Rochester studies, indicating the extent to which the results are biased or peculiar to the region. They could also increase the reliability of findings on rare conditions and by so doing extend the existing picture of the complex and regular patterns in which diseases afflict the human population.



AGE AT DIAGNOSIS of parkinsonism increased far less between 1935 and 1965 in Rochester (*color*) than it did in a series of cases at Massachusetts General Hospital (*black*). The increase observed among cases at the Boston hospital led to the hypothesis that a single exposure to a virus in about 1918 was responsible for all these cases of parkinsonism after a variable latency period. The more complete records in the Mayo Clinic archive indicated little increase in age of diagnosis with time, throwing doubt on the single-exposure hypothesis. Studies based on the archive have uncovered similar artifacts in the rates of several other chronic diseases.

An Early Habitat of Life

Microorganisms may have flourished on tidal mud flats 3.5 billion years ago. The environment can be reconstructed from sedimentary deposits in Australia, but the evidence of life remains tentative

by David I. Groves, John S. R. Dunlop and Roger Buick

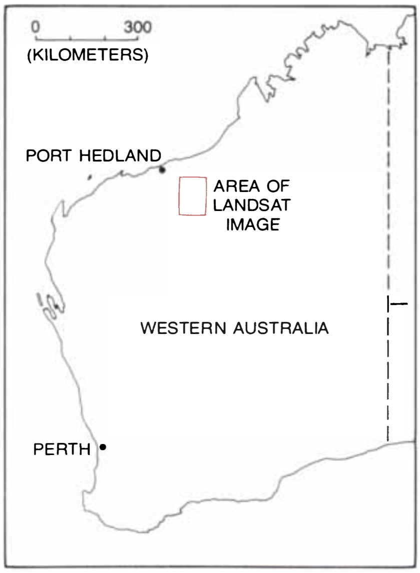
The history of the earth is written in its rocks, the history of its inhabitants in the fossil record. Geology and paleontology complement each other. Through detailed examination of sedimentary rocks it is possible to reconstruct past environments, to determine whether or not they might have supported life and, if they did, what kind of life it was.

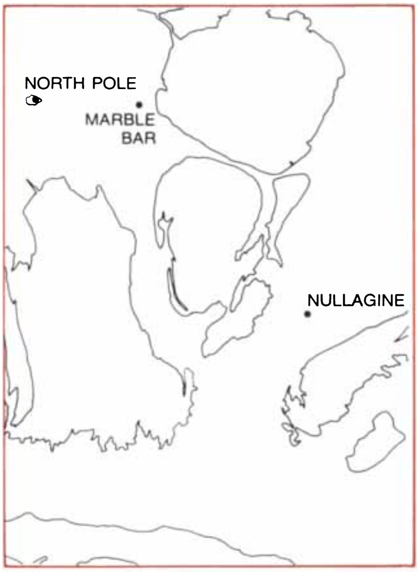
For the past several years we and a number of colleagues have been examining sedimentary rocks 3.5 billion years old that are exposed at a very hot place in Western Australia called North Pole. We have discovered what is probably evidence of early life: objects that resemble microfossils, which are the remains of ancient microorganisms, and also probable stromatolites, which are layered structures formed as the result of the accretion of fine grains of sediment by matted colonies of microorganisms. These things are found in well-preserved sediments, and our study of the sediments indicates that the early Precambrian environment at North Pole was one in which primitive organisms could have flourished. Taken together, the geological and the paleontological evidence suggest that North Pole was indeed the habitat of some of the earth's earliest living things.

The macroscopic fossil record goes back some 650 million years, to just before the beginning of the Cambrian period. Over the past two decades the remains of flourishing communities of microscopic algae and bacteria have been found in much older sedimentary rocks, and gradually the age of the oldest life forms known has been pushed back further into the Precambrian era. The course of evolution is now fairly well established back to about 2.7 billion years ago. Objects that appear to be microfossils and stromatolites have been discovered in even older sediments, but they are rare and poorly preserved, and for most of them a biological origin cannot be proved. A thorough knowledge of the ancient environment can help in assessing the likelihood that such equivocal remains are true fossils.

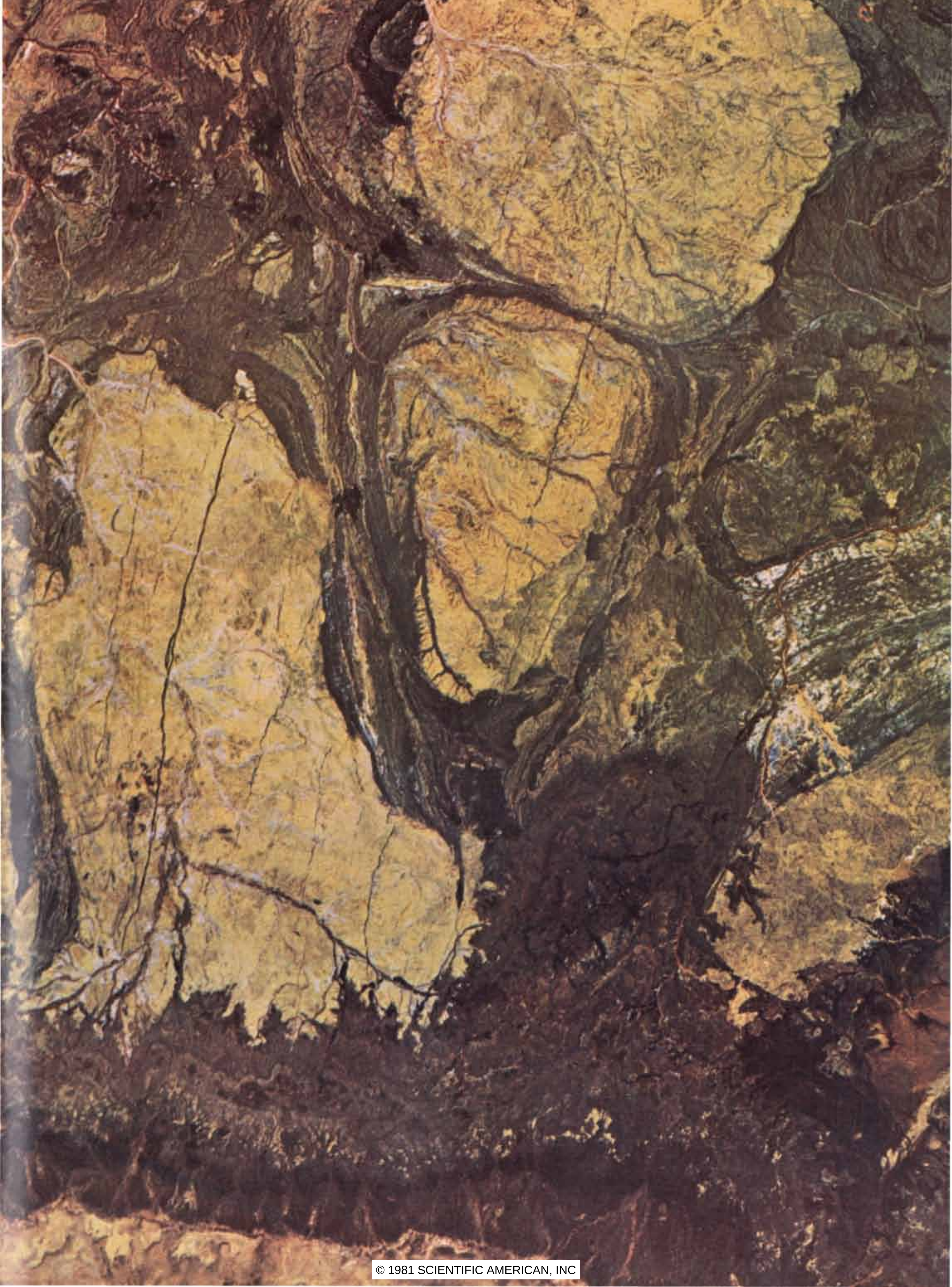
Reconstructing an environment can do more than show whether or not a sedimentary basin was suitable for colonization by ancient life forms and for the preservation of relics of those forms. The topography of the site can be determined, and so can the strength and direction of currents, tides and waves. Analysis can reveal much about the geology of the source of the sediments and can show how the sediments changed physically and chemically after they were deposited. It may even be possible to deduce the composition and the temperature of the early atmosphere and oceans. Such conditions determine what organisms exist, influence their way of life and act as selective forces in their evolution. The pathways favored by natural selection were particularly significant during the early Precambrian era because the basic biochemistry of all later organisms was determined by the outcome of this earliest interaction of organisms and environment. The geology that enables us to reconstruct the North Pole environment is the key to understanding its life forms and their preservation. We shall therefore discuss the geological findings before considering the evidence of life.

The North Pole area in northwestern Australia is part of a complex terrain, the Pilbara Block, in which ovoid granite domes are surrounded by folded sequences of both volcanic and sedimentary rock called greenstone belts. The age of about 3.5 billion years assigned to the North Pole rocks was established by radioactive-isotope dating. The rare-earth element samarium has a





POSSIBLE HABITAT of early life is in northwestern Australia at a site called North Pole. The site and its setting in the Pilbara Block are seen in the Landsat satellite image on the opposite page, which was processed by the Division of Mineral Physics of the Commonwealth Scientific and Industrial Research Organization in Sydney. Ovoid granite domes (*light color*) are enveloped by greenstone belts (*dark colors*). The map shows the location of the area included in the image and the drawing relates the North Pole site to major features of the image.



SEDIMENTS AT NORTH POLE are well preserved, unlike those in most other ancient beds. This sequence of chert (fine-grained silica) and paler barite (barium sulfate) has not been folded or appreciably tilted. The reddish tone is from iron-rich surface deposits.

MOUND-SHAPED STRUCTURE discovered at North Pole is considered to be an exposed vertical section of a stromatolite: a layered sedimentary structure formed by the accretion of mineral grains by matlike colonies of microorganisms. The arched nodule, about 20 centimeters high and 25 centimeters wide, is composed of a number of smaller nodules, each one made up of stacks of arched layers.

radioactive isotope, samarium 147, that decays to yield neodymium 143. Because the rate of decay is constant the age of a rock can be determined by comparing its relative content of the two isotopes. This method has been applied to rocks in a part of the Pilbara Block adjacent to the North Pole site and lying within the same broad geological unit (called the Warrawoona group). Stratigraphically equivalent rocks from other areas, including some from North Pole itself, have been dated by means of similar radioactive-isotope systems. All the analyses support the approximate age of 3.5 billion years.

Considering the great antiquity of the North Pole sequence, its degree of preservation is remarkable. Most early Precambrian rocks have been subjected to heat and pressure during periods when they were buried deep in the earth's crust; as a result they exhibit considerable metamorphism, or changes in mineralogical structure. The North Pole rocks show little evidence of metamorphism. Two minerals present in North Pole basalts, prehnite and pumpellyite, would not have been preserved if the basalt had ever been heated to more than about 300 degrees Celsius, even at a pressure of less than some 3,000 atmospheres. Whereas most rocks as old as these have undergone several episodes of folding, North Pole strata are tilted only some 30 degrees from the horizontal; folding is significant only near fault lines.

In some sedimentary rocks pressure and deformation have dissolved the more soluble components, modifying or obliterating structural details that show how the sediments were deposited. Such processes have had only a minor effect at North Pole because most of the more soluble minerals were replaced early in their history by silica (silicon dioxide) or barite (barium sulfate), which are resistant to solution under pressure. Both the volcanic and the sedimentary structures and textures at North Pole have therefore been preserved with comparatively little alteration since the time of deposition, even though the original minerals that constituted the rock are gone.

The sedimentary rocks bearing evidence of biological activity are minor constituents of the Warrawoona group. Most of the rocks that make up the group are volcanic; they form a sequence several kilometers thick that extends over most of the Pilbara Block. At North Pole the major volcanic rock is basalt, much of it characterized by tubular or bun-shaped structures. Such "pillows" are formed only when lava erupts underwater. The North Pole pillows have vesicles (holes left in the cooling lava by gas bubbles) whose abundance and large size show the lava was erupted in water less than 100 meters deep.

Although basalts dominate at North Pole, in other parts of the Pilbara Block there are thick sequences of paler felsic rocks, which are fine-grained volcanic rocks with a chemical composition similar to that of granite. The felsic rocks are interdigitated with the same basalts seen at North Pole, indicating that they were deposited at about the same time. As in the case of modern felsic volcanoes, most of the felsic material in the Warrawoona group was the product of explosive volcanism. (The basalt eruptions at North Pole were less violent.) The felsic material was erupted on land or in very shallow water.

From findings such as these one can reconstruct the volcanic environment, which was the setting for the deposition of sediments. Evidently the scene was dominated by extensive shallow seas into which basaltic lavas were erupted; small basaltic land masses were formed by local uplift. Scattered islands formed by felsic volcanoes rose higher, and volcanic products eroded from their flanks were deposited between successive basaltic lava flows.

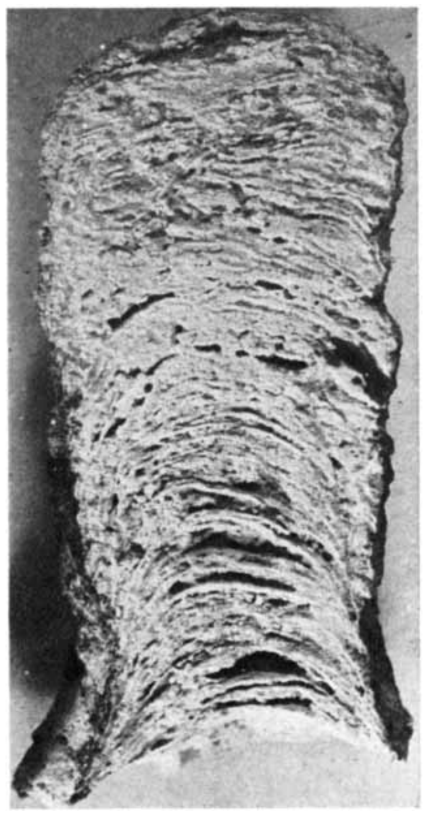
The lowest sedimentary layer interpolated into the volcanic sequence at North Pole is a chert-barite unit about 30 meters thick, and it is this unit that holds evidence of early life forms. The unit is dominated by cherts: rocks composed of microscopic grains of silica. Most early Precambrian cherts have been considered to be chemical sediments, precipitated from silica-rich water in deep oceans. The abundance of chert in deposits older than 2.5 billion years has sometimes been held to indicate that the crust of the earth was deeply submerged until then.

The clear evidence at North Pole of volcanic eruption into shallow water rules out such an origin for these cherts. Moreover, examination of the cherts under the microscope reveals the presence of clasts, which are fragments of rock that have the appearance of perfectly preserved grains of sand, silt and mud; the clasts are formed into structures characteristic of sediments deposited from running water. A two-stage process can account for the formation of the cherts. First the sand and other shallow-water materials were deposited; later silica-rich water circulated through the porous sediment, dissolving the original minerals and precipitating silica in their place when the temperature and acidity were favorable. The silica took on the form of the original grains.

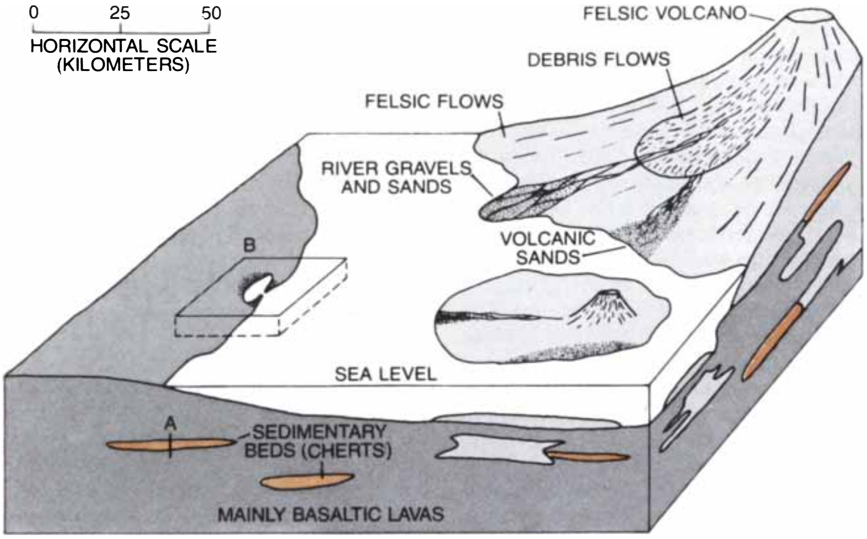
The presence of silica-rich water is readily explained. Modern ocean water is deficient in silica because organisms such as sponges and diatoms extract silica to build their skeletons. In the early Precambrian there were no silica-fixing organisms as far as is known. The seas must have contained much more dissolved silica, leached out of volcanic rocks by volcanically heated water.

One can reconstruct the environment in which these ancient sediments were deposited by comparing their structure and texture with those of sediments being deposited today along coastlines, in lakes, estuaries and rivers, on the ocean floor and elsewhere. Each modern sedimentary environment gives rise to a characteristic sequence of sand, silt, mud and other sediments having specific relations to one another in time and space. The composition, structure and texture of each kind of sediment reflect the conditions under which it was deposited: the depth of the water, the velocity of the current, the source of the clasts and so on. To interpret the North Pole rocks one needs to study distinctive sedimentary structures and the relation of one body of sediment to another both vertically and laterally. The vertical sequence of sediment types records the evolution of the environment over time; lateral variations (which are called facies changes) indicate the range of environmental conditions at any one time.

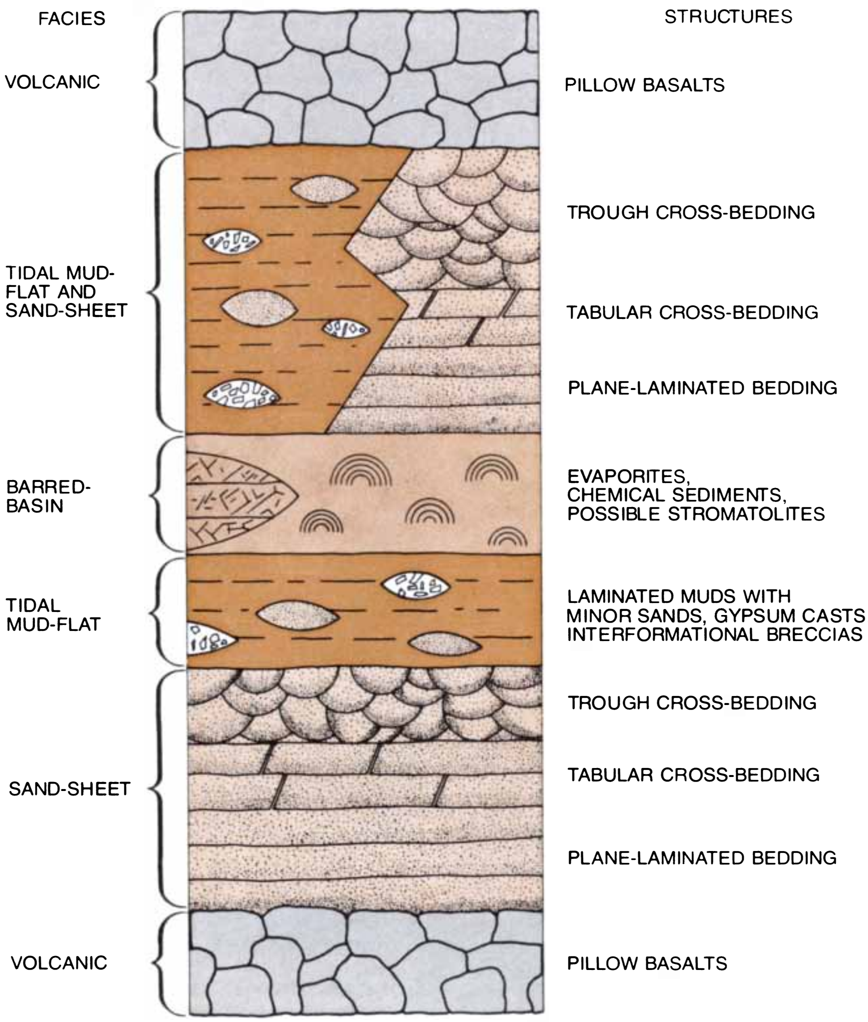
The most telling indicators of the North Pole sedimentary environment are beds of evaporative sulfates, which are now made up mostly of barite. Evaporites are chemical precipitates that crystallize slowly from brines dur-



MODERN STROMATOLITE from Shark Bay in Western Australia is about 15 centimeters high. The arched structures are formed when layered colonies of microorganisms are covered by accreting mineral grains. New growth is then fastest where the most light is available for photosynthesis, so that the colonies tend to build a pile of draped layers.



VOLCANIC ENVIRONMENT of the Pilbara Block some 3.5 billion years ago is reconstructed. The major rock type is basalt (*dark gray*), which was erupted into shallow seas and later was uplifted. Felsic volcanoes erupted paler rock (*light gray*) that became interdigitated with the basalts. In this volcanic setting sedimentary sequences (*color*) were laid down from time to time; one sequence is exposed at North Pole. A section (*A*) through a sedimentary sequence is diagrammed in the illustration below; locale of North Pole sediment deposition is shown at *B*.



STRATIGRAPHIC SECTION through the North Pole sediments shows a typical sequence of facies, or rock types (*left*), and the structures seen within each facies (*right*). The original minerals have largely been replaced by chert or barite. The entire chert-barite unit is a minor component within a thick sequence composed of pillow basalts and other volcanic rocks.

ing the evaporation of shallow ponds and lakes, usually in arid regions. The barite beds at North Pole are composed of upward-radiating groups of crystals as much as 20 centimeters thick; in some specimens sediments are either draped over the crystals or between them. These features indicate that the crystals grew upward in standing water, either on the sediment surface or under a thin cover of sediment. The crystals are almost identical in appearance with younger evaporative sulfates from the Australian coastline and the rim of the Mediterranean. Like most modern sites of evaporite deposition, the North Pole site was probably isolated from the sea, but the scoured surfaces and the broken tops of some crystals suggest that from time to time the usually tranquil site was occasionally exposed to the action of storms.

Modern evaporites generally consist of calcium or calcium-magnesium carbonates, gypsum (hydrated calcium sulfate) or halite (rock salt). Because barite is not a common evaporite mineral we had to consider the possibility that at North Pole it had replaced some other mineral. The similarities between the ancient barite and modern evaporitic gypsum are striking. Some of the crystals have a swallowtail shape that does not develop in primary barite but is a distinctive feature of gypsum. The angles between crystal faces are typical of gypsum, not of barite. In some cases, then, the barite must be a replacement for what was originally gypsum (although there is indirect evidence that some of the barite was a primary precipitate).

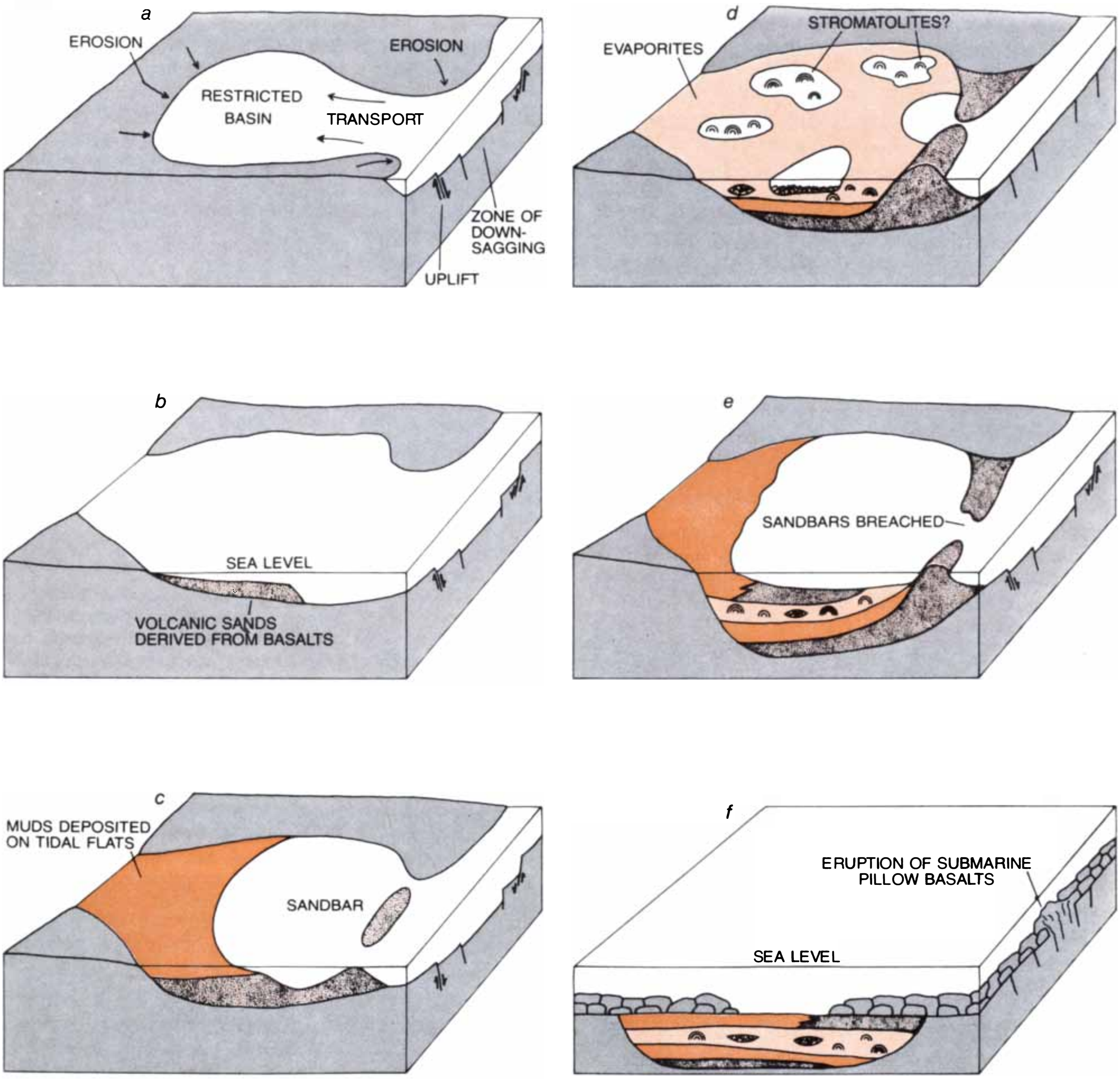
We conclude that the North Pole barite beds were formed on the bottom of shallow, calm ponds. During rare storms, when waves breached the barrier between the ponds and the sea, crystals were eroded, other sediments were deposited and the supply of sulfate was replenished. Analysis of the ratios of several stable (nonradioactive) isotopes of sulfur in the barite suggests that the site of deposition had access to a large and homogeneous reservoir of dissolved sulfate, presumably the open sea.

Distinctive units of silicified sand, silt and mud lie below and above the evaporites. At the base of the sequence thick sand beds were built up on the underlying basalt. The sands are derived mainly from the outer surface of basalt pillows that were eroded by waves. Since the pillow lavas were erupted underwater, the erosion must have taken place after either tectonic uplift or a fall in the sea level had exposed the volcanic rocks.

The grains of the sand are angular and are poorly sorted by size, indicating that they did not travel far between erosion and deposition. The sand beds generally have no silt or mud fraction. Each sand unit has flat, laminated beds at the base and large cross-beds (formed by the movement of large ripples) at the top. The sand changes progressively from coarse to fine as one moves upward within a bed and from one bed to the next. All these features are consistent with deposition in a tidal inlet, into which a dominant flood tide swept sand that had been eroded from basalt headlands; fine material was removed by a much weaker ebb current. As the sands built up shoals, finer particles gradually settled under increasingly sheltered conditions.

Fine-grained, greenish gray rocks overlying both sequences of sand beds are made up of silicified mud and silt. They display a variety of distinctive structures that are good environmental indicators. The indicators include cross-bedding in silts, suggesting that waves and currents moved ripples over the sediment surface, and gravity-induced "flame" structures, which tend to form in semiliquid oozes. In finely laminated muds there are also small crystals of what was once gypsum, some of them in daisy shapes, which suggests a high rate of evaporation. Lenses of fragmental rocks (conglomerates and breccias) are composed of pebbles of desiccated, consolidated mud ripped up by storm waves. Sediments and structures of this kind are characteristic of mud and silt deposited on tidal flats, such as those found today along the margins of the North Sea.

What was the source of the North Pole muds and silts? Some were probably finer grains of the same basaltic volcanic material that supplied the sands. Lenses of sand are found in the mud, indicating that sand and mud were being deposited at the same time in different places, depending on local conditions. Some of the mud may originally have



SEDIMENTARY SEQUENCE at North Pole evolved in an interlude between episodes of volcanism. A tidal basin was formed by uplift and downsagging along faults in the volcanic rock (*a*). Sand eroded from the basalt was deposited in the basin (*b*); then mud was deposited, gradually filling the basin, and a sandbar developed (*c*). As subsidence slowed, evaporites were precipitated from the shallow, briny water of the barred basin (*d*); what appear to be stromatolites were built by microorganisms in shallow ponds. The basin subsided again; sandbar was breached and more sand and mud was deposited (*e*). More basalt was erupted, covering and preserving sediments (*f*).

been fine, airborne felsic ash from distant volcanoes. In addition both the presence of structures resembling those seen in modern carbonate muds and the preservation in these deposits of some carbonate minerals suggest that some of the fine material is silicified carbonate mud that was precipitated during evaporation.

Let us now summarize what can be deduced about the depositional environment and its development over time. It is likely that the sedimentary sequence was marine because it is both based on and overlain by pillow basalts erupted into an extensive but shallow sea. Uplift and faulting of the basalts formed a low land mass with a tidal inlet in which wave-eroded sand was deposited. As the sequence built up, the currents and waves became less vigorous, so that they could carry only fine-grained material; tidal mud flats developed at the head of the inlet and gradually filled the basin. Then slow subsidence formed isolated ponds below sea level, within which evaporites were precipitated. With further subsidence the barrier to the sea was breached and more sand and mud were deposited. The entire sedimentary sequence was rapidly silicified, preserving both the evidence of its origins and the problematical evidence of life that we shall discuss below. Finally, rapid subsidence and renewed volcanism covered the sedimentary sequence and prevented its erosion.

BASALT is the dominant volcanic rock at North Pole both below the chert-barite sedimentary sequence and above it. Most of the basaltic lava flows solidified in distinctive tubular and bun-shaped structures called pillows, which form only when lava is erupted underwater.

BARITE BED at North Pole (*left*) is compared with modern gypsum evaporites photographed by Marjorie D. Muir at Marion Lake in South Australia (*right*). Both beds are composed of upward-radiating crystals; some of the barite crystals resemble gypsum in shape and the characteristic angles between crystal faces. Resemblance can be explained if crystals were formed as gypsum, which was replaced by barium sulfate as water percolated through beds.

What was it like, there on the mud flats 3.5 billion years ago? Let the reader imagine he is standing on the flats at low tide. The outline of high volcanic cones is visible in the distance through a haze of felsic volcanic ash and of steam from hot lava erupting into the shallow sea; thunderclouds hover near the peaks. Closer at hand the basalt cliffs of the coastline are pounded by waves whipped by high winds. Inland the scene is dominated by hummocks of black basalt lava, their surface covered with a rubble of pillow fragments. The observer is surrounded by a flat expanse of gray mud that glistens when intermittent sunlight is reflected from small crystals of gypsum. A few tidal streams meander across the flats, draining into the sea through a breach in a low bar of black sand. Elsewhere there are scattered pools, shallow and highly saline.

Could such a place have been inhabited by primeval microorganisms? Certainly bacteria and algae thrive in modern environments much like it. North Pole is a logical place to look for evidence of early life.

Three kinds of evidence of life are to be found in very ancient rocks: microfossils, stromatolites and certain chemical and isotopic properties characteristic of organic matter. Microfossils are the preserved remains of actual cells, usually of the cell wall. They can be extracted from sediments by dissolving silicate, carbonate and sulfide minerals to leave a carbonaceous residue that includes the microfossils, or they can be identified by examining transparent thin sections of rock under the microscope. Minute carbonaceous objects that have been thought to be microfossils have been found in several early Precambrian sediments, including the oldest sedimentary rocks known, those from Isua in Greenland, which are about 3.8 billion years old, and rocks about 3.5 billion years old from the Swaziland supergroup in southern Africa. There is reason for skepticism, however, about the origin of these objects, which may not be biological. Microfossils are not likely to survive temperatures higher than about 250 degrees C., above which the carbonaceous material is transformed into crystalline graphite. The Isua sequence in particular has been subjected to temperatures higher than that. The North Pole sequence, with its history of low metamorphic temperatures, includes some of the few ancient sediments known that could still harbor recognizable microfossils.

Even in mildly metamorphosed rock it is difficult to show conclusively that an ancient carbonaceous object is a microfossil. The form of the objects is usually simple: most are spheres with few surface features. They could be composed of particles of inorganic carbon squeezed into spherical aggregates by the growth of mineral grains deposited around them. Proof of biological origin is easier if the objects have more complex forms, which are less likely to be produced by inorganic means, but then it must still be shown that the microfossils were deposited at the same time as their host sediment. Many microfossils are found to be relatively recent contaminants.

Stromatolites constitute less direct evidence of life, being the remains not of microorganisms themselves but of sedimentary structures they built. Modern stromatolites are built up by bacteria or algae. Either they trap sediment grains in a jellylike ooze they secrete or their metabolic activity causes sediment to precipitate. Living stromatolites are rare except in shallow-water evaporative environments, where they may be quite common. Fossil stromatolites similar to modern ones are known from as far back as about three billion years. Still older structures that are proposed to be stromatolites, including those at North Pole, are merely domed or undulating irregularities in sedimentary bedding, very different from the complex columnar and branching forms seen in younger rocks. The simpler contortions of bedding could have been caused not only by living organisms but also by inorganic processes such as deformation or evaporative precipitation from geysers, from splashing water or in caves. They cannot be judged to be conclusive evidence of life unless it can be shown they were indeed built by microorganisms. That is usually possible only if undoubted microfossils are preserved within the structure itself.

Chemistry can also provide evidence of life. The biochemical reactions by which living things harness energy give rise to distinctive ratios among the isotopes of some elements. The ratios of carbon isotopes in some early Precambrian rocks have been thought to indicate that photosynthesis was going on then. Similarly, bacterial reduction of sulfate ions to obtain energy alters the normal distribution among the isotopes of sulfur. Kerogen, an amorphous carbonaceous material that is common in some very early sediments, includes many compounds that could have a biological origin. The problem with chemical evidence is that inorganic reactions can give rise to the same products as some biological reactions. Moreover, the isotopic and chemical "signatures" of modern microorganisms are not well understood, making it hard to apply them in the analysis of ancient rocks. As in the identification of microfossils, it is also difficult to be sure the material be-

CRYSTALS OF BARITE are seen here in the orientation in which the original gypsum crystals grew. Sand is draped over and between tops of crystals, indicating that crystals grew in standing water or under thin cover of sediment. Field of view is about seven centimeters wide.

LAMINATED FINE SAND AND SILT in oblique strata called cross-beds are indicative of ripples moved by waves or currents. In this case a current flowed from left to right, moving a ripple in that direction and then scouring off the top; flat laminae were deposited as current's speed increased. Like most other sediments in the unit, the sand and silt were later silicified.

GYPSUM DAISIES (radiating clusters of gypsum crystals) grew in the silt and mud, indicating a high rate of evaporation, which is typical of deposition on tidal flats. Quartz veins, intruded after deposition, cross at the bottom right. Width of field is 2.5 centimeters.

LENSES OF BRECCIA, or angular pebbles, are found in silicified carbonate muds in the sedimentary sequence. Consolidated, dried mud on the tidal flats was broken up by wave action to form the small fragments. The width of this field is about three centimeters.

MORE EVIDENCE OF STROMATOLITES was found in an extensive bed about 50 centimeters thick. As seen in section (*left*) the bed has arched and wrinkled laminations, some forming columns. As seen from above (*right*) the tops of some structures are dome-shaped.

ing analyzed does not represent contamination long after the time of deposition.

North Pole emerged as a key early Precambrian paleontological site in 1976, when the attention of investigators at the University of Western Australia was drawn to the bedded barite there. It had been thought such sedimentary sulfate minerals were found only in much younger rocks, deposited when the concentration of oxygen in the atmosphere had reached a value close to its present one. It has long been assumed that when the North Pole sediments were laid down, the atmosphere and the oceans were strongly reducing, having almost no oxygen. Because sulfates are rich in oxygen the presence of sedimentary sulfates at North Pole suggested that local conditions there were more oxidizing than was normal for the time. Why? One possibility was that oxygen-generating organisms were responsible.

Detailed examination of the sediments by one of us (Dunlop) revealed that some structures and textures resemble those seen at Shark Bay, a modern shallow-water evaporative environment in Western Australia. At Shark Bay cyanobacteria (the photosynthesizing bacteria formerly called blue-green algae) thrive and build stromatolites. The North Pole material was found to include mound-shaped structures similar to Shark Bay stromatolites and also some conglomerates and breccias composed of curved, laminated fragments (intraclasts) with a microstructure like that of stromatolites. Some of the intraclasts had concentric overgrowths of kerogen-rich layers resembling the spherical stromatolites called oncolites.

In work done with Marjorie D. Muir of the Australian Bureau of Mineral Resources, Geology and Geophysics, the silicate, carbonate and sulfide minerals in some of the rocks were dissolved with acids. The residue included several kinds of hollow, carbonaceous microspheroids ranging from one micrometer to 12 micrometers in diameter. Some were linked in pairs or chains and a few were grouped to form tetrads; splits in the surface of many of the microspheroids made it clear they were hollow. The same structures could be discerned in thin sections of intact rocks, demonstrating that they were not laboratory contaminants or artifacts of the dissolution process. The size distribution and other statistical properties of the microspheroid assemblage were similar to those of modern spherical bacteria. It therefore seemed likely that the North Pole microspheroids were microfossils—the oldest ever discovered.

On a brief visit to the North Pole site Stanley M. Awramik of the University of California at Santa Barbara collected specimens of carbonaceous chert. In one specimen he later found microscopic carbonaceous objects resembling filamentous bacteria. A group headed by J. William Schopf of the University of California at Los Angeles found identical microscopic filaments in other samples from North Pole. On the basis of the morphological complexity of these filaments and their close resemblance to some modern cyanobacteria and other bacteria, they have been identified by Schopf and his colleagues as unquestionable microfossils from the early Precambrian era.

Under the supervision of Malcolm R. Walter of the Bureau of Mineral Resources the largest and most complex mound-shaped structure among the original discoveries at North Pole was collected and studied in detail. It is composed of several smaller nodules. Each nodule in turn is made up of arched laminae (some as thin as 20 micrometers) that form hummocks over patches of intraclasts and ridges (shaped by erosion) in the base sediments. Within a nodule columns are formed by the superposition of small domes in successive laminae; these wrinkled layers are in marked contrast to the sedimentary deposits that enclose them, which are clearly of inorganic origin and are flat-bedded. Although a few nonbiological processes can give rise to such wrinkles, none is consistent with all the features of the observed form. Both the similarity of the mound-shaped structure to modern ones made by sediment-accreting organisms and the fact that it cannot be accounted for by any inorganic process suggest that it is a stromatolite.

It is unlikely that only one colony of stromatolite builders would inhabit an environment favorable to their growth and multiplication, and so we looked for similar structures nearby. Last year a number of good examples were found in another bed about two kilometers from the initial discovery site. Their forms range from nodules through linked hemispherical domes to almost flat laminated structures. They are better preserved than the original discoveries and some of their layers contain kerogen; the kerogenous laminae are stacked in semidiscrete columns, which are most unlikely to have arisen nonbiologically.

Unfortunately none of these four discoveries—the original mound-shaped structures, the microspheroids, the filamentous microfossils or the second group of stromatolite-like structures—provides conclusive proof of early life. The microspheroids are so simple that one can only consider them to be possible microfossils. The more complex filamentous carbonaceous objects are almost undoubtedly fossils, but they have not been proved to be 3.5 billion years old. The rocks containing filamentous material appear to have been contaminated with carbonaceous matter at some undetermined time after sedimentation took place, and that may have been when the filamentous microfossils were incorporated; none have yet been discovered in sediments that are definitely uncontaminated.

To be convinced of the biological origin of a simple stromatolite-like structure one needs to see proof within the laminae themselves that the structure was accreted by microorganisms. That can be demonstrated if microfossils wrap around or grow over sediment particles (showing that the organisms trapped or bound the particles) or if material different from that of the rest of the rock encrusts the fossils (suggesting that its precipitation was caused by the organism's metabolic activity). In the absence of these microscopic features one can say only that the North Pole structures are probable stromatolites.

Even if no single item of evidence from the North Pole site is conclusive, it is difficult to conceive of an acceptable nonbiological origin for all these discoveries. It is unlikely to be merely a coincidence that all the data point in the same direction: the geological evidence depicts an environment suitable for bacterial colonization and a number of distinct lines of paleontological evidence seem to reflect the existence of primitive life forms. The size and shape of the probable microfossils and the form of the stromatolites suggest that the organisms were bacteria. Their presence in a sulfate-rich environment at a time when sedimentary sulfates were rare suggests that they were either oxygen-releasing or sulfur-oxidizing photosynthetic organisms; their shallow-water habitat is consistent with dependence on the energy of light. Whatever the nature of their photosynthesis, free oxygen need not have been present in the atmosphere because sulfur and ferrous iron, which were particularly abundant in the early Precambrian seas, could have combined with any oxygen that was generated by the organisms.

It seems reasonable, then, to conclude there is a high probability of the existence of life at North Pole 3.5 billion years ago. We may have to be content with that statement. After five years of intensive searching we consider it unlikely that more definitive evidence of life will be discovered at North Pole, although it is possible that filamentous microfossils may yet be found in demonstrably uncontaminated sediments or that the known filamentous microfossils will be shown to have been incorporated into the cherts within a short time after sedimentation.

It appears to us to be just as unlikely that conclusive evidence of life will be discovered elsewhere in rocks as old as or older than those at North Pole. This is not to say that life did not exist before 3.5 billion years ago, but only that unequivocal evidence of its existence has probably not been preserved.

Gravitational Waves from an Orbiting Pulsar

Einstein's 1915 prediction that an accelerating mass should radiate energy in the form of gravitational waves is supported by evidence that a pulsar's orbit around a companion star is slowly shrinking

by Joel M. Weisberg, Joseph H. Taylor and Lee A. Fowler

Einstein's general theory of relativity, published in 1915, makes an extraordinary prediction: an accelerating mass should radiate energy in the form of gravitational waves. Yet the waves are so weak and their interaction with matter is so feeble that Einstein himself questioned whether they would ever be detected. In 1974, however, an object suitable for testing the prediction was found: the binary pulsar PSR 1913 + 16. The designation indicates that the radio-emitting pulsar and its radio-silent co-orbiting companion are positioned in astronomical sky charts at right ascension 19 hours 13 minutes and at declination +16 degrees, which places the binary pulsar in the constellation Aquila. In observations made with the largest optical telescopes no periodically blinking object can be seen at the pulsar's radio location. This is not surprising, because among more than 300 radio pulsars identified since the first was discovered in 1967 by Jocelyn Bell and Antony Hewish of the University of Cambridge only two are observable as visible pulsars.

Astronomers today generally agree that pulsars are very small, rapidly spinning, extremely dense stars, composed mostly of neutrons, that are the remnants of supernova explosions. These spinning neutron stars emit a highly directional radio beam that sweeps around the sky once per stellar rotation. An observer receives a pulse of radio waves each time the star's radio beam points at the earth, hence the name pulsar. A pulsar consists of approximately as much material as the sun yet has a diameter of only 20 to 30 kilometers because its atoms are literally crushed out of existence by intense gravitational forces. Pulsars are observed to be spinning at rates of up to 30 times per second. Some of this vast reservoir of rotational kinetic energy is converted (by a mechanism that remains obscure) into radio emission.

The binary pulsar is a unique tool for testing fundamental physical laws. Both the pulsar and its silent companion are more massive than the sun. They travel at velocities that range up to 400 kilometers per second in a tight orbit with a minimum separation about equal to the radius of the sun. These circumstances make the binary pulsar system an ideal laboratory for studies of strong gravitational fields. In particular the strong, gravitationally induced accelerations experienced by the pulsar and its companion should give rise to gravitational radiation. Until the discovery of the binary pulsar the best available gravitational laboratory was the solar system, where there is only one body of stellar mass (the sun) and where the nearest sizable object (Mercury) lies more than 65 solar radii away and is moving at an orbital velocity of less than 60 kilometers per second. Although there are other binary star systems in tight orbits, it is the presence of a pulsar in this particular system that makes possible a powerful test of gravitational phenomena.

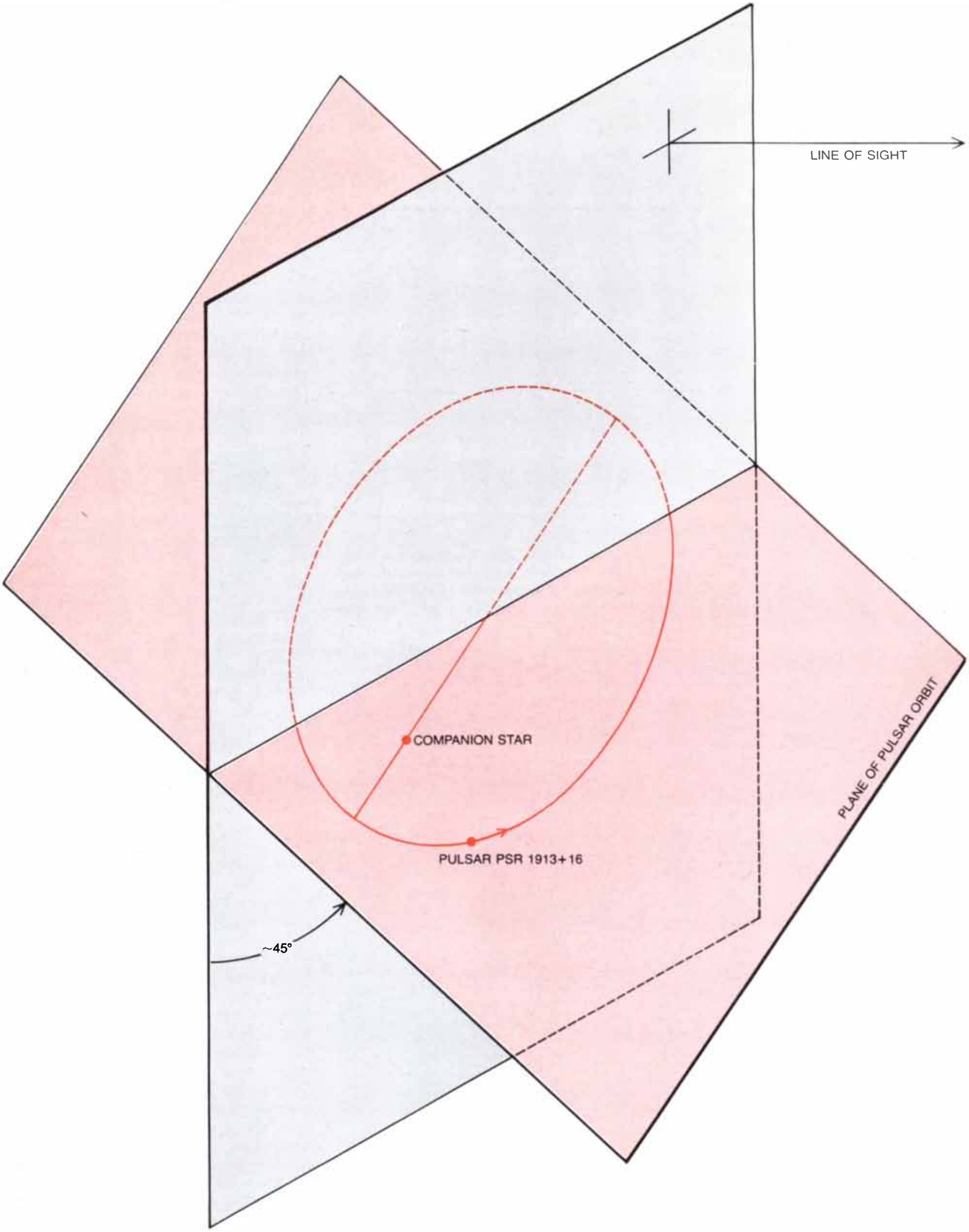
A pulsar is uniquely suited for the task because the pulse-repetition frequency (identical with the number of rotations the star makes each second) is so precisely stable that the pulsar's pulses are like the "ticks" of an extremely accurate clock. The pulse rate is stable because the pulsar acts as a massive, freely spinning flywheel that tends to rotate smoothly for an indefinite time. The binary pulsar is therefore a precise clock orbiting in the strong gravitational field of another massive body. By carefully measuring the times at which its pulses arrive at the earth one can use the pulsar clock to map the orbit and to probe subtle gravitational effects with an accuracy that is not possible in any other known system.

Our measurements of the pulses from PSR 1913 + 16, carried out over the past six years, show that the system is in fact losing orbital energy at a rate close to that expected from gravitational radiation according to the general theory of relativity. Our observations thus provide the first strong evidence for the existence of gravitational waves as well as further confirmation of the validity of general relativity.

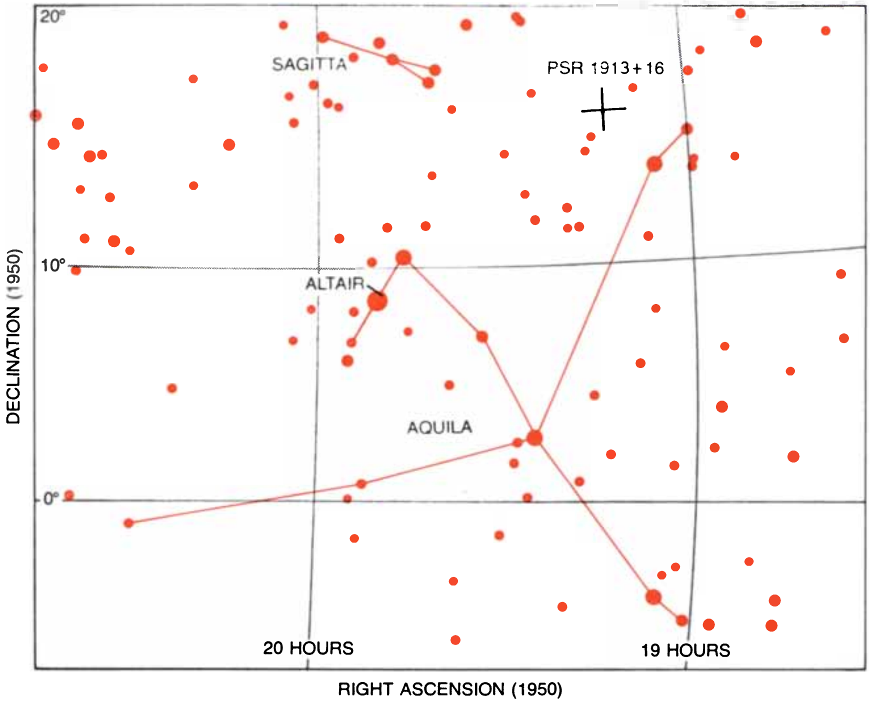
The laws of gravitation and motion developed by Isaac Newton more than 300 years ago have provided a remarkably accurate description of the motions of most orbiting objects. It has been known for many decades, however, that Newton's formulation begins to break down near very massive bodies. For example, there are slight irregularities in the orbital motion of Mercury that cannot easily be explained within the Newtonian framework.

The orbit of Mercury, like the orbit of all other planets, is an ellipse. The point on the ellipse where the planet passes nearest the sun is termed the perihelion. With the passage of time the perihelion advances, or rotates slowly in the same direction as the planet's direction of motion around the sun. Looking down on Mercury's orbit from above (that is, from above the planet's north pole), the advance of the perihelion is counterclockwise. There is a small discrepancy, amounting to only 43 seconds of arc per century (equivalent to less than twice the planet's diameter), between the advance in perihelion predicted by Newton's theory and the advance actually observed. This small anomaly was measured accurately even in the 19th century, before there was any theory to explain it.

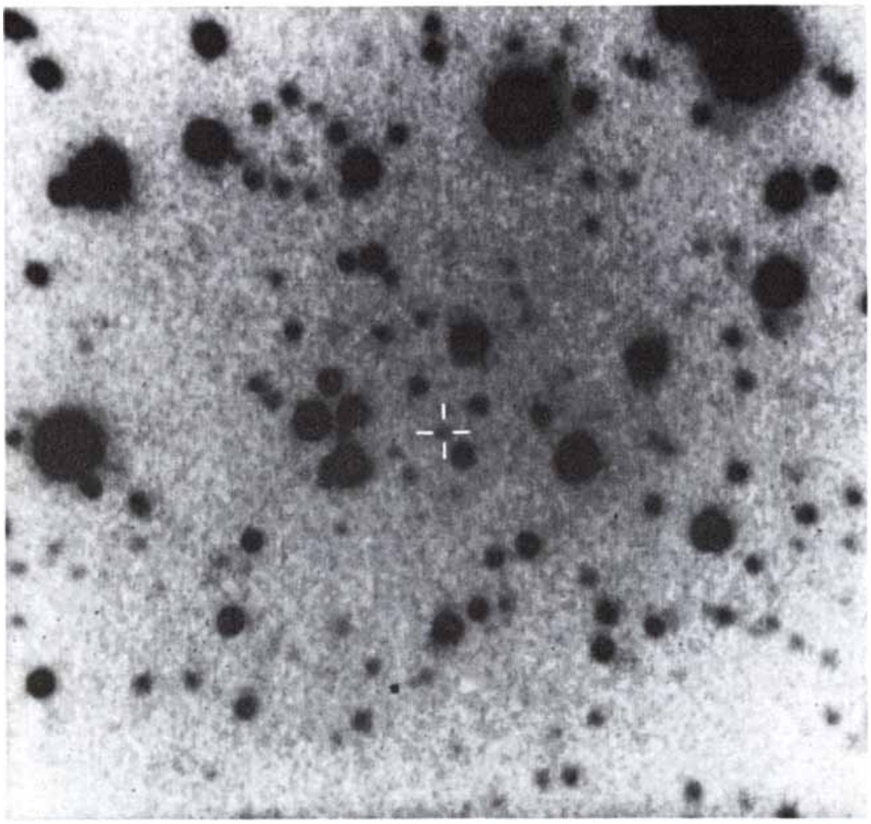
The cause is now generally ascribed to Mercury's being sufficiently close to the sun for Newtonian gravitational theory to begin to fail. The discrepancy is taken as strong evidence that the general theory of relativity is correct, since it predicts a perihelion advance of exactly the right magnitude. Nevertheless, a certain amount of controversy has continued, and attempts have been made to ascribe the discrepancy to other causes. For example, some physicists (most nota-



ORBIT OF PULSAR PSR 1913 + 16 lies in a plane tilted about 45 degrees from the line of sight. Like the more than 300 other pulsars discovered since 1967, PSR 1913 + 16 is thought to be a neutron star, 20 to 30 kilometers in diameter, that emits a radio beam that sweeps past the earth at precisely spaced intervals synchronized with the star's rate of spin. For PSR 1913 + 16 the spin rate is 16.94 revolutions per second. Unlike the large majority of other pulsars, PSR 1913 + 16 travels in an orbit around a companion star whose presence was inferred from a Doppler shift in the arrival time of the pulsar's "beeps." The beeps arrive slightly more frequently when the pulsar is traveling toward the earth and less frequently when the pulsar is receding. A complete picture of the pulsar's orbit around the center of mass of the binary system was derived through careful measurements of the Doppler shift in combination with an analysis of subtle gravitational effects predicted by the general theory of relativity. The theory made it possible to calculate that the pulsar and its companion are both 1.4 times as massive as the sun and that the separation of the two stars varies from 1.1 to 4.8 times the radius of the sun.



BINARY PULSAR PSR 1913 + 16 is in the constellation Aquila at the coordinates that supply its designation: right ascension 19 hours 13 minutes and declination +16 degrees. Its position is marked by the reticle. It is estimated that the binary pulsar is some 15,000 light-years away, too far for it to be observed optically even with the most powerful existing telescopes.



POSSIBLE SILENT COMPANION OF PSR 1913 + 16 is marked by a reticle in this computer plot of visible-light photons recorded with a video camera on the four-meter telescope of the Kitt Peak National Observatory. The observation was made by J. A. Tyson of Bell Laboratories. It has been suggested that the object is a helium-core star: a star near the end of its life that has ejected its outer layers into space, leaving behind a dense core consisting mostly of helium. If the companion is actually another neutron star, as the authors suspect, its visible radiation could not be detected by existing optical telescopes. In that case the object here would probably be a faint star that happens to lie at nearly the same position as the binary pulsar.

bly Robert H. Dicke of Princeton University) have suggested that if the sun were slightly oblate instead of perfectly spherical, at least part of the effect could be accounted for. The discovery of the binary pulsar was therefore particularly intriguing to those interested in fundamental physical laws, because the slight failure of Newtonian gravitation observed in the case of Mercury should be enormously magnified in a binary system where the orbiting bodies are both somewhat more massive than the sun and locked in close proximity.

In Einstein's general theory of relativity the Newtonian concepts of an absolutely definable space and an absolutely definable time are replaced by a single absolute quantity: space-time. Gravitational forces arise from local distortions of space-time caused by massive bodies. The trajectories of orbiting bodies are then seen as being merely the shortest paths that can be taken in warped space-time. Einstein himself showed that the advance of Mercury's perihelion is a natural consequence of the curvature of space-time in the vicinity of the sun. A related prediction of Einstein's theory, recently confirmed, is that radio signals traveling between the earth and a spacecraft on the far side of the sun should be slightly delayed as they pass close to the sun. A comparable time delay, again much magnified, could be expected if the two members of the binary pulsar system are oriented so that radio waves from the pulsar graze the silent companion body on their way to the earth. Therefore it is clear that the equations of general relativity are indispensable to a detailed analysis of the orbit of the binary pulsar. As we shall see, such analysis makes it possible to get much more information about the size and orientation of the orbit and the masses of the pulsar and its companion than could be got with Newtonian theory.

Most exciting of all was the opportunity presented by the binary pulsar system to verify a prediction of general relativity that had never been tested anywhere else in the universe: the prediction that accelerating masses (in this case the orbiting pulsar and its companion) should emit gravitational waves. Such waves, ripples in the curvature of space-time that travel at the speed of light, should be emitted by masses that are accelerating much as electromagnetic waves are emitted by electrically charged particles that are accelerating. The most sensitive laboratory techniques currently available are not nearly sensitive enough to directly demonstrate the existence of gravitational waves from the binary pulsar. According to general relativity, however, gravitational waves should carry a certain amount of energy away from the binary system. That energy should show up as a decrease in the system's orbital energy, resulting in a slight shrinkage in the size of the orbit and a corresponding decrease in the time required for the pulsar to circle its companion. It is this latter change that has now been measured, and with considerable accuracy.

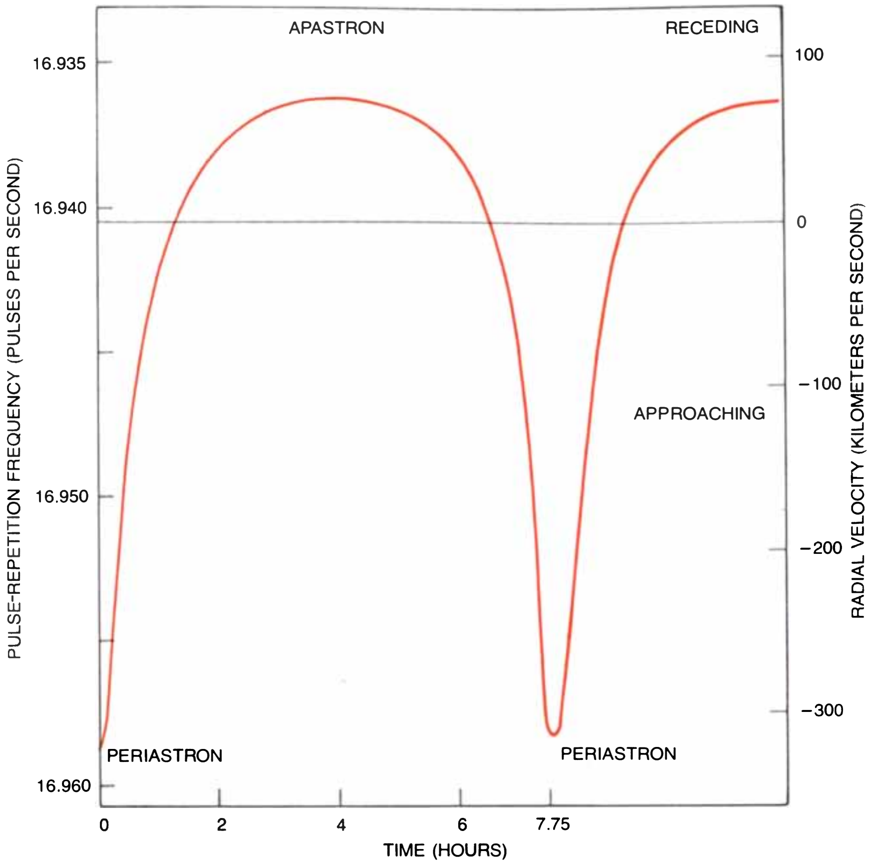
We shall now describe the binary pulsar system in some detail, beginning with the discovery of the pulsar and proceeding to the most recent measurements, which have made it possible to completely specify the orbital geometry and the masses of the two binary components, and also to test the general theory of relativity and other gravitational theories.

In 1974 Russell A. Hulse (then a graduate student at the University of Massachusetts at Amherst) and one of us (Taylor) began a search for new pulsars with the 1,000-foot radio telescope at Arecibo in Puerto Rico. Large areas of the sky were scanned for regularly pulsed signals. Since there are many terrestrial sources of radio noise (such as lightning, radar transmitters and automobile ignition systems) that can give rise to spurious pulsed signals, several procedures were devised to identify actual pulsars. The ultimate method of verification was to make subsequent observations of a candidate pulsar some days after an initial tentative discovery. If a pulsed signal with the same repetition frequency was detected on both days, the pulsar was accepted as real.

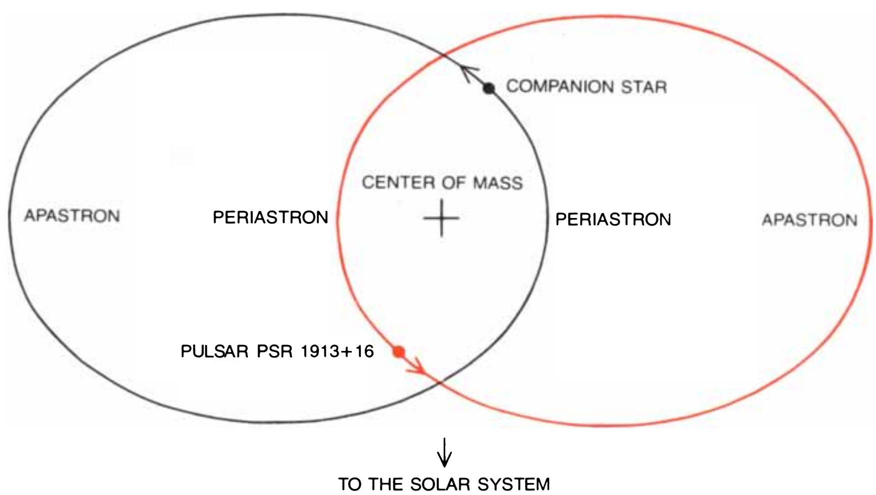
The Arecibo search revealed 40 new pulsars. Certainly the most remarkable was PSR 1913 + 16, whose pulse-repetition frequency was curiously not quite constant from day to day. After some detective work Hulse determined that the variation was cyclical, repeating itself every 7.75 hours. A natural explanation was that the pulsar was moving around another body in an orbit with a period of 7.75 hours. When the pulsar was traveling toward the earth in its orbit, the pulses were crowded together and the pulse-repetition frequency was slightly higher than the average 16.94 pulses per second; by the same token when the pulsar was receding from the earth, the pulse-repetition frequency was lower than average.

This behavior is simply a description of the well-known Doppler effect, where the observed frequency of any clock (including the regular waves of sound or light) is increased if the source and the observer approach each other and is decreased if they recede. The frequencies of the spectral lines observed in many binary stars exhibit a similar Doppler effect and for the same reason: orbital motion. Whereas the observable clock in a normal star of a binary pair consists of atoms in the star's atmosphere emitting or absorbing light of specific frequencies, the pulsar's clock is its rate of spin. In either case the same kind of information is obtained.

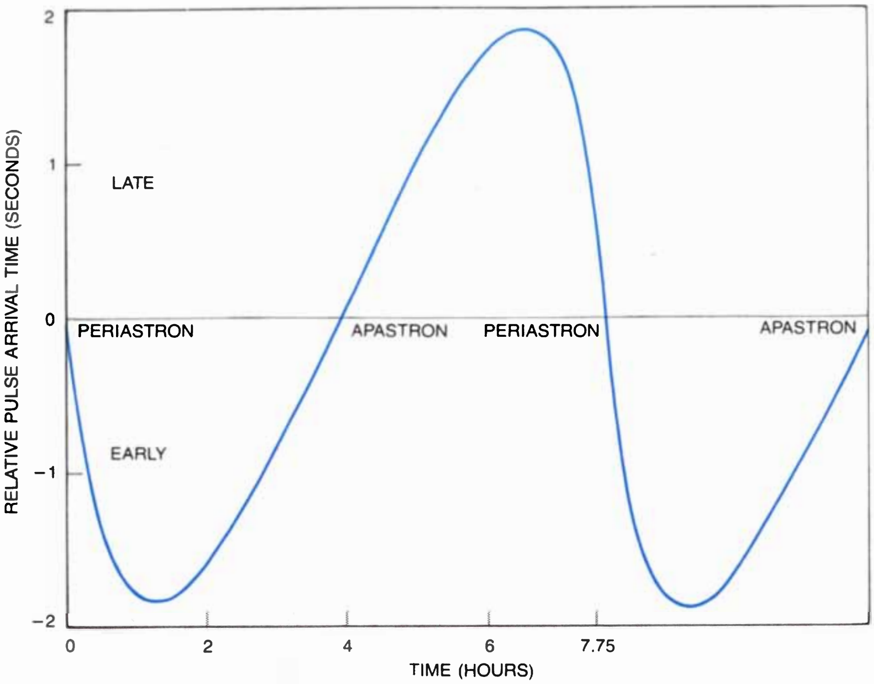
The standard techniques of binary-



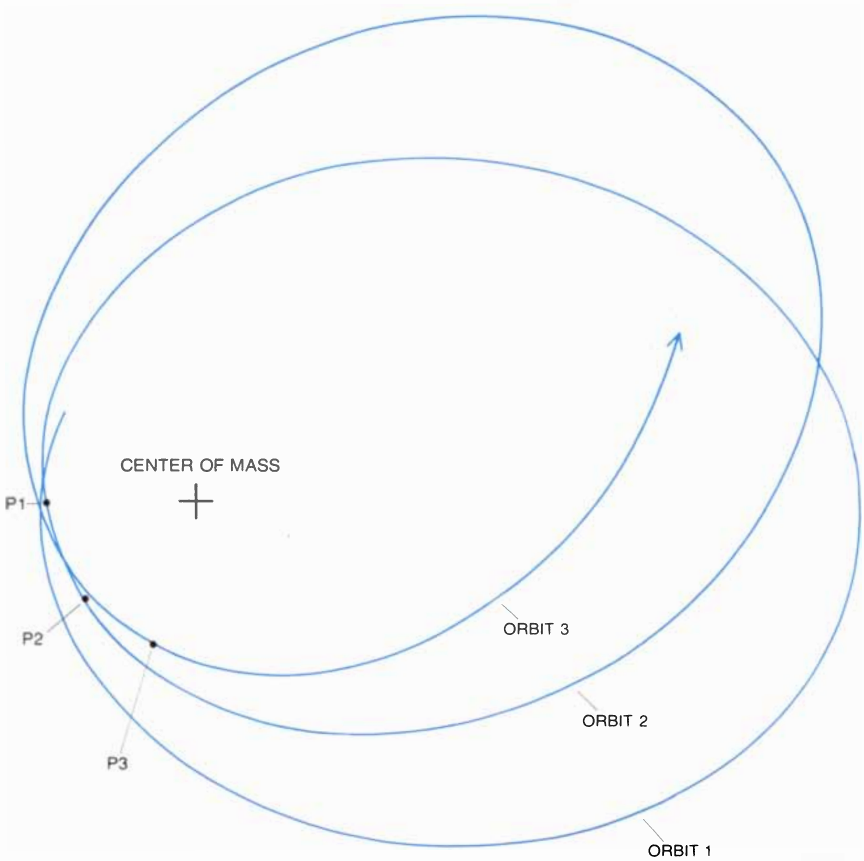
RADIAL VELOCITY OF PSR 1913 + 16 was calculated from changes in pulse-repetition frequency throughout the pulsar's orbital period of 7.75 hours by conventional Doppler-shift analysis. The frequency is lowest when the pulsar is moving away from the earth and is near apastron, the point where the co-orbiting bodies are farthest apart. The highest repetition frequency comes at periastron, when the bodies are closest together. The radial velocity is the component of the pulsar's velocity along the line of sight. Negative values indicate motion toward the earth. The existence of negative values larger than positive ones shows that the orbit is highly eccentric, with the pulsar gaining speed as it moves from apastron to periastron.



SHAPE OF THE PULSAR'S ORBIT, depicted in color, was determined from the radial-velocity curve. In this highly eccentric elliptical orbit the pulsar is about four times farther from its companion at apastron than it is at periastron. When the pulsar was discovered seven years ago, the major axis of its orbit was nearly perpendicular to the line of sight, as is shown here. Conventional Doppler analysis cannot determine several interesting parameters, such as the tilt of the orbital plane to the line of sight, the absolute size of the pulsar's orbit and the companion's orbit, and the masses of the two bodies. Knowledge of these parameters had to await an analysis on the basis of the general theory of relativity. The analysis revealed that the companion's orbit (*black ellipse*) is, to within an uncertainty of a few percent, the same size as the pulsar's. Major axis of the pulsar's orbit was found to be 4.5 times the radius of the sun.



TIME-DELAY CURVE depicts the variations in the arrival time of radio pulses from PSR 1913 + 16 as it travels in its orbit around its companion. When the pulsar is on the side of the orbit nearest the solar system, the pulses arrive more than three seconds earlier than they do when the pulsar is on the far side, indicating that the orbit is about a million kilometers across.



ADVANCE OF PERIASTRON in the orbit of PSR 1913 + 16 has provided one of the first clear observations of a general-relativistic effect involving bodies outside the solar system. The periastron advances, or rotates, as the elliptical orbit of PSR 1913 + 16 itself rotates in a plane because of the curvature of space-time in the vicinity of the pulsar's massive companion. In this diagram the effect is greatly exaggerated. The general theory of relativity predicts a periastron advance of about four degrees per year in the orbit of PSR 1913 + 16, the exact value depending on the total mass of the pulsar and its companion. The authors' measurements show the periastron is advancing 4.2 degrees per year, in good agreement with the prediction.

star analysis were therefore applied to the binary pulsar. The Doppler shifts were expressed as radial velocities (which are simply the components of the pulsar's orbital velocity that lie along the line of sight between the pulsar and the earth) and were plotted as a function of time. The radial velocity when the pulsar is approaching the earth, which is assigned a negative value, reaches a maximum of slightly more than 300 kilometers per second. The maximum radial velocity when the pulsar is receding from the earth is only about 75 kilometers per second. If the orbit were a perfect circle, the two extremes would be equal. Since they are not equal, one can conclude immediately that the orbit is highly elliptical and that the pulsar is approaching its companion and gaining speed in one segment of the ellipse and receding from its companion and losing speed in the opposite segment.

From the radial-velocity curve one can extract precise information such as the eccentricity of the orbital ellipse and its orientation in its plane. The calculations show that at apastron, the point of greatest separation, the two bodies are four times farther apart than they are at periastron, the point of closest approach. At the time of the initial observations periastron came near the time of maximum radial velocity when the pulsar was moving most directly toward the earth.

The standard method of radial-velocity analysis is not able to supply several interesting quantities such as the absolute size of the orbit, the tilt of the orbital plane with respect to the line of sight and the masses of the pulsar and its companion. For this information more accurate observations and a more sophisticated post-Newtonian analysis of the data were needed.

As a first step the absolute arrival times of pulsar pulses at the earth were carefully recorded; originally only the repetition frequency of the pulses had been measured. With this information the "phase" of the coherent train of pulses could be traced in addition to simply determining the pulse rate. The pulsar orbit could now be mapped more precisely than it could with the Doppler-shift data alone.

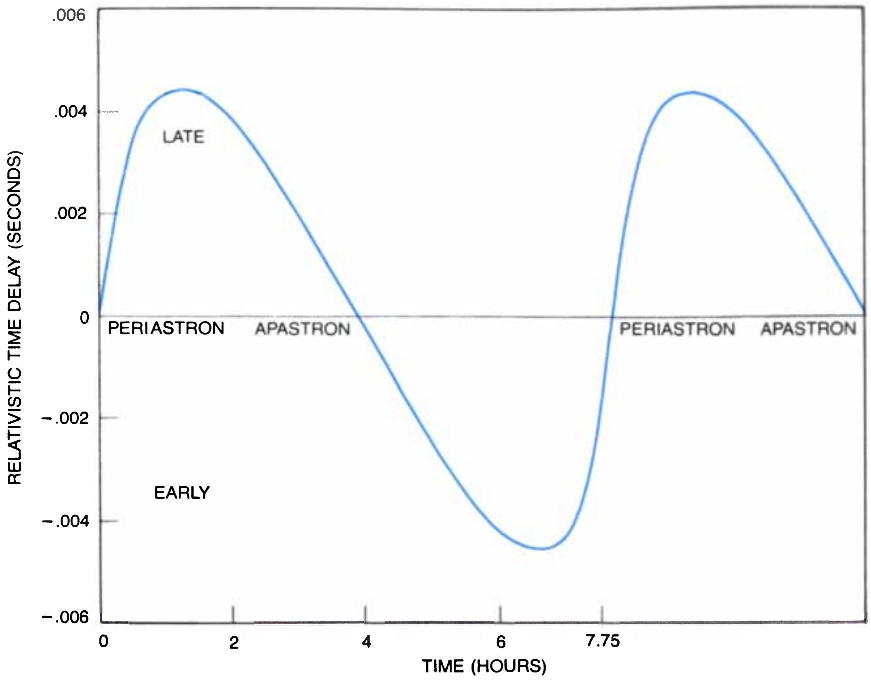
A plot of the time-delay curve, or variations in the arrival time of the pulses, shows that they arrive about three seconds earlier when the pulsar is on the near side of its orbit than they do when it is on the far side. Since radio waves travel at the speed of light (300,000 kilometers per second), the delay shows that the orbit is about a million kilometers in diameter. With the Arecibo telescope it is possible to measure the pulse arrival times to an accuracy of about 20 millionths of a second, thereby locating the pulsar in its orbit with a fractional accuracy of about one part in 150,000.

Let us now consider the subtler relativistic effects observable in the emission from the binary pulsar. The first measurable departure from Newtonian gravitation was an advance of the pulsar's periastron: a rotation in the orbital plane of the pulsar's point of closest approach to its companion analogous to the advance of the perihelion in Mercury's orbit. According to general relativity, the rotation should be about four degrees per year, an advance equivalent in one day to the advance of Mercury's perihelion in a century. There was some uncertainty because the magnitude of the effect depends on the masses of the pulsar and its companion, which were known only approximately at the time of the first calculation. Our first measured value of the advance of the periastron was 4.2 degrees per year. The success of this first attempt to utilize the binary pulsar as a gravitational laboratory encouraged us to carry out the other tests as well.

The next expected deviations from Newtonian theory have to do with the relativity of time. An observer's measurement of the time intervals marked by a moving clock depends on such factors as the velocity of the clock relative to the observer and the relative positions of the clock and the observer within a gravitational field. The orbital motion of the binary pulsar gives rise to two such phenomena: time dilation and the gravitational red shift. They combine to create an observable effect in the binary pulsar system known as the relativistic clock variation.

Time dilation is the apparent slowing of time measured on a clock that moves relative to the observer. Specifically, if an observer has two identical clocks and sends one of them off on a rocket moving at high speed while keeping the other clock with him, the rocket-borne clock will appear to measure the passage of less time than the stationary one. This idea is often expressed in the form of the "twin paradox": a man sent off on a spaceship traveling at nearly the speed of light discovers on his return that his twin who had remained on the earth has aged more than he has.

A clock running in a gravitational field of a given strength also appears to run slow when it is read by an observer in a weaker part of the field. This phenomenon is termed the gravitational red shift. The reality of these strange phenomena has been directly verified in a number of investigations. In one experiment Carroll O. Alley and his colleagues at the University of Maryland at College Park set up extremely accurate cesium-beam atomic clocks on aircraft and on the ground. After 15-hour flights the airborne clocks were found to be 47 billionths of a second ahead of the ground-based clocks. This result is in excellent agreement with the predictions of the general theory of relativity. According to the theory, the clock advance is explained by a combination of both kinds of relativistic clock variation: an advance of about 53 billionths of a second owing to the flying clocks' being in a weaker gravitational field because of their altitude (the gravitational red shift) and a retardation of about six billionths of a second owing to the airplane's speed (time dilation).



RELATIVISTIC CLOCK VARIATION has been observed in the arrival times of pulses from PSR 1913 + 16. According to the general theory, moving clocks and clocks in gravitational fields should appear to lose time when they are monitored by a distant stationary observer. The fourfold variation in the pulsar's orbital speed, combined with the highly elliptical orbit that carries the pulsar through a gravitational field with a strong gradient, provides a unique example of relativistic clock behavior. In accordance with the general theory the pulsar clock loses time while it is traveling fastest in the strongest part of the gravitational field of its companion. The maximum lag (compared with a hypothetical clock moving at constant speed and at constant distance from the companion) is a little more than .004 second. The pulsar clock then gains an equal amount as it travels more slowly in the weakest part of the field.

The speed of the binary pulsar varies by a factor of four as it moves in its elliptical orbit. The pulsar also passes through stronger and weaker regions of its companion's gravitational field as the distance between the two stars changes. As a result the observed pulse-repetition frequency of the pulsar clock varies from point to point in the pulsar's orbit owing to changes in both time dilation and the gravitational red shift. The two effects combine to advance or retard the pulses of the pulsar clock by as much as four thousandths of a second in different parts of the orbit.

As one can imagine, it is no easy task to distinguish the tiny relativistic variations in the pulsar clock from the much larger variations in pulse arrival time that result merely from the pulsar's changing distance from the earth. Nevertheless, in the time since the discovery of the pulsar its orbit has rotated sufficiently in its plane (owing to the relativistic advance of the periastron) to enable us to determine the magnitudes of the two kinds of variation separately. After six years of observations we have measured the amplitude of the relativistic clock variation to an accuracy of about 10 percent. As the orbital ellipse continues to change its orientation we expect to reduce the uncertainty in that measurement significantly.

The magnitude of the advance of the periastron and of the relativistic clock effects depends on the size and shape of the pulsar's orbit and on the masses of the two stars. Therefore by measuring these two relativistic effects we have obtained information that is useful in measuring the pulsar's orbit and in determining the total mass of the system. Indeed, the measurements of the two effects, along with the nonrelativistically determined orbital parameters, provide enough information to completely specify all the interesting orbital parameters as well as the masses of the pulsar and its companion. These calculations seem to be the first in which the general theory of relativity has been exploited as a tool for astrophysical measurement rather than simply a physical theory to be tested. We find that the orbital plane is tilted about 45 degrees from the line of sight, that the distance separating the two stars as they travel around each other varies from 1.1 to

4.8 times the radius of the sun and that both bodies have a mass about 1.4 times the mass of the sun.

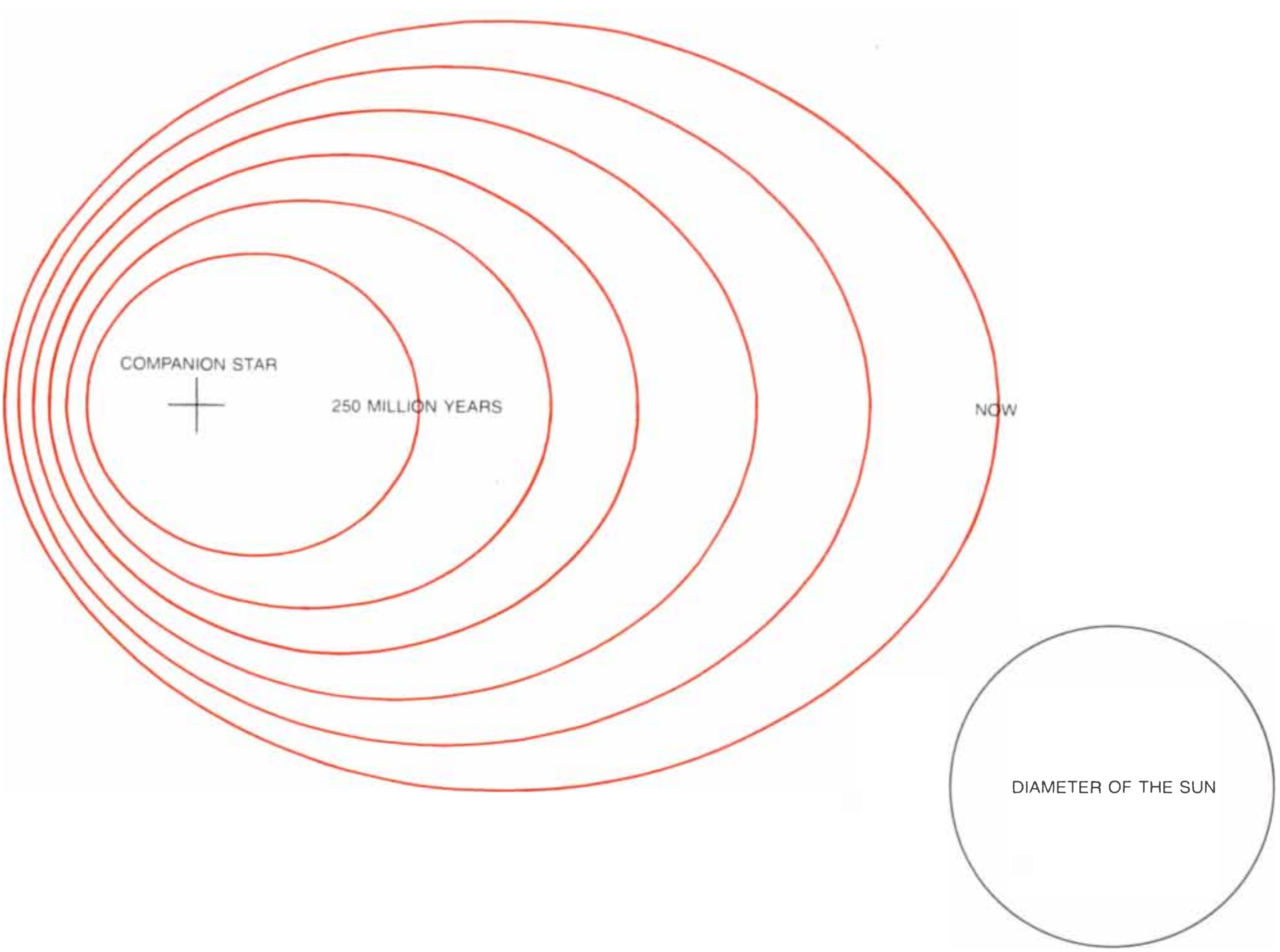
The binary pulsar is the first radio pulsar whose mass has been determined. The only ordinary stars whose masses have been measured are also members of multiple-star systems in which knowledge of gravitational laws can be combined with orbital measurements to yield mass values. The crucial difference, of course, between the mass measurements in the binary pulsar system and those in more typical binary star systems is that general relativity was invoked to provide some of the orbital information for the binary pulsar, whereas classical Newtonian laws suffice for ordinary binary stars. For example, measurements of the Doppler shifts of the spectral lines of two ordinary stars in orbit around each other (if both are visible) and observations of eclipses (for a system whose orbital plane is nearly edge on as it is seen from the earth) yield enough information to completely specify the orbital parameters and the masses of the stars.

The nature of the binary pulsar's silent companion can only be inferred; the object has not been observed directly. Our determinations of its mass and orbital parameters, however, coupled with stellar-evolution theory, have made possible some plausible guesses about its nature. The strongest constraint on the nature of the companion is that it must be small enough to fit inside the pulsar's orbit. This condition is satisfied by four known types of collapsed, dense stellar object.

The first and, according to stellar-evolution theory, most probable type of companion is another neutron star. Conceivably the companion is even another pulsar, one whose radio beacon does not happen to sweep across the earth. We shall return to this possibility when we describe a possible history of the binary system.

A second candidate is a black hole, a star whose gravitational collapse did not become stabilized at the dimensions of a neutron star but continued until all its matter was crushed to infinite density. It appears, however, that for a star to proceed to a state of unrestrained collapse at the end of its evolution it must be at least two or three times as massive as the sun. Since our calculations show that the companion has only about 1.4 times the mass of the sun, the companion is unlikely to be a black hole.

As a third possibility, the companion could be a white-dwarf star. Such an object is the remnant of a dying star, not more than 1.4 times as massive as the sun, that has collapsed to roughly the size of the earth. If the mass of the dying star is more than 1.4 times the mass of the sun, the collapse cannot be stabilized at the size of a white dwarf because the star's constituent atoms cannot resist being crushed by gravitational pressures to the next-smallest stable configuration, that of a neutron star. Since our estimate of the companion's mass is close to the instability point of 1.4 solar masses, a white dwarf cannot absolutely be ruled out, although present theories of stellar evolution indicate that it is unlikely for a system to form with a neutron star and a white dwarf

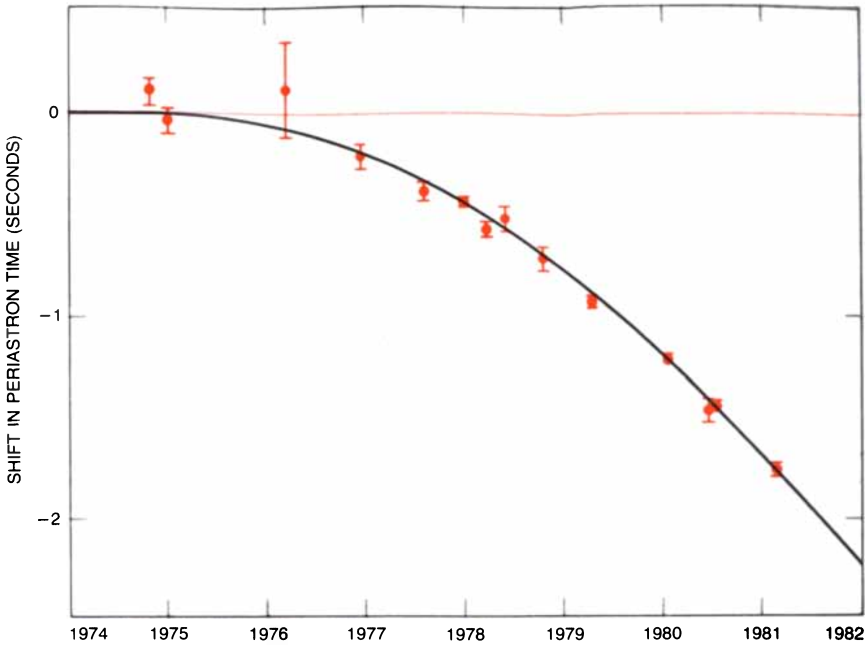


SHRINKING OF PULSAR'S ORBIT is projected on the basis of evidence that orbital energy is being converted into gravitational radiation, as is predicted by the general theory of relativity. According to the theory, the orbit of PSR 1913 + 16 should shrink by 3.1 millimeters per orbital revolution, or 3.5 meters per year. The orbital period should decrease accordingly by 6.7×10^{-8} second per orbit, or 7.6×10^{-5} second per year. This tiny change is measurable because it leads to a constantly accumulating deviation in the time of periastron passage. Here the orbit is drawn to scale as it will appear every 50 million years in the future until the two stars coalesce 300 million years from now. For comparison the sun is shown at the same scale. PSR 1913 + 16 is thought to be a 50,000th the diameter of the sun.

in such a tight orbital configuration. A white dwarf, like a neutron star, would be too faint to be observed with current optical telescopes at the estimated distance of the binary pulsar system: 15,000 light-years.

The fourth possible companion is a helium-core star. This rare kind of star consists of roughly one solar mass of matter compressed to a few tenths of a solar diameter. A helium-core star is the remnant, consisting mostly of helium, of a star that was originally much more massive. The helium, created by the thermonuclear fusion of hydrogen during the star's active lifetime, is left behind when the outer layers of the aging star expand tremendously and are either lost into space or are captured by the other member of a binary system. If the companion of the binary pulsar were a helium-core star, it might be visible with optical telescopes and might also give rise to measurable orbital effects. In fact, optical astronomers have recently detected a faint star close to the position of the binary pulsar that could be a helium-star companion. So far the only evidence linking the optical object with the binary pulsar system is the near coincidence in position. This coincidence may, of course, be accidental. Observations are continuing in an effort to clarify the nature of the optical object, which is near the limit of detectability for even the largest telescopes.

If the companion is indeed a helium star, its shape should be distorted by the pulsar's strong gravitational field. That distortion in turn could account for all or part of the observed advance in the pulsar's periastron, in the same way that an oblateness of the sun could theoretically account for part of the anomalous advance in the perihelion of Mercury. In addition the frictional energy loss associated with the distortion of a helium star might shorten the pulsar's orbital period, thereby mimicking the energy loss that we attribute to the emission of gravitational waves. Little is known, however, about the internal structure of helium stars, and estimates of the amount of change in orbital period that would be induced by a distorted helium star vary tremendously. It seems most unlikely that the dissipation of energy owing to a helium-star companion would just happen to equal the dissipation expected from gravitational radiation. We therefore consider it unlikely that the companion is a helium star.



EMISSION OF GRAVITATIONAL RADIATION BY PSR 1913 + 16 leads to an increasing deviation in the time of periastron passage compared with a hypothetical system whose orbital period remains constant. Solid curve corresponds to the deviation predicted by the general theory of relativity. Colored dots represent the deviation. The pulsar now reaches periastron more than a second earlier than it would if its period had remained constant since 1974. The data provide the strongest evidence now available for the existence of gravitational radiation.

Until 1979, PSR 1913 + 16 was the only binary pulsar known. Since then only two more binary pulsars have been discovered, so that such systems represent about 1 percent of the 330 pulsars now identified. It is estimated that approximately half of all the ordinary stars in our galaxy are members of binary systems (or systems consisting of more than two stars). Since pulsars are thought to represent a late stage in the evolution of ordinary stars in the range of masses from intermediate to large, one must explain why it is that roughly half of all pulsars are not found to be members of binary or other multiple-star systems. We shall first describe briefly what we consider a plausible history of the PSR 1913 + 16 system, one of the few systems that remained binary into the pulsar stage. Then we shall be able to comment on the general scarcity of binary pulsars.

We suggest that the system now observed as PSR 1913 + 16 began as a pair of ordinary stars, one star considerably more massive than the other and both together about 20 times as massive as the sun. The more massive of the two stars depleted its supply of nuclear fuel, chiefly hydrogen, much sooner than its companion did. With the depletion of fuel the outer layers of the more massive star expanded greatly and were captured by the less massive companion, leaving behind a helium-core star. After the star had consumed all its thermonuclear fuel it collapsed to an object a few tens of kilometers across and its outer layers rebounded in the tremendous explosion of a supernova. (The luminosity of such an explosion can rival that of the entire galaxy for a week or so.) Left behind as a remnant of the explosion was a spinning neutron star.

The companion star continued its evolution, eventually also expanding to the point where some of its matter was being pulled down onto the neutron star's surface, being heated in the process to temperatures so high that X rays were emitted. At that stage in the system's life it would have been observable as a binary X-ray pulsar, similar to those pulsars that have been observed by X-ray telescopes aboard artificial satellites in orbit around the earth. The outer layers of the companion continued to expand until they enveloped the neutron star. The friction of the neutron star plowing through the extended atmosphere of the companion was so great that orbital energy was rapidly dissipated and the orbit shrank drastically. As the companion's outer layers were heated by friction they were driven off into space, leaving behind a neutron star and a helium-core star circling each other in a very tight orbit. Finally the helium-core star exploded as a second supernova, leaving as a remnant a second neutron star.

We now observe one neutron star as a pulsar. The other neutron star may or may not emit radio waves; if it does, they apparently are not directed toward the solar system. The present highly eccentric orbit is evidence that the second explosion almost disrupted the system. Apparently most normal multiple-star systems are blown apart when one member explodes as a supernova, which explains the scarcity of pulsars in binary systems.

Let us now describe more fully the observational evidence that PSR 1913 + 16 is emitting gravitational radiation: the previously untestable prediction about accelerating masses made by

FIVE THOUSAND CONSECUTIVE PULSES from the pulsar are added together every five minutes, yielding a pulse profile such as the one shown here. The absolute time of arrival of the pulses at the Arecibo Ionospheric Observatory in Puerto Rico can be measured from the profile with an accuracy of 20 millionths of a second. The double peak suggests that the pulsar beam is a hollow cone; the peaks could be the two sides of the cone sweeping past the earth.

Einstein's general theory of relativity. Although gravitational radiation from the binary pulsar system is far too weak to be detected directly with present technology, the phenomenon should nonetheless be detectable indirectly.

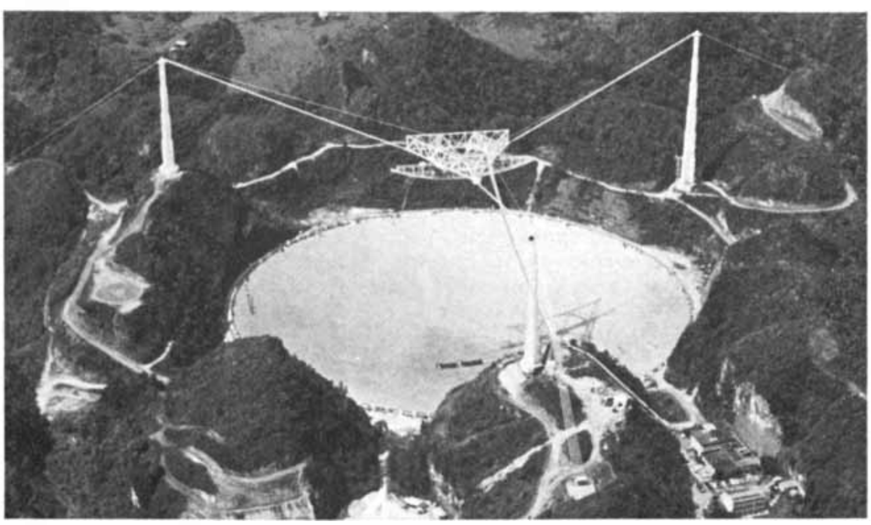
The source of energy for gravitational radiation from the binary pulsar is the energy of orbital motion. Therefore if gravitational waves exist and carry energy away from the binary system, the orbital energy should gradually decrease, causing the pulsar and its companion to spiral closer to each other and the orbital period to decrease. The orbit of an artificial earth satellite decays in the same way, although the satellite loses orbital energy not through the emission of gravitational radiation but through collisions with molecules in the upper atmosphere. (Strictly speaking, the satellite must also emit gravitational radiation, but it does so in negligible amounts.)

With our knowledge of the masses and orbital parameters of the binary pulsar system we can draw on the equations of general relativity to calculate the exact strength of expected gravitational radiation and hence the precise rate of contraction of the orbit and the decrease in the orbital period. We find that on each orbit the orbit should shrink by 3.1 millimeters and the orbital period should decrease by 6.7×10^{-8} second. There is no possibility of detecting the orbital shrinkage, amounting to 3.5 meters per year, even if the pulsar and its companion were as close to the earth as the sun and Mercury are and as readily observed. We can, however, measure the decrease in orbital period because it leads to an accumulating shift in the time of periastron passage compared with a hypothetical system whose orbital period remains constant. At the end of a year the pulsar should arrive at periastron .04 second early, and after six years it should arrive more than a second early. Thus the pulsar behaves like a poorly regulated clock that at first displays the correct time but begins to gain and then continues to do so at an ever faster rate. The cumulative error builds up quite rapidly: for the binary pulsar the shift should accumulate in proportion to the square of the elapsed time interval.

After six years of measurements we have found that the binary pulsar is indeed "gaining" in its orbit and that the acceleration is proceeding at almost precisely the rate predicted by general relativity [*see illustration on preceding page*]. This observation provides the strongest evidence now available for the existence of gravitational radiation. Moreover, the rate of decrease in the orbital period is not consistent with the predictions of several other modern gravitational theories that have been put forward as alternatives to Einstein's theory. We believe these alternative theories must now be rejected.

Hence 66 years after Einstein predicted the existence of gravitational waves an experiment has been done that yields clear evidence for their existence. Although the waves themselves remain elusive and undetected, their signature is plainly written in the orbital behavior of PSR 1913 + 16. A number of laboratory experiments designed to detect extraterrestrial gravitational waves are now in progress or being prepared, but even the best of them are not nearly sensitive enough to detect the waves from the binary pulsar. The experimenters seek instead to observe such cataclysmic (and rare) astronomical events as supernova explosions and the presumed formation of black holes.

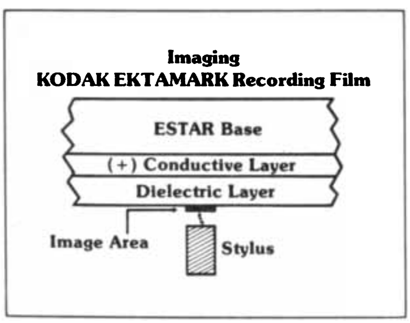
The reason for the difficulty of directly observing gravitational waves is straightforward: the interaction between the waves and any detector is so weak that it is overwhelmed by many possible contaminating effects, not least the unavoidable thermal motions of the atoms of the detector. Nevertheless, the binary-pulsar experiment should encourage the investigators who are developing gravitational-wave experiments. It now seems assured that what they are looking for does in fact exist.



INSTRUMENT USED IN DISCOVERY OF PSR 1913 + 16 is the 1,000-foot radio telescope at Arecibo. The pulsar system was found in 1974 in a survey conducted by Russell A. Hulse, then a graduate student at the University of Massachusetts at Amherst, and one of the authors (Taylor). The survey also turned up 39 other pulsars, none of them a binary. Of the 300-odd pulsars observed since the first one was discovered in 1967 only three are in binary systems.

Nice film without silver

Silver had better be saved for such ornaments of high civilization as elegant teaspoons, radiography of fractures, and the amplification of a light photon's kick by 10^9.



When energy with which to write in tiny dots can be supplied like this, you don't need that kind of amplification. Estar base is nicer than paper for carrying the conductive layer and the dielectric layer that receives the toner to visualize the electrostatic image. (No light sensitivity, no chemicals.)

In the great hunt for formations likely still to be holding hydrocarbon treasure, the geophysical industry's plotters have been drawing pictures something like the young lady is holding, but on dielectric recording paper. Paper gets pretty conductive when the humidity is up. That makes it hard to hold a fine-structured, crisp, electrostatic image. Estar base doesn't have that problem. And the film is quite transparent, not just translucent. Therefore that fine detail can be reproduced in the diazo machine without the services of somebody's copy camera.

On the chance that there are instrument people quite outside the geophysical industry who haven't heard about Kodak Ektamark recording film SO-222 and are feeling some excitement, Arnold Sullivan of our Graphics Markets Division is standing by with a fresh supply of exposed film. To obtain your samples, drop him a note at Kodak, Dept. GB-014, 343 State Street, Rochester, New York 14650.

Kodak

SCIENCE AND THE CITIZEN

National Titer of Antibodies

Among the innovations in medical technology that have contributed to lengthening the human life span, perhaps the most significant has been the practice of vaccination. It has been 185 years since Edward Jenner introduced a smallpox vaccine. (The word vaccine is derived from vaccinia, or cowpox, the active agent in Jenner's inoculant.) Since then smallpox has been permanently eradicated worldwide, and other vaccines have greatly reduced the incidence in the industrialized countries of diphtheria, measles, mumps, pertussis (whooping cough), poliomyelitis, rubella and tetanus.

The current status of immunization programs in the U.S. is evaluated in a recent review issued by the U.S. Public Health Service. The review, which is part of a report titled *Promoting Health/ Preventing Disease: Objectives for the Nation,* catalogues the achievements of immunization in the U.S. Diphtheria has been reduced from approximately 160,000 cases and 10,000 deaths annually in the early 1920's to 59 cases in 1979 and four deaths in 1978. (These are the most recent years for which the data are available.) Pertussis was diagnosed in some 200,000 cases and caused 5,000 deaths per year in the early 1930's; there were 1,617 cases in 1979 and six deaths in 1978. Some 21,000 cases of paralytic poliomyelitis were reported in 1952, an epidemic year for the disease; 26 new cases arose in 1979. Vaccines for measles, mumps and rubella have been introduced more recently, and the reduction in the incidence of these diseases is not yet as great. At various times in the 1960's there were 480,000 cases per year of measles, 152,000 of mumps and 60,000 of rubella. In 1979 between about 12,000 and about 14,000 cases of each disease were reported.

One of the most effective measures for ensuring an adequate level of immunity in the population has been the requirement that children show evidence of immunization before they can be admitted to school. Such legal requirements are now in force in all 50 states. One potential deficiency of relying on school enforcement is that it could leave a susceptible population of preschool children. The report therefore recommends extending the immunization requirement to "organized preschool settings," such as kindergartens and daycare centers. By 1990, it is suggested, at least 90 percent of all children should have completed a basic series of immunizations by the age of two.

Certain other legal and economic issues can have an influence on immunization practices. The report points out that as protection against disease has become more widespread "the rare adverse effects of immunization have become increasingly visible.... Liability associated with vaccines, and compensation of those injured as a result of immunization, have emerged as issues in the effective delivery of services." Economic measures that might encourage immunization include payment for vaccination as a Medicare benefit and supplying vaccine free to physicians if the physicians in turn will agree not to charge their patients for it.

The review sets several objectives for improving the health status of the U.S. population, which the Public Health Service suggests could be achieved by 1990 or sooner. The incidence of measles should be fewer than 500 cases per year, all of them imported by travelers. Mumps, rubella and pertussis should each be below 1,000 cases per year. Diphtheria and tetanus should each be reduced to fewer than 50 cases per year and poliomyelitis to fewer than 10.

A somewhat different account of immunization practices is presented by Albert B. Sabin of the Medical University of South Carolina, who developed the oral vaccine for poliomyelitis. Writing in *Journal of the American Medical Association,* he notes that "the almost complete elimination of paralytic poliomyelitis in the U.S." has been achieved even though "virulent polioviruses continue to be imported...and...a large proportion of economically disadvantaged children receive either no vaccine or not more than one dose." One explanation of the effectiveness of the immunization program, he suggests, is that "the oral polio vaccine strains spread from vaccinated to unvaccinated children and to others in the home and community and thereby immunize a significant proportion of unvaccinated persons."

Sabin thinks it should be possible to eradicate measles in the U.S. and in other developed countries. The strategy that would have to be adopted, however, is not the present one of routinely vaccinating each child at a specified age. What is needed instead is mass vaccination of all children within a short period of time in order to break the chain of transmission of the virus. Sabin is pessimistic about the possibility of greatly reducing the incidence of influenza by vaccination, since during an epidemic "the vaccines that are available contain the viruses of previous years and not the viruses that are causing the epidemic."

Sabin also discusses viral vaccines that are under investigation for certain diarrheal diseases of infancy, for hepatitis and for herpes zoster, the virus that causes both varicella, or chickenpox, in childhood and shingles in later life. He is optimistic about the prospects for a live-virus vaccine for varicella developed in Japan. He thinks, however, "the effort to confirm the excellent results reported by the Japanese investigators has been too small, and too slow and uncoordinated when one considers how much misery could be prevented by the eradication of this disease."

The Neutrino Shortage

Neutrinos, the elementary particles with no electric charge and little or no mass, radiate from the sun in such a copious flux that some 10^{14} neutrinos pass through the human body every second. The neutrinos are generated deep within the sun by the thermonuclear reactions that transmute hydrogen into helium. Indeed, the neutrinos may offer the only direct information about the reactions in the sun's core. Up to now, however, the information obtained has been ambiguous. Since 1968 a detector has been counting the highest-energy solar neutrinos and has consistently recorded fewer than were expected. Now certain theoretical innovations suggest new ways of interpreting the discrepancy, and experiments under consideration promise to resolve the problem.

The detector that has been operating since 1968 was designed by Raymond Davis, Jr., and his colleagues at the Brookhaven National Laboratory. It is a tank filled with 615 tons of the cleaning fluid tetrachloroethylene, a substance chosen because it is rich in chlorine. About 25 percent of all natural chlorine consists of the isotope chlorine 37, and it is this isotope that serves as the neutrino-sensitive medium of the detector. An atom of chlorine 37 can capture an energetic neutrino, whereupon the atom is transformed into an atom of the radioactive isotope argon 37. The argon is isolated by flushing it out of the tank with helium gas, and the transformed atoms are counted by detecting the electron that is emitted when an argon-37 atom decays. The entire apparatus is installed almost a mile underground in a mine in South Dakota.

The great mass of the detector is needed because even in the deluge of neutrinos from the sun a neutrino is only rarely captured. The detector includes some 10^{30} atoms of chlorine 37, and yet only two or three atoms per week are converted into argon 37. Calculations made by John N. Bahcall of the Institute for Advanced Study, based on the standard model of the sun's internal structure, indicate that the flux of solar neutrinos should be great enough to cause four times as many events.

Many proposed explanations of the missing neutrinos have cited possible


One Stamford Forum, Stamford, Conn. 06904


"Solar powered telephone calls? GEE!"
(No, GTE!)
What do a remote Navajo outpost, the Colombian jungle and a snow-capped mountain in Alaska have in common?
They're all in the middle of nowhere. Virtually impossible to run telephone cable through.
However, we've found a way to provide phone service to areas like these. Using an innovative device called a "solar repeater."
Since the repeater gets its power from the sun, no power lines are necessary.
How does it work? By relaying telephone calls, via radio wave, from one repeater to another and eventually to their destinations.
These solar repeaters, which can each handle six hundred conversations at once, can be flown by helicopter into the most remote locations.
And presto! Just like that, people in the middle of nowhere can speak to people anywhere.
GTE

flaws in the standard solar model; some 30 modifications of the model have been suggested in the past decade. For example, the theoretical flux might be made to match the observed flux if the abundance of heavy elements in the core of the sun were assumed to be less than it is in the standard model, or if the temperature were assumed to be lower. It has even been suggested that the discrepancy might be accounted for if the sun were now "coasting," without ongoing nuclear fusion, or if there were a black hole at its center.

More recently attention has shifted from the complexities of neutrino emission in the sun to the properties of the neutrinos themselves. Neutrinos are known to come in three flavors, or types; one flavor is associated with the electron, one with the muon and one with the tau particle (a massive cousin of the electron and the muon). It once seemed that the flavor of a neutrino could never be changed; an electron-type neutrino was destined always to remain an electron-type neutrino. This assumption has become less secure with the emergence of unified theories of the interactions of elementary particles. The theories suggest that the identity of a neutrino may oscillate, or change continuously.

The highest-energy solar neutrinos, which are the ones detected in Davis' experiment, are emitted as electron-type neutrinos. During one step in the solar synthesis of helium, boron 8 is created and quickly decays by the mechanism called beta decay. A proton in the boron-8 nucleus is converted into a neutron with the release of a positron (or antielectron) and an electron-type neutrino. The neutrino-induced reaction in Davis' detector is inverse beta decay: a neutron in chlorine 37 absorbs an electron-type neutrino and is thereby converted into a proton in argon 37. At the same time an electron is emitted.

The fact that the inverse beta decay of chlorine 37 can be induced only by an electron-type neutrino suggests a possible explanation of the solar-neutrino deficit. The flavor of an electron-type neutrino could change during its transit from the sun to the earth, so that it would arrive at the detector as a muon-type neutrino or a tau-type neutrino, which would not be counted. If the three known neutrino flavors oscillate maximally, the observed flux at Davis' detector should be one-third the flux predicted in the absence of oscillations. If there is a fourth neutrino flavor, a possibility that is not excluded by theory, the flux might be reduced to one-fourth the currently predicted value, which would be consistent with Davis' observations.

The true magnitude of the solar-neutrino flux may ultimately be determined only through the construction of more sensitive detectors. The drawback of Davis' experiment is that the high-energy neutrinos it primarily aims to detect constitute less than a ten-thousandth of all solar neutrinos. Only neutrinos with an energy of more than 814 kiloelectron volts (keV) can convert chlorine 37 into argon 37. What is needed is a detector with a lower threshold, sensitive to the less energetic neutrinos generated more prolifically in the solar reactions. The construction of such a detector has been proposed by a consortium of investigators from Brookhaven, the University of Pennsylvania, the Institute for Advanced Study, the Max Planck Institute for Nuclear Physics in Heidelberg and the Weizmann Institute of Science in Israel. The active medium of the detector would be 50 tons of the isotope gallium 71. The gallium would be employed as a solution of gallium chloride in order to benefit from Davis' experience with a liquid target medium.

The properties of the gallium nucleus make it a good candidate for a neutrino detector. The capture of a neutrino by gallium 71 is well understood. As in chlorine 37, the process is an inverse beta decay: a neutron absorbs a neutrino, emits an electron and thereby becomes a proton, converting gallium 71 into the radioactive isotope germanium 71. The advantage of gallium is that the neutrino energy needed to trigger the conversion is only 236 keV. As a result a gallium detector would be sensitive to virtually all solar neutrinos. It is expected that a 50-ton gallium detector would record one solar neutrino per day, or about two and a half times the event rate of the chlorine detector.

The economics of the gallium detector seem more problematical than the physics. The Department of Energy, which is considering funding the project, estimates that 25 tons of gallium is produced annually in the non-Communist countries, chiefly from aluminum ores; the detector calls for twice the annual tonnage. In acquiring the gallium the project would compete with the semiconductor industry, where gallium arsenide is a substrate for electronic devices. The gallium would cost roughly $25 million. Although it would not be degraded in the detector and could be resold, its price could fluctuate widely.

One stage of the gallium-detector project has already been completed. For the past year the investigators have been extracting germanium from 4,000 liters of a gallium-chloride solution in which the gallium content is 1.3 tons. The next stage is to measure the rate at which neutrinos convert gallium 71 into germanium 71. The plan is to bring 12 tons of gallium into contact with 70 kilograms of the highly radioactive isotope chromium 51, which emits neutrinos at a known rate.

The Department of Energy has appointed a committee of physicists, headed by Glenn T. Seaborg of the Lawrence Berkeley Laboratory, to evaluate the merits of the plan to build a gallium detector. The committee is expected to issue a favorable report. If the Department of Energy decides to fund the full-scale project, money could be appropriated for it as early as fiscal year 1983. The Max Planck Institute for Nuclear Physics has agreed to buy 25 percent of the gallium for the project.

Work along similar lines is under way in the U.S.S.R., where an underground neutrino laboratory is being constructed in the Caucasus. The plan calls for a 50-ton gallium detector and for a chlorine detector five times the size of the one in the U.S.

A limitation of both chlorine and gallium detectors is that they register neutrinos only by counting them, without measuring any of their properties. A detector based on the isotope indium 115, proposed by Ramaswamy Raghavan of Bell Laboratories and Martin Deutsch of the Massachusetts Institute of Technology, could measure the energy of a neutrino and perhaps even its direction. The observation of the direction might determine whether a detected neutrino actually came from the sun and hence might eliminate some extraneous, background events.

The neutrino-induced reaction in an indium detector would also be an inverse beta decay. When a nucleus of indium 115 captures a neutrino, a neutron in the nucleus becomes a proton and so the indium 115 is transmuted into tin 115; simultaneously an electron is released. The neutrino needs to have an energy of only 128 keV to trigger the reaction. The electron carries away any energy above the threshold value, and so the energy of the neutrino can in principle be determined from the energy of the electron.

The chief difficulty in constructing an indium detector is a technological one. The electron whose energy is to be measured often has an energy of only a few tens of thousands of electron volts, which is below the range of energies in which most electron counters are sensitive. Raghavan and Deutsch have developed a new electron counter based on a hollow quartz fiber filled with a liquid scintillator. The counter will be tested soon. If it functions as expected, work can proceed on the full-scale detector, which will consist of four tons of indium and will cost roughly $2 million.

Dinner at the Rift

The discovery of dense populations of animals living on the sea floor near vents of hot water was a considerable surprise. Food is scarce deep in the ocean, largely because photosynthesis, the ultimate source of food in every other ecosystem known, is impossible in the darkness that prevails at a depth of some 2,500 meters. Nevertheless, colonies have been found at widely separated sites, all of them associated with hydrothermal vents along mid-ocean ridges, where plates of the earth's crust are separating and new ocean floor is forming. It was suspected from the outset that the vents supply a source of nutrition for the organisms, but it was not immediately apparent how.

For the bivalve mollusks in the populations near the vents the source of nourishment was soon understood. The water issuing from the vents is rich in hydrogen sulfide, which certain types of bacteria oxidize to yield sulfate compounds and elemental sulfur. By this means the bacteria obtain the energy needed to liberate carbon atoms from carbon dioxide and manufacture organic molecules. The bivalves are filter feeders that ingest the bacteria.

The oddest organisms in the populations near the vents are giant worms, up to 1.5 meters long and 3.8 centimeters in diameter, with a flowerlike plume at one end. Identifying their food supply has proved to be much more difficult than it was for the mollusks. Indeed, before one could address the question of what the worms eat it was necessary to determine how they eat: they lack a mouth, a gut and an anus.

The worms cannot ingest bacteria. Moreover, it appears they cannot absorb organic molecules directly from seawater fast enough for sustenance. Laboratory experiments imply that the rate of such absorption is insufficient even for smaller worms. How then do the worms near the vents get their food? The work of a number of investigators now centers on the trophosome, a tissue composing more than half of the body of the worm. It is proposed that the trophosome is actually a colony of symbiotic bacteria. Like the free-living bacteria at the vents, the symbiotic ones are chemoautotrophic: they sustain themselves on inorganic substances, and thereby they sustain the worm.

Several lines of evidence combine in suggesting that the trophosome is the source of food for the worm. The evidence begins with measurements of the relative abundance of the two stable isotopes of carbon (carbon 12 and carbon 13) in tissues from various marine organisms. The ratio of the isotopes in filter-feeding clams and mussels near the vents differs from the ratio in animals living elsewhere in the sea. Each metabolic pathway by which carbon from carbon dioxide is incorporated into organic matter gives rise to a slightly different ratio of carbon 12 to carbon 13. Hence the observation of different isotopic ratios in two organisms suggests they have different sources of organic molecules. The ratio in tissue from the worms near the vents is reported by Greg H. Rau of the University of California at Los Angeles to be different from the ratio in the clams and mussels with which the worms share their habitat. The discrepancy could be explained if the mollusks and the worms are nourished by different kinds of bacteria, and so Rau proposes "that more than one type of autotrophy exists at the vents."

A second line of evidence comes from chemical assays of specimens of the trophosome done by Horst Felbeck of the Scripps Institution of Oceanography: the assays reveal the activity of five enzymes. Three of the enzymes chemically reduce sulfur compounds and store the energy released in molecules of adenosine triphosphate (ATP), which serves for energy storage in a multitude of organisms. The other two enzymes have previously been found in plants, where they take part in a biochemical cycle that exploits the energy of photosynthesis for the manufacture of organic molecules. Frequently the trophosome contains particles of sulfur, a fact that attracted Felbeck's attention. Electron microscopy of the tissue of the trophosome shows that it consists of densely packed spherical cells, which resemble bacteria in that they have no nucleus.

The third finding in support of the hypothetical symbiosis was made by Alissa J. Arp and James J. Childress of the University of California at Santa Barbara. They found that the hemoglobin in the blood of the worms has a higher affinity for oxygen than that of most invertebrate animals. The affinity depends only moderately on the ambient temperature and is almost unaffected by the ambient concentration of carbon dioxide. If the worms are to support an internal colony of bacteria, the worms must supply the bacteria with both oxygen (for reducing sulfur compounds) and carbon dioxide (as a source of carbon for organic molecules). The worms' hemoglobin seems to be well suited to carrying out both tasks in a place where the temperature varies erratically between 22 degrees Celsius, the temperature of the water issuing from the vents, and 2 degrees, the temperature of the surrounding seawater.

The extensive findings on the physiology of the worms are reported in *Science* in a series of articles by 12 investigators from eight institutions, including Meredith L. Jones of the Smithsonian Institution, who named the worm *Riftia pachyptila.* What now remains to be determined is how a colony of autotrophic bacteria becomes established in a newborn worm whose body surface is closed. Electron microscopy of the eggs of the female worm show that no such bacteria are inside them.

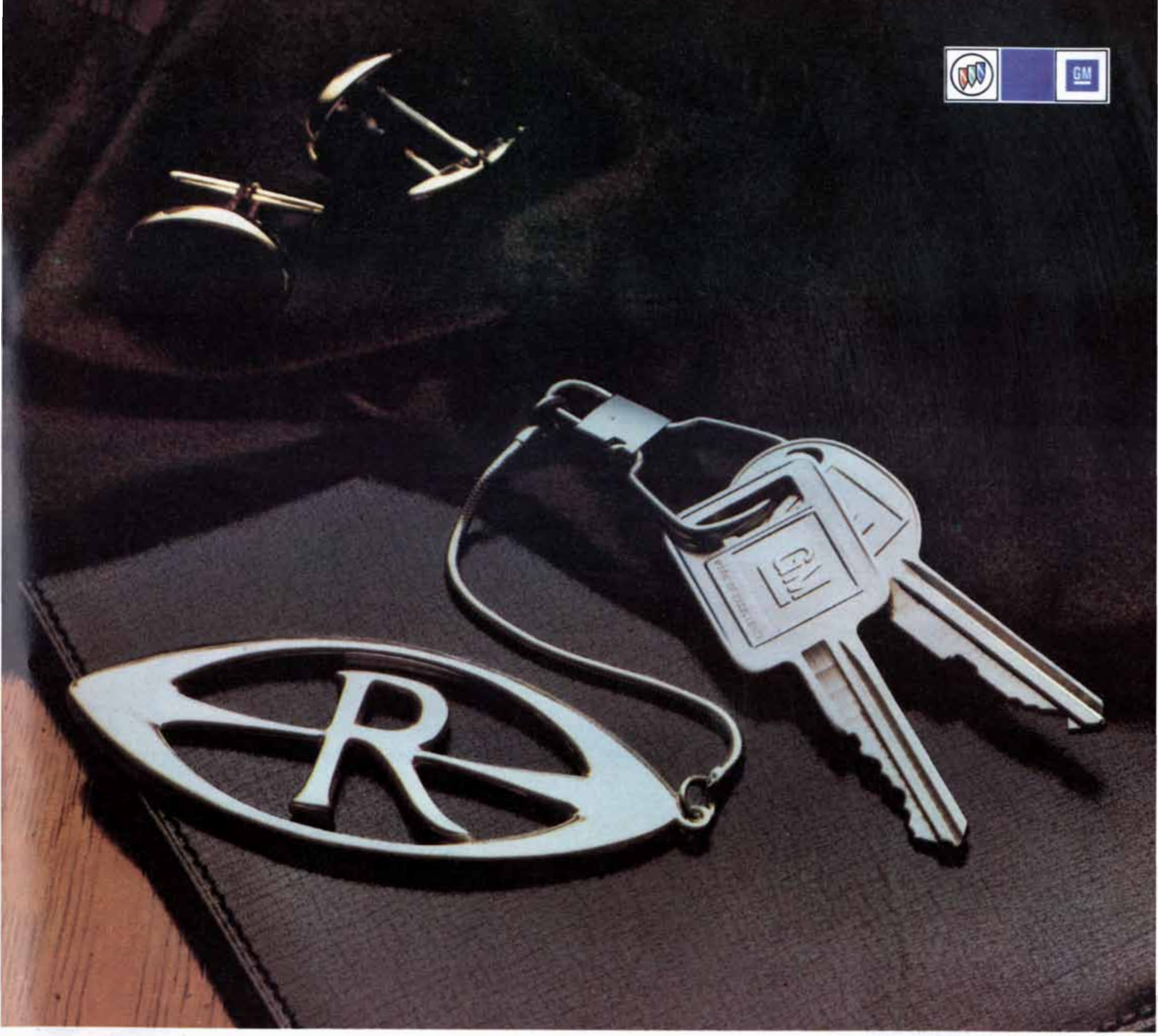
Xaser Beam

Devices that emit coherent electromagnetic radiation have been extended to ever higher frequencies in the past few decades. The first such devices, the masers, operated at microwave frequencies. In 1960 laser light, or coherent radiation at optical frequencies, was generated for the first time. Because higher-frequency radiation carries higher energy there have been efforts recently to stimulate coherent radiation in the far-ultraviolet and X-ray regions of the electromagnetic spectrum. Workers at the University of Hull in England have now achieved some success in the far-ultraviolet region, and there have been unconfirmed reports that a group at the Lawrence Livermore National Laboratory has built a xaser, a device capable of producing coherent radiation at X-ray frequencies.

Radiation is temporally coherent when all its waves are in phase, so that the crests coincide with crests and the troughs with troughs. An instrument capable of generating coherent radiation at X-ray frequencies could profoundly alter the scientific understanding of the structure of matter, particularly the structure of biological macromolecules. The standard technique for determining the structure of such a molecule is to pass a beam of X rays through a crystallized specimen of the substance and to record the interference pattern formed as the beam is diffracted by the atoms in the crystal. For the structure to be resolved in atomic detail the radiation in the beam must have a wavelength not much longer than the diameter of an atom; it is for this reason that X rays are employed.

In ordinary X-ray diffraction the phase of the incident waves is random and the information carried by the phase of the diffracted waves is lost. The diffraction pattern that results can be likened to a shadow of the atoms cast on a two-dimensional screen. The three-dimensional structure of the molecule must be guessed from the shadow, which is often highly ambiguous. Reconstructing the third spatial dimension therefore calls for great ingenuity; moreover, the ambiguities must be resolved anew for every molecule investigated. Information about the phase of the diffracted X rays could make the reconstruction a routine matter.

According to George Chapline and Lowell Wood of the Lawrence Livermore National Laboratory, an X-ray laser, like an ordinary laser, could be employed to record a hologram: a pattern created when radiation reflected from an object is allowed to interfere with a reference beam that has not been reflected. The X-ray hologram could then be illuminated by light from a laser operating at a visible wavelength and viewed through a microscope. The three-dimensional form of the object would be seen, magnified by a factor equal to the ratio of the optical wavelength to the X-ray wavelength. Biological structures such as DNA in the cell nucleus could thus be observed in three



For the car that has everything.

Buick Riviera. You say "everything" seems a bit of an overstatement? Perhaps.... but.. you should inspect a new Riviera for yourself. And see how close it comes.

BUICK

Wouldn't you really rather have a Buick?

dimensions. In principle a holographic motion picture could be made of DNA in the process of replication or of proteins being synthesized by the ribosomes of the cell.

The impediment to such applications is that coherent X-radiation is extraordinarily difficult to produce. In a maser or a laser a population of atoms is "pumped" into an excited state; each atom can then fall to a lower-energy state by emitting a photon, or quantum of electromagnetic radiation. The photon can be emitted spontaneously, but coherent radiation is observed only when the emission is stimulated by an incident photon of the same wavelength as the emitted one. In the maser and the laser stimulated emission is ensured by an arrangement of mirrors, which reflect the photons so that they pass through the active medium many times.

For an atom to emit an X-ray photon it must first be raised to an excited state of very high energy. The lifetime of such a state is brief, however. Spontaneous X-ray emission takes place within about 10^{-15} second; during that interval a photon moves less than a micrometer. Moreover, no efficient X-ray mirror is known, so that the multiplication of photons must take place during a single transit of the initial photon. For these reasons the population of excited atoms must be quite dense. An extreme concentration of energy is needed to maintain such a population.

To pump a far-ultraviolet laser to the necessary energy density the group at the University of Hull employs a maser that emits a strong flux of microwave radiation. The microwave pulse vaporizes a fiber of carbon, and some of the carbon atoms are stripped of all but one of their electrons. As the plasma of carbon ions expands, the ions recapture electrons and emit ultraviolet photons at a wavelength of 18.2 nanometers. Some amplification of the ultraviolet photons along the axis of the fiber has been detected, but the investigators do not yet claim to have succeeded in stimulating true laser emission.

The reports concerning the Lawrence Livermore National Laboratory project (which have not been confirmed by the laboratory) describe an X-ray laser pumped by the energy released from a small nuclear explosion. Apart from the question of whether or not such a pumping method has actually been demonstrated to work, there remains the question of its cost, practicality and propriety in scientific applications. Of course, the intended applications may be military rather than scientific.

Menarcheal Misunderstanding

A demographic finding that has lately seemed to be gaining the status of folk wisdom holds that children have been reaching sexual maturity at a progressively earlier age for more than a century. The apparent trend is based on historical estimates of the average age at menarche, or first menstruation. It has occasionally been cited in explanation of a supposed increase in the sexual activity of adolescents. Now the consistency of the trend itself is being called into question. It has been pointed out that individual variation in the age of puberty is so great that a trend attributed to the human population in general may have little meaning.

Many of the data supporting the earlier-maturation hypothesis were compiled by J. M. Tanner of the Institute of Child Health of the University of London. A graph published by Tanner gives the age at menarche as 17.1 years in 1840 and 12.5 years in the 1960's. A simple interpretation of the graph would be that the average age at menarche has decreased by from three to four months per decade since the mid-19th century.

Tanner estimates that the data given in the graph are "accurate to within 1.1 years" for the populations studied. Not all interpretations of Tanner's findings, however, have been confined to the populations represented in the data. Those populations are drawn exclusively from northern European countries and the U.S. Tanner did not intend the graph to show the rate at which the world menarcheal age has declined but rather to give an indication of general trends in some European and American populations. Indeed, he points out that there is much more information available on the age of menarche, but he adds that "other European and American data, though not quite so regular, agree well with these figures."

Even the reliability of inferences applied to the populations included in the data has now been questioned by Vern L. Bullough of the New York State University College at Buffalo. Writing in *Science,* Bullough points out that 19th-century methods of data collection were highly subjective. He later added that "even modern data collection is not so accurate."

The earliest datum on the graph, and the only one to give an average age of more than 17 years, is from Norway in 1840. According to Bullough, this age is based on the report of a single physician. Most of the other pre-20th-century data are similarly based on individual observations. Moreover, almost all the 19th-century data come from the Scandinavian countries, largely because more complete health records were kept there. On the average Scandinavians mature later than Americans and therefore 19th-century Scandinavian data and modern American data cannot be compared on the same basis.

Bullough also notes that 1840 was a famine year in Norway. It has been established that menarche is delayed by poor nutrition, and so the 1840 Norwegian average of 17 years is likely to be skewed. Bullough cites other data from Germany, Scotland and England that give a substantially lower average menarcheal age of from 14 to 15 years for the same period.

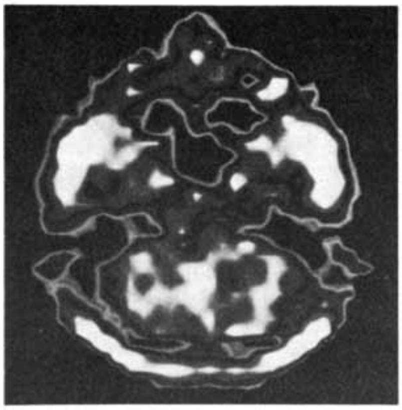
At the modern end of the graph the statistical reliability of a part of the sample is also dubious. All the data on the graph for the U.S. in the 1950's and 1960's come from a sample of women students at the University of North Carolina. The sample may not be representative of the population as a whole. Furthermore, the results of such a survey could be distorted by errors of recall, since most college-age women are well beyond the age of menarche.

Another impediment to the interpretation of historical records of menarcheal age has been pointed out by Rose E. Frisch of the Center for Population Studies of Harvard University. The difficulty is that the data are not usually correlated with environmental factors, such as level of nutrition, to which menarcheal age is sensitive. For example, Frisch has shown that for every year a young female athlete is in training (doing strenuous physical exercise for more than an hour a day) her first menstruation is delayed by five months. Stress has also been linked to delayed menarche, whereas greater consumption of foods high in fat may cause earlier menarche. Frisch concludes that age at menarche is so dependent on the environment that it is of little use in characterizing a population but can serve as a sensitive indicator of an individual's state of health and growth.

Cold Tooth, Hot Tooth

Among the processes of decay and degradation that commence with the death of an organism, one distinctive chemical transformation provides a means of assaying the age of biological materials. Every amino acid has two stereoisomers, or mirror-image forms, designated L (for levo, or left-handed) and D (for dextro, or right-handed). In life only the L forms are incorporated into proteins, but after death the isomers are randomly interconverted. If the random exchanges are allowed to continue long enough, the L and D amino acids come to be present in equal quantities. The process is called racemization, a name taken from racemic acid, a form of tartaric acid made up of L and D isomers in equal proportions.

The speed of amino acid racemization depends mainly on two factors: the chemical nature of the amino acid in question and the temperature of the postmortem environment. The rate is measured in terms of the racemization half-life; because the interconversion proceeds in both directions at once the half-life is defined as the time required



This November, you'll learn about new products and processes that could make some people very rich.

Here's one. It's called nuclear magnetic resonance tomography. It's the basis of a non-invasive imaging device that can actually "see" inside everything from the human body to plant machinery. Next year it will make headlines. But you can learn all about it *now* in TECHNOLOGY magazine.

NUCLEAR MAGNETIC RESONANCE IMAGE OF HUMAN HEAD.

ANNOUNCING TECHNOLOGY MAGAZINE

(You can sample the first issue FREE)

Technology. It's probably the toughest free enterprise arena in the world. You either win or lose . . . seize opportunities or face obsolescence.

And, to stay competitive, you've got to stay on top of all the new developments. Know when to commit to a new process or application . . . and when you'd be smarter to lay back and learn from someone else's mistakes.

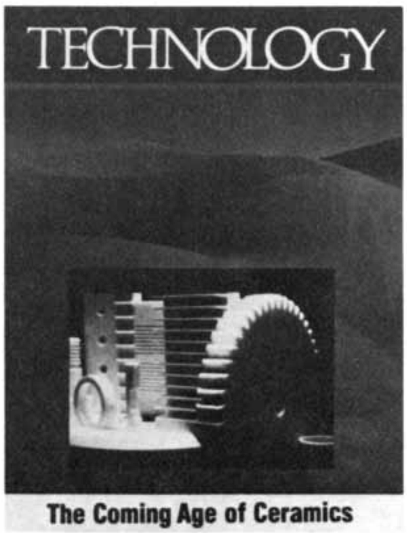
And now there's a magazine that gives you, all in one place, the vital intelligence you need to make those decisions.

TECHNOLOGY. A new kind of magazine that lets you peek into R & D labs around the world. So you can learn what's working. What's not. And what's the coming state-of-the-art in everything from ceramics that can replace metals and plastics to enzymes that could provide a bio-industrial alternative to high priced petroleum.

You won't read TECHNOLOGY. You'll *use* it!

While TECHNOLOGY will be edited to deliver the most information in the most accessible style — and will be a joy to read — it is not intended for the hobbyist or the merely curious.

It will be a *tool* to be *used* by people who have a tangible stake in technology. Each major article will appear in a unique 5-part format consisting of a general interest overview, a business outlook report, a core article for professionals and specialists, an applications side bar, and



an extensive bibliography. You'll see quickly and precisely . . .

- How a specific technology originated and in response to what needs
- How it works
- Its potential uses and limitations
- Who (and where) the experts are
- What applications are practical right now and what they cost
- How it fits or connects with other techniques, methods, processes and systems

To find out everything you need, you could comb through the dozens of magazines, trade journals, and newsletters on the subject of technology — at considerable time and expense.

Or you can subscribe to TECHNOLOGY and hook into one of the most extensive networks of information and technological expertise ever assembled. So that whether you are a technology executive, a specialist seeking new breakthroughs, an entrepreneur or investor . . . TECHNOLOGY will bring you the information you need — ahead of the crowd. So you can act on it profitably.

Get TECHNOLOGY's first issue FREE!

TECHNOLOGY will be available only by subscription — it will not be sold on newsstands.

But you can receive the premiere issue FREE simply by returning the accompanying card or the coupon below.

We'll send you Vol. 1, No. 1 of TECHNOLOGY for a free examination . . . and reserve a Charter Subscription in your name at the special rate of only $18 for 1 year (that's 6 bi-monthly issues at a saving of $6 off the regular yearly rate). If, after evaluating your first issue, you don't wish to continue, just write "cancel" on your bill and mail it back. You'll owe nothing and may keep TECHNOLOGY's first issue with our compliments.

Return the coupon below or the accompanying card to receive TECHNOLOGY's first issue FREE!

Send me the Premiere Issue of TECHNOLOGY

FREE

And enter my 1-year Charter Subscription (6 bi-monthly issues) for only $18 . . . a $6 saving off the regular subscription rate. *(TECHNOLOGY will not be sold on newsstands.)* I understand that if I'm not satisfied with my first issue, I need only write "cancel" on my bill. I may keep Vol. 1, No. 1 of TECHNOLOGY and I'll owe nothing. Otherwise, I'll return your invoice with payment and receive 5 more issues.

Name____________________
(PLEASE PRINT)

Address ____________________

City ____________________

State __________ Zip________

MAIL TO: TECHNOLOGY Magazine, P.O. Box 2528, Boulder, CO 80321

5101

to attain a D-to-L ratio of .33. The half-life is a few thousand years for aspartic acid at 25 degrees Celsius, but it is many millions of years for isoleucine at zero degrees C. The sensitivity of the process to changes in the ambient temperature adds to the difficulty of deducing the age of a biological material from the extent of racemization.

Racemization actually begins as soon as a protein is synthesized, but most protein molecules are replaced continually, so that no significant degree of racemization can be detected in living tissues. A few proteins, however, are not regenerated throughout life. In mammals the commonest such stable proteins are found in the enamel and dentine of teeth.

Perhaps the most remarkable of all mammalian teeth is the "tusk" of the male narwhal. The tusk is actually an upper canine tooth that grows through the whale's lip in a counterclockwise spiral; it can reach a length of two meters. Because the narwhal's habitat is the Arctic the tusk is continuously exposed to water at a temperature close to zero degrees C. The narwhal's other upper canine tooth develops at the same time as the tusk but remains inside the mouth, where it is exposed to a body temperature of 37 degrees C. The two teeth might therefore offer a means of calibrating the racemization age scale. The effect of the temperature difference has recently been investigated by Jeffrey L. Bada of the Scripps Institution of Oceanography and his colleagues Steven Brown and Patricia M. Masters.

The three investigators summarize their findings in a recent issue of *Report of the International Whaling Commission* that was devoted to problems of cetacean age determination. Their analysis began with a tusk that had 40 concentric layers; each layer was presumed to represent one year's growth. Samples were taken from the tip of the tusk (the first part of the canine to erupt), from the base of the tusk (which developed later) and from individual layers (which developed successively). The ratio of D- to L-aspartic acid was determined for each sample. No racemization and thus no clue to the narwhal's age could be detected. When Bada and his colleagues analyzed enamel from the tip of the same whale's other canine, the D-to-L ratio indicated that the animal was approximately 40 years old.

The current whalers' method of estimating narwhal age, which is done in order to conserve the younger animals, is to judge by the length of the tusk. Because females do not grow tusks it has heretofore been impossible to estimate their age accurately. The work of Bada and his colleagues suggests that a racemization study of narwhal teeth from female specimens of different sizes and weights could lead to improvement in age estimation.


067
077
ROBO SENSOR
Matsushita sensors work like the
human senses to let this experimental robot
measure everything from your body temperature
to the strength of your grip.

JAPANESE TECHNOLOGY TODAY

THE AUTHORS

GENE ADRIAN GREGORY and AKIO ETORI are, respectively, Professor of International Business in the Department of Comparative Culture at Sophia University, Tokyo, and managing editor of *Saiensu,* the Japanese-language edition of *Scientific American*.

Gregory received his BS in Foreign Service, Georgetown University in 1948, his MA in International Studies, Johns Hopkins University in 1950. He is a former associate editor of *U.S. News & World Report,* a contributor to *Far East Economic Review,* author of *The Japanese Challenge in Europe*.

Etori, an award-winning writer on science and technology, has covered these areas for over 20 years in Japan.

THE COVER

Photographer Masaaki Nakagawa has contrasted ancient and modern Japan by placing a milestone in VLSI technology in the hand of Japan's famous Kabuki actor, Tokizo Nakamura V. The 64K Static RAM from Matsushita Electric, Japan's largest consumer electronics company, has been achieved by superimposing the load resistors on the MOS transistors. This circuit has 402,500 elements integrated into an area of $31.6mm^2$.

Art direction, supervision of text and illustration and coordination of production were provided by Ted Bates Advertising/New York.

Less than thirty years ago the electronics industry was very much an Atlantic affair. Few national leaders paid much attention to it, except in wartime when radio communications and radar became vital to national interests, and what might pass for electronic equipment was hardly identifiable in the foreign trade accounts of the advanced industrial nations. In most countries, effort was devoted to the design of communications systems, whose terminal equipment and transmission networks constituted the measurable whole of the industry. Much of the effort went to assure autarchical exclusivity to national manufacturers, and what could not be attained through technological design was achieved by intricate cartel arrangements. Quietly, between themselves, national communications organizations and their main suppliers agreed on what equipment to manufacture, at what prices; and equipment manufacturers on both sides of the Atlantic contrived, through patent pools and marketing arrangements, to satisfy those requirements. Consumers took what was available from their respective national industries, and manufacturers exported only to their own country's preferential markets.

Now all that has changed. Electronics has moved to the forefront of industrial activity, replacing steel as the keystone of the industrial system and the symbol of national economic prowess. Development of the electronics industry has thus been granted priority on the agenda of industrial policy-making in advanced and developing countries alike. New technology has obliterated old monopolistic structures. The epicenter of electronics manufacture and innovation has shifted convulsively from the Atlantic to the Pacific Basin, with the emergence of Silicon Valley and Japan as the bipodal foci of a protean technological revolution. Among the momentous aspects of this revolution has been the emergence of Japanese manufacturers as leaders in new product development and in production systems which combine advanced microelectronics and highly developed human resources with cooperative institutional arrangements to set consummate standards of quality, reliability and efficiency for the industry. A country reputed for the ability of its industry to imitate modern technology has now emerged as a major source of innovation in advanced electronics.

Three signal developments in July 1981 underscore this revolutionary change in the configuration of the global industrial and technological order. During high level consultations on security matters, the U.S. government asked Japan to provide advanced very large scale integration (VLSI) technology to enhance air and anti-submarine defense capabilities. A week later, it was revealed in Tokyo that General Motors had requested the Toshiba Corporation to help in developing an electronic control system for automobile engines; compared with GM's best device using nearly 20 large-scale integration elements, the Toshiba system performs as well or better with only four or five more highly integrated circuits. At the same time, it was learned that General Electric was negotiating with Hitachi, Ltd. for extensive technological assistance in the manufacture of VLSI circuits in pursuance of its strategy to make a full-scale reentry into the semiconductor business, from which it had withdrawn ten years previously. According to informed sources, GE was seeking not only Hitachi's advanced technology, but the formation of a joint VLSI manufacturing company in the U.S.

Evidence that Japan was overtaking the American lead in integrated circuit (IC) technology had been mounting ever since Japanese semiconductor manufacturers came from nowhere in 1976 to take a 40 percent share of the world market for 16K random access memories (RAM), the then-ultimate in the art of packing circuits onto the 31-square-millimeter standard chip. In 1978, Fujitsu Ltd. was the first to announce the commercial production of 64K RAM circuits, which pack 4 times as many circuits on the standard chip. The Japanese industry thereupon grasped the lead in global competition in this new generation of devices, with Hitachi, Nippon Electric, Toshiba, Matsushita and Mitsubishi following in force, all bidding with Fujitsu for a share of the action. (See Matsushita Electric's 64K RAM chip on finger of Kabuki actor in the cover picture of this report on "Japanese Technology Today.")

Some industry observers have predicted that Japanese makers will take anywhere from 60 to 70 percent of the world market, given their ability to deliver reliable parts in quantity; and even U.S. industry executives have stated that it will be difficult to hold the Japanese industry's world market share below 40 percent.

Any lingering doubts that Japan had caught up with the American semiconductor industry in key areas of advanced semiconductor technology were dispelled early in 1980 when representatives of Matsushita, NEC-Toshiba and Nippon Telephone and Telegraph's Musashino Laboratory provided a solid-state-electronics conference in San Francisco with detailed descriptions of memory chips that could store still another four times as many bits of information—256K—as the 64K RAM chip. At a time when U.S. semiconductor firms were wrestling with 64K RAM production problems, Japanese manufacturers were gearing up for the next generation of 256K dynamic RAMs. Although there is no hard evidence that Japanese semiconductor makers will begin production before 1984-85, indications are that market demand, rather than technical restraints in manufacturing, will dictate the timing for introduction of these new tightly packed VLSI microcircuits.

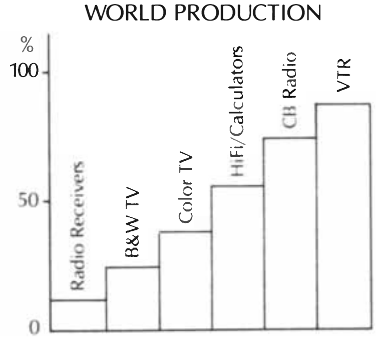
JAPANESE TECHNOLOGY TODAY

It is in this transition to the stage of VLSI production that the current lead of Japanese manufacturers will assume critical importance. Their advance in 64K RAM production, and the strong market position which this assures Japanese makers, signifies more than just another in a series of industrial triumphs dating back to the 1950's when Sony introduced the first commercially produced transistor radios. This combination of production and market prowess is expected to play a pivotal role in determining the strength of the Japanese industry throughout the VLSI era, which will extend through the 1980's and well into the 1990's.

The probable consequences of this structural change in the global electronics industry are of epochal import. If the Japanese electronics industry has been so successful in product and production innovation during the 1960's and 1970's—a period when it was largely dependent on foreign sources of technology—there are reasons to expect that leadership in advanced microelectronics technology will enable Japanese firms to move even further out in front, pioneering the substitution of electronics for mechanical systems in a broadening range of products, introducing entirely new product generations and gaining global market share in the process.

Reasons for this expectation are not based upon logic alone, however compelling that may be. In this case, syllogism is dramatically buttressed by past experience as well as by the observable dynamics of technological and industrial change.

JAPANESE SHARE OF WORLD PRODUCTION



Source: Japan Electronics Industry Development Association.

EXTRAORDINARY MEASURES

Especially significant is the Japanese record of industrial policy-making in this critical sector. As early as 1957, policy-makers identified the electronics industry as a priority sector for development, passing into law what was appropriately called Extraordinary Measures for the Promotion of the Electronics Industry. At the time, few Japanese firms could be classified as electronic equipment manufacturers, and not one among them had distinguished itself as an exporter of electronic products. Total production of the industry, such as it was, had reached barely 60 billion yen (US$167 million) at the end of fiscal 1955, and growth rates had not yet reached the point of takeoff.

The record since then is now legendary. The Japanese electronics industry emerged as the world's leader in the production of color television sets and a broad range of home entertainment products, electronic organs, electronic watches, electronic calculators, numerical controls for machine tools, robots, automotive control systems, medical electronics and virtually every component used in these various types of equipment. By the mid-1970's, Japan's output of computers, communications equipment, office machines and semiconductors was second only to that of the United States. Opto-electronics was emerging as a new field for Japanese special development, and an increasing number of products such as video tape recorders (VTRs) were the creatures of a newly discovered—or rediscovered—genius for innovation.

Significantly, during the last half of the 1970's, when other industries were languishing in the doldrums of cyclical recession or suffering from structural depression, the output of the Japanese electronics industry continued to develop at a fast pace. The total value of its production, second only to that of the U.S., reached 8,683 billion yen in 1980—or almost 150 times that of 1955—making electronics one of Japan's major industrial sectors, on a par with the automotive industry.

Present parity appears, however, to be preliminary to tomorrow's preeminence. Since with each successive generation of more highly integrated circuits the diffusion of electronic technology is extended to new products and applications, permeating one sector of the industrial system after another, the electronics industry will soon surpass all traditional industries in size and diversity.

During the post-oil-shock recession years of 1975-1978, for example, output of electronic products increased 50 percent, compared with a 33.5 percent rise in automobile production. At the same time, the automotive industry became an important new market for electronic products. With the increasing importance of fuel economies, electronically controlled fuel injection systems were developed using new generations of microcomputers. And to save both energy and labor in manufacturing processes, the automotive industry invested massively in robots and other automatic-production equipment with microcomputer controls.

If not to the same extent, the steel industry—which electronics is rapidly replacing as the basic sector of Japan's industrial edifice—also became an important market for electronic equipment during the latter 1970's.

Toshio Ikeshima, executive vice president of Sumitomo Metal Industries reviews the introduction of computer and electronics technology into post-war Japanese steel operations:

"It was 20 years ago that we started studying computer utilization both for handling of clerical work and plant operation. The first introduction of computers was made in 1968 for controlling of basic oxygen furnaces. The technology was acquired from the United States, where computer control technology was almost monopolized by General Electric. Application of domestically developed technology was considered when work on construction of the Kashima steel mill was started in 1969. We decided to computerize the plant's operation as much as possible and assigned hundreds of systems engineers to develop the control technology while the plant was in the making. The difference from the United States in this respect was that we developed the system entirely on our own. In the United States, such work would usually be commissioned to a computer company, which often fails to fully recognize the plant's potentials.

"Today, computers are serving to systemize all sorts of work in steel-making, ranging from control to energy-saving to labor-saving. To date we have invested 30 billion yen in computers. In addition, we annually spend about 6 billion yen for leasing of computers plus 8.4 billion in other related expenses. The sum amounts to 1.2 percent of annual sales. On the other hand, the annual amount of benefit from all this is calculated to be 50 billion yen, of which 40 percent is from labor-saving (reduction of workforce), 23 percent from improvement in product yield, and 37 percent from improved efficiency, energy-saving, reduction of inventories and other factors.

"Computerization proved relatively easy for producing sheets and plates, but not so for operation of basic oxygen furnaces and blast furnaces, of which the workings are not scientifically clarified.

"Cost reduction remains a principal consideration—over new product devel-

opment, for example—in our industry. A cost cutback of 0.1 percent means a lot in terms of profit performance. Improvement of product yield and energy-saving carry big weight in cost reduction. In this respect, we are attaching importance to the continuous casting technology, which could not be managed at all without computerized control."

LONG RANGE VISION

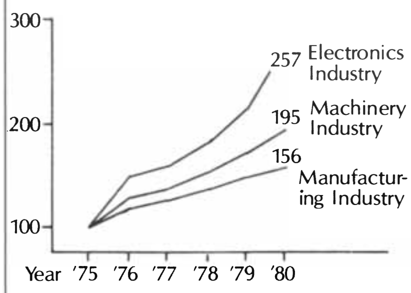
As obvious as these benefits might seem, diffusion of electronics technology and products in Japan has not been left entirely to the mechanics of the market. In 1974, after the oil crisis had made clear the necessity of reassessing Japan's industrial policy, the Industrial Structures Council published its revised *Long Range Vision* of Japanese industry, designating the electronics industry as the key sector around which virtually the entire future production structure was to be developed. Since Japan is almost wholly dependent upon imported energy and other raw materials supplies, rising crude oil prices and the uncertainties which followed in commodity markets seriously endangered the competitiveness of many key export products. Continued competitive power required a massive, coordinated effort to allocate and use resources more efficiently, and to achieve economies in production through increased productivity as well as improved quality. As electronics technology is "energy-saving, resource-saving, labor-saving and space-saving," in the words of Michiyuki Uenohara, senior vice president of Nippon Electric Co. Ltd., its rapid development was catapulted to the highest levels of national priority.

Not only was the electronics industry of central importance for increasing value-added in industry to sustain higher standards of living in an ever more affluent Japan, but competitiveness in world markets became more dependent upon innovation in advanced electronic export products and the most rapid possible diffusion of efficiency-inducing electronic technologies throughout the industrial system. In the apt words of Singapore's Premier Lee Kuan-Yew: "There is nothing like the threat of hanging in the morning to concentrate one's attention the night before." It was just such a perception of danger that gave a massive added thrust to the dynamic development of the Japanese electronics industry in the post-oil-crisis period.

But crisis was not the sole impetus to the now-renowned cooperative VLSI research and development project which was launched in 1975, following the directions indicated in the new industrial "vision." Nor was necessity itself sufficient cloth from which to fashion success in the intended bid for technological leadership in advanced electronics technology.

In the two decades since its birth, the Japanese electronics industry had, of course, accumulated a vast stock of technology. Innovation had come early and often from the industry, largely as the result of cooperative research and development—albeit mostly in applied technology—within enterprises and industrial groups. It was mainly because of this intense pursuit of product and production innovation that Japan had become the world's leader in consumer electronics and was moving out front in other product lines as well.

LABOR PRODUCTIVITY INDEX
(1975=100)



Calculation Method:

$$= \frac{\text{Mining and Man. Prod. Index}}{\text{Average of Total Actual Manpower}}$$

Figures for manufacturing from Japan Productivity Center. 1980 figures for manufacturing are averages from January to October.

Source: Electronics Industries Association of Japan.

Makoto Kikuchi, director and manager of the Research Center of the Sony Corporation, has these reflections on the role of the Japanese home market in preparing Japanese manufacturers for ventures abroad:

"We first of all should be aware that, while Japan is ahead in certain areas, Japan is lagging in some others. I reckon the U.S. and Japanese levels are pretty close to each other. For example, Japan is ahead of the U.S. in reliability so far as many semiconductor devices and electronic circuits are concerned. When it comes to pioneering research in areas rarely taken up by others, however, the U.S. continues to strongly lead Japan. So it is hard to say simply which country is ahead of the other, although it is a fact that the gap between the two countries has become so narrow that it is difficult to make a simple comparison.

"When Sony tried to make the transistor radio, the idea was laughed off in the United States. At that time, the potential of transistors was still totally uncertain. But the top management of Sony has always been quick in making such decisions. Once a decision is made, however, the company forges toward the goal. Setting a goal at an early point encourages everybody to devote himself to attainment of the goal.

"When the target is clearly defined and zeroed in on, Japanese workmanship flourishes in that specific area. The tape recorder and video tape recorder are examples. Along with electronics, good technology for machinery is important to achieve results. A combination of technologies in the two areas holds the key to production of good products.

"I suppose there is better communication and harmony in work among Japanese technicians than among their Western counterparts. Provided that motivation is adequate, the Japanese tend to willingly do the same work that other people do. This means that the more technology becomes mature, the easier it becomes to develop technology. In Europe, on the contrary, if someone is engaged in work A, he feels that he must take up work B at all costs because otherwise he cannot maintain his identity. In Japan, everyone will take up work A because of the strong inclination to join other people in work in the most current area. It's hard to say which way is better, but the Japanese way seems to be advantageous from the viewpoint of commercial technology.

"For a long time, the struggle to catch up with the international level constituted a major motivation for the Japanese. To this end, mastering technology that played a key role in that catch-up process has been highly valued. The result is that a great number of people in this country become engrossed in mastering new technology, even before they bother themselves about how such technology can be utilized for practical purposes. This is a tendency which is attributed to the Japanese situation and the environment in which Japan has been placed, and unless this point is properly understood, it will be difficult for foreigners to understand why the whole nation of Japan rushes to develop LSI's, for example.

"To give you another example, French intellectuals say they don't understand why the Japanese habitually turn on the television every morning and watch it to know the time. 'They have watches and clocks to know time, don't they?,' the French ask. I think this represents the coldness of the European market toward television.

"The Japanese society is warm, in every

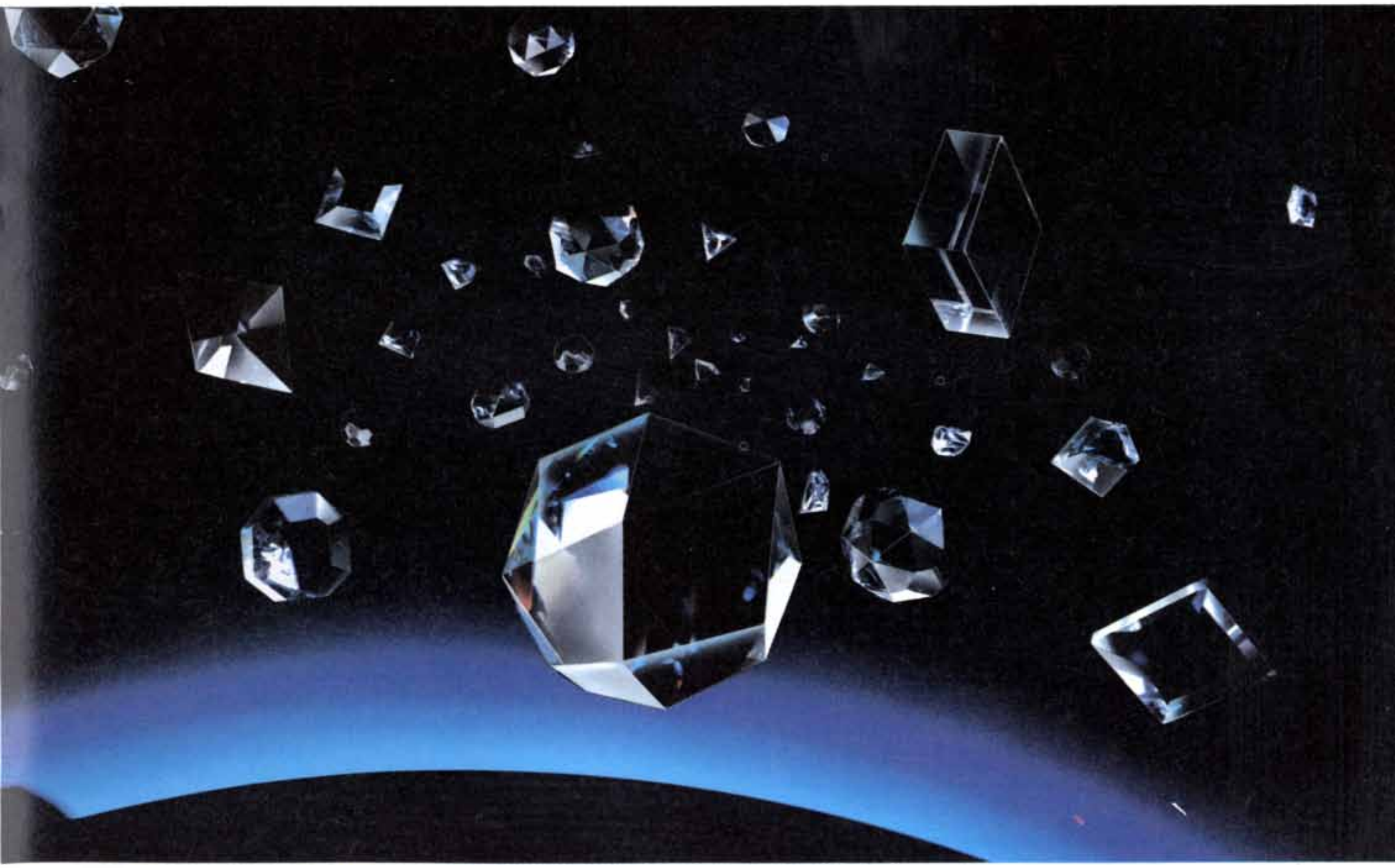
Farming perfect dodecahedrons in space.

Below, the earth looms large and blue. Above, the stars blaze brilliantly. And in weightless silence, molten molecules coalesce slowly into a multitude of crystalline shapes—cubes, octahedrons, dodecahedrons, and even more complex forms. Unfettered by gravity, the atom-by-atom structure of these crystals grows in perfect replication of the geometric lattice of the original seed.

At NEC, we're already working to turn this scenario into reality. In fact, the weightlessness of space may make possible a new world for progress: alloys, such as lead-aluminum, which cannot be produced on earth. Completely new drugs and chemicals. And, yes, acres of astonishingly perfect crystals, growing free of the disturbances and imperfections caused by gravity. For faster transistors, higher-quality large-scale integration, increasingly powerful solar batteries—and more.

And NEC will be there. The National Space Development Agency of Japan will place a special NEC image furnace aboard an early spacelab. This furnace, carried into orbit by the Space Shuttle, will grow crystals and create alloys free of the forces on earth.

Development of advanced techniques such as manufacturing and "farming" in space is just another reason why NEC has achieved such a remarkably effective synthesis of computers and communications. And why we have a following in over 130 countries around the world.



THE COMPUTER & COMMUNICATIONS COMPANY.

NEC

Nippon Electric Co.,Ltd.
P.O. Box 1, Takanawa, Tokyo, Japan

aspect—be it either research or consumers. Mental temperature is an expression which I frequently use, and this mental temperature, it seems to me, is high among the Japanese. This is a factor that makes Japan a market of tremendous strength and scale. The booming discount trade at camera and microcomputer shows, packed to capacity, are good examples of that. Hot markets like these are something you don't find outside Japan. If such enthusiasm on the part of consumers disappears, Japanese technology may instantly die away. There exists such a hot market at home and consumers there have a high mental temperature, with everybody convinced that using something new is a virtue—these are factors of extraordinary importance."

The Japanese consumer has played an equally central role in the fortunes of Suwa Seikosha Co., Ltd., and its Seiko watch line, as Hideaki Yasukawa, director, states:

"If Japan now surpasses the United States when it comes to technology to turn out consumer goods, that is because Japanese industry must compete for the favor, in the first place, of Japanese consumers. The users of electronic products, be they manufacturers or ultimate consumers, are more demanding customers than their counterparts in Europe and the United States. This is an advantage of Japan not often mentioned. But the existence of knowledgeable and demanding users must lead to high-quality products.

"Japanese electronics would not have achieved such a phenomenal growth if the level of consumers' utilization of products had not been so high or so demanding. It is Japan's great strength that it has a big domestic market.

"By contrast, Switzerland, for example, lacks a domestic market, and product users lack basic knowledge of electronics. Under the circumstances it often happens that a Japanese IC manufacturer negotiating with a Swiss watchmaker about specifications of products to be supplied faces difficulties in reaching conclusions because the Swiss tend to leave everything to the discretion of the Japanese IC manufacturer."

If this is becoming increasingly apparent even to the casual observer of the Japanese scene, the dynamic processes which have made it so remain somewhat more elusive. The critical interactive factors which determine that process, and which go far towards explaining the Japanese advance in VLSI technology, are not, however, difficult to identify. The basic ingredients of progress in microelectronics technology in Japan—simply stated—are: men, money and enterprise functioning in a fiercely competitive business environment. These, in turn, have fashioned particular corporate strategies and structures which serve to accelerate the pace of technological change and assure the efficient management of scarce resources.

STRENGTH IN HUMAN RESOURCES

History is eloquent witness that the electronics industry is one that will spread on a global scale in the quest of high-calibre human resources most adept at efficient enterprise. In brief, highly educated people who work well together are the *sine qua non* of electronic technology. Herein lies the principal explanation of Japanese ascendancy to leadership in consumer electronics production, the rapid overtaking of the American industry in LSI production, and the advantage so far attained in VLSI technology. It is likely to be the prime determinant of the role of Japanese firms in the industry for decades to come.

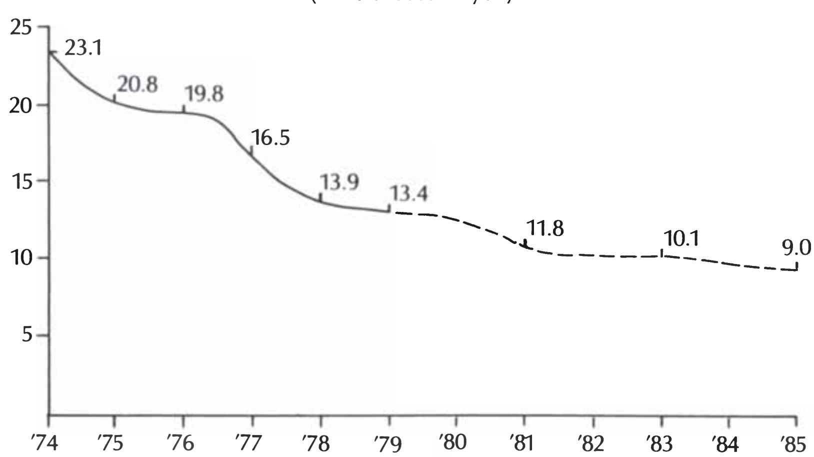
Given the paucity of Japan's natural resources, the affinity of Japanese industrialists for a knowledge- and skill-intensive industry such as electronics is derived from no special insight. It is plain common sense. Just as comparative advantage was originally obtained by wise use of abundant sources of educated low-cost manpower in the earlier stages of Japan's industrialization, the prospect for continued technological and industrial progress in its advanced stages is to be found only in technologies and production which require abundant sources of educated high-cost manpower.

Electronics technology is appropriate to Japan not only because of the relative scarcities of natural resources and the abundance of manpower, however. The quality of human resources makes the difference. Traditional arts and crafts, as well as the light industries typical of Japan's pre-World War II economic development, relied heavily upon dexterity of hand, concern for detail, a penchant for intricacy and a quest for perfection. Skills which found their supreme manifestation in the carving of ideographs on individual grains of rice (as many as 5000 kanji characters have been inscribed on a single rice grain), were ideally suited to the tedious tasks of designing and producing high-density microelectronic devices.

For a country totally reliant on its human resources for survival and welfare, education commands primacy of place among societal values. It follows, too, that for industry in such a country, education—the development of human resources—is of foremost concern. Not by accident does Japanese industry recruit workers directly and systematically from high schools, the fast-growing and most technically advanced firms hiring the best graduates from the best schools. Nor is it accidental that managers and engineers are just as systematically recruited at graduation from university. And since the reputation and strength of educational institutions depends to a great extent on their ability to place their graduates in positions with leading companies, there is competition among schools to prepare students for lifetime employment in the best industrial, financial and commercial enterprises.

Minoru Suzuki, managing director of Asahi Optical Co. Ltd., has this to say about the role of Japan's educational sys-

VTR PRICE TRENDS IN THE 1980s.
(in 10 thousand yen)



Source: Japan Electronics Industry Development Association.

hexa·photo·cybernetic
The Possibilities are Endless.

Stopped-Down 90

Six-mode exposure control. System versatility. Newer electronics for wider applications.

The Canon A-1 is one of the world's most advanced automatic SLR cameras. Combining the finest in optical and mechanical engineering with the most sophisticated electronics, it's technology applied to give you the ultimate in creative control. At the touch of a button.

Depending on your subject, you can choose from six independent exposure modes to achieve the results you want:

1 Shutter-Priority: You select the shutter speed, to freeze the action and prevent camera shake or create an intentional blur. The A-1 automatically selects the appropriate lens opening.

2 Aperture-Priority: Control the area in focus by selecting the lens opening for the effect you want. The A-1 matches with the right speed.

3 Programmed: When you need to shoot fast, just focus. The A-1 will select *both speed and aperture* for great results.

4 Stopped-Down: For extreme close-up or specialized photography, a bellows, a microscope or almost anything can be attached to the A-1. It's still automatic.

5 Flash: Totally automatic flash photography, of course, with a wide variety of Canon Speedlites to choose from.

6 Manual: Yes. For those times when you absolutely want to do it all yourself. To experiment. To explore the possibilities.

There are over forty fine Canon lenses ranging from Fish Eye to Super Telephoto, plus accessories to meet every need. If you can't photograph your subject with a Canon A-1, it probably can't be photographed.

From the sophistication of its LED viewfinder display, to a ruggedness that allows up to five-frame-per-second motor drive, the Canon A-1 represents an incredible technology. At a price that makes owning one a definite possibility.

Canon A-1

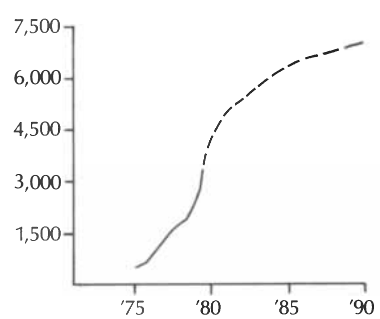
Canon USA, Inc., One Canon Plaza, Lake Success, New York 11042 • 140 Industrial Drive, Elmhurst, Illinois 60126 • 6380 Peachtree Industrial Blvd., Norcross, Georgia 30071 • 123 Paularino Avenue East, Costa Mesa, California 92626 • Bldg. B-2, 1050 Ala Moana Blvd., Honolulu, Hawaii 96814 • Canon Optics & Business Machines Canada, Ltd., Ontario *tem in the country's accelerating economic development:*

"Of basic importance is popularization of education. Today most youngsters go to high school where they acquire a fairly high level of knowledge on an average. As a result of an increase in the number of universities in the postwar period, more people receive higher education. Another important factor is the Japanese hunger for advancement and willingness to work. A factor behind such a national trait is perhaps the existence of a number of enterprises of a similar size in each area of industry and subsequent severe competition in a nation of limited physical size. Japanese enterprises have been undertaking technical development and innovation while competing hard with each other. Before they are aware, they have gone ahead of foreign competitors.

"There is the need for greater cooperation between national research organizations and private enterprises in the area of basic research. Research budgets for Japanese universities are too small—the state should provide more financial assistance to them. Since Japan has a scarcity of natural resources, development of technologies geared to efficient use of such resources ought to be stressed. In terms of flow of economic wealth, Japan has grown rich at all levels—individuals, enterprises and state—but it still lags behind in many raw materials."

Since the electronics industry offers the best opportunities for human knowledge and skill development, as well as for rapid advance in companies having the fastest growth, students show a strong preference for work in electronics, and electronics manufacturers can select the best students. It only remains for educational institutions to make the selection process possible, and the most fruitful, for both students and employers.

VTR OUTPUT IN THE 1980s.
(in 100 million yen)



Source: Japan Electronics Industry Development Association.

It is not surprising, therefore, that Japanese universities graduate about 50 percent more electronics engineers than universities in the United States, which works out to approximately three times as many per capita. In the information industry *par excellence*—electronics—people are the critical ingredient, the more so the more knowledge-intensive electronics technology becomes. There is, therefore, a keen, and quite unique, recognition in Japan that microelectronics and its progeny—computers, automated equipment, robots—do not in fact replace people. Electronics makes people more essential. It renders the development of their most important attributes—brainpower and cooperative endeavor—imperative. In software alone, Japanese industry sees the need for highly trained manpower as insatiable.

On this point, Sakae Shimizu, director of Toshiba Corporation, observes:

"If we have to increase the number of software engineers at the present rate, all the people on the planet will have to be turned into software engineers in a few years. This, of course, is impossible, and so we have to develop software technology to avoid it. We already have made significant headway in this field.

"Technologies developed so far include application of pattern recognition. It is already in practical use to sort out postal codes and bills. Also, machines can read kana (Japanese alphabet) with a 99.9999 percent accuracy and a limited number of the kanji (Chinese) characters as used in newspapers.

"Research is also progressing on voice recognition, demand for which has been rising recently. An advantage of voice recognition is that it makes interaction between man and machine easier, allowing anybody to make use of machines. Man-machine interface made easier will contribute to further popularization of electronics technology.

"Electronics is a very effective means of providing homes with comfortable and safe living conditions. The use of built-in microcomputers in air conditioners, washing machines and microwave ovens is just one example. If further progress is made in sensor technology, a completely automatic control of room temperature and humidity will be a reality. It will also bring about better burglar- and fire-prevention systems. Another possibility is a broad range of communications between the inside and outside of homes, pending the completion of communications and information processing systems for home use. This will provide a free access to outside information, satisfying intellectual desires and closing the distance between residence and work, even making it possible to do one's job without reporting to the office."

Obviously, given the pace of technological change in microelectronics with a new generation of IC devices emerging every two to five years, the requisite perpetual development of human resources cannot be left to the universities, however efficient they might be in providing raw, inexperienced recruits. It is in meeting this vitally important need for continuing education or training that Japanese industry excels over all others.

Lifelong employment makes perpetual and intensive investment in human resources possible. Japanese companies can provide training to their employees without fear that they will leave the firm in response to offers of higher remuneration by competitors. Workers, engineers and managers, for their part, can devote their full energies to self-development, knowing that they will have an opportunity to use their acquired knowledge to good advantage—their own, their company's and their nation's—for the whole of their working careers. And, of course, just as technological innovation in the electronics industry is the key to competitive advantage, technological skills are a vital element in career advancement, even in companies where seniority sets the order in which employees mount the corporate ladder.

A vivid insight into the "tenure" system in Japanese industrial companies is supplied by Minoru Morita, managing director, Corporate Production Engineering Division of Matsushita Electric Industrial Co., Ltd.:

"The lifetime employment system may give foreigners the impression that a man is bound to a company all through his life, but it is never so. It does not mean that you are not allowed to quit. The truth is that you can stay with the company all through life if you so wish. Thus the managements do not resort to layoff at such difficult times, for example, as the oil shock, nor does your company expect you to stay all through your life.

"New college graduates who join the company each year are not hired as an immediate workforce. They are first put through a long orientation—lasting six months or one year—aimed at versing them in the organization of the company. During this period, those with an engineering background, for example, learn marketing on the job, while researchers experience jobs on the line. They are paid full salaries and other benefits, including transportation expenses for commuting, which means a big investment from the standpoint of the company. But this investment pays off in years to come, as these employees, thanks to the orienta-

The most famous name in Japanese quality control is American.

His name is Dr. W. Edwards Deming, and he's a quality control expert.

In 1950, the Union of Japanese Scientists and Engineers (JUSE) invited Dr. Deming to lecture several times in Japan, events that turned out to be overwhelmingly successful.

To commemorate Dr. Deming's visit and to further Japan's development of quality control, JUSE shortly thereafter established the Deming Prizes, to be presented each year to the Japanese companies with the most outstanding achievements in quality control.

Today, Dr. Deming's name is well known within Japan's industrial community, and companies compete fiercely to win the prestigious Demings.

In 1953, Sumitomo Metals was fortunate enough to win the Deming Prize For Application. In retrospect, we believe it may have been the single most important event in the history of quality control at Sumitomo. By inspiring us to even greater efforts, it helped us to eventually become one of the world's largest and most advanced steel-makers.

Sumitomo Metals owes a great deal to the American quality control expert who became one of Japan's greatest inspirations. On that point, the management and employees of Sumitomo Metals would like to take this opportunity to say simply, "Thanks, Dr. Deming, for helping to start it all."

SUMITOMO METALS

SUMITOMO METAL INDUSTRIES, LTD. Tokyo & Osaka, Japan
Sumitomo Metal America, Inc. Head Office: 420 Lexington Ave., New York, N.Y. 10017, U.S.A. Phone: (212) 949-4760 Offices in Los Angeles, Chicago, Houston
MAIN PRODUCTS: Tubes & pipe, flat rolled products, bars, shapes & wire rods, rolling stock parts, steel castings & forgings, and a wide range of steel products

tion training, develop the ability to understand each other in immediate communication and broaden the sphere of activities in accordance with competence.

"To use the metaphor of a Go (oriental game) board covered by stones, the system of a European or American enterprise assigns each stone a job which is neatly defined, with the entire work parcelled out, man by man. In the case of a Japanese company, however, stones are placed on every third line or every other line, and the space left vacant is taken care of by surrounding stones in cooperation. Such a system leaves room for freedom of action, enabling a capable, positive-minded person to make full use of his competence.

"Such an environment exists also for low-level workers. Since room is left for freely expanding one's job within the organization, it seems that individual capabilities have a good chance of being developed.

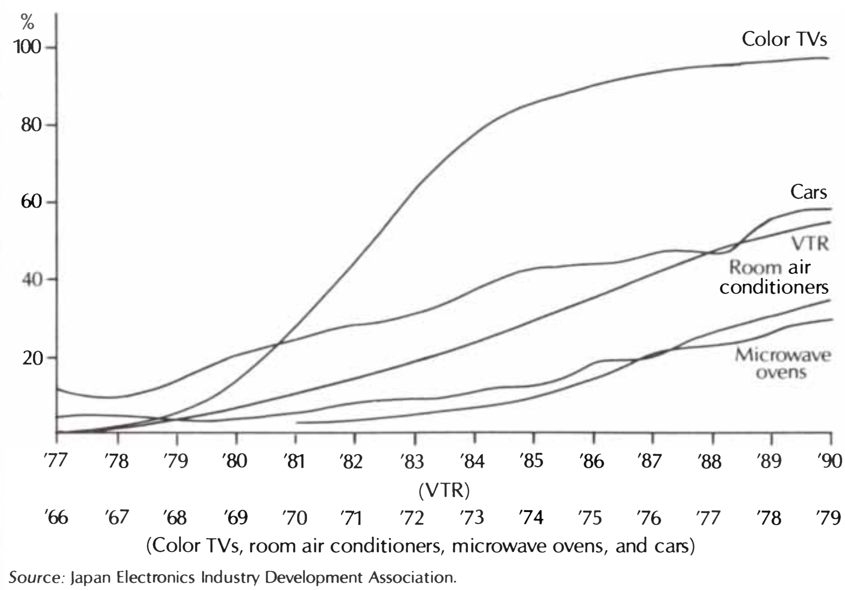
"At one point in the past, we tried to learn the rational personnel organization system of foreign countries. It seems to have turned out, however, that in the Japanese environment, the lifetime employment system—this sort of 'roomy' personnel allocation system—fits better, despite the ambiguity as to where responsibility rests.

"Additional proof of the superiority of the Japanese system is Japan's success in adopting Deming's quality control system, which was unsuccessful in the United States. This quality control system now contributes greatly to the enhancement of workers' morale, not to speak of improvement of product quality.

"A major problem at present is white collar workers. There is no accumulation of know-how for technology development and design which can be utilized by these workers. We need to work out a system that bonds white collar workers better."

Lifelong employment adds a further differential factor to the equation of human resource development in Japanese industry, a factor which has special import for prowess in microelectronics production. In the economics of all semiconductor device production, experience is a critical ingredient of competitive power. Each time a company doubles its production of a given device, cumulatively over time, it does the job better, achieving cost reductions of from 20 to 30 percent. It is important to note that the experience involved here is collective: the experience of a group working together at a given task. It follows, therefore, that those companies in which the members work together the longest have the most cumulative experience and the most cooperative systems of working together, are likely to move down the experience/cost curve faster and further than their competitors.

COMPARATIVE DIFFUSION RATES OF VTR AND OTHER CONSUMER DURABLES.



Source: Japan Electronics Industry Development Association.

As Shou Masujima, managing director of TDK Electronics Co., Ltd., explains that company is able to set itself the seeming utopian goal of zero-defect quality control:

"To supervise and encourage workers to minimize errors is what Zero Defect Quality Control—ZQC—is all about. The point here is to see how far we can go in trimming human errors.

"We are nearing the limit in the present approach. This means that if we try to continue to reduce errors to a point close to zero on the basis of existing equipment, a vast amount of additional cost will be required.

"The capabilities of the Japanese in a sense, we may say, have reached their limit. Machinery will have to be introduced to achieve quality control beyond this point.

"Let me explain: Suppose there are two products which belong to the same category in the existing product control system. Suppose the products are examined at the level of molecule and atom through an electron microscope. One of them is found with a normal distribution of size of crystal structure and the other with a broad difference in the size of crystal structure. Yet the two are the same in terms of product character. The product with varying sizes of crystal structure is likely, however, to undergo changes in crystal structure, hence in product character under a slight impact from outside, such as exposure to heat. In other words, such a product is easily affected by a minor human error in the production process. By contrast, the product with a normal distribution of crystal structure size is less liable to be affected.

"The basic concept of ZQC is to go as far upstream as possible in effecting quality control. In the example just cited, it means employment of machinery to check products on the level of molecule and atom. If quality control is strictly performed at an upstream level, human errors are permissible to some extent without worrying about their effects on the quality of products.

"In a total system of ZQC, the first thing required is to put machinery of the highest precision in the production line and allow as much time as possible for adjustment. Moreover, measuring instruments at the final stage should be arranged in such a way as to make double and triple checks possible.

"At present, TDK Electronics is experimenting with a thorough use of ZQC quality control system in one section of production. The system will be employed in the entire company in the future.

"Manpower reduction is not the a primary objective of ZQC. But, if it is pushed, the number of workers will certainly decrease."

But, however appropriate the institution-

al arrangements, they do not work by themselves. Like a symphony orchestra, a company's performance depends ultimately on the quality of its director. The finest musicians badly directed cannot render great performances. Likewise, the most highly trained human resources cannot produce quality electronic products without expert direction.

This would seem self-evident, and would not merit attention here were it not for the fact that in many advanced countries, the managers of manufacturing enterprises are more concerned with managing assets than in managing people and technology. Management tends to be impersonal; people and technology are expendables, traded in the marketplace of corporate assets by wizard-conglomerators. But manufacturing is people working together to optimize the results obtainable from a given stock of technology and to develop new technologies which will yield better results from those common endeavors. This being so, Japanese manufacturing firms tend to be managed by engineers who have had a lifetime of experience in working together with colleagues in the same enterprise, and one is usually chosen to manage the corporate team for his skills in leading cooperative effort in the management and development of the company's technological stock, skills he has manifested over a 25-30 year period. As a result, Japanese electronics companies have been less prone than some others to lose their technological way. They are likely to be more responsive to changes in technology, and to changes in needs for technology. As Hideo Sugiura, executive vice-president at Honda Motor, puts it: "Our philosophy of technology puts needs above anything else—technology is subordinate to needs. Needs here means not only users' needs but those of workers and dealers." As for Honda, so for most Japanese electronics companies: corporate philosophy is a vitally important aspect of management which states the common purposes of the men and women working together in the enterprise and the role of technology in that endeavor. Technological advance, human development and social progress are perceived as being consistent and essential aspects of corporate purpose. And the effect tends to be synergistic, providing the main thrust for rapid technological change in future-oriented industries such as electronics.

MONEY AND MICROELECTRONICS

If the Japanese electronics industry has been successful in mobilizing and developing human resources, it is in no small part because the financial system has worked to provide the necessary capital resources on terms which are conducive to sustained growth and rapid technological change. High savings and investments have been a major factor in the phenomenal growth of all Japanese industry, especially since 1960; from 1960 to 1977, according to the American Productivity Center, gross private investment averaged 35 percent in Japan, compared to 25 percent in West Germany and about 17 percent in the United States. Moreover, financial resources have been available to Japanese manufacturers on terms and conditions which are particularly conducive to continuing investment over longer time frames. And for no industry has adequate and appropriate financing been more important than for the semiconductor industry, where, as a U.S. industry leader puts it, "You almost have to invest a dollar to sell a dollar's worth of integrated circuits."

In the years 1966-1970, when Japanese semiconductor manufacturers undertook heavy investments to begin IC production, their plant and equipment investments totaled 43.2 billion yen, or 59 percent of total IC sales in that period. At the same time, R&D outlays for the development of IC technology ran as high as 21 billion yen, or an additional 29 percent of revenues from sales.

While production of integrated circuits in Japan grew tenfold during the 1970's and the race for innovation intensified, the cost of silicon wafer fabrication lines rose five or six times and R&D costs increased proportionately. A recent Ministry of International Trade and Industry (MITI) survey shows that between 1973 and 1978 Japanese IC makers reinvested an average of 17.8 percent of their sales in plant and machinery. Then, in fiscal 1979, total capital spending of the nine major IC makers increased by 77.5 percent compared with the previous year, rising to 109.4 billion yen, or more than twice the total amount invested by the industry in startup 1966-70 period. In fiscal 1980, investment of the nation's top 11 semiconductor manufacturers rose another 52.2 percent to 169.5 billion yen, most of which was for new integrated circuit production facilities. In the meantime, R&D expenditures by IC manufacturers rose 100 percent from 17 billion yen in 1973 to 38 billion yen in 1978, and continued advancing at a high rate into the 1980's as the industry geared for entry into VLSI production.

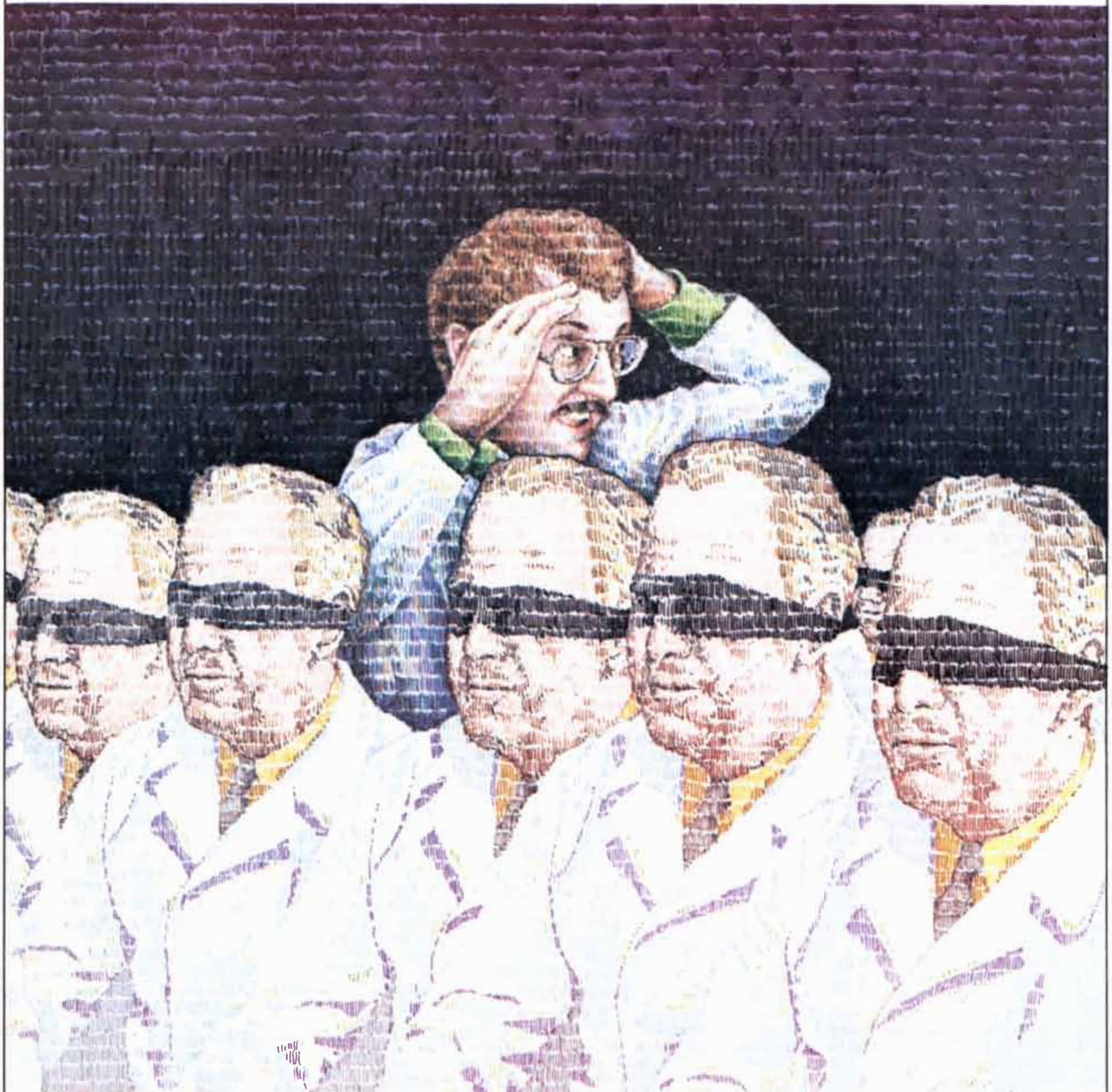
By way of comparison—one which explains the closing of the gap between the American and Japanese industries—U.S. semiconductor manufacturers invested on the average only about 12 percent of their gross revenues from sales of integrated circuits between 1973 and 1978; between 1979 and 1980 U.S. manufacturers increased their outlays by just over 20 percent. The reasons for the higher rate investments in Japan are to be found in lower costs of capital and in the availability of bank loans which are relatively insensitive to irrational money market forces and the epicycle of savings in the business cycle. The typical Japanese semiconductor manufacturer's cost of capital in 1979 was 9.3 percent, or about half the 17.5 percent cost confronting U.S. counterparts. Moreover, the Japanese capital market's demands in the form of profit margins or return on investments were more in keeping with long-term financial requirements and performance of industry. Profits of Nippon Electric, the leading Japanese IC maker, were 1.2 percent on a consolidated basis in 1978, compared with 5.5 percent for Texas Instruments and 5.6 percent for Motorola. During the same period, Fujitsu's consolidated profits of 2.6 percent were a small fraction of IBM's 14.8 percent. These lower returns on sales deliver an entirely satisfactory rate of return, however, on the highly leveraged stockholders' equity in the Japanese firms.

But greater availability and lower cost of capital are not the only reasons why Japanese manufacturers are able to invest more heavily in semiconductor technology and

VTR MARKET GROWTH IN THE 1980s.
(in thousands of units)

		1979	1981	1983	1985	1990
Domestic demand	Fresh-buying demand	480	1,360	1,780	2,310	1,480
	Carryover & add. buying demand	—	40	120	230	1,090
	Total	480	1,400	1,900	2,540	2,570
Diffusion ratio (%)		459	10.4	19.3	30.0	51.9
Exports		1,720	2,650	3,800	4,300	3,500
Gross demand		2,200	4,050	5,700	6,840	6,070

Source: Japan Electronics Industry Development Association.



TDK
SUPER
AVILYN
T-120

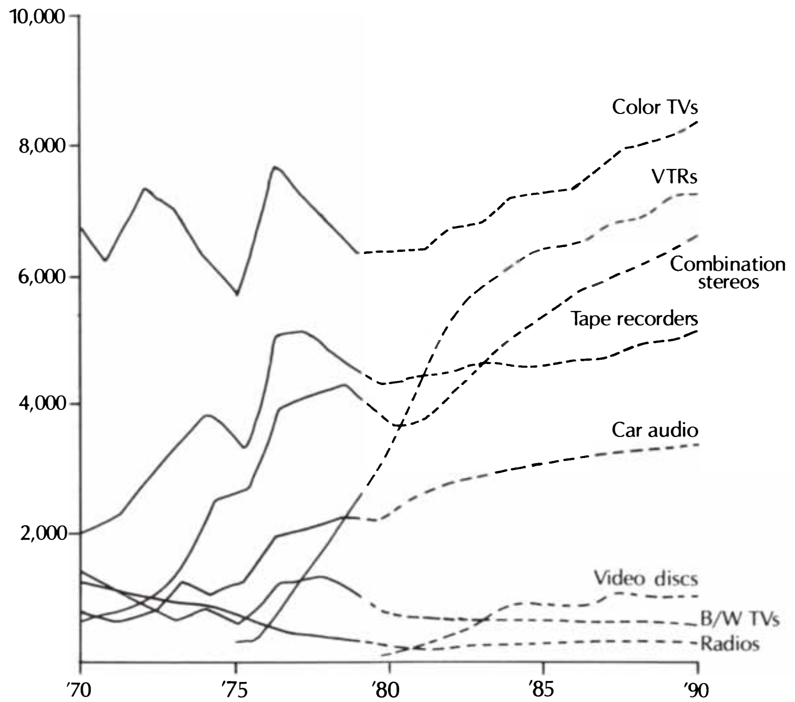
production facilities. All major Japanese semiconductor makers are also highly diversified, vertically integrated electronic equipment manufacturers, including in their various product mixes a wide variety of home appliances, data processing, telecommunications, automotive electronics and medical equipment. Semiconductor sales account for less than 20 percent for Nippon Electric and far less than 10 percent for other major makers; this compares to approximately 30 percent for Texas Instruments and Motorola and over 70 percent for specialized semiconductor manufacturers in Silicon Valley. Because of their vertical integration, Japanese makers can draw on cash flow generated by downstream products to sustain heavy capital requirements and R&D costs of the IC operations. At the same time, highly diversified downstream production provides a ready in-house demand for IC's, which reduces risks of capital outlays, assures scale economies of production and enables firms to obtain learning-curve economies more rapidly.

The net result—in an industry where capital costs are high and product life cycles of successive generations of integrated circuits are quite short—is some very important advantages for diversified Japanese manufacturers over cash-hungry, manpower-short specialized U.S. manufacturers. Economies of scale and learning, plus lower capital costs, enabled Japanese manufacturers to sell 16K RAMs in the U.S. market at 20-30 percent below the prices asked by American makers, according to testimony before the U.S. Senate Finance Committee. During the post-oil-crisis recession, while U.S. semiconductor makers cut back sharply on capital spending, Japanese manufacturers could continue to invest in new plant, equipment and technology.

The more rapid product diffusion as a result of integrated production, combined with these advantages, gives Japanese makers greater flexibility in responding to changing market needs, competitive pressures and technological change. Thus at the end of fiscal 1981, seven Japanese companies will be manufacturing a total of 2.5 million 64K dynamic RAM chips monthly, or about three times the output in June 1981. This monthly production volume compares with the total world production for all of 1980 of a mere 440,000 64K dynamic RAM chips.

Such dramatic increases in production have far-reaching implications for downstream applications and upstream supply of materials and machinery. While the diffusion rate of 64K dynamic RAMs remains still largely a matter of conjecture, the heavy investment by Japanese makers in new plant and equipment to produce this new generation of devices has already been a boon to makers of IC manufacturing machinery and specialized materials suppliers, catapulting an entire new generation of export industries into world markets which until now have been virtually the exclusive domain of American manufacturers.

JAPANESE CONSUMER-ELECTRONICS PRODUCTION IN THE 1980s.
(in 100 million yen)



Source: Japan Electronics Industry Development Association.

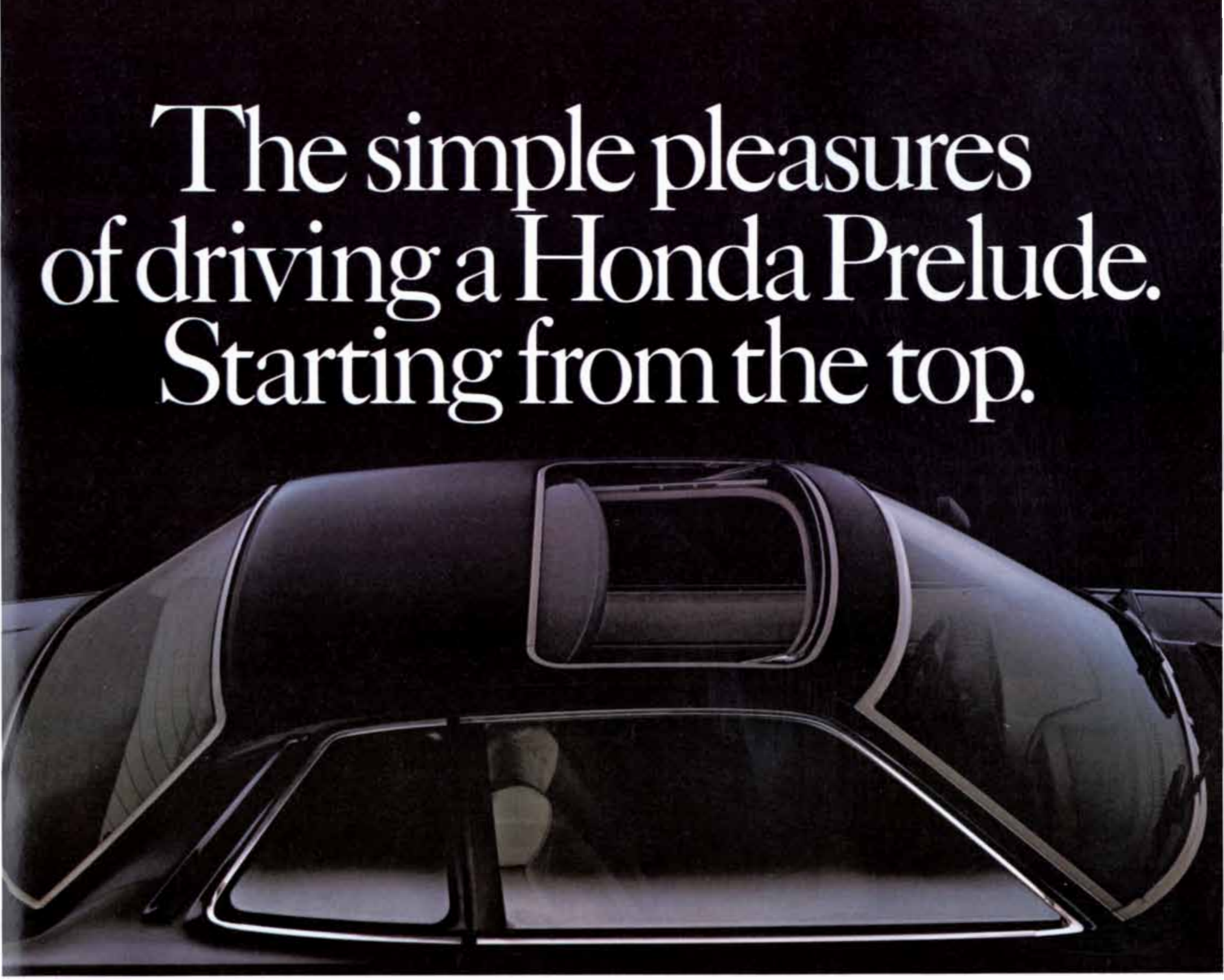
THE ENTERPRISE AND TECHNOLOGICAL CHANGE

Formidable as are the advantages which accrue separately from a finely tuned industrial policy, superior human resources and higher capital investment, it is the capacity of the Japanese enterprise to combine these forces in cooperative endeavor which gives it its ultimate synergistic power. Yet, although it is widely appreciated that there are many aspects of Japanese corporate culture which give to the management system it has produced some especially effective features, worthy of emulation, the fundamental difference between the Japanese and Western enterprise which will ultimately determine the rate of progress in high-technology industries such as electronics is generally given scant attention.

If, for example, Japanese electronics manufacturers have succeeded in attaining higher productivity than their American or European counterparts, it has been found that this competitive advantage cannot be co-opted simply through higher investment in robots and other automated equipment. There are firms abroad which are as automated as those in Japan, but they usually do not get the same results. Similarly, although Japanese semiconductor makers have obtained a strong competitive advantage from higher quality products, the American and European firms have failed to obtain the same high degree of reliability in the end product.

The distinguishing feature of the Japanese corporate enterprise which has made possible the combination of men and money for higher productivity, better quality and more rapid technological change is its cooperative mode of management, as opposed to the zero-sum management systems characteristic of most Western enterprises. This cooperative management system, in turn, is derived from the absence of those dichotomous adversary relationships which pit capital and labor, management and workers, against each other in eternal conflict within Western enterprise.

Ultimately, the Japanese cooperative



We think you'll agree, it's a real pleasure to look at.

But our sportiest Honda also sports the kind of features that make it a thrill to drive.

THE POWER-OPERATED MOONROOF. FOR OPENERS, IT'S STANDARD.

At the push of a button, the tinted-glass moonroof slides back. At the same time, an automatic deflector helps keep the wind out.

When you're behind the wheel, you'll be pleased at what you see in front of it. This year, the instrument panel houses a separate tachometer and speedometer, a quartz digital clock, a maintenance reminder and an electronic warning system.

Of course, the Prelude is just as much of a pleasure to sit in.

Its luxurious interior features comfortable bucket seats and adjustable headrests.

SOME SIMPLE ENGINEERING FACTS.

Our sportiest car has front-wheel drive, four-wheel independent suspension, rack and pinion steering and a responsive 1751cc CVCC® engine. The fact is, the Honda Prelude is designed to give you years of good performance.

POWER STEERING COMES AUTOMATICALLY WITH THE AUTOMATIC TRANSMISSION.

The 5-speed stick shift is standard. But with the optional automatic 3-speed, you also get variable-assist power steering.

And after you've weighed all your options, we hope you drive off in the Honda Prelude.

HONDA
We make it simple.

management system and its underlying philosophy is an expression of fundamental corporate purpose. Unlike Western enterprise, the primary purpose of which is seen as profit maximization, the Japanese enterprise has as its principal objective the creation of wealth, adding value through production. Since from added value must come all rewards to capital and labor, the "social partners" in the Japanese enterprise share a common interest in maximizing its performance.

Workers, managers and shareholders are all duly informed through regular detailed statistical reporting services of the performance of their company, as well as of the performance of competitors at home and abroad, in terms of value added in production. All recognize that if they take more than their share from the wealth created, either someone else must pay or the strength of the enterprise will be dissipated as a result of inadequate investment. Also, there is common recognition that higher investment in new technology and new production facilities means higher productivity, higher value added, and ultimately higher rewards to all members of the enterprise in an equitable sharing of the results of future performance.

To attain this objective, resort to external organizational power is unnecessary. Since there is general agreement and shared interest in the purpose of enterprise, company unions dedicated to those same purposes are at once adequate and desirable as appropriate instruments of communications and negotiation between managers and workers. Workers, as well as management and shareholders, have a stake in quality control and in the mechanization of production, and will voluntarily devote much of their own time and effort to those ends without immediate or direct compensation. Not surprisingly, therefore, today's labor union leaders in most companies are likely to be found among tomorrow's managers.

The implications of such an enterprise system for the electronics industry is immediately apparent. Not only does the resultant cooperative management mode assure higher quality and productivity, which provide the fine cutting edge of competitive power, but it provides the appropriate environment for maximum expression of creative forces within the enterprise. Innovation, therefore, flows naturally, producing a steady stream of new electronic products. Moreover, since rapid technological diffusion within the enterprise is essential for maximal wealth creation, application of microelectronics where savings of labor, energy, materials, space or time are possible, meets little resistance. Quite to the contrary, reason dictates that it be promoted vigorously as a common good.

At the same time, since higher value-added increases the tax take, the state has a vested interest in enterprise efficiency and rapid technological innovation. It is also true, of course, that the gross national product is nothing more than the sum of the value-added by all enterprises, which means that there is an identity of public and private interest in enterprise performance.

Japanese business leaders are quick to cite the quality of their domestic, home market as an element in their disciplining for competition in markets overseas. *To this effect, Michiyuki Uenohara, senior vice president of National Electric Co., Ltd., testifies:*

"The Japanese masses are of a very high quality. Their average level of education is the highest in the world. Consequently, competition in our home market is severe. This is a vital factor in the marvelous development of technologies for consumer electronics in Japan."

To the same effect, Hideo Sugiura, executive vice president of Honda Motor Co., Ltd., declares:

"If there is any gap between the auto industry of Japan and those of the United States and Europe, it may be attributable to differences in how much thought carmakers give to what consumers expect of cars. The technological gap is negligible. The key point is which area of existing consumer demand manufacturers choose to zero in on and care to try to meet.

"Nowadays it is quite popular to use robots in car production. But I don't think the use of robots is the secret of high productivity in the Japanese auto industry. If the robot is the answer for high productivity, it would be easy for General Motors and Ford to catch up with the Japanese—all they need is to invest in robots and bring them in on factory floors in large numbers. It's not that simple, however.

"Our philosophy of technology puts needs above anything else—technology is subordinate to needs. Needs here means not only users' needs but also those of workers and dealers. We must always be careful to see that technology does not have its way and disregard needs it is supposed to serve.

"Booming electronics and computers are meaningless unless technology in these areas is directed at clearly defined needs. Yet advanced technology tends to be toyed with in the absence of such considerations.

"This viewpoint is important particularly for us manufacturers of automobiles—products for use by a large number of people."

There is nothing unnatural or contrived in the manifest cooperation between government and business in the joint promotion of the electronics industry. The resultant higher value-added is necessary for the attainment of the respective purposes of both. And, as it is recognized that higher value-added by industry is the prerequisite of higher wages, there is a broad consensus among all social partners of the inherent wisdom of devoting all needed resources to the rapid development of the electronics industry.

This broad consensus gives to the continuing electronics revolution in Japan its special quality, and its dynamic thrust which makes possible much greater efficiency in innovation as well as in production.

In one respect, however, Hiroshi Watanabe, executive managing director of Hitachi, Ltd., finds the industry-university-government consensus lacking:

"The edge Japanese industry has over the rest of the world at the moment boils down to high productivity and reliability—not only in electronics but also in other areas. On these two points, we are definitely in the lead. If we go one step inside, however, Japanese ability seems still very shallow—as in the design of and basic research on circuits and subsystems, for example.

"By basic research I mean to include areas which belong to sciences, such as materials and properties of matter, which provide significant support to technology development. One area of industry that sustained the high economic growth from mid-1950's through mid-1960's was nuclear energy, and its growth was the result of the nuclear fission science that was developed in the mid-1930's.

"Thus, also, quantum mechanics was the science that emerged in the mid-1920's; it provided the foundation for electronics that made its debut around 1945, developing into today's prosperity.

"These cases show that it takes 20-30 years for a science to develop into a big industry. So when we look ahead for an age beyond electronics, the first important thing is to identify sciences which will be supporting the post-electronics era.

"Unfortunately, Japanese universities today keep themselves in many respects within the tradition of the Meiji Era. At present there is not a single university in Japan that has facilities for college and graduate students to make LSI's. This is in sharp contrast with American institutions where students can design and make LSI's as they like and use them for their experiments.

"Electronics is a hot area, and the VLSI is drawing keen attention today. Yet, our universities are quite indifferent about providing help to basic research in these

THIS IS WHAT WE GET FOR NOT FOLLOWING THE CROWD.

A camera with a combination of features you won't find on any other 35mm SLR in its class.

A camera so extraordinary that in just a few short months it has become one of the best-selling in America.

More than a camera. An achievement. The Minolta XG-M.

What kinds of features make a camera this outstanding?

Optical precision for sharp, clear pictures.

Convenience features like automatic exposure to make taking pictures as easy as point, focus, shoot.

Important features like metered manual and exposure override that let you give full rein to your creative instincts.

Electronic features to help prevent mistakes.

Handling features like a built-in, textured grip shaped to fit the curve of your hand.

Even professional features like the optional motor drive that automatically advances the film. One frame at a time. Or at a blistering 3.5 frames per second to capture fast action. Like a touchdown pass rifling into the end zone.

And when you want to unleash your imagination, there are more than 45 Minolta computer-designed interchangeable lenses to help you do it. Along with a full system of accessories.

The Minolta XG-M is only the latest in our over 50-year tradition of creating cameras that are not simply cameras, but outstanding achievements.

See it at your Minolta dealer. To find the XG-M, just follow the crowd.

WAIT 'TIL YOU SEE HOW GOOD YOU CAN BE.

MINOLTA

For more information, write Minolta Corporation, 101 Williams Drive, Ramsey, N.J. 07446.
In Canada: Minolta, Ontario, L4W 1A4.

Product appearance and/or specifications are subject to change without notice.
© 1981 Minolta Corporation.

fields. This fact demonstrates more than anything else how shallow Japan's base in electronics technology is. University professors seem to be scoffing at the manufacture of LSI's as something for craftsmen, not a serious subject of academic learning. As compared especially to America, academia and technology are too removed from each other in Japan.

"The government's contribution to the advanced research that breaks ground for practical development is much too small. According to the white paper on science and technology, 60-70 percent of development costs for a project commissioned and subsidized by the state is provided by the government in the United States, but the proportion is reversed in Japan for a similar project, with only 30 percent provided by the government. We at Hitachi spend 100 billion to 120 bilion yen a year for research; of this sum, government subsidies and entrusted public money amounts to less than 1 billion yen. Financial assistance from the state for research and development is meager."

MARKET SIZE, STRATEGIES AND STRUCTURES

If men, money and enterprise are critical ingredients in the development of the Japanese electronics industry, fierce competition at home and abroad has been the force fashioning corporate strategies and structures to obtain full advantage of these basic assets. Although mainstream economic theory suggests that capital intensiveness, large fund requirements, and rapid technological change which characterize the industry should lead to oligopolistic market structures, this has not been the case in the Japanese electronics industry. Rather, the nature of industrial groupings in Japan and the rapid diffusion of electronic technology in advanced, newly industrialized and developing countries throughout the world have made the global electronics marketplace the scene of increasing competition.

Essentially a large-scale technology, requiring the mobilization of massive resources for development, electronics requires optimal markets for fullest utility. The market for most electronic products is, perforce, inherently global, and the application of basic electronic technologies will tend to seek those products for which there is the highest income elasticity of demand in the global marketplace. To compete in this market, large-scale organization is imperative, not only to assure maximal economies of scale and learning in production or to sustain the heavy outlays necessary for continued innovation, but because global structures are necessary to sell the output of large production units and to manage information flows which enable the fine-tuning of innovation and production to catch the full force of market demand on each successive product upswing.

Leading Japanese electronics manufacturers are, therefore, characteristically large-scale enterprises. Although there is a vast number of smaller electronics firms in Japan, most of these are subcontracting production which is inherently small-scale and has been less conducive to mechanization in the past. Others fill special niches, some in global markets which require custom or batch labor-intensive production technologies. In the main, however, major electronics firms in Japan are more diversified and highly integrated than their competitors abroad.

This is especially remarkable in semiconductor manufacture. Japanese firms competing in this sector are significantly larger in total sales and assets than their U.S. counterparts, roughly two to four times larger than Texas Instruments and Motorola, and much larger than National Semiconductor, Fairchild and Intel in both sales and assets.

But, as important as corporate size and the extent of diversification and vertical integration are for funding capital needs, these structural features have other equally important advantages. Not only do they greatly facilitate financing of capital-intensive microelectronic and computer operations, they also speed the process of technological diffusion and enhance the reactive mechanisms which keep technologi-

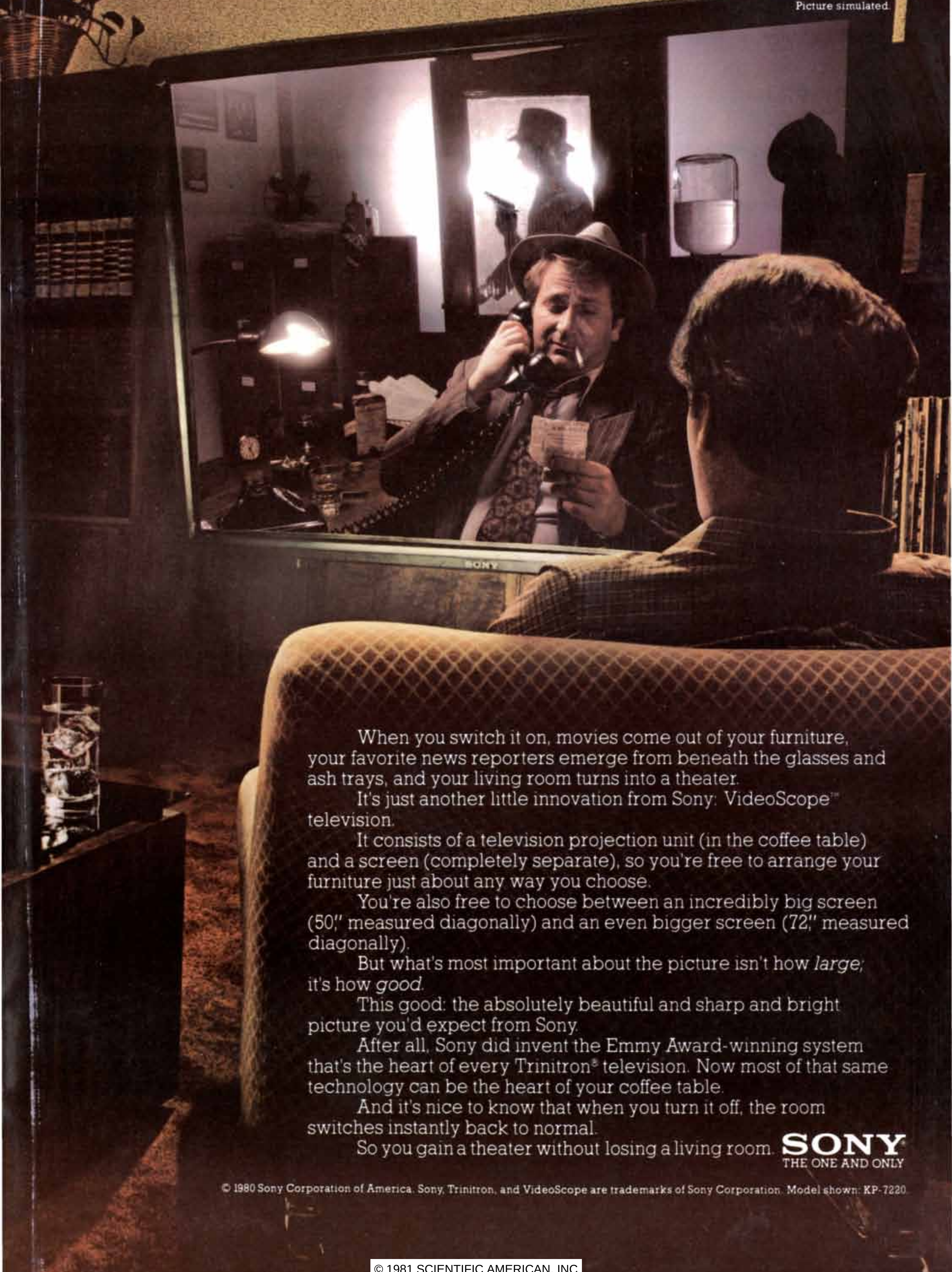
OUTLOOK FOR CONSUMER-ELECTRONICS PRODUCTION IN THE 1980s.

Units: in thousands Value: in billions of yen		1976	1977	1978	1979	1980	1981	1983	1985	1990	Average annual growth rate (%) 1985/1978	1990/1985	1990/1978
Color TVs	Units	1,115	987	888	941	930	940	950	950	950	1.0	0	0.6
	Value	7,681	7,008	6,173	6,180	6,175	6,430	6,897	7,315	8,480	2.5	3.0	2.7
B/W TVs	Units	540	534	505	392	361	320	307	280	180	⊖ 8.1	⊖8.5	⊖ 8.2
	Value	1,035	1,064	973	689	650	595	610	596	452	⊖ 6.8	⊖5.4	⊖ 6.2
Tape recorders	Units	2,961	2,919	2,793	2,863	2,809	2,752	2,609	2,476	2,256	⊖ 1.7	⊖1.8	⊖ 1.8
	Value	5,044	5,079	4,716	4,305	4,354	4,431	4,540	4,680	5,166	⊖ 0.1	2.0	0.8
Combination stereos	Value	3,919	4,076	4,137	3,700	3,912	4,299	4,937	5,575	6,741	4.4	3.9	4.2
Radios	Units	895	909	793	514	420	371	349	335	314	⊖11.6	⊖1.3	⊖ 7.4
	Value	590	526	467	277	233	212	212	216	234	⊖10.4	1.6	– 5.6
Car audio	Units	1,871	1,879	1,957	1,953	2,044	2,067	2,096	2,099	2,022	1.0	⊖0.7	⊖ 0.3
	Value	1,868	1,994	2,056	2,144	2,353	2,477	2,713	2,931	3,438	5.2	3.2	4.4
VTRs	Units	29	76	147	215	306	408	583	715	701	25.4	⊖0.4	13.9
	Value	607	1,310	2,113	2,881	3,978	4,814	5,888	6,464	6,996	17.3	1.6	10.5
Video discs	Units					17	25	90	140	190		6.3	
	Value					145	198	625	820	950		3.0	
Other	Value	1,700	1,531	1,407	1,624	1,768	1,912	2,142	2,319	2,632	4.3	2.6	5.5
Total value of consumer electronics		22,408	22,538	21,970	21,800	23,568	25,368	28,564	30,916	35,089	5.0	2.6	4.0
(Domestic supply & export)		(14,079)	(14,625)	(13,627)	(14,770)	(15,008)	(15,710)	(17,591)	(18,182)	(18,677)	(3.9)	(0.5)	(2.2)

Source: Japan Electronics Industry Development Association


See all you imagine.
FUJI
VIDEOCASSETTE
T-120
VHS
Fine grain
BERIDOX
FUJI
VIDEOCASSETTE
L-750
β type
Fine grain
BERIDOX
Suddenly, the dream is real.
Color as vivid as your imagination. Images that capture a richness of detail rarely seen before. Beauty, as fresh as tomorrow, today.
Fuji videocassettes. In all the most popular Beta and VHS lengths.
For all there is worth seeing. And all there is yet to see.
fuji
© 1981 Fuji Photo Film U.S.A., Inc., Magnetic Tape Division, 350 Fifth Avenue, NY, NY 10118.


Tonight, turn on the coffee table.
SONY


Picture simulated.
SONY
When you switch it on, movies come out of your furniture, your favorite news reporters emerge from beneath the glasses and ash trays, and your living room turns into a theater.
It's just another little innovation from Sony: VideoScope™ television.
It consists of a television projection unit (in the coffee table) and a screen (completely separate), so you're free to arrange your furniture just about any way you choose.
You're also free to choose between an incredibly big screen (50," measured diagonally) and an even bigger screen (72," measured diagonally).
But what's most important about the picture isn't how large; it's how good.
This good: the absolutely beautiful and sharp and bright picture you'd expect from Sony.
After all, Sony did invent the Emmy Award-winning system that's the heart of every Trinitron® television. Now most of that same technology can be the heart of your coffee table.
And it's nice to know that when you turn it off, the room switches instantly back to normal.
So you gain a theater without losing a living room.
SONY
THE ONE AND ONLY
© 1980 Sony Corporation of America. Sony, Trinitron, and VideoScope are trademarks of Sony Corporation. Model shown: KP-7220.

cal change in tune with market demand. These in turn have the synergistic effect of priming the processes of downstream product innovation, thus providing an immediate in-house market for new microelectronic devices, new materials, new automated machinery and new software.

Thus "mechatronics" has emerged in Japan as a new field of technological management. Committees within enterprises, within industry associations, and within and among government agencies, systematically examine the possibilities for application of new generations of microelectronics throughout the broad but shrinking range of mechanical devices. Once again, the results tend to be synergistic. On the one hand, added impetus is given to the market for new microelectronic devices, making possible earlier market entry and more rapid return on investment, which in turn make continuing technological change the more feasible and timely. On the other hand, the broader and more rapid introduction of microelectronics in mechanical equipment saves vast amounts of energy, materials, manpower, time and money—all resources in scarce supply. The effect is to reduce cost, improve functions and enhance the reliability of end products—with all this means for competitive power in world markets for sophisticated high value-added products.

Japan's new command of microelectronics technology goes to the enhancement of the value of those products themselves. For Asahi Optical Co., Ltd., Minoru Suzuki says:

"We aspire to make the camera more of a precision and multi-functional machine. To achieve this, electronic circuits must be used abundantly. There are fears that multi-functional cameras may be costly and be more cumbersome operationally. But pursuit of greater convenience is a human instinct which will never be satisfied."

Shigeo Ono, general manager of the camera designing department of Nippon Kogaku K.K., says of the electronics in present and future Nikon cameras:

"From the viewpoint of camera-making, the electronics technology in which we are interested concerns such areas as fixed image sensor, LSI's as fine electronics, and hybrid technology for a very high degree of integration. Sooner or later, a microcomputer will be fully utilized in a camera. There are already some such cameras, but they are being produced only on a minor scale and for limited purposes, such as advertising. If a microcomputer is to be built into a camera, the existing technology is not enough to accommodate peripheral mechanisms and circuits into compact space. Therefore, the time when camera-makers buy packaged IC's from IC makers will soon be over. I mean that the IC's used in cameras will mostly be customized in the near future.

"Installing various elements within a compact mechanism is an area in which Japan has advantages. Such compound technology may be considered peculiar to Japan. Japanese dexterity in manual work has a lot to do with this."

With microelectronics Olympus Optical Co., Ltd., has developed an entirely new kind of "camera," the ultrasonic microscope that looks into the insides of things. Tsunesaburo Watanabe, managing director of the company, says:

"The ultrasonic microscope was made primarily for use in research work by average users. Ultrasonic waves differ from electromagnetic waves in that they arise from mechanical vibrations of matter. An image shows up as a result of interaction between them and waves of matter. What these images have to tell us is itself a new subject of study. In the case of semiconductors, a coating of aluminum gives off a strong reflection which makes it totally impossible for an optical microscope to see anything below the coating. But, using an ultrasonic microscope, you can see below the aluminum coating. An optical microscope provides information concerning the surface, while, by contrast, an ultrasonic microscope can present information concerning the inside."

Microelectronics has, of course, brought Japan at least abreast of Switzerland in the world watch market. Hideaki Yasukawa, director of Suwa Seikosha Co., Ltd., says:

"When Seiko first marketed quartz watches over 10 years ago, we joined Intercel of the United States to develop semiconductors. At that time we imported hundreds of thousands of semiconductors from the United States. But these days all are supplied domestically. Japan's semiconductor technology has become advanced enough and domestic products are easier to get.

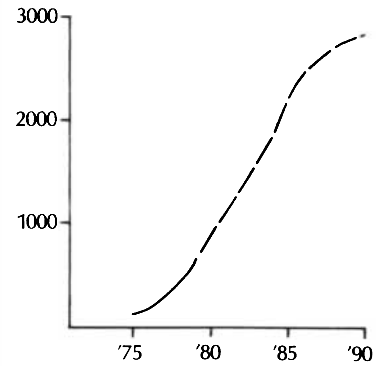
"The key item is the semiconductor. Technology in this area is at the top level in Japan and the United States. Other important things are crystal oscillators, materials for super-small magnets (used for motors) and condensers. Japan's technology is advanced in all these areas.

"You see: a watch consists of a part that moves things and generates motion and a part that controls these moves and motions. The latter part also used to depend on mechanics and therefore was not reliable enough. It is into this part—the equivalent of the brain in a man—that electronics has been introduced. The accuracy of watches thus has been ensured.

"Now the VLSI technology opens up unlimited possibilities. To give an extreme example, it may go as far as putting everything conceivable into such a tiny mechanism as a watch. But the question is whether average consumers can make full use of such watches. Will they not be too complex and made only for the sake of the technicians' own pleasure?

"Should that happen, consumers may turn their backs on them. Such a possibility represents a problem we may face in making adequate use of VLSI's. Technology assessment will become necessary. If additional functions increase difficulties for users, the product will not be welcomed. Even among existing watches, some digital types are faulty in this respect."

FACSIMILE EQUIPMENT PRODUCTION IN THE 1980s.
(in billions of yen)

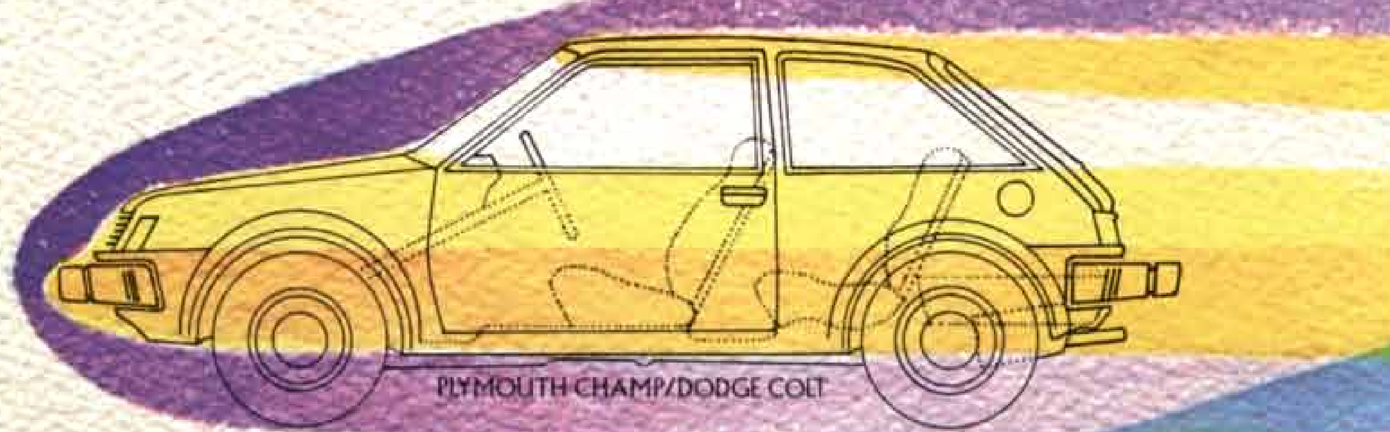


Source: Japan Electronics Industry Development Association.

Not only does this spur investment in advanced electronics at home, but, with the increasing sophistication of technology and markets and given the mounting forces of protection, major Japanese electronics companies have shifted their strategies to increasing investments in overseas production facilities. Following a decade of investment in offshore facilities in low-labor-cost developing countries, after 1975, Japanese consumer electronics equipment makers began manufacturing color television receivers in the United States and Western Europe. More recently, since the end of the 1970's, some of these same firms, along with communications equipment and computer manufacturers such as Fujitsu and NEC, have begun to move integrated circuit manufacture to both advanced and developing countries to better meet the demand of local markets and avoid trade conflicts which were experienced over

MITSU

INNOVATION FROM THE INSIDE OUT.



PLYMOUTH CHAMP/DODGE COLT

Mitsubishi puts high technology to work to bring you the kinds of cars and trucks you want.

For example, Mitsubishi builds the car rated #1 in gasoline fuel economy in the size America wants most, the subcompact.

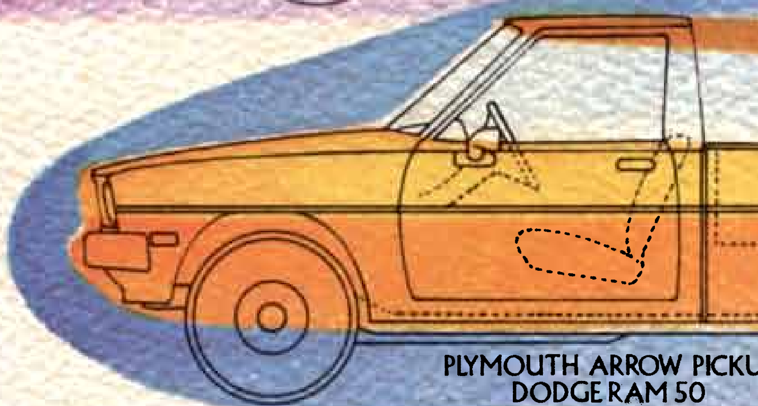
Mitsubishi also builds a peppy small pickup with plenty of power, payload *and* personal comfort.

Then, for those who want a luxury sports sedan, Mitsubishi builds one with standard high-technology luxury features not available in any other car.

No maker of cars and trucks is more aware of the practical needs of drivers in the 1980's. Yet Mitsubishi designers and engineers realize that people also want performance, handling and comfort.

At Mitsubishi, our answer is to put high technology to work—with innovation from the inside out.

MITSUBISHI MOTORS



PLYMOUTH ARROW PICKU
DODGE RAM 50

EXCLUSIVE MCA-JET SYSTEM

Standard on all Mitsubishi-built vehicles, engine design features a third, or "jet," valve injects an extra swirl of air into the combust chamber to provide lively performance wh allowing us to exceed today's tough emissio standards. Nothing like it exists in any othe car or truck.

EXCLUSIVE SILENT SHAFT ENGINE

Standard on Mitsubishi-built pickup truck and sports sedan, this engine features two counter-balancing shafts which rotate in opposite directions to cancel out noise and vibration inherent in conventional four-cylinder engines, for a smooth, quiet ride.

MITSUBISHI'S HIGH-TECHNOLOGY CARS AND TRUCKS. SO

color television receivers.

Early this year, Japan's four largest semiconductor manufacturers—Nippon Electric, Hitachi, Toshiba and Fujitsu—began strengthening their production operations in the U.S. to cope with the coming VLSI era. In July, Mitsubishi Electric revealed that it, too, was considering erecting a VLSI production facility in the U.S., as a joint venture with Westinghouse Electric. In each of these facilities, production will focus immediately on the 64K dynamic RAM, consolidating and adding to the substantial power the industry has already obtained in the market for this new generation of devices and setting the stage for the timely introduction of the first generation of VLSI devices.

Similarly, in Western Europe, these major semiconductor makers are following the strategies and experience developed in color television manufacture with manufacturing facilities strategically located for optimal results in regional markets. NEC, Fujitsu and Hitachi have all finalized plans for European production and will soon be producing 64K dynamic RAMs to strengthen their position in the UK and throughout the Continent.

In sum, the Japanese electronics industry has strong structural and strategic advantages calculated to assure a position of leadership in world markets for a wide range of microelectronic devices and the new generations of electronic products which will emerge during the remainder of the 1980's and in the 1990's. Major Japanese firms have the necessary size to manage effectively and competitively large-scale electronic technologies. Their degree of diversification and vertical integration is matched by few competitors throughout the world. And they have the added power which comes from appropriate global strategies and structures in virtually all product areas.

To a very large degree, the success of Japanese electronics manufacturers depends on this global outlook, supplemented by an implicit if not always explicit commitment to adapt both strategies and structures to the realities of world markets in order to optimize available material and human resources. As markets change, technology changes to meet new needs, new resources are developed and experience gained. Strategies and structures are being altered accordingly to optimize new realities.

Optimal advantages from large-scale electronics technology are clearly not to be obtained through national technological independence or superiority, but by the development of global objectives pursued through what Dr. Michiyuki Uenohara, senior vice-president of Nippon Electric and long-time general manager of Central Research Laboratories at Nippon Electric, likes to call "symbiotic cooperation."

Within an increasingly interdependent technological system, firms and nations are challenged to find and develop new modes of cooperation which will permit full exploitation of production economies. Ultimately, of course, the main stimulus for this internationalization of production is competition. But competition and symbiotic cooperation (within enterprises, national markets or the world marketplace), rather than being opposites, are very much two sides of the same coin. The very nature of electronics technology is such that competition must inevitably beget new patterns of cooperation. The autarchic option is tantamount to technological isolation and economic suicide.

Especially in consumer electronics, Japanese manufacturers have in the past played a key role in setting the mode of both competition and cooperation, and they are likely to continue to do so for the foreseeable future. In response to international competitive forces and national priorities in host countries, Japanese consumer product makers have extended technical assistance to, and invested in, the electronics industries of developing countries throughout Asia, the Middle East, Africa and Latin America. Moreover, through competition, followed by cooperation, Japanese makers have contributed to the restructuring of consumer electronics industries in Western Europe and the United States, with remarkable results in increased productivity, quality and export potential.

But as they have advanced technologically and their weight in the global electronics industrial system increases, Japanese electronics manufacturers have been called upon to devise new organizational approaches which, while assuring the benefits of cooperation, are designed to achieve a higher level of cooperation in world markets, including, of course, the Japanese market itself.

This will become steadily more important since, for the reasons advanced above, the Japanese electronics industry will continue to progress more rapidly than most others in advanced countries. Social and economic institutions, cultural values, and a pervasive recognition of economic dependence all conspire with geopolitical imperatives to assure a remarkably high structural adaptability to the opportunities of technological change and contingent changes in comparative advantage in the marketplace.

Since industrial structural change is likely to be more rapid in Japan than in other advanced countries, where ossification of structures is induced by social trends toward oligarchies and by the reduced capability of governments to manage economic and social policies, it follows that the focal point of innovation and invention will continue to shift gradually to Japan—especially in key technologies such as electronics.

There are clear indications that this shift is well under way. While R&D investment in electronics in advanced industrial countries has generally been declining in recent years, it has been rising steadily in Japan. More important still, during this period an increasing share of R&D expenditures by other industries in Japan has been devoted to applications of electronics and electronic capital equipment in their respective sectors.

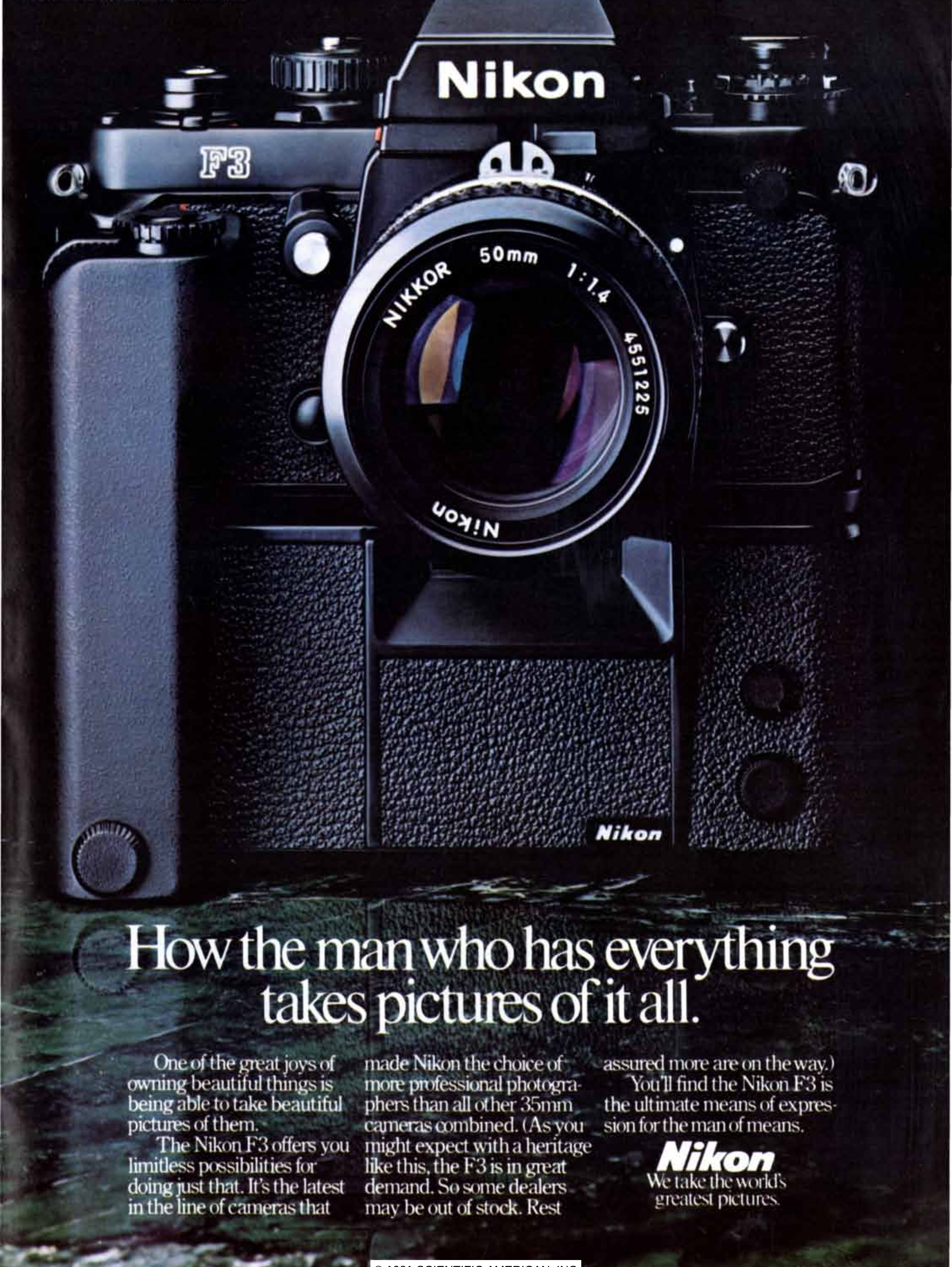
The Japanese automobile industry has taken the world lead in the application of advanced electronics technology to the manufacture and operation of this pivotal artifact of industrial civilization. *Hideo Sugiura, executive vice president of Honda Motors Co., Ltd., noting that Japanese manufacturers have been the first to put programmable robots to work on their production lines, declares:*

"When we put robots to use, we do not simply mean to replace human workers with robots. We first of all divide work into what can best be done by man and what is better left to robots. We then calculate how many workers' jobs can be taken over by robots, by putting together a workload to be done by robots. Efficiency never rises if robots are simply placed where men work. We give serious study to ways of making the best use of robots, in regard to arrangement of space and procedure of work. Introducing robots to the work floor after such careful examination and study makes a difference not only in the amount, but also in the quality, of the work done. If we establish a lead in quality, competitors behind will find it difficult to catch up."

For Mitsubishi Motors Shinji Seki, managing director, reports:

"At our Okazaki plant, all stages of production are thoroughly labor-saving—from press lines to body welding, painting and assembly. These areas are equipped with the most modern facilities and integrated in a linear layout. Thanks also to use of robot welders and painters, 75 percent of the process is automated. The plant is capable of turning out four types of passenger cars in an integrated production system.

"The current monthly output is 16,000 vehicles (which means one car is turned out every one and a half minutes), but surprisingly, a plant of such capacity has no warehouses for parts. All the parts are delivered from subcontractor factories at


Nikon
F3
50mm 1:1.4
NIKKOR
4551225
Nikon
Nikon
How the man who has everything takes pictures of it all.
One of the great joys of owning beautiful things is being able to take beautiful pictures of them.
The Nikon F3 offers you limitless possibilities for doing just that. It's the latest in the line of cameras that made Nikon the choice of more professional photographers than all other 35mm cameras combined. (As you might expect with a heritage like this, the F3 is in great demand. So some dealers may be out of stock. Rest assured more are on the way.)
You'll find the Nikon F3 is the ultimate means of expression for the man of means.
Nikon
We take the world's greatest pictures.

scheduled times. Parts made by distant suppliers are first transported to a depot near our plant. Smooth production control is possible, thanks to cooperation from these suppliers.

"When it comes to installing electronics in the automobile, however, the pace has been slow primarily because of high costs and reliability problems.

"For example, quartz watches brought about cost reduction of crystal oscillators and this enabled us to install quartz clocks in cars. Since quartz can function as a timer, it can also be used for flashers and alarm devices. The process to this point has been considerably slower than the application of electronics in other industries. Use of microcomputers in cars is still limited to control of fuel injection aimed at pollution control. A lot is yet to be done.

"The future direction is focused on two goals. One is greater safety and pollution control to meet various regulations, which are getting tighter, as in exhaust control and anti-skidding. Fuel economy is also a factor. The other is greater comfort and convenience, as in air conditioning, and a system of what may be called service-free mechanism for conducting fine adjustments when the car's mileage increases, or the possibility of disorder or breakdown emerges."

Hiroshi Takahashi, executive vice president of Nissan Motor Co., Ltd., sees some acceleration in the building of electronic technology into the automobile:

"Electronic control of fuel consumption, which at the moment is confined to gasoline engines, will spread in the future to diesel engines as well. Also, if electronic control of automatic transmission is made possible, automatic cars can be kept in a most fuel efficient condition all the time.

"Next, safety devices such as brakes with an anti-skid system, are still very simple and leave ample room for improvement. In the future, electronics is expected to make it possible for a car to detect an obstacle and warn the driver or even brake automatically. Detecting an obstacle is very difficult, though. Steering is another area which will be electronically controllable. This involves a concept of adjusting the response of the steering wheel in accordance with varying conditions under which it is turned. For example, when you are driving at a high speed, a wheel with too light a response is dangerous, because it may turn too quickly and excessively. Suspension and related parts can also be electronically operable.

"In addition, voice synthesis and recognition are of significant importance. Another requirement is streamlining of wiring, for which optical communications will be convenient.

"If voice recognition is perfected, you simply say 'wiper,' when you want to move wipers. It also enables the driver to use a car telephone (which only recently has been put into practical use in Japan, but for safety reasons, the driver is legally prohibited from using it while the car is in motion).

"It is certain that electronics will play a big role in car technology in the future."

Masatoshi Morita, senior managing director of Toyota Motor Co. Ltd., declares:

"Application of electronics ought to be made in such a way as to maintain balance among three factors—man, automobile, and environment. For example, use of microcomputers to provide the driver with certain information is justifiable if it lessens burdens on him and increases safety. It is unacceptable if information made available is more than necessary and adds to psychological and physiological burdens on the part of the driver.

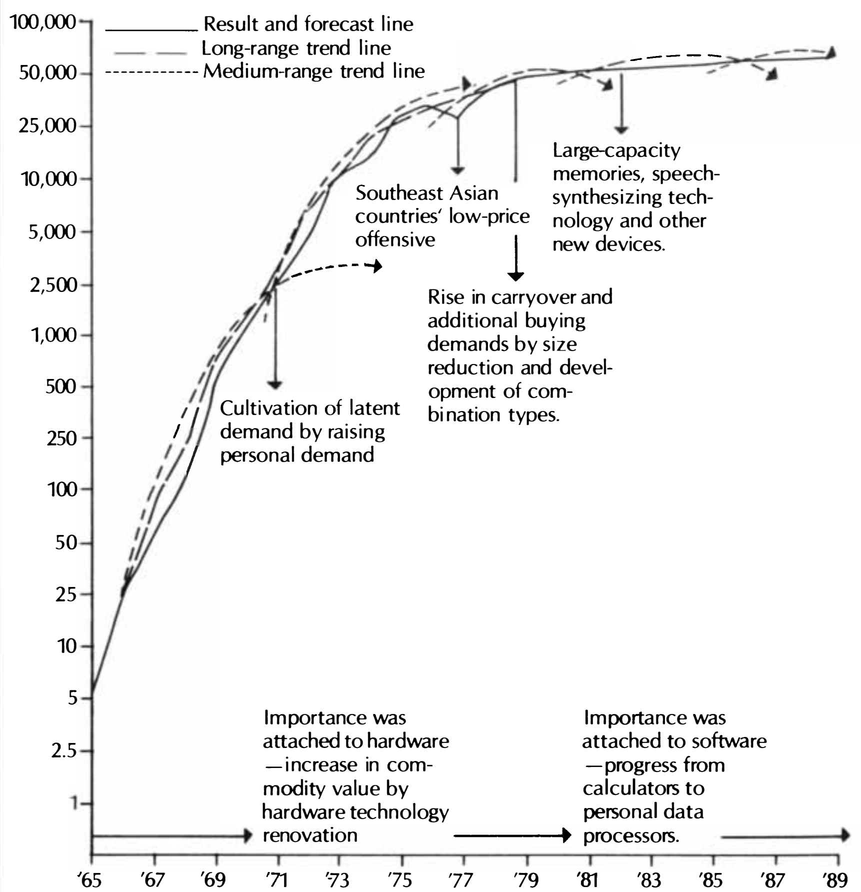
"Electronics to reduce burdens on the driver and increase the fun of driving is correct usage of technology in cars. What is important about car technology is the ends it serves. Car technology is a synthesis of various branches of technology, with electronics being just one of them.

"One problem about electronics at the moment is a shortage of software engineers. Unless software development is stepped up, application of electronics to automobiles will not really make headway."

Although much still remains to be done in the domain of basic research, the consensus necessary for innovation and technological progress is present in Japan to an extent to which it no longer exists in other advanced industrial countries.

Back in the mid-1950's, Japan's industrial policy-makers targeted the 1980's as the decade when the Information Age would emerge in full bloom, with electronics as the basic technology that would power the economy into the 21st Century. Right on schedule, in 1980 the Japanese electronics

EFFECTIVE DEMAND FOR ELECTRONIC COMPUTERS 1965-1990.
(in thousands of units)



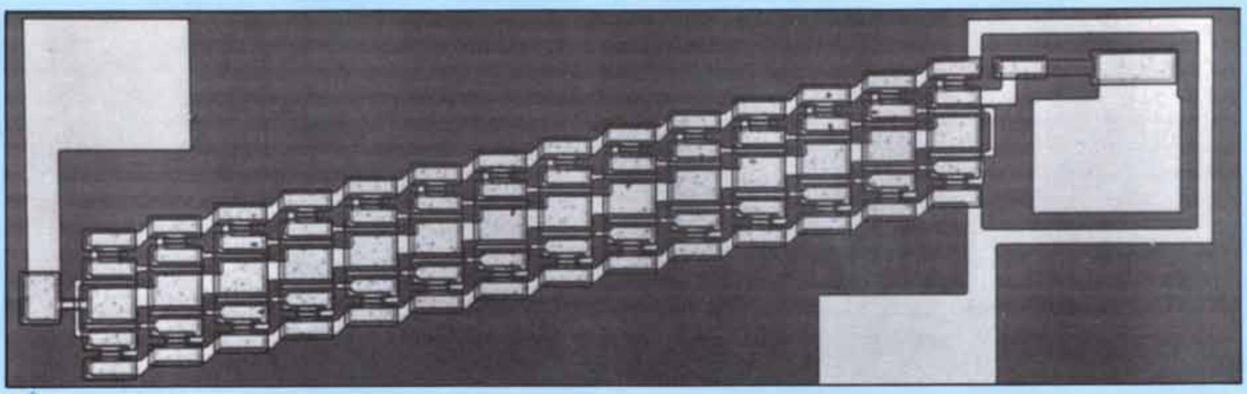
Source: Japan Electronics Industry Development Association.

It's called a HEMT. A High Electron Mobility Transistor. Permitting an electron mobility 30 times greater than a conventional silicon device, and 5 times greater than a GaAs MESFET (at 77K), it will be used in future LSIs to create a new generation of "super computers" – revolutionary systems that will process vast volumes of data at undreamed of speeds. This remarkable HEMT was developed by the company that makes the technology that shapes the future. By Fujitsu.

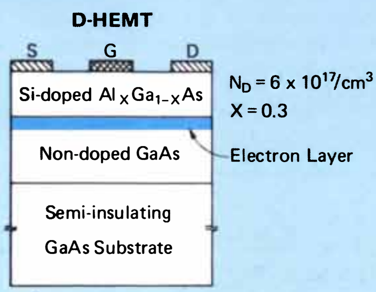
Who is Fujitsu?
Fujitsu is the number one computer maker in the second largest computer market in the world — Japan. Fujitsu is also one of the world's leading manufacturers of telecommunications systems. And a major producer of semiconductors and advanced electronic components. Fujitsu is, in short, a high technology company, with the resources and the proven capacity to develop tomorrow today. It has more than 34,000 employees, operations all over the globe, and annual sales of over $2.7 billion. It is a company that in the future you will be hearing much more about.

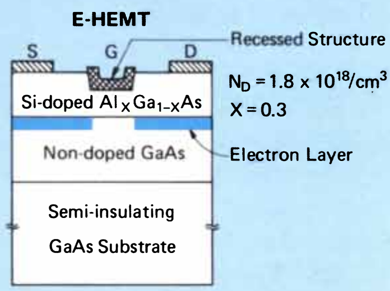
30 TIMES FASTER
than conventional silicon devices

Fujitsu brings you the fastest transistor on earth!



This ring oscillator integrates 27 D-HEMTs (Depletion Mode High Electron Mobility Transistors) and 27 E-HEMTs (Enhancement Mode HEMTs). The D-HEMT, the first HEMT we created to demonstrate theoretical feasibility, features fast speeds but somewhat higher power dissipation characteristics than the E-HEMT, which we developed for practical application. The E-HEMT's power dissipation characteristics are extremely low – two orders of magnitude lower than those of silicon transistors. Which means that very large-scale integration is possible; HEMTs will be used in the high-speed, low-power VLSIs of the coming generation of super computers. HEMTs will also find many applications in the telecommunications field. They make a low-noise amplifier, which is particularly important in microwave technology.
HEMT technology is not yet mature. To date, only small-scale integration, such as that of the ring oscillator, has been achieved. We are working rapidly toward large-scale integration, however, and the future for HEMT looks bright.





The HEMT epilayers, which consist of non-doped GaAs and Si-doped AlGaAs, are grown on a semi-insulating GaAs substrate by molecular beam epitaxy.
The key feature of the HEMT is the heterojunction composed of the two epilayers. The high mobility electrons come from the Si-doped AlGaAs donor. Because the electrons and the donor impurities do not inhabit the same space, impurity scattering is almost totally eliminated. Thus the electron trajectory becomes straight. High speeds result.
At room temperature the switching delay time of the E-HEMT is 56.5 picoseconds. As a result, power dissipation characteristics are one order of magnitude lower than those of GaAs MESFETs. At 77K, the switching delay time is 17.1 picoseconds — making the E-HEMT faster than any conventional semiconductor device on earth.

Computers. Communications.
The thrust of technology into tomorrow.

FUJITSU
Fujitsu Limited·Tokyo, Japan

industry soared into orbit as the leading growth sector in a distinctly new stage of Japan's industrial revolution.

Mitsubishi Electric Corporation is a company that has led Japan into the Information Age, and its managing director, Takashi Kitsuregawa, observes:

"The products of an electrical equipment manufacturer have traditionally been of an energy conversion type. The potential energy of water is converted into electricity and electric energy is converted into kinetic energy which runs machine tools, or into thermal energy which is used to make various things. The emergence of electronics has opened up the new business of transmitting and processing information. In the past, communications were limited to transmission, but today they involve processing as well. Processing even seems to be the area where electronics is the most useful. More recently, electronics is being applied to the function of control on the strength, among other things, of the progress in the technology of the sensors which, although transducers, are aimed at the communication of signals, not the conversion of energy.

"My reservation about our success is that the development of Japanese electronics has been due primarily to information processing and consumer electronics; our industry lags behind foreign countries in the area of pure science. Thus, while our exports of consumer appliances have been running quite strongly, as in the case of TVs and VTRs, the original basic inventions that produced these devices in the first place were all developed in foreign countries.

"Under such circumstances, should Japan be forbidden to use foreign patents, it would be dead.

"The Information Age opens up whole new frontiers for basic research. Building on its strength in existing technologies, Japanese industry must develop a comprehensive scientific base for creation of the next generation of technologies."

As had become increasingly apparent during the 1970's, and as was fully demonstrated by last year's performance, the Japanese electronics industry is not only destined to play a leading role in sustaining Japan's continued economic growth, but has also emerged as an important participant in the revitalization of economic structures around the globe.

Overall, the Japanese electronics industry grew by an impressive 22.9 percent in 1980 as output reached 8.7 trillion yen—almost twice the total production of the industry in 1975 and three times that of 1971. Despite world recessions which followed the oil crises of 1974 and 1978-9, Japan's output of electronic products continued to expand in all sectors at double digit rates, demonstrating a remarkable resistance to downswings in the business cycle. Significantly, too, this growth was sustained in the face of creeping protectionist measures in advanced countries that forced major changes in patterns of the industry's exports. Moreover, this increased home output reflects only indirectly the recent rapid extension of manufacturing to major overseas markets for consumer products and components—a development which over the past five years has entailed global reorganization of production and reallocation of resources on a massive scale.

However, throughout this period, in home production alone, the Japanese electronics industry outpaced the American industry and in 1980 grew twice as fast in all sectors except computers, miscellaneous semiconductor devices and military electronics.

By the most conservative estimates, total output of electronic products will at least double during the 1980's, but indications are that the industry will continue to grow at the pace of 1975-1980 which would mean production in the range of 35 trillion yen.

Japanese IC output, spurred by rapid technological advance and a strong competitive position in world markets for 16K RAM circuits, rose a phenomenal 48.9 percent in 1980, which was almost 2.5 times the growth in U.S. production. Admittedly, the Japanese growth was from a smaller base, and, more important, exports were largely concentrated in the high growth segments of the markets. Total production of all semiconductors by Japan's 11 leading manufacturers rose 37.1 percent to 939 billion yen in 1980 and will surpass the 1 trillion yen level during the current year.

Among the 11 majors, Oki Electric Industry Co. has emerged as the most aggressive, with a new plant on stream in August 1981 expected to produce 300,000 of its 64K dynamic RAMs each month. On the strength of this plan, Oki's semiconductor production is expected to rise 62 percent in fiscal 1985 to 35 billion yen. Hitachi will increase its 64K dynamic RAM production capacity 3.5 times to 700,000 per month in the second half of fiscal year 1981; Toshiba will boost its monthly output of the same device four times to 300,000 and NEC will triple its production to the same 300,000 level. Texas Instruments, the sole foreign-affiliated semiconductor company having production facilities in Japan, has completed a 64K production plant in Miho, Ibaraki Prefecture, from which it will export part of its production to the United States. The Miko works has a rated capacity of 100,000 chips monthly.

More important, the 11 leading Japanese semiconductor manufacturers have scheduled huge plant and equipment investments in fiscal 1981, amounting to approximately 200 billion yen. These new facilities are designed to swing into VLSI production with minimal alterations once the new generation of 256K RAM chips is marketable. Meanwhile, the five leading IC makers are continuing to expand their production facilities in Europe and the United States.

Assuming an average growth rate of 25 percent per annum for the remainder of the 1980's, total IC production in Japan will reach 900 billion yen by 1985 (or approximately the 1980 level for all semiconductor devices) and 1,900-2,000 billion yen in 1990.

In Japan itself, integrated circuit consumption will continue to be spurred by sharp increases in output of end products using large quantities of advanced semi-

DEMAND FOR ELECTRONIC COMPUTERS IN THE 1980s.
(in billions of yen)

Items \ Year	1978	1981	1983	1985	1990	Average growth rate (%)		
						1985/1978	1990/1985	1990/1978
Gross demand (gross supply)	10,212	16,680	20,940	26,000	46,300	14.3	12.2	13.4
Domestic demand	9,516	15,060	18,510	22,670	39,890	13.2	12.0	12.7
Overseas demand (exports)	697	1,620	2,430	3,330	6,410	25.0	14.0	20.3
Domestic production	9,102	15,260	19,260	23,960	42,710	14.8	12.3	13.7
Imports	1,111	1,420	1,680	2,040	3,590	9.1	12.0	10.3

Source: Japan Electronics Industry Development Association.


OH
WHAT
A
FEELING
!
TOYOTA

conductor devices. In 1980, for example, VTR production rose by a dramatic 2.23 times, robot output by 85 percent, and medical electronic equipment by 22.2 percent, while mechanical systems were increasingly replaced by microcomputers in electrical appliances, cameras and automobiles.

Contrary to predictions that consumer electronic production in Japan has substantially reached maturity and will therefore level off, total output of home entertainment equipment rose 28.1 percent to almost 30 billion yen in 1980.

The takeoff of VTR sales was, to be sure, a major factor propelling the continued high growth of this sector. Last year, 11 Japanese VTR producers turned out a total of 4.4 million units, more than twice as many as in 1979. Demand, which has continued to grow at a high rate in 1981, is expected to reach 8.4 million units this year, making VTR a trillion yen business. Not only all major consumer electronics manufacturers, but also camera and sewing machine makers, have entered the VTR market or have announced their intentions to do so.

With this succession of new entrants, Japan's monthly production of VTRs will be increased to a million units by the end of this year. At the same time, to avoid possible trade friction while expanding overseas sales, leading Japanese makers are planning joint production in several European countries. Japan Victor plans to join with AEG-Telefunken, Thomson-Brandt and Thorn-EMI in a regional production scheme. And Matsushita Electric has been negotiating with Robert Bosch to manufacture VTR's jointly in Germany.

The outlook for VTR production for the remainder of the 1980's can be described in nothing less than superlative terms. Production has already far surpassed industry projections, published less than one year ago. Among other developments, introduction of new miniature VTR gear requires a total revision of market estimates. Whereas the industry had previously assumed a diffusion ratio limit of 60 percent—since VTR's were considered to be non-essential to daily life—new miniature models raise the possibility of multiple sets per household.

With the introduction of VLSI, further refinements in VTR functions are expected while size may be further reduced and prices are likely to fall below the original forecast of under $500 as economies of scale and rising competition combine with technological innovation to force production costs downward.

Moves to invest in overseas manufacturing early in the product life-cycle, a radical new departure from past practices of firms in the industry, will change the pattern of production followed by other consumer electronics products. During the 1960's and 1970's, Japanese consumer electronics equipment makers shifted production overseas only after governments had imposed restrictions on imports or production in Japan had lost its comparative advantage due to maturing technology and rising labor costs. In the case of VTRs, however, rapid growth of demand abroad makes it difficult to supply markets exclusively from Japan-based output, and political realities militate against such a strategy.

Thus, overseas production is not likely to impinge on domestic output for the foreseeable future. Rather, since VTR's use from 3,000 to 5,000 electronic parts, many of which are not made abroad, increased production abroad will be an additional boon to semiconductor and other component output in Japan.

But VTR production was by no means the sole stimulus to the consumer electronics sector in 1980, nor will it be alone among the rapid growth products for the remainder of the 1980's. Tape recorder production rose 31.5 percent, in terms of value, over the previous year's level due to new stereo recording and playback features and the application of high value-added components such as microcomputers. For similar reasons, output of component stereo sets also rose 25.8 percent during the year.

Meanwhile another home-entertainment product, the videodisc, makes its entrance from off-stage. Whether it will collide in competition with the VTR or be accepted in parallel (because the disc proper is so inexpensive), two kinds of videodisc are aimed by their makers head-on at each other. Speaking first for the needle-activated disc, Toshiya Inoue, senior managing director of Victor Co. of Japan, Ltd., says:

"We cannot yet predict whether the videodisc is going to succeed in the marketplace. Whether it will become popular or not depends on software. In Japan, the popularization of television was touched off by the wedding of Crown Prince Akihito and by pro wrestling. For videodiscs to be accepted widely by the public, it is essential to develop software which only videodiscs can make the best use of.

"The hardware is already completed. Our schedule is to put it on the market in October in Japan, in January next year in the United States and in June in Europe. In the beginning, the optical and needle types will coexist while competing, and the better one will gradually overwhelm the other and survive.

"Electronics is already playing a significant role in audio records. First of all, it is electronics that picks up delicate sounds from the record, and technology for processing of such a mechanism must be on a high level. The videodisc may be said to be VLSI technology itself. While the disc measures as much as 30 centimeters in diameter, precision required is to the order of 0.5 micron—something surprising even by semiconductor standards."

Speaking for the laser-activated "optical" disc, Teruhiko Isobe, director of Pioneer Electronics Corporation, says:

"Generally speaking, the feature of the optical type is the reliability of discs. Since a surface containing signals of several tenths of a micron is tracked, scouring it with a needle could cause wear. The consequences of wear may not be serious because the needle runs on a very smooth surface, but the optical type may have greater assurance of reliability.

"The great question is software: what is available on the disc. Software for the time being will center on old movies and television programs but they will not be really appealing. Development of software that best suits videodiscs is the future problem. We plan to set up a subsidiary to develop and market such software."

Even production of color television sets rose by 18.1 percent in volume to a record-breaking 11.7 million units, despite protectionist measures in some key markets and apparent saturation of the market in Japan. In addition, output rose abroad, where the industry now operates 45 manufacturing or assembly facilities. In the U.S. alone, subsidiaries of Japanese manufacturers accounted for 30 percent of total color television production, or 3 million sets, bringing the total worldwide output of the industry to well over the 15 million mark.

Thus, ironically, protectionism abroad seems to be good for the Japanese electronics industry. Indeed, the main effects of protectionist measures in the U.S. have been to give American consumers the privilege of paying more for television receivers, to force the Japanese industry to be more efficient through extensive automation and to hasten the pace of rationalizing production on a global scale by major Japanese makers.

Protectionism has also served to speed the processes of technological change, giving to the Japanese consumer electronics sector a renewed vitality, catapulting production into a new stage of high growth. In the past, leading Japanese consumer electronic products—particularly transistor radios, monochrome TV receivers and color TV receivers—were manufactured with technology originally imported from abroad and sold in world markets in direct competition with the original U.S. or Euro-


maxell XLII-S
135m / 440ft.
maxell.
Epitaxial
VIDEOCASSETTE
HG T-120
HIGH GRADE
HIGH CHROMA HIGH QUALITY PICTURE
VHS
maxell.
Epitaxial
VIDEOCASSETTE
HG L-500
HIGH GRADE
MADE BY PERFECTIONISTS
FOR PERFECTIONISTS.
FD1
maxell
Floppy Disk

JAPANESE TECHNOLOGY TODAY

pean makers. Now the leading consumer electronic products—as well as advanced microelectronic devices and industrial electronic equipment—manufactured by Japanese industry are original, home-grown products of Japanese technology.

The immediate effect has been to give Japanese manufacturers a major lead in world markets at the outset, when new products are first introduced, with competition confined largely among Japanese makers themselves. In the case of VTR, Japanese makers currently account for 90-95 percent of total world output, with the Matsushita VHS format accounting for most of that output. As yet there are relatively few signs of anxiety abroad simply because there is virtually no production to be protected.

But an increased tempo of product diversification and innovation has been accompanied by a third major strategic response of Japanese manufacturers of consumer electronic products to rising protectionism. A concerted effort has been made to diversify markets, with resultant rapid increases in shipments to Asian, Latin American and European countries. By 1980, Asia was the largest market for Japanese color television sets, accounting for 33.2 percent of total exports, or 1.5 million units.

In parallel with the diversification of export markets, production has undergone a major relocation. Manufacture of products losing international price competitiveness has been moved to developing countries from which both the Japanese and world markets are increasingly supplied. And, beginning in 1974, manufacture of products which have evoked trade friction has been shifted to import substitution production in the United States, regional supply production in Western Europe, and export-oriented production in developing countries.

Throughout the 1980's, domestic and overseas market expansion for audio and video equipment is expected to continue strong, with videodisc sales mounting sharply after 1982. Also, by 1985, personal computers and word processors are expected to become important consumer durable items, with growth rates determined by cost reductions achieved through VLSI application and production volume attained as a result of the global marketing potential of major manufacturers.

The success of the Japanese electronics industry's multi-directional strategies and the emerging global structures is the stuff of which epoch-making industrial change is made. As a result, Japan has entered the Information Age as the epicenter of the world's consumer electronics industry and is better equipped than any other country to play a leading role in the diffusion of sophisticated information technologies from the factory, to office, to home.

And the timing could hardly be better. As the boom in home computer usage gathers momentum, the Japanese electronics industry is positioned to move into the vanguard of this vast new market. Since the development of VLSI integrated circuit technology, virtually every major Japanese consumer electronics equipment and computer manufacturer has introduced personal computers to the market with an explosive burst of manufacturing and marketing zeal reminiscent of the advent of electronic calculators a decade ago.

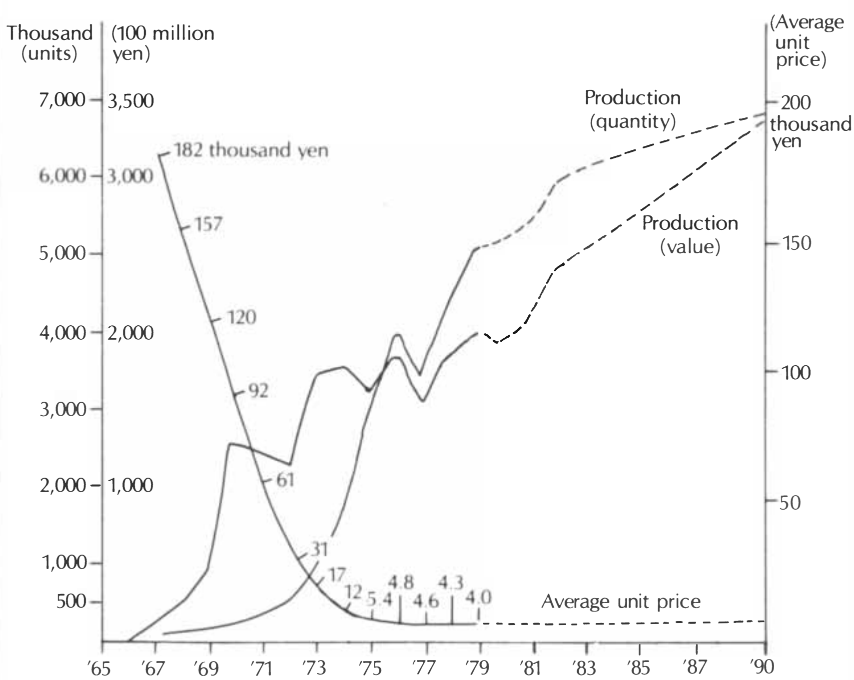
NEC, the Japanese leader in microcomputers, expects to sell 120,000 personal computers in 1981—ten times its 1979 sales volume. Not all of these sales will be made in the domestic market, of course. NEC has already exported 5,000 of its best selling model to the U.S. market in the first half of this year and expects to sell 15,000 here before year's end.

Following rapidly on the heels of NEC, six more leading manufacturers—Canon, Hitachi, Oki, Sharp, Toshiba and Matsushita Electric—are entering the U.S. personal computer market this year. Industry sources estimate that the U.S. market for personal computers has already surpassed the $2 billion level, with annual sales of 500,000 units. And by 1982, the U.S. Department of Commerce predicts, Japanese manufacturers will account for 30-40 percent of that market. If this forecast proves accurate, sales by Japanese makers in the U.S. market next year will skyrocket seven times over the expected 30,000 units to be shipped in 1981. And this is but a foretaste of expected expansion of this market for the remainder of the decade. Nothing better illustrates the magnitude of the Japanese electronics industry's potential for flexible and rapid response to market opportunities created by the incipient generation of VLSI technology.

Just two years ago, Japanese personal computer makers were barely able to hold 20 percent of their own domestic market in the face of U.S. competition. By 1980, Japanese personal computer sales accounted for 80 percent of the home market. In 1981, Japanese manufacturers entered the U.S. market, and within a year's time they are expected to capture a substantial share of that market in the face of fierce competition from already established suppliers.

The explanation of this dramatic shift in market share is quite clear. The potential market for personal computers is great and global, and Japanese electronic equipment makers are better positioned than others to take advantage of this new potential. Not only is the mass production technology required for production of personal computers, the strong suit of the Japanese electronics industry, but the extensive global mar-

ELECTRONIC CALCULATOR OUTPUT AND PRICES, 1965-1990.



Source: Japan Electronics Industry Development Association.

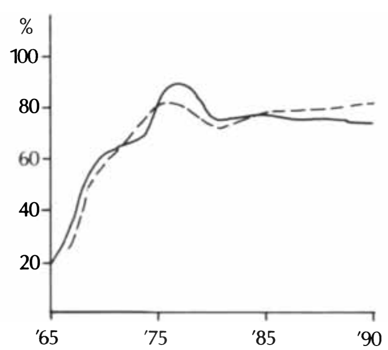

OLYMPUS
OM10
MANUAL ADAPTER
AUTO-S
F.ZUIKO
1:1.8 f=50mm
OLYMPUS

keting networks developed for the sale of calculators and other office equipment are tailor-made for a rapid takeoff in the race for worldwide market share. Moreover, Japanese semiconductor manufacturers that supply the integrated circuits for personal computers have the multiple advantages of massive IC output capabilities, vertically integrated production from chips to final products, and the lower costs which accrue from moving rapidly down the learning curve as production increases over time.

The industry's strength is not confined to mass-produced consumer products, however. In fact, since 1978, production of industrial electronic equipment has surpassed output of mass-produced consumer electronic products, reflecting the steadily growing demand at home for automation and rationalization of both factory and office operations. By 1990, industrial electronics output—led by double-digit growth rates of computers and peripheral equipment, office automation equipment and industrial robots—is expected to be more than double that of consumer electronics.

Since the oil crisis, the use of computers in manufacturing has increased sharply, resulting in a rapid proliferation of robots, automated production systems and unmanned warehouses. In 1980, designated by the Japan Industrial Robot Association (JIRA) as "Robot Diffusion Year One," production of robots increased 80 percent in value over the previous year, adding 19,900 new units to Japan's burgeoning robot park.

ELECTRONIC CALCULATOR EXPORT RATES, 1965-1990.



—— Export ratio in terms of quantity (export/production).

– – Export ratio in terms of value (export/production).

Source: Japan Electronics Industry Development Association.

Of the rapid progress in the diffusion of robots in Japanese industry, Hiroshi Watanabe, executive managing director of Hitachi Ltd., says:

"Among the kinds of work left to human beings today, those consuming the most manpower are for sorting things out by recognizing anything that can be read. There is strong social demand for labor-saving in this kind of work—as in the mechanical processing of postal codes, office computers in place of card punchers, or the assessment of a photograph of an eyeground or the evaluation of an electrocardiogram. Success in pattern recognition is bound to have a far-reaching impact on the entire society.

"As a manufacturing concern, we have never experienced trouble about the introduction of robots in our factories. They have always been welcomed. In view of the fact that factory jobs consist of repetitive work which ignores human nature, use of robots that take over such work is a humane consideration. I don't see why this is opposed in Western countries."

JIRA and the Nomura Research Institute predict that Japan's robot market will continue to grow at least 20 percent annually, which will mean installations of 55,000 new units worth 290 billion yen in 1985, and as many as 94,000 units valued at 600 billion yen in 1990.

COMPUTER'S PROGENY: THE ROBOTS

Employing the latest developments in microcomputers, lower cost robots appeared on the Japanese market this year, opening a vast new potential for their use by medium- and small-scale manufacturing companies. The cheapest robot available on the Japanese market previously sold for about 15 million yen, which is beyond the reach of most firms. Now Sankyo Seiko Manufacturing Co., an electronics parts maker, has developed a new compact assembly robot priced as low as 3.6 million yen, and Okamoto Seisakusho Co., a steel furniture maker, is offering a relatively inexpensive model selling for just 3 million yen for use in loading metal sheets on press machines.

Although the fastest growing segment of industrial electronics production, robots still represent a small proportion of total output. Computers, of which robots are but the progeny, accounted for 42.1 percent of last year's production of industrial electronic equipment and 14.9 percent of the overall electronics industry.

Despite substantial growth in computer output in 1980, the 15.2 percent growth rate was considerably lower than that of the larger U.S. industry. Japanese computer output is still almost entirely directed toward the domestic market, which remains dependent on imports to a relatively high degree. In 1980, computer imports rose to 16 billion yen, exceeding exports by 25 percent.

Indications are that this picture will change radically in the 1980's, however, as Japan's major computer makers have been steadily expanding their share of overseas markets with increasing investments and new links with foreign partners. As a result, exports of computers rose 31.1 percent in 1980 and are expected to grow at higher rates in coming years.

Shoichi Ninomiya, director and general manager of Fujitsu Ltd., says of their campaign to claim a place in the van of computer progress:

"For 15 years, we have strived to catch up with IBM and at long last, somehow, we have achieved the target. But my honest feeling is that our victory is only on the surface. Commercialized products represent only the tip of an iceberg of accumulated technologies. IBM's hidden technological resources are still no small consideration. We have many things to learn from them.

"We may say semiconductors are what computer technology is all about. At present, we are in an age of silicon so far as semiconductors are concerned and are headed for a higher level of integration. This means that semiconductors in themselves are becoming computers, rather than being parts of computers. This in turn requires a maximum degree of refinement in the original design of LSI's and VLSI's, because it is impossible to modify them later."

Exports of low cost computer printers, which have risen meteorically in recent years, promise to remain strong throughout the 1980's. Japanese makers have been especially strong in the market for printers that cost less than $1,000 in the U.S. Once again, U.S. companies created a market for dot-matrix printers used primarily with personal computers, but began losing market share when production problems and subsequent delayed deliveries gave Japanese makers the opportunity to show their mettle once again in the mastery of high-technology mass production. From virtually zero-base in 1979, Japanese manufacturers such as Ricoh, Okidata, Shinshu Seiko and Itoh were supplying half of all units selling in the U.S. by mid-1981 and were expected to raise their share of the $200 million U.S. market to 75 percent by the end of the year. By 1985, Dataquest estimates this market will be pushing $1 billion and still rising sharply.

A similar pattern can be expected in white-collar, office-function products, such as word processors, especially once Japanese manufacturers overcome some of the remaining technical barriers to low-

JAPANESE TECHNOLOGY TODAY

cost voice recognition systems. Here again, Japanese makers can be expected to excel with their strength in manufacture and distribution of mass products. The key technology in this sector will be mass production with built-in quality control, a field in which Japanese performance is without peer.

THE LANGUAGE INTERFACE

The competitive strength of Japan's computer industry is therefore expected to increase further with the recent lead it has taken in the development of VLSI technology. During the 1960's and 1970's, the initial phase of Japan's information revolution, computers were used mainly in the industrial sector, confined largely to manufacturing purposes and operated mainly by experts, and were limited in application because computers are still designed to use Western languages. But this pattern is now undergoing a radical change. New generations of computers, using advanced semiconductor technology currently available, will be faster, smaller, cheaper and easier to use, being operated by voice instruction. Vast new areas of utility will be opened once the language interface problem has been resolved. Improved versatility and greater accessibility will accelerate diffusion of computer usage to broad new segments of Japanese society.

Japanese computer manufacturers are working hard at teaching computers their language. Among them is Fujitsu, Ltd., for which Shoichi Ninomiya speaks:

"It is an absolute necessity to develop computers to handle the Japanese language. This is difficult if we depend on English for internal processing and use Japanese only for the output. We should use Japanese for internal processing as well. We are doing research on this problem.

"If processing of Japanese becomes possible, however, voice recognition may be easily done, given the phonetic structure of Japanese which consists entirely of a combination of a vowel and a consonant."

For Nippon Electric Co., Ltd., Michiyuki Uenohara declares:

"People using European languages can engage in data-exchange and dialogue with machinery through typing with almost the speed of conversation. But this is impossible in the Japanese language. If oral input becomes possible, the handicap will be totally eliminated. Because of the phonetic simplicity of the Japanese language relative to European languages, oral input will give the Japanese an advantage, reversing the present situation. The world is made in such a way that advantages and disadvantages are always offsetting relationships.

"The ultimate mission of technology is to make up for disadvantages of human beings and society and bring them progress. If it ruins advantages that human beings have, it does not deserve to be called technology.

"A local culture is something that has been developed through the long history of the region and human life itself. It should not be altered because of technology. Rather technology must be altered to fit the local culture. If the Japanese language is abolished for the sake of convenience of usage of computers, the Japanese will be deprived of their identity. I know there is an idea of a world federation (world federalism), and I should hope it embraces the idea of the coexistence of a variety of cultures having different geographical and historical roots. This is a source of charm of the human world.

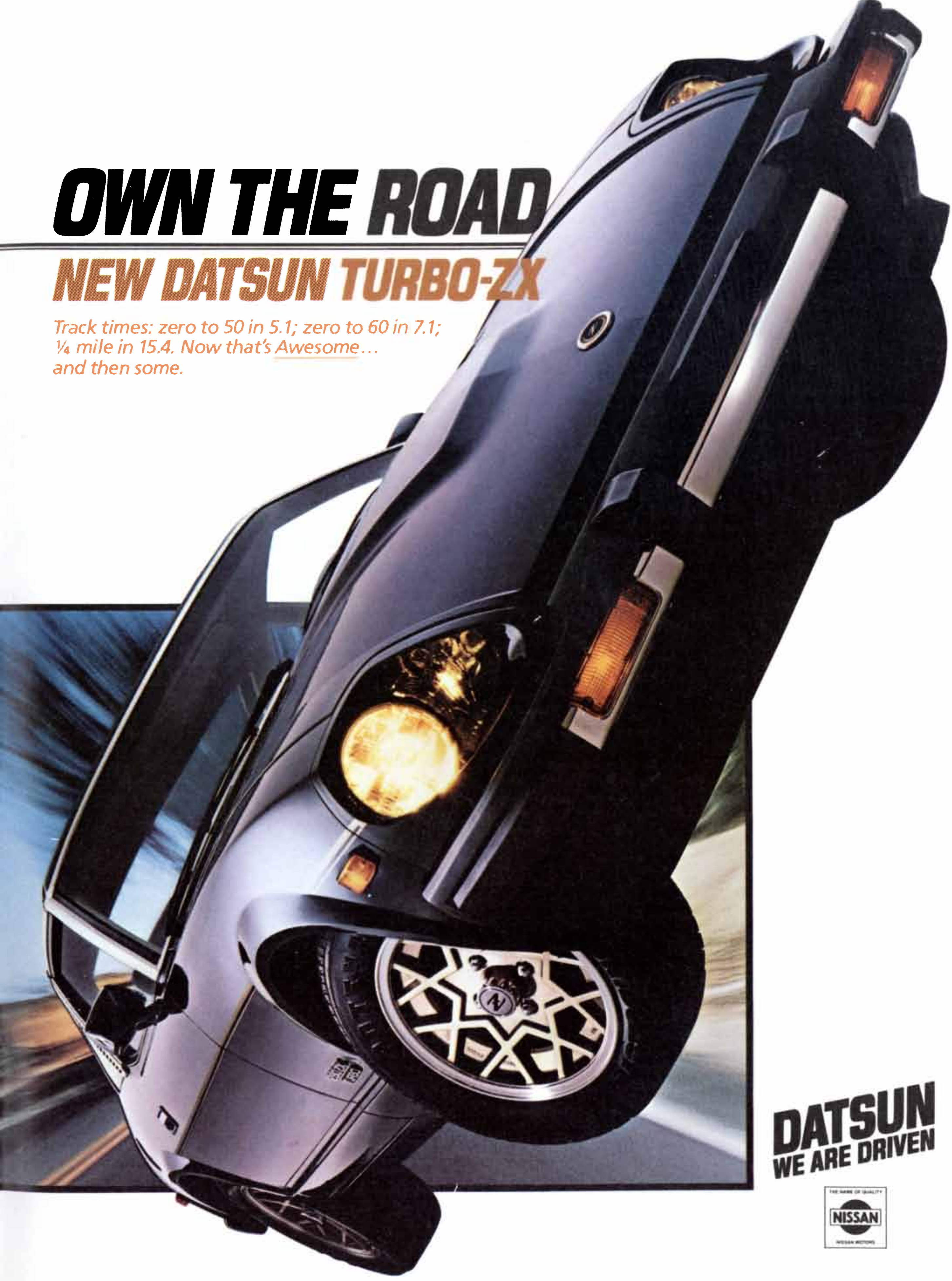
"On this assumption only can technology of Japan's own, with ample creativity, have a true meaning. Up to the present time, scientific technologies have not been really mature. For that reason man has given way to machinery in an attempt to make the best use of machinery and raise the standard of living. In other words, man has had to accept the trouble and sacrifice himself in using machinery. From now on, however, it must be that machinery adapts itself to man, not the other way around.

"Reasons for the recently controversial trade frictions between Japan and the United States may be traced to this problem. Given its economic and social setup, it has been possible for Japan to develop its industries only on the basis of technologies that cater to needs of the masses of consumers, first of all. By contrast, European countries and the United States have been able to do a good deal of business in the limited areas of advanced technologies or high technologies because of the existence of huge military, industry and government procurement markets. For technicians and engineers, advanced and high technology areas are far more exciting to work in than those catering to the masses' needs. But the ultimate objective of technical progress is to enable the masses to make use of and benefit from new technologies. Discovery of a new thing is only the beginning. The essence of technological advancement is to improve the new thing in such a way as to make it economically and socially acceptable to the public. In this reasoning, if Japan limits itself to production technologies and Eu-

THE ELECTRONIC CALCULATOR REVOLUTION, 1965-1990.

Year	Gross production quantity (in thousands)	Ratio to preceding year (%)	Exports (in thousands)	Ratio to preceding year (%)	Imports (in thousands)	Ratio to preceding year (%)
1965	4	—	0.8	—	—	—
1966	25	625	6	750	—	—
1967	63	252	20	333	—	—
1968	163	259	74	370	—	—
1969	454	279	237	320	—	—
1970	1,423	313	731	308	—	—
1971	2,040	143	1,267	173	—	—
1972	3,866	190	2,612	206	9	—
1973	9,960	258	6,366	244	54	600
1974	15,453	155	10,215	160	145	269
1975	30,040	194	25,727	252	88	61
1976	40,426	135	35,192	137	43	49
1977	31,835	79	28,168	80	458	1,065
1978	42,319	133	32,549	116	677	148
1979	49,500	117	36,000	111	1,600	236
1981	52,000	—	38,000	—	2,000	—
1983	60,000	—	47,000	—	2,200	—
1985	63,000	—	49,000	—	2,500	—
1990	66,500	—	52,000	—	4,000	—

Source: Production—MITI "Dynamic Statistics of Production."
Imports & exports—Finance Ministry "Customs Clearance Statistics."


OWN THE ROAD
NEW DATSUN TURBO-ZX
Track times: zero to 50 in 5.1; zero to 60 in 7.1; ¼ mile in 15.4. Now that's Awesome… and then some.
DATSUN
WE ARE DRIVEN
NISSAN

rope and the United States limit themselves to development technologies, both sides lack balance.

"It seems to me that progress of technology requires a sense of balance."

In a related development Matsushita Electric has announced a voice synthesizer incorporated in a single LSI chip. Shunkichi Kisaka, senior managing director, says:

"We feel machines that can talk will personalize and facilitate the use of products for the consumer. The potential for consumers includes home appliances giving verbal instructions, educational games and equipment, monitoring and alarm systems—a variety of home electronic products."

THE NEW WAVE OF COMPUTER DEMAND

In its most recent report, published in June, the Information Industry Committee of the Industrial Structures Council, described in considerable detail the forthcoming explosion of computer use in offices, education, medical care and the home. Anticipating this new wave of demand, which is expected to rise at an annual rate of 40-60 percent in the coming years to become a multi-trillion-yen business in 1990, all computer builders have developed a variety of new office automation equipment, including personal computers, small business computers and word processors.

A pacesetter in this new open field, Toshiba increased its sales of computers by 59.3 percent in 1980 by concentrating mainly on office computers and expects to increase deliveries to the market by another 43 percent in 1981. Mitsubishi Electric also hopes to boost sales of office computers by 30 percent this year, and both NEC and Hitachi have introduced new low-priced desk-top business computers, while Matsushita Electric will introduce a unique portable computer later this year, all in a bid for market share in what promises to be the fastest growing segment of the business.

Japanese language word processors are likely to be another product leader in the office automation market. This year Toshiba plans to sell 2,100 word processors in the home market, up 102 percent from 1980, and a rash of new products have been introduced by Canon, Hitachi, Matsushita Electric, NEC, Nippon Univac, Sharp and Ricoh.

Prospects are that as computer production moves from custom and batch systems suitable for a narrow industrial market at home to a mass market with global dimensions, the Japanese computer industry will in all probability close the growth-rate gap with the U.S. industry by the end of the decade, erasing the last major advantage left to American electronic equipment manufacturers.

Although the American electronics industry is considerably larger than the Japanese industry in terms of total output, as much as 45 percent of its production is geared to special military requirements. The Japanese industry, on the other hand, is totally market-oriented, offering a far wider range of products—ranging from electronic games to super computers and robots—than the more narrow-based U.S. industry. This broad and balanced mix of consumer, office, industrial and medical electronic products provides the Japanese industry with extraordinary flexibility in the marketplace and enables the industry to sustain relatively high R&D expenditures focused on the needs of specific market segments.

Norio Tashima, director of the research division of Minolta Camera Co., Ltd., reflects on the "segmenting" of the Japanese home market for his company's products:

"Competition for the market requires flexibility to meet special demands. Toyota, for example, uses the term 'limited production of a large number of varieties.' The idea is that flexibility of design is increasingly required in meeting demands. This is reflected in the arrangement of production lines, which nowadays tend to be short. If the production line is long, mass production becomes mandatory. So the line is cut short in favor of the so-called flexible manufacturing system which can turn out a wide variety of models.

"Of course, this seems to run counter to the so-called ideal of mass-production and lowest cost. But, in fact, by good management the flexible or variable production line can show competitive costs. We may even call it 'labor saving' in view of the fact that the workers employed are not increasing in proportion to the increase in the variety of models. I think workers involved in production will further decrease in the future if we start product designing which incorporates the entire stages of production. Such a designing system, which will soon become popular, depends on a computer-aided design and simulation.

"Surplus manpower thus made available then will be put into such new areas as microfilm and video equipment into which we plan to branch out in the future. These areas will require more electronics technology, but this will not nec-

JAPANESE OVERSEAS COLOR TELEVISION PRODUCTION

	Hitachi	Matsushita	Mitsubishi	Sanyo	Sharp	Sony	Toshiba
Hong Kong				x			
Indonesia				x			
Malaysia				x	x		x
Philippines	x	x		x			
Singapore	x				x		x
South Korea		x					
Taiwan	x	x		x	x		
Thailand	x			x			
Australia		x		x	x		
New Zealand				x			
Canada		x		x			
United States	x	x	x	x	x	x	x
Argentina				x			
Brazil		x		x	x		
Mexico	x						
Spain		x		x			
United Kingdom	x	x	x		x	x	x
Nigeria				x			

Source: Annual Reports

THE PENTAX LX

Sixty years of research and human engineering have given rise to a remarkable photographic instrument.

The Pentax LX is a photographic instrument of such quality that it will exhilarate you. In your work. In your art.

Integrated Direct Metering means less-than-perfect exposures are a thing of the past.

The LX system's advanced metering system changes shutter speed as light changes —automatically. Even during exposure.

In the automatic mode, the IDM system measures light directly off the film plane. An advanced Silicon Photo Diode computes the initial shutter speed.

During exposure, if there is a light change, the IDM system can adjust shutter speed through an infinitely variable range from 125 seconds to 1/2000th second.

Flash is also integrated with the IDM system. In addition to flash synchronization at 1/75th of a second, the LX offers through-the-lens measured automatic flash control with two dedicated TTL automatic electronic flash units.

The two integrate with the IDM system to read both ambient light and electronic flash at the film plane. The system is so advanced that with sufficient ambient light, the flash won't fire.

The hard-working camera for the hard-working pro.

The LX operates under some of the roughest location conditions. That's because gaskets at all points and panels where controls enter the LX body help prevent the passage of sand, dirt, dust, and moisture. Even the film advance has sealed ball bearings. If you work on location, this is the camera to take to work.

Another in-the-middle-of-nowhere advantage: Even with battery failure, the LX operates at all manual speed settings, from 2000th to 1/75th, plus B.

Here's more than you and your colleagues have asked for:

Six interchangeable viewfinders and nine focusing screens. One is a waist-level, conventional finder and sports finder all in one. Five others let you choose among: built-in diopter correction, waist-level viewing with high magnification, and compact, fold-up waist-level viewing.

The Winder LX, with speeds up to 2 fps, rewinding a 36-exposure cassette in less than 10 seconds.

The Motor Drive LX, to shoot sequential photographs at speeds up to 5 fps, rewinding a 36-exposure cassette in less than 10 seconds.

But enough of features. Now...

...Feel the excitement.

The Pentax LX is available now at your camera dealer's. Get your hands on one, and feel the excitement.

THE PENTAX INSTRUMENT

PENTAX LX

PENTAX 10 million

COMMEMORATING THE TEN MILLIONTH PENTAX CAMERA. PENTAX HAS MADE MORE 35 MM SLR'S THAN ANY OTHER CAMERA COMPANY.

essarily mean development of new technology. Improvement on existing technologies and more of a systematic thinking will result in a major change ten years from now."

At the same time, as the world's most highly integrated electronics manufacturers, Japanese firms have demonstrated a unique capacity for rapid diffusion of new technologies. Vertical integration from semiconductors to end consumer and industrial electronic products enables manufacturers to obtain much faster return on investment in basic integrated circuit R&D, for example, by rapid application of new devices in all products of the company for which they are suitable. Since returns on investment in new technology are generally greater and faster, integrated Japanese electronics manufacturers are able to allocate massive resources to investments in R&D as well as new plant and equipment.

Mitsubishi Electric, thus, regularly spends 12 percent of integrated circuit sales on R&D. And during the recent wave of new VLSI research and development, some makers spent as high as 21 percent of current sales of semiconductors on R&D and another 22 percent on new plant and equipment. This kind of capital commitment, even with a financial system reputed for its capacity to sustain long-term investment, can only be made by widely diversified and vertically integrated manufacturers which can assure the rapid diffusion of all new technology once it has become operational.

Mitsubishi's Takashi Kitsuregawa remains concerned, however, by the thinness of Japan's base in fundamental research and the need to build that base for the renewal of the country's technological initiative:

"In this pioneering task, government agencies and private enterprises must cooperate closely. Each side is as important as the other. It's fine that government agencies lead private industries, but at the same time the private sector should inspire the government. Although it may seem to foreigners that 'Japan, Inc.' is dominated by the government, in reality public and private sectors are supplementing each other. It is in the best interest of the nation that such close cooperative relationship continues in the future.

"On the other hand, however, as much as we rely upon cooperative relationships, we must reckon with the bet that people are inherently disposed to competition. One evidence of this is the fact that when all material needs are satisfied people will still compete for decorations. Generally speaking, moreover, the Japanese do not like to praise fellow members of the group. When a Japanese invention is held in great esteem abroad, it may be given an entirely cool reception in Japan. The Yagi antenna and magnetron are good examples.

"One reason for this is the tendency to put down anybody who is likely to distinguish himself and come above the others. An old proverb says: 'A nail that sticks out is hammered down.' Another factor is insufficient Japanese capability for judgment and evaluation. Perhaps the Japanese have not been psychologically able to afford to take interest in creative aspects, although they have been keen about raising productivity. It also has been more economical to buy technologies from abroad.

"Given the historical fact that the Japanese were an agricultural people, it has been difficult for any individual member of a social group to differ from others. If you look back at history, emergence of unique characters took place at times of upheavals when the society had to depend upon unique brains for survival. On the other hand, when peace prevailed, somewhat eccentric figures were tolerated. On balance, throughout most of history, the Japanese people have had to be average to be socially acceptable.

"Although Japanese productivity and quality control are given a high mark today, the reputation of Japanese goods in the past was very poor. It has taken the Japanese a long time to win worldwide praise for their products. By the same token, if they are allowed a little more time, they surely will come to be recognized for their inventiveness and creativeness as well."

THE ADVANTAGE OF VERTICAL INTEGRATION

At Matsushita Electric, 10,000 researchers and engineers working in 24 research laboratories, beginning with the most fundamental level of materials development and embracing all aspects of the production process, provide technological backup for 110 operating divisions producing a wide range of home appliances, industrial equipment, components and basic materials. As a result, Matsushita Electric is Japan's single largest patent holder with more than 98,745 industrial property rights at home and abroad. To sustain this level of activity, in 1980 Matsushita's investment in R&D exceeded 100 billion yen, adding 8,671 patent and other industrial rights covering 65 countries throughout the world.

Significantly, at Matsushita, diversified and integrated production is a basic organizational and operating principle. Moreover, this basic principle has been given a vitally important global dimension. Typical of the entire Japanese electronics industry, Matsushita Electric allocates production of its broad range of products among some 40 manufacturing divisions at home and 39 manufacturing companies abroad.

Unlike the electronics industry in the United States, where global production is mainly practiced by semiconductor and computer manufacturers, diversified and integrated Japanese electronics firms have developed worldwide manufacturing and sales networks for a broad spectrum of final products, components and materials.

Thus Nippon Electric Co., Ltd., long-established in communications found a natural place in its laboratories for computer technology and its underlying solid-state physics for the development of a new synthesis of computer and communications technology. As Michiyuki Uenohara said recently:

"Combining computers and communications (C & C) was conceived as a form of business best suited for our expansion into new areas, making the most of technological resources we have as a company that started out as a communications equipment maker. It was in 1975 that the concept was put into practical business programs, and in 1977, we came up with a clear-cut expression of 'C & C.'

"First of all, we achieved the ability to put on the market a digital switching system which can serve as a link between the communications and computer networks. Digital switching has made the combination of computers and communications not only possible, but natural. NEC announced its digital switching system in 1977, one year after Northern Telecom did.

"A technical problem that had to be overcome in commercializing a digital switching system was that of continuity with existing analogue devices of telephone communications. Since the switchboard plays an integral role as the control center of a communications system, it has a decisive influence on the mode of operation of the entire system.

"Now communications systems themselves must become increasingly digital. This means they will have greater compatibility with computers which are digital machines. Indeed, the switching system itself is close to being a computer. Multifaceted communications services will become possible. This development will also lead to labor-saving. What's more, a digital switching system is capable of performing services for clerical work which hitherto have been handled by separate computers, such as computation of charges and other accounting services, while simultaneously performing data

communications, image communications or exchange services. In this way, computer technology and switching system technology are beginning to advance hand in hand."

At Hitachi Ltd. microelectronics similarly found its place in a diversified electrical and electronics manufacturing company, as described by Hiroshi Watanabe:

"What we term our electronics sector—computers, semiconductors, electronic parts, measuring instruments and communications equipment—accounts for one third of sales, with heavy electric machinery and home appliances accounting for another third, each. But since electronics is used in almost all home appliances, electronics products' share runs to 70 percent of all when they are included. Among heavy electrical machinery, control panels, for example, are also electronics. Control panels are installed in power generating equipment, and in the process of manufacture of power generators, various devices involving electronics technology are used.

"Hitachi's two major target areas are energy and electronics and one specific challenge we face is to apply electronics to areas where it has been little utilized so far and revolutionize things there.

"Innovation of products which have been in existence already will also be important. To this end, we will raise technicians with perfect command of use of personal computers. Personal computers will play an ever growing role in future product development."

FROM ELECTRONS TO PHOTONS

It is this global organization of highly diversified and integrated manufacturing and marketing operations that gives to Japanese electronics firms the optimal market size and the versatility which will be critical to success in the development of coming generations of VLSI technology. As computers change in basic character, function and design from industrial products mainly used by specialists to mass-produced consumer products, the global structures and strategies of Japanese electronic companies will give them a decided advantage over more specialized American manufacturers. At the same time, the application of VLSI to consumer electronic products will continue to reduce their cost and increase their function, with extensive use of microcomputers strengthening the position of the consumer electronics industry in world markets. Similarly, VLSI technology will proliferate the applications of electronics in automobiles, medical equipment and new generations of robots, and all areas where Japanese industry has already taken an impressive lead.

Beyond the territory already encompassed by command of the electron there beckons for all of these enterprises another frontier, here described by Hitachi's Hiroshi Watanabe:

"Optical technology, centering on laser, has the potential to grow dimensions on a scale with solid-state electronics, involving, for example, application to nuclear fusion. At the moment, its growth has been taking place primarily in the field of electrical transmission, capitalizing on its capability of transmitting large quantities of information at a relatively low cost. The growth, moreover, has been quite rapid.

"Another application area of optical technology which is now drawing keenest interest is linkage of computers and office automation—both quite close to everyday business life. The technology is already in practical use for control of industrial plants and a community information system. In sum, optical technology appears to have great potentiality in such areas as signal transmission from the computer.

"Thanks to the semiconductor laser, the conversion device that transduces signals from electrons to photons can be manufactured very easily and cheaply, and be compact in size. Conversion would be difficult with the gas laser of the past. It will become cheaper and handier in the future. If, moreover, electrical signals with which computers operate now are replaced entirely by optical pulses, op-

JAPANESE TECHNOLOGY TODAY

tical computers and optical switching systems will no longer be a dream.

"Light has various advantages over electricity, like being stronger against noise and faster in speed. I think an optical LSI, as against the LSI we have at present, is a possibility in the future.

"Light is highly advantageous as a transmission medium. The most significant point is that the raw material for LSIs is silicon, which exists on the earth virtually without limit. While electricity requires metal as the carrier, light can do with fiber glass or crystal, which is quite an advantage from the viewpoint of resources and costs. Lighter weight and greater noise-proof capability are also significant. As for elements, electricity uses silicon, while light requires a semiconductor compound of elements in the so-called III, V groups, or one of the elements in the II, VI groups, which are less advantageous than silicon. But since the crucial factor in an electric transmission is weight, light will have greater advantages on balance.

"Whatever the case, optical technology is an extension of electronics and its areas of application will coexist with those of electronics, with neither side totally replacing the other. It should be construed that light is added to the conventional silicon technology, broadening the field of electronics."

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

Aiwa America, Inc.

American Honda Motor Co., Inc.

Canon USA, Inc.

Fuji Photo Film USA, Inc.

Fujitsu, Ltd.

Matsushita Electric Industrial Co., Ltd.

Maxell Corporation of America

Mazda Motors of America, Inc.

Minolta Corporation

Mitsubishi Electric Corp.

Mitsubishi Motors Corp.

Nikon Inc.

Nippon Electric Co., Ltd.

Nissan Motor Corporation in U.S.A.

Olympus Camera Corporation

Pentax Corporation

Pioneer Video Inc.

Seiko Corporation of America

Sony Corporation of America

Sumitomo Metal Industries, Ltd.

TDK Electronics Corporation

Toshiba America, Inc.

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

US JVC Corp.

In addition, we acknowledge the significant contributions to this section by the following distinguished business leaders who were interviewed:

Mr. A. Harada
Matsushita Electric

Dr. T. Ikeshima
Sumitomo Metal Industries

Mr. T. Inoue
Victor Co. of Japan

Mr. T. Inouye
Anelva Corporation

Mr. T. Isobe
Pioneer Electronics

Mr. M. Jablow
Nihon Teksel Co.

Dr. M. Kikuchi
Sony Corporation

Dr. S. Kisaka
Matsushita Electric

Dr. T. Kitsuregawa
Mitsubishi Electric

Mr. A. Kohno
Daiwa Securities Co.

Mr. S. Masujima
TDK Electronics

Mr. T. Matsumura
Nippon Electric Co.

Mr. M. Morita
Toyota Motor Co.

Mr. M. Morita
Matsushita Electric

Mr. S. Ninomiya
Fujitsu

Mr. K. Ogawa
Daiwa Securities Co.

Mr. S. Ono
Nippon Kogaku

Dr. A. Ouchi
Nippon Electric Co.

Mr. S. Seki
Mitsubishi Motors

Mr. S. Shimizu
Toshiba

Dr. M. Shinoda
Fujitsu

Mr. H. Sugiura
Honda Motor Co.

Mr. M. Suzuki
Asahi Optical Co.

Mr. H. Takahashi
Nissan Motor Co.

Dr. N. Tashima
Minolta Camera Co.

Dr. M. Uenohara
Nippon Electric Co.

Dr. J. Ujiie
The Nomura Securities Co.

Mr. A. Watanabe
Systems Formulate Corp.

Dr. H. Watanabe
Hitachi

Mr. T. Watanabe
Olympus Optical Co.

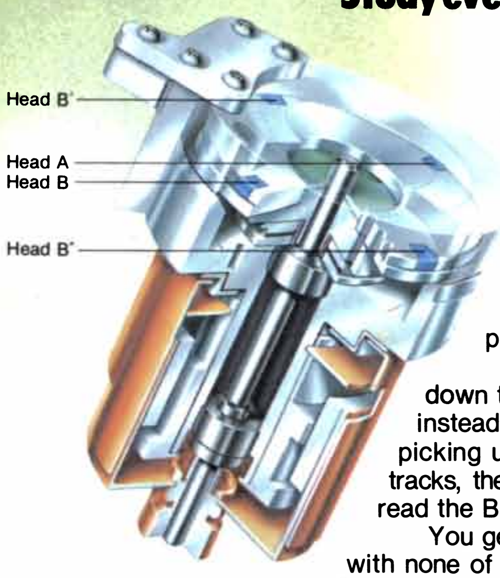
Mr. T. Yanase
Sanyo Electric Trading Co.

Mr. H. Yasukawa
Suwa Seikosha Co.

Special acknowledgement to the officers and staff of Japan Society for their contributions to "Japanese Technology Today." Founded in 1907, the Japan Society is the primary organization in the United States dedicated to enriching and strengthening the ties between Japan and the U.S. by providing a binational forum for the exchange of views and dialogue. For more information about Japan Society write to 333 East 47 Street, New York, N.Y. 10017 or call (212) 832-1155.

Food for Thought

Study every crystal-clear detail at Super Slow and Super Still with Toshiba's unique 4-head VTR.



Faster than the eye can see, churning at almost 100 beats per second, the tiny hummingbird's wings are just a whizz and a blur. But not on a Toshiba VTR.

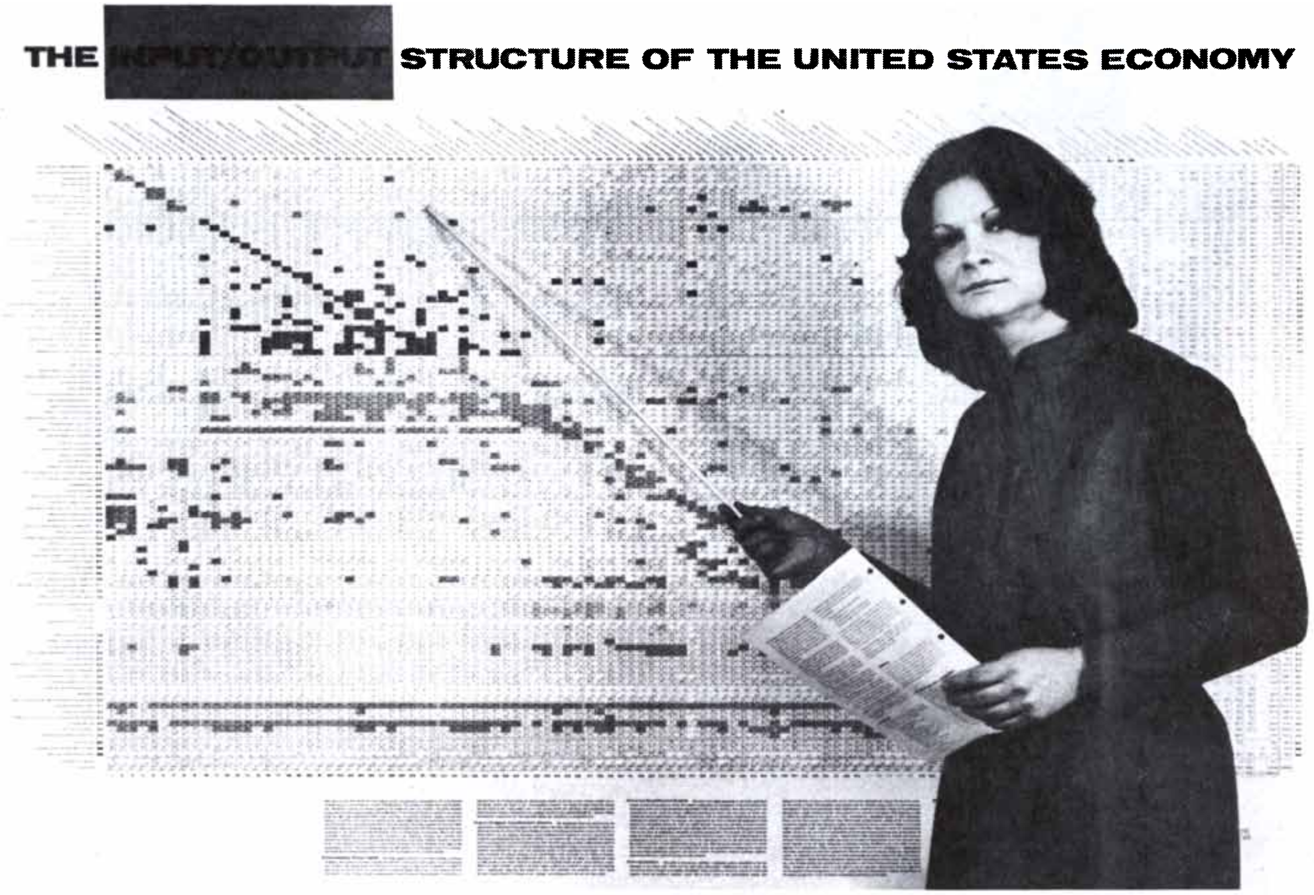
Because Toshiba knew that to catch every movement in pinpoint focus, especially at the tricky modes of slow motion and freeze frame, something out of the ordinary had to be done.

Their extraordinary solution was 4 heads in the video drum. Two (like most other VTR's) to pick up signals at normal playback, and two more (unlike anyone else) for special playback effects only.

So when you slow the action to any speed down to a minimum of 1/30, or stop it altogether, instead of two different A and B heads picking up cross-signals from two different A and B tracks, the two new B′ and B″ heads read the B track only.

You get a razor-sharp image, with none of the jiggles or fuzziness or bar-noise so annoying with most convenient 2-head video recorders. At every video mode, including frame-by-frame reverse and 30X speedier Picture Search. In fact, you get such clear insight into the very nature of things, it may make you pause to think.

Just exactly how can a hummingbird remain still in mid-air while its wings are beating so fast?



WHAT MAKES THE U.S. ECONOMY TICK?

The editors of SCIENTIFIC AMERICAN have prepared a wall chart displaying for the 1980's the Input/Output Structure of the U.S. Economy based on the latest interindustry study from the U.S. Department of Commerce.

The SCIENTIFIC AMERICAN Input/Output wall chart does for economics what the table of elements does for chemistry. It answers at a glance questions about the linkage between the microeconomics of the firm and the macroeconomics of the system; about the web of technological interdependencies that tie industry to industry; about the industry-by-industry direct and indirect consequences of swings in public and private spending; about the impact of change in technology, and about any other topic you can think of. You are rewarded by surprise as well as by confirmation of your hunches. For teaching and practical and theoretical studies, here is a powerful, graphic tool.

In the familiar format of the SCIENTIFIC AMERICAN Input/Output wall charts for the 1960's and 1970's, the wall chart for the 1980's measures 65″ × 52″ and is printed in eight colors. Each of the nearly 10,000 cells in the 97-sector interindustry matrix shows (1) the interindustry commodity flow, (2) the direct input/output coefficient and (3) the "inverse" coefficient. Where the direct input/output coefficient exceeds .01, the cell is tinted in the color code of the industrial bloc from which the input comes. This device, combined with triangulation of the matrix, brings the structure of interindustry transactions into graphic visibility.

A supplementary table displays, industry by industry, the capital stock employed; the employment of managerial, technical-professional, white-collar and blue-collar personnel; the energy consumption by major categories of fuel, and environmental stress measured by tons of pollutants.

The editors of SCIENTIFIC AMERICAN are happy to acknowledge the collaboration, in the preparation of this wall chart, of Wassily Leontief, originator of input/output analysis—for which contribution to the intellectual apparatus of economics he received the 1973 Nobel prize—and director of the Institute for Economic Analysis at New York University.

Packaged with the chart is an index showing the BEA and SIC code industries aggregated in each of the 97 sectors.

SCIENTIFIC AMERICAN

415 Madison Avenue, Dept. I/O, New York, N.Y. 10017

YES, send me the new SCIENTIFIC AMERICAN 1980's INPUT/OUTPUT wall chart—printed in eight colors on heavy paper measuring 65″ × 52″.

I understand my chart(s) will be shipped prepaid, uncreased*, flat-rolled and in a sturdy mailing tube. (Allow 4-6 weeks for delivery)

I enclose a check for $_____for_____chart(s) at $35 each, saving me $5 each in handling charges.

Bill me for $_____ for_____chart(s) at $40 each.

Name____________________

Company____________________

Street Address____________________

City, State, Zip Code____________________

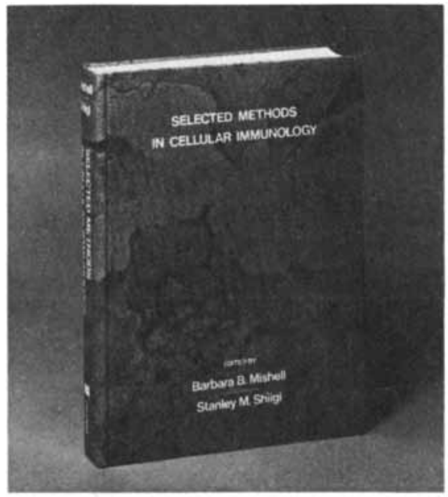
10/81

*Charts shipped international require one fold.

Make check payable to SCIENTIFIC AMERICAN. Please add sales tax for California, Illinois, Massachusetts, Michigan, Ohio and New York.

Selected Methods in Cellular Immunology

Edited by
Barbara B. Mishell, *University of California, Berkeley*
Stanley M. Shiigi, *Oregon Regional Primate Research Center*



"*Selected Methods in Cellular Immunology* by B. Mishell and S. Shiigi is almost all that one could ask from a methods book. . . .This is a book that every laboratory doing or planning to do *in vitro* cellular immunology will want to have, and don't bother finding a spot for it on your shelf; this book will be on the bench being used." —*Cell*

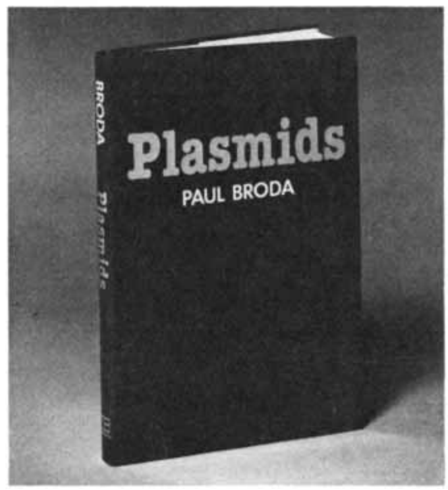
"An excellent book. . . . No laboratory with an interest in cellular immunology should be without it." —*Bioscience*

"An added bonus is that the book is written with such style that it is quite enjoyable to read for general knowledge, as well as for specific procedures."
—*American Society for Microbiology News*

1980, 486 pages, 36 illustrations

Plasmids

Paul Broda, *U.M.I.S.T., Manchester, England*



"An up-to-date, comprehensive and comprehensible guide to the subject was badly needed. . . . Here it is—Paul Broda has provided just the thing, and it is not only newcomers to the subject who will welcome his *Plasmids*." —*Nature*

"This book satisfies a real need for current information on plasmids for university lecturers, as a reference text in libraries and for use in advanced graduate student courses on microbial genetics. The author's writing style makes easy reading and his experimental familiarity with plasmids is clearly displayed." —*Search*

1979, 197 pages, 26 illustrations

Please send me the following books. I enclose payment or credit card information with order. Publisher pays postage and handling.

Copies

_____ *Selected Methods in Cellular Immunology*
Hardbound: 1106 $31.95

_____ *Plasmids*
Hardbound: 1111 $19.95

Guarantee: Examine these books for fourteen days. If for any reason you are not satisfied, you may return them for a full and prompt refund.

W. H. Freeman and Company
660 Market Street, San Francisco, California 94104

Name ______________________

Address ______________________

City/State ______________ Zip ________

I enclose payment with order.
(California residents add appropriate sales tax.)

Charge my ☐ Visa/BankAmericard ☐ Master Charge.

Account Number ______________________

Expiration Date ____________

Signature ______________________
All credit card orders must be signed. SA 61

W. H. Freeman and Company is the book-publishing affiliate of *SCIENTIFIC AMERICAN*.

Neuropeptides

They are short chains of amino acids that are active in the nervous system. In some cases they transmit signals between nerve cells and also serve the body as a hormone

by Floyd E. Bloom

The two major systems that coordinate the activity of cells with the needs of the body have long been thought to function in quite different ways. In the nervous system each nerve cell has been taken to affect its set of target cells at synapses: specialized sites at which a chemical messenger—a neurotransmitter—is released by one cell and received by another. In many instances the neurotransmitter is a monoamine: a substance the nerve cell synthesizes by making minor changes in an amino acid. Sometimes the neurotransmitter is the amino acid itself. In any case its release and reception take milliseconds. In short, such communication is point-to-point and fast. In the endocrine system, on the other hand, each gland cell has long been known to release its chemical product—a hormone—into the circulating blood. The hormone is often a peptide: a short chain of amino acids. The release of the hormone, its circulation in the blood and its influence on its target cells throughout the body takes minutes or hours.

These distinctions between the two systems are now in disarray. For one thing norepinephrine, a monoamine neurotransmitter, turns out to be also a hormone: it is released by gland cells in the medulla of the adrenal gland. Conversely, vasopressin, a peptide hormone, turns out to be also a neurotransmitter: nerve cells in the hypothalamus, a part of the brain, rely on vasopressin to signal other nerve cells in the brain. Today more than a dozen cell-to-cell messengers are known to be capable of relaying signals either between nerve cells or between gland cells and their targets. Typically each messenger is discovered first as a factor: a substance of unknown chemical composition that has a physiological effect such as the dilation of arteries or the contraction of muscle. Often it emerges that the factor is made up of amino acids. Then it emerges that the factor is active in the brain. Thus the factor is revealed as a neuropeptide.

The discovery of a peptide messenger begins with the suspicion that a chemical substance is responsible for the interaction of two groups of cells. For example, it has long been known that if an animal is anesthetized with ether, the cortex of the adrenal gland releases steroid hormones in quantity. If the anterior lobe of the pituitary gland is first removed, no such release is measured. Presumably, therefore, a substance released by the anterior lobe causes the adrenal cortex to act. The substance turns out to be adrenocorticotrophic hormone, or ACTH, also called corticotropin. In a different experiment the blood vessels linking the hypothalamus to the anterior lobe of the pituitary are interrupted. Again the etherization of the animal does not give rise to the release of steroids. Evidently the hypothalamus governs the release of pituitary corticotropin. The substance by which it does so is called corticotropin releasing factor. Its molecular structure has not yet been worked out.

The Process of Discovery

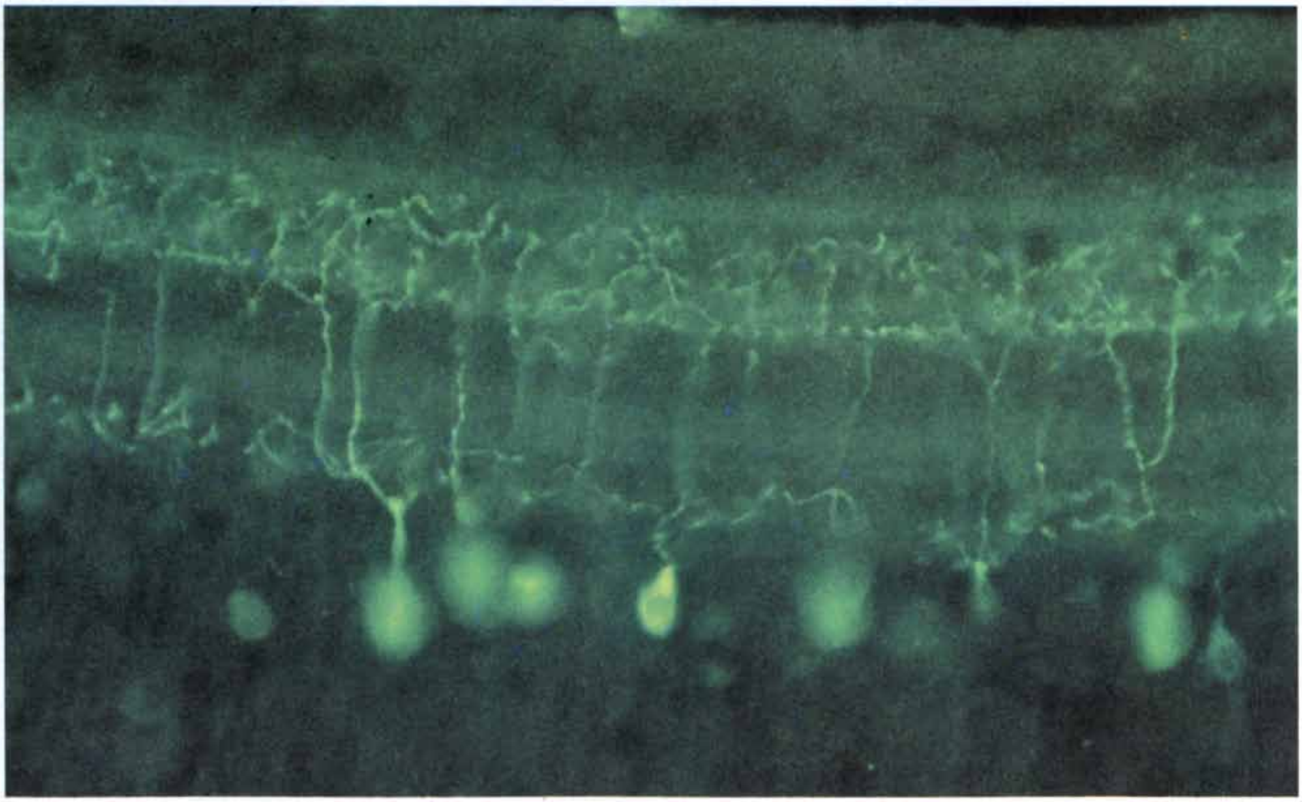
The suspicion that two groups of cells interact by means of a chemical messenger motivates a sequence of experiments that is well established today. The sequence was devised largely by Vincent du Vigneaud of the Cornell University Medical College as he worked in the 1940's and 1950's to identify the hormones secreted by the posterior lobe of the pituitary. It was elaborated in the 1960's and 1970's by Roger Guillemin, who is now at the Salk Institute for Biological Studies, and by Andrew V. Schally, who is now at the New Orleans Veterans Administration Hospital. The objects of their search were the hormones by which the hypothalamus regulates the anterior lobe of the pituitary.

First an extract is prepared from the group of cells that presumably release a factor. This can be done, for example, by homogenizing the cells. The extract is applied to the tissue whose cells the factor controls. The potency of the extract is noted; then the extract can be refined by passing it through a chemical sieve consisting of a gel that selectively filters molecules on the basis of their size or their electric charge. The refined extract is applied to the tissue. With skill, and often with luck, one finds that it now has greater potency than the original extract had. The factor has therefore been concentrated. Further refinements of the extract are tested for potency. Ultimately the factor is purified. In some instances the process begins with extracts from hundreds of thousands of fragments of brain and ends with a few nanograms of an unknown chemical substance.

At this point a different technology is applied to the factor to determine its chemical structure. Suppose the factor is inferred to be a peptide because it loses its biological activity when it is treated with enzymes that cleave peptide chains at the linkages between two successive amino acids. A chemical assay then shows the proportions of the various amino acids. The precise sequence of amino acids remains to be determined. The classical technique is to treat the factor with a number of enzymes each of which cleaves the peptide chain at the linkages between two specific amino acids.

The resulting fragments are collected. Individual amino acids are successively cleaved from each one and identified by properties such as the rate at which they travel in a column of gel that filters molecules according to their electric charge. In this way the sequence of amino acids in each fragment is worked out. Next the order of the fragments themselves is determined by finding overlapping sequences of amino acids in the sets of fragments cleaved from the factor by enzymes that cleave a peptide at different sites. The newest technique removes amino acids from the full peptide chain one by one; a mass spectrometer then identifies each amino acid in the sequence by its weight.

When a sequence has been determined, its proportional content of the various amino acids is matched against the composition of the factor as it has been determined by chemical assay. Then the peptide is synthesized. Today this can be done readily; indeed, a

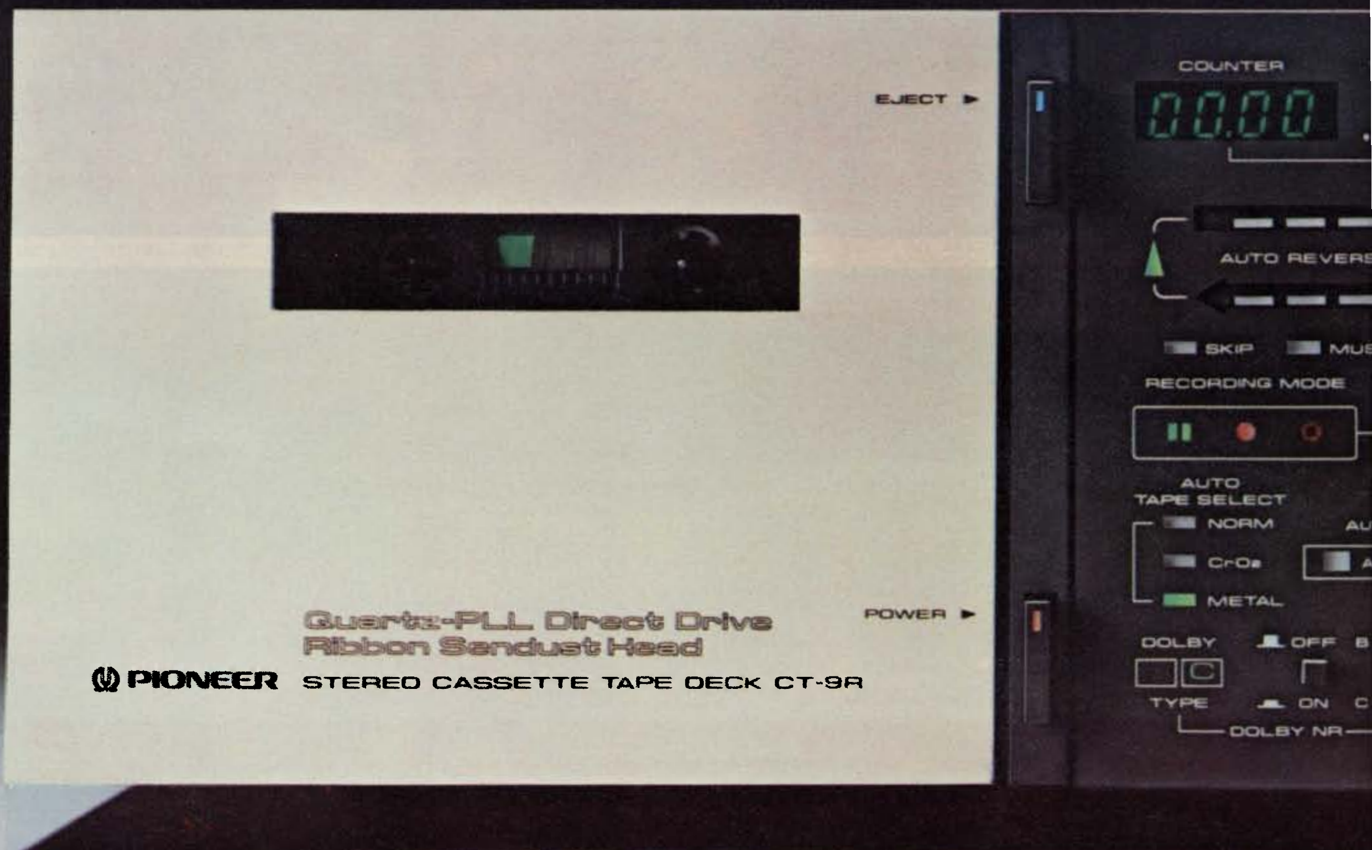


PRESENCE OF ENKEPHALIN in the organ of sight of several species is demonstrated by a technique that tags the enkephalin with an antibody and then tags the antibody with a second antibody to which a fluorescent molecule is bound. The enkephalin itself is a neuropeptide consisting of five amino acids. The upper photograph shows a retinula, or sight organ, of the lobster. The fluorescence marks sprays of cells. Each spray is an ommatidium, a cluster of light receptors some 10 micrometers in diameter. The photograph was made by Jorge Mancillas and Jacqueline F. McGinty of the Salk Institute. The lower photograph shows a cross section of the retina of a chick. Here the cells containing enkephalin are nerve cells, not light receptors. Each cell belongs to the class of neurons called amacrine cells, which make local connections in the retina by means of filaments that the technique also renders fluorescent. Some of the amacrine cells not marked by fluorescence in this preparation turn out to contain other neuropeptides: neurotensin, substance *P*, somatostatin or vasoactive intestinal peptide. The photograph was made by Nicholas Brecha and Harvey J. Karten at the State University of New York at Stony Brook.

High Fidelity for Humans:

NOW WHEN

IT DOESN'T HAV



Anyone who records on tape knows what a pain it is to run out of tape before running out of music.

Pioneer has relieved this pain. Along with quite a few others inherent in the designs of practically all components being built today.

We've done it through a concept we call *High Fidelity for Humans.* A design and engineering idea so far reaching, that for the first time components are as pleasant to live with as they are to listen to.

For example, our new CT-9R cassette deck shows you a digital readout of the precise amount of recording time left on a tape.

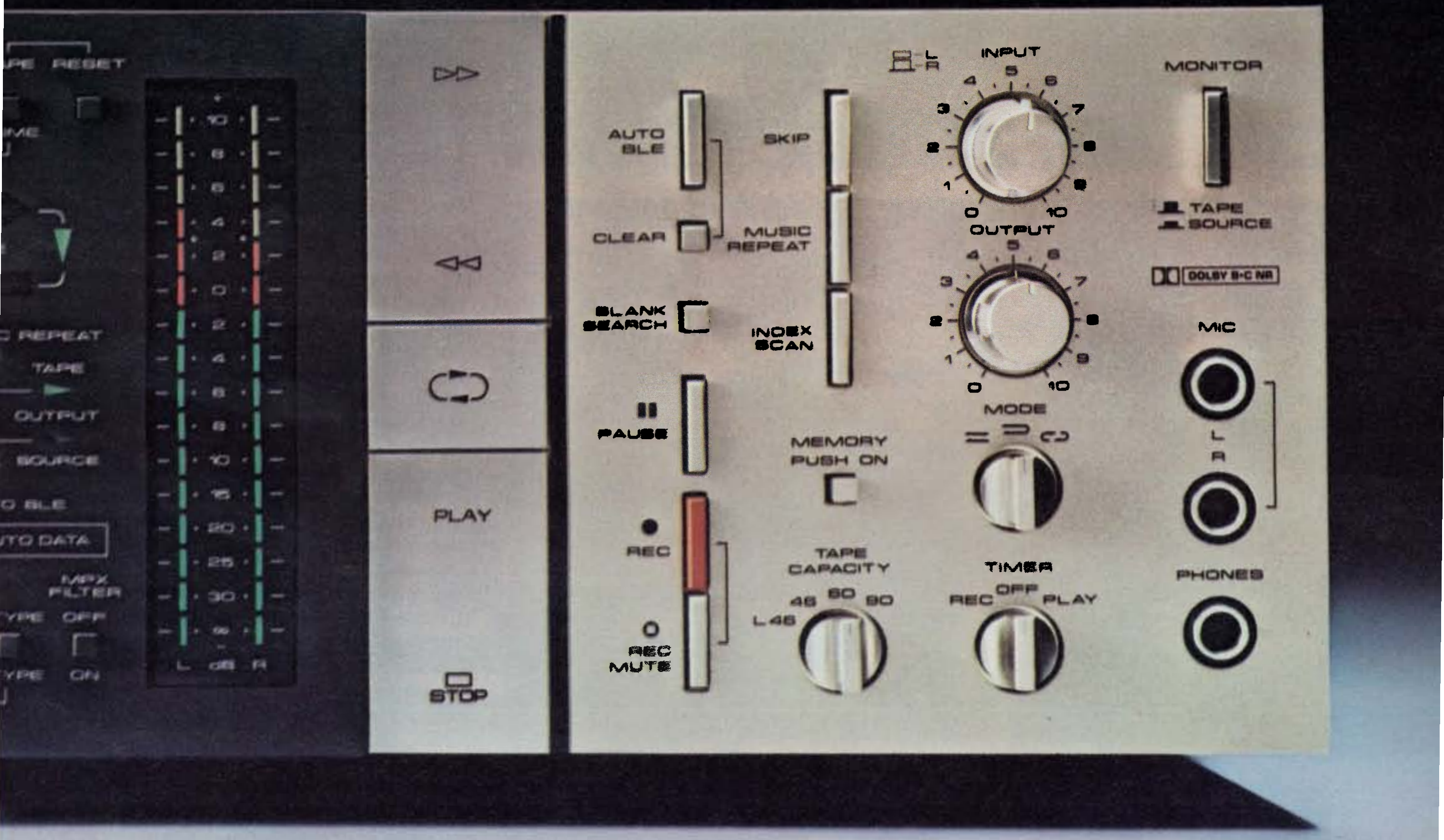
Touch a button and find your favorite song. Because the CT-9R Index Scan breezes through your tape, automatically stopping to play the first five seconds of each piece of music.

If you want to hear a song over, you don't press REVERSE. STOP. PLAY. REVERSE. STOP. PLAY, until you find the beginning. Instead, you simply press the Music Repeat button. The deck does the rest.

The CT-9R even plays both sides of a cassette, automatically.

But don't get the idea that we've produced a cassette deck that is just a lot of fun to play with. It's also a lot of

YOU RECORD,
E TO END LIKE TH



fun to listen to.

Our signal-to-noise ratio and high frequency response set a standard in state of the art electronics due to the creation of totally unique record and play heads. They're called RIBBON SENDUST heads and they're only on Pioneer cassette decks.

We've also attained extraordinary record and playback accuracy. Because we've seen to it that the drive capstan and both the take up and supply spindles are driven directly by their own motors. We call it our 3 Direct Drive motor transport and it, too, is exclusively Pioneer's.

Plus, we have Dolby C. The latest in Dolby engineering, designed to once and for all rid you and your tape of hiss.

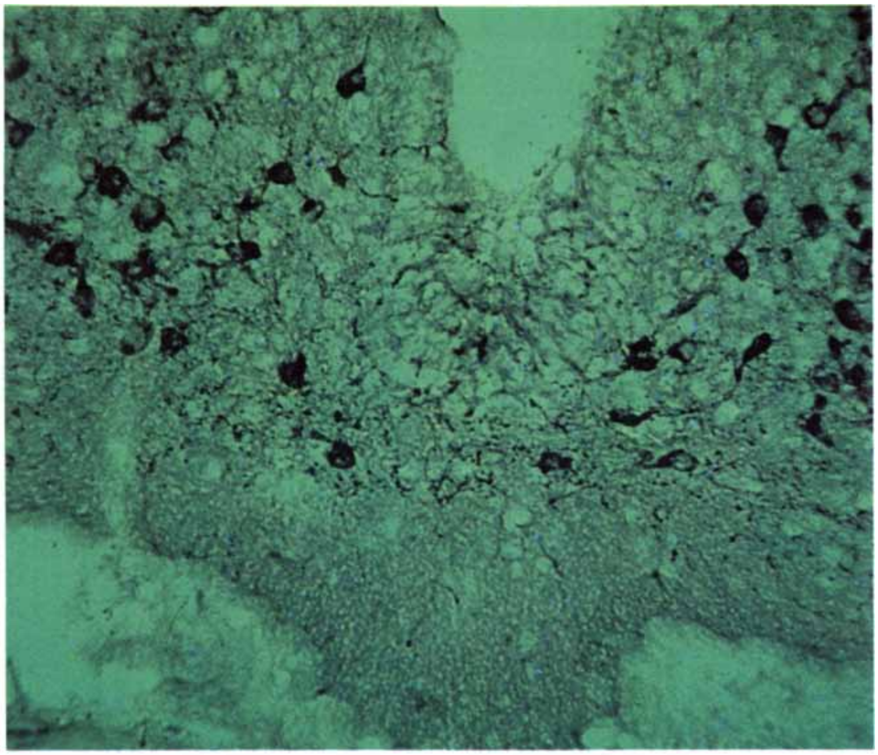
If you're the least bit skeptical that a cassette deck could do so much so well, we suggest you visit your nearest Pioneer dealer.

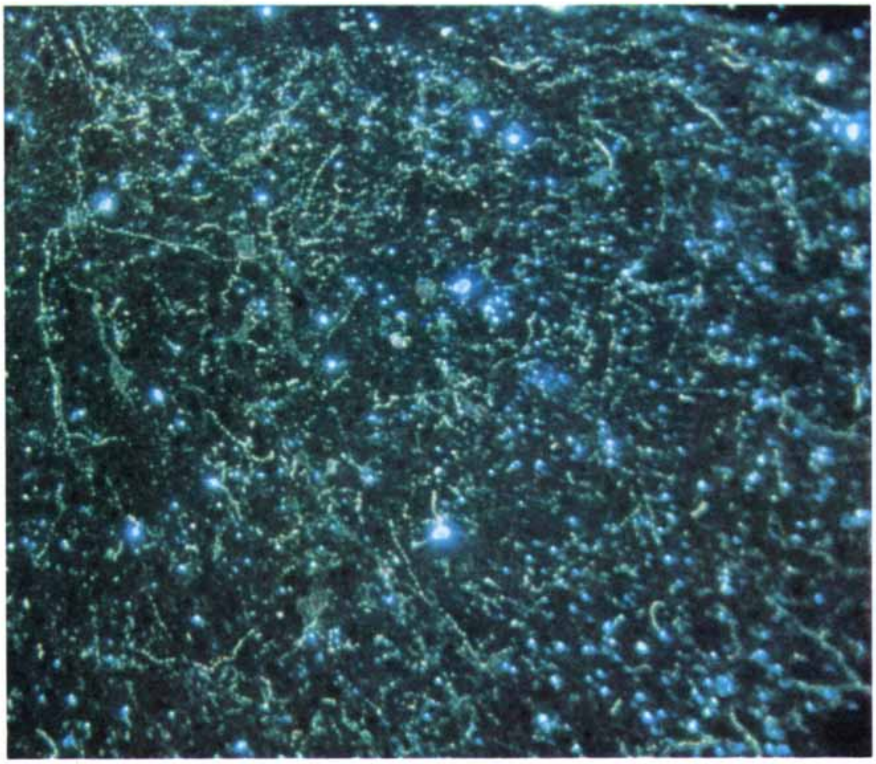
You can see the CT-9R for yourself, as well as an entire line of new Pioneer cassette decks.

But be forewarned. After seeing these, you'll begin to see cassette decks that just play music for exactly what they are.

Somewhat less than adequate.

PIONEER®
We bring it back alive.





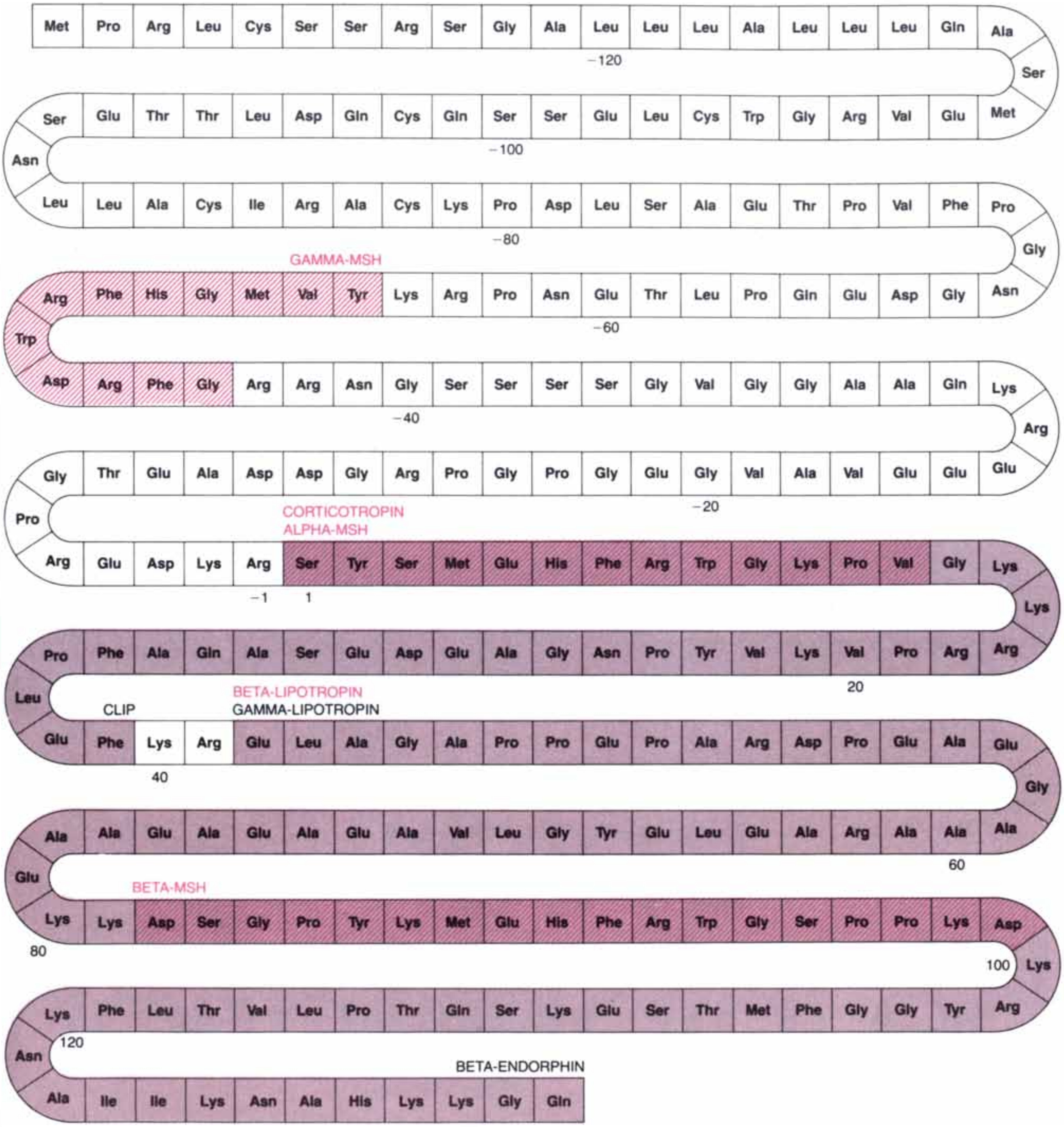
PRESENCE OF BETA-ENDORPHIN, a neuropeptide consisting of a chain of 31 amino acids, is demonstrated in the cell group in the rat hypothalamus called the arcuate nucleus by a technique that employs antibodies to tag the cells containing beta-endorphin with the enzyme horseradish peroxidase. The enzyme catalyzes the polymerization of the brown pigment diaminobenzidine. The upper photograph shows the preparation under the light microscope. The brown pigment marks numerous nerve cells. They appear to be the only cells in the brain of the rat that contain beta-endorphin. Each cell is some 25 micrometers across. The space at the top of the field is the third ventricle, a fluid-filled chamber inside the brain. The lower photograph shows the same tissue under dark-field illumination. Light passes obliquely through the tissue and is deflected onto the photographic film by the particles of the pigment. The varicosities that show up in gold are axons and dendrites: the filamentous extensions of the cells marked by the staining technique. The cell bodies themselves are small, elongated triangles. Bluish pinpoints of light are irregularities at the tissue's surface. Some of them are capillaries cut in cross section. The photographs were made by Elena Battenberg and the author at the Salk Institute.

peptide is often commercially available within weeks of its discovery. The quantity of the synthetic replicate is much greater than the quantity investigators could hope to produce by progressive purifications of a cell extract. Hence the replicate can be tested to see if it matches the purified natural peptide in both action and potency. Moreover, the replicate can be tested for biological activity on tissues other than the one it is known to act on.

Further still, the replicate can be injected into an animal of a species other than the one from which the peptide was purified. The immune system of the animal will prime antibodies against the peptide. The peptide itself can be quite short; thus its sequence of amino acids may be nearly identical in many species. In such instances the priming of antibodies is encouraged by linking the peptide to a large carrier molecule before it is injected. In any case the antibodies have several uses. In the technique called the radio-immunoassay, developed by Solomon A. Berson and Rosalyn S. Yalow of the Bronx Veterans Administration Hospital, they are mixed with a replicate of the peptide in which some of the atoms are radioactive. The antibodies bind to the replicate. Then a tissue extract is added to the mixture. The native peptide in the extract will displace a certain amount of the replicate. The degree of displaced radioactivity is a sensitive measure of the amount of native peptide.

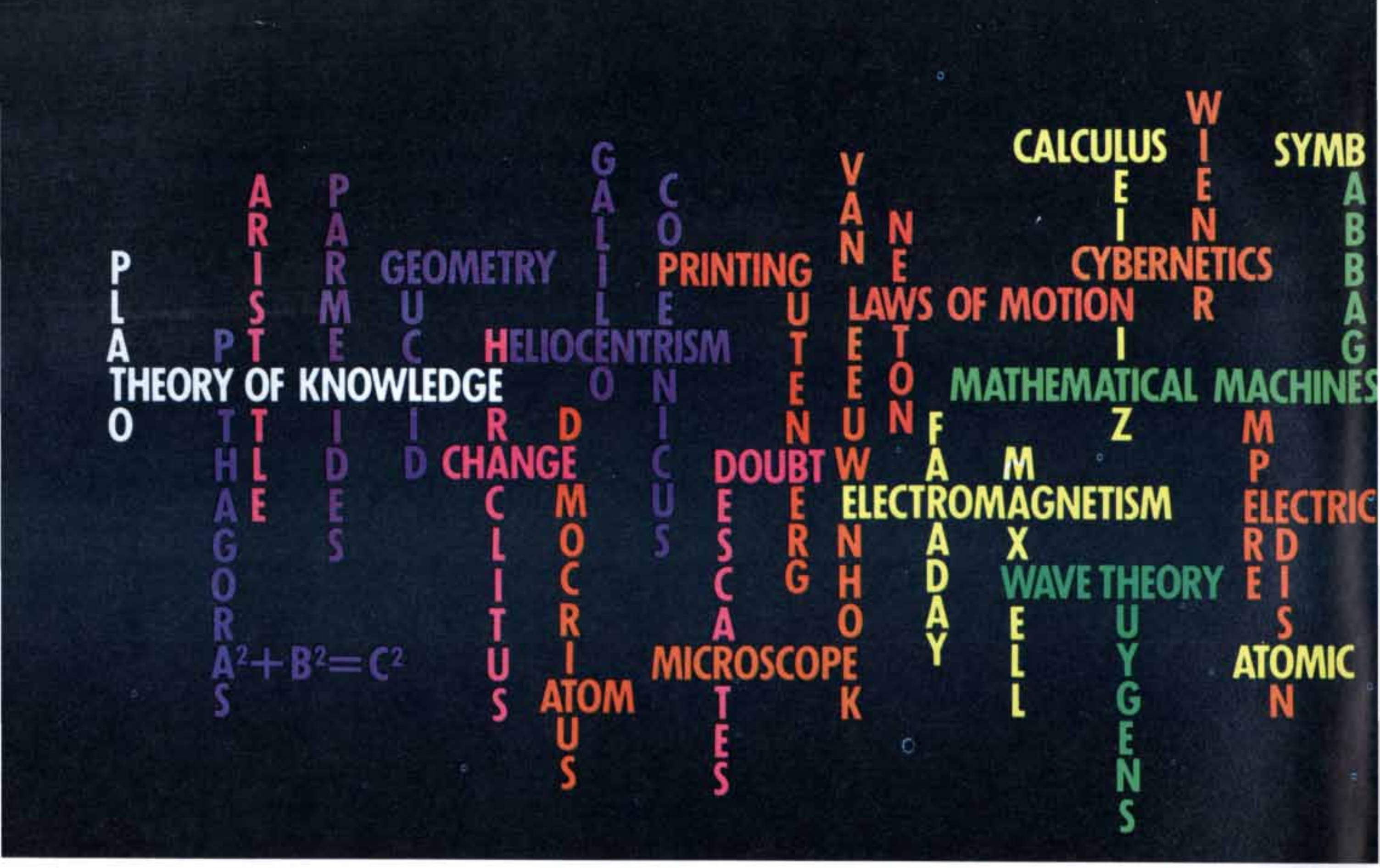
The antibodies can also serve as a microscopic stain to reveal the location of cells that store the peptide and presumably utilize it. In some techniques the antibody is labeled with a chemical group that is fluorescent or one that is radioactive. In still other techniques the antibody is linked to an enzyme that can manufacture a pigment inside cells. The most elaborate staining technique calls for animals of three different species. Suppose a peptide has been purified from the brain of a rat and a synthetic replicate of the peptide has been prepared. The replicate is injected into a rabbit. The resulting rabbit antibodies are applied to sections of the brain of a rat. There they bind to the native peptide. The rabbit antibodies are also injected into a goat. The result is goat antibodies that act against rabbit antibodies. These goat molecules are labeled and applied to the sectioned rat brain. They bind to the rabbit antibody, which is already bound to the peptide. The double-antibody technique has the advantage that the antibody prepared against the synthetic peptide is not subjected to a chemical labeling reaction, which might alter its ability to react with the native peptide molecule.

The employment of antibodies to seek out the peptide often reveals that many more cells contain the peptide than had been thought. Indeed, certain peptides



Ala Alanine
Arg Arginine
Asn Asparagine
Asp Aspartate
Cys Cysteine
Gln Glutamine
Glu Glutamate
Gly Glycine
His Histidine
Ile Isoleucine
Leu Leucine
Lys Lysine
Met Methionine
Phe Phenylalanine
Pro Proline
Ser Serine
Thr Threonine
Trp Tryptophan
Tyr Tyrosine
Val Valine

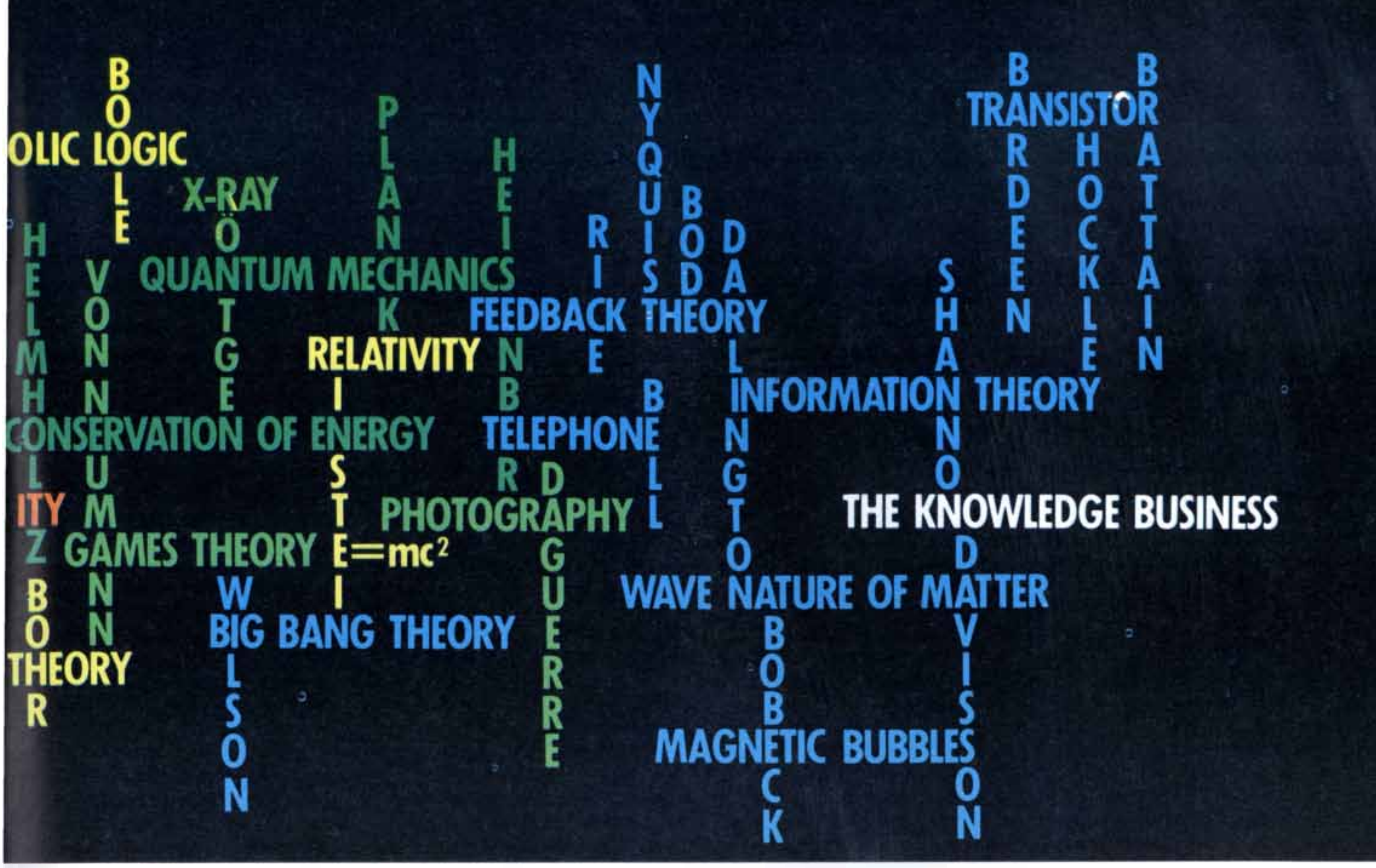
SEVERAL PEPTIDE MESSENGERS lie in the single protein (a long peptide chain) whose amino acid sequence in cattle is diagrammed. The protein, pro-opiomelanocortin, is synthesized by cells in the arcuate nucleus of the hypothalamus and also by cells in an endocrine gland: the anterior lobe of the pituitary. Evidently it serves as a precursor from which any of several intercellular messenger substances can be cleaved. Specifically the pro-opiomelanocortin chain incorporates the chain of corticotropin (*1–39*), a hormone released by the anterior lobe that stimulates the cortex of the adrenal gland to release other hormones, and the chain of beta-lipotropin (*42–132*), a weak hormone released by the anterior lobe that causes fat cells to break down their lipids. The corticotropin chain in turn incorporates the chain of alpha-melanocyte-stimulating hormone, or alpha-MSH (*1–13*), which in the frog is known to change the color of the skin. It also incorporates the chain of CLIP, or corticotropin-like intermediate-lobe peptide (*18–39*), a weak form of corticotropin. At the same time beta-lipotropin incorporates gamma-lipotropin (*42–99*), a weak form of beta-lipotropin. It also incorporates a version of MSH: beta-MSH (*82–99*). Further still, it incorporates endorphins. Beta-endorphin (*102–132*) is thought to be a neurotransmitter: a substance by which a nerve cell signals other cells. Five amino acids at one end of the endorphin chain (*102–106*) duplicate the sequence of an enkephalin. A third version of MSH (gamma-MSH) lies in a part of pro-opiomelanocortin not identified as a messenger. The ends of gamma-MSH may lie at a bond between the amino acids arginine and lysine (*–57 and –56*) and a bond between two arginines (*–43 and –42*), places where enzymes that cleave peptide chains typically act. The amino acid sequence in pro-opiomelanocortin was deduced by Stanley N. Cohen of the Stanford University School of Medicine and Shosaku Numa and Shigetada Nakanishi of Kyoto University.



Enter the information age.

From the beginning, knowledge has shaped the future.

Now, America is entering a new age, the Information Age. Bell scientists have used their knowledge to invent the tools for managing information in this new age: the transistor, magnetic bubbles and other breakthroughs shown above in blue. We apply this knowledge to advanced systems for business by drawing on our knowledge of information networks and of each industry's specific information management needs.


BOOLE
OLIC LOGIC
X-RAY
PLANCK
QUANTUM MECHANICS
NYQUIST
BODE
FEEDBACK THEORY
RELATIVITY
BARDEEN
TRANSISTOR
SHOCKLEY
BRATTAIN
INFORMATION THEORY
SHANNON
CONSERVATION OF ENERGY
TELEPHONE
PHOTOGRAPHY
THE KNOWLEDGE BUSINESS
GAMES THEORY
E=mc²
EINSTEIN
VON NEUMANN
WAVE NATURE OF MATTER
DAVISSON
BIG BANG THEORY
WILSON
DAGUERRE
THEORY
MAGNETIC BUBBLES
BOBECK

have been found in neurons that were known to contain a monoamine and so had been taken to be specialized for the release of that neurotransmitter alone. At the same time the elucidation of the chemical structure of the factor often makes possible an improvement in the method used to purify the natural factor from the cells that contain it. If the factor is a peptide, for example, and the original method of purification began with a homogenate of brain tissue, it may well be that the disruption of the cells by the homogenization allowed the peptide to be attacked by the enzymes called peptidases. A method can then be devised in which such enzymes are inactivated before the peptide is extracted. Sometimes the result is surprising: a larger form of the peptide is discovered. It had gone undetected because the first extraction procedure had destroyed it. In some instances the larger peptide is more potent than the smaller one.

Vasopressin and Oxytocin

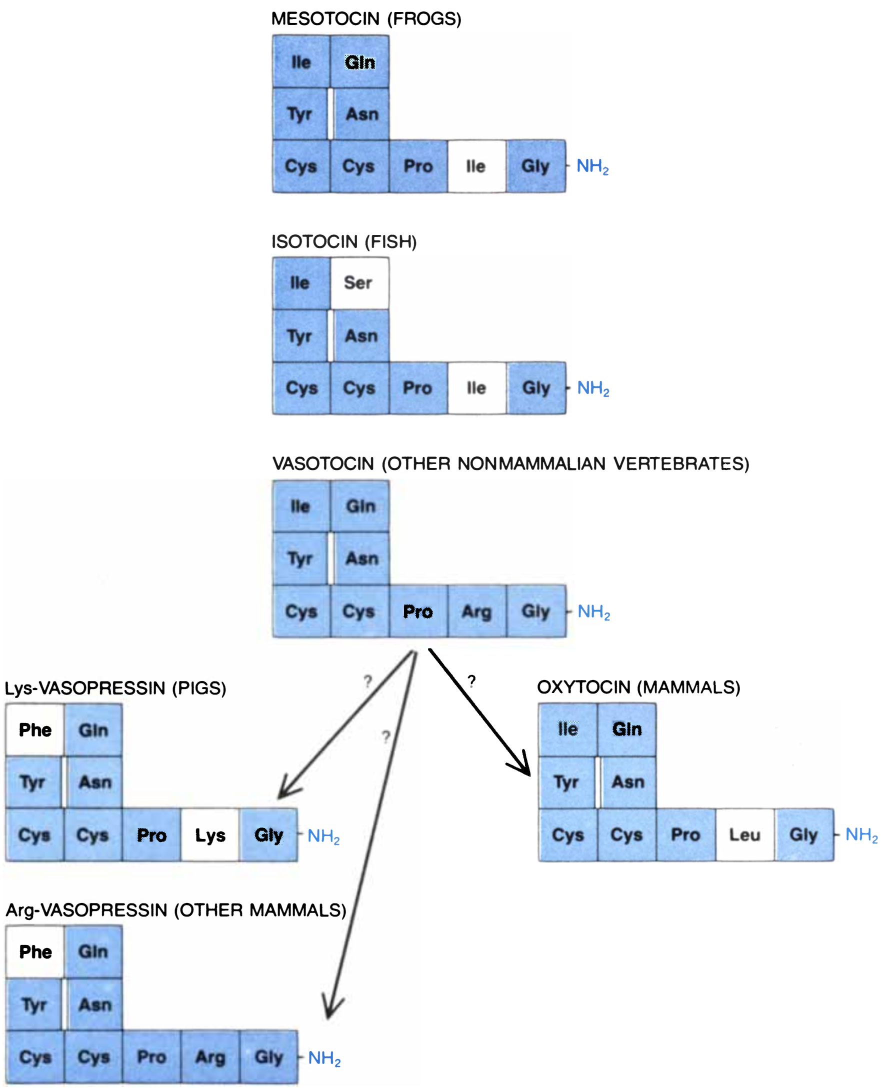
As more and more peptides have been purified and their structure has been determined several groups of chemically related substances have emerged. Two types of group can be distinguished. In one type the peptides purified from cells across a wide range of species include almost identical long sequences of amino acids. In the second type a number of different large peptides purified from the cells of a single species include identical short sequences of amino acids. Presumably the sequences that are identical have proved particularly well suited for intercellular signaling through long periods of evolution.

Consider the peptide messengers first identified by du Vigneaud and his colleagues as factors released by the posterior lobe of the pituitary gland, an appendage to the brain. They can be seen in the gland as fatty droplets inside the terminals of the axons, or long fibers, of nerve cells in the hypothalamus. To Ernst A. Scharrer, who was then at the University of Colorado School of Medicine, the cells seemed to be typical neurons: they appeared to get information from their synapses with other cells. Scharrer thus advanced the surprising hypothesis that the cells were neurosecretory neurons: like any other nerve cell they were controlled by synaptic connections, and yet on suitable command they could secrete a hormone into the bloodstream. It is now recognized that their axons also project to several levels of the brain. Hence they evidently rely on the hormone as a neurotransmitter. Working at the Albert Einstein College of Medicine in New York, Berta V. Scharrer, Ernst's widow, found similar neurons in the nervous system of a number of invertebrate species.

Du Vigneaud named the factors in accord with their physiological action. Specifically, he distinguished a factor he called antidiuretic hormone, which acts on the kidneys of an animal deprived of drinking water to lessen the loss of water in the urine, and a factor called oxytocin, which promotes the contraction of the muscle of the uterus and thereby speeds birth. When it emerged that antidiuretic hormone could elevate the blood pressure by causing certain arteries to constrict, the factor was given the second name vasopressin.

Vasopressin and oxytocin turn out, then, to be an example of a family of peptides whose structure is well conserved across a wide range of species. In most vertebrate animals less advanced than the mammals the family is represented by a single substance, which is called arginine-vasotocin because of the presence of the amino acid arginine at position No. 8 in its chain of amino acids. In mammals the hormones oxytocin and vasopressin appear. Oxytocin is identical with vasotocin except that the amino acid leucine takes the place of arginine at position No. 8. Vasopressin is slightly more idiosyncratic. In all mammals except the pig it is identical with vasotocin except that phenylalanine substitutes for isoleucine at position No. 3. In the pig one finds two substitutions: lysine takes the place of arginine at position No. 8 and phenylalanine takes the place of isoleucine at position No. 3.

Wilbur H. Sawyer of the Columbia University College of Physicians and Surgeons has conjectured that a gene duplication may have allowed a peptide



VASOPRESSIN AND OXYTOCIN are the peptide hormones synthesized by nerve cells in the hypothalamus of mammals and then released by those cells from the ends of their axons in the posterior lobe of the pituitary gland. Vasopressin (*lower left*) constricts arteries and causes the kidneys to recapture water from the urine. Oxytocin (*lower right*) differs only slightly in structure from vasopressin, yet its action is quite different: it promotes contractions of the uterus and thereby speeds birth. Both vasopressin and oxytocin are almost identical in structure with vasotocin, mesotocin or isotocin (*top*), the hormones released from the posterior lobe of the pituitary gland in nonmammalian vertebrate animals. The similarities (*color*) suggest that vasopressin and oxytocin arose through mutations in duplicate copies of a single gene.

modified at position No. 8 to evolve from arginine-vasotocin. The modification would lessen the peptide's antidiuretic action and enhance its oxytocic action. In any case the alteration in the action of a peptide hormone because of a minor change in the structure of the peptide shows that each peptide acts on different target cells in spite of the small degree of difference. It follows that the receptors on the target cells (that is, the sites on the membrane of the target cells that "recognize" the messenger) must be precise enough to accept the one molecule and reject the other. Perhaps some further minor modifications of the molecules will yield synthetic peptides whose specificity of action is greatly enhanced. Such substances might well be valuable pharmaceuticals.

Endorphins and Enkephalins

Perhaps the most striking example of the second type of family relationship among peptide messengers—the presence of certain short sequences of amino acids common to several coexistent peptides—is provided by the endorphins and the enkephalins. The terms have emerged from a decade of almost frenzied research; they are now taken to signify classes of brain peptides whose effects on cells resemble those of opiates such as morphine. (The word endorphin is a contraction of endogenous morphine.) The saga of the endorphins and the enkephalins began with the discovery in several laboratories, including those of Solomon H. Snyder at the Johns Hopkins University Medical School, Eric J. Simon at the New York University School of Medicine and Lars Terenius at the University of Uppsala, that certain cells in the brain have receptors that bind opiates. Some of the cells lie in brain structures implicated in the perception of pain; others, however, do not. Then Hans W. Kosterlitz and John R. Hughes of the University of Aberdeen purified from brain-cell extracts a fraction that occurs naturally in the brain and suppresses the contraction of muscle tissue from the intestine, much like morphine itself. The molecules with this action turned out to be two pentapeptides. One of them has the sequence tyrosine-glycine-glycine-phenylalanine-methionine; it is known as met-enkephalin. The other, which is identical except that leucine substitutes for methionine, is leu-enkephalin.

Soon the met-enkephalin sequence was found in a number of longer peptides purified from an endocrine gland: the anterior lobe of the pituitary. Then the longer peptides—the endorphins—were found also in nerve cells. Some investigators surmised that such nerve cells shorten an endorphin to make an enkephalin. The hypothesis was contradicted, however, by the preparation of antibodies to the endorphins and the enkephalins and the subsequent mapping of the nerve cells that contain them. The

PEPTIDES RESEMBLING GLUCAGON form a family of four. Glucagon (*a*) is a hormone synthesized by the alpha cells of the islets of the pancreas. It stimulates the liver to break down glycogen into glucose; hence it counteracts insulin. Gastric inhibitory peptide (*b*) has been purified from the lining of the stomach. It inhibits the contraction of smooth muscle in the small intestine, perhaps to allow time for intestinal enzymes to act on food. Secretin (*c*) has been purified from the lining of the small intestine. It causes the acinar cells of the pancreas to secrete fluid that neutralizes the acidity of gastric secretions. Vasoactive intestinal peptide, VIP (*d*), was first purified from the lining of the small intestine. Evidently it dilates the blood vessels there; thus it increases the flow of blood in the intestinal wall. Recently VIP has turned up in nerve fibers that innervate the intestine. It has also been found in the brain. The amino acid sequences shown in the illustration are those of peptides purified from the pig. Duplications of amino acids in the glucagon family are shown in color.

staining patterns reported for the enkephalins by Tomas G. M. Hökfelt and his associates at the Karolinska Institute in Sweden show that cells containing enkephalins are widespread in the brain. They lie, for example, in the spinal cord, the brain stem, the hippocampus and the corpus striatum. In contrast the staining patterns my colleagues and I at the Salk Institute have found for the endorphins place them exclusively in nerve cells at the base of the hypothalamus and in endocrine cells in the anterior lobe of the pituitary.

The family relations being discovered within the endorphins and within the enkephalins are becoming complex and distinctive. For the endorphins the first hint of complexity came when Richard Mains and Betty Eipper of the University of Colorado and Nicholas Ling of the Salk Institute discovered why endorphins are found in the cells of the anterior pituitary that release the hormone corticotropin. Each such cell makes a long precursor peptide from which a molecule of endorphin and a molecule of corticotropin are cleaved. The precursor was given the provisional name pro-opiocortin. Yet it was twice as long as it would be if it consisted of a molecule of corticotropin joined to a molecule of endorphin.

Quite recently the entire sequence of amino acids that make up the precursor has been determined by Stanley N. Cohen of the Stanford University School of Medicine and Shosaku Numa and Shigetada Nakanishi of the Kyoto University Faculty of Medicine. These workers determined the sequence without attempting to purify the peptide as in the orthodox process of discovery. Instead they employed the new techniques of genetic engineering to produce multiple copies of the messenger RNA that specifies the sequence. Then they worked out the structure of the messenger RNA. The RNA is a strand of the units called nucleotides. A particular triplet of nucleotides encodes the identity of a particular amino acid.

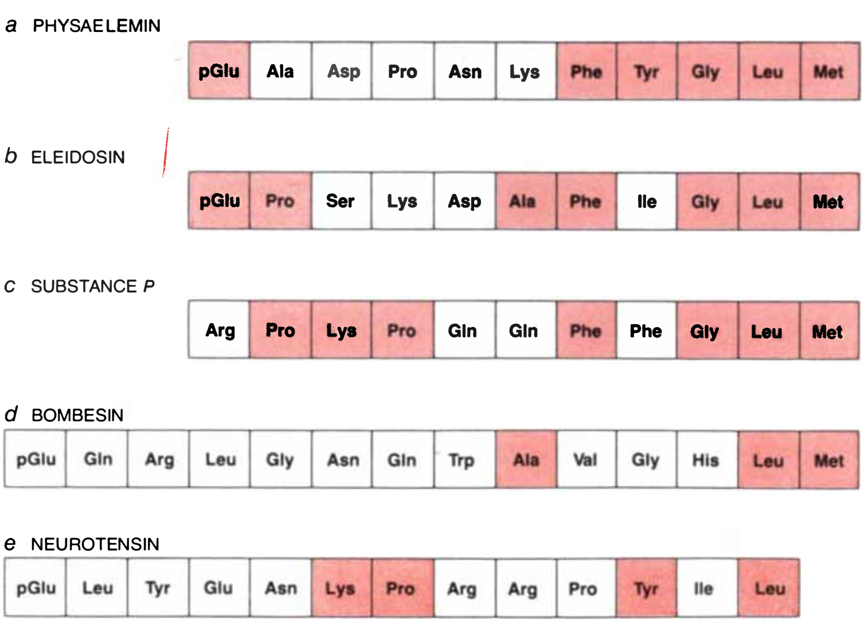
In the wake of the work of this group it proved advisable to call the precursor pro-opiomelanocortin. The added term melano refers to a sequence of seven amino acids. One copy of the sequence is included in corticotropin; it is called alpha-melanocyte stimulating hormone (alpha-MSH) because it is known to disperse the pigment melanin in the pigment cells (melanocytes) in the skin of the frog. It thereby modulates the color of the skin. (When a green frog is placed in the dark, its skin turns brown.) A nearly identical sequence lies next to the endorphin sequence in pro-opiomelanocortin; it is called beta-MSH. It was discovered by Choh Hao Li of the University of California at San Francisco, who purified a number of pituitary peptides and determined their sequences of amino acids by classical methods in the late 1950's and early 1960's.

What Cohen, Numa and Nakanishi discovered was a third sequence almost identical with the other two. It lies in the section of the pro-opiomelanocortin chain whose sequence of amino acids had not previously been determined. The third sequence was named gamma-MSH in advance of any evidence that it has its own action in the body. Work done at the Salk Institute by Ling and Guillemin and their colleagues and by my colleagues and me suggests, however, that gamma-MSH is stored in cells of the anterior lobe of the pituitary. Moreover, when gamma-MSH is injected into a cerebral ventricle of an experimental animal (a fluid-filled space in the brain), the animal's body temperature decreases. Evidence is also accumulating that certain cells of the hypothalamus rely on gamma-MSH as a neurotransmitter.

Molecules Incorporating Enkephalin

The complexities in the enkephalin branch of the family are also increasing now that investigators in the U.S. and Japan have begun to detect enkephalin sequences in longer molecules that are otherwise unlike the endorphins. The differences lie in the amino acids next to the end of the enkephalin pentapeptide designated the *C* terminus. Specifically, the endorphins in pro-opiomelanocortin have serine and threonine in that position, whereas the recently discovered molecules that incorporate the enkephalin sequence have lysine and arginine there. Lysine and arginine are special amino acids in that they each have two amino (NH_2) groups, one of which projects from the amino acid's molecular structure. Studies of a large number of peptides show that the sites where two such amino acids are adjacent in the peptide chain are the places where enzymes most often cleave the peptide. The two adjacent amino groups protruding from the peptide chain may well be the feature the enzyme recognizes when it binds to the peptide and cleaves it. The occurrence of the feature next to the enkephalin pentapeptide suggests that the large molecules are precursors of enkephalin.

One version of a large molecule that includes leu-enkephalin was discovered recently by Avram S. Goldstein and his colleagues at the Stanford School of Medicine. They call it dynorphin because it is more potent than enkephalin in the standard biological assay for the effect of an opiate. (In the standard assay the substance is applied to a sample of smooth muscle from the intestine of a guinea pig. The muscle is induced to contract by electrical stimulation, and the degree of the substance's ability to suppress the contraction constitutes the assay.) Other large molecules that include enkephalins are also reported to be more potent than the enkephalin se-



PEPTIDES RESEMBLING PHYSAELEMIN form a family of perhaps as many as five. Physaelemin (*a*) has been purified from the skin of the frog. Eleidosin (*b*) has been purified from the salivary glands of the octopus. The other three peptides—substance *P* (*c*), bombesin (*d*) and neurotensin (*e*)—have been purified from the nervous system of mammals. Although all five substances promote the contraction of muscle tissue taken from the viscera, it is not known whether they act in that way in the body. Again the sequences are those of peptides in the pig, and color denotes duplications of amino acids. The abbreviation pGlu stands for pyroglutamate, a form of glutamate in which a side chain on the amino acid binds to one end of the amino acid to form a small ring. Pyroglutamate appears at the same end of four of the peptides.

themselves more closely to where the world is going—and the products, talents, and techniques it needs to get there.

That's the selfish reason Sperry's committed to listening.

To lead, you need a lucid vision of the future.

SPERRY

We understand how important it is to listen.

Sperry is Sperry Univac computers, Sperry New Holland farm equipment, Sperry Vickers fluid power systems, and defense and aerospace systems from Sperry division and Sperry Flight Systems.

For a booklet on listening, write to Sperry, Dept. H4, 1290 Avenue of the Americas, New York, N.Y. 10104.

quence itself. Thus it is conceivable that the first estimates of the potency of the enkephalins were based on studies of the partially degraded larger forms.

Still another complication in the enkephalins arose when Hökfelt and his colleagues applied the immunohistochemical stains for the enkephalins to the cells of the adrenal medulla. The cells were previously known to secrete only the monoamines epinephrine and norepinephrine. Now Hökfelt's group found that these cells show a large degree of what Hökfelt was careful to call "enkephalin-like" reactivity to the stain. He was cautious in his conclusions because he considered it possible that the antibody employed for the staining procedure had bound itself not to enkephalin but to a similar sequence in some undiscovered peptide.

The caution was well placed. A series of discoveries made in laboratories including that of Sidney Udenfriend at the Roche Institute of Molecular Biology has now revealed that several large peptides from extracts of the adrenal medulla include the enkephalin sequence. One of them is at least 10 times larger than enkephalin. It has several copies of the met-enkephalin sequence and at least one copy of leu-enkephalin. The large polyenkephalin peptides seem to be secreted in tandem with epinephrine and norepinephrine but at less than a hundredth the concentration. Moreover, no interaction between the peptide messengers and the monoamine messengers on a common target cell has yet been discovered.

Gut Peptides in the Brain

The nerve cells that include the pro-opiomelanocortin peptide present a special problem: Do they pare that precursor down into an endorphin, into corticotropin or into gamma-MSH, or do they pare it down into other peptide molecules not yet known to be messengers? Perhaps the signals arriving at such a cell can direct the cleaving of the precursor so that certain axon terminals release specific products.

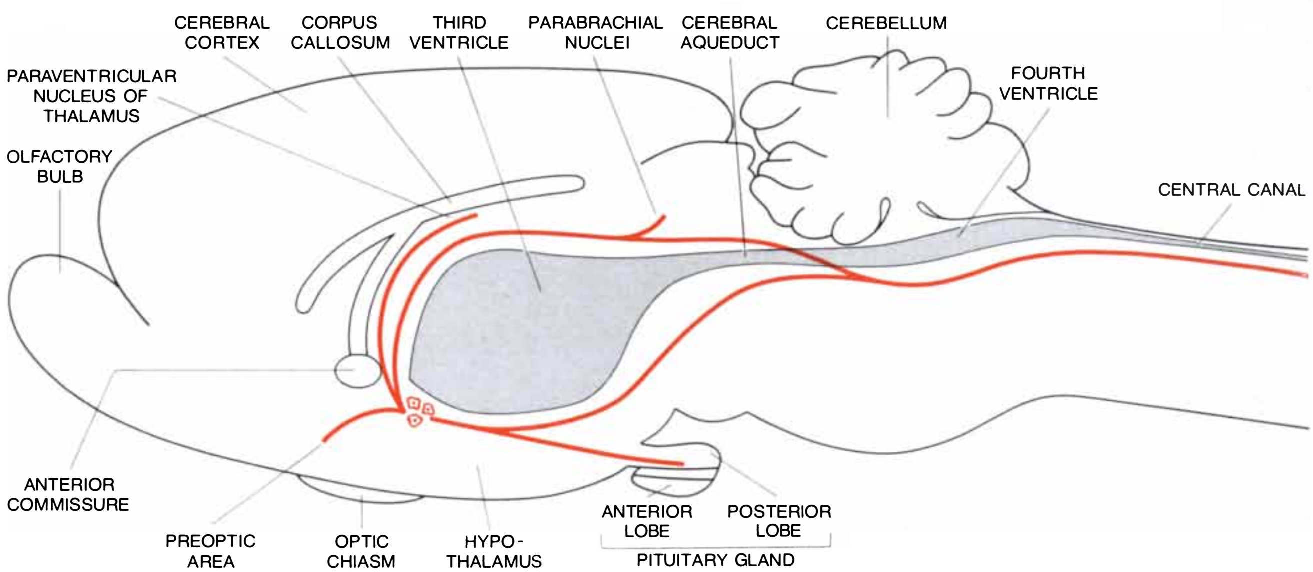
A related problem then arises. The enkephalins and some of the endorphins suppress the perception of pain. Hence it has sometimes been advanced that the modulation of the perception of pain is their primary function. The enkephalins and the endorphins are now found, however, in brain circuits that are implicated in a wide range of functions, including the control of blood pressure and body temperature, the regulation of the secretion of hormones and the governance of body movement. This wide range of functions has confounded the effort to attribute to the messengers a single domain of function.

To be sure, it might be advanced that several independent systems depend on messengers that share the amino acid sequence of an enkephalin. Presumably the messengers share the sequence because they evolved from some single messenger early in animal evolution. Today each such system would act on a different set of target cells; thus each set of target cells would require its own type of receptor, and a given type of receptor would have to be able to distinguish between two molecules that differ only slightly in structure.

The view that the various systems exploiting enkephalins and endorphins might be independent was presaged by William D. Martin of the University of Kentucky Medical Center well before the endorphins were discovered. His study of the actions of opiates on the spinal cord of the dog had shown that different actions of morphine are simulated or blocked by different drugs. This suggested distinct classes of opiate receptors. More recently the work of Kosterlitz at Aberdeen, Albert Herz at the Max Planck Institute for Psychiatry in Munich, Goldstein at Stanford and several other investigators shows that the various peptides containing an enkephalin pentapeptide differ considerably in their potency in suppressing the perception of pain or the contractility of smooth muscle. This too suggests several classes of receptors.

Adopting (at least for the moment) the view that the cells secreting structurally similar peptides need not be functionally related may help to explain why many peptides once thought to serve only the gut or the endocrine system are also detected in the brain. Often they are found in parts of the brain that have nothing to do with the realm in which the substance acts in the periphery of the body. Among these substances are the peptides of the gut (and now the brain) called substance *P*, vasoactive intestinal peptide and cholecystokinin. Each one has been found in neurons that make local synaptic connections ("local-circuit neurons") in the cerebral cortex and in the hippocampus. Moreover, nerve fibers containing substance *P* or vasoactive intestinal peptide have been found innervating visceral tissues such as parts of the lung. They have even been found innervating the thyroid gland, where no nerve fibers were known to terminate.

A still more striking example of an almost ubiquitous peptide messenger is somatostatin. It is a substance that Paul Brazeau, Wyley Vale, Roger C. Burgus and Ling and Guillemin of the Salk Insti-



NERVE CELLS EMPLOYING OXYTOCIN (evidently) as their neurotransmitter are found in the rat in the cell groups of the hypothalamus called the supraoptic nucleus and the paraventricular nucleus. The axons of the cells project into the posterior lobe of the pituitary gland, from which they release the oxytocin as a hormone. They also project, however, into cell groups in the brain and the spinal cord.

Your prayers have been answered.

If you own or use a micro-computer, then chances are that from time to time, you've wished that someone could simplify programming.

Because as useful as micro-computers are, they can only ever be as good as the programs they run.

Well then, how does this sound?

No more program-coding. No more debugging. And no more time wasting.

Arguably more comprehensive and advanced than anything else of its kind, The Last One is a computer program that writes computer programs. Programs that work first time, every time.

By asking you questions in plain English about what you want your program to do, The Last One uses your answers to generate a ready-to-use program in BASIC.

What's more, with The Last One, you can change or modify your program as often as you wish. Without effort, fuss or any additional cost. So as your requirements change, your programs can too.

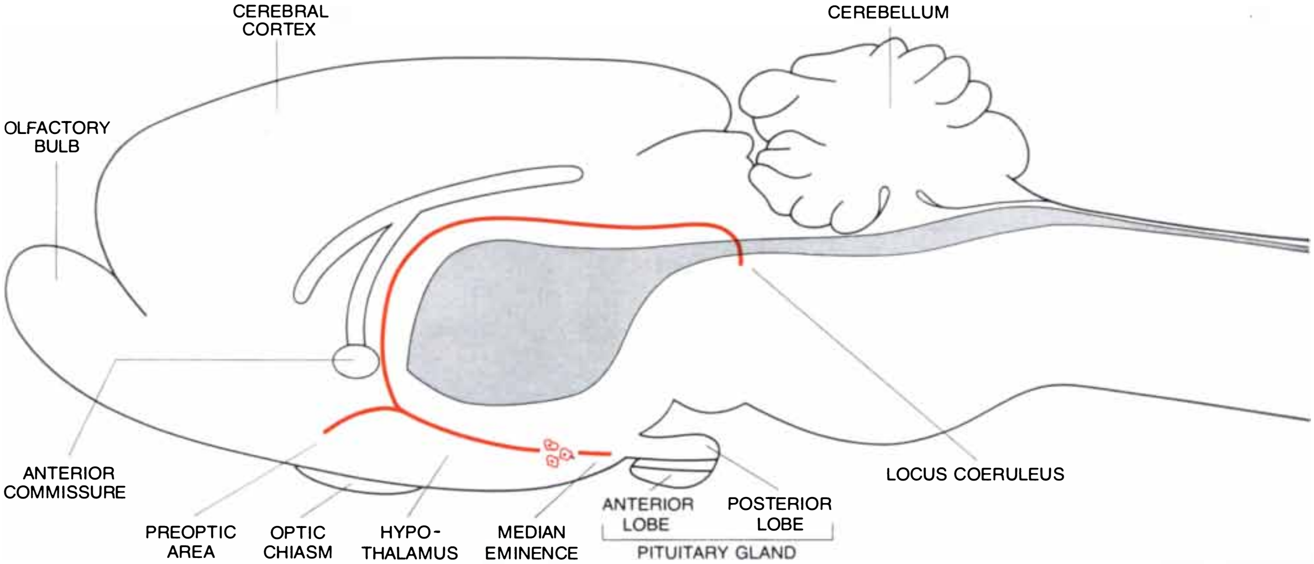
And if, because of the difficulties and costs of buying, writing and customising software, you've put off purchasing a computer system up to now, you need delay no longer.

Available now.

The Last One costs $600 plus local taxes where applicable and is now available from better computer stores.

For further information, write to D.J. 'AI' Systems Ltd.,
Two Century Plaza, Suite 480,
2049 Century Park East,
Los Angeles, CA 90067.
Tel: (213) 203 0851.

THE LAST ONE®



NERVE CELLS CONTAINING ENDORPHIN are found in the rat in the arcuate nucleus of the hypothalamus. The axons of the cells (*also shown in the photographs on page 152*) project along pathways in the brain similar to those of the cells containing oxytocin. The pattern shown emerges also for corticotropin and gamma-MSH, two other peptide messengers in the precursor pro-opiomelanocortin.

tute discovered by applying the classical process of discovery to extracts from the hypothalamus. The biological assay they resorted to as a criterion for successive purification of the extract was the ability of each successive extract to suppress the secretion of growth hormone by cells these workers had cultured from the anterior lobe of the pituitary. Soon it was determined that somatostatin is a peptide consisting of 14 amino acids. Then a replicate of somatostatin was synthesized and antibodies to it were developed. The employment of the antibodies as a stain placed somatostatin in certain cells of the hypothalamus positioned so that they can secrete the substance into a system of capillaries that carries it into the anterior lobe.

Further studies of the distribution of somatostatin have been little less than astounding. In the nervous system somatostatin is found in nerve cells in almost every neural tissue from the cerebral cortex to the autonomic ganglia: the nests of nerve cells in the periphery that govern the tissues of the viscera. In the gut somatostatin is found in the cells that line the intestine. In the endocrine system somatostatin is found in the delta cells of the islets of Langerhans. Its action in the pancreas appears to suppress the secretion of the hormones insulin and glucagon. Quite recently larger forms of somatostatin have been extracted from the pancreas and the brain that are more potent than the sequence of 14 amino acids.

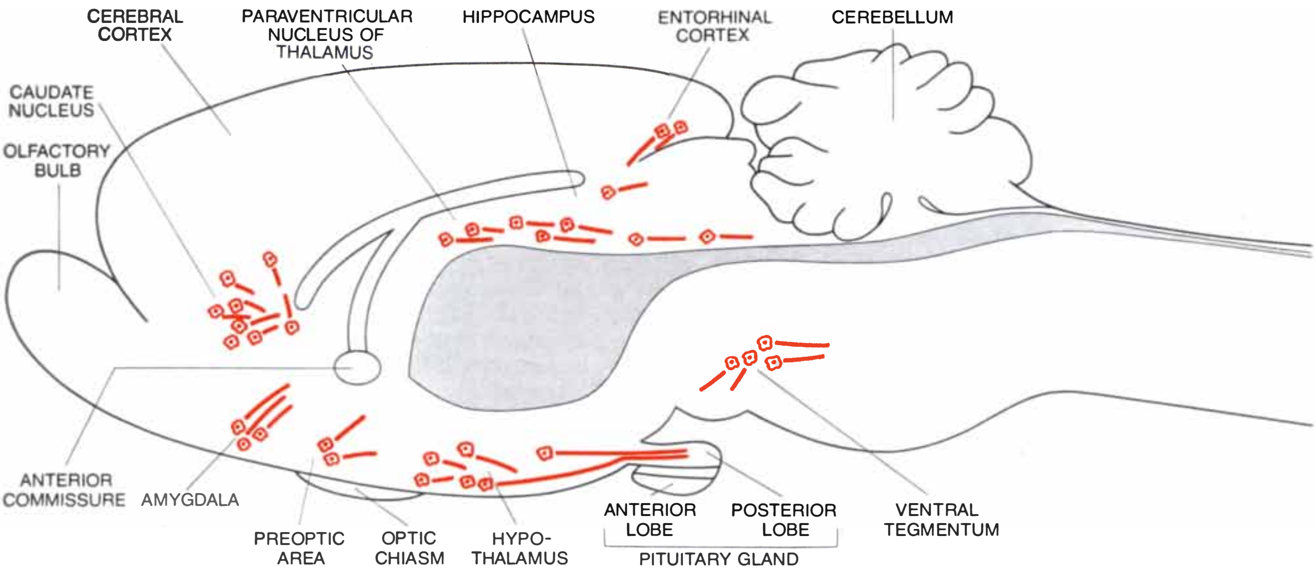
Peptides as Neurotransmitters

The utilization of a peptide by a nerve cell as a neurotransmitter differs from the utilization of an amino acid or a monoamine in a fundamental respect: the way the nerve cell synthesizes the substance and conserves it. A neurotransmitter such as gamma-aminobutyrate, serotonin or dopamine is made in a short series of steps from an amino acid in the diet. For each step an enzyme in the cytoplasm of the cell acts as a catalyst. In general each nerve cell includes the enzymes that synthesize a single transmitter. The active form of the transmitter molecule is stored in the sacs called synaptic vesicles until the nerve cell is called on to release it. After the transmitter is released the cell can reabsorb some of it. In that way the need for the synthesis of the transmitter is somewhat reduced.

In a cell that releases a peptide the process is more involved. In the first place the peptide can be synthesized only by ribosomes, the specialized intracellular organelles that synthesize all peptides, including the long ones: proteins. This means the sequence of amino acids that make up the peptide must be encoded by a gene, a strand of DNA in the nucleus of the cell. The gene must be transcribed into a strand of messenger RNA, which carries the code to the ribosome. In the second place all the peptide neurotransmitters examined so far appear to follow the pattern of the endorphins in that it seems they are synthesized first as a larger peptide. Then the active form of the molecule is produced through progressive cleavings by enzymes.

The ribosomes in a nerve cell lie only in the cell body and in the filamentous extensions of the cell body called dendrites. In general, however, the dendrites and the cell body of a nerve cell receive signals from other cells. A longer filament, the axon, transmits signals to other cells. Hence the release of a peptide neurotransmitter from an axon terminal is remote from the place where the peptide is made. The active form of the peptide (stored in synaptic vesicles) must be transported, then, to the place of release. It may follow that a cell releasing a peptide is unable to act on another cell repeatedly in a short span of time. In contrast, the axon terminals of a cell that releases an amino acid or a monoamine may well be satellite factories that make their own transmitter. Such terminals may stand less in need of replenished supplies of fresh neurotransmitter from moment to moment.

A further way in which a peptide neurotransmitter may differ from a monoamine neurotransmitter lies in the molecular details of how the transmitter influences its target cells. The classical neurotransmitters are said to be excitatory or inhibitory. The arrival of molecules of an excitatory neurotransmitter makes the target cell more likely to release its own neurotransmitter from its own axon terminals; the arrival of an inhibitory neurotransmitter has the opposite effect. Investigators who have studied these actions emphasize that the neurotransmitter alters the permeability of the target cell's membrane to ions of potassium, sodium or calcium. As a result of such changes the concentration of these ions inside the target cell changes with respect to their concentration outside the cell. The differing concentrations give rise to a voltage gradient across the membrane. The action of an excitatory neurotransmitter tends to diminish the voltage gradient; it "depolarizes" the membrane. The action of an inhibitory transmitter tends to increase



NERVE CELLS CONTAINING ENKEPHALIN are widespread in the brain of the rat, which makes them distinct from the nerve cells containing endorphin. Here the cells containing enkephalin are projected onto the midplane of the brain and displayed in proportion to the actual numbers of such cells at various locations. The trajectories of the axons arising from the cells are suggested by line segments.

the voltage gradient; it "hyperpolarizes" the membrane.

In experiments with the enkephalins, substance *P* and other peptide messengers a different action appears. The peptide makes the target cell less likely to respond to other signals. One might consider this to be an example of inhibition. In some instances, however, the peptide messenger keeps an excitatory transmitter from depolarizing the membrane; in other instances it keeps an inhibitory transmitter from hyperpolarizing it. Moreover, the arrival of the peptide itself often produces no notable change in the voltage gradient across the membrane. This curious effect of a peptide messenger is one I like to call disenabling. It has now been reported by Roger Nicoll of the University of California at San Francisco, who studied peptides in the spinal cord of the frog; by Jeffrey L. Barker and his colleagues at the National Institute of Neurological and Communicative Disorders and Stroke, who studied enkephalin in the mammalian neurons they cultured from the spinal cord; by Walter Zieglgänsberger and his colleagues at the Max Planck Institute for Psychiatry, who studied enkephalin in the spinal cord of the cat, and by Zieglgänsberger and George R. Siggins at the Salk Institute, who studied enkephalin in slices of hippocampus of the rat.

Since the peptide and the disenabled transmitter must each act on the target cell at a different set of receptors, one imagines the two sets of receptors might interact. Indeed, it seems that certain small populations of nerve cells can release both a monoamine and a peptide. Some are thought to release both acetylcholine and vasoactive intestinal peptide, others to release both dopamine and cholecystokinin and still others to release both serotonin and substance *P*. The arrival of such a pair of messengers at a target cell might give rise to a complex sequence of effects.

The Mediation of Behavior

The ultimate question about any intercellular messenger is how it integrates the activities of cells in a way that is appropriate to the circumstances in which the organism finds itself. Briefly, how does it mediate behavior? One answer is suggested by the work of Donald Pfaff and his colleagues at Rockefeller University and that of Robert L. Moss and Samuel M. McCann at the Southwestern Medical School of the University of Texas Health Science Center in Dallas. In each case the work concerns the decapeptide called luteinizing hormone releasing hormone (LHRH). LHRH is a substance purified from the hypothalamus that causes the release of luteinizing hormone from the anterior lobe of the pituitary. Luteinizing hormone in turn induces ovulation. LHRH has now been found in nerve cells in the autonomic ganglion that innervates the reproductive organs. Its utilization there instead of some other messenger may be fortuitous. As it happens, however, the injection of LHRH into a male or female rat either subcutaneously or into a cerebral ventricle evokes the posture required for copulation. It is as if several activities that constitute reproductive behavior were coordinated by the same peptide messenger.

A further example of the mediation of behavior by a peptide is the ability of angiotensin II to bring on drinking behavior. The work of James T. Fitzsimons and his colleagues at the University of Cambridge shows that the injection of a few nanograms of angiotensin II under the skin or the injection of a few picograms into a cerebral ventricle causes behavior indistinguishable from spontaneous drinking in species ranging from lizards to primates. The animal's water balance or salt balance seems not to matter.

The discovery of angiotensin's behavioral effect adds to a considerable list of its physiological effects. Basically the events that activate angiotensin begin when any of three circumstances (low blood pressure, a low local concentration of sodium or direct stimulation by nerve fibers) causes certain cells in the kidney to secrete the enzyme renin. In the bloodstream renin acts on a protein manufactured in the liver and liberates a decapeptide called angiotensin I. In the bloodstream or in any of several organs including the brain a second enzyme cleaves two amino acids off from angiotensin I. An octapeptide remains. It is angiotensin II. Its physiological effects include the constriction of blood vessels supplying the skin and the kidneys, the dilation of blood vessels supplying muscles and the brain and an increase in blood pressure.

In addition angiotensin II causes the adrenal cortex to increase its output of aldosterone, a hormone that acts in turn on the kidneys to cause the reabsorption of sodium from the urine. Further still, angiotensin II increases the secretion of vasopressin from the posterior lobe of the pituitary. Vasopressin then promotes the reabsorption of water from the urine. All these actions tend to reverse the trends that triggered the secretion of renin. Thus they regulate three aspects of the blood: its volume, its pressure and its content of sodium. The fact that angiotensin II elicits drinking behavior is consistent with all its other effects.

According to the findings of Ian R. Phillips, who is now at the University of Florida, the cells in the brain that detect angiotensin are curious neurons in the hypothalamus and the brain stem that line the ventricles. Each one sends an extension of its cell body into a ventricle and yet each maintains typical synaptic connections with other neurons. The neurons they contact presumably include the neurons in the hypothalamus that secrete vasopressin. Remarkably, the subcutaneous or intraventricular injection of vasopressin does not elicit drinking behavior even when the experimental animal is dehydrated and presumably thirsty. On the other hand, the effects of circulating angiotensin do not account for every episode of drinking behavior. If they did, the behavior and the blood level of angiotensin would be linked. The work of Edward M. Stricker of the University of Pittsburgh suggests that this correlation is sometimes lacking. Doubtless the motivation to drink and the motor acts involved in seeking water and drinking it call for the simultaneous activity of several linked brain systems.

Although vasopressin does not cause animals to drink, it is reported to have effects on behavior that are even more dramatic. Over the past decade David de Wied and his colleagues at the State University of Utrecht and Abba J. Kastin and his colleagues at the New Orleans Veterans Administration Hospital have reported observations that vasopressin and other peptides affect learning and memory. In one experimental arrangement rats were taught not to enter a dark box by being given an electric shock whenever they did. They were reported to continue to avoid the box for a longer period after the end of the training if they were given a subcutaneous injection of a minute amount of vasopressin. Molecular segments of corticotropin or of endorphin that have no known neural or endocrine actions were said to have the same property.

A Test of Vasopressin

Together with Michel Le Moal of the University of Bordeaux, George F. Koob and I set out three years ago to test the effects of vasopressin and other peptides on behavior. We resorted to an arrangement of de Wied's in which rats were trained to jump onto a pole whenever a light bulb lit up in warning that a painful electric shock was about to be delivered through the floor. When the training was complete, the light would be lit from time to time but no shock would be delivered. Before the training began we gave one group of rats an intraventricular injection of a saline solution; in general they stopped jumping onto the pole within four hours after the end of the training. We injected a second group of rats with a solution containing a nanogram of vasopressin; they continued to jump for eight to 10 hours after the end of the training. Strangely, a third group of rats, each given 50 nanograms of vasopressin, stopped jumping before the control group. It was as if the extra vasopressin had helped them to see through our ruse.

The change in behavior brought on by vasopressin seems, then, to be unquestionable. It is not yet clear, however, that the change represents the action of vasopressin on the unknown cellular processes that underlie learning and memory. The change may have a simpler explanation. In view of the effects of vasopressin on the circulatory system we were not surprised to find that even low doses of vasopressin caused brief but immediate elevations of blood pressure. We were sent a synthetic analogue of vasopressin by Sawyer of the Columbia College of Physicians and Surgeons and Maurice Manning of the Medical College of Ohio in Cincinnati. The analogue prevents vasopressin from elevating blood pressure. The rats to which the analogue was administered along with vasopressin showed no change in blood pressure, and they behaved like the control group.

The altered performance of the rats under the influence of vasopressin could mean, therefore, that the rats were aroused by an unnatural and unnecessary elevation of blood pressure and remained tense and alert for hours. The rats were aroused, that is, by a change in the body that the brain did not request. This seems less impressive than the hypothesis that vasopressin mediates learning and memory directly. On the other hand, a mismatch between the brain's commands and the body's responses may be sufficient to invoke behavioral strategies for dealing with novel situations. People given long-lasting analogues of vasopressin report enhanced attention to their surroundings, and their performance on tests of memory improves.

Moreover, even on the simpler hypothesis a neuropeptide apparently serves to signal that the survival of the animal is challenged and that the animal had best be attentive to its surroundings. Conversely, the absence of such a signal may suggest that the animal is momentarily safe. Survival signals such as these could represent a means by which the neuropeptides guide the evolution of complex forms of behavior. Surely the various aspects of research on peptide messengers are likely to advance understanding both of cellular regulation and of the means by which certain types of animal behavior result from decipherable cell-to-cell interactions.

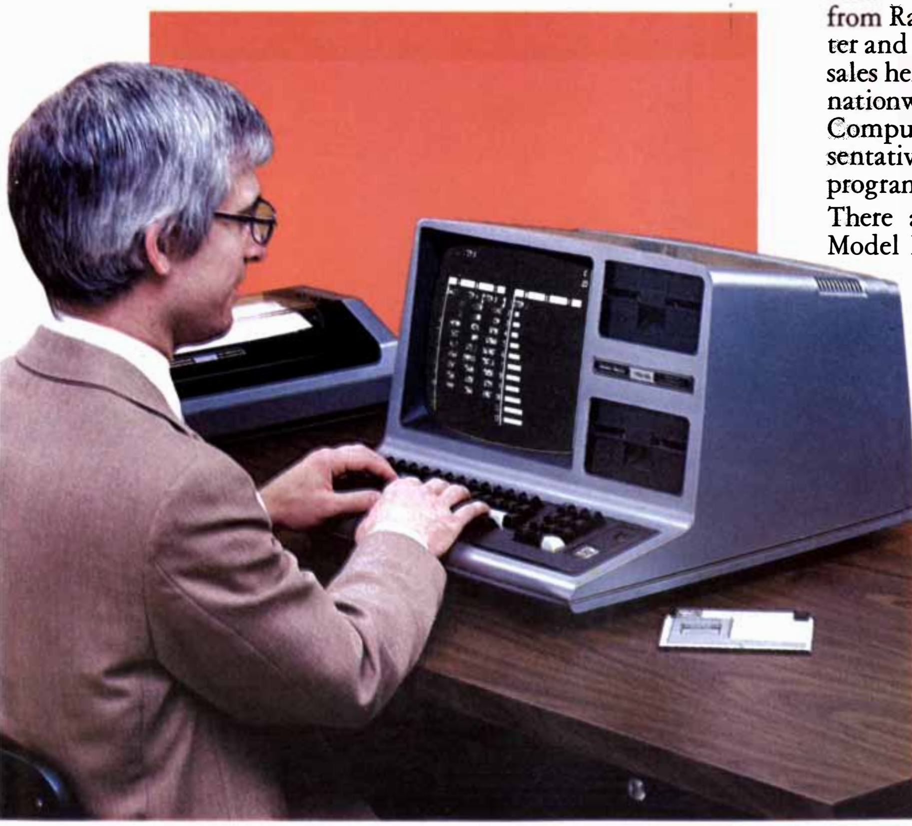
You'll Find More Software for the TRS-80 than Any Other Microcomputer!

That's Why Radio Shack's $2495 Desktop Business Computer Belongs in Every Office!

Radio Shack currently offers over 80 ready-to-run programs for our TRS-80 Model III Desktop Business Computer. That's more than four times the selection our nearest competitor offers! More software means you can put a Model III to work solving a variety of problems, without ever having to write a single program.

Increase Your Personal Productivity

Radio Shack's business software is designed to streamline time-consuming day-to-day tasks. Just use our VisiCalc® program for professional forecasting and modeling. Perform "What if . . ?" calculations quickly and easily. Create your own customized filing system with our Profile program and your Model III can find just the records you need—in seconds! Add a printer and our SCRIPSIT ™ program and you have a personal word processing system for correction-free memos and reports. But there's more!



Unlike some companies, Radio Shack doesn't believe in being your only source for software. Our $2.95 Applications Source Book lists over 2,300 programs for the TRS-80 that are available from outside vendors. It's a goldmine of software!

Industry-Wide Support

Because more people use TRS-80's, the computer industry is supporting the TRS-80 with a wide range of accessories and specialty items. Just thumb through any computer magazine to discover how widely accepted our TRS-80 is. (You'll even find magazines devoted entirely to the TRS-80!)

And all of this is in addition to the support you get from Radio Shack, including a monthly user newsletter and toll-free technical "hotline". There's friendly sales help at over 8,000 Radio Shack stores and dealers nationwide. And a network of over 185 Radio Shack Computer Centers staffed by company-trained representatives who can tell you all about our training programs, leasing plans and service agreements.

There are still more reasons to choose a TRS-80 Model III Desktop Computer. Like all the "extras" you get at no extra cost: 48K internal memory, two mini disk drives with 306,000 characters of storage space, powerful Model III BASIC language, a 12″ monitor, a 12-key data pad, full expansion capabilities—all in a beautiful, compact, *1-piece* package.

See it Today

For a hands on demonstration, visit your nearest Radio Shack Store, Computer Center or authorized dealer. Or write Radio Shack, Department 82-A-110, 1300 One Tandy Center, Fort Worth Texas, 76102.

Radio Shack®

The biggest name in little computers ™

Retail prices may vary at individual stores and dealers.
VisiCalc is a trademark of Personal Software Inc.

Radio Shack's TRS-80® —The Best Supported Small Business Computer in the World.


TOP AND BACK of a finished violin made by the author are shown from the outside, as opposed to the painting on the cover, which portrays the inside of unattached top and back plates being tested acoustically. A top plate for an instrument of the violin family is traditionally made from two pieces of straight-grain, quarter-cut spruce. A back plate is usually made from two joined pieces or a single piece of curly maple. The outer surfaces of the plates are treated with sealer and varnish, but the insides are left "in the white," that is, untreated.

The Acoustics of Violin Plates

Modern tests of the vibrational properties of the unassembled top and back plates of a violin reveal something of what violinmakers do by "feel" and lead to the making of consistently good violins

by Carleen Maley Hutchins

One of the great mysteries of music is how the renowned violinmakers of past centuries, apparently having had no more than a practical knowledge of the physics and acoustics of their instrument, could turn out violins that are still cherished today for the beauty of their sound. For some 30 years a small but worldwide group of us in an organization named the Catgut Acoustical Society have applied modern methods to the study of the physics and acoustics of violins and other stringed instruments. I described the early stages of the work in these pages nearly 20 years ago [see "The Physics of Violins," by Carleen Maley Hutchins; SCIENTIFIC AMERICAN, November, 1962]. Now the work has progressed to the point where a good deal can be said about the properties of the top and back plates (the "belly" and the "back") of a violin before they are assembled into an instrument. Moreover, it is possible on the basis of these findings to construct violins and other members of the violin family with consistently fine tone and playing qualities.

The two plates are carved traditionally from solid blocks of wood, the top plate from two adjacent pieces of straight-grain spruce (*Picea abies*) joined down the middle, and the back plate from either a single piece or joined pieces of maple (*Acer platanoides*) that usually have a "flame," or curl, across the grain of the wood. Yet because of the variations in wood, even between two adjacent pieces from the same tree, it is not possible to reproduce measurement for measurement the parts of a fine-sounding violin and so to create an instrument with the tone and playing qualities of the original. The way to duplicate a fine violin lies not in geometrical measurements alone but must include measurements related to the vibrational properties of the wood.

The long-term investigation described here draws heavily on the experience of violinmakers and provides some new answers to a question asked in 1830 by Félix Savart, a physician and physicist: "What sounds ought the top and back of a violin have before they are joined?" Through the generosity of the eminent French luthier Jean Baptiste Vuillaume, Savart was able to test the disassembled top and back plates of a dozen or so violins that had been made by Antonio Stradivari and Giuseppe Guarnieri. (Imagine!) He applied a measuring machine he had devised, together with a technique worked out by his friend Ernst F. F. Chladni. By the Chladni method the eigenmodes, or normal mode patterns of vibration, of a horizontally mounted flat plate can be demonstrated by sprinkling the plate with a fine powder and causing the plate to vibrate. At certain frequencies (the eigenfrequencies) the vibration bounces the powder into the nonvibrating nodal areas, thereby outlining the nodal and antinodal configurations of the plate at its specific resonance frequencies. These plate resonances, or normal modes, are created by the physical properties of stiffness and mass, which cause standing-wave patterns to be formed in response to vibration at discrete frequencies unique to each plate. In answer to Savart's question he reported: "We have found that the sound varies in good violins between $C^{\#}$ 3 [the 3 indicates the octave] and D 3 for the belly, and for the back between D 3 and $D^{\#}$ 3, so that there is always a difference between them of a half or a whole tone."

Over the years other investigators have made vibrational measurements of violin plates, both free and in the assembled violin, and have assessed the resulting tonal characteristics. Particularly notable is the work of the acoustician and violinmaker Hermann F. Meinel in Berlin during the 1930's, which documents the correlation between the thickness of the plates and the vibrational modes, the volume of sound and the timbre of sound. Meinel also recognized the limits of constructing violins on an empirical basis and noted the effects of the properties of the wood, the arching of the plates and the varnish. He explored the possibility of improving a particular violin in a given frequency range by removing wood from a specific area, following the work of Hermann Backhaus, but concluded that improvement does not always result because it depends on the physical state of the violin. This early work highlights a basic problem in violinmaking: a small change that will markedly improve one instrument may adversely affect another owing to the widely varying configuration of the vibrational modes and the stiffnesses throughout the plates.

In 1950 the Harvard physicist Frederick A. Saunders and I began a collaborative study aimed at verifying Savart's findings and at developing other vibrational measurements that relate the unique bending characteristics of each pair of free top and back plates to the particular tone and playing qualities of the assembled instrument. By 1960 the results of some 200 tests on violins and violas in the course of construction had confirmed Savart's main finding: the instrument has good musical qualities when the principal plate tone of the top and of the back are from a tone to a semitone apart. Violinmakers call the principal plate tone the tap tone because it is the principal tone heard when the plate is tapped. Our findings showed in addition that the actual frequencies can vary considerably and that the top plate's tap tone can be higher than the back's or vice versa and still result in an instrument with good tone.

These observations did not go far enough toward explaining our finding that when the pairs of free plates were assembled, the resulting instruments did not always possess the expected tone and playing qualities. Every now and then an instrument proved without apparent reason to be much better than the others or worse. Seeking to explain these inconsistent results, I have continued the research since Saunders' death in 1963, building and testing 160 more instruments of the violin family. (The family includes the traditional violin, viola, cello and bass. Some new and revised in-

AFTER YOU TEST DRIVE YOU THINK IS BEST, WHY NOT THE SPORTS SEDAN "ROAD &

Astute as you may be at weighing the relative merits of automobiles, there is still a lot to be said for the people who make a living at it.

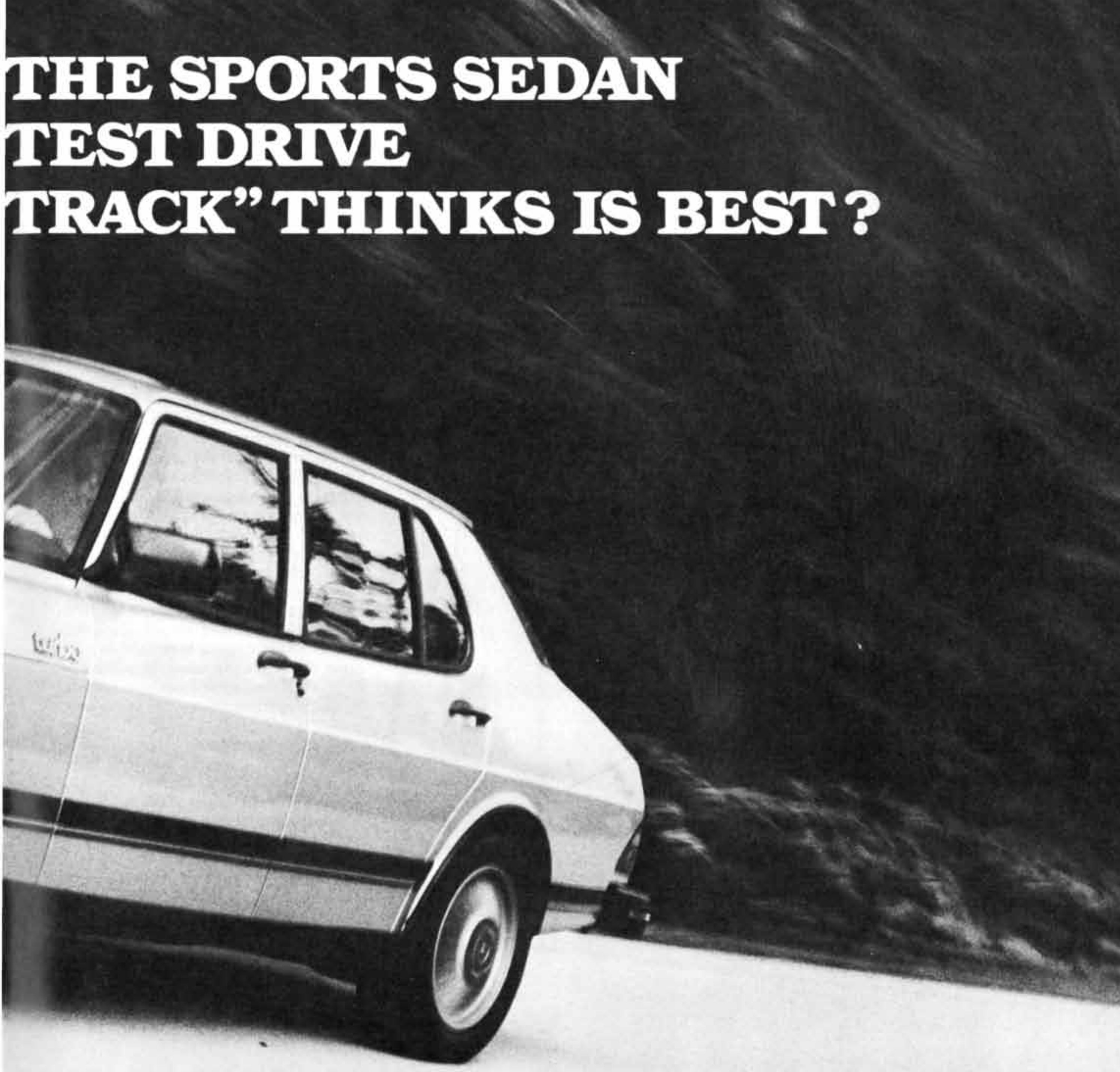
And people who make a living at it, the editors of *Road & Track*, recently chose the Saab 900 Turbo as the Sports Sedan for the Eighties. Unanimously.

Ho hum, a few cynics might say, having been inundated with Car of the Year, Car of the Month, Car of the Century designations over the last few years.

Resist the urge to join the cynics.

First of all, the category *Road & Track* considered is the very category where discriminating car buyers are heading in droves.

Because Sports Sedans are the only cars that seem to combine the imperatives (economy, room, reliability, safety) with



the discretionaries (performance, handling, enjoyment).

Second of all, the editors did have to go out on the limb a little. Nobody's eyebrows would have been raised if they had chosen Audi or BMW or Mercedes or Alfa.

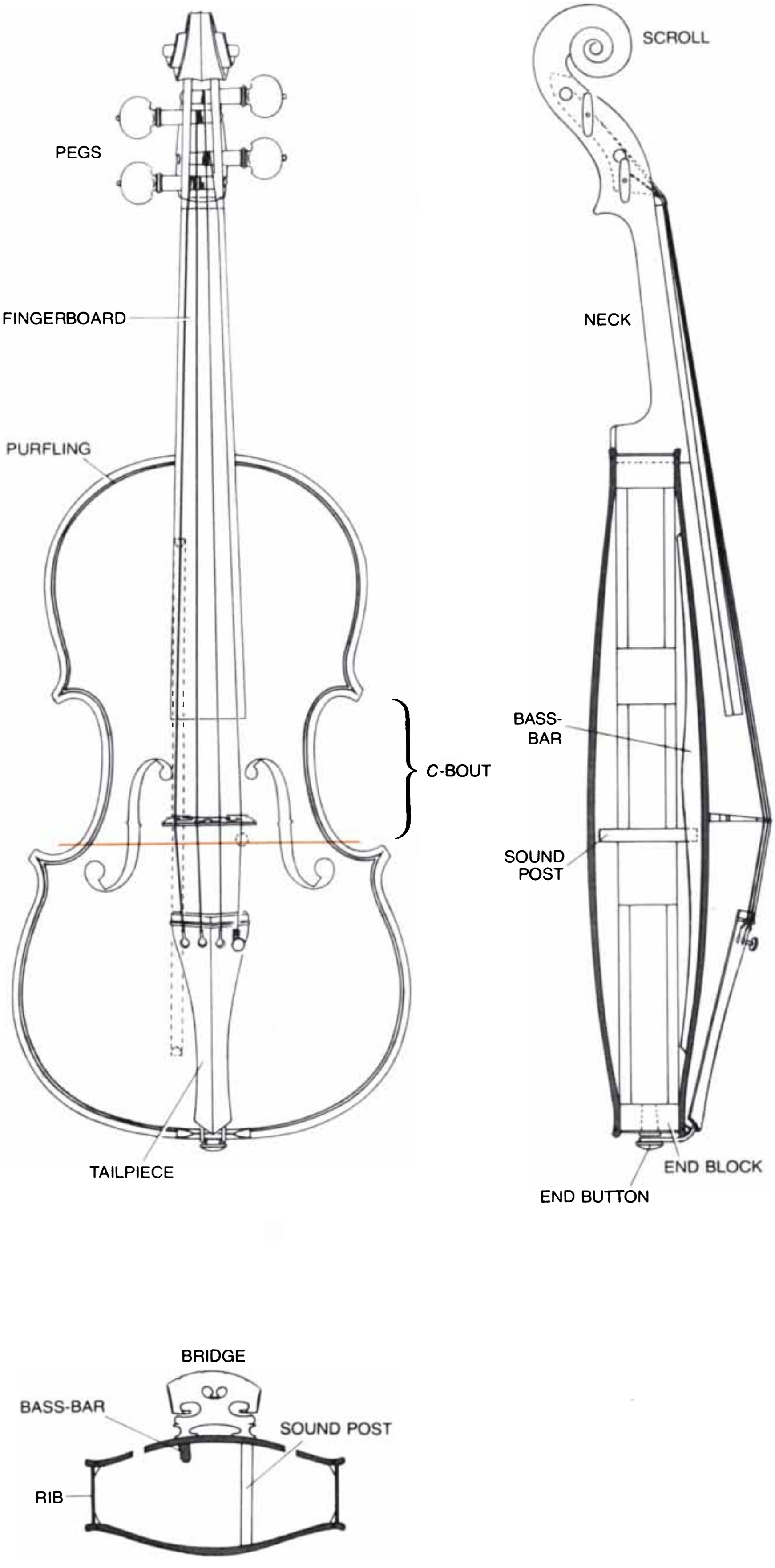
But Saab? The cult car? The car with the ignition key on the floor? The car that offered front-wheel drive when it was an oddity? The car that pioneered turbocharging before OPEC made it a necessity for rational performance?

At the very least, a long and thoughtful test drive would seem to be indicated.

Either that, or resign yourself to never knowing just what came over those *Road & Track* editors when they got behind the wheel of a Saab.

SAAB

THE MOST INTELLIGENT CAR EVER BUILT.



COMPONENTS OF A VIOLIN are depicted in views from the top, in cross section and from the side. The colored line in the view of the top shows the location of the cross section. Except for size and proportions the components of a violin, a viola and a cello are much the same.

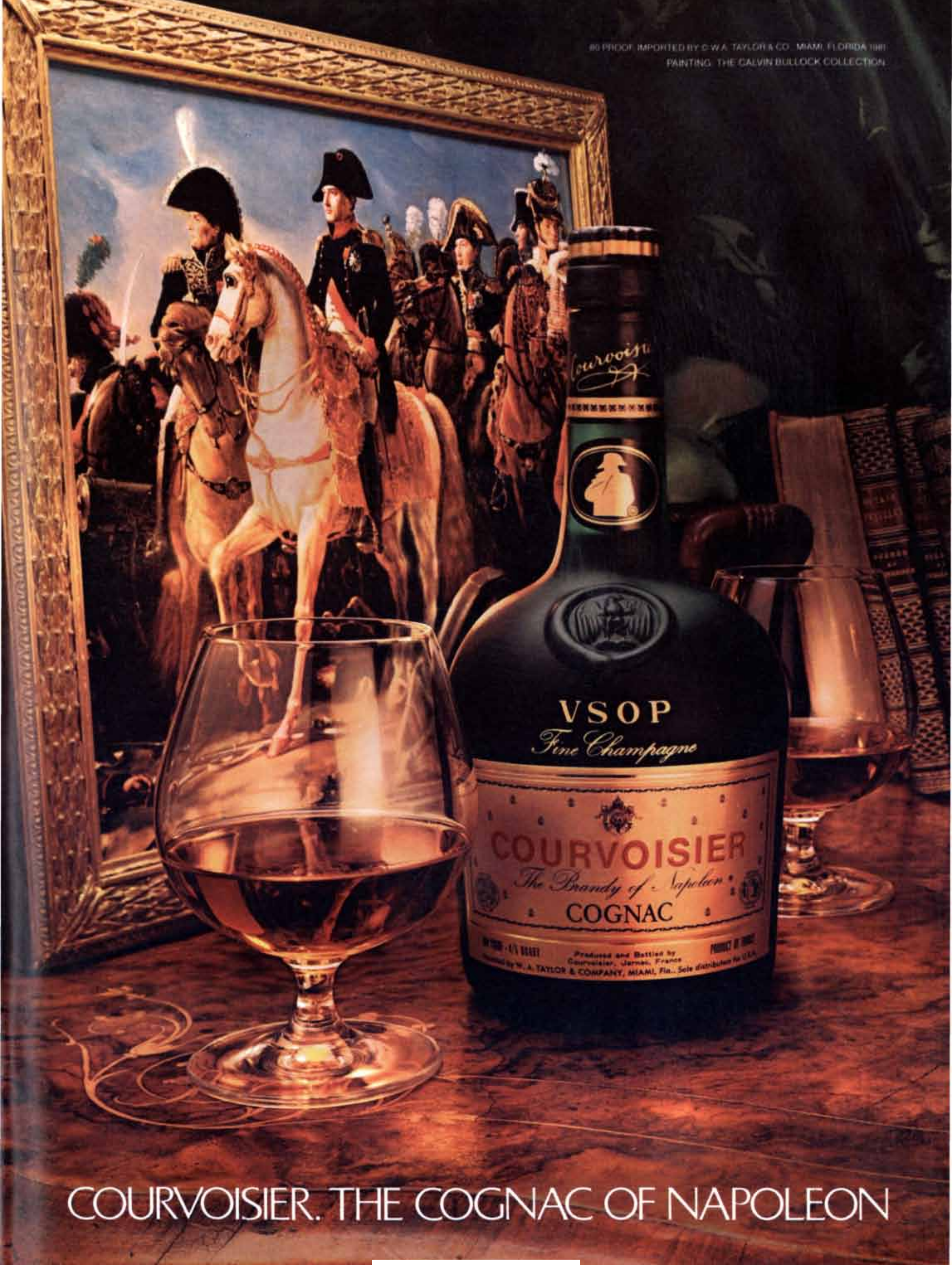
struments, developed with the test methods described here, make up the "violin octet," all designated as violins: treble, soprano, mezzo, alto, tenor, baritone, small bass and contrabass.) I have examined the instruments not only by the Chladni method but also by such new techniques as hologram interferometry and real-time analysis. It cannot be emphasized too strongly that before one can apply such methods one must learn the violinmaker's craft, so that the basic instrument is built according to the principles of fine violinmaking. I learned violinmaking in the 1950's under the tutelage first of Karl A. Berger and then of Simone F. Sacconi with Rembert Wurlitzer's encouragement. It was eight years of slow, painstaking work.

The vibrational properties of each finished top and back plate are the result of the life history of that particular piece of wood. (The care and lore a violinmaker brings to the selection of a suitable tree are a story in themselves.) Tradition also calls for long seasoning of the flitches (log-size pieces of wood of appropriate length split or sawed longitudinally in "quartered" sections from the trunk) in stacks in outdoor sheds, spruce for from five to 10 years and maple somewhat longer. Some makers maintain that the wood should be seasoned for at least 50 years. This judgment may be supported by the suggestion of several wood technologists that the ratio of crystalline areas to amorphous areas in the cell structure of wood seems to increase as the wood seasons. This concept fits nicely into violinmaking traditions, because amorphous material absorbs and loses water readily but crystalline material does not. Perhaps the relation helps to explain why many older instruments are less susceptible to changes in moisture than new ones.

What physical properties of top and back wood, so carefully selected by the violinmaker, are most important to the sound of a fine instrument? Technical workers in this area generally agree on five: elasticity along and across the grain; shear; internal friction (damping) resulting in the dissipation of energy; density, and the velocity of sound in the wood.

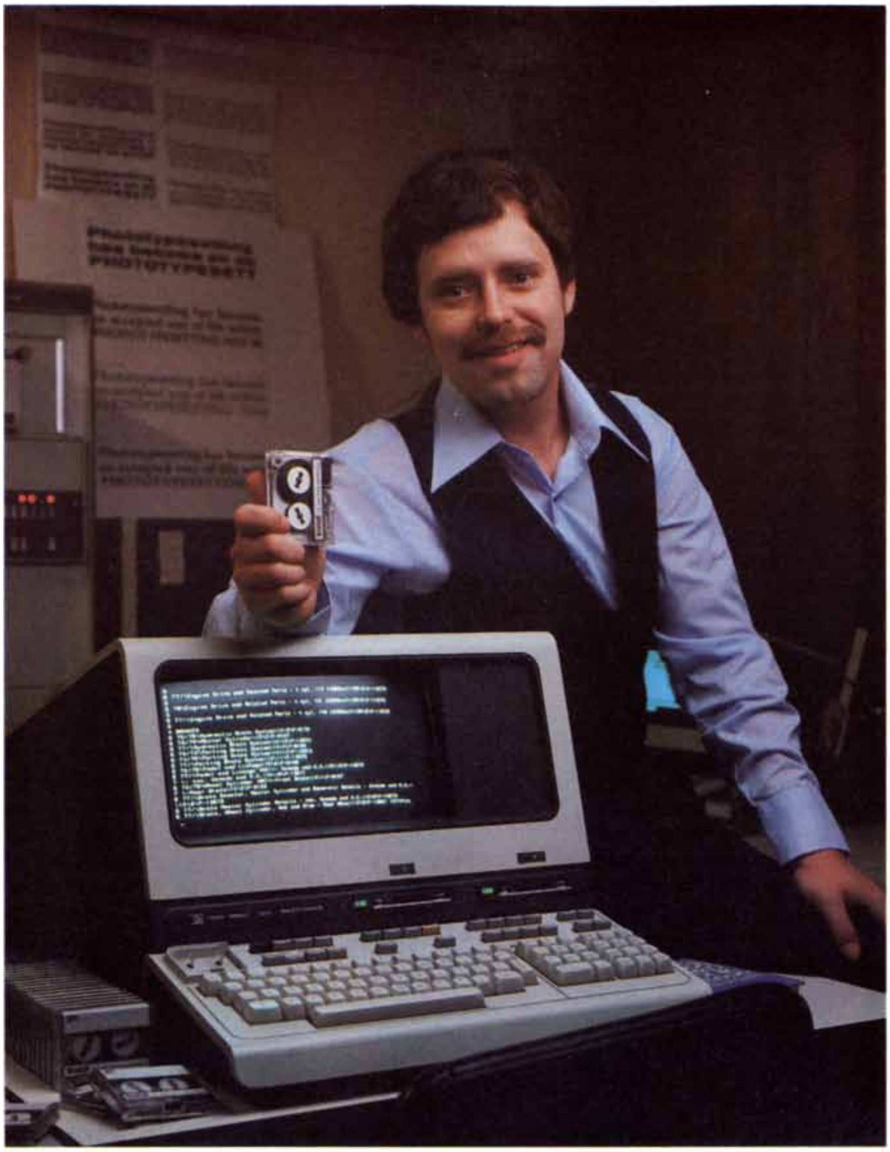
The most important aspects of elasticity are the values of Young's modulus along and across the grain. Young's modulus is a measure of the material's resistance to local bending and stretching; it is the ratio of the local force applied per unit area to the resulting relative change in length. Shear is a measure of resistance to angular distortion of the kind one sees when the top of a thick book lying on a flat surface is pushed sideways, shifting the upper surface with respect to the lower one.

Internal friction or damping is a measure of the ratio of energy dissipated to



PAINTING THE CALVIN BULLOCK COLLECTION
VSOP
Fine Champagne
COURVOISIER
The Brandy of Napoleon
COGNAC
Produced and Bottled by
COURVOISIER. THE COGNAC OF NAPOLEON

"Our reputation rests on digits, decimal points, and details. We wouldn't trust them to anything less than Scotch® Brand Data Cartridges."



Bill Birkett, Vice President,
Trade Graphics, Inc.,
Livonia, Michigan

The unique design of a data cartridge provides great reliability, high storage capacity and long tape life. And where could you possibly get better data cartridges than Scotch Brand, made by 3M, the people who invented the data cartridge system itself?

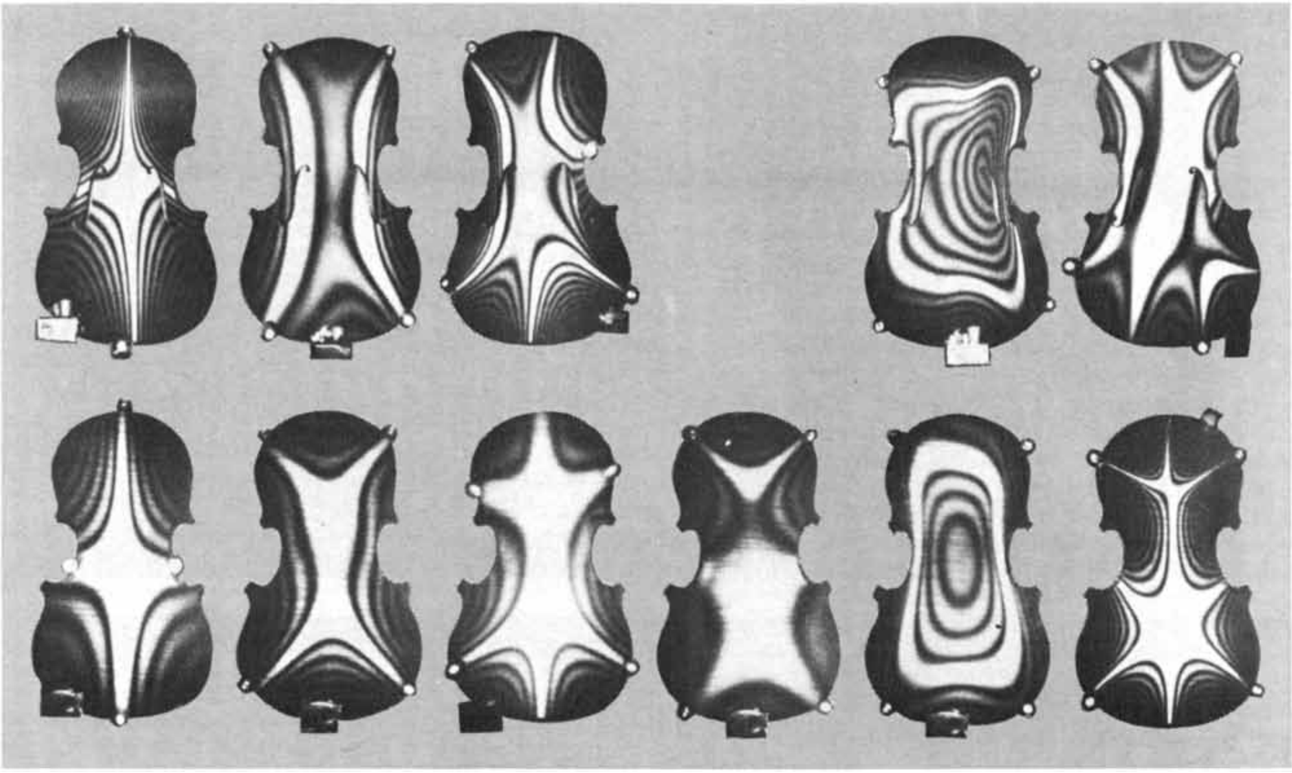
3M controls every step in manufacturing. Top quality magnetic tape and precision components are part of every Scotch Data Cartridge. Over twenty-five years of service to the computer industry assure you of the utmost reliability.

Scotch Data Cartridges are available in miniature DC 100A, the standard-size DC 300A and now, an extra-length DC 300XL with 50% more storage capacity. They are compatible with most cartridge systems including Hewlett-Packard, IBM, NCR, Tektronix and TI.

To find out where you can find Scotch Data Cartridges or virtually any other data recording medium, call toll-free: 800-328-1300. (In Minnesota, call collect: 612-736-9625.) Ask for the Data Recording Products Division.

If it's worth remembering, it's worth Scotch Data Recording Products.

3M



VIBRATIONAL PATTERNS of a pair of unattached top and back plates for a violin (the top complete with f-holes and bass-bar) are made visible by hologram interferometry. Starting with the mode of lowest frequency, they are designated mode 1, mode 2 and so on. The upper row of interferograms shows the mode shapes at the particular frequencies where they occur in this top plate: respectively 80, 147, 222, 304 and 349 hertz. (Mode 4 is missing.) The lower row shows the first six modes in the back plate at 116, 167, 222, 230, 349 and 403 hertz. Many other resonance modes occur at higher frequencies. In all the instruments of the violin family the configuration of the lower resonance modes is quite consistent, but the frequencies differ depending on the dimensions, mass and stiffness of the plate.

energy stored elastically. It can be expressed in several ways. One is by the decay time, or the time during which vibration persists after excitation is cut off; a violinmaker listens for a long decay time in the tap tone as he tunes a violin plate. Another is by the width of the frequency interval within which there is a response to continuous excitation as the frequency is varied about a resonance. Damping is often expressed as the "quality factor," or Q. The higher the Q value, the lower the damping.

Density is weight per unit volume and is found by multiplying the length, width and thickness of a rectangular strip of wood and dividing the product into the weight of the strip. Velocity is found by dividing Young's modulus by the density and taking the square root. One of the desirable characteristics of spruce for the soundboard of a musical instrument is its high ratio of stiffness to density, reflecting a high sound velocity.

Two main problems arise in the scientific investigation of violin plates. First, what physical mechanisms are activated in free top and back plates as they are flexed, held and tapped? Second, can measurement of these mechanisms and their component factors yield practical information about the resulting tone and playing qualities when a given pair of plates is assembled into a finished instrument?

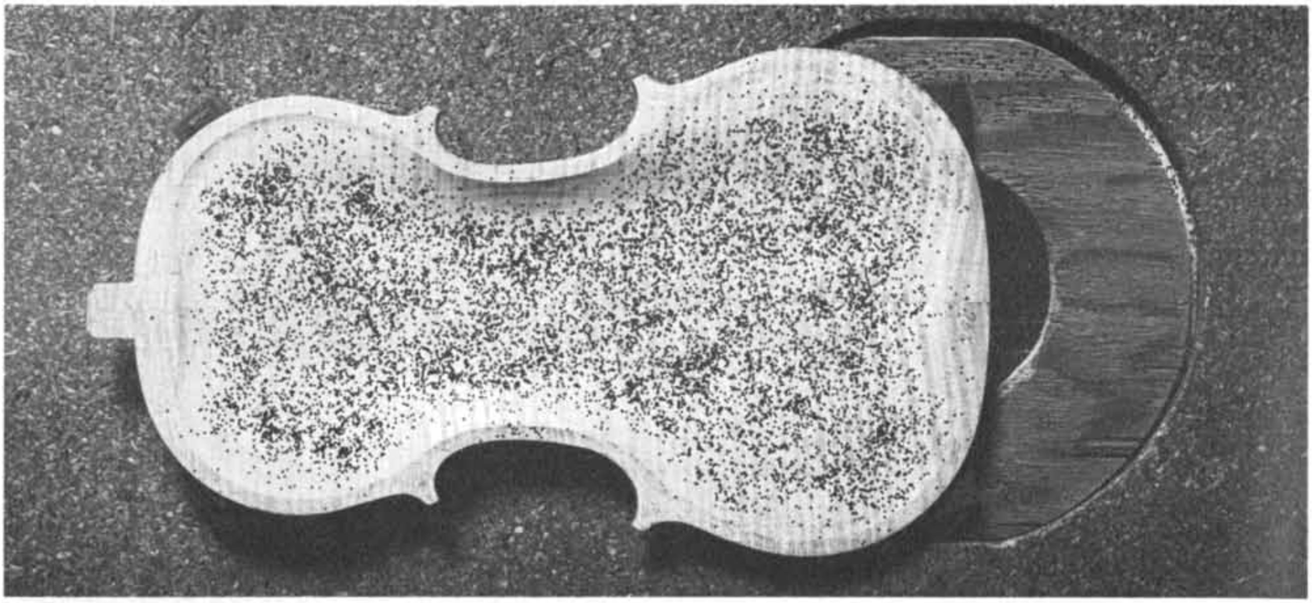
In our effort to find answers to these questions we have explored the plate mechanisms and made several thousand comparisons of the eigenmodes of free plates with the tone and playing qualities of the instrument assembled from each pair of plates. The main characteristics of the eigenmodes are easily made visible by our modified Chladni technique. A free plate is placed horizontally over a loudspeaker, with the inside of the plate facing upward like a dish. A sine-wave signal (a single-frequency signal) is directed through the speaker, sweeping across the frequency range of interest and causing powder sprinkled on the plate to assume characteristic patterns at discrete frequencies that are unique to each plate.

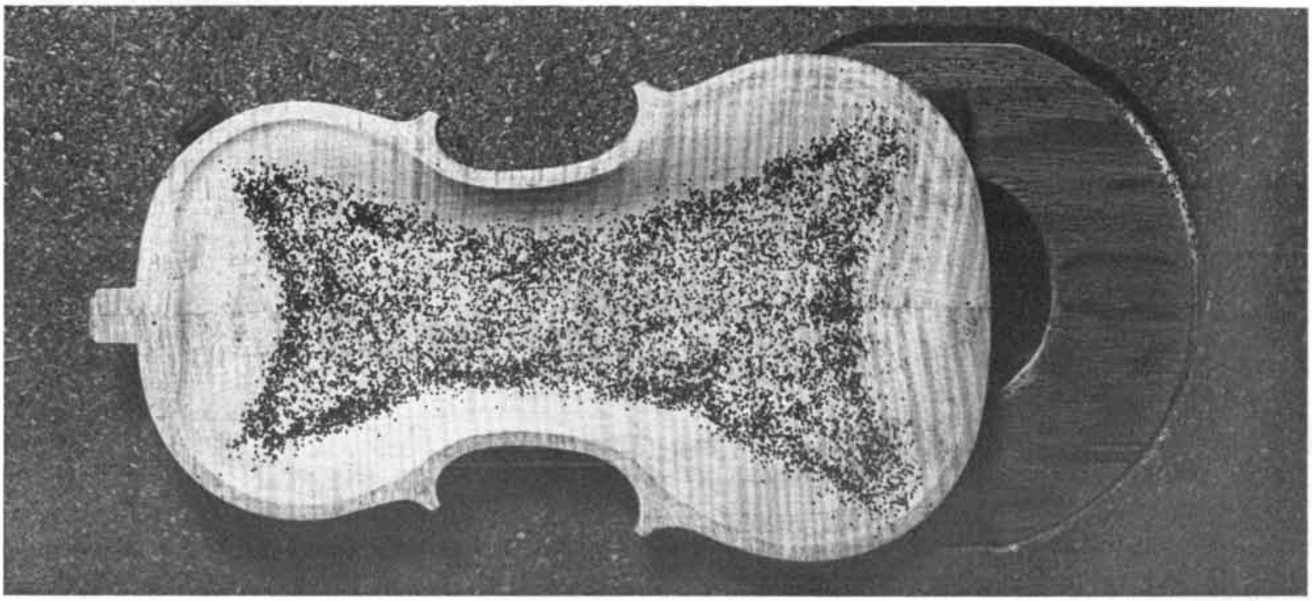
An even clearer understanding of the configurations of the modes emerged when lasers made it possible to apply hologram interferometry to violin plates. This line of experimentation was first pursued with free violin plates in the late 1960's by Karl A. Stetson, who made interferograms showing the bending modes (some with amplitudes of only a few micrometers) as the plates were vibrated at their discrete frequencies of resonance.

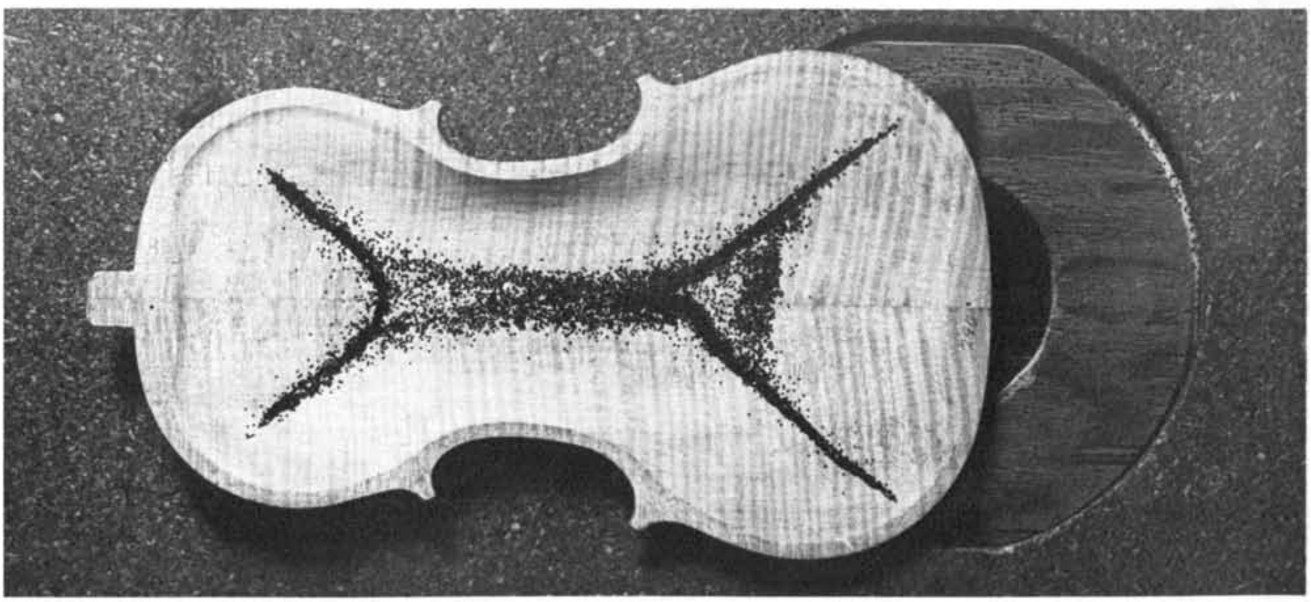
By each method the mode shapes for instruments of the violin family are found to follow a similar sequence in all sizes of free plates. They have therefore been designated mode 1, mode 2 and so on, starting with the mode of the lowest frequency. Although the mode shapes are similar throughout the violin family, the frequencies at which they occur are unique to each plate. In general the larger the plate, the lower the mode frequency, but even between plates of the same size there is considerable variation in the mode frequencies.

The modes that have so far been found to be most important in tuning violin plates are 1, 2 and 5. Mode 1 entails a twisting of the plate, with one corner up and the other down in opposite phase. Thus when a violinmaker holds a plate at each end, twisting it between his hands to feel its "resistance," he is actually sensing the main stiffness characteristics of mode 1.

When a maker holds one end of a plate in both hands with thumbs on top and fingers spread out underneath across the wood, squeezing it and bending it slightly to assess the cross-grain stiffness of first one end and then the other, he is comparing the relative stiffness of mode 2 in the two ends. Some





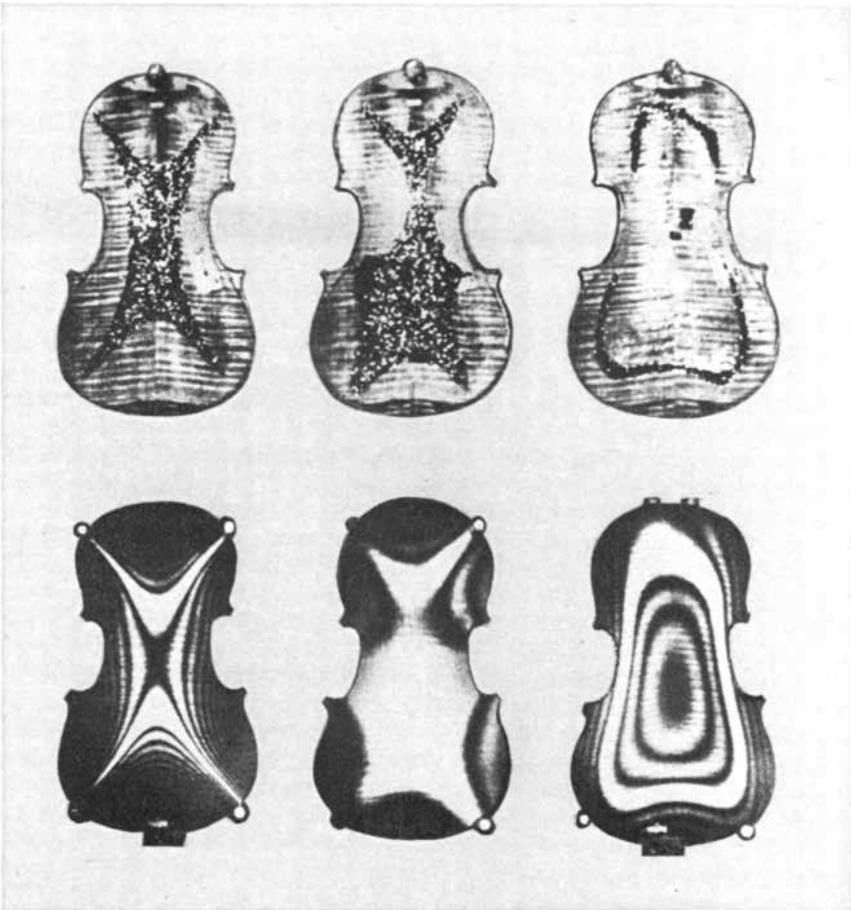


makers achieve essentially the same result by laying a plate arch side up (the plates are arched outward in the finished instrument) on a flat surface and putting a shallow dish of water first on the upper area of the plate and then on the lower area, pushing gently on the plate to compare how much the water in the dish moves in each case.

When a luthier holds a plate around the two ends in his fingertips and pushes down in the middle with his thumbs, he is actually checking the principal stiffness of mode 5. The same test can be made by holding the plate around the edges and gently pressing the top of the arch against a flat surface to feel the bending.

Holding the plate at the midpoint of one end and tapping with the soft part of a finger around the upper and lower edges will activate the sound of mode 1 quite clearly, because the holding point is a node and the curves of the upper and lower edges are antinodes for that mode. Holding at one of the four points where the nodal lines of mode 2 intersect the edges and tapping on the antinodal area near the midline of either end of the plate activates primarily mode 2. Holding at a point along the nearly oval nodal line of mode 5 and tapping in the center of the plate causes the sound of mode 5 to predominate.

All the modes, however, contribute in greater or lesser degree to the sound heard when the plate is tapped, depending on the place of excitation and the effect of holding. For example, when a violinmaker holds a plate between thumb and forefinger near one end just off the midline and taps it in the center, listening for a clear and full ring, he often finds it necessary to move his holding point a bit to attain optimum sound. The best holding point is at the crossing of the nodal lines of modes 2 and 5; then tapping in the center activates primarily mode 5 and tapping at the lower or upper end of the plate activates primarily mode 2. In a well-tuned plate contributions are heard from both modes at each tapping place, with the resulting sound's being particularly clear if modes 2 and 5 are an octave apart. In a plate that is not well tuned it is often difficult to identify by ear the pitch of the sound that predominates. These variations can help to explain why there are so many different interpretations among violinmakers of the sounds in a violin plate and so many views of what to do about them.



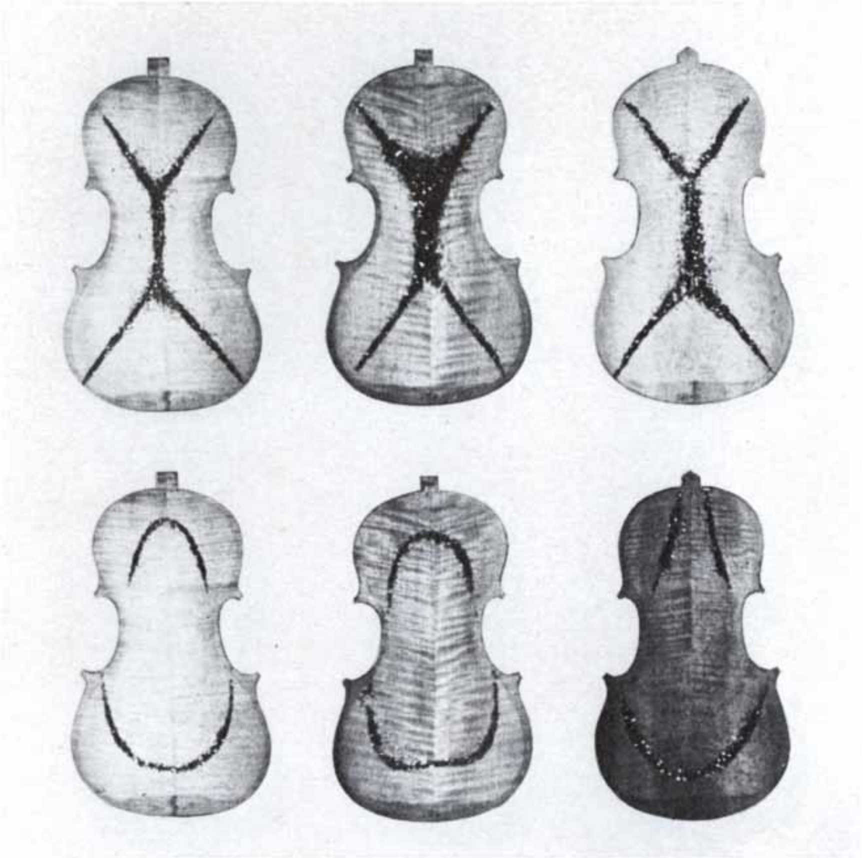
TWO TECHNIQUES for demonstrating the eigenmodes of a free violin plate are compared. In the photographs at the top the Chladni patterns of a back plate are shown at frequencies of 165, 225 and 357 hertz. Laser interferograms of the same plate at a different relative humidity are shown in the photographs at the bottom at frequencies of 165, 222 and 348 hertz. The nodal patterns, which show in the interferograms as wide white areas, are indicated by the quite similar dark shapes in the Chladni patterns. The laser technique not only is more sensitive than the Chladni one to small vibrations of a top or back plate but also indicates motion in the antinodal areas of a plate by the narrow and dark lines of the characteristic interference fringes.

The process of tracing the evolution of the eigenmode characteristics from a pair of free violin plates to the assembled instrument in playing condition is extremely complicated and not yet clearly understood. The theoretical analysis of even one free violin plate must take account of at least nine parameters, the calculation of which would take a great deal of technical skill, to say nothing of more time and money than are currently available.

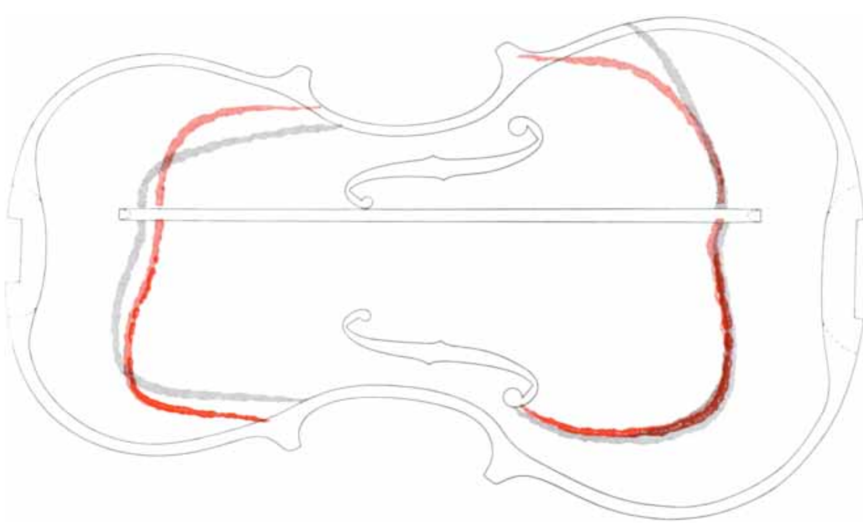
In the assembled violin the two plates are glued to the ribs (sides) of the instrument. The resulting edge constraints alter the modes of the plates in many ways. Moreover, a new set of coupled resonances is created from the interaction of the two plates through the ribs and the sound post (a pencil-size rod of spruce held by friction between the top and the back and positioned almost under the foot of the treble side of the bridge). Additional resonances are created by couplings between the wood of the box and the modes of vibration of the air mass within the box.

To discover the effect of various free-plate mode characteristics on these complicated constraints and couplings we have employed the long, slow process of making violin-family instruments of all sizes. One carefully selects the wood, forms it into plates, tunes the

CHLADNI METHOD of displaying the eigenmodes of a free (unattached) violin plate is demonstrated in the author's workshop. In the photographs on the opposite page the plate, a violin back, is mounted inside upward on four soft foam pads over a loudspeaker. Each pad is adjusted to support the plate at a nodal (nonvibrating) point and the speaker is centered under an antinode for the mode being tested. Particles of thin aluminum flake or some other powder are sprinkled on the plate (*top*). When the plate mode resonates in response to the appropriate single-frequency sound from the speaker, the particles begin to bounce (*middle*). The vigorous bending motion of the antinodal areas bounces the particles into the nonvibrating nodal areas of the plate, thereby outlining the nodal and antinodal configurations characteristic of that particular mode. In these photographs one sees the development of a mode-2 pattern (*bottom*). The technique, a modified version of a testing method devised in the 18th century by Ernst F. F. Chladni, was also used to make the patterns of modes 1, 2 and 5 shown on the cover.



WELL-TUNED AND POORLY TUNED PLATES are revealed by means of the Chladni technique. Each pair of photographs, top and bottom, shows modes 2 and 5 respectively of a violin back plate. In the plate at the left both modes are well tuned. In the plate at the center the nodal areas of mode 2 are too wide in the upper section, indicating that the entire upper area of the plate is too stiff. In the plate at the right the nodal areas of mode 5 extend straight to the upper edge instead of closing into a ring shape. This usually happens when the plates are too thick through the upper section of the center between the C-bouts and inside the upper corner.



EFFECT OF VARNISH on the outer surface of the detached top plate of a viola is evident from the change in the Chladni pattern of mode 5 on its inner surface. The colored lines show the pattern of mode 5 in the tuned plate before assembly, when it already had sealer and two coats of varnish on it. The gray lines show the pattern in the detached plate after the viola had been completely finished (with a total of seven coats of oil varnish) and had been played for two years. Varnish and sealer help to protect the wood and reduce somewhat the effects of varying humidity, but they also change the tonal characteristics of the instrument. The tonal characteristics continue to change over a period of approximately two years as the varnish hardens.

eigenmodes of the plates, assembles the instrument, evaluates it and then in many cases takes the plates off, retunes them and reassembles the instrument, repeating the technical and musical evaluations.

The possibility of testing the free plates of a fine concert violin always looms large as a way to obtain information on the free-plate eigenmodes that are coupled into such instruments. It is highly desirable to test the top and back plates at the same time because of changes in the wood with changes in temperature and relative humidity, but obtaining the detached top and back plates of a fine violin simultaneously is almost impossible. Even in an extensive repair job a violinmaker seldom removes both top and back unless the instrument is in very bad condition.

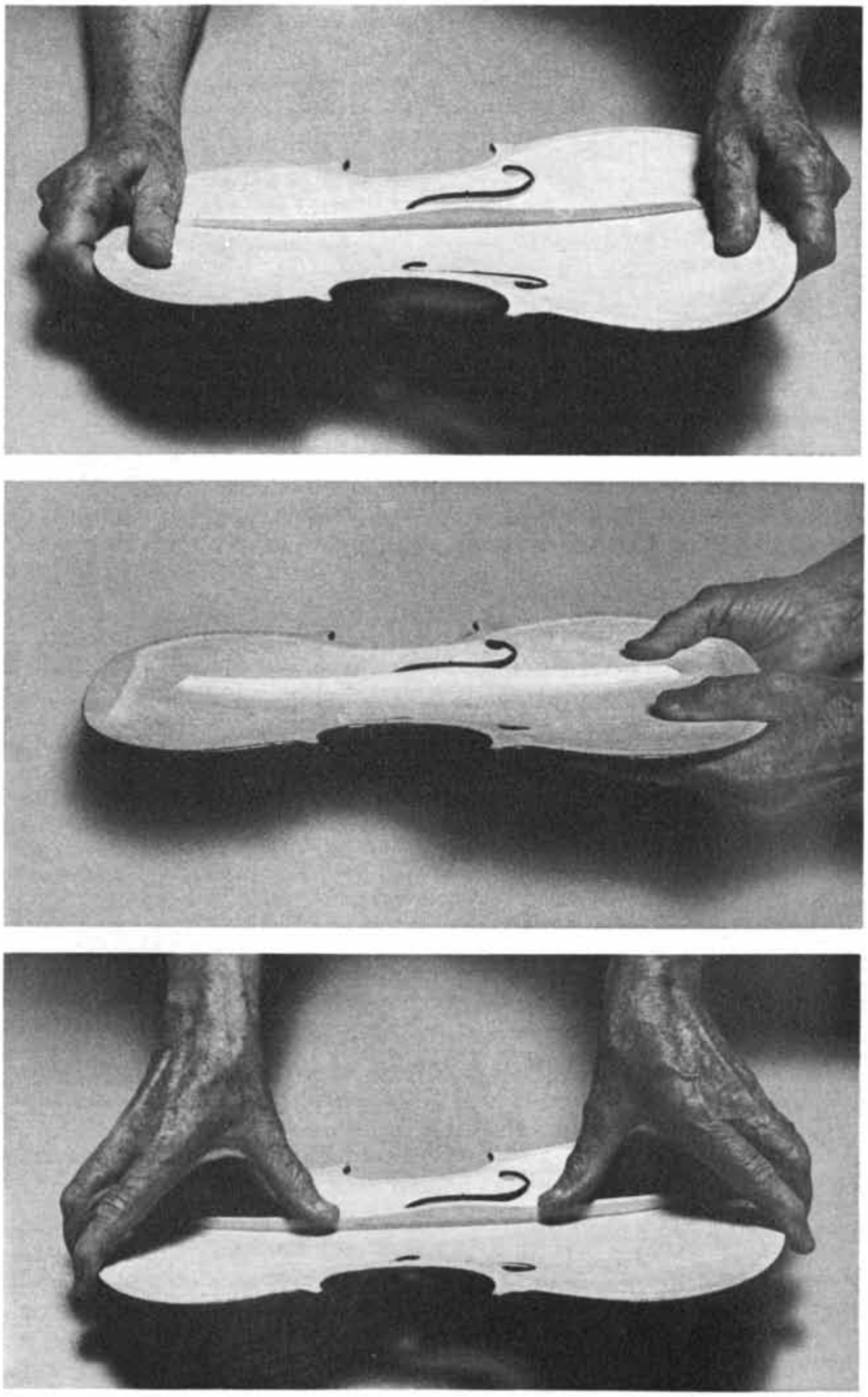
Through the kindness of two violinmakers, however, we have been able to test the free-plate pairs of two concert violins, a Stradivarius of 1713 and a Guarnerius del Gesù of 1737, with comparison tests after the instruments were reassembled. Considerable information was gained from these two violins, but in each case the repairs were so extensive that gaining any knowledge of their original condition was out of the question. Moreover, in view of the changes that have been made in violins built before 1800 in order to increase their tonal output (longer neck, increased fingerboard angle, higher bridge and appropriately heavier bass-bar, with invariably some scraping of the inside of the top plate) there is little chance of ascertaining what their original makers intended.

In the course of our studies we have tested (in playing condition) many fine modern instruments as well as early ones, thanks to the interest and cooperation of their owners. In this way we have accumulated a body of knowledge based on more than 800 tests of all types of instruments of the violin family with a wide variety of musical potential. For example, the response curve of a famous Guarnerius del Gesù violin made in 1731 and that of a violin recently constructed on a Stradivarius pattern show quite similar characteristics. The curves reflect a decrease in amplitude of the resonances through the 1.5-kilohertz range and an increase in the range from two to three kilohertz. This characteristic has been reported by Meinel and others as being typical of the response curves of some of the most musically desirable violins.

Our tuning process for free plates starts with a pair of nearly finished top and back plates for an instrument of the violin family. The outside archings are planed and scraped to their final contours. For the top plate the f-holes are cut and the bass-bar is installed (but not shaped); the purfling (the three strips of

black and white wood inlaid around the edges of the two plates) is glued in, and the edges are finished. It has also been found desirable for the sealer, or filler, and at least two coats of varnish to have been on the outside of each plate for several months.

As the violinmaker planes and scrapes away the wood from the inside of a pair of violin plates so that the top is from three to four millimeters thick and the back from three to six millimeters thick, he begins to flex each plate in his fingers, holding it and tapping it in various ways.



MANUAL TESTING for the characteristics of wood that give rise to modes 1, 2 and 5 entails a flexing of the plate in various ways by the violinmaker. When the maker holds the plate at each end and twists one corner up and the other down several times (*top*), he is testing for the stiffnesses that produce mode 1. When he holds one end of the plate in both hands with his thumbs on the top and his fingers spread out underneath (*middle*), squeezing and bending the wood slightly across first one end and then the other, he is checking and comparing some of the stiffness characteristics of mode 2 in the upper and lower areas. When he holds the plate in his fingertips and pushes down in the middle with his thumbs (*bottom*), he is checking some of the bending characteristics of mode 5. The hands in these photographs are the author's.

He feels the stiffnesses of the wood and listens for certain sounds as he continues to thin the wood in different areas a few tenths of a millimeter at a time. The process of learning the proper feel of the wood and the sounds to listen for in the two free plates is crucial to the art of fine violinmaking; years of experience are needed to become good at it.

As an aid to understanding what the violinmaker feels and hears in his beautifully carved plates our tests of free plates explore modes 1, 2 and 5, mainly with the Chladni-pattern method. In a given pair of top and back plates each mode is checked for frequency, amplitude, Q and the conformation of its nodal pattern, with adjustments of the frequencies of the three modes to certain relations as far as is possible in each plate and between the two plates. We test the finished instrument by means of the response curve, the loudness curve and comments by players.

Five findings stand out in our results. First, an instrument of good quality results when mode 5 has a relatively large amplitude and its frequency in the top plate lies within a tone of the frequency in the back plate. If the frequency of mode 5 in the top is higher than that in the back, the tone quality is usually "brighter." If this relation is reversed, the tone is somewhat "darker."

Second, smooth, easy playing characteristics result when the frequency of mode 2 in the top plate lies within 1.4 percent (about five hertz in violin and viola plates) of that of mode 2 in the back plate. Third, if mode 5 is at the same frequency in a pair of top and back plates, the frequency of mode 2 in the top should be within 1.4 percent of that in the back, otherwise the resulting instrument is hard to play and has a harsh, gritty tone. Fourth, violins of exceptional quality have resulted when modes 2 and 5 are placed approximately an octave apart in each plate and at corresponding frequencies with high amplitudes in both plates. Fifth, a further refinement is to place the frequency of mode 1 in the top plate an octave below that of mode 2, so that modes 1, 2 and 5 are in a harmonic series. It is possible but difficult to adjust the frequency of mode 1 to this relation in the top plate; the adjustment cannot be made in the back plate because of the different structures of the two plates.

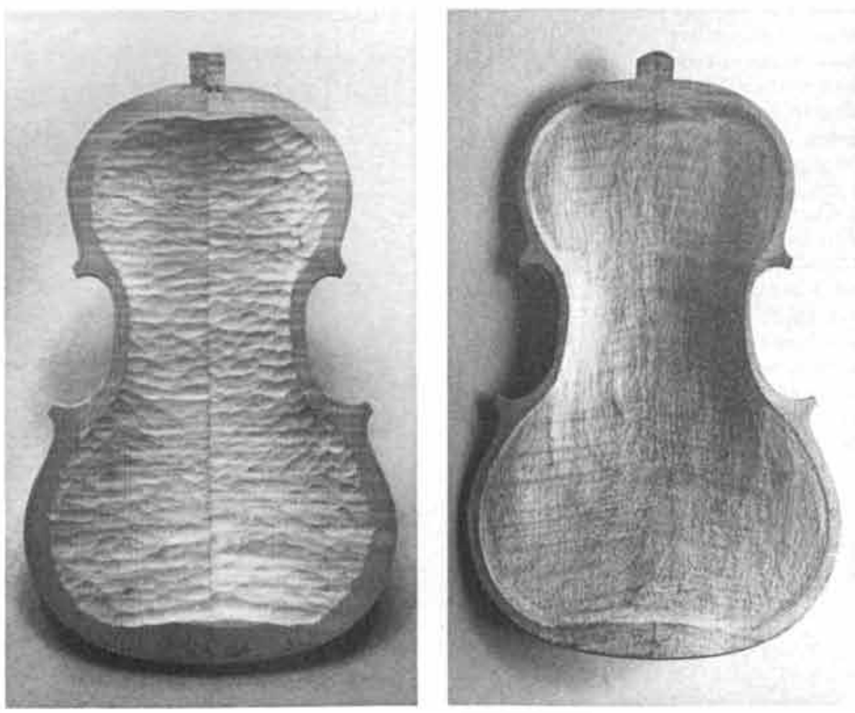
It is easier to state these conclusions than it is to achieve them in the actual tuning of plates. Many problems arise in attaining optimum eigenmode and eigenfrequency relations. The problems are largely related to four factors: the selective thinning of the plate to adjust for desired mode characteristics; the effects of the coatings (the sealer and the varnish); changes in relative humidity and temperature, and certain physical properties of the wood selected for the top and back plates.

Shaving the surface of a wood plate clearly reduces both mass and stiffness and also alters the plate's capacity for absorbing energy. Thus the frequency and shape of a given mode can be adjusted selectively to a limited extent by thinning the plate a few tenths of a millimeter at a time on the inside of the arch when it is within a millimeter or so of its final thickness. The general rule is that removing wood from a region of pronounced bending of a particular mode will tend to lower its frequency; removal from an area of little bending will raise the frequency. Thinning the wood in an area of pronounced bending reduces stiffness more than mass, so that the frequency goes down. Removing wood from an area of little bending reduces mass more than stiffness, so that the frequency goes up.

The bending areas (areas of strong curvature of motion) of a violin plate can be identified on the interferograms by lines that change from close to wider spacing. The process is somewhat like reading a topographical map where the curved slope of a hill is indicated by contour lines of equal elevation that become more closely spaced as the slope gets steeper. A straight slope is indicated by lines of equal spacing. In the interferograms the evenly spaced lines of the interference fringes indicate translation, or motion without pronounced curvature, just as the two ends of a seesaw move up and down without bending.

Hence removing wood a few tenths of a millimeter at a time from a crescent-shaped area around the two ends of the plate beginning just inside each corner will tend to lower the frequency of mode 5 more than it does that of mode 2. Removing wood from the center of the plate, reducing mass where the amplitude of motion of mode 5 is high, will tend to raise the frequency of the mode slightly. On the other hand, thinning the wood through the center of the upper and lower areas of a plate tends to lower the frequency of mode 2 because they are usually the bending areas for mode 2. Since the effective stiffnesses of a particular mode in one plate are not necessarily the same as those in another, it is important to assess them by the violinmaker's methods of feeling the bending areas as well as by observing the characteristics of the nodal patterns.

The sealer and the varnish influence plate tuning because they add mass, stiffen the outermost fibers of the wood and increase the damping. The lower the Young's modulus of the bare wood is, the more pronounced the increases in stiffness and damping are with added coatings. The effects are somewhat different for spruce and maple, with the result that the addition of sealer and var-



CARVING INSIDE OF PLATE after the outside archings have been shaped is the means by which the violinmaker achieves the desired acoustical characteristics. At the left a back plate for a violin is shown at an early stage of carving; at the right a different plate is shown in a nearly finished state. The marks were made by toothed planes. The plate is now ready for final scraping and acoustical testing by the Chladni method. A finished top plate for a violin usually varies in thickness from two to 3.5 millimeters, a back plate from two to six millimeters.

nish tends to detune the modes of the top plate considerably more than those of the back plate. Daniel W. Haines has reported that sealer and varnish increase Young's modulus and damping in cross-grain spruce much more than they do in cross-grain maple, with an associated rise in frequencies. Maple is two and a half times stiffer in cross grain than spruce.

Our tests indicate that sealer and varnish can indeed be detrimental to the sound of an instrument but that precautionary measures can be taken in the tuning of the free plates. For example, if the aim is to match the frequency of mode 2 in the top and back plates of a violin or a viola, mode 2 in the top should be left from five to 10 hertz lower before varnishing than it is in the back. Thus the varnish can help the sound of the instrument. If, however, the frequency of mode 2 in the top plate matches or is higher than that in the back plate when the plates are "in the white" (before varnishing), the discrepancy will be even greater after the addition of the coatings, and the instrument will probably have a harsh, gritty playing quality.

Violinmakers often say that a violin sounds better in the white than it does after it is varnished, and many of them have learned to compensate for the effect. Louis Condax, who experimented with sealers and varnishes for years, reported that when he removed the varnish from a violin that had a "harsh, gritty, tight sound," the instrument "came to life." The late John C. Schelleng showed that the acoustical properties of varnish coatings continue to change over a period of more than two years, which is doubtless one of the reasons it takes a newly varnished violin several years to settle into its true playing qualities.

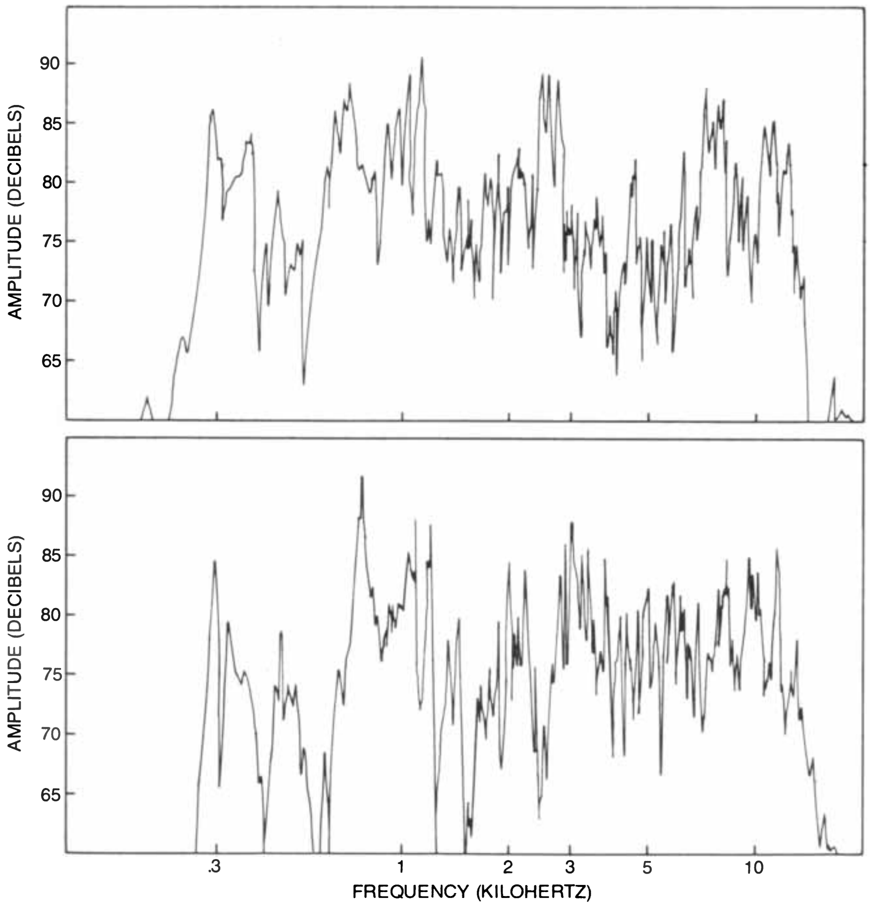
Violinmakers have long been troubled by the complaints of players whose instruments begin to sound dull and unresponsive in hot, humid summer weather and those whose instruments get harsh and gritty in the dry, heated houses of the Temperate Zone in winter. Adjustments of the bridge and the sound post can help to mitigate these problems, but a given instrument will generally sound best under the conditions of temperature and relative humidity in which it was made.

Wood is a hygroscopic material, taking on water and losing it readily in response to ambient conditions. The coatings of sealer and varnish on the outer surfaces of the violin help to retard this process somewhat, but to our knowledge there is no satisfactory treatment of the bare inner surfaces that is not detrimental to the tonal qualities of the instrument.

Experiments done in recent years by Robert E. Fryxell indicate that wood of various ages (as well as varnished and unvarnished violin plates) absorbs moisture quite slowly over a period of several months but loses it in a few hours, with maple slightly more absorbent than spruce. He also found that plates coated with sealer and varnish showed appreciably more stability than those in the white. Rex P. Thompson in Australia has found that the frequencies of mode 5 in a pair of varnished (for two years) and tuned plates varied as much as 18 hertz in the back and 23 hertz in the top over the range from 15 to 79 percent relative humidity. At a constant relative humidity the difference did not exceed five hertz. He also found that at a constant relative humidity of 50 percent the variations in frequency were only about 1 percent at temperatures from 15 to 25 degrees Celsius (59 to 77 degrees Fahrenheit). He concludes that for precise plate tuning either the temperature and humidity should be controlled or the humidity should be at 50 percent if the temperature cannot be controlled.

Our study has entailed work with many kinds of wood and many sizes of violin plate. We have found that carefully selected spruce of both European and American varieties can be employed successfully in making top plates. It is important, however, that the cross-grain stiffness be high enough to help maintain the desirable relation of an octave between modes 2 and 5. Various species of American maple have been tuned effectively by the Chladni method to make back plates for instruments with excellent tone and playing qualities. We have also found that other species of wood with characteristics fairly close to those of maple (pear, apple, cherry, sycamore and teak) can serve effectively for violin and viola backs, although the resulting tone qualities differ slightly from one instrument to another depending mostly on the high-frequency properties of the wood.

In the light of the understanding that our group and others have gained of the enormous number of variables encompassed in the making of a fine concert violin it is even harder than before to comprehend how the early luthiers managed to build instruments that were sophisticated and beautiful in both tone and appearance. The work I have described indicates that it is most desirable (but often quite difficult) to have the free-plate modes 1, 2 and 5 of a finished violin top lie in a harmonic series, with mode 5 having a large amplitude and a frequency near 370 hertz, and to have the frequencies of modes 2 and 5 in the top plate match those in the back. The shape, arching contours and thickness distributions of the top and back plates are crucial in achieving these relations. Moreover, the physical characteristics of the wood must lie within a narrow range of variables, and many other components of expert violinmaking must be held to close tolerances. In these conditions, and with the application of the Chladni method to aid in the determination of desirable eigenmode and eigenfrequency relations in each pair of unassembled top and back plates, violins and other members of the violin family can be made with consistently fine tone and playing qualities.



RESPONSE CURVES of a famous Guarnerius del Gesù violin made in 1731 and a violin made by the author in 1979 are compared. The curves reflect the same test procedure: a constant-current sine wave to the bridge, with the response of the violin (hung on rubber bands) picked up by a microphone 14 inches away in a fairly nonreverberant room. The upper curve is from the Guarnerius, the lower one from the modern instrument, which was made according to the plate-tuning principles described in this article. Note the decrease in the amplitude of the resonances through the 1.5-kilohertz range and the marked increase in amplitude from two to three kilohertz. This characteristic has been reported by Hermann F. Meinel and other investigators as typical of the responses of some of the most musically desirable violins.

Although the knowledge of certain characteristic relations in the eigenmodes and eigenfrequencies of free violin plates makes possible the construction of consistently fine instruments, it does not explain what happens to those modes when the plate pairs are assembled into the extremely complex vibratory system of the completed violin. In an analytical sense, however, the eigenmodes and eigenfrequencies of the parts fully define those parts. Therefore one can hope for the eventual recognition of a thread of generic information that will link the known vibratory characteristics of the free plates to the vibratory characteristics of the completed instrument.

Any structure such as a violin can be considered as essentially an arrangement of materials with its own characteristics of geometry, stiffness, mass and dissipation of energy. The assembly of the various parts into a completed instrument creates an additional set of potentially recognizable properties, namely the eigenmodes and eigenfrequencies of the instrument itself, each mode with its associated damping. Even though the process is a highly complex one, the eigenmodes of the parts can be regarded as elements that will ultimately determine the eigenmodes of the whole. The challenge this investigation presents for future investigators is: Can the tools and methods of measuring and analyzing vibrations be applied to a study of how the characteristics of free violin plates affect the vibrations of the coupled top and back plates and the enclosed air mass of the violin body as they respond to the forces generated by the bowed violin string?

Vitruvius' Odometer

A machine for measuring mileage that the Roman engineer described but may never have seen proved puzzling to Leonardo 1,500 years later. It may have been invented by Archimedes during the first Punic War

by André Wegener Sleeswyk

What links Archimedes, the Syracusan wizard of mechanics of the early Punic Wars, with Marcus Vitruvius Pollio, the military engineer and architect of the Caesars, and Leonardo da Vinci? It may be a wheeled device for measuring road distances, invented by the first of these three, described by the second and unsuccessfully reconstructed by the third. To demonstrate this possibility I shall begin by showing how a description of the device, written in about 20 B.C. by Vitruvius, apparently left Leonardo puzzled and at a loss when he read it some 1,500 years later. Then, profiting from Leonardo's incomplete analysis, I shall show how the measuring device can indeed be reconstructed according to Vitruvius' description. Finally, I shall argue that the device, which Vitruvius himself rather faintly praised as one "made in times of peace and quiet for profit and enjoyment," had at least in part a military origin and purpose and that it can be traced to none other than Archimedes.

The measuring device is known as an odometer (from the Greek *hodos*, way, and *metron*, measure). A similar mechanism is found in every automobile today. Before the advent of motorized transportation, however, the odometer served only special purposes. For example, during the Peninsular War of 1808–14 the British general Arthur Wellesley (later the Duke of Wellington) insisted on having his army's marches measured in this way, and similar mile measurers were often attached to hired coaches in order to determine what fare should be charged. Indeed, once gearing had been invented the essential simplicity of such devices would have offered little challenge to the engineers of antiquity. In the first century A.D. Hero of Alexandria described one. His account has come down to us in an Arabic translation of the original Greek: actuated by the turning of a chariot wheel, a series of worm gears moved a pointer along a scale that showed the distance traveled. The odometer described by Vitruvius, however, was not quite as simple. Let us turn to his text (which we find in that volume of his *Ten Books of Architecture* devoted to machinery) to see what gave Leonardo in the late 15th century and successive generations of commentators and translators so much trouble.

Some of Vitruvius' obscurity is minor. For example, his Latin is not the easiest and his style is uneven. Sometimes he is verbose; at other times he is so succinct that one has to guess at his meaning. A larger difficulty comes from the fact that in his day technical terminology was not well developed. To him gears are "drums" with teeth, and the gears are held in place by a *localumentum*. Now, *localumentum* was a general word in Latin (we have no exact equivalent of it) for a structure of planks designed to keep things in their place. The word might describe anything from a bookshelf to an open box for storing grain. Here I shall translate *localumentum* as "container."

The invention of the odometer must predate the time of Vitruvius. Indeed, it is not even certain that he ever actually saw the device he describes; he starts his account with the statement that he is now turning "to a matter transmitted by our forebears." The late A. G. Drachmann, both a classical philologist and a student of ancient technology, drew my attention to the fact that Vitruvius customarily wrote in the indicative mood in his other descriptions of machinery but wrote in the subjunctive mood throughout his description of the odometer. He may have done so to show that he was quoting from an earlier source.

In Book X, Chapter IX, of Vitruvius' *Architecture* he describes two versions of the odometer. One version is for service on land and is actuated by one of the rear wheels of a four-wheeled carriage, a *raeda*. The other, a marine version, is driven by a paddle wheel mounted on shipboard. This forerunner of a ship's log will not concern us here. As for the land odometer, Vitruvius, having told us that the wheels of the *raeda* were four feet in diameter and would advance 12.5 feet per revolution, goes on to describe the gearing and the counting mechanism as follows:

"A drum is firmly mounted on the hub on the inside of the [carriage] wheel, having a single little tooth beyond the rim of its circumference. But above it is a container, firmly attached to the carriage box, with a drum that can turn, standing on edge and held in position on a little axle. In the rim of this drum 400 equally spaced teeth have been cut that engage the little tooth on the lower drum. Furthermore, to the side of the upper drum another little tooth is attached, projecting beyond the [400] teeth.

"Above it stands another horizontal [drum], toothed in the same way, held by another container. The teeth engage the small tooth that has been attached to the side of the second drum. Holes are made in this [horizontal] drum, as many as the number of miles the carriage can go in a lengthy journey. More or less does not matter. And in all these holes small round pebbles are placed. And in the drum's cover, or rather container, a single hole is made that has a tube through which, when they have come to this position, the small pebbles that were placed in the drum can fall one by one into the carriage box and the bronze vessel placed underneath.

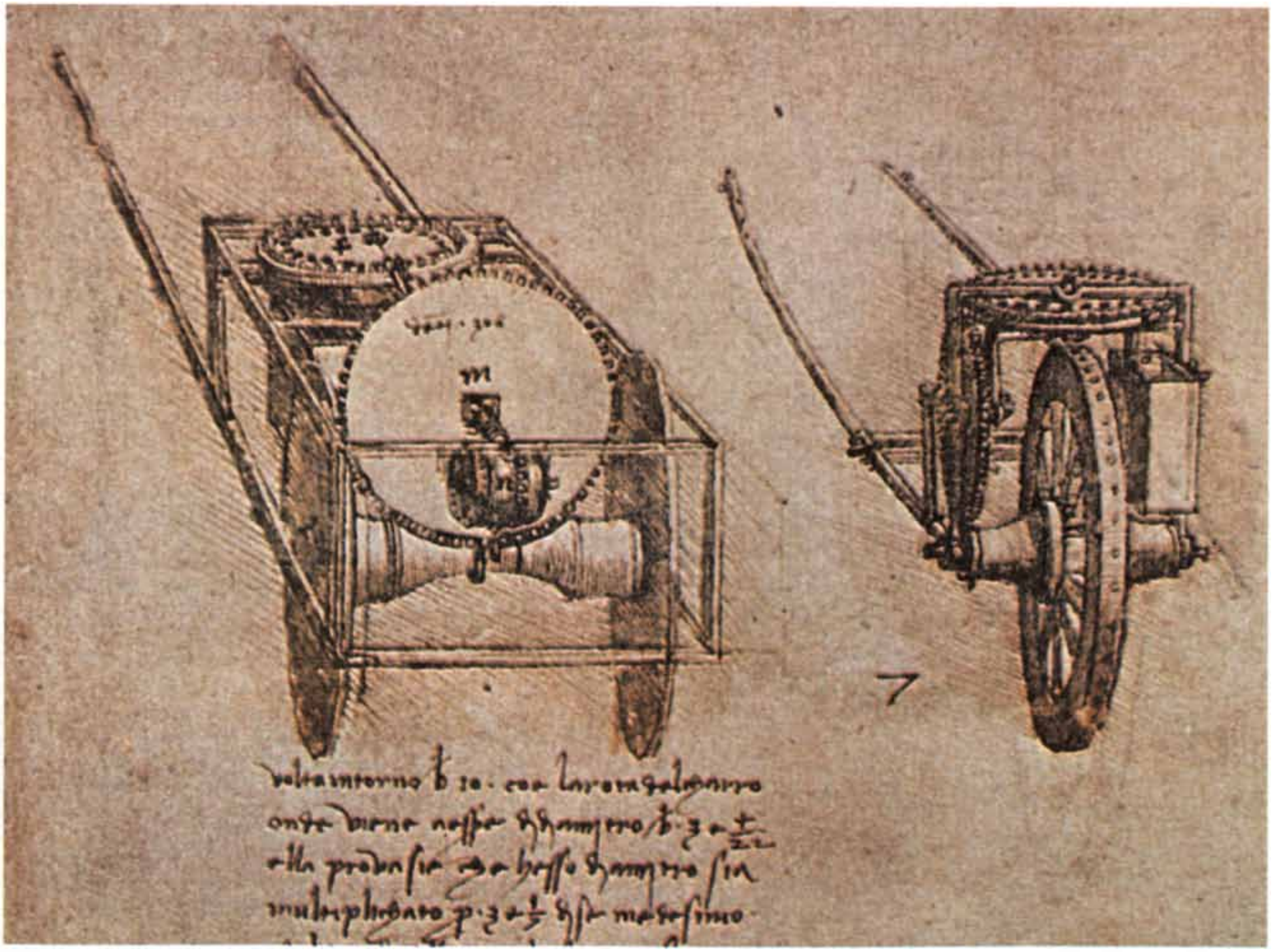
"When the [carriage] wheel goes forward, it takes the lower drum with it, and its little tooth causes the teeth of the upper drum to go forward by pushing them at every turn. [Thus] it is brought about that when the lower drum has been turned 400 times, the upper one has turned once, and the small tooth attached to its side will turn the horizontal drum one tooth forward. So that when by 400 turns of the lower drum the upper one rotates once, the distance covered will have a length of 5,000 feet, which is 1,000 *passus*. It follows that the falling pebbles will mark by their sound every [Roman] mile that has been accomplished. The number of pebbles collected below gives the total number of miles in a day."

Let me add a few words of explanation. The "carriage box" is a box that was fastened to the rear axle of some Roman carriages; its lid extended on both sides over the rear wheels. One would suspect from the description that both 400-tooth gears were mounted on the carriage box, the vertical gear on one of its sides and the horizontal gear on top of the lid. To repeat the sequence of events, the gear with the single tooth moved the vertical 400-tooth gear forward one tooth with every turn of the carriage wheel. By the same token, the single tooth mounted on the side of the vertical 400-tooth gear moved the horizontal 400-tooth gear forward one tooth with every 400 turns of the carriage wheel (a distance of 5,000 feet), causing one of the pebbles stored in the holes provided in the horizontal gear to fall with a clang into the bronze beaker inside the carriage box.

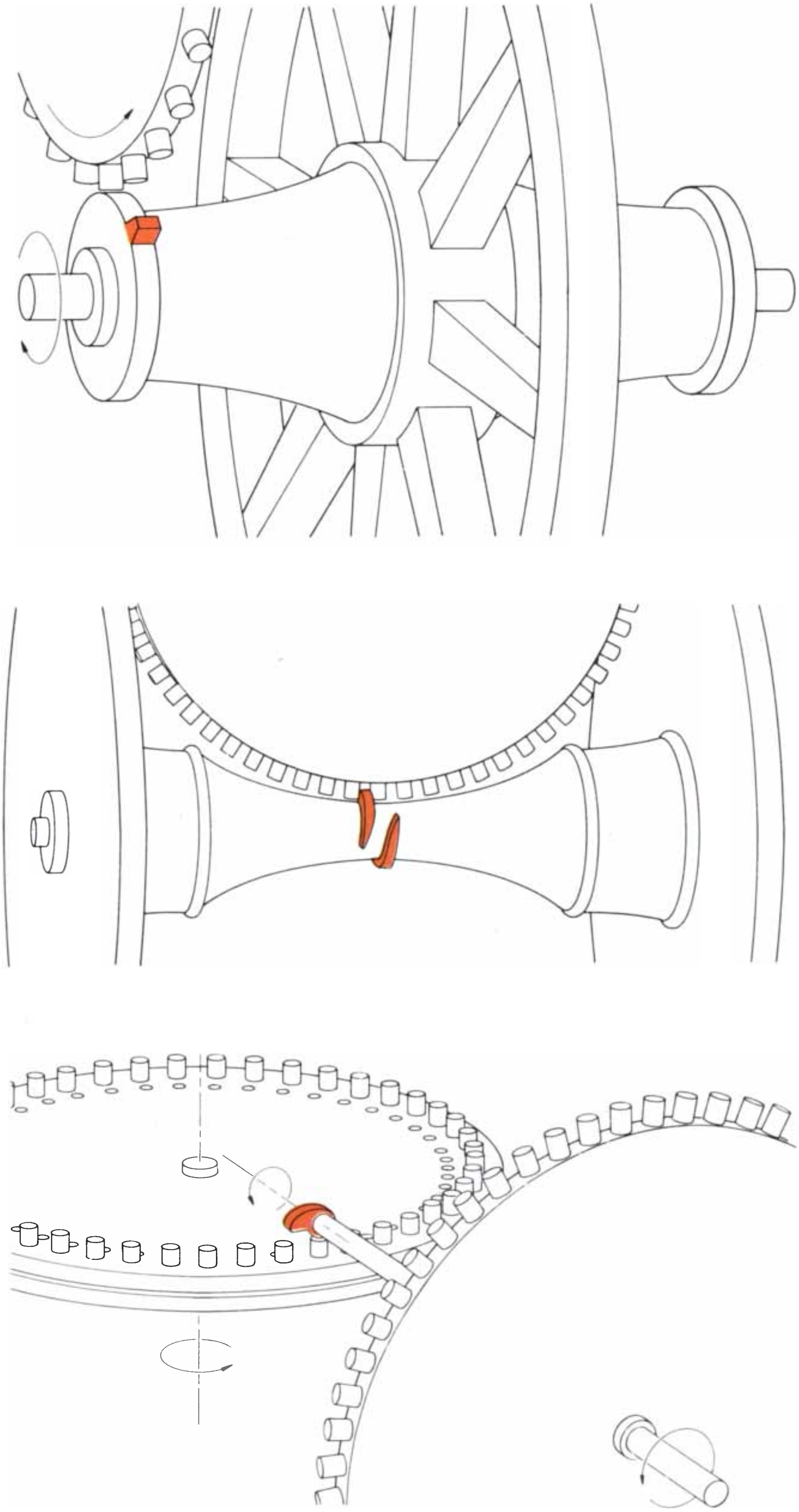
Now let us move forward to Leonardo. His famous odometer sketches appear on the first page of the *Codex Atlanticus.* Leonardo notes that he has adapted the mechanism to the measurements of his day. He required the circumference of the wheel to be 10 yards and the number of teeth in the two large gears to be 300 because the mile of his day was 3,000 yards. In all other respects, however, he seems to have followed Vitruvius' description.

Of the two odometer sketches, the left-handed Leonardo presumably first drew the one at the right side of the page: an odometer of the wheelbarrow type. Here a small tooth on the hub of the wheel intermittently advances the vertical gear at the right (which for the sake of simplicity he drew with some 40 teeth rather than 300). For every full turn of the vertical gear the horizontal gear advanced one step and dropped a pebble into the rectangular box at the left side of the wheel.

With its elegantly drawn 40-tooth gears Leonardo's odometer might have worked. With 300-tooth gears, however, the device would have been mechanically hopeless. This seems to have been first pointed out in print by Claude Perrault, a 17th-century French physician, the architect of the Louvre and incidentally a brother of the author of *Puss in Boots.* It would have been hopeless because with such a large number of teeth (whether it was 300 or 400 does not matter) the spacing between them becomes so small that the single tooth supposed to move the large gear forward simply cannot mesh effectively. Perrault, who published his commentary in 1673, estimated that a 400-tooth gear would have to be at least two feet in diameter (and hence more than six feet in circumference) to have teeth strong enough to support the stress of meshing. Even at



TWO ODOMETERS, sketched by Leonardo da Vinci on the basis of a description written some 1,500 years earlier by the Roman engineer Vitruvius, appear on the first page of the *Codex Atlanticus.* The wheelbarrow version was shown for the sake of clarity with vertical and horizontal gears having only 40 teeth rather than the 400 required by Vitruvius' description. The "single tooth" gear on the axle, which would advance the vertical gear one step with each turn of the wheel, is shown near the hub in the wheelbarrow version but is a worm gear at the center of the axle in the two-wheel version at the left. The details of both versions are illustrated on the next page.



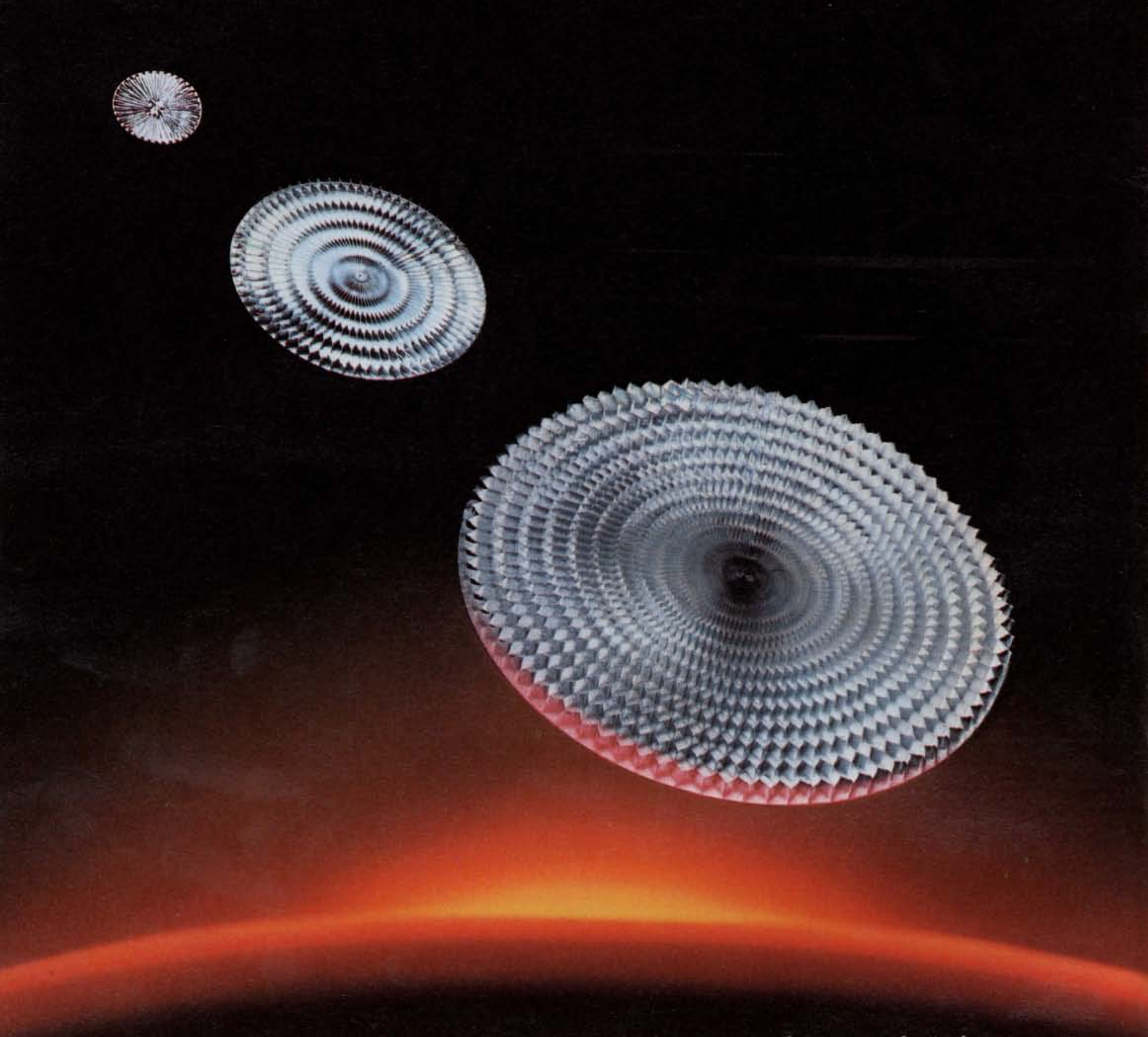
SINGLE-TOOTH GEAR (*color*) **attached to the rotating wheelbarrow hub might have meshed successfully with a 40-tooth vertical gear but, as was pointed out two centuries after Leonardo's time, could not have meshed with a 300- or a 400-tooth gear. The worm-gear "single tooth" of the two-wheel odometer** (*color*) **might have meshed successfully with a 400-tooth gear, but the rest of the gear train, particularly the single tooth on the axle of the vertical gear** (*color*)**, does not match Vitruvius' description. Dissatisfied, Leonardo stopped here.**

that circumference the meshing depth would be only half a millimeter. This was obviously not a workable gear train.

The problem seems to have escaped the attention of many later commentators, but there is reason to believe Leonardo became aware of it after he had sketched the wheelbarrow odometer. In his second drawing he circumvented the meshing difficulty with characteristic ingenuity. He chose to assume that the first "single tooth" of Vitruvius' description meant an endless screw, or worm gear (which may indeed be regarded as an elongated single tooth wrapped around a cylinder). Leonardo knew that Vitruvius was familiar with the endless screw, an invention of Archimedes'. Thus in his second sketch he fitted the axle of the two-wheel vehicle with a worm gear to drive the vertical gear. In addition he mounted the second single tooth on the axle of the vertical gear in order to drive the horizontal gear. This tooth, at least, rotated on a very small radius, so that the meshing depth attained a realistic value. Both innovations were geometrically correct ways out of the gear-train difficulty, and Leonardo's second reconstruction of the Vitruvian odometer might have been a workable device even if it had had gears with 400 teeth.

Leonardo's two-wheel odometer departs considerably, however, from Vitruvius' description. The two large gears are no longer "toothed in the same way." The horizontal gear has been given contrate teeth, that is, teeth that lie parallel to the gear axle; only the vertical gear has retained radial teeth. What is worse, one can by no stretch of the imagination make the single tooth attached to the axle of the vertical gear project "beyond the teeth" of that gear. It is therefore not surprising that classical scholars, with their regard for the written evidence, did not accept Leonardo's proposal. Their rejection was perhaps most clearly formulated by Drachmann, who in 1963 declared that he considered the Vitruvian odometer to be the ancient Roman's "armchair invention" rather than a description of a workable machine. Drachmann's judgment seemed at the time the only one possible for a serious scholar; no one had gone further than Leonardo in search of a valid interpretation of Vitruvius' words.

Must Drachmann's verdict be the final one? Once I began to look into the matter I came to doubt it. Whatever shortcomings Vitruvius had as a stylist, he was certainly a practical and thoroughly realistic engineer, not given to armchair inventions. Moreover, nowadays Leonardo's reconstructions, for all their ingenuity, cannot be regarded as exhaustive explorations of the possibilities implicit in Vitruvius' text. After all,



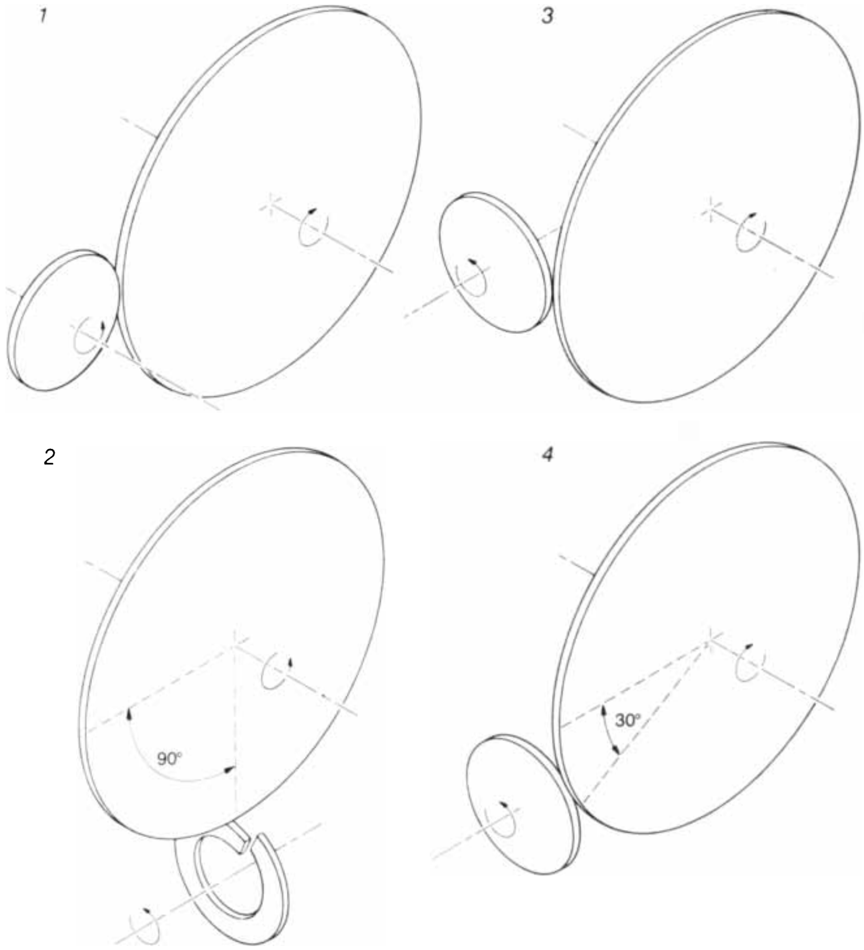
Leonardo lived at a time when machine technology was only first unfolding; today we have the advantage of half a millennium of further progress. What is more, in these five centuries our knowledge of the past has also greatly increased. For example, neither Leonardo nor Perrault knew that gears made of thin bronze plates were not uncommon around the beginning of the Christian Era. This has been demonstrated only by 20th-century archaeological finds, such as the fragment of a large bronze disk made in the second century A.D., found at Salzburg, and the remains of a complex train of gears made in the first century B.C., found by sponge divers off the Greek island of Antikythera.

The spectacular Antikythera mechanism, successfully reconstructed in 1974 by Derek J. de Solla Price of Yale University, was an analogue computer for calendrical calculations that employed more than 30 bronze gears. Its gear teeth were in the shape of equilateral triangles, a configuration that continued in use well into the Renaissance for "spur" gears, named after their resemblance to the rowels of spurs [see "An Ancient Greek Computer," by Derek J. de Solla Price; SCIENTIFIC AMERICAN, June, 1959].

A comparison of Vitruvius' text with Leonardo's interpretations of it shows clearly that Leonardo was aware of the gaps in the description. For example, Vitruvius does not mention whether the vertical 400-tooth gear was mounted parallel to the carriage wheel or perpendicular to it. To cover both possibilities Leonardo showed the gear parallel to the wheel in his first drawing and perpendicular to it in his second.

Actually, when Vitruvius says that the horizontal 400-tooth gear was "toothed in the same way," he does provide a hint, although not much more, about the mounting of the 400-tooth vertical gear. The single tooth attached to the side of the vertical gear rotates in a plane perpendicular to the horizontal gear. One would then expect that if the gears were toothed in the same way, the vertical gear would in turn be perpendicular to the plane of rotation of the single tooth on the wheel hub, that is, perpendicular to the carriage wheel.

Some tangible information about the gears themselves is also hidden in Vitruvius' text. This emerges when one compares his description of the marine odometer with the description of the carriage-mounted one. In the carriage-mounted device the single tooth attached to the vertical gear is said to project "beyond" the gear's 400 teeth. The text describing the single tooth in the marine odometer is more specific; it states that the tooth projects "beyond the circumference." One may infer that the single tooth could only have pointed in a radial direction; the gears were spur gears, as apparently all ancient bronze gears were.

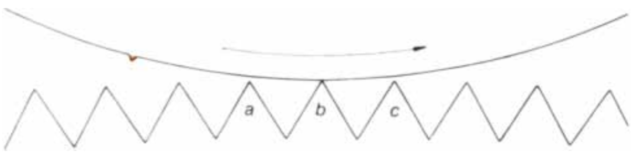


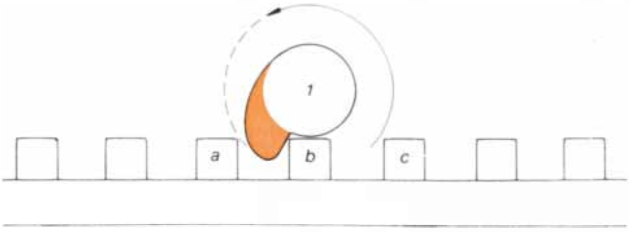
FOUR MEANS OF GEAR MESHING, three of them perpendicular and one in the same plane, are shown schematically. The spur-gear mesh (*1*) is the one proposed by Leonardo for the wheelbarrow odometer and the right-angle worm-gear mesh (*2*) is the one he proposed for the two-wheel odometer; its offset angle is 90 degrees. A conventional right-angle mesh (*3*) has a zero offset angle; the radius perpendicular to the intersection of the gear planes and the radius to the point of interaction are identical. The right-angle mesh with a 30-degree offset (*4*) is the arrangement tested with a scale model. It was proposed by the author to fit Vitruvius' description and at the same time to allow a single-tooth gear to mesh with a 400-tooth one.

This is as far as Vitruvius and Leonardo can take us. What is missing is information about the geometry of the interaction of the single-tooth gears and the 400-tooth gears. That information in turn would determine how the gears fitted on the carriage box. The dimensions of the carriage box, and hence the space available for the gears, may be estimated on the basis first of the four-foot diameter of the carriage wheels and second of the wheel gauge. The gauge can be estimated on the basis of paired ruts in Roman roads; in the excavated ruins of Pompeii, for example, the ruts are from four and a quarter to four and a half feet apart.

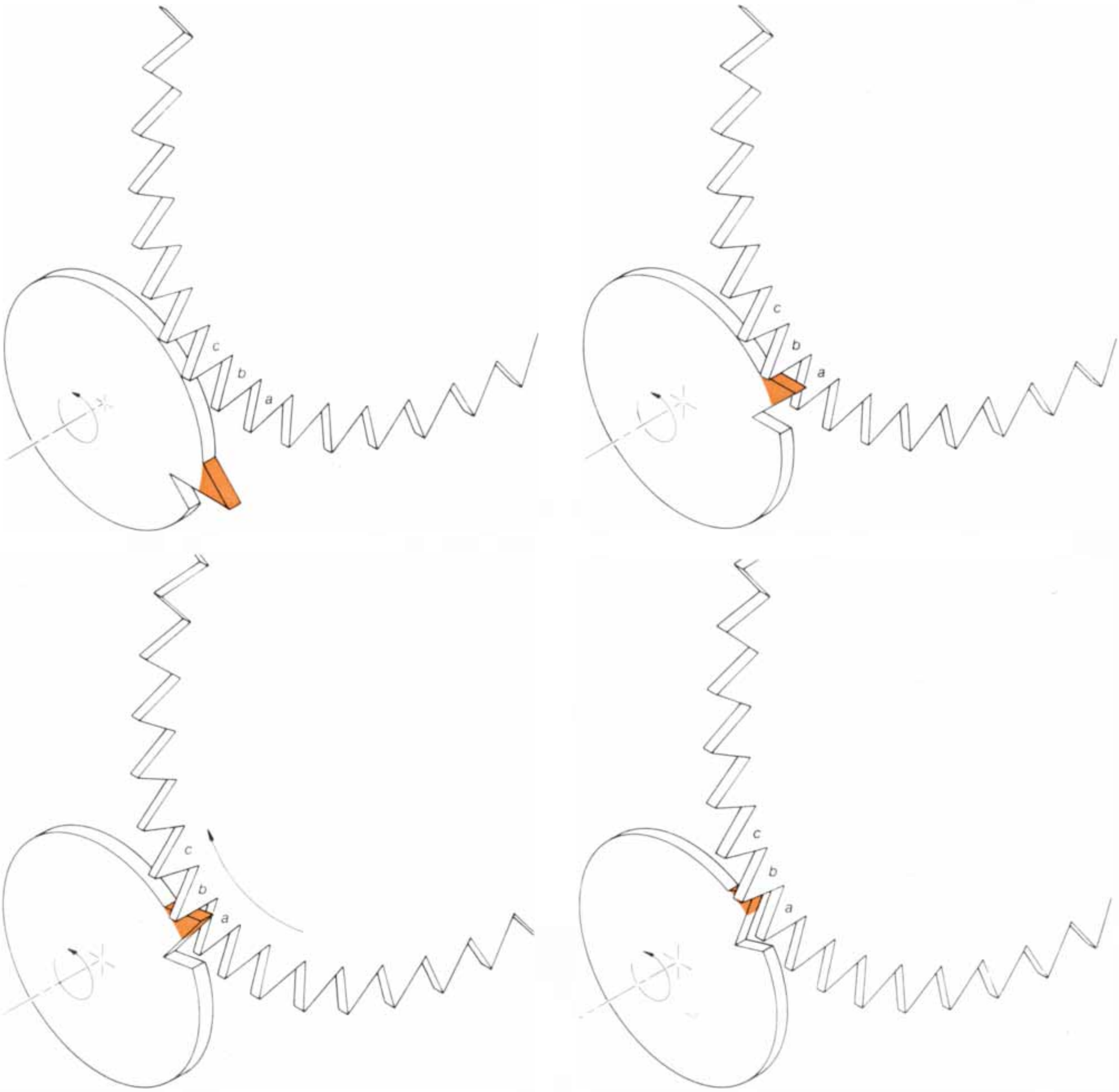
The missing link in all of this is provided by an elementary but systematic exploration of the geometry of gear interaction. To keep the number of different possibilities within manageable bounds certain reasonable simplifying assumptions must be introduced. For one, let the exploration be limited to right-angled pairs of gears. For another, assume that the gears are disks with their rims touching at the point of interaction. Assume further that the tooth shapes of the interacting gears are the same; therefore the interacting gears will have the same offset angle. Finally, on the basis of the archaeological information, assume that the tooth has the shape of an equilateral triangle. Given all of this, the number of independent variables has been reduced to only two: the offset angle and the ratio of gear-disk diameters. On this set of assumptions one obtains the basic geometry of a range of gearing that includes the bevel, hypoid and worm gears of today.

Of course these modern gears, with their sophisticated tooth shapes, are





INTERACTION OF A SINGLE TOOTH with a 400-tooth gear, the practical difficulty first pointed out by Claude Perrault in 1673, is shown schematically at the left. In order to clear tooth *a* before engaging tooth *b* (which must then be advanced to the position now occupied by tooth *c*) the single tooth must be so short that it cannot effect the desired movement. Leonardo's solution, as shown in his two-wheeler sketch, appears schematically at the right. Fixed to the end of the rotating axle of the vertical gear (*1*) is a larger single tooth. Because the total radius is very small the tooth clears tooth *a* of the horizontal gear, successfully engages tooth *b* and advances it as desired.



SINGLE-TOOTH GEAR of substantial dimensions can drive a 400-tooth gear in a way that agrees with Vitruvius' description, as this sequential illustration shows. At the upper left the rim of the gear is seen to pass between teeth *c* and *b* as the single tooth approaches tooth *b*. At the upper right the interaction of the single tooth with tooth *b* begins; the faces of the two teeth are approximately parallel when they touch because of the 30-degree offset of the single-tooth gear. At the lower left the single tooth is advancing tooth *b* toward the position formerly occupied by tooth *c;* the cutout in the gear rim behind the single tooth prevents interference with tooth *a*. At the lower right the advance of tooth *b* is completed; now the rim of the gear will pass between teeth *b* and *a* until a full turn brings the tooth to *a*.

beyond consideration here. One must think rather about the interaction of elementary spur gears, very simply cut or filed around the circumference of bronze disks. A good workman would have striven for minimum wear, which calls for a maximum area of contact at the point of interaction. It would have been achieved by having the faces of the two gear teeth approximately parallel where they touched. Given equilateral triangular teeth, this requirement would have been met if the offset angle was 30 degrees. The resulting arrangement would leave the single-tooth gear on the hub close to the carriage wheel. This position is in marked contrast to Leonardo's second arrangement, where the worm gear is most inconveniently situated in the middle of the axle, at an offset angle of 90 degrees.

Having arrived at this point in my reasoning I decided to test my gear-train deductions by building a simple model with gears cut out of thick paper. The single-tooth gear was positioned to drive a gear with 120 teeth when both were pinned to a piece of Styrofoam. Provided the multitooth gear was guided in a slot near the point of interaction, it proved entirely feasible for the single-tooth gear to advance the multitooth gear one step with each revolution.

In the absence of such a slot, however, the teeth of the multitooth gear tended to be deflected sideways when they were in contact with the single tooth. Suddenly this deflection made clear to me what the purpose of the *localumentum* might have been. Each of the containers would have provided the guiding slot essential for the proper functioning of the Vitruvian gear train.

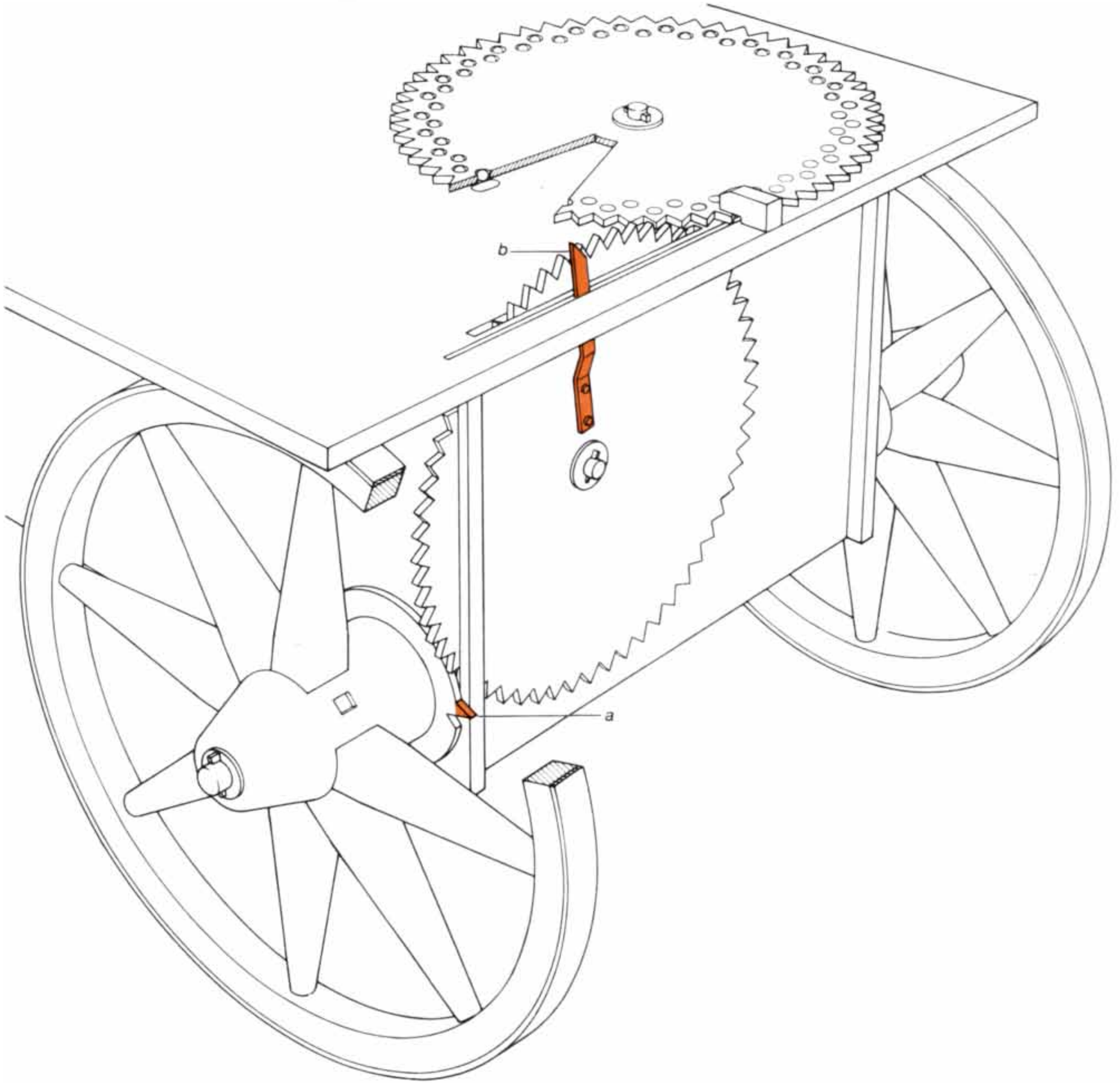
Other variables, such as the ratio of gear diameters, remained to be selected. In order to have adequate ground clearance the gear attached to the hub of the carriage wheel must have been approximately a fourth the diameter of the wheel. And if the horizontal multitooth gear was to lie on the lid of the carriage box above the wheels, the diameter of the vertical multitooth gear would have to be at least 2.7 feet. Other constraints included the need for the two multitooth gears to clear each other and the need for the single tooth attached to the vertical gear to clear both the carriage wheel and the single tooth attached to the hub. On paper the geometry of all of this proved to be just feasible.

To determine the validity of my paper reconstruction of the Vitruvian odometer I built a quarter-size scale model. In the large horizontal gear I drilled two concentric arrays of 200 holes each, to hold ball bearings in lieu of pebbles. The steel balls fell one at a time through an oblong hole in the lid of the carriage box, making an audible thump every quarter mile as they landed in a plastic beaker inside. Because my model agrees well with the details of Vitruvius' text and also because no equally probable alternatives presented themselves in the course of the reconstruction I do not think my odometer can be much different from the one Vitruvius may never have seen but painstakingly described.

Even if this successful reconstruction settles the question of whether Vitruvius was describing a real mechanism or an imaginary one, many other questions remain. As Drachmann had remarked earlier, why trouble oneself with a 400-tooth gear when a gear train with two successive 1 : 20 reductions would achieve the same effect? (This, by the way, is surely how Leonardo would have designed an odometer if he had started from scratch rather than trying to follow Vitruvius.) Again, to what purpose was this curious device put, considering that in the time of Vitruvius the distances along all the highways in Italy were well marked by milestones? The practice of setting up milestones goes back at least to 252 B.C. and was made compulsory by the tribune Caius Sempronius Gracchus in 123 B.C. This leads to still another question: Just when was it that the "forebears" cited by Vitruvius as the inventors of the odometer actually lived?

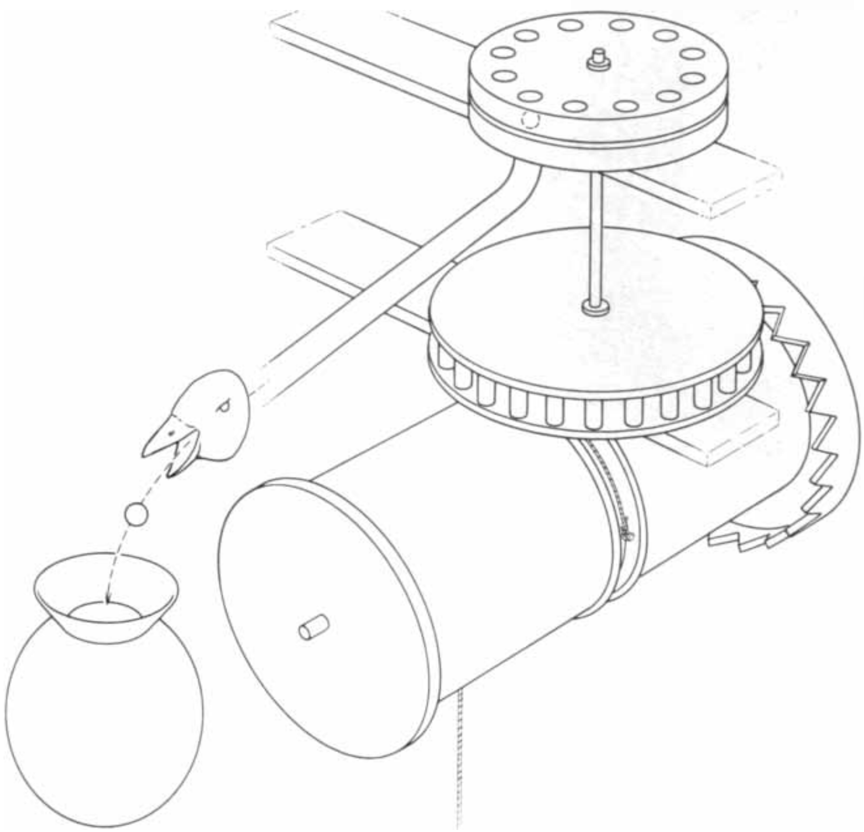
I shall return to these last questions below. As for the 400-tooth gears, I think the large size of the gear mounted on top of the carriage box could actually have been advantageous to the operator of the odometer. If the central part of the big disk was coated with black wax, as Roman writing tablets were, it could have provided a convenient informal note pad for jotting down particulars observed along the road. This information might have been used to compile the kind of classical travel guide of which a few, such as the *Tabula Peuteringiana,* still survive.

As for the when and the why of the Vitruvian odometer, answers to these questions cannot, of course, be given with any certainty. I shall nonetheless venture my own answer. It suggests a possibility that agrees with what we know about early Roman history, al-



RECONSTRUCTION of the carriage-box odometer described by Vitruvius is seen in this drawing; for the sake of clarity fewer than 400 teeth are shown on the two large gears. The single-tooth gear (*a, color*), fixed to the hub of the carriage wheel, advances the vertical gear fixed to the back of the carriage box one step with each rotation of the wheel. Another single tooth (*b, color*), attached to the vertical gear, advances the horizontal gear one step at every 400th rotation of the carriage wheel; this is a distance of 5,000 feet (one Roman mile).

QUARTER-SCALE MODEL of the carriage-box odometer, built by the author with 400 teeth on each gear, dropped a ball bearing (in lieu of a pebble) into a beaker every quarter mile.



PEBBLE-DROPPING CLOCK, attributed to Archimedes, is seen in part in a version drafted by Donald R. Hill, based on E. Wiedemann and F. Hauser's 1918 reconstruction. The drum, with contrate triangular teeth at its right end, is turned by a float (not shown) that follows the level of a descending water column. The drum teeth drive a second gear that is set at a right angle above the drum. Set above the second gear on the same axle is a disk with 12 holes, each hole containing a pebble. One hole aligns with a release tube each hour and the pebble falls out, to roll down the tube and drop, on the exterior of the clock, from a bird's beak into a vase.

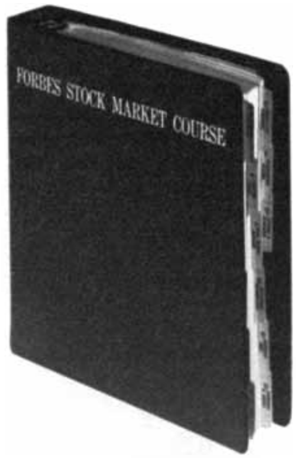
though such agreement does not rule out equally valid alternative answers. Mine is largely based on the fact that the unusual method of marking the miles—releasing pebbles from a storage disk to fall into a resonant container—is identical with the one employed in a water clock attributed to Archimedes (287–212 B.C.).

The clock is known to us because it is described in an Arabic manuscript that was probably composed in its final form in about A.D. 1150. Several copies of this treatise, *Kitāb Arshimīdas fī 'amal al-binkamāt* (*The Book of Archimedes on the Construction of Water Clocks*), exist. Although the title attributes the treatise to Archimedes, that surely cannot be true of the entire work because many of its details date from later times. For this reason the author is known to scholars as pseudo-Archimedes.

Donald R. Hill, a student of early Arab mechanics who translated and commented on the treatise a few years ago, came to the conclusion that it marks an important stage in the development of machine technology. He sees it as probably the earliest synthesis of Greek, Byzantine, Persian and Arab ideas about mechanical devices. In addition it provided a source of inspiration for later Arab writers on technology.

Hill points out that Arab commentators unanimously attribute the water machinery and reckoning by the release of pebbles to Archimedes, and he too believes this attribution is the most likely one. Pebble-release mechanisms for marking the time were at least moderately common in the Islamic world. The remains of a water clock of the 14th century A.D. that employed this system still stand at Fez in Morocco. Price, who investigated the device in 1961, found that some of the pebbles used to mark the time were still present. They were irregular spheres of pyrite, each one about five millimeters in diameter. In the Roman world, however, this mechanical recording system seems to have been limited to the two odometers described by Vitruvius. Was Archimedes the inventor of the Vitruvian odometer? Is there other evidence for or against that conclusion?

To me such a conclusion appears to fit the internal evidence remarkably well. Archimedes is known as the inventor of the endless screw, and as we have seen the method of gear meshing employed in the odometer is halfway between the endless screw and the meshing of spur gears at right angles to each other. This offset meshing system would not seem out of place among the other known inventions of Archimedes. The fact that Vitruvius failed to credit Archimedes with having invented the odometer is not strong evidence that Archimedes did not invent it; in the section of *Architecture* that deals with water-raising ma-


FORBES STOCK MARKET COURSE

ROMAN HIGHWAY SYSTEM, shown here as it existed at the time of Archimedes' death in 212 B.C., had undergone a great expansion following the outbreak of the first Punic War. The Via Appia, which then went only from Rome to Venusia, was extended to Tarentum and Brundisium in 264 B.C. The Via Aurelia was extended in 241 B.C. and the Via Flaminia in 220 B.C. Archimedes may have invented the odometer for setting milestones along the extensions.

chinery Vitruvius does not mention Archimedes in a detailed description of what is otherwise universally known as "the screw of Archimedes."

This conclusion, furthermore, ties in well with a reasonable guess about when and where the odometer was put to use. What originally was a local network of roads around Rome underwent its first major expansion when Archimedes was still a young man. That was in 264 B.C., when the first Punic War began. At that time the great southeastern highway, the Via Appia, already reached Venusia, some 320 kilometers from Rome. In that same year the Via Appia was extended an additional 140 kilometers through Apulia to Tarentum in Calabria and then another 70 kilometers across the heel of the Italian boot to Brundisium on the Adriatic. This brought the total length of the great highway, from Rome to Brundisium, up to 530 kilometers (358 Roman miles). Similar extensions were soon undertaken elsewhere.

Archimedes was a citizen of Syracuse, a powerful Greek city-state in Sicily, and was reputedly a kinsman of its king, Hiero II (306–215 B.C.). At the outbreak of the war between Carthage and Rome, Hiero first chose the Carthaginian side. Two years later, however, Syracuse went over to the Romans, and for the remainder of Hiero's long reign he was one of Rome's most loyal allies.

Thus for 36 years of Archimedes' adult life he worked in a city-state closely allied with the Roman Republic. During this same period the Romans built an unprecedented 750 kilometers of major highways. If the odometer was a tool for measuring distances along these roads, I think its original purpose was to ensure the correct emplacement of milestones along the Via Appia. For example, this would explain the correlation between the length of the Appian Way and the odometer's 400-mile "radius of action."

All of this further implies that Archimedes, one of the most ingenious inventors of antiquity, designed the odometer for this specific purpose in order to help his Roman allies. Historically that seems entirely plausible. Archimedes' hand in the matter seems all the more probable when one considers that the odometer design, as reconstructed according to Vitruvius' description, not only occupies a logical place among Archimedes' known inventions but also recorded distances by a system of pebble dropping that is persistently attributed to Archimedes in the ancient Arabic literature. It seems to me, however, that the odometer is more interesting for its ingenuity than for its practical utility and most interesting of all for the evidence associating it with great engineers of the past: Leonardo, Vitruvius and, so it now seems, Archimedes.

Quantum Logic

In the quantum theory classical patterns of inference break down. Mathematical structures called lattices can model alternative roles for the words "and" and "or" that may fit the world more coherently

by R. I. G. Hughes

If Aunt Agatha is dead and either the butler did it or the gardener did it, then surely either Aunt Agatha is dead and the butler did it or Aunt Agatha is dead and the gardener did it. If a quarter is in a box and it shows either heads or tails, then either the quarter is in the box and it shows heads or the quarter is in the box and it shows tails. The reasoning in such sentences compels assent, although the conclusions seem rather trivial, given the premises. Moreover, the subject matter appears to be entirely incidental to the validity of the inferences. Neither the wealth of Aunt Agatha nor the greed of the butler nor the concept of death has any bearing on the conclusion reached in the first argument; neither the monetary value of a quarter nor the definition of heads or tails could alter the conclusion of the second argument. Because of the irrelevance of the subject matter the validity of the inferences depends only on the rules of logic. In particular it depends only on the structural patterns displayed by connectives such as *and* and *or*.

Consider now certain physical phenomena of microscopic scale as they are described by the quantum theory. According to the theory, the electron (like a number of other elementary particles) has an intrinsic angular momentum, or spin. The spin is quantized: it is always found to assume one of only two values, either up or down, along any direction in which it is measured. It is impossible, however, to specify the spin of an electron along two spatial axes simultaneously. For example, if the spin of an electron measured along the x axis is up, it is not possible to assign any definite value to the spin along the y axis.

Suppose a beam of electrons is completely spin-polarized along the x axis, which means that all the electrons in the beam are found to have the same spin value (say spin-up) whenever the spin is measured along the x axis. Because the beam has not been polarized along the y axis, one can say of each electron in the prepared beam that its spin along the x axis is up and that its spin along the y axis is either up or down. Following the pattern of reasoning I employed in analyzing Aunt Agatha's murder, the statement about the electron implies either that the spin along the x axis is up and the spin along the y axis is up or that the spin along the x axis is up and the spin along the y axis is down.

Both clauses of this assertion, however, violate the principle of quantum mechanics stating that spin cannot be specified simultaneously along two axes. Since neither clause can be accepted, the assertion itself must be rejected. One must therefore either reject the initial statement about the prepared beam of electrons or disallow a logical procedure for defining the consequences of the statement, a procedure that seemed quite innocuous in ordinary reasoning. There is no motive at hand for rejecting the initial statement, and so it seems at least one law of classical logic cannot be applied to quantum phenomena.

Any proposal to revise the laws of logic, or even to regard them as being open to revision, runs counter to centrally and universally held beliefs. Nevertheless, the alternative of retaining classical logic by denying the validity of the quantum theory has little appeal. Fifty years after its inception quantum mechanics is one of the most successful of scientific theories. Its versatility and its predictive power are such that it has no serious rivals. It is applied routinely and with high precision to the understanding of the interactions of elementary particles. The theory also provides a means of describing a wide range of other phenomena, including the physics and the chemistry of atoms, molecules and solid materials.

In the opinion of some philosophers of science there is precedent in the history of physics for the kind of conceptual shift entailed by a modification of classical logic. It is now commonplace to point out that in the development of physics since Einstein the notions of space, time, energy, momentum and mass have been profoundly altered. The change can be expressed in what philosophers call the formal mode of speech by stating that the roles of words such as *space, time, energy, momentum* and *mass* are different in the language of modern physics from what they were in the language of classical physics. Similarly, one might suggest that if the conceptual structure of physics is embodied in the language of physics, then even the roles of words such as *and, or* and *not* are not exempt from revision. Because these words have traditionally been associated with logical investigations, and because the shift in their roles is motivated by the development of quantum mechanics, it makes sense to call the result of the shift quantum logic.

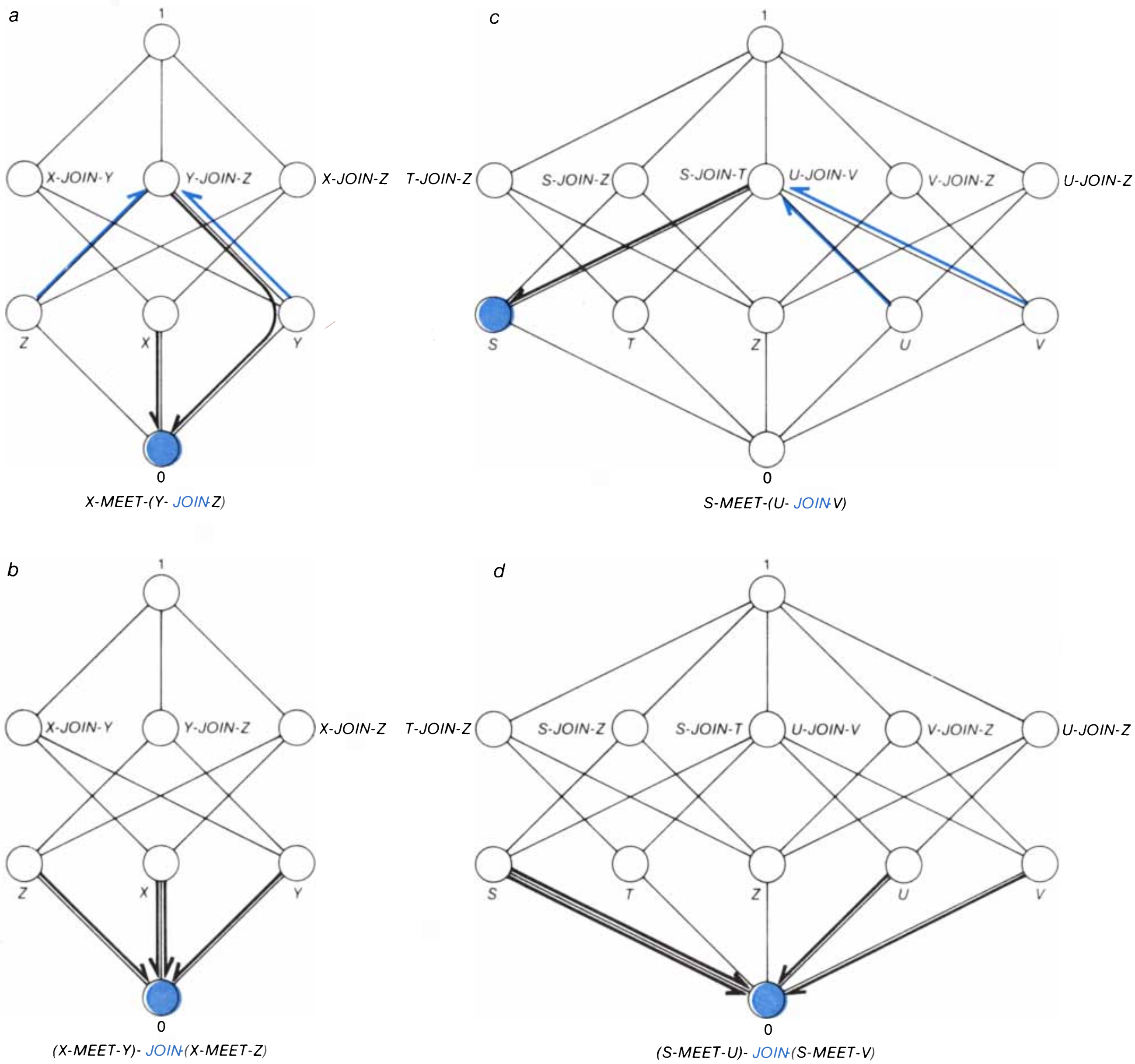
How might one characterize the role played in a language by the words that serve as logical connectives? Suppose P designates the sentence "Aunt Agatha is dead," Q designates the sentence "The butler did it" and R designates the sentence "The gårdener did it." Then the logical structure of my first inference can be set forth as follows:

From *P and* (*Q or R*)
one infers (*P and Q*) *or* (*P and R*).

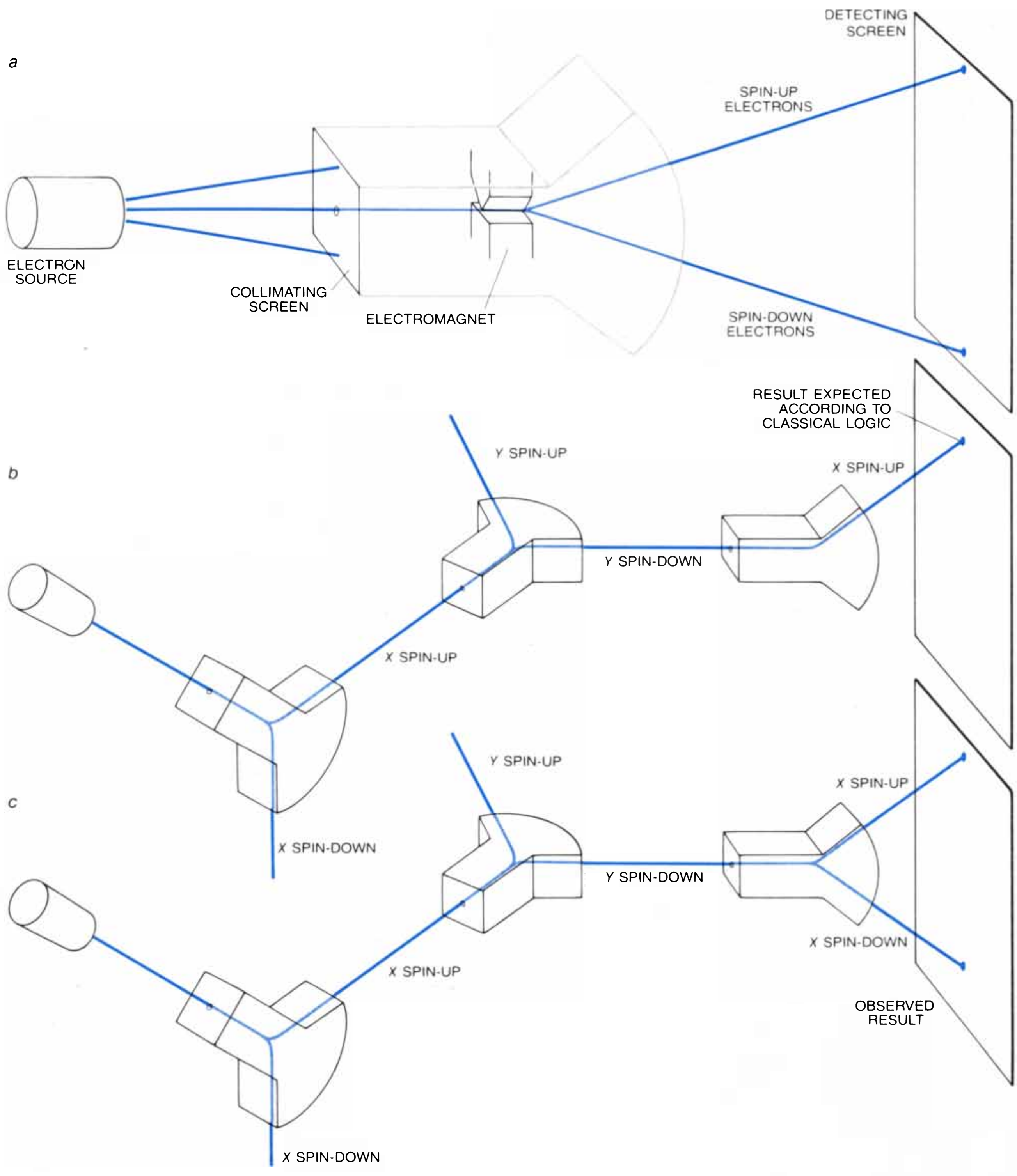
Writing the logical form of the reasoning in this way is meant to indicate that the inference is valid no matter what simple declarative sentences are substituted for the letters P, Q and R. The formula is called a distributive law of logic. In another distributive law of logic the words *and* and *or* are interchanged. Such laws of logic closely resemble the distributive law of arithmetic, which states, for instance, that the expression $2 \times (3 + 4)$ is equal to $(2 \times 3) + (2 \times 4)$.

In classical logic the roles of the logical connectives *and* and *or* are at least partially and implicitly defined by the patterns they form in the distributive law. In the quantum-mechanical description of electron spin, however, the logical step from the first line of the distributive law to the second line is disallowed. Suppose P designates the sentence "The spin along the x axis is up," Q designates the sentence "The spin along the y axis is up" and R designates the sentence "The spin along the y axis is down." The formula P *and* (Q *or* R) may then be true, but the formula (P *and* Q) *or* (P *and* R) cannot be maintained. Hence the suggestion that quantum mechanics may demand the revision of a law of logic amounts to the proposal that the roles of the connectives *and* and *or* must be altered so that statements about quantum mechanics no longer combine logically to satisfy the distributive law.

How can one be sure that such a change in the language preserves some sense of the traditional logical structure and does not lead to inconsistency or paradox? The traditional response to the question has been to build a mathematical model that incorporates the change but still exhibits reasonable behavior. (What is meant by "reasonable behav-



DISTRIBUTIVE LAW OF LOGIC states that for any sentences *P*, *Q* and *R*, if the compound sentence *P and* (*Q or R*) is true, then the compound sentence (*P and Q*) *or* (*P and R*) must also be true. The distributive law can be modeled by a lattice: an array of points and a network of lines that connect lower points with upper ones. The points represent sentences and the lines represent entailment relations. A sentence represented by a given point entails all the sentences represented by points higher in the lattice that can be reached from the given point along upward-moving lines. Two operations called *meet* and *join* can be defined on the lattice. The *meet* of two points is the highest point to which they are both connected by lines moving downward from at least one of them. The *join* of two points is the lowest point to which they are both connected by lines moving upward from at least one of them. If the operation *meet* is identified with the word *and* and the operation *join* is identified with the word *or*, the lattice can model the logical relations among sentences. The structure of the lattice determines whether or not distributive relations hold for the operations *meet* (*black arrows*) and *join* (*colored arrows*). For the lattice in panels *a* and *b* the operations *x-meet-*(*y-join-z*) and (*x-meet-y*)*-join-*(*x-meet-z*) lead to the same point, so that the operations satisfy a form of the distributive law. Such a lattice models classical, or distributive, logic. For the lattice in panels *c* and *d* the two operations do not lead to the same point. The lattice models the nondistributive logic that seems to be required for the description of quantum-mechanical phenomena. (The *join* operation in panels *b* and *d* is not shown by an arrow because it does not lead away from the bottom point, 0.)



OBSERVED PROPERTIES OF SPIN, or intrinsic angular momentum, suggest that the distributive law of logic cannot be applied to the description of an atomic or a subatomic particle. According to classical mechanics, particles whose spins are randomly oriented should be deflected along a continuous range of paths by a magnetic field that varies from point to point. Otto Stern and Walther Gerlach found in 1921 that such a field deflects particles along only two paths, indicating that the spin is quantized (*a*); the component of the spin measured along any designated axis is invariably either up or down. In each of the deflected beams all the particles have a spin-up component along the axis of the magnetic field (say the *x* axis), and they have either a spin-up or a spin-down component along an axis perpendicular to the field (say the *y* axis). In classical logic this is equivalent to saying that some particles in the deflected beam are *x* spin-up and *y* spin-up and some are *x* spin-up and *y* spin-down. A second magnetic field at right angles to the first field can split one of the beams into its two *y* components. One would expect that a third magnetic field, oriented along the *x* axis like the first field, would give rise to only one beam, since the particles had previously been exposed to spin-polarizing fields first along the *x* axis and then along the *y* axis (*b*). Instead two beams emerge from the third field; each beam is polarized along the *x* axis, but the *y* component is randomly oriented (*c*). The result indicates it is impossible to simultaneously assign exact values to the *x* and *y* components of the spin of a particle. The equivalence derived from the distributive law of classical logic does not hold because the *x* and the *y* spins cannot be defined simultaneously.

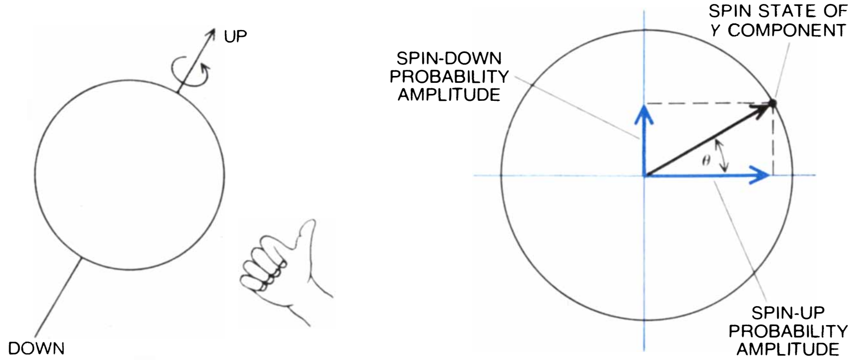
ior" probably cannot be specified in advance.) An instructive example of this process is provided by the development of non-Euclidean geometry.

The development began when mathematicians came to doubt the self-evidence of the fifth postulate of Euclid. The fifth postulate states that through any point in a plane that does not lie on a given line one and only one line can be drawn that is parallel to the given line. By denying the fifth postulate and assuming instead that either many parallel lines or no parallel lines can be drawn through the given point, 19th-century geometers were able to construct rich formal systems. The systems were composed of postulates and theorems that at first could not be interpreted as statements about geometry at all.

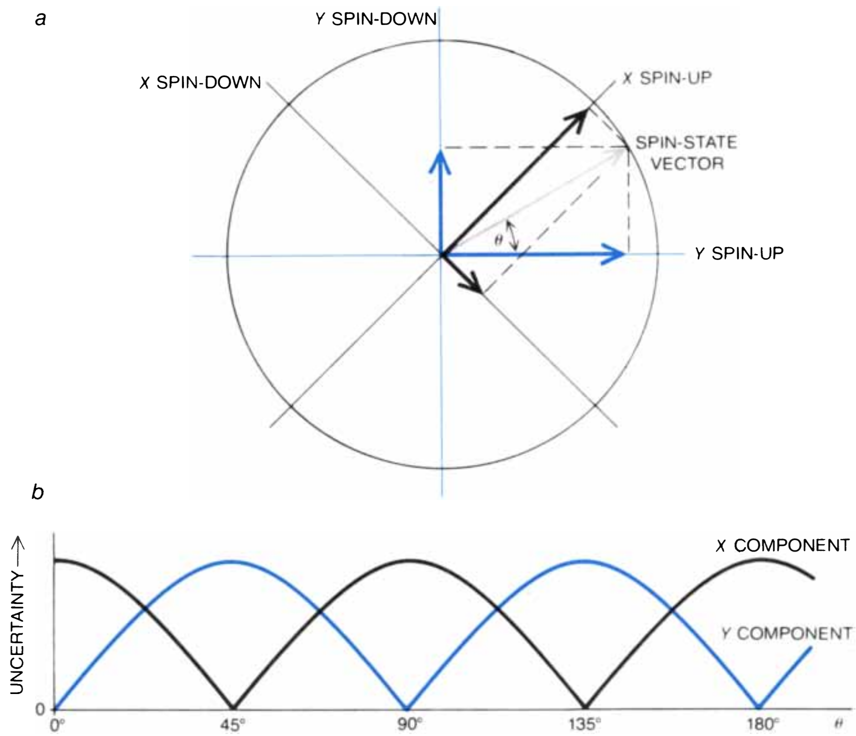
In spite of the intrinsic interest of such formal systems to mathematicians, it is difficult to work with the systems if there is no way to interpret their theorems and postulates. Initially geometers could not guarantee that all their results were free of inconsistencies, nor did they perceive that the various formal systems are deeply related. After some time, however, it was recognized that statements in the formal systems could be interpreted as statements about the geometry of curved surfaces, such as the surface of a sphere or the saddle-shaped infinite hyperbolic surface.

The construction of geometric models made possible a more intuitive understanding. Abstract patterns in the formal systems could be visualized as geometric relations, and one could see at a glance whether or not a theorem made sense. Moreover, mathematicians came to regard each geometric model as being merely one case in the more general study of curved surfaces; the geometry of a particular model could be characterized by specifying the curvature of the surface. Euclidean geometry, for example, is the study of flat surfaces, that is, surfaces whose curvature is zero. In this way the geometric interpretation of the various formal systems showed that they are not unrelated but are members of a single family. The theorems within each system came to be understood as analogues of one another, playing similar roles in the geometries of different surfaces.

In 1932, in order to construct a similar interpretive model for quantum logic, John von Neumann of the Institute for Advanced Study began investigating the properties of mathematical structures called lattices. Von Neumann elaborated his approach in 1936 in an article written with Garrett Birkhoff of Harvard University. They showed that the lattice structure of a physical theory can be regarded as a mathematical model of the system of logic appropriate to the theory. The notion of a lattice is quite



SPIN STATE OF A PARTICLE gives the probability that the spin measured along a given spatial axis is either up or down. By convention, if the fingers of the right hand curl in the same sense as the spin, the right thumb points in the spin-up direction. The probability of finding that the particle is spin-up along a given axis is the square of the magnitude of a vector called the spin-up probability amplitude. Similarly, the probability of finding that the particle is spin-down along a given axis is the square of the spin-down probability amplitude. The probability that the particle is either spin-up or spin-down is 1, so that the sum of the squares of the two probability amplitudes is also 1. If the probability amplitudes are graphed at right angles to each other, their vector sum is the hypotenuse of a right triangle; the length of the hypotenuse is 1. Hence any spin state of a particle can be represented by a vector that is the vector sum of the two probability amplitudes, and the set of all possible spin states corresponds to a circle of radius 1 in an abstract space called phase space. A complete analysis of spin states requires the introduction of complex numbers, that is, numbers with both a real and an imaginary part.

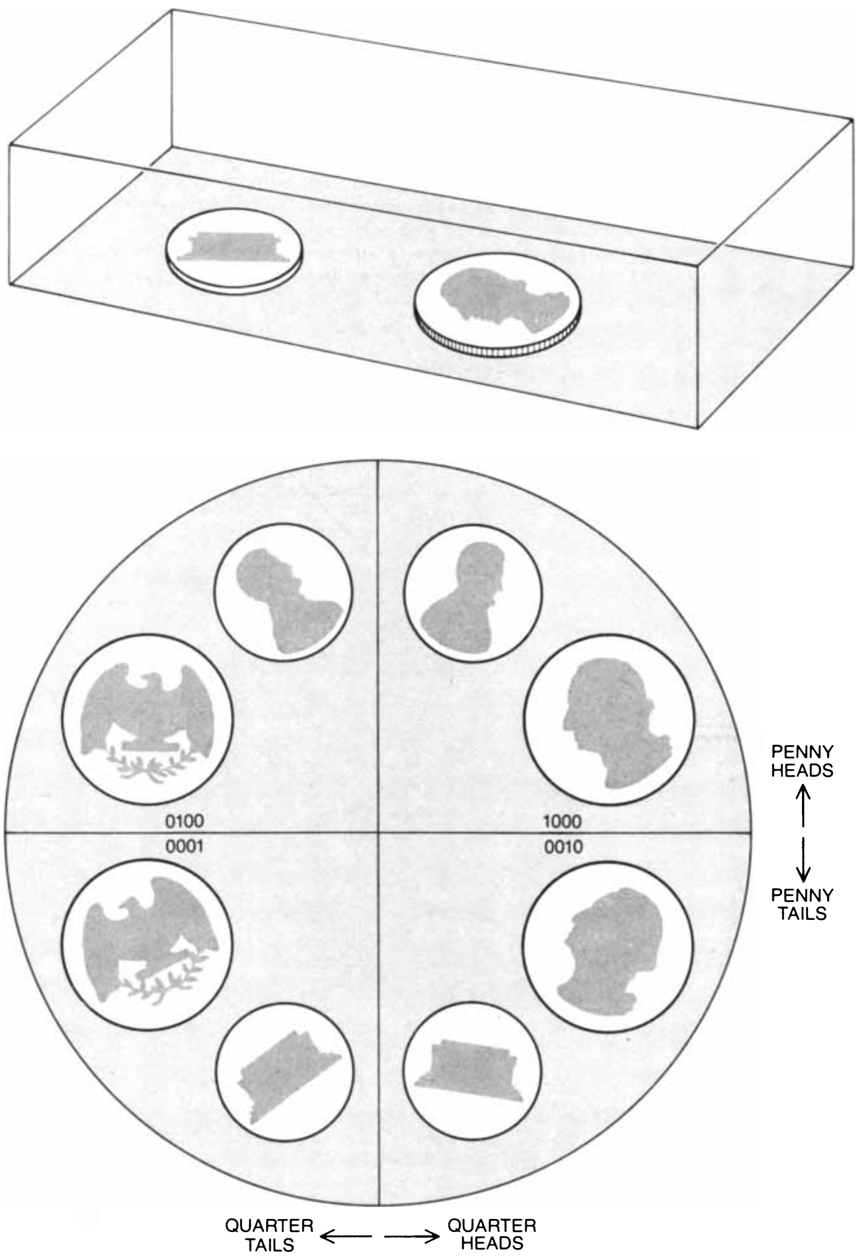


UNCERTAINTY PRINCIPLE of Werner Heisenberg states that the values of specified pairs of variables that characterize the state of a particle cannot be known simultaneously with unlimited precision. The x and y components of the spin of a particle form such a pair: they are said to be incompatible. Here the spin-up and spin-down probability amplitudes are given for the x and y components of the spin (a). The spin-state vector (*gray arrow*) can be treated as a vector sum of either the two x-component amplitudes (*black arrows*) or the two y-component amplitudes (*colored arrows*). The diagram is only an approximation; a full representation of the spin state would have to include the z component and could be drawn only in a space whose points represent complex numbers. As the vector rotates away from the y spin-up axis to the x spin-up axis the probability of finding that the x component of spin is up increases, but the result of any measurement of the y component becomes more uncertain. By imagining that the spin-state vector continues to be rotated counterclockwise, a graph of the uncertainties associated with the two spin components can be constructed (b). When the uncertainty in the value of one component falls to zero, the uncertainty in the value of the other component is maximal.

general. Here I shall apply it to the logical structure of a simple theory contrived for the purpose of illustration, but it can also capture the structures of classical physics and quantum mechanics. In the past dozen years physicists and philosophers have been returning to the lattice analysis introduced by von Neumann and Birkhoff.

The interpretation of a logical structure by means of a lattice is analogous to the interpretation of a formal system by means of a particular geometry. The roles of logical connectives in the physical theory can be identified with the roles of some of the operations and relations defined on the lattice associated with the theory. The result is a comprehensive view analogous to the more general understanding of geometry attained by introducing the idea of curvature. The abstract logic modeled by different lattices in different physical theories can encompass the change in the distributive law required by quantum mechanics while retaining the distributive law for the theories of classical physics.

Before proceeding with a description of lattice structures it is useful to examine more closely the assertions of the quantum theory that have precipitated such abstract logical investigations. Although quantum mechanics makes a number of puzzling assertions about physical events at microscopic scale, such as the duality of waves and particles, I shall confine my exposition to the effects of spin. The concepts required for an understanding of the quantum nature of spin reflect in an uncomplicated but nontrivial way the basic conceptual structure needed throughout quantum mechanics.



SIMPLE PHYSICAL SYSTEM made up of a box, a penny and a quarter has four states: both coins heads, both coins tails, penny tails and quarter heads, and penny heads and quarter tails. Each state can be represented by a quadrant of a disk, labeled by a four-digit binary number. Areas of the disk made up of various combinations of the quadrants correspond to every possible collection of the system's states and so to every "theoretical expression" that makes reference to some state or some combination of states. The disk is the phase space of the system.

The quantum-mechanical spin of a particle is analogous to the rotation of an ordinary object such as a top. The spin has a component along each of the three axes of three-dimensional space; the components are called the *x, y* and *z* components. By convention, if the fingers of the right hand curl in the same sense as the spin, the right thumb points in the spin-up direction.

The spin of a particle that carries an electric charge can be detected and manipulated by means of the magnetic moment generated by the spinning charge. The method of detection consists in passing the particle through a region in which a magnetic field varies greatly from point to point. According to classical physics, such a field gradient will deflect a moving particle by an amount that depends on the magnetic moment of the particle.

In an experiment done by Otto Stern and Walther Gerlach in 1921 silver was vaporized in a furnace and the silver atoms were directed by a series of baffles into a strongly varying magnetic field. Because the magnetic moment of the silver atoms is effectively randomized by the experimental procedure, Stern and Gerlach expected the atoms to be spread out uniformly by the magnetic-field gradient and to be detected as a smear on a photosensitive plate. They found instead that the plate recorded two well-defined patches where most of the atoms were concentrated. The magnetic moment of the atoms appeared to take on only two distinct values. The result has since been verified in many other experiments with more sensitive apparatus; in all cases the magnetic moment of an elementary particle or of a composite structure such as an atom is found to be quantized.

Because of the quantization of magnetic moment it is possible to segregate particles whose spins have a certain orientation by selecting just one of the beams that emerge from the magnetic field. The selected beam is spin-polarized in a direction parallel to the field gradient. For convenience I shall assume that the beam is moving along the *z* axis and that after passing through the magnetic field all the particles have their

"ALLIED纤维塑料公司在北卡罗来纳州的MONCURE城培训了我国的一批工程技术人员来支援在北京附近计划建造的一座大型聚酯纤维工厂。"

"He means, we mean business."

Mr. Yang Zhiding of the People's Republic of China is saying "The Allied Fibers and Plastics Company trained engineers and technicians from my country in Moncure, North Carolina for a large polyester complex to be built by China near Beijing."

Mr. Yang and his associates traveled around the world to us for training because in polyester technology, we're a world leader. And that's because, as Jim Dunbar of Allied is explaining, when we do something, we mean business.

But pioneering in polyester technology is only one way we mean business.

Carpets that hate dirt.

We meant business when we invented the first carpet fiber that actually rejects stains and spills. It's named Anso® IV. A carpet made with it is easier to keep clean because liquids just bead up on the surface. Which means no more crying over spilt milk.

We can stand the heat.

We meant business when we invented Halar® plastic wire insulation. It's a fire-resistant alternative to costly metal-encased wiring, and it can warm a contractor's heart by bringing down the cost of wiring buildings by up to 50%.

Victory at sea.

The old salts at Allied meant business when they produced polyester fiber so seaworthy it was chosen to be woven into the official foul-weather gear and lines aboard the America's Cup racing yachts.

We mean good business.

Last year, Allied Fibers and Plastics had over $1 billion in sales. It's been one of the most consistently profitable companies of its kind in the U.S., earning an average return on assets of over 20% for the past 5 years. See what we mean?

We are not alone.

We're woven together with other major businesses like chemicals, electrical products, and oil and gas to form the fabric of Allied Corporation.

You'll find Allied's technology threaded through everything from firemen's rope to high fashions. When we say fibers and plastics, we mean business. Or, as Mr. Yang puts it, "我们很认真。"

ALLIED Corporation

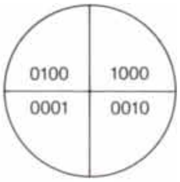
We mean business.

spin component along the x axis pointing up. The spin-polarized beam can then be passed through a second magnetic field at right angles to the first one and parallel to the y axis. One would suppose the beam would thereafter be polarized in both the x and the y directions. If either of the beams emerging from the second magnetic field were passed through a third field oriented in the x direction, one would expect to detect only the up value for the x component of the spin. Instead two patches again appear on the screen. The beam has somehow lost its spin-polarization along the x axis, and the measurement yields a random mixture of up and down values [*see illustration on page 204*].

This somewhat idealized experiment illustrates several fundamental principles of quantum mechanics. The passage of a beam polarized along the x axis through a magnetic field oriented so that it polarizes the y component destroys the polarization of the x component. It can be stated more generally that any operation on an elementary particle that determines the value of some quantum-mechanical variable must simultaneously randomize the value of at least one other variable; two variables that are linked in this way are said to be incompatible. If the y component of the spin is known with certainty, the value of the x component must remain completely unknown: it is therefore a maximally randomized quantity, with an equal probability (namely 1/2) of being either up or down.

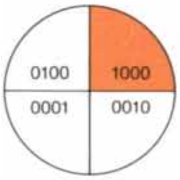
In general as the probability of measuring a particular value of one variable increases, the probability of measuring a particular value of an incompatible variable decreases. For example, if the y component of the spin became only partially polarized by the passage of the beam through the second magnetic field, the probability that the x component is up would be less than 1 but greater than 1/2. The uncertainty principle formulated by Werner Heisenberg specifies in a quantitative way how the probability of detecting a particular value of a variable depends on the probability assigned to an incompatible variable.

I shall now describe how the theoretical structure of a simple physical system can be represented by a lattice and how a logical structure can be grafted onto the lattice. Imagine a box containing a penny and a quarter and covered by a transparent lid. One can regard the box as a physical system whose state is determined by the upper face of each coin. The system can be in any one of four states: both coins heads, both coins tails, the penny heads and the quarter tails and the penny tails and the quarter heads. (The idea of discussing a system with just four possible states, although not this example, I owe to Ariadna



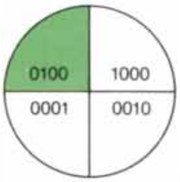
0000

P	Q	P AND NOT-P
1	1	0
1	0	0
0	1	0
0	0	0



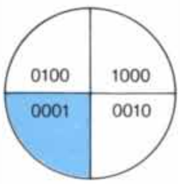
1000

P	Q	P AND Q
1	1	1
1	0	0
0	1	0
0	0	0



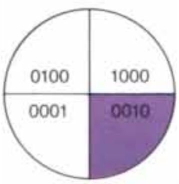
0100

P	Q	P AND NOT-Q
1	1	0
1	0	1
0	1	0
0	0	0



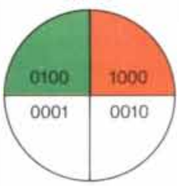
0001

P	Q	NOT-P AND NOT-Q
1	1	0
1	0	0
0	1	0
0	0	1



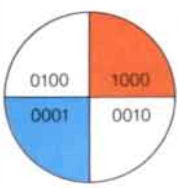
0010

P	Q	NOT-P AND Q
1	1	0
1	0	0
0	1	1
0	0	0



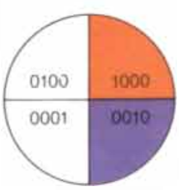
1100

P	Q	P
1	1	1
1	0	1
0	1	0
0	0	0



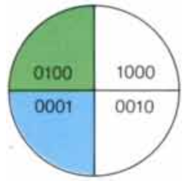
1001

P	Q	(P AND Q) OR (NOT-P AND NOT-Q)
1	1	1
1	0	0
0	1	0
0	0	1



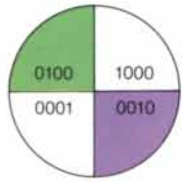
1010

P	Q	Q
1	1	1
1	0	0
0	1	1
0	0	0



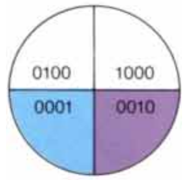
0101

P	Q	NOT-Q
1	1	0
1	0	1
0	1	0
0	0	1



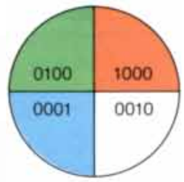
0110

P	Q	(P AND NOT-Q) OR (NOT-P AND Q)
1	1	0
1	0	1
0	1	1
0	0	0



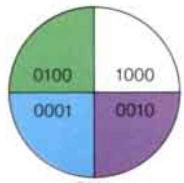
0011

P	Q	NOT-P
1	1	0
1	0	0
0	1	1
0	0	1



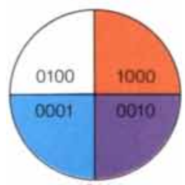
1101

P	Q	P OR NOT-Q
1	1	1
1	0	1
0	1	0
0	0	1



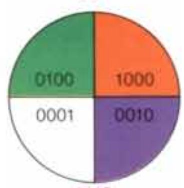
0111

P	Q	NOT-P OR NOT-Q
1	1	0
1	0	1
0	1	1
0	0	1



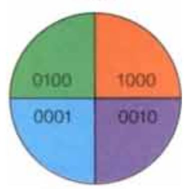
1011

P	Q	NOT-P OR Q
1	1	1
1	0	0
0	1	1
0	0	1



1110

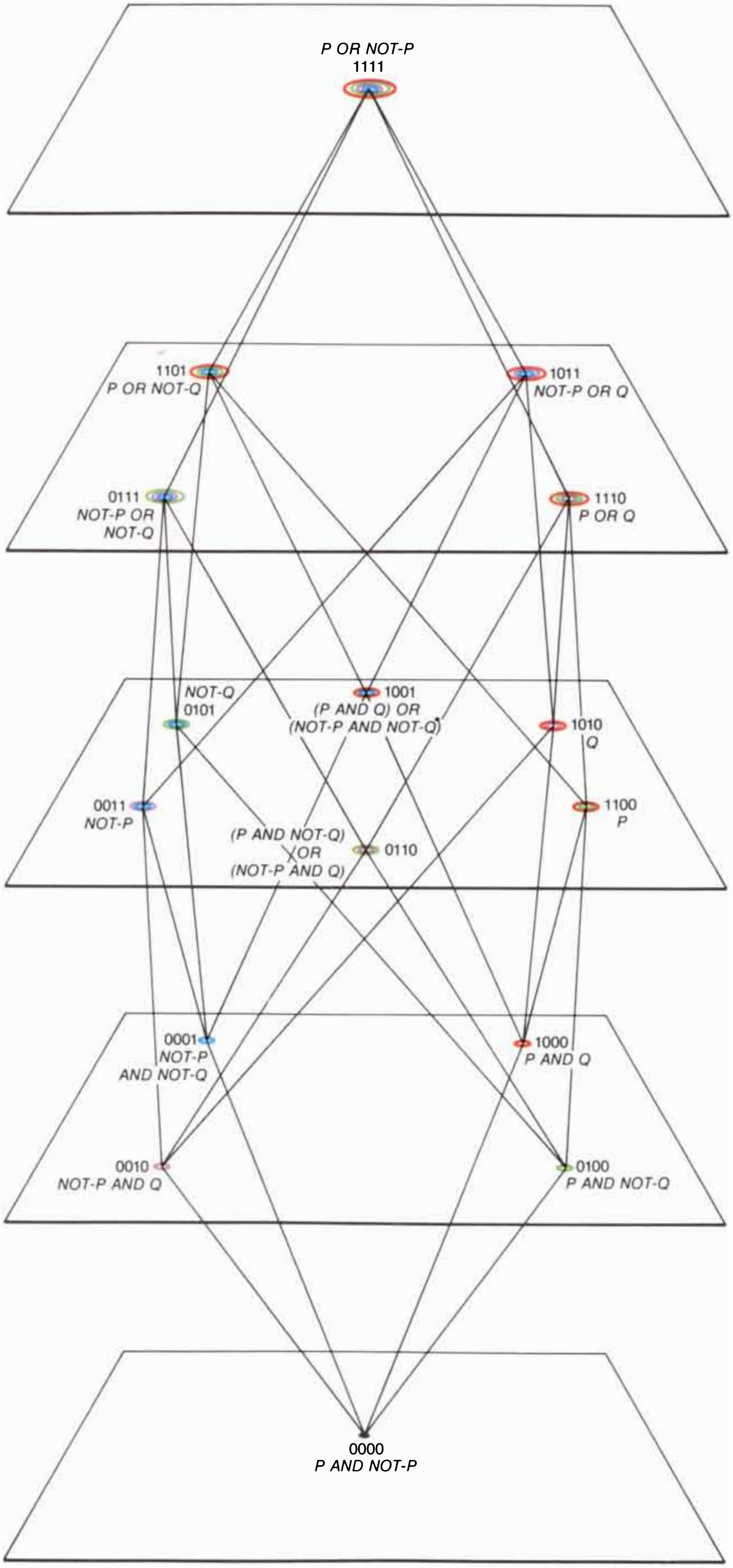
P	Q	P OR Q
1	1	1
1	0	1
0	1	1
0	0	0



1111

P	Q	P OR NOT-P
1	1	1
1	0	1
0	1	1
0	0	1

TRUTH TABLES show how the truth values of sentences can be associated one to one with the binary numbers that correspond to areas of a disk. Each quadrant of the disk represents a single state of the penny-quarter system, and every possible area is represented uniquely by the binary sum of the numbers assigned to the quadrants that constitute the disk. For any sentences *P* and *Q* a truth table shows how the truth (represented by a 1) or the falsity (represented by a 0) of a derived sentence composed of *P*, *Q* and certain logical connectives depends on the truth or the falsity of *P* and *Q* alone. There are four ways in which the sentences *P* and *Q*, taken together, can be either true or false. For each of the four ways the derived sentence can be either true or false. Hence there are 2^4, or 16, possible truth tables. For each truth table it is possible to find a sentence whose dependence on the truth values of *P* and *Q* is given by the table. For example, the sentence *P or Q* is true when either *P* or *Q* is true or both *P* and *Q* are true; it is false only if both sentences *P* and *Q* are false.



Chernavska of the University of British Columbia.)

In a simple representation each possible state of the system can be assigned a quadrant of a disk. If the disk is divided by a horizontal and a vertical line, the upper right quadrant can correspond to the state in which both coins are heads, the upper left to the state in which the penny is heads and the quarter is tails, the lower left to the state in which both coins are tails and the lower right to the state in which the penny is tails and the quarter is heads. The quadrants can be labeled in this sequence with the binary numbers 1000, 0100, 0001 and 0010. Any area of the disk that includes an integral number of quadrants can also be designated by a four-place binary number. In each place the digit is a 1 if any quadrant included in the area has a 1 in that place; otherwise the digit is a 0. For example, the area of the disk that includes all but the lower right quadrant is labeled 1101. Each quadrant must be defined so that it has no points in common with the other quadrants except the point at the center of the disk. I shall therefore assume that only the radius on the clockwise side of each quadrant belongs to that quadrant.

Any physical theory can be viewed as a statement about the possible or the actual states of a physical system. The theory may single out states or collections of states for commentary. In the penny-quarter system I shall suppose any combination of the four states can be referred to by a "physical theory" pertaining to the system. There are 16 combinations of the states, formed by combining the quadrants of the disk in all possible ways.

Suppose each combination of states is represented by a point. The 16 points can then be connected by a network of

LATTICE OF POINTS AND LINES displays the subset relations among the areas of the four-part disk. Each point represents a distinct area of the disk. The colors of the concentric circles around each point match the colors of the disk quadrants that make up the corresponding area (*see illustration on preceding page*). Points in the lattice that lie in a given plane all represent areas of the disk that include the same number of disk quadrants. The area represented by a lower point is a subset of any area represented by a higher point to which the lower point is connected by upward-moving lines. Binary numbers associated with the points correspond to the areas of the disk and to the truth tables of the sentences that label the points (*see illustration on preceding page*). The lines in the lattice also have a logical interpretation. Wherever they represent subset relations for the areas of the disk, they also represent entailment relations. A sentence represented by a lower point entails all sentences represented by higher points that can be reached by upward-moving lines.

lines to show how some areas of the disk can be considered subsets of other areas [*see illustration on opposite page*]. A line connecting two points indicates that the lower point is a subset of the higher point. More precisely, the area represented by the lower point is included in the area represented by the higher point.

The lowest point in the network represents the zero area (0000), the area of the point at the center of the disk that all the quadrants have in common. From the zero point, lines in the network travel upward to the four points (1000, 0100, 0001 and 0010) that represent the areas of each of the four quadrants of the disk. As the connecting lines indicate, each of these areas includes the central point as a subset. Above the four points representing the quadrants are points representing larger areas of the disk, each of which includes at least two quadrants and the central point. At the top of the network is the point 1111, which represents the area of the entire disk. Subset relations can be indicated not only by a direct upward line but also by a line that passes through intermediate points. Thus the uppermost point is linked to all the points below it. One can also think of each point as being connected to itself.

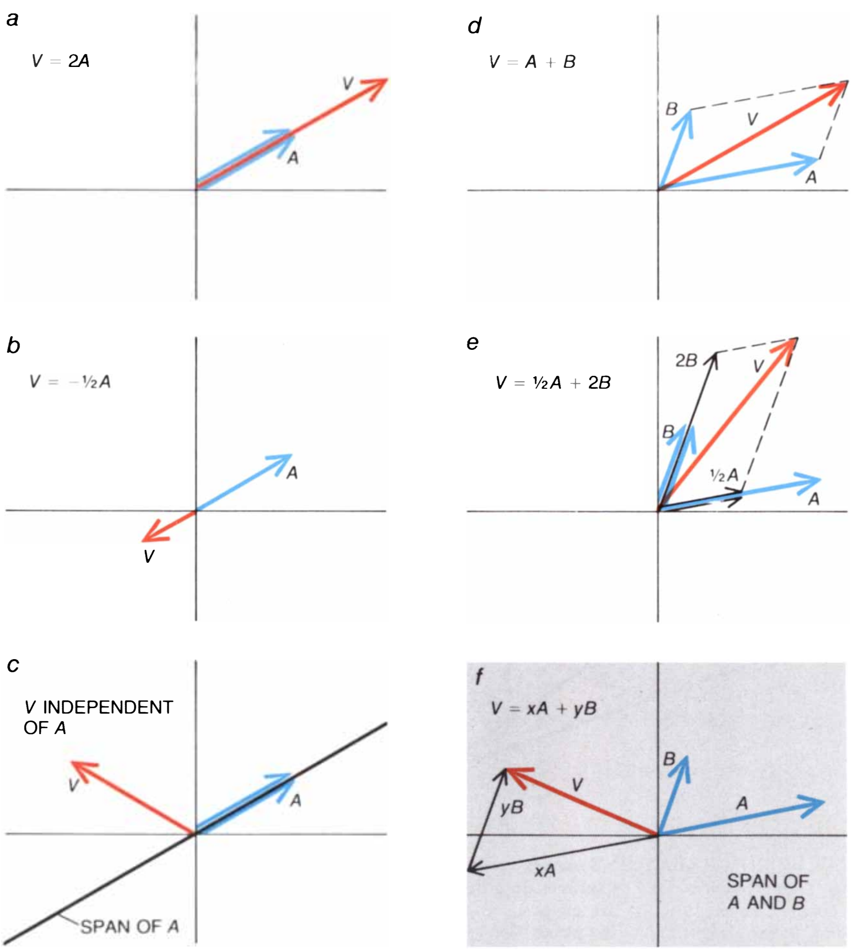
It is now possible to define two operations on the network of points and lines that make the network a lattice of the kind defined by von Neumann and Birkhoff. Given any two points *a* and *b* in the network, the lowest point in the lattice to which *a* and *b* are both connected by lines moving upward from at least one of them is referred to as *a-join-b*. If both *a* and *b* are directly linked to the same higher point, that point is *a-join-b*. If *a* and *b* are connected to each other, then *a-join-b* is the higher of the two points. For the lattice representing areas of the disk the *join* of two areas is the point representing the smallest area that includes both of them; in set theory the same concept is called the union. The *join* of 0000 (the central point of the disk) and 1000 (the upper right quadrant) is 1000, since the quadrant includes the central point. The *join* of 1000 and 0100 (the upper left quadrant) is 1100: the upper half of the disk.

A second operation on the lattice is defined by selecting the highest point to which both *a* and *b* are connected by lines moving downward from at least one of them; such a point is called *a-meet-b*. On the disk the *meet* of two areas is the largest area they have in common; in set theory it is called the intersection. The *meet* of 1100 (the upper half) and 1010 (the right half) is 1000: the upper right quadrant. The *meet* of 1000 and 0000 is the point 0000.

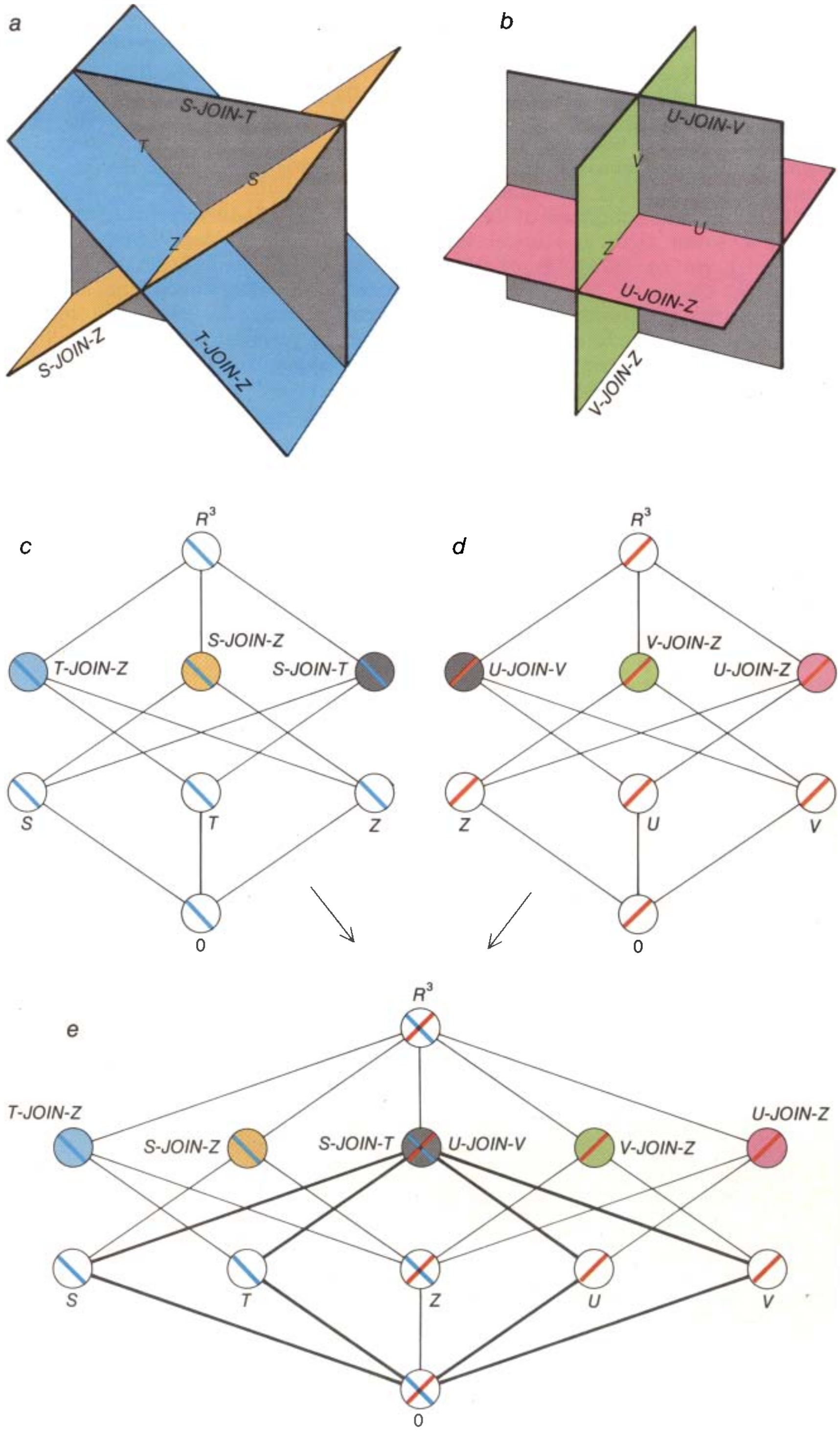
The lattices I shall discuss are called complemented ones: for any point *a* in the network it must be possible to find its complement *a'*, which has the property that *a-join-a'* is the top point of the network (1111) and *a-meet-a'* is the bottom point (0000). On the disk the complement of an area is another area that has no points except the center of the disk in common with the first area, but whose area taken together with the first area makes up the entire area of the disk. The points 0100 (the upper left quadrant) and 1011 (the area composed of the upper right, lower right and lower left quadrants) are complements of each other. The analogous operation in set theory is also called complementation.

The behavior of the lattice operations *join* and *meet* depends on the lattice on which they are defined. For the lattice representing the areas of the disk, however, they function in a familiar way. They obey the same laws as their counterparts in set theory: union and intersection. In particular the distributive laws hold. For any points *a, b* and *c* in the lattice, *a-meet-*(*b-join-c*) is the same point as (*a-meet-b*)*-join-*(*a-meet-c*). Similarly, *a-join-*(*b-meet-c*) is the same point as (*a-join-b*)*-meet-*(*a-join-c*). The resulting mathematical structure is called a complemented distributive lattice; it is also known as a Boolean algebra, after the 19th-century British logician George Boole.

I have described the properties of an abstract representation of the penny-quarter system, namely the set relations among the areas of a disk. What kinds of statement might be employed to describe the physical system itself? Suppose *P* stands for the statement "The penny is heads," *Q* stands for "The quarter is heads," *not-P* stands for "The penny is tails" and *not-Q* stands for "The quarter is tails." Any property of the penny-quarter system that a physicist might want to describe by means of a theory about the system can then be defined by writing a compound sentence. For example, the compound sentence *P*



SPAN OF A VECTOR SPACE is the set of all vectors that can be obtained by adding any vectors in the space or by multiplying them by a scalar quantity (a quantity that has only a magnitude and not a direction). The span of the single vector *A* (*blue arrows*) is a straight line passing through *A*, because additional vectors (*red arrows*) can be generated only by scalar multiplication or by the addition of vectors pointing along the line (*panels* a *and* b). Vectors that cannot be generated by such operations are independent vectors (*panel* c). Two independent vectors *A* and *B* combine by vector addition and scalar multiplication according to the parallelogram law (*panels* d *and* e). Their span is a plane because any point in the plane can be reached by vectorially adding some scalar multiple of *A* to some scalar multiple of *B* (*panel* f).



INCLUSION RELATIONS among vector subspaces of the three-dimensional physical space (R^3) can be represented by lattices. In *a* the subspaces represented by the lines *s* and *t* span the subspace represented by the plane *s-join-t*. Similarly, the lines *s* and *z* span the plane *s-join-z* and the lines *t* and *z* span the plane *t-join-z*. In *b* the lines *u*, *v* and *z* span corresponding planes, where the lines *u* and *v* are in the same plane as the lines *s* and *t*. The points in lattices *c* and *d* represent the vector subspaces formed by the lines and planes, by the zero subspace (the single point where the three lines in each diagram intersect) and by the vector space R^3 itself. Vectors in a subspace represented by a lower point in a lattice also lie in any subspace represented by a higher point to which the lower point is connected by upward-moving lattice lines. The two lattices can be linked at points that represent the same subspaces, namely the points R^3, *s-join-t* (equivalent to the point *u-join-v*), *z* and 0. The resulting lattice (*e*) is a nondistributive one, identical with the lattices at the right in the illustration on page 203. The compound lattice displays the inclusion relations that would hold among the vector subspaces if the two planes *s-join-t* and *u-join-v* were coincident and the lines *s*, *t*, *u*, *v* and *z* all crossed at the origin.

and Q makes the assertion "Both coins show heads." The sentence *not-P or Q* says that in the present state of the system "Either the penny is showing tails or the quarter is showing heads"; in other words, it specifies that the state is anything but the penny showing heads and the quarter showing tails.

In ordinary circumstances (unlike those encountered in quantum mechanics) logical intuition is a fairly reliable guide to the ways in which the truth or falsity of a compound sentence depends on the truth or falsity of its constituents. The sentence *P and Q* is true only if both *P* is true and *Q* is true, and hence only if both coins show heads; otherwise the compound sentence is false. The truth value of the compound sentence must be evaluated separately for each possible combination of truth and falsity of the constituent sentences.

A complete account of how the truth value of a compound sentence is determined by its constituent sentences is called a truth function, and logicians customarily display the evaluation of the truth function in an array called a truth table. Suppose 1 stands for true and 0 for false. Then any compound sentence made up of the two constituent sentences *P* and *Q* can be represented as a four-digit binary number in the same way as the areas of the disk were classified. There is a 1 or a 0 in the truth table for each of the four possible pairs of truth values for *P* and *Q*. Thus the truth function for every possible sentence having two constituents is given by one of 16 possible truth tables. The truth values of the sentence *P and Q*, for example, are given by the truth function 1000; the sentence is true only in the quadrant of the disk labeled 1000.

The relation between the truth function of a compound sentence and the number for the area of the disk in which the compound sentence is true is a general one. The binary numbers are the same [*see illustration on page 207*]. Hence to each area of the disk, and so to each point on the lattice, one can attach some sentence made up of the elements *P*, *Q*, *not*, *and* and *or*. More strictly one can associate with each point in the lattice an equivalence class of sentences, made up of all the sentences that have the same truth table. The sentence *not-(P and Q)*, for example, is associated with the same point in the lattice as the sentence *not-P or not-Q*.

The most important outcome of the association of sentences with lattice points is that it shows how the lattice can display logical relations. If the sentences *A* and *B* are represented by the points *a* and *b* in the lattice, then the sentences *A and B* and *A or B* are represented respectively by the points *a-meet-b* and *a-join-b*. Moreover, the lines connecting lower points in the lattice to higher points represent the logical relation of entailment.

Sentence A entails sentence B provided that B is true whenever A is true. The truth table for B must include a 1 on every line where a 1 appears in the truth table for A. (The truth table for B may include other 1's as well.) Therefore entailment on the lattice that describes the logical structure of the penny-quarter system is represented by the same lines that represent the subset relation.

The lattice for the penny-quarter system shows the same logical relations as the lattice that corresponds to Newtonian, or classical, mechanics. The state of a Newtonian particle is characterized by the position of the particle at a specified time and by its momentum (the product of its mass and its velocity). For simplicity suppose the Newtonian system consists of a single particle constrained to move in only one dimension. Just as a state of the penny-quarter system can be represented by an area on a disk, so the state of the Newtonian particle can be represented by a point on a plane. The coordinate axes of the plane are labeled with values of position and momentum, so that every point on the plane corresponds to some pair of values for position and momentum. The plane is called the phase space of the system.

What is the mathematical structure of the phase space to which the simplified Newtonian theory conforms? The statements of Newtonian physics make reference to regions, or subsets, of the phase space, just as statements about the penny-quarter system make reference to areas of the disk. A statement corresponds to a region of the plane whenever the statement holds true for every point in the region and for no other points. For example, the statement that a particle has position a corresponds to a line of points in the phase space; the line is a units from the origin along the position axis and parallel to the momentum axis.

The class of regions of Newtonian phase space is an infinite class, and so the corresponding lattice has an infinite number of points. In other respects, however, the mathematical structure of the lattice is the same as the structure of the lattice for the penny-quarter system. The regions of the phase space and hence the points in the Newtonian lattice are ordered by set inclusion, and as before the operations on the lattice are equivalent to set union and set intersection. Moreover, the logical relations defined by the structure of the lattice and by the operations *meet* and *join* remain the relations of classical logic. The distributive laws are valid. The extension of the phase space and the lattice to three-dimensional Newtonian systems with many particles is complicated but not different in principle from the one-dimensional, single-particle case.

In quantum mechanics the situation is altogether different. The state of a particle is no longer specified by its position and its momentum because position and momentum are incompatible variables, which cannot be determined simultaneously. Instead the state of a particle is defined by the theoretical construct called a wave function, which gives the probability of finding that the particle has a certain value of a physical variable. For example, the spin state of an electron is given by a wave function that specifies the probabilities that the x, y and z components of the spin are either up or down.

Consider the y component of the spin. The probability that the y component is spin-up can vary from 1 to 0 while the probability that it is spin-down varies from 0 to 1. For any spin state the sum of the two probabilities is equal to 1. It is mathematically convenient to represent the spin state as a vector: a quantity that has both a magnitude and a direction. The spin vector is the vector sum, calculated according to the parallelogram law of vector addition, of two other vectors. The latter two vectors lie along two perpendicular lines that represent the two possible values the spin can assume along the y axis. They are called the spin-up and the spin-down probability amplitudes. The square of the magnitude of the vector representing the spin-up probability amplitude gives the probability of finding the y component of the spin in the up state. Similarly, the square of the magnitude of the vector corresponding to the spin-down probability amplitude gives the probability of finding that the y component of the spin-state vector points down [*see top illustration on page 205*].

The abstract vector space with its perpendicular axes labeled "spin-up probability amplitude" and "spin-down probability amplitude" is a phase space of quantum mechanics. By superposing on the phase space an additional pair of perpendicular axes at a 45-degree angle to the first pair, it is possible to represent the essential features of the uncertainty principle as it applies to spin [*see bottom illustration on page 205*]. The additional axes can be designated the spin-up and spin-down probability amplitudes for the x component of the spin, just as the original axes designate the amplitude of the y component. If the spin-state vector lies along the spin-up axis for the x component (and so predicts the result of a measurement of the x component with certainty), the vector lies halfway between the spin-up and the spin-down axes for the y component. The length of the projection of the vector onto the spin-up axis for the y component is therefore $\sqrt{2}/2$, and so the probability of finding that the y component of the spin is up is the square of $\sqrt{2}/2$, or 1/2. Similarly, the probability of finding that the y component of the spin is down is also 1/2. Thus, in accord with the principle of uncertainty, there is no spin state in which a probability of 1 can be assigned both to some value of the x component and to some value of the y component.

The kind of inclusion relation that is defined on a vector space is the relation of being a subspace, rather than of being merely a subset, of the given vector space. It is the subspaces of a vector space that correspond to the propositions of the quantum theory.

What is a vector subspace? The subspace must itself be a vector space. The most important property of a vector space is that adding any two vectors in the space generates a third vector that lies in the same space. Similarly, multiplying any vector by a scalar quantity (one that has only a magnitude, not a direction) also yields another vector in the space. In mathematical terms the vector space is said to be closed under vector addition and the multiplication of a scalar.

The closure criterion can be employed to construct a vector space. The complete vector space can be generated from vectors called basis vectors if every vector in the space can be obtained by adding the basis vectors according to the parallelogram rule, perhaps after each basis vector has been multiplied by a suitable constant. (Scalar multiplication can lengthen or shorten a vector and can make it point in the diametrically opposite direction, but it cannot otherwise alter its direction.) The vector space generated in this way by two basis vectors A and B is called the span of A and B [*see illustration on page 209*].

A subspace is said to be a proper subspace if there is at least one vector in the vector space that does not lie within the span of any basis vectors of the subspace. For example, the u-v plane is a vector space that can be spanned by two vectors directed along the positive u and v axes. There is no way to add the two vectors to get a third vector that points out of the u-v plane. Hence the u-v plane is a proper subspace of the vector space constituted by three-dimensional space.

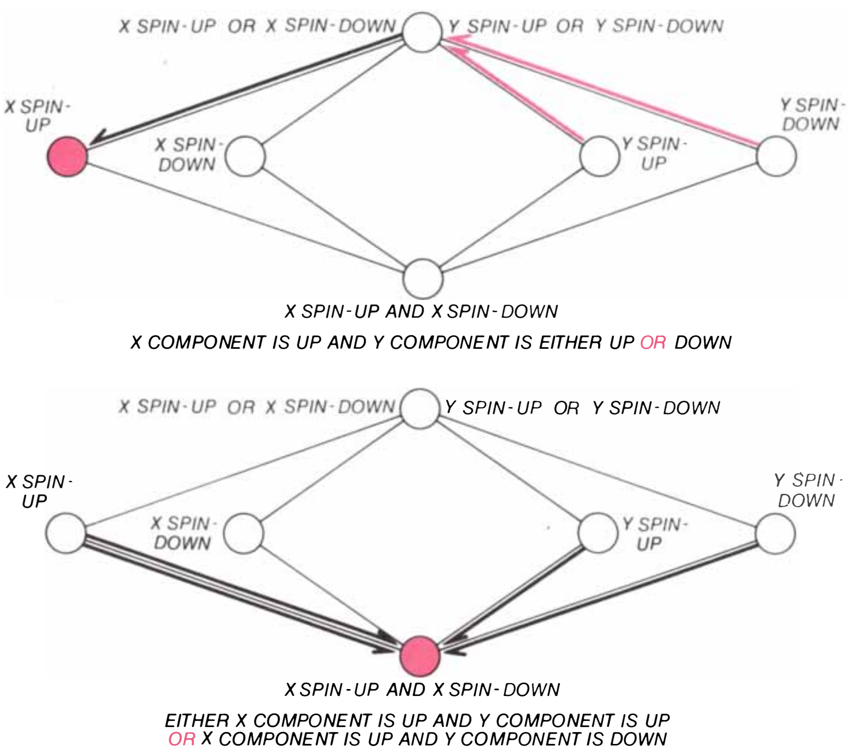
In order to form a lattice whose points correspond to the propositions of quantum mechanics one must form the lattice of vector spaces and their proper subspaces. Consider three-dimensional physical space with a designated point 0 as the origin. The subspaces of the space include the origin itself, all straight lines that can be drawn through 0, all planes that include 0 and the entire three-dimensional space. Some of the subspaces are ordered by inclusion, so that logical entailment on the lattice can still be a matter of subspace inclusion.

A crucial difference, however, is that the *join* operation on the lattice of subspaces is not the union operation of set theory but rather the span of two subspaces. For example, the *u* axis and the *v* axis are both one-dimensional subspaces. Their union is the set of all the vectors that begin at the origin and terminate at any point on one of the two lines. The set is not a subspace, however, because it is not closed under vector addition. Therefore the union of the axes cannot be represented by a point in the lattice of subspaces. The *u-v* plane is a subspace, namely the span of the two one-dimensional subspaces. Moreover, the span of *u* and *v* is the smallest subspace that includes all the vectors on both lines. The point in the lattice of subspaces that is the result of the operation *u-join-v* is therefore the point that represents the *u-v* plane.

Because of the peculiar properties of the operation of vector spanning, the lattice of subspaces is not distributive. Suppose the lattice includes points representing the four lines *s, t, u* and *v;* all the lines lie in the *u-v* plane and they all pass through the origin. Then *s-join-t* and *u-join-v* are both represented by the same point in the lattice, namely the point that corresponds to the *u-v* plane.

The point *s-meet-*(*u-join-v*) is by definition the largest subspace that *s* and *u-join-v* have in common; since *u-join-v* is the *u-v* plane and *s* lies in the plane, the largest subspace they have in common is the line *s* itself. On the other hand, *s-meet-u* and *s-meet-v* both include only the origin, or point 0 on the lattice, and the *join* or span of these expressions—that is, (*s-meet-u*)*-join-*(*s-meet-v*)—is also the point 0. It follows that the point designated *s-meet-*(*u-join-v*) and the point (*s-meet-u*)*-join-*(*s-meet-v*) are not identical in the lattice of vector subspaces.

If the logical connectives *and* and *or* and the relation of entailment are defined on the lattice of subspaces as they were on the lattice of subsets for the penny-quarter system, the structure of the lattice is the same as the structure of nondistributive quantum logic. The subspace *s* can be identified with the sentence "The *x* component of the spin is up" and the subspace *t* with the sentence "The *x* component of the spin is down." Similarly, the subspaces *u* and *v* correspond to the analogous sentences for the *y* component. The compound sentence "The *x* component of the spin is up and the *y* component is either up or down" is associated with the point in the lattice *s-meet-*(*u-join-v*). The compound sentence "Either the *x* component of the spin is up and the *y* component is up or the *x* component of the spin is up and the *y* component is down" is associated with the point (*s-meet-u*)*-join-*(*s-meet-v*). Since these lattice points are not identical, the logical structure associated with the lattice is nondistributive. Moreover, the lattice shows that (*s-meet-u*)*-join-*(*s-meet-v*) entails *s-meet-*(*u-join-v*) but not vice versa [*see illustration on this page*].



LOGICAL RELATIONS among the statements describing the spin of a particle can be modeled in a highly simplified form by lattices identical with the one outlined in heavy black lines at the bottom of the illustration on page 210. The lattice points correspond to sentences describing pure spin states. On each lattice the word *and* is identified with the *meet* of two points (*black arrows*), and the word *or* is identified with the *join* of two points (*colored arrows*). The sentences below the lattices, which are logically equivalent according to the distributive law, are modeled by the operations shown by the arrows. Because the structure of the lattices derives from the structure of the inclusion relation for vector spaces, the lattices are nondistributive. The two compound sentences are not represented by the same point in the lattice. Hence the lattices model the failure of the distributive law of logic to hold for statements about spin.

Confronted by the success of lattice theory in modeling logical relations within various physical theories, one tends to forget that the approach presupposes the resolution of an important philosophical issue. A dominant theme in philosophy during the past 200 years has been the thesis that there are two kinds of true statement: contingent facts about the world and truths of logic that would hold no matter what the condition of the world. The Scottish philosopher David Hume called the two kinds of statement "matters of fact" and "relations of ideas." The notion that the quantum theory might call for a revision of logic presupposes a denial of Hume's thesis that there are two kinds of truth.

Nevertheless, philosophers are now inclined to agree with Willard Van Orman Quine of Harvard University that no sharp distinction of this kind can be maintained. The laws of logic, Quine argues, have a central place in our web of beliefs, but a strong pull on that web at the periphery of observation can cause even the center of the web to become distorted. Logical considerations cannot themselves justify a revision of logic (how could they?), but one can (again to adopt a metaphor of Quine's) gerrymander the language when the fit becomes too strained between the natural world and the language we have inherited to describe it.

Even if it is granted that logic can be revised, there are many possible responses to quantum logic. One extreme is to deny that quantum logic is a logic and to say that what is going on is merely algebra under another name. The other extreme is to say that because quantum mechanics deals with particles that are the fundamental constituents of the universe, one should replace classical logic with quantum logic and learn to "think quantum logically," hard though that might be.

With respect to the first response one can say the following. Logic, although notoriously hard to define, deals with certain kinds of relations between sentences: what follows from what, what is consistent with what, and so on. Quantum logic does this too. What makes quantum logic peculiar is that it deals entirely with sentences stating that some vector lies in some subspace. The oddi-

CLASSICAL TWO-VALUED LOGIC

P	T	T	F	F					
Q	T	F	T	F					
NOT-P	F	F	T	T					
NOT-Q	F	T	F	T					
P OR Q	T	T	T	F					
P AND Q	T	F	F	F					
P IMPLIES Q	T	F	T	T					
P IS EQUIVALENT TO Q	T	F	F	T					

REICHENBACH'S THREE-VALUED LOGIC

P	T	T	T	I	I	I	F	F	F
Q	T	I	F	T	I	F	T	I	F
NOT-P (CYCLICAL)	I	I	I	F	F	F	T	T	T
NOT-P (DIAMETRICAL)	F	F	F	I	I	I	T	T	T
NOT-P (COMPLETE)	I	I	I	T	T	T	T	T	T
P OR Q	T	T	T	T	I	I	T	I	F
P AND Q	T	I	F	I	I	F	F	F	F
P IMPLIES Q (STANDARD)	T	F	F	T	T	T	T	T	T
P IMPLIES Q (ALTERNATIVE)	T	F	F	T	T	T	T	T	T
P IMPLIES Q (QUASI)	T	I	F	I	I	I	I	I	I
P IS EQUIVALENT TO Q (STANDARD)	T	I	F	I	T	I	F	I	T
P IS EQUIVALENT TO Q (ALTERNATIVE)	T	F	F	F	T	F	F	F	T

THREE-VALUED LOGIC proposed by Hans Reichenbach in 1944 maintains that some sentences describing quantum-mechanical phenomena are neither true nor false but indeterminate. For example, if the *x* component of the spin of a particle is known to be up, a statement asserting that the *y* component has a specific value would be classified as indeterminate in Reichenbach's scheme. Instead of focusing on the distributive law, Reichenbach defined enriched patterns of logical connectives by means of truth tables. The tables assign one of the three truth values—true (*T*), false (*F*) or indeterminate (*I*)—to each of the nine possible combinations of truth values for the sentences *P* and *Q*. Ten truth functions are designated as having a special place in the logic of quantum mechanics. The truth functions of ordinary two-valued logic are special cases of the extended truth functions (*color*). In three-valued logic there are 3^9, or 19,683, possible truth functions for derived sentences with two constituents.

ties of quantum logic are consequences of two requirements that must be met by any sentence to which the logic pertains. First, the sentences must be among those that ascribe quantum-mechanical properties to individual systems. Second, when two such sentences are linked by means of the quantum-mechanical analogues of *and* and *or,* the resulting sentence must still be descriptive of the physical system. In dealing with this closed set of sentences the logical relations among them are not those of classical logic.

Is one therefore forced to adopt the second stance and think quantum logically? In part the answer is yes. Certain kinds of statements met with in a physical theory fit together in ways that do not conform to classical logic. By no means, however, does this finding imply that classical logic should everywhere be replaced by quantum logic. Such a drastic revision of everyday ways of thinking could be justified only if it brought with it a vast simplification of the overall theory of the world; it is doubtful that such a simplification would be achieved. Even within quantum mechanics the logico-algebraic approach, although valuable, has not cleared away all the perplexities. Furthermore, even if quantum logic were totally successful in its own domain, the extension of it to other realms would seem peculiar. One arrives at quantum logic by considering the mathematical structure of the formalism of quantum mechanics. That mathematical structure, however, is based on the deductive patterns of classical logic. Hence classical logic is presupposed in the development of quantum logic.

One is left, therefore, with a family of logics that includes classical logic, quantum logic and perhaps other logics as well. Among them one logic still has priority. Although the nonclassical logics may have specialized applications, the logic employed for abstract reasoning, including reasoning about logic, will probably continue to be classical logic.

Image Processing by Computer

When the information in an image is expressed in digital form, it can be manipulated mathematically rather than optically. By such methods a blurred photograph can sometimes be restored to clarity

by T. M. Cannon and B. R. Hunt

Readers of *Scientific American* are familiar with some of the more spectacular products of digital image processing: enhanced satellite photographs of the planets (including the earth), colorful representations of more distant celestial objects and images assembled from X-ray data to show a cross section of the human body. The digital computer is now an essential tool in other areas of image processing as well. For example, an out-of-focus photograph or one streaked by movement of the camera can now be digitally processed to improve its resolution and to restore lost details; in many cases the cause of the blur can be determined from the blurred image itself. The techniques for deblurring images can be applied not only in scientific research and medicine but also in fields such as criminology and military intelligence. In addition investigators are currently working on digital methods for compressing the information in an image, a capability that could lead to greater efficiency in television transmission. Another potential application is the automatic recognition of certain recurrent patterns in large numbers of images, which would make it possible, for instance, to extract more information from satellite images of the earth.

What is responsible for the recent surge of activity in digital image processing? First, an image embodies a prodigious amount of information, and so most image processing is done with a large and powerful computer. As the continuing revolution in microelectronics has reduced the cost of such machines while increasing their capacity they have become more widely available. Second, the same advances in microelectronic technology have also led to the construction of improved devices for converting an image into digital data and vice versa. Third, the development of sophisticated algorithms, or mathematical procedures that can be executed by a computer, has made it possible to carry out image-processing operations that once would have been impossible or impractical.

We shall review here the basic principles of digital image processing and describe several of the more important techniques currently employed for the enhancement and the restoration of digitized images. In so doing we shall give particular attention to our own area of interest: the formulation of more effective algorithms for the deblurring of photographs that are out of focus or streaked by camera motion.

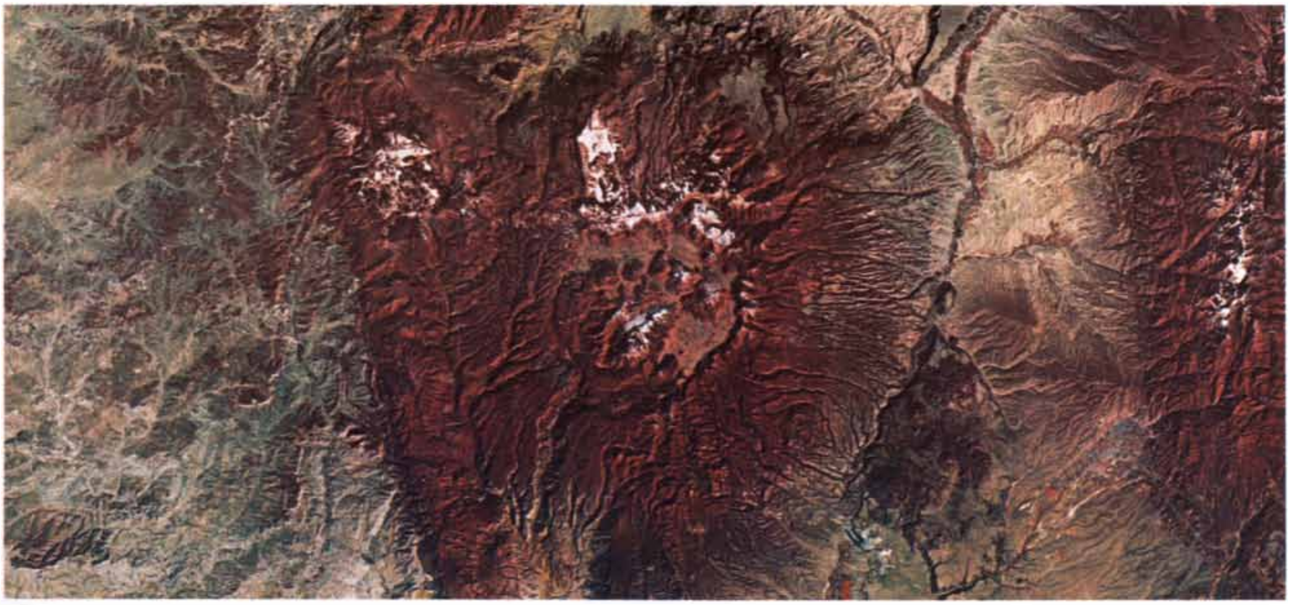
A digital computer processes information in discrete numerical units: digits. Most images, of course, do not come in such units. An ordinary photograph is an analogue representation of a scene; the information is recorded in continuous gradations of tone (and in some cases of color) across the two-dimensional surface of the film. Processing a photograph by means of a computer therefore requires that the analogue image first be converted into a digital one. (Certain photographs, such as Landsat images of the earth, are recorded digitally and so are already in a form suitable for computer processing.)

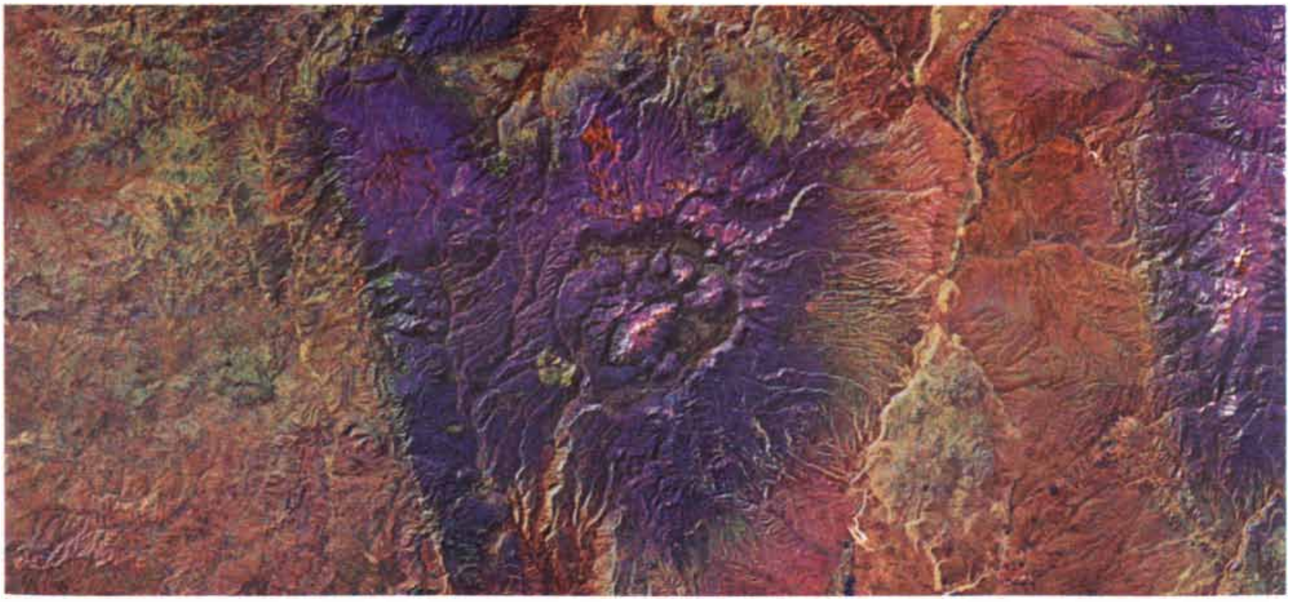
One method of converting an analogue image into a digital one is employed in broadcast television. In a television camera the image formed by a system of lenses is broken down into a set of horizontal scan lines. The changes in brightness along each scan line are transmitted as a continuous electrical signal. Hence a television image is partially digitized: it is reduced to discrete information in one dimension but not in the other. To go from this representation of the information in the image to one that can serve as the input to a digital computer requires one further step: each horizontal scan line must be broken down into a set of discrete values. In this way the entire area of the image is represented by an array of picture elements, which are called pixels. The brightness of each pixel is set equal to the average brightness of the corresponding area in the original image.

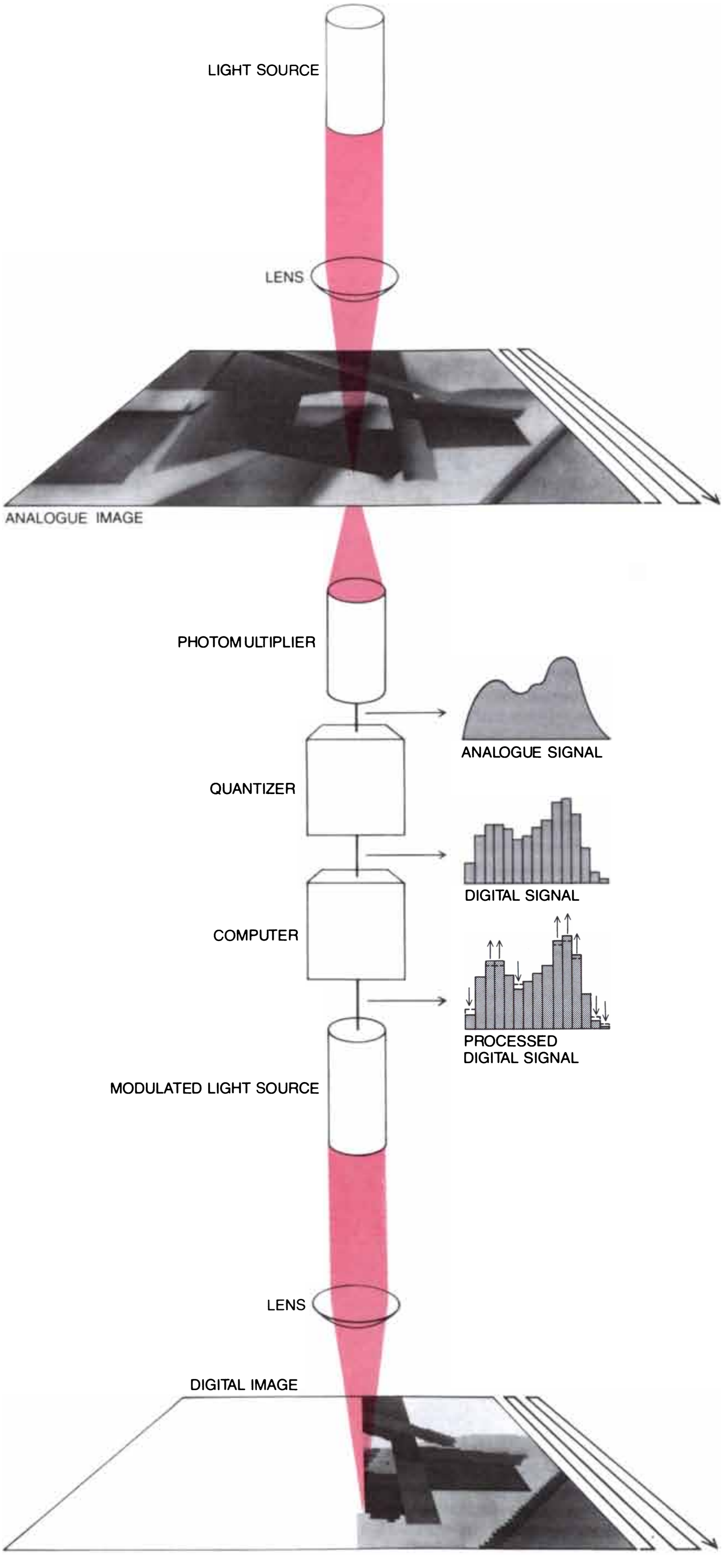
A number of more direct ways of digitizing an image are now available. A photographic image can be sampled by a scanning microdensitometer. In this device a small spot of light is projected onto the photograph (or through it in the case of a transparency) and a photomultiplier tube collects the light reflected from or transmitted through the film. The electrical output of the photomultiplier is compared with the output recorded when no image is in place, and the relative brightness of the image at each sampled point is calculated.

A similar device, the flying-spot scanner, differs from the microdensitometer only in certain mechanical details. In the microdensitometer the spot of light is stationary and the film moves back and forth on a gear-driven table. In the flying-spot scanner the film is stationary and the spot of light moves in a raster pattern. In the flying-spot scanner the light spot is usually generated on the face of a cathode-ray tube, and an optical system focuses an image of the moving spot onto the film as the electron beam sweeps back and forth inside the tube. More advanced image-sampling

LANDSAT VIEW of the Jemez Caldera in northern New Mexico, recorded on November 26, 1974, is presented in three strikingly different versions in a demonstration of digital image processing. The three photographs were prepared by workers at the Earth Satellite Corporation. The unenhanced image at the top is a photographic color composite made by combining data collected at three wavelengths by the satellite's multispectral scanning system. In such a false-color composite growing vegetation usually appears red. The digitally enhanced false-color image in the middle has been processed by a computer to correct certain defects in the original data (such as the gaps between scan lines) and to sharpen the edges of objects in the scene. The enhanced image at the bottom has been subjected to a color "ratioing" process to accentuate the contrast between features of the landscape. Santa Fe is in the lower right-hand corner. The river running from the top to the bottom of each image is the Rio Grande. The Los Alamos National Laboratory is in the eastern foothills of the ancient collapsed volcano.







systems, based on semiconductor technology, have been devised in recent years. In a charge-coupled device, for example, the entire image falls directly on a rectangular array of sensors. Each sensor records the brightness of its own sample of the image for transmission directly to the computer.

Semiconductor image-sampling devices offer another advantage besides their direct digital output. It is usually possible to make the response of the device linearly proportional to the intensity of the light falling on the semiconductor array. A linear response is not characteristic of a photographically recorded image. In photography the density of the silver grains in the developed film is proportional to the logarithm of the intensity of the incident light. The logarithmic response must sometimes be taken into account in the digital processing of a photographic image.

The process of sampling divides the image into discrete pixels, but the brightness of each pixel is still continuously variable; it must be quantized, that is, converted into a discrete numerical value. In this step the possible range of sample values is divided into a number of adjoining intervals called quantization levels. All sample values that fall within a given interval are replaced by the value assigned to that quantization level. The total number of quantization levels defines the gray scale: the number of shades that can be represented. If the gray scale has only two levels, every pixel must be either black or white. A gray scale with 64 levels can yield a fairly realistic reproduction.

An image that has been reduced to a set of binary numbers can be stored on magnetic tape or transmitted directly to the computer. The data can then be processed to obtain a new set of digital values, which in turn can generate a revised image. The new image is usually displayed on a cathode-ray tube, where

DIGITAL PROCESSING of an image such as a photographic transparency requires that the image first be converted from an analogue form into a digital one. This can be accomplished by the device called a scanning microdensitometer, shown here in a schematic diagram. A small spot of light is projected through the transparency, which moves in a raster pattern on a gear-driven table. A photomultiplier tube collects the light transmitted through the film and generates an electrical analogue signal proportional in amplitude to the sampled brightness of the image. The analogue signal is digitized in a device known as a quantizer before being transmitted to the computer, where the digital signal is processed. The processed signal can in turn modulate the light source of the microdensitometer; in this way the device can be operated in reverse to make a new photographic image, exposing the film one pixel, or picture element, at a time.

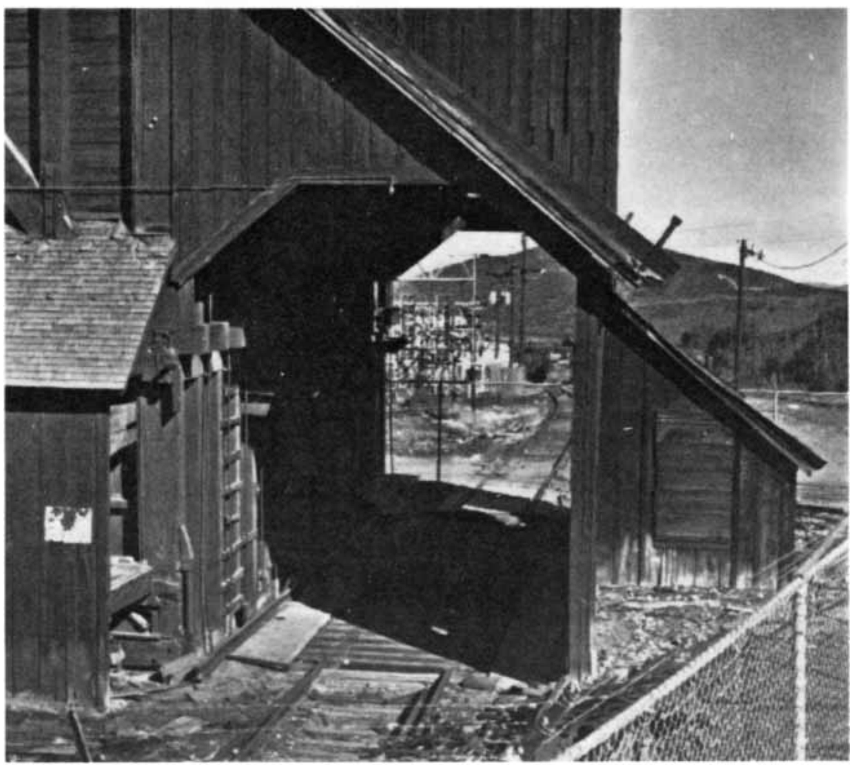
it is either viewed directly or photographed. Alternatively, an image-sampling device such as a microdensitometer or a flying-spot scanner can be operated in reverse to make a new photographic image. For the reverse operation an unexposed piece of film is mounted in place of the image to be scanned. The data from the computer modulate the light source, which exposes the film pixel by pixel. In general the reproduction quality of the resulting photograph is superior to that of a photograph made from the screen of a cathode-ray tube.

A display on a cathode-ray tube is created by transferring the data to a digital "refresh memory." All the data in such a memory can be extracted 30 times per second, which is the rate at which a complete image is "repainted" on a television screen. A series of digital values is extracted from the refresh memory, converted into an electrical signal and made to generate a scan line on the screen. The refresh memory provides a way of converting a static entity—the fixed values of the digital data in the sampled image—into the continuously varying signal required for viewing on a cathode-ray tube. One advantage of such an arrangement is that the image can benefit from interaction with the viewer; processing routines in the computer can be invoked to produce a new image, and the effect of the change can become visible on the screen in a few seconds.

Accurate and reliable devices for sampling and quantizing an image or for displaying an image reconstructed from digital data have been commercially available for more than a decade. Without them digital image processing would not be the active field it is today. Indeed, one way to gauge the growth of the field is to plot the increase in the number of companies that specialize in the manufacture of devices for getting images into and out of a computer.

Once an image has been digitized and transmitted to a computer, various mathematical operations can be carried out on the data in order to enhance the visual quality of the image. The operations generally fall into two categories: point-by-point processes and Fourier-domain processes.

A point-by-point operation on the elements of a picture is one in which the operation on any given point is done without reference to the surrounding points. The simplest operation of this kind consists in adding a constant value of brightness to each pixel. If the original image was a photograph, the addition of such a constant corresponds to increasing the exposure time or (equivalently) increasing the brightness of the illumination. Accordingly such an operation might be performed to compen-



EFFECTS OF DIGITAL PROCESSING on the visual quality of a black-and-white photograph can be seen by comparing this pair of digitized images. Because of the high contrast of the original photograph, shown in its unenhanced digital form at the top, it is difficult to print it without losing either fine detail or visual highlights. Computer processing can separate the two components that determine the brightness of each element of the scene: its illumination and its reflectance. It is therefore possible to apply a different degree of amplification to each component. The result of such a "crispening" procedure is shown in the enhanced image at the bottom. Some differences between the two images are readily noticeable, such as the apparent increase in illumination in the tunnel. Other differences are subtler, such as the added sharpness of the shingled roof and the increased clarity of the scrap-iron pile behind the fence.

sate for underexposure of the film. Conversely, the subtraction of a constant value of brightness from each pixel would compensate for overexposure.

Another point-by-point process is called image stretching. Suppose the brightness values of the pixels in a digitized image all lie in a narrow range (as is sometimes the case with certain X-ray images, for example). A narrow range of digital values means the original film lacked adequate contrast. By subtracting the minimum brightness value from each pixel and then multiplying the brightness of each pixel by a constant, the digital data are spread over a wider range and the resulting image has greater contrast. In this way subtle variations in brightness that were barely perceptible in the original film become more readily apparent.

A closely related point-by-point process is known as pseudocoloring. Instead of stretching the data to make fine variations in film density more noticeable, one assigns a different color to each shade of gray in the digital image. Areas that could be distinguished only by a subtle difference between two shades of gray may now be strikingly different colors. This technique, which has found application in fields as diverse as galactic astronomy and cell biology, is becoming so popular that several manufacturers are making modified closed-circuit television systems that do the pseudocoloring without the aid of a general-purpose computer.

A final example of point-by-point processing consists in calculating the antilogarithm of the brightness of each pixel. This operation essentially undoes the logarithmic encoding of information in photographic film and returns a measure of the actual light intensity that exposed the film. It is the digitized record of light intensities that is most often of value, since it is the light intensity that might have been altered by passage through the atmosphere or through a maladjusted optical system, resulting in a degraded image. In order to correct the degradation one must represent the image by the set of light intensities that

COMPUTER-ENHANCED COLOR IMAGE was processed by a technique similar to the one employed to make the enhanced black-and-white image on the preceding page. In this case the color and the brightness components of the digitized version of the original photograph (*left*) were processed separately, just as the reflectance and the illumination components of the black-and-white image were. The

formed it rather than by the response of a recording medium. Because most of the computer processing needed to correct such degradations is not of the simple point-by-point type we shall postpone our explanation of the methods until after we have introduced the concept of Fourier-domain processing.

Fourier-domain processing is perhaps most easily understood by considering what one hears when listening to a phonograph record. The music being played may be perceived as the sounds of many independent instruments, but it must all be embodied in a one-dimensional time signal generated by the motion of the stylus in the record groove. Although the listener may not be aware of it, he is making use of a theorem put forward in 1822 by Baron Jean Baptiste Joseph Fourier. The theorem states that any signal (in this case the music) can be represented by a sum of sine waves and cosine waves of various frequencies and amplitudes. The listener has some control over the amplitudes of the component waves. Turning the bass control changes the amplitude of the low-frequency sine and cosine variations in the music; the treble control affects the high-frequency components in a similar way. The volume control uniformly alters the amplitudes of all the sine and cosine components, regardless of their frequency.

Fourier introduced a mathematical procedure, now called the Fourier transform, for analyzing a signal into its fundamental frequency components. When the Fourier transform is applied to a time-varying signal such as music, it returns the amplitudes of the sine and cosine components of which the signal is composed in the frequency domain. These amplitudes, or Fourier coefficients, uniquely define the signal, and by adjusting them one can alter the makeup of the signal.

Just as a musical signal is made up of differences in air pressure that vary in time, so a picture is made up of differences in brightness that vary in space. The Fourier transform can be applied to

enhanced product (*right*) also benefits from a digital color-balancing process. For this demonstration the pixels of the images were made large enough for them to be seen individually. The images were "painted" on a cathode-ray tube and transferred to photographic film. These images and the ones on the preceding page were prepared by Brent S. Baxter and Oliver Faugeras of the University of Utah.

Lockheed knows how.

Advances in research tend to make headlines, but breakthroughs in manufacturing can be as important. The scene above depicts a manufacturing breakthrough that can result in easily fabricated buildings in space. That ingenious advance and others more Earth-bound are reported below.

Making building blocks for space.

Someday, astronauts will build a structure in space, and they may well use a new Lockheed manufacturing process.

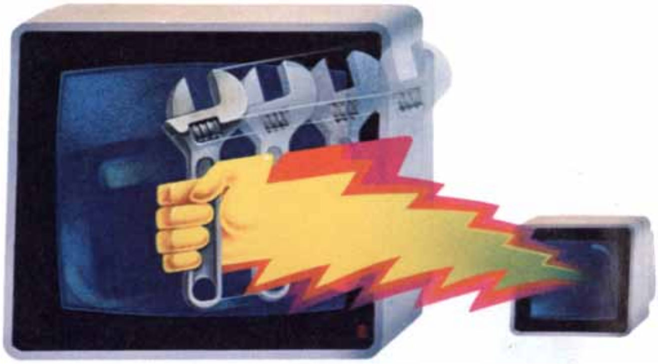
Light and very strong, composites—fibers of graphite embedded in epoxy—are gaining increasing use in aircraft and spacecraft. But making composites usually involves costly "cooking" in a huge oven called an autoclave. However, Lockheed now eliminates that process. It winds graphite fibers around a tube that tapers from 4 to 2 inches over its 8-foot length. Then the tube is heated—this avoids autoclaving—and epoxy is injected around the fibers. Result: an 8-foot structure weighing about 1½ pounds, ⅓ the weight of an aluminum structure of equal strength.

These tapered structures can be packaged inside one another like paper cups to save precious space when transported in the Space Shuttle. A special automated assembly machine designed by Lockheed under contract to the NASA Langley Research Center will enable astronauts to assemble them in space without tools.

Aspiration? Far from it. Lockheed has tested, proven and manufactured dozens of these remarkable structures.

Adding a work force by wire.

To increase productivity—virtually a nationwide goal—you simply make more efficient use of existing staffs and facilities. So logical, yet so difficult. However, Lockheed has invented a remarkable new computerized system that does just that.



It begins with another unique Lockheed System—CADAM. The computer-based CADAM basically gives engineers a three-dimensional drawing board as big as 48 football fields on which to design parts. CADAM's computer then electronically transmits the design to the plant's manufacturing experts.

Now Lockheed has linked CADAM computers at divisions across the country by a telecommunications network—COMCAM.

If one plant is overloaded, it calls on COMCAM and transmits a CADAM design to another division that can free staff and tools to make the part. This eliminates the cumbersome task of physically moving actual drawings, specifications and tapes from plant to plant. Changes in the

Astronauts could assemble buildings in space without using any tools, thanks to a Lockheed process described below.

design can be made at one plant and instantly transmitted to another. Put simply, COMCAM makes better use of people, machines and time—increased productivity.

Calling up the best brains in manufacturing.

The CADAM system, which was invented at Lockheed, enables engineers to design aircraft parts on a computer terminal with greater precision and imagination than ever before.

To help its manufacturing experts translate those designs into actual airframes, Lockheed has given them an even more remarkable computer-based system—GENPLAN.

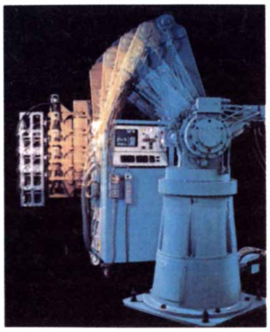
Exclusive to Lockheed, GENPLAN actually makes decisions such as which tools and manufacturing processes should be used. It determines the best manufacturing sequence. It considers factors such as labor standards and factory rules. Within GENPLAN's huge data base, Lockheed has captured the accumulated knowledge of thousands of manufacturing experts. Moreover, Lockheed divisions thousands of miles apart can share this knowledge.

Working at a computer terminal, the manufacturing planner can create, review and edit the manufacturing plan.

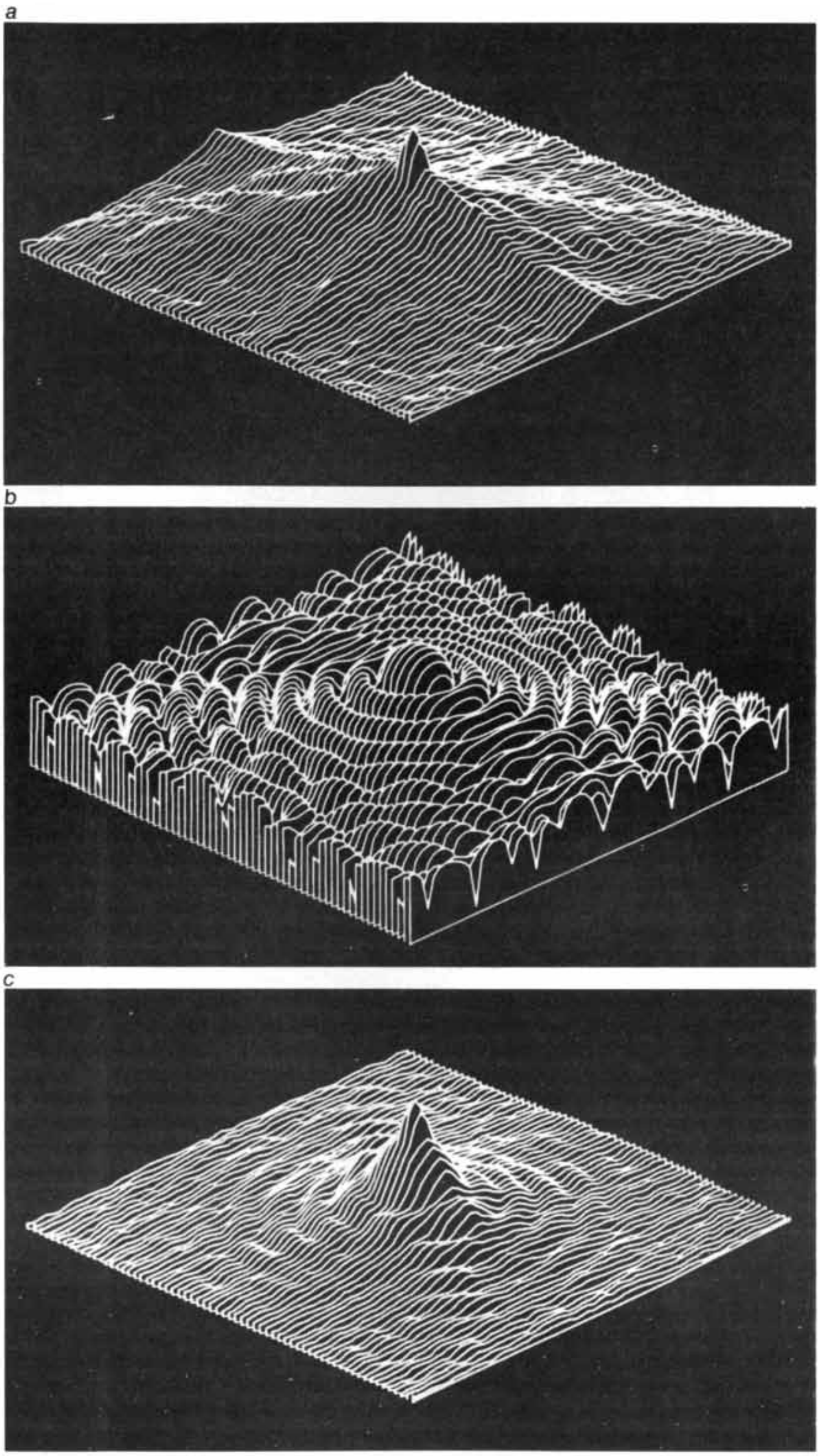
The time savings of GENPLAN are immense. Plans that took hours now are made in minutes. The steps required to release the plan to the assembly line have been reduced by 75%. Another key benefit flows from GENPLAN. The increase in precision and reliability of manufactured parts is as dramatic as the time savings.

Robots roll up their sleeves.

The extended "arm" that you see in motion is a robot being "trained" to place rivets with great speed and precision. Once programmed—and programming is critical in the effective use of a robot—it may cut riveting costs 50%. Other robots are at work at Lockheed, including machines that handle onerous painting duties and that can "see" and choose among different parts.



Lockheed



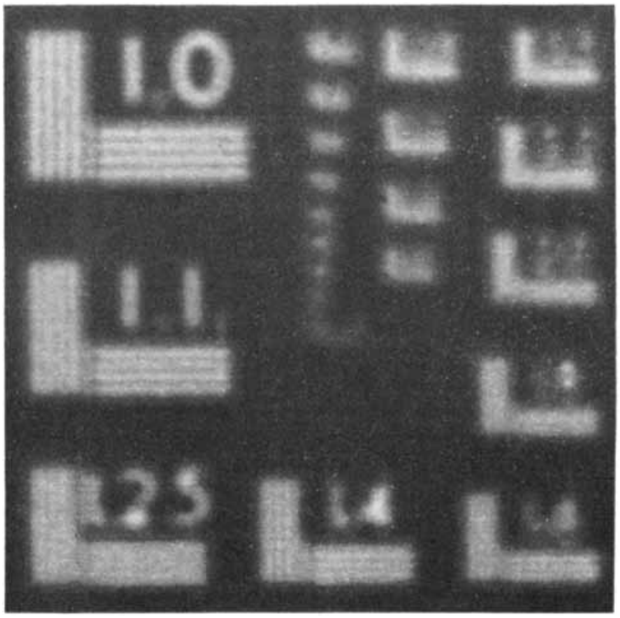
COMPUTER-GENERATED GRAPHS correspond to the Fourier transforms of an in-focus photograph (*a*), the optical transfer function of a defocused lens system (*b*) and an out-of-focus photograph of the same scene made through the defocused lens system (*c*). A Fourier transform represents the spatial variations in the brightness of an image as the sum of the amplitudes of the sine waves and cosine waves that constitute the image in the spatial-frequency domain. In such a two-dimensional graph of the spatial frequencies in an image the low-frequency Fourier coefficients are at the center and the high-frequency coefficients are farther out. The magnitude of each coefficient corresponds to the height of the surface above the base plane. The pattern in *c* can be obtained by multiplying *a* and *b;* in the process some of the Fourier coefficients of the blurred image go to zero and others are attenuated, particularly at higher frequencies. The digital restoration of the image is effected essentially by multiplying *c* by the reciprocal of *b* to obtain *a*. The degree to which the lens system was defocused can be deduced from the concentric rings of zeros in the frequency-domain representation of the image.

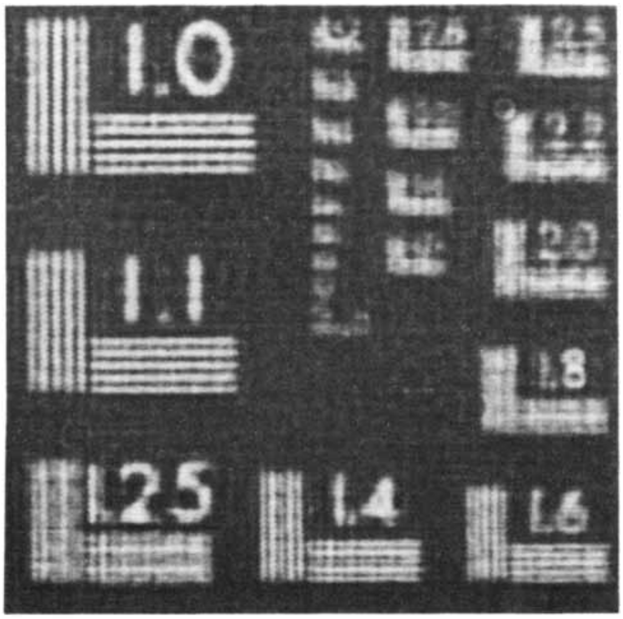
the image to yield a set of spatial frequencies that indicate how rapidly the brightness changes from one pixel to the next. The Fourier coefficients define the amplitudes of these two-dimensional, spatial sine and cosine waves. Moreover, in the same way as one can vary the high- and low-frequency components of a musical signal, one can also vary the frequency components of an image.

One of the simplest Fourier-domain image-processing techniques is edge enhancement. Edges are usually characterized by an abrupt change in the brightness of an image, and hence they are represented by the high-frequency coefficients in the Fourier transform. Increasing the magnitude of these coefficients increases the sharpness of the edges of objects in the image. The sharpening process consists in computing the Fourier transform of the image, changing the Fourier coefficients and then regenerating an image by means of an inverse Fourier transform.

Thomas G. Stockham, Jr., working at the Lincoln Laboratory of the Massachusetts Institute of Technology, incorporated many of the enhancement techniques mentioned above into a process known as crispening. Stockham considers an image to be the product of two quantities: the reflectance of the objects in the scene and the illumination falling on the scene. The computer is programmed to calculate the logarithm of the brightness of each pixel in the image. Because multiplying two numbers corresponds to adding their logarithms, the value obtained by this procedure represents the sum of the logarithms of the reflectance and the illumination. It has been determined empirically that the Fourier transform separates these two quantities (roughly speaking) into two sets of Fourier coefficients. The variation of the illumination from place to place tends to be rather slow, and so it is represented by the low-frequency Fourier coefficients. The reflectance of the scene tends to include rapid spatial variations such as changes in color and surface texture that are embodied in the high-frequency coefficients. By altering the magnitudes of the coefficients in the two groups one can, for instance, moderate changes in illumination while simultaneously causing the reflective properties of the scene to be more sharply defined. Earlier techniques usually achieved one of these goals at the expense of the other.

The crispening technique is particularly interesting in that it parallels a similar process carried out by the human visual system. The eye responds in an approximately logarithmic way to the intensity of the light that stimulates it. In addition the visual system is more sensitive to higher spatial frequencies than it is to lower ones. Therefore when one





OUT-OF-FOCUS PHOTOGRAPH of a test chart was deblurred with the aid of a computer algorithm. The algorithm first determined from a digital version of the blurred image (*left*) the extent to which the camera lens was defocused when the photograph was made. The algorithm then proceeded to generate the restored image (*right*). In the blurred image some groups of five bars appear to have only four bars; in the restored image the five bars show clearly. The restoration was done by Joel Trussell of the Los Alamos National Laboratory.

looks at a scene, what one sees is the logarithm of the scene with the higher frequencies emphasized. Digital crispening does the same thing, and so it assists the human visual system.

The crispening technique has also been employed to improve the color quality of an image. The procedure consists in separating the image into brightness and color components, which are then processed in a manner analogous to that employed for the illumination and reflectance components of a black-and-white image [*see illustration on pages 218 and 219*].

Up to now we have been concerned exclusively with computer-processing techniques that enhance an image and thereby allow the information in the image to be more readily and accurately perceived by the human visual system. Enhancement may be carried out even though there is nothing inherently wrong with the image; it is not necessarily a corrective operation.

Image restoration is undertaken as a remedial measure. Some images are created under adverse circumstances; for example, a scene might be photographed through a turbulent atmosphere or through an out-of-focus lens system resulting in a blurred image. In such cases the task is to restore rather than merely to enhance the image. The key to a successful restoration of a blurred image is an adequate mathematical description of what is wrong with the image; without such prior knowledge computer processing is usually of little value.

Some blurs are easy to describe mathematically. For example, many nearsighted people may have noticed that when they remove their glasses, a small point of light such as a star becomes a small disk of light. A defocus blur of this type is described by a point-spread function: each point of light entering the optical system emerges as a spread-out disk of light. If an image is passed through a defocused lens system (either in the human eye or in a camera), each point in the image is spread out over neighboring points. The analytic description of the result, which mathematicians call a convolution, is beyond the scope of this article, but there is a Fourier-domain representation that is fairly easy to understand.

The Fourier transform of the point-spread function is referred to by optical designers as the optical transfer function of the lens system. The Fourier transform of an image that passes through such a system is multiplied by the optical transfer function (which is equivalent to the image itself being convolved with the point-spread function). The product of the multiplication is the Fourier representation of the blurred image. When the optical transfer function is that of an out-of-focus lens system, some of the Fourier coefficients of the blurred image may be zero, which means that some information has been irretrievably lost. Other Fourier coefficients are attenuated, particularly at higher frequencies, which accounts for the lack of sharpness in an out-of-focus image. The attenuation is mathematically similar to having adjusted the treble control of an amplifier to muffle the higher frequencies of a musical signal. Many other image blurs have a similar mathematical description and a corresponding loss of high-frequency image information.

Let us now return to the problem of restoring a blurred image. If the image is out of focus and the degree of defocus is known, one can effect a restoration by multiplying the Fourier transform of the blurred image by the reciprocal of the optical transfer function of the defocused lens system. The result should be the Fourier representation of the unblurred image. Since some of the points of the optical transfer function may be zero, however, the reciprocal is undefined at those points. A possible solution to this problem is to modify the reciprocal to eliminate the zeros. This approach is conceptually simple, but it requires much trial and error.

Most successful restoration methods take into account not only the optical transfer function but also other forms of degradation, such as visual noise. (A good example of visual noise is what is seen on a television set tuned to a channel where no signal is received.) The fine grains of silver that constitute a photographic image contribute a noise component to the image. The problem often encountered in deblurring such an image is that part of it, namely the noise, did not pass through the degrading system and therefore is not blurred. The noise is particularly troublesome at spatial frequencies where the transfer function approaches zero. At these frequencies the blurred picture may receive a greater contribution from the noise than it does from the original scene. It is at precisely these frequencies that the reciprocal of the transfer function has very large values and hence gives rise to great amplification.

A restoration procedure that does not take this fact into account will excessively amplify the unblurred noise and produce a restored image that may well resemble the aforementioned television image. Because of these difficulties and others most restoration methods rely on three sources of information: the original scene, the noise and the transfer function. The resulting formulations can sometimes be solved only by advanced methods of numerical analysis based on iterative computer algorithms, but they yield a restoration that is optimal in a statistical sense.

The procedure outlined above for restoring an out-of-focus image relied on the subtle assumption that the degree of defocus is known. If one did not have the aid of a digital computer, this would be a difficult item of information to obtain. It would require painstaking measurements of the camera, assuming that it had not been readjusted since the defocused photograph was made. It would also require that one have access to the camera, which in some instances is impractical. For example, the camera might be on board a space vehicle. In the case of a blur caused by movement of the camera one must know not only the extent of the camera motion but also the direction.

Fortunately clues to the type and severity of the blur can be obtained from the Fourier transform of the blurred photograph. The clues are found in the areas where the Fourier coefficients are zero. Such zeros in the Fourier transform, it will be remembered, result in a loss of information in the image, but they also enable the cause of the loss to be determined.

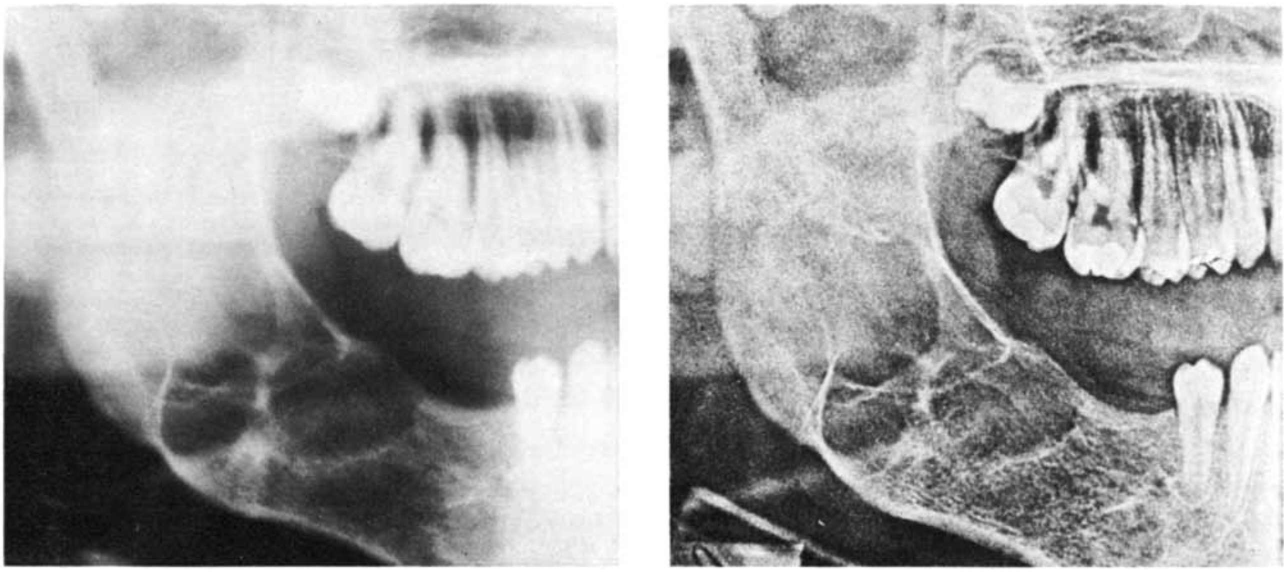
The illustration on page 222 is a two-dimensional graph of the Fourier transform of an in-focus photograph, a graph of the optical transfer function of a typical maladjusted lens system and a graph of the Fourier transform of an out-of-focus photograph. The valleys in the transfer function are concentric rings of zeros. It can be seen that the Fourier transform of the blurred photograph resembles the product of the transfer function and the Fourier transform of the original image. In particular the rings of zeros are reproduced in the transform of the blurred image. If one knows that the photograph was made with an out-of-focus lens, the spacing of the rings of zeros indicates the degree of the defocus. A camera-motion blur also leaves a telltale signature in the form of zeroed information in the Fourier domain. In this case the zeros in the Fourier domain form parallel lines.

Utilizing the zeroed spatial frequencies in the Fourier domain to determine the cause of a blur is analogous to finding a faulty piano key by listening to a sonata. Information at certain sound frequencies will be missing from the sonata, and from the missing frequencies one can identify the defective key. Similarly, the missing spatial-frequency information in a blurred photograph reveals the cause of the blur.

Detecting a pattern of zeros in the Fourier domain can be difficult. Indeed, the clarity of the rings in the illustration on page 222 is exceptional. Usually the rings are obscured by random fluctuations in the Fourier transform, and identifying them calls for a more sophisticated approach. Suffice it to say, however, that when blurring is caused by defocusing or by movement of the camera, a generally accurate description of the cause of the degradation can be obtained by digital processing. Methods of correcting other kinds of blur remain more elusive. For example, it is usually not possible to restore a picture that has been degraded by atmospheric turbulence. The reason is that it is not yet possible to obtain an accurate description of the blur. Indeed, one of the most active areas of investigation in digital image processing is the attempt to devise a general method of determining the nature of a blur from the degraded image itself.

Perhaps the most important application of digital image processing is in the area of remote sensing, particularly in the analysis of satellite images of the earth. Consider the images transmitted back to the earth by the satellites in the Landsat series. The satellites are equipped with scanning devices that record images of the earth passing under them in a series of strips. Each strip is 185 kilometers wide, and it is usually

MOTION-BLURRED PHOTOGRAPH was made by intentionally moving the camera during the exposure. A computer program determined the direction and the extent of the camera's motion from the digitized photograph (*top*) and then produced the restored image (*bottom*).



DENTAL X-RAY IMAGE was enhanced with the aid of a computer in this demonstration of the applicability of digital image processing to medicine. Deficiencies such as low resolution and a narrow gray scale (or lack of contrast) are alleviated by the enhancement.

subdivided (by computer processing on the ground after the data are received) into a series of images 185 kilometers on a side. The rotation of the earth brings new ground coverage into view on each successive orbit, with each satellite returning to its initial point of coverage every 18 days. The first two satellites in the series, *Landsat 1* and *Landsat 2*, are nine days apart; in other words, nine days after a point on the earth is photographed by *Landsat 1* it is photographed by *Landsat 2*. The volume of images produced by the satellites is impressive: as many as 1,500 scenes per week may be processed, with each scene being represented by a matrix of more than 7.5 million pixels. The images have a resolution of 57 by 79 meters (the area of a single pixel). More than 90 percent of the earth's surface has been surveyed under cloud-free conditions at this resolution.

The great volume of data in the Landsat images would require computer processing even for strictly clerical functions such as the annotation of the time and geographical position of each observation. Clearly it is desirable to explore ways of enlisting computer processing to present the images with the greatest possible efficiency and to enhance them for the extraction of the maximum amount of information.

A demonstration of the visual information that can be gained through computer processing is provided by the Landsat multispectral images. Each ground area photographed by the Landsat satellites is recorded through a series of narrow-band color filters. The multispectral scanner records two images at wavelengths in the visible spectrum and two in the infrared spectrum. Areas of the terrain that differ in certain physical properties reflect different amounts of radiation in each of the four wavelength bands of the multispectral scanner. Thus a forest yields a strong signal in the green band, whereas a desert yields a weak green signal.

By combining images made at different wavelengths it is possible to enhance the information display. For instance, consider an image of a forest growing in partially open and rocky land. An image made with green light shows intense reflections from the vegetation. An image in the red region may be more sensitive to the earth tones and rocks, and it will show these details. A computer can be programmed to divide the brightness of each point in the green image by the brightness of the corresponding point in the red image; the result is a "ratioed" image that accentuates the forest more than either of the original images did. Points with a strong signal in the green band (the trees) are divided by points with a weak signal in the red band (also trees) and hence their brightness is magnified. Weak signals in the green image (the rocks) are decreased further in strength by dividing them by strong signal values (the rocks) in the red image. The illustration at the bottom of page 215 shows the result of such a ratioing process. Three pairs of multispectral Landsat images were ratioed, and the three quotient images served as color primaries for a pseudocolor composite image. The color combinations in the composite may assist geologists and agricultural scientists in identifying areas of interest.

The potential of remote sensing for helping to solve human problems is great. For example, the growth of a crop such as wheat is accompanied by predictable changes in its multispectral image. The onset of various crop diseases, such as blight in corn or rust in wheat, can be detected from changes in the multispectral image. Experiments of this kind have led to the suggestion that it might be possible to monitor food crops throughout the world and to predict the world harvest from satellite images. Testing this hypothesis is the purpose of the Large Area Crop Inventory Experiment, which is currently being undertaken by the U.S.

The volume of images from a fully operational series of satellites could overwhelm a team of human interpreters. Ultimately the successful utilization of such photographs may depend on the development of methods for automatically extracting pertinent information from the images. Accordingly there is much interest in the development of computer programs for the recognition of patterns in images. Programs have already been demonstrated in which the computer processes multispectral images of an urban area and classifies the environment into various categories of land use such as streets and parking lots, parks and greenbelts, factories and industrial districts. Similar work is under way on the automatic recognition of various crops and land uses in rural areas. Because most deposits of mineral resources are associated with characteristic geological features, efforts are also being made to automatically isolate and extract from remotely sensed images such geological details as faults and escarpments. The natural-resource companies have asked for more Landsat images than any other group has. The expectation is that the automatic extraction of information from such images will someday be one of the most active areas of digital image processing.

THE AMATEUR SCIENTIST

The aerodynamics of the samara: winged seed of the maple, the ash and other trees

by Jearl Walker

The winged seeds that spin so gracefully to the ground from ash, elm and maple trees are called samaras. Watching them in flight might arouse your curiosity about their aerodynamics. It is not as simple as it looks.

The first detailed analysis of the aerodynamics of single-wing samaras such as those of the maple was done in the early 1970's by R. Åke Norberg of the University of Göteborg. Further studies were carried out by Charles W. McCutchen of the National Institute of Arthritis, Metabolism, and Digestive Diseases and by F. M. Burrows of the University of North Wales. I shall first follow the work by Norberg, who applied to samaras the aerodynamics of helicopters. Then I shall briefly review McCutchen's studies of samaras other than the maple's.

Norberg made two major assumptions. One was that the samara is a flat wing. The other was that the mass of the wing can be considered as lying on the long axis from the seed to the wing tip.

A spinning maple samara does not accelerate downward, in spite of the effect of gravity, because an aerodynamic force is applied by the air through which it falls. For a simple explanation of the force one can regard the samara as spinning about its center of mass, which falls along a vertical axis. (The center of rotation is actually slightly off the center of mass.) The wing is essentially horizontal and sweeps out a disk in which the flow of air is uniform. One can visualize the flight from either of two viewpoints: the usual one of an observer watching the seed fall or one in which the observer imagines himself falling along with the samara.

In the first viewpoint the samara drops into a column of still air. Each horizontal layer of air in the column is accelerated downward as the samara passes through it. The column of air left above the samara is moving with a final downward speed lower than the sinking speed of the samara. The air in the column reflects the effects of a downward force because it has been accelerated from a stationary state to a final downward speed. The corresponding reaction force on the samara is upward.

From the viewpoint of the observer falling with the samara the column of air below the samara moves upward with a speed equal to the samara's true sinking speed. Each layer of air is accelerated downward as it passes through the disk swept out by the spinning wing. After the acceleration a layer of air is moving upward at a lower speed. Thus the layer reflects the effects of a downward force that has decreased the layer's ascent. The corresponding reaction force on the samara is upward. In each case the samara is affected by an upward force.

This aerodynamic force cancels the weight of the samara, preventing it from accelerating downward. For the observer on the ground the samara moves downward at a constant speed. For the observer falling with the samara the upward-moving column of air approaches at a constant speed equal to the true sinking speed of the samara. The sinking speed is proportional to the square root of a ratio called the disk loading. This ratio is equal to the weight of the samara divided by the area of the disk swept out by the wing during a full spin about the vertical axis of the fall.

Different maple samaras descend at different speeds because they do not have the same disk loading. One obvious reason for the difference in disk loading is the variation in the size and weight of the samaras. A less obvious reason is that a spinning samara does not actually have a horizontal wing. The angle between the plane of the wing and the horizontal differs from one samara to another. Some seeds descend with their wing almost flat; in others the wing is tilted upward at an angle of 45 degrees or more to the horizontal. The disk loading is larger for a tilted wing than it is for a more horizontal one, and the samara falls faster.

Analyzing the aerodynamic force on a maple samara with a tilted wing is difficult for an observer on the ground, and so I shall take the viewpoint of the falling observer. It is portrayed in the illustration on page 230; the relative motion of the air passing the wing is upward and toward the thinner trailing edge of the wing. This motion provides the aerodynamic force to counter the samara's weight.

From the viewpoint of the falling observer several types of data are needed to study the force on the wing. The vertical axis about which the samara is truly spinning is the axis of rotation. I shall assume that it passes through the overall center of mass of the samara, which is near the end containing the seed. The span of the wing is the long axis that runs from the seed to the outer tip of the wing. The chord is a shorter axis between the leading edge and the trailing edge of the wing; it is perpendicular to the span. The angle the span forms with the horizontal is called the coning angle.

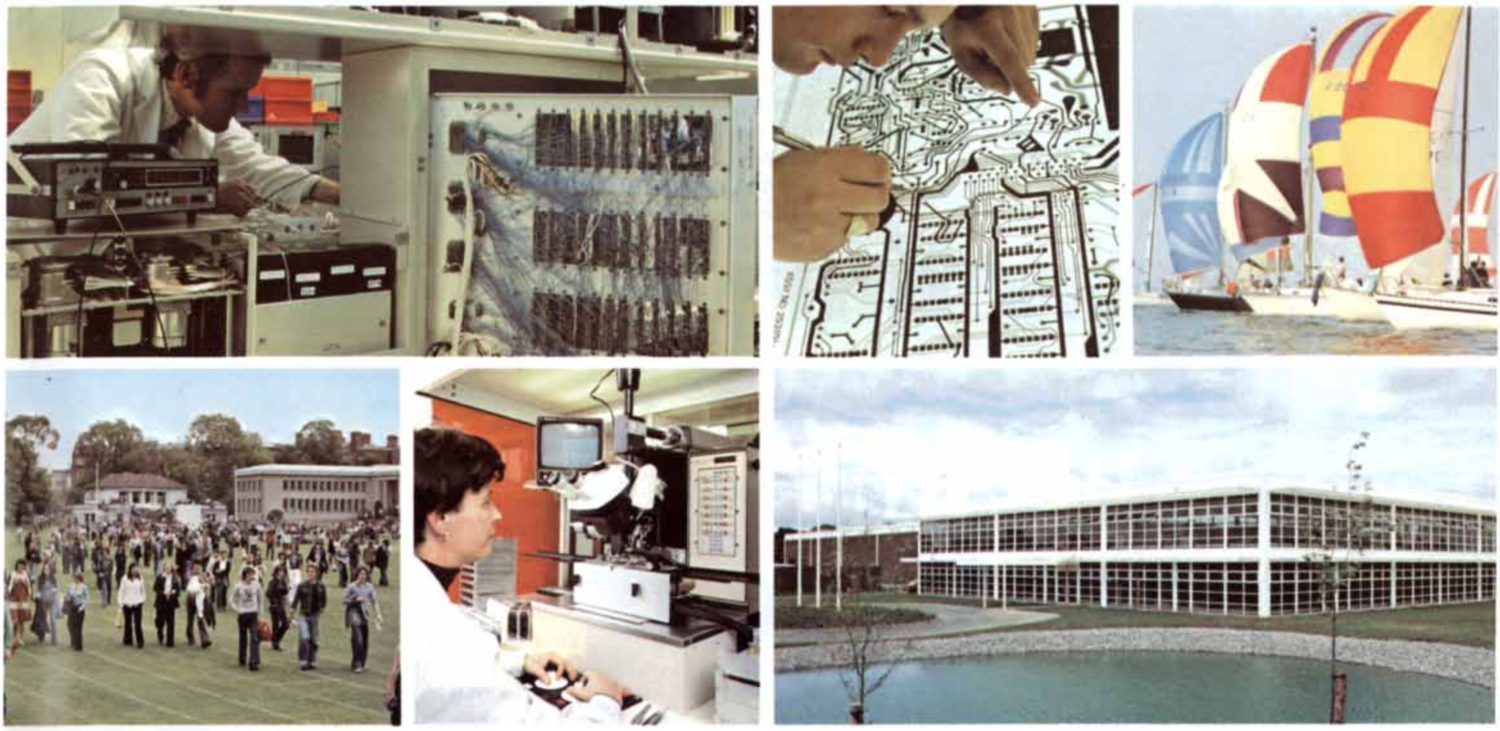
The aerodynamic force can be calculated by assuming that the wing is divided into thin strips running from the leading edge to the trailing edge. The force on each strip is computed by multiplying the area of the strip by the square of the speed of the air moving past it. The net force on the wing is found by adding the forces on the individual strips.

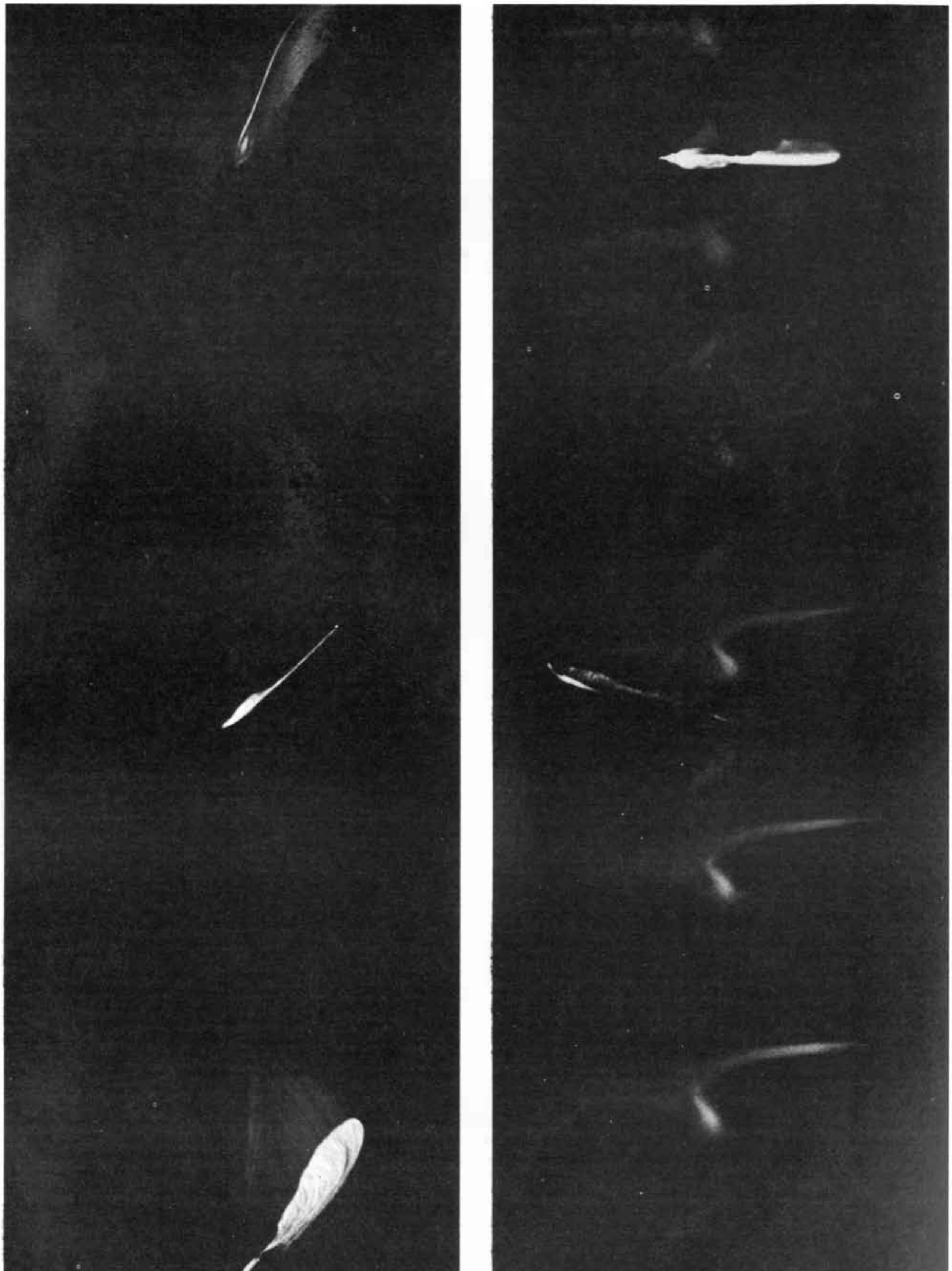
The air moving vertically past the wing can be divided into two components, one perpendicular to the span and one parallel to it. Only the perpendicular component contributes to the aerodynamic force supporting the wing. The movement of air parallel to a chord from the leading edge to the trailing edge of the wing is more difficult to evaluate because its speed depends on the distance between the axis of rotation and the strip being considered. The air moving along the chord of a strip near the axis of rotation has a relatively low speed. (The reason is that such a strip moves comparatively slowly as the wing spins.) The air moving along the chord of a strip farther from the axis has a higher speed. To demonstrate the effect of the movement of air along a chord I shall concentrate on only one strip positioned about midway on the wing.

Air moves past this strip both vertically and backward along a chord. Of the vertical motion only the component perpendicular to the wing matters. The net velocity vector of the air moving past the strip lies somewhere between a line perpendicular to the wing and the chord running through the strip.

If the contribution from the vertical motion of air is relatively large (if the samara is sinking fairly fast), the net velocity vector is nearly vertical. If instead it is relatively small, the net velocity vector is closer to being parallel to the chord. A subtle shift between these two possibilities is a key to some of the stability of a descending samara.

The magnitude of the net velocity vector determines the magnitude of the aerodynamic force on the strip. Most of the force on the wing as a whole arises





Falling maple samaras photographed by repetitive flash from the strips farther out on the wing. One reason is that the chords are longer there, and so the strips have more area. Moreover, the net velocity of the passing air is larger because the speed of the air moving parallel to the chord is greater. Hence a properly designed samara is narrow in chord near the center of mass, where little aerodynamic force is available, and wider in chord farther out.

The net aerodynamic force on the wing is perpendicular to the wing and also lies in a vertical plane that passes through the span. The vertical component of the force balances the downward weight of the samara. If you toss a maple samara into the air, it must somehow adjust the orientation of its wing so that the vertical component of the aerodynamic force matches its weight. It is impressive to discover that a descending samara achieves this adjustment automatically.

Once a samara has achieved the proper orientation it must rapidly counter any chance perturbations by breezes. Most samaras are inherently stable in at least four respects: the wing's angle of attack, the coning angle, the tilt of the circle traced out by the outer tip in relation to the horizontal, and the possibility of sideways motion.

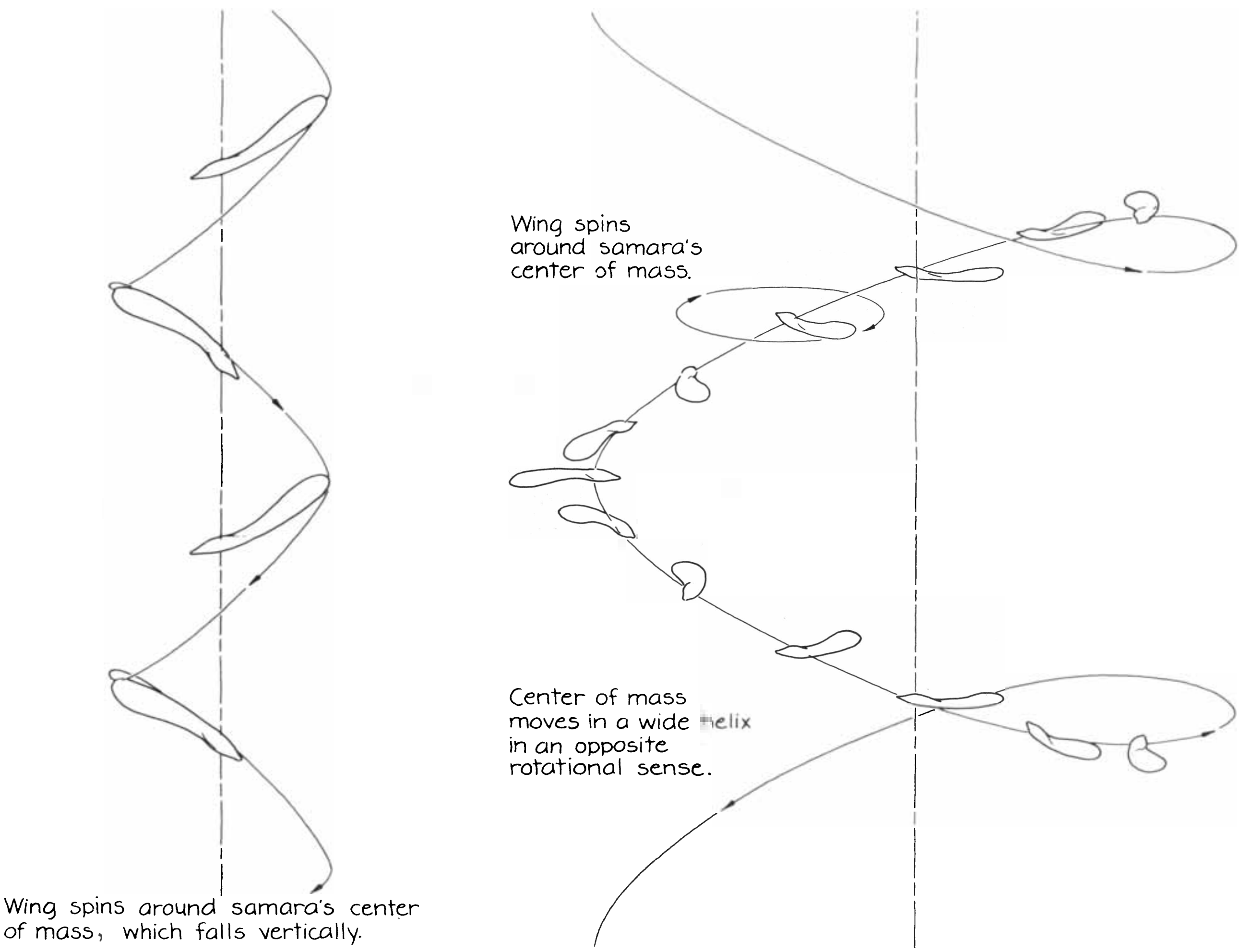
The angle between the chord of a wing and the velocity vector of the passing air stream is called the angle of attack. The bottom illustration on page 233 shows a cross section of a samara's wing at about the midpoint of the span. The net aerodynamic force on this section of the wing is indicated by a single force vector operating through a point called the center of pressure. The effects of the aerodynamic forces on each small part of the section are the same as is indicated by this one vector. The position of the center of pressure is determined by the shape of the section and the direction of the passing air.

The section also has a center of mass reflecting the weight of the section. The weight is insignificant, however, compared with the centrifugal force on the section. This force, resulting from the spin of the samara, also operates through the section's center of mass.

The orientation of the section with respect to the velocity vector of the passing air is stable if the center of pressure coincides with the center of mass. Thus a certain angle of attack is desirable for stability. Suppose the wing is shifted suddenly in such a way that the center of pressure moves toward the trailing edge of the wing. This nose-up orientation results in a larger angle of attack, which generates a different aerodynamic force on the wing. The vertical portion of the force no longer balances the samara's weight as is needed for a stable, prolonged flight.

The samara is designed to correct this situation by regaining the proper angle of attack. The force operating through the center of pressure creates a torque that rotates the wing section about the center of mass and back to its original orientation. The center of pressure returns to the center of mass and the samara is again stable. Something similar happens if the perturbation reduces the angle of attack from its optimum value, moving the center of pressure forward from the center of mass and sending the wing into a nose-down orientation. Again the force through the center of pressure rotates the section about the center of mass until the proper orientation is regained.

According to research on the aerodynamics of gliding flat plates, such an au-



The flight paths a maple samara might take

tomatic adjustment of a wing's angle of attack is possible if the mass is distributed in a certain way along the chord of the wing. Consider again a cross section through the wing. The center of mass of this section must be behind the leading edge by a distance that is between 27 and 35 percent of the chord length of the section. The center of mass must also be behind the most forward position to which the center of pressure could move, otherwise the wing would not survive a nose-down perturbation. The samara has evolved with just such a distribution of mass along the chord. The ribbing in the wing is bundled near the leading edge and spread near the trailing edge.

Another stability requirement has to do with the angle between the horizontal and the glide path of the wing. The samara actually spirals, but the angle can still be defined if the glide path is taken as a tangent to the downward spiral at any given instant. The glide angle is important because it determines the orientation of the aerodynamic force on the wing. At the proper angle, called the natural glide angle, the force is in a vertical plane that passes through the span of the wing. If the glide path is too steep, the force vector inclines toward the leading edge of the wing. If the path is too shallow, the force vector inclines toward the trailing edge of the wing. Neither of these orientations will give rise to a stable flight.

As a samara begins its fall to the ground it must rapidly achieve the proper glide angle so that the aerodynamic force on it lies in a vertical plane. Only then is the vertical component of the force able to cancel the weight of the samara to provide a constant speed of descent. If in the early stages of descent the vertical component of the force is smaller than the weight (because the velocity of the passing air is lower than it should be), the samara must adjust its glide angle accordingly or it will accelerate all the way to the ground.

The steps in the correction are as follows. Since the samara is now accelerating downward, the passing air has a velocity vector that is more vertical than it should be. The angle of attack is too large; the wing of the samara noses down to gain the proper angle of attack. The aerodynamic force shifts from being vertical (as it should be) to being canted toward the leading edge of the wing. This leaning of the force vector propels the wing into a greater spin about the axis of rotation.

From the viewpoint of the observer falling with the samara the extra spin amounts to an increase in the speed of the air moving along a chord toward the trailing edge of the wing. The net velocity vector of the passing air (which includes both the vertically moving air because the samara is falling and the air moving along the chord because the samara is spinning) was initially too vertical. With the increase in airspeed along the chord the velocity vector becomes less vertical. Now the angle of attack is too shallow. The wing noses up. The movement restores the aerodynamic force to a vertical plane, and the spin no longer increases.

In the process of adjustment the samara has increased its speed of descent (because it initially was accelerating downward) and its rate of spin (because the tilt of the aerodynamic force during the early part of the adjustment propelled the wing around the axis of rotation). Hence the velocity of the passing air is now higher and the force is stronger, just strong enough for the vertical component of the force to match the weight of the samara. Thereafter the samara falls with a constant speed and spin, with the proper angle of attack and with a suitable glide angle.

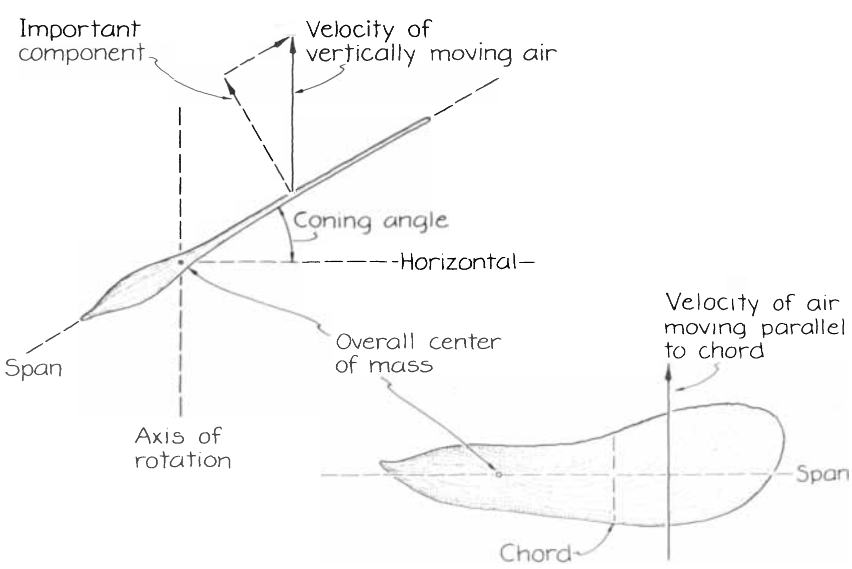
If the initial glide path results in too high a velocity for the passing air, the vertical component of the aerodynamic force exceeds the weight. The samara executes a similar kind of adjustment to reduce the aerodynamic force. As the samara continues to fall the excessive force reduces the rate of descent. From the viewpoint of the falling observer this reduction amounts not only to a decrease in the velocity of the passing air but also to a shift in the net velocity vector so that it is more nearly parallel to the chord. The angle of attack is then wrong, being too shallow. The wing noses upward to stabilize.

This movement tilts the aerodynamic force back toward the rear of the wing. The tilt opposes the spin of the samara, and the spin begins to decrease. To the falling observer the decrease in spin amounts to a reduction in the speed of the air that moves parallel to the chord. Again the overall velocity vector of the passing air is reduced. Its direction is changed too; it is now more vertical. The angle of attack is now too steep, and so the samara noses down to reduce it. This shift brings the force back to a vertical plane through the span, and the reduction in the spin ceases.

The net result is a reduction in the rate of descent (because the initially overly large aerodynamic force slowed the descent) and a decrease in the spin (because of the backward tilt of the force during the first part of the adjustment). The overall velocity vector of the passing air is now lower and therefore so is the force. The vertical component of the force now equals the samara's weight. Thereafter the samara falls with constant speed and spin and with an appropriate angle of attack and an appropriate glide angle.

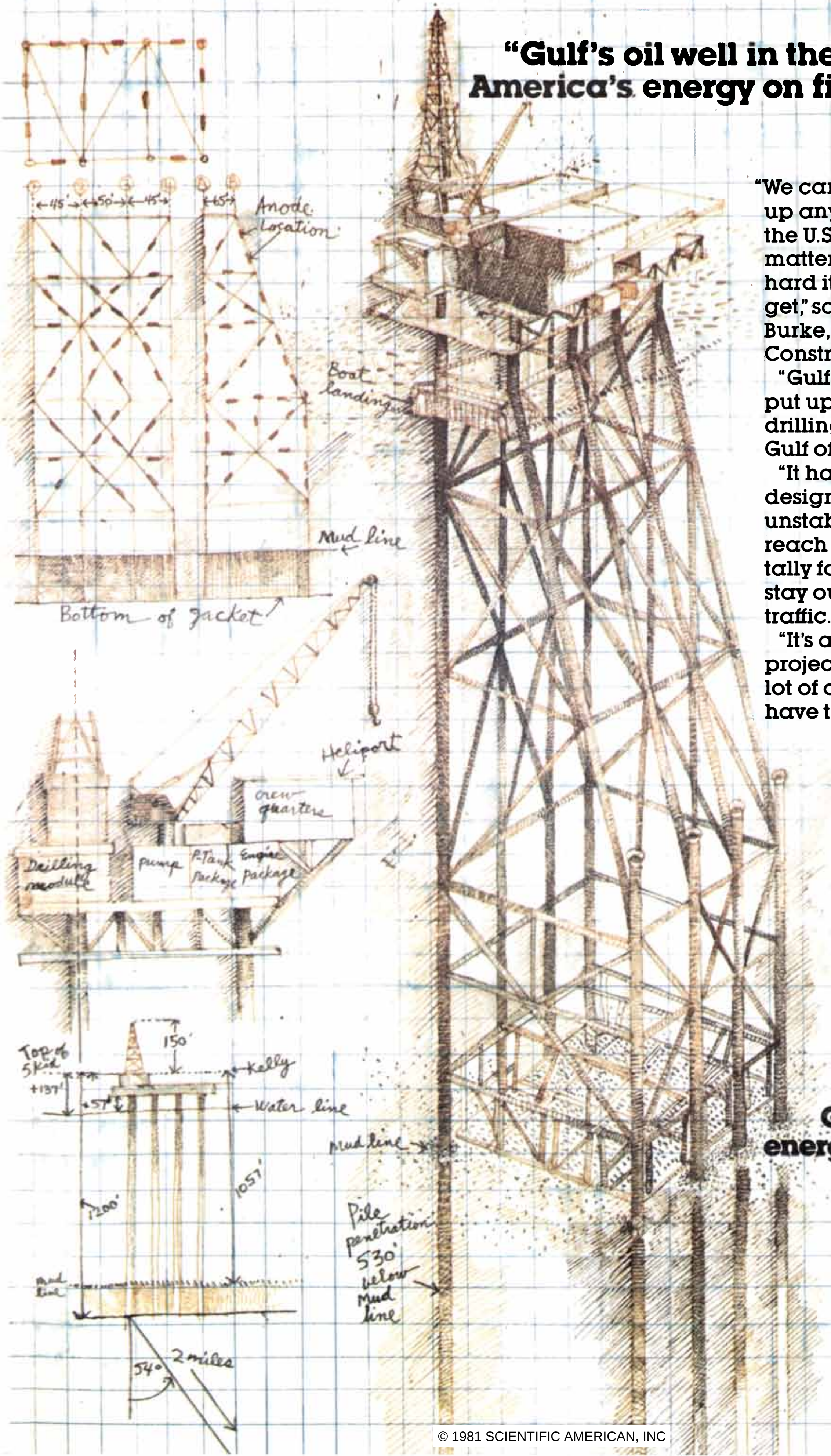
A chance perturbation of the glide angle is countered by similar responses. For example, if the glide path suddenly gets too steep, the net velocity vector of the passing air is too vertical. Since the angle of attack is then wrong, the wing noses down to correct it. The aerodynamic-force vector tilts forward. The spin rate increases and makes the net velocity vector more horizontal. Again the attack angle must be corrected; this time the wing must nose upward, moving the force vector back to the vertical plane. In principle the samara should now have the proper angle of attack and glide angle. It probably overshoots the proper glide angle, however, and must make several progressively smaller corrections in its spin, rate of descent and angles of orientation before it finally gains the proper values.

The coning angle too is automatically adjusted against small perturbations. The angle is set by a balancing of two kinds of torque. One torque results from the centrifugal force on the strips along



Vectors of the air acting on a samara

"Gulf's oil well in the mud will put America's energy on firmer ground."



"We can't pass up any oil in the U.S., no matter how hard it is to get," says Tom Burke, Chief Construction Engineer.

"Gulf Oil Corporation has put up a very special new drilling platform in the Gulf of Mexico.

"It has been especially designed to stand in deep, unstable mud, and to reach two miles horizontally for the oil, so as to stay out of the way of ship traffic.

"It's a big, challenging project, but it will supply a lot of oil that we won't have to buy from overseas."

Gulf

Gulf people: energy for tomorrow.

Hood sails. Rolex watches. Ultimates each.

For over twenty years, Frederick E. "Ted" Hood has designed and manufactured the most desirable, durable and efficient sails ever to drive the competitive yachts of the oceans of the world.

He's raced and won with the best. His list of prestigious victories include the Mallory Trophy in '56, the Bermuda Race in '68, and the SORC and the America's Cup in '74.

He knows his sails and he knows his sailing. His philosophy contains no room for anything less than maximum application of ability.

Durability. Excellence. Exactness. The words fit the man...the way he works. With ultimate care. With attention to detail. With the pride that comes from making the best that can be made.

The result is that every skipper worth his salt wants a suit of Hoods on his mast. And, chances are, he wants a Rolex on his wrist. Because uncompromising performance counts.

Rolex. The totally uncompromised machine. Handcrafted step by step from a solid block of gold or stainless steel. Mechanically unequaled. Esthetically unparalleled. The world's idea of what a watch should be.

Indeed, Mr. Hood and Rolex have precisely the same opinion toward accomplishment. The only true measure of achievement is result.

ROLEX

Pictured: The Rolex Submariner Date Chronometer in stainless steel with matching Fliplock bracelet. Pressure-proof to 660 feet. Also available in 18 kt. yellow gold.

Write for brochure. Rolex Watch, U.S.A., Inc., Dept.403, Rolex Building, 665 Fifth Avenue, New York, New York 10022. World headquarters in Geneva. Other offices in Canada and major countries around the world.

the span of the wing. Consider a strip about midway along the span. The centrifugal force is outward from the axis about which the samara is rotating. The torque generated by this force tends to rotate the wing about its overall center of mass to decrease the coning angle. The same force affects all the strips in the wing.

The wing does not become horizontal because another torque, arising from the aerodynamic force on each strip, tends to increase the coning angle. At one optimum value of coning angle the two opposing sets of torques balance each other. If the wing is deflected into a different coning angle, these torques bring it back to the proper angle.

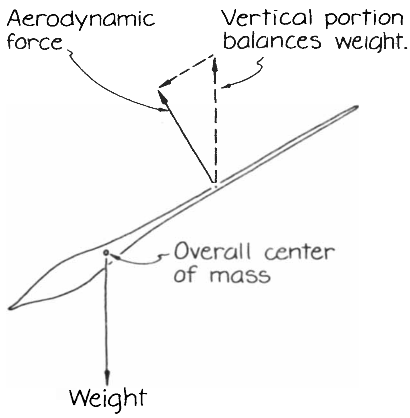
In general the coning angle should be shallow, so that the wing is more nearly horizontal. Then the disk loading is lower (since the wing sweeps out a larger disk) and the sinking speed of the samara is also lower. I have two maple samaras that are approximately the same in weight and overall appearance but descend at different coning angles. The one with the shallower coning angle falls slower.

Finally, a stability is related to the direction of travel of the samara. Most of my samaras spin while falling along a vertical axis. A falling observer could describe a plane traced out by the outer tip of the wing. The majority of my samples trace out a horizontal plane. If I blow against a samara to deflect it momentarily from the horizontal, it quickly adjusts its motion to bring the low side upward so that the plane is again horizontal.

Some of my prize samaras do something else. Even when I do not disturb their descent, they move in a large helix while still spinning in the normal way. Norberg calls this additional motion sideslip. The aerodynamic force on one of the peculiar samaras causes it to tilt the plane traced by the outer wing tip. Once the plane has been tilted it is forced not back up, as with most samaras, but rather to the side, so that the samara's center of mass follows a helical path to the ground. The sense of rotation of the center of mass in this path is opposite to the sense of spin of the samara about its axis of rotation. If from an overhead viewpoint the samara spins clockwise, the center of mass moves around in a helix in a counterclockwise direction. The helical motion is much slower than the spin.

If a maple samara is dropped with its center of mass downward and its flat wing vertical, it may not spin at all. If the wing is initially tilted from the vertical or is indented, it will begin to spin after a short drop. Consider a flat wing tilted from the vertical. As the fall begins, the aerodynamic force causes the wing to lag behind the descent of the center of mass, with the result that the wing rotates toward the vertical over the center of mass. Continuously adjusting its angles of attack, gliding and coning, the wing eventually reaches the equilibrium values. Thereafter the samara spins as it descends. When I toss a samara into the air with the center of mass leading, it almost always starts to spin once it stops moving upward and begins to fall. It is then unlikely to be exactly vertical, and spin is almost certain to begin.

In a good samara the distance of the center of mass from one end is between zero and 30 percent of the full length of the span. The distance of the center of mass of maple samaras is usually between 10 and 20 percent. The center of mass along the chord of a strip on the wing must be at a distance from the leading edge that is approximately 27 to 35 percent of the chord's length. The disk loading must be low, and so a thin wing that sweeps out a large disk is a desirable combination of properties. Most of the aerodynamic force is generated at the outer end of the wing because the relative velocity of the passing air is larger there. Since the force also depends on the area of a strip on the wing, the strips at the outer end of the wing should have the longest chords.

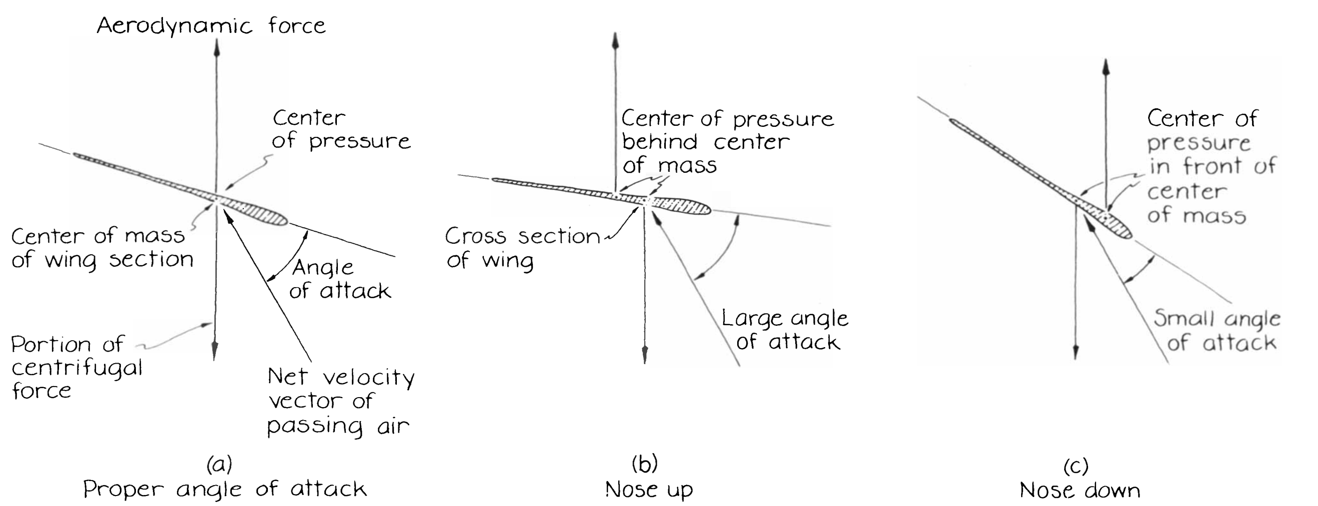


The balance of vertical forces

An astute observer might disagree with this last because the outer tips of maple samaras are somewhat tapered and flared toward the trailing edge. This natural design takes into account the vortexes shed by the outer tip as the samara spins about the vertical. The vortexes decrease the lift and increase the drag on the wing tip. The taper and the flare at the tip minimize the strength of the vortexes.

Most of my samples of maple samaras are several centimeters long in the span and about a centimeter wide in the chord. A few of them follow wide helical paths to the ground, but most fall with the center of mass moving along a vertical line. I cannot distinguish any physical difference between the two sets of samaras. The difference in flight pattern must be the result of subtle differences in the samaras.

The illustration at the lower right on page 235 summarizes my analysis of the mass distribution of a typical samara. To find the overall center of mass I



Angles of attack of a falling and spinning wing

A Boeing 727 is washed down after construction.

This glistening bird could use less fuel than ever before—thanks to a sleek aluminum skin developed by Alcoa.

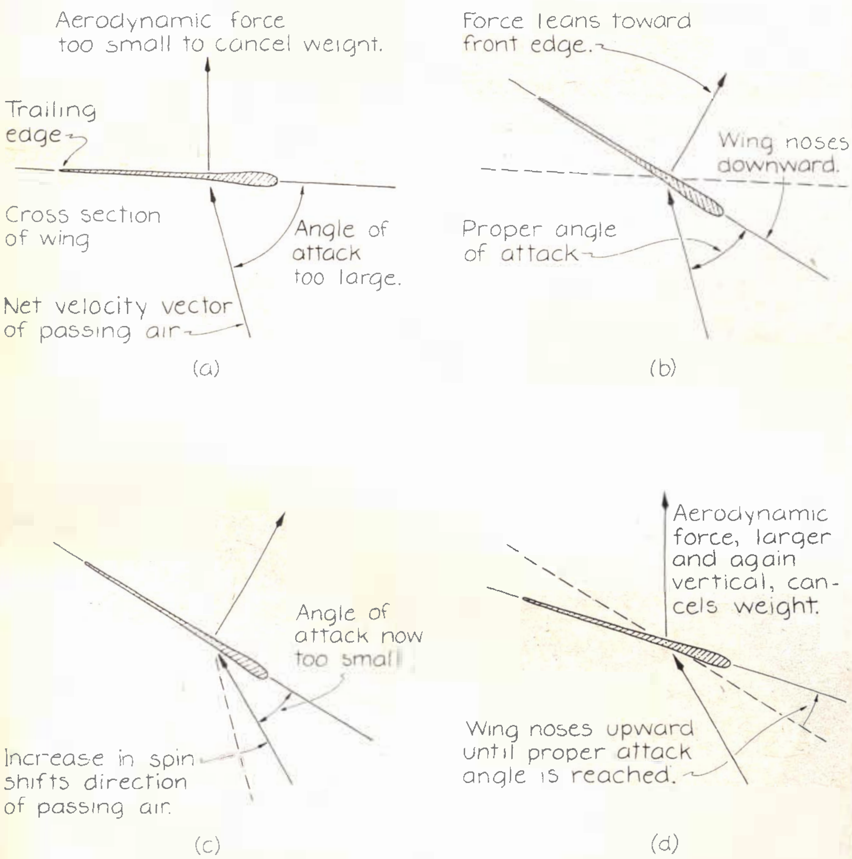
Alcoa's glistening bright-skin sheet can be left unpainted to increase fuel savings. And more airlines are doing just that. One has estimated that the additional savings in eliminating downtime for repainting their fleet of 130 727's is equivalent to having an extra plane. In addition, the new energy-efficient Boeing 767 will increase fuel savings even more, helped by lightweight, strong Alcoa® alloys, rolled or extruded in giant one-piece shapes. One more way Alcoa's unmatched production capability in aluminum is helping America conserve energy today. For more information, write Alcoa, 523-K Alcoa Building, Pittsburgh, PA 15219.

We can't wait for tomorrow. ALCOA

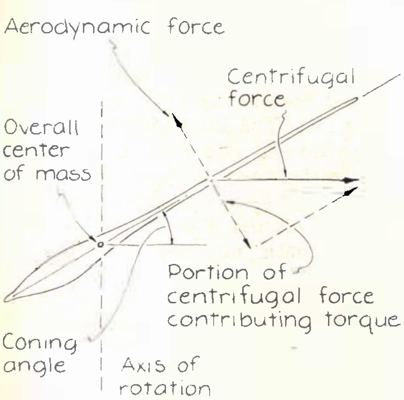
balanced the samara on a sharp edge, positioning it so that the span was perpendicular to the edge. When the samara was in balance, I drew a line on it just above the edge. Then I balanced the samara with a chord perpendicular to the edge and drew a line just above the edge. The intersection of the two lines marked the overall center of mass.

I was also interested in the distribution of mass along individual strips in the wing. I traced the outline of the samara on a sheet of graph paper. With a razor blade I sliced the wing into strips that ran parallel to chords through the wing. The position of each strip was indicated on the traced outline. To determine the approximate center of mass of a strip I balanced it on a sharp edge with the chord perpendicular to the edge.

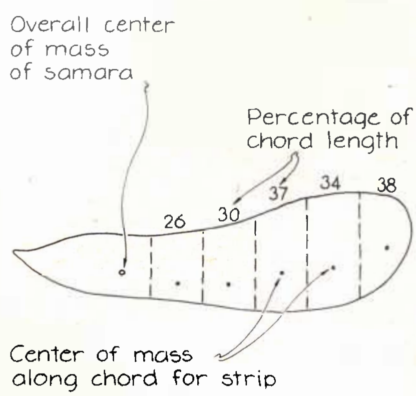
Having achieved a balance, I marked the position of the edge of the strip. Then I measured the distance between the mark and the leading edge of the strip. I also measured the average chord length of the strip. (Since the trailing edge of the samara curves, the strip had a small range of chord lengths.) My results are shown in the illustration of mass distribution. The distance between the center of mass of a strip and the leading edge of the wing is given as a



How a samara's flight pattern is adjusted



Components of the coning angle



Centers of mass of chord-length strips percentage of the average chord length of the strip.

If the samara were a flat wing, the distance of the center of mass along the chord of a strip should be between 27 and 35 percent of the chord length in order for the wing to be stable as it glides. The samara I analyzed had strips in which this distance was between 26 and 38 percent.

To measure the spin rate of a sample I directed a stroboscopic light downward on the path taken by a falling seed. With the room lights off I varied the flash rate of the stroboscope until each flash revealed approximately the same orientation of the descending samara. This effective freezing of motion can be accomplished by many settings of frequency on the stroboscope, including the frequency that matches the spin frequency of the samara. For example, a flash frequency of half the spin frequency reveals the same orientation on every other rotation of the samara. A flash frequency of a third the spin frequency reveals the same orientation on every third rotation.

To determine the spin frequency I begin at a low flash frequency and gradually increase the rate. The highest frequency that effectively freezes the motion of the samara is the one that matches the spin frequency. Any higher flash frequency does not freeze the motion. Instead I see images of the samara at various stages of each rotation. I found that most of my samples had a spin frequency of between 10 and 12 hertz (cycles per second). If I want to see the samara at, say, four stages during each rotation, I set the flash rate at about 40 hertz.

A stroboscope can be employed to make a photograph of a descending, spinning samara. With the room lights off the repetitive flashing of the stroboscope will illuminate the samara at various stages of its descent, leaving a permanent record of the stages on film. McCutchen made a similar photograph with a continuous light that shone upward into the path of a falling tulip-tree samara. That type of samara not only spins about the vertical axis along which its center of mass descends but also spins about the axis of its span. As the samara fell in the beam of light it exposed to the camera alternately an illuminated bottom surface and a darker top surface. The succession of separate images on the film recorded the fall. Contrast between the two sides can be obtained if one side is painted white.

I have discussed only the common maple samara. You might want to follow up McCutchen's work with other kinds of samara. Tulip-tree and ash samaras are particularly interesting. You could also experiment with cardboard models of samaras that you could modify in various ways to see how the changes affect the flight patterns.

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

MATHEMATICAL GAMES

EUCLID'S PARALLEL POSTULATE: ITS NATURE, VALIDITY, AND PLACE IN GEOMETRICAL SYSTEMS. John William Withers. The Open Court Publishing Company, 1905.

THE ELEMENTS OF NON-EUCLIDEAN GEOMETRY. D. M. Y. Sommerville. Dover Publications, Inc., 1958.

NON-EUCLIDEAN GEOMETRY. H. S. M. Coxeter. University of Toronto Press, 1965.

EUCLIDEAN AND NON-EUCLIDEAN GEOMETRY: DEVELOPMENT AND HISTORY. Marvin Jay Greenberg. W. H. Freeman and Company, 1974.

THE PATIENT RECORD IN EPIDEMIOLOGY

POPULATION STUDIES IN ROCHESTER AND OLMSTED COUNTY, MINNESOTA, 1900–1968. Leonard T. Kurland, Lila R. Elveback and Fred T. Nobrega in *The Community as an Epidemiologic Laboratory: A Casebook of Community Studies,* edited by Irving I. Kessler and Morton L. Levin. The Johns Hopkins Press, 1970.

25 YEARS OF NEUROEPIDEMIOLOGY IN THE AMERICAS: PROCEEDINGS OF THE FOURTH PAN-AMERICAN CONGRESS OF NEUROLOGY, OCTOBER, 1975. Edited by Leonard T. Kurland, 1975.

CONTRIBUTIONS TO NEUROLOGY FROM RECORDS LINKAGE IN OLMSTED COUNTY, MINNESOTA. Leonard T. Kurland and Dana D. Brian in *Advances in Neurology,* Vol. 19; 1978.

THE CONTRIBUTION OF THE MAYO CLINIC CENTRALIZED DIAGNOSTIC INDEX TO NEUROEPIDEMIOLOGY IN THE UNITED STATES. Leonard T. Kurland in *Clinical Neuroepidemiology,* edited by F. C. Rose. Pitman Medical Publishing Co., Ltd., 1980.

AN EARLY HABITAT OF LIFE

A NEW MICROFOSSIL ASSEMBLAGE FROM THE ARCHAEAN OF WESTERN AUSTRALIA. J. S. R. Dunlop, M. D. Muir, V. A. Milne and D. I. Groves in *Nature,* Vol. 274, No. 5672, pages 676–678; August 17, 1978.

SEDIMENTARY EVIDENCE FOR AN ARCHAEAN SHALLOW-WATER VOLCANIC-SEDIMENTARY FACIES, EASTERN PILBARA BLOCK, WESTERN AUSTRALIA. M. E. Barley, J. S. R. Dunlop, J. E. Glover and D. I. Groves in *Earth and Planetary Science Letters,* Vol. 43, No. 1, pages 74–84; April, 1979.

STROMATOLITES 3,000–3,500 MYR OLD FROM THE NORTH POLE AREA, WESTERN AUSTRALIA. M. R. Walter, R. Buick and J. S. R. Dunlop in *Nature,* Vol. 284, No. 5755, pages 443–445; April 3, 1980.

GRAVITATIONAL WAVES FROM AN ORBITING PULSAR

PULSARS. Richard N. Manchester and Joseph H. Taylor. W. H. Freeman and Company, 1977.

THE SEARCH FOR GRAVITY WAVES. P. C. W. Davies. Cambridge University Press, 1980.

NEUROPEPTIDES

PEPTIDES: INTEGRATORS OF CELL AND TISSUE FUNCTION. Edited by Floyd E. Bloom. Raven Press, 1980.

PEPTIDERGIC NEURONES. Tomas Hökfelt, Olle Johansson, Åke Ljungdahl, Jan M. Lundberg and Marianne Schultzberg in *Nature,* Vol. 284, No. 5756, pages 515–521; April 10, 1980.

BRAIN PEPTIDES AS NEUROTRANSMITTERS. Solomon H. Snyder in *Science,* Vol. 209, No. 4460, pages 976–983; August 29, 1980.

THE ACOUSTICS OF VIOLIN PLATES

REGARDING THE SOUND QUALITY OF VIOLINS AND A SCIENTIFIC BASIS FOR VIOLIN CONSTRUCTION. H. Meinel in *The Journal of the Acoustical Society of America,* Vol. 29, No. 7, pages 817–822; July, 1957.

THE VIOLIN AS A CIRCUIT. John C. Schelleng in *The Journal of the Acoustical Society of America,* Vol. 35, No. 3, pages 326–338; March, 1963.

THE HAZARDS OF WEATHER ON THE VIOLIN. R. E. Fryxell in *American String Teacher,* Vol. 15, No. 4, pages 26–28; Fall, 1965.

ACOUSTICAL EFFECTS OF VIOLIN VARNISH. John C. Schelleng in *The Journal of the Acoustical Society of America,* Vol. 44, No. 5, pages 1175–1183; November, 1968.

BENCHMARK PAPERS IN ACOUSTICS. MUSICAL ACOUSTICS: PART I, VIOLIN FAMILY COMPONENTS; PART II, VIOLIN FAMILY FUNCTIONS. Edited by C. M. Hutchins. Dowden, Hutchinson & Ross, Inc., 1975–76.

VITRUVIUS' ODOMETER

THE ORIGIN OF CLOCKWORK, PERPETUAL MOTION DEVICES, AND THE COMPASS. Derek de Solla Price in *Contributions from the Museum of History and Technology,* Paper 6; U.S. National Museum Bulletin No. 218; 1959.

ARCHIMEDES. Marshall Clagett in *Dictionary of Scientific Biography,* edited by Charles Coulston Gillispie. Charles Scribner's Sons, 1970.

THE MECHANICAL TECHNOLOGY OF GREEK AND ROMAN ANTIQUITY. A. G. Drachmann. Munksgaard, 1963.

THE UNKNOWN LEONARDO. Edited by Ladislao Reti. McGraw-Hill Book Company, 1974.

VITRUVIUS' WAYWISER. André W. Sleeswyk in *Archives internationales d'histoire des sciences,* Vol. 29, pages 11–22; 1979.

QUANTUM LOGIC

TWO DOGMAS OF EMPIRICISM. Willard Van Orman Quine in *From a Logical Point of View: Nine Logico-Philosophical Essays.* Harper & Row, Inc., 1963.

FOUNDATIONS OF QUANTUM MECHANICS. Josef M. Jauch. Addison-Wesley Publishing Co., 1968.

DEVIANT LOGIC: SOME PHILOSOPHICAL ISSUES. Susan Haack. Cambridge University Press, 1974.

THE PHILOSOPHY OF QUANTUM MECHANICS: THE INTERPRETATIONS OF QUANTUM MECHANICS IN HISTORICAL PERSPECTIVE. Max Jammer. John Wiley & Sons, Inc., 1974.

THE LOGICO-ALGEBRAIC APPROACH TO QUANTUM MECHANICS, VOL. I: HISTORICAL EVOLUTION. Edited by Clifford A. Hooker. D. Reidel Publishing Co., 1975.

CURRENT ISSUES IN QUANTUM LOGIC. Edited by E. Beltrametti and B. C. van Fraassen. Plenum Press, 1981.

IMAGE PROCESSING BY COMPUTER

SPECIAL ISSUE ON IMAGE PROCESSING. *Computer,* Vol. 7, No. 5; July, 1974.

DIGITAL IMAGE PROCESSING. R. C. Gonzalez and P. Wintz. Addison-Wesley Publishing Co., 1977.

DIGITAL IMAGE RESTORATION. H. C. Andrews and B. R. Hunt. Prentice-Hall, Inc., 1977.

DIGITAL IMAGE PROCESSING. William K. Pratt. John Wiley & Sons, Inc., 1978.

SPECIAL ISSUE ON IMAGE PROCESSING. *Proceedings of the IEEE,* Vol. 69, No. 5; May, 1981.

THE AMATEUR SCIENTIST

AUTOROTATION, SELF-STABILITY, AND STRUCTURE OF SINGLE-WINGED FRUITS AND SEEDS (SAMARAS) WITH COMPARATIVE REMARKS ON ANIMAL FLIGHT. R. Åke Norberg in *Biological Reviews of the Cambridge Philosophical Society,* Vol. 48, No. 4, pages 561–596; November, 1973.

THE SPINNING ROTATION OF ASH AND TULIP TREE SAMARAS. C. W. McCutchen in *Science,* Vol. 197, No. 4304, pages 691–692; August 12, 1977.

Before this RCA silicon image sensor, these little stars were just a twinkle in the astronomer's eye.

Stars that used to be no more than a dim blur in the telescope are now having their pictures taken with astonishing clarity and brilliance by a tiny silicon chip produced by RCA engineers and scientists.

The chip, called a CCD or Charge-Coupled Device, is built into a camera attached to a telescope. It makes today's most powerful telescopes 5 to 10 times more powerful.

The CCD is capable of reading light from stars too dim to be captured by conventional photography. It transforms the faintest visual image into thousands of electrical signals—digital information that can be displayed on a TV screen or analyzed by computer.

CCD technology allows the astronomer to see what could never be seen before, giving us a clearer picture of our universe than we have ever had in the past.

This CCD image sensor is a product of RCA research—the kind of dedicated research that has been a tradition at RCA for over 60 years. Today, many years after Nipper started listening to "His Master's Voice," research and development of electronic products are funded at the highest level in our company's history. Our commitment is, as it has always been, to the advancement of technology—to the creation of products that expand the human horizon.

RCA

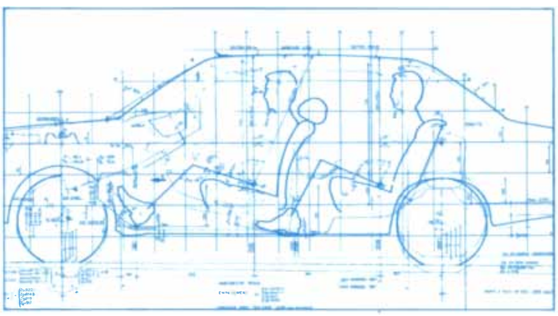
A tradition on the move

WHEN THERE IS A CAR FROM A COUNTRY THAT DESIGNED THE CONCORDE...

....a car designed by the first people to fly so far into the future.... a car engineered and crafted with a rare feeling for luxury...why drive an ordinary car? When you can drive the remarkable Renault 18i.

It begins with a design philosophy that has made Renault Europe's leading builder of cars: a fine automobile is not a collection of parts; it is a series of interdependent systems based on total system design.

Example: Renault designed the 18i independent front suspension as part of a system that includes chassis design, front-to-rear axle weight ratio and passenger seating. And even the passenger himself. Every component of that system—from bio-mechanically tuned seats (that are "infinitely" reclining), to long-travel suspension, to front and rear anti-sway bars and rack and pinion steering—all component matched for driver comfort, fatigue resistance and road control.



Example: The 18i total system design matches a lightweight front wheel drive system to a four-cylinder, 1.6 liter engine with pressure cast aluminum alloy block, hemispherical combustion chambers, pistons matched to within 7/100 of an ounce and a Bosch L-Jetronic fuel injection system. The result is surprisingly responsive performance with remarkable fuel efficiency.*

To complete the total system design, no one outdoes the French in their command of interior luxury. From Renault's optional nutmeg leather seating and full color-keyed carpeting to the tasteful design of the instrumentation.

37 HWY* EST | 26 EST MPG

*Compare these 1981 EPA estimates with estimated mpg for other cars. Your actual mileage depends on speed, trip length, and weather. Actual highway mileage will probably be lower.

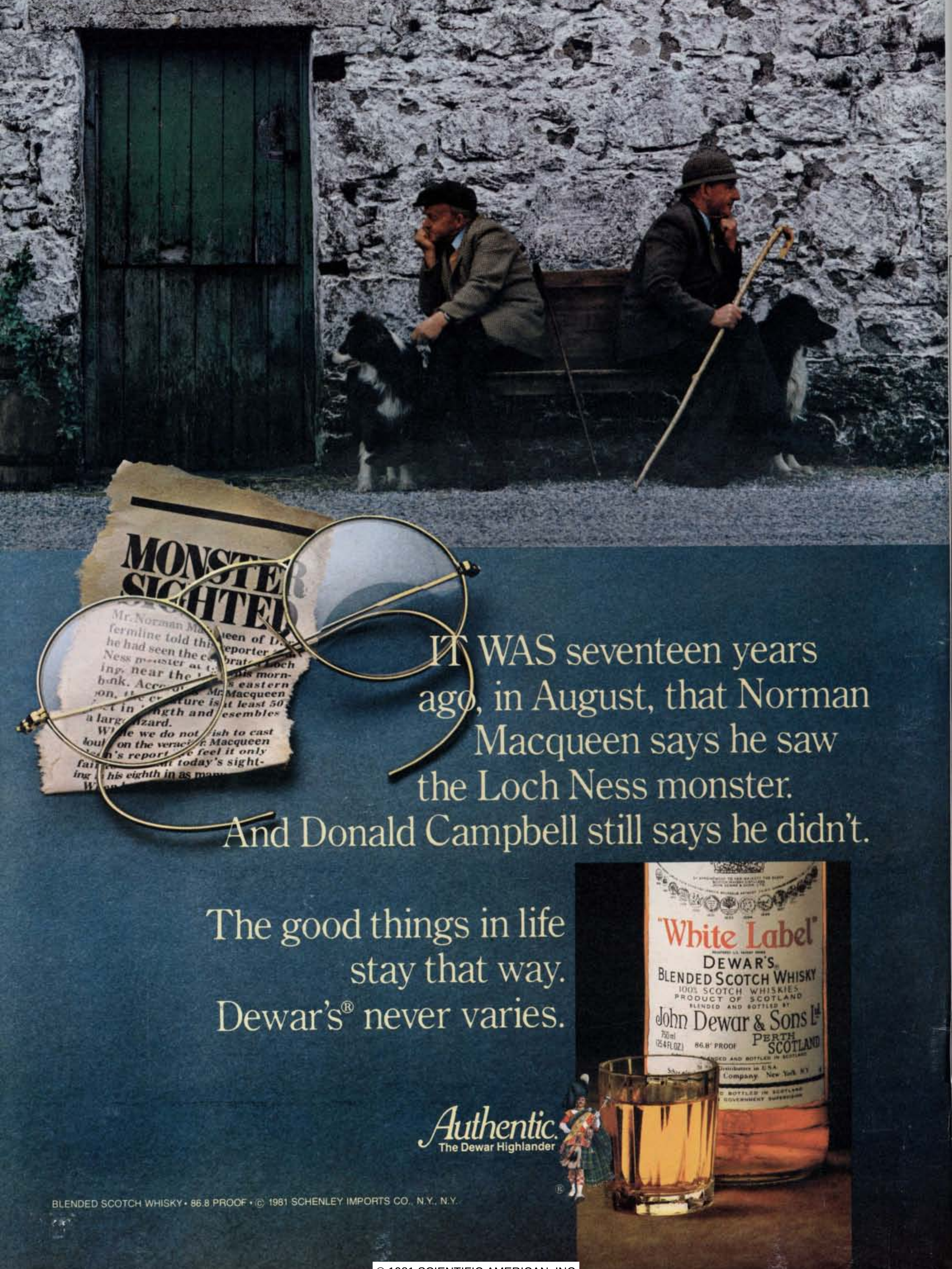
And now every Renault is covered by American Motors' exclusive Buyer Protection Plan®—the plan with the only full 12-month/12,000-mile warranty.

Knowing these facts, there's no excuse for you to drive an ordinary family car when you could drive the Renault 18i, the newest achievement from Europe's leader. Remarkable.

RENAULT

American Motors

Where great engineering lives in great design



IT WAS seventeen years ago, in August, that Norman Macqueen says he saw the Loch Ness monster.
And Donald Campbell still says he didn't.
The good things in life stay that way.
Dewar's® never varies.
"White Label"
DEWAR'S
BLENDED SCOTCH WHISKY
100% SCOTCH WHISKIES
PRODUCT OF SCOTLAND
BLENDED AND BOTTLED BY
John Dewar & Sons Ltd
PERTH SCOTLAND
86.8° PROOF
Authentic.
The Dewar Highlander
BLENDED SCOTCH WHISKY • 86.8 PROOF • © 1981 SCHENLEY IMPORTS CO., N.Y., N.Y.