SCIENTIFIC AMERICAN

THEORY OF KNOTS

$1.50

June 1979


turbo
SAAB
SAAB
turbo

Nothing performs like a Saab.

turbo The world is filled with cars that seem to be stamped out by a cookie cutter. But there's still a chance to own a thoroughly unique automobile. The new Saab 900 Turbo. Something about the Turbo makes people who drive it buy it.

The performance car perfected.

At the heart of this excitement is the Saab turbo engine. Saab engineers took the kind of power that dominates the big race tracks like Indianapolis and LeMans and harnessed it to work at everyday driving speeds. They did this by designing a turbocharging system that puts out more torque at lower engine speeds. So when you have to pass, or merge into fast moving traffic, there's an incredible surge of power at your disposal.

A feel for the road.

In addition to having extraordinary power, the new Saab Turbo is uniquely equipped to control it. A longer wheelbase, new suspension geometry and front wheel drive give the car cat-like agility – even on rain, snow and ice. Bilstein gas shocks and rack and pinion steering help you cling to the road. Power assisted 4-wheel disc brakes help the low profile Pirelli P6 tires come to smooth, sure stops.

Sports car excitement/Comfort for 5.

While the new Saab Turbo will give you and four passengers the driving thrill of a lifetime, the interior environment is tasteful, restrained and superbly organized. The relationship between the steering wheel, windshield and the hood provide an ideal field of vision. The driver faces a handsome new instrument panel where all controls are located in specialized zones for optimum convenience.

Upholstery is soft velour, front seats are fully reclining and the driver's seat is heated. A sliding sunroof made of steel is standard.

Engineered in Sweden, this is one of the world's most unique cars. For sheer comfort, excitement and driving confidence nothing performs like a Saab Turbo.

SAAB 900

You can never thank your father enough, but at least you can give him Chivas Regal.

12 YEARS OLD WORLDWIDE • BLENDED SCOTCH WHISKY • 86 PROOF • GENERAL WINE & SPIRITS CO., NEW YORK, N.Y.

Established 1845 **SCIENTIFIC AMERICAN** June 1979 Volume 240 Number 6

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $18 PER YEAR, U.S., ITS POSSESSIONS AND CANADA; $22 PER YEAR, ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.

King Carl XVI Gustav of Sweden—on the right in these photos—presents the 1978 Nobel Prize in Physics to Bell Laboratories scientists Robert Wilson (top photo) and Arno Penzias.

What does the Nobel Prize have to do with your telephone?

The two scientists on the opposite page are receiving the highest honor a scientist can earn—the Nobel Prize. They are the sixth and seventh laureates who did their prize-winning research at Bell Telephone Laboratories. These scientists shared a common goal—the search for new knowledge to further advance the art of telecommunications.

Clinton Davisson shared the Nobel Prize in 1937 for demonstrating the wave nature of matter. In 1956, John Bardeen, Walter Brattain and William Shockley were honored for their invention of the transistor. Philip Anderson's theoretical work on amorphous materials (such as glass) and on magnetism led to a Nobel Prize in 1977. And in 1978, Arno Penzias and Robert Wilson received the Prize for detecting the faint radiation from the "big bang" explosion that gave birth to the universe some 18 billion years ago.

The search for knowledge

These scientists and their colleagues at Bell Labs, given the freedom to explore, have proved time and again the value of investment in research—not only for telecommunications but for society in general. The transistor, for example, revolutionized communications and brought into being entire new industries—indeed, a new industrial society—based on solid-state electronics.

Other Bell Labs advances—products of this same research environment—have included high-fidelity recording, sound motion pictures, long-distance television transmission in the United States, the electrical digital computer, information theory, the silicon solar cell, and the laser. The impact of this work—on almost every field of commerce, industry, education and even medicine—has been incalculable.

The innovation process

Research done at Bell Labs in the past is the basis for the products and services the Bell System offers its customers today, just as the research going on now is the foundation for tomorrow's telecommunications.

Bell Labs scientists—specialists in physics, chemistry, mathematics and many other disciplines—team their efforts with those of our systems, development and design engineers. They, in turn, work closely with Western Electric manufacturing engineers and with the people of the Bell System operating telephone companies.

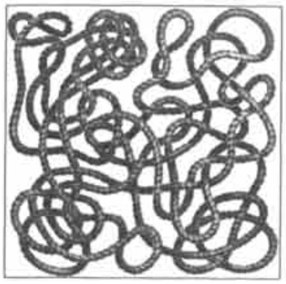
This technical integration is the foundation for true innovation. One idea feeds another. A basic scientific discovery can make possible entire new technologies and products for telecommunications, and a concept for a new product or system can stimulate the research to find even more new knowledge. That interaction, that teamwork, has been extremely productive: Bell Labs people have received 18,645 patents between our founding in 1925 and the end of 1978.

Sometimes, the search for knowledge may lead to a Nobel Prize. Often, it benefits all of society. And always, its ultimate aim is better service for Bell System customers.

Bell Laboratories
600 Mountain Avenue
Murray Hill, N.J. 07974

Bell Laboratories

Keeping your communications system the best in the world.



THE COVER

The picture on the cover shows a piece of twisted, interlaced rope whose ends have been joined so that it forms a physical model of the abstract mathematical object called a knot. In mathematics a knot is defined as a one-dimensional curve traced in three-dimensional space in such a way that it begins and ends at the same point and does not intersect itself (see "The Theory of Knots," by Lee Neuwirth, page 110). Knot theory, which seeks to distinguish differently knotted curves, is a part of topology, the branch of geometry concerned with the properties of spaces that remain unchanged when the spaces are deformed. For example, the model shown on the cover can be deformed into a simple loop. In other words, this apparently intricately knotted figure is in fact unknotted.

THE ILLUSTRATIONS

Cover illustration by George V. Kelvin

Page	Source
20–34	Ilil Arbel
42	National Aeronautics and Space Administration
46	Ralph Morse, courtesy of Princeton University Library
48	L. E. Roth and Y. Shigenaka
58–61	Ilil Arbel
62	R. G. Edwards, University of Cambridge
63	R. G. Edwards, University of Cambridge (*top*); Davis Division, Department of Embryology, Carnegie Institution of Washington (*middle*); Ilil Arbel (*bottom*)
64	Ilil Arbel
65	Davis Division, Department of Embryology, Carnegie Institution of Washington
66	Ilil Arbel
67	Dominick P. Purpura, Albert Einstein College of Medicine, New York
69	J. S. Chen, Laboratory of Cell Biology, Rome
70	Rita Levi-Montalcini, Laboratory of Cell Biology, Rome
71	Carol Donner
72	Rita Levi-Montalcini, Laboratory of Cell Biology, Rome
73	Albert Miller
74	Albert Miller (*top*); Rita Levi-Montalcini, Laboratory of Cell Biology, Rome (*bottom*)
75–76	Carol Donner
77	Luigi Aloe, Laboratory of Cell Biology, Rome
79	Intel Corporation
80–90	Gabor Kiss
111–124	George V. Kelvin
126–128	Tom Prentiss
129	Peter W. Bosch, Netherlands Geological Survey, Heerlen
130–132	Tom Prentiss
135	Institut Max von Laue–Paul Langevin
136–154	Allen Beechel
157–169	Andrew Tomko
170–171	Albert Miller
172–173	Andrew Tomko
174–177	Albert Miller
178	Ralph Morse, courtesy of Beinecke Rare Book and Manuscript Library, Yale University
180	Albert Miller
182	Library of Congress
184	Henry Stommel, Woods Hole Oceanographic Institution
188	Macbeth (*top*), Michael Goodman (*bottom*)
190–195	Michael Goodman
198	Michael Goodman (*top*), Ralph C. Eagle (*bottom*)

ROLEX

*The Rolex crown.
For the innovators
of our time.*

*Lamar Hunt.
Founder of the American
Football League.
Co-founder of World
Championship Tennis.
Member of the Pro
Football Hall Of Fame.
A dynamic sports
entrepreneur.*

*Rolex.
Dynamic innovator in
watchmaking.*

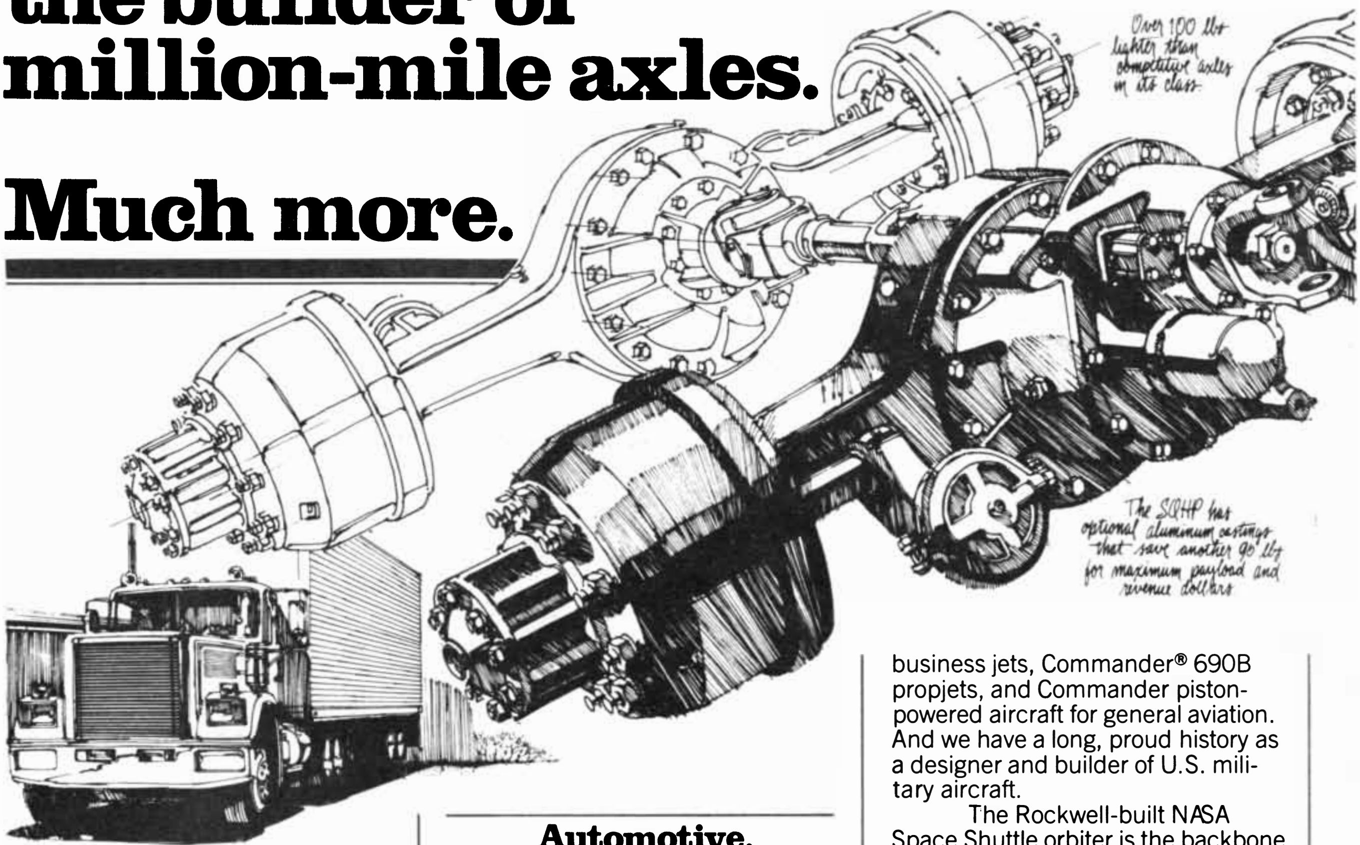
Mr. Hunt's Rolex is the Rolex Day-Date Oyster Perpetual Superlative Chronometer, self-winding, handcrafted from a solid block of 18kt. yellow gold, pressure proof to 165 feet, with matching hidden clasp President bracelet. Also available in 18kt. white gold and platinum.

For brochure, write Rolex Watch U.S.A., Inc., Dept. 116, Rolex Building, 665 Fifth Avenue, New York, New York 10022.

Geneva · Bombay · Brussels · Buenos Aires · Caracas · Cologne · Dublin · Hong Kong · Johannesburg · London · Madrid · Mexico City · Milan · New York · Paris · Sao Paulo · Singapore · Sydney · Tokyo · Toronto

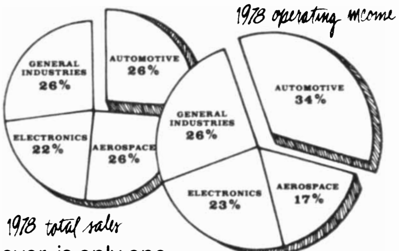
Rockwell is more than the builder of million-mile axles.

Much more.



You may have heard Rockwell International referred to as "the company that makes heavy-duty truck components." Our SQHP tandem truck axle is a prime example of our automotive capabilities; it's designed to go over a million miles in line-haul service, and half a million miles without a major overhaul. We can produce products like this because we operate the country's largest independent automotive components research facility.

Providing components for the land transportation industry, however, is only one of Rockwell's businesses. Rockwell International is a major multi-industry company, applying advanced technology to a wide range of products — in automotive, aerospace, electronics and general industries. Following are some examples of our balanced diversification.

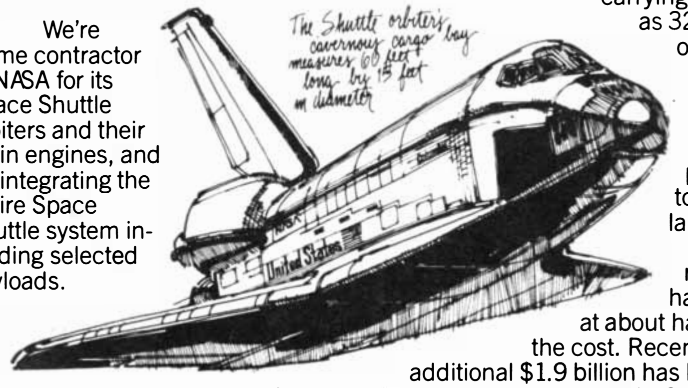


Automotive.

One half of the highway tandem tractors in North America are equipped with Rockwell axles — and more than half of the heavy-duty trucks stop with Rockwell brakes. We're also a major supplier of drivelines, steel and styled aluminum wheels, mechanical devices, castings, stainless steel wheel covers, reinforced plastic and other components for trucks, trailers, buses, vans and passenger cars.

Aerospace.

We're prime contractor to NASA for its Space Shuttle orbiters and their main engines, and for integrating the entire Space Shuttle system including selected payloads.



We also build rocket engines for other applications, and satellites such as Navstar. We manufacture Sabreliner® business jets, Commander® 690B propjets, and Commander piston-powered aircraft for general aviation. And we have a long, proud history as a designer and builder of U.S. military aircraft.

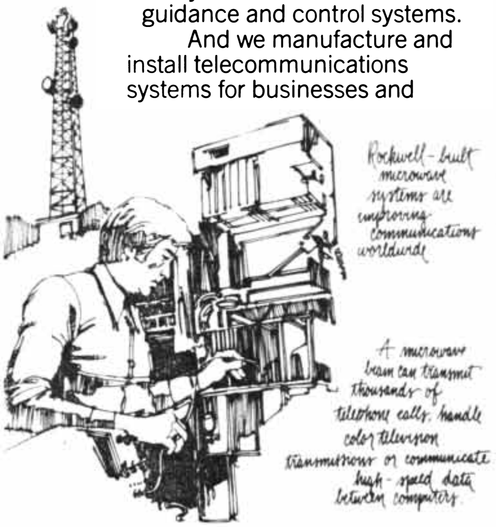
The Rockwell-built NASA Space Shuttle orbiter is the backbone of the world's first reusable space transportation system. Our Space Systems Group and Rocketdyne Division began work on it over eight years ago. Rockwell brings to this program the technology and systems management experience gained as a prime contractor to NASA for its Apollo program.

The Space Shuttle will launch like a rocket, haul like a truck, work like a space station, then return to Earth, landing like a glider. Each Shuttle is designed to make at least one hundred trips to space and back, carrying as much as 32½ tons of payload into orbit each time — twice the payload today's largest expendable rocket can handle — at about half the cost. Recently an additional $1.9 billion has been added to our existing NASA contract to complete the initial fleet of Shuttle orbiters.

Electronics.

We're one of the world's leading suppliers of avionics — communications, navigation and flight control equipment for air transport, general aviation, and government aircraft. We also make microelectronic systems and devices and guidance and control systems.

And we manufacture and install telecommunications systems for businesses and



governments worldwide.

As a matter of fact, we're one of the nation's largest independent suppliers of microwave systems for transmission of telephone calls, television pictures and computerized data.

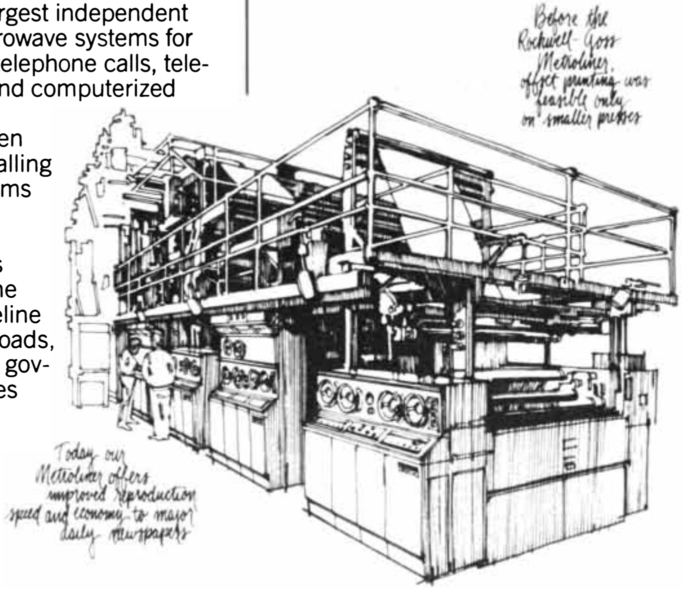
We've been making and installing microwave systems for over 25 years, and our many customers include telephone companies, pipeline companies, railroads, U.S. and foreign government agencies and the Public Broadcasting Service.

General industries.

Rockwell is one of the world's largest suppliers of high-technology valves for the energy market and for general industry. Our extensive technology is also being applied to the world's growing need for alternate sources of energy. We're involved in projects for nuclear energy, coal gasification, flue gas desulfurization, and solar, wind and geothermal power.

We also make printing presses, textile equipment, industrial sewing machines, power tools and products for utilities.

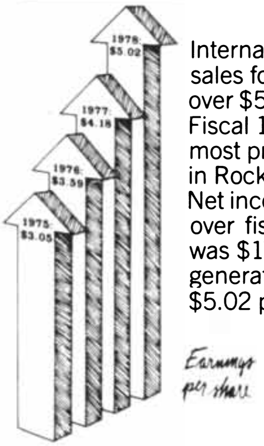
Speaking of our printing presses, most of America's big newspapers are printed on them. The list of major American dailies that are printed on Rockwell-Goss Metroliner presses alone reads like a Who's Who of newspapers. Another Rockwell-Goss press, the Signature HV, is designed to give smaller newspapers and commercial printers unprecedented speed and flexibility. Worldwide, Rockwell-Goss presses are serving publishers in more than 80 countries.



Over 14,300 scientists and engineers.

Of our 114,000 employees, one in eight is either a scientist or an engineer. They constitute about one percent of America's total scientific-engineering community. This technological base positions us for leadership in each of our product areas. It also makes our corporate slogan, ". . . where science gets down to business," a fact.

Business is good.



Rockwell International's total sales for 1978 were over $5.6 billion. Fiscal 1978 was the most profitable year in Rockwell's history. Net income, up 23% over fiscal 1977, was $176.6 million, generating a record $5.02 per share.

For more of the Rockwell story, please write us:
Rockwell International,
Department 815S,
600 Grant Street,
Pittsburgh, PA 15219.

Rockwell International

...where science gets down to business

LETTERS

Sirs:

"The Medical Influence of the Stethoscope," by Stanley Joel Reiser [SCIENTIFIC AMERICAN, February], describes one of the simplest and yet greatest tools invented for use in the practice of medicine. Laënnec's treatise published in 1819 was a brilliant work, a paradigm of clinical investigation for all who have followed.

I must take issue, however, with the author's general tone implying that the use of the stethoscope in clinical diagnosis was the beginning of the depersonalization of medicine, that physicians have become less concerned with the importance of the physical examination and that the doctor has lost his self-reliance thereby. These statements are simply not true.

Certainly there is far more specialization in medicine now than there was 50 or 100 years ago, a trend neither all good nor all bad. Specialization and growth of new knowledge have gone hand in hand. Each specialty has developed certain components of the physical examination of the patient to a high art as clues are sought at the bedside to guide the diagnostic hypotheses and the selection of laboratory tests and treatment.

Furthermore, medical educators are well aware of the importance of physical diagnosis and the bedside skills of clinical medicine in the curriculum applied to the education of budding young physicians. Chapters on percussion and auscultation occupy a significant proportion of standard textbooks on physical diagnosis. Considerable time in lectures, demonstrations and experience with patients is devoted to teaching these techniques in medical school. I am not aware of any medical school in this country that fails to emphasize these basic skills.

Some 50 percent of the current U.S. medical-school graduates are embarking on initial training for careers in family practice, internal medicine and pediatrics. These specialty areas are termed "primary care" because they offer a physician of first contact and the possibility of continuous care for the vast majority of chronic and acute medical problems. Throughout the undergraduate (pre-M.D.) and postgraduate (house officer) years the education and training of specialists in primary care emphasize the skills of history-taking and physical diagnosis, including particularly the use of the stethoscope. Indeed, lack of such emphasis would be deemed an unequivocal demonstration of an inadequate educational program. It is not possible to perform a complete physical examination, the first step in delivering medical care, without the intense application of the techniques of percussion and auscultation with the use of the stethoscope.

I fear that Dr. Reiser has attempted to demolish his own straw man. The facts are that a demonstrated ability to obtain a health history and perform a physical examination including the use of the stethoscope is an essential part of undergraduate medical education. These skills are a major asset of those physicians who devote themselves to primary care. In my view, the greatest value of Laënnec's invention is that the stethoscope has brought the doctor closer to the patient, not farther away.

THOMAS KILLIP, M.D.

Northwestern University Medical School
Evanston, Ill.

Sirs:

There is ample evidence of a declining skill in using the stethoscope and other techniques of physical examination. This trend can be traced, for example, through observations of some of the most eminent bedside clinicians of the 20th century. In 1931 James B. Herrick, professor of medicine at Chicago, complained that medical teachers are leading undergraduates "to place the emphasis on the instrument of precision rather than on the eye, ear, hand of the physical examiner." In 1945 a debate at the Royal Society of Medicine in London concluded with the general agreement that the stethoscope has lost its preeminence in the diagnosis of chest diseases, and it records one of the debaters asking why doctors should trouble themselves with "vague mysterious sounds, when something quite definite could be seen on the X-ray film." In 1952 Samuel A. Levine, professor of medicine at Harvard, wrote his essay "A Plea for the Stethoscope," in which he urged colleagues to reconsider their growing disregard of the instrument. In 1976 George L. Engel, professor of medicine at Rochester, reported that discussions with students in more than 50 North American medical schools visited in the preceding 15 years had revealed to him that "with a few notable exceptions relatively little time is devoted in the average undergraduate curriculum to supervised instruction in interviewing and physical examination."

The teaching of patient interviewing (as Engel also notes) suffers the same benign neglect as physical examination. A statement by the cardiologist Paul Dudley White summarizes the modern problem: "Too often the history of the patient is relegated to a relatively untrained person when it would be safer to turn over any other part of the examination." Taking a patient's history is principally learned in medical school from the personal experiences and random circumstances that are part of hospital life. Relatively few medical students receive systematic, formal, expertly supervised and continuing training in how the patient's sensations or thoughts can be effectively elicited and then interpreted as key medical data.

The relative neglect of physical examination and history-taking in modern medicine is not the result of a deliberate decision to ignore evidence that (and here I fully agree with Dr. Killip) is extremely valuable. Rather it is the result of a complicated historical process, more than three centuries old, during which the doctor's attention has been directed at different ways of viewing illness, a process that is hard to epitomize but that I document fully in my book *Medicine and the Reign of Technology* (Cambridge University Press, 1978). In this evolution the physician has changed his diagnostic emphasis from one centered on facts derived from the patient's history (18th century) to one concentrated on physical examination (19th century) to one stressing technology (20th century). This changing focus on different diagnostic techniques affects not only the character of the medical evidence doctors receive but also their view of themselves and their patients. The history inevitably draws the doctor into the subjective life of the patient, into the human drama of illness. Such person-centered data contrast with the impersonal evidence revealed by physical di-

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 40 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 ____________

____

____

NO RUM REFLECTS PUERTO RICO LIKE RONRICO.

From Puerto Rico, the Rum Island, comes Ronrico, *the* Puerto Rican rum. Smooth, light tasting Ronrico has been the pride of six generations of Puerto Rican rum masters (since 1860). One sip and you'll agree... rum lovers never had it so good.

RONRICO: AUTHENTIC RUM OF PUERTO RICO.

General Wine & Spirits Co., N.Y.C. 80 proof agnosis (such as chest sounds heard through a stethoscope), on which the 19th-century doctor relied, and technologically generated data (such as laboratory tests), in which contemporary doctors place so much credence.

Twentieth-century medical literature reveals that excessive reliance on modern technology by physicians, based on a belief in the superiority of its objective determinations, erodes their capacity for making subjective clinical judgments and decreases their confidence in evidence gained from history-taking and physical examination. In 1944 Tinsley R. Harrison, a medical educator and the senior editor of a standard textbook of medicine, commented on "the present-day tendency toward a five-minute history followed by a five-day barrage of special tests in the hope that the diagnostic rabbit might emerge from the laboratory hat." By the 1970's the rate of laboratory testing was doubling about every five years, and at a growing cost. The 1976 laboratory-testing bill in the U.S. was about $12 billion. The spate of "routine" testing has just caused the Blue Cross and Blue Shield associations to announce that they plan to stop paying for the admission batteries of laboratory tests for nonsurgical patients going into the hospital, unless the doctor orders them specifically and on an individual basis. This policy, the associations estimate, could save as much as $200 million a year.

A problem for this medical era, and for preceding ones, has been to understand the process by which useful but older diagnostic techniques are imprudently disregarded in an enthusiasm for the new and to recognize that the one-sided dominance of a given mode of fact-finding can limit the doctor's perception of disease and cloud his medical judgment.

STANLEY JOEL REISER, M.D.

Harvard Medical School
Boston

Sirs:

I was very interested in the timely article "Ion Channels in the Nerve-Cell Membrane" [SCIENTIFIC AMERICAN, March], written by my friend Richard D. Keynes. But I am surprised that he ascribes the origin and the first application of the well-known and now widely used voltage-clamp concept to Alan L. Hodgkin and Andrew F. Huxley.

In 1941 I realized that the negative-resistance limb of the iron-wire nerve-analogue characteristic should be stable in a low-resistance circuit and persuaded James H. Bartlett, Jr., to use the concept in his work. I first used the idea on the squid giant axon at Woods Hole in 1947, changing the wiring of the equipment my colleague George Marmont had developed for the membrane current control. Those records showed all the characteristics that have become classical.

I told Hodgkin of my work in a letter of October 7, 1947, from which he quotes in his Physiological Society centennial address. He tells of coming to Chicago in the spring of 1948 for a tutorial in what was later dubbed the "voltage clamp." He also saw the results that he, Huxley and Bernhard Katz fully confirmed with the improved techniques I showed were needed. Then Hodgkin and Huxley went on to produce their epoch-making series of papers in 1952.

In 1962 Alexander von Muralt was chairman of a small symposium at Leiden. Before the second paper he said: "I've heard it said that the voltage clamp is an invention of the Devil. It is now my pleasure to present the Devil himself—"

K. S. COLE

San Diego, Calif.

Sirs:

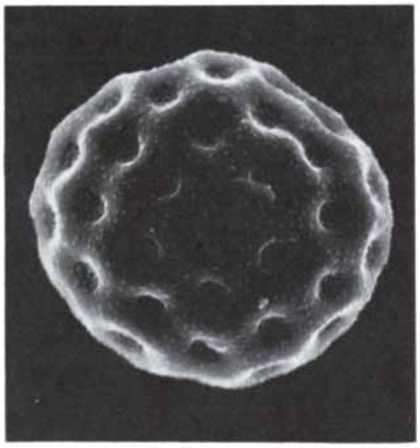
In "The Amateur Scientist" for April, Jearl Walker commented on the functionality of dimpling as it relates to the flight aerodynamics of golf balls: "The dimpled ball has less drag."

He further states that this phenomenon was serendipitously discovered by early golfers who noticed that rough, scarred balls traveled farther than the smooth golf balls in vogue at the time.

Enclosed is a scanning electron micrograph of a single pollen grain of sugar beet, *Beta vulgaris* L., at a magnification of 3,000 diameters. *Beta vulgaris* utilizes wind as a pollen vector. It is readily apparent that natural selection preceded man by more than a few million years in designing a sphere that would travel long distances.

MICHAEL BURGETT

Oregon State University
Corvallis



A pollen grain is dimpled like a golf ball

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JUNE, 1929: "Dr. Hubble's pioneer work on the distance of the Andromeda nebula is well known. By similar methods he now gives the distances of 24 nebulae. It appears that the large Magellanic cloud is 110,000 light-years off and the Andromeda nebula 840,000. The next nearest nebula is at 1,500,000 light-years. Then follow a dozen with distances varying from 1,600,000 to 4,500,000 light-years. The great cluster of nebulae in the constellation Virgo appears to be at 6,500,000 light-years. These data alone would make Dr. Hubble's results noteworthy, but his main point involves a comparison of the distances of the nebulae with their spectroscopically observed velocities. The remotest nebulae have the highest velocities, and all of them are receding from us. This suggests that the speed of a nebula is nearly proportional to its distance. If this 'distance effect' holds for still remoter nebulae, their spectra would be very strange. A nebula 250,000,000 light-years away could still be easily photographed as an object of the 17th magnitude, but its expected velocity would be nearly 40,000 kilometers per second. This would shift the sodium lines from the yellow part of the spectrum far into the red! The chance of photographing the spectrum of such a faint object with existing instruments appears to be very small, but velocities a good deal higher than those that have been recorded should be observable."

"At about the time this magazine reaches the reader there will be in operation in Jersey City, N.J., a high-powered broadcasting station dedicated to television. It is obvious that the erection of a large transmitter for television purposes would be useless unless there was a move afoot to supply the means of converting radio receivers designed for broadcast reception for use in television. In the Jersey City plant of the Jenkins Television Corporation production is going forward on a television unit, or 'radiovisor,' which is designed to be connected to a standard multi-tube receiver capable of tuning to this wavelength. Radio experimenters and others interested in television must be warned not to look for perfect results or large pictures that can be shown to a group of admiring friends. The pictures so far reproduced with the apparatus available for home use are small—in the neighborhood of 1 by 1½ inches—and often are not too clear. This, however, is no deterrent to progress. Pictures can be received, and it will not be long before they will be bigger and better."

"An electric room cooler, designed by the Frigidaire Corporation for use in homes and offices, has just been announced. In experimental tests this cooler lowered the temperature of an average-size room as much as 10 degrees in 40 minutes and brought about a 10 per cent reduction in humidity. When the cooler is in operation, the air passes into it through the bottom of the cabinet and out of it through louvers in the front panel. This air is passed over cooling coils, and a fan with a capacity of 450 cubic feet per minute distributes the cold air into the room. The room cooler stands approximately four feet high and weighs 210 pounds."

SCIENTIFIC AMERICAN

JUNE, 1879: "Six months ago popular attention was strongly drawn to the development of the electric light, and a panic prevailed among the holders of gas stocks. That flurry has blown over. The electric light has not fulfilled its promises, and Mr. Edison's assertion that his latest lamp is a complete success falls on indifferent ears. The world is not so eager for the change as it appeared, and on all sides the disposition is to await developments patiently."

"At a meeting of the Royal Society Mr. William Crookes, F.R.S., exhibited a series of experiments illustrating the curious behavior of electrified molecules in extremely rare media. Mr. Crookes's experiments show that there exists in nature a fourth and higher condition of matter much more ethereal than the gaseous condition. With the improvements made in the Sprengel pump by Mr. C. H. Gimingham it is now possible to produce vacua in which the pressure is measured in millionths of an atmosphere. It is with vacua so produced that Mr. Crookes's investigations have been conducted. His experiments led to the discovery that the dark space around the negative pole within a vacuum tube is a field of molecular activity. In one apparatus Mr. Crookes illustrated in a very beautiful manner the four principal phenomena connected with molecular physics in high vacua: first, the green phosphorescent light of molecular impact; second, the projection of molecular shadows; third, the magnetic deflection of the trajectory of molecules, and fourth, the mechanical action of the gaseous molecules projected from the negative pole."

"An international canal congress for discussing projects for the construction of an interoceanic ship canal across the American isthmus met in Paris May 15. M. Ferdinand de Lesseps was fitly chosen president. Up to this writing six routes have been under examination and discussion, namely the Nicaragua route, the Panama route, the San Blas route, the Tiati-tolo route, the Tuyra-Caquirri-Atrato route and the Atrato-Napipi route. At first the Tiati-tolo route seemed to have the brightest prospects. The Sub-Committee on Tunnels, however, found that its probable cost had been greatly underrated and that under the most favorable conditions it would be $160,000,000. The choice seems to be narrowed to two projects, the Nicaragua route and the Panama route."

"It is generally conceded that the nearest approach yet made to a rational explanation of the formation of the earth is to be found in the bold hypothesis that was conceived by Kant and elaborated by Laplace. On this assumption the earth and all the other planetary bodies have resulted from the condensation of nebulae. Thousands of these faintly luminous cloud-like bodies have been detected in the heavens, and the spectroscope has shown that some of them contain glowing hydrogen. On the condensation of a nebula, by the attraction of its particles, great heat would necessarily be developed. Chemical forces would then come into play during the contraction, and such compounds would be formed as were capable of existing under the given conditions of temperature and pressure. On the increase of pressure by contraction, and on the reduction of heat by radiation, a liquid magma would eventually be formed. The present tendency among most men of science seems in the direction of a return to the old-fashioned views according to which the earth has a moderately thin crust that rests on a spheroid of molten matter in a more or less viscous condition."

"At the late meeting of the British Iron and Steel Institute in London the Bessemer Medal of the Institute was presented to the venerable Peter Cooper as 'the father of the iron trade in America.' In his presentation speech the President spoke of Mr. Cooper's half-century connection with the iron trade, his Baltimore rolling mill in 1830, his building and running the first American locomotive, his extensive iron works at Trenton, and particularly the founding and direction of the great Cooper Institute of New York. In view of the fact that it is through the efforts of Mr. Cooper and other leaders in the American iron trade that England's greatest rival in iron production has almost reached supremacy, this recognition of his labors by the English iron and steel producers is particularly handsome."

THE AUTHORS

CLIFFORD GROBSTEIN ("External Human Fertilization") is professor of biological science and public policy in the Program in Science, Technology and Public Affairs at the University of California at San Diego. He received his Ph.D. in zoology from the University of California at Los Angeles and went on to teach at Oregon State College. In 1946, after three years in the Air Force as an aviation physiologist, he joined the staff of the National Cancer Institute and did research in developmental biology for the next 10 years. He then joined the faculty of Stanford University. After nine years as professor of biology at Stanford Grobstein moved to San Diego, where from 1967 to 1973 he was dean of the School of Medicine.

RITA LEVI-MONTALCINI and PIETRO CALISSANO ("The Nerve-Growth Factor") are neurobiologists at the Laboratory of Cell Biology of the National Research Council of Italy in Rome. Levi-Montalcini received her M.D. from the University of Turin in 1936. With the invasion of Italy by the Germans in 1943, she fled to Florence and lived underground until the liberation of Italy in 1945. She thereupon returned to Turin and became a research assistant in the department of anatomy under Guiseppe Levi. In 1947 Viktor Hamburger, chairman of the department of zoology at Washington University in St. Louis, invited her to spend a year in his laboratory, and soon afterward he offered her a permanent staff position. In 1961 she was invited back to Italy to direct a small neurobiology research unit in Rome, and in 1969 she was offered the directorship of the newly founded Laboratory of Cell Biology. Calissano is head of the research group in neurobiology at the Laboratory of Cell Biology. He was graduated from the medical school of the University of Genoa in 1964 and began to do research in enzymology. In 1966 he joined Levi-Montalcini's group at Washington University. On his return to Italy in 1968 he became a permanent member of the Neurobiological Unit in Rome, moving to the Laboratory of Cell Biology in 1974.

STEPHEN KAHNE, IRVING LEFKOWITZ and CHARLES ROSE ("Automatic Control by Distributed Intelligence") are systems engineers at the Case Institute of Technology of Case Western Reserve University. Kahne is professor and chairman of the systems-engineering department. He was educated at Cornell University and the University of Illinois, earning his Ph.D. in electrical engineering in 1963. After three years in the Air Force he spent 10 years on the faculty of the University of Minnesota and was consulting partner and director of InterDesign, Inc., a Minneapolis design firm. Kahne moved to Case in 1976. Lefkowitz is professor of systems engineering. He was graduated from the Cooper Union School of Engineering and worked for 10 years in industry. He then went to Case to obtain his Ph.D. in control engineering. Since 1957 Lefkowitz has directed Case's research program in the control of complex systems. Rose is associate professor of computer engineering. He received his bachelor's and master's degrees in electrical engineering at Duke University in 1963. After four years as a senior engineer at the Texas Instruments Corporation he did his doctoral work at Case and received his Ph.D. in computing and information science in 1970.

LEE NEUWIRTH ("The Theory of Knots") is director of the Communications Research Division of the Institute for Defense Analyses in Princeton, N.J. He received his B.S.E. in chemical engineering at Princeton University, his M.A. in applied mathematics at Columbia University and his Ph.D. in mathematics from Princeton in 1959. After a postdoctoral fellowship at the University of California at Berkeley he joined the Institute for Defense Analyses in 1961. He was appointed deputy director in 1965 and director in 1977. In his spare time Neuwirth plays squash and jogs about 40 miles a week; he completed his first (and last) marathon in 1977.

PETER W. BOSCH ("A Neolithic Flint Mine") is a cartographer in the department of geological mapping at the Dutch Geological Survey in Heerlen, the Netherlands. He is responsible for mapping the province of Limburg, in the southernmost part of the country, and has specialized in the geological history of the Maas (Meuse) River. For nine years he worked as one of the volunteers who excavated the prehistoric flint mines in Rijckholt, which are described in his article.

R. GOLUB, W. MAMPE, J. M. PENDLEBURY and P. AGERON ("Ultracold Neutrons") are nuclear physicists who have worked together for several years on the development and installation of ultracold neutron facilities at the Institut Laue-Langevin (ILL) in Grenoble, France. Golub received his Ph.D. from the Massachusetts Institute of Technology in 1967. After two years on the faculty of Brandeis University he went to the University of Sussex, where he has remained as a "temporary" research fellow. Mampe received his Ph.D. in physics in 1971 from the Tech-

A LUXURY SEDAN BASED ON THE BELIEF THAT ALL OF THE RICH ARE NOT IDLE.

In every age, there have always been those among us whose primary activity is inactivity.

And for those people—members of what has been variously dubbed the "idle rich" and the "leisure class"—there has always been an abundance of suitably sedentary luxury sedans.

Cars designed primarily for sitting. Cars that reach their performance peak preening in the country club driveway.

The BMW 528i, on the other hand, is a luxury sedan built to satisfy an opposite mentality. People who believe that activity—mental as well as physical—is what makes living worthwhile.

THE BMW 528i. A LUXURY SEDAN FOR SERIOUS DRIVERS.

At the Bavarian Motor Works, it is our contention that the things that seem to preoccupy the luxury car makers of the world have little to do with the one thing that makes an expensive car worth the money: extraordinary performance.

First and foremost, the engineers at BMW are racing engineers by nature and by profession.

So, while the BMW 528i provides all the creature comforts one could rationally require in any expensive automobile, it also provides a driving experience so unusual, so exhilarating it will spoil you for any other car.

Its six-cylinder engine has been called "...the most refined in-line six in the world."

Its suspension system—independent on all four wheels—provides the driver with a tactile sense of control, an uncanny feel of the road that is positively unique to BMW.

Its interior is biomechanically engineered to the nth degree.

The front seats are shaped to hold their occupants firmly in place, and so thoroughly adjustable that it is all but mathematically impossible to not find a proper and comfortable seating position.

All controls are within easy reach. The tachometer, speedometer and all ancillary instruments are large, well marked and totally visible—set in a dashboard designed in the manner of an airplane cockpit.

THE RARE OPPORTUNITY TO SHIFT FOR YOURSELF.

To some, there is simply no alternative to the singular pleasure of a manual transmission.

For the automotive purist, shifting up through the gears, smoothly and precisely, is an experience that borders on the esthetic.

And it is particularly significant that—while a three-speed automatic transmission is available as an option—the BMW 528i offers a four-speed manual transmission as standard equipment.

If the thought of owning such a car intrigues you, call us anytime, toll-free, at 800-243-6000 (Conn. 1-800-882-6500) and we'll arrange a thorough test drive for you at your nearest BMW dealer.

THE ULTIMATE DRIVING MACHINE.

Bavarian Motor Works, Munich, Germany.

nical University of Munich and then went to ILL to do research in nuclear spectroscopy. Pendlebury is lecturer in experimental physics at the University of Sussex. He studied mathematics, physics and chemistry at the University of Cambridge, staying on to obtain his Ph.D. in atomic-beam spectroscopy in 1963. Ageron studied at the École Polytechnique in Paris. In 1960 he joined the French Atomic Energy Commission to work in the field of research reactors, in particular on the design of the high-flux reactor at ILL.

JOHN G. SCLATER and CHRISTOPHER TAPSCOTT ("The History of the Atlantic") are geophysicists interested in the past configurations of the earth's tectonic plates. Sclater is professor of earth and planetary sciences in the joint program in oceanography sponsored by the Massachusetts Institute of Technology and the Woods Hole Oceanographic Institution. He was graduated from the University of Edinburgh with a degree in physics and did graduate work in geophysics at the University of Cambridge, obtaining his Ph.D. in 1966. From then until 1972 he did research at the Scripps Institution of Oceanography in La Jolla, Calif. In 1972 he left Scripps to join the M.I.T. faculty; he was appointed full professor in 1977. Sclater is the 1979 recipient of the Rosenstiel Award in Oceanography. Tapscott is a research geophysicist with the Exxon Production Research Company in Houston, Tex. He was graduated from Swarthmore College in 1972 and began graduate study in nuclear physics at Princeton University. The following year he took a leave of absence to pursue an interest in marine geology and geophysics. In the fall of 1974 he entered the joint program in oceanography and did his doctoral research in Sclater's laboratory. Tapscott received his Ph.D. in March of this year.

HENRY STOMMEL and ELIZABETH STOMMEL ("The Year without a Summer") are husband and wife. Henry Stommel is senior scientist at the Woods Hole Oceanographic Institution. He obtained his B.S. from Yale University in 1942 and taught mathematics and astronomy there for the next two years. He then became a research associate in oceanography at Woods Hole, where he remained for the next 16 years. In 1960 he was appointed professor of oceanography at Harvard University; he moved to M.I.T. in 1963. Last year he returned to Woods Hole as a resident scientist. In recent years Henry Stommel has codirected the large U.S.S.R.-U.S. POLYMODE research project on ocean eddies. Elizabeth Stommel was graduated from Bennington College with a B.A. in English. Their article is based on part of a book they have written on the same subject, which is soon to be published.

Intecolor's 8110 vs. IBM's 5110. Draw your own conclusions.

	ISC 8110	IBM 5110
PRICE	$14,300	app. $18,000
COLOR	YES	NO
GRAPHICS	YES	NO
SCREEN	19″ diagonal 3,840 characters	5″ diagonal 1,024 characters
NUMBER OF LINES/CPL	48 lines x 80 CPL	16 lines x 64 CPL
LANGUAGE	Scientific BASIC w/16-digit precision	BASIC w/ 13-digit precision
LANGUAGE, OPT.	FORTRAN	APL

	ISC 8110	IBM 5110
USER MEMORY	8K bytes*	16K bytes
STORAGE	1,182K bytes Dual double-headed 8″ floppy disk drive	300K bytes
PRINTER	150 lpm impact matrix printer/ full graphics plotter	80 cps bi-directional matrix printer
KEYBOARD	with numeric & color pads	with numeric pad

* Expandable to 24K user memory.

The Intecolor 8110 is a stand-alone scientific microcomputer system with color graphics. It gives you more than three times the storage of the IBM 5110 computing system, more than four times the display area, plus a faster printer and plotter.

And with the Intecolor 8110, you also get the important advantages of color, graphics and Scientific BASIC. All for *$14,300, backed by a 6-month warranty.

Why pay thousands more for less efficiency? Contact your Intecolor representative today for a demonstration. **Color Communicates Better.**

*U.S. Domestic Price

(Furniture not included)

Unretouched photo of screen.

ISC SALES REPRESENTATIVES: **AL**: 205/883-8660, **AK**: (GA) 404/449-5961, **AZ**: 602/994-5400, **AR**: (TX) 214/840-2169, **CA**: Alhambra 213/281-2280, Goleta 805/964-8751, Irvine 714/557-4460, Los Angeles 213/476-1241, Mountain View 415/964-9300, San Diego 714/292-8525, **CO**: 303/759-0809, **CT**: (GA) 404/449-5961, **DE**: (GA) 404/449-5961, **DC**: (GA) 404/449-5961, **FL**: Orlando 305/425-5505, Ft. Lauderdale 305/776-4800, Melbourne 305/723-0766, Tallahassee 904/878-6642, **GA**: 404/455-1035, **HI**: 808/524-8633, **ID**: (UT) 801/973-7969, **IL**: (No.) 312/564-5440, (So.) (MO) 816/765-3337, **IN**: (IL) 312/564-5440, **IA**: (MO) 816/765-3337, **KS**: (MO) 816/765-3337, **KY**: (OH) 513/434-7500, **LA**: 504/626-9701, **ME**: (GA) 404/449-5461, **MD**: (GA) 404/449-5961, **MA**: (GA) 404/449-5961, **MI**: 313/588-2300, **MN**: 612/822-2119, **MS**: (AL) 205/883-8660, **MO**: 816/765-3337, **MT**: (CO) 303/759-0809, **NB**: (MO) 816/765-3337, **NH**: (GA) 404/449-5961, **NJ**: (GA) 404/449-5961, **NV**: (AZ) 602/994-5400, **NM**: 505/292-1212, **NY**: (GA) 404/449-5961, **NC**: 919/682-2383, **ND**: (MN) 612/822-2119, **OH**: Cleveland 216/267-0445, Dayton 513/434-7500, **OK**: (TX) 214/840-2169, **OR**: 503/620-5800, **PA**: (East) (GA) 404/449-5961, (West) 412/892-2953, **RI**: (GA) 404/449-5961, **SC**: 803/798-8070, **SD**: (MN) 612/822-2119, **TN**: 615/482-5761, **TX**: 214/840-2169, El Paso Area (Las Cruces, NM) 505/523-0601, Houston Only 713/780-2511, **UT**: 801/973-7969, **VT**: (GA) 404/449-5961, **VA**: (GA) 404/449-5961, **WA**: 206/455-9180, **WV**: (GA) 404/449-5961, **WI**: (IL) 312/564-5440, **WY**: (CO) 303/759-0809, **EUROPEAN EXPORT SALES**: **EUROPE**: (MA) 617/661-9424, **BELGIUM**: Brussels 02-242 36-04, **FRANCE**: Rueil Malmaison 749-40-37, **GREECE**: Athens 642-1368, **ITALY**: Roma 805-647/872-457, **THE NETHERLANDS**: Poeldijk 01749-7640, **SPAIN**: Barcelona 204 17 43, **SWEDEN**: Vallingby 08-380-370, **SWITZERLAND**: Mutschellen 057-54655, **UNITED KINGDOM**: Bournemouth 0202-293-115, **WEST GERMANY**: München 089-31881, **AUSTRALIA & NEW ZEALAND**: Melbourne 543-2077, Meadowbank 808-1444, Chermside 59-6436, Wellington 64-4585, Auckland 876-570, **CANADA**: Datamex, Ltd. (Distributor) Dorval 514/636-9774, Ottawa 613/224-1391, Toronto 416/787-1208, Vancouver 604/684-8625, **CENTRAL & SOUTH AMERICA & CARIBBEAN**: (GA) 404/394-9603, **MEXICO**: Monterrey 564-876, **FAR EAST**: (CA) 213/382-1107, **HONG KONG**: 5-742211, **JAPAN**: Tokyo 402-8596, **TAIWAN**: Taipei 02-7026284, **MIDDLE EAST**: **IRAN**: Tehran 891148, **ISRAEL**: Tel Aviv 266-291, **KUWAIT**: Kuwait 438 180/1/2, **LEBANON**: Beirut 221731 260110, **SAUDI ARABIA**: Jeddah 27790, Ryadh 25083-39732, **UNITED ARAB EMIRATES**: Sharjah 24068.
For sales and service in other countries contact ISC headquarters in Norcross, GA., U.S.A.

Intelligent Systems Corp. □ 5965 Peachtree Corners East □ Norcross, GA 30071 □ Telephone 404-449-5961 □ TWX 810-766-1581

MATHEMATICAL GAMES

Chess problems on a higher plane, including mirror images, rotations and the superqueen

by Martin Gardner

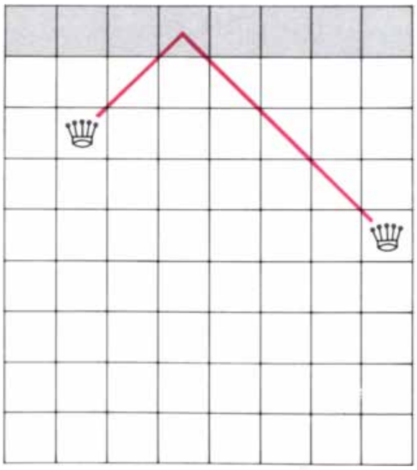
"It is a beautiful, complex and sterile art related to the ordinary form of the game only insofar as, say, the properties of a sphere are made use of both by a juggler in weaving a new act and by a tennis player in winning a tournament. Most chess players... are only mildly interested in these highly specialized, fanciful, stylish riddles."

—VLADIMIR NABOKOV, *Speak, Memory*

I last devoted an entire column to chess seven years ago (in May, 1972). This month the subject is again chess, in the form of some unusual chess puzzles I have not dealt with before. Here "chess puzzle" does not refer to the conventional type of chess problem that asks how a player can checkmate in a certain number of moves. The tasks to be considered are of a more mathematical nature, so that to work on them readers need only know the rules of the game.

I shall begin with a remarkable discovery made by David A. Klarner, a mathematician at the State University of New York at Binghamton. In the early 1960's pure-number theorists were working on a problem concerning any set of consecutive integers from 1 through $2n$, where n is a positive integer. Such a set can always be divided into two equal subsets, one made up of the consecutive integers 1 through n and the other made up of the consecutive integers $n + 1$ through $2n$. The problem asks whether for different values of n it is possible to pair each number in one subset with a number in the other so that the $2n$ sums and absolute, or unsigned, differences of the numbers in each pair are all distinct. For example, if $n = 1$, then the original set is $\{1, 2\}$, and its subsets are $\{1\}$ and $\{2\}$. In this case the only pairing provides a trivial solution, because the sum $1 + 2$, or 3, is distinct from the absolute value of the difference $1 - 2$, or 1.

Klarner's clever discovery was that this problem is isomorphic with a variation of the old chess problem of how to put n chess queens on a board of side n in such a way that no two queens attack each other. He modified the task by placing an extra row at the top of the chessboard. He calls the row a reflection strip, because, as is shown with a colored line in the illustration on this page, one queen can attack another by reflection in the strip. In other words, a queen is allowed to move diagonally into the strip and out again on the opposite diagonal to attack another queen. In Klarner's paper "The Problem of Reflecting Queens" (*The American Mathematical Monthly,* Vol. 74, No. 8, pages 953–955; October, 1967) he proved that the number-theory problem for a given n has a solution if and only if it is possible to place n queens on an n-by-n board with a reflection strip in such a way that no two queens attack each other either directly or by reflection.



Two queens attacking by reflection

The top illustration on page 25 shows how easily a solution of this variation on the queen-placement problem (in this instance for the case $n = 4$) can be transformed into a solution of the number problem. Beginning at the upper left on the board, number the rows consecutively from top to bottom and then continue the sequence from left to right below the columns, as is shown in the illustration. Note that the two solutions shown (the only ones for this case) are left-to-right mirror images. Now make a list of the row and column numbers, or coordinates, for each queen, as is shown below the boards in the illustration. Taking the sums and differences of the eight pairs of numbers obtained in this way, it is easy to verify that they provide a solution to the number-theory problem for the case $n = 4$.

It is not hard to understand why the two problems are isomorphic. It is obvious that the set of numbers listed around the outside of the chessboard is equal to the set of numbers 1 through $2n$ and that the row numbers for the queens in a solution form one subset of the set and the column numbers form another subset. If two queens did attack each other in the same row or column, then a number would be duplicated in one of the subsets. Since this cannot happen, the subsets must respectively be made up of the numbers 1 through n and the numbers $n + 1$ through $2n$. Moreover, if two queens attacked on a diagonal slanting up and to the right, one of the sums would be repeated. If two queens attacked on a diagonal slanting down and to the right, an absolute difference would be repeated. And if two queens attacked by reflection in the strip, a sum and an absolute difference would be equal. Since no two queens may attack each other directly or by reflection, however, the sums and absolute differences of the pairs of numbers derived from the chess solution must all be distinct, and so they provide a solution to the number-theory problem.

Solutions to the classic queens problem (with no reflection strip) on boards of up through order eight have long been known and are discussed in Chapter 16 of my book *The Unexpected Hanging and Other Mathematical Diversions* (Simon & Schuster, Inc., 1972). There are no solutions for orders two and three and, ignoring rotations and reflections, only one "basic" solution for order four, two basic solutions for order five, one for order six, six for order seven and 12 for order eight. To derive all the solutions for the corresponding number problem from one of these sets of solutions it is necessary to consider the four rotations of each basic pattern and their left-to-right mirror images, ruling out all cases in which queens attack by reflection in a strip added to the board. The remaining cases provide a complete set of solutions to the number problem.

For the case $n = 4$ the pattern of the four queens that is the single solution to the basic problem has rotational symmetry, so that there is no need to test rotations. For $n = 5$ the five queens can be arranged in the two essentially different ways shown in the bottom illustration on page 25. The pattern at the left contains a reflection attack (shown by the colored line) and has rotational symmetry. Therefore it cannot in any

JUSTERINI & BROOKS Founded 1749
Why more dads prefer J&B than any other scotch:
Rare taste.
RARE
BLENDED SCOTCH WHISKY
100% SCOTCH WHISKIES
JUSTERINI & BROOKS LTD
KING GEORGE III
THE PRINCE OF WALES (1921-1936)
PRODUCT OF SCOTLAND
Here's to you, Dad
To the father who always gave us the best.
To my favorite dad
J&B
RARE
SCOTCH
Send a rare gift—a bottle or a case of J&B Rare Scotch—to Dad, almost anywhere in the United States. Just call, toll free: 800-528-6050

The Happy Truth About Th Tales Of The Unexpected

Droughts. Water shortages. Deserts encroaching on once-fertile land.

It's about time America had some good news about water. It's about time we found a supply that could take the burden off the country's overtaxed and often polluted lakes, rivers and reservoirs.

Well, this'll be the best news you've heard about water in a long, long time. The happy truth is America has a virtually endless supply just waiting to be tapped.

All we have to do is look a little deeper.

AMERICA'S BEST WATER IS ALWAYS GETTING UNDERFOOT.

The earth has as much water today as it had in prehistoric times. The water may sometime change forms, but the supply itself is constant. So where is all the water today?

An amazing percentage is where it's always been. Underground.

An incredible 95% of all the fresh water on earth is ground water. In the United States, there's 20 or 30 times as much water underground as there is in all the lakes, streams and rivers *combined.*

What's more, it's a completely renewable supply: Although the rate of replenishing varies greatly, water is seeping into the ground constantly, every hour of the day, every day of the year.

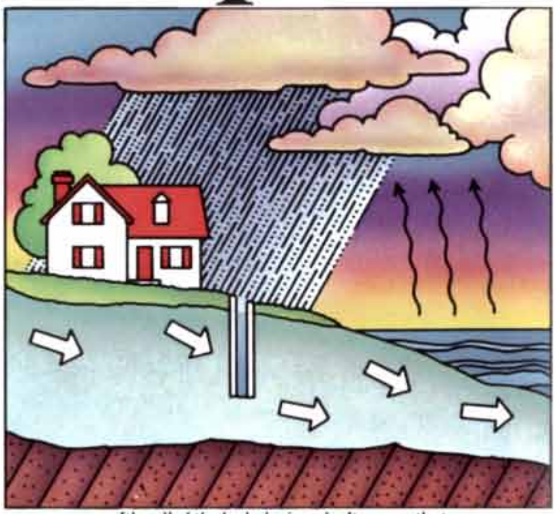
WELL WATER IS A NATURAL DELIGHT.

The American public is using more than 90 *billion* gallons of well water a day.

Approximately 700,000 new wells are drilled every year. Wells now provide almost 35 percent of the water used by municipalities—and over 80% of the water used in rural areas for homes and livestock.

Why is it so incredibly popular now? Part of the reason is the very fact that it is so plentiful—and that once tapped, the water is available for as long as it's sensibly used.

But part of the reason, too, is that well water is part of America's healthy drive to get back to its roots. After all, wells tap America's purest supply of natural water.



It's called the hydrologic cycle. It means that billions of gallons of water are entering underground reservoirs every day.

Does that mean this water's 100% pure? Absolutely not.

Although it rarely suffers from the pollution that so often afflicts surface water, nine times out of ten, well water is even better than totally pure water. It's enriched with dissolved natural minerals. Those minerals give well water in every part of the country its distinct, regional flavor. (If those minerals impart too strong a flavor, a simple filtering device can usually purify the water automatically.)

AMERICA'S MOST FERTILE FIELDS ARE ALL WET.

Many of the irrigation systems in the arid and semi-arid regions of the United States rely exclusively on huge underground reservoirs that have been collecting water for centuries. Without those rich water supplies, many of the most fertile fields in the country would soon become barren. In the 17 western states, for example, about 70 percent of the water used for irrigation comes from wells.

e Endless Well And Other From The Underground.

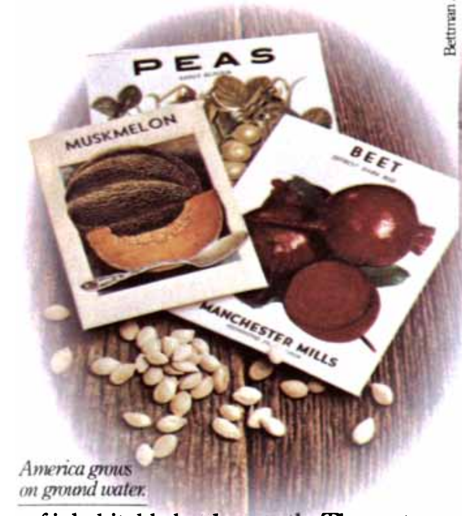
As long as that water isn't pumped faster than it's naturally replaced, there will always be plenty for the fields—and for the factories, homes and entire cities that rely on our underground resource.

WHY YOU SHOULD ALWAYS LOOK BELOW THE SURFACE.

Let's say you're looking for a new water supply. Maybe you're building a new home, factory or office building. Maybe you're planning a multi-acre farm or a quarter-acre garden. Maybe you're on the board of a regional or municipal water authority. Or maybe you'd just like the secure feeling that comes with knowing that no one can turn off or limit your water supply.

The first place to look is your own backyard.

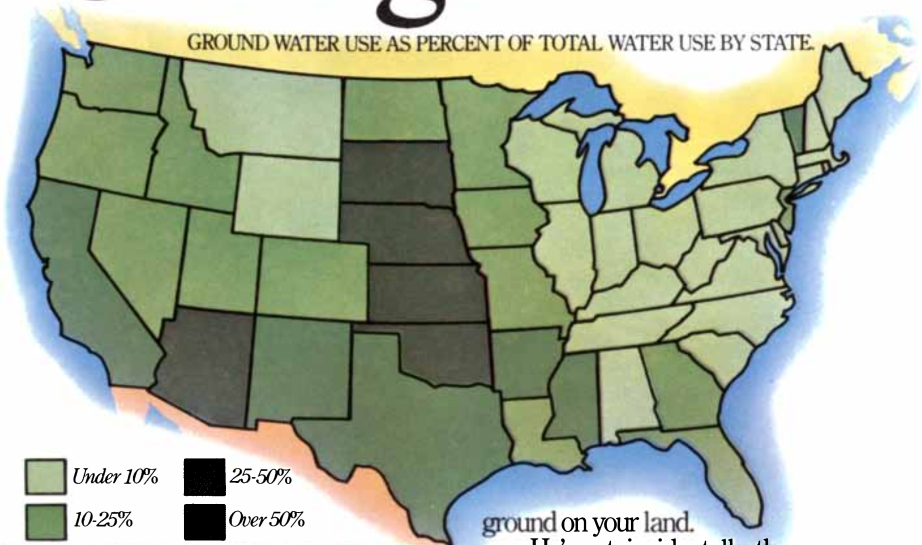
Ground water in some quantity can be found under almost every tract



America grows on ground water.

of inhabitable land on earth. The cost of getting that water is almost always lower than the cost of relying on any other water supply.

If you have a professional, experienced water well contractor inspect your site to determine the feasibility of using ground water, you'll be doing yourself and your locality a favor, because you'll be relying on a water storage system that doesn't drain the public reservoirs. A system that doesn't require major



Bettman Archive

Wells can help lessen the horrible effects of drought.

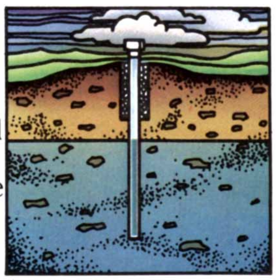
construction. A system that never becomes clogged with silt or weeds. One that never loses water by evaporation. One that doesn't remove productive land from other beneficial uses.

The fact is, ground water is a virtually untapped, unfailing resource. A priceless resource that's so vast, its potential has just begun to surface.

IT'S TIME YOU WENT TO THE PROS.

A professional water well contractor is a water specialist, an engineer and a craftsman all rolled into one. He knows the right techniques for finding the best site and building the best kind of well. He knows what equipment to use.

There's water under almost every inhabitable tract of land on Earth.



He's the man you should call to find out more about what's going on underground on your land.

He's not, incidentally, the man paying for this message. It's brought to you instead by Franklin Electric, the company that makes the motors for most of the submersible pump manufacturers.

We're doing it because we think it's time you knew the good news about one American resource that's not in trouble.

The story's been kept underground long enough.

To find out more about well water, ground water contracters and ground water pumping techniques, just send this coupon to Franklin Electric, 400 E. Spring St., Bluffton, IN, 46714. We'll work with the National Water Well Association to send you more good news from the underground.

Name ____________________

Address ____________________

City ________ State ______ Zip ______

Franklin Electric

Bluffton, Indiana 46714

"13 YEARS AGO I BOUGHT THIS VOLVO BECAUSE IT WAS ADVERTISED AS THE 11 YEAR CAR."

—William Stiles, Bronx, New York

13 years ago, William Stiles, an expert in American Indian history and artifacts, discovered the treasure you see here: a 1966 Volvo.

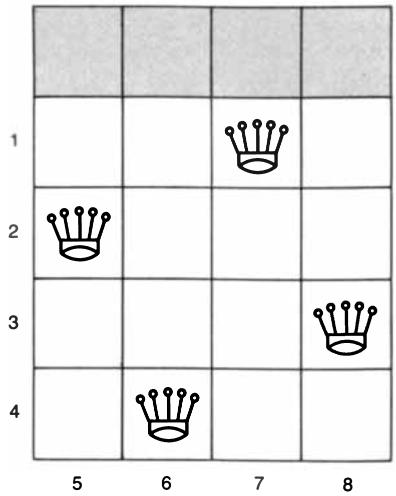
He bought it because ads of the time said Volvos were so durable they lasted an average of 11 years in Sweden.

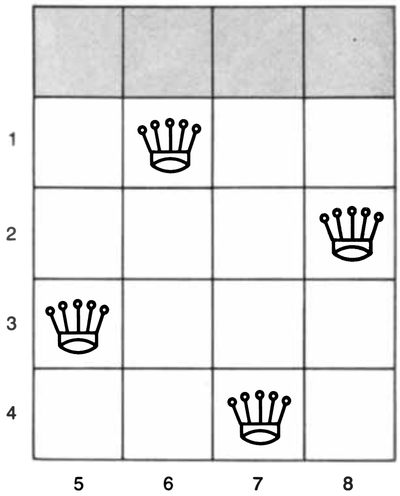
As Mr. Stiles recalls: "One ad said that a Volvo was so tough, you could 'Drive it like you hate it.' I did exactly that. In my field work I've driven this car 295,000 hard miles, much of it through former Indian territory. It's held up even better than promised. Driving it like I hated it made me love it."

Expressions of love are not uncommon among Volvo owners. In fact, 9 out of 10 people who have bought new Volvos are happy.

So if you're unhappy with your current car, do what Mr. Stiles once did after reading one of our ads. Buy one of our cars.

VOLVO
A car you can believe in.





PAIRS	SUM	DIFFERENCE
1,7	8	6
2,5	7	3
3,8	11	5
4,6	10	2

PAIRS	SUM	DIFFERENCE
1,6	7	5
2,8	10	6
3,5	8	2
4,7	11	3

Four queens placed so that no two attack each other directly or by reflection

of its rotations or reflections generate a solution to the number problem. The second pattern generates one solution in the orientation shown and another when it is rotated clockwise 90 degrees. The mirror images of these two patterns give two other solutions, making four in all.

The single pattern of nonattacking queens in the case $n = 6$ has reflection attacks in all rotations (and in their mirror images), so that there is no solution to the corresponding number problem. Therefore the following playing-card puzzle cannot be solved. Place in a row cards with values 1, 2, 3, 4, 5 and 6. The object of the puzzle is to place below each of these cards a card from the set 7, 8, 9, 10, jack and queen so that the 12 numbers obtained by taking the sum and the absolute difference of each vertical pair of cards are all distinct. (The jack and the queen are assumed to have values of 11 and 12.)

For the case $n = 7$ only one of the six basic patterns in only one of its four rotations generates a solution. This pattern and its mirror image generate two solutions to the number problem. Of the 12 basic patterns of nonattacking queens on the standard order-eight chessboard three have one rotation that avoids reflection attacks and another pattern has two. These five patterns and their mirror images generate 10 distinct solutions to the number problem in the case $n = 8$. Interested readers can find all of them by exploring the rotations and reflections of the 12 basic order-eight patterns shown on page 192 of *The Unexpected Hanging*.

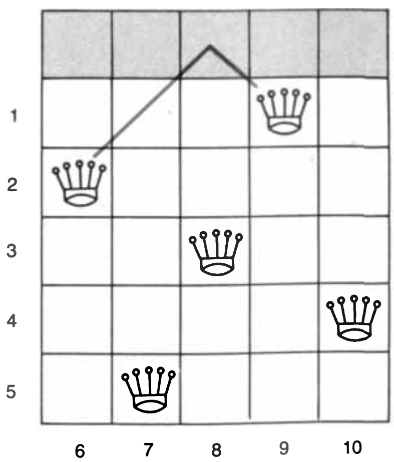
For square boards of orders higher than eight it is much easier to write a computer program for calculating the total number of solutions (including all rotations and reflections) to the variation on the nonattacking-queens problem than to write one that eliminates the rotations and reflections of the solutions to the basic problem. Some recent programs have, however, extended the number of basic patterns for orders nine through 16. The results obtained are respectively 46, 92, 341, 1,787, 9,233, 45,752, 285,053 and 1,846,955. The numbers for orders nine, 10 and 11 were correctly determined before computers came into wide use, but the higher-order results contradict all the claims and conjectures I have seen in older books.

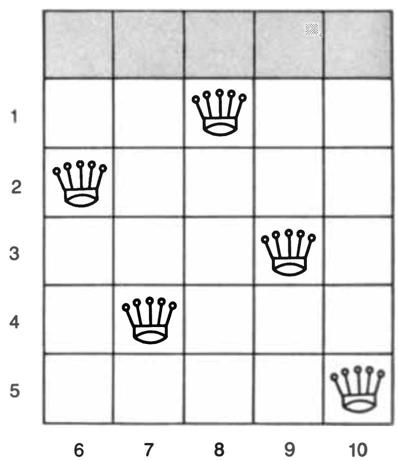
Because the number of basic solutions increases rapidly as n increases it seems unlikely that none of them would furnish a solution when reflection strips are added. For this reason Klarner conjectures that for all n greater than 6 the number problem has at least one solution. In *The American Mathematical Monthly* (Vol. 76, No. 4, pages 399–400; April, 1969) J. D. Sebastian reported on a computer program that found a solution to the number problem for every value of n from 9 through 27. The number of distinct solutions for values of n greater than 8 remains unknown.

At the start of a chess game the asymmetry introduced by the positions of the king and queen makes the pattern of white chessmen a mirror image of the pattern of black ones. The upper illustration on the next page shows how the Japanese graphic artist Mitsumasa Anno has extended this mirror-reflection symmetry to the players themselves as well as to the chessmen of a game. Because of Anno's passion for mathematical themes, he has been called the M. C. Escher of Japan. The illustration is reproduced from his book *Anno 1968–1977,* which has been published in Tokyo by Kodansha but is not yet available in an English-language edition.

Anno's picture calls to mind two whimsical chess tasks that Sam Loyd, America's greatest composer of puzzles and chess problems, published in 1866. In both tasks it is assumed that every move by Black will be a mirror-image duplicate of the preceding move by White. (Obviously we are concerned here only with possible, or legal, play, not with competent play.) The first task is to design a mirror-play game in which White checkmates on his fourth move. The second and much more difficult task is to design a mirror-play game in which White makes an eighth move that forces Black to checkmate him with a nonmirror move. (Students of chess problems call such a move a selfmate.) Next month I shall describe both of these outrageous but legal games.

How many different moves can be made on a standard chessboard? The answer is 1,840. Each move is represented by a line segment in the graph shown in the lower illustration on the next page. The graph is taken from statistician I. J. Good's 1972 Christmas card. The 64 cells of the chessboard have been replaced by 64 dots, so that the graph

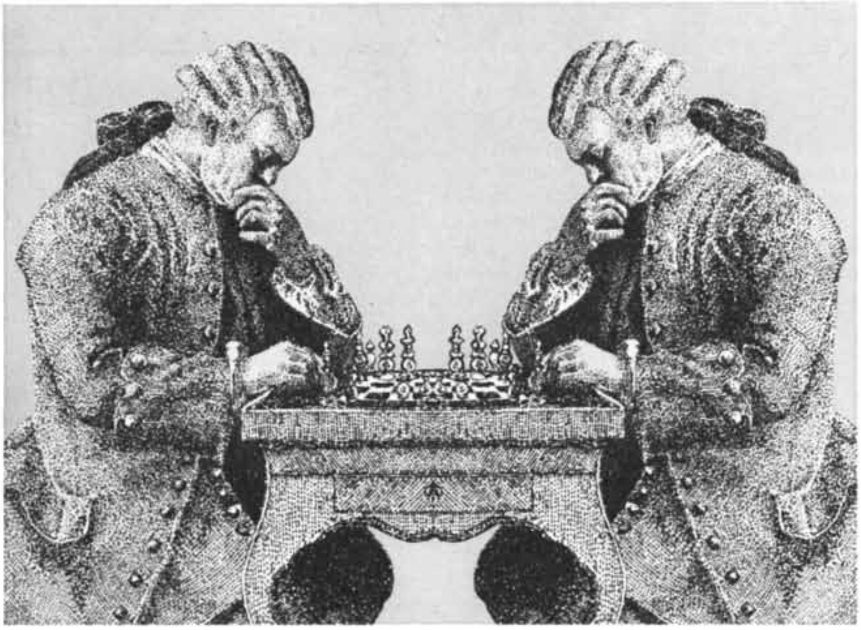




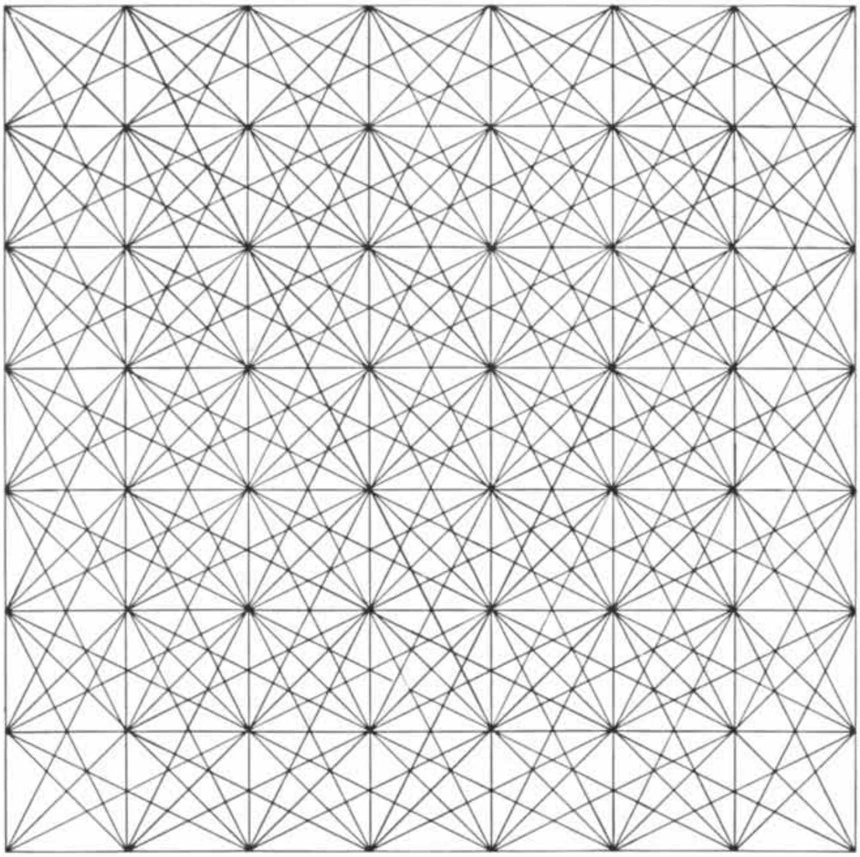
Classic nonattacking-queens problem has two basic solutions for the case **n** = *5* appears to display a seven-by-seven board.

Good's graph calls to mind a nontrivial graph-coloring problem that has been applied to square chessboards. Given any one of the five different chess pieces, what is the minimum number of colors needed for coloring the cells of a board so that no matter where the piece is placed on the board it can only move to a cell of a different color? This minimum number is called the chromatic number of the order-*n* board for the piece.

The smallest chromatic number is 2, the number of colors needed for the knight on all boards of side three or

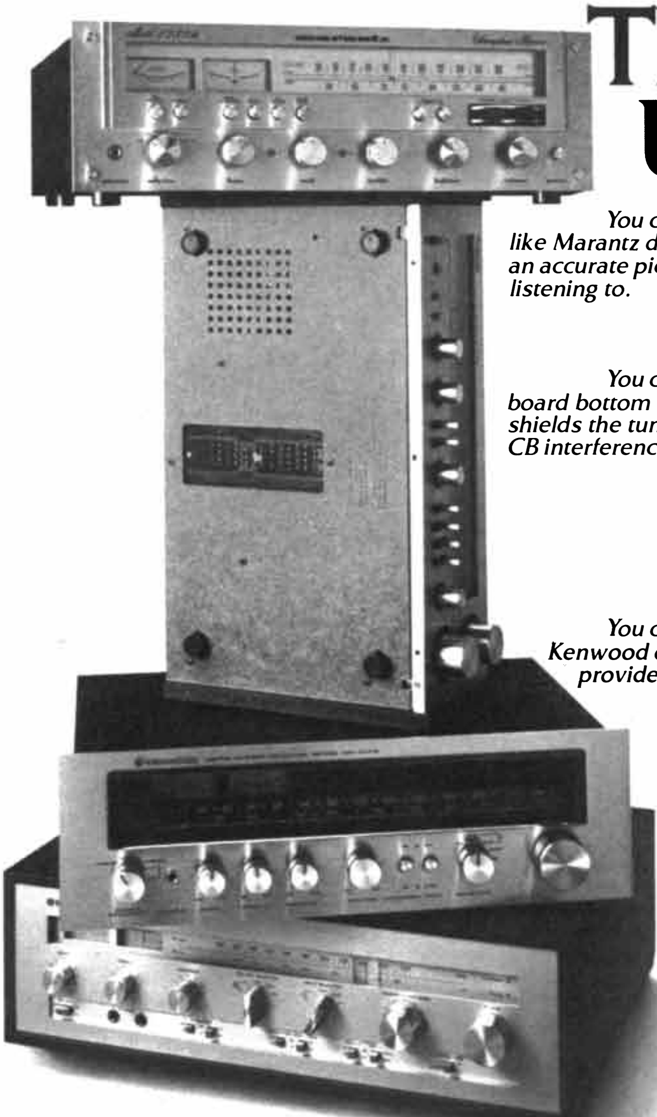


Mitsumasa Anno's mirror chess



The 1,840 different moves in a chess game

THERE ARE A LOT OF WAYS TO BUILD A RECEIVER THAT SELLS FOR UNDER $400.



You can leave out dual wattage meters like Marantz did, instead of providing them to give you an accurate picture of what you're listening to.

You can install an inexpensive press board bottom like Technics did, instead of a metal one that shields the tuning section from spurious noise and CB interference.

You can use a conventional power amplifier like Kenwood did, instead of an advanced DC amplifier that provides cleaner, more natural sound.

You can use standard high band filters for FM stereo reception like Yamaha did, instead of a special integrated circuit that cancels out the unwanted FM stereo pilot signal.

PIONEER DID IT THE RIGHT WAY.

At Pioneer, we build a moderately priced high fidelity receiver somewhat differently than our competitors.

We build it the same way we build a receiver that sells for over $1000. Without compromising quality, features, or sound.

You see, what *really* separates our SX-780 from others is more than just a matter of things like wattage meters, metal bottoms, DC power, advanced circuitry, or even price.

It's Pioneer's commitment to giving you a quality hi fi receiver, no matter how much, or how little you plan to spend.

PIONEER'S SX-780.

So if you're planning to spend less than $400 on a receiver, you couldn't ask for more than the SX-780.

PIONEER® We bring it back alive.

© 1978 U.S. Pioneer Electronics Corp.
85 Oxford Drive, Moonachie, N.J. 07074

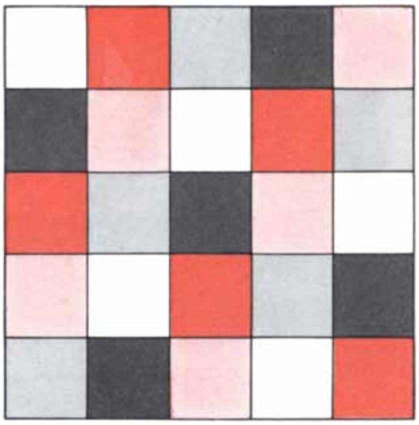
more. The graph-coloring solution for any of these boards is simply to color its cells like those of a checkerboard. This finding gives a quick solution to the following brainteaser: If every cell of a chessboard is occupied by a knight, can all the knights simultaneously move to another cell? The answer is yes if and only if the board's side length is even, because only then are there equal numbers of cells of each color. If the side length is odd, there is an extra cell of one color. In that case, since every knight must move to a cell of the other color, there must be at least one knight with no place to go.

The chromatic number for the king on all square boards of side two or more is 4. If the four colors are labeled *A, B, C* and *D,* the top row of the board is colored *ABABAB*..., the second row *CDCDCD*..., the third row *ABABAB*... and so on. For a rook the chromatic number of an n-by-n board is n. To color the board begin at a corner and color parallel diagonal lines, using color *A* for the one cell of the first diagonal, *B* for the two cells of the second diagonal, *C* for the three cells of the third diagonal and so on until the nth color is used for the main diagonal. Starting again with *A* for the next diagonal, use the same sequence of colors for the remaining smaller diagonals.

For a bishop the chromatic number is n or $n - 1$, depending on whether the longest diagonal on which the bishop can move has n or $n - 1$ cells. If the bishop can move n cells, color each row a different color. If the bishop's longest move is $n - 1$ cells, the color of the first row can be repeated on the nth row.

The case of the queen is the most interesting because it corresponds to the ancient combinatorial problem of forming "diagonal Latin squares," or n-by-n matrixes with cells colored so that no color is repeated within a single row, column or diagonal of any length. The minimum number of colors needed will obviously be equal to the chromatic number of an order-n chessboard for the queen. More than 60 years ago George Pólya proved that such n-by-n squares can be formed with n colors if and only if n is not divisible by 2 or 3. Pólya's method of coloring the five-by-five square is shown in the illustration on this page. Note that the individual colors mark parallel chains of knight moves. (These chains are more evident in the pattern for the 11-by-11 square in the Dover reprint of Maurice Kraitchik's *Mathematical Recreations,* page 252.) Similar patterns generate solutions for any order-n board when n is not a multiple of 2 or 3.

Consider what happens on any such board if red queens are placed on all the red cells, blue queens on all the blue cells, green queens on all the green cells and so on. It will obviously be possible to place n differently colored sets of n queens on the n-by-n board in such a way that no queen of one color attacks another of the same color. The smallest board of this type (other than the trivial one-by-one board) is the order-five board. It has often been sold as a commercial puzzle with 25 counters divided into five differently colored sets of five counters each. The task is to place the counters so that no counter shares a row, a column or a diagonal with another counter of the same color.



A pandiagonal Latin square of order five

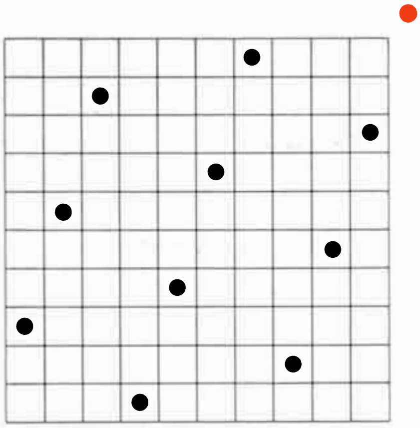
As far as I know the chromatic-number problem for the queen has not yet been solved for all order-n boards when n is a multiple of 2 or 3, although the answer is known for many low-order boards. For example, the six-by-six board requires seven colors and the eight-by-eight requires nine. It has been established that the queen's chromatic number, which cannot be less than n, cannot be greater than $n + 3$.

Pólya's coloring method has an additional attribute that is often called the wraparound property: if either pair of opposite edges of the colored board are joined to make a cylinder, it will still be true that every diagonal contains n colors. In a different terminology each "broken diagonal" of the square has n colors. Squares with this property are called pandiagonal Latin squares.

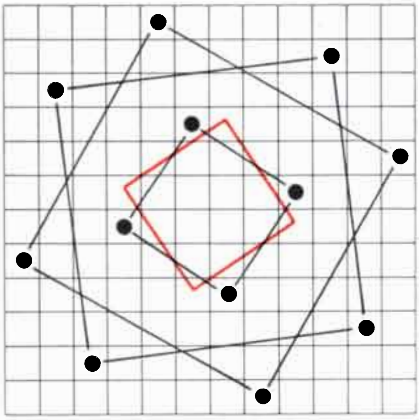
A superqueen, or amazon, is a special chess piece that combines the moves of a queen and a knight. It is not possible to place n superqueens on an order-n board so that they do not attack one another when n is less than 10. The one solution in the case $n = 10$ is shown in the top illustration on page 30. Solomon W. Golomb has shown that such patterns do exist when n is greater than 9 and is either a prime or one less than a prime. Ashok K. Chandra has shown that a wraparound pattern of n nonattacking superqueens exists for an order-n board if and only if n is greater than 11 and is not a multiple of 2 or 3. I know of no general solution of the superqueen problem.

Sin Hitotumatu of Kyoto wrote a computer program that found six solu-

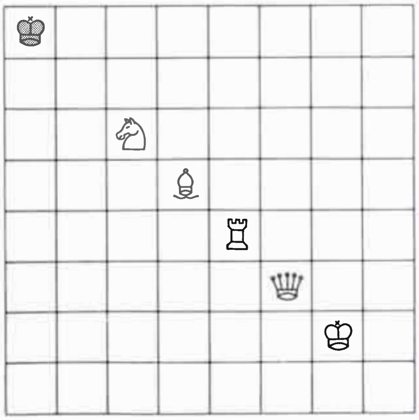
tions for the order-11 board, although all of them are identical if the board is wrapped around in both directions to make a torus. These solutions were found earlier by Golomb and are based on his construction method. To find one of them simply extend the order-10 solution shown in the top illustration below by adding a row at the top and a column at the right; the 11th superqueen is placed in the upper right-hand corner of the board, as is shown in color. Hitotumatu also found 23 basic solu-



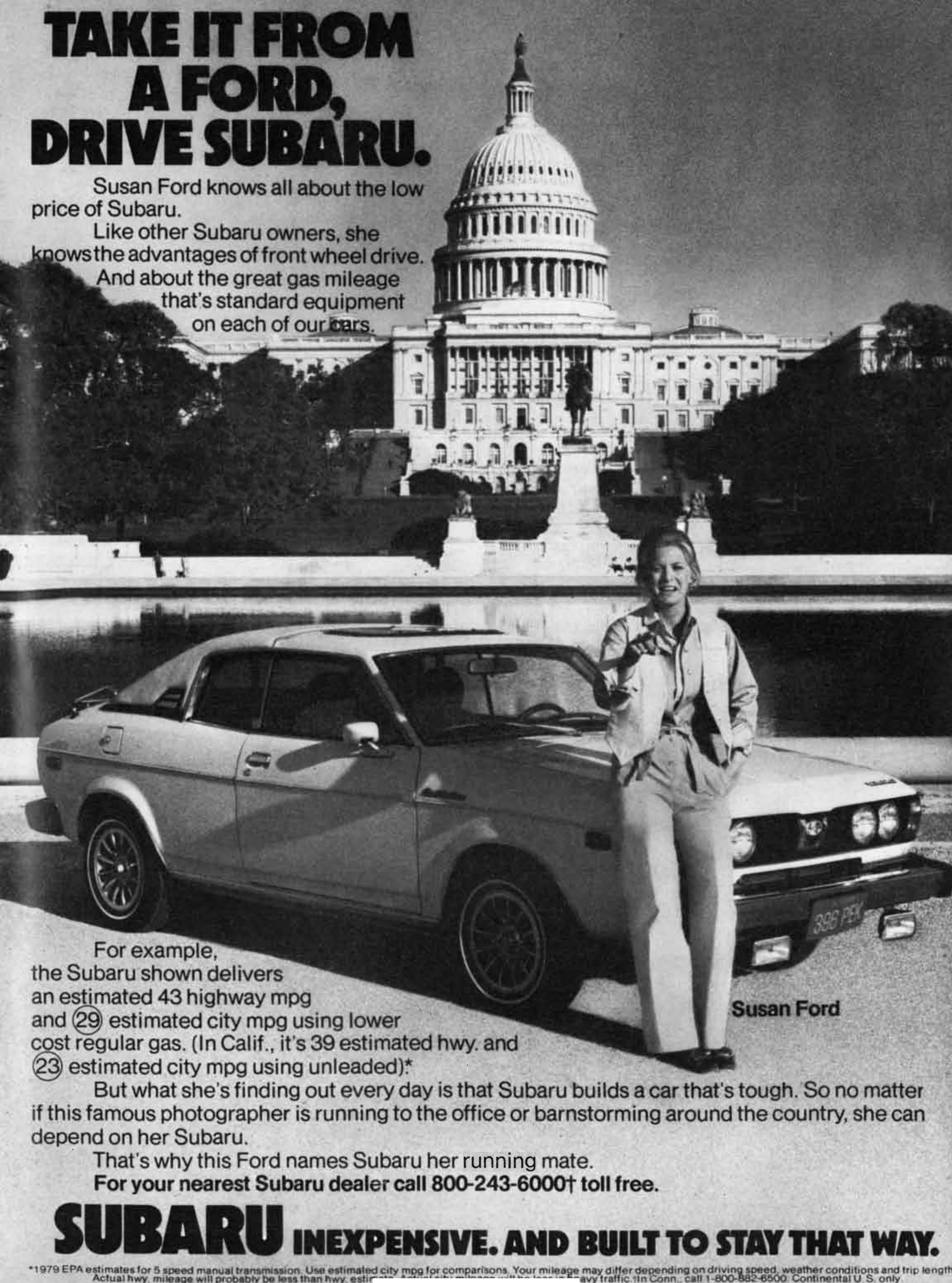
Smallest n*-by-*n *board for* n *superqueens*



Twelve superqueens on a 12-by-12 board



How can White mate in less than one move?


TAKE IT FROM A FORD, DRIVE SUBARU.
Susan Ford knows all about the low price of Subaru.
Like other Subaru owners, she knows the advantages of front wheel drive.
And about the great gas mileage that's standard equipment on each of our cars.
396 PEK
Susan Ford
For example, the Subaru shown delivers an estimated 43 highway mpg and (29) estimated city mpg using lower cost regular gas. (In Calif., it's 39 estimated hwy. and (23) estimated city mpg using unleaded).*
But what she's finding out every day is that Subaru builds a car that's tough. So no matter if this famous photographer is running to the office or barnstorming around the country, she can depend on her Subaru.
That's why this Ford names Subaru her running mate.
For your nearest Subaru dealer call 800-243-6000† toll free.
SUBARU INEXPENSIVE. AND BUILT TO STAY THAT WAY.
*1979 EPA estimates for 5 speed manual transmission. Use estimated city mpg for comparisons. Your mileage may differ depending on driving speed, weather conditions and trip length. Actual hwy. mileage will probably be less than hwy. esti [illegible] avy traffic. †In Conn., call 1-800-882-6500. Continental U.S. only.

tions for the superqueen problem on the order-12 board. The one in the middle illustration on page 30 is the prettiest, with its symmetrical arrangement of two equal squares and a smaller interior square. (The colored lines show how the smaller square can be modified by rotation.)

Donald E. Knuth has recently found a method for solving the order-14 and -15 cases that generalizes to certain higher orders. It is now known that if the Knuth and Chandra constructions are combined, solutions to the superqueen problem exist whenever n is greater than 9 and does not have the form $12k + 8$ or $12k + 9$ for some integer k. Chandra's method provides solutions for the cases $n = 6k$, $n = 6k + 1$, $n = 6k + 4$ and $n = 6k + 5$; Knuth's method provides solutions for half of the remaining cases. Hence the smallest unresolved cases are $n = 20$, $n = 21$, $n = 32$ and $n = 33$.

I shall conclude with a joke. How can White, in the position shown in the bottom illustration on page 30, checkmate Black by making a fraction of a move?

The first problem last month concerned Robert Connelly's card betting game Say Red. In the game a banker deals cards face up from a standard deck, and at some point (before the last card is dealt) the player must say "Red." If the next card is red, he wins the game; otherwise he loses. At all times the player may inspect the cards already dealt. From this information can he devise a strategy that will in the long run raise his probability of winning above 1/2?

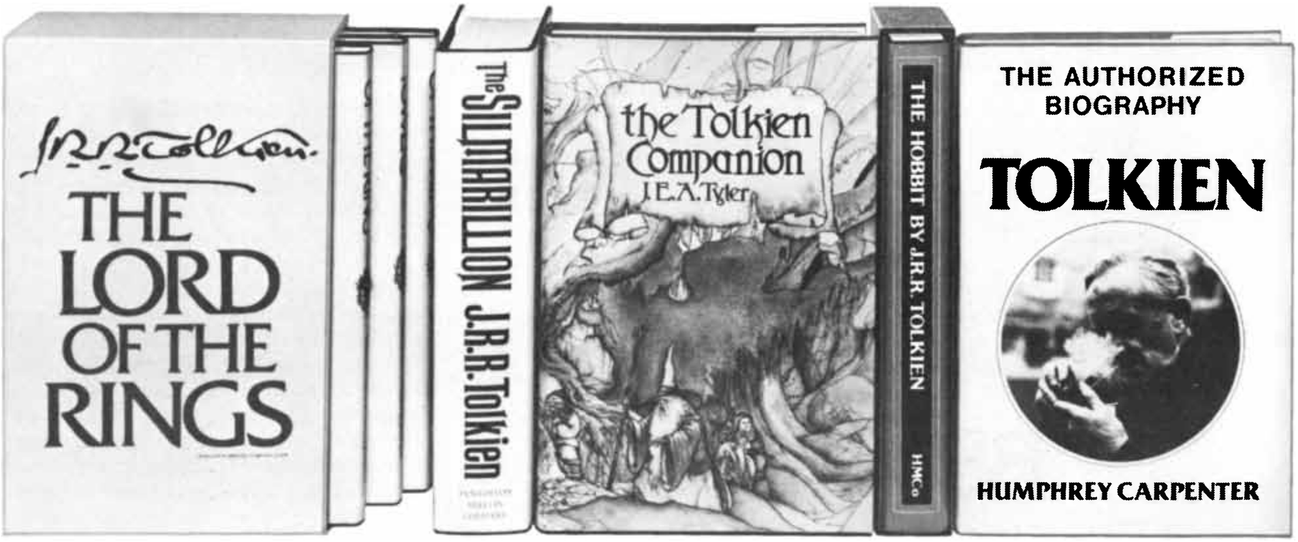
Surprisingly, he cannot. There are many formal proofs of this fact, but perhaps the best way to "see" why it is true is to analyze a deck consisting of two black and two red cards. Labeling the black cards B and the red cards R, the six equally likely deals are as follows.

RRBB
RBRB
RBBR
BBRR
BRBR
BRRB

Glancing down the column, it is obvious that the probability the player will win is 1/2 regardless of whether he calls before the first, the second, the third or the last card has been dealt. On the other hand, suppose two black cards are dealt. One could argue that the third must be red and therefore the player cannot lose. That is true, but in the long run it happens only once in six deals. On the other five deals the probability that the player will win is $5/6 \times 2/5$, or 1/3, and so the overall probability of winning on the third card is $1/6 + 1/3$, or 1/2.

Now try a different tack. Suppose the player's strategy is to call red after the first card is dealt if that card is black, and otherwise to call red before the last

All the riches of Middle-earth.



A complete Tolkien library for only $12⁵⁰

Publishers' list prices total: $85⁸⁰ You simply agree to buy 4 books within a year.

No writer in our century has created a more enduring epic fantasy than J.R.R. Tolkien in his classic novels set in the mythical kingdom of Middle-earth. These enthralling works of high adventure, dark treachery and glorious deeds comprise a lifework of superlative entertainment and a triumph of imaginative genius.

The Lord of the Rings (Second revised edition): *The Fellowship of the Ring • The Two Towers • The Return of the King*. 3-volume boxed set. **Publisher's list price: $32.95**

The Silmarillion edited by Christopher Tolkien. The Elder Days of Middle-earth. **Publisher's list price: $10.95**

The Tolkien Companion by J.E.A. Tyler • A complete guide to Middle-earth. **Publisher's list price: $12.95**

The Hobbit (Collector's Edition) • Boxed, with paintings and drawings by the author. **Publisher's list price: $18.95**

Tolkien: A Biography by Humphrey Carpenter • The only authorized biography. **Publisher's list price: $10**

Facts About Membership

You receive the *Book-of-the-Month Club News®* 15 times a year (about every 3½ weeks). Each issue reviews a *Main Selection* plus scores of Alternates.

If you want the *Main Selection* do nothing. It will be shipped to you automatically. If you want one or more Alternate books—or no book at all—indicate your decision on the reply form always enclosed and return it by the date specified.

Return Privilege. If the *News* is delayed and you receive the *Main Selection* without having had 10 days to notify us, you may return it for credit at our expense.

Cancellations. Membership may be discontinued, by either you or the Club, at any time after you have purchased 4 additional books.

Begin your Book-of-the-Month Club membership with this special opportunity to own a basic Tolkien collection in seven beautiful volumes. Then continue to choose the books you want from the best fiction and nonfiction published today.

And continue to enjoy *real* savings. Because unlike the inferior, all-of-a-kind editions that some clubs offer, all of our volumes are as varied, handsome and well-made as those you'd find in a bookstore.

After a brief trial period, you'll also benefit from our Book-Dividend® system, which lets you save at least 70% on current art books, reference sets, literary classics, records, children's books and more. Join today for the finest in reading pleasure tomorrow.

BOOK-OF-THE-MONTH CLUB®
AMERICA'S BOOKSTORE®

Book-of-the-Month Club, Inc., Camp Hill, Pennsylvania 17012

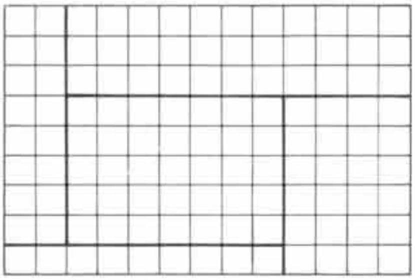
Please enroll me as a member of Book-of-the-Month Club and send me the TOLKIEN LIBRARY. Bill me $12.50, plus shipping and handling charges. I agree to buy 4 books during the coming year. A shipping and handling charge is added to each shipment. **A170-6**

Mr.
Mrs. 9-19
Miss (Please print plainly)

Address.. Apt.........

City..

State.. Zip.................



A rectangled rectangle

card is dealt. If the first card is black, the player does have a 2/3 chance of winning on the next card, but this is counterbalanced by the fact that if the first card is not black, he has a 2/3 chance of losing on the last card.

This reasoning generalizes to decks with b black cards and r red ones. The probability that a player will win by any strategy whatsoever, Connelly points out, is always $r/(b + r)$. Therefore knowing which of the cards in a deck have been dealt gives a player no advantage at all, regardless of the ratio of the black cards to the red cards or the size of the deck. No strategy is better than always calling red on the first card, or on the last card, or indeed on any card. (As I stated last month, the player must specify his bet before each deal; otherwise he could come out ahead by, as in a system for playing blackjack, betting high when the odds favor him and low when they do not.)

The second problem was to design a set of four cards for a magic trick similar to the one described in this department last month except that the four numbers apparently picked at random must always have the same product rather than the same sum. The secret is to choose pairs of numbers to be placed on opposite sides of the cards that all have the same ratio. For example, on the cards prepared by Shigeo Futigawa, the Tokyo magician who invented both variants of the trick, the number pairs are 26/34, 39/51, 52/68 and 65/85. In each case the ratio is 13 to 17.

The trick is handled almost exactly the way the one I described last month is, except that the four selected numbers are multiplied instead of added. The predicted product in this case is 5,860,920. Note that this number is equal to $13 \times 13 \times 17 \times 17 \times 2 \times 3 \times 4 \times 5$. As with last month's version, I leave it to readers to prove algebraically that the trick cannot fail.

In this department for March the rectangled rectangle shown in the illustration at the right on page 22 was drawn incorrectly. The rectangle at the lower right in the larger rectangle should not have been three units wide but four, as is shown in the illustration above.

LCD Alarm Chronograph II

Twice as Good As the Original

5 Major Improvements

Now, with five major improvements the Advance II is twice as good as the original—at the original low price.

The differences you'll enjoy with our new LCD Alarm Chronograph are real and make it the absolutely best value on the market today.

1. Perfect Accuracy Imagine, at the touch of a button you can correct the slightest variation—to a tenth of a second. So, in addition to guaranteed quartz crystal accuracy of ± 5 seconds a month, you can instantly correct, or "trim the crystal" electronically, actually have to-the-second accuracy.

The correct time all the time.

2. Now the Alarm's an Automatic Reminder. The alarm can be set in an instant for any minute of any hour. And now features a snooze alarm, personal reminder system you can repeat every seven minutes. Great, because you can't always leave for a meeting or your plane when the alarm rings, make that phone call at the time you promised. With the reminder you can be sure you don't forget.

Also, the new alarm has a warning or "executive tone." This sounds once and pauses—so only you hear it and can shut the alarm off if you want to before the regular 52 second beeping cycle begins.

When the alarm's set, "ALARM" appears. To check for the time, touch the alarm button.

3. Solar Rechargeable 3-year Battery. The battery in the Advance II is recharged—automatically by any kind of light. So, instead of the usual one year, you'll get guaranteed (by Mallory) 3-year battery life. And the bank of seven solar cells makes a handsome addition to the bold, impressive appearance.

4. Digital Counter—for laps when you're running, golf score and more. Now, when someone asks, "What'd you shoot?" you can look at your watch instead of your score card. And it's terrific for keeping track of your laps, for counting cards, your ski runs, telephone calls, customers (daily, weekly, monthly cumulative comparisons), the people at a meeting, and in taking inventory. Touch a button to count, then store the tally in memory if you're interrupted or want to refer to the total any time after.

Exclusive touch button counting. (To 9,999)

5. Memory Bank for every mode. Which brings us to another major difference and advantage: every mode can be put into memory. For example, with the imitators you can't get the time when you're using the chronograph. It would wipe out timing of the event!

The Chronograph System. As to the chronograph, or split-second timer, its precision is so fine, it borders on the infinitesimal—to one-hundredth of a second for the first 20 minutes . . . to the second for a full 1 hour . . . and in a variety of ways unequaled by any other instrument.

Still Only $100.00

The best known LCD Alarm Chronograph, of course, is the Seiko, which regularly sells for $299.95. While Seiko pioneered in this area, its chronograph is limited and still times only to 1/10th of a second, its alarm has no reminder, its battery has to be replaced at least once a year.

Copies abound, of course. But Advance is a manufacturer in its own right. Forging ahead, building on existing technology, creating a reputation of its own for extraordinary quality, which we're sure the almost 30,000 people who ordered the first LCD Alarm Chronograph from us will verify.

(Actually, as one of the oldest and largest mail merchandisers, our only concern is to assure all these people that the refinements in the Advance II in no way minimizes the fact that the watch they already own does more and does it better than anyone else's.)

The LCD Alarm Chronograph II is still only **$100.00** which includes shipping, handling, insurance and a handsome gift case. It's available in either chrome (white) or gold-plate (yellow) on a forged, hand-finished brass case. The back is stainless steel as is the band, which adjusts instantly to a perfect and extremely comfortable fit.

Then, so you can see when it's dim or you're in the dark, the face lights up.

The Watch and Electronic Calendar. Most important, the basic display gives you the exactly right time. The hour, minutes and running seconds, plus the day of the week. At a touch you can replace the time with the month, date and day, always right with no adjustment necessary.

Thirty day trial. Now we invite you to put everything we say and promise to the test—without risk or obligation for 30 days. And to send the LCD Alarm Chronograph II back to us for a complete refund if it isn't the equal of any watch at any price for accuracy, versatility and usefulness. A watch that wins the admiration of your friends as well as your own pleasure and lasting satisfaction.

CREDIT CARD BUYERS

CALL 800-621-5554

(Illinois residents call 800-972-5858)

To order by credit card, call the toll-free number above. Send your check to Douglas Dunhill at the address below. Be sure to specify gold or chrome. (Illinois residents add the appropriate sales tax.)

The Advance LCD Alarm Chronograph II comes with a full ONE YEAR limited warranty and our assurance it will provide you with years of unmatched, trouble-free performance. Order now and enjoy the satisfaction that comes with wearing a watch that's second to none.

Three-way Chronograph System

No other instrument, at any price, gives you greater precision than the LCD Alarm Chronograph II, with its one 1/100th of a second accuracy. Or greater versatility and flexibility. Three separate chronographs that work in memory—combine in function.

Timed to 18 minutes, 14 seconds and 85/100ths of a second.

1 Add time . . . the stopwatch, which you can stop (for a time-out, say) and have pick up where you left off when the action resumes.

2 Split time . . . lets you stop to get a reading (of the ¼, the ½, each contestant crossing the finish line) while overall timing continues in memory until displayed.

3 Lap time (Taylor Split) . . . gives you the time of individual segments of an event, automatically starting again from zero so you can time the following sequence. It literally splits elements that can't be separated any other way.

All this, plus a digital counter, in a sleek, superbly styled timepiece. Order now and take 30 days to prove how easy it all is to master and how useful when you have.

Douglas Dunhill
INC. AFFORDABLE QUALITY
Dept 40-2576
Ten Douglas Dunhill Dr., Oak Forest, IL 60452

BOOKS

Visions of Mars and the terrestrial deeps, celestial clockwork, microtubules, grasses

by Philip Morrison

THE MARTIAN LANDSCAPE, by the Viking Lander Imaging Team. National Aeronautics and Space Administration, NASA SP-425, U.S. Government Printing Office ($12). FAMOUS, PHOTOGRAPHIC ATLAS OF THE MID-ATLANTIC RIDGE: RIFT AND TRANSFORM FAULT AT 3,000 METERS DEPTH, by Arcyana. Gauthier-Villars/ C.N.E.X.O., 17 Rue Remy-Dumoncel, 75014 Paris (150 francs). It is more than a century since Jules Verne described so delightfully the extravagant voyages of princely Captain Nemo, restless wanderer of the deep, and of those who sailed to the moon and back, projected there by the super-Colombiad the Gun Club promoted so well. Today even more extravagant voyages have become real, and the clever verisimilitude of Verne's word pictures has given way to the meticulous photographic image and the sober technical summary. Just as Captain Nemo's dedicated crew of exiles anonymously manned the *Nautilus* and the Gun Club joined in wide fellowship the colleagues of artillery science, so in our time do these explorations put an entire collectivity at work (not to mention legions of taxpayers).

These two large, thin volumes with hundreds of big halftones, maps and diagrams are Jules Verne come true. It was the Viking Imaging Team that specified and operated the twin cameras each Viking Lander carried to the dry, airless deserts of Mars ("Did we really want to send a one-eyed traveler to Mars?"); Arcyana is a collective name (based on the names of the French deep submersibles *Archimède* and *Cyana*) for seven scientist-adventurers of the French-American Mid-Ocean Undersea Survey (FAMOUS) of 1973 and 1974, including one of the pioneers of continental drift, Xavier Le Pichon.

Everyone who can see has had a chance to view by photograph the landscape of Mars. Here it is shown in full, close up and far away, in colors real and false, even with the frost of winter shadowy on the land. Perhaps freshest of all are the 19 carefully controlled stereo pairs. A cardboard viewer is enclosed in the book. If you ask your friend whether he "sees the third dimension and he responds noncommittally 'yes,' then you know he has not." When he utters "an exclamation of surprise and wonder... you will know he has seen it"; the effect is so unusual "that very few people come upon it without excitement." The remark is just; the reality of the undulating landscape transcends all the postcard scenes of that rock-strewn desert. A single scene shown in color (under three assumptions for determining true color balance) is also most informative. Thomas A. Mutch, the Brown University geologist who was the team leader, supplies such an open and detailed anecdotal chronicle of his decade of work that we gain empathy with the members of the team of diverse experts, men and women, young and old, whose skills secured these wonderful scenes.

Points of special interest are the *B* on a rock (a graffito of chance), the ditches dug by the Lander as a child might but over a gap of half an astronomical unit, the rationale of the slow-scan camera design and the picture of a fossil trilobite some wag sneaked in during the rehearsal stage in Pasadena to try the skill of the interpreters. The story of the bids, proposals, committees and all the other things that end in the assembly of a NASA team and its instruments bears out Mutch's feeling that "it is a supremely democratic arrangement"—for those who can offer enough expertise and commitment. It is reassuring to read that the contract for the cameras was not blindly let to the lowest bidder but was awarded to the bidder with the most compelling and reliable design. One hopes the General Accounting Office will rest content with this splendid album of the remotest desert we have yet seen, a land where change is all but absent. Only a little dust and frost are there, blowing in the unheard wind.

The sea floor of our own planet is not so strange; it resembles familiar volcanic regions in the open air even though it lies in endless dark under two miles of water. It has been only a few years since visions of such depths first materialized; now submersible vessels, with bright searchlights and thick portholes in their strong steel spheres, have sought out a particular geological feature in the deep Atlantic. The tracks of the French submersibles (their American counterpart the *Alvin* supplied the data for a similar atlas published earlier) wandered like a sea worm for hundreds of hours at depth. Ten thousand photographs and many samples were the take, given strength by the long mapping voyages of surface ships making extensive sonar soundings and calibrated by careful acoustical triangulation to absolute geographical positions to within tenths of a kilometer. Relief maps show the bottom extremely well: three square patches about five kilometers on an edge patrolled by these geologists, hammer and sample having been replaced in a sea change by powerful new technology.

What they were exploring was, of course, the Atlantic suture, a fully developed segment of the boundary at which the African and the American plate spread apart and slide transversely, 20 kilometers in a million years. The region is volcanic, strewn with pillow lavas, caved and fissured. The glassy surface slowly darkens and weathers to allow age measurements, which are fully confirmed by the fission tracks that form in the glass. The newest portions of this land are only centuries old; this is a submarine visit to a lava field that was active, so to speak, in the memory of the oldest inhabitants.

There on the edge of a deep crevasse, into which the searchlight beam fades before we can see bottom, grows a sea-tulip cup, the peduncle of a *Hexactinellida* sponge. A few photographs in color show the somber tones of hydrothermal flow. The same vessels are hard at work today in another rift: the other edge of the American plate, in the Gulf of California. We can look forward to the detailed exploration of the submarine crust, the most active of large-scale geological structures, a result the *Nautilus* somehow missed.

A warning to those who would decipher French acronyms: C.N.E.X.O. in full ends with the words "pour l'Exploitation des Océans."

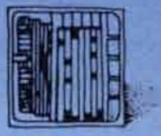
GEARED TO THE STARS: THE EVOLUTION OF PLANETARIUMS, ORRERIES, AND ASTRONOMICAL CLOCKS, by Henry C. King, in collaboration with John R. Millburn. University of Toronto Press ($50). The cathedral town of Wells lies across England from Canterbury. The noble sculptured facade was no longer new in Geoffrey Chaucer's time, when (inside the transept and more for the use of the clergy than for the public) the splendid astronomical clock still on view there today was built. Its beautifully decorated two-plate concentric dial displayed the sun, moving once around in 24 hours, and the moon, its age indicated both by a pointer and by its changing shape, waxing and waning through all the phases in 30 days (with some hand adjustments implied). One of the earliest extant astronomical clocks in Europe, this fine instrument declares

BY MAKING ENGINES THINK.

miniature, with a memory and the ability to manipulate what it learns in terms of what it already knows. It works so well that car and heavy-duty-equipment manufacturers in America and Europe plan to use it, some as early as the 1980 model year.

IMPOSSIBLE WITHOUT ELECTRONICS.

Such precise, continuous engine management would be impossible without the integrated circuit, an electronic microcosm that contains the equivalent of twelve thousand transistors and measures about 5mm square. These small miracles are the central nervous system of Motorola's electronic engine-management system, and they're a remarkable but not unique demonstration of the kinds of things Motorola is doing with microelectronics today.



A microcomputer, drawn larger than life.

IMAGINATIVE ELECTRONICS.

Motorola is not only one of the world's largest manufacturers dedicated exclusively to electronics, but also one of its foremost designers of custom and standard semiconductors.

We've come a long way from the time when we put radios into cars fifty years ago, and TV sets into America's living rooms. Now we make hundreds of models of two-way radios, and we no longer make home TV sets here at all.

We make microelectronics carry telephone services to places where there are no phone lines. Transmit electrocardiograms and voice messages simultaneously from the scene of an accident to a nearby hospital. Help the energy industries develop the resources of the earth, the sea and the sun.

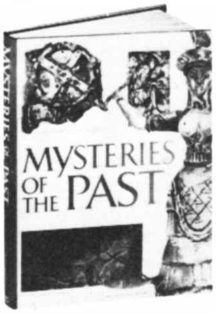
And we help make automobile engines think about how they use precious fuel.

MOTOROLA

Making electronics history since 1928.

For further information, write Public Affairs Office, Corporate Offices, Motorola, Inc., 1303 E. Algonquin Road, Schaumburg, Illinois 60196.

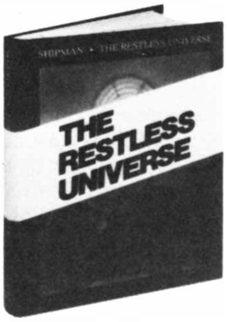
Motorola and Ⓜ are registered trademarks of Motorola, Inc.



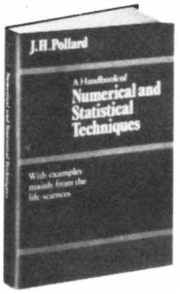
63340-2. MYSTERIES OF THE PAST. Captivating investigation of prehistory and great riddles. Outsized. Sumptuously illustrated. *Counts as 2 of your 3 books.* **$34.95**

34730. THE ATLAS OF EARLY MAN. *Jacquetta Hawkes.* A panoramic overview of concurrent events in history through charts, graphs, and illustrations, from prehistory to the Roman Empire. **$20.00**

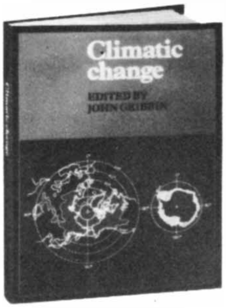
50010. FUTURES RESEARCH. Professional insights by 23 prominent researchers in such areas as general systems research, mathematical modeling and the future of culture. **$22.50**



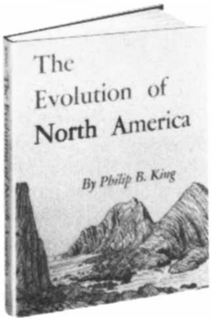
73940. THE RESTLESS UNIVERSE. *Henry L. Shipman.* A highly acclaimed 450-page tour of the ideas, principles, and process of modern astronomy. **$16.95**



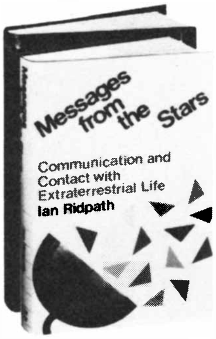
52292. A HANDBOOK OF NUMERICAL AND STATISTICAL TECHNIQUES. *J. H. Pollard.* **$24.95**



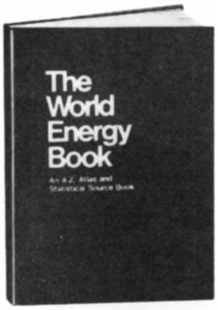
38362-2. CLIMATIC CHANGE. *Dr. John Gribbin.* Provides an A to Z report on the basic pieces and patterns of the earth's rich climatological puzzle. *Counts as 2 of your 3 books.* **$32.50**



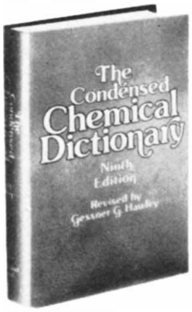
46550. THE EVOLUTION OF NORTH AMERICA. *Dr. Philip B. King.* Sweeping geological survey of our continent. Extensively illustrated. **$25.00**



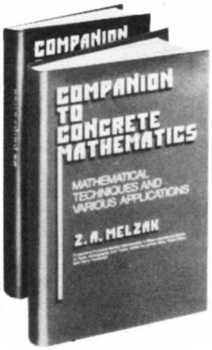
74771. MESSAGES FROM THE STARS/THE RUNAWAY UNIVERSE. Interstellar communication. Plus, the birth and death of the cosmos from the Big Bang to titanic future holocaust. *The 2 count as one book.* **$19.95**



87610. THE WORLD ENERGY BOOK. *Edited by Crabbe and McBride.* Global energy sources and the complex issues surrounding them. Includes 35 world maps and 40 pages of tables, diagrams, and charts. **$25.00**

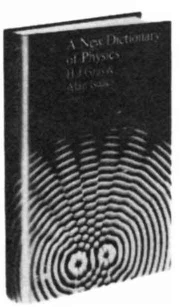


40167-2. THE CONDENSED CHEMICAL DICTIONARY. Updated to meet today's needs. Over 18,000 entries. *Counts as 2 of your 3 books.* **$32.50**

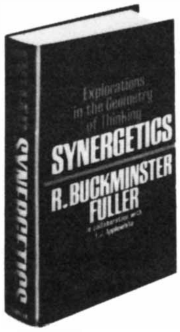


39746-2. COMPANION TO CONCRETE MATHEMATICS. VOLUMES I AND II. *Z. A. Melzak.* Mathematical techniques and various applications. *The set counts as 2 of your 3 books.* **$46.45**

62915. MONUMENTS OF CIVILIZATION: Ancient Cambodia. *Mazzeo and Antonini.* Portrays the magnificence of ancient Khmer civilization—over 100 color photographs. Oversized volume. **$25.00**



64145-2. A NEW DICTIONARY OF PHYSICS. Updated reference contains hundreds of illustrations, graphs, charts and tables. *Counts as 2 of your 3 books.* **$35.00**



81840. SYNERGETICS. *R. Buckminster Fuller.* **$25.00**

37347. CASTE AND ECOLOGY IN THE SOCIAL INSECTS. *Oster and Wilson.* **$20.00**

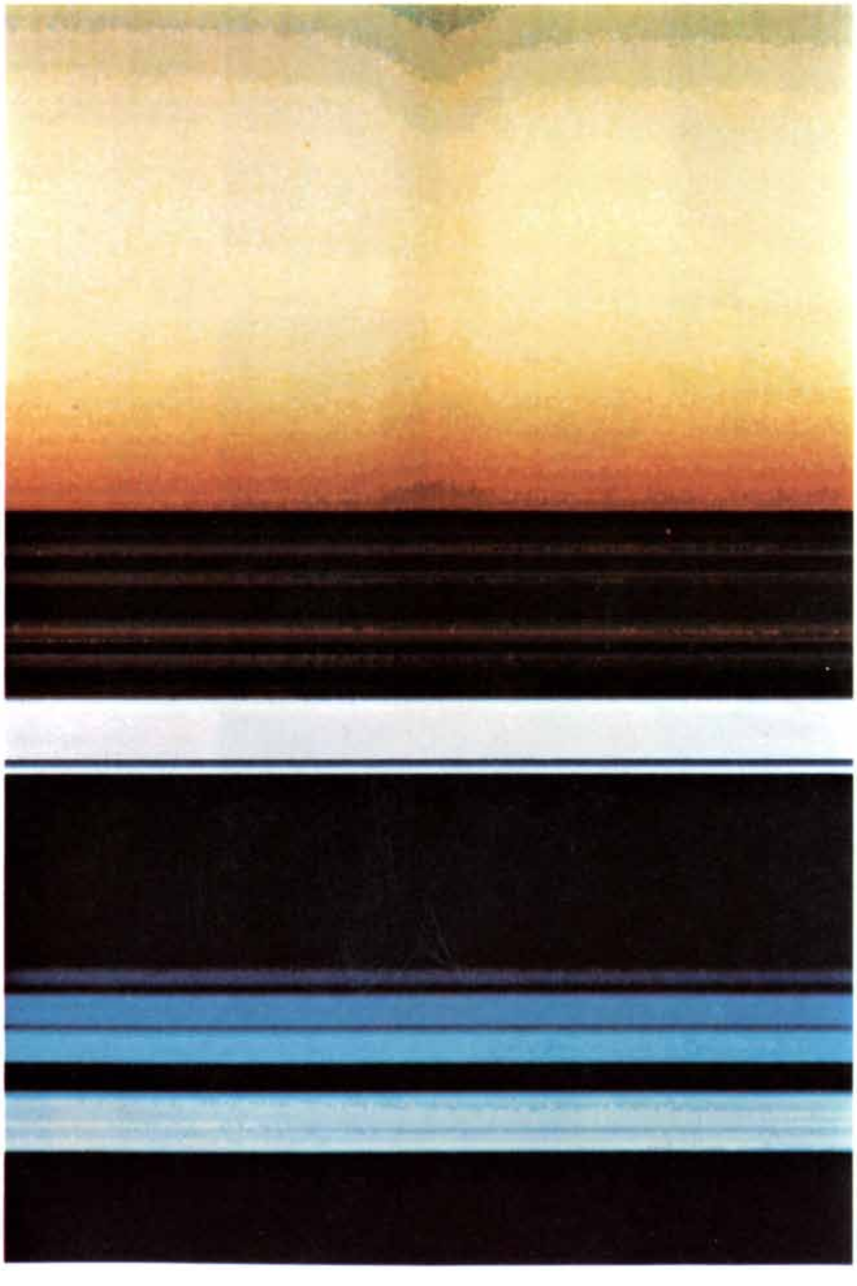
(Publishers' Prices shown)

If the reply card has been removed, please write to
The Library of Science
Dept. 2-B5A, Riverside, N.J. 08370
to obtain membership information and application

its high purpose in a Latin motto that can be construed as "This circular dial represents the universe in miniature."

In a comprehensive volume (so meticulous and circumstantial as to reduce most general readers to happy samplers) a Toronto historian of astronomy and his expert English horologist partner have assembled a serious text and a rich collection of images to spell out the entire history of mechanisms "geared to the stars." It was Derek de Solla Price who made clear how much clockwork is owed less to the usefulness of marking the hour than to the exalted joy of modeling the motions of the cosmos. This theme is pursued here in authoritative detail, from the corroded bronze fragments of Antikythera to the projection planetariums of our time, those hemispherical theaters served worldwide by ingenious mechanisms made in Jena, Osaka or the Delaware Valley.

Wells had a big wrought-iron movement, hand-forged and hand-cut and innocent of screw threads, which is still to be seen in going order in the great Science Museum of South Kensington. (The dials in Wells are now turned, and their assisting knights and hammermen are driven, by a modern movement concealed within the stonework.) Perhaps more people know the later public clocks of Prague, Strasbourg, the Piazza San Marco and Rouen, among many others across Europe. All are given some account here, and the story is extended to the elegant brasswork of Giovanni de'Dondi, as old as Wells but subtler and more elegant; its many dials vivified the Alfonsine tables, recording the motion of all the planets for the admiration of the Duke of Pavia. Wheelwork grew by successive approximations, and timekeeping itself prospered through the centuries, until Jost Bürgi seems to have supplied Tycho Brahe himself with dials reading in seconds and accurate to within a second or so per hour.



SHADOW OF PHOBOS is the vertical streak in this repeated-line scan image made from *Viking Lander 1* as it rested on the surface of Mars. Blue and white horizontal stripes at the bottom of the picture correspond to color-test patches on the spacecraft. The brownish stripes in the middle correspond to the Martian surface. The picture appears in *The Martian Landscape.*

The earth globe began to turn in one or another crude model earlier, but the first truly Copernican models seem to have followed the precepts of Kepler, who directly addressed the mechanical problem of modeling his own laws. Now the cast becomes star-studded: Galileo, Huygens and Roemer all execute one or another partial version of the new cosmos—perhaps Jupiter and its four moons, or the earth moving about the sun—as clocks improve and mechanism is mastered. In the 1650's, at the Gottorp castle of Duke Frederick III of Schleswig-Holstein, there was built a hollow copper globe three meters in diameter (it can still be seen restored in Leningrad) within which one could sit with a few friends to watch the sky roll by, while for outside viewers the globe of the earth rotated slowly. The Sun King commissioned a still grander pair of such globes, and another was set up more thriftily a century later by Roger Long for students and savants at Pembroke Hall in Cambridge. It bore painted stars that were later replaced by holes pierced through the tin coating of the sphere.

The next chapters celebrate the period of the orrery, the multigeared and many-axled movements that by successive clever approximations made the cosmos of Newton clear to the gentlemen of the Enlightenment. A welcome chapter on the American orreries of David Rittenhouse and one or two chapters on the elaborations of the French craftsmen lead to the 19th century. Now showmen take control, and the orrery reaches theater size, caught in limelight. (Even then the big ceiling planetarium was not new: Ole Roemer had built one in Paris that he moved to the round tower of Copenhagen before 1700.) In these later times gear-drunk devotees made miracles of clockwork, like those of Jens Olsen in the City Hall of Copenhagen and of Auguste Vérité in the Beauvais cathedral; these outdo each other in complexities and accuracies. "Vérité had a passion for dials." There are 52 at Beauvais, and more in other examples. Much design detail is given. In our digital world even dials are lost to abstraction.

Our own epoch, dazzled by images and tired of gears, has brought an illuso-

Today's Bunker Ramo is making it a lot easier for millions of Americans to put money in the bank.

If you use a commercial bank, savings and loan association, mutual savings bank or credit union, you're probably used to standing in line.

But if you bank at one that uses Bunker Ramo's Bank Control or Thrift System 90, you probably haven't stood in line in ages.

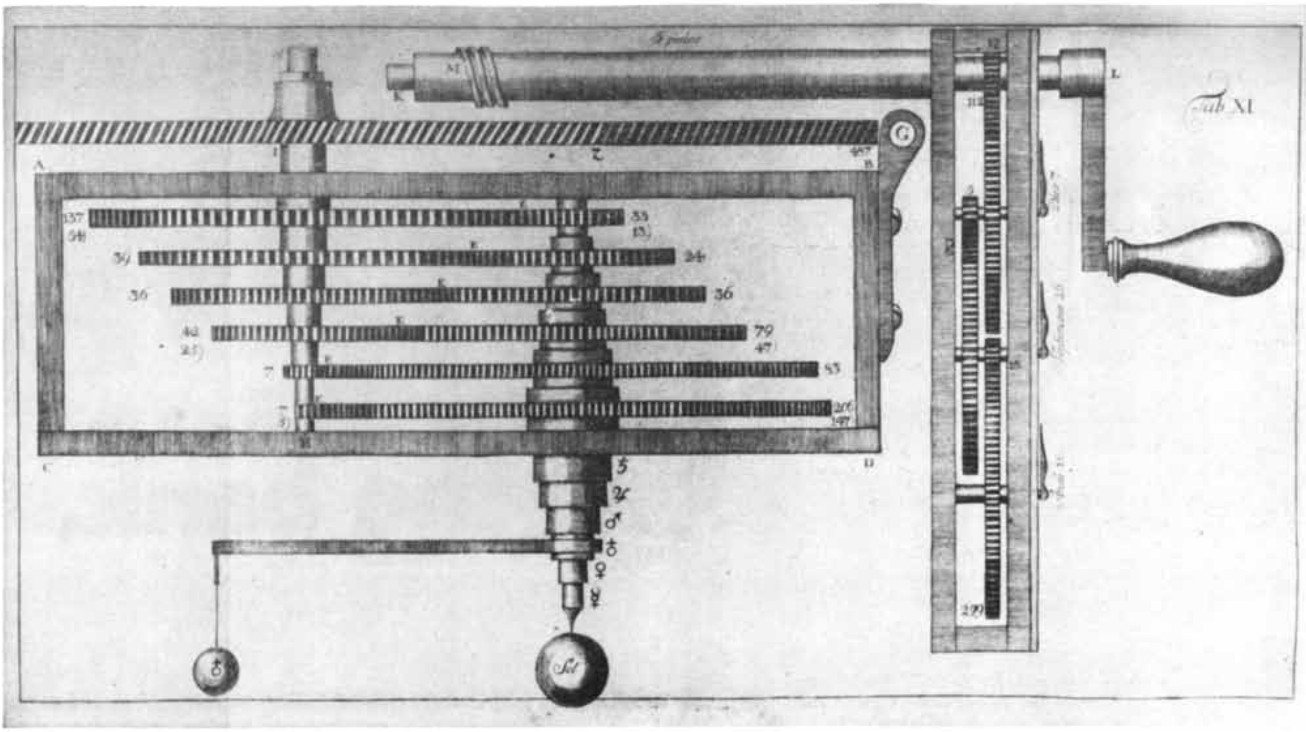
That's because Bunker Ramo's System 90 family operates in real time and provides the automated devices required for superior customer service—from the simplest of teller terminals to the most sophisticated of teller equipment to administrative terminals for management control. Because the system is modular, it provides the commercial bank or thrift institution with future growth and expansion without wasting money. Because the system completes transactions in seconds, it helps customers make their deposits or withdrawals without wasting time.

Today, Bunker Ramo is a major producer of the data terminals and minicomputers that comprise the information systems widely used by commercial banks, thrift institutions, insurance companies, and brokerage houses. In addition, through our well-known Amphenol® connectors, we have a leading position in the aerospace and aircraft industries, a strong foothold in telecommunications, and a growing role in fiber optics.

Most important, we are a company with a reservoir of high technology, a solid base of innovative products, and a strong management team.

Which is why today's Bunker Ramo is a company that's on the move. For our latest financial reports, write Bunker Ramo Corporation, Dept. FR, 900 Commerce Drive, Oak Brook, Illinois 60521.

BUNKER RAMO



MECHANISM OF A PLANETARIUM designed by the Danish astronomer Ole Roemer late in the 17th century appears in this engraving reproduced in *Geared to the Stars,* by Henry C. King in collaboration with John R. Millburn. The numeral beside each gear gives the number of its teeth. The symbols beside the nested shafts below the horizontal gears represent the planets, from the top Saturn, Jupiter, Mars, the earth, Venus and Mercury. The arm carrying the ball of the earth is shown; the other arms are omitted. At the bottom is the sun.

ry sky, projected star by star and planet by planet on a dome, with a growing diffusion of examples and a growing wealth of theater. This trend was started in 1919 by Walther Bauersfeld of the Zeiss company in Jena, as a result of a request from the Deutsches Museum in Munich. (There was a predecessor in Chicago that was not known in Jena.) The instruments of today seem more interesting in their studded complexity than the sky theater they project. This history still unrolls.

The book is sure to be a lasting reference; its apparatus, notes and bibliography are research tools and its design is attractive and workable.

MICROTUBULES, by Pierre Dustin. Springer-Verlag ($74). A hundred tiny rays issue radially from the little cell body of one aquatic organism, each ray .3 millimeter long. The old microscopists dubbed the group the sun animalcules, and Ernst Haeckel himself drew them splendidly. Wonderful enough, but under the electron microscope each of those rigid spines in cross section is resolved into complex regular twelvefold spiral tessellations of tiny protein tubes, each tube itself a helical bundle of a dozen or so protofilaments arranged in a cylinder about 24 nanometers in diameter. These are the microtubules, and they are not a special feature of the heliozoans; they are found, singly or formed into marvelous networks, in every cell of the eukaryote kingdoms. (The anucleate mammalian red blood cell is the only well-known exception; perhaps this is compensated for by the fact that the disk form of the blood platelets seems to be the result of their having a stiff peripheral hoop of microtubules.) Cilia and flagella, from protozoon to sperm cell, are bundles of microtubules; the remarkable fibrous spindles that choreograph the dance of the chromosomes in mitosis are complex arrays largely of the same modules, and every neuron is lined with very long microtubules (together with still smaller tubules of a different character).

This heroic, up-to-date, one-man review of a wide and active field of biology at the organelle level surveys the work recorded in some 2,000 papers, most of them published in the past dozen years. Here is the molecular biology not of information transfer but of structure, even of movement. The heart of the matter is the biochemical isolation of two related proteins, the alpha and beta tubulins: globular molecules some 50 angstrom units in diameter that can assemble spontaneously, in vitro and in vivo, into the long helical tubes that mark these fibrous structures. The conditions of spontaneous assembly are becoming defined. A few other molecules are needed, and the right ionic and thermal environment. Disassembly too is not hard to achieve; here biochemistry controls structure, just as it does in the tiny cylinders and boxes that coat the viruses. One can arrange for long ribbons of the stuff to wrap into giant tubes. Marvelous hybrid tubules are shown, including a bacterial flagellum smoothly extended by the assembly of more tubulin extracted from the cells of the chick brain. The full flavor of universality is here, as it is in the genetic-information stores, but marked by a still more complex set of protocols and programs.

The chapters survey the discovery, biochemistry and physiology of individual microtubules and of many more complex assemblies such as those described above, and the function of such structures in cell shape and movement, in secretion (the tubules may serve as macromolecular plumbing) and above all in the ballet of cell division. The topic has yet to match in clarity what is known of the informational side of the life of cells, surely because here we must deal with structures richer and larger than the tape and copying gear of the world of the nucleic acids, although perhaps less subtle ones.

The story of the gradual elucidation of the microtubules is a fascinating one. It really begins with the ancient medical use of the extract of the autumn crocus, colchicine. First it was an antirheumatic and then a specific for alleviation of the pain of gout, perhaps still the best. Overdose caused symptoms of poisoning, particularly dysfunction of the rapidly dividing cells of the small intestine, like that familiar in acute radiation disease. By the 1930's it was recognized (one of

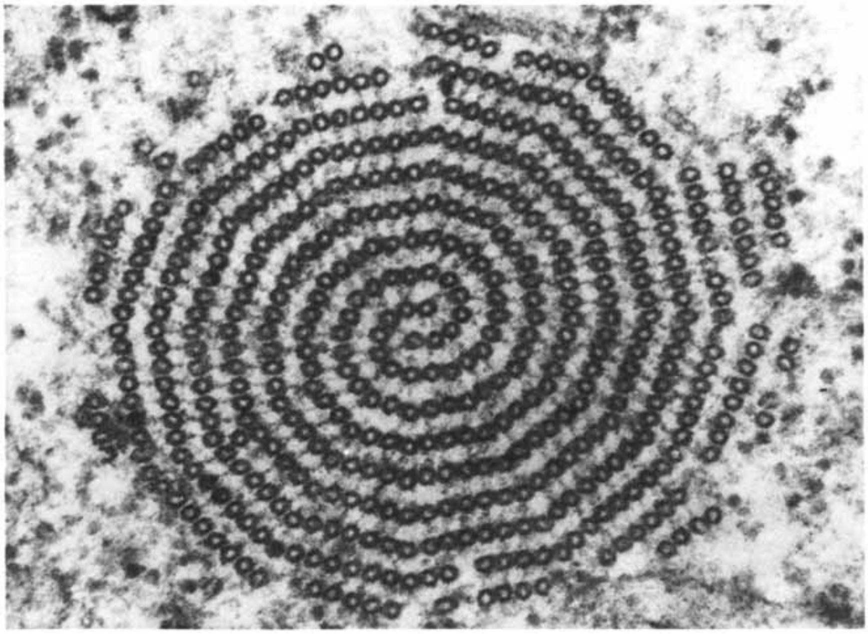
the early findings was apparently made by the author's father) that colchicine deranged the events of mitosis, and it became a powerful tool of plant genetics. The observation by E. W. Taylor in 1965 that radioactively labeled colchicine bound specifically but reversibly to the mitotic spindle was the turning point for the field. The specific tubulins were shown to be present outside of mitosis in many cells, and thus the scaffolding of the cell was laid open to study. It is still far from clear that the benefits to the gouty come from the specific "spindle poison" effects of colchicine. Other spindle poisons are now known and used, some as mitotic inhibitors in cancer chemotherapy. High pressure dissassembles the microtubules; heavy water stabilizes them and may favor assembly. There is a lot to be learned.

The Belgian author does not offer much of the quantitative geometric and thermodynamic material (still rather uncertain) on assembly and disassembly processes or much on the still speculative story of microtubule evolution. The tubulins are of ancient pedigree as proteins go, remaining, it seems, about the same in amino acid sequence over a billion years of evolution. Their rapid assembly into helical tubes—which can grow by microns in an hour—and their still faster disassembly is under delicate control in space and time, and it continues at a level beyond mere helical assembly to provide the scaffolding of the intricate tentacles of protozoa and of the octopuslike spindle apparatus of all higher cells.

It is not so much the stability of the structures as their "remarkable plasticity, demonstrated by their disassembly in a few minutes and their rapid assembly," that fits these little protein balls for the scaffolding of the cell. They serve functions as different "as ... the rays of heliozoa, the force-transmitting organelles of arthropodia ..., the flagella of the spermatozoa with all their structural varieties, the modified cilia of sensory cells, and the several forms of mitotic [microtubule].... Reality is still more fantastic than the long story which has been gathered in these pages."

GRASSES: AN IDENTIFICATION GUIDE, written and illustrated by Lauren Brown. Houghton Mifflin Company ($9.95). The rough little cylinder heads of timothy have offered good hay to fuel the horses of New England since Timothy Hanson promoted the crop in 1720; it is an English species, there called cat's tail. On the coast, everywhere north of Cape Fear the circles traced in the sand by the windblown leaves of *Ammophila* beach grass are stored among the poignant memories of most idlers along the beach. Meanwhile we all live by grass: oats, barley, rye, wheat and corn, not to mention the rice, sorghum and millet of other regions.

The subtle complexity of the family of grasses and their kin, the sedges and rushes, covers a third of the green lands of the earth. True grasslands crown the U.S. from Indiana to the Rockies. In the region bounded by the Mississippi, the Ohio and the Great Lakes and their outlet, grass abounds today. Once those were woodlands (except along the floodplains and in the salt marshes). The early settlers needed meadowlands for their

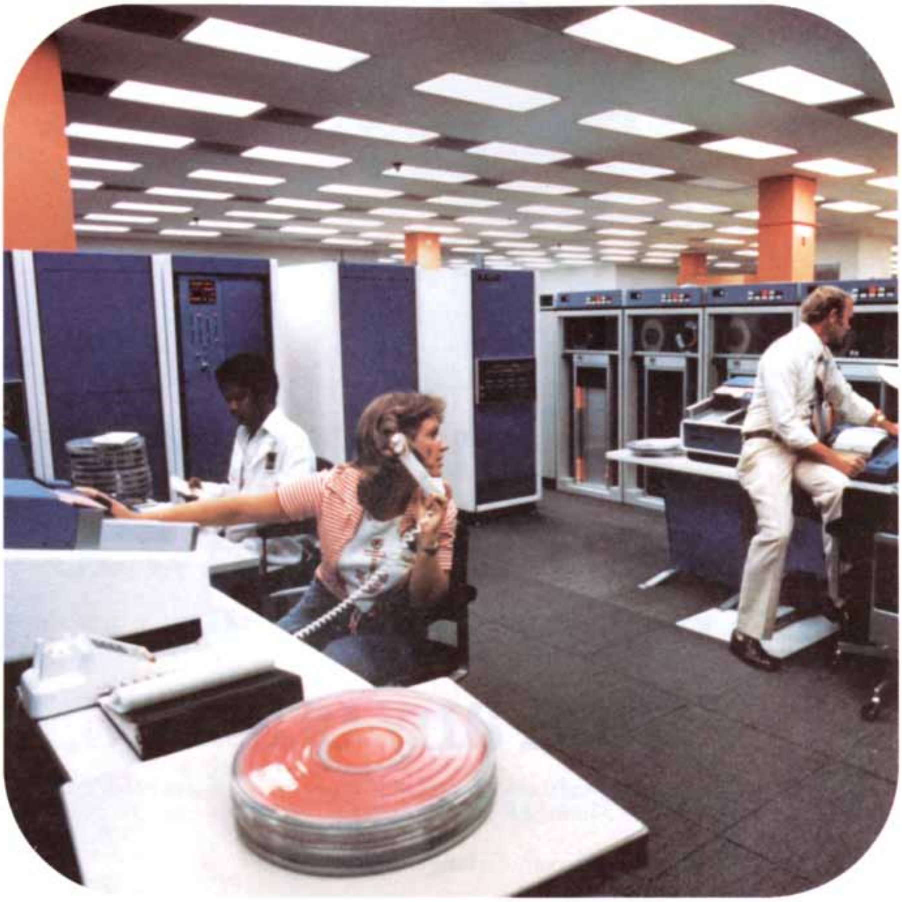


DOUBLE SPIRAL OF MICROTUBULES is seen in cross section in an electron micrograph in *Microtubules*, by Pierre Dustin. The array forms a shaft in an extension of the protozoon *Echinosphaerium nucleofilum*. The micrograph, which enlarges the structure 100,000 times, was made by L. E. Roth of University of Tennessee and Y. Shigenaka of Hiroshima University.

Harris technology on the job

In information systems,
Harris produces data processing terminals, general-purpose computers, word processing systems and supervisory control systems.

Harris offers a complete family of high-performance, general-purpose computer systems. Providing virtual memory, data base management and a multitude of languages for scientific and commercial applications.



Harris computer system at Shell Oil's Science and Information Center

Harris technology works worldwide—in communication equipment, information systems, government systems, semiconductors and printing equipment. For information, write: Harris Corporation, Melbourne, Florida 32919.

HARRIS

COMMUNICATION AND INFORMATION PROCESSING

The Pentax MX. From the people who've made more major advances in 35mm SLR's than any other company.

At Pentax, major breakthroughs are a matter of routine. In fact, the actual look of today's 35mm SLR is the direct result of Pentax photo-engineering.

Which makes the Pentax MX part of a long tradition of innovation.

We were first to mass-produce 35mm SLRs combining the pentaprism and instant-return mirror. We were first to use TTL (through-the-lens) light metering. We introduced the first Super-Multi-Coating for 35mm lenses. And we were first to offer automatic electronic exposure in a 35mm SLR.

So it's no coincidence that Pentax is first in total sales of 35mm SLRs worldwide.

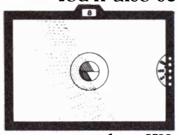
And it's no surprise, that in the Pentax MX, we've built a totally reliable system camera with features and capabilities far beyond its price range.

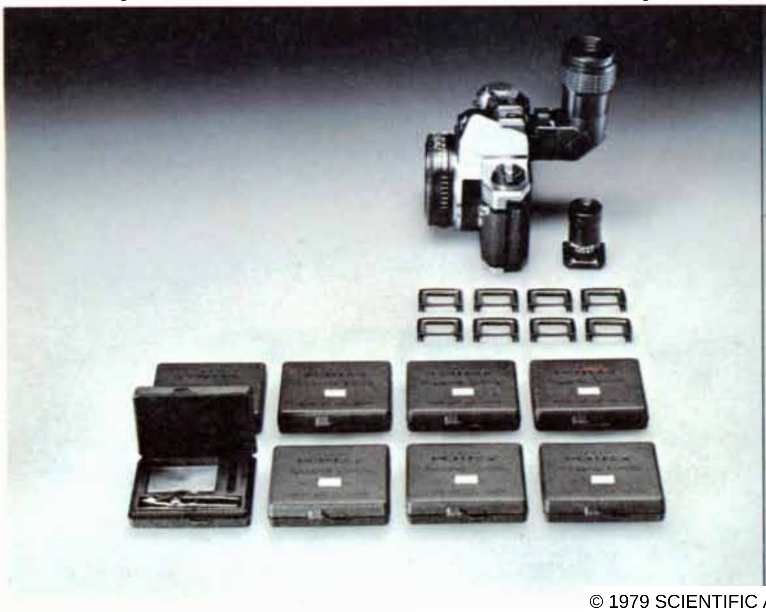
The MX was built for pros by pros to be the smallest and lightest total-system 35mm SLR in the world. It's so light, you can add either the 2-fps Winder MX or the 5-fps Motor Drive MX, and the whole system will still weigh less than other SLRs *without* a winder or a drive unit.

THE PANORAMIC VIEWFINDER

The image in the viewfinder is so big and bright it snaps right into focus regardless of lighting conditions. Inside the finder you'll see aperture setting, shutter speed, and the two adjacent shutter speeds.

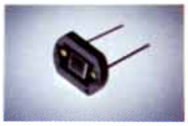
You'll also see our incredible five-dot LED exposure readout. When the green dot in the center is lit, you have perfect exposure. When one of the two yellow dots is lit, you're a half stop under- or over-exposed. And when one of the two red lights is on, you're a full stop (or more) over or under. With all that information right in the viewfinder,





you can work easily without having to move the camera away from your eye. And the solid-state LEDs are rugged. You can't jar them out of alignment, as you might with many matched needle systems.

GPDs—A BETTER WAY TO METER LIGHT



The light sensors in the MX are a major technological breakthrough in themselves, developed by Pentax in cooperation with Nippon Electric. They're called Gallium Arsenide Phosphide Photo Diodes, or GPDs. They react 1,000 times faster than CdS cells in low light situations, and they respond instantly to any radical light change, no matter how rapid. And yet, unlike SPD cells, GPDs are virtually immune to infrared light. In the future, you'll probably see GPDs in most other SLRs. At Pentax, we use them right now.

COMPACT LENSES

A lot of SLRs have been made smaller lately, but unfortunately, their lenses haven't. Pentax decided to design 18 ultra-thin, ultra-light M-series lenses for use on the MX. Of course, our regular Pentax lenses also fit the MX. So there are over 40 interchangeable Pentax lenses available, from 1,000mm telephoto to 17mm fisheye. All have the Pentax bayonet mount, which is becoming the standard of the industry. And all have Super-Multi-Coating, a process we invented and perfected.

MORE INNOVATIONS

Loading the MX is a real treat. Our "Magic Needle" takeup spool is probably the most efficient loading system you've ever used.

Also, we make nine interchangeable focusing screens which can be changed in seconds.

MOTOR DRIVE MX

When attached to your Pentax MX, the Motor Drive MX lets you fire off consecutive exposures at up to 5fps. And it's available with three different power sources.

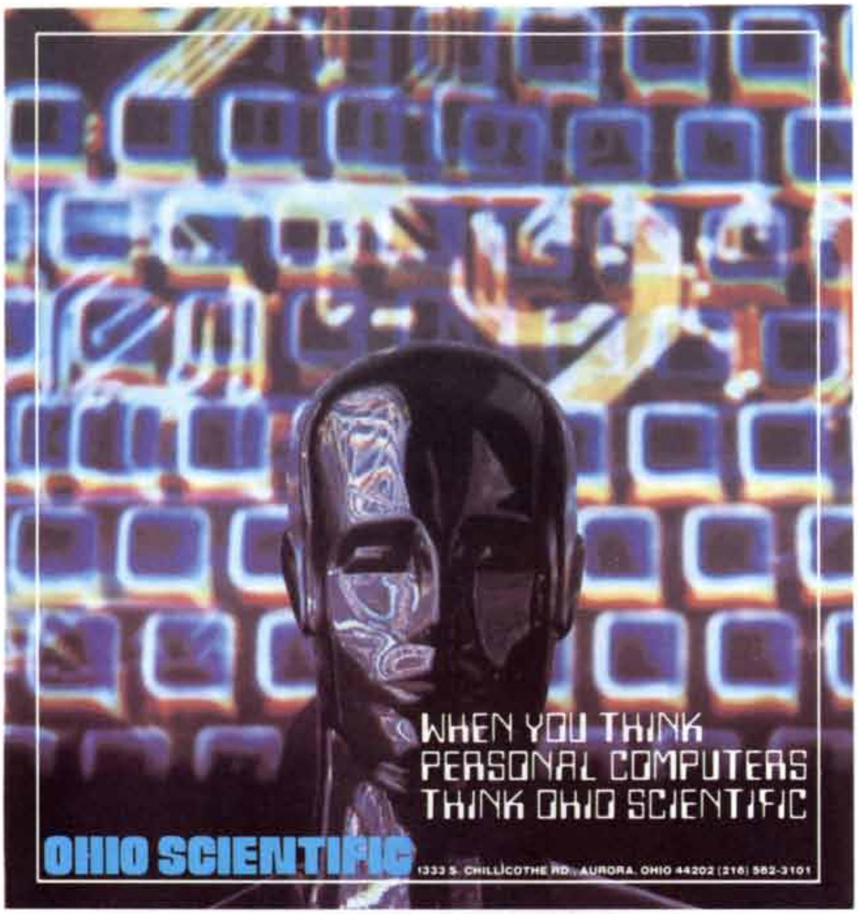
DATA BACKS AND 250 EXPOSURE BACK

The back of the MX accepts the Pentax Bulk Film Magazine and data back. If you have a lot to shoot, the Bulk Film Magazine holds 250 frames. With the Dial Data MX, your MX becomes a data camera. On each exposure, your choice of alpha/numeric data is imprinted automatically. And our Data MX camera comes complete with a built-in clock, so you can imprint the time, date and more on each frame.

All in all, the Pentax MX is a formidable camera, the heart of a superb, sophisticated, professional system. It's the result of years of meticulous research and exhaustive testing. We wouldn't have it any other way. And we don't think you would either.

For more information, write: Pentax MX, P.O. Box 2585, Littleton, CO 80161

PENTAX®

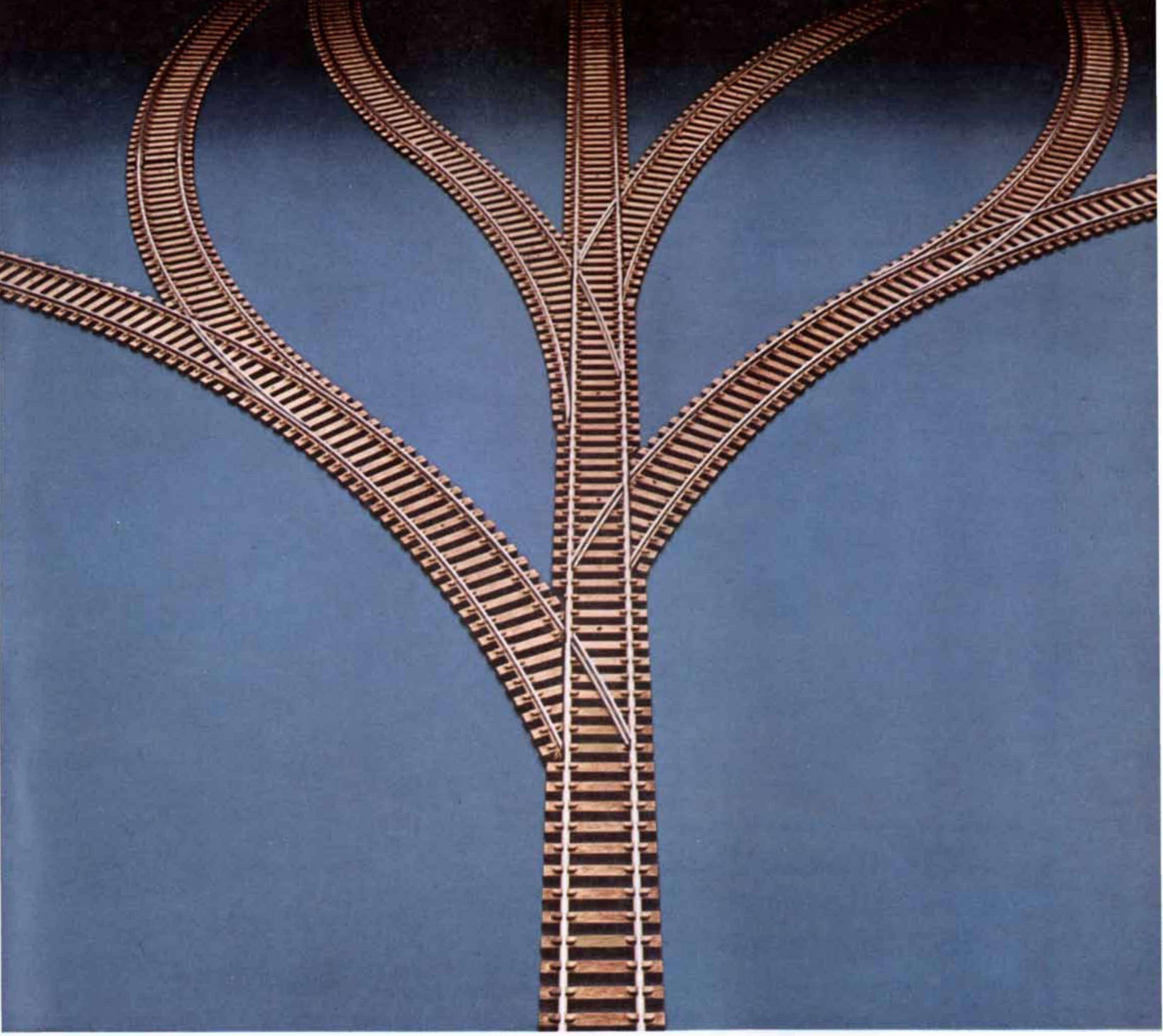


cattle and began to bring in the forage plants of their homelands. Today, with the forests diminished, lawns and wastelands all over the Northeast support the strong root mats and sturdy basal growth of the grass family. Yet these most important and most numerous of wild flowers are neglected by many who love to know in detail the showier plants of wood and field. The reason is surely one of scale. The grasses are adapted for windblown pollen, independent of animal pollinators; their flowers are small, no longer formed, fragrant and colored to wheedle some service from insect, bat or bird. And so we look uncomprehendingly at the grassy fields, unable to read the rich message.

The botanists do much better. Their method is sure but not easy: a hand lens and careful, close study of the tiny grass flowers and their intricate parts with strange names such as glume and lemma. No doubt it is a source of pleasure to come to know the grasses in this classical way, but there is no doubt either that it places a real burden on the part-time naturalist. Lauren Brown, the botanist and artist whose sharp eye and evocative drawing pen have given us this original book, offers new help. In a page or two of drawings and text for each major species she opens the way to knowing the grasses (and a few more grasslike plants) as we recognize the other wild flowers, visually by plant form, texture, color, habitat and comparisons. She does not omit the small detail where it helps, but she organizes the grassy world (including the chief crop plants) on a more familiar and less taxing basis.

In part Brown can offer success because she sensibly limits the universe under study: we read here not of thousands of species but of 135 common ones, enough to include most that are to be found in the Northeast and some of the most important ones countrywide. Her ink drawings are as beautiful as they are exact; they seem quite real on the pages of the book—more than ink. There are keys by appearance, by habitat, by common properties and by name. *Spartina patens* forms "big cowlicks in the salt marshes"; sweet vernal grass scents the air as you walk through it. *Grasses* is a long-needed aid to the Eastern outdoors. Although the task still requires effort, the guide should help those who turn to it to know the indispensable filigree of the grassy world as now they know the brighter ornaments of flowery fields. One hopes also that some readers will follow the author in one more way: she does not mow her lawn!

ALBERT EINSTEIN, THE HUMAN SIDE: NEW GLIMPSES FROM HIS ARCHIVES, selected and edited by Helen Dukas and Banesh Hoffmann. Princeton University Press ($8.95). ALBERT EINSTEIN, AUTOBIOGRAPHICAL NOTES: A

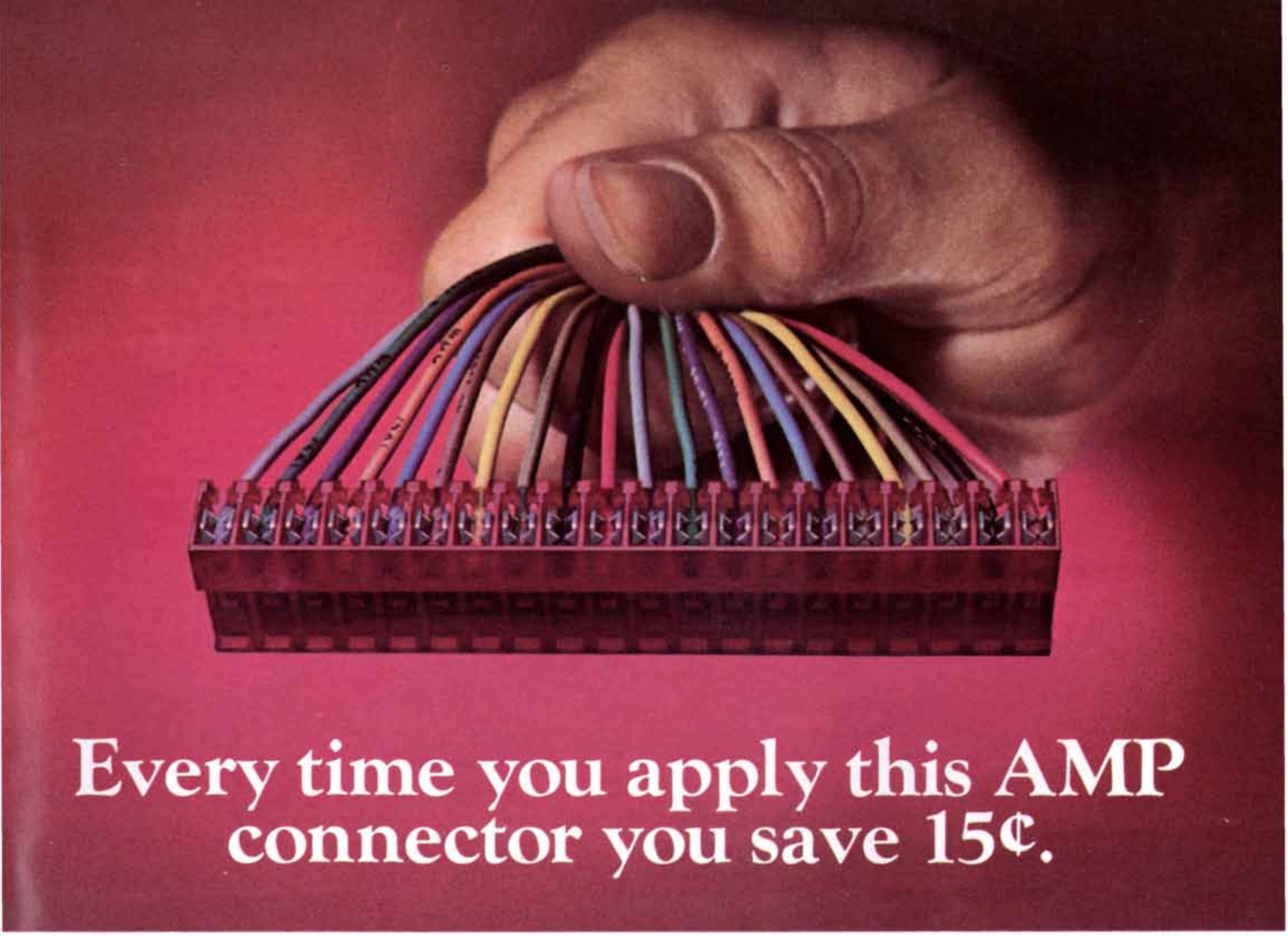


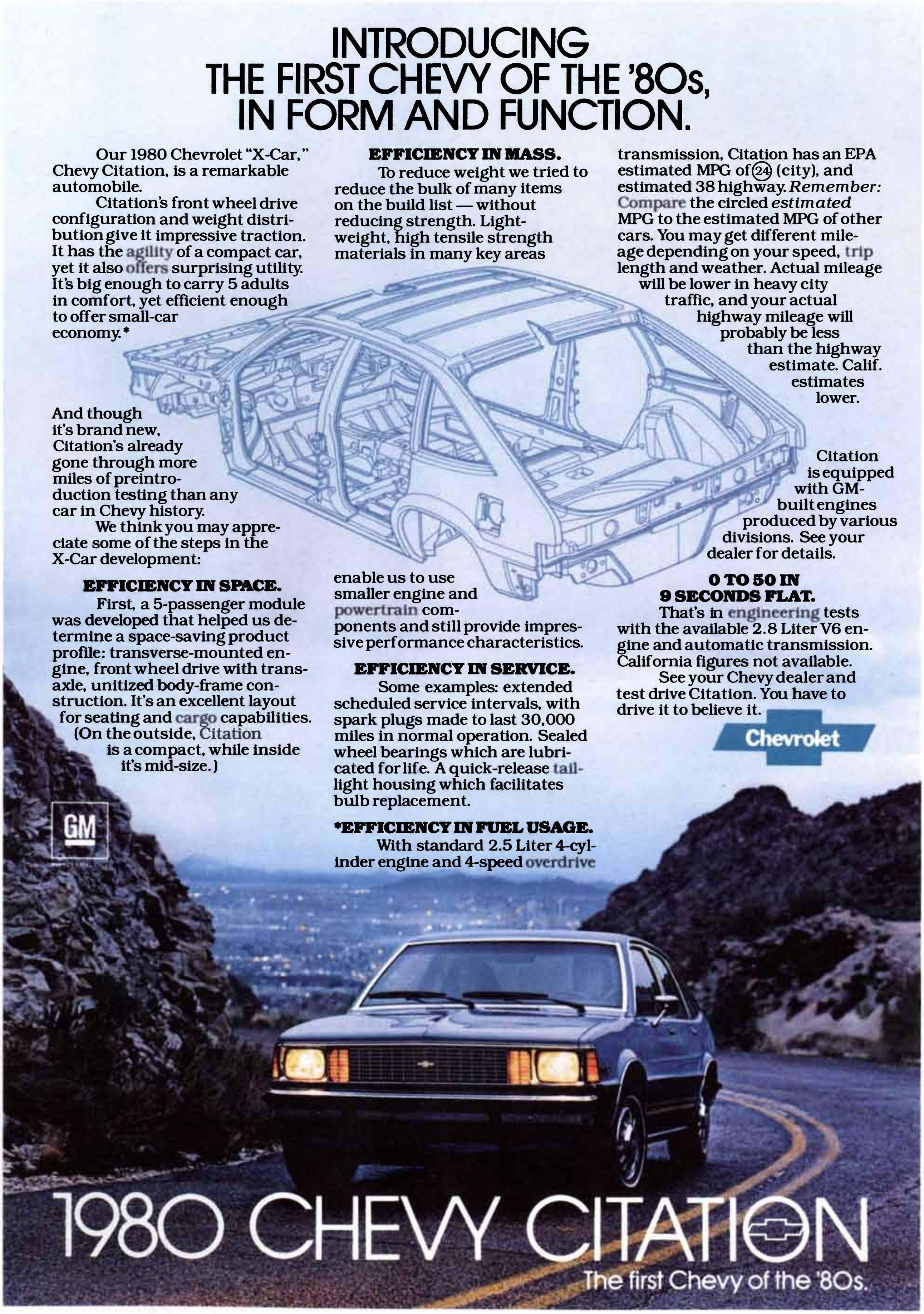
CENTENNIAL EDITION, translated and edited by Paul Arthur Schilpp. Open Court Publishing Company ($9.95). EINSTEIN'S THEORY OF RELATIVITY, by Max Born. Revised edition. Dover Publications, Inc. ($4). These monthly pages take about 10 weeks to pass from the reviewer's typewriter to nearly 700,000 printed copies. The worldwide tide of text and image that celebrated the centenary in mid-March of the birth of Albert Einstein had therefore to be marked here rather elliptically, in advance. Almost as anticlimax (surely interest has been saturated) it seems proper still to note very briefly three books that offer the greatest value for those who would even now join the feast.

Albert Einstein, the Human Side presents itself in such a modest and loving tone that it is fitting for the memory of the man it lets us hear. It is a fresh and delicious little anthology of citations from the body of Einstein's letters, journal entries and other written comment that is held in Princeton. Most of them are published here for the first time. Here is one, a "pious wish" written by the patient and generous physicist to store in a library cornerstone set down by an American publisher in 1936: "Dear Posterity," Einstein writes wryly, "If you have not become more just, more peaceful, and generally more rational than we are (or were)—why then, the Devil take you." These varied, penetrating, warm and open remarks to queens and schoolchildren, friends and antagonists, philosophers and sophomores have been sensitively chosen by two old friends of Einstein's and well translated. The German originals are included.

Autobiographical Notes is a reissue of the most important of Einstein's conscious examinations of his own mind and work. "Here I sit in order to write, at the age of sixty-seven, something like my own obituary." Written in about 1946 and first published in 1949, this 43-page essay recalls the wonder of the magnetic compass for the child of five: "Something deeply hidden had to be behind things." We read of the "holy geometry booklet" he found at 12, of the paradox of chasing the light beam that preoccupied the boy of 16, and on to the unified-field attempts. The entire inner history of his ideas is reflected in these few pages; the German original is again included.

There remains one old best buy, *Einstein's Theory of Relativity.* It is a simple but deep exposition by a gifted physicist, a lifelong friend of Einstein's. It is Max Born's introduction, for serious study but with little formal mathematics, to the physical principles of both the special and the general theory of relativity. First published in German in 1920, this clear translation was revised by Born himself in 1962, and the exposition has yet to be bettered.


Every time you apply this AMP connector you save 15¢.


INTRODUCING THE FIRST CHEVY OF THE '80s, IN FORM AND FUNCTION.
Our 1980 Chevrolet "X-Car," Chevy Citation, is a remarkable automobile.
Citation's front wheel drive configuration and weight distribution give it impressive traction. It has the agility of a compact car, yet it also offers surprising utility. It's big enough to carry 5 adults in comfort, yet efficient enough to offer small-car economy.*
And though it's brand new, Citation's already gone through more miles of preintroduction testing than any car in Chevy history.
We think you may appreciate some of the steps in the X-Car development:
EFFICIENCY IN SPACE.
First, a 5-passenger module was developed that helped us determine a space-saving product profile: transverse-mounted engine, front wheel drive with transaxle, unitized body-frame construction. It's an excellent layout for seating and cargo capabilities. (On the outside, Citation is a compact, while inside it's mid-size.)
EFFICIENCY IN MASS.
To reduce weight we tried to reduce the bulk of many items on the build list — without reducing strength. Lightweight, high tensile strength materials in many key areas enable us to use smaller engine and powertrain components and still provide impressive performance characteristics.
EFFICIENCY IN SERVICE.
Some examples: extended scheduled service intervals, with spark plugs made to last 30,000 miles in normal operation. Sealed wheel bearings which are lubricated for life. A quick-release taillight housing which facilitates bulb replacement.
*EFFICIENCY IN FUEL USAGE.
With standard 2.5 Liter 4-cylinder engine and 4-speed overdrive transmission, Citation has an EPA estimated MPG of (24) (city), and estimated 38 highway. Remember: Compare the circled estimated MPG to the estimated MPG of other cars. You may get different mileage depending on your speed, trip length and weather. Actual mileage will be lower in heavy city traffic, and your actual highway mileage will probably be less than the highway estimate. Calif. estimates lower.
Citation is equipped with GM-built engines produced by various divisions. See your dealer for details.
0 TO 50 IN 9 SECONDS FLAT.
That's in engineering tests with the available 2.8 Liter V6 engine and automatic transmission. California figures not available.
See your Chevy dealer and test drive Citation. You have to drive it to believe it.
Chevrolet
GM
1980 CHEVY CITATION
The first Chevy of the '80s.

Established 1845 SCIENTIFIC AMERICAN June 1979 Volume 240 Number 6

External Human Fertilization

The procedure whereby a new human being is conceived in a glass dish has raised issues that must be faced by investigators and by the public. The debate needs to be informed by known facts of human development

by Clifford Grobstein

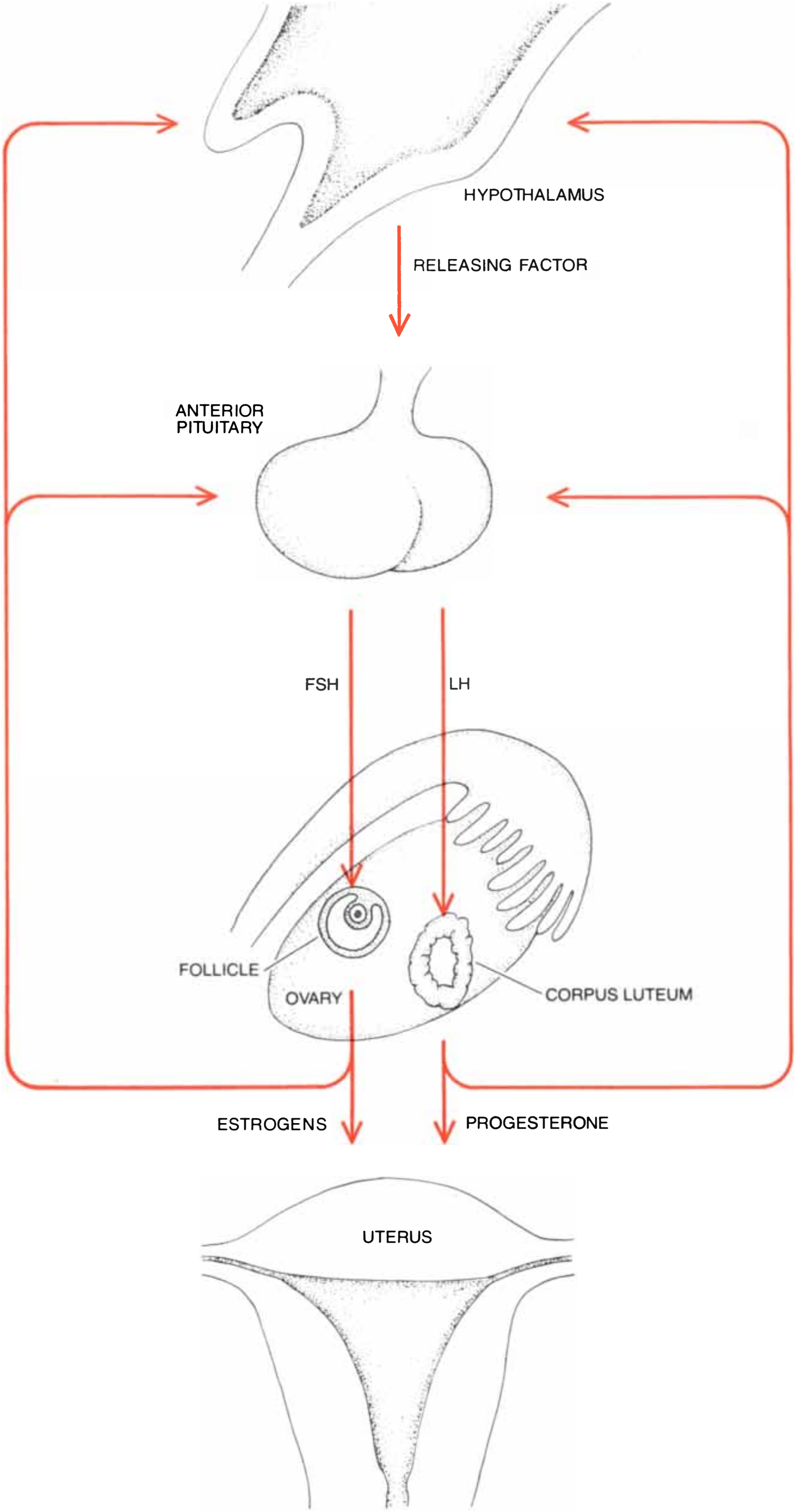
Louise Joy Brown (five pounds 12 ounces) was born last July 25 to Mr. and Mrs. Gilbert John Brown of Bristol, England. The birth in Oldham General Hospital near Manchester was noted around the world as marking the beginning of a new era in human reproduction. It was the first recorded case in which a human egg had been taken from a woman's ovary, had been fertilized externally and then implanted in the woman's uterus, and had developed to term. The success of the procedure gave couples suffering from several forms of sterility hope that they might have a child of their own. Such a clear practical benefit of biomedical research is normally an occasion for general satisfaction. This particular advance, however, gave rise to mixed reactions, and in some quarters to grave concern.

External human fertilization, involving the laboratory manipulation of cells with the potential to become a human being, raises issues that engage those associated with abortion and with experiments involving human beings, both of which have been subjects of controversy in the U.S. for some years. When Louise Joy Brown was born in England, research on external fertilization had been suspended for about five years in the U.S. while first Congress, then a national commission on human experimentation and finally a new Ethics Advisory Board established to counsel the Secretary of Health, Education, and Welfare considered the propriety of various forms of "fetal" research. In March of this year the board concluded that research on external fertilization and subsequent implantation "would be acceptable from an ethical standpoint." It referred back to the National Institutes of Health, for reconsideration in the light of various expressed reservations, a long-standing proposal for laboratory investigation of human fertilization (but not for implantation) by Pierre Soupart of the Vanderbilt University School of Medicine. Several clinical groups are reported to be preparing to provide external fertilization and implantation for infertile women.

Debate on the procedure certainly will continue. The birth of a human being conceived in a glass dish is enough of a novelty, with social and ethical implications as well as biological ones, to raise questions that need to be faced by the biomedical community and the public. One set of questions has to do with whether or not external fertilization has been adequately tested and shown to be safe enough to warrant its application as a therapy for infertility. Even if the procedure is accepted as a valid one with wholly beneficial results, does it represent the crossing of a kind of biological and ethical Rubicon into uncharted territory? What further possibilities does it open up in the manipulation by human beings of human development, and are those possibilities desirable ones? This article will outline the scientific basis for the hopes and the fears attending this major new development and consider some of the policy issues it raises.

The fertilization of a human egg in vitro (in glass, literally) appears first to have been accomplished in England in 1969 by the physiologists R. G. Edwards and B. D. Bavister and the gynecologist P. C. Steptoe. It was Edwards and Steptoe who went on to develop the technique and apply it to achieve the birth last July and then a second successful birth in January of this year. Research on external fertilization in nonhuman species, primarily marine animals, dates back, however, to 1893. Two decades ago convincing evidence was presented for the successful external fertilization of rabbit eggs. Since then the process has been established in half a dozen other mammalian species, yielding considerable information that was applied to develop the technique in human beings. It was the growing understanding, in the past 25 years or so, of the complex processes by which internal fertilization and development are normally accomplished in mammals and specifically in human beings that made external fertilization possible.

A detailed understanding of the interrelated hormonal processes that control the female reproductive cycle has been of particular importance. The major anatomical elements involved in the cycle are the hypothalamus near the base of the brain, the pituitary gland just below it, the ovaries, the oviduct, the uterus and the vagina. In each turn of the usual human female reproductive cycle one egg matures, is discharged from the ovary and passes into the expanded end of the oviduct, down which it moves to the uterus. In the absence of insemination the egg disintegrates and the cycle is repeated roughly at monthly intervals. If insemination has occurred at the right time in the cycle, the egg may meet sperm in the oviduct and may be fertilized. The fertilized egg then moves along the oviduct, dividing as it goes. The egg usually reaches the uterus about three days later as a morula, a spherical cluster of eight or more cells. The division of the cells continues in the uterus to produce a blastocyst, a hollow ball with a group of cells, the inner cell mass, at one pole. The blastocyst adheres to the endometrium, the lining of the uterus, and a complex implantation process embeds the blastocyst in the uterine wall, where development continues. The inner cell mass becomes the embryo; the outer cells of the blastocyst, together with neighboring uterine



INTERRELATED HORMONE ACTIVITY controls the female reproductive cycle. A releasing factor secreted by the hypothalamus at the base of the brain stimulates the adjacent anterior pituitary gland to secrete follicle-stimulating hormone (FSH) and luteinizing hormone (LH) into the circulation. The FSH causes one or more follicles in the ovary to grow and to produce estrogens, notably estradiol; the LH triggers ovulation and transformation of the follicle into a corpus luteum, which secretes progesterone. The estrogens and progesterone circulate to the uterus and prepare the uterine wall for the implantation of the egg; the ovarian hormones also feed back to inhibit or stimulate the secretions of the hypothalamus and the pituitary.

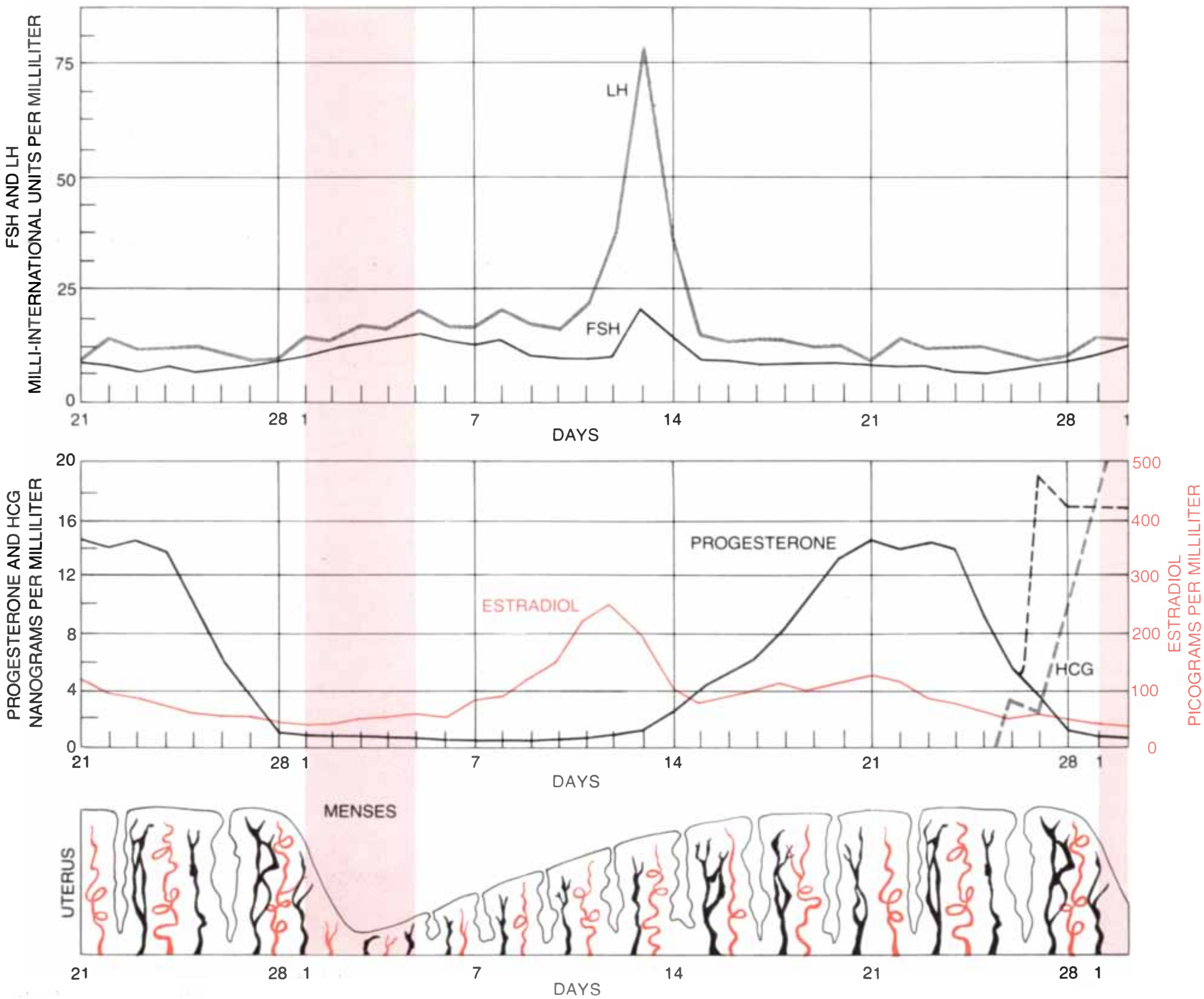
cells and blood vessels, form the placenta, the organ that provides for the exchange of materials between the later embryonic circulation and the maternal circulation. With implantation, pregnancy may be said to be established.

What organizes these events is the interplay of hormones produced in the hypothalamus, the pituitary, the ovaries and the placenta. The single egg discharged each month from one or the other ovary matures in a follicle made up of many cells. The follicle grows in size and in number of cells during the early part of the cycle under the influence of the follicle-stimulating hormone (FSH) secreted by the pituitary. As the follicle grows, its cells secrete the estrogenic hormones, chief among them estradiol, that act on both the pituitary and the uterus. The feedback to the pituitary reduces the output of FSH; the effect on the uterus is to stimulate the growth of the uterine wall and its glands.

As the follicle approaches full size, FSH output decreases and a second pituitary hormone, luteinizing hormone (LH), becomes predominant. LH completes the maturation of the egg and triggers the rupture of the follicle, discharging the egg into the oviduct. It also causes the residual follicle cells to be transformed into the corpus luteum, which secretes a second ovarian hormone, progesterone, that acts on the uterine wall to induce further glandular development and to sensitize the reaction of the uterus to contact with the blastocyst cells. In the absence of pregnancy (if a blastocyst does not arrive in the uterus) a new set of ovarian follicles begin to grow, and the corpus luteum derived from the ovulated follicle begins to decline. The resulting change in the estradiol-progesterone balance precipitates degenerative changes in the uterine wall, and menstruation begins.

If, however, an egg has been fertilized and a blastocyst arrives in the uterus, the outer cells of the blastocyst secrete human chorionic gonadotropin (HCG), a hormone that has the same effect as pituitary LH. The corpus luteum in the ovary is thereby maintained and continues to secrete progesterone; no new follicles grow in the ovary and menstruation does not ensue. In about the fifth week the placenta takes over, secreting progesterone to maintain pregnancy.

These hormonal effects controlling the menstrual cycle have several implications for the success of external fertilization. The administration of FSH beginning soon after menstruation can lead to the maturation of several follicles rather than of a single one, resulting in "superovulation" and multiple births. The increased frequency of production of human quintuplets and even sextuplets in recent years reflects the more frequent administration of FSH to stimulate the production of eggs in treating



BLOOD LEVELS OF HORMONES that govern the reproductive cycle are charted along with changes (*bottom*) in the thickness of the lining of the uterus and in the proliferation of its blood vessels and glands. A slow, early rise in the FSH level causes the ovaries to secrete estradiol, which promotes thickening and vascularization of the endometrium. An LH peak brings on ovulation and formation of a corpus luteum, which secretes progesterone, further preparing the endometrium for implantation. In the absence of pregnancy, progesterone secretion falls and the endometrium is sloughed off. If an egg is fertilized, however, and reaches the blastocyst stage, first the blastocyst and then the placenta secrete HCG, a hormone similar to LH. The HCG (*broken gray curve*) causes the corpus luteum to continue secreting progesterone (*broken black curve*), as is indicated by data from Christopher A. Adejuwon of the Population Council.

female infertility due to reduced egg production.

The technique can also increase the availability of eggs for in vitro fertilization. The eggs are recovered from the ovary by laparoscopy, a minimally invasive procedure developed for diagnostic purposes. A tiny telescope and a fiber-optical illuminator are introduced through a very small incision in the region of the navel; the ovary and any mature follicles can be examined and eggs can be removed by aspiration from the follicles. There is considerable evidence that only about one in four of the eggs ovulated and fertilized in a healthy female yield normal offspring carried to term, and so it is theoretically an advantage to obtain multiple eggs for studies of in vitro fertilization.

Some patients have therefore been treated with FSH before laparoscopy to increase the number of eggs obtained during a single procedure. It may also be desirable to treat the patient with LH to ensure that the eggs are fully mature. During the development and testing of the procedure the induction of superovulation could have the additional advantage of producing a number of eggs for study, and particularly for the development of improved techniques for maintaining the eggs externally. External maintenance is no trivial problem. The egg is normally suspended in the complex biological fluids that are present in the oviduct and the uterus. Initially there was little information about which constituents of such fluids were essential for continued health of the eggs. It took experience with the culture of other mammalian eggs and cells in general to get a basis for formulating the requirements for human eggs.

A similar problem had to be faced with the sperm, the other component of normal fertilization. The sperm are obtained in large numbers from ejaculated semen. An important obstacle had to be dealt with, however, before external human fertilization could be accomplished. In the initial ejaculate the sperm are not prepared to penetrate the egg. They must first undergo capacitation, a process that normally is induced by conditions in the female genital tract. These conditions must be simulated externally if the sperm are to be in the proper state to penetrate the egg. Taking into account the requirements of both the egg and the sperm, an artificial environment appropriate for their union had to be carefully defined, and the adequacy of the environment had to be validated by proving that a sperm enters an undamaged egg and that both the egg and the sperm then function normally.

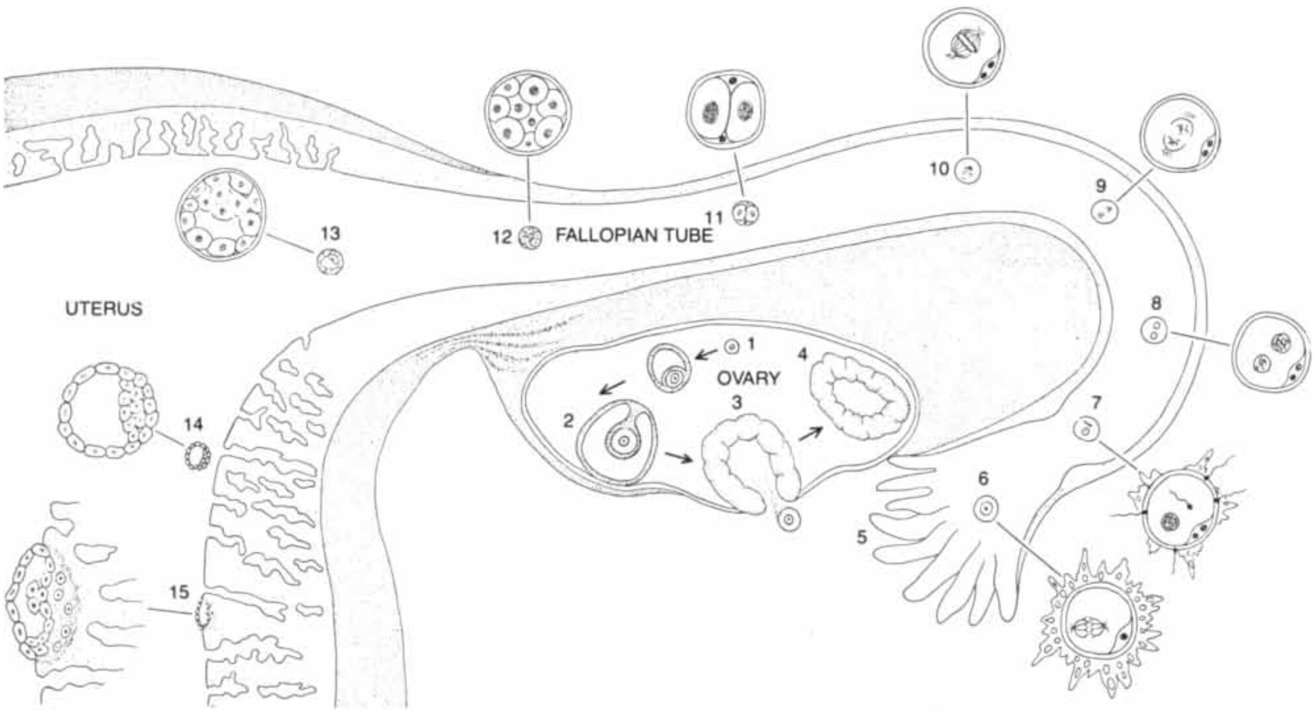
Although there are intermediate steps in the validation, the final test is the normal development of the zygote, or fertilized egg. In the case of eggs that normally develop externally, for example those of starfish in seawater or of frogs in pond water, this is no great obstacle, but for eggs that normally develop internally the problem becomes difficult. Two options are available: the return of the egg to its normal internal environment or the development of a suitable external environment. As a therapeutic procedure in human beings there is no question but that the first option is biologically simpler and currently far more acceptable socially.

Given the close hormonal control of the normal events in ovulation, fertilization, oviductal transport and implantation of the egg, however, it is clear that if implantation of an externally fertilized zygote is to be successful, the state of the zygote must be closely matched to the state of the uterine wall. If, for example, an egg that is fertilized after removal from a follicle develops more slowly outside the oviduct than within it, the egg might be introduced into the uterus of the donor in too premature a state to stimulate an implantation reaction, or it might be introduced too late because the uterus is already past its optimal time for the reaction. This was one of the variables Steptoe and Edwards had to consider carefully, and the full details of its management have yet to be published. The probability of effective matching can theoretically be influenced by administering suitable hormones to the mother, but Steptoe has indicated that this was not done in their successful cases because unusual hormone levels had adverse effects.

External fertilization followed by implantation, as reported by Steptoe and Edwards, is based on a confluence of new understandings: of reproductive biology, of hormonal effects, of cell (and particularly nuclear) function and of the maintenance of cells, organs and even entire embryos under artificial conditions. The procedure is not an isolated trick ingeniously conceived by two men to meet an individual problem. Rather it is a specific realization—a particular application—of new knowledge accumulated in many investigations by many people over many years. If human in vitro fertilization were to be regulated or even prohibited, the growing capability for this procedure and further developments would not thereby be erased.

This is the crux of the policy questions that are raised. Secretary of Health, Education, and Welfare Joseph A. Califano, Jr., cited the major questions in a memorandum to the ethics board last September. "Can techniques of in vitro fertilization and transplantation of the embryo damage the resulting fetus and lead to abnormal children? Will this research lead to selective breeding, to attempts to control the genetic makeup of offspring or to the use of 'surrogate parents,' where, for example, rich women might pay poor women to carry their children?"

To take up the first of Secretary Califano's questions, there is no doubt that when any new technique is first applied in human beings, it carries some element of risk. In the present instance the medical judgment is that there is only a slight risk to the mother and a risk of somewhat less certain magnitude to the fetus. The mother is exposed to the risks of any hormonal treatment, of laparoscopy and of transplantation of the embryo into her uterus and to the normal risks of pregnancy. The hormonal and surgical procedures are regarded as being relatively minor. Perhaps the greatest risk (although it is not a large one) is the pregnancy itself, and that risk presumably is understood by the subject and is taken willingly.



INTERNAL FERTILIZATION of the human egg is one event in a complex process that begins when a primary follicle in the ovary (*1*) develops to become a mature follicle (*2*), which ruptures, releasing the egg (*3*). The follicle becomes a corpus luteum (*4*). (Follicular development is arbitrarily shown at different sites in the ovary for clarity.) The ovulated egg is swept from the surface of the ovary by the open end of the oviduct, or fallopian tube (*5*). The egg has just completed meiosis, a division whereby its chromosomes are reduced to half the normal complement and a polar body is extruded to the margin of the egg; a second meiosis has begun (*6*). A sperm penetrates the egg cytoplasm, activating the egg to complete the second meiosis, extrude a second polar body and form the female pronucleus (*7*). The sperm head swells to form a male pronucleus (*8*) and the two pronuclei fuse, mingling male and female chromosomes (*9*). The chromosomes replicate and divide (*10*) and the fertilized egg, or zygote, undergoes cleavage (*11*). Successive cleavages as the zygote moves along produce a morula (*12*), an early blastocyst (*13*) and, after some $4\frac{1}{2}$ days, a late blastocyst with an inner cell mass, or embryoblast, which gives rise to the embryo, and an outer trophoblast (*14*). On the sixth or seventh day the blastocyst is implanted in the uterine wall (*15*).

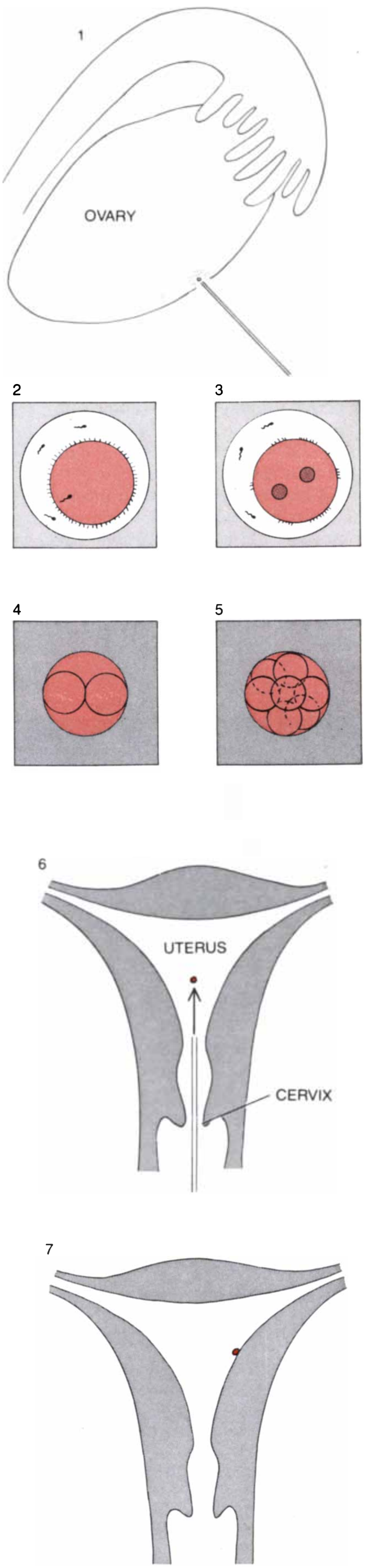
The risk to the fetus is at present more difficult to assess, and it must be evaluated in the light of the high risk to which any developing egg is exposed in the course of natural fertilization and implantation. Of 100 human eggs exposed to sperm it is estimated that only 84 are fertilized and only 69 are implanted. A week later only 42 of the 100 are estimated to be still alive; 37 of them survive to the sixth week of pregnancy and only 31 survive to birth. The normal process of internal fertilization and implantation, then, involves a loss of about 70 percent—and of the surviving newborn infants a few percent will have developmental abnormalities of varying seriousness. This high degree of embryonic loss and abnormality is thought to be largely the result of chromosomal aberrations, produced primarily during the formation of eggs and sperm and to a lesser degree during fertilization itself. The natural risk provides a complicated background against which to measure the consequences of in vitro fertilization followed by implantation.

The two reportedly successful births do not eliminate concern that the procedure may not be safe for the fetus, although they do indicate that some of the worst fears—of grossly abnormal infants—at least are not realized in every instance. The intrauterine progress of the developing fetuses was reportedly followed carefully by means of several standard monitoring procedures, which are said to have revealed no evidence of abnormality at any stage. And yet there are unanswered questions. Steptoe said last winter that a total of 79 women had entered on the treatment since he and his colleague began clinical trials; 68 underwent laparoscopy, of whom 44 yielded eggs at the correct stage of maturity. Fertilization and subsequent cleavage of the egg were successful in 32 cases; four women became pregnant, of whom two carried the fetus to term and two miscarried. Little information is available on the causes or consequences of the failures and on how their frequency and nature compare with the pattern in normal reproduction. Moreover, the successful products of the procedure are still in their infancy, and it is possible that some long-term effects may yet be manifested. Finally, a mere four pregnancies are not enough on which to base a statistical conclusion about the level of risk to the developing fetus.

A conservative approach therefore justifies caution, as in the case of any innovative therapy. At the moment there is some difference of opinion among experts. Steptoe and Edwards clearly are convinced that previous experience with both animals and human beings justified clinical trials.

Other experts are not yet fully persuaded. The Soupart application was for an investigation explicitly aimed at gaining more information on the process of fertilization, implying some residual uncertainty. At the first board meeting to consider the application last September, Luigi Mastroianni, Jr., of the University of Pennsylvania School of Medicine argued for more research on fertilization of human eggs on the ground that "the risk of returning an in vitro fertilized [human] embryo to the uterus must be scrutinized with great care." He noted that whereas animal tests are a prerequisite, they do not "completely guarantee that the procedures would work equally well in the human." In particular, he held that the animal work is not as yet statistically valid on the question of the normality of offspring; it is so far limited to only two laboratory species, not including a primate. He called for "additional extensive studies in laboratory animals" and with human eggs before therapy is initiated on human beings in the U.S.

The safety issue for the fetus, therefore, is not fully resolved and can only be further assessed through additional testing and experience. The other questions raised by Secretary Califano are no better resolved, and they require analysis of a very different kind. They are what Sissela Bok, a member of the review board, has called the "slippery slope" questions, that is, questions about developments that may tend more or less inevitably to follow the newly won capability for performing human external fertilization. There are several possibilities: (1) fertilization involving eggs and sperm not obtained from married couples; (2) manipulation other than simple implantation of the preimplantation embryo; (3) transplantation of embryos into the uterus of a woman other than the one who provided the egg; (4) ectogenesis, or maintenance of embryos in vitro beyond the implantation stage, and (5) transplantation of embryo parts into other individuals. Let us consider these questions with special reference to their immediacy, rationale and likely consequences.



EXTERNAL-FERTILIZATION procedure developed by R. G. Edwards of the University of Cambridge and P. C. Steptoe has not been published in detail; the procedure outlined here schematically is based on reports of talks and interviews given by Steptoe. When a rise in the LH level signals imminent ovulation, an egg is removed from a mature follicle by laparoscopy (*1*), washed through a culture medium and placed in a droplet containing the husband's sperm (*2*). Fertilization is revealed by the appearance of the two pronuclei (*3*). The fertilized egg is transferred to a different culture medium and is observed as it cleaves (*4*). At either the eight-cell (*5*) or the 16-cell stage the cellular aggregate is introduced, by means of a thin cannula, into uterus (*6*), where, if all goes well, it implants (*7*).

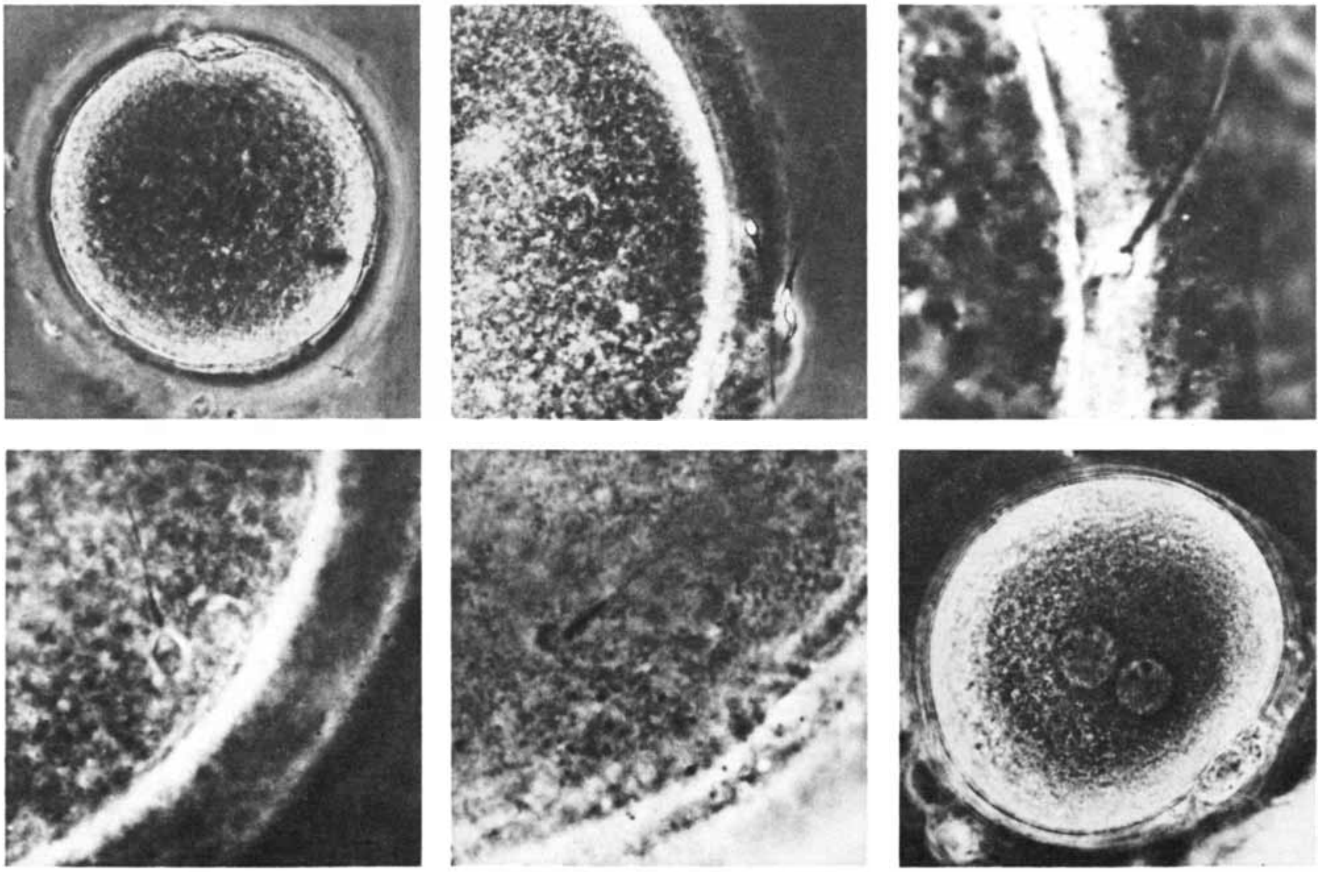
With reference to nonspousal fertilization, any healthy source of human sperm is likely to be as effective as any other in fertilizing the egg and providing a preimplantation embryo. The biological consequence would be the same as artificial insemination with sperm from someone other than the husband, which is already widely practiced as a remedy for male infertility. The ethical and legal aspects should also be very similar as long as the fertilized egg is implanted in the woman who supplied the egg. The procedure would indeed offer an opportunity for more selective breeding if donor sperm were provided by "desired" sources, but the opportunity is no greater than it is in the case of artificial insemination, and the technical difficulties are greater. (There is also the possibility of less selective breeding, which would be the case if sperm sources were deliberately randomized to ensure anonymity.) The immediate beneficiaries of nonspousal external fertilization would be women with blocked oviducts who do not have a fertile husband; the procedure would enable such a woman to have a child that is biologically at least half her own.

With reference to the second question, external fertilization does provide a technical first step toward the manipulation of early human development. Observation of the early stages of human development has been very limited by the great scarcity of suitable material. Controlled experimentation has been confined largely to external fertilization and the brief maintenance that is necessary to bring the zygote to the proper stage for implantation. With this success, however, Secretary Califano's question about the possibility of "attempts to control the genetic makeup of offspring" becomes at least pertinent.

Genetic intervention is one experimental possibility that comes to mind immediately because of recent discussion of recombinant-DNA techniques for changing the content of the genetic material and the theoretical possibility of "cloning" human beings by introducing the full complement of one person's chromosomes into an enucleated egg. If suitable means are developed for the transfer of foreign DNA into human cells, or if techniques are evolved for introducing nuclei into enucleated mammalian eggs, the Edwards-Steptoe procedure would provide a gateway for applying those techniques to the early stages of human development. The prior steps still to be taken are not, however, simple technically, and human eggs, given all the ethical and logistic problems attendant on their use, are hardly favorable material on which to undertake the necessary studies.

Nevertheless, there have been recent early reports of successful transfers of recombinant DNA into cultured mammalian cells and of the renucleation of a human ovum with the 46-chromosome nucleus of a sperm-cell precursor. Both maneuvers are conceivably applicable to intervention in the human genetic constitution. The window opened by external fertilization in the formerly black box surrounding early human development is therefore a relevant step, even if not a long one, toward control of "the genetic makeup of offspring." Yet the new potential in this direction is probably more usefully studied in nonhuman species, particularly in applications to animal breeding.

The third question has to do with surrogate motherhood: the transplantation of a fertilized egg derived from one



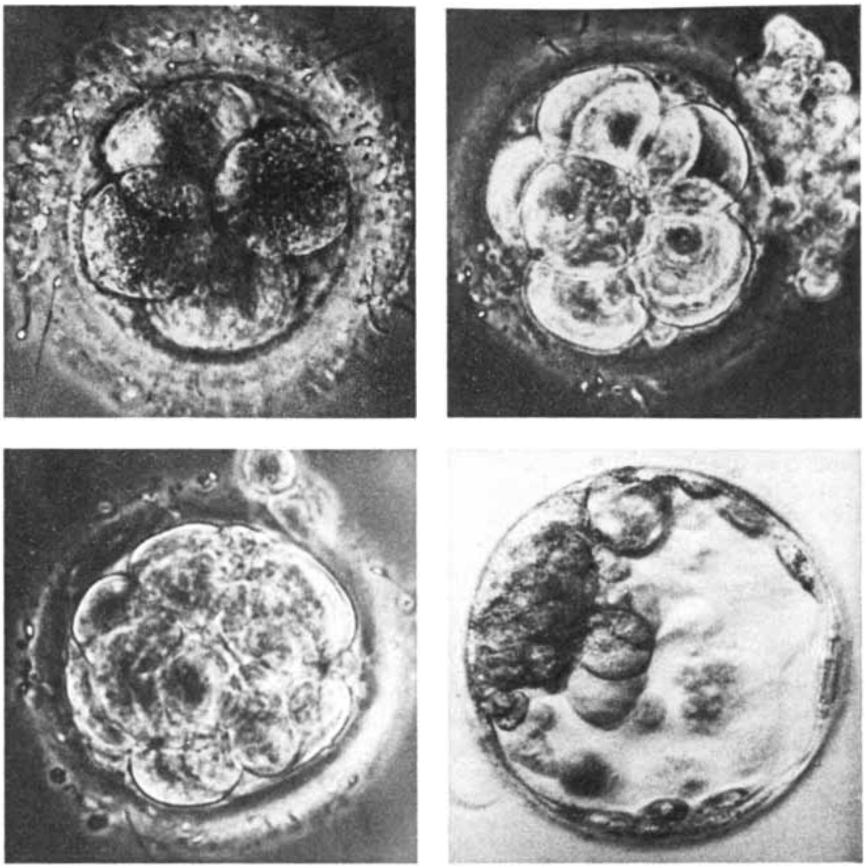
HUMAN EGG is fertilized in a glass laboratory dish in these micrographs made by Edwards. The unfertilized egg, with its protective glycoprotein layer, the zona pellucida, is about .2 millimeter in diameter (*top left*); the polar body extruded in a meiotic division just before ovulation is at the egg's top margin. Segments of the egg are shown, first with two spermatozoa penetrating the zona pellucida (*top middle*) and then with one sperm in the perivitelline space (*top right*). A sperm is shown inside the cytoplasm (*bottom left*). A sperm mid-piece is seen in the cytoplasm of an egg fixed and stained after fertilization began; the sperm head is beginning to swell to form a male pronucleus (*bottom middle*). The male and female pronuclei, which will fuse before the first cleavage, are seen in the living zygote (*bottom right*).

woman into the uterus of another woman. This possibility has already been realized in other mammalian species and there appear to be few theoretical or technical obstacles to accomplishing it in human beings. The technical problem is simply one of matching the developmental stage of the foreign embryo to the reproductive stage of the recipient. The developmental stage has been controlled in animal embryos by freezing them, which appears to have no serious ill effect. The appropriate receptivity of the uterus in a woman can be predicted by following her menstrual cycle, and it can be controlled in some measure by the administration of hormones.

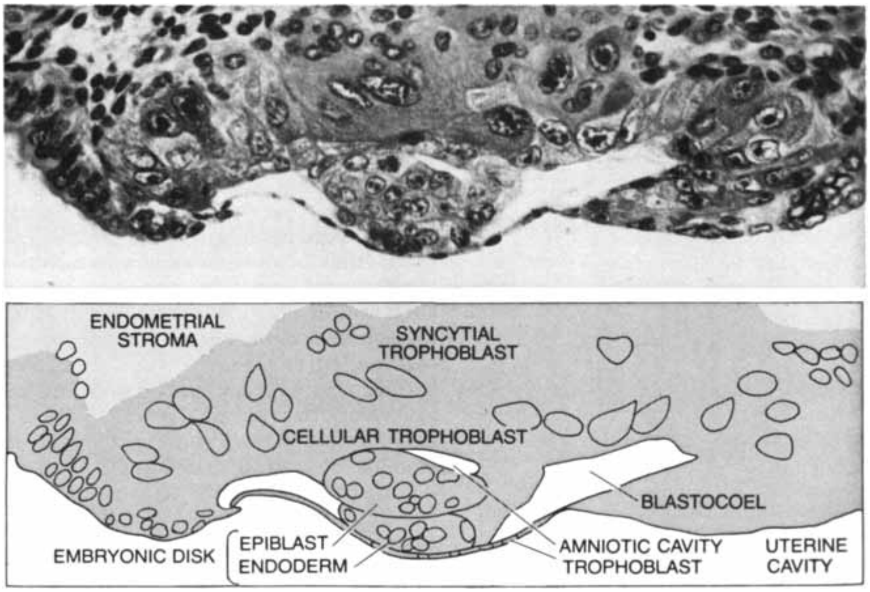
If a favorable ethical and social consensus develops, the role of wet nurse (which is accepted in many cultures) could probably be expanded fairly rapidly to include surrogate childbearing. The social, emotional and ethical aspects of such a practice clearly call for careful assessment before research or trials in this direction proceed. Consideration should be given not only to the practical value of the procedure but also to the fact that there is probably much to be learned about the genesis of important human characteristics through the study of offspring subjected to genetically unrelated maternal environments.

As for the fourth question, on the continued external development of human embryos beyond the preimplantation stage, research on ectogenesis is already in process in other mammalian species and limited success has been reported. The motivation for such animal experimentation is improved understanding of early mammalian development, both for itself and for its relevance to human development. Much can be learned about the role of genetic factors and about developmental processes themselves from animal studies, by monitoring whole mammalian embryos and their parts as they develop in external isolation. There are some things, however, such as specific human drug sensitivities and higher brain functions, that can be understood only in part from animal studies, and important information might be gained by studying human embryos maintained in external culture beyond the implantation stage. There are equally important reservations, which I shall discuss below.

Whatever conclusions are reached concerning research on early embryos, it should be emphasized that there are many large practical obstacles to any attempt to prolong early human development externally all the way to the stage of independent viability and thus actually eliminate intrauterine development from the human life cycle. Such true "test-tube babies" and the human hatcheries envisioned in Aldous Huxley's *Brave New World* are not feasible in the current state of knowledge, nor are they known scientific objectives. Wheth-



SUCCESSIVE STAGES in the cleavage of human eggs fertilized in the laboratory are shown in four photomicrographs made by Edwards. As the number of cells, called blastomeres, increases, the individual blastomeres become smaller, so that the overall size of the cellular aggregate does not increase at first. At the four-cell stage (*top left*) spermatozoa are still seen within the zona pellucida. The next stage has eight cells (*top right*). At the morula stage, with between 16 and 32 cells, the boundaries of individual blastomeres are less distinct (*bottom left*). At the blastocyst stage (*bottom right*) the zona pellucida has disappeared. The cells have divided to form an inner cell mass and a thin trophoblast and to leave a cavity, the blastocoel.

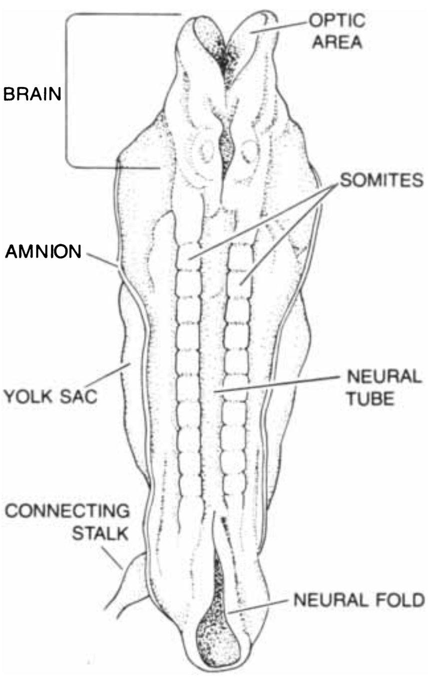


IMPLANTATION takes place when the blastocyst is fixed to the endometrium, erodes through its epithelium and becomes embedded in the uterine wall. The photomicrograph, from the collection of the Department of Embryology of the Carnegie Institution of Washington, shows a partially implanted blastocyst about seven days after normal (not external) fertilization. As is indicated in the accompanying drawing, the embryoblast has become an embryonic disk with two layers: the epiblast, which will form ectoderm and mesoderm, and the endoderm.

er they will ever become human objectives is for the future to decide.

The fifth question emphasizes the importance of careful consideration of the fourth. The value of organ and tissue transplantation in the therapy of degenerative and traumatic disease has been demonstrated by the spectacular (if not fully successful) surgical efforts involving skin, kidneys and the heart. Transplantation from cooperative donors and from victims of accidental or other premature death is limited by the availability of sources and by the recipient's immunological rejection of foreign tissue. Human embryonic tissue and organs could constitute an additional resource, with the potential for greater abundance and conceivably less vigorous immunological resistance. The culturing of preimplantation human embryos to the stage of early organ formation might provide a practical therapeutic source, since many early organs of other mammalian embryos have been shown to continue to develop externally in isolation from the rest of the embryo.

I have so far discussed these five questions from a strictly biological point of view, but clearly each of them also gives rise to important ethical concerns. What is scientifically and technically possible may not be socially or individually acceptable. There is controversy over such issues as "the right to life," the right to death, genetic engineering and informed consent in human experimentation because in each of these areas advancing biomedical knowledge and capabilities have impinged on deeply held political, social, ethical and religious concepts. The debate waxes and wanes as new knowledge agitates particular issues. External human fertilization is an advance that is sure to heat up the general debate and to focus attention on particular aspects of it. The aspect that looms largest is the question of the stage in human development at which a "person," in the ethical and legal sense, comes into being. It is important, given the very different views of the matter that currently are being asserted, to understand what can be said scientifically about this central issue.



EARLY EMBRYO, at the "10-somite" stage reached during the fourth week after fertilization, is formed by the folding of various cell layers. The somites, ranged along the dorsal aspect of the embryo, are blocks of cells that will develop to form muscle, skeletal elements and the dermis, the lower layer of the skin.

Some people argue on nonscientific grounds that the genesis of a person is not developmental, in the sense of gradual appearance over a period of time; that the moment of fertilization—conception—marks the origin of a person because in that moment the maternal and paternal contributions are combined and a new individuality is established. Scientifically, however, fertilization is itself a process in time. Such successive events as the following are distinguishable in the process: contact of the sperm with the egg surface; "activation" of the egg's cortex, or outer layer; formation of the fertilization cone, extrusion of the cytoplasm toward the sperm; entry of the sperm into the egg cytoplasm; completion of meiosis and final maturation of the egg nucleus; formation of the sperm and egg pronuclei; fusion of the pronuclei. These events, occurring over a number of hours, culminate in the first cleavage, the division of the zygote to form two cells.

A particular stage such as the entry of the sperm head or the fusion of pronuclei might be thought of as the critical moment of emergent individuality, but neither of them is a step that is essential to subsequent development; in various animal eggs other kinds of activation of the egg cortex are sufficient to initiate development without a paternal contribution (parthenogenesis), and even when paternal chromosomes are present, they have been shown to have no effect until well beyond fertilization. Hence fertilization, like the steps in the formation of the egg and the sperm that precede it and the cleavage stages that follow it, is a complex process that is continuous in time. Selection of a particular moment as the initiation of individuality is necessarily arbitrary.

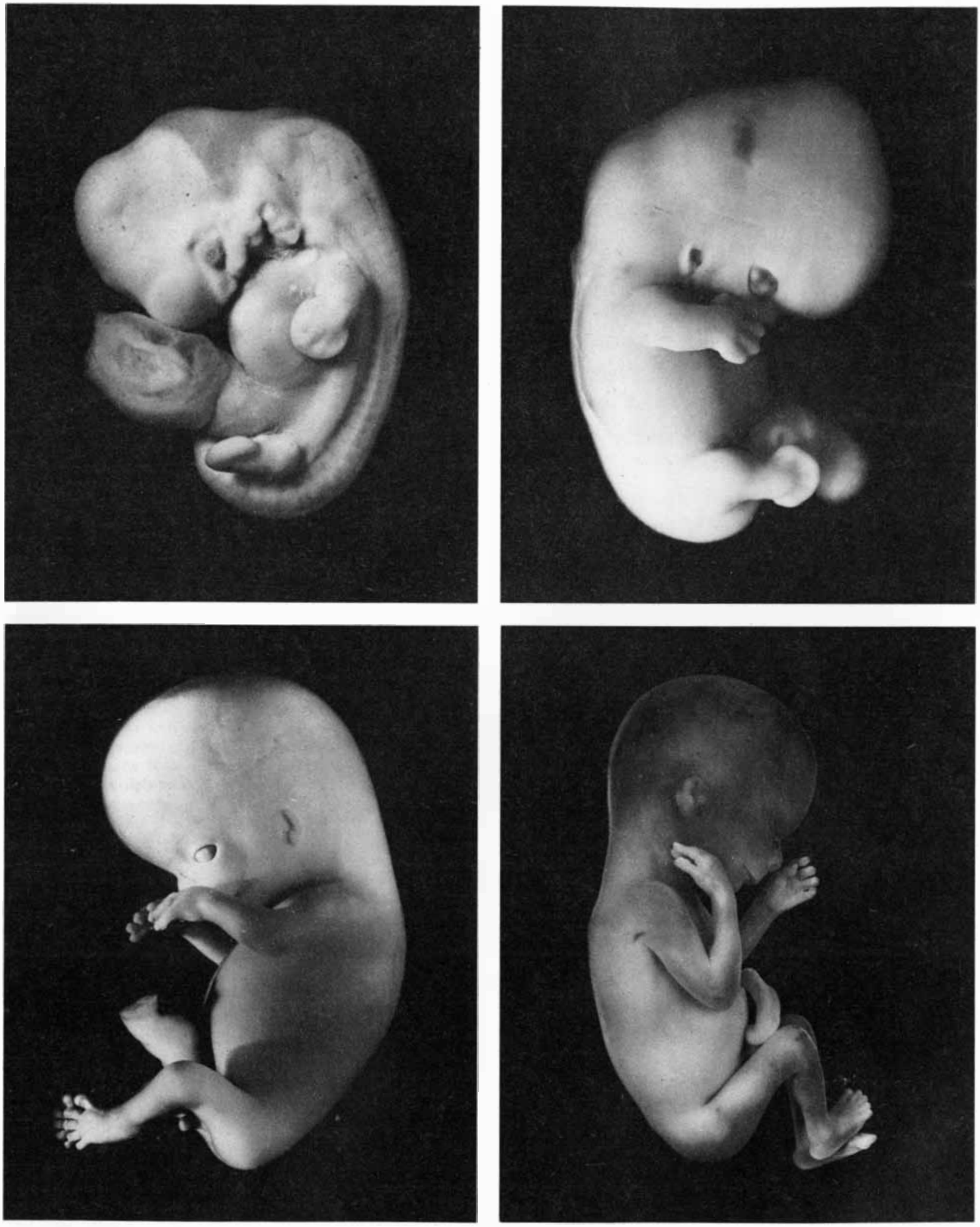
Scientifically there are two essential events in fertilization. First, the long developmental arrest that is characteristic of the mature egg is broken by a series of progressive changes that initiate a new life cycle. Second, in human beings a new genetic constitution contained in two homologous sets of chromosomes is formed. The real question is whether a "person" exists as soon as the progressive developmental changes of fertilization begin. The scientific answer is negative, unless one holds to the naive preformationism that antedated modern embryology: the notion that a homunculus—a complete but miniature human being—is present in either the egg or the sperm before fertilization.

Today the overwhelming evidence is that this notion is false. What is present in the fertilized egg is genetic information, inscribed in chemical language in molecules of DNA (and in cytoplasmic elements formed under the influence of maternal DNA). Through complex processes that information will be translated and elaborated to yield a new person. The evidence also indicates that the translation and elaboration will be influenced by many factors throughout the course of development and that some of those factors were generated in the egg long before fertilization. Under some circumstances it is possible for two individuals to be produced from one fertilized egg (twinning) or even for no true person to be produced at all (in the case of anencephaly, or the absence of a cerebrum). Moreover, minute examination of an egg just before or for some time after fertilization reveals few of the characteristics by which we recognize a person. What it reveals are all the characteristics of a single human cell.

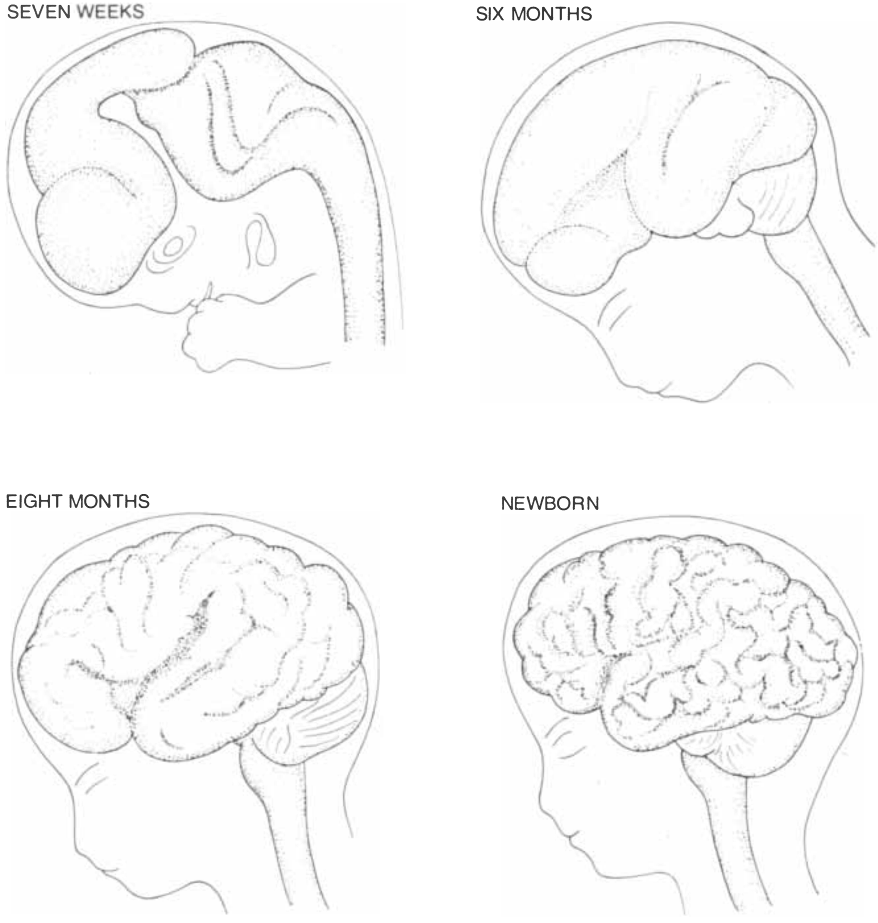
Fertilization does not change the single-cell character of the egg. Only after the initiation of cell division and the multiple and complex changes that ensue does the zygote begin to evolve the set of characteristics of a multicelled organism and finally of a highly organized individual. What is special about a human egg cell is its capacity, under favorable circumstances, for giving rise to a person; there is no scientific evidence that it actually is a person.

If a person cannot be identified in the egg, when can one be identified in the developing cellular system? It is important to recognize that there are two aspects of the concept of person as it relates to human development: that of external perception and that of internal experience. The two aspects do not necessarily correspond in the course of development, and yet both have an important bearing on the issue of the time of appearance of personhood. The external aspect is objectively diagnosable; the internal one is subjective and can at best only be inferred objectively. The external judgment can rely on structural and behavioral criteria associated with personhood. Any judgment about internal personhood is thwarted because sentience, pain and capability for conscious self-awareness are essentially impossible to establish without some means of communication.

Relying entirely on external signs, therefore, one can distinguish several intergrading phases of human devel-



LATER EMBRYOS AND FETUS are from the Carnegie collection. The specimens are embryos at about six (*top left*), seven (*top right*) and eight (*bottom left*) weeks after ovulation and a fetus (*bottom right*) at about 13 weeks; they measure (from crown to rump) about 12, 18, 31 and 100 millimeters in length. In the six-week embryo the vestiges of the somites are still visible along the spine. Primitive hand and foot paddles have formed on the limb buds; a prominence above the umbilical cord (which extends to the left) indicates the heart and liver. At seven weeks the beginning of eye formation is evident and the head is more clearly defined. Elements of the arms and legs are also defined and ridges represent the developing digits. At eight weeks the cerebral hemispheres have grown, eyelid folds begin to cover the eyes and the nose becomes prominent. The limb parts are well defined, the fingers and toes differentiated. Differences in the external genitalia make sexual identification possible. The 13-week fetus is clearly recognizable as human by a layman. The head is still relatively very large. The ear, nose and mouth approximate in appearance those of a full-term infant. The eyes have been covered by the eyelid folds.



DEVELOPMENT OF BRAIN is shown at four embryonic and fetal stages. The tubelike early brain begins to achieve its mature form as the result of a series of bends, or flexures. Different regions of the brain become differentially enlarged, and the cerebral hemispheres overgrow the cerebellum and other regions. Although the main outlines of the brain are established by the sixth month, the convolutions and fissures of the surface, which enormously increase the area and volume of the cortex, do not proliferate until about the eighth intrauterine month.

opment: the cellular or preembryonic phase, the embryonic phase, the fetal phase and finally the phase of vital autonomy, particularly with respect to respiration. These are continuously intergrading phases, but it is nonetheless important to distinguish among them. The first two phases, preembryonic and embryonic, can be recognized as being human only by experts, and when experts recognize the human character of these stages, they rely on criteria different from those generally applied to a person. The reason is that the properties of human beings (as well as the properties of life) are not the same as the properties of persons. The preembryonic and embryonic stages are alive and are human, but they are not externally recognizable as to personhood. (The same is of course true of cells and tissues that are separated from an adult: they may remain alive and human, but they do not constitute a person.)

The fertilized egg cell and the cellular aggregate to which it gives rise prior to implantation have the capability, under favorable circumstances, of giving rise to a person, but they often do not become one, even under natural conditions. It is only after implantation that increasing size, the genesis of shape and the appearance of rudimentary structures and organs mark the transformation of the cellular aggregate into an embryo. Not until after about eight weeks of development does the embryo have characteristics that begin to be externally recognizable by nonexperts as being human. By 20 weeks the external appearance of what by now is a fetus is such that even nonexperts are likely to react to it as a developing person.

Hence the transformation of an embryo into a fetus, although it is not precisely bounded, is a transformation from external recognizability only as human to increasing recognizability as an emergent person. Whether these externally perceived changes are accompanied by inner ones that establish some internal aspects of personhood is not known, but the state of development of the brain makes that seem unlikely until at least late in the period.

Considerable progress has been made in recent years in understanding the development and activity of the brain. Several things seem reasonably clear. Not only complex behavior but also the state of awareness is intimately tied to brain function. An essential aspect of human brain function is the number of neurons (in the billions) and the rich connections among neurons (in multiples of billions) in the cerebral cortex. The number of cortical neurons continues to increase into the sixth month of fetal life, but the connections among these neurons increase enormously not only well beyond the sixth fetal month but also for a number of years into postnatal life. The immature state of development of the brain during fetal life corresponds well with observations on fetal behavior. Responsiveness, various reflexes and even spontaneous movement are manifested in the last third of pregnancy, but there is minimal evidence for cognitive function even at birth.

Nevertheless, the external changes during late fetal life are significant for defining the genesis of a person because they evoke recognition and affective emotions in other persons—and these emotions matter, because personhood is a social status as well as a state of the individual. In these terms the fetal period is the one during which the first external signs of personhood appear, as is indicated by the increasing emotional empathy elicited in other persons. When such empathy comes to be sufficiently great, the fetus has become an infant—a person—regardless of whether it is still in utero or has been delivered. This determination is independent both of the time of birth, which may vary for biological and medical-technological reasons, and of the internal aspect of personhood, the determination of which is fraught with difficulty. The intention should be to so define personhood as to ensure that the status is conferred well before there can be any reasonable possibility that the internal aspect has in fact been attained.

In assessing external fertilization, therefore, the major point is that at least until eight weeks the human embryo can be safely considered to still lack the two essential aspects of personhood: affective recognizability by other persons and internal conscious awareness. The Edwards-Steptoe procedure manipulates not persons but human cells. The stages involved not only are prepersons but also are preembryos. They are preembryos because separated cells from the early cleavage stages of the eggs of nonhuman mammals (and most likely of human eggs) can each give rise to a whole embryo in twinning, and because cells can be deleted from the early clusters (or two clusters can be combined) and a single embryo can still result. Therefore in the cellular stages prior to implantation a single embryo has not yet been established. Since even in the embryonic stage the criteria for personhood have not yet appeared, the Edwards-Steptoe procedure allows an ample margin of safety to ensure that existing persons are not being manipulated and that the rights of persons are not being violated.

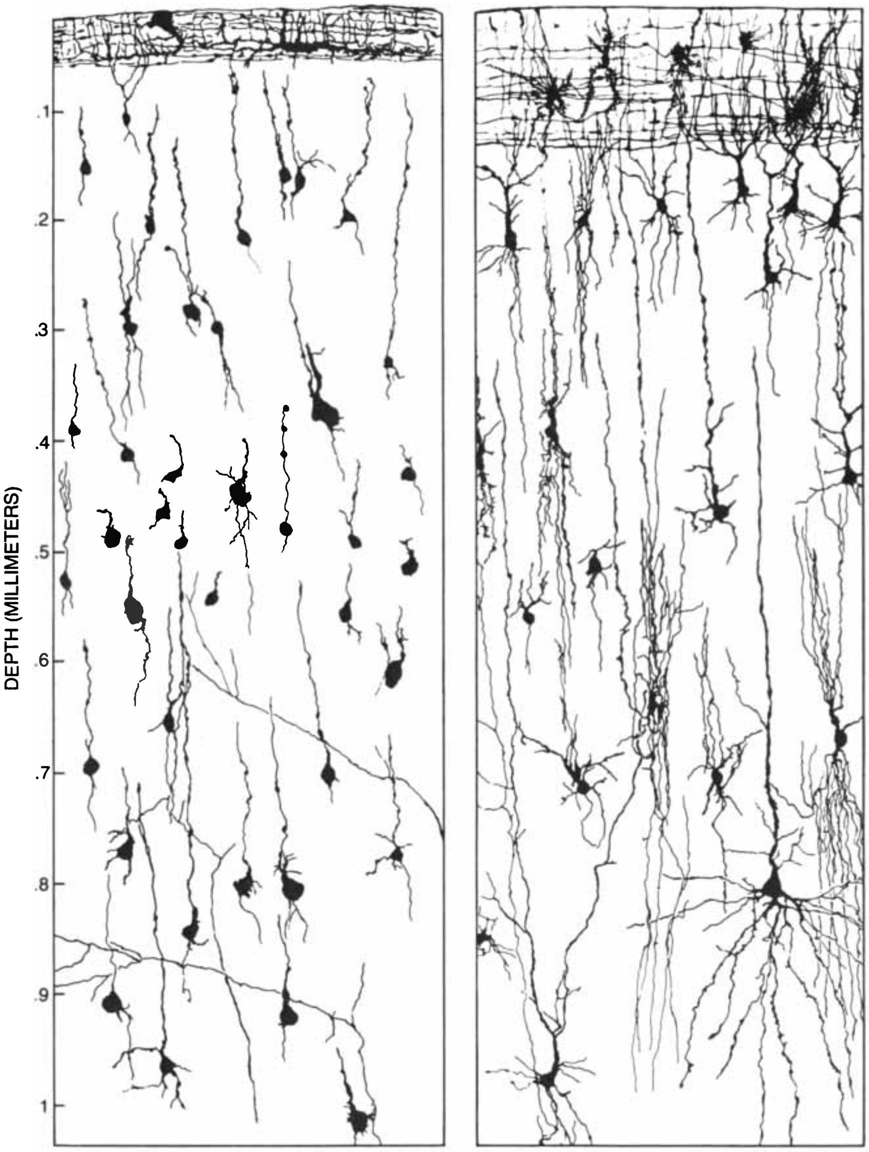
This does not exclude concern for the consequences if the products of the procedure are to be implanted or otherwise nurtured so that they can give rise to a person. Any cell or cell cluster that is a progenitor of a person requires protection against interventions or influences likely to limit the life quality of that person. Hence there are good grounds for insisting that eggs to be implanted for continued development must be appraised carefully for any possible impairment. It does not follow, however, that eggs that are not to be implanted have any special status, beyond the special status normally assigned to human cells, tissues or parts. The capability of producing a person under appropriate circumstances should not be confused with realization of the capability.

An important question that certainly will arise before too long is whether research should be undertaken to prolong the external development of cellular stages of human development to embryonic stages. There is concern that such studies, if they were sufficiently extended, could lead to complete ectogenesis, but that must currently be considered beyond reach, even if it were desired. What is not beyond reach is a better understanding of the conditions necessary for transformation of the cellular stage to the embryonic stage, which considerably antedates the attainment of the criteria of personhood; short of ectogenesis, such stages cannot possibly give rise to a person. Once such stages are reached externally, however, they could provide a source for the continued development of rudiments of organs by techniques that have already been worked out in laboratory mammals. Such rudiments, like the whole embryos from which they came, could not, however, possibly give rise to persons. Policy with respect to such possibilities is much in need of full consideration.

The advent of arenas for discussion of these matters in the Ethics Advisory Board and in the Office of Technology Assessment, which is beginning an assessment of applied genetics, is to be welcomed. One would hope that this consideration, with full public discussion, will produce guidelines that can be reviewed and modified as issues arise and change. Given the complexity of the issues and the deep social and emotional reactions they engender, it seems sound to consider first the establishment of guidelines for problems that are at the margin, both in immediacy and in level of consensus. A frontal attack on the entire complex would only generate frustrating controversy among opposed general views.

As examples, laboratory studies of human fertilization and preimplantation stages call for immediate consideration, and appropriate guidelines should be set for the procurement of eggs and for the utilization of fertilization products. Less pressing items on the agenda include the question of the status of embryonic cells, tissues and organs, whether they are obtained from abortuses or from preimplantation stages developed further in the laboratory. Still other subjects for consideration are various manipulations of preimplantation stages prior to implantation, including cell fusion, genetic intervention and cloning.

Placing such items on the agenda does not presuppose their desirability; considering them may, however, help to establish the appropriate limit beyond which current policy should not be permissive. That limit seems likely to be set considerably short of the external production of sentient stages of development or the production of persons who are limited or altered in terms of the quality of life.



NEURONS OF VISUAL CORTEX of spontaneously aborted 25-week (*left*) and 33-week (*right*) fetuses are shown in composite drawings made from photomicrographs by Dominick P. Purpura of the Albert Einstein College of Medicine. The tissue was stained by the Golgi method. Depth from the surface is shown in millimeters. Only the upper half of the cortex is represented in the older specimen (*right*); actually the visual cortex virtually doubles in thickness from 25 to 33 weeks. The obvious difference between the two samples is the great increase in the number and length of the dendrites, the shorter processes that mediate the transmission of nerve impulses between neurons. The dendrites are also more mature, displaying (under higher magnification) numerous "spines," where transmission of nerve impulses takes place.

The Nerve-Growth Factor

This protein plays a key role in the formation of the nervous system. It has also been employed as an "Ariadne's thread" to explore mechanisms of nerve-cell growth and differentiation

by Rita Levi-Montalcini and Pietro Calissano

The human nervous system is a vast network of several billion neurons, or nerve cells, endowed with the remarkable ability to receive, store and transmit information. In order to communicate with one another and with non-neuronal cells the neurons rely on the long extensions called axons, which are somewhat analogous to electrically conducting wires. Unlike wires, however, the axons are fluid-filled cylindrical structures that not only transmit electrical signals but also ferry nutrients and other essential substances to and from the cell body. Many basic questions remain to be answered about the mechanisms governing the formation of this intricate cellular network. How do the nerve cells differentiate into thousands of different types? How do their axons establish specific connections (synapses) with other neurons and non-neuronal cells? And what is the nature of the chemical messages neurons send and receive once the synaptic connections are made?

This article will describe some major characteristics and effects of a protein called the nerve-growth factor (NGF), which has made it possible to induce and analyze under highly favorable conditions some crucial steps in the differentiation of neurons, such as the growth and maturation of axons and the synthesis and release of neurotransmitters: the bearers of the chemical messages. The discovery of NGF has also promoted an intensive search for other specific growth factors, leading to the isolation and characterization of a number of proteins with the ability to enhance the growth of different cell lines.

The peripheral nervous system of vertebrate animals includes three kinds of nerve cells: sensory neurons, which transmit impulses from sensory receptor structures to the brain; motor neurons, which innervate the striated, or skeletal, muscles, and autonomic neurons, which regulate the functional activity of the circulatory system, the organs, the glands and the smooth muscles (such as those of the intestine). Autonomic neurons are of two kinds: sympathetic and parasympathetic. The sensory neurons and some of the sympathetic neurons are situated in chains of ganglia flanking the length of the spinal cord. Because these neurons are uniquely accessible to experimental manipulation much of the research on the development of the nervous system at the cellular level has focused on how the nerve fibers projecting from the sensory and sympathetic ganglia make connections with their corresponding target organs.

In the first half of this century the new science of experimental embryology seemed to offer the best way to study the intimate bond that interlocks the growth of the peripheral neurons and their target organs. Ross G. Harrison of Yale University challenged the nervous system of the larva of amphibians to solve problems it would ordinarily never confront, such as providing nerves for limbs and organs grafted from other species. He wanted to see how sensory and sympathetic ganglia, sending out their nerve fibers to these peripheral "fields of innervation," would adjust to the different dimensions and configurations of the foreign organs.

Harrison's results demonstrated that the developing amphibian nervous system is remarkably flexible in adapting to such novel situations, even to the point of accelerating the growth of nerve fibers in a host species to keep pace with the faster-growing limb of a smaller donor species. He concluded that the embryonic nervous system is highly receptive to influences exerted by the peripheral field. Such influences are not species-specific, however, since they can be evoked by organs or rudimentary limbs that have been transplanted from one species to another.

Viktor Hamburger of Washington University extended Harrison's work but chose to study the chick embryo because its nervous system, although more complex than that of an amphibian, lends itself better to experimental analysis: its nerve centers are more clearly delineated and their strong affinity for silver stain enables the experimenter to visually examine the nerve structures more easily. Hamburger grafted limb buds onto chick embryos at very early stages of development and observed how the modified peripheral field was innervated by sensory and sympathetic fibers. Unfortunately the responses were often so complex that they defied interpretation.

In 1948, in an effort to get more straightforward results, Elmer D. Bueker of Georgetown University modified Hamburger's experimental approach. He had the ingenious idea of replacing one limb bud of a chick embryo with a fragment of a bird or mammalian tumor. The tumor cells were all undifferentiated and hence provided a homogeneous peripheral field, in contrast to the cells of the normal limb bud, which were destined to differentiate into many types of tissue. Bueker's experiment was therefore expected to reveal how a homogeneous but fast-growing peripheral field would be innervated.

Of three different tumors implanted in the body wall of three-day-old chick embryos only one, a mouse tumor of connective-tissue cells called sarcoma 180, grew vigorously and was invaded by nerve fibers growing out from adjacent sensory ganglia. In embryos sacrificed after five days the sensory ganglia innervating the tumor were 33 percent larger than those innervating the normal limb bud on the opposite side of the embryo. At first these findings seemed to suggest that the size of a sensory ganglion depends on the size and rate of growth of its field of innervation. According to this hypothesis the rapidly growing tumor provided a more favorable peripheral field for the innervating sensory fibers than the slowly growing limb bud did.

A reexamination of Bueker's findings by our group at Washington University revealed new aspects of the phenomenon that obliged us to revise his conclusions. We found not only that the growth of the sensory ganglia innervating the sarcoma-180 tumors increased but also that the sympathetic ganglia increased enormously in volume, becom-

HALO OF NERVE FIBERS radiates from a sympathetic-nerve ganglion that was removed from a chick embryo and then incubated for 12 hours in a semisolid medium containing the nerve-growth factor (NGF). The factor was originally isolated from a mouse tumor designated sarcoma 180 but has subsequently been found to be secreted in trace amounts by a variety of normal and neoplastic (cancer) cells. In 1953 one of the authors (Levi-Montalcini) and Hertha Meyer found that NGF could induce the outgrowth of nerve fibers from an isolated sympathetic ganglion in tissue culture. This discovery led ultimately to the isolation of NGF and determination of its chemical structure. Micrograph, which magnifies ganglion 135 diameters, was made by J. S. Chen at Laboratory of Cell Biology in Rome.

ing five to six times larger than they were in control animals. This increase in size of the sympathetic ganglia was considerably greater than that exhibited by the sensory ganglia. Together with the sensory fibers the sympathetic fibers branched all over the peripheral field provided by the tumor but did not form any synapses with the tumor cells.

In addition, and indicating an even more striking departure from normality, the viscera of the embryos with the transplanted tumors were flooded with excessive numbers of sympathetic fibers long before the embryos of the control animals were even sparsely innervated. The sympathetic fibers forced their way into large and small veins, impeding and sometimes completely obstructing the flow of blood. These extraordinary effects suggested that the overgrowth of the sympathetic ganglia was more than simply a response to the rapidly growing peripheral field of innervation provided by the tumor, as Bueker had proposed. Rather it seemed the tumor was releasing some chemical factor that was in turn inducing the remarkable growth of the sympathetic ganglia and the exuberant branching of their nerve fibers.

This hypothesis was tested by transplanting sarcoma-180 tumors into the respiratory membranes of the chick egg, which are permeated with blood vessels from the embryo. The tumor and the developing chick embryo therefore shared the same blood supply, although they were not in direct contact. We found that when the tumor was transplanted into the respiratory membranes, it elicited the same growth-promoting effects on the sympathetic ganglia as it did when it was implanted in the embryo itself, providing convincing proof that the tumor was releasing a soluble factor that was carried in the bloodstream to the embryo.

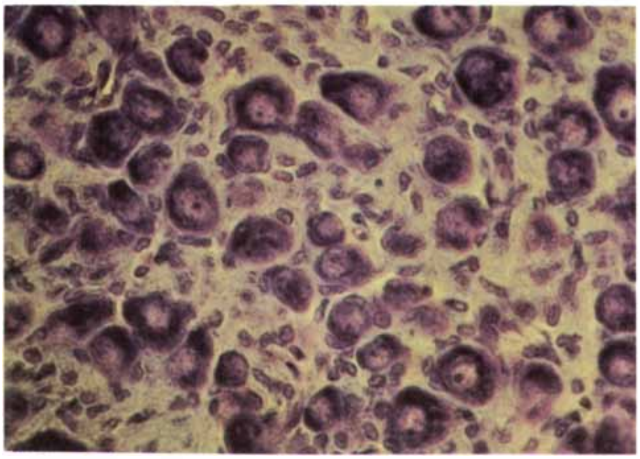
The next challenge was to identify the postulated nerve-growth factor being released by the tumor. For this purpose we needed a system much less complex than the developing embryo. Tissue culture (which in the early 1950's had yet to become the universal biological tool it is today) seemed to offer a useful alternative. We reasoned that if sarcoma 180 was releasing a chemical factor with the ability to enhance nerve growth, the same effect should appear when an isolated sympathetic ganglion was incubated with the tumor in laboratory glassware. In 1953 one of us (Levi-Montalcini) and Hertha Meyer did this experiment at the Biophysics Institute in Rio de Janeiro. Sensory and sympathetic ganglia were dissected out of eight-day-old chick embryos and cultured in a semisolid medium in proximity to fragments of mouse sarcoma-180 tumors. Within 10 hours of incubation the isolated ganglion gave rise to a dense halo of nerve fibers, radiating out from the explant like rays from the sun. Control ganglia cultured for the same length of time in the absence of the sarcoma-180 cells displayed only a sparse and irregular outgrowth of nerve fibers.

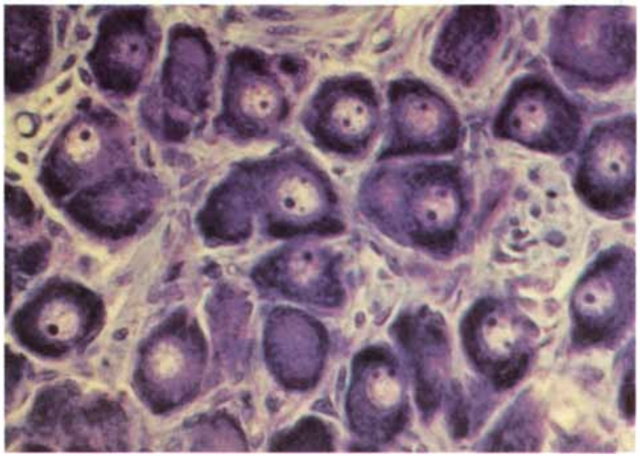
The discovery that the tumor could exert its growth-enhancing effects on an isolated ganglion in tissue culture was the turning point of the investigation. Whereas our earlier experiments in developing chick embryos had required weeks of painstaking work, we could now in a few hours screen a large number of tissues, organic fluids and chemicals to determine if they were sources of the growth-promoting activity. Furthermore, it now became possible to attempt to isolate the nerve-growth factor from our simplified tissue-culture system.

Stanley Cohen, a biochemist, agreed to join our group at Washington University to undertake the task of identifying the active agent. He soon managed to restrict the growth factor to a fraction of the tumor cells containing both protein and nucleic acid. This finding opened the way to a more precise biochemical analysis, but we would not have made much progress if it had not been for a fortuitous event two years later. In order to determine whether the nerve-growth factor was a protein or a nucleic acid, Cohen and one of us (Levi-Montalcini) treated the tumor-cell extract with snake venom containing a high concentration of the enzyme phosphodiesterase, which degrades nucleic acids. To our surprise we found that adding minute amounts of the venom to the active fraction of the sarcoma-180 cells enhanced the growth-promoting effect of the fraction rather than reduced it. That the snake venom itself was the source of the increased activity was readily confirmed. The addition of a small amount of venom to the culture medium in the absence of the extract of sarcoma 180 elicited the growth of the same dense halo of nerve fibers around an isolated sensory or sympathetic ganglion.

In fact, the nerve-growth factor in the snake venom turned out to be much more abundant and potent than that in the sarcoma-180 cells. Cohen was therefore able to purify the factor from the venom and demonstrate that it was a protein. Injections of the purified snake-venom NGF into growing embryos resulted in the same excessive sympathetic innervation of the viscera and the blood vessels as that induced by the sarcoma-180 cells.

The discovery that two unrelated sources—mouse sarcoma and snake venom—harbor NGF suggested the possibility that still other tissues might secrete the factor. The search focused on the submaxillary salivary gland of rodents, which is similar in certain respects to the venom gland of snakes. Indeed, Cohen isolated an NGF from mouse salivary glands that was about 10,000 times more active than that puri-





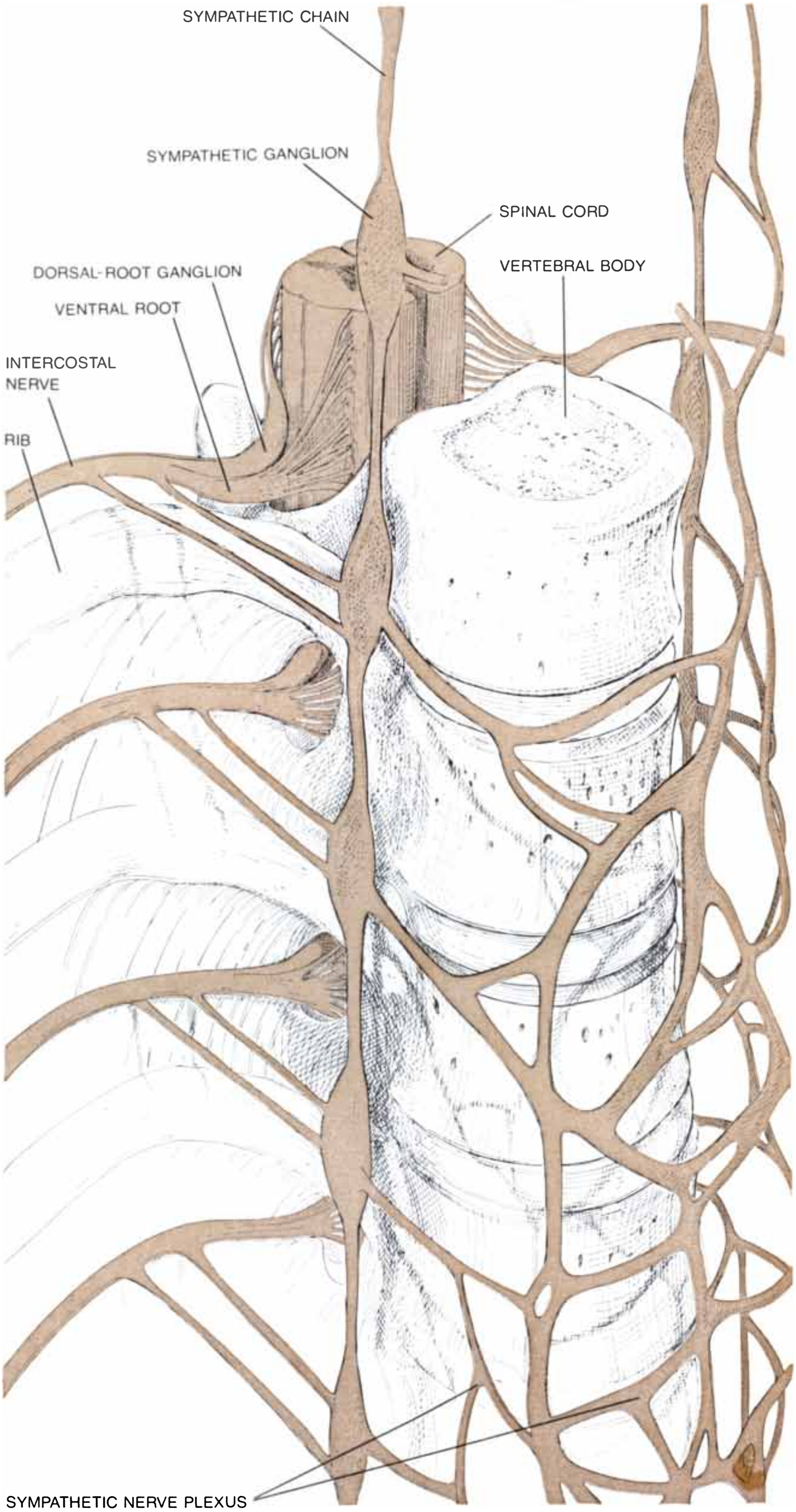
INCREASED SIZE AND NUMBER of neurons in the sympathetic ganglia of NGF-treated mice result in an increase in the volume of the ganglia. These micrographs are of sections through two ganglia, one ganglion from a three-week-old rat injected daily with saline solution only (*left*) and the other from a littermate mouse injected with NGF (*right*). Neurons in NGF-treated rat are substantially enlarged. Note also that cells in ganglia of animals treated with NGF exhibit a greater affinity for toluidine blue, the dye used to stain the cells.

fied from mouse sarcoma 180 and about 10 times more active than that purified from snake venom. Over the next two decades smaller amounts of NGF were found to be secreted by a wide variety of normal and neoplastic (cancer) cells.

In 1969 V. Bocchini and Pietro U. Angeletti at the Laboratory of Cell Biology in Rome devised a method for purifying NGF from mouse salivary glands. With the large quantities of NGF made available by their technique Ruth Hogue Angeletti and Ralph A. Bradshaw of the Washington University School of Medicine were able to determine the sequence of amino acids in the protein. Biologically active NGF is a dimer, or complex of two identical polypeptide chains, each of which has a molecular weight of 13,250 daltons. The NGF molecule has a marked predominance of positively charged amino acids and has a net positive charge at neutral *p*H. In addition the folded polypeptide chain is stabilized by three covalent sulfur-sulfur bridges between units of the amino acid cysteine at different positions along the chain. Such disulfide bonds are common in proteins that are actively secreted from cells, such as insulin and antibodies. The restraint these molecular bridges impose on the conformation of the polypeptide chain is apparently necessary to prevent its denaturation and inactivation in an environment that is much more subject to adverse conditions than the interior of the cell.

A comparison of the sequence of amino acids in NGF with that of several other polypeptides by William A. Frazier of Washington University revealed that NGF and insulin have certain sequences in common. This observation led to the hypothesis that the gene for NGF evolved from an ancestral gene for proinsulin, the large precursor polypeptide that is cleaved to yield the active molecule of insulin. Frazier suggested that the ancestral proinsulin gene had duplicated itself and the two copies had subsequently evolved divergently, giving rise to proinsulin and a postulated precursor polypeptide for NGF. Although this intriguing possibility remains theoretical, the similarities in the amino acid sequences of NGF and insulin are not great enough to result in any similarity of function: the two molecules have completely different target cells and biological activities.

In 1967 Silvio S. Varon and Eric M. Shooter of the Stanford University School of Medicine isolated NGF by a different procedure and found that the NGF dimer was present as a stable complex with two copies each of two other proteins, which they designated the alpha and gamma subunits. This finding was puzzling, since it was known that the NGF dimer by itself was biologically active. Further investigation revealed that the gamma subunit is a specific en-

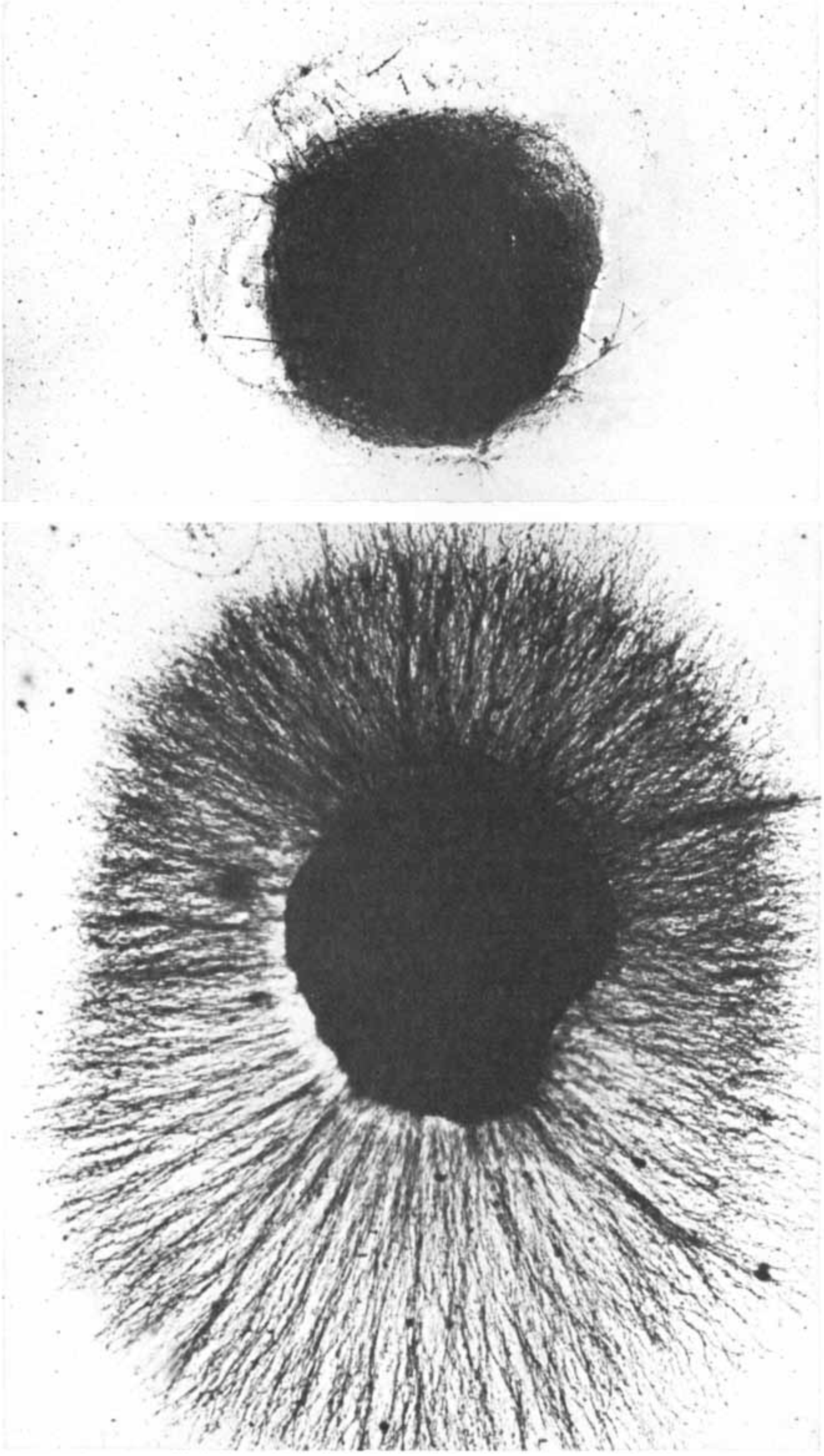


CHAIN GANGLIA flanking the spinal cord of mammals (those of a human being are shown here) contain the cell bodies of the sympathetic neurons that innervate the various organs, blood vessels, smooth muscles and glands of the body. The irregularly shaped nodules are held together by nerve fibers and are also connected to the roots of the spinal cord. Each ganglion innervates the blood vessels and organs in the corresponding segment of the trunk and legs. During the development of the nervous system the neurons in the ganglia send out fibers to innervate their specific target organs. Because the ganglia are readily accessible they are easy to study.

zyme that cleaves polypeptide chains only at a point adjacent to the amino acid arginine. The alpha subunit, on the other hand, has no detectable biological activity. This odd association of three disparate proteins demanded an explanation.

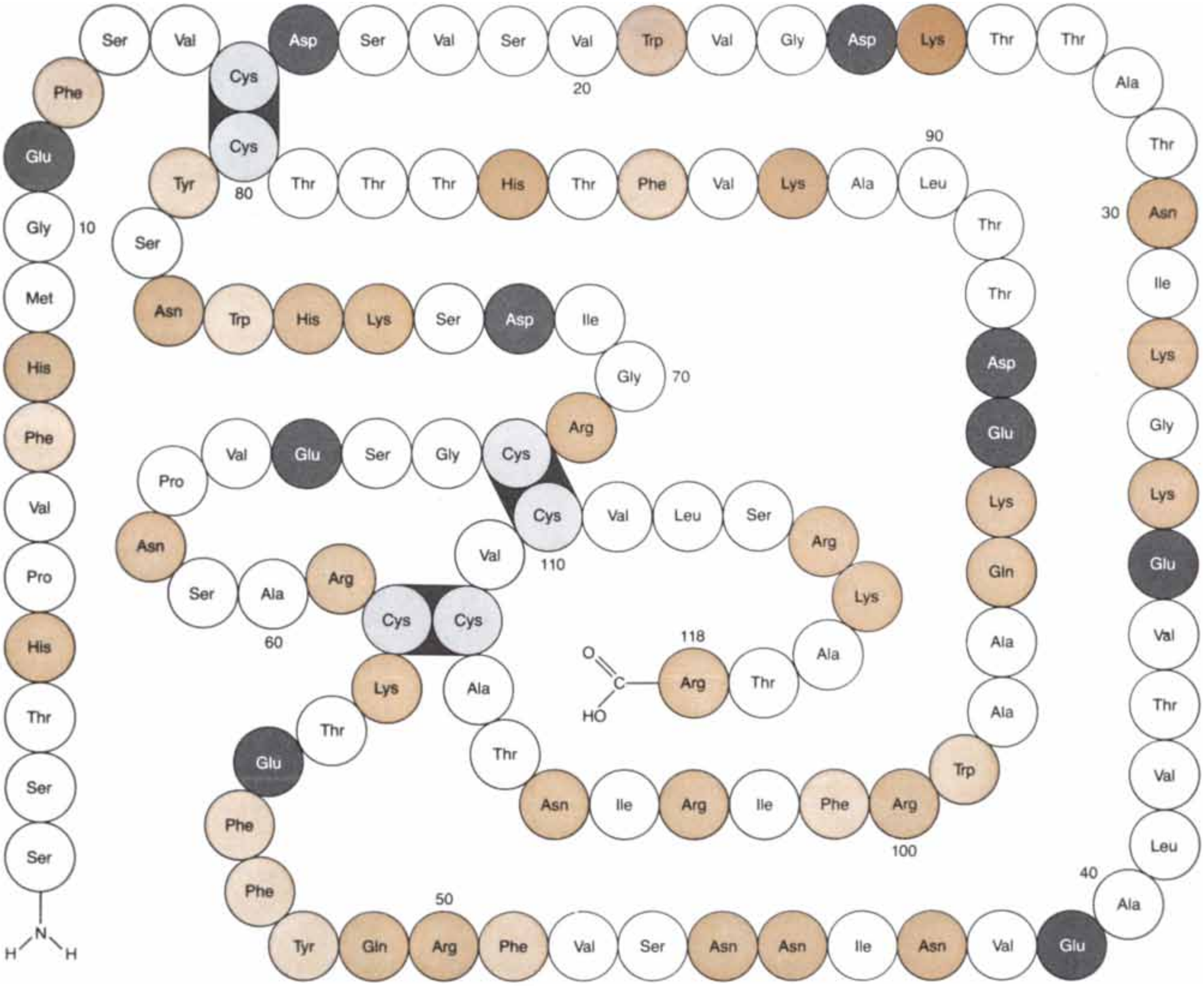
Shooter hypothesized that the alpha and gamma subunits serve to activate, store and protect the NGF molecule. According to his scheme the initial gene product in the manufacture of NGF is a large precursor polypeptide called proNGF, two copies of which form a dimer. Two gamma subunits then associate with the proNGF dimer and cleave the precursor chains to generate the active NGF dimer. Unlike the typical complex of an enzyme and its product, which rapidly dissociates, the gamma subunits remain bound to the NGF dimer after the cleavage of the proNGF chains. (The two alpha subunits may bind either to the proNGF dimer before cleavage or to the complex of the gamma subunit and the NGF dimer after cleavage.) The association of the gamma and alpha subunits with the NGF dimer apparently serves to protect it from further degradation by other protein-cleaving enzymes in the body fluids. Shooter's hypothesis recently received support when he and Edward A. Berger isolated the proNGF molecule and demonstrated that incubation of this precursor with the gamma subunit resulted in its complete conversion into active NGF.

The biological activity of a protein resides in its three-dimensional structure, which consists of not only its amino acid sequence but also the precise folding pattern of the polypeptide chain (the secondary structure) and the suprafolding of two or more such chains to form a unique globular entity (the tertiary structure). The only reliable way to determine the secondary and tertiary structure of a protein molecule is to crystallize it and mathematically analyze the crystal's X-ray-diffraction pattern. The recent crystallization of NGF by A. Wlodawer, Keith O. Hodgson and Shooter is therefore an important first step in determining the three-dimensional structure of the molecule by X-ray crystallography.



BIOLOGICAL ASSAY for NGF is based on the ability of the factor to induce the outgrowth of nerve fibers from an isolated ganglion. These micrographs show sensory ganglia dissected out of an eight-day-old chick embryo and cultured for 12 hours either in the absence (*top*) or the presence (*bottom*) of 10 micrograms of pure NGF. After incubation the fibers were stained.

The first real insight into the function of NGF in the living animal was provided by experiments done in our laboratory at Washington University in 1959. Cohen obtained specific antibodies to NGF by injecting purified mouse NGF into rabbits. Small amounts of rabbit serum containing the NGF antibodies were then injected into newborn mice. A month later the experimental and control mice were sacrificed. In the mice treated with the NGF antibodies the sympathetic ganglia were reduced to such diminutive size that they were barely visible in the dissecting microscope! Nevertheless, the NGF antibodies had no ill effect on any other organs or tissues and for unexplained reasons did not damage the peripherally located sympathetic ganglia that control the sex organs in both sexes. It has therefore been possible to raise to maturity entire colonies of mice entirely lacking in sympathetic nervous function but normal in all other respects. These animals have provided a most valuable model system for studying how the lack of sympathetic innervation interferes with a multiplicity of body functions.



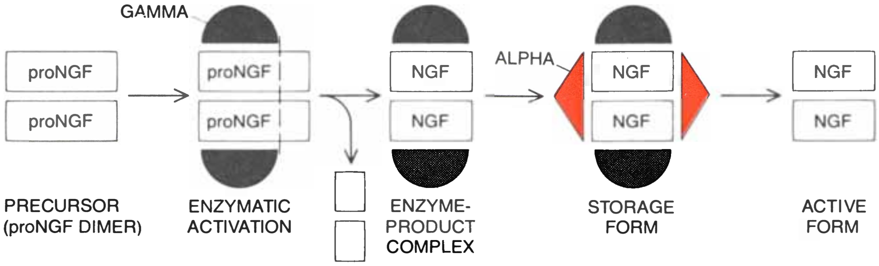
Ala	ALANINE	Leu	LEUCINE
Arg	ARGININE	Lys	LYSINE
Asn	ASPARAGINE	Met	METHIONINE
Asp	ASPARTIC ACID	Phe	PHENYLALANINE
Cys	CYSTEINE	Pro	PROLINE
Gln	GLUTAMINE	Ser	SERINE
Glu	GLUTAMIC ACID	Thr	THREONINE
Gly	GLYCINE	Trp	TRYPTOPHAN
His	HISTIDINE	Tyr	TYROSINE
Ile	ISOLEUCINE	Val	VALINE

PRIMARY STRUCTURE OF NGF is shown schematically as a necklace of amino acid units. The chain is bridged at three points by covalent bonds between the sulfur groups of cysteine units. As is true of all proteins, the sequence of amino acids in the chain gives rise to a unique three-dimensional structure, which is the result of weak interactions among the amino acid units. The precise folding pattern of NGF has not yet been determined. The various categories of amino acid are indicated by shading: positively charged (*dark color*), negatively charged (*dark gray*), incorporating an aromatic ring (*light color*), incorporating sulfur atoms (*light gray*) and uncharged (*white*). Above a critical concentration two monomeric chains of NGF, each having a molecular weight of 13,250 daltons, associate to form a dimer. It is not yet known whether the monomer or the dimer is the form of the NGF molecule that is biologically active.

It is not yet known whether the dramatic and selective effects of the NGF antibodies are due to a direct toxic action of the antibodies on immature sympathetic neurons or to an inactivation by the antibodies of circulating molecules of NGF, thereby indirectly causing the death of the sympathetic neurons by depriving them of the NGF they need in order to survive. This second alternative seems the more likely one; convincing evidence in its favor would provide tangible proof that NGF is an absolute requirement for the survival and growth of immature sympathetic neurons in the living animal.

That NGF is essential for the survival of sympathetic neurons in cell culture has been known for some time. In 1963 Pietro Angeletti and one of us (Levi-Montalcini), who were then at the Istituto Superiore di Sanità in Rome, dissected sensory and sympathetic ganglia into their cellular components: neurons, glial cells (which support and nourish the neurons) and fibroblasts (embryonic connective-tissue cells). After 24 hours of culture the glial cells and fibroblasts had survived and multiplied but the neurons had undergone a massive degeneration. The daily addition of minute amounts of NGF to the culture medium, however, enabled the neurons to survive for indefinite periods and to form a dense meshwork of nerve fibers that after a few days completely covered the surface of the culture dish. This ability of NGF to sustain sympathetic neurons in culture has since made possible some ingenious and revealing experiments on neuronal differentiation [see "The Chemical Differentiation of Nerve Cells," by Paul H. Patterson, David D. Potter and Edwin J. Furshpan; SCIENTIFIC AMERICAN, July, 1978].

The availability of milligram quantities of pure NGF has made it possible to test its effects in the living organism. All mammals respond to NGF the same way, although for practical reasons rodents are the experimental animals of choice. When 10 micrograms of NGF per gram of body weight is injected into newborn rodents for periods of up to three weeks, their sympathetic ganglia become 10 to 12 times larger than those of control animals. This excessive in-



MANUFACTURE OF NGF requires a number of "processing" steps. The initial form of the molecule is a precursor chain called proNGF, which forms a dimer. The proNGF chains are cleaved by an enzyme known as the gamma subunit to yield the NGF dimer. The two gamma subunits remain associated with the dimer after it is generated, and two additional proteins, the alpha subunits, bind to the complex. The association of the alpha and gamma subunits with the NGF dimer apparently protects it from further degradation by other enzymes.

crease in size of the sympathetic ganglia has been traced to three separate processes: (1) an increased rate of differentiation of the sympathetic neurons, (2) an increase in the total number of neurons in the ganglia and (3) an increase in the size of the fully differentiated neurons.

NGF does not increase the number of neurons in the sympathetic ganglia by enhancing their multiplication. Instead, as was first proposed by I. A. Hendry of the Australian National University, the marked increase in the number of neurons is due to the survival of redundant immature neurons that would ordinarily die off in the course of development. Cell death is a common event in the formation of the nervous system: entire populations of immature neurons die off or are sharply reduced in size. In fact, in the early developmental stages of the chick embryo the dead cells in the sensory and sympathetic ganglia often outnumber the live ones. The generally accepted hypothesis is that the immature neurons that do not establish functional connections with their target cells are doomed to die; the redundant number of neurons ensures that all the appropriate connections are made. The effect of experimentally administered NGF is to enable the redundant neurons to survive and differentiate in spite of their failure to establish connections. It is this effect of NGF that gives rise to the substantially increased numbers of neurons in NGF-treated sympathetic ganglia.

In addition to apparently being essential to the survival of immature sympathetic neurons NGF seems to play a vital role in guiding nerve fibers toward their corresponding target organs. Three basic mechanisms have been proposed over the years to explain the formation of specific neural circuitry: (1) an elaborate predetermined program encoded genetically in each neuron that unfolds according to rigid and unmodifiable rules, (2) a random process of trial and error in which growing nerve fibers that make the right connections are consolidated and those that fail are reabsorbed and (3) a general program of circuit formation that is brought to completion by an interplay between genetic and extrinsic factors.

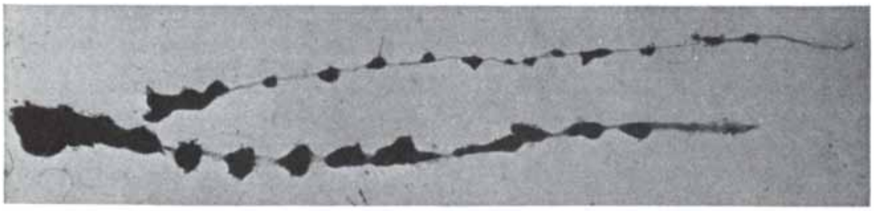
The first mechanism can be discounted because it would require very large amounts of genetic information—much more than could be encoded in the entire complement of DNA in the cell nucleus of each neuron. The second mechanism is also unlikely, because such a random process of circuit formation would be extremely time-consuming and wasteful of energy and resources. The third mechanism therefore seems the most probable: the circuitry of the nervous system is established through some combination of genetic and extrinsic factors.

The existence of such extrinsic factors was first suggested by the Spanish neurologist Santiago Ramón y Cajal, who visualized them as chemical signals emanating from peripheral tissues that would direct the growing nerve fibers toward their matching target cells. Cajal named the process neurotropism. His hypothesis was neglected for many years because the methodology for detecting such chemical factors in the living embryo was not yet available. The discovery and isolation of NGF, however, has provided an opportunity to reconsider the concept of neurotropism under more favorable experimental conditions.

Convincing evidence for the role of NGF in the formation of neural circuits in the sympathetic nervous system has come from both animal and test-tube experiments. At the Laboratory of Cell Biology in Rome, NGF was injected into the brain of newborn rodents. An unexpected effect of the treatment was that nerve fibers sprouted from the sympathetic ganglia and invaded the brain and spinal cord. Apparently the NGF injected into the brain diffused through the motor and sensory roots of the spinal cord and reached the sympathetic chain ganglia flanking the cord, where it induced the outgrowth of nerve fibers. This finding implies that the tip of a growing sympathetic nerve fiber elongates along a diffusion gradient of NGF released by the target organ. The NGF gradient does not entirely determine the course of the nerve fibers but helps to orient them in the right direction.

Further evidence for a neurotropism mediated by NGF was provided by some elegant experiments done by Robert B. Campenot of the Harvard Medical School. He devised a three-chambered cell-culture system in which the chambers were separated by impermeable barriers of sterile silicon grease, preventing the diffusion of fluid from one chamber to another. Immature sympathetic neurons were seeded into the central chamber in the presence of NGF and allowed to send out growing fibers under the silicon barriers along scratches made on the bottom of the culture dish. Campenot filled one side chamber with a nutrient solution containing NGF and the other chamber with the same nutrient solution devoid of NGF. He observed that the growing nerve fibers migrated from the central chamber only into the side chamber containing NGF. When the NGF-rich solution was withdrawn from that side chamber, the nerve fibers that had extended into it began to degenerate, even though the neuronal cell bodies in the central NGF-containing chamber persisted in good condition.

These experiments resolved the longstanding question of whether extrinsic chemical factors play a role in guiding the growth of nerve fibers. Since NGF is released in minute amounts from peripheral tissues receiving innervation from sympathetic ganglia, it now seems clear that a diffusion gradient of NGF directs the fibers toward their corresponding target tissues. When the growing nerve fiber and the target cell finally come in contact, the provisional adhe-



NGF ENHANCES THE GROWTH of sympathetic chain ganglia. The two sets of ganglia shown here were dissected from littermate three-week-old mice. The control mouse (*top*) was injected daily with a saline solution, whereas the test mouse (*bottom*) received 10 micrograms of NGF per gram of body weight. Chain ganglia showed a substantial increase in volume.

sion consolidates into the structural and functional organization of the synapse.

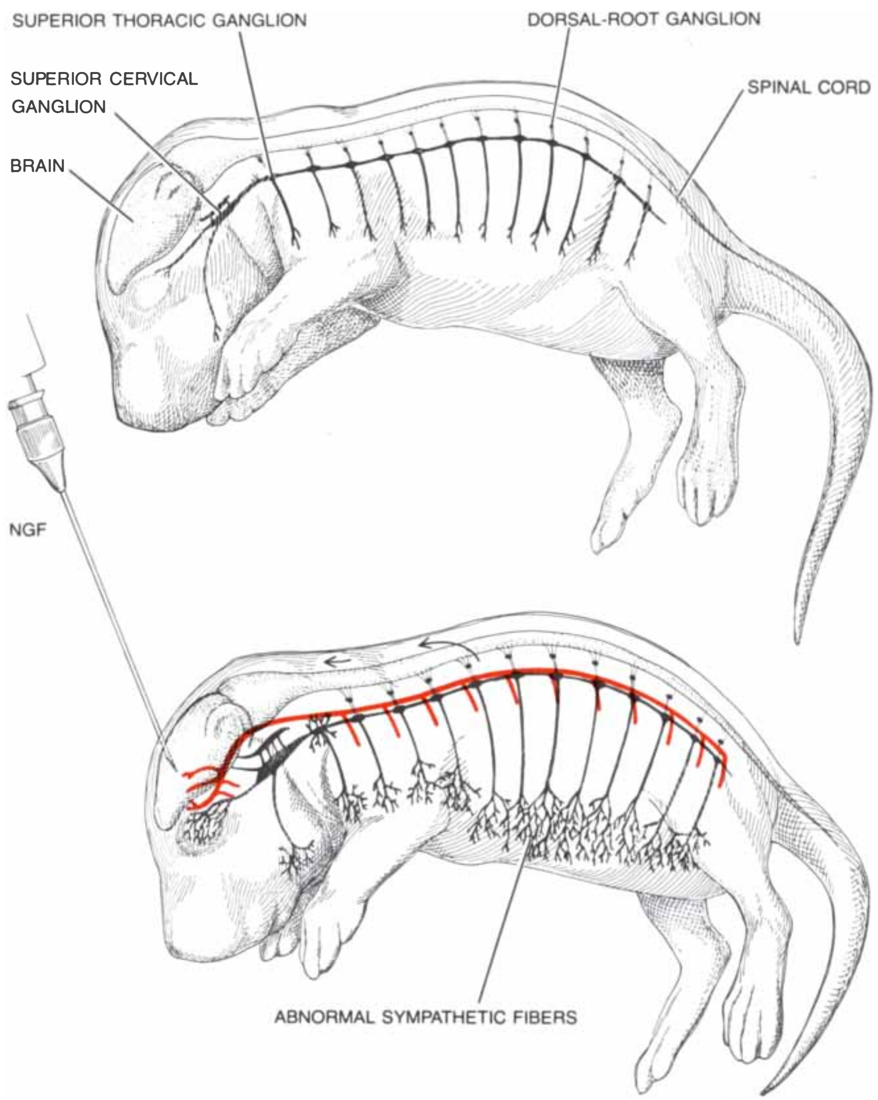
Once the elongating fibers have established the appropriate synaptic contacts with the target cells, the continued survival of the innervating cells in the ganglion appears to depend on the availability of NGF. Studies conducted by Hendry at the Australian National University and by K. Stöckel and H. Thoenen at the Basel Institute for Immunology have demonstrated that NGF is taken up at the terminal nerve endings of the sympathetic fibers and transported back to the neuronal cell body along the axon. This retrograde axonal transport of NGF is absolutely essential for the survival of the innervating neurons. When it is experimentally prevented (either by severing the projecting axons, by treating them with the drug vinblastine, which blocks axonal transport, or by administering 6-hydroxydopamine, which destroys the nerve endings), the innervating sympathetic neurons in the ganglion die off. The lethal effects of blocking the axonal transport of NGF can be completely overcome, however, by supplying the cell bodies with externally administered NGF. In this case the external NGF makes up for the NGF that would normally be transported back inside the axon to the cell body from the innervated cells.

Work in several laboratories has shown that the retrograde axonal transport of NGF follows its interaction with specific receptor sites on the nerve terminals of the newly established fibers. Receptors are proteins that are usually located on the external surface of the cell membrane; they provide specific recognition sites for messenger substances such as hormones, neurotransmitters and growth factors. The existence of such specific receptors on the neuronal surface makes it possible for NGF to exert its effects at exceedingly low concentrations (about 2.8 micrograms per liter). The binding of NGF to its receptors triggers a chain of biochemical events that leads ultimately to the outgrowth of the nerve fiber.

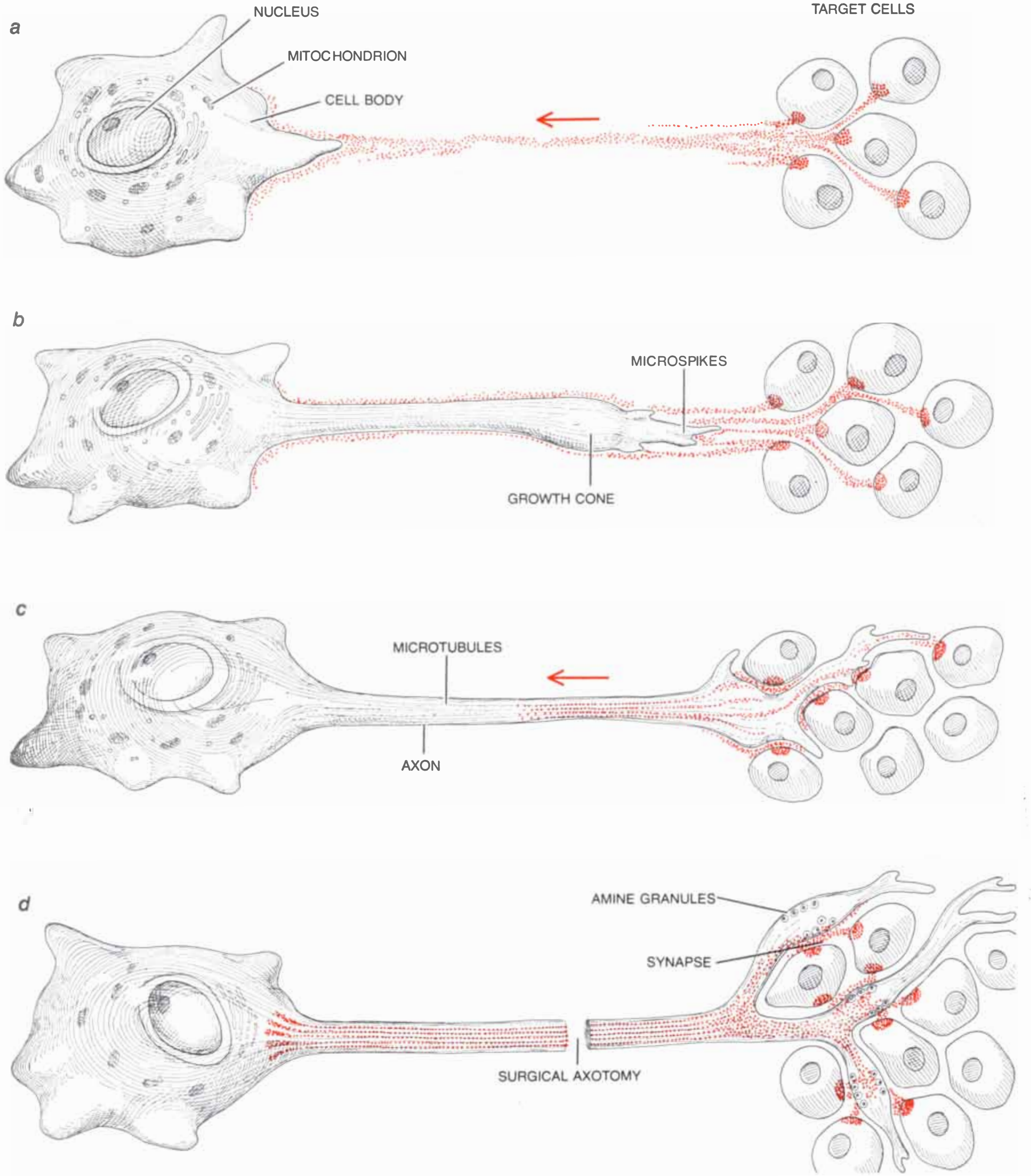
Immature sympathetic neurons respond to NGF with a burst of metabolic activity that provides the material necessary for the growth of the nerve fiber and the manufacture of molecules of neurotransmitter. The cells make more proteins and lipids, take up amino acids from the surrounding medium and burn glucose and other energy-rich compounds at a faster rate. These effects can be seen as a general response to the primitive signal conveyed by NGF through its specific receptors on the cell surface.

Within a short time after the binding of NGF to its receptors the protein constituents of the cytoplasm of the immature sympathetic neuron are profoundly rearranged. In particular the filamentous structures called microtubules and microfilaments come to fill all the available space between the cell nucleus and the cell membrane. These filaments play a key role in the growth of the nerve fiber by providing a structural framework and the propulsive force for its elongation.

How does NGF control the assembly of these filamentous proteins? One possibility is that it acts directly to enhance the polymerization of tubulin and actin, the monomeric proteins that give rise respectively to microtubules and microfilaments. Working at the Laboratory of Cell Biology in Rome, we tested the hypothesis by measuring the rate of assembly of tubulin and actin in the test tube in both the presence and the absence of NGF. In the absence of NGF dilute solutions of the proteins were unable to polymerize into filaments (or did so at a very low rate) because the thermodynamic tendency of the monomers to stay far apart in solution was greater than their tendency to aggregate. The addition of NGF to the solution, however, induced a rapid and massive polymerization reaction. Subsequent investigation revealed that NGF joins together the minimum number of monomers needed to initiate polymerization, after which the reaction proceeds spontaneously to completion. On the basis of these test-tube findings we have hypothesized that the massive formation of mi-



NEUROTROPIC EFFECT of NGF was demonstrated by injecting the factor into the brain of newborn rats for 10 days; the control animals were injected with saline solution only. In the NGF-treated animals (*bottom*) the growth factor diffused from the injection site to the spinal cord and reached the adjacent sympathetic ganglia through the motor and sensory roots. The NGF induced the outgrowth of sympathetic nerve fibers (*color*), which gained access to the spinal cord through the spinal roots and extended as far as the site of injection in the brain. This abnormal projection of sympathetic fibers (which usually innervate peripheral tissues) in response to NGF provided evidence for the hypothesis that a gradient of NGF guides the direction of growing sympathetic fibers in the periphery. In addition injection of NGF into either the brain or the systemic circulation enhanced the peripheral branching of the sympathetic fibers. Experiment was done by M. L. Menesini Chen and J. S. Chen in authors' laboratory.



EXTENSION OF THE NERVE FIBER from an immature sympathetic neuron is induced and guided by NGF. Under normal conditions the target cells that are innervated by the sympathetic neurons or that are present along the path of the growing axon manufacture and secrete small amounts of NGF. The growth factor diffuses into the intercellular space and binds to specific receptor molecules on the surface of the immature sympathetic neuron (*a*). The interaction of NGF and its receptor triggers (by a still unknown mechanism) the assembly within the neuron of filamentous proteins called microtubules and microfilaments. These structures are essential constituents of the growth cone, the locomotory organelle of the growing axon. Situated at the tip of the extending fiber, the growth cone possesses motile, exploratory fibers called microspikes, which confer movement and directionality to the growing axon (*b*). Once the axon has established definitive connections with its target cells, the contact consolidates into the physical and functional link known as the synapse. NGF supplied by the target cells is then channeled directly along the axon to the cell body by a mechanism called retrograde axonal transport, which is most likely achieved by means of filamentous structures within the axon (*c*). Blockage of retrograde axonal transport (by cutting the axon, destroying the synapses with 6-hydroxydopamine or preventing axoplasmic flow with vinblastine) results in the degeneration of the sympathetic neuron in the ganglion (*d*). These different treatments all exert their lethal effects by preventing the normal access of NGF to the sympathetic neurons by way of retrograde axonal transport from the peripheral innervated tissues. Antibodies to NGF also kill immature sympathetic neurons by inactivating NGF so that it can no longer bind to neurons and ensure their continued survival.

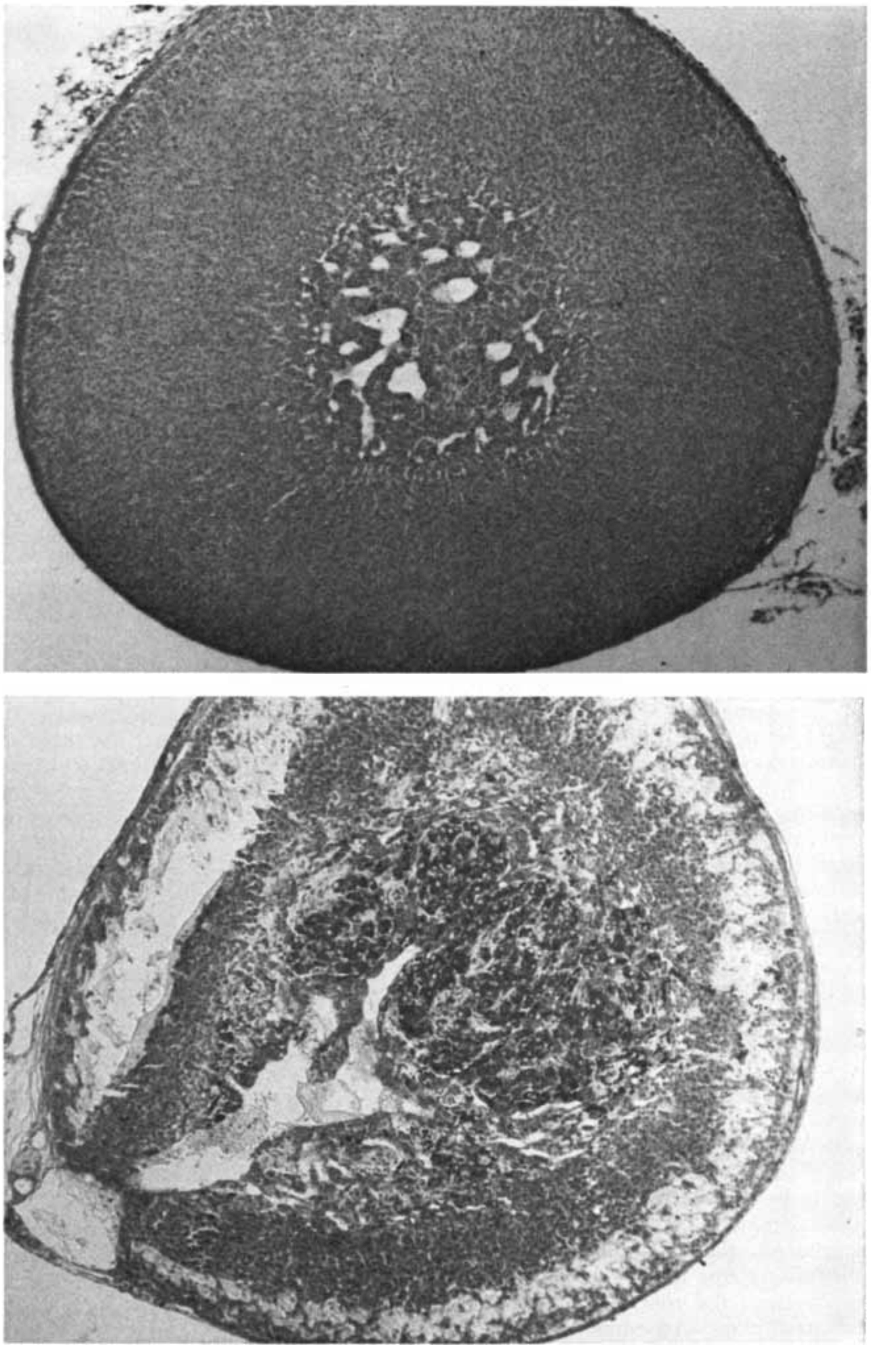
crotubules and microfilaments in the developing sympathetic neuron is triggered by NGF, and that this effect may lead directly to the growth and elongation of the nerve fiber.

An entirely different function of NGF came to light recently with the discovery that certain non-neuronal cells are able to respond to the protein. Lloyd A. Greene and Arthur S. Tischler of the Harvard Medical School demonstrated in 1977 that cells of the line designated *PC*12 (which are derived from a rat tumor) respond to NGF by acquiring properties characteristic of sympathetic neurons. The NGF-treated cells send out fibers, become electrically excitable and store and release neurotransmitters of the catecholamine type. When NGF is withdrawn from the culture medium, the cells retract their fibers, lose their other neuronal properties and resume the uncontrolled proliferation characteristic of neoplastic cells.

Shortly after this discovery K. Unsicker and his colleagues at Johns Hopkins University found that immature chromaffin cells obtained from the medulla (inner part) of the adrenal gland and cultured in the presence of NGF acquire the biochemical and morphological properties of sympathetic neurons. Subsequent experiments carried out at the Laboratory of Cell Biology in Rome by Luigi Aloe and one of us (Levi-Montalcini) demonstrated that this remarkable transformation can also take place in the intact animal: repeated injections of NGF into rat fetuses, continued for two to three weeks after birth, resulted in the differentiation of chromaffin cells into sympathetic neurons in the core of the adrenal gland! These findings indicate that NGF plays a much broader role in the living organism than had been supposed.

It should be obvious to the reader that the investigation of NGF, which began more than two decades ago, is far from over. Among the many unanswered questions is why NGF is manufactured and secreted by the venom gland of snakes and the salivary gland of rodents, even though neither of these glands is essential for the life of the organism and the sympathetic neurons that depend on NGF for their survival. The molecule is also manufactured and released in minute amounts by a wide range of normal and neoplastic cells.

The discovery of NGF has made it possible to experimentally manipulate sympathetic neurons with considerable ease, both in tissue culture and in the living animal. With the aid of purified NGF and NGF antibodies the neurobiologist is now able to increase the number of sympathetic neurons in an animal tenfold or to eliminate them altogether. Moreover, the catecholamine-secreting neurons in the brain are strikingly similar to the sympathetic neurons in the peripheral ganglia in both their morphological and biochemical properties. Already information gained from the study of sympathetic neurons has proved applicable to the treatment of disorders of the central catecholamine-secreting neurons, such as Parkinson's disease. A. Björklung, B. Bierre and U. Stenevi of the University of Lund recently obtained preliminary evidence that the catecholamine-secreting neurons in the brain respond to NGF with a profuse branching of their nerve fibers. If these findings are confirmed, a powerful tool will become available for modulating the function of such brain circuits, which play a crucial role in many kinds of behavior.



NGF TRANSFORMS ADRENAL CELLS into sympathetic neurons when it is injected into fetal rats before and after birth. These equatorial cross sections of the adrenal glands from two-day-old rats show the effect of injections of saline solution alone (*top*) and an NGF solution (*bottom*). The normal adrenal gland consists of a dense cortical layer of cells surrounding a central medullary area made up of adrenal cells and blood vessels. NGF treatment results in a total rearrangement of the gland. The central medulla dramatically increases in volume as a result of the transformation of the adrenal cells into sympathetic neurons, which sprout a large number of profusely branching fibers. These fibers form a large bundle that replaces the inner cortical layer and cuts across the cortex, running in a circular fashion around the external zone. The experiments were performed by Luigi Aloe of the Laboratory of Cell Biology in Rome.

Automatic Control by Distributed Intelligence

Microelectronics makes it possible to run an industrial plant with a multitude of "chips" interconnected in a hierarchy extending from the control of one machine to the integration of the entire plant

by Stephen Kahne, Irving Lefkowitz and Charles Rose

The advent of the digital control computer in the mid-1950's initiated a revolution in the control of industrial plants. The revolution is, however, still in its infancy. It nonetheless promises to have as profound an impact on technology, economics and society as the Industrial Revolution had. The early expectation was that the computer would make it possible to go beyond the simple objective of maintaining process variables at specified fixed values toward the more interesting objective of determining how those variables should be changed with time or with respect to other variables, so that the performance of the entire system could be continuously optimized. The trouble was that the early control computers were seriously limited in speed, in memory capacity and above all in the sophistication of their programs. As a result many of the initial attempts at optimization through computer control, although boldly conceived, fell far short of expectations. In that period the ultimate goal was the automatic control of an entire plant through a large central computer, which would receive inputs from a multitude of sensors, digest their readings and issue the appropriate commands to actuators controlling process variables of every kind. The goal proved to be too ambitious.

With the advent of the minicomputer and more recently the microprocessor (the "computer on a chip") the entire approach to the computer control of industrial systems has changed. Today's approach to the control of a complex industrial plant is based on the concepts of decentralized control, distributed data acquisition and distributed information processing. The implementation of these concepts calls for a hierarchy of computers, each performing tasks appropriate to its position in the hierarchy. It now appears that this approach may make it possible to reach the elusive goal of fully integrated plant control.

Distributed processing is a term that refers to any computing environment in which multiple loosely coupled computer systems implement a given application. The term is commonly used in the context of data processing, where multiple medium- or large-scale computers are interconnected to satisfy the data-processing needs of an organization. Although the linking of closely cooperating multiple microprocessors to realize the control of an industrial process can also be called distributed processing, it is qualitatively different. To describe a plant environment controlled by multiple microprocessors we shall therefore rely on the term distributed intelligence.

Automatic control by distributed intelligence has, among other advantages, great design flexibility together with improved reliability and performance. As a result the application of computers in industrial control systems has grown apace, and today it is estimated that nearly half of the minicomputer installations in the U.S. are associated with control.

With the rapid parallel advances in computer hardware, computer software (programs) and control theory it is now feasible to design control systems in which all the factors influencing plant performance—interactions and complex feedback paths of all kinds—are taken into account to achieve an overall optimum performance. An advanced integrated system will include, in addition to the normal process-control functions, such higher-level functions as startup and shutdown procedures, diagnostic programs for responding to unexpected events, monitoring and alarm systems, and devices that make possible rapid and efficient interaction between man and machine.

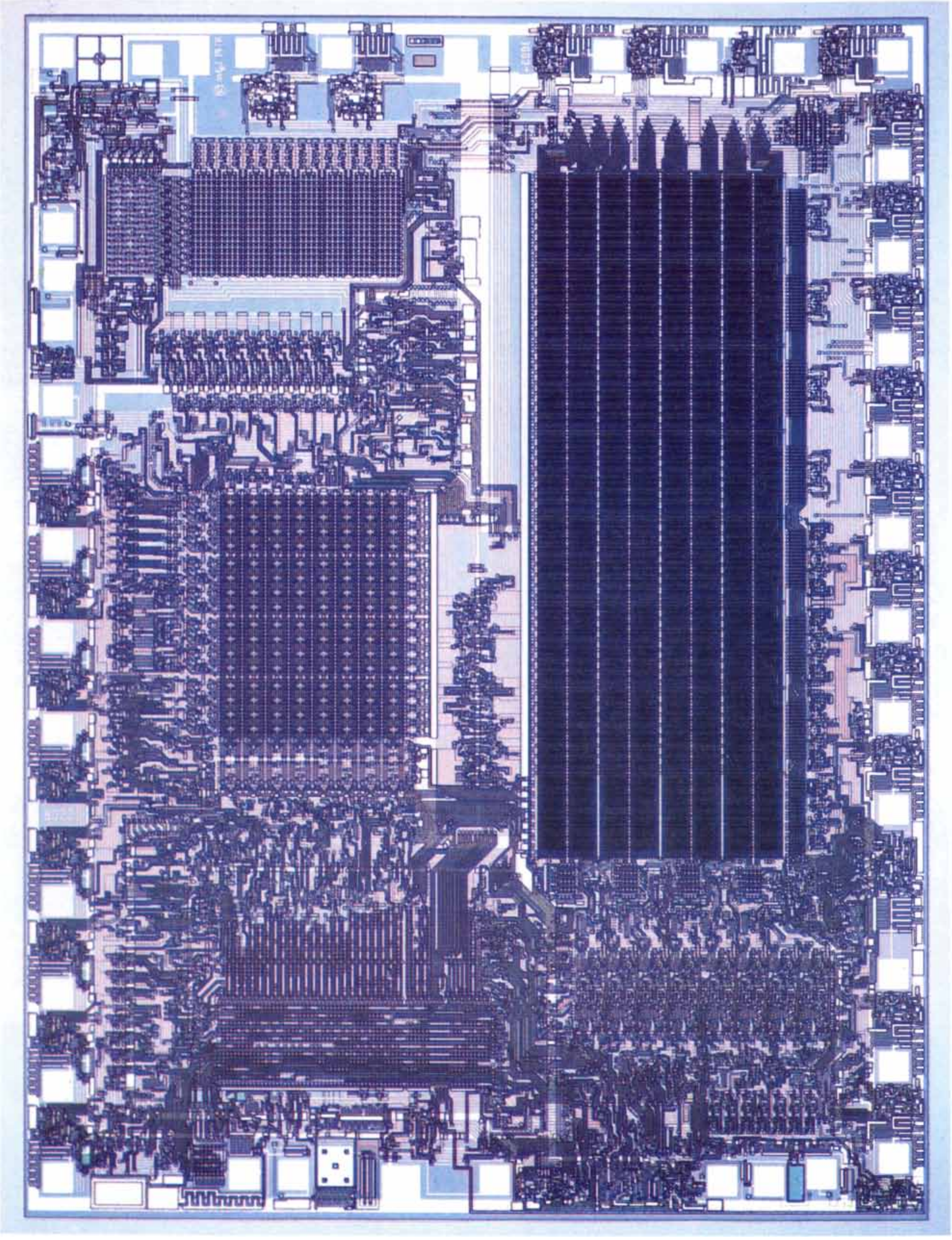
We shall use the term control in its broadest sense to embrace all aspects of information processing and decision making that result in actions affecting the controlled system in a desired way. Here, however, the object of control is the industrial plant. The plant may be a facility producing goods or services or may be a complex of production units making up an entire company.

The factors that affect the operation of the plant system are referred to as inputs. We distinguish between two types of inputs: disturbance inputs and control inputs. Disturbance inputs enter the system from external sources and tend to interfere with the desired performance of the system. They usually represent the effects of interactions with other plant units and with the external environment. Control inputs are those generated by the controller or decision maker. They are usually derived from measurements of the outputs of the plant and are designed to drive the outputs to desired values.

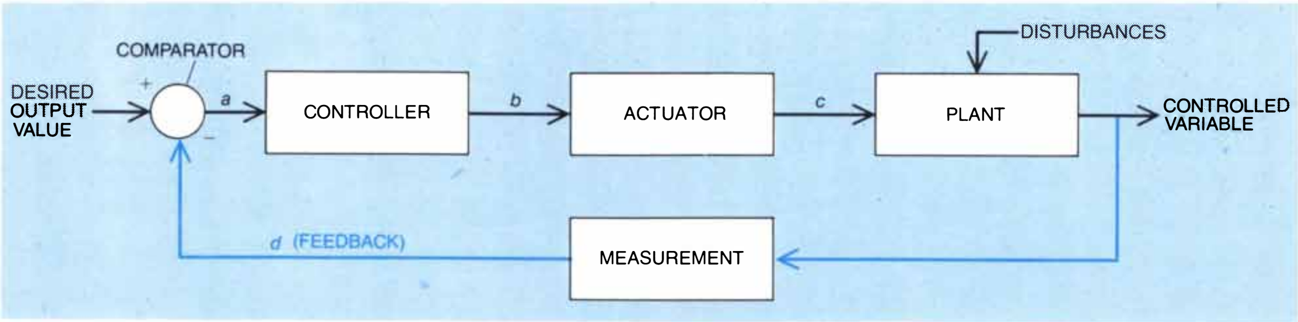
Outputs are defined as the variables in the operation of the plant that result from plant inputs and that relate to the overall objectives of the system. The objectives are usually more complex than simple maintenance of plant production at a constant high level. For example, a steel mill may have the goal of maximizing productivity while minimizing energy consumption and pollution. Moreover, the goal must be accomplished within some set of allowable constraints such as those imposed by equipment capacity, wage levels and Government regulations.

Perhaps the biggest challenge in the control of a system is how to define the goal in a mathematical form that appropriately represents the desired behavior of the system and that can be evaluated from measurable quantities in the system. The task is particularly difficult in large systems where it is necessary to deal not only with the usual technological variables but also with social ones.

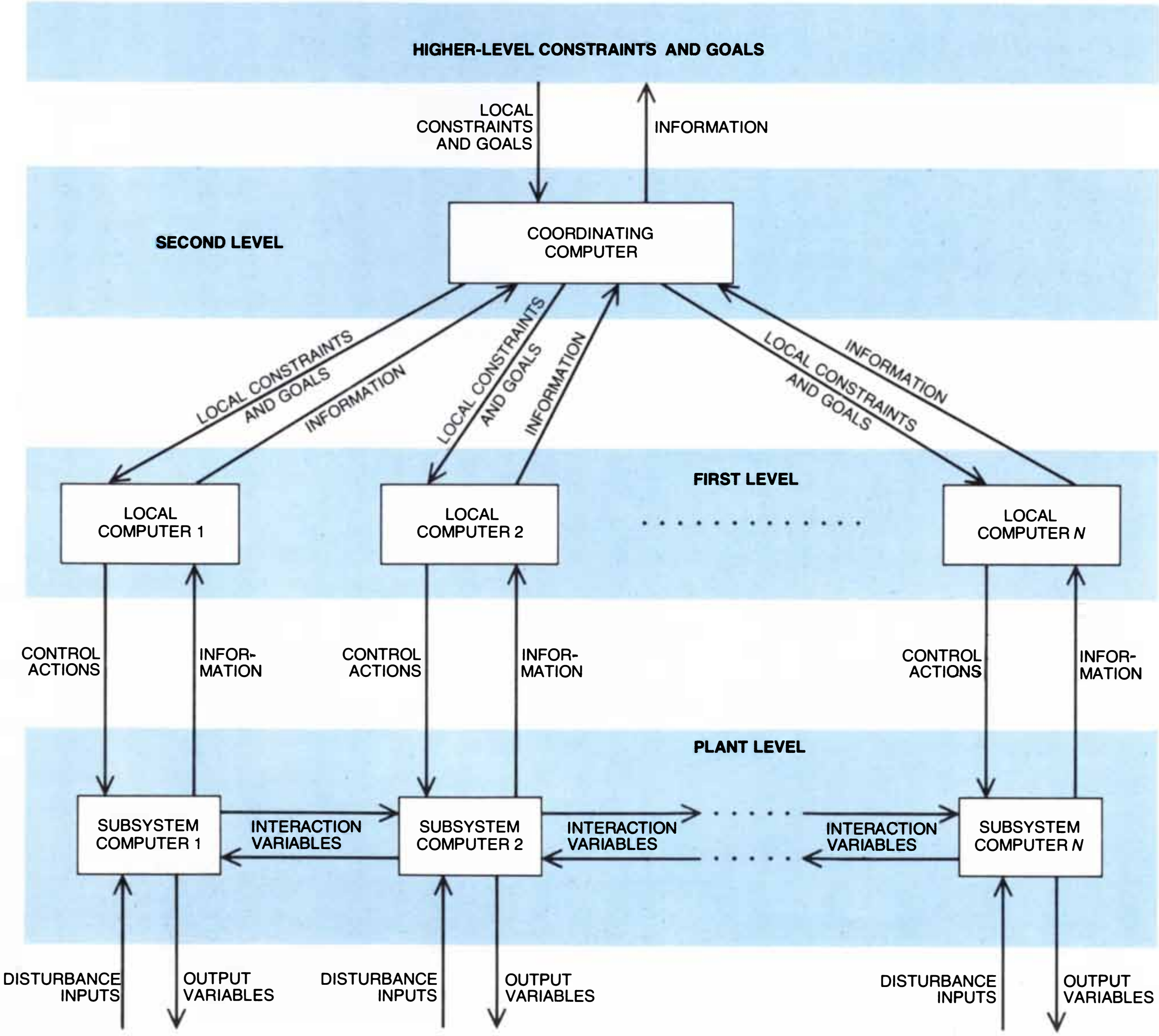
As the plant or system to be controlled increases in size the problem of control increases rapidly in complexity. Most large companies have production facilities at several locations. The production facilities of multinational com-



MONOLITHIC MICROCOMPUTER, the Model 8022 recently introduced by the Intel Corporation, is one of a new generation of computers on a chip that are accelerating the trend toward the distributed, hierarchical control of industrial processes. Measuring .41 by .54 centimeter, the Model 8022 will sell in quantity for less than $10. The eight-bit (eight-binary-digit) microcomputer includes a 16,000-bit read-only memory (occupying most of the right half of the chip) and a 512-bit random-access memory (the large rectangle in the middle of the left half of the chip). To facilitate interfacing with analogue signals the device includes an analogue-to-digital converter, together with provision for interrupts to handle the unscheduled events that frequently arise in industrial-plant operations. The central processing unit of the chip, which can perform 100,000 arithmetic and logic operations per second, fits into the small area at the lower right.



FEEDBACK CONTROL SYSTEM provides a general method for maintaining a physical variable at some desired value. The variable to be controlled may be temperature, pressure, velocity, thickness or anything else that can be quantified. The actual output is measured and converted into a signal (*d*) that is fed back and compared with the set point, or desired value, to generate an error signal (*a*). The controller is a data processor that analyzes present and past error signals to generate a control signal (*b*), which drives an actuator. The actuator is a device that converts the control signal into an action on the plant (*c*) that alters the controlled variable in the desired direction. The plant is also driven by disturbances that enter from outside the control loop and tend to cause the plant output to deviate from the desired value. The desired value is selected outside the control loop either by a human operator or by a supervisory automatic controller.



DISTRIBUTION OF CONTROL FUNCTIONS can be achieved by dividing a plant into a number (*N*) of subsystems and subjecting each to the control of a local computer, representing the first control level. The local controllers are responsible for maintaining subsystem variables at values consistent with local objectives and constraints. The subsystems generally interact with one another, and it is necessary to coordinate actions of local controls so that overall plant goals are satisfied. This is the role of the computer at second level.

panies are distributed worldwide. Many facilities consist of multiple production units, with each unit having many components calling for hundreds of control elements. The geographical distribution of facilities of course implies the distribution of other corporate activities, including managerial and financial activities. It is therefore reasonable to distribute the functions of data gathering, information processing and control in a way that matches the distribution of plant facilities and functions. Since the control function is to an increasing degree being implemented by digital hardware, it is expedient that the hardware too be distributed.

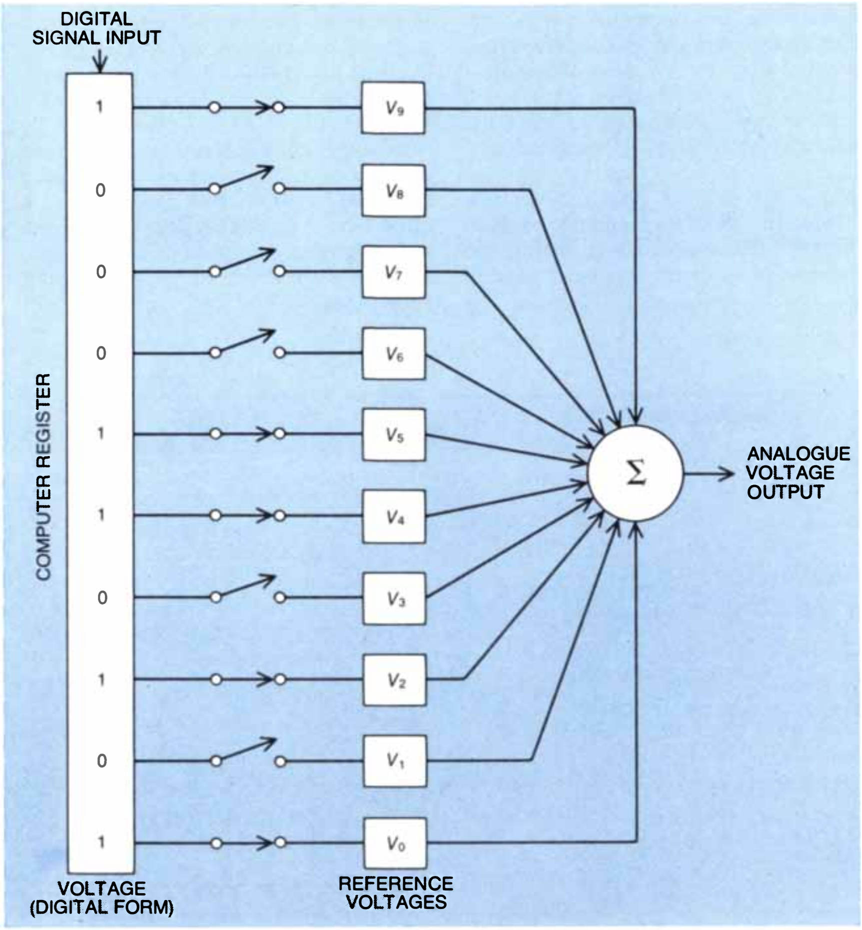
A central concept in control is feedback. Feedback signals communicate with the control element to indicate how the controlled variable is deviating from the desired setting. The control element responds by trying to eliminate the deviation. Feedback loops are incorporated in distributed control systems both locally and globally, that is, systemwide. Local feedback controllers serve to improve dynamic response and to maintain plant variables at the desired values in spite of disturbing inputs. Globally, feedback information from various parts of a large plant serves to modify production policies and to compensate for unexpected changes in the behavior of the system beyond those caused by local disturbances.

A major difficulty in the control of complex systems is that overall performance is affected by the behavior of each element of the system and the elements interact. This means that control actions imposed on one part of the plant often have an effect on other parts. In order to integrate the operation of all the components into a goal-seeking unit the distributed control functions must be coordinated through a hierarchy of control computers organized into a multilevel system.

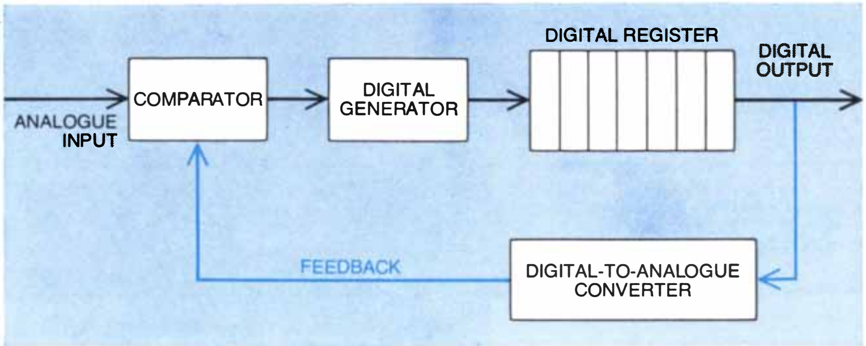
One of the more advanced industrial applications of computers for real-time information processing and control is in steelmaking. We shall therefore use the steel industry for most of our examples. It also happens that our research group concerned with the control of complex systems at the Case Institute of Technology has directed a number of studies to the special control needs of the steel industry. Steelmaking has four basic steps: the reduction of iron ore to produce molten iron, the purification of the iron to form steel, the rolling of the steel into various forms and the metallurgical treatment of the steel to meet product specifications.

The control requirements of the rolling of steel in a hot-strip mill are particularly demanding because of the high speed of the process and the close tolerances in the operating conditions. The hot-strip mill turns out the sheet steel needed for automobile bodies, household appliances and myriad other products. The steel enters the mill in the form of red-hot slabs about a foot thick. The slabs pass through a sequence of stands, or sets of rolls, that squeeze the steel progressively thinner. As the steel leaves the last set of rolls, reduced to a thickness of .01 inch or less, it can be traveling at 35 miles per hour.

A typical hot-strip mill may have a dozen or more stands. The thickness of the steel leaving each stand is governed by the gap between the rolls. The gap is regulated by a controller that positions



DIGITAL-TO-ANALOGUE CONVERSION OF SIGNALS is commonly required in industrial control systems at the point where a low-level digital control signal needs to be converted into a high-power signal in order to control a motor or some other kind of actuator in the plant. The digital signal is converted into analogue form by associating each bit in the digital signal with a reference voltage. If reference voltages increase by a factor of two from one level to the next, V_9 has 512 times the value of V_0. The voltages are summed in a voltage adder (Σ).



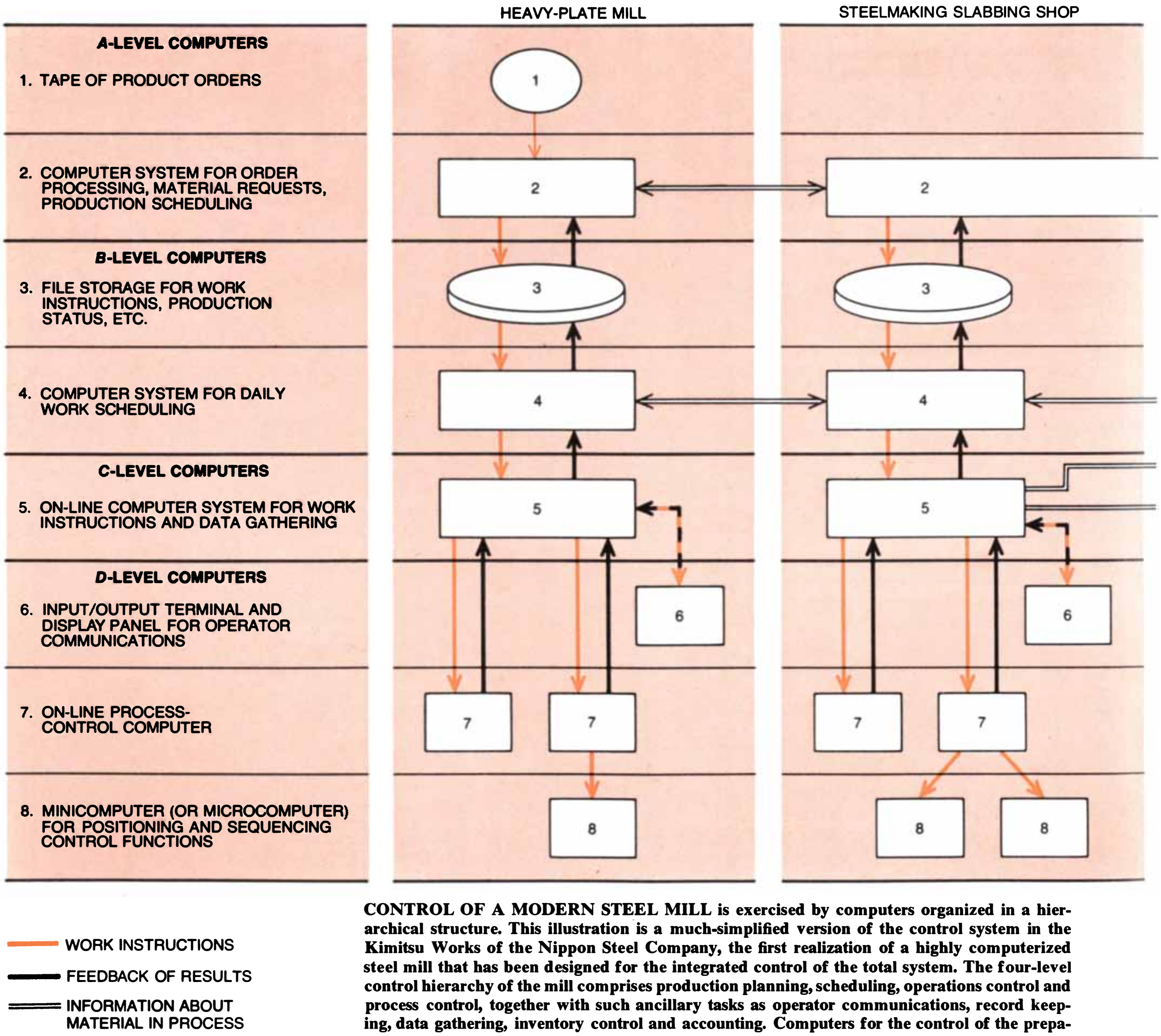
ANALOGUE-TO-DIGITAL CONVERTER in one design incorporates a feedback loop so that the digital output can be compared with the analogue input. If the converted signal is too small, the comparator output triggers the digital generator to increase the digital register. In this scheme, which is highly simplified, the loop incorporates a digital-to-analogue converter.

the upper roll by means of powerful motor-driven screws. The forces developed in compressing the steel are so large that they stretch the frame supporting the rolls and bend the rolls themselves. Accordingly the thickness of the strip leaving a set of rolls will depart significantly from the setting of the gap between the rolls. For the precise control of the thickness of the strip leaving the final stand of the hot-strip mill a measurement of the strip's thickness (supplied, for example, by an X-ray thickness gauge) must be fed back so that the roll positions can be adjusted to meet the thickness specification of the product.

In practice it is not necessary to provide this kind of feedback at each mill stand. Elsewhere along the line the thickness of the strip can be estimated with sufficient accuracy from the gap between the rolls, the pressure on the rolls and other operating measurements. The important point is that the way the reductions of the thickness of the strip are distributed among the sets of rolls affects the quality of the product, the total amount of energy required and the rate of production. The distribution needed for good mill performance is determined by a computer at a higher level in the hierarchy that takes into account the type and grade of steel being rolled, the final thickness of the strip, the condition of the rolls and a variety of other factors.

In addition to controlling the gap for each set of rolls it is necessary to closely regulate the rolls' rotational speeds. A mismatch of speeds between two successive sets of rolls would cause the steel strip between them to buckle or tear. Either event would lead to a costly shutdown of the mill. The problem is complicated by the fact that as the strip is squeezed thinner it gets proportionately longer. As a result each set of rolls has to run faster than the preceding set. Synchronizing the roll speeds therefore involves interaction with the roll-gap settings. One common means of adjusting the settings of the successive speed controllers is to measure the tension (or the degree of slack) in the strip between the two sets of rolls and adjust the speed settings to keep the tension constant.

Another important variable is the temperature of the steel at various points in the mill. Slabs entering the mill must first be heated to bring them up to rolling temperature. Controllers adjust the flow of fuel to the furnace so that each slab achieves its required temperature at the scheduled time for the rolling operation. The heating cycles can be



CONTROL OF A MODERN STEEL MILL is exercised by computers organized in a hierarchical structure. This illustration is a much-simplified version of the control system in the Kimitsu Works of the Nippon Steel Company, the first realization of a highly computerized steel mill that has been designed for the integrated control of the total system. The four-level control hierarchy of the mill comprises production planning, scheduling, operations control and process control, together with such ancillary tasks as operator communications, record keeping, data gathering, inventory control and accounting. Computers for the control of the prepa-

managed so that fuel consumption is held to a minimum.

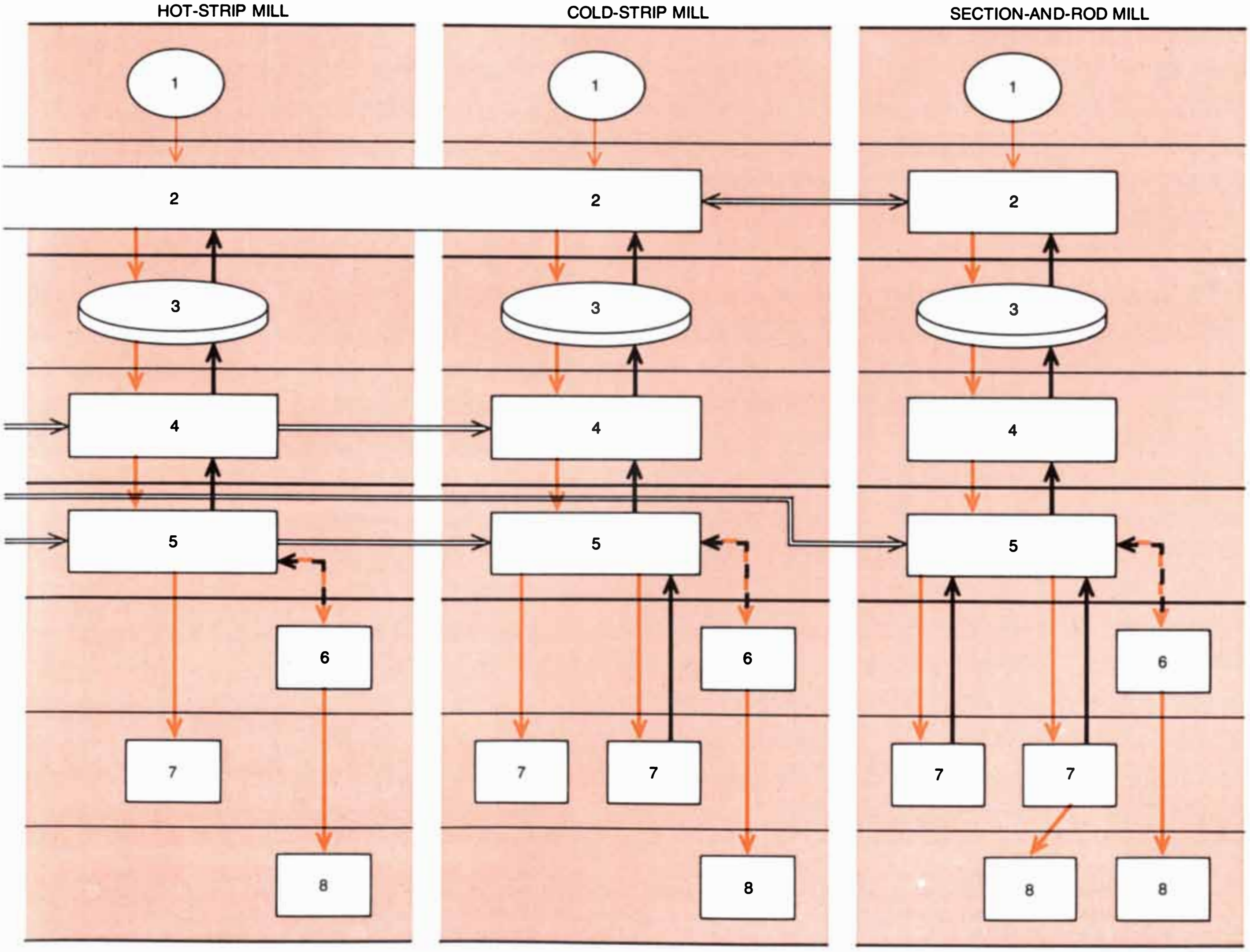
The temperature of the steel strip during rolling affects the metallurgical properties of the strip, and so it is necessary to monitor the temperature of the strip at many points along the mill. The temperature just before the strip is coiled up at the end of the process is particularly critical. Feedback control is therefore provided for varying the spray of coolant applied to the strip in order to maintain the temperature within the permissible range.

Many other kinds of sensors are needed to provide a comprehensive picture of the rolling process. For example, there are position sensors that signal the computer system on the arrival of each new slab so that the appropriate sequence of operations can be initiated. Other information is gathered to diagnose the condition of the mill equipment. When something goes wrong, information is needed to determine the source of the problem and to specify what remedial actions can be taken until repairs can be made. Not least, the sensors provide data for evaluating the performance of the mill, for bringing up to date the coefficients in the various control algorithms (procedural rules) and for recording the history of the operation for accounting, analytical and legal purposes.

All of this has much to do with what is widely regarded as a critical situation in the U.S. steel industry, a situation largely brought about by competition from modern, highly automated steel plants in other countries. In our opinion the introduction of new concepts of distributed intelligence and integrated systems control holds substantial promise for improving the situation by making American steel plants more efficient and more productive.

A digital computer was first installed in an industrial plant some 20 years ago. Its role was to determine the set points of key analogue controllers in order to improve plant performance. The development aroused great interest among control engineers in industry and the universities. The Case Institute's own research program in the control of industrial systems, with which the three of us are associated, originated just before that period.

The next step was what was called direct digital control, in which many conventional process controllers were replaced by a single digital computer. The speed of the computer enabled it to deal with each feedback control loop in se-



ration of the ore, the blast furnaces, energy production and other activities are not shown. Each of the process-control computers at the *D* level in the illustration performs a variety of tasks comparable to those carried out by the on-line control computer in the hot-strip mill. That computer controls, among other things, the charging, the extraction and the temperature of the slab-reheat furnace in addition to controlling the width, thickness and temperature of the steel strip as it travels through the rolls of the mill. The *D*-level computers are situated on the floor of the mill. The computers for the higher levels are in the computing center of the works. The order-entry system in the main office distributes order tapes for each group of products to the order-processing systems at the *A* level. Such division of "intellectual" labor among multiple cooperating processing units is becoming known to computer workers as control by distributed intelligence.

quence, with the entire cycle held within the time constraints of the controlled process. The software held the potential for realizing more sophisticated algorithms. For various reasons, however, including high cost and low reliability, direct digital control was not widely adopted.

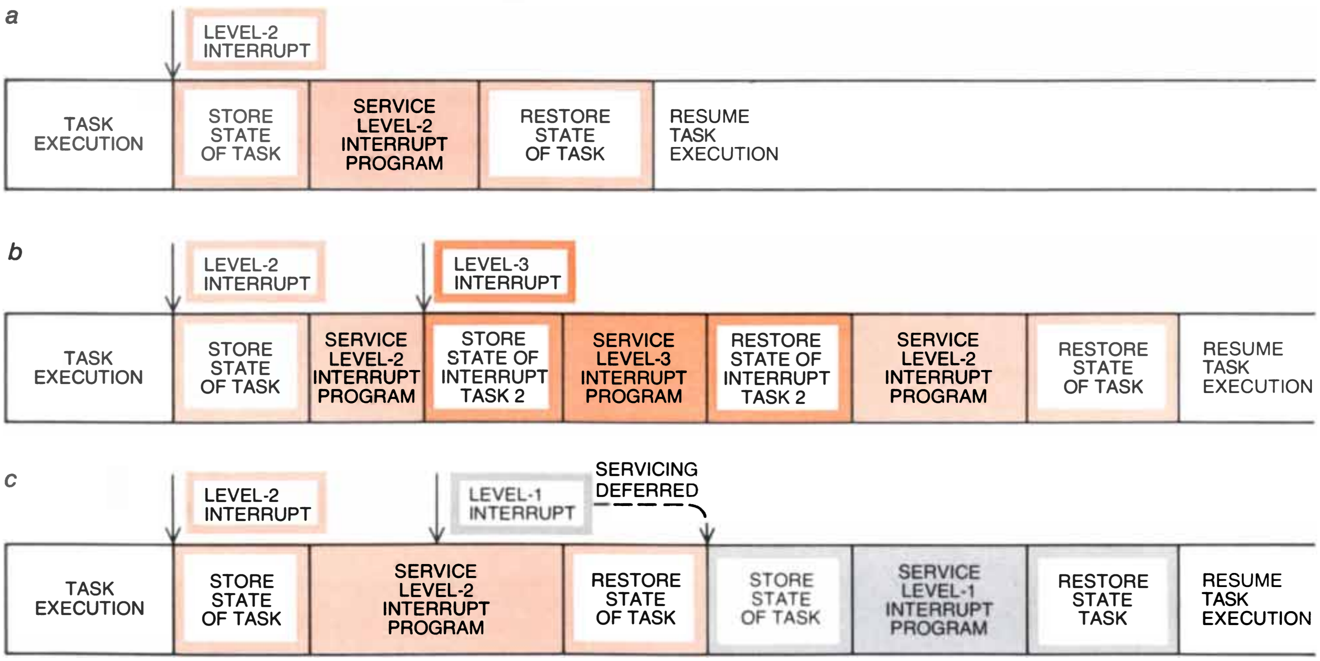
Since then there have been many extraordinary developments in computer technology. With a somewhat arbitrary index (which summarizes price, binary-digit word length, speed, instruction repertory, input-output facilities and software) one can show that the ratio of performance to price for medium-size computers and minicomputers has increased at a rate of more than 50 percent per year since the early 1960's. Between 1966 and 1978 the performance/price index for minicomputers climbed by more than two orders of magnitude. In addition user-oriented programming languages have made it much easier for plant engineers to program, debug and modify control algorithms. Reliability has greatly improved as a result both of the increased reliability of the computer components and of the introduction of fault-tolerant designs, component redundancy and new routines for diagnosing trouble. These multiple advances in computer hardware and software have vastly enlarged the domain of what is technologically and economically feasible in the application of computers to control of industrial systems.

Let us now look more closely at the basic tasks that must be performed by a computer control system and at some of the software problems that must be solved in the control of a single plant process. A computer control system has three primary tasks: data acquisition, control and actuation. The value of a process variable (such as temperature, pressure or flow rate) is transformed by a transducer into an electrical signal. The data-acquisition task, implemented in software, is to convert the signal into usable form, check the resulting quantity to ensure that it is within the expected limits and make the data available to a control task. The control task combines the measured value with other information about the system to compute a value for the control input signal, which is transmitted to the actuator (such as a flow-control valve or a positioning-motor drive).

Since measurements of process variables are usually in analogue form, and since the digital computer can process data only in digital form, the analogue data must be converted into binary digits, or bits (0's and 1's). The computer's output must then be reconverted from digital into analogue values so that it can be accepted by the actuators. Such transformations are carried out by special hardware: analogue-to-digital and digital-to-analogue converters.

The frequency with which the process is interrogated by measuring devices and subjected to control actions can vary widely. For example, the temperature of steel slabs in the exit zone of the reheat furnace of a hot-strip mill need be sampled only once every few minutes. On the other hand, the tension of the steel strip traveling between rolling stands must be measured many times a second. Still other quantities, such as the number and type of slabs to be rolled in the mill, need be summarized and sent to a higher-level computer no oftener than once per work shift.

Unscheduled events that abruptly change the state of the system call for an immediate and specific response from a designated computer in the system. Typical of such events are the malfunction of a piece of equipment, the need to change a set of rolls in the mill and the receipt of an urgent order to be scheduled into the production system. The computer control system is equipped to handle such events by means of an interrupt facility. The program currently being executed is interrupted and the information contained in various controllers at that moment is saved, usually by



INTERRUPT ROUTINES are a necessary part of the programs in computer control systems. Such routines enable a computer to set aside the execution of its normal task to respond to an unscheduled event, such as a malfunctioning piece of equipment or a power failure. Such emergencies are assigned priorities so that a more urgent event will be able to interrupt the servicing of a less urgent one. In the scheme that is depicted in this illustration the program responds to a priority-2, or level-2, interrupt event (*a*) by storing the "state" (the information content) of the normal task that is being executed. After the level-2 interrupt program has been serviced the state of the normal task is restored and its execution is resumed. The servicing of a level-2 interrupt can itself be interrupted by a higher-priority level-3 event. In that case (*b*) the state of the interrupt task in level 2 is stored while the level-3 interrupt is being serviced. The servicing of the level-2 interrupt is then resumed, followed by a resumption of the execution of the normal task. If in the course of the servicing of a level-2 interrupt an event of lower, or level-1, priority occurs, no action is taken until the servicing of the level-2 interrupt is completed (*c*).

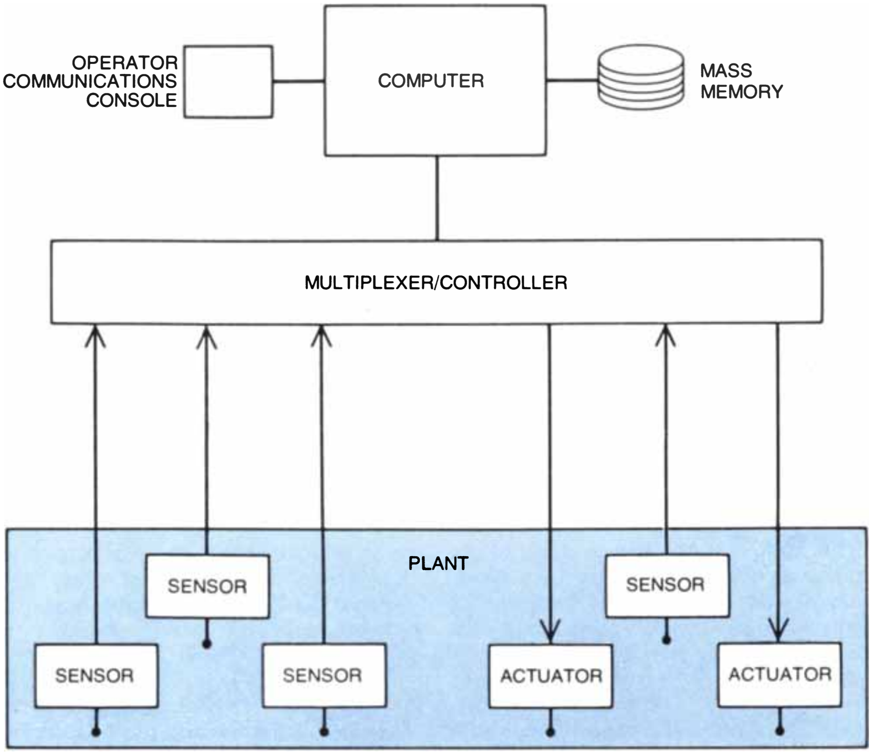
the operating system. The program responsible for servicing the external stimulus is activated, and when it is completed, the execution of the interrupted program is resumed. In this way the unscheduled external event is synchronized with the control-system software.

The software of most computer control systems incorporates a means of categorizing unscheduled events, so that priorities can be assigned to them. Typical of these categories are breakdowns in operator communications, power failures, alarms that certain process conditions are seriously in error and routines to coordinate the peripheral devices of the computer. Events in such categories are assigned priorities according to their urgency and are given access to the control system through a multilevel interrupt system. Thus interrupt tasks can themselves be interrupted by events of higher priority.

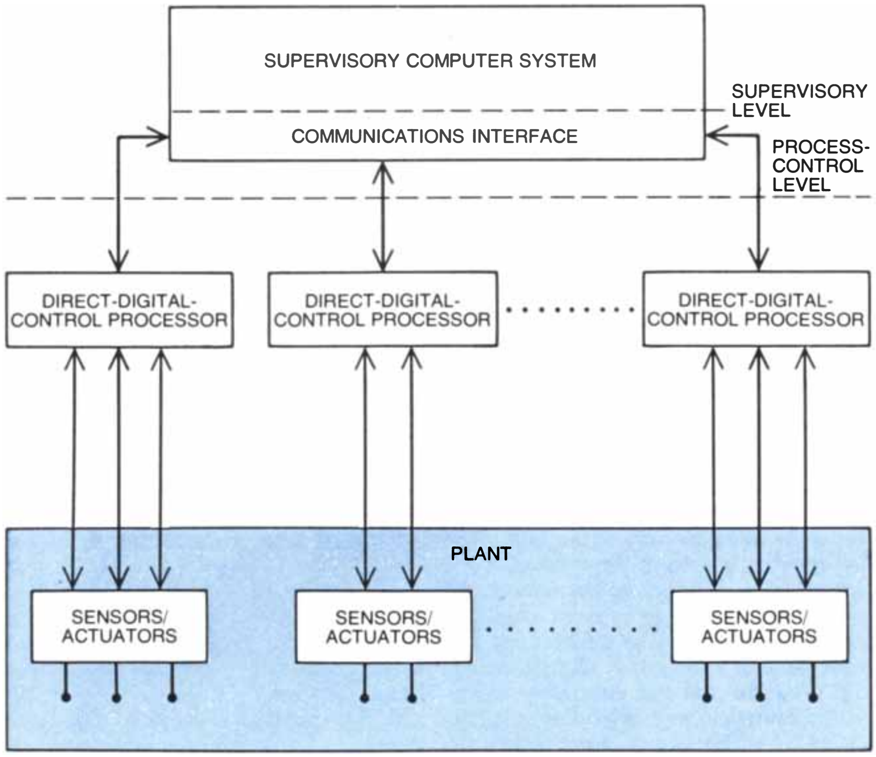
The timekeeping process in a real-time computer control system is realized by connecting the computer to a "clock," or counter, that interrupts the computer at specified intervals, usually every few milliseconds. By keeping a running sum of the clock events in internal computer memory the control program can determine arbitrary intervals and even times of day.

In large-scale industrial processes thousands of variables must be controlled and therefore many different inputs and outputs must be sensed and acted on. It was because of the complexity of the interactions among these variables that early efforts toward the direct digital control of industrial processes were hampered by the limited speed and storage capacity of the available computer control systems and by the lack of adequate software. The first direct digital control system (at the Texas City plant of the Monsanto Chemical Company in the early 1960's) could handle only 10 control loops because of hardware limitations, and at that time computer manufacturers had developed virtually no software support programs. Hence several completely independent computer systems were needed to handle the number of control loops in even a fairly simple plant. It was largely because of the absence of effective software to coordinate the control efforts of the several computers that the performance of the early plantwide direct-digital-control systems was disappointing. These installations were costly not only because several computers were needed but also because plant managers insisted on having conventional analogue controllers as backup systems.

At that time many thought a large-scale central computer with the speed, memory capacity and computational power to handle all the process



CENTRALIZED DATA ACQUISITION AND CONTROL seemed a promising approach in the 1960's with the introduction of large mainframe control computers by IBM, General Electric, Bunker Ramo and other manufacturers. One central computer facility receives data from all sensors in the plant, together with operator commands and other data previously stored in mass memory. The computer processes the data and returns commands to actuators. Multiplexer/controller arranges data for funneling into computer and for dispersion of commands.



EARLY HIERARCHICAL CONTROL SYSTEMS became feasible with the advent of the minicomputer in the late 1960's. The direct digital control of plant operations could be split up among minicomputers while supervisory control was retained in a large central machine.

variables would finally solve all the problems encountered in direct digital control. It was argued that the centralization of control would also achieve the desired optimization of plant performance because all the plant data would be concentrated in a single computer site. The introduction of a generation of large mainframe control computers in the late 1960's (such as the IBM 1800, the General Electric 4060, the Bunker Ramo 430 and many others) provided an opportunity to test the concept of centralized control. Since the new machines cost hundreds of thousands of dollars, they were suitable only for large installations.

Although the large computers proved that the integration of control was feasible and did improve plant performance, they introduced perplexing new problems. One was that since much of the plant was being controlled by a single machine, the reliability of that machine was a major concern. In most cases the installation of a standby computer was considered to be prohibitively expensive. (Redundant computers were, however, installed in the control room of a few large electric utilities.) Moreover, the diagnostic software needed to identify incipient component failures was virtually nonexistent. As a result the central computer was usually backed up by analogue systems and even manual ones. Another problem with centralized control was the complexity of the input and output data systems, which called for the multiplexing (sequential switching) of hundreds of analogue signals into a common computer input channel. The problem is not easily solved even with the solid-state electronics of today; the solution was exceedingly cumbersome and unreliable with the relay technology of the 1960's.

A bigger problem, the seriousness of which has become clearer with the passage of time, is the complexity of the software needed to control large systems centrally. In order to manage the resources of the control computer and to allocate them among the competing control tasks in real time, elaborate executive programs had to be developed. Such real-time executive programs have five major functions: scheduling the central processing unit of the computer among the principal control tasks, managing the main memory of the computer by dividing it among the tasks as they are scheduled, handling the input and output operations to prevent conflicts among competing tasks, maintaining the data base of the system (consisting of the program and the data files stored in the memory) and providing mechanisms by which two or more tasks can communicate.

The size and complexity of the executive program grows, of course, with the size and complexity of the system to be controlled. Executive programs for today's generalized multitask systems commonly need 18,000 to 25,000 words (consisting typically of 16, 24 or 32 bits each) of main memory and account for 30 to 40 percent of all the computer instructions executed. This large "overhead" dramatically affects the performance of the entire system. Typically hardware can respond to an interrupt signal in between four and 10 microseconds, but the executive program can increase the response time to between 200 and 300 microseconds. Therefore the time from the sensing of an external stimulus to the time the control task is able to respond is between 20 and 80 times longer than the basic response time of the computer. Such a delay is unacceptable in some operations, and so some executive programs incorporate dedicated scan routines that "amortize" the overhead time needed for servicing one interrupt over a large number of input or output operations.

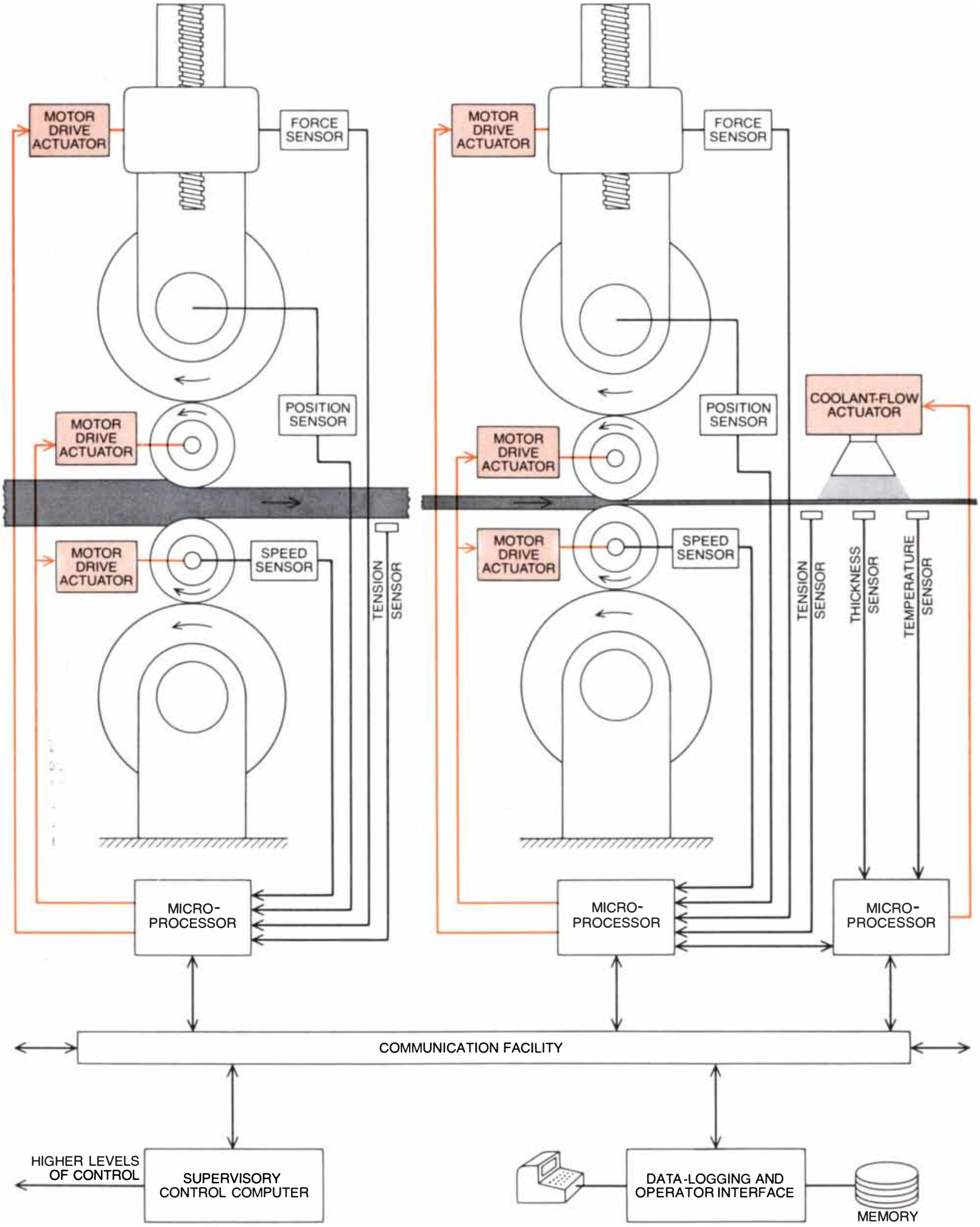
In the early 1960's several groups of investigators advanced the concepts of multilevel and hierarchical control of industrial processes. Among them were Mihajlo Mesarovic, Donald Macko, James D. Schoeffler, Leon Lasdon, Coleman Brosilow and one of us (Lefkowitz) at the Case Institute and Theodore J. Williams and his group at the Monsanto Company and later at Purdue University. It was not, however, until the advent of low-priced minicomputers in the late 1960's that it became feasible to seriously consider the implementation of a hierarchical computer control system. In such a system the supervision of plant operations was divided among a number of minicomputers that reported to and carried orders from a large supervisory computer. The decreasing prices of computer hardware exposed the true cost of developing the necessary software, which in many instances was found to be 10 to 20 times as expensive as the hardware.

Then in 1970 the first microprocessor, the Intel 4004, was introduced. Fabricated on a chip of silicon a quarter of an inch on a side, the 4004 was a four-bit processor: a small computer central processing unit capable of performing arithmetic operations on four bits at a time. It was originally designed for hand-held calculators. The 4004 was followed by the 16-bit Texas Instruments 9900 and the still more powerful models offered by half a dozen manufacturers today. Within only seven years, between 1970 and 1977, three full generations of microprocessors came on the market. Current models are true microcomputers, with a central processing unit, a memory and input and output devices on the same chip. Within the past six months three manufacturers have announced microprocessors with between 15,000 and 20,000 gates on a single chip. (A gate is a group of circuit elements that can perform one logic operation.) The processing power of the new chips exceeds that of many high-performance minicomputers. Within three years such chips will sell for less than $30 in quantity, and a computer consisting of one of them together with a million eight-bit words of memory will probably sell for well under $20,000. Today that amount of memory alone sells for about $35,000 as part of a minicomputer system.

The new microprocessor technology captured the imagination of computer control system designers, who immediately began considering the impact that low-cost and relatively powerful computing could have on software, input/output and reliability problems. Since the absolute computing power of each microprocessor is not great, a multiple architecture based on distributed intelligence suggested itself. If such an architecture were applied to the control of a hot-strip steel mill, each set of rolls could be controlled by its own microprocessor. Inputs to the processor could report the roll position and pressure; outputs could control the motor drives that determine the position and speed of the rolls. Other processors could be assigned to determining the thickness, temperature and tension of the strip and controlling the flow of the coolant spray. The set points of all the control loops might be calculated by a supervisory processor on the basis of algorithms designed to optimize the performance of the mill. Since each processor would be handling fewer control loops than the direct-digital controllers of the earlier systems, the multiplexing of input and output signals would no longer be a major problem.

Much thought has been given to the kinds of communication facilities that might be most effective for interconnecting the microprocessors of a distributed control system. The size of the plants and the cost of wiring have dictated that practically all distributed systems be built around a high-speed communication facility in which each bit of information is transmitted sequentially over a single channel. At the direct-control level all the processing elements usually communicate through one such channel. The channel can consist of a simple twisted pair of wires or, if high transmission rates are required, a coaxial cable or a fiber-optics link. Alternative schemes, in which several streams of bits are transmitted simultaneously over several channels, are much costlier.

Some 10 architectures have been proposed for distributed-intelligence sys-



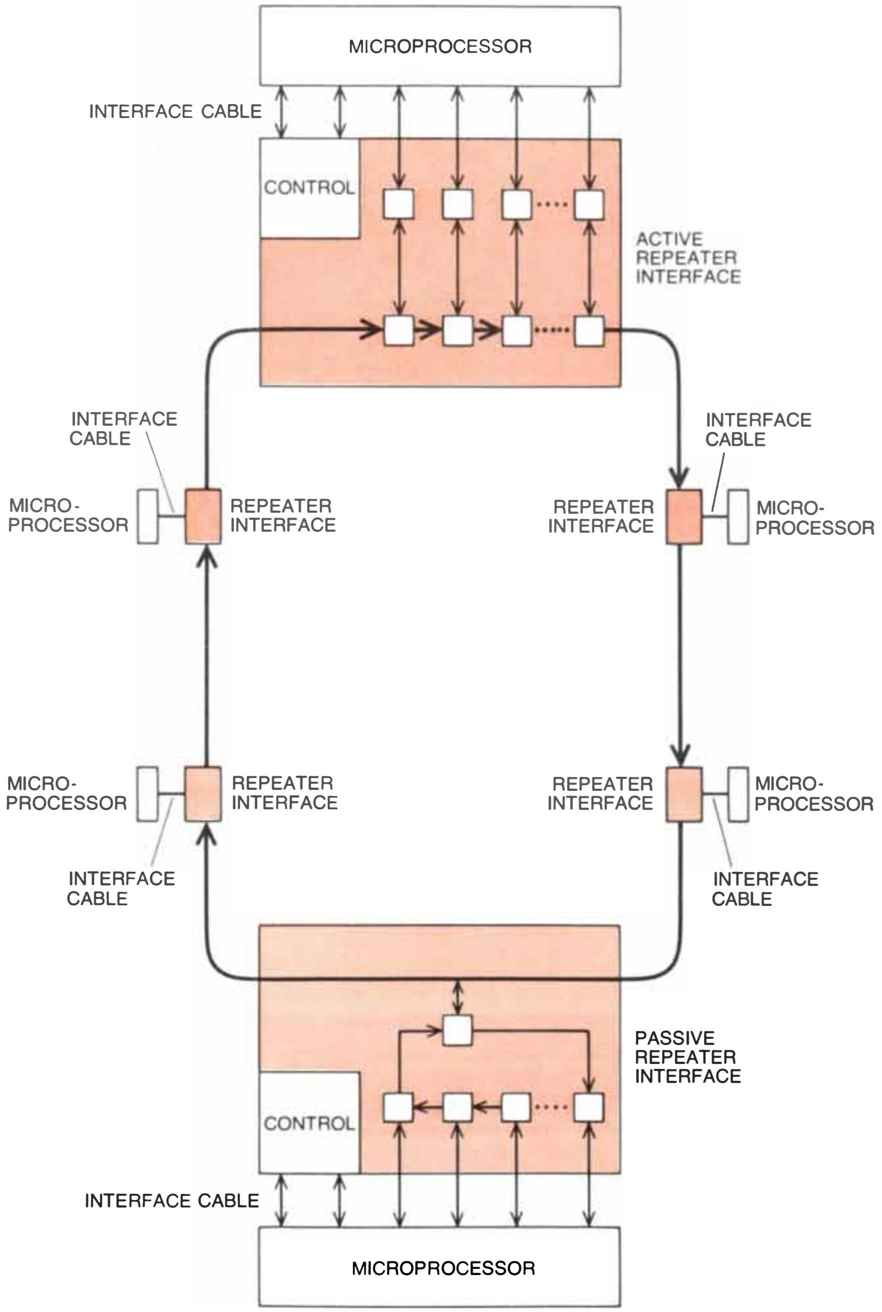
A POSSIBLE APPLICATION of distributed intelligence, which is made feasible by the development of microprocessors, is depicted for two "stands" of a hot-strip steel mill. Each mill stand is serviced by its own microprocessor, which controls the speed and position of the rolls in the stand on the basis of information on the desired and actual speeds and positions. The actual values are provided by speed and position sensors; the set points (desired values) are determined from measurements of strip tension and roll force and from inputs received from the supervisory control computer based on production specifications and on the production rate of the mill. A separate microprocessor (*right*) can be used to control the flow of coolant spray and the roll gap on the final stand in order to maintain at specified values the temperature and thickness of the strip leaving the mill. A typical mill will require many microprocessors for local control, each of which is capable of communicating with other units on the same control level and with other computers that are operating on higher levels.

tems, but only three have been widely applied in data acquisition and control: the ring (or loop) system, the shared global bus and the switched global bus. In the ring system the processing elements and the input and output devices are connected to the ring by simple interface devices that intercept any messages carrying the appropriate address. The interface devices can be either active or passive. Active interfaces are part of the ring, and so every bit of every message passes through them. Passive interfaces are not part of the ring; they simply copy properly addressed messages as they go by.

The interface devices must also be able to insert messages into the ring. Three general schemes are in service. In one scheme each interface has an assigned time slot during which it can insert a message. In a second scheme "null" bytes of data (sequences of eight 0's) are transmitted on the ring when there is no other traffic on it. On detecting a null byte an interface can insert a message. In a third scheme an interface that has a message for the ring "broadcasts" the message repetitively. The system is designed in such a way that within a statistically guaranteed period each message will be transmitted free of interference from any other message. The third scheme requires the ring to operate at a bit rate much higher than that required by the first two schemes if it is to achieve the same message rate, because many messages will be garbled by overlapping with other messages. The scheme has the great advantage, however, that the interfaces can be quite simple and therefore reliable.

The shared global bus is open rather than being closed in a ring. It has an important advantage over the ring in that it transmits messages in two directions. Interfacing is achieved as it is in the ring. The broadcast insertion of messages is common practice. In some systems extra lines are included to allow the automatic arbitration of requests for access to the bus. Alternatively the arbitration device can supervise access to the bus by polling the interface devices on the bus periodically to determine whether any of them have messages to transmit.

In the switched bus, messages from one microprocessor to another are usually sent to a switch, which directs them to the proper destination. Polling is common practice. In an entirely different approach one recently developed system has as its communication path a high-capacity coaxial cable. The cable can carry a broad band of frequencies, so that each processor can transmit messages to the switch at a unique frequency and circumvent the arbitration problem. The switch stores the messages and retransmits them to the destination processor at that processor's own unique frequency. In this way many channels of data, television and voice can be transmitted simultaneously over the same bus. The technique is called frequency-division multiplexing.



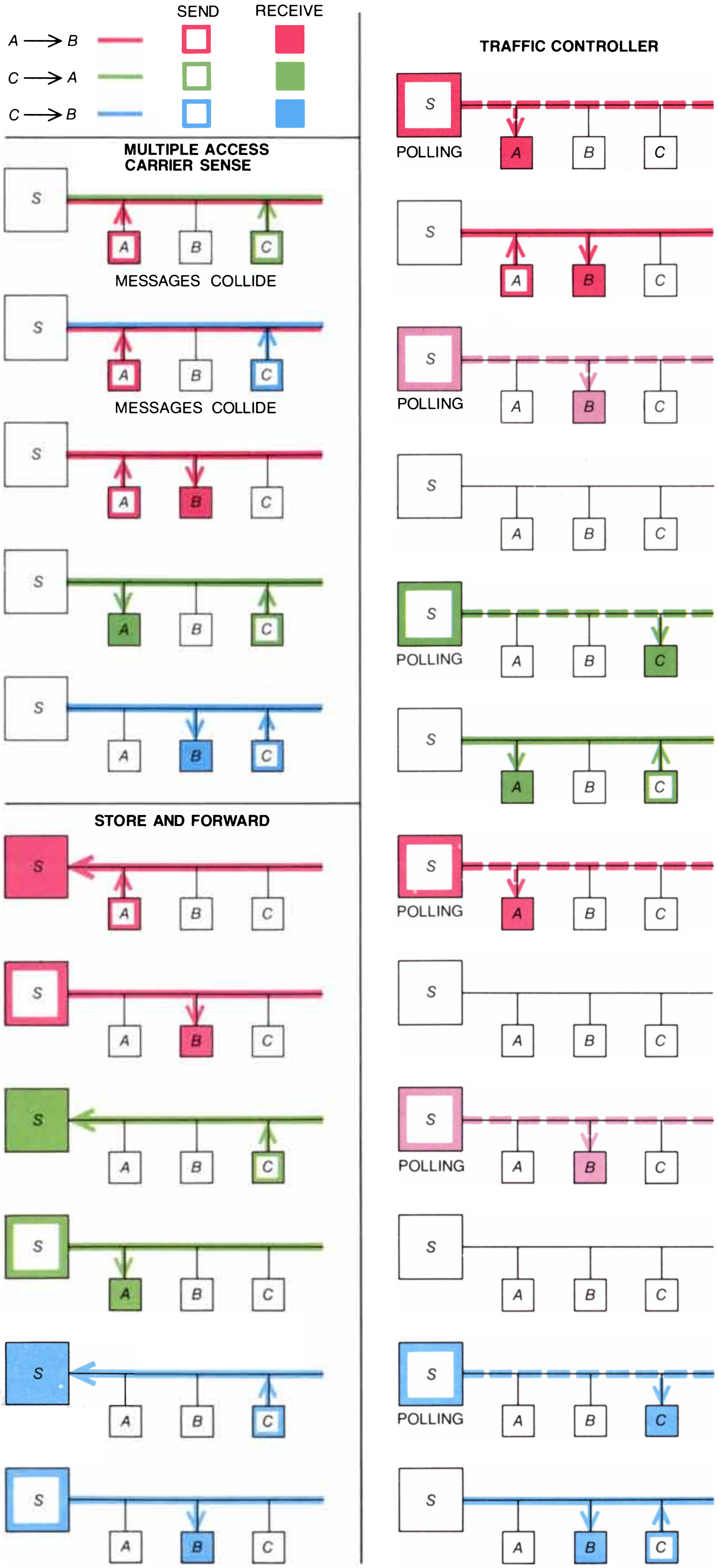
COMMUNICATION FACILITY IN THE FORM OF A RING is widely used for the high-speed transmission of data among microprocessors. Processing elements in the ring have access to the communication channel through repeater interfaces, of which there are two types. Active interfaces are actually part of the ring. Data bits are shifted sequentially through a set of storage elements in the interface as they circulate around the ring. Control circuitry can read messages passing through and can add brief new messages by inserting more storage elements into the path and temporarily lengthening the ring. Passive repeaters, lacking storage elements, simply monitor passing messages. In one scheme they insert messages at fixed intervals. In another they can "broadcast" a message until it has been transmitted without "colliding" with another.

Among the important attributes of any distributed-intelligence control system are modularity, fault tolerance and reliability. By modularity is meant that if the ring or bus system has an adequate address capacity, additional sensors, actuators and microprocessors can be added to it almost anywhere at little cost. Interface devices have a simple logic and are inexpensive compared with the devices they interconnect. Modularity is therefore not difficult to achieve.

As for fault-tolerance and reliability, the buses and interfaces of a distributed-intelligence system are usually installed in duplicate to guard against a failure in communication. Then in the event of a failure the backup facility can be switched in automatically. Today's microprocessors are inherently reliable, but they too can fail. One scheme for dealing with the possible failure of a processor is to install sensors and actuators in duplicate and to link them to two different processors only one of which is actively operating. If the operating microprocessor fails, the backup processor takes over, relying on its own linkages to the sensors and actuators. An alternative is to provide a switching system, so that in the event of a processor failure the sensor data are switched to an alternate processor, which is simultaneously given access to the actuators of the failed unit.

If a microprocessor at any level in the hierarchy above the lowest level of processors fails, its load can be assumed by another processor (or more than one) on its own level, provided the memory of the backup unit contains a copy of the tasks to be assumed. Since messages between tasks in a hierarchical system are sent over the communication facility, each message must include the address of the destination task. The address can be that of the processor discharging the task or the name (the numerical designation) of the task itself. In the first instance the failure of a processor and the subsequent reconfiguration of the system would change the addresses of some tasks. All the tasks would then have to be notified of the changes because the messages would have to be directed to new addresses. This solution is clearly a clumsy one.

The preferable alternative is to provide each communication interface device with the names of all the tasks assigned to the microprocessor to which it is connected. As the messages go by on the line each interface recognizes and



SHARED GLOBAL BUS, with several access variations, is also widely used as a communication facility for linking microprocessors. The bus provides a two-way serial data path. Three access schemes are depicted. In the example shown here the desired sequence of communications from one microprocessor to another is *A* to *B*, *C* to *A* and *C* to *B*. In the simplest scheme, known as carrier-sense multiple access, passive interfaces with messages to send check the bus for the presence of a message in transit. If no message is present, they broadcast either immediately or after a random interval. If two or more interfaces transmit at the same time, there is a collision and both messages are lost. Each interface then waits for a random interval and retransmits. In some designs an "access arbitrator" is included to determine which interface may use the bus according to some priority rule. In a second scheme a switch can be used to store and forward messages from one device to another (*A*→*S*, *S*→*B* and so on). In a third scheme a "traffic controller" polls each interface to learn if it has a message to transmit. If the controller polls *A* (*S*→*A*) and finds *A* has a message, *A* is given access to bus (*A*→*B*).

intercepts those pertaining to its tasks. This scheme will also handle failures in the data-acquisition processors. By sending to one another periodic "Are you well?" messages processors can determine when a neighboring processor has failed and initiate reconfigurations that redistribute the work load. This approach is called soft task addressing.

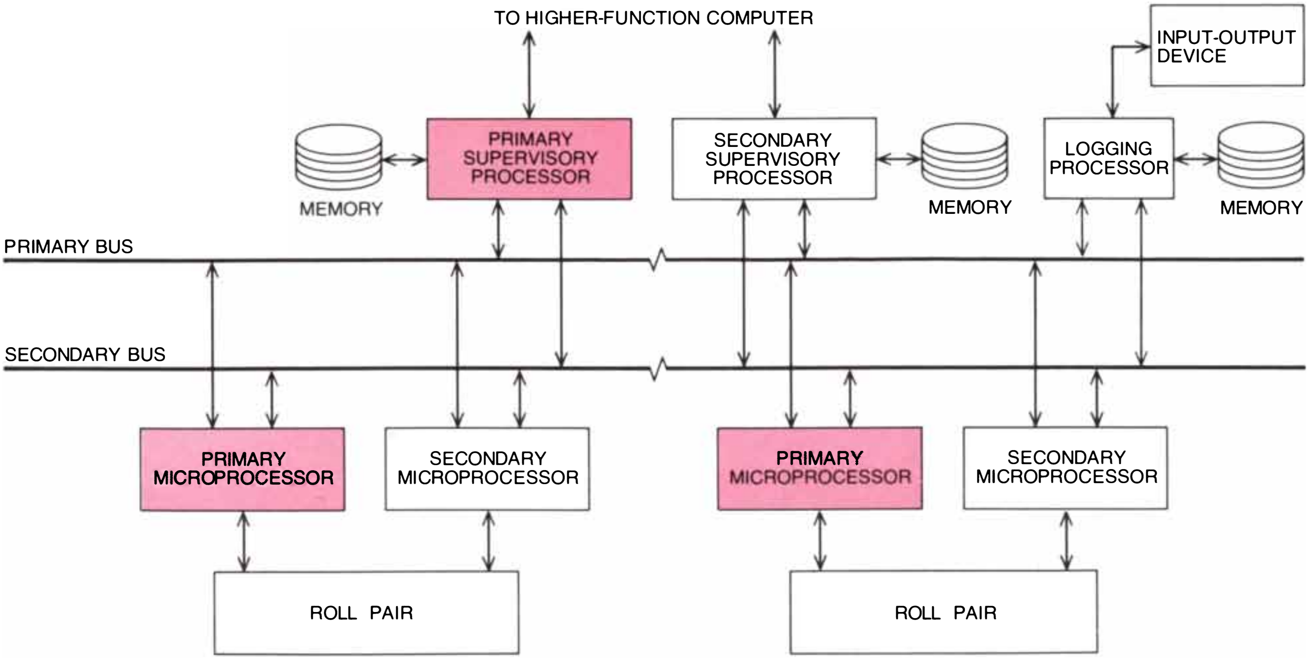
The information contained in a microprocessor at the moment of failure is known as the processor's state. How can the state of the failed processor be recovered so that adequate functioning of the system can be maintained? No general solution to the problem has been found. Specialized solutions have nonetheless been devised, and in specific instances they work.

The communication paths in a distributed-intelligence control system are often exposed to hostile environments (electric motors, magnetic fields and so on), so that the messages of the system are often garbled because one or more bits have been altered by electrical or other noise. There are various schemes for adding redundant bits to each message so that the receiver can detect an error in the message on its arrival. If an error is detected, the sending device is asked to repeat the message until an uncorrupted version is received. Additional schemes protect against the loss or duplication of messages. They even make it possible not only to detect errors at the receiver but also to correct them.

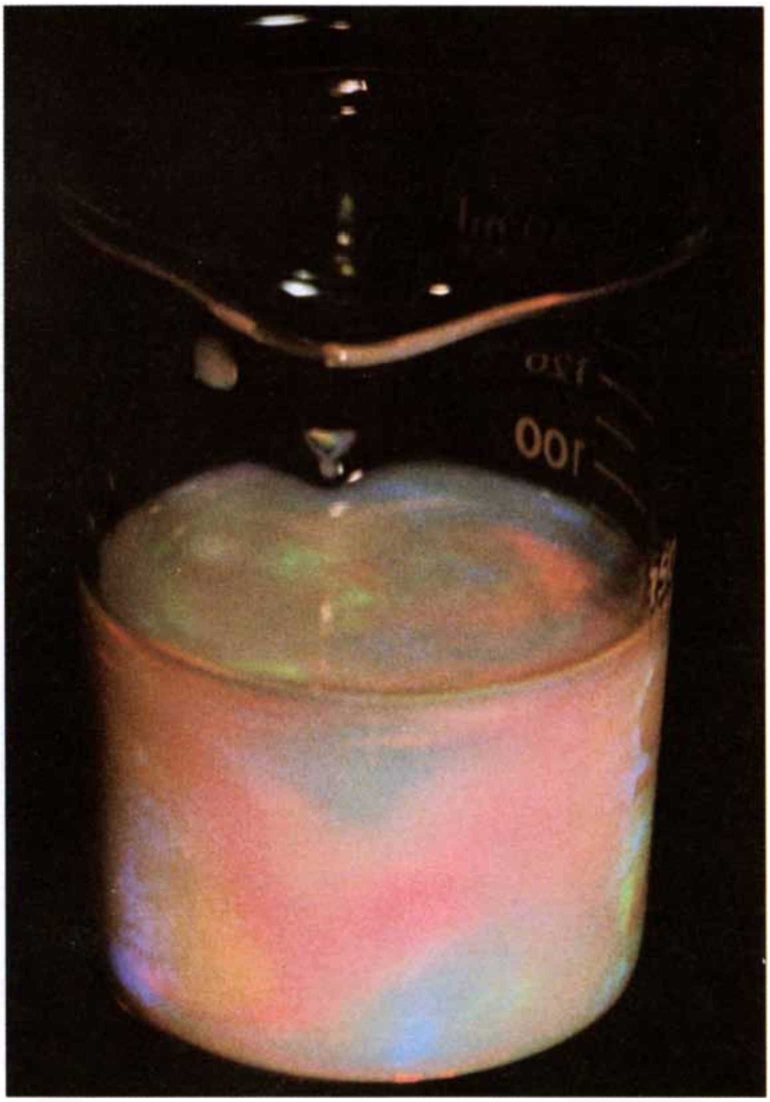
Providing adequate capacity in a distributed-intelligence control system is usually not difficult because the message-traffic rates and the response-time requirements are normally well known to the system designer in advance. The communication facility, the number of microprocessors and the allocation of tasks to the processors can be designed with the amount of excess capacity needed. The overall performance of the system is affected in interesting ways by the distribution of processing. By limiting the functions of the low-level processors to the direct digital control of a small number of control loops, the number of tasks each processor must execute can be reduced to a level that allows each set of tasks to run to completion without being preempted by more important or more urgent tasks. Hence the structure of the operating system can be significantly simplified. The processor capacity needed by the operating system is reduced and performance is enhanced. Performance is also enhanced by the fact that tasks in a distributed system can be done in parallel, or concurrently, rather than sequentially as they would be done on a single large computer.

On the other hand, since the tasks are no longer centralized, messages between tasks incur delays in transmission and processing. Moreover, when a low-level data-acquisition processor detects a "noxious" external event, a substantial fraction of a second may elapse before the alarm can be passed upward to initiate an interrupt program on a higher-level processor, if indeed that is where the response must be generated. Such disadvantages can be minimized by assigning to the same processor tasks that interact frequently. This is made much easier by dedicating processors to low-level data-acquisition and direct-control functions, so that only alarms and summary information need be transmitted to higher-level processors.

The explosive advance of semiconductor technology will continue to drive the evolution of control systems and enable them to take advantage of the benefits to be had from distributed intelligence. With this technology one can begin to exploit the existing theory of hierarchical control, but the fact remains that the theory available for the design of such control systems has not progressed to the point where it can be applied to achieving optimum control of complex industrial facilities. To help remedy this deficiency we, with the aid of our colleagues and students at the Case Institute, are developing methodologies for modeling and designing distributed-intelligence systems applicable to various industrial processes. A new laboratory for studying distributed computer control established with the support of the National Science Foundation and the Foxboro Company is playing a central role in this challenging effort.



REDUNDANT MICROPROCESSORS AND BUSES are customary in the digital control of industrial processes. This illustration shows schematically a possible arrangement of duplicate facilities for controlling two stands of a hot-strip steel mill. Redundant microprocessors are assigned to each pair of rolls in a stand. The secondary processor periodically asks the primary unit in effect "Are you well?" It also receives duplicate set-point messages from the supervisory processing element and periodically assures the supervisor that it is indeed well. If a primary microprocessor should fail, the secondary unit notifies the supervisor and assumes control of the rolls. Similarly, the secondary supervisor takes over if the primary one fails. If the primary bus should fail, the secondary bus is then switched in immediately.



Bragg diffraction in the visible

After all the millenia of inquiry and meditation on the nature of life, Bragg diffraction of x-rays spotted and streaked some photographic film and soon we hardly needed to be told what the initials DNA stand for.

The Bragg diffraction principle is not confined to x-rays. The phenomenon of critical opalescence occurs with visible light when particles of a very uniform size pass from a gas-like state of Brownian dance to a state where, even at very moderate concentration, isotropic repulsive forces impose three-dimensional order.

A company that lives by silver halide emulsions (actually dispersions), dye coupler dispersions in color film and paper (actually emulsions), and dye and pigment dispersions in man-made polymers for fibers and plastics needs people like Pranab Bagchi.

But it also ought to share with other scientists as much of Dr. Bagchi's thought (on matters short of the nature of life) and work as interests them, yet without dissipating the stockholders' assets to the four winds.

For a reprint of any of the papers at right on the characteristics of colloidal systems, write to Pranab Bagchi, Research Laboratories, Eastman Kodak Company, Rochester, N.Y. 14650.

"Differences Between Fact and Theory in the Stability of Carbon Suspensions in Heptane," P. Bagchi and R. D. Vold, *Journal of Colloid and Interface Science* **33**, 405 (1970).

"Entropic Repulsion Between Two Identical Spherical Particles Coated with a Polymeric Adsorption Layer in an Athermal Solvent," P. Bagchi, and R. D. Vold, *ibid* **38**, 652 (1972).

"Enthalpic Repulsion Between Two Identical Spherical Particles Coated with a Polymeric Adsorption Layer," P. Bagchi, *ibid* **41**, 380 (1972).

"Theory of Stabilization of Spherical Colloidal Particles by Nonionic Polymers," P. Bagchi, *ibid* **47**, 86 (1974).

"Nonionic Denting and Mixing Potentials Between Two Flat Plates," P. Bagchi, *ibid* **47**, 100 (1974).

"On the Assumption of a Step Function for the Adsorption-Layer Thickness in the Nonionic Stabilization of Dispersions," P. Bagchi, *ibid* **50**, 115 (1975).

"Kinetic Theory for the Slow Flocculation at the Secondary Minimum," P. Bagchi, *ACS Symposium Series "Colloidal Dispersions and Micellar Behavior"* **9**, 145 (1975).

"Test for Particle Size Dependence in the Theory of Slow Flocculation at a Secondary Minimum," P. Bagchi, *Colloid and Polymer Science* **254**, 890 (1976).

"A Simple Method for Determination of the Average Particle Size of Coarse Suspensions from Measurements of Apparent Specific Turbidity," P. Bagchi and R. D. Vold, *Journal of Colloid and Interface Science* **53**, 194 (1975).

"Preparation of Model Poly(vinyl toluene) Latices and Characterization of Their Surface Charge by Titration and Electrophoresis," P. Bagchi, B. V. Gray, and S. M. Birnbaum, *ibid* in press.

"Let's give Dad ~~Scotch~~." Jameson Irish

If he likes fine Scotch, he'll love light, imported Jameson Irish.

Let Dad try a glass of Jameson Irish Whiskey the way he would his favorite Scotch.

He'll notice how much it tastes like fine Scotch—only lighter and more delicate.

Chances are, as a dedicated Scotch drinker, he will instantly appreciate this flavor difference.

Though it may take a little time for him to get used to saying, "Jameson Irish on the rocks, please."

Jameson. World's largest-selling Irish Whiskey.

E Pluribus Unum

An enduring hope (and prejudice) of physicists is that the world is so simple that all the elementary particles of matter and all the forces acting between them can be explained by a single, comprehensive theory. For many years such a "grand unification" seemed a distant prospect, but new tools for the construction of a unified theory have recently come to hand, and there have been a few attempts to build one. Although the proposed theories are still tentative, they are taken seriously enough for experimental tests to be planned.

There is already a measure of simplicity in the physicist's view of nature. Matter is considered to be made up of the particles called fermions, which interact with one another by exchanging particles of another kind called bosons. The fermions are the bricks of which the world is built; the bosons are the mortar that binds them together. The description grows more complicated, however, when the details are filled in. Two kinds of fermions must be recognized as being elementary, or in other words as having no internal structure: leptons, such as the electron, and quarks, which are the supposed constituents of the proton and the neutron and of a great many other particles related to them. The bosons are still more varied; they are associated with the fundamental forces of nature, of which there are four: the strong, the weak, the electromagnetic and the gravitational.

A simpler universe could be imagined. There is no obvious need for two kinds of elementary fermion, nor is it apparent why there should be four kinds of force. One kind of each would do. The forces in particular are baffling in their diversity: each force has its own characteristic strength, and the strengths differ by almost 40 orders of magnitude.

If a single theory is to describe all the particles and forces, these complications must somehow be shown to be only superficial. A first step toward that end, which has guided all the subsequent efforts, was taken in the 1960's with the theoretical unification of the weak and the electromagnetic interactions. These two forces are quite different in their outward manifestations. For example, electromagnetism has an infinite range, whereas the range of the weak force is for all practical purposes only 10^{-15} centimeter. What does it mean, then, to say that the forces have been unified? They are said to be unified because all the properties of both forces can be derived from a single underlying theory. By way of analogy, a man is different from a crocodile, but the two species are unified in the sense that the provenance of both can be explained by a single theory, namely evolution by natural selection.

The unification of the weak and electromagnetic interactions can be understood by imagining that the laws of nature have evolved from a simpler state to the present complicated one. In the original state the weak force and the electromagnetic force were indistinguishable: they had the same strength and the same infinite range, and they were transmitted by four bosons, all of which were massless. The forces exhibited a symmetry in that they could be freely interchanged; no matter which force was applied the effect was the same. The differences that are observed today arose from the process called spontaneous symmetry breaking, which destroyed the equivalence in range and strength. Moreover, when the symmetry was broken, only one of the bosons remained massless: the photon, the quantum of electromagnetic interactions. The other three bosons, which are collectively called intermediate vector bosons, acquired large masses. When the masses are expressed in energy units, they approach 100 billion electron volts, or 100 GeV. (For comparison, the mass of a proton is a little less than 1 GeV.)

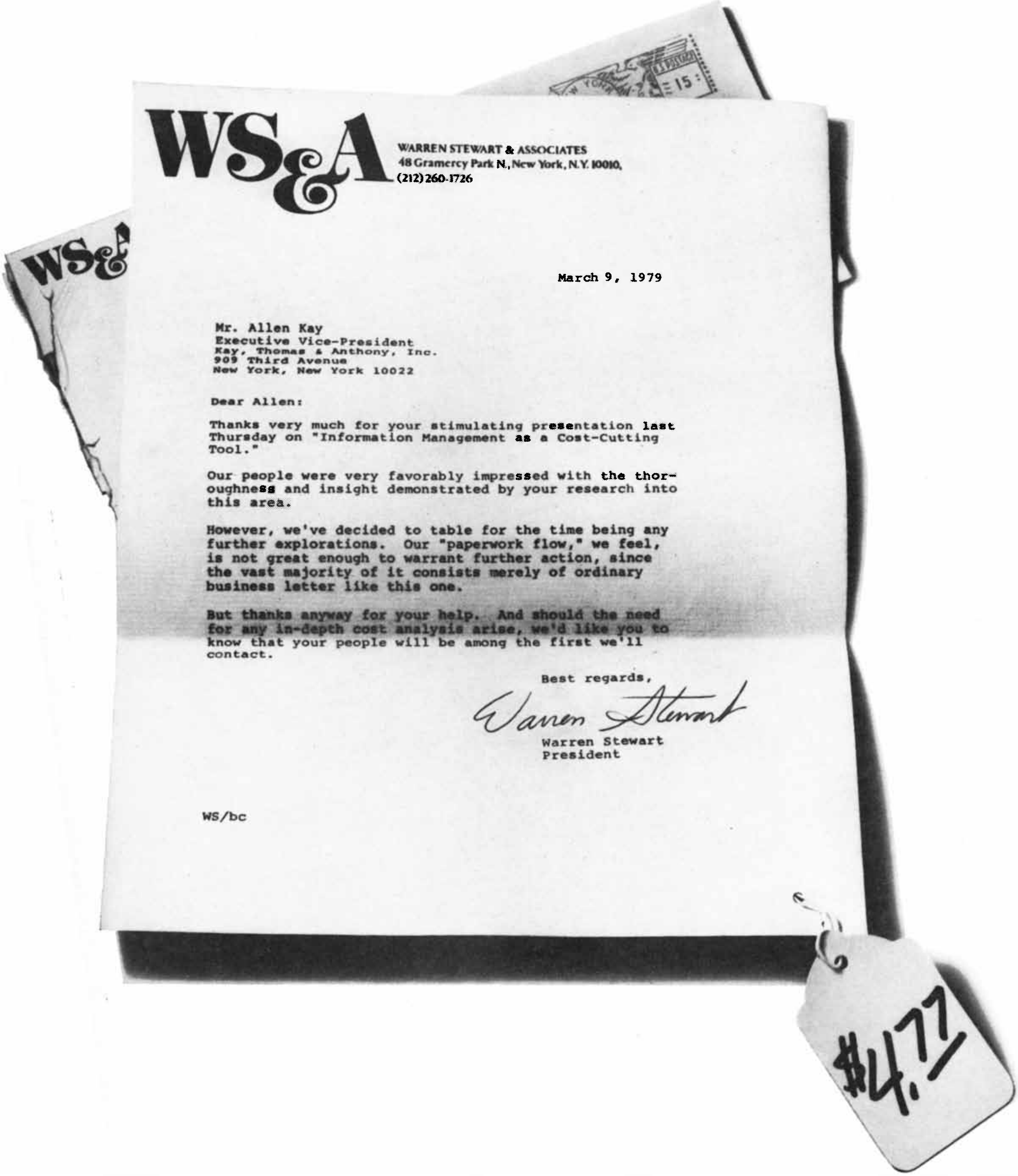
A theory has also been worked out for the strong force, which acts only between quarks (and in a different form between the particles made of quarks). It now seems there are at least six "flavors" of quarks, which have been given the labels up, down, strangeness, charm, bottom and top. Each of the six quark flavors comes in three varieties, which are given the whimsical name colors; it is the colors that govern the strong interactions. The massless bosons that transmit the strong force between colored quarks are called gluons, in direct recognition of their role as binding material. Eight gluons are needed, each one a different combination of the three primary quark colors. The theory that describes the interactions of colored quarks and gluons has been given the name quantum chromodynamics.

The grand unifications now under consideration combine the unified theory of weak and electromagnetic interactions with quantum chromodynamics and subsume both in a single, larger structure. In these grand unified theories three of the basic forces of nature (the strong, the weak and the electromagnetic) are treated as different manifestations of a single underlying phenomenon. Moreover, quarks and leptons are incorporated into the same families.

Actually three "generations" of fermions are needed, each generation including two quark flavors and two leptons. The first generation consists of the up and down quarks (each flavor in three colors) and the electron and the electron-type neutrino. These four particles (or eight particles if each color is counted separately) are all that is needed to explain the composition of ordinary matter; the remaining particles appear only in high-energy experiments. In the second generation the strange and charmed quarks are combined with two more leptons: the muon and the muon-type neutrino. The third generation includes the bottom and top quarks and the lepton called tau and its associated neutrino.

The structure of a grand unified theory is much like that of the weak-and-electromagnetic unification, but it has a larger scope. It can be imagined that in some primitive state all three forces were of identical strength and that all the leptons and quarks in each generation were massless. The distinctive properties of the three forces were introduced by the spontaneous breaking of a symmetry, which also imparted masses to most of the fermions while leaving one fermion in each generation (the neutrino) with zero mass. In this scheme there is a hierarchy of broken symmetries. The first symmetry breaking was responsible for the separation of the strong force on the one hand from the still unified weak and electromagnetic forces on the other. The breaking of the second symmetry distinguished between the weak and the electromagnetic interactions. Of course, the temporal sequence of events imagined here is entirely artificial; both symmetries were broken from the beginning.

If quarks and leptons are included in the same families, there must be some mechanism for converting one kind of particle into the other. In the absence of a grand unification the transformation of a quark into a lepton (or vice versa) is absolutely forbidden. A gluon can change the color of a quark, and an intermediate vector boson can change the flavor of a quark or a lepton, but the boundary between quarks and leptons is inviolable. Any grand unified theory introduces a fifth basic force of nature, called the hyperweak, that mediates the quark-lepton transition. Associated with the new force is a new set of bosons, which as a result of the symmetry


WS&A

WARREN STEWART & ASSOCIATES
48 Gramercy Park N., New York, N.Y. 10010,
(212) 260-1726

March 9, 1979

Mr. Allen Kay
Executive Vice-President
Kay, Thomas & Anthony, Inc.
909 Third Avenue
New York, New York 10022

Dear Allen:

Thanks very much for your stimulating presentation last Thursday on "Information Management as a Cost-Cutting Tool."

Our people were very favorably impressed with the thoroughness and insight demonstrated by your research into this area.

However, we've decided to table for the time being any further explorations. Our "paperwork flow," we feel, is not great enough to warrant further action, since the vast majority of it consists merely of ordinary business letter like this one.

But thanks anyway for your help. And should the need for any in-depth cost analysis arise, we'd like you to know that your people will be among the first we'll contact.

Best regards,

Warren Stewart

Warren Stewart
President

WS/bc

breaking acquire an enormous mass, perhaps as much as 10^{15} GeV. Given this almost palpable mass, the bosons of the hyperweak force have been nicknamed intermediate vector baseballs.

Another way of expressing the substance of unification is to note that the strength of each of the forces is not constant but depends on the energy with which particles interact. As the energy of an interaction increases, the weak force grows stronger and so does the electromagnetic force but at a lower rate. The strong force, on the other hand, declines in strength. At sufficiently high energy all three forces must have the same strength.

The energy at which the strengths of the three forces are expected to converge is about 10^{15} GeV, or the same as the mass of the intermediate vector baseballs. For now there is no hope of reaching that energy in the laboratory (the largest accelerator has yet to achieve 10^3 GeV). There was an era in the history of the universe, however, when particles bearing such energies must have been common: the first millisecond after the "big bang." In that brief interval the symmetries of the basic forces and particles must have been fully apparent and quarks and leptons must have been in thermal equilibrium, so that one kind of fermion could be freely converted into the other. The symmetries are obscured now only because the universe is quite cold; it has undergone a kind of freezing.

Although the energy regime where unification might be observed directly is out of reach, the unified theories explain a number of phenomena that are observable. For example, it is an enigma of long standing why the electron and the proton have electric charges of exactly the same magnitude but opposite sign. The world would be a much different place if they did not, but that is not a compelling argument. When the electron and the quarks that compose the proton are placed in the same family, however, the quantization of electric charge emerges naturally from the theory. The masses of the quarks, which have seemed rather arbitrary, can also be calculated in a grand unified theory; for the heavier quarks these calculations are in good agreement with the values deduced from experiments, but discrepancies that have yet to be explained arise with the lighter quarks. A grand unification also offers the possibility of explaining how the universe evolved with an excess of matter over antimatter.

The most remarkable prediction of the unified theories is that matter itself is unstable. If quarks can be converted into leptons, then so can the particles that are made up of quarks. Most notably, the proton, which in less comprehensive theories is absolutely stable, can decay through pathways that yield as final products only photons, electrons, muons and neutrinos. It follows that atoms also disintegrate, and the entire universe is evanescent. The time required for the world to evaporate, however, is reassuringly long: the mean lifetime of the proton is estimated to be some 10^{30} years, which is more than 10^{20} times the present age of the universe.

Detecting the decay of the proton would be the most dramatic and conclusive way of confirming a grand unified theory. A lower limit on the lifetime has already been set, and it is near the predicted lifetime. More sensitive searches are now being planned by at least three groups of investigators. If the typical lifetime is 10^{30} years, observing 10^{30} protons should reveal one decay per year. The planned experiments call for monitoring some 10,000 tons of water with instruments capable of registering proton decays and distinguishing them from the numerous extraneous events.

If one of the present schemes for unifying strong, weak and electromagnetic forces should turn out to be correct, it would certainly be a grand theory, but it would not be the ultimate theory of nature. One source of dissatisfaction is that three separate generations of fermions are required, two of which appear to be superfluous. Another problem might prove to be more disturbing: one of the basic forces has been left out of the account entirely, namely gravitation. The characteristic energy of the grand unified theories is not far short of a quantity called the Planck mass, which is the mass at which the gravitational interactions of particles begin to have an important influence on their behavior. The Planck mass has a value in energy units of some 10^{19} GeV, compared with the mass of 10^{15} GeV at which the other forces are unified. The near correspondence of these large numbers may not be a coincidence; indeed, a theory of elementary-particle interactions that treats masses this large may have to include gravitational effects if it is to give accurate results. Certain preliminary steps toward this end have been discussed; they involve the concepts of supergravity and supersymmetry, where fermions and bosons are included in the same families. Constructing a quantum-mechanical theory of gravitation, however, is itself a problem of notorious difficulty, and for now it may be necessary to undertake one task at a time.

Portrait of a People

"Facts of Life and Death," a compendium of vital and health statistics published by the National Center for Health Statistics of the U.S. Pub-

Bank behind this symbol.

We've got the answers.

You work hard for your money. And you want to do the right things with it. But when you consider all of the options that exist in today's complex money world, you're bound to have some questions.

That's why you need A FULL SERVICE BANK. It's a lot different than all the other institutions that advertise so-called "bank-like" services. Because we're ready with the answers.

Over the years, FULL SERVICE BANKS have developed over 100 different financial services for individuals, businesses, and municipalities. And only at A FULL SERVICE BANK can you find *America's Bankers*, highly trained professionals who can answer those confusing money questions. Even if your question is better handled by some other financial advisor, your banker can suggest where you can be helped.

So when you keep your savings and checking accounts at A FULL SERVICE BANK, you've established a relationship with all the bank you'll ever need. America's FULL SERVICE BANKS. We've got the answers.

AMERICA'S FULL SERVICE BANKERS

lic Health Service, provides a statistical portrait of the American people as of 1976, the bicentennial of the country's independence and the last year for which complete statistics are available. The picture it paints is of a society with greatly improved living standards but lingering inequities in the distribution of health resources.

The estimated U.S. population as of July, 1976, was 214.7 million, with a growth rate of .76 percent per year. This growth rate, which has stayed approximately constant since 1973, declined from 1.21 percent in 1970 and 1.75 percent in 1960. The decline is the result of several factors. The birth rate dropped from its modern high of 24.6 births per 1,000 population in 1955 to 15.6 per 1,000 in 1972 and then leveled off at about 14.8 per 1,000 through 1976. Fertility rates of women between the ages of 15 and 44 declined from a peak of 118 live births per 1,000 in 1960 to 65.8 per 1,000 in 1976.

At the same time the infant-mortality rate and the overall death rate continued to decline. In 1976 the death rate was 8.9 per 1,000, down from 9.5 per 1,000 in 1960 and 10.9 per 1,000 in 1935. A child born in 1976 could expect to live for 72.8 years, approximately 24 years longer than a child born in 1900 and five years longer than one born in 1950.

Now that the swollen cohort of the postwar "baby boom" are entering their late 20's and early 30's the U.S. population is slowly getting older. In 1970, 45.9 percent of the U.S. population were under 25, but by 1976 the fraction had shrunk to 43.4 percent. The declining birth and death rates are also expanding the ranks of the older people. Those 65 or older, who were only 6.9 percent of the U.S. population in 1940, constituted 10.8 percent in 1976.

Chronic and degenerative illnesses are the major causes of death now that fatal infectious diseases are largely under control. The only infectious disease still present at epidemic levels is gonorrhea, which is rarely fatal if treated. Still, between 1969 and 1976 the annual number of cases of gonorrhea nearly doubled, with a total of 1,001,994 reported cases in 1976. The four leading causes of death in 1976 were, in descending order: heart disease, cancer, cerebrovascular disease and accidents. Motor-vehicle accidents alone accounted for 47,038 deaths in the bicentennial year. One surprising figure is that between 1970 and 1976 there was a significant decline in the number of deaths due to major cardiovascular disease, from 340.1 per 100,000 to 284.4 per 100,000. Much of this decline has been attributed to the improved detection and management of hypertension. At the same time, however, the death rate due to cancer rose from 129.9 per 100,000 in 1970 to 132.3 per 100,000 in 1976.

The sexual revolution has had a significant impact on American life in more ways than just the current epidemic of venereal disease. Births to unmarried women between the ages of 15 and 44 have increased dramatically over the past 40 years. In 1940 only 7.1 of every 1,000 live births were illegitimate; the number rose to 14.1 in 1950, 21.6 in 1960 and 26.4 in 1970. The trend seems to have peaked, however: in 1976 there were 24.7 illegitimate births per 1,000, a decline from the 1970 figure. Marriage rates have remained fairly steady over the past decade at about 10 per 1,000 population per year, although they were somewhat higher in the 1950's and somewhat lower in the 1960's. There are nonetheless more single women today. Marriages by previously unmarried women over the age of 15 declined from a high of 90.2 per 1,000 in 1950 to 65.2 per 1,000 in 1976. At the same time the divorce rate almost doubled, increasing from a rate of 2.6 per 1,000 in 1950 to 5.0 per 1,000 in 1976. One in every two marriages begun in 1976 was likely to end in divorce.

The fact that American society is still divided along racial lines is evident in many statistics. A male infant born in 1976 had a life expectancy of 69.7 years if it was white but only 64.1 years if it was nonwhite. The mortality rate was 13.3 per 1,000 births for white infants but 23.5 per 1,000 for nonwhite ones. Maternal mortality was also much higher for minority women: 26.5 per 100,000 for nonwhites compared with 9.0 per 100,000 for whites. Although the overall rates of infant and maternal mortality for both whites and nonwhites have declined dramatically over the past 30 years, the historical discrepancies still exist.

Death rates due to violence in the U.S. also differ strikingly between whites and members of racial minorities. Nonwhites are less likely to commit suicide than whites but are much more likely to be murdered. The suicide rate in 1976 was 13.3 per 100,000 for whites and 7.0 per 100,000 for nonwhites. Yet the homicide rate was 5.4 per 100,000 for whites and 33.2 per 100,000 for nonwhites. For nonwhite males the incidence of murder was extremely high: 55.8 per 100,000.

Cloud Nine

The chemistry of interstellar clouds has turned out to be curiouser than anyone had imagined. The first molecules to be discovered in interstellar clouds of gas and dust were, with few exceptions, reasonably mundane: water, ammonia, formaldehyde, carbon monoxide, methyl alcohol, acetaldehyde, dimethyl ether, silicon sulfide and ethyl alcohol, among others. All told some 50 molecular species have now been identified in space, virtually all by radio astronomers. The latest discoveries involve carbon atoms linked into exotic chains so unstable that it is extremely hazardous even to attempt their synthesis in the laboratory. In at least three recent instances there are no laboratory examples at all; the interstellar species have been identified from theoretical calculations of their spectra. "We are seeing a quite marvelous marriage of astronomy and chemistry," says Patrick Thaddeus of the Goddard Institute for Space Studies, who has participated in some of the new discoveries.

Thaddeus and his colleagues have searched particularly for a family of radicals in which two or more carbon atoms form a chain that is linked either to a single hydrogen atom or to a single nitrogen. The chains rotate end over end as rigid sticks. In the envelope of a cool but luminous red-giant star that is shedding mass, IRC + 10216, Thaddeus and his co-workers have now identified the cyanoethynyl radical C_3N and the butadiynyl radical C_4H. The structural formulas of the two radicals are $C\equiv C-C\equiv N$ and $C\equiv C-C\equiv C-H$. The identification of both substances must be considered tentative since their spectra have not been observed in the laboratory. Limited information, however, has been available on C_4H, which has been studied frozen in a matrix of inert gas.

Theory predicted that C_3N and C_4H would emit microwave radiation at several frequencies corresponding to successive transitions between rotational levels of a rigid molecule. In the case of C_3N Thaddeus and Michel Guélin observed three of the predicted rotational transitions in the radiation from IRC + 10216. Subsequently Thaddeus, Guélin and Sheldon Green identified four of the predicted transitions of C_4H in the envelope of the same star. The temperature of the star appears to be less than 600 degrees Kelvin (or less than 330 degrees Celsius).

The heaviest molecule yet detected in space is cyano-octatetra-yne, a molecule consisting of a chain of nine carbon atoms with a hydrogen atom at one end and a nitrogen at the other: HC_9N. Not yet synthesized on the earth, its chemical structure is $H-C\equiv C-C\equiv C-C\equiv C-C\equiv C-C\equiv N$. The lowest members of the series, hydrogen cyanide (HCN) and cyanoethyne (HC_3N), proved to be quite abundant in interstellar clouds. The next member of the series, cyanobutadiyne (HC_5N), could not be searched for successfully until it had first been synthesized in the laboratory. The delicate synthesis was achieved in

Free. 11 ounces of potent medicine for companies sick of outrageous business travel problems.

It's a simple fact: the cost of business travel today is outrageous.

But there's something simple you can do about it.

Write or call for the free Beechcraft Business Flying Kit. It's the kit that's already provided first aid for hundreds of companies, large and small, that were sick and tired of seeing travel costs cutting deeper into their profits.

And this year the kit has been updated and expanded with pertinent new information to make it even more helpful. Information like:

Up-front answers to your most important questions about owning an airplane: Do I need one? What size should it be? Who will fly it?

An inside look at dozens of creative ways companies use business airplanes to solve problems and create new opportunities.

An independent study of the relationship between business flying and corporate profits.

And you can even figure the net capital cost of a company airplane, like the new T-tailed Beechcraft Duchess, shown here.

Truth is, a company airplane isn't always the answer. But with this kit you'll know the right questions. And that's the first step to making a rational decision on a company airplane.

Find out for yourself why the company airplane is the fastest-growing mode of business travel today.

Send for the Beechcraft Business Flying Kit. Ounce for ounce, it's the most potent information on business flying today.

Write on your company letterhead to: Beech Aircraft Corporation, Dept. Y, Wichita, Kansas 67201. Ask for our free Business Flying Kit, and please mention if you're a pilot. If you'd rather call, make it collect and ask for Dick Schowalter, Jr. 316-681-7072.

Member of General Aviation Manufacturers Association

The Beechcraft Duchess. The practical 4-seat twin that can get you to where you're needed at speeds over 190 mph. In quiet, solid comfort.

Know any friends who want to learn to fly? Tell them about the General Aviation Manufacturers Association **Take Off** Sweepstakes. They could win a $50,000 airplane just for earning their private pilot license. Have them call TOLL FREE, 24 hours a day, any day and ask for the BEECH "TAKEOFF" operator: USA 800-447-4700 (in Illinois, 800-322-4400); Canada 800-261-6362 (Toronto, 445-2231).

Now Polaroid giv
instant advantage.

es professionals an

The new Polaroid Model 600 SE is the professional camera you've been waiting for. Because it doesn't make you wait for your pictures.

Instead, you see them instantly. Full frame. Without adding any special backs. Or losing any flexibility.

You see, even though it's an instant camera, the Model 600 SE still gives you complete control: You set the aperture. You select the shutter speed. You pick the lens you want for the job—standard, wide angle or telephoto.

Plus, you get the added advantage of seeing your pictures on the spot. Instead of in the darkroom.

No other camera offers you this kind of versatility.

The Model 600 SE lenses are interchangeable, bayonet-mounted, precision glass. Each one is fully coated and color corrected to produce sharp images over the full picture area. And each one is equipped with a Seiko between-the-lens shutter that's X-synchronized for electronic flash and has 10 speeds, ranging from 1 second to 1/500 second.

What's more, with the Model 600 SE you get a bright-image, coincidence-type rangefinder designed for accurate viewing and framing. Plus an interchangeable film back. And you can choose from six different Polaroid 3¼ x 4¼ inch self-developing films designed to handle virtually any photographic assignment.

The new Polaroid Model 600 SE. It's the professional camera with the instant advantage. And in this profession, you need every advantage you can get.

For more details, see your Polaroid photographic dealer. Or from the continental U.S., call toll-free: 800-225-1384 (from Massachusetts, call collect 617-864-4568).

Polaroid's Model 600 SE

Professional Pack Film Camera

1976 by A. J. Alexander, H. W. Kroto and D. R. M. Walton of the University of Sussex, who measured the rotational spectrum of the molecule. Interstellar HC_5N was discovered the same year by L. W. Avery, N. W. Broten, J. M. MacLeod and T. Oka of the Herzberg Institute of Astrophysics of the National Research Council of Canada. The rotational emission lines of HC_5N were found in Sagittarius *B*, a dark interstellar cloud believed to lie near the center of the galaxy.

Last year cyanohexatriyne (HC_7N) was synthesized by Kroto and his coworkers. Once its microwave spectrum had been measured in the laboratory the molecule was detected by Avery's group in a dark cloud known as Heiles' Cloud 2, which is only a few hundred light-years from the solar system. By now enough was known of the cyanopolyyne series to predict the rotational spectrum of the next member, HC_9N. A search by the Herzberg group showed that HC_9N is also present in Heiles' Cloud 2.

Roughly speaking, every time two more carbon atoms are added to a cyanopolyyne the abundance of the resulting molecule is less by a factor of four. Since the emission of HC_9N is already very weak, Avery is doubtful that the next members of the series, $HC_{11}N$ and $HC_{13}N$, can be detected with existing instruments. There is still no theory to explain how such long carbon chains are synthesized in interstellar clouds. The absence of carbon ring compounds, which are plentiful on the earth, is also a puzzle. In spite of intensive searches no carbon rings have yet been discovered in space.

Male Pill

A chemical antifertility agent that could turn out to be the first safe and effective oral contraceptive for men is currently undergoing clinical trial in China. The chemical is gossypol, a yellowish phenolic pigment isolated from cottonseed (and named for *Gossypium*, the genus of the cotton plant). Recent visitors to China have been hearing about gossypol from Chinese physicians, and now a preliminary report on studies of the chemical's antifertility effects, side effects and potential as a contraceptive has been published in the *Chinese Medical Journal* by a national coordinating group on male antifertility agents.

As long ago as the 1950's, according to the article, investigation of instances of widespread infertility in localized populations had suggested that ingestion of food cooked in crude cottonseed oil (rather than in oil purified by boiling) could lead to male infertility. In 1971 the active agent was shown to be gossypol, a compound that had been isolated in Russia at the end of the 19th century and was known to be toxic in animal feeds. Large amounts of purified gossypol were prepared, and experiments were undertaken at a number of research centers to determine the site and mechanism of gossypol's antifertility effect, the movement of the substance through the body and its toxicity. Rats given oral doses of from 15 to 40 milligrams of gossypol per kilogram of body weight per day for from two to four weeks became infertile; the infertility continued for from three to five weeks after the treatment ended, following which fertility was gradually reattained.

The chemical appeared to have its effect in the seminiferous tubules of the testis, where the spermatozoa are produced. Various precursors of the sperm were damaged. The sperm themselves were consequently malformed; with increasing duration of treatment the sperm count decreased until there were no live sperm at all. The secretory cells of the testis, which manufacture the male sex hormone testosterone, seemed not to be damaged, and the rats' libido seemed not to be affected. Radiographic studies showed an initial concentration of the chemical in the liver, but by the ninth day after administration the gossypol was concentrated in the testis. Although the peak concentration was higher in several organs than it was in the testis, only the testis was damaged, implying that gossypol's selective action is due to a special vulnerability of the testis. There were said to be no toxic effects at antifertility dosages.

The results in rats were confirmed in other laboratory animals, and in 1972 clinical trials were begun. More than 4,000 men have been treated for longer than six months, of whom more than half have been followed up for two years. The initial dose was 20 milligrams per day of gossypol, or of gossypol linked to acetic acid or formic acid. The drug's efficacy is said to have been 99.89 percent, with infertility defined as a semen analysis showing no live sperm or fewer than four million sperm per milliliter, compared with a normal 50 to 100 million. That level of infertility was usually achieved in about two months, after which subjects were put on a maintenance dose of from 150 to 220 milligrams per month in divided doses.

The major side effect was a "transient weakness," when the drug was first administered, that subsided without treatment. About 6 percent of the subjects complained of some decrease in libido, but their potency was not impaired; blood levels of testosterone remained normal. A few subjects were found to have low blood-potassium levels, but the connection with gossypol is not well established. When treatment was discontinued, the sperm count returned to a normal level, usually in about three months. The report does not cite any follow-up studies of fertility in men after discontinuance of gossypol treatment, nor of the condition of children they subsequently begot.

The report concludes: "Gossypol, used as a male antifertility agent, is reliable and relatively safe" at proper dosages. "The supply is relatively abundant and it is relatively economical. It therefore seems to be a male antifertility agent of promise." Investigations of the effect and safety of long-term administration, ways to reduce side effects, the precise mechanism of the drug's action and its effect on potassium metabolism are continuing.

Superconducting Supercomputer

The fastest computer now contemplated would be no bigger than a grapefruit and would operate only when it was cooled to the temperature of liquid helium. It would have no transistors or other semiconductor components; instead it would be built of Josephson-junction devices, which are superconducting switches. Building such a computer would entail introducing an entirely new electronic technology, but the machine would operate several times faster than the much larger mainframe computers of today.

The first requirement of a fast computer is that the active circuit elements, which are essentially switches, change their state very quickly. The Josephson junction is the fastest switch known, and so it is a likely candidate for a high-speed electronic technology on this basis alone.

A second reason for adopting the Josephson junction is less obvious but equally compelling. Any computer that would run appreciably faster than the present standard must be quite compact. If it were not, then even with circuit elements that switched instantaneously there would be unacceptable delays for signals to pass over the long wires that would be needed to connect one part of the machine with another. In a hypothetical high-speed computer defined as a design goal by the International Business Machines Corporation the longest path within the central processing unit, where all arithmetical and logical operations take place, is 7.5 centimeters. The longest path between the central processing unit and the main memory is 30 centimeters. If these requirements are to be met, the entire computer, including all its memory, must fit in a box with dimensions of eight centimeters by eight centimeters by 10 centimeters.

One can imagine shrinking a main-

DP Dialogue

Notes and observations from the IBM Data Processing Division that may prove of interest to the engineering community

This automatic stacker crane travels to the selected bin and retrieves the required items. Computer simulation enables Harnischfeger to design efficient industrial material handling systems.

Harnischfeger Simulates Material Handling Systems in a Computer

"For analysis and design of material handling systems," says Dale A. Baker of Harnischfeger Corporation, "we have found computer simulation with GPSS V to be invaluable."

Baker is manager of computer systems for the firm's Material Handling Systems Division, where General Purpose Simulation System V (GPSS V), an IBM computer program, is used to model industrial lifting, moving and storage systems.

The stacker crane is the heart of a modern high-rise storage system. Mounted on a track, it runs horizontally in an aisle while rising vertically to reach bins from floor level to as high as 100 feet. Often, there is also a conveyor carrying material to and from the storage area.

Range of Choices

Computer simulation is helpful because of the broad range of choices to be investigated in the design of an overall system. The givens are size and geometry of the storage space, speed of crane movement in each axis, conveyor speed, the total number of stockkeeping units, and the distribution of activity over the stockkeeping units. The question is how many cranes, in aisles of what length, will do the job? Overdesign, and the price is too high. But performance must be guaranteed, so the system must not be underdesigned.

GPSS V can be used to test a proposed design under all sorts of conditions specified by the customer, Baker points out. "For example," he says, "suppose a storage system receives material from manufacturing processes, truck docks and rail docks. When will the system be operating at capacity? Where are the bottlenecks? GPSS V simulation develops this kind of performance data quite reliably."

Demonstrates the Design

When a customer submits a request for a proposal that specifies system performance, Harnischfeger uses simulation on its IBM System/370 Model 148 to arrive at a general configuration. Later, it may be used to demonstrate to the customer the capabilities of the proposed design. Even after a system has been built, Baker notes, it may be difficult to test all the conditions cited in the acceptance procedure. "There are extreme conditions the real world won't present in five years of operation, but which the model can create at any time.

"GPSS V highlights problem areas for the engineer and offers a quick means of testing design changes," Baker adds. "It can give us a 'rough cut' on a system in a hurry, or remarkably accurate performance figures under a wide variety of operating conditions."

Portland GE Tests Planning Decisions with Engineering Models

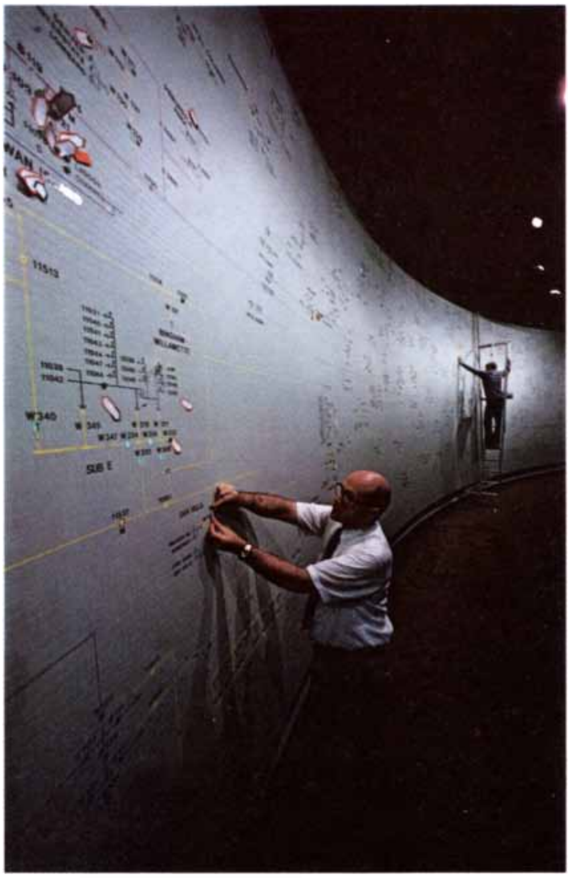
"Today we use engineering models of our system to help us plan projects, set priorities, and forecast the impact of change," says Preston N. Adkins, Chief of System Planning Development for Portland General Electric Company.

"With the cost of power rising rapidly, we vitally need these new and better methods of evaluation."

PGE engineers operate the models through terminals located throughout the electric utility's headquarters in Portland, Oregon. They use Virtual Storage Personal Computing (VSPC), an IBM program product, to activate programs stored in the IBM 3032 Processor or to work interactively at the terminal to solve current problems.

"Simulations usually take a number of trial runs," Adkins notes. "With VSPC, we can see sample output in the display and ask for printout only for the final run. We can try more alternatives, while saving time and expense.

"For example, to improve service or voltage levels anywhere in the system, an engineer uses a mathematical model of the distribution network. He enters topological factors, loads, impedances, and other parameters of the feeder system, then varies the model to find the most economical and effective modification."



Energy Control Center at Portland General Electric. A technician uses a tag to make a notation on the automated "mimic board," which graphically represents the sub-transmission system.

Another program used every day is power flow analysis, Adkins adds. A mathematical model of a very large, complex network, it reveals the weak spots and high-flow segments of the transmission system. For contingency analysis, planners study the effects of running the model with one system element at a time removed.

"An important feature of VSPC has been easy access to the corporate data base," Adkins points out. "No keypunching or other manual handling is required to use such historical data as load profiles and operating costs. This avoids errors and yields great savings in time and cost."

Precision at Simmonds: Computer-Aided Document Typing

"ATMS saves us hours and hours of typing," says Paul Schroeder, proposal supervisor at Simmonds Precision Products, Incorporated. He is describing the use, by the company's Instrument Systems Division, of the Advanced Text Management System (ATMS), which stores, processes and types textual documents through an IBM System/370.

"We produce hundreds of technical documents each month: operating, training and overhaul manuals, proposals, parts bulletins and engineering specifications," Schroeder adds. "They vary widely in format and content, and usually require immediate turnaround. Typically, a 200 page technical manual must be typed and produced in one week. Acquiring ATMS 18 months ago enabled the department to meet a rising volume of work without adding to the three typists in the word processing section.

"We often have to add or delete paragraphs, move material around, or change the format," Schroeder notes. "Once the text is in the memory we need only key in a couple of commands and the system does the rest. If a change affects all the following paragraph or page numbers, ATMS renumbers them automatically."

The Instrument Systems Division of Simmonds Precision, in Vergennes, Vermont, manufactures electronic instruments for commercial and military aircraft, such as fuel gauge systems, engine pressure-ratio computers, asymmetry detectors and tachometers. One recent product is the Performance Advisory System (PAS), an in-flight computer system that allows a pilot to fly more economically and saves airlines thousands of dollars per aircraft in fuel consumption.

"Airlines may each have a slightly different configuration of a complex product like the PAS," Schroeder explains. "ATMS keeps track of which airline requires which version of the instruction manual, helping to insure that all users receive the proper revision pages."

To use ATMS, an operator keys the text into the System/370 Model 138 via an IBM 3277 Display Station. Corrections are made by moving a cursor to desired points in the cathode ray tube display and keying in the change. Non-printing characters imbedded in the text by the operator control such things as column widths, right justification, spaces for illustrations, type fonts, and page numbering. By changing these codes, the operator can reformat a document.

"Our biggest productivity gains – often by a factor of eight – are during revision of existing documents," Schroeder continues. "We no longer need to retype

An engineer at Simmonds Precision tests a prototype of the Performance Advisory System. Used on commercial airliners to recommend climb and descent profiles, engine settings, cruise altitude and airspeed, it can reduce fuel consumption by as much as eight percent.

Helping Teachers Teach at U. of Evansville

There's a teaching aid at the University of Evansville, Indiana, that's available morning, noon and night, seven days a week. It's the university's computer, which is helping students of nursing, medicine and other health-care specialties in three of their courses covering the human body and how it works.

A student taking anatomy and physiology, for example, can work at a computer terminal in the Computer Assisted Instruction Lab – typically in ten review sessions per school term, of about 50 minutes each. Dr. William Snively, who directed the development of the learning system, emphasizes that computer-aided instruction is one of six carefully integrated modes of study at the university. The five others are lectures, classroom work, textbook study, laboratories and other programmed instruction including written and audio/visual programs. Each has its own advantages in helping students keep their interest up, while meeting the demands of a heavy workload.

During a study session, the IBM System/370 poses questions that the student answers by typing at the terminal keyboard. "The questions and answers are designed so the program can distinguish whether a wrong answer, in context, is an educated miss or a wild guess," Professor Snively notes. "And the system responds accordingly: 'You're almost right but here's a point you missed.' Or: 'Where were you last night?'

"Because these are undergraduate students, we've used humor and informal language in the system responses, and the computer addresses each student by name. In addition to making a session more enjoyable, this also makes for more productive involvement by the student.

"We don't require anyone to use the computer," Snively says, "but 50 percent of the students do so voluntarily." One course on acid/base disturbances, which deals with changes in the chemistry of body fluids, is a systematic coverage of the subject. Those for anatomy/physiology and patho-physiology are computer-assisted reviews. All three were adapted from programmed learning texts by Scott Kincaid of the university's data processing department.

The computer keeps track of each student's performance and reports on progress to the student and to the instructor. However, the computer score never goes into the student's grade or record.

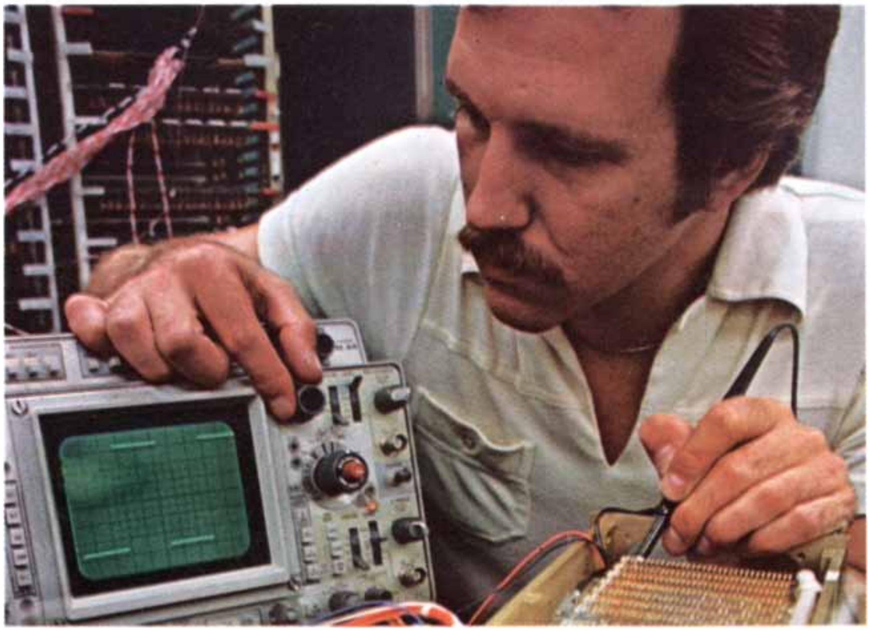
"It seems that everybody benefits from this system," notes Snively. "The constant use of the terminals, 15 in all, attests to its popularity with the students. And it lets us make better use of our time in preparing material and meeting personally with students who have a particular question or problem that they have trouble understanding."

Dr. William Snively was the moving force behind the development of computer-based learning aids for medical and health-care subjects at the University of Evansville.

---

unchanged material or sections in which paragraph numbers, part numbers or a piece of nomenclature are the only changes.

"Manuscript is usually submitted in parts, by several people," Schroeder says. "ATMS accepts these inputs in random order and assembles them in sequence. Once the text is keyed in, the system can produce a first draft for engineering review. With less time required to produce a finished document, there is more time to work on it and check it over.



"We used to have a great deal of overtime when a big piece of work was due," Schroeder reports. "But since installing ATMS we have met our commitments while reducing overtime considerably. The typing staff loves the system, the product looks better, it is more accurate, and we now have time to make sure it contains everything the customer wants."

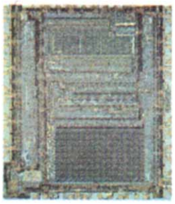
DP Engineering Dialogue is designed to provide you with useful information about data processing applications, concepts and techniques. For more information about IBM products or services, contact your local IBM branch office, or write Editor, DP Dialogue, IBM Data Processing Division, White Plains, N.Y. 10604.

IBM®
Data Processing Division

HOW THE NEW EATON PROFITS FROM CHANGE

Detroit needs a billion dollars worth of miracle chips by 1981. And that's just for starters.

The wafer on the hood holds up to 300 chips. Each chip is a mini-computer (photo at left). A computer that will change the way you drive. The first major application: the ignition system of your car.



We profit from change in semi-conductors…

Computer chips will also direct air and fuel mixtures, govern speed, signal when anti-freeze is low, when to change oil. They'll shift gears and balance suspension when you corner. They'll increase performance and save gas.

The companies making these chips are growing fast. So the companies that make the equipment that makes the chips have to grow just as fast. And we're in the thick of that growth – through our Cutler-Hammer group, a leader in electronics and a recent Eaton acquisition.

automated storage and retrieval systems…

and hydrostatic transmissions for farm equipment.

Our base is broad and balanced. Our sales have grown to an annual rate of $3 billion. Earnings per share have increased 45% during the last five years. And with our growing involvement in advanced technology for industry, we expect to profit from change even more in the years ahead.

For the complete story write: Eaton Corporation, World Headquarters, 100 Erieview Plaza, Cleveland, Ohio 44114.

EATON

frame computer based on semiconductor technology to this size. The individual chips of silicon on which the circuits are inscribed might fit in the allotted volume, and means might be found for interconnecting them, but such a computer would not work. One reason is that for all the interconnecting wires to fit they would have to be reduced not only in length but also in cross section, with the result that their electrical resistance would remain approximately constant. The transmission time for a signal depends not only on the distance to be traversed but also on the resistance of the path, and so the miniature computer would operate little faster than the full-size one. Actually it would not work at all. Power consumption would be only slightly reduced by the change in scale, so that the grapefruit-size computer would have to dissipate several kilowatts of heat. When it was turned on, it would promptly melt.

The adoption of superconductors would eliminate both of these problems. The resistance of a superconductor is zero even when the cross section is very small, and signals would propagate with a speed about half that of light. For the same reason the power consumption of the superconducting components is zero. In any workable design some circuit elements must always be in the nonsuperconducting state, but even so the total power dissipated would be only a few watts.

A Josephson junction consists of two superconductors separated by a thin barrier of insulating material; a third superconductor carries a control current. Electrons can cross the insulating barrier by the quantum-mechanical process called tunneling, and in the absence of a magnetic field the current of tunneling electrons passes without resistance; hence the voltage measured across the junction is zero. A current in the control line gives rise to a magnetic field that destroys the superconductivity of the barrier; electrons still tunnel through it, but there is now an effective resistance opposing their passage, and a voltage proportional to that resistance appears across the junction. Thus the junction has the two states necessary in any device that is to implement the binary logic of a digital computer.

Experimental chips bearing arrays of Josephson-junction circuits have been constructed at IBM laboratories in Yorktown Heights, N.Y., and in Zurich. Logic circuits suitable for a central processing unit have been fabricated at a density such that from 500 to 1,000 circuits would fit on a single chip. The logic delay (the time required for a signal to pass from the input of one circuit to the input of the next) is about .05 nanosecond. Two kinds of memory element have been tested, one for a fast "cache" memory of limited capacity and the other for a much larger main memory. For the cache memory the access time, or the time needed to retrieve an item from the memory, is one nanosecond. The main memory is slower but also denser; it is estimated that a complete chip will hold some 16,000 bits of information and will have an access time of about 15 nanoseconds. In the main memory each bit is represented by a single quantum of magnetic flux, generated by a persistent current, one that flows indefinitely even in the absence of an applied voltage.

The superconducting traces that define the circuits are made of an alloy of lead, indium and gold, which is laid down on a substrate of silicon. The choice of silicon had nothing to do with its electrical properties, which have made it the basis of the modern semiconductor industry. It was adopted because the technology for fabricating microscopic structures in silicon has been refined by more than 20 years of experience in making semiconductors. In a working computer the Josephson-junction chips would be mounted directly on small cards also made of silicon. (Building the entire machine of a single material would avoid thermal strains when the computer was cooled down to its operating temperature.) Connections between cards would be made by "solder" joints, the solder being mercury, which would solidify when it was immersed in a helium bath.

The devices built so far by IBM are comparatively small arrays of Josephson-junction circuits, and much further work will be needed before a complete computer can be assembled. The design goal calls for a machine with a central processing unit, a cache memory of 256,000 bytes (one byte being equivalent to eight bits of information) and a main memory of 64 million bytes. The basic cycle time of the processor, which is approximately equal to the time required to execute a single instruction, is expected to be four nanoseconds. (The speed of light is roughly one foot per nanosecond.) Mainframe computers today have a cycle time of more than 50 nanoseconds.

It is worthwhile asking why a faster computer is needed. There are a few applications in which absolute speed is essential, mainly those in which the computer must keep up with events as they are happening. The major impetus, however, is an economic one. A Josephson-junction computer might cost about the same as a mainframe computer today but it could do between 10 and 20 times as much. The cost per computation would therefore be reduced by a similar factor.

The Theory of Knots

A piece of knotted string with its ends joined is a good physical model of an abstract mathematical knot. The theory of knots seeks to analyze the different ways of tracing a knotted curve in space

by Lee Neuwirth

Knots aid sailors and challenge boy scouts, and they also provide mathematicians with a serious topic for investigation. In mathematics a knot is defined abstractly, as a one-dimensional curve situated in ordinary three-dimensional space so that it begins and ends at the same point and does not intersect itself. Yet the theory of knots, which seeks to classify and analyze the different ways of tracing such a curve, is not completely unrelated to sheepshanks and half hitches: a piece of knotted string with its ends spliced together so that it cannot be untied is an excellent physical model of the abstract mathematical object. Moreover, some of the deepest problems in this rich and various theory arise when mathematicians try to answer the same questions about a knotted curve that one might ask about a knotted string. Is it really knotted? Can tying a second knot untie the first? In particular, is one knot equivalent to another?

The last question raises the fundamental problem of knot theory. How can different knots be distinguished? To understand the issues involved it might be well to begin by thinking about the problem not mathematically but intuitively. Mathematicians are free to construct a theory any way they like, but there is an obvious practical value in adhering to physical reality. Consider the knotted curves shown in the illustration on the opposite page. If those curves were made of string, most people would agree that the top two figures are knotted differently. The circle is clearly unknotted (indeed, it is technically called the trivial knot), but there seems to be no way to unravel the figure next to the circle without cutting it. On the other hand, although the figures in the remaining pairs have been drawn differently, there does not seem to be any essential difference in the way they are knotted. The bottom figures, which differ only in the position of the loop in color, seem to be equivalent, as do the straight-edged figure and the smooth one next to it and so on. How can these intuitive judgments be supported? Knot theory seeks to supply some way of proving that the curves in the top pair are different, while ignoring the "superficial" differences between the curves in the other pairs.

The area of mathematics where qualities such as smoothness, size and even shape are ignored is the branch of geometry called topology. Topology is concerned with those properties of spaces that remain unchanged no matter how the spaces are deformed. One way topologists investigate such properties is by embedding an object in a space and seeing how the empty part of the space behaves. To see how knots arise in this type of investigation it is important to understand that "knottedness" is a property that is intrinsic not to a curve but to the embedding of the curve, that is, the way the curve lies in three-dimensional space. An imaginary insect crawling along in the one-dimensional space of a knotted curve would not be able to tell that the curve was knotted; from the insect's point of view the curve would be a circle. (In fact, a one-dimensional curve can be knotted only if it is embedded in three-space. In two-space there is not enough room for it to be knotted and in four-space there is too much room, that is, any reasonably embedded curve can be unraveled.)

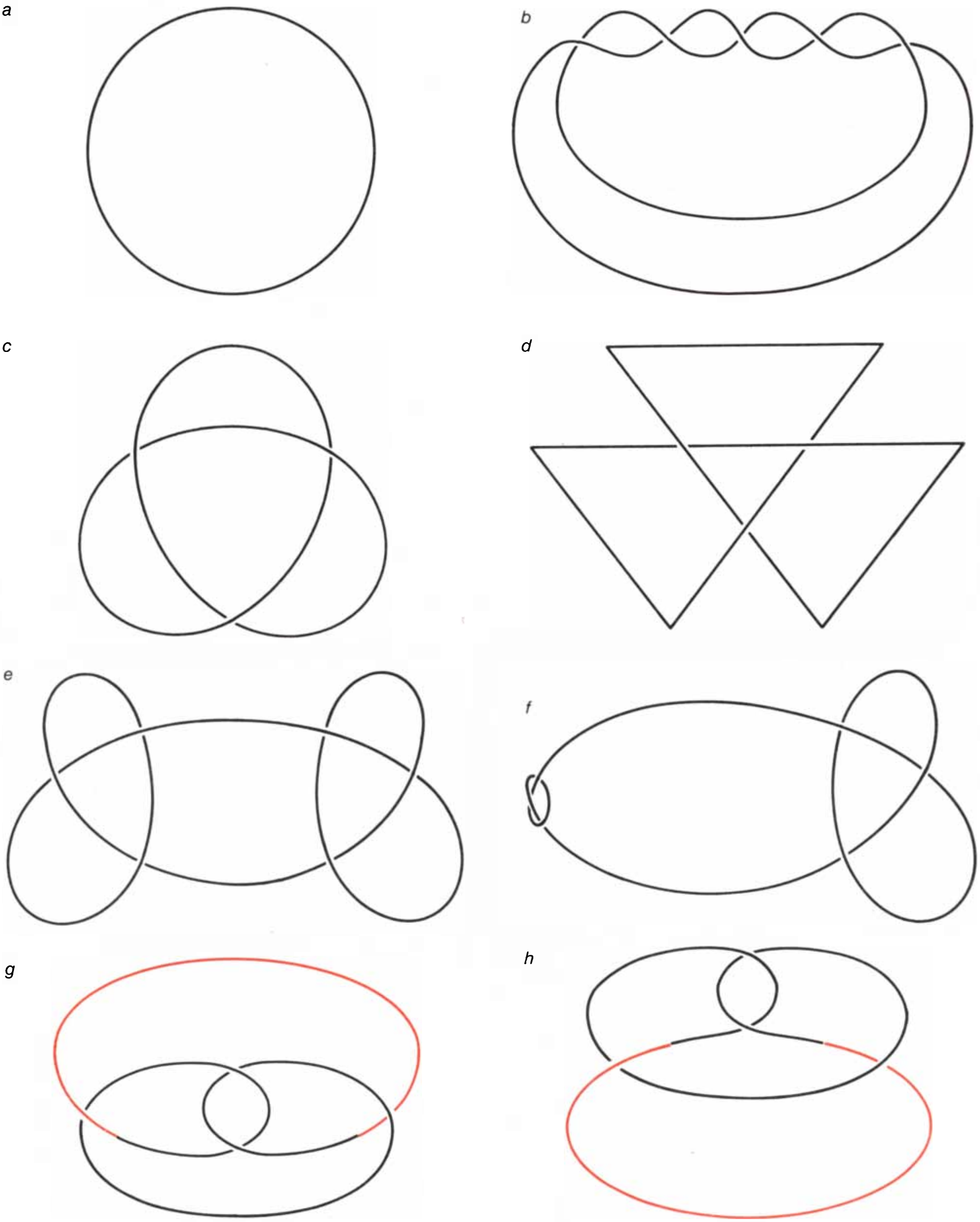
Equivalent Knots

Topologists, then, view knot theory as a rather simple example of what is known as a placement problem: What are the different ways of embedding a circle in three-space? The study of knots and the spaces in which they are embedded is a relatively new area of mathematics, dating from about the turn of the century, but much progress has been made in it. Before describing some of the methods that have been developed for proving that knots are different, however, it is necessary to decide when they can be regarded as being the same.

There is a precise mathematical definition that could be given here to explain which knots are equivalent, but it would call for a good deal of preliminary mathematics. Fortunately resorting to a physical analogy will have almost the same result. To determine whether knots are equivalent, begin by visualizing them as knotted strings. To be more exact, consider a model of each knot obtained by thickening the knot slightly so that it appears to be encased in a solid, flexible three-dimensional tube. Two knots will be called equivalent if the tube model of one can be deformed—pulled, pushed or twisted—into the shape of the other without breaking the tube or pushing it through itself. (Certain "wild" knots, such as the one shown at the bottom of the lower illustration on page 112, cannot be encased in a three-dimensional tube, but they are not dealt with in classical knot theory and I shall not discuss them.)

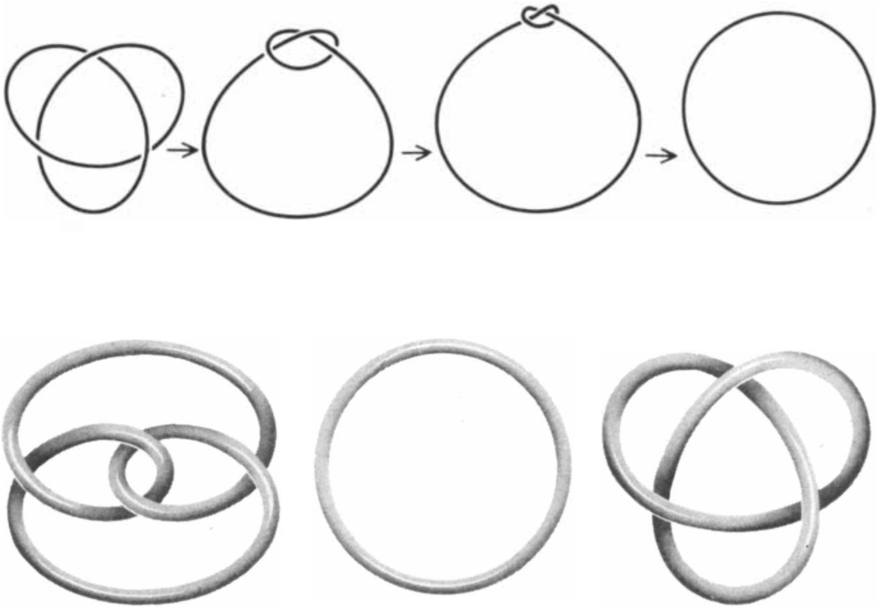
To prove that two knots are equivalent it will suffice to deform the tube model of one until it matches the other. (Knotted one-dimensional curves cannot be handled in this way because nontrivial knots can always be squeezed out of them, as is shown in the upper illustration on page 112.) Returning to the pairs of intuitively equivalent knotted curves shown in the illustration on the opposite page, it is obvious that their tube models can indeed be made to match. On the other hand, it quickly becomes evident that there is no way to twist the model of the circle into the shape of the knot shown next to it. The failure of the trial-and-error method cannot be considered a proof that these two knots or any others are different, however, because the possibility always remains that there is some clever trick for twisting the models that will deform one into the other. Therefore in order to prove that the knots are different it is necessary to find some property that distinguishes them. Such a property of a knot is called an invariant.

An invariant is a characteristic prop-

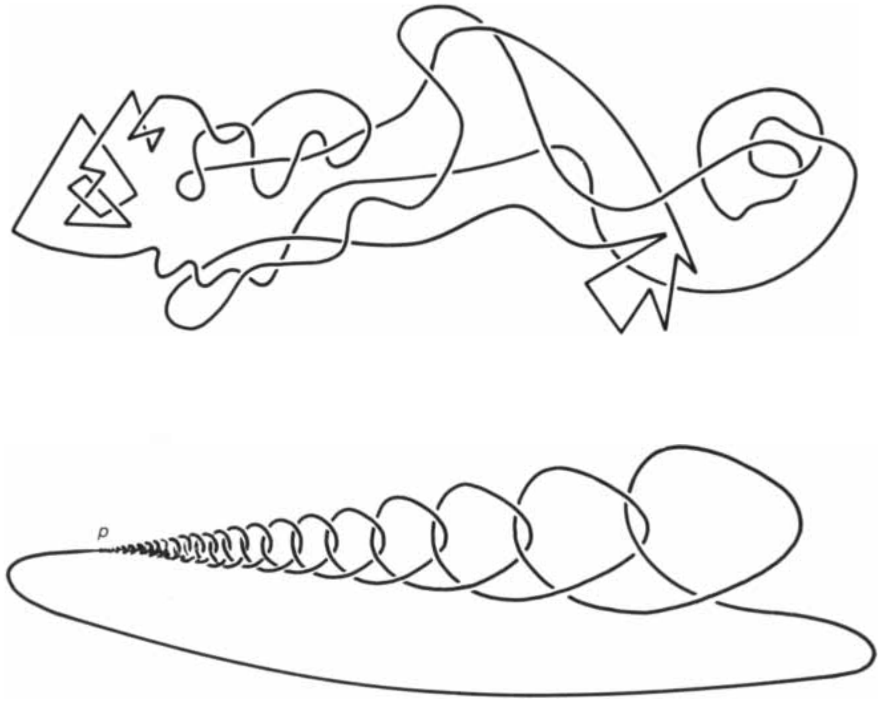


A KNOT is a one-dimensional curve traced in three-dimensional space in such a way that it begins and ends at the same point and does not intersect itself. All the figures shown here represent knots, even the circle *a*, which is called the trivial, or unknotted, knot, and the straight-edged figure *d*. The theory of knots is a part of topology: the branch of geometry concerned with the properties of spaces that remain unchanged when the spaces are deformed. Thus knots *c* and *d*, which differ only in the shape of their loops, and knots *e* and *f*, which differ only in the size of their loops, are considered to be equivalent, as are knots *g* and *h*, which differ only in the position of a loop (*color*). On the other hand, the unknotted circle *a* seems to be intrinsically different from the knotted curve *b* shown next to it. The problem of proving that such knots, or knotted curves, are indeed distinct from one another is the fundamental task of the theory of knots.

erty of a knot that is unchanged by deformations of the tube model of the knot. For example, the minimum number of crossing points required in a drawing of a knot is an invariant that frequently serves as the basis for organizing tables of knots. In some ways, however, the crossing-point number is an unsatisfactory invariant. To begin with it is not always obvious whether a knot has been drawn with the minimum number of crossing points, and so it is not easy to calculate the number. Moreover, many different knots have the same crossing-point number, so that the number is not a very powerful invariant. In other words, the crossing-point number does not distinguish most knots. Finally, because a simple number holds little information about knot structure, the invariant seems to provide little insight into other properties of knots.



A NONTRIVIAL KNOT in a one-dimensional curve can always be "squeezed out," or deformed to a point, as is shown at the top. Because the same is not true of knotted strings, it is convenient to consider models of knotted curves such as the ones shown at the bottom. These are constructed by thickening each curve slightly so that it seems to be encased in a solid, flexible three-dimensional tube. Two knots can be considered to be equivalent if tube model of one can be deformed into shape of other without breaking tube or pushing it through itself.



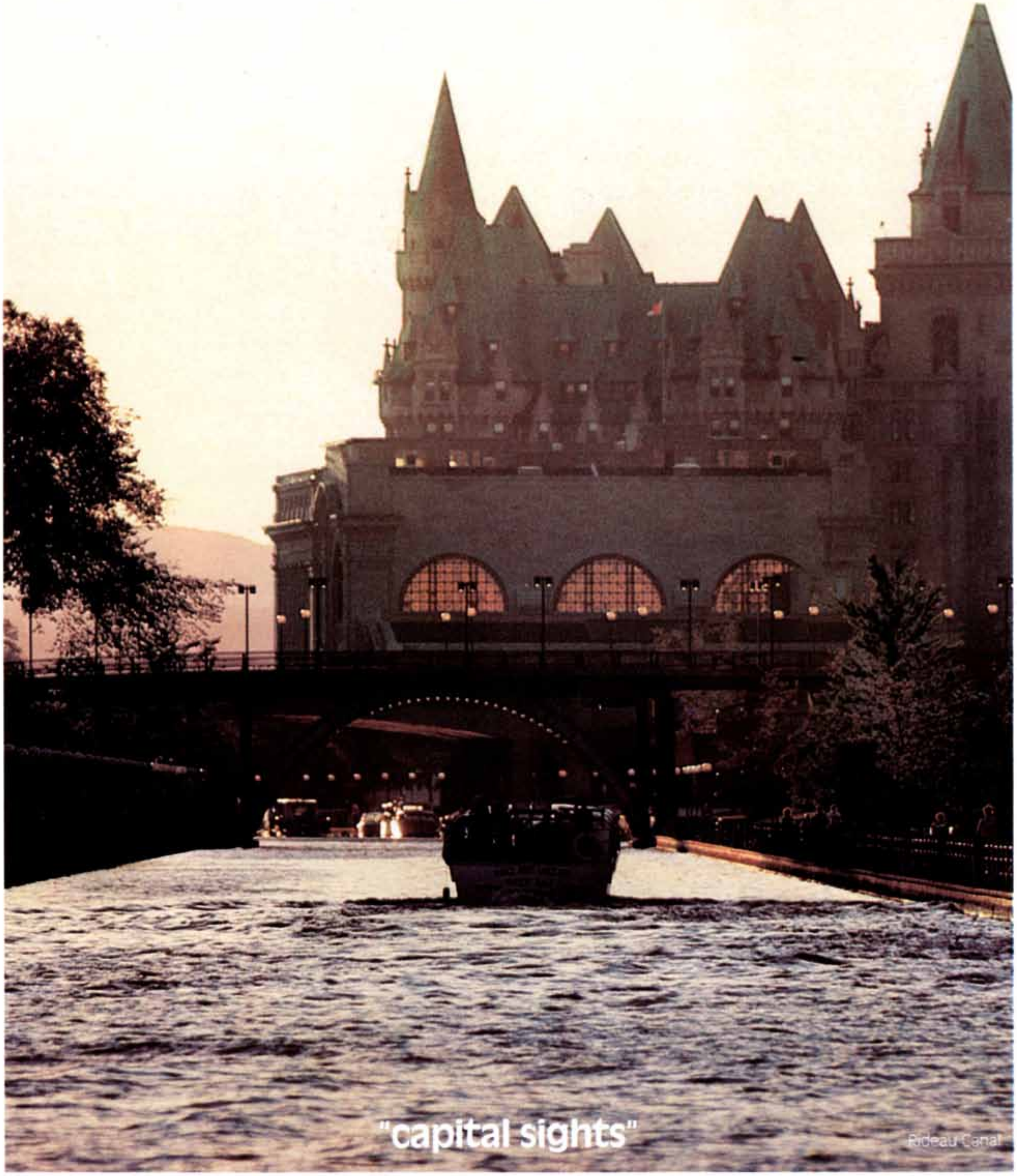
IRREGULAR KNOTTED CURVE such as the one shown at the top is considered to be an acceptable knot, since it can be encased in a solid, three-dimensional tube. The curve shown at bottom cannot be encased in such a tube because its interlaced loops converge to single point *p*. Curves of this type are called "wild" and are not generally studied in classical knot theory.

At present, although there are several sophisticated techniques that in principle could serve to determine whether any two knots are equivalent, there is no practical way to make such a determination. In other words, there is no easily derived invariant that completely characterizes knots. There are, however, several invariants that come close to providing a complete classification. One of them, which I shall describe in some detail, is called the knot group. Roughly speaking, the knot group describes the different ways of moving through three-dimensional space without running into an embedded knot.

Constructing the Knot Group

The space around an embedded knot and the paths through it are most conveniently described in terms of geometry, which investigates problems with spatial representations. Some 80 years ago Henri Poincaré realized that important characteristics of this geometric configuration could also be described in terms of algebra, which investigates problems with symbolic representations. The knot group is only one of the algebraic invariants of knots that have been devised since then. Imposing an algebraic structure on an inherently geometric space presents a formidable (and not always worthwhile) mathematical challenge. In the case of knots, however, most problems have proved to be more easily solved by the manipulation of symbols. Indeed, much of the beauty and depth of knot theory arises from the counterpoint between geometry and algebra, two of the oldest branches of mathematics that appear here in new forms invigorated by new ideas.

To understand how an algebraic structure—a group—can be associated with an embedded knot consider the tube model of a cloverleaf knot shown in the illustration on page 117. The area of three-space outside an embedded knot is called the complement of the knot, and for the sake of convenience I shall call the complement of this knotted tube S and fix a base point b in S. (The structure of the knot group will be independent of the choice of b.) The illustration shows several elements of Ω: the set consisting of all the one-dimensional paths that begin at b, wander through S (thereby avoiding the knotted tube) and finally return to b. Note that the set Ω is made up of directed closed loops in S, which may or may not inter-



Savour the old world flavours, the gracious walks and parkways, the double-decker bus tours and canal cruises, the art, architecture and ambience of Canada's capital – Ottawa. For all you need to plan your vacation, call us COLLECT (416) 965-4008, or write: Ontario Travel, Dept. C.S., Queen's Park, Toronto, M7A 2E5.

Ontario Canada

We treat you royally

Complete Accord.

In 1976 we introduced the Honda Accord® Hatchback. With its stunning array of standard features, it was an instant success. But some of you told us you'd like to have an Accord that was even more luxurious. We responded in 1978 with our elegant Accord LX.

Was that the end of the line? Not quite. We now began hearing from people who loved the Accord but needed a slightly bigger car. So this year we are introducing our brand-new 4-Door Sedan.

That gives us a complete line of Accords. And, more important, the way we equip our cars, every Accord is in its own way a complete Accord.

Besides a roomy interior, the new Accord 4-Door Sedan gives you a spacious trunk that is equipped with an interior light, two stowage boxes for small items, and a lever that permits you to release the trunk lid from the driver's seat. Other standard

I've finally found a personal computer I respect. Compucolor II.

It's not surprising that professionals get excited about the Compucolor II. It's a totally-integrated 8080A system with full color graphics display, built-in 51K mini-disk drive, and the best cost performance ratio available in a personal computer.

The complete system is only $1495.* And that price includes 8K user RAM, RS-232C compatibility and random access file capabilities.

Our 8 foreground and background colors will boost your comprehension, while introducing you to an exciting new dimension in BASIC programming. The vector graphics have 16,484 individually-accessible plot blocks. And the 13″ diagonal measure screen gives you 32 lines of 64 ASCII characters. You also have the flexibility that comes with 16K Extended Disk BASIC ROM.

Compucolor II offers a number of other options and accessories, like a second disk drive and expanded keyboard, as well as expandability to 32K of user RAM. Of course we also have a whole library of low-cost Sof-Disk™ programs, including an assembler and text editor.

Visit your nearest computer store for details. And while you're there, do some comparison testing. With all due respect to the others, once you see it, you'll be sold on the Compucolor II.

Compucolor® Corporation

Unretouched photo of screen

*U.S. Domestic price

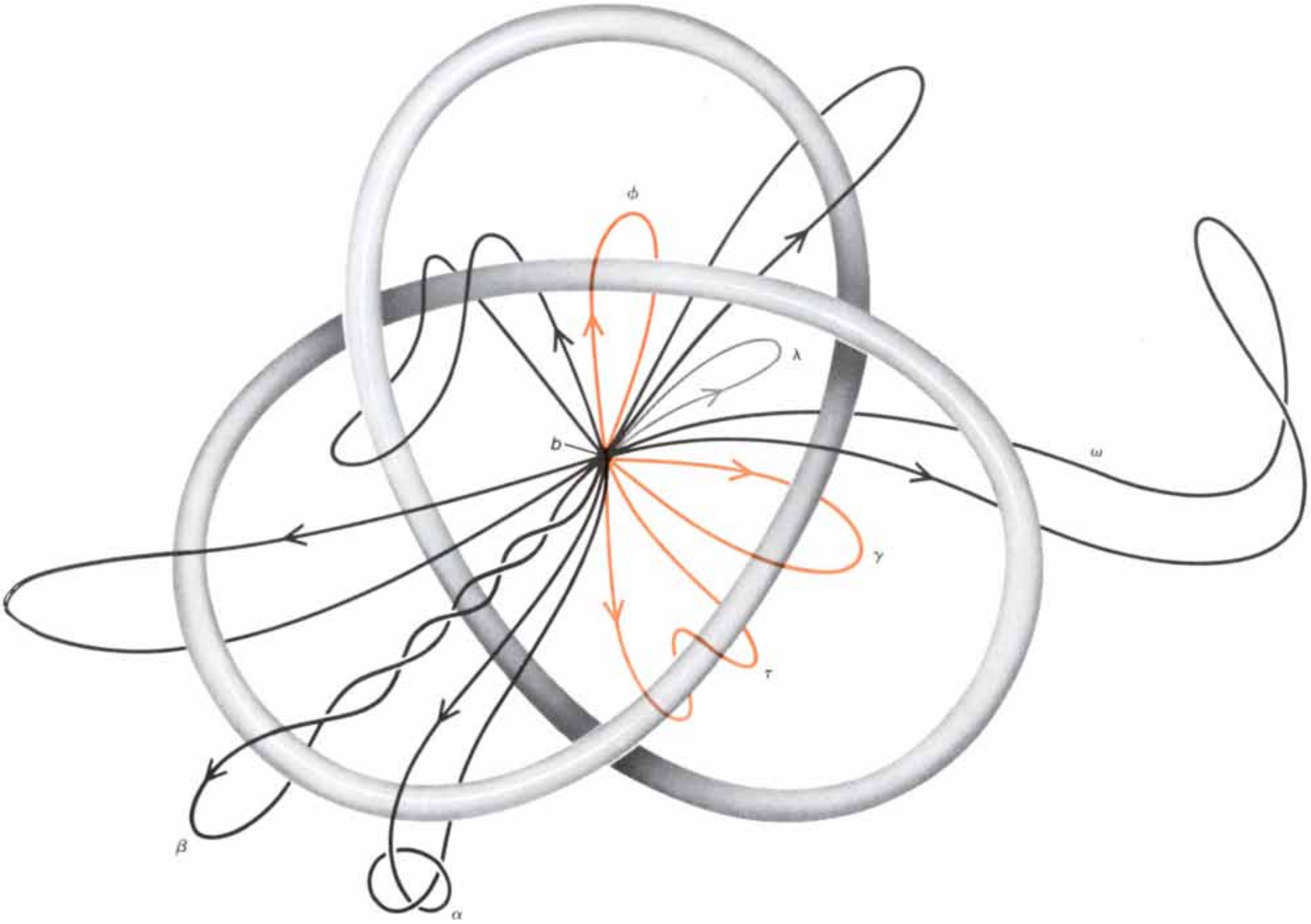
Compucolor Corporation Post Office Box 569 Norcross, Georgia 30071 Telephone 404/449-5996 sect themselves, and includes the constant, or trivial, loop *e*, which never leaves *b*. Although the set Ω is clearly quite large, there is a way of classifying its elements that will make it far more manageable.

The set Ω can be divided into classes of equivalent paths by calling any two paths α and β equivalent if α can be deformed into β. In the process α may be pulled, pushed or even crossed over itself, but neither its beginning point nor its end point may be moved, and it may not be broken or moved out of *S*. (In other words, α may not come in contact with any segment of the knot.) For example, among those elements of Ω shown in the illustration the path λ can be shrunk back to the base point *b* (so that λ is equivalent to the constant path *e*), but the path β, which winds around a segment of the knot, cannot be pulled back that far. Path β can be untwisted, however, and the "knot" in path α can be removed by crossing the path over itself, and so those paths are equivalent. Deformations of this type are called homotopies, and paths such as α and β that differ only by a homotopy are said to be homotopic. In the illustration all the paths in black are homotopic, whereas the paths in color are not homotopic to one another or to the paths in black.

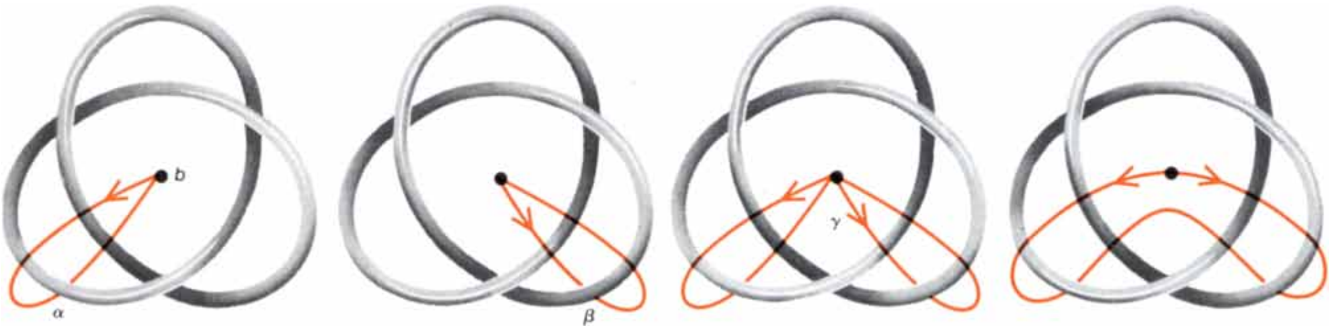
The class of paths homotopic to a particular path α is written $[\alpha]$. It should be evident, however, that the composition of the class $[\alpha]$ does not depend on the choice of α. If α is homotopic to β, the classes $[\alpha]$ and $[\beta]$ will be identical. In this way the set of paths Ω has been partitioned into a collection of clearly defined homotopy classes. Moreover, there is a very natural form of multiplication that can be defined for these classes: the product $[\alpha][\beta]$, or $[\alpha]$ times $[\beta]$, of any two classes of paths in Ω is equal to the class $[\gamma]$, where γ equals $\alpha\beta$, the path that begins at *b*, follows α back to *b* and then follows β back to *b*. It is not difficult to see that the class $[\gamma]$, or $[\alpha\beta]$, is also well defined in the sense that its composition does not depend on the choice of representative paths α and β.

The Knot Group

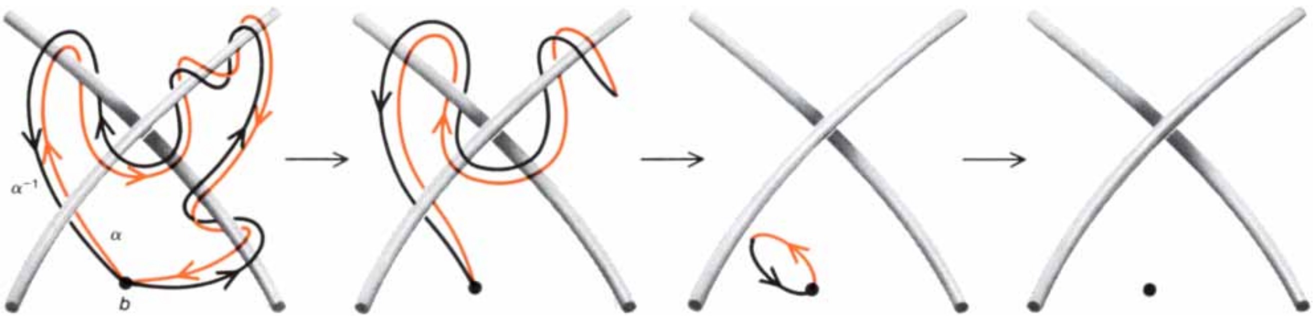
Up to this point the description of the set of paths through the complement of an embedded knot has been basical-



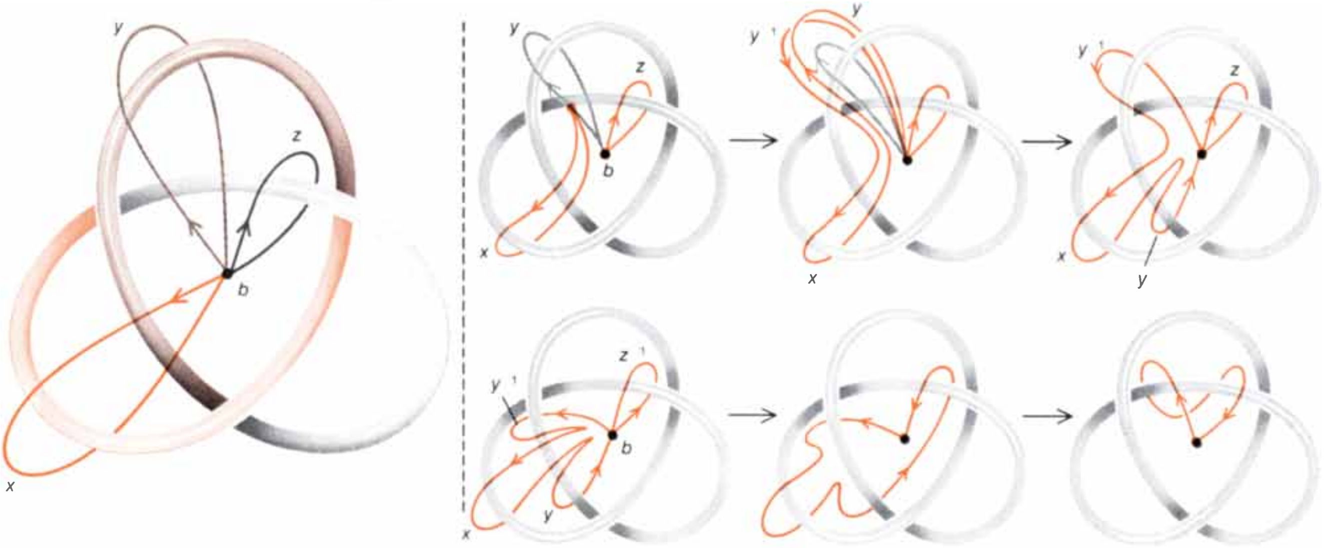
KNOTTEDNESS is a property not of a curve but of the way a curve is embedded in three-space. Thus one way to distinguish a particular knot type is by characterizing the ways of moving through three-space without running into the embedded knot. For example, several of the closed paths through the space around the cloverleaf knot from a fixed (and arbitrary) base point *b* are shown. (In knot theory the area of three-space outside an embedded knot, which is called the complement of the knot, is always assumed to be empty.) Not all the paths from *b* represent topologically different routes through the complement space, and so two paths are considered to be equivalent if there is a way of deforming one into the other. In such a deformation, called a homotopy, a path may be stretched, contracted or even crossed over itself, but its end points may not be moved, and it may not be pulled or pushed through a segment of the knot. The paths shown in black are all homotopic; for example, the "knot" in path α can be removed by crossing the path over itself, and path β can be untwisted. The paths shown in color are not homotopic to one another (or to the black paths); for example, paths γ and ω wind around the same segment of the knot in different directions, path τ goes in the same direction around the same knot segment as ω but winds around one time too many, and path ϕ goes around a different segment. Path λ, shown in gray, can be deformed back to base point *b*, making it homotopic to the constant path *e*, which never leaves *b*. Introduction of homotopy equivalence relation transforms set of paths through complement of knot into collection of classes of homotopic paths. Class of paths homotopic to particular path α is written $[\alpha]$.



COLLECTION OF HOMOTOPY CLASSES of paths through the complement of the cloverleaf knot can be treated algebraically, or symbolically, by imposing on the classes the form of multiplication shown here. The product $[\alpha][\beta]$ of the classes represented by α and β is defined as the class $[\gamma]$, where γ is the path $\alpha\beta$ that begins at b, follows α and then follows β, ending back at b. Another path in $[\gamma]$ (that is, a path homotopic to $\alpha\beta$) is shown at the far right. It is not hard to see that this product is associative, so that $([\alpha][\beta])[\tau]$ equals $[\alpha]([\beta][\tau])$ for any classes $[\alpha]$, $[\beta]$ and $[\tau]$. On the other hand, the product is not commutative, that is, $[\alpha][\beta]$ is not usually equal to $[\beta][\alpha]$. Class of constant path $[e]$ acts as identity element, so that for any class $[\alpha]$, $[e][\alpha]$ equals $[\alpha][e]$, which equals $[\alpha]$.



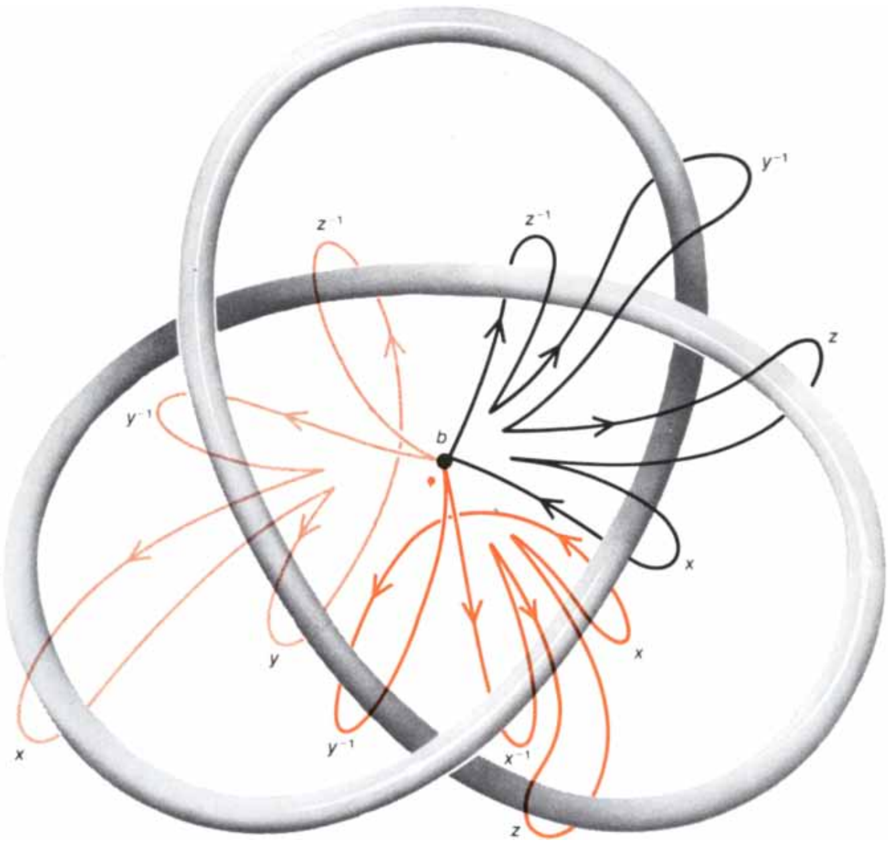
INVERSE OF A HOMOTOPY CLASS $[\alpha]^{-1}$ is defined as the class of paths homotopic to α^{-1} (*black*), or α (*color*) traversed in the opposite direction. As is shown here, $\alpha\alpha^{-1}$ is homotopic to e, so that $[\alpha][\alpha]^{-1}$, which equals $[\alpha][\alpha^{-1}]$, equals $[e]$. In algebra any set and associative product with an identity element and an inverse for every element in the set is called a group. The group whose elements are the homotopy classes of paths through the complement of a knot is called the knot group. It is not hard to see that no matter how a knot is deformed its knot group remains the same. In other words, the group is an invariant characteristic of a knot. Moreover, many (but not all) knots are distinguished by their group (*see top illustration on page 120*), and so group is powerful tool for differentiating knots.



KNOT GROUP IS CALCULABLE in the sense that it is always possible to construct a finite, explicit description, or presentation, of the group. In constructing a presentation of the group of the cloverleaf knot shown at the left, for example, one should first note that the crossing points of the knot divide it into three segments, taken from underpass to underpass. Then, choosing an arbitrary base point b in the complement of the knot, one can draw three paths x, y and z, each one looping around a different knot segment. The homotopy classes $[x]$, $[y]$ and $[z]$ are called generators of the knot group because any element of the group can be expressed as a product of these three classes and their inverses. Moreover, at each of the crossing points of the knot the three segments come together in a different way, defining a different relationship among the generators. In expressing the relationship at, say, the crossing point at the upper left in the knot, one should observe that pulling the end points of path x along path y creates a path that is homotopic to z. As is shown in the sequence at the top right, however, the path is also homotopic to $y^{-1}xy$. Therefore at this crossing point $[y]^{-1}[x][y]$ equals $[z]$, or, as is shown in the sequence at the lower right, $[y]^{-1}[x][y][z]^{-1}$ equals $[e]$. This equation (in either form) is called a relation of the knot group. The relations among the generators at the other two crossing points are obtained in a similar manner (*see illustration on opposite page*).

ly geometric. With the imposition of a form of multiplication on the homotopy classes of those paths, however, there begins to emerge an algebraic structure, a way of treating the paths symbolically. Some observations about that structure will help to identify it. First of all, note that the multiplication of classes is an associative operation: for any paths α, β and γ in Ω, $([\alpha][\beta])[\gamma]$ will always be equal to $[\alpha]([\beta][\gamma])$. On the other hand, a little experimentation shows that $\alpha\beta$ is not usually homotopic to $\beta\alpha$. Therefore the multiplication of homotopy classes is not commutative, that is, $[\alpha][\beta]$ is not usually equal to $[\beta][\alpha]$ unless either $[\alpha]$ or $[\beta]$ is equal to the class of the trivial path $[e]$. The class $[e]$ acts as an identity element in the multiplication of classes of paths in Ω (just as 1 does in the multiplication of integers), so that for any class $[\alpha]$, $[\alpha][e]$ equals $[e][\alpha]$, which equals $[\alpha]$. Finally, it is evident that for every element $[\alpha]$ there is an inverse element $[\alpha]^{-1}$ such that $[\alpha][\alpha]^{-1}$ equals $[\alpha]^{-1}[\alpha]$, which equals $[e]$. (As is shown in the middle illustration on the opposite page, $[\alpha]^{-1}$ is the class of paths homotopic to α^{-1}: the path in Ω consisting of α traversed in the opposite direction.)

Any set with an associative operation defined on its elements in such a way that there is an identity element and an inverse for every element in the set is an algebraic object called a group. It is not hard to see that the group of the knot—its collection of homotopy classes of paths and the imposed product—is indeed an invariant of a knot. Consider what happens to a knot group when the knot, or knotted tube, for which it was defined is deformed into an equivalent knot. In the course of the deformation all the paths in the complement of the first knot will be turned into paths in the complement of the second knot. In particular, any pair of homotopic paths in the complement of the first knot will be turned into a pair of homotopic paths in the complement of the second knot. Because the same is true for the deformation of the second knot back into the first knot the homotopy classes and thus the groups of two knots will be equivalent. With a little effort these observations can be transformed into a rigorous proof that the group of a knot is indeed an invariant of the knot.



KNOT-GROUP RELATIONS arising at the three crossing points of the cloverleaf knot are $[y]^{-1}[x][y][z]^{-1} = [e]$ (*light color*), $[z]^{-1}[y]^{-1}[z][x] = [e]$ (*black*) and $[x]^{-1}[z][x][y]^{-1} = [e]$ (*dark color*). From these three relations any equation among the elements of the knot group can be derived. Therefore since any element of the group can be derived from the generators and their inverses, the list of generators and relations of the cloverleaf knot group describes the group completely. There are many other algorithms for deriving presentations, or lists of (possibly different) generators and relations, that explicitly describe the group of a particular knot.

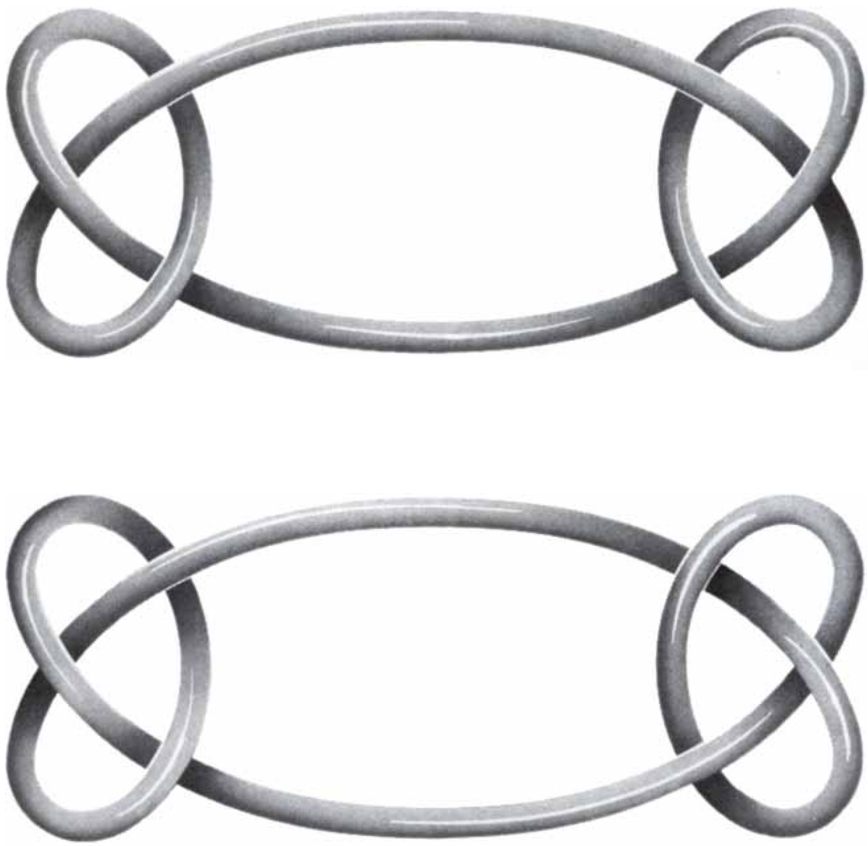
Presentation of the Knot Group

A knot group is a rather unwieldy mathematical object, consisting of an infinite number of abstract equivalence classes of paths. Therefore this invariant can be used for differentiating knots only if there is some way of explicitly describing it. The knot group can, in fact, always be calculated: it is always possible to construct a finite list of concrete objects that will completely present, or describe, the group. Such a list will consist of a number of group elements called generators and a number of equations called relations.

To see how such a list can be derived consider the knot shown at the left in the bottom illustration on the opposite page. The three crossing points of the knot divide it into three segments, taken from one underpass to the next. (These segments have been shown with differently colored shadings in the illustration.) An arbitrary base point b has been fixed in the complement of the knot, and for each knot segment a path has been drawn from b that loops around the segment and returns to b. The three paths x, y and z obtained in this way give rise to the three homotopy classes $[x]$, $[y]$ and $[z]$ in the knot group. These classes are called generators of the group.

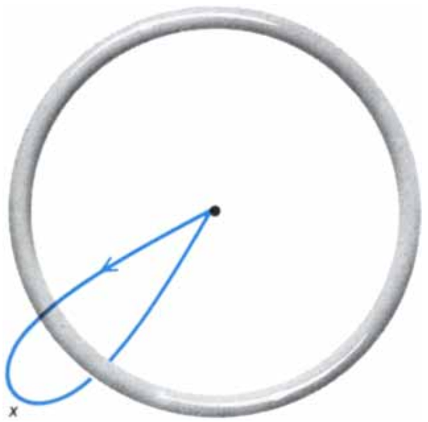
Observe that at each crossing point the three segments of the knot come together so that one is on top and the other two meet below it. For example, at the crossing at the upper left the segment associated with the generator $[y]$ is on top and the segments associated with $[x]$ and $[z]$ meet below it. Thus each crossing point brings together the generators in a different way, determining some relationship among them. The way the generators are related at this particular crossing point can be seen by moving the end points of path x along path y, as is shown at the top right in the illustration. In this way a path is created that is homotopic to z and also homotopic to $y^{-1}xy$. Thus the equation $[y]^{-1}[x][y] = [z]$, or, as is shown at the bottom right, $[y]^{-1}[x][y][z]^{-1} = [e]$, expresses the way the generators meet at the crossing point. This procedure can be repeated at the other two crossing points to obtain the equations $[z]^{-1}[y]^{-1}[z][x] = [e]$ and $[x]^{-1}[z][x][y]^{-1} = [e]$, as is shown in the illustration above. These equations are called relations of the knot group. In 1910 the German mathematician Max Dehn proved that for any knot the generators and relations obtained in this way provide a complete description of the knot group in the sense that any group element can be expressed as some product of the generators and their inverses, and any equation among the group elements can be derived from the relations. Any list of generators and relations that satisfies these requirements is called a presentation of the group.

Now the transformation of a geomet-



NOT ALL KNOTS ARE DISTINGUISHED by their group. For example, the square knot shown at the top and the granny knot shown at the bottom have identical groups. More sophisticated techniques must be applied in order to prove that these two knots are indeed distinct.

ric problem, concerned with manipulating knots in space, to a purely algebraic one, concerned with manipulating symbols, is complete. The method I have described for constructing a presentation of the cloverleaf knot group can be generalized for treating more compli-



TRIVIAL, OR UNKNOTTED, KNOT has the simplest knot group: the group presented by only one generator $[x]$ and no relations. No other knot is described by this group, and so in determining whether a particular knot is really knotted it is sufficient (although not always easy) to determine whether its group is made up only of powers of $[x]$ and $[x]^{-1}$.

cated knots, and there are a number of other algorithms for calculating explicit group presentations. Thus for every knot there is a calculable presentation of its knot group, and the relations in the presentations can be treated almost exactly like ordinary algebraic equations with several variables. The only difference is that the order of the homotopy classes in the relations must be maintained. But even though group elements cannot be commuted, both sides of a relation can be multiplied by the same symbols, and a symbol and its inverse can be canceled when they are contiguous. (It should be noted that a new symbol must be entered on the same side of each half of an equation. In other words, if $[x] = [y]$, then $[a][x] = [a][y]$, but it is not usually true that $[a][x] = [y][a]$. Similarly, because the paths in the homotopy classes of a group have direction it is necessary to reverse the order of the classes in a product when its inverse is taken: $[xy]^{-1} = [y]^{-1}[x]^{-1}$.)

Therefore new symbols can be introduced into group presentations and superfluous ones can be eliminated, and as a result quite different-looking presentations of a single group can be obtained. For example, in the presentation of the cloverleaf knot group $[z]$ is expressed as the product of $[x]$, $[y]$ and their inverses, so that by simple substitution and some manipulation of equations $[z]$ can be eliminated, yielding another perfectly valid presentation: $[x]$, $[y]$; $[x][y][x] = [y][x][y]$. (In this way at least one of the generators can always be eliminated from a Dehn presentation of a knot group.)

The presentation of the cloverleaf group can be transformed further by setting $[u]$ equal to $[x][y]$ and $[v]$ equal to $[x][y][x]$. It is then not difficult to show that any product of $[x]$, $[y]$ and their inverses can be written as a product of $[u]$, $[v]$ and their inverses, and that the relation $[x][y][x] = [y][x][y]$ can be transformed into $[u]^3 = [v]^2$. Therefore the cloverleaf group can be described in yet another presentation, which discloses a subtle, unsuspected relation among the elements of the group: $[u]$, $[v]$: $[u]^3 = [v]^2$.

Applications of the Knot Group

Since the aim of knot theory is to distinguish knots, the realization that a single knot group may have such dissimilar presentations is disappointing. One may well wonder what advantage has been gained by the substitution of knot groups for knots or, more broadly, of algebra for geometry. The advantage is unquestionably substantial. To begin with, the knot group turns out to be a very powerful invariant, distinguishing far more knots than, say, the crossing-point number. Not all knots are distinguished by their groups, but only a little more information is needed to determine whether any two knots are equivalent. In other words, knot groups, for which a presentation can always be calculated, are almost strong enough to completely characterize knots.

Moreover, the fact that the presentations of identical groups can appear in such different guises is not as big a problem as it first seems to be. The problem of proving algebraic objects to be equivalent (by, say, matching up their members) can often be solved by calculation, and there is a wide range of purely algebraic theorems that can be exploited for the purpose. On the other hand, it may be possible to distinguish groups by identifying characteristic properties in which the groups differ. Much of the algebraic side of knot theory is devoted to solving such problems.

The strength of the knot-group invariant flows from the fact that unlike the crossing-point invariant it can be easily calculated and it captures a large measure of the complexity and structure of the knot. For this reason it can be used to answer a variety of basic, practical questions about knots, some of which I posed above. For example, is it possible to untie a knotted curve—to make it equivalent to a circle—by tying an additional knot in it and somehow unraveling the result? To answer the question consider the knot group defined on the original knotted curve. Tying an additional knot in the curve will create a knot group that is bigger and more complicated than the original one. On the other hand, as is shown in the bottom illustration on the opposite page, the trivial, or unknotted, knot has the smallest and least complicated group: the group presented by one generator and no relations. Therefore the group resulting from the added knot cannot be equal to the group of the trivial knot, and so it is not possible for the second knot to "cancel" the first.

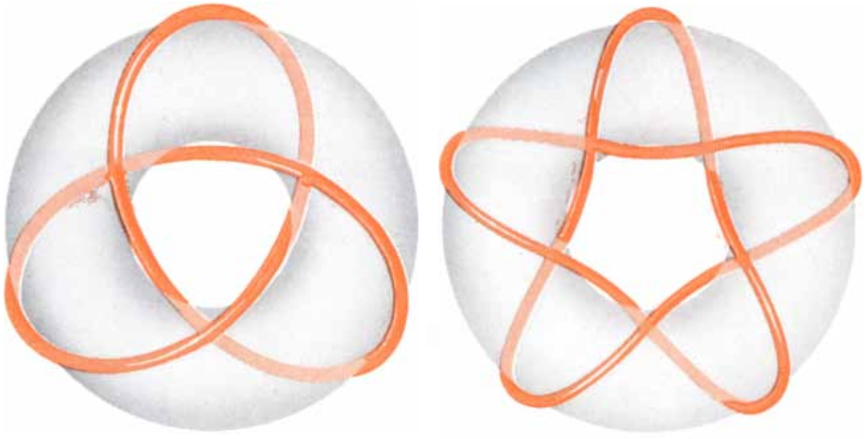
The observation that the trivial knot has such an uncomplicated group presentation raises a deceptively simple question. Are all knots with that presentation unknotted? The deep issues implicit in this very natural question took many years to resolve. In one of the first papers on knot theory, published in 1910, Dehn maintained he had proved that any knot whose group is presented by a single generator is unknotted, but in 1928 the German mathematician H. Kneser discovered that the demonstration of an intermediate result in Dehn's proof of this theorem was incomplete. Another 29 years passed before C. D. Papakyriakopoulos of Princeton University was able to prove the result (which ironically had come to be known as Dehn's lemma) and thus the theorem.

Dehn's theorem is an important and useful result. It does not, however, solve the problem of deciding whether an arbitrary knot is knotted, because it is not always easy to determine whether a knot group is equivalent to the group of the trivial knot. It does establish that the trivial knot has the distinction of being completely characterized by its group. There is another class of knots for which almost the same claim can be made. The knots in this class are those that can be traced on the surface of a torus, the geometric form resembling an inner tube that is generated by rotating a circle around a line in the same plane as the circle but outside it.

Torus Knots

Torus knots arise in the study of equations in two complex variables. (The branch of mathematics in which geometric forms that arise from algebraic considerations are studied is algebraic geometry.) Because these knots are restricted to the surface of the torus, which is considered to be embedded in three-space in the most natural way, their knot groups have simple, stereotyped presentations. Moreover, torus knots are almost completely characterized by their groups.

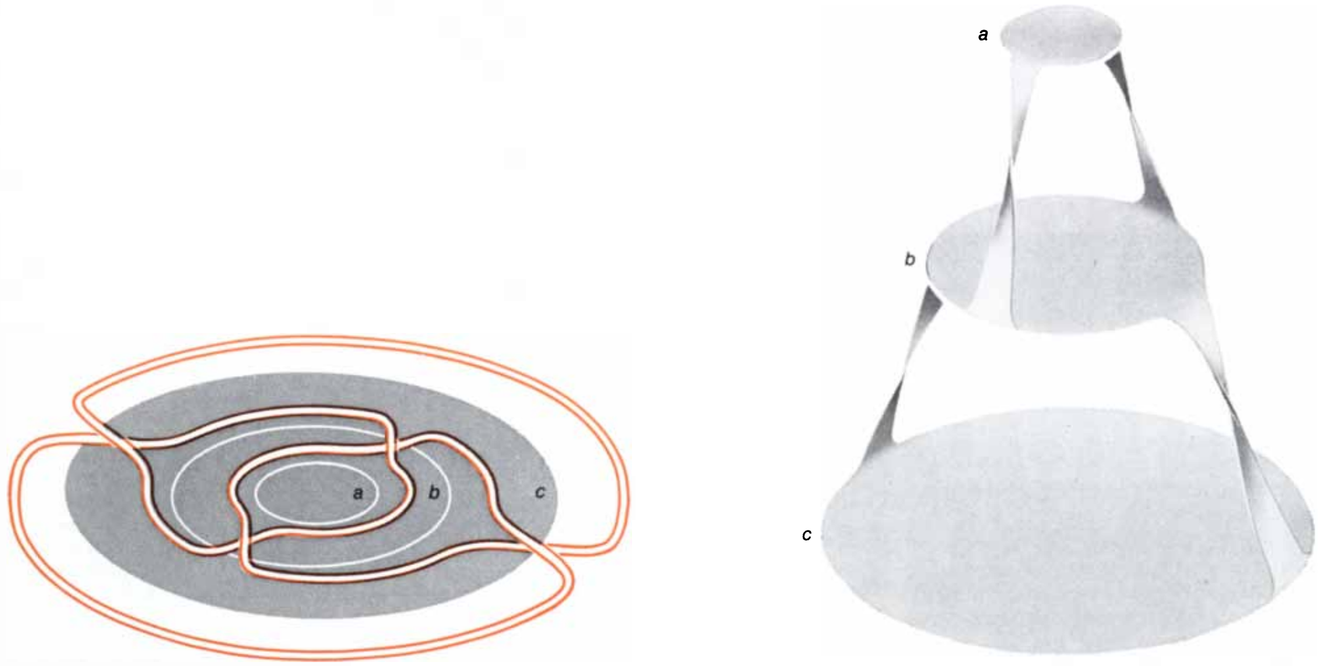
Examples of torus knots appear in the



TORUS KNOTS are a class of knots for which the group is a particularly effective invariant. These knots are restricted to the surface of a torus (which is considered to be embedded in three-space in the most natural way), and so they have extremely simple, regular groups that characterize them almost completely. Only a torus knot and its mirror image (the knot obtained by changing all undercrossings of the original knot to overcrossings and all overcrossings to undercrossings) cannot be distinguished by the torus knot group (*see illustration below*).

PUSHING A KNOT INSIDE A TORUS (*lower left*) shows that the knot winds three times around the inner circumference marked by path *a*. Pushing the knot to the outside of the torus (*lower right*) shows that it winds five times around the outer circumference marked by path *b*. More generally, any torus knot will wind *p* times around the inner circumference and *q* times around the outer circumference. For example, in the illustration at the top of this page, for the knot shown at the left *p* equals 2 and *q* equals 3, and for the knot shown at the right *p* equals 2 and *q* equals 5. It is this characteristic design of all torus knots that gives rise to their stereotyped group presentations. Note that paths *a* and *b* lie in the complement of the knot and have a common base point on the torus. The group of any torus knot can be presented by the homotopy classes [*a*] and [*b*], which are related by the single, simple equation $[a]^p = [b]^q$. This equation means that in the complement of the torus knot a path that is homotopic to *a* wound *p* times around the inside of the torus can be deformed into a path homotopic to *b* wound *q* times around the outside of the torus. It is not difficult to prove that the groups of two torus knots are the same if and only if the knots have the same pair of values *p* and *q*. Thus the knot group alone can serve to determine whether any two torus knots are equivalent (up to mirror images).

KNOT CAN BE VIEWED GEOMETRICALLY as the single edge, or boundary, of a two-dimensional surface. Two surfaces bounded by the cloverleaf knot are shown here. Both were constructed by using disks to "fill in" loops in the knot and ribbons with a half twist to fill in the crossing-point areas and connect the disks. Bottom construction is said to be orientable, or two-sided, because in traveling from level to level one always remains on the same side of surface. (The same is not true of top construction, which is said to be nonorientable.)



ORIENTABLE SURFACE can be constructed for any knot so that the knot is the only edge of the surface. The first general method for constructing such an orientable surface, demonstrated here for the figure-eight knot, was developed by the German mathematician H. Seifert in 1935. The significance of Seifert's construction, which led to the definition of another knot invariant called the genus, is that any orientable surface is topologically equivalent to a disk with a number of "handles" attached to it (*see illustration on opposite page*).

top illustration on page 121. As is shown in the bottom illustration on page 121, when such a knot is traced around the torus, it winds p times around the inner circumference of the torus and q times around the outer circumference, for some positive integers p and q. It is this essential characteristic of all torus knots that gives rise to their highly regular presentations. From a fixed base point in the complement of any torus knot (and on the torus) draw one path inside the torus the long way around and another path outside the torus the short way around. Calling these paths a and b, as is shown in the illustration, one may select as the generators of the group of the torus knot the homotopy classes $[a]$ and $[b]$. These generators are related by the single equation $[a]^p = [b]^q$. In other words, in the complement of the knot any path homotopic to the one consisting of a wound p times through the torus can be deformed into a path homotopic to the one consisting of b wound q times around the torus. (A little doodling will suggest that p and q must be relatively prime, that is, it is impossible to draw a knot on the torus for which p and q have a common divisor.)

The torus knots are the only knots with a group presented by two generators and this simple defining relation. Moreover, it is not difficult to show that two such groups are equivalent only if they have the same pair of values p and q. Hence the torus knots can indeed be completely classified according to their group. It should be noted, however, that the group does not differentiate between a torus knot and its mirror image, which is obtained by changing all the undercrossings of the knot to overcrossings and all the overcrossings to undercrossings.

The group of a torus knot is probably the least complex of all knot groups other than the trivial knot. Perhaps as a result the group seems to be more tightly organized than most knot groups. For example, it is the only nontrivial knot group in which there is an element other than the identity element that commutes with every other element in the group. In other words, there is a nontrivial element $[c]$ such that $[c][x] = [x][c]$ for every homotopy class $[x]$ in the knot group. (In view of the fact that powers of $[a]$ commute with $[a]$ and powers of $[b]$ commute with $[b]$ it is not hard by examining the defining relation $[a]^p = [b]^q$ to determine the identity of one such $[c]$: it is equal to $[a]^p$.) Thus among nontrivial knot groups the group of a torus knot comes closest to displaying complete commutativity.

HANDLE IS ATTACHED TO A DISK by cutting two holes in the disk and stretching a tube between them. In these examples handles are arranged so simply that edges are not knotted. Handles of surface bounded by truly knotted curve would be intertwined in more complex way.

A Geometric Invariant

The torus knots provide a striking demonstration of how effective an algebraic invariant can be, but there are other types of invariants that illuminate knots and knot theory. In particular a geometric invariant has been utilized to answer some theoretical questions concerning knots. This invariant is called the genus of the knot, and it is defined through the construction of a surface whose only edge is the knot.

To understand how this invariant is derived begin by considering a two-dimensional surface with a single edge that is embedded in three-space. This type of surface may be either two-sided as a disk is or one-sided as a Möbius band is. The two-sided surface is said to be orientable and the one-sided surface is said to be nonorientable. Orientable surfaces are particularly agreeable because they are all topologically equivalent to a disk with a certain number of "handles." (A handle is added to a disk by cutting two holes in the disk and stretching a tube from one hole to the other.) Returning to the construction of a surface with a knot as its edge, the top illustration on the opposite page demonstrates that such a surface may be either orientable or nonorientable. In 1935, however, the German mathematician H. Seifert devised a way to construct for any knot an orientable surface with the knot as its only edge.

Seifert's construction proved that a knot must bound at least one surface consisting of a disk with handles. Actually for any given knot there will be many such surfaces with different numbers of handles. Of those surfaces the one with the fewest handles is called a minimal surface of the knot. The genus is defined as the number of handles on a minimal surface. Deforming the knot cannot reduce the number of handles on a minimal surface, and so the genus is indeed an invariant of the knot.

The genus and the related minimal surfaces have proved to be valuable tools for treating a variety of knot problems. For example, although the genus is not nearly as powerful an invariant as the knot group, it has been shown to be somewhat more effective than the knot group in determining whether an arbitrary knot is knotted or unknotted. It is obvious that the trivial, or unknotted, knot—the circle—bounds a disk with no handles and therefore has a genus of zero. In addition it can be proved that the trivial knot is the only knot with that genus number. Hence it is possible to determine the true "knottedness" of any knot whose genus can be calculated. In general, however, the genus is difficult to calculate, and so it would appear to offer no advantage over the knot group.

In 1962, however, Wolfgang Haken of the University of Illinois at Urbana-Champaign found an algorithm for calculating the genus of any knot from a picture of the knot. Moreover, last year Friedhelm Waldhausen of the University of Bielefeld used results obtained by Haken and G. Hemion of Bielefeld to construct an algorithm for deciding whether two pictures represent equivalent knots. These algorithms are actually far too unwieldy to be applied, so that practically speaking there is still no way to tell real knots from trivial ones or one knot from another. The algorithms nonetheless demonstrate that the problems are decidable; one could, at least in principle, build a machine that would accept a picture of any knot, calculate

GENUS OF A KNOT is defined as the minimum number of handles required in an orientable surface with the knot as its only edge. The surface shown is a disk with one handle, but it has been deformed so that the knottedness of its boundary is evident. In other words, the boundary of the surface is a knot of genus 1 or less. This particular knot cannot be placed on the boundary of a disk, so that its genus is indeed 1. The genus number is an invariant of a knot, and although it is less powerful than the knot-group invariant, it has proved to be useful in settling a number of important questions in the theory of knots.

its genus and decide whether or not the knot was knotted. Or one could build a machine that given the pictures of two knots would decide whether or not they were equivalent.

The genus is a powerful geometric invariant of a knot, just as the knot group is a powerful algebraic invariant. An interesting recent result in knot theory is the discovery of a relation between these mathematical constructs. In fact, Charles D. Feustel of the Virginia Polytechnic Institute and Wilbur Whitten of the University of Southwestern Louisiana have shown that the genus, which was defined purely geometrically, is an algebraic invariant that is determined by the group. Once again the interplay between algebra and geometry supplies knot theory with depth and excitement.

In this article I have tried to show how some fundamental problems in knot theory have been solved and how others are being attacked. At the present time, however, mathematicians are taking many other directions in the study of knots. For example, knots are currently being applied to the construction and study of a variety of interesting three-dimensional spaces. Moreover, the subject of higher-dimensional knots—two-dimensional surfaces knotted in four-space, three-dimensional surfaces knotted in five-space and so on—is attracting much attention. Many of the results obtained in classical knot theory can be generalized to these higher dimensions, opening richer and more complex subjects for mathematical investigation.

Working with one of the most modern fleets in the world is a source of pride for us.

Iberia's fleet—made up by Boeings 747 and 727, Douglas DC-10, DC-8 and DC-9, Fokkers F-27 and, soon, the latest Airbuses—is one of the youngest, most modern commercial fleets of our times.

We are 5,452 engineers, technicians, specialists and other professionals working together to keep our fleet in top shape. With jobs that range from the simplest "end of journey" inspection, through the different periodic ones, to what we call the Major Overhaul, which implies the dismantling and total renewal of the plane.

And, since each of those specialists knows his job inside and out, we are proud to say that not even the smallest detail ever escapes us.

Keeping one of the most modern, better maintained fleets in the world is yet another way of showing our passengers that we care.

This is Iberia today. But we want to be better.

IBERIA
INTERNATIONAL AIRLINES OF SPAIN

MORE THAN 50 YEARS MAKING FRIENDS

A Neolithic Flint Mine

The start of farming in prehistoric Europe increased the demand for flint tools. Excavation at one flint-rich valley in the Netherlands shows how miners dug up flints by the ton more than 5,000 years ago

by Peter W. Bosch

The unique form of silica called flint has been a mainstay of mankind for at least half a million years. Harder than steel, it is also brittle and so is easily trimmed into any wanted shape by hammering and flaking. Nodules of flint, some of them as much as a meter in diameter, are particularly abundant in beds of chalk that were deposited on ocean floors around the world during the Upper Cretaceous period, some 80 million years ago. In Europe uplifted beds of Cretaceous chalk lie exposed here and there from France and the south of England eastward to Poland and Czechoslovakia.

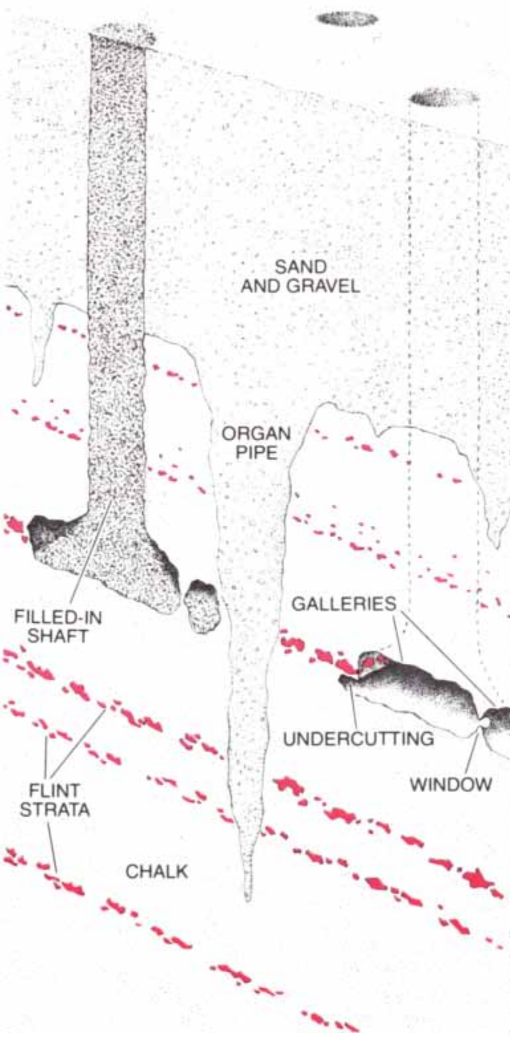
During the Old Stone Age the hunters of Europe favored as a raw material for making choppers, scrapers and other tools loose nodules of flint that had naturally weathered out of the chalk beds. Only a few thousand years ago, with the rise of farming as a way of life, an increasing demand for flint tools turned the Europeans of the New Stone Age from gathering flint nodules on the surface to mining them. Here I shall relate what has recently been learned about one such flint-mining site where Neolithic workers some 5,000 years ago produced more than 150 million tools. The site, near Rijckholt in the Netherlands, was discovered in 1881, was studied on and off thereafter for three-quarters of a century and was finally excavated in detail by a group of amateur archaeologists who did their work one night per week over a nine-year period.

The river Meuse rises in France and on its way to the North Sea flows north through Belgium and then through the southernmost corner of the Netherlands near Maastricht. Traveling by rail to Maastricht from the Belgian frontier town of Visé in 1881, Marcel de Puydt, a prominent Belgian archaeologist, noticed that the land sloping up from the Meuse valley looked much like terrain where he had located prehistoric sites in his own country. The wooded land rose gradually from the valley floor and eventually leveled off into forest-clad tableland. De Puydt decided to reconnoiter the area, and that same year he came on a surface deposit of flint artifacts near Rijckholt.

De Puydt must therefore be considered the discoverer of the Rijckholt site, even though it was not until later that he found its major surface work area. This was a natural basin, filled with flakes of waste flint to a depth of one and a half meters, that was later named the Grand Atelier. In any event, even as others came to investigate the site, de Puydt continued to work at Rijckholt from time to time over a period of nearly 35 years, expanding his collection of flint implements such as axheads, scrapers, projectile points and knife blades, and of miners' tools such as picks fashioned from the antlers of red deer.

The landowner on whose property the Rijckholt mines were situated, René de Geloes, also investigated the site. Count de Geloes began his work in 1886; among his associates was a young medical student, Eugène Dubois, who four years later in Java discovered *Pithecanthropus erectus* (now *Homo erectus*). Other archaeologists from both Belgium and France also worked at Rijckholt, among them Joseph Hamal Nandrin of the University of Liège, who began his excavations in 1903 with the help of local villagers. In 1910 Nandrin came on several ancient mine shafts filled with chalk rubble. Nandrin's discovery provided the first firm evidence that prehistoric flint mines had existed in the Netherlands (although the Grand Atelier at Rijckholt, with its great quantities of flint waste, gave strong evidence that flint nodules had been mined nearby).

It was not until 1923 that any Dutch investigator showed an interest in the Rijckholt flint mines. In that year A. E. van Giffen of the University of Groningen began a systematic study of the site. He was the first to excavate a series of trenches in the Grand Atelier and took care to plot the position of all his finds. The trenching quickly demonstrated that the Neolithic workers had subdivided the hollow into a number of separate areas. For example, axheads had been made in some places, knives in others and scrapers in still others; each specialized flintworking area could be identified from the presence of waste flakes characteristic of a particular production



BLOCK DIAGRAM of the flint-mining zone at Rijckholt, near Maastricht in the southernmost Netherlands, shows at the left the overburden of sand and gravel covering a deep de-

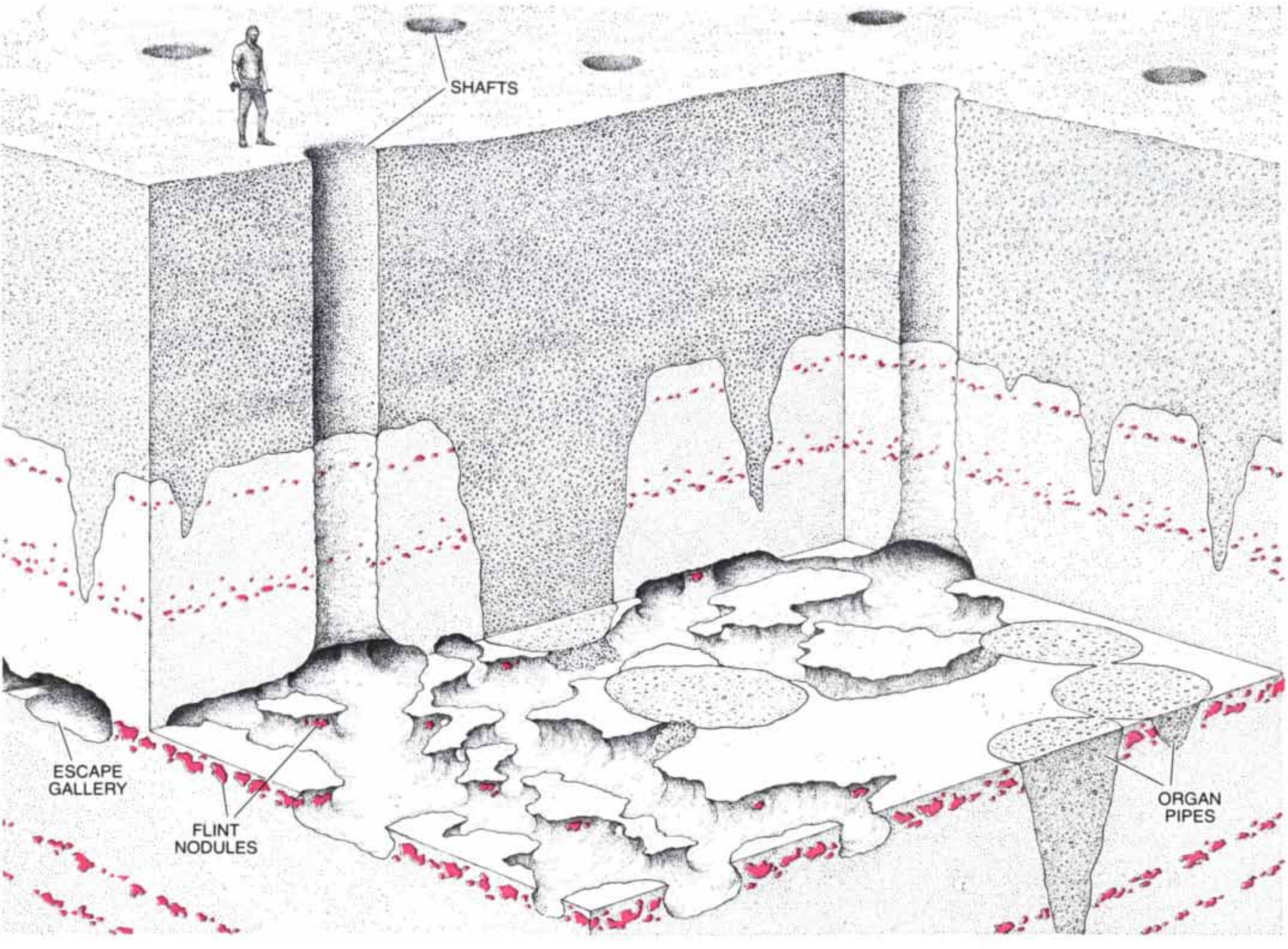
process. Van Giffen also investigated the Rijckholt site in general. Where the shoulder of the tableland sloped down to meet one edge of the Grand Atelier he found traces of mine shafts. On the tableland itself he found potsherds, the remains of hearths and what appeared to be the outlines of temporary huts, suggesting that the miners had camped there while they dug for flint.

Van Giffen completed his fieldwork in 1926, but at the suggestion of certain Belgian archaeologists the monks of the Dominican monastery at Rijckholt continued to investigate the site in their spare time. Between 1929 and 1932 they too found traces of mine shafts on the slope near the Grand Atelier. In addition they came on collapsed mine galleries (the horizontal tunnels that are dug outward from the bottom of a shaft) and collected a quantity of mining tools: more than 100 deer-antler picks and some 1,200 flint axheads. Throughout this time Nandrin came from Liège at intervals, collecting Rijckholt artifacts for his university. He continued to do so until 1953.

No further work was done at Rijckholt during the next decade. In 1964 a Dutch investigator, H. T. Waterbolk of the University of Groningen, came to the site as the head of a group from the university's Biological and Archaeological Institute. Waterbolk's group moved onto the tableland where van Giffen had reported the presence of hut outlines. With a dragline they cleared away a meter of topsoil over a surface area of about 4,000 square meters. This removal of overburden revealed that van Giffen's hut outlines were in fact subsidences of the topsoil over a large number of backfilled mine shafts of the kind first found by Nandrin in 1910. The number of shafts (eventually 66 of them were found within a space of 3,000 square meters) demonstrated that the Rijckholt mines had occupied a much larger area than any earlier workers had supposed.

Waterbolk and his co-workers, like van Giffen and the Dominican monks before them, found evidence of collapsed mine galleries along the slope adjacent to the Grand Atelier. They tried to dig a passage into the Neolithic workings, but they were unsuccessful because none of them had had any mining experience. The investigation came to a halt. It looked as though no one would ever learn what the Rijckholt flint mines had been like.

At this point a professional geologist and enthusiastic amateur archaeologist approached Waterbolk with a novel proposal. The newcomer was Werner M. Felder. Employed by the Netherlands Geological Survey, Felder was an active member of the Limburg chapter of the Dutch Geological Society. He suggested digging a tunnel into the slope above the Grand Atelier in order to follow the stratum of flint-rich chalk the Neolithic miners had exploited by driving shafts down from the tableland above. Such a tunnel would surely intersect many of the horizontal galleries the



posit of Cretaceous chalk that contains seams of flint nodules (*color*). "Organ pipes" are areas where percolating ground water has dissolved the chalk, allowing the overburden to intrude. The vertical section reveals one filled-in Neolithic mine shaft and the horizontal galleries associated with two other shafts. The horizontal section at the right shows two shafts and the lower halves of the galleries associated with them, as mapped by the amateur archaeologists who reopened the Neolithic mines in 1964. The human figure at the top gives scale.

MAJOR FLINT MINES of Neolithic Europe include Grimes Graves and Cissbury in England, Grand Pressigny and Saint-Mihiel in France, Spiennes in Belgium, Hov in Denmark, Krzemionki in Poland and Sumeg in Czechoslovakia. Outline marks area in illustration below.

NETHERLANDS FLINT MINES are situated between Maastricht and Aachen near the center of an extensive chalk deposit (*color*). The fact that prehistoric men had made quantities of flint tools in this area was determined by a Belgian archaeologist, Marcel de Puydt, in 1881.

miners had dug through the chalk to extract the flint nodules. It was Felder's estimate that such an exploratory tunnel need not be more than 140 meters long. He further suggested that the first 10 meters of each prehistoric gallery intersected by the tunnel should also be explored.

Waterbolk endorsed Felder's plan and offered to act as scientific supervisor. The Rijckholt mines, no longer owned by the heirs of Count de Geloes, lie within a local nature reserve; the keeper of the reserve gave his approval for a trial excavation, and 13 members of the Geological Society's Limburg chapter soon formed a body they named the Work Group on Prehistoric Flint Mining. The group included geologists, workers, students, a teacher and, not least, some experienced miners from nearby coal mines. Although the members of the group were from different backgrounds, they shared three interests: archaeology, mining and a desire to run their own excavation. Thus was born the smallest and least official mining company in the Netherlands.

When we started work on June 6, 1964, no one had any idea how big the job would be, how long it would take or how much it would cost. For that matter no one knew what the Neolithic mines would look like when we reached them. The consensus was that the prehistoric galleries would be high enough to allow us to walk upright; some optimistic members of the group even came equipped with what they would need for making maps. Things turned out differently.

Our first step was to build a weatherproof shed at the entrance point on the slope. This proved to be a wise precaution because the weather was bad when we began the actual digging, with pickaxes and shovels, to clear a tunnel mouth about two meters wide and one and a half meters high. As unpaid volunteers we could devote only weekends to the digging. Before the tunnel was more than a few meters long, however, we had encountered the first Neolithic mine galleries, and after some weekends of work we had reached the first shaft.

Both the galleries and the shaft were filled with chalk rubble; in addition the galleries were partially collapsed. As we removed the rubble we also brought in timbers to shore up the gallery roofs and shaft entrances. While we were clearing the galleries we found our first flint axheads. During removal of the rubble fill from the first shaft we also found enough charcoal, apparently firewood fallen from a prehistoric hearth at the top of the shaft, to send off for a carbon-14 age determination.

We awaited the results anxiously. The Rijckholt site had been under investigation for 80 years but no one knew exactly how old the mines were. The reading that came back gave the date 3150 (±60 years) B.C. Evidently flint mining had begun in the southern Netherlands somewhat more than 5,000 years ago.

The trial excavation had carried our tunnel 15 meters into the chalk. The results were encouraging enough to warrant a full-scale project, although still an amateur one, and so official government permission was needed. It was forthcoming, conditional on the use of steel supports rather than timber ones. The Netherlands Foundation for Scientific Research provided a subsidy that helped us to meet the added expense.

Now that the project was official some formal rules had to be made. The trial-excavation work groups had come to the site on Saturdays, but most of us were married men who wanted to be with our families on the weekend. It was decided to push the working time forward: from 7:00 P.M. on Friday nights until 3:00 A.M. Saturday mornings. It was also agreed that each member of the group would work a minimum of 20 nights per year. Eventually 20 such volunteers formed the basic cadre of the Limburg chapter's mining group, so that separate 10-man teams did the night work at Rijckholt for 40 of the 52 weeks in the year. Even when Friday fell on Christmas Eve or New Year's Eve most of the volunteers turned up for work.

Werner Felder and his brother Joseph were respectively chosen as leader and as technical director of the project. Both had worked in the Dutch coal mines and therefore had practical mining experience. Other work-group members undertook the task of corresponding with the various institutions and museums that had taken an interest in the project.

In brief, our progress was as follows:

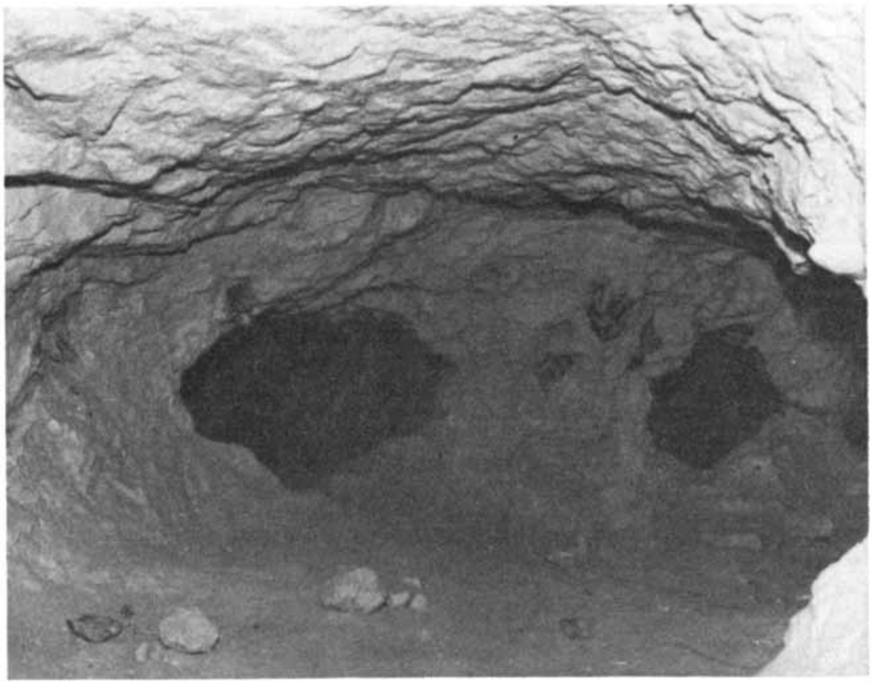
In 1965, 50 tons of steel were brought to the site by hand to substitute for the wood supports that had been put in place during the trial excavation. Rails were laid so that waste chalk could be moved out of the mine in small carts rather than in wheelbarrows. By the end of the year the tunnel had reached a length of 18 meters.

In 1966 one of the companies interested in the project lent us an air compressor and pneumatic hammers so that we could make better headway against the increasingly hard chalk. During the preceding year our only source of electric illumination had been a 12-volt battery; now we were able to borrow a small generator and set up a more extensive illumination system. As the work progressed it became necessary to split up our Friday-night groups into smaller teams; following coal-mine procedure, each team was directed by a leader who was responsible for safety and who also recorded all finds. By the end of the year the tunnel was 38 meters long.

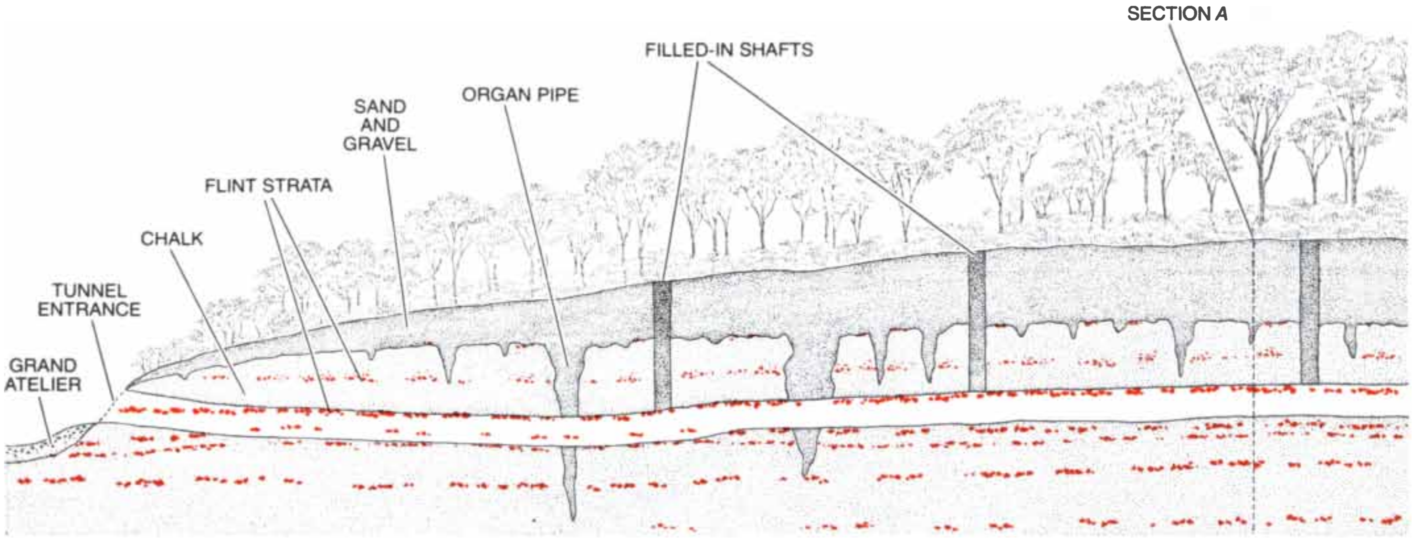
In 1967 the transport of the waste chalk by mine cart became increasingly difficult. The Dutch coal-mining industry, which had shown much interest in our project, lent us a conveyor with a total length of 150 meters. Because no power line extended to the site we bought a tractor and connected one rear wheel to the conveyor drive shaft with a belt. Money subsidies and gifts of equipment now provided each of us with a set of warm mining clothes, a safety helmet and a battery lamp. Every member of the group was also insured against accidents, since the work was not without danger. By the end of the year the tunnel was 67 meters long.

In 1968 the team system was contin-

MINE GALLERY, still not completely cleared of the waste chalk stacked there by Neolithic miners, is about 60 centimeters high. Black inclusions in the chalk wall (*left*) are flint nodules.



TWO "WINDOWS" appear in the walls of another mine gallery. Too small to allow passage from one gallery to another, they were probably accidental. Gallery is 80 centimeters high.



EXPLORATORY TUNNEL, excavated by weekend volunteers between June, 1964, and December, 1972, is illustrated in elevation and in transverse section (*far right*); again the human figure gives scale. Extending for 150 meters from the tunnel entrance, adjacent to the Grand Atelier, the tunnel followed the 30-centimeter layer of chalk that the Neolithic miners had found to be particularly rich in nod-

ued. One team worked to extend the tunnel. A second team saw to supplies, maintenance of the equipment and clearance of the site. A third team and sometimes a fourth cleared the side galleries of rubble, mapping the new areas and recording the artifacts and charcoal they discovered. By the end of the year the tunnel was 108 meters long.

In 1969, in consultation with Waterbolk, the group decided to extend the tunnel the full length of the Neolithic mining area. This called for the excavation of an additional 40 meters, and all the members of the group were enthusiastic at the prospect. Some of the newly discovered galleries were left unexamined as the work teams concentrated on digging the tunnel. Nevertheless, before the end of the year nearly three-fourths of all the mine galleries had been cleared of rubble to a distance of 10 meters. Some 10,000 flint implements had now been recovered. At the end of the year the tunnel was about 110 meters long.

In 1970, 1971 and 1972 work continued both at the excavation and at home. The homework included fund raising (at this time the government had begun to close down many of the coal mines that were supporting our work, thereby ending their assistance), correspondence, the preparation of site maps and the cataloguing of finds. The few meters of tunnel still to be excavated presented us with an old problem on a new scale.

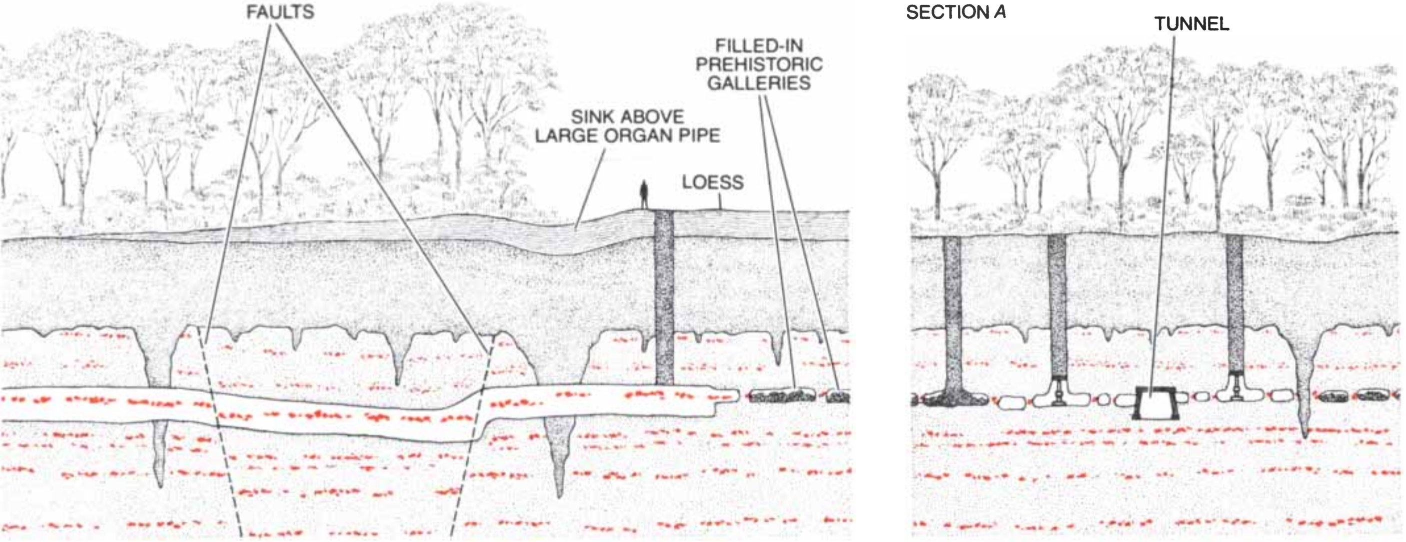
"Organ pipe" is the name given to a sinkhole formed by the percolation of ground water through the chalk. The organ pipes vary in diameter from a meter to more than 10 meters; the space that had been occupied by the chalk is filled with sand and gravel. If a digger happens to break into an organ pipe while he is tunneling, there is great danger that the loose fill will pour into the tunnel, blocking the passage and possibly burying him. Organ pipes were present throughout the Rijckholt area, but we had fortunately encountered only small ones. Now, almost at the conclusion of our work, we came on perilously large ones. Nevertheless, before the end of 1972 the tunnel had reached the planned length of 150 meters.

In 1973, having come to the end of the mine area, we were able to spend more time clearing out the remaining galleries. Even these activities had been completed by the end of the year. Joseph Felder estimated that we had opened an area of mine floor totaling 3,000 square meters. All that remained for us to do was to remove our machinery and equipment and restore the surface of the site to its original condition. The staff of the nature reserve had helped in this work over the years by carting away the waste chalk as we removed it from the mine galleries. After nine years of weekend labor underground we could now turn to working up the final data.

What had the Rijckholt mines been like in Neolithic times? We know from the discovery of two other datable charcoal samples that flint mining was in progress here somewhat more than 5,000 years ago. (The two additional dates, 3120 and 3050 B.C., are in good agreement with our initial carbon-14 finding.) We do not know how long mining continued or how many miners were at work. Joseph Felder has nonetheless been able to compile some overall statistics for the operation. First, the Rijckholt mining area covers some 25 hectares (62 acres). In the course of exploring only 3,000 square meters of the 25-hectare area we had encountered 66 mine shafts. If our sample was a representative one, the entire Rijckholt area must have had some 5,000 mines. (Each shaft, together with its cluster of galleries, is counted as a single mine.) It is Felder's estimate that the 5,000 Rijckholt mines together yielded 41,250 cubic meters of flint nodules. That would be enough raw material for 153.3 million axheads. If one supposes the miners worked at Rijckholt for 500 years, the average production rate would have been 1,500 axheads per day, assuming that both the miners and the flint shapers at the Grand Atelier were active year round and not just seasonally.

The length of the shafts leading from the tableland down to the flint-bearing chalk stratum varied from 10 to 16 meters, depending on the thickness of the sand, gravel and loess overlying the chalk. The shafts ranged in diameter from a meter to a meter and a half. While digging down through the loose overburden, the miners evidently prevented cave-ins with a retaining wall of plaited tree branches: hollows resembling the imprint of branches were visible in the wall of the single shaft we cleared all the way to the surface. (We ourselves guarded against a cave-in by installing successive cylindrical steel sections as we went along.)

On the average the galleries the Rijckholt miners excavated at the bottom of each shaft had a total surface area of 25 square meters. The height of a gallery was about 60 centimeters (two feet); its width varied depending on the location of the flint nodules. The extraction method duplicated the modern mining



ules of flint. The excavators reached the first Neolithic mine galleries when the tunnel was only a few meters long; at the end of nine years they had located 66 mine shafts, each with radiating galleries that had an average surface area of 25 meters. It is estimated that overall the Rijckholt flint-mining area could have yielded 41,250 cubic meters of flint, which is enough to make 153.3 million axheads.

method known as the room-and-pillar system. The flint-bearing layer of chalk, about 30 centimeters thick, was undermined and removed a block at a time; at intervals the chalk was left untouched to provide pillars that helped to support the gallery roof. To supplement these natural pillars the miners also backfilled the undermined area with waste chalk. Once a gallery was mined out it served as a dump for the waste chalk being excavated from the next gallery; 95 percent of the galleries we discovered had been refilled in this way.

How did the miners hew their way through these uncountable tons of chalk? They worked with axheads their flint-shaping colleagues manufactured at the Grand Atelier overhead. Judging by certain hollow spaces we found in the piles of waste chalk, the axheads had been mounted on wood handles some 80 centimeters long. The heads were remarkably uniform: between 15 and 18 centimeters long, five centimeters wide and 3.5 centimeters thick.

Mining the chalk was hard work that called for considerable skill; we estimate that a miner would wear out five axheads in the course of removing one cubic meter of chalk. As the axheads were blunted the miners sharpened them on the spot, using chalk hammerstones to retouch the cutting edge. A particularly hard type of recrystallized chalk was selected for the purpose. We found most of the hammerstones in association with groups of 10 or 20 axheads at what we came to call ax depots. These resharpening sites were generally in the vicinity of a shaft and were marked by numerous flakes of waste flint.

We found more than 15,000 artifacts in the mines. Most of them were either blunted or broken axheads. On the basis of Joseph Felder's extrapolations this suggests that 2.5 million additional axheads, abandoned by the Rijckholt miners, remain in the unexplored galleries. If that is true, the miners reinvested less than 2 percent of their total flint production in the extraction effort, the cost per mine being about 350 axheads. We also found a few antler picks resembling those de Puydt had found on the surface at Rijckholt during his 19th-century investigations. The fact that we did not find many suggests they were seldom used in excavating galleries because the chalk there was harder than antler.

It appears that the Rijckholt miners worked more than one mine at a time. This conjecture is given weight by our discovery of certain "escape galleries." These passages were backfilled with waste chalk only partly or not at all. The galleries connected two adjacent mine shafts; some connected more than two. Making the connections did not require much additional work by the miners because the shafts themselves were dug quite close to one another, probably in order to limit the distance the flint nodules had to be hauled from the galleries to the shafts. Apparently the last gallery of each mine was extended to reach the next mine shaft. The regular use of the escape galleries is evident from the compaction of the chalk rubble on their floors and the scars and scratches on their walls and roofs caused by the frequent passage of crawling miners.

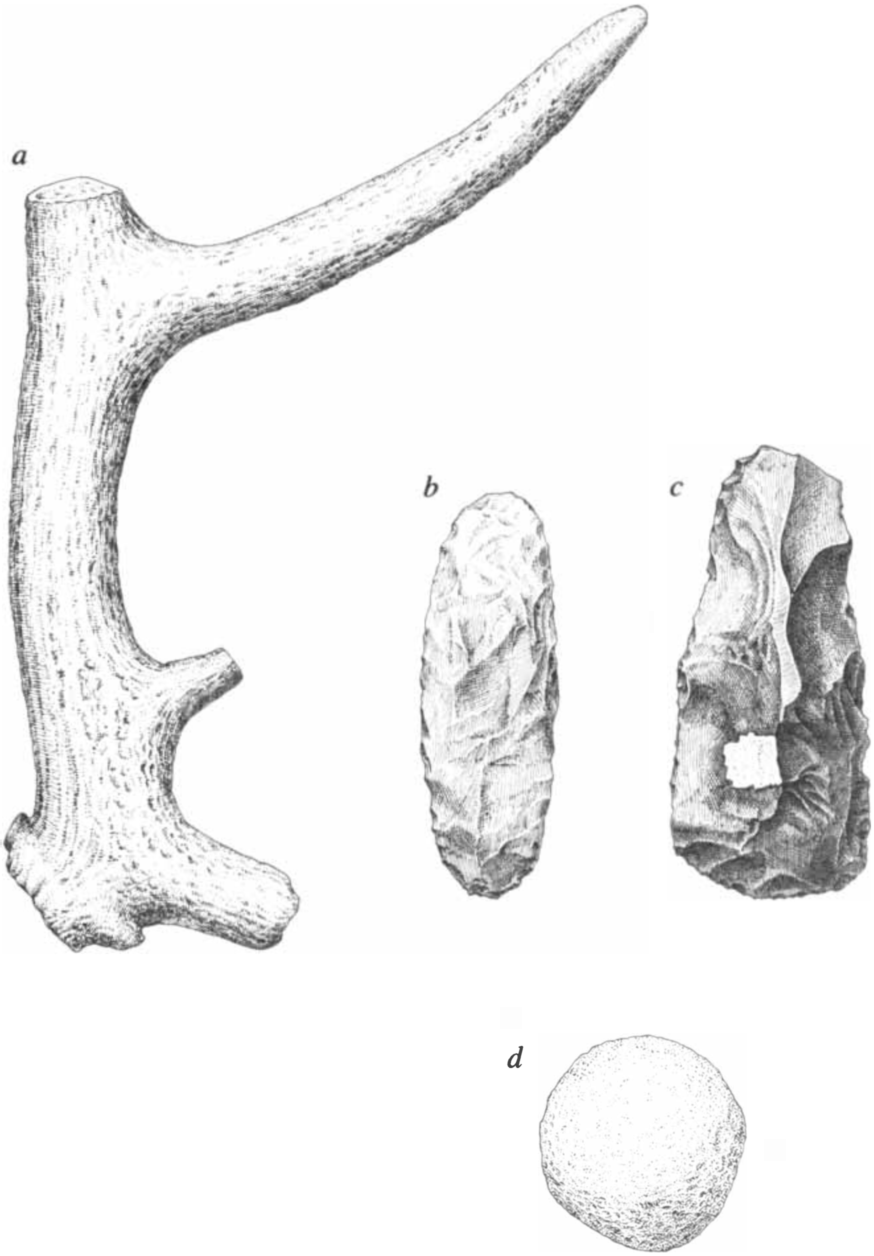
How the miners descended from the surface to the bottom of their shaft and ascended again is not known for certain. Neither is the method of raising the flint nodules to the surface. It seems reasonable to conjecture that ropes served both purposes, but the only evidence for it is the presence of deep grooves worn into the chalk where the shafts end and the galleries open inward.

We found no evidence of lamps or of torches underground; such flames would probably have depleted the oxygen the miners needed more than they did illumination. The ax depots were near the bottom of the shafts, perhaps because the miners needed light to sharpen their axheads. The miners working in the galleries, however, apparently made do with whatever indirect illumination filtered through to them from above.

Cryptic evidence that the Neolithic miners engaged in cult practices was encountered during our first year of full-scale tunneling. On November 5, 1965, we found at the end of a newly opened gallery a shallow pit containing a human skull without a lower jaw. Except for the area where the skull was found the gallery had been completely backfilled with chalk rubble. Obviously the skull and only the skull had been placed in the pit, possibly as a gesture of secondary burial. The skull was examined by G. van Vark of the Laboratory for Anatomy and Embryology at the University of Groningen; he identifies it as that of a male about 40 years old. The skull exhibits injuries at the left temple and above the left eye, but the fact that the bone had regenerated indicates that the injuries were not the cause of death. Exactly what practice was responsible for the presence of the skull in the mine is not easy to imagine. One can only guess that some form of ancestor worship provided the motive.

In 1975 the second in a series of international Flint Symposia was held at Maastricht, and a visit to Rijckholt was part of the agenda. By then the excavation had been deserted for three years, the steel tunnel supports were rusting and the galleries also showed signs of deterioration. The symposium participants forwarded a resolution to the Netherlands Ministry of Culture pointing out that the Rijckholt mines were a prehistoric monument of the greatest importance and asking government action for their preservation. Both the government and several private enterprises came forward with offers of support. Beginning that same year a number of steps were taken to preserve the excavated area.

Today the tunnel floor has been dug deeper so that visitors can walk upright for its full length, and the tunnel walls and roof have been waterproofed. A small shelter for the reception of visitors stands at the tunnel entrance and electric lights illuminate both the tunnel and the galleries. Tourists and scholars alike can now come to see how their ancestors mined the flint needed for their tools 5,000 years ago. In retrospect it seems remarkable that 20 amateur archaeologists made this possible by their work on Friday nights.



MINER'S TOOLS found in the Rijckholt galleries included antler picks (*a*), axheads (*b, c*) and hammerstones for resharpening blunted axheads (*d*). Some hammerstones were made from quartz and some from extra-hard chalk; the one seen here is chalk. It is estimated that five axheads were worn out per cubic meter of chalk removed, making the cost per mine 350 axheads.

How to Profit from The Revenue Act of 1978

The Ultimate Tax Shelter

by
TED NICHOLAS

Tax experts are now referring to a small, privately owned corporation as "The Ultimate Tax Shelter." This is especially true with the passage of the Revenue Act of 1978. This law makes most former tax shelters either obsolete, or of little advantage. Investments affected include real estate, oil and gas drilling, cattle feeding, movies, etc. These former tax shelters have lost their attractiveness. Aside from that, these tax shelters required a large investment. Only a small segment of the population could benefit from them.

I've written a book showing how you can form your own corporation. I've taken all the mystery out of it. Thousands of people have already used the system for incorporation described in the book. I'll describe how you may obtain it without risk and with a valuable free bonus.

A corporation can be formed by anyone at surprisingly low cost. And the government encourages people to incorporate, which is a little known fact. The government has recognized the important role of small business in our country. Through favorable legislation incorporating a small business, hobby, or sideline is perfectly legal and ethical. There are numerous tax laws favorable to corporate owners. Some of them are remarkable in this age of ever-increasing taxation. Everyone of us needs all the tax shelter we can get!

Here are just a few of the advantages of having my book on incorporating. You can limit your personal liability. All that is at stake is the money you have invested. This amount can be zero to a few hundred or even a few thousand dollars. Your home, furniture, car, savings, or other possessions are not at risk. You can raise capital and still keep control of your business. You can put aside up to 25% of your income tax free. If you desire, you may wish to set up a non-profit corporation or operate a corporation anonymously. You will save from $300 to $1,000 simply by using the handy tear-out forms included in the book. All the things you need: certificate of incorporation, minutes, by-laws, etc., including complete instructions.

There are still other advantages. Your own corporation enables you to more easily maintain continuity and facilitate transfer of ownership. Tax free fringe benefits can be arranged. You can set up your health and life insurance and other programs for you and your family wherein they are tax deductible. Another very important option available to you through incorporation is a medical reimbursement plan (MRP). Under an MRP, all medical, dental, pharmaceutical expenses for you and your family can become tax deductible to the corporation. An unincorporated person must exclude the first 3% of family's medical expenses from a personal tax return. For an individual earning $20,000 the first $600 are not deductible.

Retirement plans, and pension and profit-sharing arrangements can be set up for you with far greater benefits than those available to self-employed individuals.

A word of caution. Incorporating may not be for you right now. However, my book will help you decide whether or not a corporation is for you now or in the future. I review all the advantages and disadvantages in depth. This choice is yours after learning all the options. If you do decide to incorporate, it can be done by mail quickly and within 48 hours. You never have to leave the privacy of your home.

I'll also reveal to you some startling facts. Why lawyers often charge substantial fees for incorporating when often they prefer not to, and why two-thirds of the New York and American Stock Exchange companies incorporate in Delaware.

You may wonder how others have successfully used the book. Not only a small unincorporated business, but enjoyable hobbies, part time businesses, and even existing jobs have been set up as full fledged corporations. You don't have to have a big business going to benefit. In fact, not many people realize some very important facts. There are 30,000 new businesses formed in the U.S. each and every month. 98% of them are small businesses; often just one individual working from home.

To gain all the advantages of incorporating, it doesn't matter where you live, your age, race, or sex. All that counts is your ideas. If you are looking for some new ideas, I believe my book will stimulate you in that area. I do know many small businessmen, housewives, hobbyists, engineers, and lawyers who have acted on the suggestions in my book. A woman who was my former secretary is incorporated. She is now grossing over $30,000 working from her home by providing a secretarial service to me and other local businesses. She works her own hours and has all the corporate advantages.

I briefly mentioned that you can start with no capital whatsoever. I know it can be done, since I have formed 18 companies of my own, and I began each one of them with nothing. Beginning at age 22, I incorporated my first company which was a candy manufacturing concern. Without credit or experience, I raised $96,000. From that starting point grew a chain of 30 stores. I'm proud of the fact that at age 29 I was selected by a group of businessmen as one of the outstanding businessmen in the nation. As a result of this award, I received an invitation to personally meet with the President of the United States.

I wrote my book, How To Form Your Own Corporation Without A Lawyer For Under $50, because I felt that many more people than otherwise would could become the President of their own corporations. As it has turned out, a very high proportion of all the corporations formed in America each month, at the present time are using my book to incorporate.

Just picture yourself in the position of President of your own corporation. My book gives you all the information you need to make your decision. Let me help you make your business dreams come true.

As a bonus for ordering my book now, I'll send you absolutely free a portfolio of valuable information. It's called "The Income Plan" and normally sells for $9.95. It describes a unique plan that shows you how to convert most any job into your own corporation. You'll increase your take-home pay by up to 25% without an increase in salary or even changing jobs in many cases. If you are an employer, learn how to operate your business with independent contractors rather than employees. This means that you'll have no payroll records or withholding taxes to worry about. And you'll be complying with all I.R.S. guidelines. "The Income Plan" includes forms, examples and sample letter agreements to make it possible.

I personally guarantee your purchase in the fairest way I know. If you feel my book is not all that I've described, return it undamaged within two weeks and I'll promptly refund your money with no questions asked. If you should decide to return it, you may keep the bonus Income Portfolio for your trouble.

To get your copy, write the words "Corporation Book and Bonus" on a plain sheet of paper, along with your name and address. Enclose a check for $14.95. There is no sales tax on your order and your purchase price is tax deductible. Mail your order to me, Ted Nicholas, c/o Enterprise Publishing Co., Dept. SA-96C, 501 Beneficial Building Wilmington, Delaware 19801.

Ultracold Neutrons

Neutrons moving so slowly that they cannot penetrate a solid surface can be confined in a metal bottle. By storing them for long intervals it may be possible to measure fundamental properties of the neutron

by R. Golub, W. Mampe, J. M. Pendlebury and P. Ageron

It often seems that progress in understanding the elementary particles of matter comes from a single advancing frontier of ever higher energy. Even at low energy, however, unfamiliar territory still waits to be explored. Indeed, John G. King of the Massachusetts Institute of Technology has remarked that there is as much to be learned between the lowest energy yet reached and zero energy as there is between the highest energy attained and infinite energy.

One outpost on the low-energy frontier is represented by neutrons with an energy of 10^{-7} electron volt. One electron volt is the energy acquired by an electron when it is accelerated through a potential difference of one volt, and so the neutrons have the energy that would be imparted to an electron by a potential difference of one ten-millionth of a volt, or .1 microvolt. That is the average energy of a particle in a gas whose temperature is one millidegree Kelvin, or one-thousandth of a degree Celsius above absolute zero. Neutrons in this energy range are similar to such a gas, and they are called ultracold neutrons. It is not easy to appreciate how small their energy is. Whereas high-energy particles have speeds near the speed of light, ultracold neutrons move at only a few meters per second; a human runner could keep up with them. Moreover, they can rise only about a meter against the earth's gravitational field before they stop and fall back.

A striking property of ultracold neutrons is their total reflection from solid surfaces. At higher energy neutrons readily penetrate matter, but when they are moving slowly enough, they bounce away from most material surfaces. As a result they can be stored in a bottle made of ordinary materials such as metal or glass. They can also be confined by a magnetic field. Most other subatomic particles must be studied on the fly, but ultracold neutrons almost stand still for examination.

The ability to confine ultracold neutrons for long periods suggests two of their applications. One is the measurement of the neutron lifetime, which is now known to an accuracy of only about 1.5 percent. Another is the search for an electric dipole moment of the neutron, a quantity that is of great theoretical interest. Ultracold neutrons can also serve as an instrument for the investigation of other physical systems, such as thin films and solid surfaces.

Neutrons bound in the nuclei of atoms make up more than half of the mass of the earth, but free neutrons are rare in nature. The reason is that outside the nucleus the neutron is an unstable particle: through the process of beta decay it breaks down into a proton, an electron and the massless particle called a neutrino. The average lifetime of a neutron is a little more than 15 minutes.

Free neutrons for experimentation are obtained in greatest abundance from a nuclear reactor, where they are released as a by-product of nuclear fission. The neutrons are emitted with energies of several million electron volts, but they soon give up a major share of their energy as a result of collisions with atomic nuclei in the material surrounding the reactor core, which is called the moderator. After only a few collisions the average neutron energy is comparable to the energy associated with the thermal motions of the moderator nuclei. The neutrons are then in thermal equilibrium with the nuclei, that is, they can be regarded as a gas with the same temperature as the moderator. If the moderator is at room temperature (about 300 degrees K.), the neutron gas has the same temperature and the average neutron energy is about 25 millielectron volts.

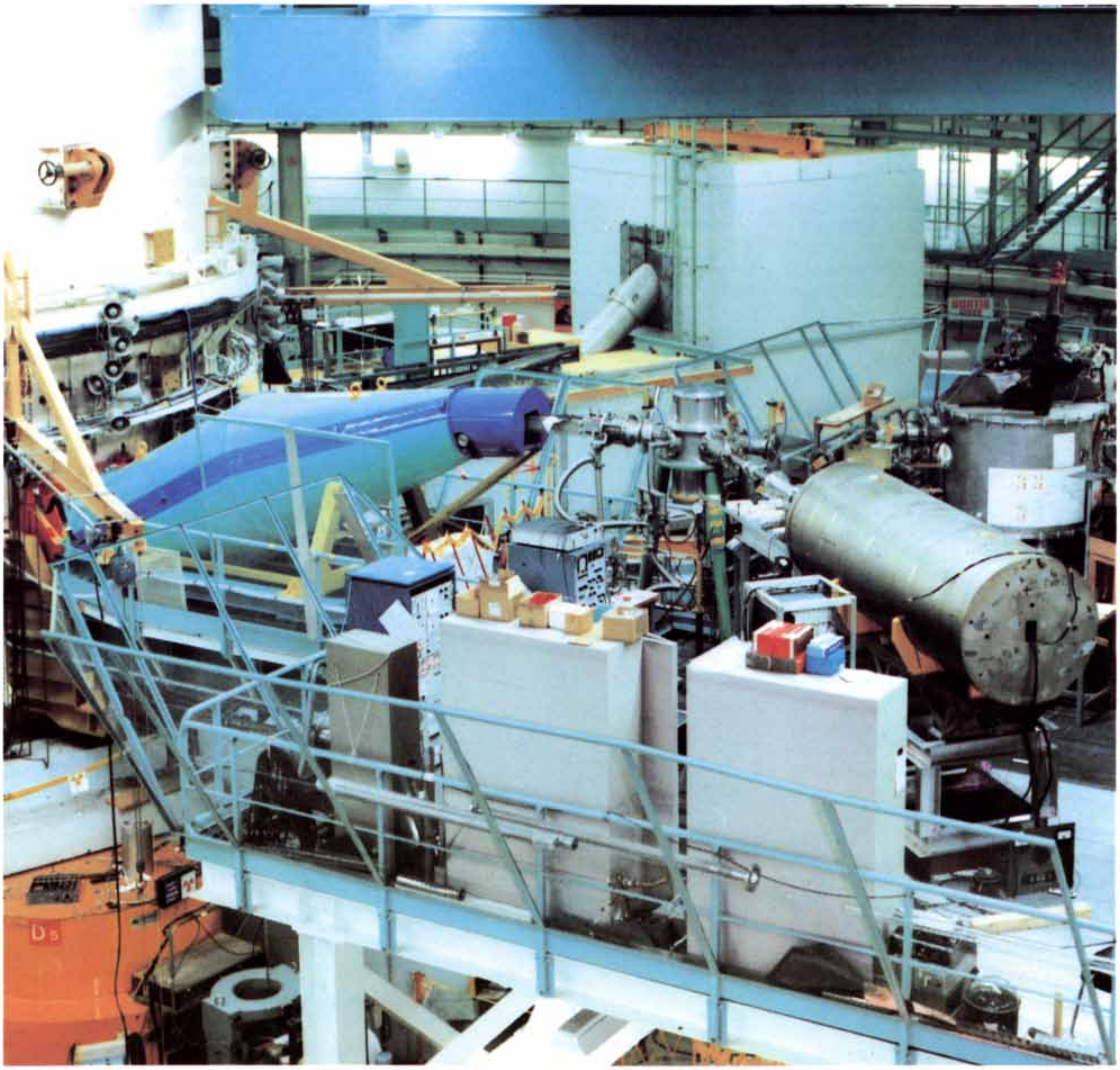
The temperature of the neutron gas defines not only the average energy but also the range over which the energies are distributed. At room temperature there are neutrons with energies substantially higher and lower than 25 millielectron volts, although the number present declines as one departs further from the average energy. At the temperature of liquid hydrogen (20 degrees K.) the average energy is smaller and so is the range of energies.

Because the energies are distributed over a range comparable to the average energy, neutrons of arbitrarily low energy can be found even at room temperature, although they make up a very small proportion of the total flux. By convention neutrons near the peak in the room-temperature distribution, with an energy of about 25 millielectron volts, are called thermal neutrons. Those with progressively lower energies are called cold, very cold and ultracold neutrons. In general the boundaries between these categories are somewhat vague, but there is an unambiguous criterion for defining ultracold neutrons: the property of total reflection.

The total reflection of neutrons from a solid surface is analogous to the total internal reflection of light inside a glass prism, and it can best be understood by regarding the neutrons as matter waves. In quantum mechanics any particle can be represented as a wave whose length is inversely proportional to the square root of the particle's energy. For a thermal neutron with an energy of 25 millielectron volts the wavelength is 1.8 angstrom units, one angstrom being equal to 10^{-8} centimeter. That distance is comparable to the spacing between the atoms in a solid. For an ultracold neutron with an energy of 10^{-7} electron volt the wavelength is 900 angstroms. To the shorter wave a solid is a lattice of atomic nuclei with empty spaces between them, which the wave readily penetrates. To the longer wave the solid is an essentially continuous barrier, and the amplitude of the wave declines steeply inside the surface, falling to zero within about 100 angstroms. The probability of finding the neutron at a given place is related to the square of the amplitude of the wave at that place, and so the neutron is unlikely to be found in the solid. It is repelled by the surface as a whole. (This effect was predicted and demonstrated in 1945 by Enrico Fermi.) The neutron can overcome the repulsion and penetrate the material only if it has a velocity component normal to the surface that exceeds some critical value.

The critical velocity a neutron needs in order to penetrate a substance is a property of the substance. For nickel, iron, beryllium and copper it is about six meters per second, but for most other materials it is substantially lower. For a copper surface a neutron moving at 600 meters per second is certain to be reflected only if it strikes the surface at an angle of less than about half a degree, so that the velocity normal to the surface is less than the critical value. Grazing reflections of this kind are employed in mirrors and in guide tubes that can transport neutron beams with little loss of intensity.

Ultracold neutrons are those whose total velocity is less than the critical velocity for a given material. They are therefore certain to be reflected from a surface no matter what the angle of incidence is, and they can be confined simply by placing them in a sealed bottle whose walls are made of an appropriate material. For definiteness the critical velocity for total reflection can be set at the six meters per second characteristic of copper. A neutron moving at six meters per second has an energy of 2×10^{-7} electron volt, which corresponds to a temperature of two millidegrees K.



SOURCE OF ULTRACOLD NEUTRONS is a nuclear reactor at the Institut Max von Laue–Paul Langevin at Grenoble in France. The reactor itself is behind the curved wall at the far left. The neutrons emerge from an inclined guide tube, which is thickly swathed in shielding. From the guide tube the neutrons pass through a distribution valve (mounted on the green tripod) to the experimental apparatus. When the photograph was made, two experiments were being prepared. The cylinder with a vertical axis houses a magnetic storage ring for ultracold neutrons, built by Wolfgang Paul and Uwe Trinks of the University of Bonn. It will be used to measure the neutron lifetime. The longer, horizontal cylinder is part of an experiment being done by Norman F. Ramsey of Harvard University, Kenneth F. Smith of the University of Sussex, the authors and others. The aim of the experiment is to search for an electric dipole moment of the neutron.

Although a straightforward calculation made from the thermal-equilibrium curve suggests that ultracold neutrons should be present in a nuclear reactor at room temperature, they are exceedingly rare. At 300 degrees K. ultracold neutrons make up only one part in 10^{11} of the total neutron flux. The first serious proposal for extracting them from a reactor and isolating them was made in 1959 by Yakov B. Zel'dovich of the Academy of Sciences of the U.S.S.R. The first successful experiments with ultracold neutrons were carried out almost simultaneously in 1968 by workers at the Joint Institute for Nuclear Research at Dubna in the U.S.S.R. and by a group at the Technical University in Munich.

At Dubna the late F. L. Shapiro and his colleagues extracted the neutrons from a low-power reactor and went on to demonstrate that they could be stored for many seconds. A sinuous, horizontal guide tube conducted the neutrons from the vicinity of the reactor core through the surrounding shielding to the experimental apparatus. Because the tube made several turns few higher-energy neutrons could travel its full length; instead, on striking the wall of the tube they passed through it and were absorbed in the shielding. Ultracold neutrons, on the other hand, being totally reflected, bounced from wall to wall and had a reasonably high probability of reaching the end of the tube.

An immediate problem with this scheme is that the guide tube must be evacuated, and so it must be closed off at the end inside the reactor. The metal window can be very thin, but even so the definitive characteristic of total reflection excludes from the tube exactly those neutrons that are being sought: the ultracold ones. Shapiro's group overcame this difficulty by installing a material called a converter inside the window. Some of the neutrons moving fast enough to pass through the window gave up a part of their energy to nuclei in the converter and were thereby transformed to the ultracold-energy range. The best converters are generally materials rich in hydrogen, such as polyethylene or water.

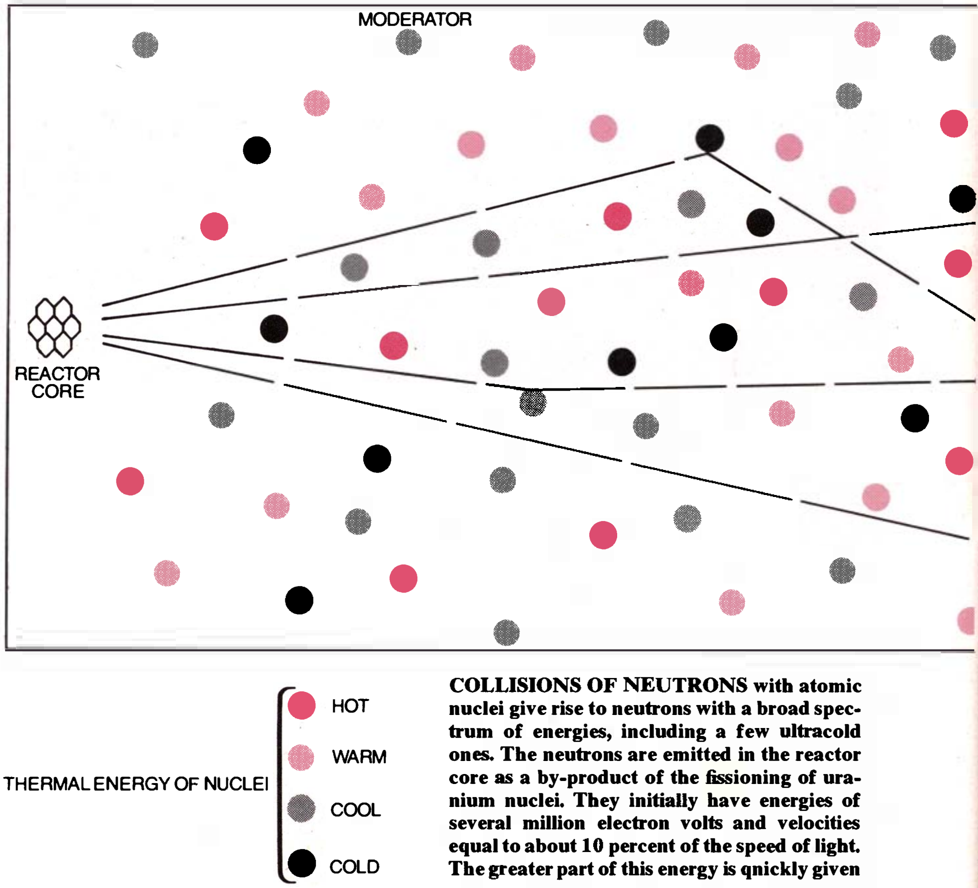
In Munich, Albert Steyerl obtained ultracold neutrons by a different method: he selected neutrons moving at speeds somewhat greater than the critical velocity (so that they can penetrate the entrance window) and employed the earth's gravitational field to decelerate them. This was accomplished by transporting the neutrons through a vertical guide tube some 11 meters long. Neutrons that happen to reach ultracold energies in the reactor cannot rise this far and so cannot reach the apparatus at the top of the tube, but slightly faster neutrons, which can just make the ascent, give up most of their energy on the way and emerge at ultracold speeds.

More recently Steyerl has devised another way to decelerate neutrons; it consists in passing them through a "turbine" that absorbs part of their energy. The turbine is a wheel 1.7 meters in diameter with some 660 curved reflecting surfaces fixed to its circumference. The wheel is spun with a tangential velocity of 25 meters per second, and neutrons moving at about 50 meters per second are directed so that they strike the receding blades at a glancing angle. Each neutron can be reflected several times from the concave surface of the blade, and with each bounce it gives up energy to drive the rotation of the turbine, just as falling water or expanding gases would. The neutrons are of course too feeble to actually turn the turbine; the energy needed is supplied by a motor.

The total reflection of ultracold neutrons is actually total only in the sense that the particles cannot be transmitted through a solid. When a neutron strikes a surface, there are three possible outcomes. One is elastic scattering, in which the neutron is reflected from the surface without any change in its energy or velocity, or in other words without exchanging any energy with the nuclei in the solid. For ultracold neutrons elastic scattering is by far the most probable result of a collision, and if it were the only possible result, the neutrons could be confined indefinitely (or at least until they decayed).

Another possibility is inelastic scattering, in which there is an exchange of energy between the neutron and a nucleus. As a result of thermal excitations the nuclei in a solid vibrate about their nominal positions. Because the neutron wave penetrates a short distance into the material, it overlaps some of the nuclei and can therefore interact with them. If the neutron happens to interact with a nucleus that is receding, the neutron may lose energy; if the nucleus is approaching at the moment of collision, the neutron's energy is augmented. In either case the neutron is still reflected from the surface but with its velocity changed. An increase in energy is much more likely than a loss, and a neutron that gains energy almost invariably exceeds the critical velocity. Thus when it next strikes the wall of the confining vessel, it is usually transmitted rather than reflected and is thereby permanently lost to the confined population.

The third possible fate of a neutron is absorption by a nucleus in the wall. For



COLLISIONS OF NEUTRONS with atomic nuclei give rise to neutrons with a broad spectrum of energies, including a few ultracold ones. The neutrons are emitted in the reactor core as a by-product of the fissioning of uranium nuclei. They initially have energies of several million electron volts and velocities equal to about 10 percent of the speed of light. The greater part of this energy is qnickly given

example, a nucleus of carbon that has six protons and six neutrons, for an atomic mass of 12, can absorb an additional neutron to become a nucleus of the isotope carbon 13. As with inelastic scattering, absorption permanently removes a neutron from the population.

The probability that a neutron will be lost by absorption or by inelastic scattering depends on the material of the walls. Inelastic scattering is more likely with lighter nuclei because at a given temperature they move faster than heavy ones. The probability of absorption depends in a complicated way on the composition of the nucleus, but the probability has been measured for all the common nuclei. From the properties of the wall material one can calculate the average number of bounces a neutron can make before it will be lost by one of these mechanisms. For copper the calculated number is about 6,000; for beryllium it is 200,000.

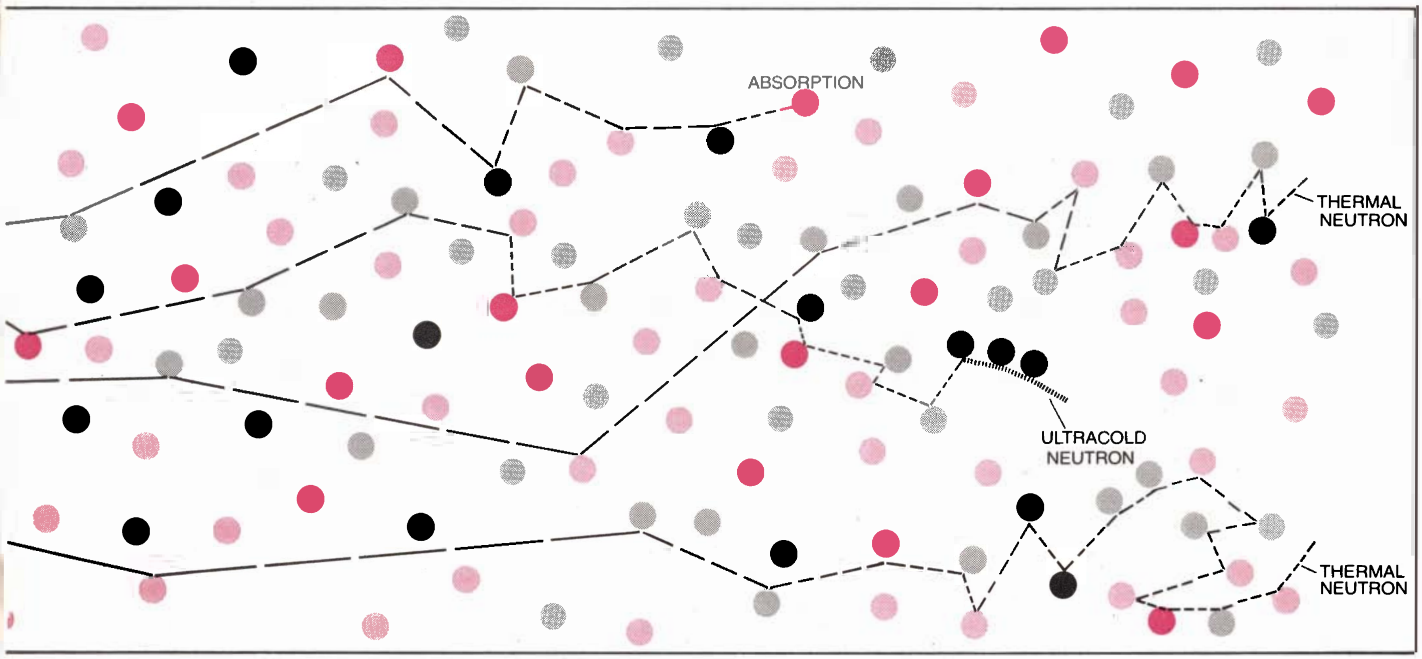
For neutrons in a closed vessel the expected storage time is directly proportional to the average number of bounces before loss. (The storage time also depends on the size and shape of the container and on the velocity of the neutrons, since these factors determine the time between collisions with the walls.) A troublesome finding is that the measured storage times are significantly smaller than had been expected. A long series of experiments done at several institutions with varied techniques and apparatus have all confirmed this anomaly. The results of the experiments can be summarized simply: in spite of large differences in the theoretical probability of loss, most materials yield about the same average number of bounces before loss, roughly 1,000. This discrepancy was discovered 10 years ago, and it is still not fully understood.

The leading hypothesis attributes the excessive loss of neutrons to impurities on the wall surfaces, most likely hydroxide ions (OH^-) or other chemical groups or substances rich in hydrogen. A hydrogen nucleus can absorb a neutron (forming a nucleus of deuterium), but what is more important, hydrogen is the lightest of all the nuclei, consisting of a single proton, and so its thermal motions are exceptionally rapid. A layer of hydrogen adsorbed on a surface could therefore greatly increase the probability of inelastic scattering.

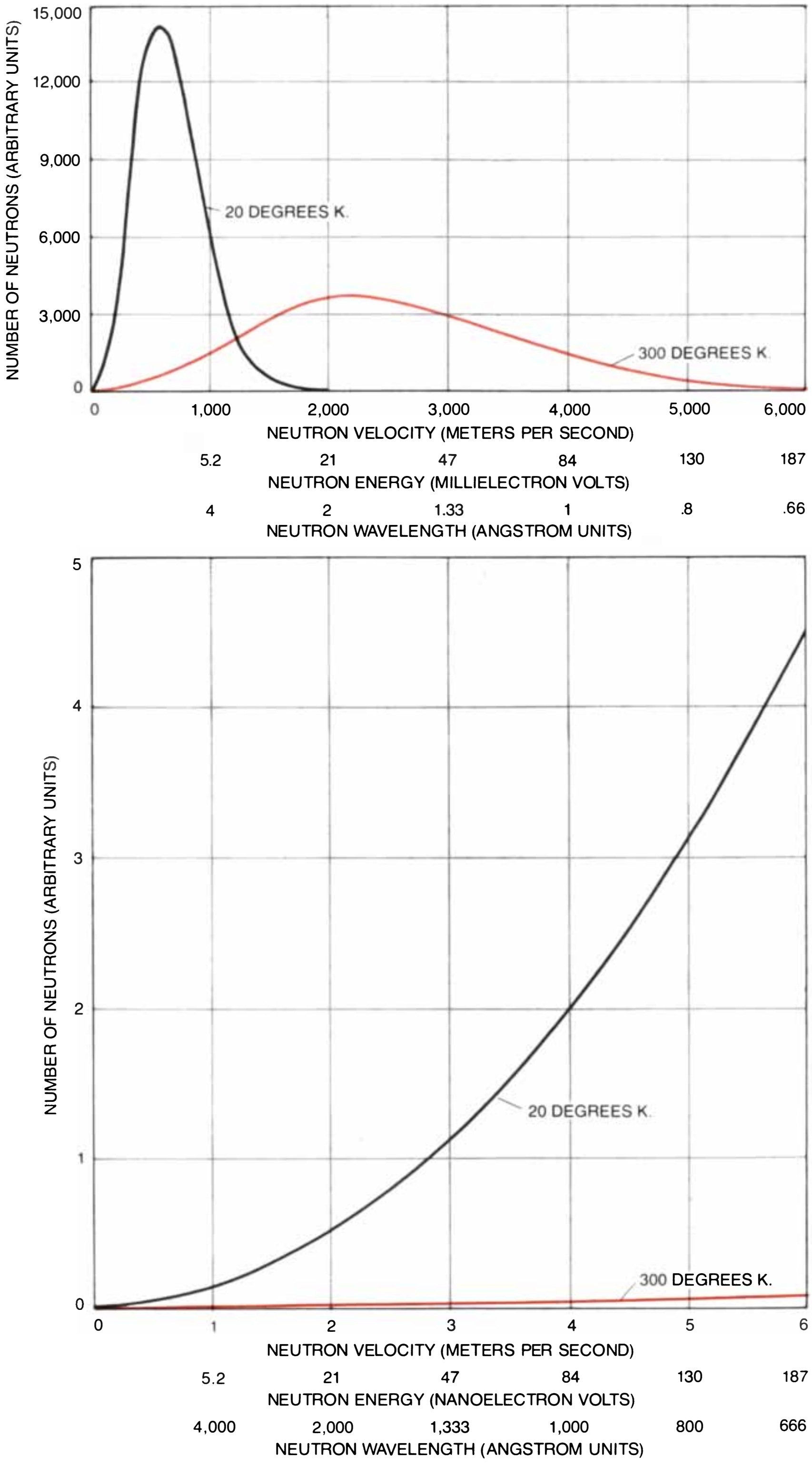
One of us (Golub), working with William A. Lanford of Yale University, has measured the hydrogen on surfaces prepared in the same way as the materials in vessels for the confinement of ultracold neutrons. We found enough hydrogen to account for the anomalously high rate of neutron loss. In another investigation A. V. Strelkov and M. Hetzelt at the Institute of Nuclear Physics at Alma-Ata in the U.S.S.R. surrounded an ultracold-neutron vessel with detectors that measured the energy of any escaping neutrons. They found that neutrons leaving the bottle were near thermal energy, which is what would be expected if the neutrons had gained velocity through inelastic scattering by impurities.

There is also evidence, however, that is not easily reconciled with the impurity hypothesis. If impurities are responsible, the rate of loss should depend on the temperature of the walls. Raising the temperature both increases the thermal motions of the atoms and drives some of them off the surface entirely, so that some cancellation of the two effects is likely. So far no dramatic temperature dependence of the storage time has been detected. Attempts to replace the adsorbed hydrogen with deuterium, which would yield a smaller probability of loss, have also been unsuccessful.

Strelkov and his colleagues have carried out an ingenious experiment in an effort to improve the maximum storage time. They employed neutrons with an energy of less than 3×10^{-8} electron



up to the material surrounding the core, which is called the moderator. At the Institut Laue–Langevin the core is immersed in heavy water (D_2O), which serves as the moderator. After only a few collisions the energy of the neutrons is comparable to that associated with the thermal motions of the nuclei in the moderator; the neutrons can then be regarded as a gas in thermal equilibrium with the moderator. The nuclei themselves have a range of energies, and when a neutron collides with a particularly fast-moving nucleus, the neutron is likely to gain energy. In a collision with a slow nucleus, on the other hand, the neutron generally gives up energy. After many collisions the average neutron energy reaches a steady value determined by the temperature of the moderator, but some neutrons are warmer and some are cooler. Ultracold neutrons are produced only in those improbable events in which a neutron loses almost all its energy in a single collision. Because such events are unlikely ultracold neutrons are a minor component of the total flux. The thermal energies of the nuclei are shown here by their colors, as is indicated in the key at the left; neutron energies are given by the length of the line segments in their trajectories.



DISTRIBUTION OF ENERGIES in a gas of neutrons at thermal equilibrium is determined by the temperature. In the upper graph the complete spectrum of neutron energies is shown, with the number of neutrons being given as a function of three related quantities: velocity, energy and wavelength. At room temperature (300 degrees Kelvin) the average velocity is some 2,350 meters per second but the distribution is very broad. At the temperature of liquid hydrogen (20 degrees K.) the range is much narrower and the average velocity is some 600 meters per second. In the lower graph the extreme-low-energy portion of the spectrum is shown magnified by a factor of 1,000. It is the neutrons in this region, with velocities of less than six meters per second, that are considered to be ultracold. Even at 20 degrees K. the ultracold neutrons are rare, and at 300 degrees they make up less than one part in 10^{11} of the total neutron flux. It is nonetheless possible to extract them from a reactor that is at room temperature.

volt, which is about a factor of 10 smaller than the critical energy for many materials. The neutrons were stored in a vessel whose interior surfaces were continuously renewed by the deposition of a metallic vapor. No increase in the storage time could be detected as fresh material was deposited. The experiment is not conclusive, however. The hydrogen content of the newly deposited layers has not been measured, and it may not be appreciably smaller than that of the original wall material.

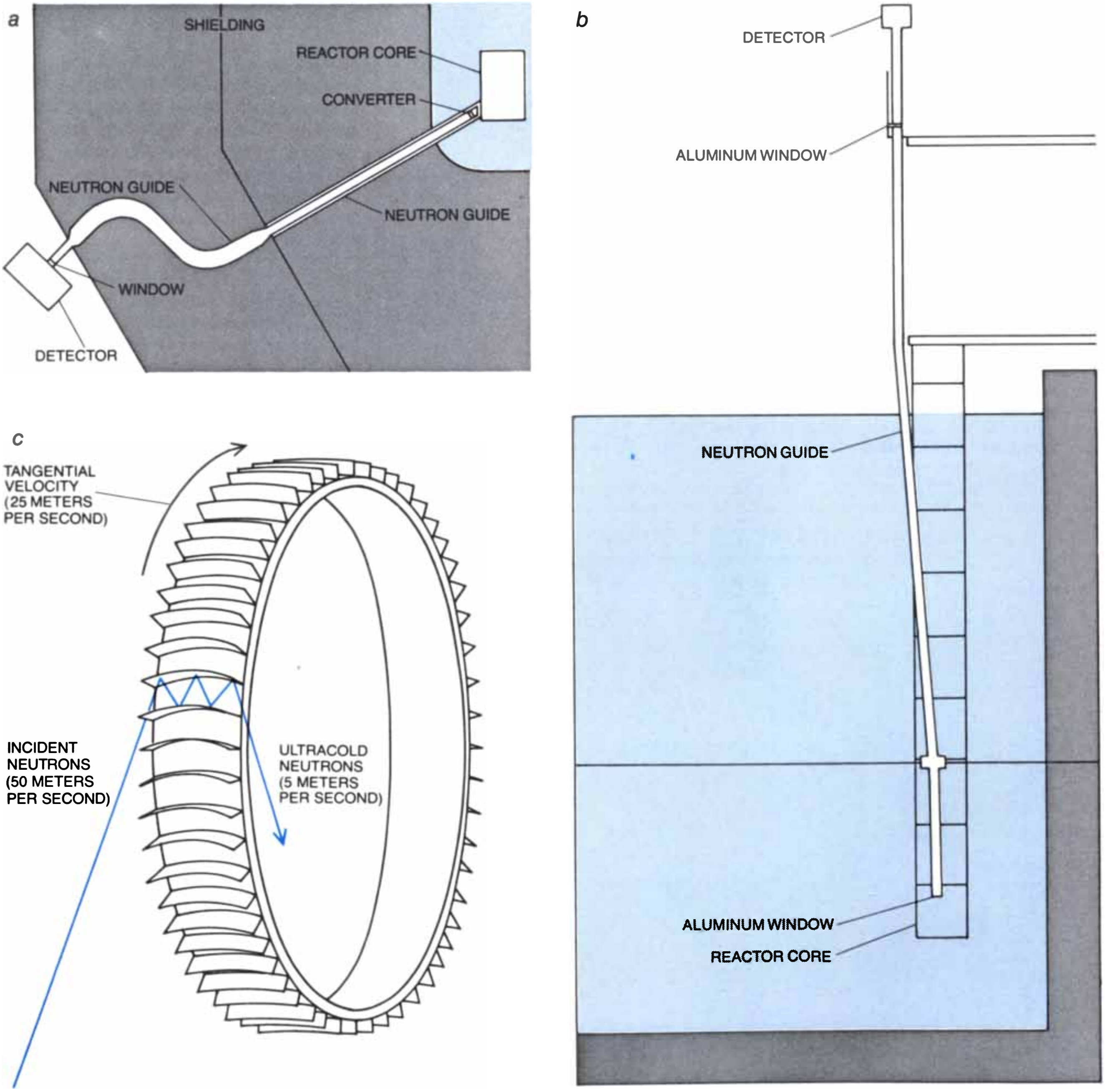
Other effects have been proposed to account for the anomalous storage times. For example, neutrons might scatter inelastically from atoms diffusing through the walls, or they might leak out through pores in the wall material. Large-scale vibration of the walls, caused by noise, could impart an additional velocity to the nuclei that would increase the rate of inelastic scattering. None of these effects, however, seems large enough to explain the rate of neutron loss, and contamination of surfaces with hydrogen remains the best hypothesis. Experiments that are now in preparation, in which the walls will be cooled to the temperature of liquid helium and contamination will be reduced by an ultrahigh vacuum, promise to settle the matter.

Complex surface interactions can be avoided entirely by another means of storing ultracold neutrons: confinement in a magnetic field. Magnetism is the usual method for controlling the trajectories of other elementary particles, such as electrons or protons, in a particle accelerator or a storage ring, but there the magnetic field acts on the electric charge of the moving particle. Since the neutron has no electric charge, it is not affected by a uniform magnetic field. The neutron does, however, have a magnetic dipole moment, that is, it has a north and a south magnetic pole. The axis of the dipole moment is lined up with the axis about which the particle spins, and indeed the spin angular momentum of the neutron may ultimately be responsible for generating the magnetic moment.

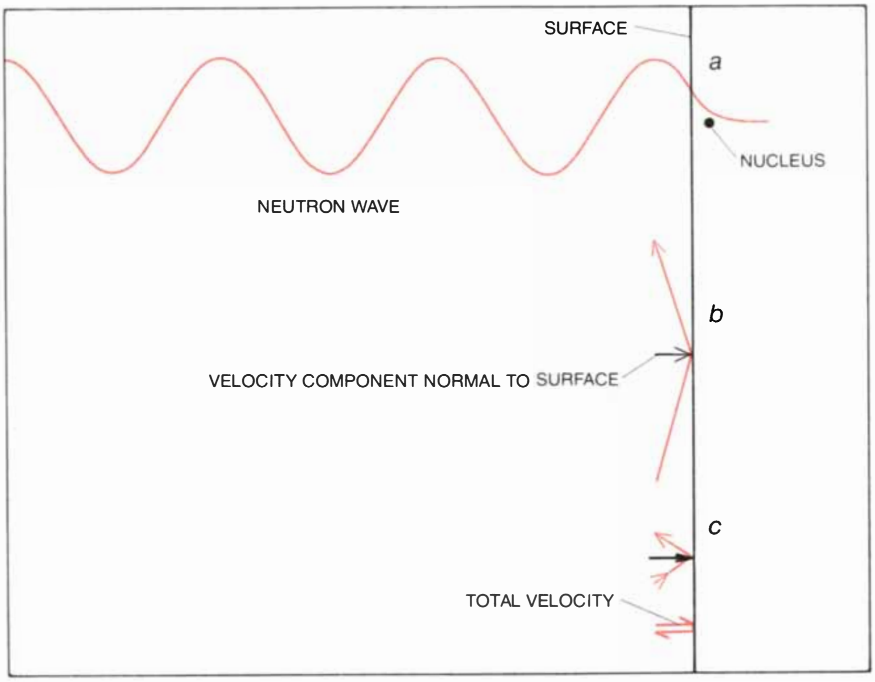
In an external magnetic field that varies in magnitude from point to point, the magnetic dipole moment provides a means of influencing the motion of a neutron. In such a field gradient one pole must be immersed in a stronger field than the other pole is, so that a force is exerted on the particle. A field configuration suitable for confining neutrons is a hexapole, which is generated by six magnetic poles of alternating north and south polarity arranged around a cylinder. The lines of force that represent the direction of the field form a quite complicated pattern, but the magnitude of the field varies in a simple way. The magnitude is zero along the axis of the cylinder, and in all directions away from the axis it increases as the square of the distance.

The response of a neutron to a hexapole field depends on the orientation of the spin axis and on that of the magnetic dipole moment. If the moment is parallel to the external field, the neutron is subject to a force directed away from the axis of the cylinder and is expelled from the apparatus. If the dipole moment is antiparallel to the field, the force is directed toward the center. (According to the rules of quantum mechanics no other orientations are possible.) If a neutron whose magnetic dipole is antiparallel to the field has a velocity perpendicular to the axis that is too small to overcome the force generated by the field gradient, then the neutron will be permanently confined. Ordinarily the orientation of the neutrons is random, half of them being parallel and half antiparallel, and so a hexapole field can confine half of the available neutrons.

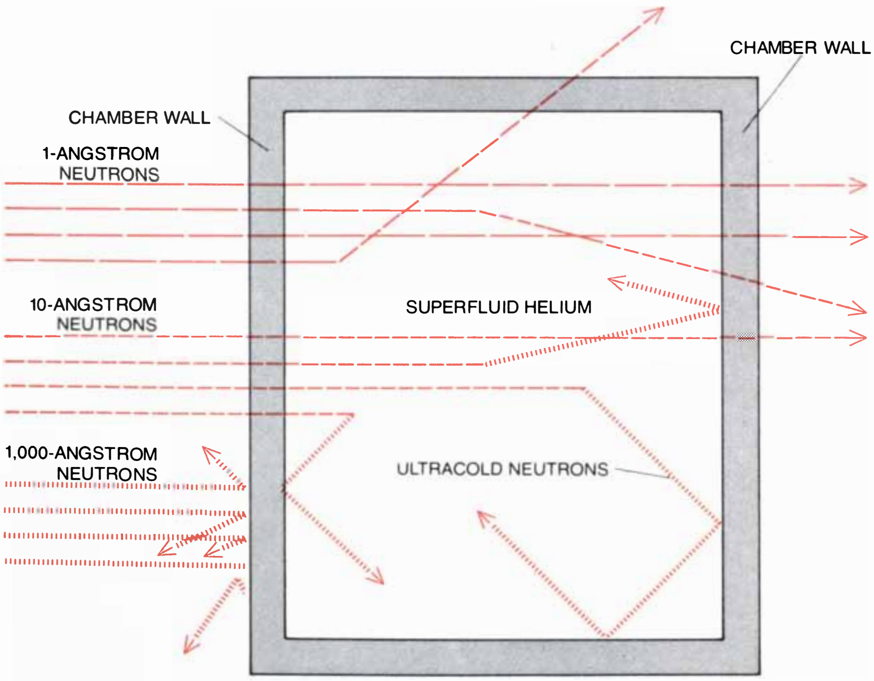
Of course, with a cylindrical hexapole field, neutrons can escape from the ends of the cylinder. A practical magnetic bottle for neutrons is constructed by curving the cylinder back on itself to form a torus. The result is a storage ring in which the neutrons follow helical tra-



EXTRACTION OF ULTRACOLD NEUTRONS from the vicinity of a reactor core can be accomplished by several means. A horizontal guide tube (*a*) conducts the neutrons through the shielding that must surround the core; if the tube curves, neutrons of higher energy pass through the walls and are lost, whereas ultracold neutrons follow the curves. A vertical guide tube (*b*) accepts neutrons of energies somewhat higher than ultracold energy at the bottom, but in climbing to the apparatus at the top they are slowed to the ultracold-energy range by the earth's gravitational field. In a neutron turbine (*c*) neutrons moving at about 50 meters per second bounce several times from the receding blades of a rapidly spinning wheel, losing energy with each bounce. They emerge with a speed of about five meters per second.



TOTAL REFLECTION of ultracold neutrons from a solid surface is a consequence of their interaction with the nuclei in the solid. The wave that describes the motion of the neutron (*a*) has a length of some 1,000 angstrom units (one angstrom is equal to 10^{-8} centimeter), and it can penetrate only about 100 angstroms into the solid. As a result the surface repels the neutron, and the repulsion can be overcome only if the neutron is moving toward the surface fast enough. A neutron of moderate velocity (*b*) is sure to be reflected if it strikes the surface at a glancing angle, so that the component of velocity normal to the surface is smaller than the critical velocity. Ultracold neutrons (*c*) are those whose total velocity is below the critical value, and so they are reflected even if they strike the surface head on. For copper the critical velocity is six meters per second; neutrons moving slower than this can be stored in a copper bottle.



COOLING OF NEUTRONS in liquid helium exploits the quantum-mechanical properties of a superfluid to confine ultracold neutrons at a density higher than their density in the reactor. The superfluid has only a few possible states of motion, which are widely separated in energy. A neutron with a wavelength of about 10 angstroms can give up almost all its energy to the superfluid, and it is unlikely to gain the energy back again. If the superfluid is put in a vessel made of a suitable material, 10-angstrom neutrons can pass through the walls, to be converted by the helium into ultracold neutrons that are then trapped in the vessel by total reflection.

jectories around the axis of the torus. The device is strictly analogous to the rings employed at several high-energy physics laboratories to accumulate protons or electrons.

As in a material vessel, the magnetic neutron bottle has a critical velocity: it is the radial velocity at which a neutron is able to escape the field gradient. The critical velocity is proportional to the square root of the strength of the field, and for practical values a very powerful magnetic field is needed. A ring of this kind has been built by Wolfgang Paul and Uwe Trinks of the University of Bonn, in which a field of more than 50,000 gauss is supplied by superconducting windings. The ring is operated at the high-flux reactor of the Institut Max von Laue–Paul Langevin in Grenoble, a laboratory under the joint administration of France, West Germany and Britain. Neutrons have been stored in the ring for as long as 45 minutes, or about three times the average lifetime of the neutron.

The same group of investigators is now preparing a spherical magnetic bottle that is expected to be still more efficient. Three superconducting windings are arranged on the surface of a sphere, with one winding at the equator and the others parallel to it. The field created by these conductors has a magnitude of zero at the center of the sphere, and the magnitude increases symmetrically in all directions away from the center. Neutrons of sufficiently low energy are confined in circular or elliptical orbits much like those followed by electrons in a simple model of the atom.

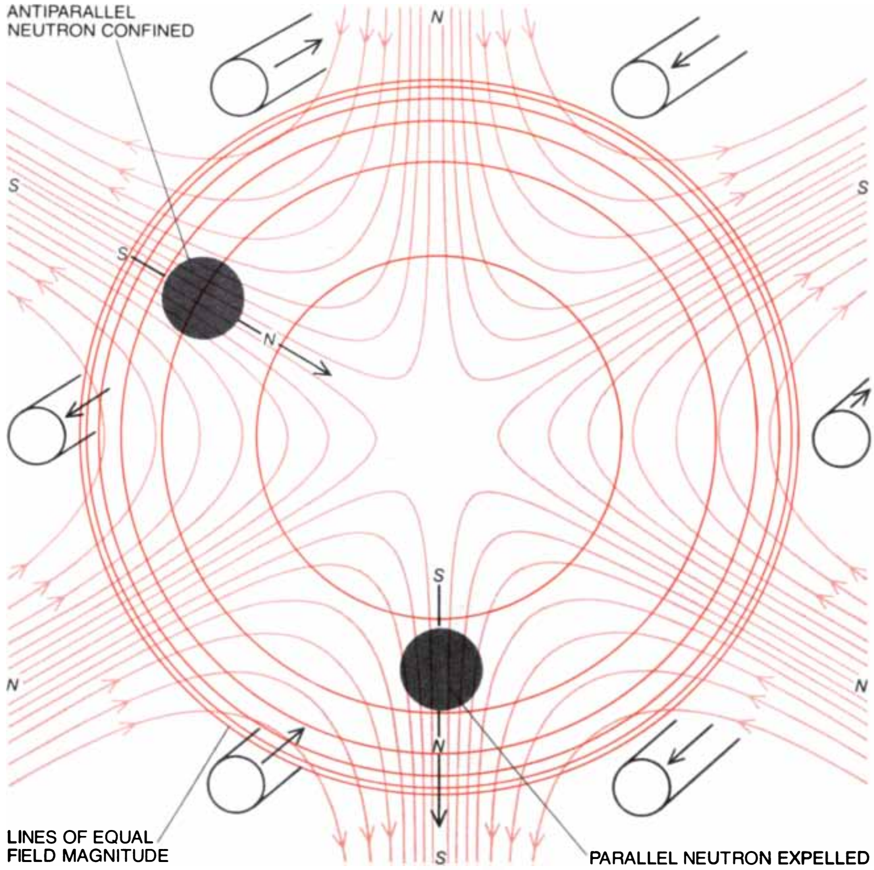
Even with a perfect containment vessel there would be a fundamental limit to the density of ultracold neutrons that could be obtained by simple extraction from a reactor. The concentration in the apparatus cannot exceed the concentration in the reactor itself; indeed, because of losses in the guide tubes and elsewhere not even equal concentration can be achieved. In this regard the behavior of the neutrons is again like that of the molecules in a gas. The system made up of the reactor, the neutron bottle and the guide tube can be represented by two gas bottles, one full and one empty, connected by a pipe. When a valve in the pipe is opened, the pressure in the empty bottle gradually rises until it equals the pressure in the full one, but it cannot rise any higher.

In order to achieve a higher density of ultracold neutrons there is no alternative but to cool the neutrons as they emerge from the reactor. Even methods of extraction that reduce the average energy of the neutrons, such as a vertical guide tube or a turbine, do not remove heat. In order to cool a gas it is not enough to reduce the average energy of the particles; the spread in energies around that average must also be made smaller. The vertical guide and the turbine merely subtract a fixed quantity from each neutron's energy, leaving the spread unchanged.

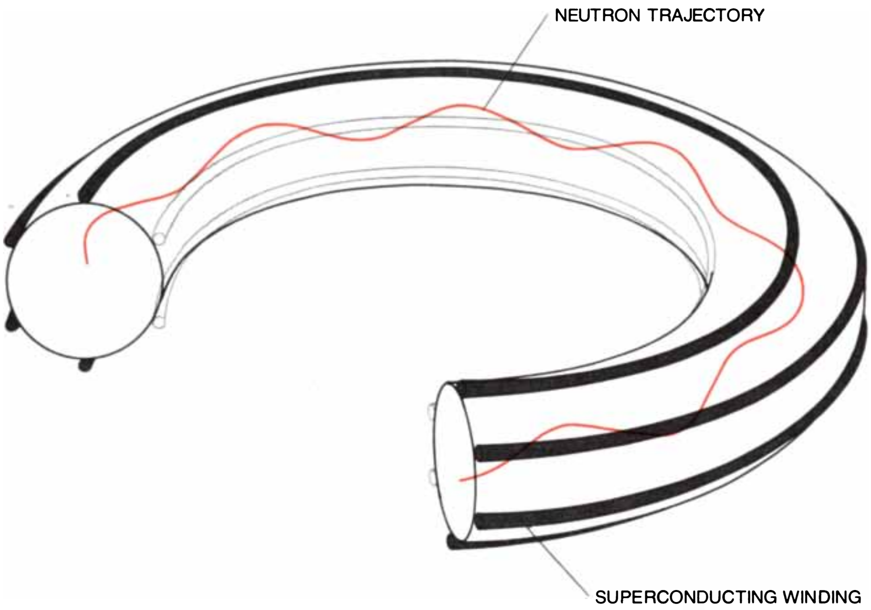
A source of ultracold neutrons proposed by two of us (Golub and Pendlebury) would provide true refrigeration, the coolant medium being liquid helium. At temperatures below about two degrees K. helium becomes a superfluid: it loses all viscosity and exhibits quantum-mechanical properties on a macroscopic scale. The superfluid has only a few possible states of motion, each of which has a definite energy and momentum. An exchange of energy between a neutron and the superfluid can take place only if the quantity of energy and momentum exchanged is equal to the difference in energy and momentum between two of these superfluid states. A neutron with an energy of about one millielectron volt (or a wavelength of 10 angstroms) can be transformed into the ultracold-energy range by scattering from the superfluid. The energy given up by the neutron creates a quantum-mechanical excitation in the helium called a phonon, which subsequently decays. Once the neutron has become an ultracold one the only way it can gain energy from the helium is by the exact reverse of this process, in which a phonon promotes the neutron to an energy of about one millielectron volt. If the temperature is kept low enough, say below one degree K., phonons are rare and scattering to higher energies is unlikely.

The neutrons not only give up energy to the superfluid but also give up heat. The final distribution of energies is confined to a narrow range. If the superfluid is placed in a flux of 10-angstrom neutrons, heat is transferred continuously to the helium. If the heat is then removed from the helium by an external refrigeration system, ultracold neutrons are created at a steady rate. At the Institut Laue–Langevin we have recently demonstrated production of ultracold neutrons by this method.

If the helium is placed in a suitable vessel, ultracold neutrons created in the superfluid should be trapped by the walls. It is expected that the neutrons can accumulate to much greater densities than can be achieved by other methods. Losses in the helium itself are negligible. We have already seen that inelastic scattering to higher energy is unlikely, and if the helium is made up exclusively of the common isotope helium 4, absorption is impossible: a helium-4 nucleus cannot absorb a neutron. The cooling of the apparatus should also greatly reduce losses at the walls caused by inelastic scattering from impurities. Another significant mechanism of loss can be eliminated entirely: the entrance

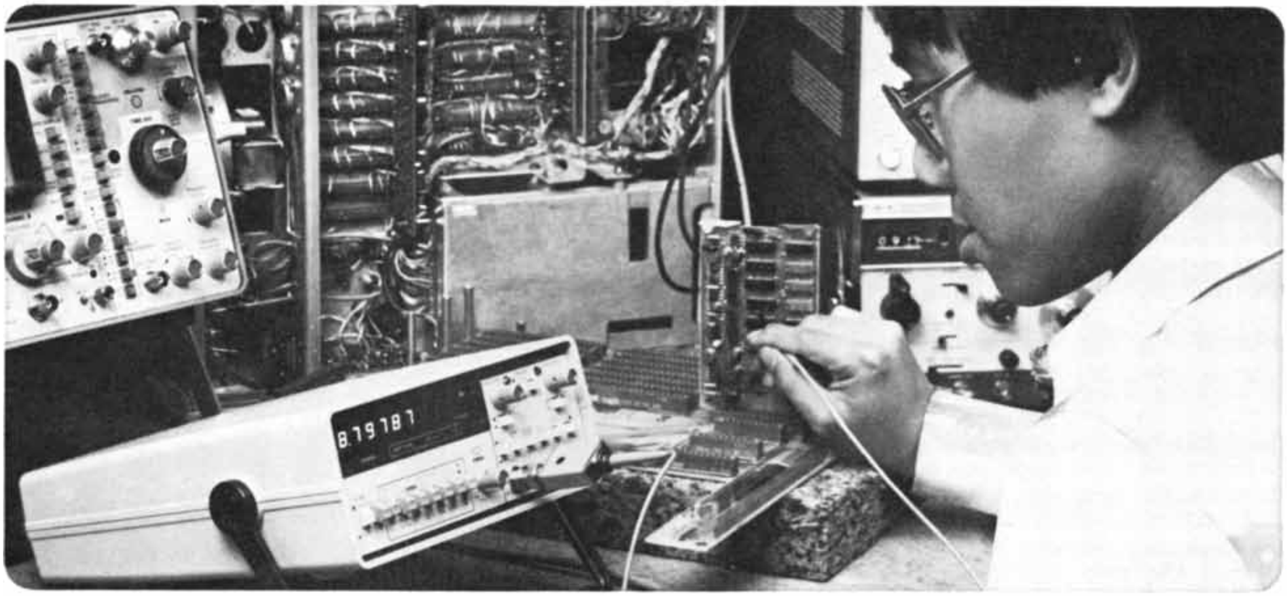


HEXAPOLE MAGNETIC FIELD can also confine ultracold neutrons. Although the form of the field is complicated, the magnitude varies in a simple way: it is zero along the axis of the field and increases as the square of the distance from the axis. The resulting field gradient acts on the magnetic dipole moment of the neutron. One end of the neutron is always in a stronger field than the other end, and so there is a net force acting on the particle; the direction of the force depends on the orientation of the neutron. A cylindrical hexapole field can be generated by six straight conductors with electric currents flowing in the directions indicated by the arrows.



TOROIDAL STORAGE RING traps ultracold neutrons in a hexapole magnetic field. The neutrons follow helical trajectories around the centerline of the torus. The two inner windings (*gray*) can be eliminated because centrifugal forces prevent the neutrons from escaping in this direction. The ring constructed at the University of Bonn and operated at the Institut Laue–Langevin follows this design. Not shown is the apparatus for injecting the neutrons.

HP measurement and computer advances



New HP counter chip plus microprocessor provide advanced capability in a low-cost universal counter.

Using a reciprocal-taking technique for high resolution and accuracy at low frequency, the HP 5315 universal counter measures frequency to 100 nanosecond resolution. With the additonal capabilities of a full range of period, frequency ratio, time interval averaging, and totalizing, the 5315 is the equal of instruments costing twice its price.

Engineers and scientists use electronic counters for precise measurements of frequency and time intervals. In labs where specific measurement requirements are varied, universal counters, those that measure frequency *and* time intervals, are the type most often desired. In the past, the relatively high cost of such counters often precluded their use—especially if one wanted to measure low frequencies rapidly and accurately. This latter requirement meant that the user had first to measure the waveform's period (which can be done rapidly and accurately by most counters), and then calculate the reciprocal of the reading to obtain frequency—a slow, inconvenient process. Or one could buy a reciprocal-taking counter that does this automatically for about $2000—a lot of money for a low-budget lab.

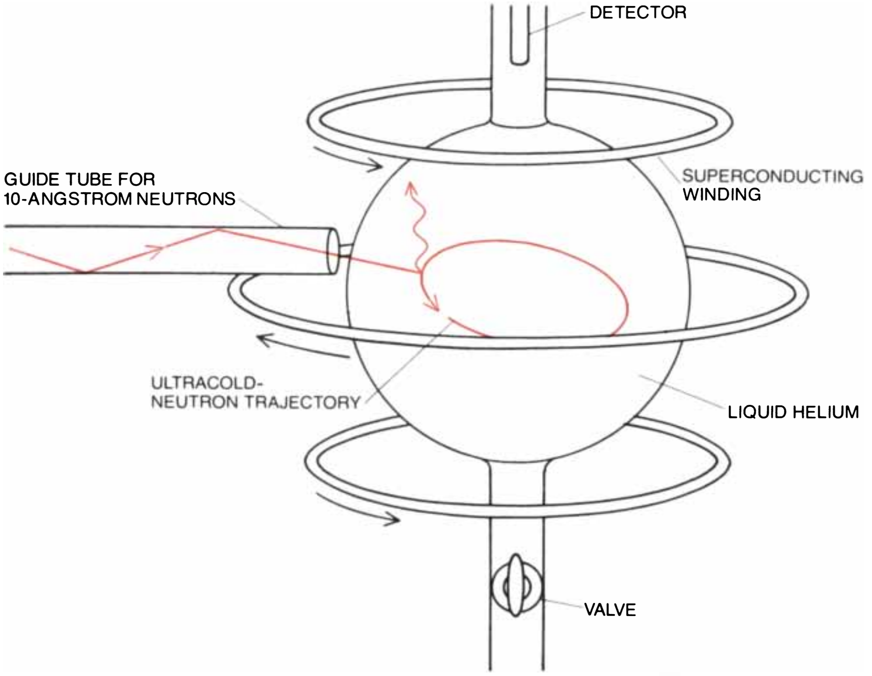
Now *computed measurement* capability, in the form of HP's large-scale integrated circuit technology plus microprocessor control and computation, brings to the new 5315 universal counter improved performance and reduced complexity at a much lower price than previous generations of counters.

Only three digital IC's are required to give this performance to the 5315: a single-chip microcomputer, an integrated display driver chip and, most important, an LSI chip known as the Multiple Register Counter. The MRC was developed by HP for the new generation of microprocessor-controlled counters, and contains all of the data acquisition functions needed by a universal counter.

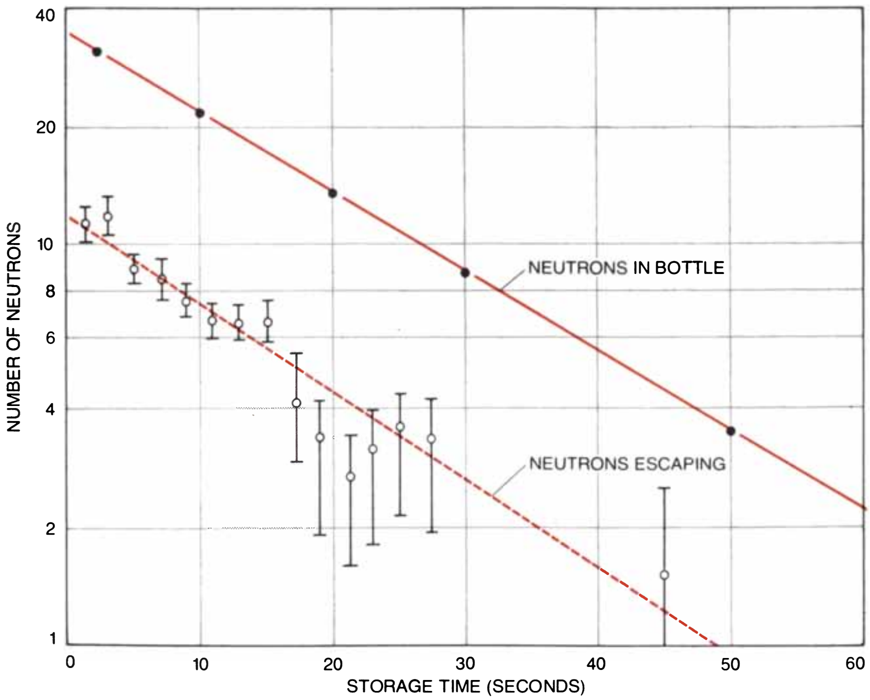
The primary feature distinguishing the 5315 from earlier moderately priced counters is that it is a computing or reciprocal frequency counter. The short measurement time and high resolution of the reciprocal-taking technique enables the 5315 to display at least seven digits of answer per second of gate time, independent of the input frequency. Gate time is continuously variable for greater convenience.

The 5315 is available in two versions. The 5315A comes in a rugged plastic case that is suitable for bench use, and can be equipped with an optional internal battery pack for portable and field applications. The 5315B is essentially the same instrument housed in a metal case for rack mounting or stacking. The 5315A's electrical quietness (RFI and EMC) suits it well for working with sensitive circuits, and the metal-cased 5315B is even better, having been designed to meet demanding military specifications on interference.

Prices: HP 5315A, $800*; HP 5315B, $950*; optional battery (5315A only), $225*.



SPHERICAL MAGNETIC BOTTLE has a magnetic field that is zero in the center and increases in all directions as the square of the distance from the center. The vessel, which is now being built by Paul and his colleagues at Bonn, will be operated at the Institut Laue–Langevin as a superfluid source of ultracold neutrons. The bottle will be filled with liquid helium, the magnetic field will be switched on and neutrons with a wavelength of about 10 angstroms will be injected through the wall. Interactions with the helium will convert these neutrons into ultracold neutrons, which will remain trapped by the magnetic field when the helium is drained.



STORAGE TIME of ultracold neutrons in a material bottle is substantially shorter than theoretical calculations suggest it should be. The probable cause of the discrepancy was identified by surrounding a copper vessel with neutron detectors so that both the neutrons remaining in the bottle and those escaping could be counted. The fact that both quantities decline at the same rate suggests that the neutrons are gaining energy as a result of inelastic scattering and being lost thereafter by transmission through the walls. The high rate of inelastic scattering is probably caused by impurities, notably hydrogen, on the wall surface. The experiment was done by A. V. Strelkov and M. Hetzelt at the Institute of Nuclear Physics at Alma-Ata in the U.S.S.R.

hole to the vessel, through which neutrons can also escape. There is no need for an opening, since the walls can be made transparent to the incident 10-angstrom neutrons and opaque to those of longer wavelength created inside the container.

Once a neutron reaches the ultracold-energy range in such a helium source, it will have a high probability of staying there until it decays. In most experiments with conventional systems of extraction and storage the density of ultracold neutrons is less than one per cubic centimeter. We estimate that a practical superfluid source, placed outside a reactor at the end of a guide tube, could increase the attainable density of ultracold neutrons by a factor of from 10 to 100. Such a source is now being built at the Institut Laue–Langevin.

The spherical magnetic bottle under construction at Bonn by Paul and Trinks will be operated with the superfluid source. The sphere will be filled with liquid helium with the magnetic field switched on; then it will be exposed to a beam of 10-angstrom neutrons. When enough ultracold neutrons have accumulated inside, the helium will be drained away but the neutrons will remain, trapped by the magnetic field gradient. The only significant mechanism of loss should be the beta decay of the neutrons themselves.

One motive for improving the storage time of ultracold neutrons is the desire to measure accurately the lifetime of the neutron. As we have indicated, a neutron outside the nucleus decays to yield a proton, an electron and a neutrino. This process of beta decay is mediated by one of the four basic forces of nature, namely the weak force. The decay is a statistical process and the period that any given neutron will survive can never be predicted, but an average or typical lifetime can be measured.

The currently accepted value of the average lifetime of a free neutron is 918 seconds, or 15 minutes 18 seconds. The accuracy of the measurement, however, is given as plus or minus 14 seconds, which represents a probable error of 1.5 percent. The lifetimes of many less familiar particles are known with much greater precision. For example, the muon, which also decays under the influence of the weak force, has a lifetime of only two-millionths of a second, but it has been measured to an accuracy of .003 percent. A more accurate knowledge of the neutron lifetime may lead to an improved understanding of the weak force.

A straightforward way of measuring a lifetime is to start with a known number of particles and monitor the number remaining as a function of time. The result should be an exponential curve from

"Chess Challenger-10 Wins Microchess Tourney"
—Personal Computing Magazine
February, 1979

Genius Offspring

"Its successor, the new Chess Challenger "7", is infinitely more powerful."
—S. Samole
President, Fidelity Electronics

Chess Challenger-10 did more than win the Penrod Memorial Microchess Tournament, it literally trounced all opponents. Personal Computing Magazine, February, 1979, reports, *"Chess Challenger-10 emerged as the easy victor with ten wins, two draws and no losses."*

All Top Name Performers

There were no amateurs in the championship playoff. Every contender bore the brand of a well-known electronic chess game, and each was accompanied by its entourage of coaches, programmers, and engineers. After each contestant had played all of the opponents in round robin fashion, the brilliant Challenger-10, stood far ahead of its second place runner-up.

Nobody Knew

Unknown to the other companies, the undefeated tournament leader was being retired after the contest. Taking its place was a far more powerful chess computer, the Challenger "7". This new micro-computer had already beaten the official undefeated champ during a series of pre-tournament warm-up games at the factory. Its engineers explain that it is simply 14 months ahead in technology, in finer algorithm sophistication and in its superb performance.

Improve Your Game to Near Brilliant

Within its seven different levels of play, you can enjoy every degree of chess competition, from beginner to tournament skill. Its total flexibility lets you change games midstream or switch sides with the computer to see how it would handle your dilemna. You can add pieces to your side or take away the computer's Queen. It is a superb teacher!

Touch the PV key and the "7's" total recall memory will verify every piece position on the board. You can even set up hypothetical encounters to test its reaction at each level.

Fidelity's Challenger "7" is able to analyze over 3,024,000 board positions. It masterfully handles over one thousand book openings and will respond to any deviation. Academic openings as Sicilian, French, Ruy Lopez and Queen Gambit Declined, are just some of the challenges to keep you on your toes.

It Knows Every Rule in the Book

The Challenger "7" will permit you to castle or perform an En Passant capture or do so itself, if that is its best move. When your pawn has reached the eighth rank, it will be automatically raised to a Queen, unless you tell the computer to promote it to another piece. It will take on any player and sharpen his skills considerably...but it won't permit illegal moves.

Sound Signals

You just cannot believe the chess I.Q. of this phenomenal unit. The Challenger "7" has a beep audio feedback which sounds to acknowledge your move, and double beeps when the computer has made its response.

At Level 1, its average response time is 5 seconds. At Tournament Level 7, the Challenger makes championship decisions in just 3 minutes.

Unbeatable in Price As Well As Play

Best of all, the Chess Challenger "7" is the most affordable electronic unit you can own. It is just $89.95 complete with Staunton designed pieces and UL approved 110V AC adaptor.

All pieces are magnetized, to stay where you place them on the permanent metal board. The set is mounted in a simulated wood-grained housing which measures 12⅛" x 8" x 1". Bright, one-half inch tall LED electronic digits, provide unmistakably clear readout. The unit is backed by a 90-day manufacturer's limited parts and labor warranty.

Enjoy It for 10 Days—At Our Expense

As a gift or for yourself, the "7" is unquestionably the finest chess computer you can select...but, if within 10 days, you are not satisfied, simply return it for a prompt, no-questions-asked refund.

Final Results

Reprinted Courtesy of Personal Computing, February, 1979. P. 66. (Darker lines ours.)

	CONTESTANTS		#1	2	3	4	5	6	7	8	9	Games Won	Games Drawn	Games Lost	FINAL SCORE	FINAL POSITION
1	MICRO-CHESS 1.0 (Heath H-8)	W	X	½	Ø	1	Ø	Ø	Ø	—	—	1	3	8	2½	7*
		B		½	½	Ø	Ø	Ø	Ø							
2	MICRO-CHESS 1.5 (TRS-80)	W	½	X	½	½	Ø	Ø	Ø	—	—	Ø	5	7	2½	6*
		B	½		Ø	Ø	Ø	½	Ø							
3	MICRO-CHESS 2.0 (PET)	W	½	1	X	1	Ø	Ø	½	—	—	3	4	5	5	4
		B	1	½		½	Ø	Ø	Ø							
4	CHESS CHALLENGER (3 Level)	W	1	1	½	X	Ø	½	½	—	—	2	5	5	4½	5
		B	Ø	½	Ø		Ø	Ø	½							
5	CHESS CHALLENGER (10 Level)	W	1	1	1	1	X	1	½	—	—	10	2	Ø	11	1
		B	1	1	1	1		1	½							
6	BORIS	W	1	½	1	1	Ø	X	1	—	—	7	2	3	8	3
		B	1	1	1	½	Ø		Ø							
7	SARGON I (TRS-80)	W	1	1	1	½	½	1	X	—	—	6	5	1	8½	2
		B	1	1	½	½	½	Ø								
8	ATARI *Did not play*	W	—	—	—	—	—	—	—	X	—					
		B														

* Note: Microchess 1.5 wins 6th place over Microchess 1.0 by virtue of the tie breaking analysis of relative strength of opponents

CREDIT CARD ORDERS CALL TOLL FREE
800-621-5809
ILLINOIS RES: 800-972-5858
24 HOURS—7 DAYS/WEEK

Please send me ________ Chess Challenger "7(s)" at $89.95 plus $3.00 for shipping and insurance. Ill. residents add 5% sales tax. If not satisfied, I can return it within 10 days for a refund.

- ☐ Enclosed please find check or money order.
- ☐ Charge My Credit Card:
- ☐ American Express ☐ Master Charge ☐ Carte Blanche
- ☐ BankAmer./Visa ☐ Diners Club

Credit Card No. ____________

Master Charge # ____ Exp. Date ________

Name ____________

Address ____________

City ____________

State ________ Zip ________

Signature ____________

SA-6 © Camelot '79

Camelot DIRECT

801 Green Bay Rd., Lake Bluff, IL 60044

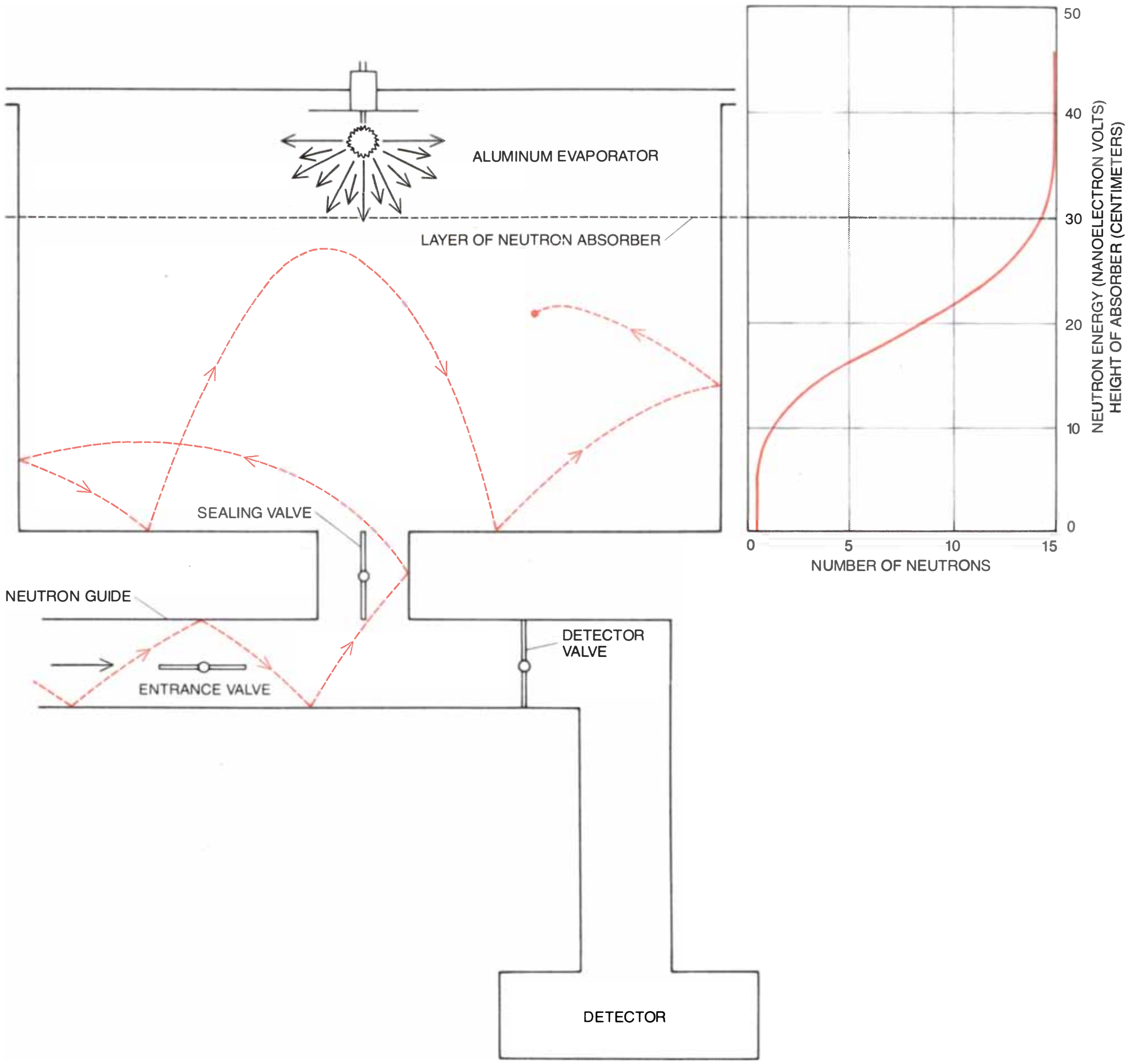
A DIVISION OF UNITED EDUCATORS, INC.

which the lifetime can easily be calculated. The main difficulty in determining the neutron lifetime by this means is that the neutron is an electrically neutral particle and cannot be detected with high efficiency. The only practical way of detecting a slow neutron is to allow it to be absorbed by a nucleus, which in response gives off excess energy in the form of energetic charged particles. It is the charged particles that are actually detected. A nucleus that is commonly employed in neutron detectors is that of helium 3, the rare light isotope of helium, which avidly absorbs neutrons.

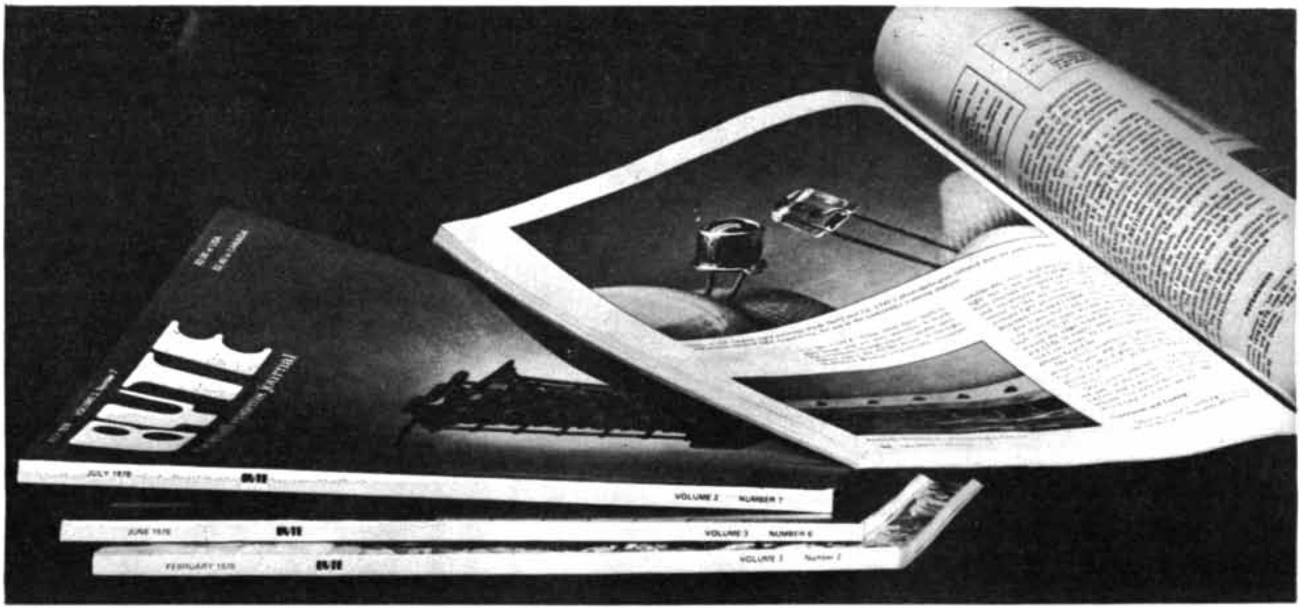
A conventional measurement of the neutron lifetime, carried out with a beam of unconfined neutrons, requires two detectors. One is a neutron detector, which counts the neutrons in the beam. The other detector is sensitive either to electrons or to protons, and so it counts the neutrons that have decayed. The calculation of the lifetime depends on the efficiencies of both detectors, which are known with only limited precision.

A reliable device for neutron storage could greatly simplify the procedure, so that only one detector would be needed. By monitoring the number of neutrons present in the storage vessel as a function of the storage time, the rate of loss could be determined. If losses resulting from the escape of neutrons could be neglected, then the measured rate would be determined entirely by the beta-decay lifetime.

Preliminary measurements of the lifetime have been made with ultracold



CONTINUOUSLY RENEWED SURFACE and neutrons moving at little more than one meter per second were employed in another experiment by Strelkov and his colleagues. After the neutrons were admitted by an entrance valve the chamber was sealed and a layer of neutron-absorbing material was lowered to a predetermined height. Neutrons capable of reaching the absorber were trapped there, so that only those moving too slow to bounce as high as the absorber could survive. After a period of storage these were counted by opening the valves to the detector. The energy spectrum of the neutrons (*graph at right*) was determined by measuring the number of neutrons remaining in the vessel as a function of the height of the absorber. In an effort to reduce surface contamination aluminum was evaporated continuously onto the walls of the chamber, but no effect of the deposited material on the storage time could be detected. The impurity hypothesis remains the favored one, however, because the hydrogen content of the evaporated-aluminum layers has not been measured.



Get the inside story...

on software, hardware, simulations, computer games, robotics, computers and calculators, languages and compilers, custom systems design and scores of applications.

Read BYTE.®

BYTE, the leading magazine in the personal computer field, will keep you aware of fast-paced changes in the ever growing world of microprocessors. You'll find **BYTE** tutorials invaluable, reports on home computer applications instructive, and the reviews of computer products stimulating.

Tutorial information in **BYTE** is of interest to both the beginner and experienced computer user. **BYTE**'s editorial content explores the fun of using and applying computers for personal satisfaction in activities from electronic music to electronic mail, from games to pragmatic personal business. In **BYTE** is found authoritative yet easily read coverage of principles of computer hardware and software design, approaches to novel applications, and essentials of proven interest to personal computer enthusiasts.

Each month serious computer users read **BYTE.** They're rewarded with timely articles by professionals, computer scientists and competent amateurs. Isn't it time you got the inside story? Subscribe now to **BYTE,** the Small Systems Journal.

Fill in and mail the coupon today. Read your first copy of BYTE. If it is everything you expected, honor our invoice. If it isn't, just write "Cancel" on the invoice and mail it back. You won't be billed and the first issue is yours at no charge.

Allow 6 to 8 weeks for processing. If you have any questions, **dial toll free 800-258-5485.** This is the number of the BYTE Subscription Department.

BYTE Subscription Dept. P.O. Box 590 Martinsville, NJ 08836
Please enter my subscription for

☐ One year $18 (12 issues) ☐ Two years $32 ☐ Three years $46

☐ Check enclosed entitles me to 13 issues for price of 12 (North America only)
☐ Bill Visa ☐ Bill Master Charge ☐ Bill me (North America only)

Card Number ____________ Expiration Date ____________

Signature ____________ Name (please print) ____________

Address ____________

City ____________ State/Province/Country ____________ Code ____________

Foreign Rates (To expedite service, please remit in U.S. Funds)
☐ Canada or Mexico ☐ One year $20· ☐ Two years $36 ☐ Three years $52
☐ Europe, one year (air delivered) $32
☐ All other countries, one year (surface delivered) $32. Air delivery available on request

7169

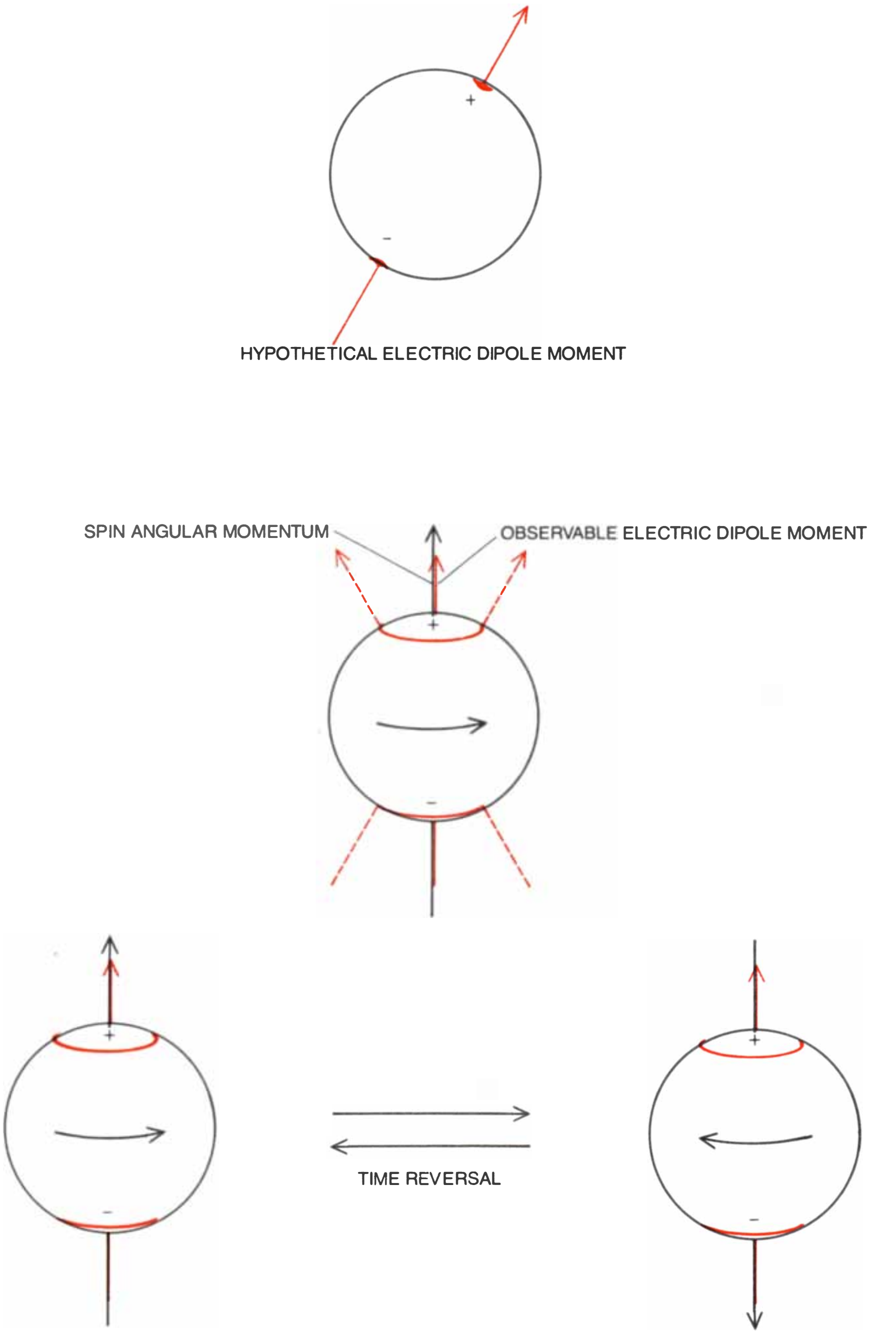
neutrons by the Bonn group, but greater precision should be attainable. Indeed, a definitive measurement should soon be possible. In the spherical magnetic bottle, which will be employed for a lifetime determination at the Institut Laue–Langevin, only statistical uncertainties (as opposed to experimental ones) are expected to limit the precision of the measurement. The statistical uncertainties depend solely on the number of neutrons available and on the patience of the experimenter.

A second undertaking that could benefit from longer storage times is the search for an electric dipole moment of the neutron. The total electric charge of the neutron is of course zero, but it may be that there are charges "inside" the particle and that the overall neutrality comes about only because the component charges happen to cancel one another, or add up to zero. (Indeed, the prevailing theory of the neutron, which states that the neutron is made up of the smaller particles named quarks, predicts a structure of exactly this kind.) If the positive and negative charges are permanently displaced, or if they are mobile but tend to spend much of the time in different regions, then the neutron will have an electric dipole moment: it will have a positive end and a negative end. The magnitude of the moment might be very small, but its consequences would not be trivial. The discovery of an electric dipole moment of the neutron would reveal a violation of the physical principle known as time-reversal symmetry.

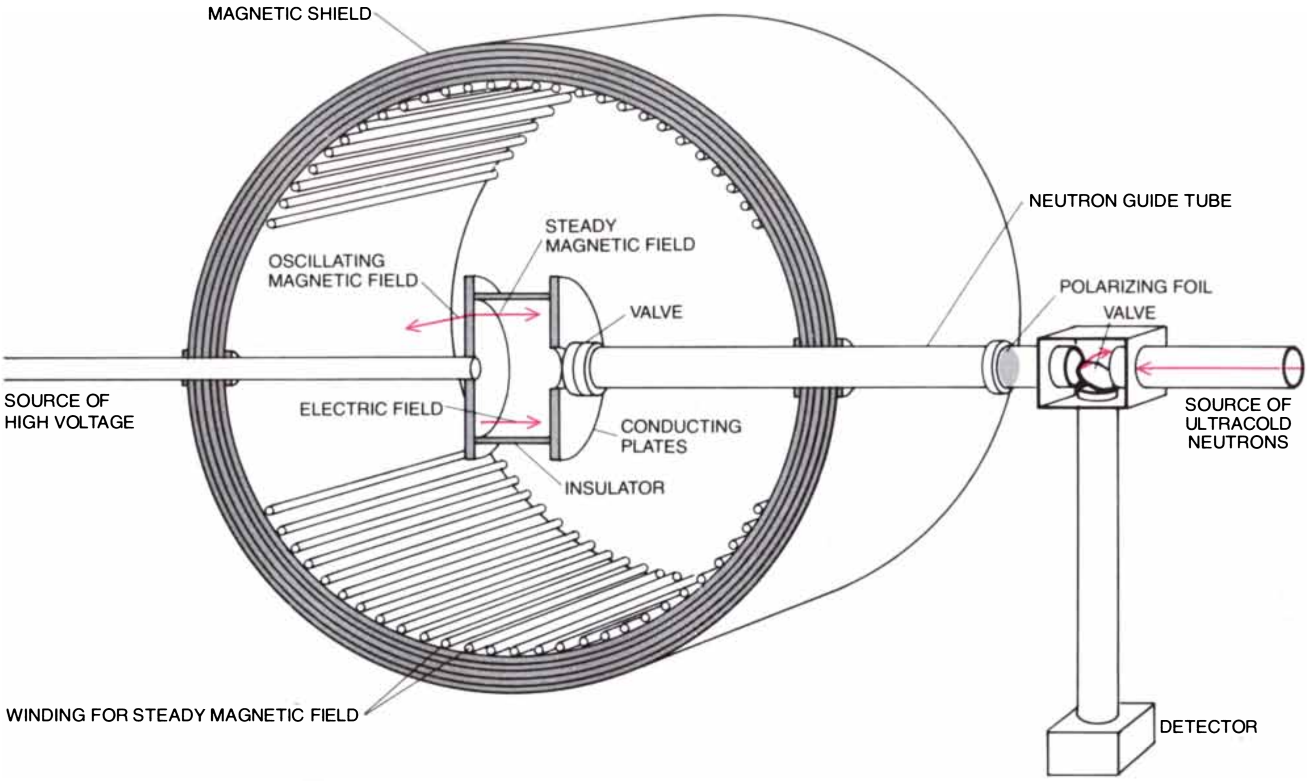
In physics the term symmetry has a broader meaning than it has in everyday usage. A symmetry principle states that an object or a process remains unchanged when some specified transformation is made. For example, the principle of symmetry with respect to spatial translation states that an experiment conducted in Alma-Ata will give the same result when it is repeated 8,000 kilometers away in Grenoble.

Two important symmetries in elementary-particle physics are those of parity and time reversal. The principle of symmetry with respect to parity implies that no physical process can distinguish left from right. If parity were a valid symmetry of nature, any event allowed by the laws of physics would also be allowed if left and right were interchanged, as in a mirror. In 1957 C. S. Wu and her colleagues at Columbia University demonstrated that parity is not a valid symmetry in all contexts. They showed that it is violated in the beta decay of cobalt-60 nuclei, and it is now known that parity is violated in all beta decays, including that of the neutron. (In some other interactions, however, parity is strictly conserved.)

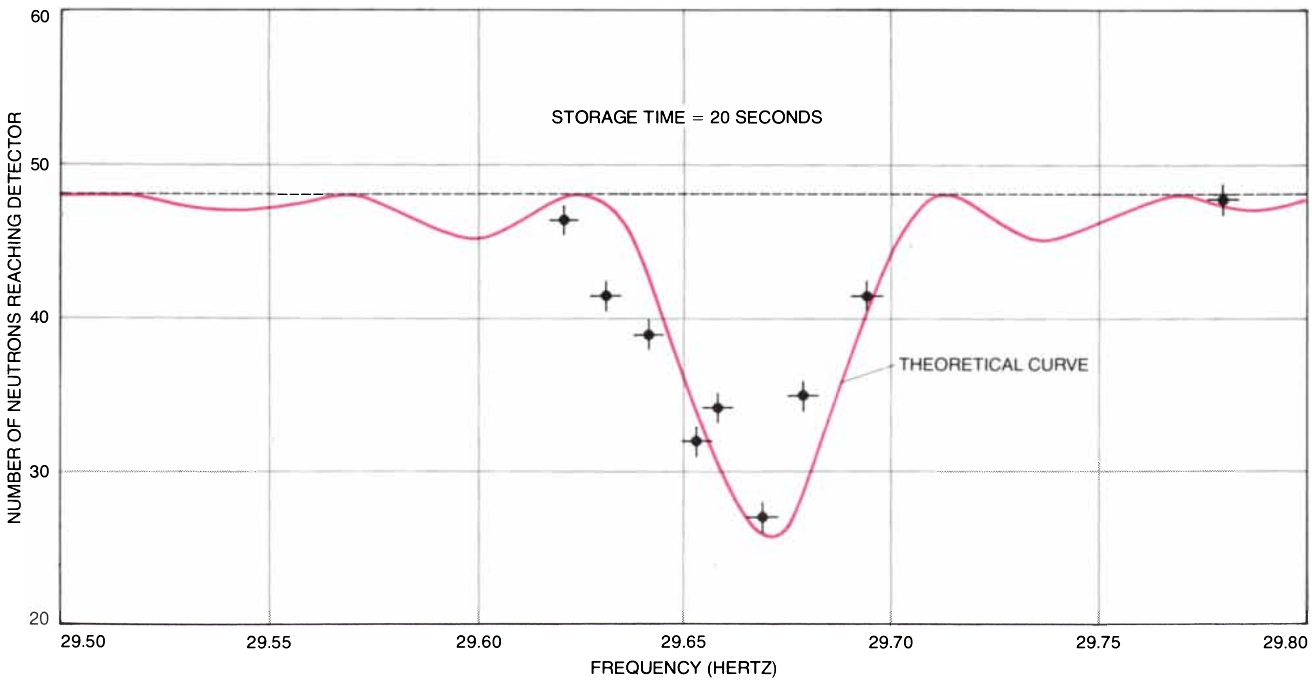
Time-reversal symmetry states that any process allowed in nature would still be allowed if all motions were reversed, or in other words if time flowed backward. On the macroscopic scale this principle is obviously violated more often than it is observed: glassware breaks but does not reassemble itself. Such a reversal, however, is forbidden only by its improbability; it would not violate any known physical laws. At the level of subatomic particles most events seem to be fully reversible. Just as the neutron decays into a proton, an elec-



ELECTRIC DIPOLE MOMENT of the neutron, if it is greater than zero, would signal a violation of time-reversal symmetry. An electric dipole moment is a separation of electric charges, giving the particle a positive end and a negative end. The moment can be represented as a vector, which is shown here pointing in the direction of the positive pole. The spin angular momentum of the neutron is also a vector, and the two vectors must be lined up along the same axis, or at least the only component of the dipole moment that can be observed is the one that is lined up with the spin vector. Time-reversal symmetry requires that the state of the particle remain unchanged if all motions are reversed. In some initial state the two vectors might be parallel. Reversing time would also reverse direction of spin vector, but it would have no effect on electric dipole moment. Two states therefore would not be equivalent and symmetry would be violated.



EXPERIMENTAL SEARCH for an electric dipole moment of the neutron attempts to detect a small change in the energy of a neutron when it is exposed to a strong electric field. The electric field is generated by a high voltage applied across two conducting plates; the chamber is also permeated by a steady magnetic field. Neutrons enter the chamber through a polarizing foil that admits only those neutrons whose magnetic moment is parallel to the steady magnetic field. An oscillating magnetic field then flips the spins and the magnetic moments of the neutrons, with the result that they cannot pass back through the foil to reach the detector. The frequency of oscillation needed to induce this spin-flip depends on the exact energy of the neutrons, which would be influenced by an electric dipole moment. Any change in the magnitude of the electric field would shift the resonance frequency if the neutron had an electric dipole moment.



RESONANCE SIGNAL recorded at the Institut Laue–Langevin with the apparatus shown at the top of the page is a sensitive indicator of neutron energy. The sharp dip in the curve results from the flipping of neutron spins and magnetic moments when the oscillating magnetic field is tuned to the resonant frequency. Neutrons whose spins have flipped are unable to pass through the polarizing foil and hence do not reach the detector. An electric dipole moment would be detected as a shift in the position of the dip. The precision with which any frequency can be measured is proportional to the counting interval available. The ability to store ultracold neutrons for comparatively long intervals should therefore make it possible to set a new upper limit on the value of the neutron electric dipole moment.

tron and a neutrino, so can a proton, an electron and a neutrino come together to form a neutron.

One violation of time-reversal symmetry has been discovered in particle physics. It concerns the decay of the neutral *K* meson, and it was found through a series of experiments that began in 1964 with the work of James H. Christenson, James W. Cronin, Val L. Fitch and René Turlay of Princeton University. Since then no second instance of a time-reversal violation has been discovered. Moreover, it is still not known which of the basic forces of nature is responsible for the violation. The weak force is a likely candidate, since it is responsible for beta decay and hence for the known violation of parity symmetry. Other possibilities are electromagnetism and a hypothetical new force, the superweak force.

The observation of an electric dipole moment of any elementary particle would be evidence for another violation of time-reversal symmetry. The reasoning behind this conclusion is as follows: If the particle has an electric dipole moment, the axis connecting the electric poles must be aligned with the axis of spin rotation. If the dipole moment were perpendicular to the spin axis, so that the electric poles were on the equator, then the spinning of the particle would yield an average moment of zero, much as a black-and-white disk appears gray when it is spun rapidly. If the dipole had some orientation intermediate between parallel and perpendicular, only the component of the dipole moment that was lined up with the spin axis would be observable.

Both the spin and the electric dipole moment can be represented as vectors, or arrows. By convention the direction of the spin vector is defined by a right-hand rule: If the fingers of the right hand are wrapped around the particle in the direction in which it is spinning, then the thumb gives the direction of the vector. We shall adopt a similar convention for the electric dipole vector: it passes through the two electric poles and points in the direction of the positive charge. Suppose in some initial state the spin vector and the electric dipole vector are pointing in the same direction, say up. If time is reversed, the particle must spin in the opposite direction and the spin vector must point down. The reversal of time has no effect, however, on the electric dipole moment, any more than it has on the overall electric charge of the particle. Thus in the one state the vectors are parallel and in the time-reversed state they are antiparallel; the two states are not equivalent and time-reversal symmetry is violated. There is a sharp contrast here with the behavior of the magnetic dipole moment, which depends directly on the spin direction; if time were reversed, the magnetic dipole would change direction along with the spin vector.

The search for an electric dipole moment has focused on the neutron mainly because the necessary experiments are easier with neutrons than they are with other particles. Experiments with higher-energy neutrons have already established a limit on the maximum electric dipole moment of the neutron: if the moment exists, it must be less than 3×10^{-24} *e*-centimeter, where *e* is the charge of an electron. In 1968 Shapiro of Dubna pointed out that a still lower limit might be set in experiments with ultracold neutrons.

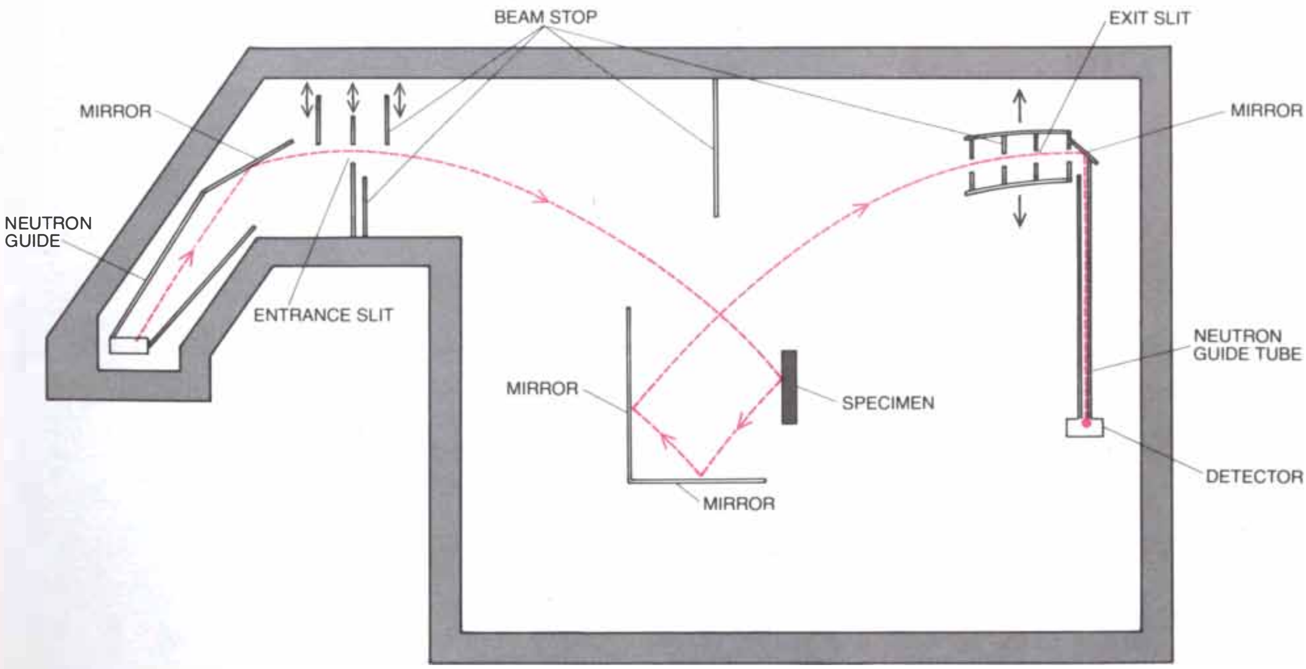
The basic experimental technique is to expose neutrons to a strong electric field; if the neutron has an electric dipole moment, its energy will depend to some small extent on the orientation of the electric dipole with respect to the applied electric field. If the neutron is made to flip end for end, the energy should change by an amount proportional to the strength of the dipole and of the external field.

In practice a somewhat roundabout technique must be adopted to make this small change in energy observable. A much larger energy difference is first established by a strong, steady magnetic field, which acts on the neutron's magnetic (not electric) dipole moment. The neutrons are then made to flip 180 degrees by an oscillating magnetic field. The frequency required to induce this spin-flip is sensitively dependent on the exact energy of the neutrons and is therefore determined in part by the interaction of any electric dipole moment with the applied electric field. If the magnitude of the electric field is changed, and if the neutron electric dipole moment exists, the resonance frequency should change. If the dipole moment is zero, the electric field should have no effect on the frequency needed to induce the spin-flip.

Any measurement of a frequency is uncertain by one oscillation, but the significance of the uncertainty diminishes as the number of cycles counted increases. Ultracold neutrons provide a sensitive test for the electric dipole moment because they allow a long observation time and hence a precise determination of the resonance frequency. V. M. Lobashov and his colleagues at the Institute for Nuclear Physics in Leningrad have recently reported the first results of a search for the neutron electric dipole moment that employed ultracold neutrons. In the experiment the neutrons were not stored but were passed slowly through the applied electric and magnetic fields for a counting interval of about four seconds. The results are preliminary, and although they did not extend the limit on the dipole moment, they did show that the technique is capable of greater precision than other methods are. An experiment we are now preparing at the Institut Laue–Langevin will employ confined neutrons and may yield a new upper limit.

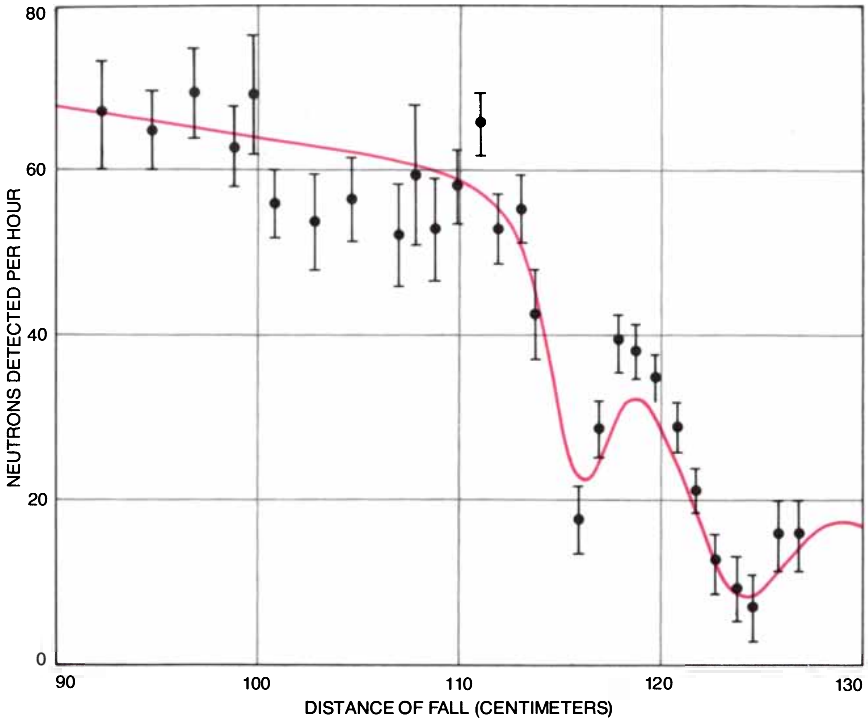
A traditional application of thermal neutrons is as a probe of atomic and molecular structures and motions, particularly in solids. Neutrons complement the X rays and other forms of electromagnetic radiation that are also employed for such studies; X rays are sensitive primarily to electrons in the solid, whereas neutrons interact mainly with atomic nuclei. Ultracold neutrons can be employed in a similar way, but their low energy and long wavelength adapt them to the examination of materials on a somewhat larger scale.

Steyerl has constructed a novel ultracold-neutron spectrometer. In an ordinary spectrometer, where the projectiles are electrically charged particles, the energy of the particles is analyzed by a magnet that bends their trajectories. In the ultracold-neutron spectrometer the role of the analyzer is played by the earth's gravitational field. The neutrons enter the spectrometer moving horizontally and then accelerate as they fall a fixed distance to a specimen. Rebounding neutrons rise as far as they can against gravitation and are then collected by an exit slit. If the interaction with the specimen is completely elastic, the rebounding neutron will rise to the same height it fell from. Any exchange of energy with the specimen will be detected



NEUTRON SPECTROMETER constructed by Albert Steyerl and his colleagues at the Technical University in Munich employs the earth's gravitational field for analyzing the energy of ultracold neutrons rebounding from a specimen. The neutrons enter the apparatus moving horizontally and then accelerate as they fall to the specimen. After reflection from the specimen and from a pair of mirrors the neutrons are allowed to rise as far as they can before they are collected by an adjustable exit slit. The height of the rebound is sensitive to exceedingly small changes in neutron energy. The device is particularly suitable for studies of thin films and of the surfaces of solids.

The GLENLIVET
AGED 12 YEARS
Discover more.
Let your mature taste lead you from the finest premium Scotch on up to the most expensive 12-year-old Scotch in the world. The Glenlivet. The ultimate in Scotch.
Most premium Scotch is blended and depends on several whiskies for taste and smoothness.
Yet the unblended character of The Glenlivet is noble enough to stand alone. In this distinctive Scotch whisky you'll experience an exceptional smoothness and full-bodied richness, unmatched in all other premium Scotch.
Try the taste beyond premium Scotch.
The Glenlivet
Unchanged since 1824.
The GLENLIVET
12 YEARS OLD
Unblended all malt Scotch Whisky
Distilled and Bottled in Scotland
GEORGE & J. G. SMITH LIMITED
THE TASTE BEYOND PREMIUM SCOTCH



SPECTRUM OF NEUTRONS reflected from a thin film of gold applied to one of the mirrors in the neutron spectrometer exhibits an interference pattern, an unmistakable indication of the wavelike nature of the neutron. The height from which the neutrons fell before striking the film was varied over a range of about 40 centimeters, which gave rise to an associated variation in the energy of the incident neutrons. Peaks in the curve, which records the number of neutrons detected at the exit slit, correspond to energies where the neutron waves interfered constructively in the gold film; at troughs in the curve the waves interfered destructively.

as a change in the maximum height of the rebounding neutrons.

Since the neutron's total energy is very small (roughly 10^{-7} electron volt), the device is exquisitely sensitive to slight changes in energy caused by interactions with the specimen. The spectrometer might provide a more precise knowledge of the motions of atoms in solids and of the forces between them. Because the ultracold neutrons penetrate only about 100 angstroms, the spectrometer should also be suitable for studies of thin films and of solid surfaces. Steyerl has demonstrated the interference of neutrons reflected from films about 1,000 angstroms thick.

The diffraction of neutrons (and of electrons and even atoms) has been demonstrated often and is the most compelling evidence of their wavelike nature. The diffraction is generally produced by passing the particles through the orderly array of atoms in a crystal, where the spacing between atoms is on the order of one angstrom. Steyerl has demonstrated the diffraction of ultracold neutrons by a ruled grating, like the gratings employed to diffract light. The experiment is a direct manifestation of a quantum-mechanical effect; it shows that the wave describing the neutron is indeed distributed over an area of some 1,000 angstroms. Such diffraction experiments could lead to a direct comparison of an optical wavelength with a neutron wavelength, and perhaps to a new determination of one of the fundamental constants of nature: the Planck constant.

Still another project Steyerl has undertaken (with G. Shütz) is the construction of a neutron microscope. The neutrons will be focused by a Fresnel lens, which depends for its effect on interference and which could not be built for shorter-wavelength neutrons. Again the instrument could complement an existing device, the electron microscope. Certain materials that are indistinguishable or transparent to electrons reveal sharp contrast when they are illuminated with neutrons.

The development of these instruments and the recent progress of experiments on the neutron lifetime and electric dipole moment suggest that a first phase in the study of ultracold neutrons may be ending. The decade since they were first detected has been dedicated mainly to learning how to extract them, store them and manipulate them. They can now be put to good use in investigations of the neutron itself and of other systems of particles.

The History of the Atlantic

The ocean is 165 million years old. Its growth and the topography of its bottom are explained by plate-tectonic theory, deep probes of the bottom and measurements of heat flow and magnetism

by John G. Sclater and Christopher Tapscott

The Atlantic Ocean did not exist 165 million years ago because the continents that now border it were joined like the pieces of a jigsaw puzzle to form one huge expanse of land. At that time molten magma rising from the interior of the earth split the land mass into the continents and began to push the continents apart at the rate of several centimeters per year, leaving the Atlantic in their wake. Over the past two decades geophysicists have mapped the topography of the floor of the Atlantic by bombarding it with sound waves and studying the echoes. The vertical structure of the ocean bottom has been determined from columnar samples of sediments collected by deep-sea drilling vessels, and measurements have been made of anomalies in the magnitude and direction of the earth's magnetic field above the bottom. The theory of plate tectonics and current theoretical models of the flow of heat out of the ocean floor can account for the topographic features, the major sedimentation trends and the magnetic anomalies. Together theory and data have enabled earth scientists to reconstruct the history of the Atlantic, particularly the bathymetric, or depth, history.

The most conspicuous topographic feature of the Atlantic is the mid-ocean ridge, a huge submerged mountain range surpassing in scale the Alps and the Himalayas combined, that runs down the middle of the ocean from Iceland in the north to Bouvet Island, 1,800 kilometers off the coast of Antarctica, in the south. The crest of the ridge marks the approximate middle of the Atlantic, and with the exception of isolated areas near Iceland and near the Azores the ocean floor at the crest has a mean depth of 2,500 meters. The floor deepens gradually and symmetrically with increasing distance from the crest, reaching a maximum depth of between 5,000 and 6,000 meters before rising abruptly at the continental shelves.

The simple north-south symmetry of the bottom of the Atlantic is broken by thin, isolated transverse ridges and by fracture zones, or deep gashes in the ocean floor, that trend away from the crest of the mid-ocean ridge in a roughly perpendicular direction. The east-west ridges rise about 2,000 meters above the surrounding ocean floor. Unlike the mid-ocean ridge, on which fall the epicenters of most of the shallow-focus earthquakes in the Atlantic, the east-west ridges are seismically inactive. The Walvis Rise and the Rio Grande Rise are the two most prominent aseismic ridges. Together they span the middle of the South Atlantic. The fracture-zone canyons (which can be as much as 500 kilometers long, 25 kilometers wide and three kilometers deep) offset the axis of the mid-ocean ridge at intervals along its entire length.

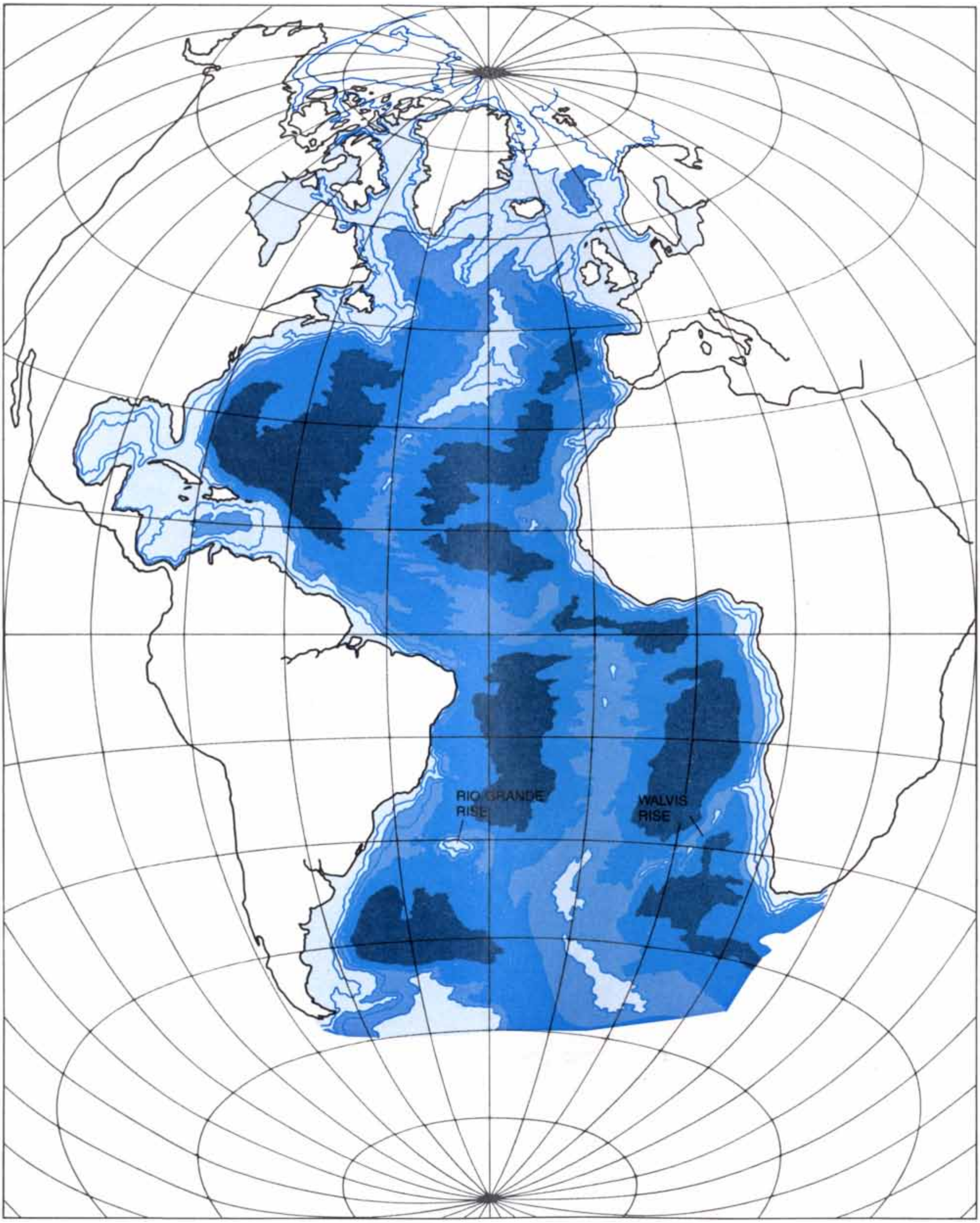
The ocean floor at the crest of the mid-ocean ridge consists mostly of hard volcanic rock. With increasing distance from the crest the rock is covered by an increasing depth of soft sediment: a combination of red clay (a light detrital material from the continents) and calcareous ooze (calcium carbonate from the decomposed skeletons of microorganisms). Near the crest of the ridge the calcareous ooze is predominant. Between 4.5 and five kilometers below the ocean surface is the calcium carbonate compensation depth: the level below which the pelagic snowfall of calcareous sediments dissolves rather than accumulates because the water is undersaturated with calcium carbonate. This means that in deep water far from the crest of the mid-ocean ridge the calcium carbonate sediments dissolve, leaving the red clay as the predominant material on the ocean floor.

Near the continents the sediments are covered by layers of detrital material that has spilled from the continental shelves. Only there is the ocean floor extremely smooth, the thick layers of sediments forming flat abyssal plains. The greatest accumulation of sediment is found at the base of the continents. The sediments are deposited by rivers and consist primarily of fine particles of clay and sand interspersed with calcareous ooze and other materials of organic origin.

Since 1968 the deep-sea drilling vessel *Glomar Challenger,* built for a consortium of American oceanographic institutions, has gathered columnar samples of sediments at hundreds of sites on the ocean floor. One of the most surprising results was the discovery of calcium carbonate sediments in samples taken near the continental shelves. Thin layers of such sediments, found just above hard volcanic rock, were covered by as much as a kilometer of red clay and other sedimentary products of continental erosion. The calcium carbonate was between six and seven kilometers below the ocean surface. Since the compensation depth is substantially less, the presence of the calcium carbonate was puzzling. The puzzle was solved when the bathymetric history of the Atlantic was reconstructed.

Many of the large-scale geophysical phenomena of the ocean floor can be accounted for by the theory of plate tectonics, in which the surface of the earth is regarded as a set of rigid plates forming the lithosphere. The plates vary in size from the six major ones, such as the plate that underlies virtually the entire Pacific, to several small ones, such as the plate that is essentially coextensive with Turkey. The lithosphere is some 100 kilometers thick and comprises the crust and upper mantle of the earth. It floats on the asthenosphere, a mobile and fairly fluid layer of the mantle a few hundred kilometers thick. The lithospheric plates float on the asthenosphere somewhat like hardened wax on a pool of molten wax. The plates are always moving with respect to one another and hence are always interacting at their boundaries.

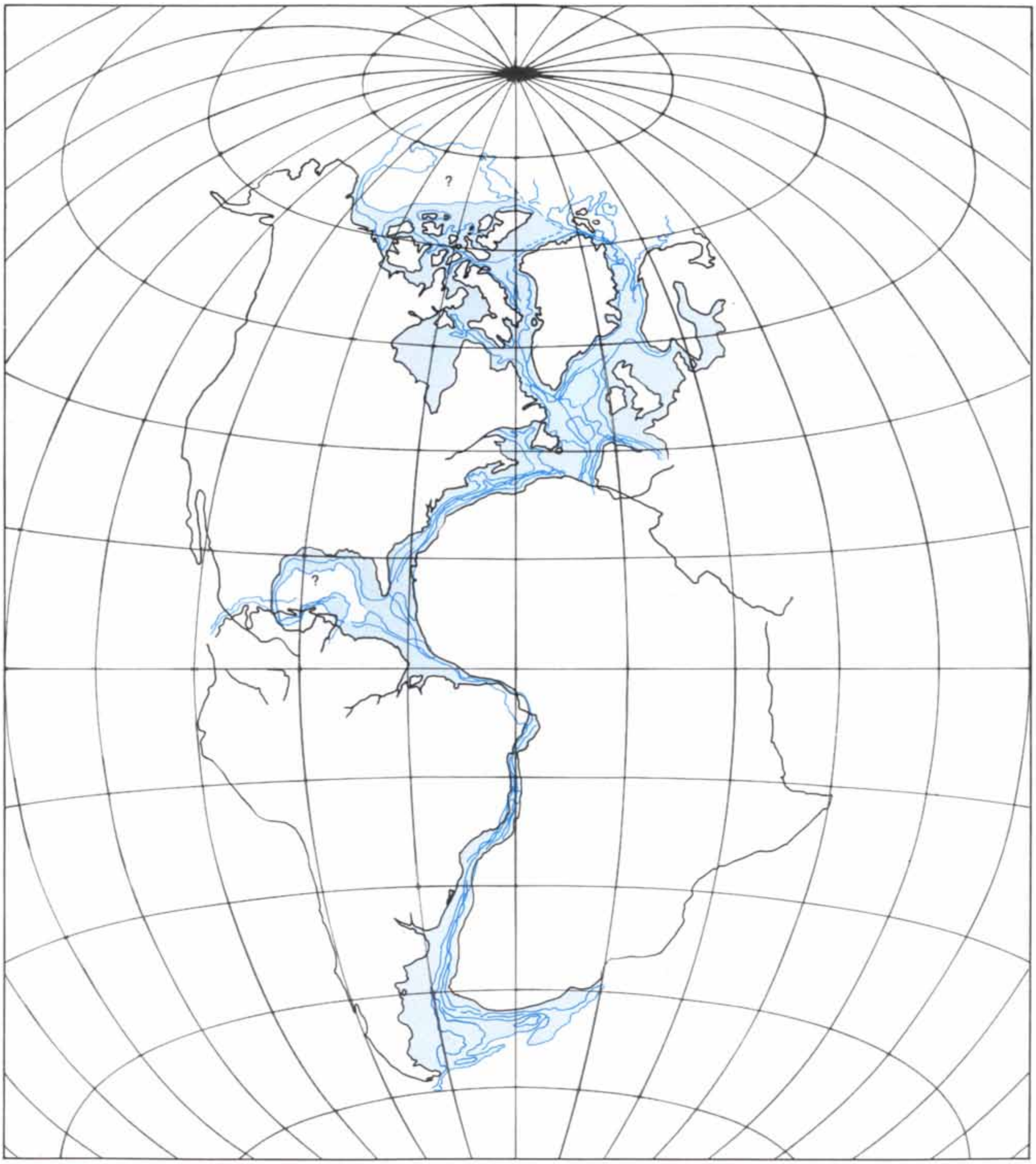
The plates can interact in three ways: they can diverge, or move apart from each other (as they do on opposite sides of the mid-Atlantic ridge), they can converge, or collide, and they can slide past each other. At the mid-Atlantic ridge molten magma continuously wells up in



0 – 3,000 METERS
3,000 – 4,000 METERS
4,000 – 5,000 METERS
5,000 – 6,000 METERS

BATHYMETRIC CONTOUR CHART OF THE ATLANTIC shows that the ocean floor deepens symmetrically with increasing distance from the mid-ocean ridge, the huge submerged mountain range that runs down the middle of the ocean from Iceland in the north to Bouvet Island, 1,800 kilometers off the coast of Antarctica, in the south. The ocean floor, which has a mean depth of 2,500 meters at the crest, reaches a maximum depth of between 5,000 and 6,000 meters before rising abruptly at the shelves of the continents. Thin aseismic ridges and fracture zones, or deep gashes in the floor of the ocean, break up the north-south symmetry of the ocean floor. The Walvis Rise and the Rio Grande Rise are the two most prominent aseismic ridges.

the gap between the diverging plates, filling the gap to create new ocean floor. As the outflowing magma cools, increases in density and sinks, it forms new lithosphere that adds to the broad, spreading swell of the ridge. Where two plates collide one may be subducted, or plunge under the other, and be reabsorbed in the mantle. Along the line where the plate descends into the mantle a deep trench forms in the ocean bottom. The melting of the subducted plate generates intense volcanic activity along the edge of the overriding plate about 100 kilometers from the trench. Such volcanic activity in conjunction with the collision of continents has been responsible for the creation of most mountain ranges. The plunging of the subducted plates also gives rise to most of the world's earthquakes, which therefore

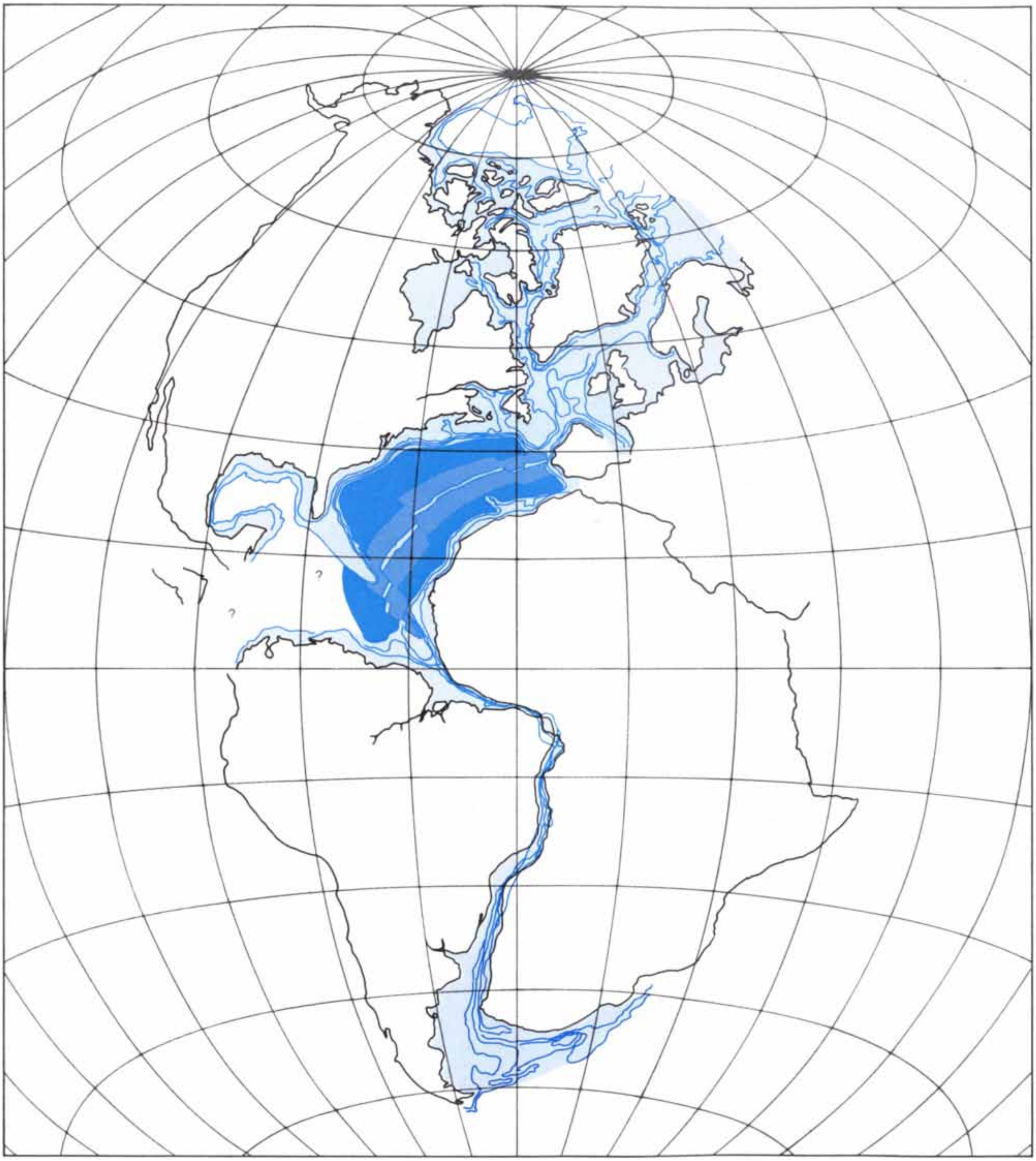


0 – 3,000 METERS
3,000 – 4,000 METERS
4,000 – 5,000 METERS

MAP OF THE CONTINENTS 165 MILLION YEARS AGO (*left*) shows that North America, South America, Greenland, Europe and Africa, which had formed one huge expanse of land, started to rift apart as molten magma from the deep interior of the earth welled up between them. The creation of new crust by the outward flow of magma separated the northern continents from Africa and South America, opening up the infant North Atlantic and Caribbean. By 125 million years ago (*right*) the North Atlantic had reached maximum depths of

are beacons marking the boundaries of the plates.

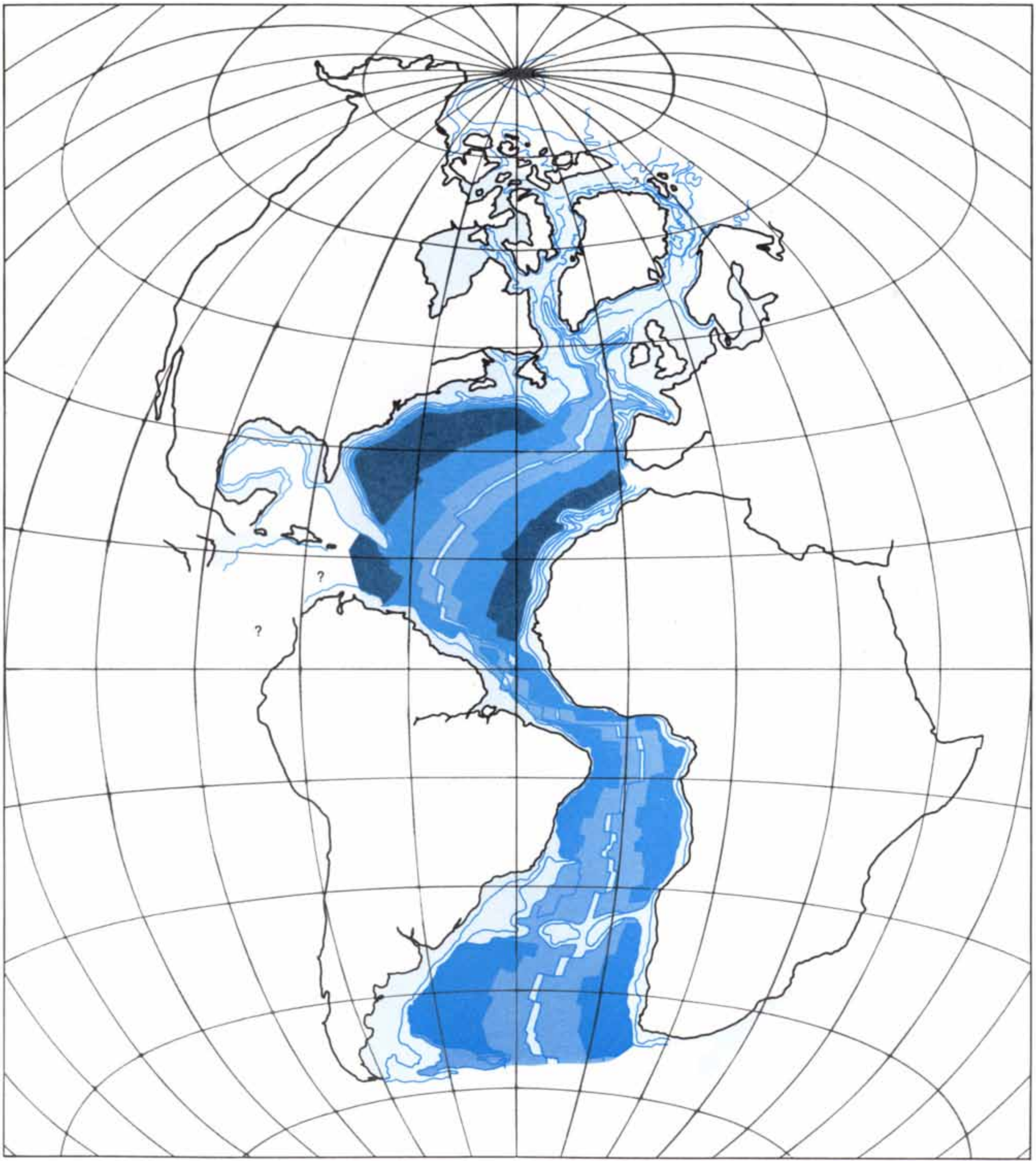
The floor of the Atlantic has been described as a giant conveyor belt that transports lithosphere at the rate of a few centimeters a year from its creation at the mid-ocean ridge to its disappearance in the deep trenches. This description calls attention to the fact that the ocean floor ages with increasing distance from the crest of the mid-ocean ridge. The motion of the ocean floor is driven by thermal convection in the mantle. This motion provides the visible half of a convection loop; the other half lies deep in the interior of the earth, where the mass of the subducted plates must be conserved by some kind of return flow of matter from the trenches to the mid-ocean ridge. The details of the



4,000 meters. Water could not yet flow to the north because the land north of Gibraltar had not yet rifted and because the channel between Spain and Africa was not deep enough. Water probably also could not flow to the south because coral reefs that had built up on top of the Bahama platform may have blocked circulation between the Atlantic and the Caribbean. (Water could nonetheless flow through the Caribbean into the Pacific because parts of what is now Central America were below sea level.) About 125 million years ago South America began to separate from Africa, and North America continued to move away from Africa and Spain. The plate north of Venezuela was subducted under the South American plate as the motion of South America with respect to North America compressed the Caribbean region.

return flow are poorly understood and are therefore the subject of continuing investigation.

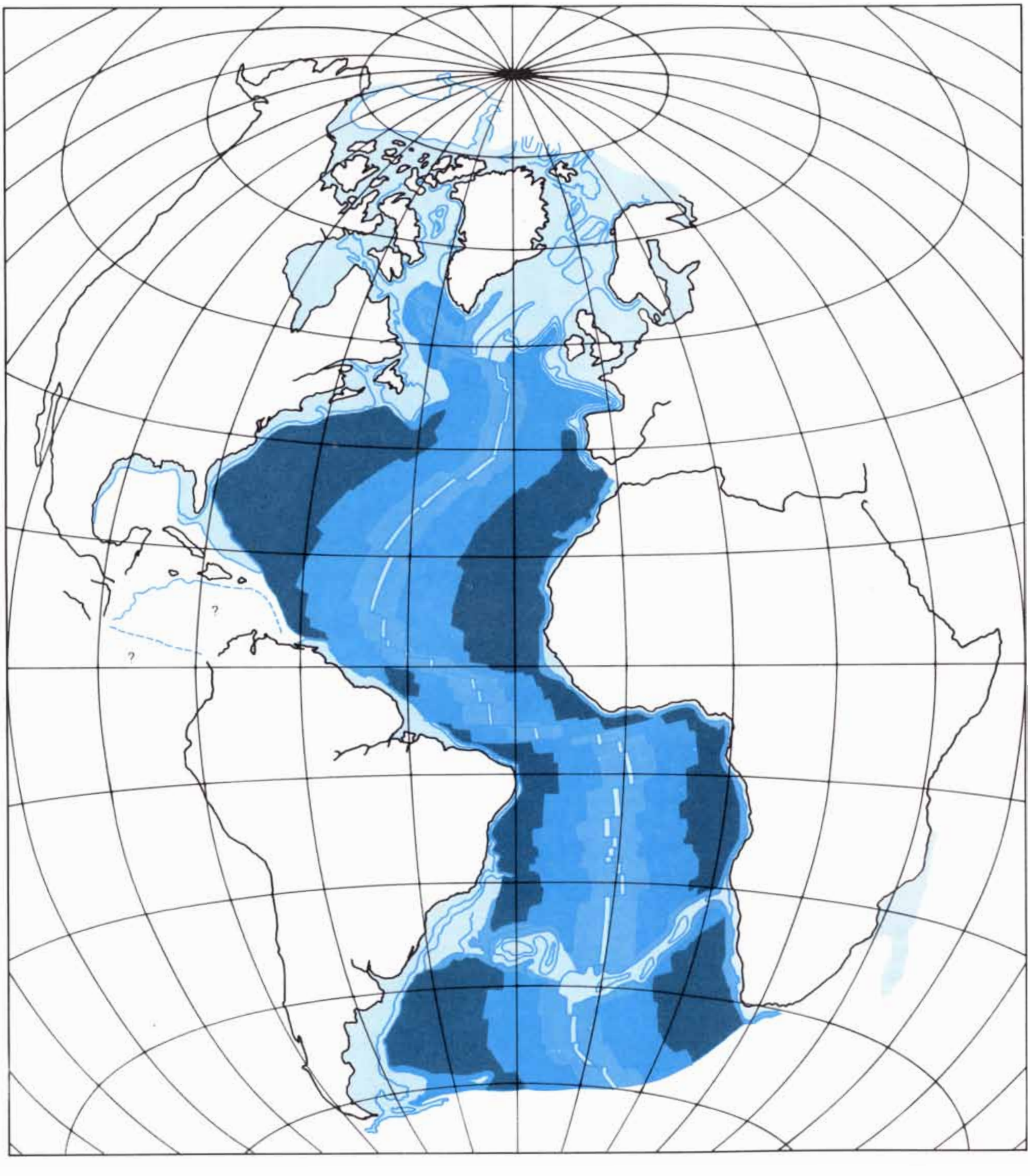
The lithosphere can be regarded as being the upper thermal-boundary layer of the mantle convection system. This boundary model leads to far-reaching predictions about the flow of heat out of the ocean floor and about the depth of the ocean. The predictions have largely turned out to be true. Consider the lithosphere of the Atlantic in vertical section. Created at high temperatures at the center of the mid-Atlantic ridge, the plate cools first at its upper surface, where it is in contact with seawater at a temperature of about zero degrees Celsius. As heat flows out of the upper surface into the ocean the plate cools and thickens. The theoretical description of thermal conduction yields a simple equation that



MAP OF THE ATLANTIC 80 MILLION YEARS AGO (*left*) shows that the North Atlantic was then a full-fledged ocean reaching depths of more than 5,000 meters. Its water could flow into other oceans through the Caribbean and through the wide Strait of Gibraltar. Greenland and North America had begun to drift apart, leaving in their wake narrow, shallow seas jutting out from the main body of the North Atlantic. The South Atlantic still had not developed to its

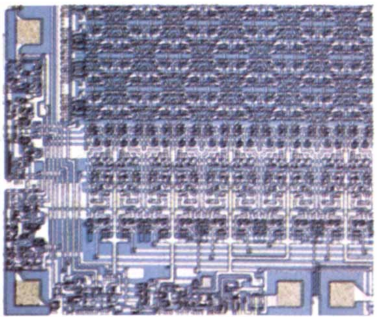
predicts the flow of heat through the plate. The equation in turn yields the expression $q = 11.3/\sqrt{t}$ for the flow of heat (q) out of the plate into the ocean as a function of the age (t) of the ocean floor. The dimension of q is the heat-flow unit, or 10^{-6} calorie per centimeter squared per second; the dimension of t is a million years.

The validity of the heat-flow expression can be tested by measuring q at various locations on the ocean floor and comparing the measured values with the predicted ones. Such measurements of the flow of heat are difficult to make, however, because of the thermal effects of other phenomena. Vertical cracks and faults at the axis of the mid-Atlantic ridge fracture the rapidly cooling magma, so that the new lithosphere is highly permeable to water. Hence thermal



full extent. Reaching a maximum depth of only about 4,000 meters, it was split into northern and southern basins by the Rio Grande Rise and the Walvis Rise. The rises strongly limited circulation between the two basins. By 36 million years ago (*right*) most of the major topographical features of the Atlantic had formed. Deep water could circulate in the South Atlantic because the Walvis and Rio Grande rises had subsided sufficiently. The major difference between then and now is that Spain was farther from Africa. Over the past 36 million years Africa and Europe have slowly come together, almost entirely closing the passage between the Atlantic and the Mediterranean.

High throughput, high yield, high reliability: Manufacturing integrated circuits with the Perkin-Elmer Micralign 200 Series



The crux of microcircuit manufacture is the printing of tiny, complicated electrical patterns on photosensitive materials. The challenge is to reproduce these patterns crisply and accurately, in production quantities, at competitive prices. This is exactly what Perkin-Elmer Micralign mask aligners do.

The newest Micralign, the Model 200 Series, is responsive to the semiconductor industry's needs. It achieves 2-micron geometries or better in production, distortion/magnification tolerance of .25 micron and four percent uniformity of illumination.

Like other Micralign systems, the Model 200 Series uses the 1:1 optical projection concept, pioneered by Perkin-Elmer, to focus light from a mask. The mask contains many repeats of a single pattern. In one fast scan the Model 200 Series exposes all of these patterns onto a silicon wafer coated with a photosensitive material. After development, the wafer is further processed—etched, doped, and recoated. By repeating this procedure a number of times circuits are built, layer on layer, on a single wafer. Finally the wafer is cut to separate each individual circuit.

Because dust, heat and vibration are major enemies of precise projection, the 200 houses its optics in a quiet, clean, wear-free world of their own. Vibration is minimized by two frames, one inside the other. All vibrating components have been mounted to the outer frame, thereby isolating the projection optics from all sources of vibration.

The 200 Series has a built-in environmental chamber. External air is blown through the top of the unit. The air is filtered and temperature regulated. A positive-pressure, Class 100 environment is carefully controlled to better than 1°F.

Options include automatic wafer loading and a revolutionary automatic alignment system. Deep UV coatings for even smaller geometries will be available soon.

"Sputtering" thin films onto microcircuits quickly, uniformly, economically: The Perkin-Elmer 4410

The thin metal films used in making semiconductors are deposited by sputtering systems. Evacuate a chamber; fill it with argon; place a high-potential cathode on the ceiling of the chamber and a lazy susan loaded with silicon wafers on the floor; attach to the cathode a target of the material to be deposited; turn on the current.

That, in essence, is sputtering. The current ionizes the argon, and the argon ions in their efforts to reach the cathode bombard the target material, knocking off atoms which settle, uniformly dispersed, on the wafers circling below. Layer by layer, the film builds up to the desired thickness.

Perkin-Elmer sputtering equipment has long set industry standards worldwide. Our latest unit, for example, the 4410, features a novel delta-shaped DC magnetron cathode and microprocessor controller. Both play an important role in producing the high quality films and high throughput essential for the economic production of semiconductors.

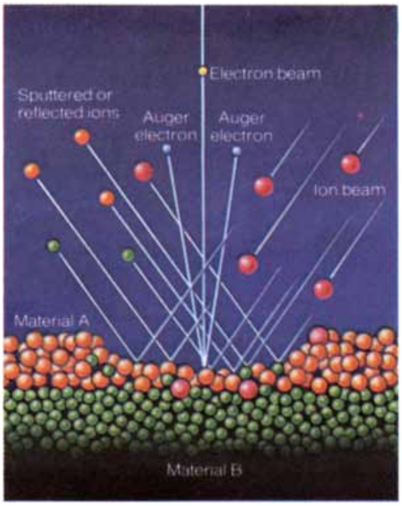
The Delta™ cathode deposits aluminum alloy and other metallic films at very high rates. This is important for building dependable microcircuits today. It's even more important for the high density semiconductor devices of the not-too-distant future.

The microprocessor automatically controls all process variables and eliminates human error. It thus assures run-to-run repeatability for high yield and throughput.

All of which help make the 4410's return-on-investment better than that of any competitive system.

The Perkin-Elmer SAM 590: Probing semiconductor quality layer by atomic layer

As the Micralign system is the world's leading producer of micro-electronic surfaces, so the leading tool for analyzing such surfaces is the Perkin-Elmer 590 Scanning Auger (pronounced Oh-djay) Microprobe, familiarly known as SAM.



Three-dimensional materials characterization using Auger analysis with ion sputter-etching.

Typical tasks for SAM include evaluating the integrity of a semiconductor's thin film interfaces and checking to be certain impurities were not introduced during manufacture. Or detecting such imperfections as microscopic aluminum spikes between semiconductor layers which can cause short circuiting. Or scrutinizing bonding pads for trace contaminants which cause poor bonding adhesion.

To do such jobs, SAM bombards a specimen with a beam of electrons, causing the emission of X-ray photons and chemically specific Auger electrons. These Auger electrons originate in the topmost two or three atomic layers of a specimen surface. By measuring the kinetic energy and number of Auger electrons emitted, SAM provides a quantitative as well as qualitative identification of surface constituents. If the electron beam is scanned, SAM can map the distribution of chemical elements over a selected area.

SAM incorporates an ion bombardment gun that continuously erodes an area of the surface so that the microprobe can analyze downward, layer by atomic layer. This makes possible a true three-dimensional analysis of thin films.

Where applications require, SAM can be expanded into a multiple-technique instrument through addition of ESCA (Electron Spectroscopy for Chemical Analysis), which measures electrons released due to X-ray stimulation of a surface, and SIMS (Secondary Ion Mass Spectroscopy), which detects ions emitted after primary ion bombardment.

For more information on these products, please write: Corporate Communications, Perkin-Elmer, Main Ave., Norwalk, CT 06856.

PERKIN-ELMER
Responsive Technology

Physician, did you miss any of these significant developments in medical science?

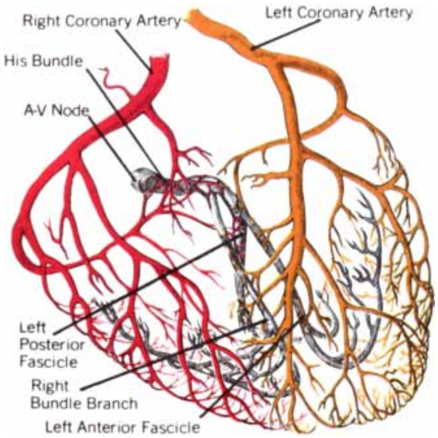
• The lidocaine level in the blood may fall below the therapeutic range when the drug is given as a single bolus followed by a continuous infusion. To avoid a "therapeutic hiatus," a second bolus should be given 20 to 30 minutes after the first bolus...

• Evidence is mounting that there is widespread over-diagnosis of pulmonary embolism at a time when heparin, in therapeutic doses, has become an important cause of drug-related death in the relatively well, hospitalized patient...

• The widely used street drug, phencyclidine, appears to cause persisting and possibly permanent brain dysfunction after repetitive use...

• There is alarming incidence of acute leukemia which may reach five percent in patients who appear to have been successfully treated with chemotherapy or radiation therapy for Hodgkin's disease...

Every day medical science evolves in ways that affect patient management and therefore patient well-being. But with more than 2,000 journals published, innumerable meetings held and hundreds of courses given, it is all too easy for developments that compel change in diagnosis and treatment to get by even the most conscientious reader and meeting-goer. And since such discoveries are today's news, they will not appear in the standard texts for years.



Branches of the right and left coronary arteries supply blood to the AV node and intraventricular conduction system.

SCIENTIFIC AMERICAN *Medicine* is lucidly illustrated with drawings and photographs. Some examples are seen here and on the facing page.

We have a solution to this problem: It is called SCIENTIFIC AMERICAN *Medicine*. It makes such developments accessible to you when and where you need them, month after month.

Here is how SCIENTIFIC AMERICAN *Medicine* works. When you subscribe, you receive a two-volume, 2,000 page text, organized according to the familiar subspecialties so that you know where to find the answers to your questions.

The authors of SCIENTIFIC AMERICAN *Medicine* are leading authorities from the faculties of the Harvard Medical School and the Stanford University School of Medicine. These seasoned scholar-practitioners, close to the creation of new medical knowledge, write for you the practitioner, not the student. Therefore, at the time you receive your copy of SCIENTIFIC AMERICAN *Medicine*, you will hold in your hands the world's most advanced text of medicine. As long as you continue to subscribe it will stay that way.

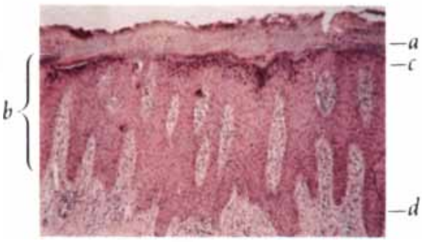
Each month you will receive several updated chapters. The new chapters which replace the old chapters contain reports of the significant advances in medical science that affect patient management. The index will also be updated each month. Furthermore, updated material will be summarized in a monthly bulletin accompanying the updated chapters so that you can quickly scan the information.

As the updated material is incorporated into the text, older more familiar material is deleted. Thus the amount of useful knowledge increases but the number of pages does not.

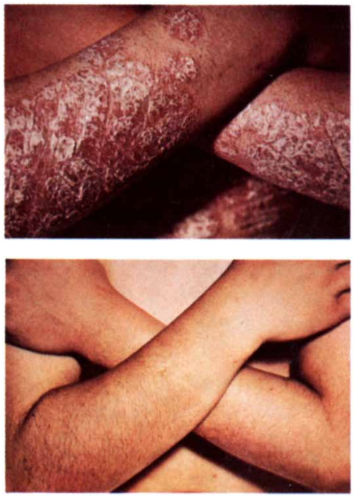
A Continuing Medical Education Program

For subscribers who wish it, SCIENTIFIC AMERICAN *Medicine* offers at no extra charge a continuing medical education program sponsored by the Stanford University School of Medicine's Office of Postgraduate Medical Education and administered by the Educational Testing Service of Princeton, New Jersey. The program consists of eight patient management problems per year, each written by a member of a Panel of Examiners drawn from leading clinical institutions. By completing all eight PMP's in a 12-month period the subscriber can earn 32 Category 1 credits per year, from the Stanford University School of Medicine, toward the AMA's Physician's Recognition Award.

This complete, concise medical knowledge service is yours for the drop of the accompanying card. Please fill it out and mail it to us. We think you will find that SCIENTIFIC AMERICAN *Medicine* is your answer to the challenge of bringing all that modern medicine has to offer to your patients.



Histopathologic changes in psoriasis.



Extensive psoriasis (top) cleared completely (bottom) after four weeks of treatment with anthralin.

THE DISTINGUISHED AUTHORS, AND THE FIFTEEN SECTIONS OF SCIENTIFIC AMERICAN *MEDICINE*.

1. Cardiovascular Medicine

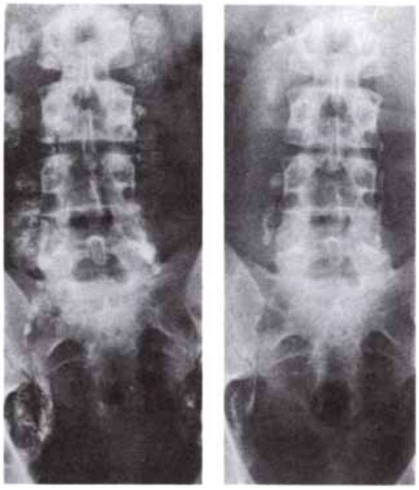
Edgar Haber, M.D., Professor of Medicine, Harvard Medical School, Physician and Chief of Cardiology, Massachusetts General Hospital

E. William Hancock, M.D., F.A.C.P., Professor of Medicine, Stanford University School of Medicine

Roman W. DeSanctis, M.D., Professor of Medicine, Harvard Medical School, and Physician and Director, Coronary Care Unit, Massachusetts General Hospital

Adolph M. Hutter, Jr., M.D., F.A.C.P., Associate Professor of Medicine, Harvard Medical School, and Assistant Physician and Associate Director, Coronary Care Unit, Massachusetts General Hospital

Eve Elizabeth Slater, M.D., Instructor in



Extensive retroperitoneal Hodgkin's disease before and after MOPP chemotherapy.

Medicine, Harvard Medical School, and Chief, Hypertension Unit, Massachusetts General Hospital

2. Dermatology

Eugene M. Farber, M.D., Professor and Chairman, Department of Dermatology, Stanford University School of Medicine

3. Endocrinology

Daniel D. Federman, M.D., F.A.C.P., Professor of Medicine and Dean for Students and Alumni, Harvard Medical School

4. Gastroenterology

Gary M. Gray, M.D., Professor of Medicine and Head, Division of Gastroenterology, Stanford University School of Medicine

5. Hematology

Stanley L. Schrier, M.D., Professor of Medicine and Head, Division of Hematology, Stanford University School of Medicine

6. Immunology

John David, M.D., Professor of Medicine, Harvard Medical School, and Physician, Robert B. Brigham Hospital

7. Infectious Disease

Thomas C. Merigan, M.D., Professor of Medicine and Head, Division of Infectious Diseases, Stanford University School of Medicine

Morton N. Swartz, M.D., F.A.C.P., Professor of Medicine, Harvard Medical School, and Chief, Infectious Disease Unit, Massachusetts General Hospital

Cyrus C. Hopkins, M.D., Assistant Professor of Medicine, Harvard Medical School, and Hospital Epidemiologist and Associate Physician, Massachusetts General Hospital.

Adolf W. Karchmer, M.D., F.A.C.P., Assistant Professor of Medicine, Harvard Medical School, and Associate Physician, Massachusetts General Hospital

Robert H. Rubin, M.D., Assistant Professor of Medicine, Harvard Medical School, and Assistant Physician, Massachusetts General Hospital

Harvey B. Simon, M.D., F.A.C.P., Assistant Professor of Medicine, Harvard Medical School, and Assistant Physician, Massachusetts General Hospital

8. Intensive And Emergency Care

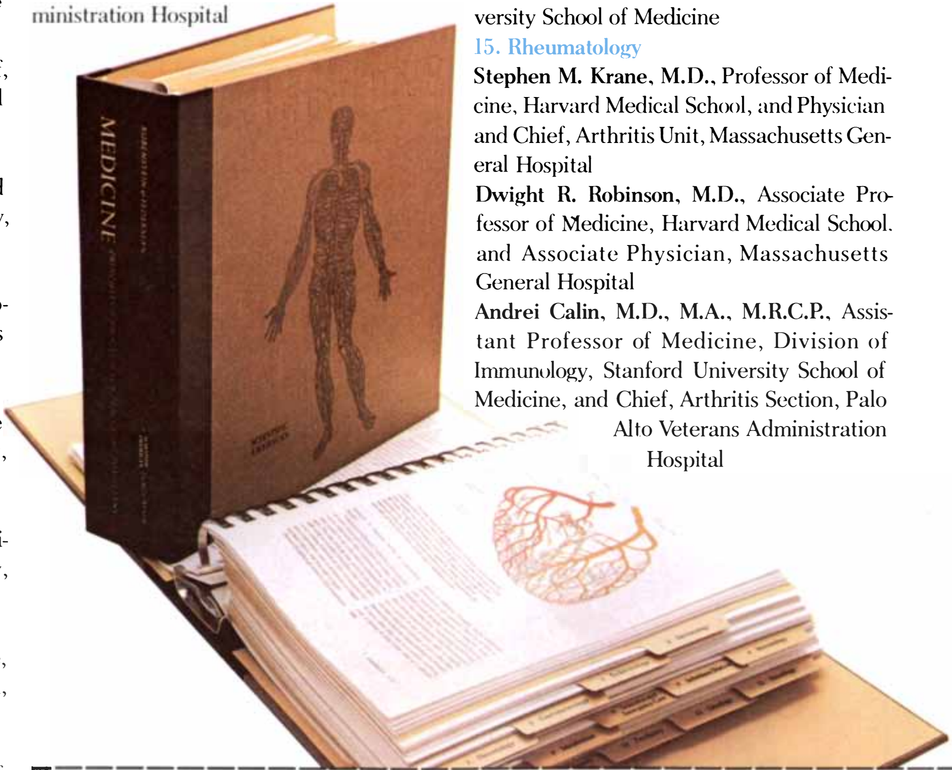
Edward Rubenstein, M.D., F.A.C.P., Associate Dean of Postgraduate Medical Education, and Professor of Clinical Medicine, Stanford University School of Medicine

9. Metabolism

George F. Cahill, Jr., M.D., Professor of Medicine, Harvard Medical School, Director of Research, Howard Hughes Medical Institute, and Physician, Peter Bent Brigham Hospital

10. Nephrology

Roy H. Maffly, M.D., Professor of Medicine, Stanford University School of Medicine, and Chief, Renal Service, Palo Alto Veterans Administration Hospital



11. Neurology

Robert W.P. Cutler, M.D., Professor of Neurology, Stanford University School of Medicine

12. Oncology

Saul A. Rosenberg, M.D., F.A.C.P., Professor of Medicine and Radiology, and Chief, Division of Oncology, Stanford University School of Medicine

13. Psychiatry

Ned H. Cassem, M.D., Associate Professor of Psychiatry, Harvard Medical School, and Chief, Psychiatric Consultation—Liaison Service, Massachusetts General Hospital

14. Respiratory Medicine

Eugene D. Robin, M.D., F.A.C.P., Professor of Medicine and Physiology, Stanford University School of Medicine

15. Rheumatology

Stephen M. Krane, M.D., Professor of Medicine, Harvard Medical School, and Physician and Chief, Arthritis Unit, Massachusetts General Hospital

Dwight R. Robinson, M.D., Associate Professor of Medicine, Harvard Medical School, and Associate Physician, Massachusetts General Hospital

Andrei Calin, M.D., M.A., M.R.C.P., Assistant Professor of Medicine, Division of Immunology, Stanford University School of Medicine, and Chief, Arthritis Section, Palo Alto Veterans Administration Hospital

SCIENTIFIC AMERICAN MEDICINE

Post Office Box 5911, New York, N.Y. 10017

Gentlemen: Please enroll me as a subscriber to SCIENTIFIC AMERICAN *Medicine*. On receipt of this coupon you will send me the advanced two-volume text described in your announcement, and update it regularly by sending me new monthly subsections. As part of this service, I understand that the annual subscription of $150 for SCIENTIFIC AMERICAN *Medicine* is tax-deductible. If I am not entirely satisfied, I may cancel at any time during the first 60 days, returning all materials for a complete refund. SA6-79

☐ Please enter my subscription for SCIENTIFIC AMERICAN *Medicine*.
☐ I shall also enroll in the CME Program
☐ I enclose a check made out to SCIENTIFIC AMERICAN *Medicine* for $150
☐ Please bill me

Name ____________________

MD Specialty ____________________

Address ____________________

City ____________ State ____________ Zip ____________

Subscribers outside of the U.S., its possessions, and Canada will be charged extra for shipping and handling; updates will be sent by surface routes unless air mail delivery is requested.

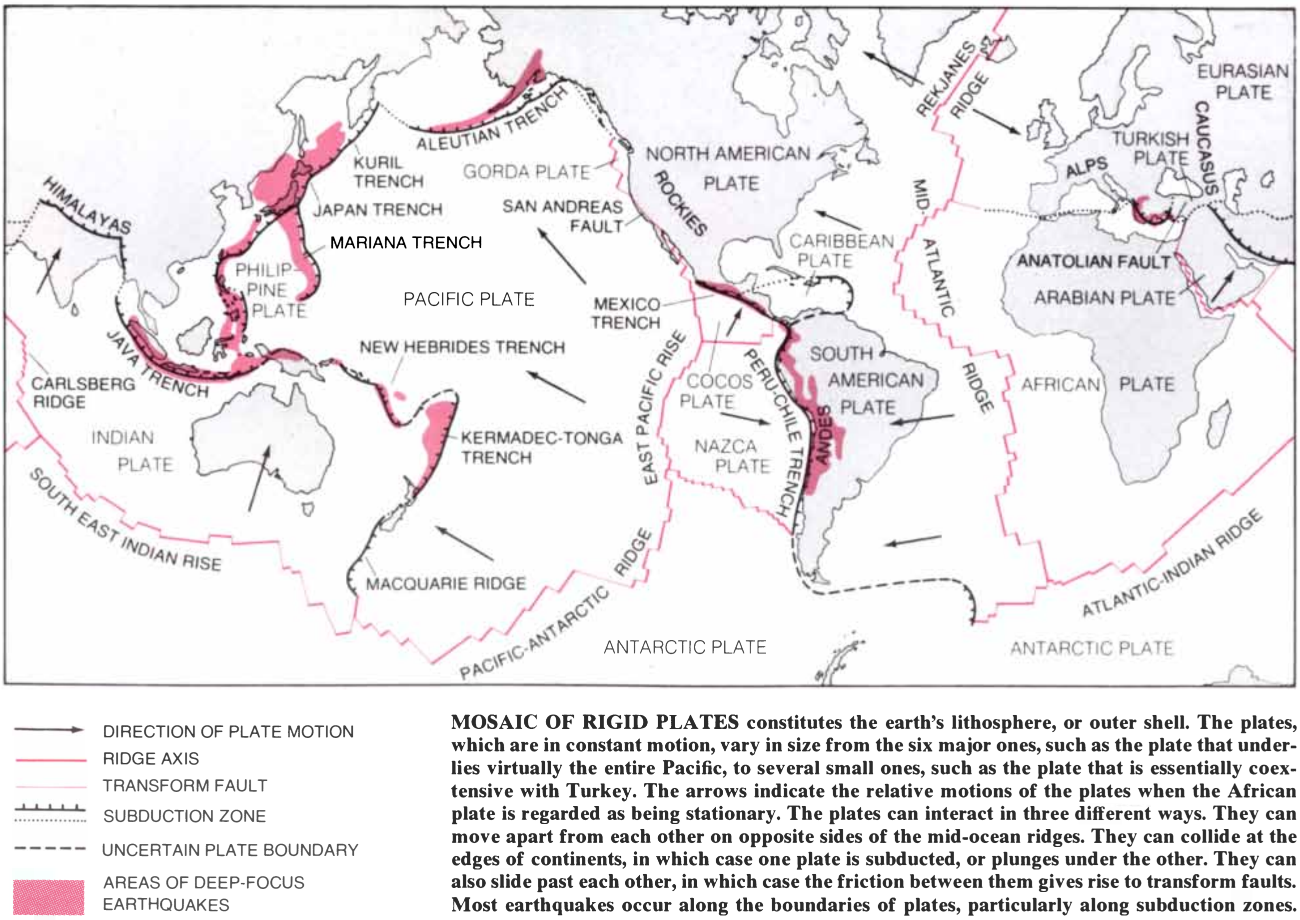
energy can be transported through the ocean-bottom crust by advection: the thermally driven flow of water. Near the axis of the ridge more heat is lost through advection than through conduction. At substantial distances from the center of the ridge advection is negligible. There the permeable layer of lithosphere has been covered by a layer of sediment that is impermeable to water, so that the flow of heat out of the ocean floor is mainly due to conduction. Measurements made at such locations are in close agreement with the predicted values. They show that the expression $q = 11.3/\sqrt{t}$ correctly predicts the flow of heat for lithosphere that is less than 120 million years old. In older lithosphere additional heat originating in the mantle makes the actual heat flow greater than the heat flow predicted by the boundary model.

As the lithosphere cools it undergoes thermal contraction. The rate of vertical contraction can be calculated from the heat flow. This contraction, in conjunction with the gravitational-loading effect of the ocean water, is responsible for the depth of the ocean floor. Again theory yields a simple mathematical expression: $D = 2{,}500 + 350\sqrt{t}$, where D is the depth in meters and t is the age of the floor in millions of years. Empirical data indicate that this expression, like the one for heat flow, yields correct values for lithosphere no older than 60 million years. As we have mentioned, the axis of the mid-Atlantic ridge is 2,500 meters below sea level. The 3,000-meter isobath, or contour of equal depth, lies on ocean-bottom crust that is two million years old, the 4,000-meter isobath on crust that is 20 million years old and the 5,000-meter isobath on crust that is 50 million years old.

The expressions for the heat flow and the depth were developed on the basis of the model that treats the lithosphere as the upper thermal boundary of the mantle convection system. A different model makes the additional assumption that heat flows out of the mantle into the lithosphere to maintain an isotherm, or surface of equal temperature, of 1,400 degrees C. at a depth of about 125 kilometers. Below the isotherm the mantle material is less solid, and so the model agrees with plate-tectonic theory in regarding the lithosphere as solid material (plates) floating on less solid material. This plate model predicts correctly the heat flow and the depth of ocean floor of any age. The plate model postulates strong heating from below the lithosphere in order to account for the lower rate of flattening of ocean floor that is more than 60 million years old. A source of heat in the mantle would decrease the rate at which the older lithosphere cools and hence decrease the rate at which the lithosphere contracts and causes the ocean floor to lie deeper.

The rate at which plates are diverging from each other on opposite sides of the mid-ocean ridge can be determined from the magnetic characteristics of the ocean floor. As magma wells up at the ridge and cools to form new ocean floor, it becomes magnetized in the direction of the prevailing magnetic field of the earth. The geomagnetic field reverses in polarity at random intervals of about a million years, so that the ocean floor, as it moves ponderously away from the ridge, is magnetized in a series of long linear stripes alternating in polarity and running parallel to the ridge axis. These stripes of alternating polarity give rise to magnetic anomalies that perturb the local field by about 1 percent.

Geophysicists have calculated the perturbation by measuring the perturbed field and subtracting from the measurement the value the field would have in the absence of the local magnetic bodies. From such calculations the width of the polarized stripes has been determined. The field reversals over the past five million years have been dated from the "signatures" they left in flows of lava on dry land. Such dates have provided the ages of the recent ocean-floor stripes, and the rate of separation of two plates can be calculated directly



MOSAIC OF RIGID PLATES constitutes the earth's lithosphere, or outer shell. The plates, which are in constant motion, vary in size from the six major ones, such as the plate that underlies virtually the entire Pacific, to several small ones, such as the plate that is essentially coextensive with Turkey. The arrows indicate the relative motions of the plates when the African plate is regarded as being stationary. The plates can interact in three different ways. They can move apart from each other on opposite sides of the mid-ocean ridges. They can collide at the edges of continents, in which case one plate is subducted, or plunges under the other. They can also slide past each other, in which case the friction between them gives rise to transform faults. Most earthquakes occur along the boundaries of plates, particularly along subduction zones.

HOW THE BELL SYSTEM OVERSEES 40 MILLION LONG DISTANCE CALLS A DAY.

ON AN EASY DAY.

The Network Operations Center, Bedminster, New Jersey.

Part of the network status board, where NOC personnel watch for possible jam-ups.

You are looking at the Bell System's Network Operations Center. Here, our technology and people work 24 hours a day to help your long distance calls go through quickly, effortlessly.

When you make a long distance call, it has several different routes it can take, automatically.

But sometimes traffic gets particularly heavy. We can get a bottleneck.

That's when the people of the Network Operations Center move in. Using the most advanced Bell System computer technology, they re-route the traffic to get your call through.

In round numbers, the Network Operations Center helps manage nearly 40 million calls, on a normal day. At busy times on busy days, the volume surges even higher.

So come Christmas or Mother's Day, hurricane or high water, virtually every long distance call you make goes through quickly and easily.

Thanks to all the people of the Bell System.

No matter when you make your long distance call, the NOC stands ready to help it get through without a hitch.

Bell System

Keeping your communications system the best in the world.

The Fisher CR5150 cassette deck. Gorgeous up close. Even better from a distance.

Great styling and state-of-the-art performance are two things this new Fisher cassette deck has plenty of.

But it's got something even more exciting: full-function remote control—*without wires!* It's the first tape deck ever to offer this sensational feature.

Think of it: by touching a button on the remote infra-red transmitter, you can control Play, Record, Pause, Stop, Fast Forward, and Rewind modes—from up to 20 feet away! You can record, edit, search, and listen to cassettes—without leaving your easy chair. And the CR5150 is just plain fun to operate.

Wireless control would probably make the CR5150 a big seller even if its performance was only average. But Fisher went all out, and gave it 3 heads for 30-19,000 Hz response, dual-process Dolby** for 68dB S/N ratio, and a servo-controlled transport with 0.04% wow & flutter (WRMS). Superb specs that only a handful of ultra-high priced cassette decks can match.

Feature-wise, there's a built-in digital clock that will turn on the CR5150 deck (or your receiver) to record anything you want at a preset time, whether you're home or not. The clock display doubles as an electronic tape counter with memory rewind. Silky-smooth, feather-touch buttons control the solenoid tape mechanism.

But considering the prices of other decks with similar performance and far less features, the Fisher CR5150 at $650* has to be one of the greatest values in high fidelity today. No matter how you look at it. Available at better audio stores or the audio departments of fine department stores.

*Manufacturer's suggested retail value. Actual selling price determined solely by the individual Fisher dealer.

New guide to buying high fidelity equipment. Send $2 for Fisher Handbook, with name and address to Fisher Corp., Dept. H, 21314 Lassen St., Chatsworth, CA 91311.

FISHER

The first name in high fidelity.®

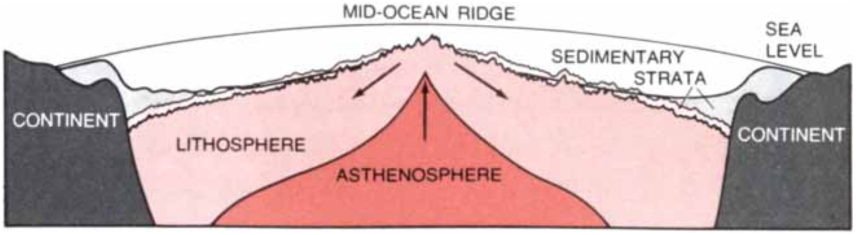
from the ages and widths of the stripes. Moreover, it has been possible to date the field reversals over the past 150 million years by comparing the magnetic anomalies of different areas of the ocean floor. In this way the ages of all the ocean-floor stripes have been determined.

Knowledge about the past motion of the ocean floor also comes from the transform faults that are left when plates slide past each other. The friction between the plates gives rise to strong shearing forces that wrench the ocean floor into canyons. Transform faults are seismically active because the grinding of the plates along them generates strong earthquakes. When the plates stop grinding, the gash in the ocean crust remains but is no longer seismically active. Such gashes are the fracture zones that break up the north-south symmetry of the Atlantic. Whereas a transform fault indicates the present direction of motion between two plates, a fracture zone indicates the direction of relative motion at times in the past.

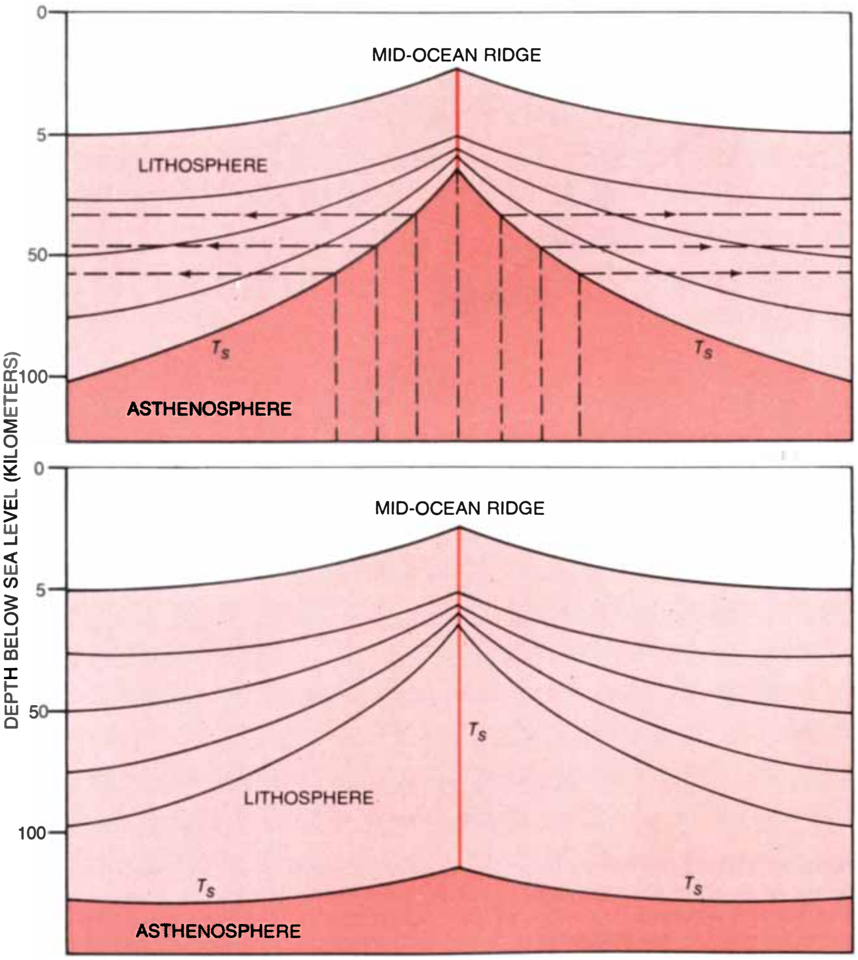
By studying the magnetic anomalies and the fracture zones it is possible to reconstruct the relative positions of the various plates (and hence the continents riding on them) at almost any time in the past. In practice geophysicists are limited to the past 200 million years. As a slab of new ocean floor is created between diverging plates at the mid-ocean ridge 2,500 meters below ocean level it ages, cools, contracts and sinks deeper. A few million years after the continents began to separate, the ocean floor at their edges was only about 3,000 meters deep. After 50 million years the floor at their edges had sunk to a depth of at least 5,000 meters. As the continents eroded, terrigenous sediments collected on the continental shelves and in the adjacent pockets left by the deepening ocean floor. The tremendous weight of these sediments depressed the ocean crust still further.

Before proceeding with the reconstruction of the bathymetric history of the Atlantic we must introduce one last concept: the formation of the aseismic ridges, such as the Rio Grande Rise and the Walvis Rise, that constitute anomalously shallow parts of the ocean. The rate of subsidence of the aseismic ridges reveals whether or not they ever acted as significant barriers to sedimentation and water circulation. Aseismic ridges are created little by little at sites of exceptionally active magmatism on or near the axis of the mid-ocean ridge. The surging magma forms volcanic piles that rise to sea level. As the spreading of the ocean floor moves a pile away from the active areas, the pile becomes merely a passive, aseismic passenger on top of the floor. Therefore the aseismic ridge, that is, a succession of such piles, subsides at the same rate as the ocean floor that supports it, with its crest always lying some 2,500 meters above the sinking floor.

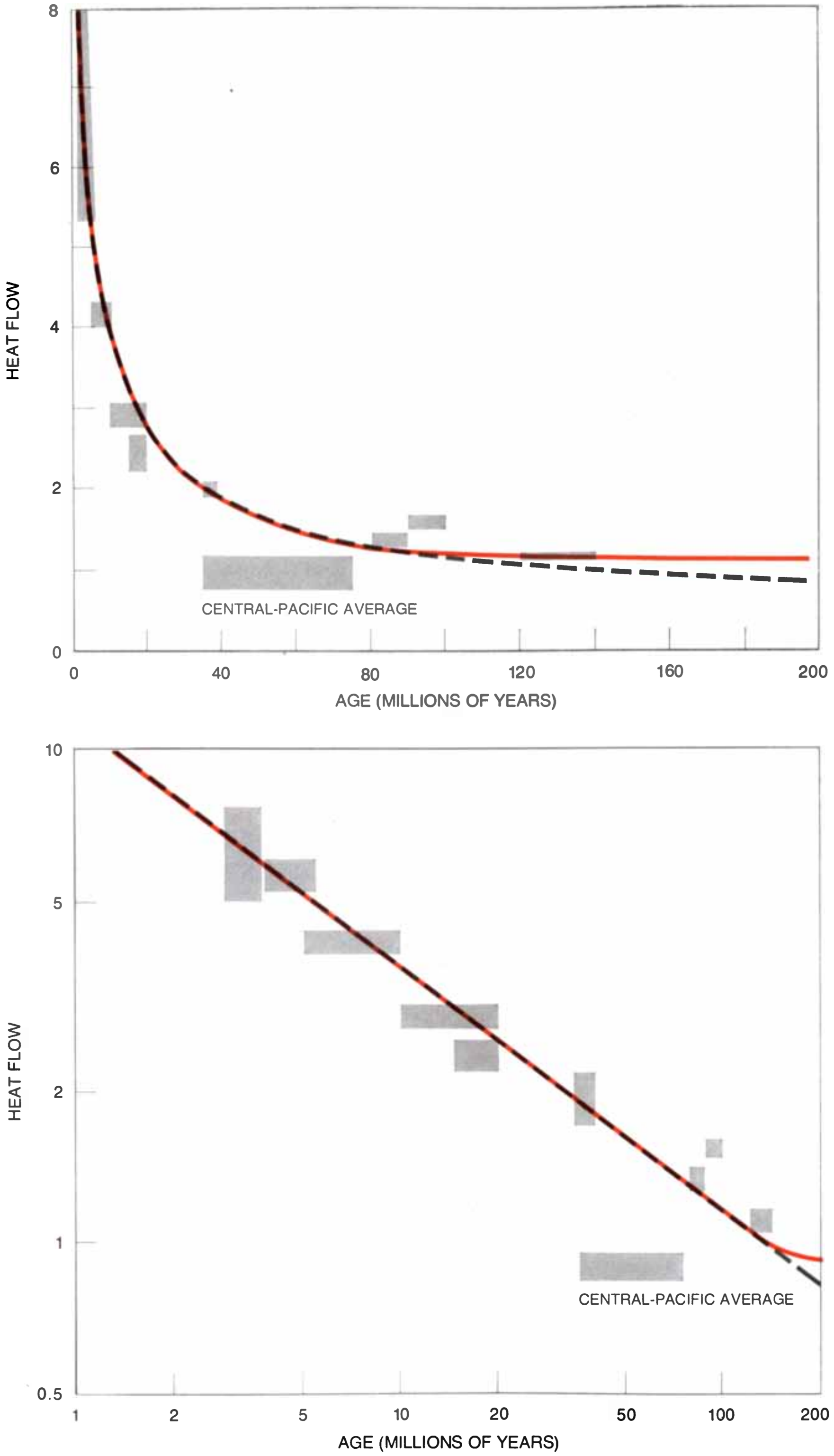
Geophysicists have reconstructed the history of the Atlantic from thermal models, plate-tectonic theory and deep-ocean measurements. In the middle of the Jurassic period 165 million years ago North America, South America, Greenland, Europe and Africa formed one expanse of land. In our reconstruction we have fitted the continents together at their present continental shelves, which are much closer to their ancient edges than the current coastlines are. It is remarkable how well the continents have retained their original shape. If they could be brought together today, Greenland would fit neatly into the coastline of northern Europe, and



CROSS SECTION OF THE MID-ATLANTIC RIDGE shows how the ocean floor deepens symmetrically with increasing distance from the crest. Near continents the sediments increase in thickness. The sedimentary strata are made of clay and terrigenous sediments (*light gray*) and calcium carbonate sediments (*light color*). The lithosphere floats on the asthenosphere.



TWO SCHEMATIC MODELS of the lithosphere and asthenosphere show the flow of matter (*broken lines*) and the isotherms, or surfaces of equal temperature (*solid lines*). The solidus temperature (T_s) is the isotherm that marks the transition from the solid material of the lithosphere to the partially molten material of the asthenosphere. The boundary model (*top*) treats the lithosphere as the upper thermal boundary of the mantle convection system. The plate model (*bottom*) makes the additional assumption that heat flows from the mantle into the lithosphere to maintain a solidus temperature of 1,400 degrees Celsius at a depth of about 125 kilometers. The other isotherms are nearly identical in the boundary and plate models.



PLOT OF THE FLOW OF HEAT (*top*) from the lithosphere into the ocean as a function of the age of the ocean floor was made at various places where the ocean floor was covered by thick layers of sediments. The sediments were impermeable to water, and so the heat flow was mainly the result of conduction (the cooling of the lithosphere), not advection (the transport of thermal energy by water motion). The solid areas indicate the range of the measured data. The broken line represents the relation between heat flow and age given by the boundary model. The model predicts that $q = 11.3/\sqrt{t}$, where q is the heat flow in 10^{-6} calorie per centimeter squared per second and t is the age of the ocean floor in millions of years. Apart from the anomalous central-Pacific average, the equation works well for lithosphere that is less than 120 million years old. The plate model, which predicts that $q = .9 + 1.6e^{(-t/62.8)}$ (*colored line*), works well for any age. The straight-line plot (*bottom*) of the logarithm of the heat flow v. the logarithm of the age of the ocean floor shows the $1/\sqrt{t}$ relation for the past 120 million years.

North America would hug the bulge of Africa and the mass of land consisting of what is now Spain and France. South America, including the sweep of the Falkland plateau, would fit tightly against the African margin.

In the middle of the Jurassic molten magma started to rift the land mass into continents and to create ocean crust between North America and Africa, separating the northern continents from Africa and South America and opening up the infant North Atlantic and Caribbean. By 125 million years ago the North Atlantic had developed an active mid-ocean ridge and had reached depths of as much as 4,000 meters. Water from the new ocean could scarcely flow to the north because the land north of Gibraltar had not yet rifted and because the passage between Spain and Africa was extremely narrow, or perhaps even still above sea level.

At that time parts of Central America were below sea level, but the evidence suggests that water from the Atlantic could not flow through the Caribbean into the Pacific. The ancient flow of water between the Atlantic and the Caribbean depends on the development of the Bahama platform, which juts out from Florida in a southeasterly direction. If coral reefs grew on top of the platform at the same rate at which it subsided, the North Atlantic might have remained separated from the Pacific for a considerable period of time. In any event circulation between the Atlantic and the Caribbean would have been almost as limited and sluggish as the circulation in a stagnant body of water. Such stagnation may have been responsible for the deposition of salt beds along the edges of Africa and North America and for the preservation in the sediments of vast quantities of organic material.

About 125 million years ago South America began to separate from Africa and Spain began to move with Africa away from North America. The motion of South America with respect to North America compressed the Caribbean region, causing the plate north of Venezuela to be subducted. By 80 million years ago the North Atlantic was a full-fledged ocean. Parts of it were more than 5,000 meters deep, and its water could flow into other oceans through the Caribbean and through the wide Strait of Gibraltar. North America and Greenland had begun to drift apart, leaving in their wake narrow, shallow seas jutting out from the main body of the North Atlantic.

The South Atlantic is younger than the North Atlantic, and it had only begun to develop 80 million years ago. Split into northern and southern basins by the Rio Grande Rise and the Walvis Rise, the South Atlantic reached a maximum depth of only about 4,000 meters.

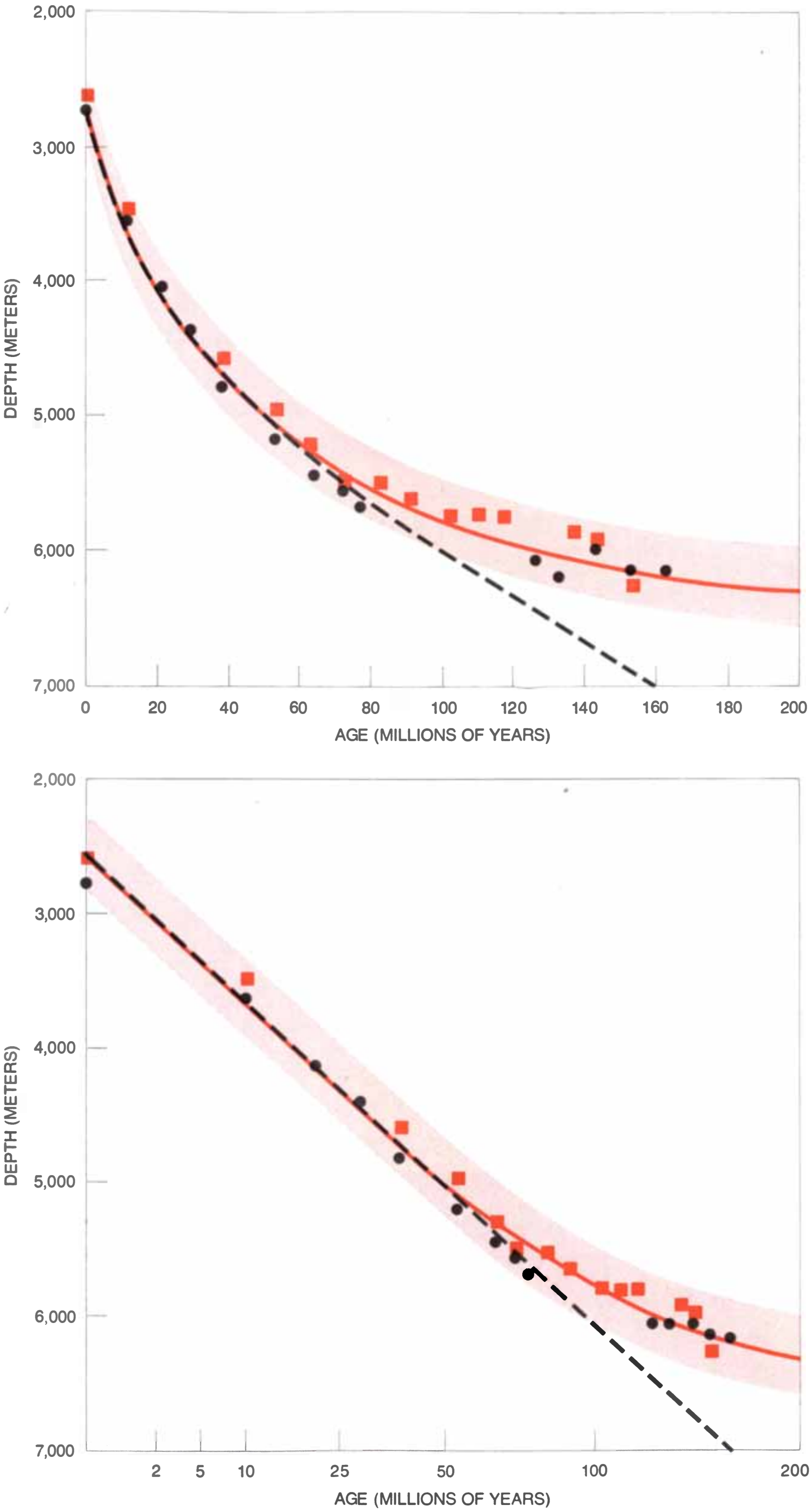
The rises strongly restricted circulation between the basins. The southern basin was linked to other oceans at its lower end, but the northern basin was kept isolated by the bulges of Africa and South America. These bulges and the Rio Grande and Walvis rises enclosed the northern basin for the first 20 or 30 million years of its history. In that period intermittent filling and subsequent evaporation led to the deposition of salt throughout the area; the salt is now found on the continental shelves of Africa and South America north of the rises.

About 65 million years ago Greenland began to move away from Europe. Until about 20 million years ago an aseismic ridge near Iceland had kept cold water from flowing from the Arctic Ocean into the Atlantic; by then the ridge had subsided enough to allow cold seawater from the north to surge into the Atlantic, giving rise to the vigorous circulation the ocean now has.

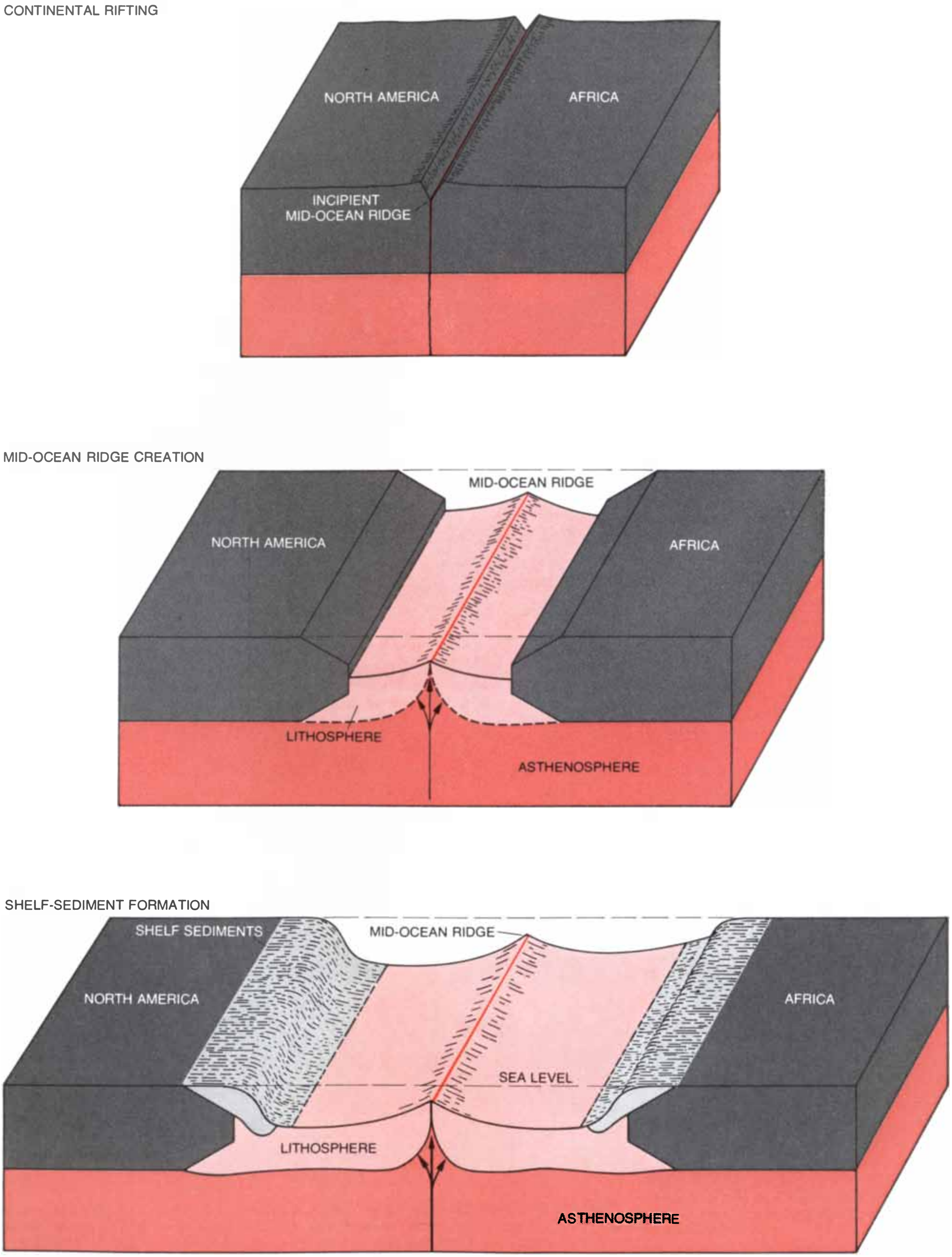
Most of the major topographic features of the Atlantic had formed by 36 million years ago. Both the North Atlantic and the South Atlantic were broad and deep, although the North Atlantic was more so. The Walvis and Rio Grande rises had subsided enough so that they did not greatly obstruct the flow of deep water in the South Atlantic. The Caribbean had almost acquired its present shape. The major difference between then and now is that Spain was farther from Africa. Over the past 36 million years Africa and Europe have slowly converged, almost completely closing the passage between the Atlantic and the Mediterranean.

We had initially hoped we could exploit the preceding bathymetric history of the Atlantic to reconstruct how the morphology of the ocean floor controlled shallow- and deep-water currents. We had expected to be able to directly relate changes in the morphology of the ocean floor to the distribution of sediments. That hope was overly optimistic, and in fact it may never be realized because small-scale local effects, such as the flow of water through fracture zones, play too large a part in determining the circulation of ocean water. The bathymetric history nonetheless provides a framework for examining the deep-ocean sediment record. From the bathymetric history we were able to determine the surface distribution of deep-ocean sediments through time. As magma wells up between the continents and the continents drift apart terrigenous sediments are deposited on the continental shelf and clay and carbonate sediments are deposited in the oceans.

Aseismic ridges act as barriers to the flow of sediments from the continental shelves. When the ridges subside, they still control the kinds of sediment deposited in the deep ocean. As we have



PLOT OF THE DEPTH (*top*) of the ocean floor as a function of its age was made for the North Atlantic (*colored squares*) and the North Pacific (*solid dots*). The tinted area represents the estimated scatter of the depth data. The boundary model predicts that $D = 2{,}500 - 350\sqrt{t}$ (*broken line*), where D is the depth in meters and t is the age of the floor in millions of years. The plate model predicts that $D = 6{,}400 - 3{,}200e^{(-t/62.8)}$ (*colored line*). The straight-line plot (*bottom*) of the depth v. the square root of the age emphasizes the $\sqrt{t}$ relation for ocean floor that is less than 60 million years old. The boundary model works poorly for older ocean floor. The plate model yields correct values for the depth of ocean of any age, as it did for the heat flow.



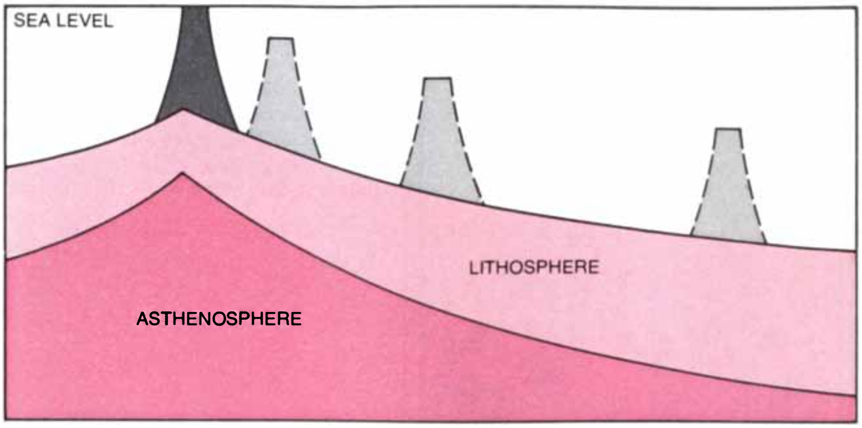
RIFTING OF CONTINENTS begins at the incipient mid-ocean ridge where molten magma from the deep interior of the earth wells up and splits an expanse of land into continents (*top*). At the ridge magma continuously flows out of the gap between the continents, pushing them apart (*middle*). As the surging magma cools, increases in density and sinks it forms new lithosphere that adds to the broad, spreading swell of the mid-ocean ridge. Sedimentary products of continental erosion start to amass on the continental shelves (*bottom*).

mentioned, the calcium carbonate concentration depth marks the level in the ocean below which the seawater is undersaturated in calcium carbonate. The compensation depth is different in each basin and even varies within a basin as a function of position. If the compensation depth and its variation with position are known, it is possible to reconstruct the distribution of carbonate and clay sediments through time. On ocean crust above the compensation depth calcium carbonate dominates, whereas on crust below that depth the surface sediments consist only of fine-grain clays, siliceous oozes and light hemipelagic sediments from the continents.

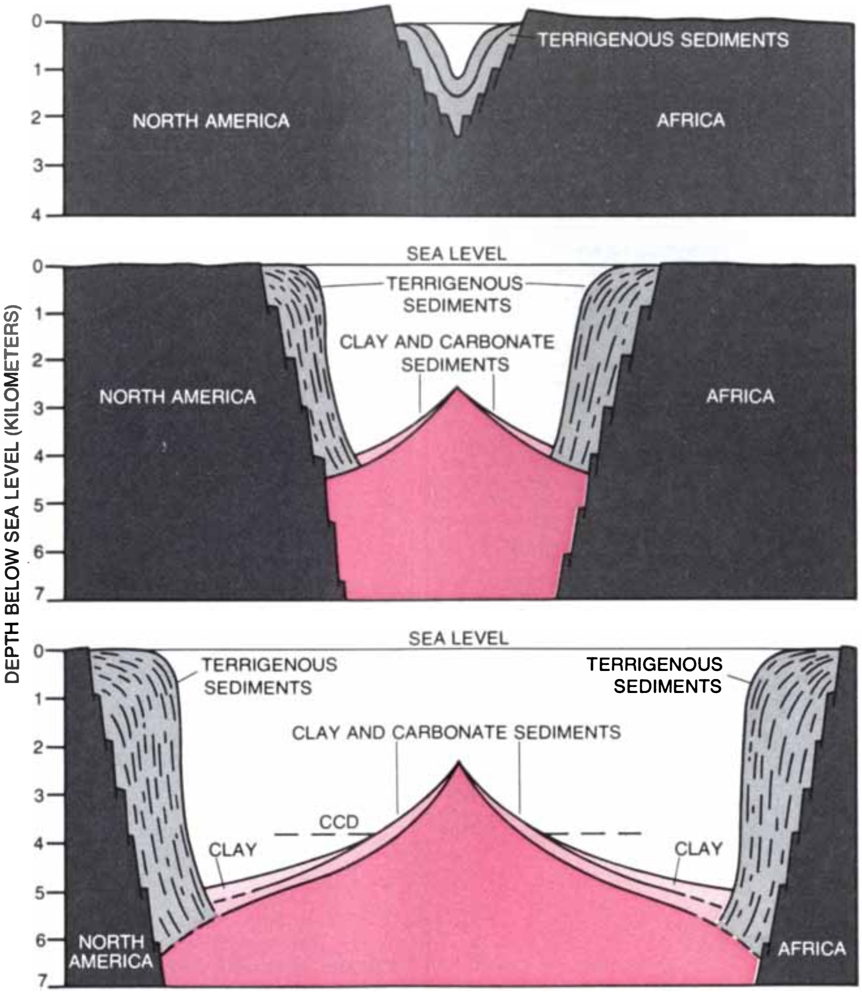
The bathymetric history of the Atlantic accounts for the paradoxical discovery of calcium carbonate sediments below the compensation depth at the base of columnar sections of sediments that were collected near the continental shelves by deep-sea drilling vessels. When the ocean crust forms, it lies above the compensation depth, and so calcium carbonate sediments are deposited on it. As the floor ages it moves away from the spreading center and subsides eventually to a level below the compensation depth. By this time the carbonate sediments have been covered by clay and siliceous ooze, which, if they are close enough to shore, are covered in turn by the overflow of terrigenous sediments from the shelf. The terrigenous, clay and siliceous sediments protect the calcium carbonate deposits from the seawater, which would dissolve them if it were in direct contact with them at those depths.

The bathymetric history of the Atlantic provides information not only about the ocean floor but also about the continental shelves. From columnar samples of sediments on the shelves geophysicists have concluded that the shelves subside at roughly the same rate as the mid-ocean ridge. The similarity in the subsidence rates has been attributed to a thermal process at the shelf that is analogous to the thermal process at the ridge. If that is the case, the shelf should lose heat at the same rate as the neighboring ocean floor. Because most of the sediments deposited on the shelves are impermeable to water the heat there will be dissipated only by conduction, not by advection. As a result the equation that describes the heat flow as a function of age for the mid-ocean ridge applies with only slight modification to the continental shelves. The temperature at any depth in the sediment is directly proportional to the heat flow multiplied by the depth and divided by the conductivity. Therefore if the thickness of the sediments and their rate of deposition are known, a complete temperature and depth history can be reconstructed for any place on the shelves.

Such histories may have large eco-

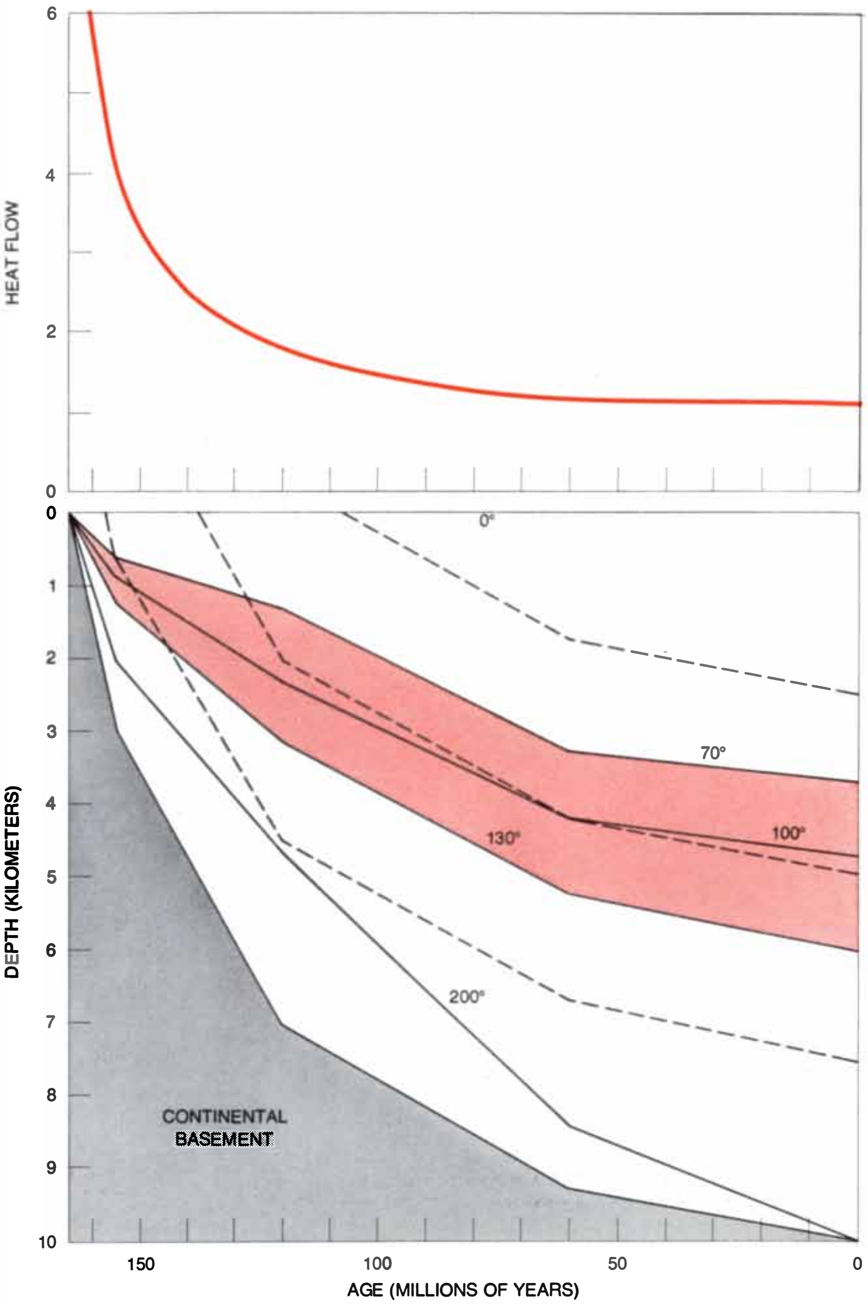


ASEISMIC RIDGES are created little by little at sites of exceptionally active magnetism on or near the axis of the mid-ocean ridge. The surging magma forms a volcanic pile (*dark gray region*) that rises to sea level. As the spreading of the ocean floor moves the pile (*light gray regions*) away from the active areas, the pile becomes a passive, aseismic passenger on top of the ocean floor. Hence the aseismic pile or ridge subsides at the same rate as ocean floor under it.



THREE-STEP MODEL OF SEDIMENTATION illustrates the development of the North Atlantic. About 165 million years ago (*top*) terrigenous sediments from the continents started to accumulate on the emerging continental shelves. By 125 million years ago (*middle*) the ocean floor was covered by red clay (a light detrital material from the continents) and calcareous ooze (calcium carbonate from the decomposed skeletons of microorganisms). Near the crest of the mid-ocean ridge the calcareous ooze is predominant. Between about 4.5 and five kilometers lies the calcium carbonate compensation depth (CCD): the level below which calcareous sediments dissolve because the seawater is undersaturated in calcium carbonate. This means that as ocean floor subsides below the compensation depth calcium carbonate sediments are no longer deposited (*bottom*). Only clay is deposited on ocean floor below the compensation depth.

nomic consequences. One of the major factors in determining whether or not sediments contain oil or gas is their degree of maturation: the length of time the sediments have remained within a certain optimal range of temperatures. Petroleum geologists have been reconstructing the thermal history of continental shelves around the world in order to determine the degree of maturation of the sediments that have been deposited on them. This approach should help in the location of untapped sources of oil and natural gas off the East Coast of the U.S. and in other continental-shelf environments.



PREDICTED HEAT FLOW from the continental-shelf sediments as a function of age (*top*) was made on the basis of the assumption that the thermal process at the shelves is similar to the thermal process at the ridge. The heat flow decreases rapidly to the equilibrium value it has maintained for the past 50 million years. From the heat-flow predictions and from knowledge of the subsidence rate of the ocean floor under the sediments, a complete temperature and depth history can be reconstructed for the sediments at any place on the shelves. The bottom diagram shows the isotherms (*solid lines*) and sedimentation trends (*broken lines*) at a hypothetical location on the outer continental shelf of the eastern U.S. Such reconstructions may have large economic consequences. One of the major factors petroleum geologists consider in determining whether or not continental-shelf sediments contain oil or gas is their degree of maturation: the length of time the sediments have remained within an optimal range of temperatures (from 70 to 130 degrees C.). Here the sediments in this temperature range are colored.

The Year without a Summer

In 1816 in New England it snowed in June, and then killing frosts continued through August. The cause was the explosion of a volcano in Indonesia. The economic and social consequences are instructive

by Henry Stommel and Elizabeth Stommel

In New England, Canada and western Europe the summer of 1816 was extraordinarily cold. A meteorological record for New Haven that had been kept by the presidents of Yale College since 1779 records June, 1816, as the coldest June in that city, with a mean temperature that would ordinarily be expected for a point some 200 miles north of the city of Quebec. The Lancashire plain in England had its coldest July, and the summer as a whole ranks as the coldest on record in the Swiss city of Geneva for the entire period from 1753 to 1960. In New England the loss of most of the staple crop of Indian corn and the great reduction of the hay crop caused so much hardship on isolated subsistence farms that the year became enshrined in folklore as "Eighteen Hundred and Froze to Death." The calamity of 1816 is an interesting case history of the far-reaching and subtle effects a catastrophe can have on human affairs.

The chain of events began in 1815 with an immense volcanic eruption in the Dutch East Indies (now Indonesia), when Mount Tambora on the island of Sumbawa threw an immense amount of fine dust into the atmosphere. Sir Thomas Stamford Raffles, who commanded a British military force in the islands, described the eruption in his *History of Java*: "Almost every one is acquainted with the intermitting convulsions of Etna and Vesuvius, as they appear in the descriptions of the poet and the authentic accounts of the naturalist, but the most extraordinary of them can bear no comparison, in point of duration and force, with that of Tomboro. This eruption extended perceptible evidences of its existence ... to a circumference of a thousand statute miles from its centre, by tremulous motions and the report of explosions; while within the range of its more immediate activity, embracing a space of three hundred miles around it, it produced the most astonishing effects, and excited the most alarming apprehensions. On Java, at a distance of three hundred miles, it seemed to be awfully present. The sky was overcast at noonday with clouds of ashes; the sun was enveloped in an atmosphere, whose 'palpable' density he was unable to penetrate; showers of ashes covered the houses, the streets, and the fields to a depth of several inches; and amid the darkness explosions were heard at intervals, like the report of artillery or the noise of distant thunder. So fully did the resemblance of the noises to the report of cannon impress the minds of some officers, that from an apprehension of pirates on the coast vessels were dispatched to afford relief."

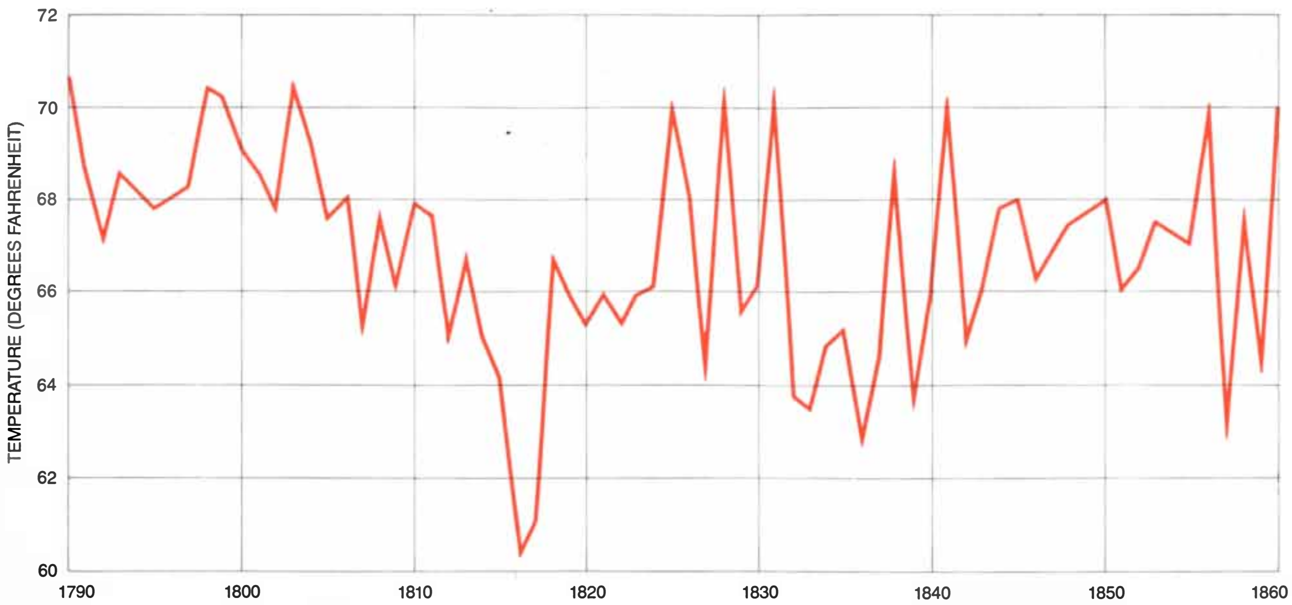
This eruption, which was considerably larger than the better-known one of Krakatoa in 1883, reduced the height of Mount Tambora by some 4,200 feet and ejected some 25 cubic miles of debris. Ash was encountered by ships at sea as large islands of floating pumice as much as four years after the event. Climatologists rank the eruption as the greatest producer of atmospheric dust between 1600 and the present. The dust circled the earth in the high stratosphere for several years, reflecting sunlight back into space and thereby reducing the amount of it reaching the ground.

The idea that dust in the upper air can result in lower temperatures at ground level is quite old. Benjamin Franklin invoked it to explain the cold winter of 1783–84. Today the idea can be confirmed more conclusively through long records of temperature from many parts of the world, which can be compared with the fairly complete record of the volcanic eruptions that have been observed during the past two centuries.

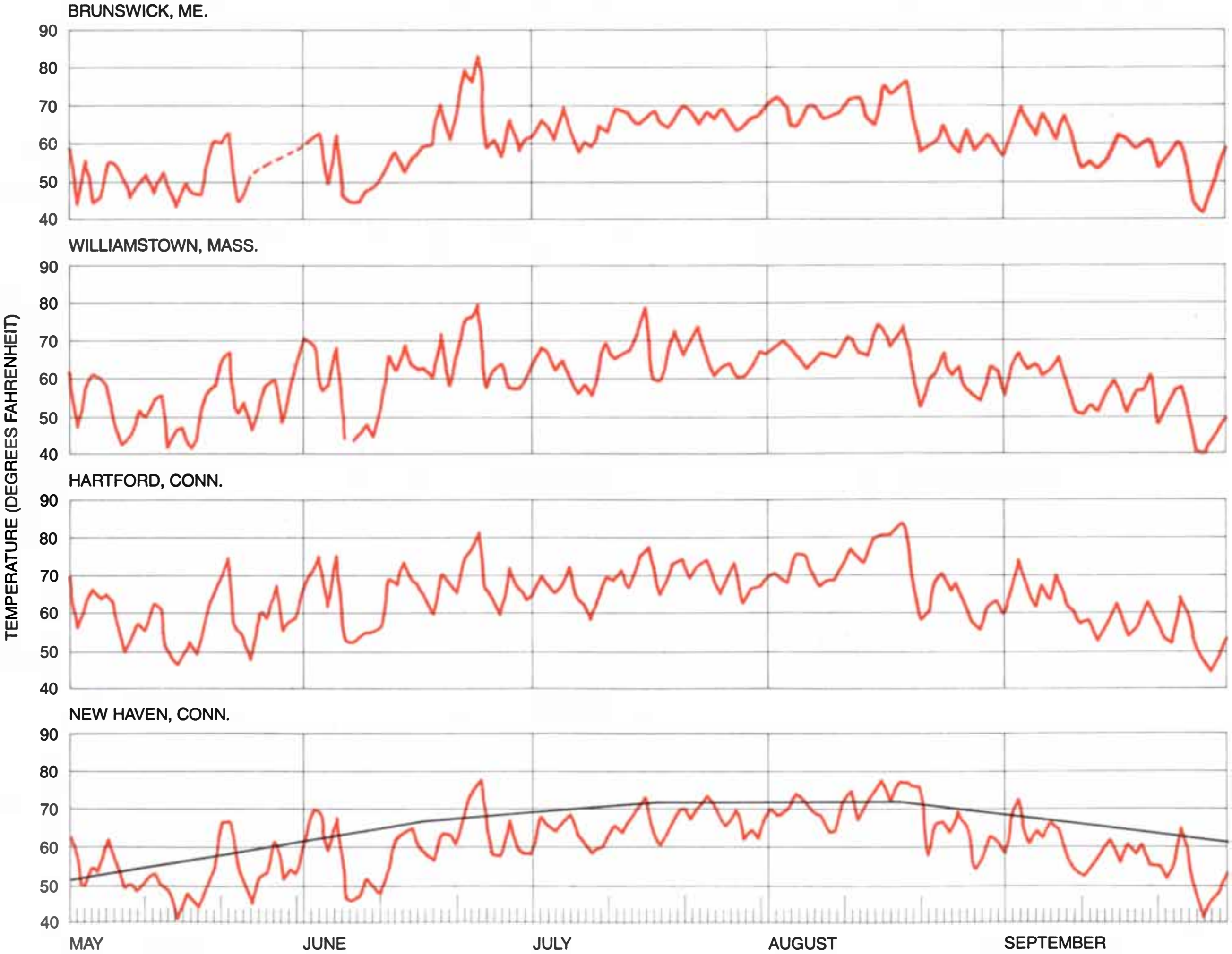
As the dust in the upper atmosphere circled the earth after the eruption of Tambora, it gradually shadowed the higher latitudes. The first two months of 1816 were not exceptionally cold in New England, but by May observers had begun to comment on the lateness of the spring. June began auspiciously, and crops that had survived the unwonted frosts of mid-May started to progress. The first of three unseasonable cold waves moved eastward into New England early on June 6. The cold and wind lasted until June 11, leaving from three to six inches of snow on the ground in northern New England. A second killing frost struck the same areas on July 9 and a third and fourth on August 21 and 30, just as the harvest of twice-ravaged crops was about to begin. The repeated summer frosts destroyed all but the hardiest grains and vegetables.

The temperatures are recorded in a number of meteorological journals kept at the time, including the one maintained by the presidents of Yale College and another kept by William Plumer, who in 1816 was governor of New Hampshire. Many personal accounts also record the unusual weather. For example, Hiram Harwood, a farmer at Bennington in the southwest corner of Vermont, remarked in his diary on the frosts of mid-May, which had hindered vegetation to the point of making it about two weeks late. When he awoke early on June 6, after a night of heavy rain and sharp northeast winds, he saw that the mountains on every side were crowned with snow. On the morning of June 7 he found his fields stiff with frost. The leaves of the trees were now blackened. On June 8 the fierce weather continued, with sweeping blasts from the north all morning and with intermittent snow squalls.

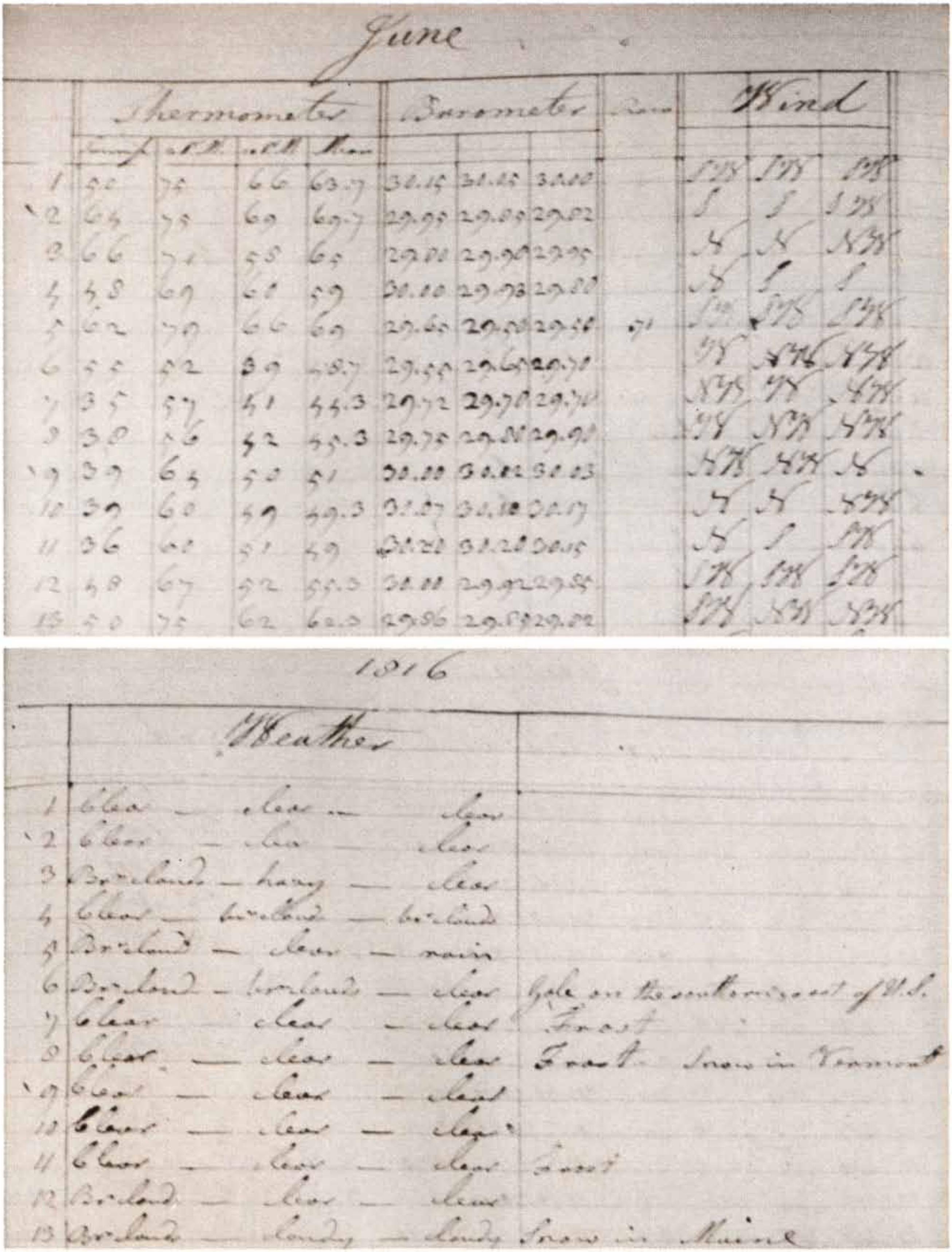
The summer made enough of an impression on the mind of Chauncey Jerome, who in 1816 was a 26-year-old apprentice clockmaker in Plymouth, Conn., so that when 44 years later he came to write his *History of the American Clock Business for the Past Sixty Years,* he said: "I well remember the 7th of June, while on my way to work, about a mile from home, dressed throughout with thick woolen clothes and an overcoat on, my hands got so cold that I was obliged to lay down my tools and put on a pair of mittens which I had in my pocket. It snowed about an hour that day. On the 10th of June, my wife



TEMPERATURE RECORDS for New Haven, Conn., convey the mean temperature in June over a period of 70 years, beginning in 1790. The low of slightly more than 60 degrees Fahrenheit in 1816 was the result of temperatures that ranged from 35 degrees at sunrise on June 7 to 88 degrees at 2:00 P.M. on June 24, according to a journal kept for many years by the presidents of Yale College.



ABNORMAL COLD in New England during the late spring, the summer and the early fall of 1816 is reflected in these charts of the average daily temperature in four towns from May through September. The black curve for New Haven indicates the normal average.



June

	Thermometer				Barometer			Rain	Wind		
1	50	75	66	63.7	30.15	30.25	30.00		SW	SW	SW
2	64	75	69	69.7	29.95	29.89	29.82		S	S	SW
3	66	71	58	65	29.80	29.90	29.95		N	N	NW
4	48	69	61	59	30.00	29.93	29.80		N	S	S
5	62	79	66	69	29.60	29.50	29.56	.71	SW	SW	SW
6	55	52	39	48.7	29.55	29.65	29.70		W	NW	NW
7	35	57	41	44.3	29.72	29.70	29.74		NW	W	NW
8	38	56	42	45.3	29.75	29.80	29.90		W	NW	NW
9	39	64	50	51	30.00	30.02	30.03		NW	NW	N
10	39	60	49	49.3	30.07	30.10	30.17		N	N	NW
11	36	60	51	49	30.20	30.20	30.15		N	S	SW
12	48	67	52	55.3	30.00	29.92	29.85		SW	SW	SW
13	50	75	62	62.3	29.86	29.80	29.82		SW	NW	NW

1816

	Weather	
1	clear — clear — clear	
2	clear — clear — clear	
3	Br. clouds — hazy — clear	
4	clear — br. clouds — br. clouds	
5	Br. clouds — clear — rain	
6	Br. clouds — br. clouds — clear	Gale on the southern coast of U.S.
7	clear clear — clear	Frost
8	clear — clear — clear	Frost. Snow in Vermont
9	clear — clear — clear	
10	clear — clear — clear	
11	clear — clear — clear	Frost
12	Br. clouds — clear — clear	
13	Br. clouds — cloudy — cloudy	Snow in Maine

METEOROLOGICAL JOURNAL for New Haven kept at Yale College for many years reflects the unusually cold weather that began on June 6, 1816, and continued through June 11. One sees here the upper part of two pages that face each other in the journal; the left-hand page is at the top. The series of journals was begun in the 18th century by Ezra Stiles when he was president of Yale and was continued by or under the direction of his successors. The president in 1816 was Timothy Dwight, but he was near the end of his life and was succeeded in 1817 by Jeremiah Day. Whether entries for June, 1816, are in hand of Dwight, Day or some other person at Yale is not known.

brought in some clothes that had been spread on the ground the night before, which were frozen stiff as in winter. On the 4th of July I saw several men pitching quoits in the middle of the day with thick overcoats on, and the sun shining bright at the time."

Newspapers carried long descriptions of the extraordinary weather. Early in June the *North Star* of Danville, Vt., reported: "Some mention was made not long since of the unusual backwardness of the spring, and the remarkable instability of the weather. Although the summer months have commenced, the weather is no more steady nor the prospects more promising. Wednesday last [June 5] was perhaps as warm and sultry a day as we have had since September—At night heat lightning was observed, but on Thursday morning the change of weather was so great that a fire was not only comfortable, but actually necessary. The wind during the whole day was as piercing and cold as it usually is the first of November and April. Snow and hail began to fall about ten o'clock, A.M., and the storm continued till evening, accompanied with a brisk wind, which rendered the habiliments of winter necessary for the comfort of those exposed to it." Noting that the wintry weather continued through the week, the newspaper said: "Probably no one living in the country ever witnessed such weather, especially of so long continuance."

In the wake of the storm good growing weather returned. From Connecticut to Quebec farmers with reserves of seed plowed their blackened fields and replanted their corn and beans and other tender crops.

The second cold wave, which was less severe, came early in July. Still, it was severe enough for ice to form in Maine on July 5, and on July 9 corn was again killed by frost except in well-sheltered places. Summer weather returned on July 12, and conditions remained warm and pleasant until August 20. Then New England was hit by a series of unusually early frosts. In *Historic Storms of New England* Sidney Perley wrote: "On the night of the twenty-first, there was a frost, which at Keene and at Chester, N.H., killed a large part of the corn, potatoes, beans and vines, and also injured many crops in Maine. It was felt as far south as Boston and Middlesex county in eastern Massachusetts, and in the western portion of the state as far as Stockbridge, where it injured vegetation. The mountains in Vermont were now covered with snow, and the atmosphere on the plains was unusually cold. In Keene, N.H., the oldest persons then living said that they never saw such a severe frost in August. It put an end to the hopes of many farmers of ripening their corn, especially in the low lands, and they immediately cut the whole stalks up for fodder, but being in the milk it heated in the shocks and spoiled. By the twenty-ninth of the month the frost had reached as far south as Berkshire county, Mass., where it killed the Indian corn in many of the fields in the low lands. The farmers there saved much of it by cutting it up at the roots and placing it in an upright position, where it ripened upon the juices of the stalks. If frost had kept off two weeks longer, there would have been a very good crop of corn in Massachusetts."

The cold was still more severe in Canada. The small lakes to the north of Baie Saint Paul on the St. Lawrence River were still covered with ice in the middle of July. In Canada even wheat, which along with other grains except corn had done well in the U.S., perished. The Halifax *Weekly Chronicle* noted that "great distress prevails in many parishes throughout Quebec Province from a scarcity of food. Bread and milk is the common food of the poorer classes at this season of the year; but many of them have no bread."

It is difficult to piece together a coherent picture of the effects of the unusual weather from ephemeral accounts in the newspapers of the time. Later accounts, such as county histories in which farmers are said to have slaughtered their flocks and even to have hanged themselves because of the privations brought on by the cold summer, appear to have been to at least some extent embellished. Fortunately there are more reliable sources on the social import of the cold year: a special study instituted by the Philadelphia Society for the Promotion of Agriculture, the record of wholesale prices for agricultural products and studies of patterns of emigration from the afflicted states to new lands opening in the West.

In response to widespread interest in the effects of the cold summer the Philadelphia Society for the Promotion of Agriculture resolved on October 30 of that year to "collect facts relating to Agriculture and Horticulture, and of all circumstances connected therewith, which have occurred through the extraordinary season of 1816; and particularly the effects of Frost on vegetation." Copies of the resolution were circulated widely. Replies soon began to arrive.

One of the first was from Samuel Latham Mitchill, a physician who was also professor of natural history at Columbia College in New York. "There will not be half a crop of maize on Long Island, and in the southern district of this state," he wrote. "Further northward there will be less. The buckwheat is so scanty, that a few days ago I paid four dollars for a half barrel of the meal, for the use of my family. The winter crop of wheat and rye was abundant.... An entomologist complained to me, a few weeks ago, that it has been a most unfortunate season for the collection of insects. That kind of game, he said, was so rare, that he had added but little to his museum. There have been at New York fewer fleas and musquitoes than ordinary."

Another reply came from General David Humphreys, a hero of the American Revolution who in 1816 was president of the Connecticut Society of Agriculture. He wrote: "The principal injury done by early and late frosts, fell on our most important crop, Indian corn. Of this, there is not more than half the usual quantity; and, in many places in this neighbourhood, not more than a quarter part sufficiently hard and ripe for being manufactured into meal. That which is unripe, mouldy or soft, when given as feed to hogs and cattle, has little tending to fatten them.... Grasses, for pasturage and hay, have been diminished by the drought about 50 per cent. The hay is estimated to be nearly 25 per cent better than it is in wet seasons; containing considerably more nutriment, and having been well cured.

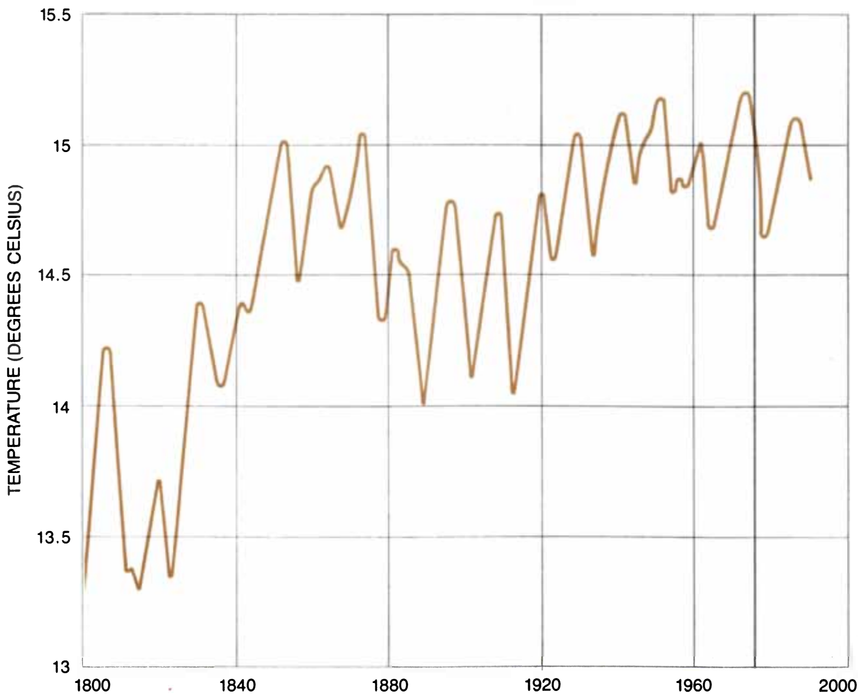
"All kinds of grain were a longer time than usual in filling and ripening; which is considered one reason, why those which came to maturity are more than commonly full and heavy. Wheat and rye never yielded more abundantly....

"Roots and vegetables, in gardens that were well tended by having the earth frequently moved round the plants, early in the morning, while the dew was on the ground, have generally been more flourishing and productive than in ordinary seasons. Some attentive horticulturists observed that more dew fell than usual."

The soaring prices are a matter of record. Newspapers regularly published the wholesale prices of most agricultural products, including wheat and flour, which serve as an index of food prices as a whole. The newspaper accounts all reported a large and sudden rise in the price of flour owing to the reports from Europe of widespread crop failures.

Surpluses of American wheat and rye could have met the foreign demand without inflation had it not been for the failure of much of the corn crop from Pennsylvania northward. As it was, the price of wheat rose in 1817 to $2.45 per bushel, far higher than it had been before or would be for long afterward. In the period 1800–1811 the wholesale price of wheat was about $1.30, and for most of the 30 years after 1817 it stayed around $1.05.

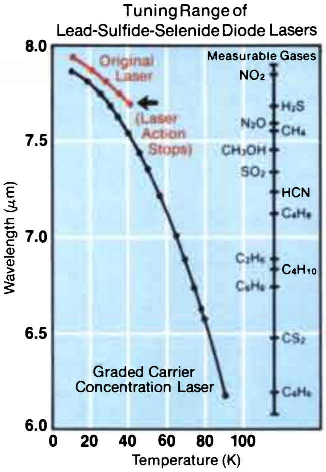
In *Historic Storms of New England* Perley recorded a number of observations about the effect of the cold summer on prices. "There was great destitution



GLOBAL MEAN TEMPERATURES were worked out by Stephen H. Schneider and Clifford Mass of the National Center for Atmospheric Research in Boulder, Colo., using a mathematical model of the global radiation balance calculated on the basis of the number of sunspots and the density of volcanic dust in the atmosphere. The curve shown here does not take into account the putative warming effect of the carbon dioxide that has been put into the atmosphere by the combustion of fossil fuels. Schneider and Mass found that the density of dust is a more important factor than the number of sunspots in contributing to fluctuations of temperature.

“Tunable semiconductor diode laser” is a long name for a speck of lead-salt crystal. When tuned (by adjusting temperature) to emit the infrared wavelength that a particular gas can absorb, this tiny device can beam through a mixture of gases and quantify even small amounts of the targeted gas.

When research on diode lasers began here at the General Motors Research Laboratories, their tuning range was limited to about 1% of the infrared spectrum. Even modest temperature increases (to extend tunability) caused energy losses in the crystal that stopped laser action.

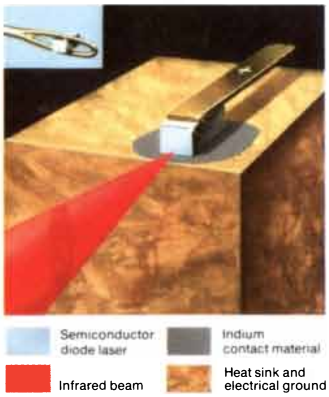


However, by producing a unique form of lead-sulfide-selenide single crystal that reduces these losses, our physicists developed a new type of semiconductor laser having 10 times the tuning range.

There were problems besides limited tunability. One was degradation from diffusion of the electrical contact material, indium, into the crystal. Interposing layers of gold and platinum ended that difficulty.

Another problem: The crystal would separate from the heat sink after repeated cycling down to the necessary cryogenic temperatures. That, too, was resolved . . . by a better chip-mounting arrangement.

As a result of these and other advances in tunable semiconductor diode lasers, even the sky won’t be the limit of their usefulness. For example, such miniature searchlights may permit the measurement of gases surrounding the planets. They are already scheduled by NASA for satellite probings of our own stratosphere. And here on earth, they’re now ready to play a significant role in monitoring pollution.



We currently have openings for Ph.D.s in engineering and the physical and mathematical sciences. If interested, please send your resume to: GMR Personnel, Dept. 610. An Equal Opportunity Employer.

More tunability, more durability for the world's smallest lasers.

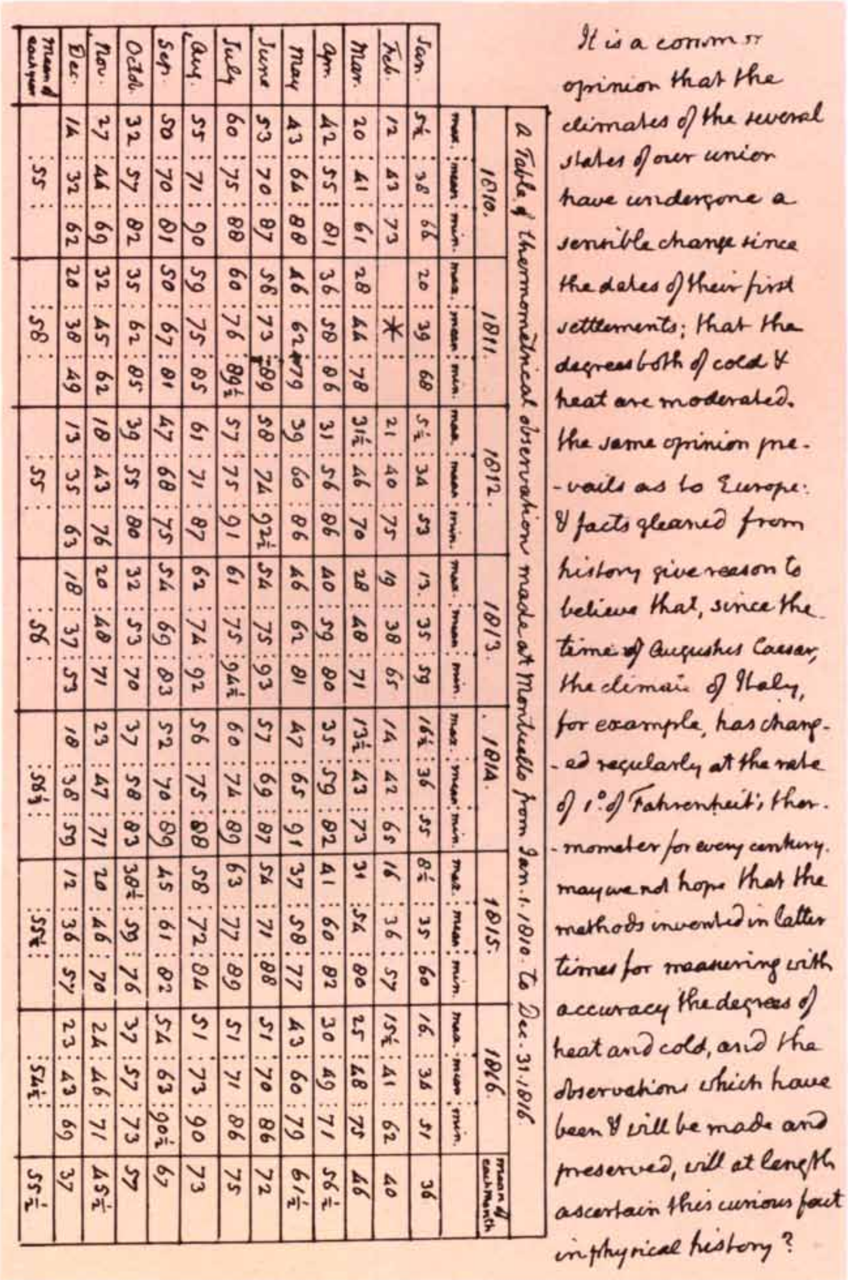
GM

General Motors Research Laboratories
Warren, Michigan 48090

among the people the next winter and spring," he wrote. "The farmers in some instances were reduced to the last extremity, and many cattle died. The poorer men could not buy corn at the exorbitant prices for which it was sold. In the autumn, stock was sold at extremely low prices on account of lack of hay and corn, a pair of four-year-old cattle being bought for thirty-nine dollars in Chester, New Hampshire....

"The next spring hay was sold in New Hampshire in a few instances as high as one hundred and eighty dollars per ton, its general price, however, being thirty dollars. The market price of corn was two dollars per bushel; wheat, two dollars and a half; rye, two dollars; oats, ninety-two cents; beans, three dollars; butter, twenty-five cents per pound; and cheese, fifteen cents. In Maine, potatoes were seventy-five cents per bushel, the price in the spring of 1816 having been forty cents, which was the usual price."

The geographer Joseph B. Hoyt extracted weekly price reports from New York newspapers of the period for wheat, corn, pork and beef. The prices of pork and beef usually went through a yearly cycle with the highest point (pork at $24 per barrel, for example) in late fall, when farmers were holding and feeding their stock, and the lowest point (pork at $20 per barrel) in late spring. The summer of 1816 completely reversed the cycle. Farmers, concerned about the poor prospects for hay, sold off their stock early enough to depress the summer and fall pork prices to $17 per barrel. The price of beef followed a similar pattern.



A Table of thermometrical observations made at Monticello from Jan. 1. 1810. to Dec. 31. 1816.

	1810. max.	mean	min.	1811. max.	mean	min.	1812. max.	mean	min.	1813. max.	mean	min.	1814. max.	mean	min.	1815. max.	mean	min.	1816. max.	mean	min.	mean of each month
Jan.	5¼	38	68	20	39	68	5¼	34	53	13	35	59	16½	36	55	8½	35	60	16	38	51	36
Feb.	12	43	73		*		21	40	75	19	38	65	14	42	65	16	36	57	15½	41	62	40
Mar.	20	41	61	28	44	78	31½	46	70	28	48	71	13½	43	73	31	54	80	25	48	75	46
Apr.	42	55	81	36	58	86	31	56	86	40	59	80	35	59	82	41	60	82	30	49	71	56½
May	43	64	88	46	62½	79	39	60	86	46	62	81	47	65	91	37	58	77	43	60	79	61½
June	53	70	87	58	73	89	58	74	92½	54	75	93	57	69	87	54	71	88	51	70	86	72
July	60	75	88	60	76	89½	57	75	91	61	75	94½	60	74	89	63	77	89	51	71	86	75
Aug.	55	71	90	59	75	85	61	71	87	62	74	92	56	75	88	58	72	84	51	73	90	73
Sep.	50	70	81	50	67	81	47	68	75	54	69	83	52	70	89	45	61	82	54	63	90½	67
Octob.	32	57	82	35	62	85	39	55	80	32	53	70	37	58	83	30½	59	76	37	57	73	57
Nov.	27	44	69	32	45	62	18	43	76	20	48	71	23	47	71	20	46	70	24	46	71	45½
Dec.	14	32	62	20	38	49	13	35	63	18	37	53	18	38	59	12	36	57	23	43	69	37
Mean of each year		55			58			55			56			58¼			55¾			54¼		55½

It is a common opinion that the climates of the several states of our union have undergone a sensible change since the dates of their first settlements; that the degrees both of cold & heat are moderated. The same opinion prevails as to Europe: & facts gleaned from history give reason to believe that, since the time of Augustus Caesar, the climate of Italy, for example, has changed regularly at the rate of 1.° of Fahrenheit's thermometer for every century. may we not hope that the methods invented in latter times for measuring with accuracy the degrees of heat and cold, and the observations which have been & will be made and preserved, will at length ascertain this curious fact in physical history?

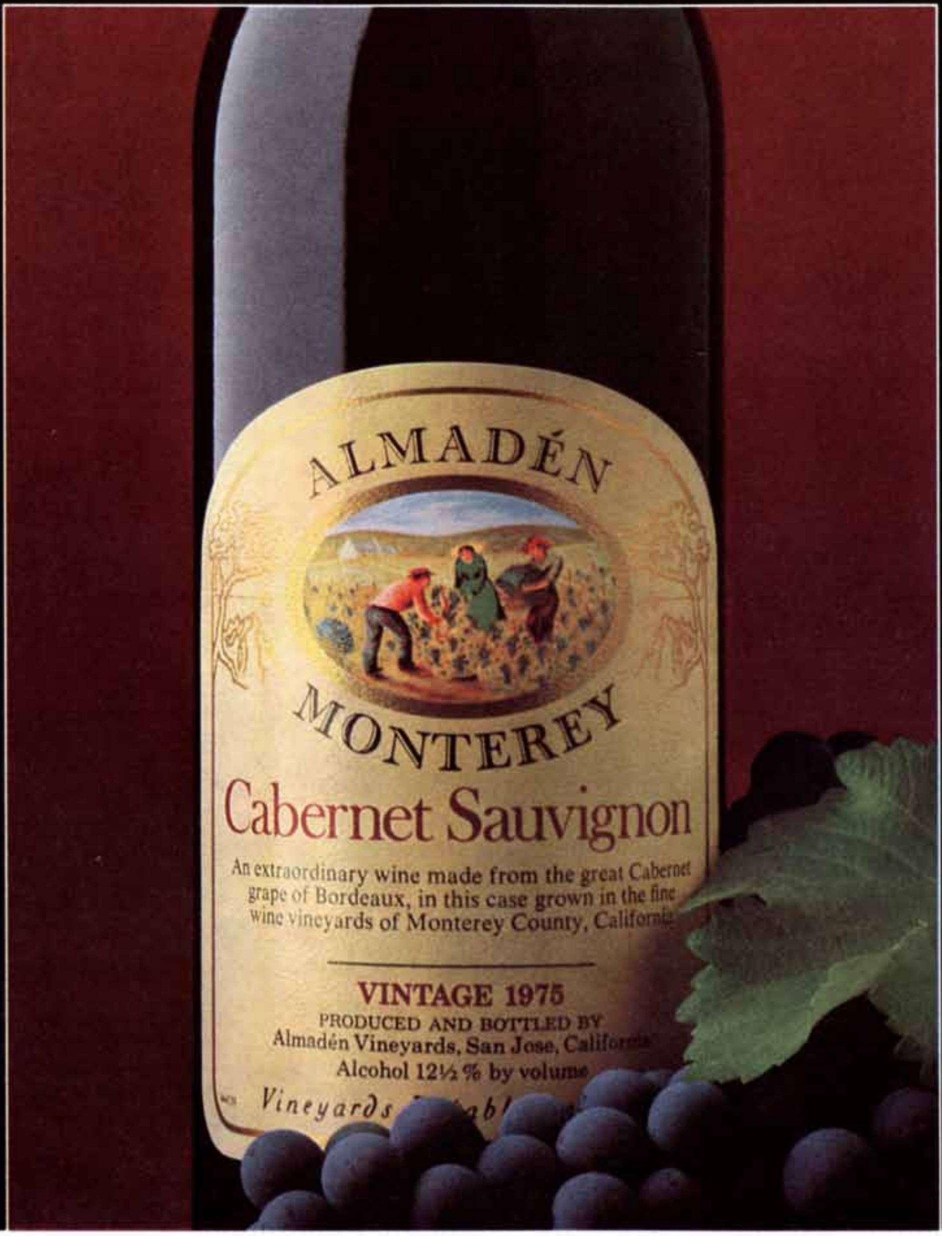
WEATHER RECORD kept by Thomas Jefferson at his home in Virginia shows that the summer of 1816 was somewhat colder than normal there, although the principal effects of the volcanic dust on temperature were felt farther north. Jefferson began this set of observations on the weather the year after he left the presidency. In the table he transposed, presumably through inadvertence, either the maximum and minimum readings or the column headings.

Exports of wheat to Canada, Britain and Europe substantially inflated the domestic price. The New York *Weekly Museum* for January 4, 1817, made mention of the subject: "Within a few weeks, it is said, more than 100,000 barrels of flour have been shipped from New York, Philadelphia and Baltimore for England; in consequence the price of flour here is now $14.00 a barrel."

The summer of 1816 marked the point at which many New England farmers who had been thinking of moving west finally decided to go. The movement had been under way for some time, and the cold weather of 1816 with its severe consequences for agriculture was certainly not the only motivating factor, but it did serve as an extra prod. Emigration was particularly heavy from Vermont and Maine. The Vermont historian L. D. Stilwell examined many accounts of individual migrations to the west from Vermont and found that in 1816–17 the number of people leaving was almost twice the number who had left in other years of the decade. At the other end of the migration the *Messenger* of Zanesville, Ohio, noted on October 31, 1816, that "the number of Emigrants from the eastward the present season, far exceeds what has ever before been witnessed."

The harsh summer of 1816 had even severer consequences for parts of Europe than it did for the U.S. The bad weather followed closely on the disruptions of the Napoleonic Wars, which had ended in 1815 with the exile of Napoleon to St. Helena. In many places the poor crops of 1816 caused grave shortages of food and conditions approximating famine. Although several countries were affected, we shall limit our discussion to Switzerland and France.

Since the Middle Ages, Zurich had been a center for the grain market. High prices chronicle the periods of scarcity: 1692, 1770–71 and 1816–17. Local publications record 1816 as being unusually cold, dry and unpleasant. Attempts to replant summer wheat were frustrated by a lack of seed in the state granaries. Swine had to be slaughtered for want of fodder. By the end of the year the shortage of food was severe, particularly in the cities. Parish registers record a num-



We are proud of the birthdates of our children, the grapes of Almadén.

Yes, we are very proud. And we show our pride by putting a vintage date on our classic varietal wines.

Because each vintage—each year's harvest—is like a new child. A child we lovingly care for as it develops and matures. And we give each its just recognition by giving it a vintage date.

For you, the vintage date adds yet another dimension to the enjoyment of our classic varietal wines. Because each vintage has a different personality. Subtle differences you may taste and enjoy.

Also, while all our wines are ready to drink when purchased, you may sometimes wish to set a few bottles aside. Here you will find the vintage date helpful in giving the wine the additional aging you desire.

We invite you to enjoy the incomparable quality of our vintage-dated wines. An incomparable quality you will find in our entire family of Almadén wines.

ALMADÉN

MEMENTO OF COLD SUMMER is seen on this tombstone, which still stands on what was evidently Reuben Whitten's farm in Ashland, N.H. The reference to the "40 bushils of wheat" that he raised appears to reflect the fact that the wheat crop, particularly winter wheat, was comparatively good in 1816, whereas the corn crop, which was the staple of most farms in New England, was severely curtailed.

ber of deaths that show a clear association with famine. The churches designated January 26, 1817, a day for the special gathering of funds to alleviate distress. All kinds of things were eaten: sorrel, moss and cat flesh. Instructions were issued to help people identify poisonous plants.

The situation in France was equally grave. Torn by military campaigns and the defeat at Waterloo in 1815 and stripped of the feudal protection that had been one of the positive aspects of the vanished aristocracy, the French peasantry had no reserves with which to buffer the bad season of 1816. The country was in a political ferment as Louis XVIII and Talleyrand struggled to maintain the constitutional monarchy against the followers of Napoleon. The political ferment brought about a decline in industrial activity, which in combination with the meager harvest of 1816 and the high cost of food (already in short supply because of the invasions of 1815) set the stage for riots and insurrection.

In Poitiers rioting broke out because of a tax of three francs per bushel imposed on wheat. Grain carts on their way to market in towns along the Loire valley had to be protected by soldiers and gendarmes, who found themselves fighting as many as 2,000 hungry and enraged citizens. Where the harvest had been good farmers were afraid to take their produce to market because so many robbers were abroad.

In August the government suspended import duties on grain. By November it was seeking supplies for importation. According to a recent study by J. P. Housel of the mean monthly prices of wheat in France from 1801 to 1912 the highest point was reached during the first half of 1817. It was then approximately twice as high as the long-term average, a pattern similar to the one that prevailed in the U.S. The political and economic difficulties continued in Europe until the harvest of 1817 brought significant relief.

One suggested consequence of the unusual weather of 1816 is intriguing and plausible, although it cannot be proved. It is the suggestion (made by J. D. Post of Northeastern University) that the anomalous weather of 1816 was responsible for the world's first pandemic of cholera.

Medical histories indicate that before the time of this great outbreak cholera had been confined to the region of the Hindu pilgrimage on the Ganges, with sporadic incursions into China. The failed harvests of 1816 and the ensuing famine in India weakened enough people to give rise to a local epidemic in Bengal. From there the disease was spread by British military operations to the Afghans and the Nepalese. On

How to buy a personal computer.

In California, a store owner charts sales on his Apple Computer. On weekends though, he totes Apple home to help plan family finances with his wife. And for the kids to explore the new world of personal computers.

A hobbyist in Michigan starts a local Apple Computer Club, to challenge other members to computer games of skill and to trade programs.

Innovative folks everywhere have discovered that the era of the personal computer has already begun—with Apple.

Educators and students use Apple in the classroom. Businessmen trust Apple with the books. Parents are making Apple the newest family pastime. And kids of all ages are finding how much fun computers can be, and have no time for TV once they've discovered Apple.

Visit your local computer store

The excitement starts in your local computer store. It's a friendly place, owned by one of your neighbors. He'll show you exactly what you can use a personal computer for.

What to look for

Your local computer store has several different brands to show you. So the salesman can recommend the one that best meets your needs. Chances are, it will be an Apple Computer. Apple is the one you can program yourself. So there's no limit to the things you can do. Most important, Apple's the one with more expansion capability. That means a lot. Because the more you use your Apple, the more uses you'll discover. So your best bet is a personal computer that can grow with you as your skill and involvement grow. Apple's the one.

It's your move

Grab a piece of the future for yourself. Visit your local computer store. We'll give you the address of the Apple dealer nearest you when you call our toll-free number. Then drop by and sink your teeth into an Apple. **800-538-9696**. In California, **800-662-9238**.

apple computer

reaching the shores of the Caspian Sea it moved slowly westward by two routes: one up the Volga to the Baltic ports and the other through the Moslem hajj (the pilgrimage to Mecca) to the Middle East. In a time before railroads and airplanes the progress of the disease was leisurely, somewhat like the movement of plant diseases today.

When the first worldwide attack of cholera struck New York in the summer of 1832, dispersing well-to-do city folk throughout the nearby countryside, neither they nor their sometimes reluctant rural hosts connected the pandemic with the weather of 1816. News dispatches of the pandemic had come in from Europe and the Near East carrying vivid accounts of the dreadful impact of the disease in Moscow, Pest, Sevastopol, Paris and elsewhere. Then came the dismaying news from Montreal that the scourge had crossed the Atlantic. By July 20 the toll of dead in New York was as high as 100 per day. The devastation was mainly among what the authorities of the time would have called the lower orders of society. The disproportionate mortality of that stratum was so clearly a matter of sanitation and water supply that it led the city to undertake the construction of the Croton Aqueduct, which still supplies the city with much of its water.

The train of events from 1816 to 1832 is tenuous. The weather certainly caused the famine in Bengal, and the famine made conditions ripe for the outbreak of cholera there, whence it spread to the West. Even in the absence of the famine in Bengal, however, European adventures in imperialism would inevitably have loosed cholera on the world eventually.

The cold summer of 1816 did not escape the notice of contemporary scientists. Some of them blamed it on sunspots. Ernst Chladni, a well-known amateur acoustician, attributed the cold to an outbreak of Arctic ice in the North Atlantic. He put forward his theory in *Annalen der Physik,* attempting to substantiate it by reporting sightings of icebergs that had been made from ships in the North Atlantic. Another view was that a significant amount of heat came from the interior of the earth by resistive electric heating and that the introduction of the lightning rods invented by Franklin had upset the natural flow, thereby bringing on the cold weather. As far as we have been able to determine no one at the time attributed the strange weather to the massive eruption of Mount Tambora the year before, in spite of the fact that Franklin's speculations on the meteorological effects of atmospheric dust were more than 30 years old and that news reports of the bad weather appeared in many newspapers side by side with reports of large floating islands of volcanic ash in the Pacific.

JPL proudly presents...

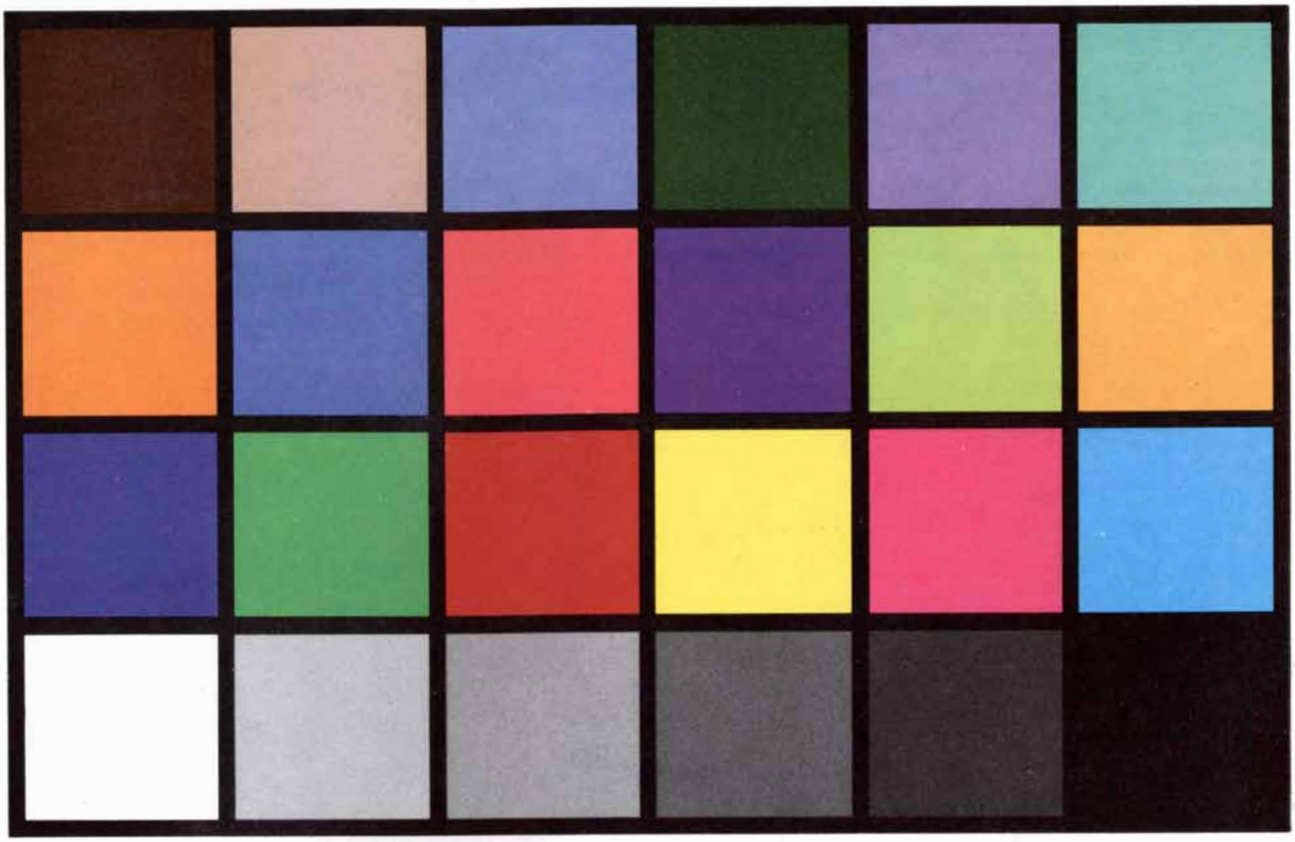
JUPITER.

At JPL, even we've been surprised and excited by the data NASA's Voyager I spacecraft has been sending back to Earth about Jupiter. Like the ring around Jupiter and the volcanoes on Io. We can't wait to see what's new on Saturn and Uranus. In fact, we're taking a close look at most of outer space. Plus, we're probing the depths of the ocean and seeking new energy sources. If you would like a career with excitement, send us a resume.

Jet Propulsion Laboratory
California Institute of Technology
Attention: Professional Staffing
4800 Oak Grove Drive
Pasadena, CA 91103

THERE'S ONLY ONE JPL

Voyager I photo of Jupiter's turbulent atmosphere and the Great Red Spot.



The ColorChecker chart photographed by Robert Szabo in his experiments

Dark skin	Light skin	Blue sky	Foliage	Blue flower	Bluish green
Orange	Purplish blue	Moderate red	Purple	Yellow green	Orange yellow
Blue	Green	Red	Yellow	Magenta	Cyan
White (.05)†	Neutral 8* (.23)†	Neutral 6.5* (.44)†	Neutral 5* (.70)†	Neutral (3.5)* (1.05)†	Black (1.50)†

* Color notation † Optical density

Identification of the colors on the chart

THE AMATEUR SCIENTIST

Experiments with Edwin Land's method of getting color out of black and white

by Jearl Walker

In the 1950's Edwin H. Land put forward a set of deceptively simple experiments designed to investigate the stimuli that give rise to color vision. In one experiment he photographed a colored scene twice with black-and-white film, once through a red filter and once through a green one. He then projected the two black-and-white transparencies so that they were superposed on a screen. In front of the transparency made through the red filter he placed the same red filter. No filter was placed in front of the other transparency.

What one would expect from this arrangement is a reproduction of the original scene in various shades of white, pink and red because of the red filter in front of one of the projectors. Surprisingly, what Land saw on the screen was a reproduction of most of the colors of the original scene. Although the colors were dull, they were there, notwithstanding the fact that Land was projecting two black-and-white slides.

For more than 20 years Land and John J. McCann and their colleagues have endeavored to explain how the human visual system (the eye and the visual cortex) perceives the subtle differences of shading in what is viewed. Land described the original work in this magazine in 1959 [see "Experiments in Color Vision," by Edwin H. Land; SCIENTIFIC AMERICAN, May, 1959]. The two decades of experimentation since then have yielded a deeper understanding of the color process, as can be gathered from Land's more recent article [see "The Retinex Theory of Color Vision," by Edwin H. Land; SCIENTIFIC AMERICAN, December, 1977].

Robert Szabo, a student at Cleveland State University, not long ago repeated Land's procedure of making black-and-white transparencies through color filters and then projecting the transparencies superposed on a screen with a single color filter. Although the results are not as sharply colored as Land's earlier results apparently were, they are easy for an amateur photographer to obtain and they do show essentially what Land saw.

On occasion Land has photographed still lifes of fruits and other such objects and at other times he has employed a complex array of colored papers that he calls a "Mondrian." Szabo chose to photograph a fairly new color chart, the ColorChecker, produced by Macbeth, a division of the Kollmorgen Corporation. The chart is apparently becoming a standard reference in all kinds of color photography because of the spectral design of the squares arranged on it. Each square is designed to closely mimic a natural color a photographer might encounter. For example, one square has a blue that closely matches a typical blue sky not only in the dominant blue but also throughout the visible spectrum.

An advantage in photographing a standard reference is that one can at any time compare the reference with the projected images. A still-life scene does not have such permanence, particularly if it includes fruit. Another advantage is that one's results might be more easily conveyed to another person if both of them employ the standard reference. The Mondrian array does not lend itself to that flexibility.

The ColorChecker may also present a disadvantage: the results of the experiment may be partly affected by the colors surrounding a particular square in the projected display. With the random array of the Mondrian's small sections, each of which has various colors and is surrounded in a variety of ways, the influence of surrounding colors can be avoided.

When Szabo photographed the color chart, he mounted his camera on a sturdy stand and triggered it remotely so that he could avoid jarring it. Two tungsten floodlights were aimed at an angle of about 45 degrees to the tabletop on which the chart was laid. Instead of high-quality color filters Szabo chose less expensive but still good filters that are available from the Edmund Scientific Company (6982 Edscorp Building, Barrington, N.J. 08007). The filters come in a booklet of 44. Each filter is five by eight inches, a convenient size for photographing and projecting. A booklet is also available from the company to show the spectral-transmission characteristics of the filters.

The best type of film is thought to be Polaroid's black-and-white transparency film because it responds to light fairly evenly across the visible spectrum. Szabo decided to try the relatively inexpensive Pan-X black-and-white film. With a Kodak direct-positive processing kit he was able to develop and mount his own slides. As he photographed the color chart he first set his camera at the correct exposure (*f*-stop and exposure speed) for the lighting conditions and then bracketed those values by photographing the chart at other nearby exposures. In order to be able to compare the slides later he carefully recorded the exposure data for each frame. Two carousel slide projectors served for projecting the slides on a conventional screen. The lights for both the projection and the photographing should both be good white lights, matched reasonably closely in their color composition.

The slide made through the filter that passes primarily the longer wavelengths is called the long record. The other slide, made with the filter that passes primarily the shorter wavelengths, is the short record. In following Land's experiments Szabo made a great many trials, employing a variety of filters and projection methods. Some of his results are given in the table on page 195.

In the simplest kind of trial there was a short and a long record; when a filter was included in the projection, it was the filter for the long record in front of the projector holding the long record. For example, in one trial the long record was made through the Edmund filter 821, which transmits primarily in the red. The short record was made through filter 856, which transmits mainly in the blue. What Szabo then had were two black-and-white slides that had been individually photographed through filters transmitting at opposite ends of the visible spectrum. The slides provided no readily apparent color information, and even if they had embodied a color code of some kind, it would presumably not have included green and yellow because neither of the filters transmitted the green and yellow wavelengths well.

When Szabo projected the slides, he put filter 821 in front of the projector holding the long record. On superposing the two projected images he found a reproduction of the ColorChecker that showed not only red but also some blue, yellow and green (but no purple).

The slides Szabo projected had been photographed at the proper exposures for the illumination provided by his floodlights. When he checked the slides he had made at somewhat longer and shorter exposures, he generally found that the projected image had about the same color characteristics, provided the slide had not been too strongly overexposed or underexposed. Land had said that the intensity level of the projectors was not critical. Whatever color coding the black-and-white slides showed did not depend critically on the brightness of the light. Szabo and I checked that result by mounting two crossed polarizing sheets in front of one of the projectors. By adjusting the angle between the polarizing axes of the sheets we could vary the brightness of the image cast on the screen by the projector. The brightness was not critical to the projection of the colors we saw.

In another trial Szabo used for the short record a slide made through filter 817 and for the long record the slide made earlier through filter 821. Filter 817 transmits more in the orange than 821. Again the long-record filter 821 was placed in front of the long-record projector. The projected superpositioning was quite faint and showed mainly pink. The transmission characteristics of the two filters were apparently too similar to give rise to a full spectrum in the projection. If the same filter had been used for photographing both slides, no colors would have been produced except the ones provided by a filter placed in front of one of the projectors.

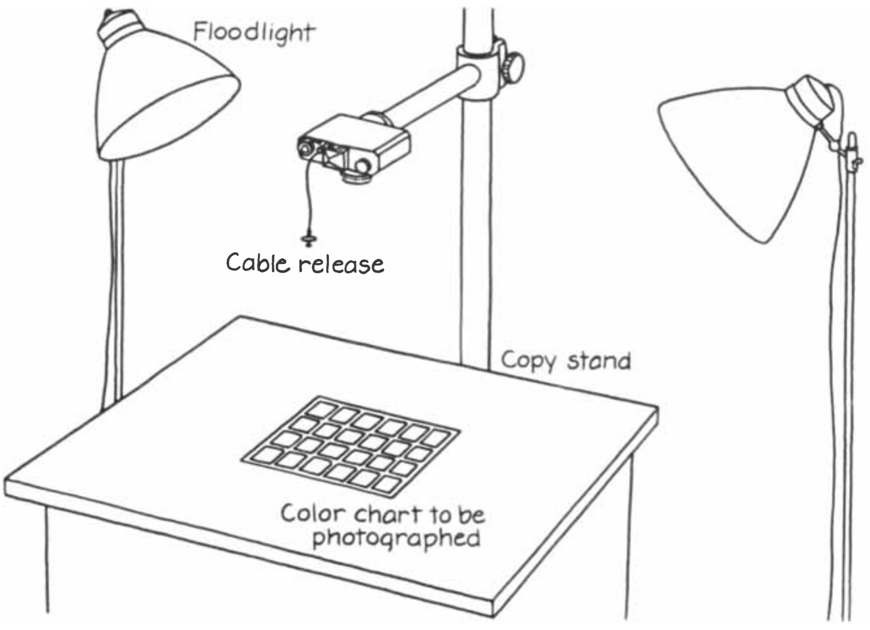
The first article by Land in this magazine included a graph with which one can predict which colors will not appear in the projected images. For example, Szabo projected his slides made through filter 821 (which has a peak transmission at wavelengths between 650 and 700 nanometers) and filter 856 (which has a peak transmission at about 475 nanometers). The corresponding point on the graph lies in a region labeled "No purple." Szabo indeed saw no purple.

In one region on the graph, running along the diagonal, the long and short records are too nearly the same to produce colors in the projected image. Szabo's filters 817 and 821 apparently corresponded to a point near the diagonal. An unmarked region below the diagonal corresponds to a situation where the filter for the long record serves in the projection of the short record and vice versa. Both Szabo and I found it amusing to hold the long-record filter in front of the long-record projector and then move it in front of the short-record projector to get a color reversal.

Szabo also tried projecting with the short-record filter in front of the short-record projector and with no filter in front of the long-record projector. The colors were usually not as good. For example, he photographed through filter 858 (which transmits blue green) and filter 863 (which transmits more in the blue); the projected image consisted of browns, blues and blue greens.

In addition Szabo tried stacking the filters, that is, using two or more at the same time in order to narrow the band pass of wavelengths being transmitted. In one trial he stacked filters 861, 863 and 866 to eliminate any light from being transmitted in the range from 550 to 700 nanometers, which corresponds to yellow, red and some green. He made the long record with filter 821 and projected the slides with that filter in front of the long-record slide. Szabo found no yellow in the projection, but the other colors were present.

In one series Szabo photographed the long record through a stack (filters 874 and 877) to give a narrow band pass centered near 500 nanometers. The short record was photographed through filter 866, which transmits more in the blue with a transmission peak at 450 nanometers. When he put the long-record filters in front of the long-record slide, he



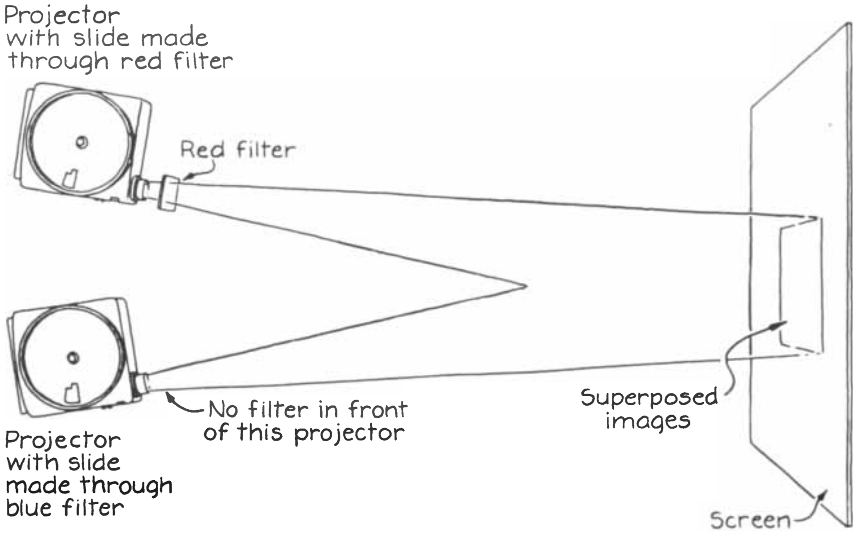
Szabo's setup for photographing the chart got blues and greens. When instead he placed the short-record filter in front of the long record, he got blues and a dull yellow (a color reversal, since the "wrong" filter was in front of the long record). Out of curiosity he then put those filters aside and placed in front of the long-record slide still another filter, 807, which transmits in a broad band ranging from 525 nanometers through 775 nanometers. Although the projection filter was not the proper one, some dull color did emerge in the projection.

Land said in his first article that the colors will emerge even if the short record is made without a filter. Szabo combined such an unfiltered slide in an unfiltered projection with some of his long-record slides. When the long-record filter was placed in front of the long-record projector, dull colors appeared on the screen.

Why do any of the slide combinations show colors in their superposed projection? The complete explanation is still not known, but Land and his colleagues have narrowed the search considerably in the past two decades. In a complex scene color is not assigned to a particular area according to the color theory in which the three visual pigments in the human retina serve as three separate photometers. Since the absorption curves of the pigments peak at 440, 535 and 565 nanometers, the photometers are said to be sensitive to different areas of the visual spectrum and to span between them the entire visual spectrum. The theory holds that when the energy falling on the retina at, say, 565 nanometers is greater than that at the other peaks, the observer perceives the area viewed as being reddish. Conversely, a preponderance of energy at 440 nanometers yields a perception of blue.

Such was the textbook explanation for a long time, at least in simple discussions of color theory. Then Land demonstrated that color was independent of the energy flux into the eye at any particular wavelength. For example, in one experiment he caused the energy from two areas to be the same at the three absorption peaks. According to simple color theory, the two areas should then appear to be the same color. When the areas were viewed as part of the complex Mondrian array, however, an observer would discover, for example, that one area was green and the other was red. Hence the energy fluxes in themselves were not the source of the color information being processed by the observer.

In a complex scene color is assigned as a result of two main processes. First the visual system compares the energies at the three absorption peaks (the receptors for the short, middle and long wavelengths) and in effect assigns a triplet of numbers accordingly. Color has not yet been determined. Next the visual system compares that area with other areas in



Szabo's arrangement for projecting slides

the scene, assigning to each area its own triplet of numbers. Finally the system can begin to assign color after comparing the triplets from a large number of areas. For example, if a triplet from a particular area is high in the long wavelengths, somewhat lower in the middle wavelengths and lower still in the short wavelengths, it will be assigned a reddish color.

In the first of the experiments I have described Szabo photographed the color chart through a red filter and then separately through a blue filter. The long record showed black-and-white shadings according to the reflectance of the various squares in the red wavelengths. Red areas appeared light in the slide, orange areas a bit less light and blue areas relatively dark. In the short record the situation was reversed: blue areas appeared light, orange areas darker and red areas darkest. These relative shadings code the color information that can be scanned and decoded by an observer.

When the long record was projected with the same red filter, the projected scene was the same (to the limits of the accuracy of the film and the whiteness of the projection lamp) as the one transmitted through the filter in the photography. Blue areas were projected dark, orange areas less dark and red areas light. The color available in the light areas was, of course, only pink or red. When that projector was turned off and the short record was then projected with no filter in front of it, the projected image showed a range of dark and light regions reflecting what had been transmitted through the blue filter. This time the projection had only shades of dark gray, light gray and white light, but they recorded the degree of blue light coming from each of the squares on the color chart.

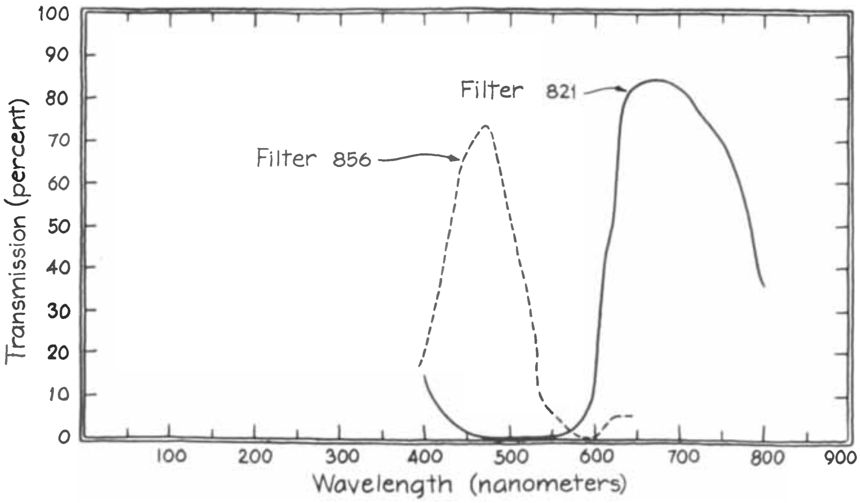
When both projectors are turned on and the images are superposed, why do you see colors other than just pink and red? First consider a square you will see as red. In the superposed images the square is light in the long wavelengths, not as light in the middle wavelengths and relatively dark in the short wavelengths. Why? Because the long record is providing relatively bright light in the long wavelengths and the short record is providing relatively little light in the short wavelengths. In a sense the visual system computes the ratios of lightness in each of the wavelength regions and assigns to the square a triplet of numbers according to the ratios.

Now consider a square that is blue in the projection. The long record is providing little of the long-wavelength light to the square, whereas the short record is providing relatively bright light. Again the visual system compares the lightness at the short, middle and long wavelengths and assigns a triplet of numbers. The color blue is not yet assigned, because the observer's visual system has not yet understood what this particular triplet of numbers means as to color.

The process is repeated for every square seen in the projection, apparently automatically and without necessarily sweeping the eyes over the images. Every square is assigned a certain triplet of numbers according to how light it is in the short, middle and long wavelengths. Then the triplets are compared. A square with a triplet indicating that it is light in equal amounts in all three wavelength regions is assigned white. Another square with a triplet indicating less light in the middle wavelengths and still less in the short wavelengths is assigned red. A third square, which in comparison with the white square has less light in the middle wavelengths and still less in the long wavelengths, is assigned blue. A square that is light in the middle wavelengths, a bit dimmer in the long wavelengths and still darker in the short wavelengths may be assigned yellow. And so it goes until as much color as it is possible to interpret from the projection has been assigned.

Suppose the chart had no reference white square. Would the process of assigning colors be destroyed? No. The triplets would still be assigned to each square and then compared. Since red light is being projected from the long record, it can act as a reference point. For example, a triplet that is relatively deficient in the long wavelengths, less deficient in the middle wavelengths and stronger in the short wavelengths could be assigned blue in comparison with the red square. The experiment works better (in the sense that the colors are more easily distinguished) if the filters employed in the photography have distinctly separate band passes. Then the light and dark shadings of the slides are more distinguishable.

The explanation of color assignment I have presented here is only a brief summary of a much richer model developed by Land and McCann and their colleagues. You should read their articles



The band passes of two filters

New TI Programmable 58C with Constant Memory™ feature. Still only $125.*

If you enter a 480-step program today, the new TI Programmable 58C will remember it tomorrow...even if you turn it off overnight!

We haven't changed the price, but we've made the TI-58 even more versatile by adding a Constant Memory feature that retains data or program information even when the calculator is turned off. You can turn off your TI Programmable 58C right in the middle of the most complex problem, for hours or even days, and pick up right where you left off when you turn it back on.

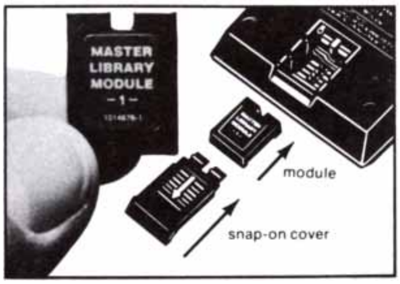
The TI Programmable 58C, with plug-in application libraries and more than 170 functions and operations in scientific, engineering, and statistical fields ... now an even greater value with Constant Memory feature at no extra cost.

More easy-to-use software.

Take advantage of the versatility of the new TI Programmable 58C and new software. Twelve optional *Solid State Software*™ libraries are now available. Each plug-in library module contains dozens of prewritten programs. Solid-state libraries are available in a variety of fields including engineering, science, statistics, and business. From $35.*

The *Professional Program Exchange* lets you exchange programs with your colleagues. Over 1,500 programs are already available to members.

Specialty Pakettes are conveniently formatted program listings available in 16 subject areas. Just follow the program listings and you are ready to tackle a wide variety of problems without the need to do any conventional programming. And, with the TI Programmable 58C's Constant Memory feature, you won't lose your program if you have to turn the calculator off. $10* per subject.



In addition, *Custom Software Modules* can be developed for specialized applications and a unique *Professional Productivity Program* is available for corporate use.

Printing and plotting capability.

Add the PC-100C printer/plotter to your TI Programmable 58C and you can plot, prompt, and print inputs and outputs automatically. Enjoy the convenience and efficiency of a whisper-quiet printing calculator for hard copy printouts. $200.*

Free literature.

We want to tell you more about our products and services. For additional information write: Texas Instruments, Programmable Calculators, P.O. Box 53, Lubbock, Texas 79408.

*U.S. suggested retail price.

Texas Instruments technology –bringing affordable electronics to your fingertips.

TEXAS INSTRUMENTS
INCORPORATED

45645

to get a better explanation. Since Szabo's experiments are inexpensive and fairly easy to do, you might want to try some of them yourself.

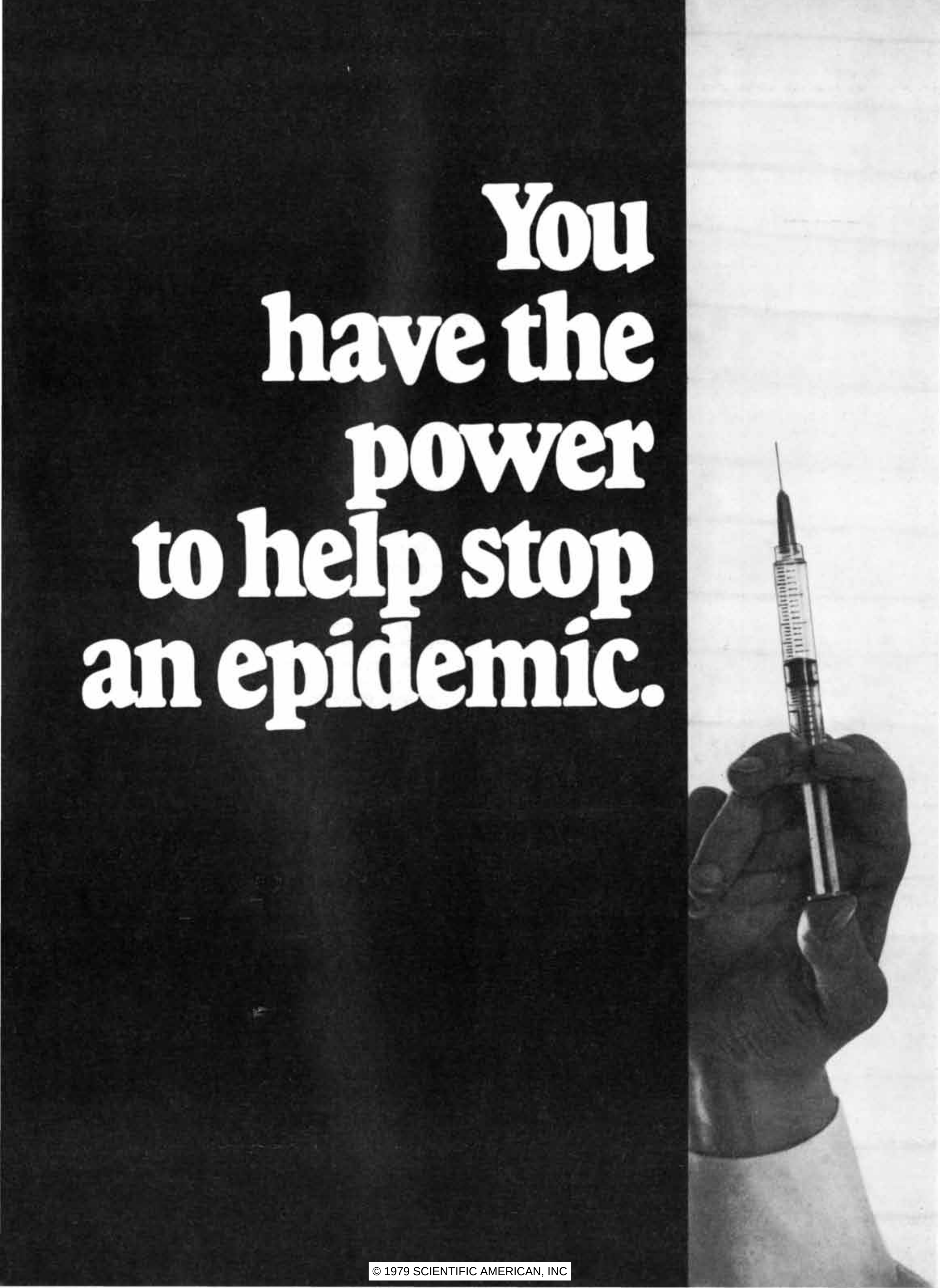
I have received quite a bit of mail about the stereoscope and stereoscopic pictures, which I described in my December column. Apparently many people are making their own stereoscope photographs and slides. One organization, the National Stereoscopic Association, offers a bimonthly journal, *Stereo World,* to its members, some 800 strong. (If you are interested, write to John Weiler, P.O. Box 14801, Columbus, Ohio 43214.) The members buy, sell and make samples of stereoscopic photography.

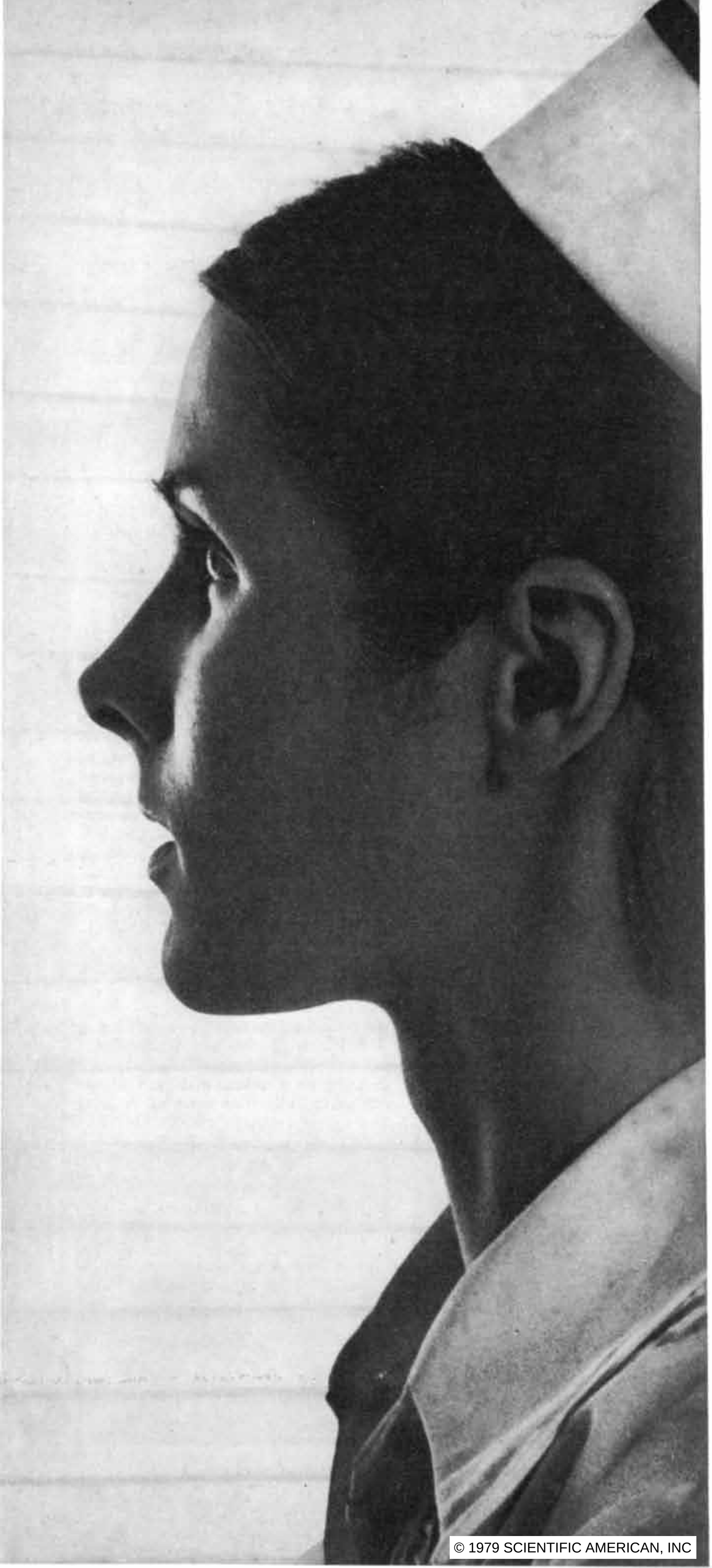
Some readers suggested that I mention two points about making your own stereoscope photographs. First, a stereoscopic pair in which the second exposure is made by rotating the camera has

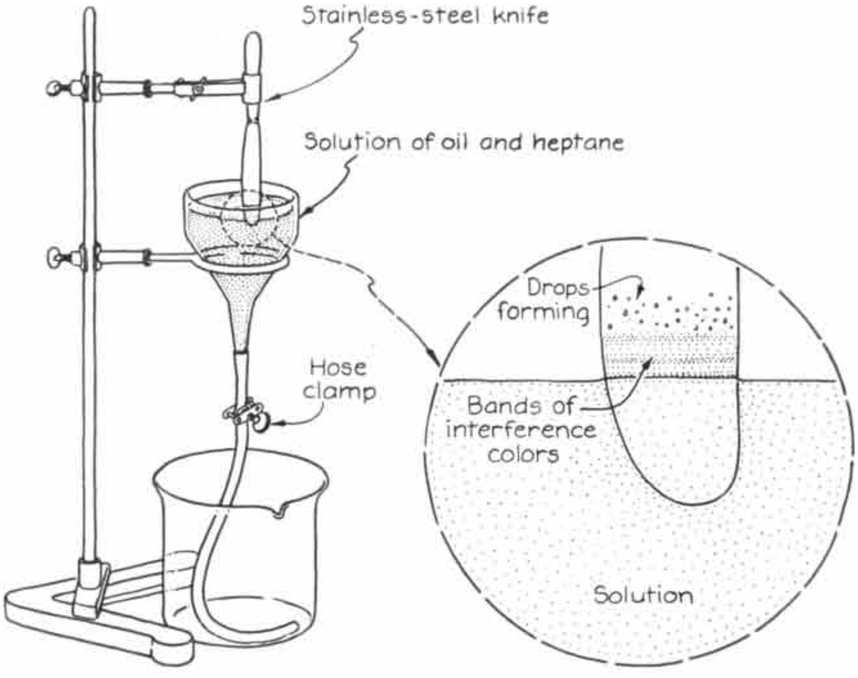
Photographing filter for short record	Photographing filter for long record	Projection filter	Comments
Green 874	Red 821		Gray scale O.K. Washed-out greens. Reds O.K. White O.K. Reversed, stronger cyan — much brighter.
Green 877	Red 821		Same as above.
Green 871	Red 821		Gray scale good; dull greens, white; faint blue green.
Green 874, 877	Red 821		Good green combination.
Green 871, 874	Red 821		Not as good as above.
Green 874, 877	Red 821, 823		Not good.
Green 874, 877	Red 823		Brownish tint, bad reds.
Green 874, 877	Orange 819		Bright oranges.
Green 874, 877	Orange 819		Darker and redder than above.
Blue 851	Orange 819		Cyan good, orange O.K.
Orange 819	Red 821		Washed out.
Orange 817	Red 821		Even lighter.
Orange 818	Red 821	Same	Very white.
Orange 818	Red 821	as for photographing	More color.
Blue 851	Red 821	for	Good green and red. No yellow or violet.
Violet 843	Red 821	long record	Brighter than above.
Blue 856	Red 821		Very good color, a little yellow.
Blue 858	Red 821		Very good red and blue greens.
Blue 859	Red 821		Lighter than above, but good.
Blue 863	Red 821		No yellow, red and greens good, no violet, a little blue.
Blue 866	Red 821		Flesh tones and gray scale very good. Magenta washed out, no yellow.
Blue 863, 866	Red 821		Average; dark colors wrong.
Blue 863, 866	Red 821		Good gray scale and flesh tones.
Blue 863, 866	Red 821		Reds and blue greens, no pure blue or green, no yellow.
Yellow, blue 807, 861	Red 821		Washed out.
Blue 861, 863, 866	Red 821		No yellow, but good color.
Blue 858	Orange 817		Orangy tint, gray scale brownish.
Blue 858	Yellow 809	Orange 817	Same as above.
Blue 866	Green 874, 877	Blue 866*	Not as good, but brings out yellow.
Blue 863	Green 874, 877	Blue 863*	Better yellow, gray scale appears slightly blue.
Blue 866	Blue 863	Blue 866*	No effect.
Blue 859	Blue 863	Blue 863	Very little color.
Blue 858	Blue 863	Blue 858*	Brown and blue greens; gray scale brownish.
Blue 856	Red 821	Red 821	Very good!
Blue 856	Red 821 (longer exposure)	Red 821	Color more saturated as long-record slide is less dense. Also very good combination.
Blue 856	Red 823	Red 823	Muted.
Blue 856 (larger stop)	Red 823	Red 823	Better than above, but dull. Gray scale brownish.
Blue 856	Orange 817	Orange 817	No red, but yellow and orange. Gray scale brownish, dull.
Blue 866	Green 874, 877	Yellow 807	Dull colors.
Blue 866	Green 874, 877	Green 874, 877	Green colors.
Blue 866	Green 874, 877	Blue 866	Blues and dull yellow.

* Filter used in front of short record

Szabo's observations on a series of experiments


You
have the
power
to help stop
an epidemic.





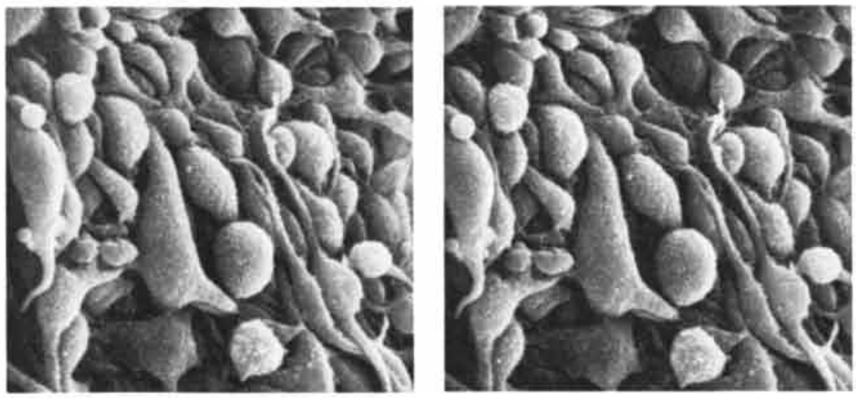
A. A. Fote's setup for laying a thin film of oil on metal

less depth than one in which the second exposure is made by displacing the camera to the side. Second, one should not think that putting a pair of identical photographs in the stereoscope can achieve the same effect as putting in a true stereoscopic pair. With both kinds of pair the eye can utilize depth clues relating to the size of objects and the converging geometry of things, but with identical photographs depth clues from the convergence of the eyes are of course not available.

An apparently controversial subject is whether displacing the camera more than the interocular distance between photographs gives rise to exaggerated depth in the stereoscopic pair. One reader, Philip R. Pennington of Portland, Ore., said he had never seen such exaggeration even in stereoscopic pairs made several miles apart. Other readers said they regularly make use of the depth distortion offered by wide base lines, particularly in aerial photography designed to enhance the contour variations of the ground.

Some of the most beautiful samples of stereoscopic pictures that I have received came from James Raymond of Fairbanks, Alaska, who photographed Mount McKinley with a base line of about 2,000 feet while flying in an airplane. An amazing stereoscopic pair came from Ralph C. Eagle of the University of Pennsylvania School of Medicine. They were micrographs made with the scanning electron microscope, showing cancer cells in a specimen of the human iris. The specimen had been rotated by seven degrees between exposures, and depth is readily apparent.

In discussing the colors in soap films in my column for last September I briefly described the problem of laying a thin film on a surface that has a higher refractive index than the film. A good



Ralph C. Eagle's stereoscopic-pair micrographs of cancer cells way to accomplish the feat has been sent to me by A. A. Fote of the Aerospace Corporation in Los Angeles. In order to investigate the behavior of lubricating oils under space conditions he prepares thin oil films on polished metal surfaces. When the films are sufficiently thin, they display interference colors similar to the ones seen in many soap films.

Fote mixes a solution of Apiezon *C* oil with heptane so that the oil makes up from 20 to 40 percent of the total volume. The solution is poured into a container that has a controlled leak at the bottom, and the flat metal piece to be coated is suspended vertically inside the container. As the solution drains from the bottom a thin layer remains on the metal. Since the heptane evaporates almost immediately, the film is oil only. The thickness of the oil film can be controlled somewhat by changing the concentration of the oil in the solution and the rate at which the solution leaks out. Thinner film layers result from lower concentrations and slower drainage.

At Fote's suggestion I replaced the Apiezon *C* with ordinary mineral oil, which is also soluble in heptane. For the bath I set up a funnel to drain into a beaker through a hose that was partly closed by a clamp of the screw type. I poured the solution into the funnel and controlled the rate of drainage by adjusting the clamp. Then I suspended a small, flat piece of metal in the solution. As the solution drained, the piece of metal was coated with a layer of mineral oil. To form a film with a uniform thickness you must be sure the solution drains at a constant rate. To that end I worked with a funnel that had vertical sides at the top, and I kept my piece of metal in the upper section.

The relative indexes of refraction of the film and the metal govern the colors one sees in thin oil films and whether with ultrathin films one sees a white film or a dark one. Some metals have a high index of refraction (steel about 2.5); others have lower values (copper about .6 and silver about .2). When an oil is applied as a layer on a surface that has an index of refraction higher than that of the oil, ultrathin layers are transparent and show no colors. If white light is shined on the layer, white light is reflected to the observer. On the other hand, when the oil has the larger index of refraction, an ultrathin film is dark.

Somewhat thicker films give interference bands and a variety of colors. You have probably noticed such colors on metal cookware. When the metal surfaces of pots and pans oxidize, the thin oxide layer acts as the thin film and contributes colors created by the interference of light.

The first metals I tried in my oil bath gave no interference colors, although I could see the oil films clearly. Fote explained that my metals were probably not sufficiently polished to make interference possible. At least three factors are involved. First, thinner layers (thin enough for perceptible interference) may be more likely on a highly smooth metal surface. Second, a rough metal surface will destroy the coherence of the light waves that is necessary for interference. Third, with a rough surface the thickness of the film layer will vary so rapidly from point to point that the interference colors will wash out into a perceived whiteness.

To understand the second factor one must follow the path of a light ray in the film. A ray incident on a film splits into two rays: one that is reflected immediately and another that passes through the film and is reflected off the back of it. If the back surface, lying next to the metal, is rough, the ray will not be reflected outward in the same direction as the ray reflected from the front surface. The waves in the rays therefore cannot interfere. If a metal surface is to show interference colors in an oil film covering it, the surface should be polished so brightly that you can see your reflection in it. I obtained clear, beautiful colors with a stainless-steel kitchen knife. The colors included pink, blue and green, with some areas of yellow and brown. The thinner the film was, the wider these bands were, as is the case with soap films.

It seems to be more difficult to predict the colors with these oil films than it is with soap films because the intensity of the reflection of light from metal surfaces varies considerably with wavelength and with angle of reflection. For a highly polished metal the reflection from the back surface of the film is so much stronger than the reflection from the front surface that the interference colors are lost in the white glare from the back reflection. The glare appears to be particularly strong when the incident light is approximately perpendicular to the film.

When I first dipped a piece of metal in the oil bath (before I began the drainage), a film climbed about .5 centimeter up the metal surface. After a while the film formed "tears" that ran back down the metal. On my kitchen knife I could see interference colors in the film except where the thicker drops descended. The climbing film and the formation of drops have been noted in other circumstances since about 1850, when the effect was observed in glasses of wine. Usually the climbing is attributed to the fact that fluid is being pulled upward by unequal surface tensions in the vertical film, an effect named after Carlo Marangoni.

Readers interested in the structures and colors of soap films might want to look into an excellent book, *The Science of Soap Films and Soap Bubbles,* by Cyril Isenberg, recently published by Tieto Ltd., 5 Elton Road, Clevedon, Avon, England.

SCIENCE/SCOPE

Apparel manufacturers have improved productivity with a computerized system that lays out pattern pieces more efficiently than manual methods. The Hughes AM-1 automatic marker has a TV-like interactive console that an operator uses to fit pattern pieces into the most economical arrangement possible. The procedure is easier and more precise than the old method of placing full-size cardboard cutouts on room-size tables. The system increases fabric utilization by 1 to 5 percent -- a margin that translates into substantial savings for companies that annually spend millions of dollars for material.

A communications terminal almost one-third the size and less than half the weight of the three pieces of equipment it replaces serves a key role in an advanced military network. The Hughes Improved Terminal (HIT) combines a transmitter-receiver, signal processor, and computer into one unit that's more reliable and less costly to build than the separate units. HIT is designed to let all four military services exchange data instantaneously and securely via the Joint Tactical Information Distribution System. The terminal can transmit coded digital data over a single channel in preassigned time slots of several milliseconds. It can receive all information sent by other units or simply select what it wants.

Communications via satellite continue to cost less every year despite inflation. The International Telecommunications Satellite Organization, crediting improved technology and efficiency of its Intelsat network of satellites, has cut its monthly charge for a full-time, two-way telephone circuit by 16 percent to $960. The same service in 1965 initially cost $5334. If that charge had risen with inflation, the cost today would be about $11,000. Intelsat, a consortium of more than 100 nations, has lowered its rates for nine consecutive years. The satellites presently providing the service were designed and built by Hughes.

Are you a graduate EE, ME, or physicist with experience in project or systems engineering, optics, product design, reliability and test? Can you fit in with a very bright scientific team working on long-term high technology projects that are advancing the state-of-the-art in: lasers, electro-optics, automatic test systems, digital and analog computers, airborne space sensors, electronic/electromechanical components and devices, and a myriad of other far-sighted technical and strategic systems? If so, and if you seek challenge and just reward, contact Hughes Aircraft Company, Professional Employment, Dept. SE, Electro-Optical & Data Systems Group, 11940 W. Jefferson Blvd., Culver City, CA 90230.

Technical papers covering a wide variety of technologies are available without charge from Hughes. Three recent reprints in the fields of lasers, radar, and communications are entitled "New System for Accurate Laser Intensity Profile Analysis," "Nonuniformly Spaced Linear and Planar Array Antennas for Sidelobe Reduction," and "Selective Frequency Spacing and the Half-Step Synthesizer." For these papers or a list of others, write to: Publications Clearance, Hughes Aircraft Company, Bldg. 6, M.S. E-110, Culver City, CA 90230.

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

ÜBER DIE „DOPPELT-PERIODISCHEN" LÖSUNGEN DES *N*-DAMEN-PROBLEMS. G. Pólya in *Mathematische Unterhaltungen und Spiele,* by W. Ahrens. Verlag und Druck von B. G. Teubner, Leipzig, 1918.

SAM LOYD AND HIS CHESS PROBLEMS. Alain C. White. Dover Publications, Inc., 1962.

ON COLORING THE $N \times N$ CHESSBOARD. M. R. Iyer and V. V. Menon in *The American Mathematical Monthly,* Vol. 73, No. 7, pages 721–725; September, 1966.

INDEPENDENT PERMUTATIONS, AS RELATED TO A PROBLEM OF MOSER AND A THEOREM OF PÓLYA. Ashok K. Chandra in *Journal of Combinatorial Theory, Series A,* Vol. 16, No. 1, pages 111–120; January, 1974.

EXTERNAL HUMAN FERTILIZATION

HAMILTON, BOYD AND MOSSMAN'S HUMAN EMBRYOLOGY: PRENATAL DEVELOPMENT OF FORM AND FUNCTION. Fourth edition, by W. J. Hamilton and H. W. Mossman. W. Heffer & Sons Ltd, 1972.

FERTILIZATION OF HUMAN EGGS IN VITRO: MORALS, ETHICS AND THE LAW. R. G. Edwards in *The Quarterly Review of Biology,* Vol. 49, No. 1, pages 3–26; March, 1974.

RESEARCH ON THE FETUS. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Department of Health, Education, and Welfare Publication No. (OS) 76–128, 1975.

THE NERVE-GROWTH FACTOR

EXCESSIVE GROWTH OF THE SYMPATHETIC GANGLIA EVOKED BY A PROTEIN ISOLATED FROM MOUSE SALIVARY GLANDS. Rita Levi-Montalcini and Barbara Booker in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 46, No. 3, pages 373–384; March, 1960.

PURIFICATION OF A NERVE-GROWTH PROMOTING PROTEIN FROM THE MOUSE SALIVARY GLAND AND ITS NEURO-CYTOTOXIC ANTISERUM. Stanley Cohen in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 46, No. 3, pages 302–311; March, 1960.

THE NERVE GROWTH FACTOR: ITS MODE OF ACTION ON SENSORY AND SYMPATHETIC NERVE CELLS. Rita Levi-Montalcini in *The Harvey Lectures,* Series 60, pages 217–259; 1966.

NERVE GROWTH FACTOR FROM MOUSE SUBMAXILLARY GLAND: AMINO ACID SEQUENCE. Ruth Hogue Angeletti and Ralph A. Bradshaw in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 68, No. 10, pages 2417–2420; October, 1971.

THE RETROGRADE AXONAL TRANSPORT OF NERVE GROWTH FACTOR. I. A. Hendry, K. Stöckel, H. Thoenen and L. L. Iversen in *Brain Research,* Vol. 68, No. 1, pages 103–121; March 15, 1974.

NERVE GROWTH FACTOR. William C. Mobley, Alfred C. Server, Douglas N. Ishi, Richard J. Riopelle and Eric M. Shooter in *The New England Journal of Medicine,* Vol. 297, No. 20, pages 1096–1104, November 17, 1977; Vol. 297, No. 21, pages 1149–1158, November 24, 1977; Vol. 297, No. 22, pages 1211–1218, December 1, 1977.

NERVE GROWTH FACTOR POTENTIATES ACTOMYOSIN ADENOSINETRIPHOSPHATASE. Pietro Calissano, Gianni Monaco, Loriana Castellani, Delio Mercanti and Andrea Levi in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 75, No. 5, pages 2210–2214; May, 1978.

AUTOMATIC CONTROL BY DISTRIBUTED INTELLIGENCE

DIRECT DIGITAL PROCESS CONTROL: PRACTICE AND ALGORITHMS FOR MICROPROCESSOR APPLICATION. D. M. Auslander, Y. Takahashi and Y. Tomiyuka in *Proceedings of the IEEE,* Vol. 66, No. 2, pages 199–208; February, 1978.

SPECIAL ISSUE ON LARGE-SCALE SYSTEMS AND DECENTRALIZED CONTROL. *IEEE Transactions on Automatic Control,* Vol. AC-23, No. 2; April, 1978.

THE THEORY OF KNOTS

KNOTENTHEORIE. Kurt Reidemeister. Chelsea Publishing Co., 1948.

KNOTS AND LINKS. Dale Rolfsen. Publish or Perish, Inc., Berkeley, Calif., 1976.

INTRODUCTION TO KNOT THEORY. Richard H. Crowell and Ralph H. Fox. Springer-Verlag, 1977.

A NEOLITHIC FLINT MINE

LA STATION NÉOLITHIQUE DE SAINTE-GERTRUDE. Joseph Hamal-Nandrin and Jean Servais in *Revue Anthropologique,* Vol. 33, pages 345–492; 1923.

DER ÄLTESTE BERGBAU IN EUROPA. Martin Jahn. Akademie-Verlag, Band 52, Heft 2, 1960.

5000 JAHRE FEUERSTEINBERGBAU IN DEN NIEDERLANDEN. F. H. G. Engelen in *Der Anschnitt,* Vol. 21, No. 2, 1969.

ULTRACOLD NEUTRONS

ELECTRIC DIPOLE MOMENT OF THE NEUTRON. R. Golub and J. M. Pendlebury in *Contemporary Physics,* Vol. 13, No. 6, pages 519–558; November, 1972.

ULTRACOLD NEUTRONS. V. I. Luschikov in *Physics Today,* Vol. 30, No. 6, pages 42–51; June, 1977.

ULTRA-COLD NEUTRONS. R. Golub and J. M. Pendlebury in *Reports on Progress in Physics.* Institute of Physics, March, 1978.

THE HISTORY OF THE ATLANTIC

AN ANALYSIS OF THE VARIATION OF OCEAN FLOOR BATHYMETRY AND HEAT FLOW WITH AGE. Barry Parsons and John G. Sclater in *Journal of Geophysical Research,* Vol. 82, No. 5, pages 803–827; February 10, 1977.

THE PALEOBATHYMETRY OF THE ATLANTIC OCEAN FROM THE JURASSIC TO THE PRESENT. John G. Sclater, Steven Hellinger and Christopher Tapscott in *The Journal of Geology,* Vol. 85, No. 5, pages 509–552; September, 1977.

THE YEAR WITHOUT A SUMMER

MIGRATION FROM VERMONT (1776–1860). Lewis D. Stilwell in *Proceedings of the Vermont Historical Society,* Vol. 5, No. 2, pages 63–241; 1937.

THE CHOLERA EPIDEMIC OF 1832 IN NEW YORK CITY. Charles E. Rosenberg in *Bulletin of the History of Medicine,* Vol. 33, No. 1, pages 37–49; January–February, 1959.

A STUDY IN METEOROLOGICAL AND TRADE CYCLE HISTORY: THE ECONOMIC CRISIS FOLLOWING THE NAPOLEONIC WARS. John D. Post in *The Journal of Economic History,* Vol. 34, No. 2, pages 315–349; June, 1974.

STATISTICAL EVIDENCE ON THE INFLUENCE OF SUNSPOTS AND VOLCANIC DUST ON LONG-TERM TEMPERATURE RECORDS. Clifford Mass and Stephen H. Schneider in *Journal of the Atmospheric Sciences,* Vol. 34, No. 12, pages 1995–2004; December, 1977.

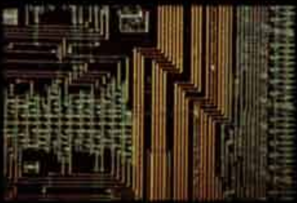
THE AMATEUR SCIENTIST

INTERFERENCE COLOR. Hiroshi Kubota in *Progress in Optics: Vol. 1,* edited by E. Wolf. North-Holland Publishing Company, 1961.

LIGHTNESS AND RETINEX THEORY. Edwin H. Land and John J. McCann in *Journal of the Optical Society of America,* Vol. 61, No. 1, pages 1–11; January, 1971.

hexa·photo·cybernetic

Technology this advanced deserves its own word.



The incredible new Canon A-1. It's on the cutting edge of technology, with state-of-the-art electronics and computerized operation.

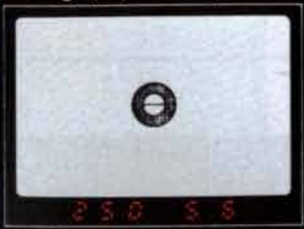
Yet, although the way we build it is complex, the way you use it is the ultimate in simplicity. If pushbutton-easy fine photography is what you're after, the A-1 can give it to you six ways.

1 Shutter Priority automation: You pick the speed to freeze or blur action or prevent camera shake, the A-1 picks the perfect aperture automatically.

2 Aperture Priority automation: Select the aperture you want for a soft background or total overall sharpness, the A-1 will give you the speed that's correct.

3 Programmed automation: On the A-1, this unique exposure mode picks both aperture and speed in ideal combinations depending on the light. You don't have to set *anything*.

4 Stopped-Down automation: This mode lets you get automatic exposure with any lens or accessory you can mount on the camera. Great for astrophotography, photomicrography, close-ups.

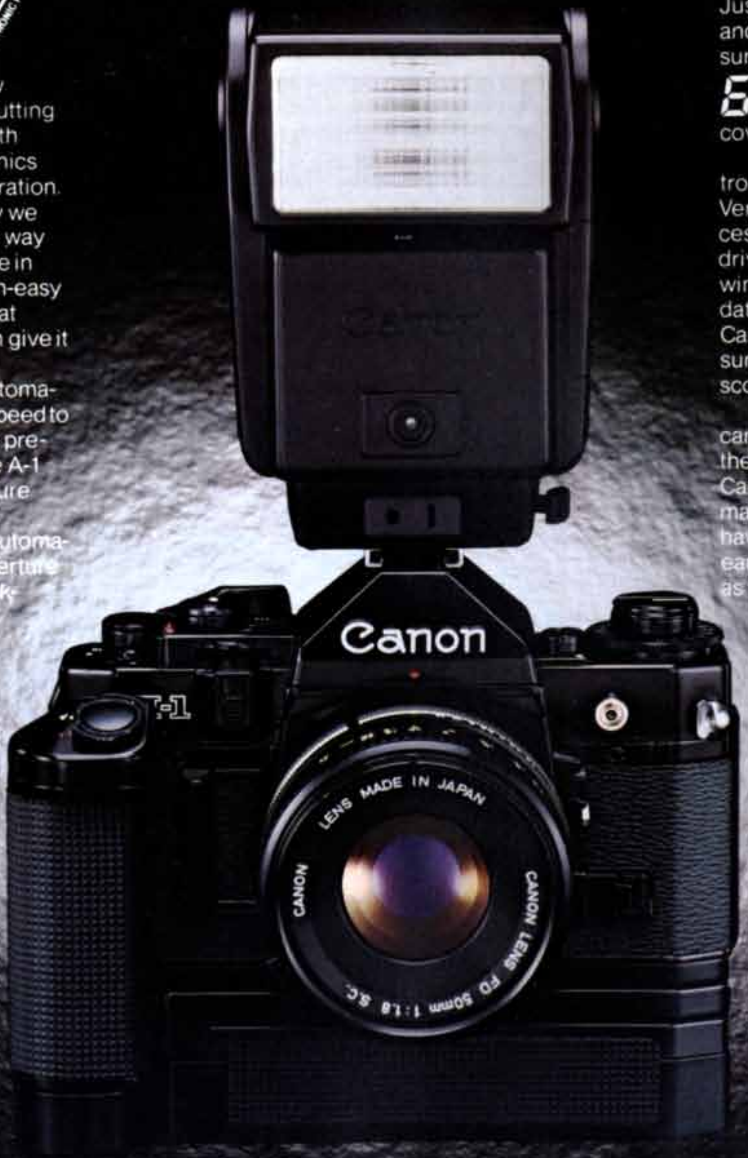


5 Flash automation: The A-1 has the easiest electronic flash of any camera of its type. With special Canon flash units, your aperture *and* speed are set automatically. Just mount the flash, turn it on and shoot for perfect exposures every time.

6 For any type of photography that's not already covered, there's a manual.

There's more. Digital control and viewfinder display. Versatile control features. Accessories like a rugged motor drive, compact motorized film winder and interchangeable date/data back. Plus forty Canon FD lenses for unsurpassed image quality and scope.

You won't find another camera like it on the face of the earth. Because only Canon has the technology to make it. Luckily, you don't have to go to the ends of the earth to get an A-1. Only as far as your local Canon dealer.



THE OFFICIAL 35MM CAMERA OF THE 1980 OLYMPIC WINTER GAMES

Canon A-1

So advanced, it's simple.

Canon USA, Inc., 10 Nevada Drive, Lake Success, New York 11040
140 Industrial Drive, Elmhurst, Illinois 60126 • 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

THE AUDI 5000:

"...IT WOULD BE HARD TO FIND A BETTER PERFORMANCE, COMFORT AND FIVE-

W.A.W., to insiders, is WARD'S AUTO WORLD, a highly regarded trade journal read primarily by engineers and executives who work in the automotive manufacturing industry.

In the November 1978 issue, WARD'S AUTO WORLD rated 24 imported cars on the basis of driving evaluations by W.A.W. editors and writers. (The Audi 5000 used in this road test was a 1978 model equipped with automatic transmission.)

Here are excerpts from the W.A.W. report:

"A winner since the day it entered this country, Audi's front-drive 5-cyl. 5000 sedan is believed by many to be the best buy in the import luxury class."

"Not surprisingly, the 5000 grabbed a clear

Suggested retail price $9,395 P.O.E. Transportation, local taxes and dealer delivery charges additional. Fully equipped "S" model $10,975.


Show Dad you've inherited good taste.
Seagram's
Crown Royal
Fine De Luxe
Blended Canadian Whisky
A PRODUCT OF
Joseph E. Seagram & Sons Limited
WATERLOO · ONTARIO · CANADA
80 PROOF 750 ML (25.4 FL OZ.)
Crown Royal
SEAGRAM DISTILLERS CO., N.Y. BLENDED CANADIAN WHISKY. 80 PROOF.