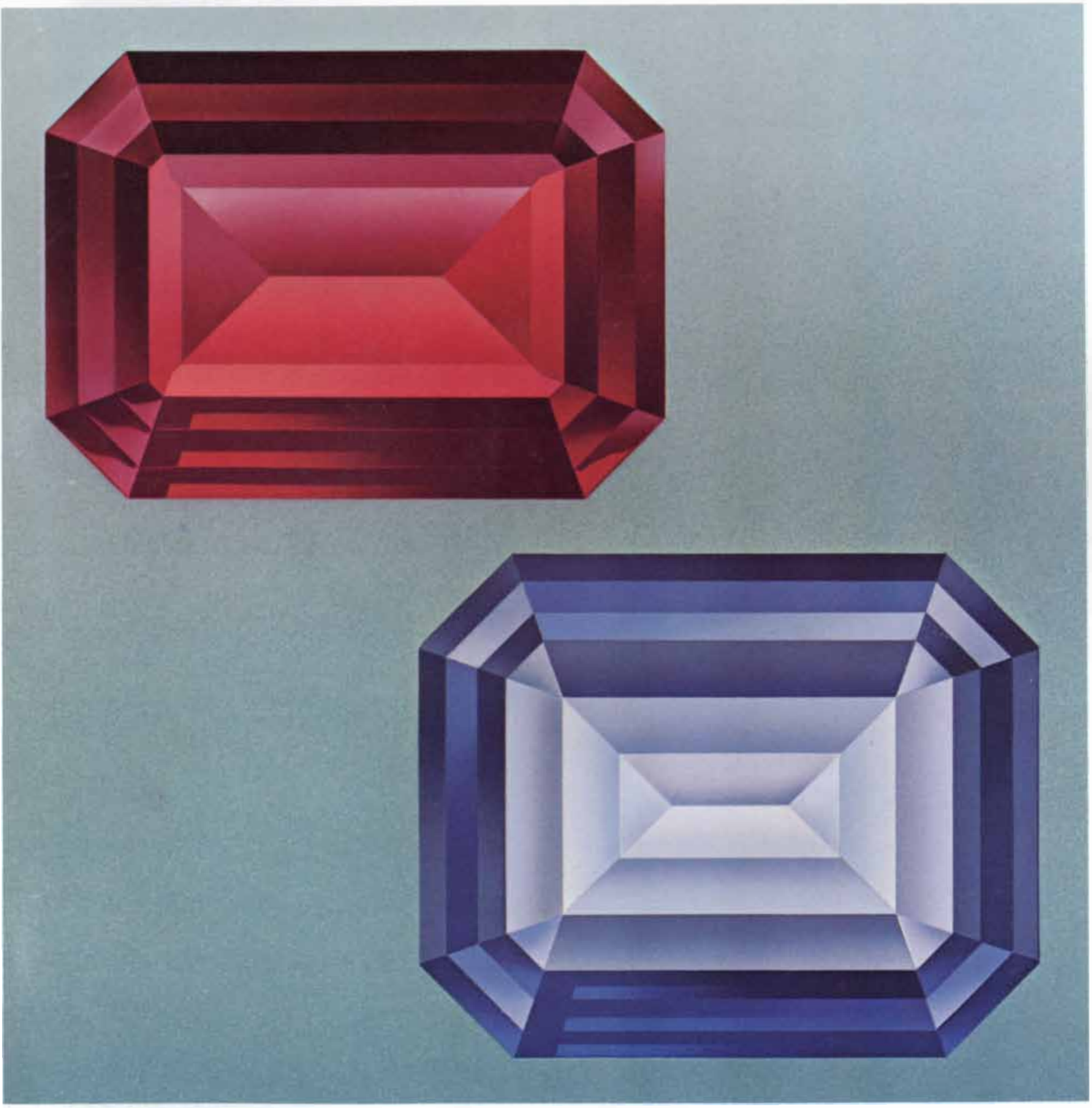
SCIENTIFIC AMERICAN

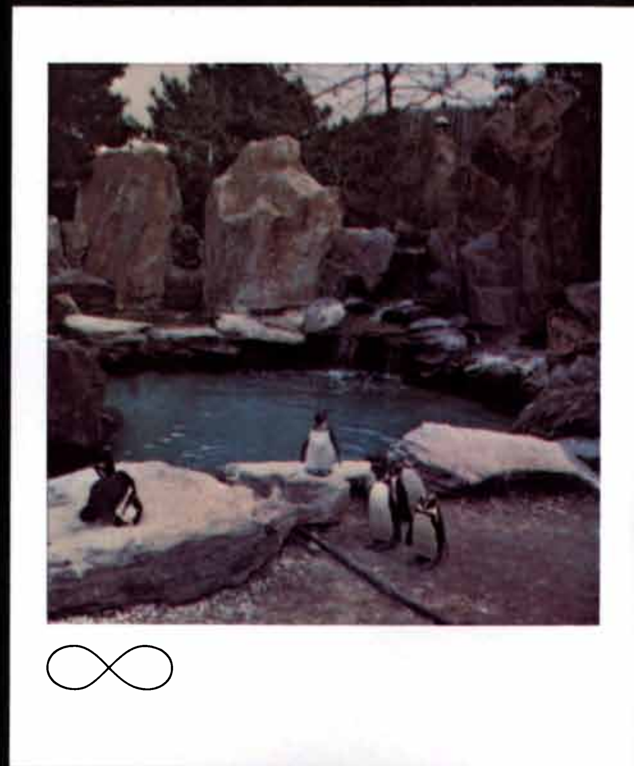


THE CAUSES OF COLOR

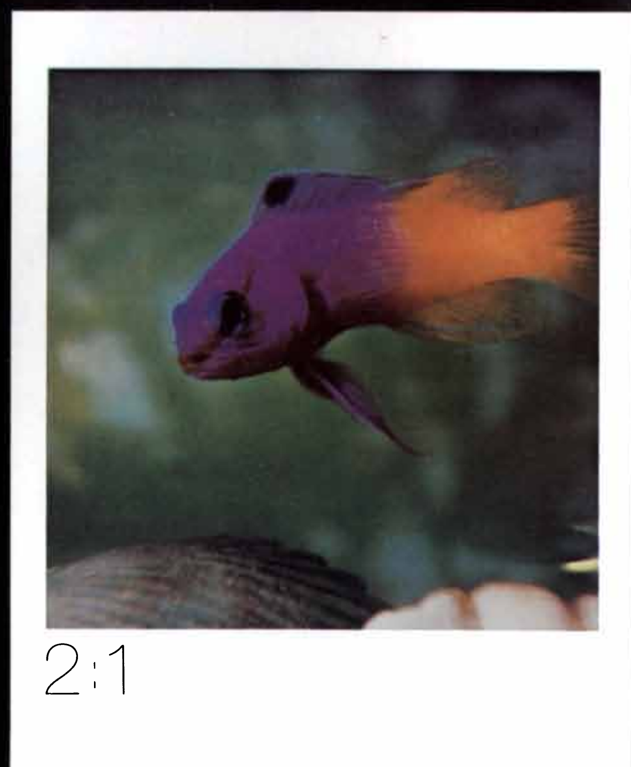
$2.00

October 1980

The new Polaroid CU-70 photographic system. It makes close-ups as easy as 1, 2, 3.



Blackfooted Penguin (Spheniscus demersus) and Humboldt Penguin (Spheniscus humboldti)



Royal Gramma (Gramma loreto)—Indo-Pacific

All photos reproduced actual size from original unretouched Polaroid SX-70 prints.
Pictures shot at New York Zoological Society—Osborn Laboratories of Marine Sciences.

In Saronno love like ours happens once in 450 years.

In 1525 the people of Saronno fell in love with Amaretto di Saronno. And this love has endured generation after generation.

Perhaps it's because we make the original drink of love with time, patience and affection. Until its flavor is rich and subtle, a gentle sensation to the palate.

You can taste Amaretto di Saronno by itself, over ice or in a delicious mixed drink. And once you do, you'll become part of a love story that never ends.

For free food and drink recipe booklets, write: Foreign Vintages, Inc., Dept. 50, 333 Jericho Turnpike, Jericho, N.Y. 11753. To send a gift of Amaretto di Saronno anywhere in the continental U.S. (except where prohibited by state law), call 800-528-6050 toll free and charge it to your credit card. Arizona residents please call 957-4923.

Amaretto di Saronno® Originale.

Since 1525, the true spirit of love.

Liqueur 56 proof. Imported by Foreign Vintages, Inc., Jericho, New York. © 1980. A unique product from Illva Saronno, Italy.

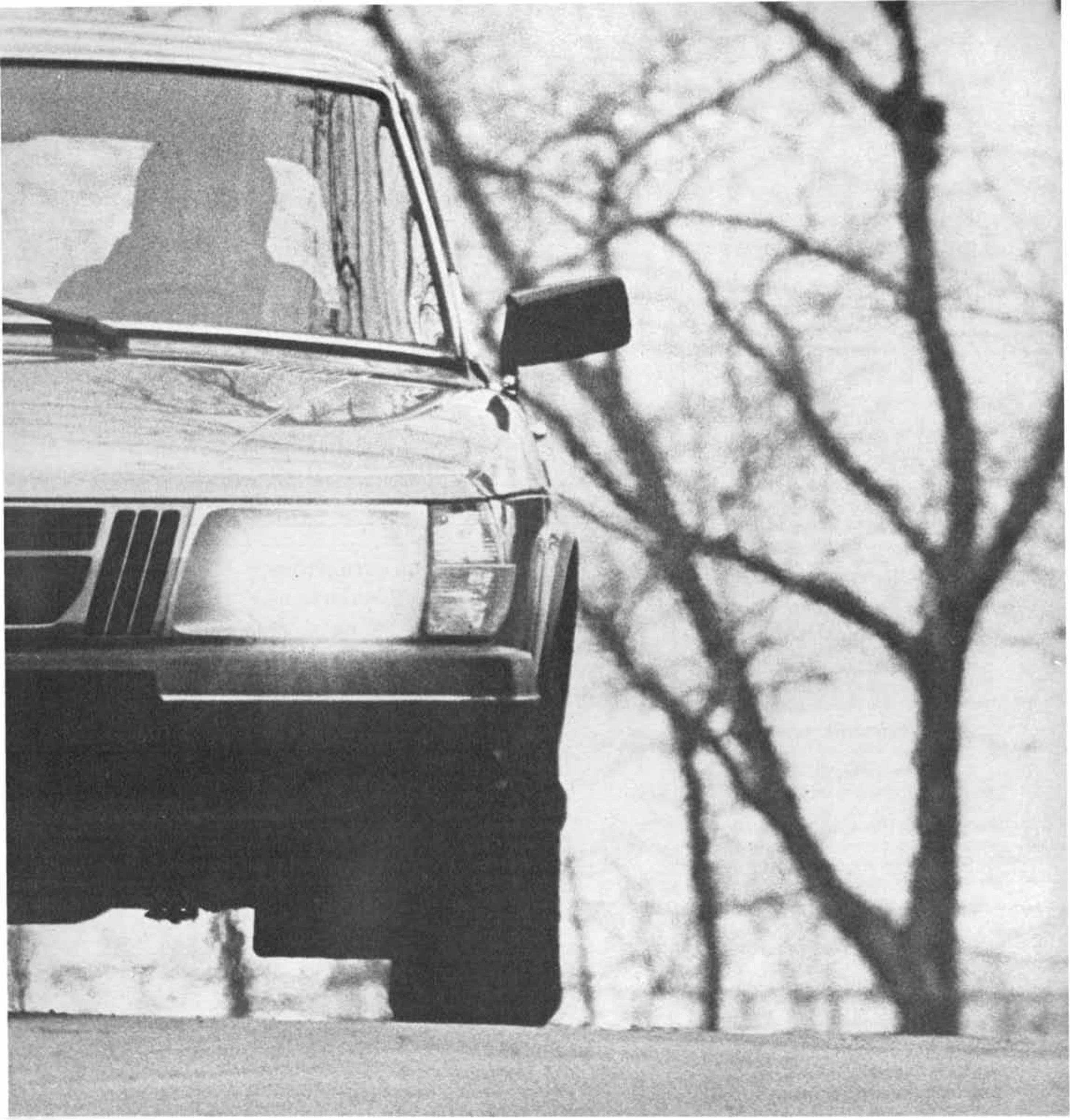
Established 1845 SCIENTIFIC AMERICAN October 1980 Volume 243 Number 4

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. CONTROLLED CIRCULATION POSTAGE PAID AT NEW YORK, N.Y. (USPS 509-530), AND AT CHICAGO, ILL. (USPS 508-030). AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $21 PER YEAR, U.S., ITS POSSESSIONS AND CANADA; $27 PER YEAR, ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.


SAAB



etter than a Volvo, feels better than
miracle? No, a Saab.

ou've always said you'd buy a big car hat gets good gas mileage, here's your hance.

(Saab 900 GLE five-speed: (22) EPA stimated mpg, 33 estimated highway npg. Remember, use estimated mpg for omparison only. Mileage varies with peed, trip length and weather. Actual ighway mileage will probably be less.)

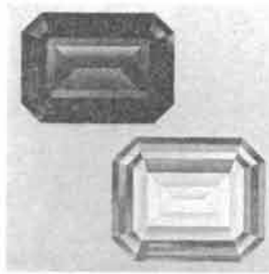
FACT: Saab 900 GLE prices start at $10,825.* Other Saabs start at $8,225.* More, perhaps than a Datsun, a Toyota or a Subaru will cost you, but less than a BMW, a Mercedes or the like will set you back.

All told, buying a front-wheel drive Saab affords you several rare opportunities: one, the chance to buy a new car without forgoing a thing. And, two, the chance to do it without spending a fortune.

These days, that is nothing to sneeze at.

SAAB

The most intelligent car ever built.



THE COVER

The painting on the cover shows two gemstones, a ruby and a blue sapphire, which have the same underlying crystal structure but differ dramatically in color. Both of the stones are basically corundum, an oxide of aluminum that in its pure form is colorless. The colors derive from the presence of impurities: chromium in ruby and iron and titanium in blue sapphire. Electrons in the impurity atoms absorb some wavelengths of light but allow others to pass unimpeded, so that the crystals are transparent but deeply colored. The mechanisms by which the electrons absorb certain wavelengths of radiation are quite different in ruby and sapphire (see "The Causes of Color," by Kurt Nassau, page 124).

THE ILLUSTRATIONS

Cover painting by Ted Lodigensky

Page	Source
18–53	Ilil Arbel
55	Eugene A. Harlan, Lick Observatory (*top*); C. Bruce Stephenson and Nicholas Sanduleak, Case Western Reserve University (*bottom*)
56	B. J. Geldzahler, Massachusetts Institute of Technology (*top*); Eric W. Greisen, National Radio Astronomy Observatory (*bottom*)
57	Frederick D. Seward, Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory
58–65	Allen Beechel
67	Cesar Milstein and Georges Köhler
68–74	Bunji Tagawa
125	Fritz Goro
128–142	Dan Todd
147	Kurt Nassau
148	Dan Todd
150	Pierre Gilson
152	Ilil Arbel
157	National Aeronautics and Space Administration, EROS Data Center
159–160	Andrew Tomko
162	George V. Kelvin
164–165	George V. Kelvin; Frederick A. Cook, Larry D. Brown and Jack E. Oliver, Cornell University (*second from top*)
166–168	George V. Kelvin
170	Michel M. Ter-Pogossian, Marcus E. Raichle and Burton E. Sobel, Washington University School of Medicine
171	Ilil Arbel
172–175	Gabor Kiss
176	Michel M. Ter-Pogossian, Marcus E. Raichle and Burton E. Sobel, Washington University School of Medicine
177–179	Gabor Kiss
180	Ralph Morse
181	Ilil Arbel
183–185	Walken Graphics
185A	Syncrude Canada Ltd.
188	Walken Graphics
191	Peter Ward, University of California at Davis
192–196	Tom Prentiss
198	Lewis Greenwald, Ohio State University
203	Peter Ward, University of California at Davis
205–209	Michael Goodman

LETTERS

Sirs:

As I examined Alexander Keewatin Dewdney's planiversal devices in Martin Gardner's article on science and technology in a two-dimensional universe ["Mathematical Games," SCIENTIFIC AMERICAN, July] I was struck with the similarity of the mechanisms to the lockwork of the Mauser military pistol of 1895. This remarkable automatic pistol (which had many later variants) had no pivot pins or screws in its functional parts. Its entire operation was through sliding cam surfaces and two-dimensional sockets (called hinges by Dewdney). Indeed, the lockwork of a great many firearms, particularly those of the 19th century, follows essentially planiversal principles. For examples see the cutaway drawings in *Book of Pistols and Revolvers* by W. H. B. Smith.

Gardner suggests an exhibit of machines cut from cardboard, and that is exactly how the firearms genius John Browning worked. He would sketch the parts of a gun on paper or cardboard then cut out the individual parts with scissors (he often carried a small pair in his vest pocket) and then would say to his brother Ed, "Make me a part like this." Ed would ask, "How thick, Jon?" Jon would show a dimension with his thumb and forefinger, and Ed would measure the distance with calipers and make the part. The result is that virtually every part of the 100 or so Browning designs is essentially a two-dimensional shape with an added thickness.

This planiversality of Browning designs is the cause of the obsolescence of most of his designs. Dewdney says in his enthusiasm for the planiverse that "such devices are invariably space-saving." They are also expensive to manufacture. The Browning designs had to be manufactured by profiling machines: cam-following vertical milling machines. In cost of manufacture such designs cannot compete with designs that can be produced by automatic screw-cutting lathes, by broaching machines, by stamping or by investment casting. Thus although the Browning designs have a marvelous aesthetic appeal, and although they function with delightful smoothness, they have nearly all gone out of production. They simply got too expensive to make.

JOHN S. HARRIS

English Department
Brigham Young University
Salt Lake City, Utah

Sirs:

In Martin Gardner's article he and the authors he quotes seem to have overlooked the following aspect of a "planiverse": any communication by means of a wave process, acoustic or electromagnetic, would in such a universe be impossible. This is a consequence of the Huygens' principle, which expresses a mathematical property of the (fundamental) solutions of the wave equation. More specifically, a sharp impulse-type signal (represented by a "delta function") originating from some point is propagated in a space of three spatial dimensions in a manner essentially different from that in which it is propagated in a space of two spatial dimensions. In three-dimensional space the signal is propagated as a sharp-edged spherical wave without any trail. This property makes it possible to communicate by a wave process because two signals following each other in short time can be distinguished.

In a space with two spatial dimensions, on the other hand, the fundamental solution of the wave equation represents a wave that, although it too has a sharp edge, has a trail of theoretically infinite length. An observer at a fixed distance from the source of the signal would perceive the oncoming front (sound, light, etc.) and then would keep perceiving it, although with intensity decreasing in time. This fact would make communication by any wave process impossible because it would not allow two signals following each other to be distinguished. More practically such communication would take much more time. This letter could not be read in the planiverse, although it is (almost) two-dimensional.

STEFAN DROBOT

Department of Mathematics
Ohio State University
Columbus

Sirs:

I was interested in Martin Gardner's article on the physics of Flatland, because for some years I have given the students in my general relativity class the problem of deriving the theory of general relativity for Flatland. The results are surprising. One does *not* obtain the Flatland analogue of Newtonian theory (masses with gravitational fields falling off like $1/r$) as the weak-field limit. General relativity in Flatland predicts no gravitational waves and no action at a distance. A planet in Flatland would produce no gravitational effects beyond its own radius. In our four-dimensional space-time the energy momentum tensor has 10 independent components, whereas the Riemann curvature tensor has 20 independent components. Thus it is possible to find solutions to the vacuum field equations $G_{\mu\nu} = 0$ (where all components of the energy momentum tensor are zero) which have a nonzero curvature. Black-hole solutions and the gravitational-field solution external to a planet are examples. This allows gravitational waves and action at a distance. Flatland has a three-dimensional space-time where the energy momentum tensor has six independent components and the Riemann curvature tensor also has only six independent components. In the vacuum where all components of the energy momentum tensor are zero all the components of the Riemann curvature tensor must also be zero. No action at a distance or gravity waves are allowed.

Electromagnetism in Flatland, on the other hand, behaves just as one would expect. The electromagnetic field tensor in four-dimensional space-time has six independent components that can be expressed as vector E and B fields with three components each. The electromagnetic field tensor in a three-dimensional space-time (Flatland) has three independent components: a vector E field with two components and a scalar B field. Electromagnetic radiation exists, and charges have electric fields that fall off like $1/r$.

J. RICHARD GOTT III

Director of Graduate Studies
Department of Astrophysical Sciences
Princeton University
Princeton, N.J.

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

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

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

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

Name ____________________

New Address ____________________

____

____

Old Address ____________________

____

____

A report on the great American forest.

Money does grow on trees.

(And vice versa.)

Consider the wonderfully versatile American forest.

It provides all Americans with useful wood and paper products, recreation, natural beauty, protection for watersheds and wildlife.

But it goes even farther, making these less obvious, but no less important, contributions to the country's economic well-being:

Jobs — millions of workers in the woods, in the factories that turn wood and fiber into thousands of essential products, and on construction sites all over this growing land.

Taxes — billions of dollars a year, going to support schools, hospitals, public safety, countless other vital public services.

International trade — wood products we export help strengthen the American dollar and improve the nation's balance of payments.

Inflation deterrent — the more productive the forest, the more likely that the nation's supply of wood and paper products will be adequate to meet sharply increasing demands — and will be available at reasonable prices. Thus, a productive commercial forest, growing trees for today *and* tomorrow, becomes a significant inflation fighter.

Your stake in the commercial forest

So every American has a stake in increasing the productivity of the commercial forests. (*Commercial forest,* as defined by the U.S. Forest Service, is *all* forestland — whether owned by individuals, government or the forest industry — that is capable of, and potentially available for, growing repeated crops of trees for harvest. It includes land in National Forests but not in National Parks or Wilderness areas.)

So far, the commercial forest has been able to cope with all the demands made on it. But we can't expect it to continue to provide its benefits automatically.

Why trees need money too

If wood supply is to keep up with the predicted doubling of demand in this nation over the next 50 years, expenditures for replanting and regeneration will have to be substantially increased.

And the greatest potential for improvement is on publicly held land.

So that means the nation needs to establish policies and take actions to encourage productivity — not only in the National Forests but in other forests as well.

If you'd like to be better informed on how important it is to keep America's forests productive, write American Forest Institute, P.O. Box 873, Springfield, VA 22150 for a free booklet, "The Great American Forest."

The great American forest. Trees for tomorrow. And tomorrow. And all the tomorrows after that.

Trees. America's renewable resource.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

OCTOBER, 1930: "President Hoover won a decisive victory in the prompt and almost unanimous ratification of the naval pact by the Senate. We believe the Senate reflected the public opinion of the country in its action; nevertheless, the outcome is a personal triumph for the President, and coming at the close of a Congressional session where his leadership had been several times repudiated, it must have been gratifying. We believe the President deserves all the prestige he has gained by his patient and tactful handling of this difficult problem. At the same time we entirely understand and sympathize with the patriotic motives that inspired the small group of senators, ably led by Hale of New Hampshire and Johnson of South Dakota, to unsuccessfully oppose the ratification. There is no gainsaying the facts they brought out in their vigorous attacks on the treaty. But most Americans felt that, taking all matters into consideration, it was a good treaty. We think they reconciled themselves to this treaty in the firm belief that President Hoover would put the whole force of the Administration behind a building program that would give the nation the maximum navy it is entitled to have."

"The 13th expedition of the Department of Tropical Research of the New York Zoological Society under the direction of Dr. William Beebe has had a very successful season in Bermuda, and our knowledge of undersea life has been greatly enlarged by the use of the 'bathosphere' designed by Otis Barton and J. H. J. Butler. This spherical steel diving chamber or tank is a single steel casting fabricated by a firm specializing in hydraulic machinery. It is four feet nine inches in diameter and its walls are more than an inch and a half in thickness to resist the enormous pressure that is found at great ocean depths. During their deepest dive of 1,426 feet the divers were comfortable and cool, although they were inside more than an hour and a half. One of the chief objects of the expedition is to determine the continuity of fish life that connects the mid-water zone of 100 fathoms with that of the ocean's floor two miles down."

"Efforts of the Brazilian government to popularize the use of alcohol motor fuel are meeting with success, according to reports to the Department of Commerce by Assistant Trade Commissioner J. Winsor Ives. Official cars are required to use this fuel, and 60 per cent of all cars in Pernambuco are regular consumers. Petroleum companies maintaining branches in Pernambuco report reduced sales of gasoline due to the lower price of alcohol fuel. Alcohol motor fuel, which consists of a mixture of cane alcohol and ether, is selling from 500 to 700 reis (six to eight cents) per liter, with gasoline approximately 18 cents."

"The news of the untimely death of Glenn H. Curtiss at the age of 52 came as a shock not only to the aviation world, of which he was one of the few outstanding pioneers, but also to the world in general. Inventor, pilot and manufacturer, he was in turn a bicycle racer and repairer, a motorcycle manufacturer, the first man to make scheduled airplane flights and finally an airplane manufacturer. He contributed more to the development of heavier-than-air flight than any other man, and much of America's success in building planes during the war was due to his genius. Together with Alexander Graham Bell he formed in 1907 the Aerial Experiment Association and built the *Red Wing,* which cracked up on its first flight of nearly 319 feet, the first public airplane flight in America. He won the first leg of the SCIENTIFIC AMERICAN trophy by a flight of one kilometer in the now famous *June Bug.* The second leg of the SCIENTIFIC AMERICAN trophy was won by Curtiss in the first exhibition flight in America at Hempstead Plains, Long Island. This flight covered 24.7 miles. Then on May 29, 1910, came his greatest triumph: a successful flight down the Hudson River from Albany to New York City by which he won the *World* prize and the third leg of the SCIENTIFIC AMERICAN trophy, which then became his permanent possession."

"During the past few months thousands of cases of paralysis from drinking bootleg Jamaica ginger have occurred, principally in the southwestern portion of the United States. Studies made by chemists indicate that the poisonous substance is a derivative of coal tar or phenol. Few of the patients have died, but practically all of them have developed forms of paralysis that seem to be fairly permanent."

SCIENTIFIC AMERICAN

OCTOBER, 1880: "A remarkable series of investigations in relation to sound transmission by light has led to the invention of the photophone by Professor Bell. In this instrument articulate speech is transmitted by means of a beam of light, without any visible or tangible connection between the transmitting and receiving stations. A beam of light from any source is concentrated on a diaphragm by a lens, and the diaphragm, which is capable of reflecting the light, is placed in such a position as to project the light through a second lens to a parabolic reflector at a distance. The parabolic reflector concentrates the light on a selenium cell. The selenium forms a part of an electrical circuit, which includes a battery and a receiving telephone. A sound in the vicinity of the transmitting instrument vibrates the diaphragm and undulates the beam of light projected through the lens, and the consequent variation in the intensity of the light concentrated on the selenium by the parabolic reflector changes the electrical conductivity of the selenium and renders the electric current undulatory. This current affects the receiving telephone in the same way as it would be affected in an ordinary telephonic circuit, and the sounds made in the transmitting instrument are reproduced in the telephone."

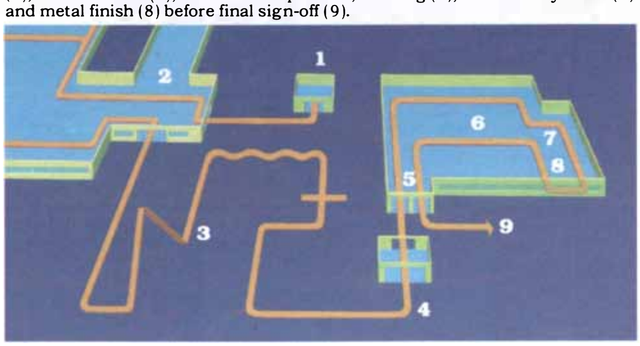
"Russian journalists, says the London *Telegraph,* appear to be painfully exercised by the announcement that two American steamers laden with grain have entered the port of Revel for the purpose of discharging their cargo, a circumstance without precedent in the annals of Russian commerce. That Russia would never need to import cereals from foreign countries has heretofore been a firmly established article of popular faith throughout the Czar's dominions. So rapid, however, has of late years been the falling off in productiveness exhibited in the agricultural districts of the empire that the seemingly impossible has at length come to pass, and northern Russia is importing wheat from the United States. It is but justice to the Russian press to acknowledge that it has been profuse of warnings with respect to the probable consequences of slovenly and unintelligent farming, persistence in old-fashioned and exploded systems of cultivation, reluctance to invest capital in modern agricultural improvements, absenteeism and other shortcomings that have practically disqualified Russian grain growers from competing for foreign customers with their trans-Atlantic rivals. But Russian buyers and peasant farmers alike were so immutably possessed by the conviction that Russia was the predestined granary of Europe that they ignored these salutary admonitions."

"Dr. Phipson takes sulphide of barium, or some other substance that is rendered phosphorescent by the solar rays, and encloses it in a Geissler tube, through which he passes a constant electric current of a feeble but regular intensity. He claims to obtain in this manner a uniform and agreeable light at a cost lower than that of gas."

Apollo program. We're an industry leader in automotive electronics, going back to electronic ignition. There are several hundred electronics experts down there in Huntsville and after we switched them from space work to commercial work, the Imperial is one of the things they went to work on.
Sinatra: What did they do?
Iacocca: Built our system of Electronic Fuel Injection.
Sinatra: Fuel injection's been around for years.
Iacocca: Not like this. The Imperial has the first continuous flow fully electronically controlled fuel injection for any production automobile built anywhere in the world.
Sinatra: Why is that good?
Iacocca: First of all, it controls both the fuel delivery and the spark advance, based on information it gets from about a dozen sensors that monitor everything from engine temperature to barometric pressure. They feed all this information into what we call the Combustion Computer. Then the computer figures the best possible spark advance for all these conditions and the best proportion of air to fuel, and meters out the exact amount you need.
Sinatra: What's in it for me?
Iacocca: Four things. First of all, it starts by just a turn of the key. Second, you get the smoothest idle, even on cold mornings. Number three, there are practically no balks. And number four, practically no stalls.
Sinatra: Sounds terrific.
Iacocca: And that's only the half of it. I tell you, we've never had electronics like this. Now you've got to admit, this instrument panel is rather unusual.
Sinatra: I've driven a lot of cars and I've never seen anything that looked like it. But why does a digital instrument panel make the Imperial a better car?
Iacocca: Because it makes you a smarter driver. Tell me, what do you need to know when you're driving a car?
Sinatra: I want to know how fast I'm going. How far I've gone. How much gas I've got left. What time it is. That's all.
Iacocca: Okay. This instrument panel shows you all of that in words and numbers.
Sinatra: No gauges?
Iacocca: No gauges. This is the only American car with a completely electronic digital instrument cluster.
Sinatra: And what do the buttons do?
Iacocca: They turn your instrument panel into an Electronic Communications Center. Push button number one and your speedometer changes from miles to kilometers.
Sinatra: Not bad.
Iacocca: Push button number two and your electronic chronometer gives you the date instead of the time. Button number three turns the whole chronometer into an electronic timer that tells you how long you've been driving.
Sinatra: What about how far I've gone?
Iacocca: That's button number four. Push it and the electronic odometer shows you how far you've gone since you started. The next button tells your average speed.
Sinatra: And it's all in miles or kilometers.
Iacocca: That's right. Button six shows how far the gas in your tank will take you under present driving conditions. Number seven shows how many miles you're now getting to the gallon. And number eight gives your mileage over your entire trip.
Sinatra: I've never had that on any car.
Iacocca: There's more.
Sinatra: More buttons to push?
Iacocca: No. This part is all automatic. If either half of the dual hydraulic brake system is defective, a light goes on. If the emergency brake is on, this lights up. If your oil pressure or battery voltage are too low, or your engine temperature is too high, these lights come on.
Sinatra: That's a lot of lights.
Iacocca: We've got chimes, too. If you don't buckle up, your Imperial will chime at you. Leave the key in the ignition and it beeps. Leave the lights on and it makes a tone.
Sinatra: Very musical.
Iacocca: I told you we built it for you, Frank.
Sinatra: What about your other customers? Is this stuff all standard equipment?
Iacocca: Frank, the only option on the Imperial is a power sliding roof. Every luxury is standard. And there are more luxuries standard than any car in America. I've got even more luxuries than *you* would ask for.
Sinatra: Try me.
Iacocca: You try me. What features do you want in your car?
Sinatra: Start with music.
Iacocca: Your choice of four sound systems, standard. All stereo. All with six speakers. What else do you want?
Sinatra: Power steering, power windows, power seats.
Iacocca: Of course.
Sinatra: Leather upholstery?
Iacocca: Frank, this is the only car in the world with an interior by Mark Cross. It's even got a built-in garage door opener and a hood ornament of Cartier crystal.
Sinatra: You're not fooling around with quality on this baby, are you?
Iacocca: I wanted a quality automobile to sell. I wanted to be able to give it a basic limited warranty twice as long as the competition's.
Sinatra: How long is that?
Iacocca: Two years or 30,000 miles, whichever comes first. It covers all parts except tires, all labor and all schedule maintenance. Our rust warranty lasts *three* years.
Sinatra: I don't even pay for an oil change?
Iacocca: That's right. The Imperial warranty has the details.
Sinatra: That kind of confidence takes guts.
Iacocca: I know what went into it. Look, every one of these Imperials gets a test drive by an expert. We built a Quality Assurance Center where the whole vehicle gets a going-over to make sure that nothing gets out that's not first-rate. When an inspector there signs off on the car, the signed papers actually come with the car.
Sinatra: I think there are a lot of people waiting for a car like this one, Lee.
Iacocca: I've got a bunch of Imperial dealers who are ready for them, Frank.

"Every luxury is standard. And there are more luxuries standard than any car in America."

Imperial Quality Assurance Program starts with preselected parts (1) and special care assembly (2), goes on to a 5.5-mile test drive (3), hot engine checks (4), water checks (5), checks of suspension, steering (6), electrical systems (7) and metal finish (8) before final sign-off (9).



"I wanted to be able to give it a basic limited warranty twice as long as the competition's."

1981 Imperial. Size and price: comparable to Cadillac Eldorado and Continental Mark VI. (16) EPA est. mpg. 26 est. hwy. mpg. Use these ests. for comparison. Mileage may vary depending on speed, weather, trip length. Acutal hwy. mileage will probably be less. Calif. ests. lower.

For the name of the Imperial dealer nearest you, call toll-free 1-800-521-7272. In the state of Michigan call 1-800-482-6838.

It's time for Imperial.

The New Chrysler Corporation.

ATARI® PERSONAL COMPUTER SYSTEMS THAT GROW WITH YOU

Start with a better computer. Atari computers have built-in capabilities you can't even add onto many other personal computers. Three programming formats (ROM cartridge, disk and cassette). A 57 key upper/lower case ASCII keyboard with 29 keystroke graphics symbols. 128 colors and hues. Four separate sound channels and a built-in speaker. Four controller ports. A built-in RF modulator and FCC approval for connection to any TV. Plus, nationwide Atari Authorized Service Centers. And more.

Add memory. The ATARI 800™ is supplied with 16K of memory. You can expand up to a full 48K of RAM with 8K or 16K Memory Modules™ you install yourself. In less than a minute. The ATARI 400's™ 8K of RAM may be expanded to 16K at Authorized Service Centers. Both may be expanded to 26K of ROM with slip-in ROM cartridge programs.

Add peripherals. The ATARI 410™ audio-digital program recorder. Single or dual density* disk drives. The ATARI 800 individually addresses up to four drives. Add the ATARI 850™ RS232 Interface Module. Add high speed 40 or 80-column printers. Add an acoustic modem for remote data access. Add a light pen.* And there are more Atari peripherals on the way.

Add programs. Choose among dozens of programs in Atari's rapidly expanding software library. Programs categories include:

- **Personal Finance & Record Keeping**
- **Personal Interest & Development**
- **Professional Applications**
- **Education**
- **Information & Communication**
- **Entertainment**
- **Programming Languages**
- **Small Business Accounting**

Add it up. With Atari, you start with more. And you can build to more. Because Atari offers you personal computer systems that grow with you. Ask your Atari retailer to give you a full demonstration of Atari computers, peripherals and programs. Complete systems. Because when other people were thinking hardware and software, Atari was thinking systems.

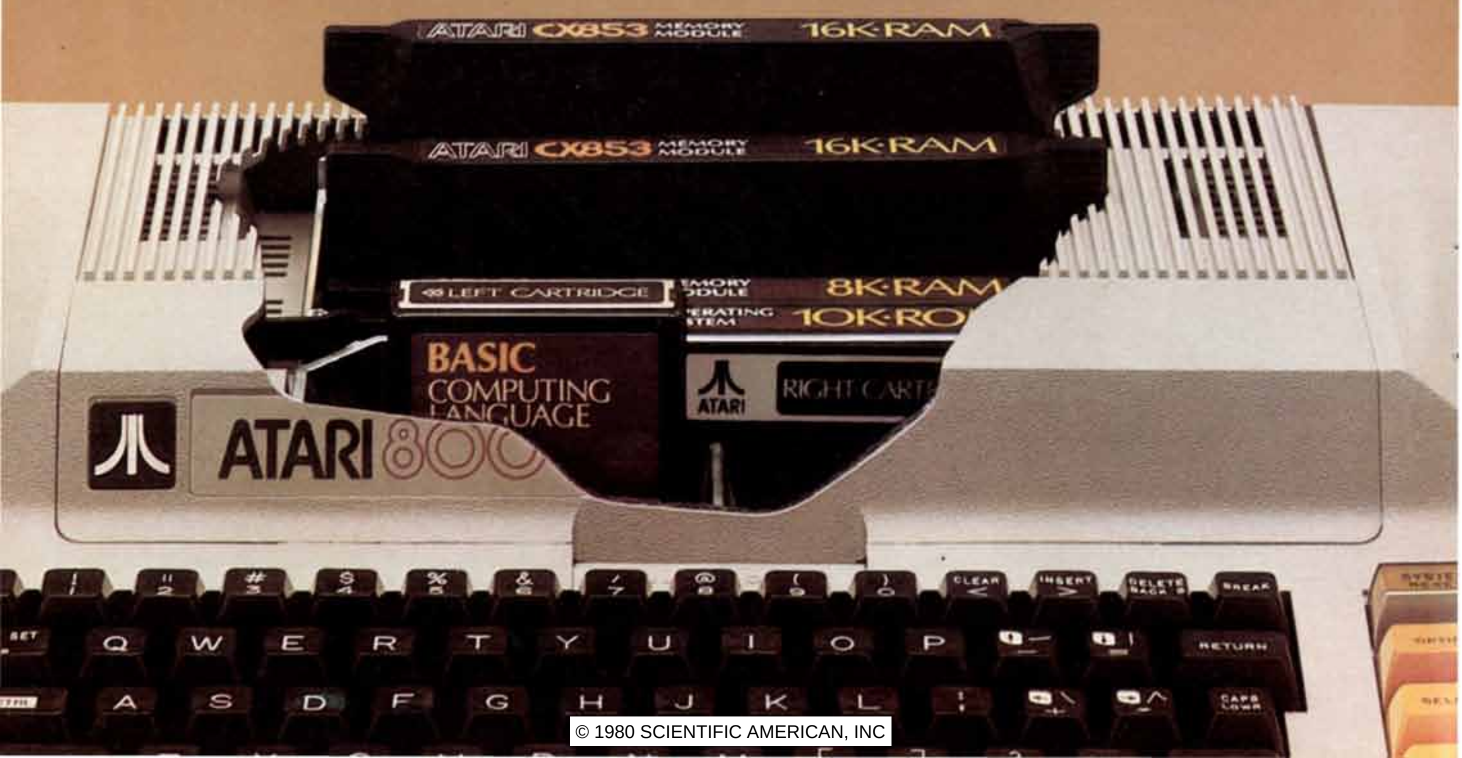
*Available Fall, 1980

ATARI®

PERSONAL COMPUTERS

1265 Borregas Avenue, Sunnyvale, CA 94086

Call toll-free (800) 538-8547 (Except Alaska and Hawaii)
(In California: (800) 672-1404) for the name of your nearest Atari retailer.



THE AUTHORS

ERNEST W. SAWARD and SCOTT FLEMING ("Health Maintenance Organizations") have each spent much of their career in the administration of the Kaiser-Permanente Medical Care Program. A graduate of the University of Rochester, Saward in 1945 initiated the Kaiser-Permanente program in Portland, Ore., and served as its medical director for 25 years. He then returned to his alma mater as associate dean and as professor of medicine and of social medicine. Fleming received his law degree from the University of California at Berkeley in 1949. He has worked in the Kaiser-Permanente program since 1955 except for a two-year leave of absence as deputy assistant secretary for Health Policy Development in the Department of Health, Education, and Welfare. He is presently senior vice-president in the Kaiser-Permanente central office in Oakland, Calif.

BRUCE MARGON ("The Bizarre Spectrum of SS 433") has recently joined the department of astronomy at the University of Washington. He received his bachelor's degree in astrophysics from Columbia University in 1968 and his Ph.D. from the University of California at Berkeley in 1973. After a year as a NATO Postdoctoral Fellow at University College London, he returned to Berkeley as a member of the university's research staff. In 1976 he was appointed assistant professor of astronomy at the University of California at Los Angeles and in 1978 associate professor.

CESAR MILSTEIN ("Monoclonal Antibodies") is a senior member of the British Medical Research Council Laboratory of Molecular Biology in Cambridge. Born in Argentina, he attended the University of Buenos Aires from 1945 to 1952, graduating with a degree in chemical sciences. He continued at the university's Institute of Biological Chemistry and in 1957 completed his doctorate in chemistry. The following year he joined the department of biochemistry at the University of Cambridge, where he received a Ph.D. in 1960. After heading the division of molecular biology at the National Microbiological Institute in Buenos Aires, Milstein returned to Cambridge at the Laboratory of Molecular Biology. A Fellow of the Royal Society, he received the Ciba Foundation medal and prize in 1978.

KURT NASSAU ("The Causes of Color") is a research scientist at Bell Laboratories. Born in Austria, he moved to Britain in 1938 and studied chemistry and physics at the University of Bristol, from which he graduated in 1948. He then came to the U.S. and worked as a chemical, biochemical and medical investigator, completing his Ph.D. in physical chemistry at the University of Pittsburgh in 1959. At Bell Laboratories since that time he has worked in the field of crystal chemistry, principally the study of crystals and glasses for lasers and optical fibers. "My interest in color," he writes, "goes back to my work with growing laser crystals. I became acquainted with the nature of one type of color and fluorescence, and this initiated an interest in minerals and gems."

FREDERICK A. COOK, LARRY D. BROWN and JACK E. OLIVER ("The Southern Appalachians and the Growth of Continents") work in geology and geophysics at Cornell University. Cook graduated from the University of Wyoming in 1973. He completed his master's degree there two years later and worked as a geophysicist for the Continental Oil Company before moving to Cornell. Brown has been assistant professor of geological sciences at Cornell since 1977. He did undergraduate work in physics at the Georgia Institute of Technology and received his Ph.D. in geophysics from Cornell in 1976. Oliver has served since 1971 as Irving Porter Church Professor of Engineering and as chairman of the department of geological sciences. He obtained a B.A. from Columbia College in 1947 and a master's degree in physics and Ph.D. in geophysics from Columbia University in 1950 and 1953. From 1955 through 1971 he was on the faculty at Columbia, spending the last three of those years as chairman of its department of geology.

MICHEL M. TER-POGOSSIAN, MARCUS E. RAICHLE and BURTON E. SOBEL ("Positron-Emission Tomography") are professors at the Washington University School of Medicine. Ter-Pogossian studied as an undergraduate at the Lycée Michlet in Paris and earned his bachelor's degree in mathematics from the University of Paris. While engaged in graduate research in physics at the university, he spent a year at the Institute of Radium studying under Irène Joliot-Curie. He continued his graduate work in nuclear physics at Washington University, receiving his master's and Ph.D. degrees in 1948 and 1950. Ter-Pogossian then joined the Mallinckrodt Institute of Radiology at the Washington University School of Medicine, where in 1963 his research in positron-emission tomography led to the installation of the first American medical cyclotron in the university's Barnard Hospital. Raichle is professor of neurology and of radiation sciences. After obtaining a B.S. (in political science) and an M.D. from the University of Washington, he developed an interest in brain circulation and metabolism while a resident in neurology at the New York Hospital–Cornell Medical Center. In 1971 he joined the medical faculty of Washington University, where he has applied positron-emission tomography to his study of brain blood flow and metabolism. Sobel is professor of medicine and director of the cardiovascular division. He graduated from Cornell University in 1958 and obtained his M.D. from Harvard Medical School in 1962. After several years as a clinical associate at the National Heart and Lung Institute he joined the faculty of the University of California at San Diego School of Medicine. In 1973 Sobel moved to Washington University.

RICHARD A. DICK and SHELDON P. WIMPFEN ("Oil Mining") have had long careers as engineers for the mining industry and the U.S. Bureau of Mines. A graduate of Michigan Technological University with a degree in mining engineering, Dick has been with the Bureau since 1965. Wimpfen began his mining career in 1934 as a graduate of the University of Texas School of Mines and Metallurgy at El Paso. He joined the Bureau of Mines in 1970, serving as Assistant Director for Mineral Supply until 1975. He then accepted the dual posts of Assistant Director of Field Operations and Chief Mining Engineer, both of which he held until his retirement from the Bureau in February.

PETER WARD, LEWIS GREENWALD and OLIVE E. GREENWALD ("The Buoyancy of the Chambered Nautilus") began their collaboration on the nautilus at Ohio State University. Ward received his bachelor's and master's degrees from the University of Washington and completed his Ph.D. in 1976 at McMaster University. He taught earth history and paleobiology at Ohio State from 1976 to 1978 and is currently assistant professor of geology at the University of California at Davis. "I worked my way through school as a professional salvage diver and scuba instructor," he writes. "I have spent part of every year since 1975 in the Indo-Pacific studying the nautilus. This has involved night diving to great depths off the fore-reef slopes of New Caledonia. To my knowledge I am the only cephalopod investigator who has actually observed the nautilus in its natural habitat." Lewis and Olive Greenwald joined the faculty of Ohio State University in 1972, where they are now respectively associate professor of zoology and adjunct assistant professor of zoology. Lewis Greenwald graduated from Syracuse University, Olive Greenwald from Rice University.

MATHEMATICAL GAMES

From counting votes to making votes count: the mathematics of elections

by Martin Gardner

With the presidential election coming up next month in the U.S. it seems a good time to discuss the mathematics of voting, not least because there are a variety of mathematically determined paradoxes and anomalies that are particularly pertinent to this presidential contest. The way these contradictions touch on plurality voting and the proceedings of the electoral college will be dealt with here, as will the relatively new system known as approval voting. A procedure of increasing interest to political scientists, approval voting manages to avoid many of the logical inconsistencies inherent in other voting schemes.

This month's discussion of the mathematics of voting is written not by me but by Lynn Arthur Steen, professor of mathematics at St. Olaf College in Northfield, Minn. The editor of *Mathematics Magazine,* Steen writes frequently on mathematical subjects and has twice been awarded the Mathematical Association of America's Lester R. Ford award for excellence in writing. He has also edited a variety of books, including *Mathematics Today: Twelve Informal Essays* (Springer-Verlag, 1978) and (with Matthew P. Gaffney) *Annotated Bibliography of Expository Writing in the Mathematical Sciences* (Mathematical Association of America, 1976).

All that follows (up to my concluding comments on previous columns) was written by Steen, who titles his discussion of voting paradoxes and anomalies "Election Mathematics: Do All Those Numbers Mean What They Say?"

Over the past year the prospect of a three-way race for president of the United States has focused public attention on the importance of strategies for voting and on the special vagaries of the electoral college. Although the complications imposed by the electoral college are unique to presidential elections, other uncertainties imposed by three-way contests for public office are not. When the public must choose among more than two alternatives, the task of making the choice is frustratingly difficult. The source of both the difficulties and the possible solutions is to be found in the little-known mathematical theory of elections.

The social contract of a democracy depends in an obvious and fundamental way on a simple mathematical concept, namely the concept of a majority. Barring the unlikely event of a tie, in any dichotomous ballot one side or the other must receive more than half of the votes. When there are three or more choices of approximately equal strength, however, it is unlikely that such a ballot will yield a majority decision. It is primarily for this reason that many people believe the two-party system is essential to the stability of democracy in the U.S., even though that system is neither mandated nor recognized by the Constitution.

Mathematical theory and political idealism notwithstanding, quite often the public does face a choice among three or more significant alternatives. The same problem that this year appeared as Carter v. Reagan v. Anderson has developed in other years. Such multiple-candidate contests are difficult to resolve fairly if there is no clear-cut majority, but they can easily arise in any free election. Indeed, it follows from some simple mathematics that there are practically no positions candidates in a two-way contest can take that are invulnerable to attack by a third or a fourth candidate.

If each issue in a two-candidate election is represented by a rating of voter preference on a one-dimensional scale, then regardless of the distribution of attitudes among the voters the optimal position for each candidate is the median: the point that divides the electorate into two camps of equal size. The same is true whether public opinion is distributed normally (so that the graph of position v. number of supporters has a single, centered hump), is split bimodally (so that the graph has two approximately equal humps), is skewed sharply to one side or is divided in a highly irregular way. An example of each of these distributions, with the median marked, is given in the accompanying illustration [*upper illustration on page 18*].

Consider a two-candidate contest in which one candidate adopts a position a little to the left of the median and the other candidate begins with a position at about the middle of the right half of the population. This would be typical of a centrist candidate *C* running against a moderate right-wing candidate *R*. In this case it is reasonable to assume that as far as this particular issue is concerned the voters whose preference lies to the left of the position held by the centrist candidate *C* will favor *C,* the voters whose preference lies to the right of candidate *R* will favor *R,* and the voters whose preference lies in between will be divided about evenly between the two candidates. Under these circumstances, in a preelection poll the centrist candidate would receive a majority of the votes.

The only way for candidate *R* to improve his standing in the poll (on this single issue) is to shift his position toward the middle of the distribution, to ensure that more voters will be to his right. Moving toward the center, or to the left, will always be advantageous for the right-wing candidate. Similarly, a left-leaning candidate can improve his standing with the voters by moving toward the center, or to the right. The median position is the only one that cannot be improved on by further shifting on the part of either candidate.

There is, of course, nothing very novel about this analysis. It is part of our common experience in presidential politics. Candidates representing the right or the left tend to begin distinctly to the right or to the left and then move progressively closer to the center as they attempt to appeal to a greater number of voters. The appeal of the median position in a two-candidate contest, however, is precisely what makes such a contest vulnerable to assault from either side by a third or a fourth candidate. In any contest with two candidates near the center a third candidate entering on the left or the right can always gain a plurality. Indeed, for practically any distribution of the electorate there are no positions in a two-candidate contest where at least one of the candidates cannot be beaten by a third. As is shown in the accompanying illustration [*lower illustration on page 18*], there is always a place along the one-dimensional continuum where a new candidate can position himself to displace one or more nearby candidates.

A single issue rarely plays a deciding role in an election. Hence election analyses based on single issues are not very helpful, unless they can be combined to show how to design a platform that will ensure a candidate's election. Shaping a winning platform is a complex business, however, because it is possible for a platform consisting entirely of winning, or majority, planks to be defeated. The

reverse side of the coin is that a majority platform can be constructed from minority planks. Hence a majority can be formed from a coalition of minorities.

To see how this paradox can arise consider the simplest possible case: a ballot to decide two unrelated, dichotomous issues, represented by resolutions *A* and *B*. In this case the voters actually have four options:

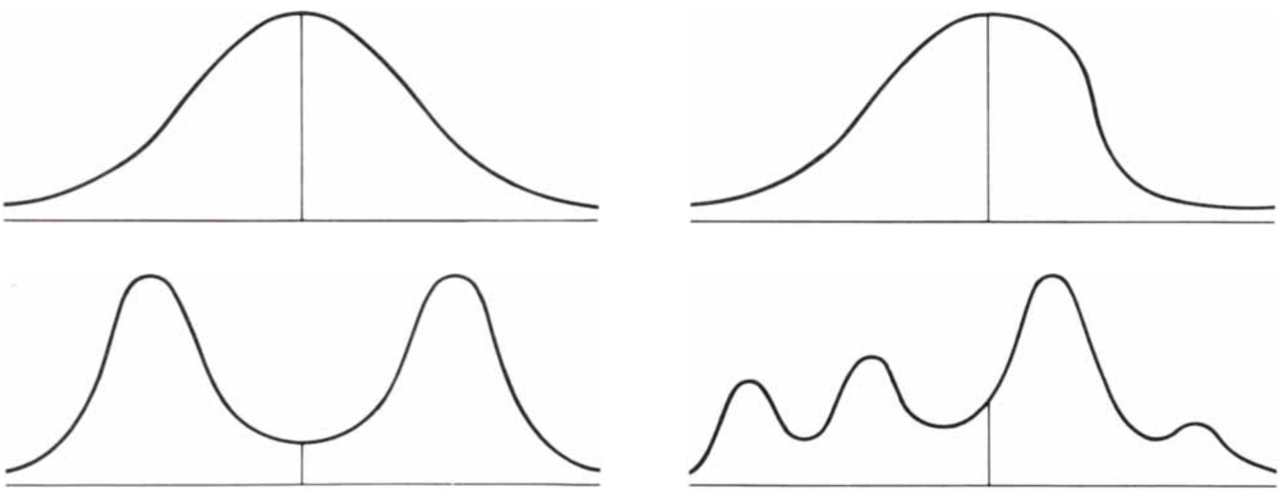
I. Approve *A* and *B*.
II. Approve *A* and defeat *B*.
III. Defeat *A* and approve *B*.
IV. Defeat *A* and *B*.

The voters who favor both *A* and *B* would choose option I as their first choice, option IV as their fourth choice and option II as their second or third choice, depending on whether they feel more strongly about *A* or about *B*. The voters who favor *A* but object to *B* might rank the four options in the order II, I, IV and III (or II, IV, I and III). In general each voter will have a preference ranking for one of the 4 × 3 × 2 × 1, or 24, possible permutations of the four available options. (The rankings are by no means equally likely; it would be hard to imagine circumstances under which many people would rank the options in the order of preference I, IV, II, III.)

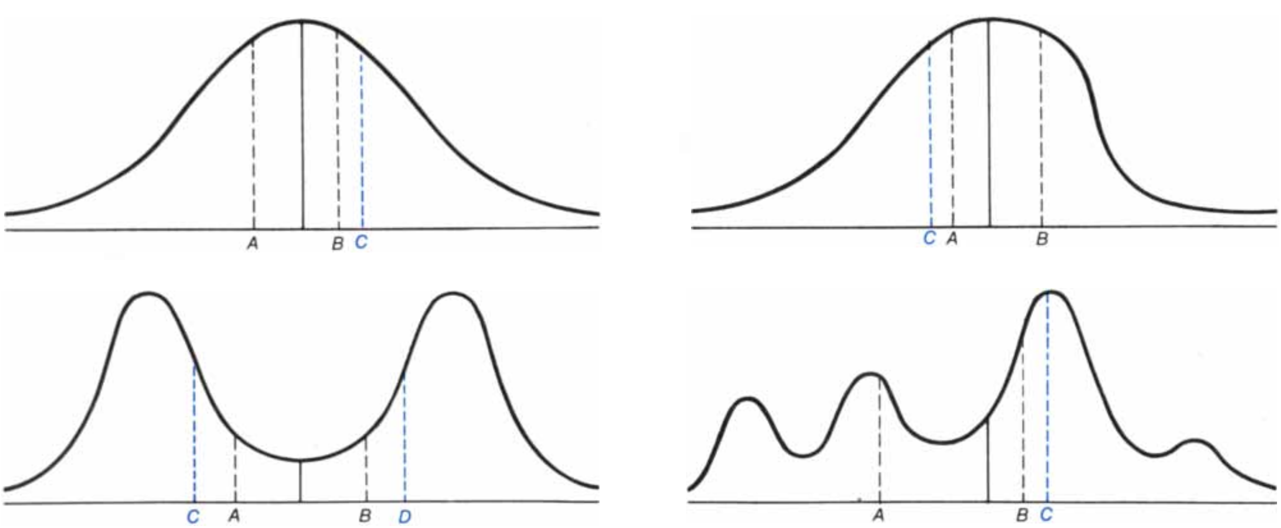
Now, for the sake of simplicity suppose 500 voters (say at a party convention) are divided into three caucuses as follows: caucus *X*, with 150 votes, ranks the four options in the order I, II, III, IV; caucus *Y*, with 150 votes, ranks them II, IV, I, III, and caucus *Z*, with 200 votes, ranks them III, IV, I, II. In this case caucuses *X* and *Y*, with 300 votes, favor the approval of resolution *A*, whereas caucuses *X* and *Z*, with 350 votes, favor the approval of resolution *B*. Because there are different voters making up these majorities, however, the platform consisting of the planks "Approve *A*" and "Approve *B*" will be defeated by the 350-vote block of caucuses *Y* and *Z!*

This surprising phenomenon is a special case of the well-known anomaly of cyclic majorities: If three voters respectively prefer *A* to *B* to *C*, *B* to *C* to *A*, and *C* to *A* to *B*, then any candidate can be defeated by some other candidate by a vote of two to one in a two-candidate contest. When the issues in an election create cyclic majorities, no set of positions on the issues is invulnerable to assault by a new coalition of minorities, another factor that encourages third- and fourth-party candidates.

The accompanying diagram [*top illustration on page 23*] shows how the four options from which party planks in the example must be constructed create a variety of cyclic majorities, thereby explaining how a platform consisting of majority planks can represent the will of only a minority. The arrows joining various platforms depict voting dominance: the platform to which an arrow points will always lose to the platform at which the arrow originates in a dichotomous contest. The winning caucuses in each case appear beside the corresponding arrow. As this distribution demonstrates, any possible platform can be defeated by some other platform, and so a real convention whose divisions resem-



Four possible shapes of public opinion



Two candidates near median (**A** *and* **B**) *can be defeated by a third candidate* (**C**) *and sometimes a fourth* (**D**)

Painting by Wilson McLean

The future is coming. And with it will come great benefits for mankind. And a whole new set of problems. Because we are a forest products company, and plant seeds that take up to 50 years to become mature trees, Champion International has to think a lot about the future. We'd like to share some of the things we've learned with you—to help you make intelligent choices in the years to come. Here is something you might want to think about.

In the future, nations will be dividing up mankind's common inheritance: the vast wealth of the oceans.

But will we do it recklessly or with vision?

The shortages facing the world today could be coming to an end sooner than we think. The seas—the last, largely untapped resource left on earth—have the potential for supplying far more energy, raw materials and food than man has ever dreamed of. But the stakes are so high, it will not be an easy matter to control.

Who owns the oceans? Which countries will get how much? Since an unthinking grab at these resources could irrevocably damage even as vast an entity as the oceans, how can a gold-rush mentality be avoided?

Consider the lures beckoning us seaward.

There is the sea as energy source.

Soon—because people will be able to work in pressurized structures—it will be possible to move off the shallow continental shelves, far out into the deep sea thousands of feet down, where there may be vast unknown fields of oil and natural gas. This alone could radically alter the economics and geopolitics of the planet.

But even the moment-by-moment heaving of the sea itself is energy: tappable, renewable, literally endless since it plugs into the kinetic energy of the solar system itself.

As research money flows down to the sea, even the temperature difference between deep cold water and surface warm water will be yet another basic energy source, fed by the sun.

There is the sea as mineral source.

One example is the mysterious bit of magic by which the sea seems to *mine itself*. As pearls form around a grain of sand, valuable metals cluster around such ocean debris as sharks' teeth. The process ends with a potato-sized chunk of manganese, iron, nickel, copper and cobalt, called a "nodule." These lie in countless millions on the sea bottom, waiting to be simply vacuumed up.

Then there is the sea as pantry.

Before long, salmon will be cultured in pens on a commercial basis. Prawns, lobsters, oysters, hardshell crabs, all are being grown experimentally. Giant kelp farms (it's rich in protein) may eventually be as important as grain farms.

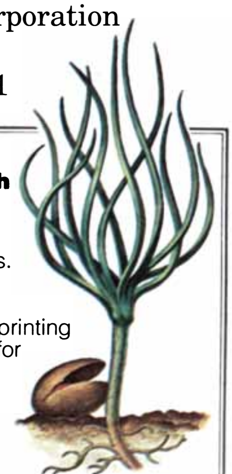
But if we are to enjoy this largesse, the sea must be handled with care, the way a forest is harvested for its wood, with pains taken to ensure its continual renewal. Because there aren't Seven Seas, there is only The Sea, everywhere interconnected, touching all shores, common to all; it belongs to the Family of Man. Indeed, it is the Family heirloom, and all should treat it with veneration.

You can learn more about some of the other critical issues we face in the future and get a bibliography for further reading by sending for our free brochure. Write:
Champion International Corporation
Dept. 200A, P.O. Box 10144
Stamford, Connecticut 06921

**Champion—
a forest products company with its roots planted firmly in the future.**

We are in the forest products business. We plant trees, grow trees, harvest trees. And from trees we make wood building products. Plus fine paper for printing and business. And paper packaging for shipping and selling.

Because we make our living from the forest, our success depends, in one way or another, on the future. And we're planning—and planting—for it.



Champion
Champion International Corporation
Planting seeds for the future

Options

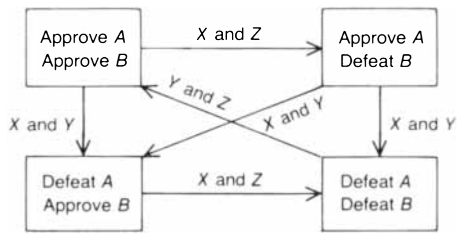
I. Approve resolutions *A* and *B*.

II. Approve resolution *A* and defeat resolution *B*.

III. Defeat resolution *A* and approve resolution *B*.

IV. Defeat resolutions *A* and *B*.

Caucus	Policy	Votes	Order of Preference
X	Favors *A* strongly and favors *B* mildly	150	I, II, III, IV
Y	Opposes *B* strongly and favors *A* mildly	150	II, IV, I, III
Z	Opposes *A* strongly and favors *B* mildly	200	III, IV, I, II



Three caucuses voting on two platform planks create cyclic majorities

ble the ones given in this example could become mired in an unending sequence of platform motions, with each motion defeating the one before.

The phenomenon of cyclic majorities is also responsible for the most famous election paradox, Kenneth J. Arrow's 1951 proof that certain generally accepted desiderata for voting schemes are logically inconsistent. If there are only two candidates, no problems arise. If three or more candidates appear on a single ballot, however, chaos reigns.

There are diverse schemes other than plurality voting for determining the winner in an election. Many were suggested by 18th-century scholars concerned about implementing the democratic ideals of the French Revolution. Although some of these proposals are so complex as to be completely impractical, several are still in common use, in particular the method of assigning points that reflect degrees of preference to the candidates in a contest (where the candidate receiving the most points is the winner) and various methods of holding runoff elections. Yet as Arrow has shown, none of these schemes—indeed, no method other than a rational benevolent dictatorship—satisfies such commonsense rules as: If *A* is preferred to *B*, and *B* is preferred to *C*, then *A* should be preferred to *C*. Cyclic majorities reduce all voting schemes to unpredictable mystery. (For a discussion of Arrow's proof see "Mathematical Games" for October, 1974.)

Another important problem with voting in three-option contests is that in many circumstances a vote for the candidate a person prefers most will increase the likelihood that the candidate he prefers least will be elected. (This dilemma was the one often seen in Anderson's candidacy. Many voters who preferred Anderson to Carter and Carter to Reagan believed most Anderson votes would be at Carter's expense.) The anomaly frequently leads thoughtful voters to what is called (depending on a voter's point of view) insincere or sophisticated voting.

If sophisticated voting is widely practiced, it can lead to a state of serious confusion where no one votes for his first choice, and so the public will is effectively camouflaged. An Anderson backer for whom Carter was a second choice might have voted for Carter instead of for Anderson in order to prevent the election of Reagan. If there were enough Anderson backers who reasoned this way, of course, some Reagan supporters might have begun to support Anderson to prevent Carter's reelection. The process of second-guessing the voting strategies of other segments of the electorate can quickly lead to an absurd hierarchy of insincerity in which the votes cast fail to reflect real preferences. Such a process, which it should be added is more a part of game theory than of classical voting theory, rarely gives a legitimate mandate to the victor.

Arrow's theorem shows there is no "perfect" voting scheme for multicandidate elections. The procedure known as approval voting, however, manages to reflect a popular will without inducing anyone to vote insincerely. In approval voting each voter marks on the ballot every candidate who meets with his approval, and the candidate who receives the most votes of approval is the winner.

With this system it is never to a voter's advantage to withhold a vote for his first choice while voting for a less preferred candidate. Indeed, if most candidates seem to have an equal chance of winning, a rational voter should vote for all the candidates he believes are above the average of those running. To vote for more candidates would give unnecessary support to individuals the voter does not endorse, whereas to vote for fewer candidates (say to vote only for one's first choice) is to withhold support from an acceptable compromise candidate and to risk victory by an unacceptable candidate.

Steven J. Brams, professor of politics at New York University, has described approval voting with the phrase "One man, *n* votes." It is an apt description because approval voting is merely a way of letting a person cast as many votes as he wishes, one for each acceptable candidate. It is easy to count votes that have been cast under this system, and no runoff elections are needed. For both theoretical and practical reasons approval voting is a good compromise between the single-vote ballot that encourages insincerity and the complete preference ordering whose complexity renders it useless in any practical situation.

The accompanying illustration [*below*] shows how approval voting might compare with plurality voting, runoff voting and point voting in an entirely hypothetical three-way contest. The number of voters supporting each of the six possible rankings of candidates are listed in the column "Total votes," and since *C* would receive the largest block of first-

		Approval votes	
Order of preference	Total votes	First choice	First and second choices
A, B, C	30	20	10
A, C, B	5	5	0
B, A, C	20	10	10
B, C, A	5	5	5
C, A, B	10	5	5
C, B, A	30	20	10
Total	100	65	35

Plurality voting	Runoff voting	Point voting	Approval voting
A 35	*A* 35 + 20 = 55	*A* 35 (3) + 30 (2) + 35 = 200	*A* 25 + 10 + 15 = 50
B 25	*C* 40 + 5 = 45	*B* 25 (3) + 60 (2) + 15 = 210	*B* 15 + 10 + 20 = 45
C 40		*C* 40 (3) + 10 (2) + 50 = 190	*C* 25 + 15 + 0 = 40

Comparison of voting methods for a three-way race

MCAUTO STRUDL helps solve design problems fast.

Whether you're performing static or dynamic analysis, or steel and concrete design—for a bridge, building, or tower—you can save valuable time with our structural analysis programs. For example: MCAUTO STRUDL. Just enter the data in common engineering terms and in the sequence of your thought process rather than the formats required by many computer programs. Its automated steel design capabilities result in time saved because the system selects the members and lets you choose any of seven design codes.

These capabilities—and many more—are right at your fingertips. And you don't have to be a computer programmer to use MCAUTO STRUDL.

Data processing ideas at work in engineering.

Take the stress out of piping analysis.

Our piping stress analysis program, PIPELINE,® is loaded with timesavers that make engineering easier. And more profitable. Now, one-directional and limited-movement restraints are analyzed automatically. The best spring hanger design to support weight and movement is selected by you. A detailed report shows that all code conditions have been satisfied.

Best of all, our graphics capability improves error detection by immediately uncovering any geometry errors. You can even "window in" for a closer look at detail, label points and display coordinates.

We have another program to reduce those nagging profit eaters—material costs, work scope changes and panic scheduling. It automates material control, management reporting and isometric drawings. Designed by experts in piping management and design, it offers you a one-two punch: Material control and automated isometrics.

We call it FASTDRAW™ for the obvious reason.

Can you model a cylinder and torus—with only radius and end points given—in 45 minutes? You can with FASTDRAW, our low-cost, interactive graphics modeling system. You enter design data, and spot and correct errors within seconds, before you make a computer run. Fully automatic, easy-to-use interactive mesh generation, complete with powerful surface options, enables you to create error-free models ready for structural analysis. That means time saved, more hours for investigating optional designs and better final designs—the work you do best.

Energy conservation starts before construction.

The best way to conserve valuable energy resources in a building is to start before you break ground. That's exactly what architects are now doing. Working with our computer programs, design engineers use weather data, peak demand estimates, lighting and other building and environmental features to simulate the proposed building's heating and cooling requirements. These needs are then factored into the final building plans.
The result is good building design combined with full energy efficiency.

MCAUTO Engineering programs have been developed with you, the engineer, in mind. They're highly enhanced programs, characterized by unique design features. Yet they're easy to use, and fully supported by the company geared for computerized engineering: MCAUTO. If you'd like more information about any of these programs, or if you'd like to discuss a particular program we've not covered here, call us at (314) 232-8021. Or mail the coupon.

MCAUTO®
MCDONNELL DOUGLAS
Setting the standard in data processing

MCAUTO is a registered trademark of McDonnell Douglas Automation Company.

W. R. Vickroy
MCAUTO, Dept. KI-500
Box 516, St. Louis, MO 63166

☐ Please send me more information on MCAUTO Engineering Programs.

☐ Have your representative phone me for an appointment.

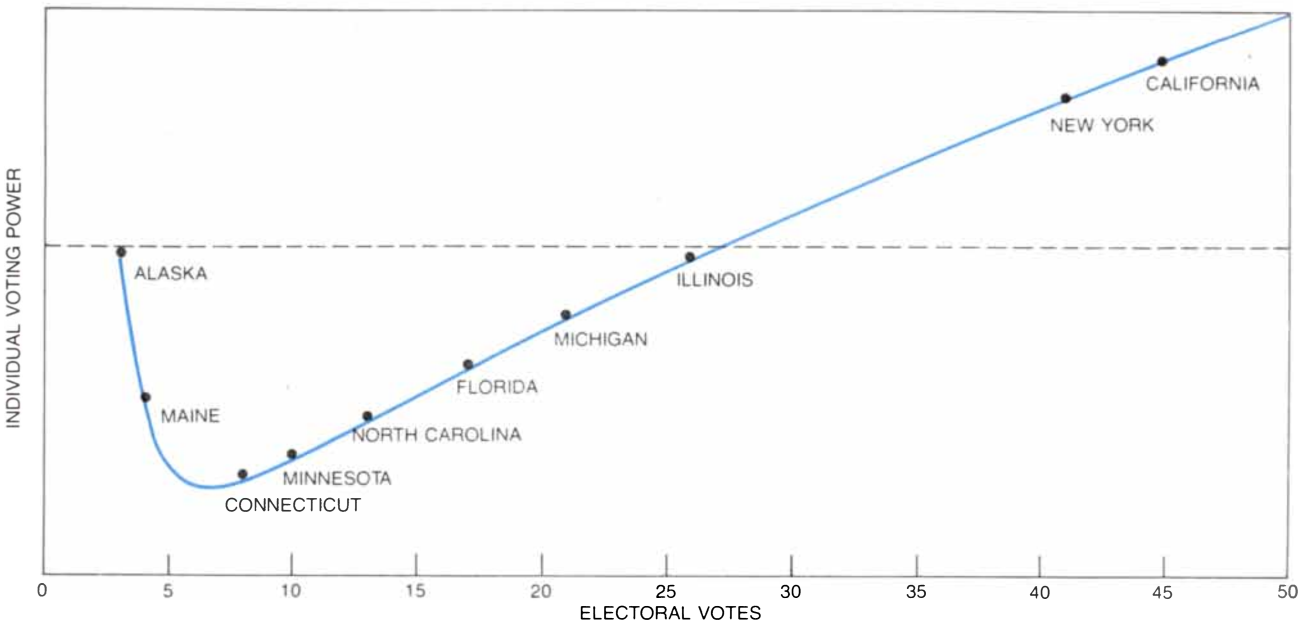
Name ____________________ Title ____________

Company Name ____________________ Phone ____________

Address ________________________________

City ____________ State ____________ Zip ____________

A JOB? LET'S SWAP RESUMES. SEND YOURS, WE'LL SEND OURS. AN EQUAL OPPORTUNITY EMPLOYER. WRITE: BOX 14526, ST. LOUIS, MO 63178.



Graph of individual's voting power in various states

choice votes, he would win in a plurality contest. In a runoff election *B* would be eliminated, and *A* would pick up enough second-choice votes (from those who had first voted for *B*) to defeat *C* by 55 votes to 45. In the simplest system of point voting first choices are assigned three points, second choices two points and third choices one point. Because of the large number of voters (60) for whom *B* is the second choice, with this voting scheme *B*, who was eliminated in the runoff, would be the winner.

The results of approval voting depend on whether voters find only their top choice acceptable or whether they could accept some other choices as well. (Because there are only three candidates in this example it is assumed that no one votes for all three; such a vote is legal, but it would be wasted since it would raise each candidate's total by the same amount.) In this case, with 65 voters approving only their first choice, *A* would receive 50 votes of approval and win the election. If some voters choose to approve two of the three candidates, however, *B* stands to gain most because of the large number of people who rank him as their second choice. With approval voting a shift in the number of candidates meeting the approval of even a small number of voters can easily change the outcome of the election. Hence the implementation of this voting scheme would necessitate a transformation of campaign strategies, from trying to convince voters that a candidate is the best choice to trying to convince them that he is acceptable.

In the U.S., of course, presidential elections are held by the totally different rules of the electoral college. Through most of U.S. history the electoral college has served mainly to impose a unit rule on individual states so that the winner of the popular vote in each state receives that state's entire electoral vote. According to the Constitution of the U.S., there are other significant consequences of this system that affect the outcome of three-candidate contests (in particular provisions for transferring the responsibility of deciding a presidential election from the electoral college to Congress), but here we shall examine only the consequences of the unit rule.

The most widely held view of the electoral college's unit rule has been that it favors smaller, or less populated, states, because the number of votes accorded to each state in the college is two more than its number of representatives. In relative terms these two extra votes, which represent the two senators from each state, do increase the voting strength of smaller states and diminish that of larger ones.

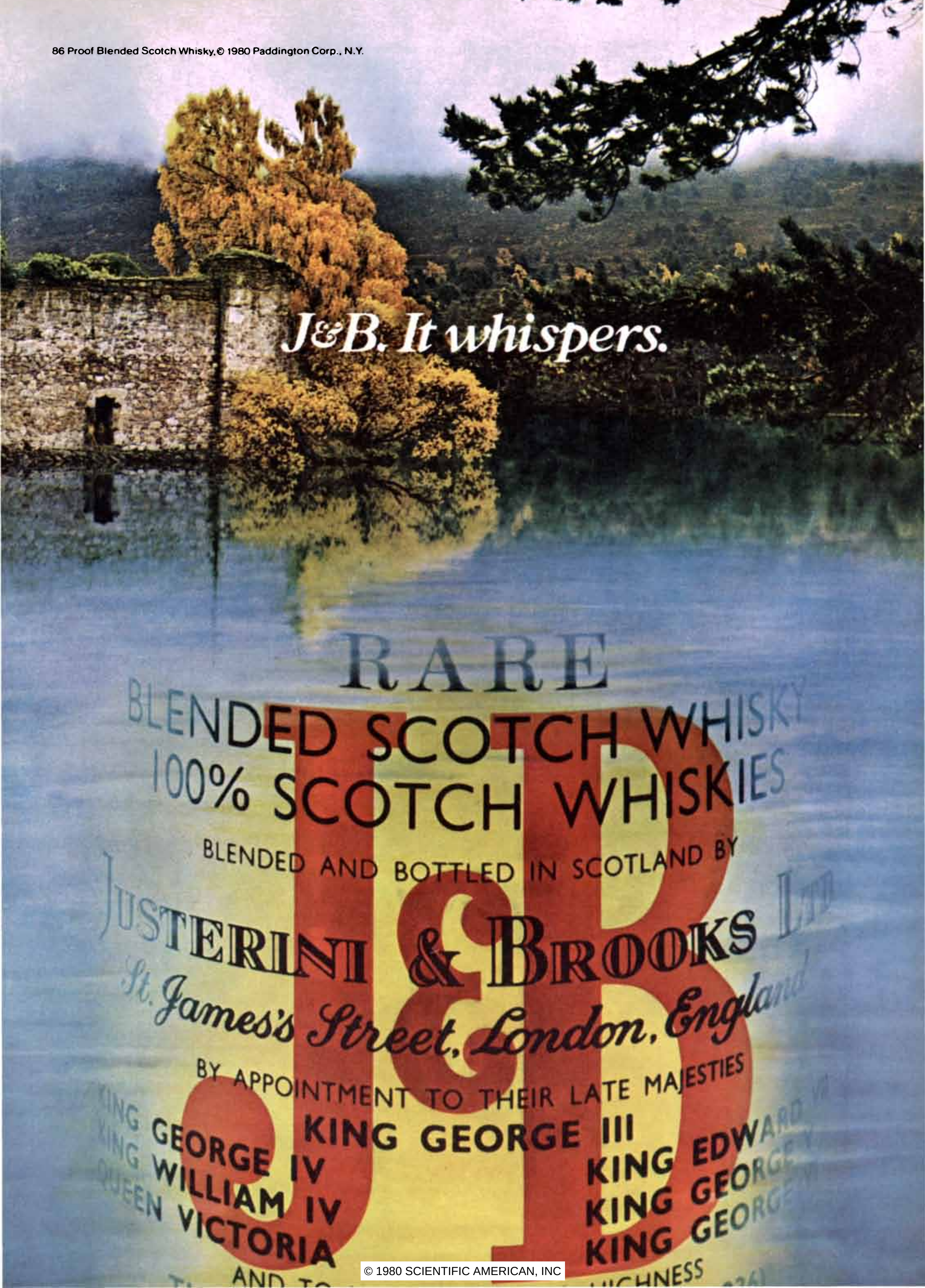
Paradoxically, however, the effective strength of a state in a presidential election is actually proportional to the population of the state raised to the 3/2 power. And as a result individual votes cast in the largest states are as much as three times as important as those cast in the smallest ones. This surprising conclusion is a direct consequence of elementary probability theory, and it is consistent with the spending record of candidates in recent presidential elections: candidates do devote disproportionate resources to the larger states at the expense of the smaller ones.

The "3/2 rule" is based on the assumption that candidates will generally match one another's campaign efforts in the various states. (Comparison of candidates' allocations of time and money in recent election campaigns shows that the assumption is entirely realistic.) The reasoning begins with the obvious: Each candidate seeks to maximize his expected electoral vote, which is the sum over all 50 states of the product of each state's electoral vote and the probability that the candidate will win a majority in that state. By expressing this relation in the form of an equation and taking into account candidates' tendencies to match one another's campaign efforts from state to state it can be shown that the optimal way to maximize the expected electoral vote is to allocate campaign resources approximately in proportion to the 3/2 power of the electoral vote of each state. Thus although California has about four times the electoral vote of Wisconsin (45 compared with 11), the 3/2 rule would suggest that candidates should devote $4^{3/2}$, or 8, times more resources to California than to Wisconsin.

Another way to understand why larger states gain power rather than lose it in electoral-college politics is to examine the likelihood that any particular vote may be decisive in swinging the state for or against a particular candidate. This measure of decisiveness is the traditional way of gauging the power of an individual voter. What is needed is a measure of the average number of votes necessary to reverse the result of an election in each state.

Calculations show that the decision power of an individual in a state with v electoral votes (to be cast as a unit in the electoral college) is proportional to $\sqrt{v}$. Since the power of a state in the electoral college is magnified by the number of electoral votes cast by the state, the contribution of each state to the presidential decision is approximately proportional to v times $\sqrt{v}$, or $v^{3/2}$.

In order to gauge the relative voting power of individuals in different states


86 Proof Blended Scotch Whisky, © 1980 Paddington Corp., N.Y.
J&B. It whispers.
RARE
BLENDED SCOTCH WHISKY
100% SCOTCH WHISKIES
BLENDED AND BOTTLED IN SCOTLAND BY
JUSTERINI & BROOKS
St. James's Street, London, England
BY APPOINTMENT TO THEIR LATE MAJESTIES
KING GEORGE III
GEORGE IV
WILLIAM IV
VICTORIA

the large-state bias created by the 3/2 rule must be weighed against the small-state bias of the two-senator electoral-college bonus. The significance of an individual's vote, instead of being equal for all voters, is determined by the individual's share of his state's power, and as is shown in the accompanying illustration [*page 26*] the different states' powers are decidedly unequal. (The broken line on the graph marks the hypothetical even distribution of power.)

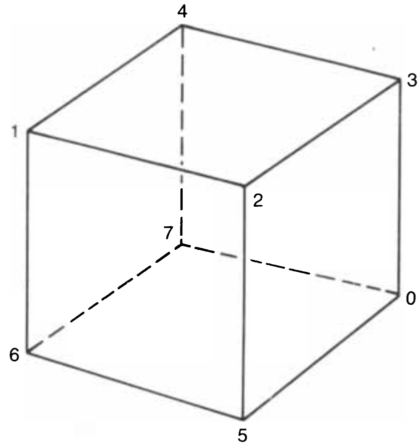
Elections will always remain a matter of passion more than of logic, based on belief more than on reason. As these examples demonstrate, however, the mathematics of elections can have subtle and unexpected consequences. As in many other realms of human experience, naive expectations can be shattered by simple mathematical structures disguised as paradoxes and anomalies.

The above concludes Steen's article. The answers to last month's problems are as follows.

The four prices Iva paid for the trinkets at the Great Bazaar are $1, $1.50, $2 and $2.25. Both the sum and the product of this set are $6.75. If it had not been specified that Iva paid $1 for the earrings, there would have been a second solution to the problem: $1.20, $1.25, $1.80 and $2.50. The problem, posed by Kenneth M. Wilke, is a variant of one devised by David A. Grossman that appears as No. 65 in L. A. Graham's *Ingenious Mathematical Problems and Methods* (Dover Publications, 1959). In Grossman's version both the sum and the product are $7.11, and the unique set of prices is $1.20, $1.25, $1.50 and $3.16.

A second way to dissect a cube into three congruent solids is to simply slice it into three parallel slabs. I know of no third way to cut a cube into three identical parts. The illustration on this page shows the unique labeling of a cube's corners with the integers 0 through 7 (ignoring rotations and reflections) so that the sum of the pair of numbers at the ends of each edge of the cube is a prime. It is not hard to prove there is no solution using eight consecutive integers greater than 0.

The paradox of the surface-to-volume ratios of the sphere and the cube is the first problem in Christopher P. Jargocki's *Science Brain-Twisters, Paradoxes and Fallacies* (Charles Scribner's Sons, 1976). The fallacy springs from the fact that two entirely different meanings are applied to the symbol d. The proper way to compare the surface-to-volume ratios of the sphere and the cube is to consider two solids of equal volume. A cube of volume v has an edge of $v^{1/3}$ and a surface-to-volume ratio of $6(v^{1/3})^2/v$, or $6/v^{1/3}$. A sphere of the same volume has a diameter of $(6v/\pi)^{1/3}$ and a surface-to-volume ratio of $\pi(6v/\pi)^{2/3}/v$ or about $4.8/v^{1/3}$.



Solution to last month's cube-labeling problem

The surface area of a sphere, Jargocki points out, is about 20 percent less than that of a cube of the same volume. Similar reasoning shows that for a given surface area the volume of a sphere exceeds that of a cube by about 39 percent.

The following surprising theorem, which was sent to me several years ago by Thomas D. Waugh, is a closely related fallacy. If a sphere of radius r is enclosed in a polyhedron in which each face touches the sphere, the surface-to-volume ratio of the polyhedron, regardless of its shape or number of faces, is $3/r$. But $3/r$ is also the surface-to-volume ratio of the sphere! The two paradoxes are good introductions to what physicists call dimensional analysis, the technique of expressing constants in dimensionless numbers independent of adopted units of measurement.

In several columns I have discussed the problem of placing n superqueens (pieces that move like knights as well as like queens) on an n-by-n chessboard so that no superqueen attacks another. This problem was unsolved for all values of n, but now the conjecture that solutions exist for all n greater than 9 has been settled affirmatively. Last year Ashok K. Chandra of the Thomas J. Watson Research Center of the International Business Machines Corporation completed a proof for all cases except $n = 20$, $n = 21$, $n = 32$, $n = 33$ and $n = 57$. As I reported in May, solutions for these cases have now been discovered, and so the problem is solved. Another general solution was sent early this year by J. Reineke and P. Päppinghaus of the University of Hannover. And in June a third solution came from Charles Zimmermann of Madison, Wis. Zimmermann and Paul Steves, also of Madison, independently wrote computer programs for determining all the basic solutions on low-order boards. There is just one solution for the case $n = 10$, and there are six for $n = 11$, 22 for $n = 12$ (not 23 as I previously reported), 239 for $n = 13$ and 653 for $n = 14$.

Minnesota... Technology Wellspring

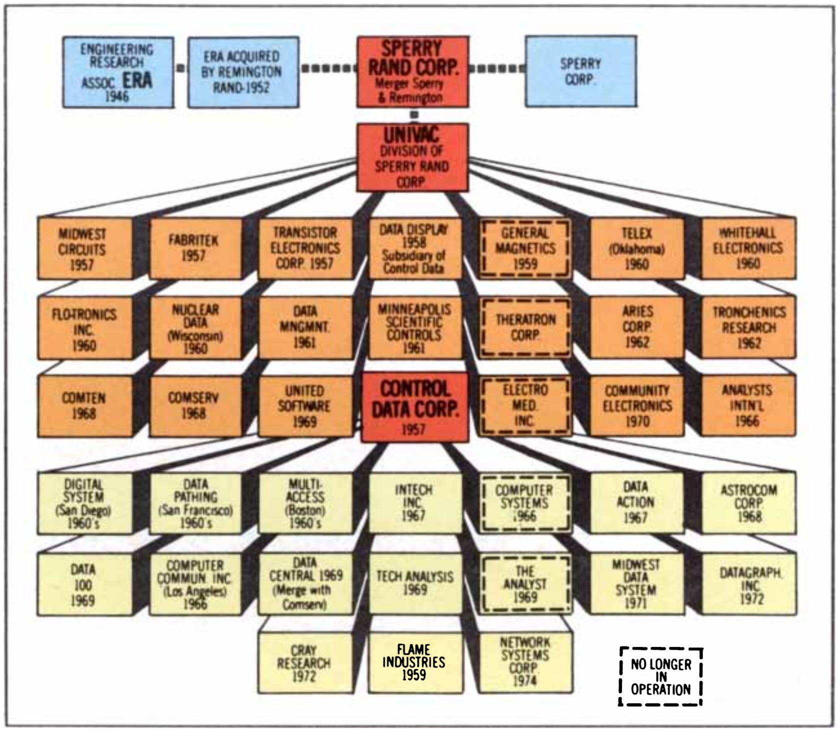
This special report on advances in high technology throughout the United States was written by Peter J. Brennan, writer on international business, finance and development in concert with Development Counsellors International, Ltd.

Minnesota. There is much unique about this North Central State, but an expected uniqueness that hides a surprise.

Most northerly of the continental 48 states (the Northwest Angle is about 120 miles further north than Maine), it is also the source of the Mississippi River, already navigable at the twin cities of Minneapolis and St. Paul. Well-named the Land of Sky Blue Waters, with its 20,000 or so lakes and numerous rivers, the state is a camping canoeist's delight. Once it was a lumberman's dream for the enormous forests that covered it. And most of the iron and steel in the U.S. came from the Mesabi Range and other Minnesota deposits.

Endowed with rich land and settled by essentially farming people, the State was an agricultural center by the time of the War Between the States. In many respects, the state much resembles its neighbors, which lacked forests and iron ore, though these in the end proved transitory.

But unlike its neighbors, Minnesota has also developed a tradition of innovative technology that is both unique and a surprise.



Examples of technical company formations and spin-offs originating with Engineering Research Associates (ERA) in 1946. (all are Twin Cities companies except as noted)

The Technology Farm.

More surprising still, Minnesota technology ventures tend to stay put and grow. Other areas that are remote from the nation's largest business and financial centers, while they may spawn new ventures, tend to lose them later. But in Minnesota, too often to attribute the phenomenon to chance, company after new company has succeeded at a rate far higher than the national average. Nationally, 4 of 5 new ventures fail; in Minnesota, 4 of 5 succeed. As these enterprises founded on ideas and advanced technology rather than on raw materials and heavy industry grew and prospered, they remained in the area. Around them grew up new companies. And thus the state has become an unlikely center for advanced technology.

The region justly claims to be the birthplace of the modern computer and the continuing source of advanced computer design. But there are also medicine, instruments and controls, materials and chemicals, word processing, communications, microelectronics and solid-state technology. Nor does this technology exist in a rarefied academic atmosphere. There is in Minnesota a social/business climate that actively fosters cooperation among companies, the educational system and the state and local government to aid new ventures and to develop new technology.

Professor Roger Staehle, Dean of the Institute of Technology at the University of Minnesota, sees the phenomenon in terms of Japan where the well known formal and informal cooperation among business, finance and government has given the world "Japan, Inc." Dean Staehle, building on the high degree of corporate interaction already present in Minnesota, would like to see "Minnesota, Inc." grow out of it.

There's a good base. With but 1.7% of the nation's population, Minnesota has 4.4% of the nation's biggest businesses. Some 440 electronics companies make their home here and employ 100,000 people. Moreover, of the world's five largest computer companies, three were founded in Minnesota and remain there. A fourth, IBM, has a major facility in the state. A sixth, Cray Research Corp., is small but claims to make the world's largest and fastest computer.

The Nurturing Factors.

Why Minnesota? William C. Norris, founder and Chairman of Control Data Corporation, which went from zero sales in 1957 to over $3.3-billion in 1979, has a theory. "I think the environment in Minnesota is best in the world for innovation," he says. "Why?" he asks. "There's nothing else to do here in the wintertime."

The Twin Cities suffer the coldest winters of any major North American metropolis. And since years of scholarly research have not found any other single reason for the state's high standing among centers of business and technology, the climate is as good a one as any.

Other reasons cited: an ethnic mix that retains a work ethic from another place and time; a spirit of self-reliance combined with a communal outlook that causes people, groups and organizations to work together uncommonly well; a personal attachment to the place that makes transferred corporate Minnesotans frantically search for ways to

stay in or return to the state even if it means starting a business; a sense of intimacy between senior people at major corporations and the community—they are not only in it, they are of it; the presence of a major university on a single campus. The financial community has also welcomed new ideas. Absent from this list is government. The communal spirit that may foster innovation also produced a Populist philosophy that has its base in the small farmer and harbors a deep suspicion of the business community. The resulting legislatures and tax structures favor neither business nor salaried employees. Change is coming, though. Business, academic and citizens' groups are working with legislators to reform the tax climate.

Innovation through Invigoration.

Minnesotans make a virtue of the climate. It was there before the Mayo Clinic at Rochester, or the Consolidated Temperature Control Company. But the climate was at least partially responsible for both since the state was once the nation's number one health resort thanks to its widely advertised "pure and invigorating air" and the Consolidated Temperature Control Company started with a device to spare householders the need to get up in a cold house.

A.M. Butz invented the automatic thermostat in Minneapolis in 1883. "The severe winters here," remarks Dr. William T. Sackett of Honeywell, "gave more incentive to invent that sort of thing here." Dr. Sackett is vice president in charge of Honeywell's Corporate Technology Center where that sort of thing still goes on.

The heat regulator was ahead of its time. Butz's Consolidated Temperature Controlling Co., founded in 1885, was failing. Reorganized, the struggling company resorted to what has long since become a classic Minnesota tactic. The company approached an 'Angel', a wheelbarrow maker named William R. Sweatt. He put $1,500 into the venture. The company again went through a rocky period and was reorganized in 1892 as the Electric Heat Regulator Co. In 1896 another financial crisis gave Sweatt 100% of the company for another $5,000.

Business was slow. People can be very obtuse about accepting new technology. But the company grew, became the Minneapolis Heat Regulator Co., and in 1927 merged with Honeywell Heating Specialties Co. to become the Minneapolis Honeywell Regulator Co., headquartered at the site of Sweatt's old wheelbarrow plant, where today's company, Honeywell, Inc., still is. The $4.2-billion company now has about 18,000 people in the Twin Cities area of whom some 25% are professionals. More than half of those are engineers.

Honeywell still makes heat regulators and thermostats. It has also become one of the world's leading instrument and control manufacturers and, more significantly, leapfrogged that to become a full-scale computer company, in 1955, when it started a computer business with Raytheon as Datamatic Corp. Since then, the company acquired or started many other computer ventures including all General Electric's computer business. That made this controls company the world's second largest computer company.

Energy, however, remains a prime interest, not least, solar energy. "We started on coating technology for solar work in the early 1960s," recalls Dr. Sackett. "We've been expanding since then." Most visible evidence of the expansion is a large demonstration solar plant at the company's headquarters. Says Sackett, "Now we've got more than a million in it. It should pay out in about 100 years. But it's an experimental test bed," he hastily emphasizes. "Our main interest is in the control aspects of solar energy."

Demonstrating the company's commitment both to solid state technology and to the further development of advanced technology in the region is its involvement with the Microelectronics and Information Center at the University of Minnesota's Institute of Technology. Honeywell has contributed $2 million as an initial infusion to the microelectronics center and another $1 million to the University's Business School "to get more going in computer-oriented management information systems," says Dr. Sackett.

Efficient application of advanced computer technology is paramount. To that end, the company started three years ago a corporate computer sciences center that has gone from none to 35 scientists. Remarks Dr. Sackett, "We have to find and teach more disciplined ways to develop software. Also, how do you get 100,000 items on a chip? We need people who understand systems, designers who know solid state electronics, people who can appreciate that technologies are merging, becoming a single spectrum from the sensor to the control computer, not a series of disconnected technologies."

Perhaps it's coincidental, but the role of the "Angel" seems to have been a large one in Minnesota technology, and remains so. The state's largest private employer, 3M, formerly Minnesota Mining and Manufacturing Co. and still called 'The Mining' by oldtimers, got off to an even shakier start than Honeywell. Shakier because the company's founding premise was wrong. Again, it was an Angel to the rescue.

The company was founded in 1902 to mine a deposit of supposed corundum. Then it developed that the deposit wasn't corundum at all. The promoters changed course and started making sandpaper with the stuff, which was no good either. Only when they bought more traditional abrasives did they have any success. But it was slow.

In St. Paul, Lucius P. Ordway, general manager of a plumbing firm, for some reason kept pouring money into Minnesota Mining as it went from crisis to crisis. He moved the company to St. Paul where he could keep a closer eye on it as he pumped in $250.000. By 1914 the sandpaper business was beginning to make money. Then a materials defect nearly ruined the company, again. An application of primitive technological innovation solved the problem and the company at last was well founded.

Today the $5.4-billion company is regarded as an innovative company with an entrepreneurial spirit. Sandpaper is still a large profit center. But other products run the gamut: business products; traffic control; safety systems; electrical and electronic products; health care products; recording materials; even household items. True to its beginnings, the company is a supplier of things that other industries use.

As such, its products must be at the leading edge of its customers' technologies. Regarding the computer business, for example, General Manager of the New Business Ventures Division, Dr. J. Tait Elder, remarks "We make exquisite connectors; we provide extraordinarily reliable recording media. But basically it's a matter of high sophistication in the low technology of painting—coatings, if you please. We believe it is an achievement to make something complicated useable in its simplest sense."

Central to the company's development is its thirty-year-old New Ventures Division. A new idea unrelated to a current business gets support from the top through Dr. Elder to bring it to a level where it appears viable, commercial or possible. Then Dr. Elder's Division offers the venture to the operating divisions.

While some of these ideas may seem to spring full blown upon the world like Venus from the sea, Dr. Elder leans to the Cinderella analogy. "We only think of Cinderella on the night of the Ball," says he, "but we tend to forget she was around being formed for eighteen years before that. So it is with new technology and new ventures."

3M, a self-contained firm, has great depth of people and promotes from within. But it does develop shortages of skilled people in new areas. "Once we get them in though," says Dr. Elder, "we have no trouble growing more of our own people. And good people attract good people."

3M's self-containment is physically evident in its complex of buildings. Relations with other companies and institutions in the region are cordial, polite and contributory, but not intimate. Dr. Elder, for example, acknowledges 3M's debt to the University of Minnesota since one of the company's early lights was induced to leave the school to help the company. "We have benefitted from it," he says, "But our relationship is basically platonic." Similarly, the company's involvement in projects such as the Institute of Technology's microelectronics project, the Minnesota Cooperation Office, Minnesota Inc. and the like tends to be pro forma. "We are almost aloof—preoccupied is a better word—without meaning to be," says Dr. Elder.

A third Angel gave rise to an entire industry. John E. Parker, a builder of Army gliders found himself with no business in 1945. A group of Navy veterans headed by William C. Norris had embryonic computer technology, ideas on how to exploit it, but no money.

Norris and his associates got together with Parker, who provided financing in return for "the magic 51%" as Datacard founder Willis Drake puts it. The new company was Engineering Research Associates (ERA). With Navy contracts in hand and space in the old Minneapolis glider factory, the computer industry began.

Other companies entered the new field. Since the only computer people around were at ERA, the new companies hired ERA experts. Eventually one of them, Remington Rand, bought ERA. Norris and his associates went to Remington Rand, which through acquisitions rapidly became the premier computer company. The company then merged with Sperry Gyroscope Co. to become Sperry Rand Corp. This company centered in Minneapolis its computer operations, which included Univac, itself an outgrowth of an earlier Pennsylvania acquisition that had built the first electronic computer, ENIAC.

Then occurred one of the really extraordinary seminal events in the area's development. Norris and some associates set up a new company, Control Data Corporation, to build very large and fast computers. They financed the new company with dollar stock that was almost literally sold door-to-door, thus giving many Minnesotans a stake in the new firm, and the firm an abiding commitment to the region.

The group soon produced the world's largest and fastest computer, the CDC6600, which staggered the industry and could ironically be responsible for IBM's leading position today. The giant could not allow this Midwest upstart to upstage it and promptly revamped its own design, manufacturing and marketing practices to assure its continued supremacy.

Control Data has wider interests than computer technology. The company and Norris have a continuing interest in the development of small companies in the area, in the advancement of developing countries, in the use of technology for technology transfer and for education. "Education has used technology far less than any other area," says Norris. "It's tragic. Not only is there a great need but a great business opportunity. Eventually, it's going to be the largest segment of our business."

William Norris has had a profound impact on Minnesota in two ways. His Engineering Research Associates and its successor companies have spawned dozens of new technology companies in competing or related areas. Since 1945, most of these have been set up by former associates or employees of ERA, Sperry Rand, Univac, Control Data and generations beyond. A recent count showed some forty companies that can trace their lineage directly to ERA. Seventeen alone were born out of Control Data, itself one of the forty.

As in any large family, some members are deceased, some are married out of the family and some have moved away. But the majority remained in the Minnesota region and prospered, in time to produce their own progeny.

Among the ERA progeny are Data 100, now part of Northern Telecom, Datacard, and Cray Research, the latter two founded and still headed by close associates of William Norris, Willis Drake and Seymour Cray.

There are not many companies having only 17 sales in ten years that one would consider successful. One that is is Cray Reseach, Inc. In a sense, Cray is the keeper of the faith since the new company was set up exclusively to do what Control Data set out to do but somewhere along the way got sidetracked—build super-big, super-fast computers.

Established in 1972, Cray had spent $8.5 million by 1976 when the company sold the first CRAY I, for about $8.8 million. And it has gone on selling them since wherever immense numbers of fast calculations are needed in a short time to model extremely complex systems.

For all its speed and cost, the CRAY computer is simple. Since the speed of light limits the ultimate speed of an electronic device, connections must be as short as possible. Component packing is dense and heat generation a problem. "The most innovative technology in the machine," says President John Rollwagen, "is the liquid Freon cooling system that took three years to develop. There were two basic problems. One was the basic architecture—how should the components be related to one another for cooling and electronics—and the other an engineering one—how to squeeze a ton of copper into the space allotted." Cray solved the problems with the system's unique cylindrical configuration.

Though Norris and associates spun out of Sperry Univac, the moves in no way stunted that firm. The company never faltered as a formidable and growing factor in the computer industry. More so than other computer companies in the area, Sperry Univac is heavily into defense systems, its Defense Systems Division producing military components and a line of standard computers for the Navy.

"We are right on the cutting edge of the technology here," says Vice President and General Manager of Sperry Univac Semiconductor Division Robert Erickson, whose Division, indeed, may be the shape of the area's future. As computers pack more power into less space, semiconductor technology becomes paramount. "As we move past large-scale integrated systems (LSI) into Very Large Scale Integrated Systems (VLSI)," remarks Dr. Erickson, "the design philosophy becomes very sophisticated and needs highly specialized technical support. More systems houses, such as ours, will have to become involved in chip design—we won't be able to just buy them off the shelf and put them in our products," he forecasts. "And for that kind of work, the people available are very, very scarce. So," he says, "we are creating a new breed of technologist who understands the entire spectrum from design through manufacture to application."

"We have an engineering prototype line now and a medium level production operation," says Erickson. Honey-

well, Control Data and IBM at Rochester also have captive semiconductor production facilities. Sperry Univac is expanding its own semiconductor operations in the Twin Cities area. As to that industry in the region: "This is one of the areas for growth," says Erickson.

Erickson sees the computer companies integrating backward to chip manufacture, supported by the Institute of Technology's new Microelectronics Center. The company also strongly supports through the Technical Advisory Committee the beefed up advanced technology programs at the Institute of Technology. The Committee advises the University on future use of resources to advance technology in the region.

Medical Technology

The region abounds with companies that apply the latest in electronic and biological technology to the medical field. It is a zone of technology that has grown in parallel with the computer and electronics technology and has generated its own corporate genealogies. Firms range from large and relatively old ones like 30-year old Medtronic Inc. through newer companies like Cardiac Pacemakers and Med General to the newest, such as Molecular Genetics, founded last year.

There are fifteen firms listed including 3M under X-ray and electromedical apparatus alone, with many more under related categories. Honeywell, for example, is not a medical technology firm yet it has significant operations in the field. It designs and manufactures a range of complex equipment from pacemakers to monitors.

Medtronic, with 4,700 total employees and 2,300 in the Minneapolis area, is the largest. The company grew out of a medical instrument repair service in a garage which, on request, built a transistorized battery-powered device for heart stimulation. The cardiac pacemaker was born.

The products have come a long way from the initial external device. Laser welding, microelectronics and microprocessors, improved batteries—"It requires the merger of a number of technologies," says George Heenan, Medtronic's Senior Vice President for Research and Ventures. "We have to integrate medical, biological, chemical, electrochemical and polymer technologies with manufacturing procedures, laboratory work and superb quality control."

Medtronic cooperates closely with other companies in the area, both in its operations and in the community. "We use PLATO (Control Data's educational data base), work with Honeywell, the medical schools and the hospitals. We are working at the leading edge of the practice of medicine so we must work closely with the medical profession."

The company has expanded beyond pacemakers, though those remain the primary business. New technologies include an electronic device to alleviate pain, deep brain stimulation, artificial heart valves and the like. "If the bionic man comes, he'll come out of Minnesota," says Mr. Heenan.

More than 10% of the company's staff is in research and development. Among these are many different types of technologists specializing in materials and polymers as well as bio-engineering. "Bio-engineers are a very important group," says Mr. Heenan.

Of more recent vintage and different emphasis is Molecular Genetics Inc., a small company founded by a physician and a professor of microbiology. They are Dr. Franklin Pass, the physician, and Prof. Anthony Faras who saw in the new technology of gene transplants and recombinant DNA a way to benefit the local agricultural industry and, eventually, the medical field.

Directly and indirectly, the University of Minnesota and particularly its Institute of Technology have been central to the area's technological sophistication. Several companies have come directly from the University.

Among these are Rosemount Inc., a $100-million instrument company started by three employees of the Aero Engineering Department founded in 1956; TSI Inc., a $5-million maker of specialized particle research and mass flow instruments founded on a doctoral thesis in the Department of Mechanical Engineering; MTS Systems Corp., successor to a company also founded out of the Aeronautial Labs in 1966, now a $45-million maker of electronic instrumentation and software; and Physical Electronics, Inc., now part of Perkin-Elmer Corp., whose first products were made of glassware in the University's electrical engineering department.

Physical Electronics founder, Dr. Roland Weber, was part of a group at the University working on surface physics. They came up with a way to determine the elemental composition of a surface by bombarding it with electrons. Dr. Weber with his students built several of these devices. Outsiders saw them and wanted them, so some were built for barter. "I decided this was not a proper use of students or university facilities," Dr. Weber says, "so in 1969 we started a company."

The embryonic company had rough going. The easy financing days William Norris found ten years before were gone. To get professional management, Weber sold his company in the first year to a local mini-conglomerate. "They promptly went belly-up," he says. Weber then raised $200,000 from local venture capitalists in return for 40% equity and got the business going again. It is now a $30 million division of Perkin Elmer employing 330 people. "It's a low volume, high cost business," says Dr. Weber, who has traded the title of founder and chairman for Division General Manager of the Surface Sciences Division of Perkin-Elmer. "But it's more than that. There's lots of art in the products.

"In most of science, it's the art in putting things together that makes things work," says Dr. Weber.

Help for Small Firms.

Norris's willingness to help a potential competitor, Cray Research, among others, get going is another aspect of Minnesota that shows the commitment of local business to the general health of the area. Norris in particular has been a moving spirit behind a number of recent organizations intended to advance technology and help smaller businesses. These include the Microelectronics and Information Center at the University of Minnesota's Institute of Technology, the Minnesota Cooperation Office, which gives management help to budding entrepreneurs, the Minnesota Seed Capital Corp., which provides seed money somewhere between an entrepreneur's own resources and the stage where he can interest a venture capitalist in his idea.

President Rollwagen of Cray Research doesn't think the small company could have succeeded anywhere as well as it did here. The willing help of other companies, potential rivals, was one factor. But so was community financial support when the company went public. "There is an infrastructure of people used to dealing with small, risky companies," he says. "In an area not familiar with the technology, we could not have got the money."

"I have Control Data to thank for this company," says Dean Scheff, President and founder of CPT Corp., a Twin Cities firm that has nothing to do with either Control Data or the University of Minnesota. "We went public in 1972 before we even had a machine, at $3 a share," he says, "and sold all the stock locally. Technical things are easier to sell in this

The Cray-1 Computer System shown here at the Cray Research, Inc. Manufacturing Facility .

town, primarily because of Control Data's success."

Mr. Scheff is neither engineer nor scientist. The native Minnesotan is a former salesman of office machines and personnelman from Univac who decided he could make a better product than he was selling. With an associate who resigned from Honeywell rather than be transferred, he built an electronic automatic typewriter in his basement.

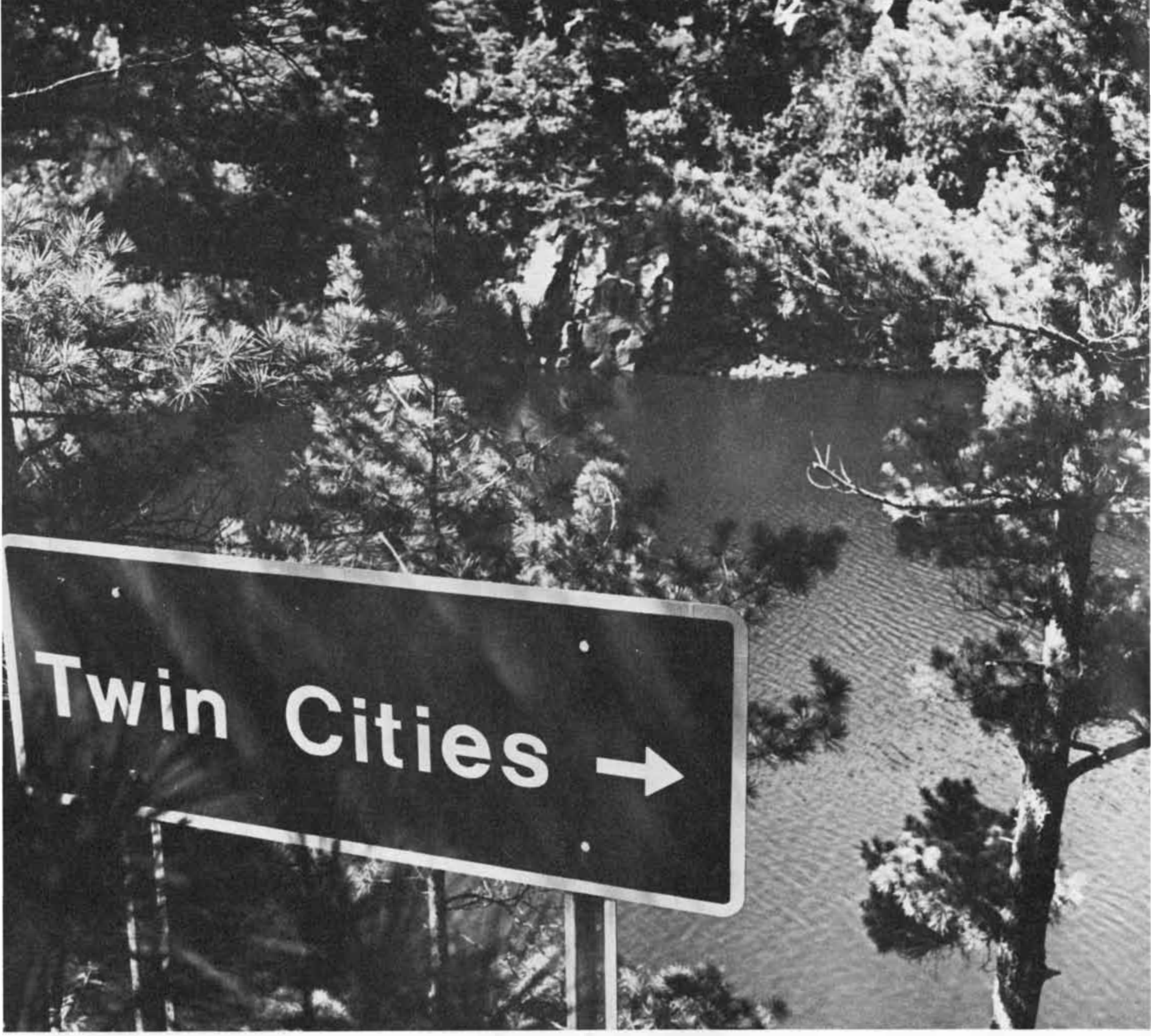
"The microprocessor, the floppy disk and inexpensive memories have made the word processor business," he says with thanks to the scientists and engineers. "And of those, the microprocessor was the biggest step." Scheff agrees with many others that the future in computer applications will be less hardware and more software. "We have as many people on the software side now," he says, "as on hardware, and it's getting more so. Good programmers and engineers are never easy to find."

Minnesota is a leader among advanced technology centers in this country. It means to hang onto that lead. And certainly, the intellectual, entrepreneurial, technical and financial base on which the technology sector stands is large and vigorous enough to generate a still larger sector. However, the growing sector needs ever larger numbers of engineers and scientists, which the state's relatively small population and suitably-scaled educational system cannot produce.

"The state's needs for engineers far exceeds the supply," says Mr. Charles Denny, president of Magnetic Controls, Inc., a supplier to the communications industry that was itself a spinoff from Honeywell. "The deficit, of course," Mr. Denny continues, "is accomodated by recruiting out of state." Central to the recruiting efforts is the University of Minnesota's Institute of Technology, a major supplier of engineers and scientists as well as a magnet for them. To improve the pulling power of the magnet, the local technology sector has established the Institute of Technology Advisory Council, with Mr. Denny as current president. The council, comprised of presidents and senior R&D vice presidents from some forty-five local technology companies, advises Dean Staehle on curriculum and, perhaps more importantly, intercedes with the community and the legislature on behalf of the Institute. In a region where agriculture is still king, technology has not had a strong enough voice despite its stature. The Council and similar efforts will remedy the lack.

Newcomers will find a pleasant place. The opportunities to work at the leading edges of electronic, computer, medical and other technologies are boundless and the need for people of a curious bent great. The state's near-wilderness areas beckon outdoors types while the less adventurous can enjoy the water closer to home on the many lakes within the Twin Cities. The climate is a point of pride. Says Charles Denny, "I find more to do in winter than summer and many bemoan the disappearance of the snow in springtime." The Cities themselves are true metropolises that work well as places to live and do business. Indeed, it is said that Minneapolis will succeed Chicago as "the city that works."

But perhaps one employee of a high technology company, a former Peace Corps worker, native of another state and former resident of still others, put it best. "It's the greatest place to raise a family."



Frontierland.

Minnesota. It's high-tech country. A country that has, in the past, provided fertile ground for rapid growth of such top corporations as Control Data, Honeywell, 3M, Sperry-Univac, plus scores of other related high-tech companies.

ADC is a high-growth company serving the communication and computer markets. Our product portfolio is diverse: precision telecommunication transformers and connectors, power supplies, and voice and data test systems. Our customers are the leaders of their industry and our products are utilized throughout the countries of the free world.

We need imaginative professionals in engineering, marketing and manufacturing to help us explore the frontiers of these fast-growing communications and computer industries. Write us for details. We're ADC/Division of Magnetic Controls Company, 4900 West 78th Street, Minneapolis, Minnesota 55435.

MAGNETIC CONTROLS CO.
An Equal Opportunity Employer

BOOKS

The origins of human cancer, sign language, stone rings and the world of small particles

by Philip Morrison

ORIGINS OF HUMAN CANCER, BOOK A: INCIDENCE OF CANCER IN HUMANS. BOOK B: MECHANISMS OF CARCINOGENESIS. BOOK C: HUMAN RISK ASSESSMENT. Edited by H. H. Hiatt, J. D. Watson and J. A. Winsten. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. ($45). CANCER: SCIENCE AND SOCIETY, by John Cairns. W. H. Freeman and Company ($8.25). In the year 1850, among those English people who reached the age of 10 by escaping the mortality then widespread among infants and children the majority would die before the age of 60. In 1972 not only was the loss of young children very much lower but also four-fifths of those who reached 10 survived to reach 60. One consequence of this demographic shift in the developed countries is that now about one person in six (in the U.S.) dies of one of the more than 200 kinds of cancer the pathologists recognize. Even if one reckons only the loss to the aggregate working life span, according to U.S. data for 1968 cancer still takes about a seventh of it. Ordinary accidents, together with suicide and homicide, then become the leading cause of work-life loss; vascular disease, the leading agency of death, like cancer occurs preferentially in the later decades.

Cancer incidence is worldwide, with variations among most populations by only a factor of two or so around the mean, some three cancers per year per 1,000 people between the ages of 35 and 64. The wide variety of cancers, however, implies multiple causes. "In the fact that . . . four or five cancer types produce half the total . . . burden" one can see hope for major relief from a few successful measures of control. In the diversity of these specific major burdens around the world "we learn that no single type of cancer is a major risk for all populations, and therefore none is irrevocably part of being human." Finally, because there is clear evidence that cancer risks can show major shifts when an emigrant group changes culture and environment one can be sure that most risk is not inborn. "Since most cancer risk is environmental and cultural, it follows that it must be avoidable."

In that special sense, then, cancer is a disease of human society. As a biological entity it is known as a disorder of development, widespread among all the organisms (including plants) whose life is an ecology of distinct classes of cells. Indeed, it is known that cancer is no general rebellion but a mutinous clone, with the cellular disobedience passed along strictly within one cell lineage by inheritance from the one ancestral mutineer. The direct study of biochemical and chromosomal genetic markers has made it sure in a couple of very different forms of cancer; it is hard to avoid the generalization.

Given that insight, the new importance of cancer for the current age pyramid, the present mastery of molecular genetics and the precipitate climb in the industrial production of novel compounds, many of which are known mutagens, there was bound to be widespread interest in the causes and nature of the disease among molecular biologists. These two books, widely cited in current papers, are not brand new; the work they describe is three or four years old. Together, however, they offer the long and the short of two decades: what is both known and surmised today. *Origins of Human Cancer,* of whose three volumes the second is mainly devoted to detailed studies of the mechanism of carcinogenesis down to the DNA level, is the record of a week's symposium in 1976, with some 120 research papers and reviews by its multinational participants. *Cancer: Science and Society* is a brief book aimed at the general reader and the medical student. The two works interlock: Dr. Cairns, now director of the Mill Hill Laboratory of the Imperial Cancer Research Fund, was before that director of the American fundamental-biology research station at Cold Spring Harbor, where the three-volume report originated. Dr. Watson, I presume, shares that interest. The coauthor of *Origins of Human Cancer* is the J. D. Watson of the double helix and current director of Cold Spring Harbor, following Cairns.

The symposium volume is both much more detailed and more limited in scope than Cairns's graceful overall introduction. The symposium concentrated on origins, although it offers Volume A on the general results and issues of cancer epidemiology; the third volume is mainly about the means for seeking out carcinogenic materials, with an interesting 100 pages on the issues of the regulation of foods and drugs, including lively discussion from the floor. You will not find there much about the nature or growth of the tumor itself. Cairns does include enough of the story of the development of tissues (skin is his fullest example) to help a reader understand what the runaway clone is like.

Just about any cell can transform into the head of a cancerous lineage. Actually, however, about nine-tenths of all cancers arise in the cells of the epithelia, the layered sheets of tissue that form the external surface of the body or that line the tubes and cavities forming the internal surfaces. This fact is understandable; epithelial cells more than others must come in contact with the substances of the outside world, say food, tens of tons in a lifetime. Of course, it is not exposure alone that matters. It is also something like growth, the number of cell divisions. Nerve cells rarely divide in adults, and they are host to tumors almost only in children. Other nonepithelial cells, within the bone marrow, divide constantly, to furnish short-lived specialized cells to the blood, and fall before dangerous circulating malignancies of acute cell overgrowth. The epithelial cells, just because they do cover the interfaces, are normally and continually damaged, shed and resupplied new-grown from layers within; these exposed and growing cells thus bear the greatest risk.

It is plain that one single outlaw cell cannot wreck the entire polity of the body. Not even a pea-size growth is very likely to cause fatal damage. Such a tumor, however, has already counted 25 generations from its outlaw founder. It is the next few divisions that make it into an invasive cancer. Many steps and much delay must intervene, and indeed the time course of cancer incidence shows a slow rise over 10 or 20 years followed by a steep increase. On the log-log plot of the death rate v. age from cancer of the large intestine, for example, a straight line fits the data quite well from 25 years up to about 80. The slope is nearly five, just what a simple model of six independent, slow, improbable steps, all occurring to transform a single cell, would predict.

Experience does show that two distinct insults greatly increase risk: both uranium miners and asbestos workers who smoke heavily are at a risk 10 or more times greater than that of their fellows in the same dangerous trade who have never smoked. What those steps

How we're putting gold into connectors you can afford.

A lot of companies that depend on electrical connectors are trying to reduce their usage of gold.

They can with AMP.

The reason? We've designed connectors with less gold that perform as well as ever.

One beauty of gold is its unique reliability in electrical contact areas. But the predictability of its price is something else. That's why we developed our own method of selective plating to put the gold only where it's needed. We call it ACCU-PLATE. It's precise. Its advantages work with other metals, as well. And it's saving our customers money.

Providing reliable platings that are cost-effective isn't new to AMP. Over the past decade, we've reduced the gold in our connectors by 33%.

What can we do for you in the '80s?

Whether you call us early with a problem, or simply call us for connectors, you'll get what the world has come to expect from AMP. Connectors and interconnection systems designed for performance. And as the industry leader in research and development, we'll continue to bring you products that reduce the cost of making yours.

AMP has a better way.

For product innovations, for new applications, for ways to save you money, AMP has better ways to help make your company more productive.

Write for our brochure, "AMP Has A Better Way." AMP Incorporated, Harrisburg, PA 17105.

Barcelona • Brussels • Buenos Aires • Frankfurt • Harrisburg • Helsinki • s-Hertogenbosch • London • Luzern
Mexico City • Montreal • Paris • San Juan • Sao Paulo • Stockholm • Sydney • Turin • Toronto • Tokyo

AMP means productivity™


Cadillac announces V8-6-4 Fuel Injection
As you drive, the 1981 Cadillac automatically goes from 8 to 6 to 4 cylinders.
GM
Cadillac

Is V8-6-4 Fuel Injection standard equipment?

Yes, the V8-6-4 fuel-injected engine is the standard gasoline engine for all 1981 Cadillacs.

How does it operate?

As you leave your driveway, all 8 cylinders in your gasoline-powered 1981 Cadillac are in operation. Then, as you reach intermediate speeds on a street or avenue and your power requirements lessen, the car automatically switches to 6 cylinders. And then, when you reach cruising speeds and your power needs decrease further, the car automatically switches to 4-cylinder operation.

Does any other carmaker offer anything like this?

To our knowledge no other carmaker offers it — foreign or domestic. It serves as a prime example of American innovation.

It sounds complicated—is it?

No. It's surprisingly simple. The on-board digital computer of the Digital Fuel Injection system senses the power requirements of the engine and closes off two, four or no cylinder valves, depending on those power needs. Then, as your power needs change, it opens fuel valves.

Can you feel the car going from 8 to 6 to 4 cylinders and back again?

Some people can. Some can't. The perceived sensation is slight. Because, in a sense there is no shifting. Fuel valves simply close or open as instructed by the computer.

The idea of 4 cylinders in a Cadillac bothers me—should it?

No. The system only goes to 4 cylinders when your power needs are relatively low. To pass another car, push down on the accelerator and the system instantly goes into 8 cylinders for added power. Then, as your power needs decrease again, the system will return to 4 cylinders.

How reliable is it?

This system has been proven in over a half-million miles of testing. It's *that* reliable. All electronic components are solid-state, including the digital computer itself.

Is this the same as overdrive?

No. Overdrive is a function of gears. V8-6-4 is a function of the number of cylinders receiving fuel. By way of interest, overdrive is a feature associated with the V6 engine available on 1981 Fleetwood Broughams and DeVilles.

Cadillacs are equipped with GM-built engines produced by various divisions. See your Cadillac dealer for details.

Can I tell how many cylinders are active at any given time?

Yes. Push a button and Cadillac's MPG Sentinel on the instrument panel shows a digital display of the number of cylinders active at that moment. The MPG Sentinel will also show instantaneous mpg and average mpg.

Is it true that this combination could help a person to become a more efficient driver?

Absolutely. Knowing your active cylinders and instantaneous mpg can help you adjust your driving habits accordingly. And thereby help you become a more efficient driver. You can further demonstrate it to yourself by resetting the MPG Sentinel before a trip. Then push a button at trip's end and the MPG Sentinel will tell you how well you did by displaying your average miles per gallon to the nearest tenth.

All this is standard in 1981?

All this and more. Buy or lease, see your Cadillac dealer for a test drive.

Trust Cadillac to lead the way.


JOHN JAMESON & SON
ESTD 1780
SINE METU
IRISH WHISKEY
A BLEND
DISTILLED IN IRELAND BY
John Jameson & Son
LIMITED
BOW STREET DISTILLERY
DUBLIN
80 PROOF
IMPORTED

are is not known. That they are general, rather than specific to certain types, is not expected, and even whether such a simple theory is valid—one can find other formal models that are equally plausible—is also not known. The notion that the first step to all cancers is an induced mutation is at least hopeful, even though it has led to somewhat unreasonable public concern. It implies one might do away with all external mutagens, particularly since there are now quick tests for carcinogens through counting mutant bacteria, reared in the presence of enzymes from mouse liver, which more or less realistically activate any premutagens in the sample.

Certainly it is sound public health to reduce the intake of such new and strange carcinogenic materials. Although there is no clear sign of disaster, there is some ground for the view that a rise of cancer incidence is impending, a surprise increase in carcinogenesis like the one from the widespread smoking of cigarettes, which pushed lung cancer to the top of the list of fatal cancers. The 20-year lag in cases that comes from the multistage form of time curve is the logical link; most of the rise in organic chemical manufacture comes after World War II. Yet as Cairns remarks: "There is little evidence that the chemical industry causes much of the current total cancer incidence.... With the exception of lung cancer, none of the common cancers are much commoner now than they were 50 years ago.... In fact, it seems more likely that the main determiner of cancer is diet rather than industry, and that we should be looking for mutagens (or perhaps promoters) that are formed in the body from the normal ingredients in our diet."

The remarkable ability growing cancers have to attract to themselves a mat of normal blood vessels that then serve to nourish the invading tumor is a decisive property of an invasive cancer. Could this factor somehow be externally inhibited? It seems clear that vitamin A and its analogues can prevent certain induced epithelial cancers in animals. These substances are essential for the complex steps of the normal replacement of epithelial cells as the outermost layers are shed; there are natural repair mechanisms that must serve to inhibit the outlaw cells. Initial genetics is not everything; development is at the heart of the matter, and it is not understood.

Meanwhile we know more than we are able to do, albeit less, much less, than we need. As Cairns writes, "we learn how we could easily prevent the most lethal of all cancers ... and people continue to smoke." Indeed, governments around the globe urge the growing of tobacco. Stomach cancer has dropped by a factor of five over five decades in the U.S., and it is now falling faster in Japan. It is not known why; in Japan there is evidence that a westernized diet (at least the use of that very non-Japanese beverage, milk) protects.

Dr. Cairns has given us a quite readable, although by no means simple, overview of the problem. The symposium is nowhere near as accessible; its biochemical papers are hard. Yet in many of the papers, and in a few fine reviews, facts and insights stand out even more clearly than they do in Cairns's more guarded and often less specific text.

The initiation of cell change by ionizing radiation and by ultraviolet finds a rational place in the picture. It is easy to infer genetic change, and to observe the close connection with cells in frequent division. But the effect of the fibers of asbestos minerals, empirically vivid and confirmed in vitro, is not easy to grasp. Heating the fibers has been reported to reduce their carcinogenic power; is some surface organic contaminant the real agent? Or is it some fragile property of the fiber surface?

In the Cold Spring Harbor volumes one striking report is given, sketchily, both by Cairns and by his Oxford colleague Richard Peto. The genetic-marker technique has shown that each of the sclerotic plaques that form on aging arterial walls is a clone of some single smooth-muscle cell within the wall. These self-limited cell colonies, all kin, sometimes grow enough to clog a vital flow of blood or to initiate a blocking blood clot. It is this mechanism that lies at the root of the fatal vascular disorders, today's chief cause of mortality. The entire kingdom of the body is thus brought down by a few malcontent cell families residing along an important artery. Some great charter is violated.

There is a page of candid photographs as a frontispiece in each of the three big volumes. No conference shots can do much better than to show, as Volume A does, Sir Richard Doll and Lady Doll. It was this British epidemiologist who first demonstrated in the 1950's both the heavy risks of inhaling cigarette smoke and the serious hazards faced by workers handling asbestos. If anyone should seek an individual who has put rational analysis to the widest service of human welfare, here is a redoubtable candidate.

THE SIGNS OF LANGUAGE, by Edward S. Klima and Ursula Bellugi, with 10 collaborators. Harvard University Press ($25). Every human utterance in natural language strings beads of meaning along a thread of time. That one-dimensional phonemic stream finds more or less direct mapping in the alphabetic or syllabic signs of most written languages. The single dimension rules; its requirements constrain the nature of words and the grammar of language. But anyone who has looked at Chinese characters, the most widely used of ideographic and pictographic scripts, knows that a different architecture is present on the two-dimensional page. Only one character in 20 or so remains pictographic, iconic; on that concrete base has been built a complex visual system all but alien to sound. Hence at a naive level if a tree can be caricatured in three quick strokes, a pair of such trees can mean a forest and three of them come to mean shade. In a more generalizable mode the radical for "water" added to "branches" yields "froth," and added to "every" it names "the sea."

This volume presents a record of a seven-year investigation by experiment and analysis into the nature of American Sign Language, that formal system of hand signs that has become over nearly two centuries, now among half a million people, a full natural language in silence. Its symbols flash smoothly in time within the spatial volume around the head and shoulders of its speakers; they are frequently to be seen these days in inset views on the television screen as the sound string of speech holds the attention of most of us. The authors, a Salk Institute for Biological Studies research pair (and their shifting colleagues), had intended only to study language learning in children, using ASL as a variant mode. They found themselves thrust into a new linguistic world, an autonomous gestural-visual system, "complexly structured with a highly articulated grammar," obeying the fundamental laws of human languages but molded by that embedding three-space.

First of all it works, and it works well. The nimble tongue is certainly quicker than the eye-hand combination. We all know how cheap talk is. No known community of hearing people has ever held a sign language primary to their spoken one. Careful experiment has shown, in fact, that the rate of sign production in ASL is roughly only half the rate of articulation of words. (The measurements were made on a few fluent bilingual speakers.) The rate of forming propositions, however, defined as some equivalent to the simple underlying sentences of a text, shows close similarity across the two modes, not only for the bilinguals but also on videotape measurements of a number of deaf native signers. There is a premium on brevity in ASL. Any loss of information in the ASL "texts" was controlled against by requiring retranslations of the videotape record. Actually nothing substantial is lost, even though a gloss of the signs used seems almost cryptic, supertelegraphic—like Chinese poetry. "The man who was washing the car became angry because they jumped all over it" becomes just five signs: MAN WASH, ANGRY, WHY, JUMP. But in the telling something was added. In the little test story in words (whose last sentence is the one just cited) the word "car" occurred twice, but in ASL it was indicated only once to begin the story, and a quick classifier sign for "vehicle" was pointed to a particular spot to the right of the signer. From then on that spot in space was the car. "The memory lingers," as if the sign still dwelt in that small volume.

The signs can also be modulated by their flow. RED is not merely "red"; repeated with tenser, sharper movements it becomes progressively stronger. A sign can be held, and the meaning too acquires duration in time. One sign, SICK, is drawn in line here to illustrate no fewer than eight of its modulations. "To be sick" is a hand to the forehead: simple enough. Variations described as tremolo make the illness chronic; a long hold plus a quick single move reports "very sick indeed," and the rest are no less clear and marked.

By placing signs before untrained viewers for their interpretation it was easy to show that most of the ASL signs chosen were more "transparent" iconically than spoken words. Still, they are only caricatures; the iconicity is itself symbolic. Thus in ASL a tree is a spread hand held upright, the finger branches twisting in the wind. In Danish Sign Language the tree is signed by a two-handed gesture moving out and down around the swelling branches, to slide down smoothly, close-circled around the narrow trunk. In Chinese Sign Language two hands move upward along the trunk. The signs are all transparent, but they represent distinct conventions. The ASL sign for CANADA—the right hand quickly grasps the clothing near the shoulder—is opaque. And yet might it not refer to the heavier clothing expected of our northern neighbors? The grammatical processing of the full language depends, however, very little on such visual maps, as a Chinese character goes far beyond a pictograph.

There are well-defined ways of enriching the vocabulary. Compounds are of course one chief path. They are of many kinds; to generalize, a list of key examples is formed and linked. Linked, CAT DOG BIRD say "pet," with an optional sign following to signal the idea "category." A clear gesture often specifies size and shape. For example, the fingers and the thumb outline a rectangle in space. That invariant sign we can gloss as RECTANGULAR. Now we link two signs to generate a set of terms: RED RECTANGULAR is a brick, SIGNATURE RECTANGULAR is a credit card, and so it goes.

Finally, language means wit, and it probably must also mean poetry. Without these there is no natural language; Fortran and Lisp have a way to go. ASL is not lacking. A new tendency in schools for the deaf is the use of "total communication," which includes both signing and speech in the classroom.

"The most complete and most scholarly dictionary of the English language"

—The Christian Science Monitor

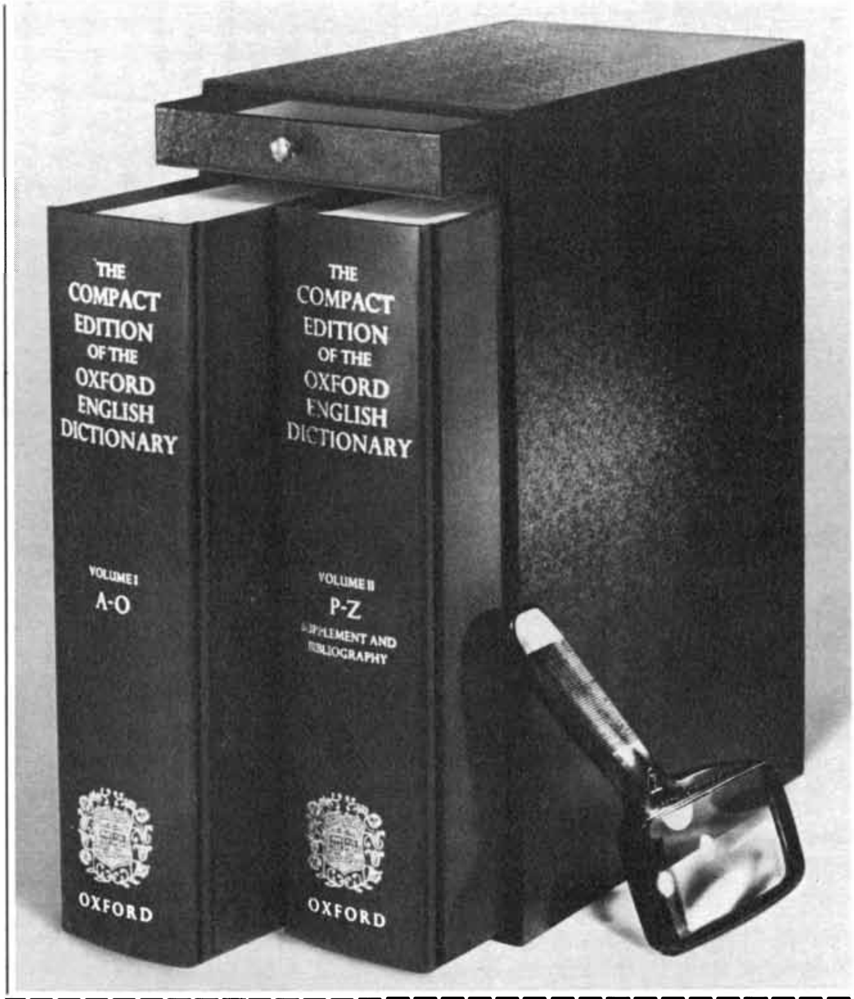
The Compact Edition of the Oxford English Dictionary.
Yours as an introduction to membership in the BOOK-OF-THE-MONTH CLUB®

YOURS FOR ONLY

$19.95

(PUBLISHER'S LIST PRICE: **$125**)

You simply agree to buy four books within the next 2 years.



FOR the price of an average dictionary, you can now treat yourself and your family to the world's best dictionary of the English language. The contents of this two-volume edition are identical to those of the original thirteen-volume set, priced at $495.

FEATURES

- Boxed set of two volumes, 9¾" by 13½" each.
- All 16,569 pages of the 13-volume original included in 4134 pages of *The Compact Edition* through a photo-reduction process which permits printing of four pages of original on one page of compact edition.
- Paper is 30-pound Special Dictionary White.
- Binding is library buckram reinforced and gold-stamped.
- Bausch & Lomb magnifying glass included in special drawer of slipcase. 2" by 3⅞" lens scientifically designed to make reduced print easily readable.

Book critics call the OED a "miracle." Join the Book-of-the-Month Club and this "miracle" can be yours now. As a Club member, you'll continue to enjoy benefits on the best and most important books published today.

Bookstore Quality at Book-Club Savings
You conveniently shop at home at considerable savings. Whether you're adding up your remarkable savings on the introductory offer, or on books you are offered as a member, these are always true savings...because every Club book is as good as, or better than, those sold in stores. You *don't* settle for the altered or inferior editions that some book clubs send their members.

Book-Dividends. When you remain a Club member after the trial period, every book you buy earns Book-Dividend® credits. These entitle you to choose from a wide variety of significant books at hard-to-believe savings of ***at least 70%***.

Book-of-the-Month Club, Inc.
Camp Hill, Pennsylvania 17012 A170-10

Please enroll me as a member of Book-of-the-Month Club and send me THE COMPACT EDITION OF THE OXFORD ENGLISH DICTIONARY. Bill me $19.95 (in Canada $22.50), plus shipping and handling charges. I agree to buy 4 books during the next two years. A shipping and handling charge is added to each shipment.

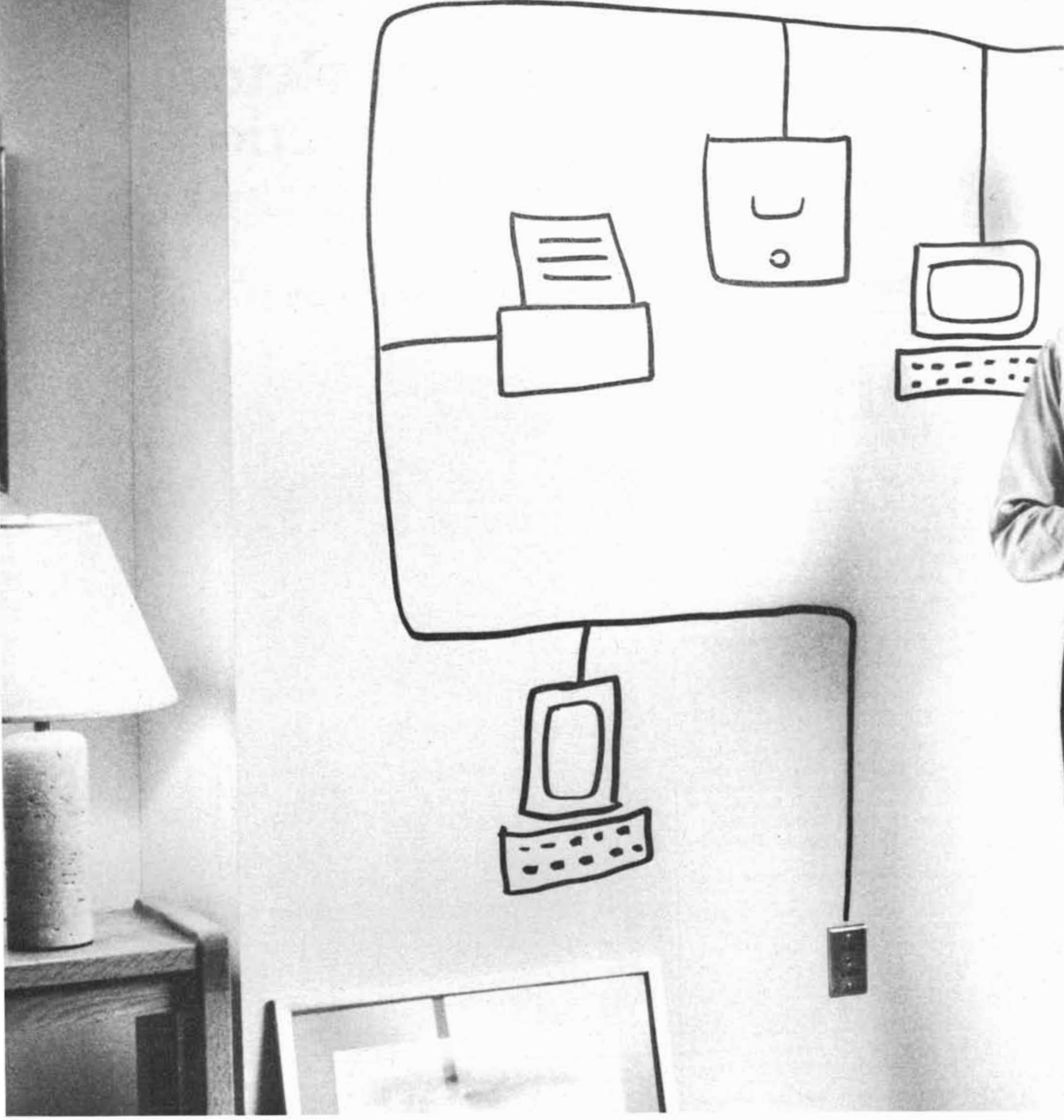
Mr.
Mrs. 0-41
Miss. (Please print plainly)

Address .... Apt. ....

City ....

State .... Zip ....

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.



Xerox introduces the

If you're wondering how business will handle information in the '80s, the handwriting is clearly on the wall.

We call it the Information Outlet — a new way for you to custom design an information management system that will give you maximum flexibility with minimum expense.

Here's how it works:

The Information Outlet gives you access to a special Xerox Ethernet cable that can link a variety of office machines. Including information processors like the Xerox 860, various electronic printers and files, and, of course, computers.

The Xerox Ethernet network will enable people throughout your company to create,

If you'd like more information on the Information Outlet, write us and we'll

Information Outlet.

store, retrieve, print and send information to other people in other places—instantaneously.

This network wasn't designed to work exclusively with our equipment. Other companies' products can be connected as well.

As your needs change, so can your network. You'll simply plug in new machines as you need them—or as technology develops better ones.

So, through the Xerox Information Outlet, you'll get to the future the way the future itself will get here.

One step at a time.

XEROX

send you a booklet: Xerox Corporation, P.O. Box 470065, Dallas, Texas 75247.

The conventional sign for this new teaching style is an unusual two-handed combination of letter forms, the letters for *t* and *c* made as a single sign, with the hands moving alternately toward the speaker and away from him. In the popular story the student was asked why everyone in the school seemed so carefree and happy. He smiled and signed TOTAL COMMUNICATION, but "his hands moved gradually closer and closer to the mouth and his head began to tilt from side to side, until the single sign had become transformed into the two highly iconic signs DRINKING (alcohol) and SMOKING (marijuana) made alternately.... The play on meaning is multileveled." It is less useful to try a summary of the poetry of Dorothy Miles, whose ASL work *Four Haiku Poems* is studied in detail here, with many line drawings and diagrams. Two performances by different artists are examined. The conclusion does not seem unjust: in these very years ASL is growing a new form of heightened signing. It is a disciplined dance of the expressing hands, as song is a melodic excursion of the speaking voice, "the hands simultaneously creating both signs and designs in space."

This set of research papers skillfully turned into a most evocative book is no guide to the learning of the sign language itself, but it is one to understanding it as an independent flowering of the capacity we call language. That is a treasure richer than speech, sign or ideograph, the springs of both art and science. It is worth mention that for the many sequential drawings and diagrams needed to make the flow of sign understandable on the static page the authors have themselves evolved a small visual language of symbol: an entire set of spiraling, swelling and dwindling arrows bridge their scenes. The two hemispheres of the brain are plainly cooperating well in this study!

Experimenters, these authors offer only a few pages on the history of sign around the world after its first systematic recording during the French Enlightenment. A fuller history is what is now needed. Some of the pleasures it promises are hinted by one regional sign that seems to evoke sunny meadows in an earlier century: the signs RED SECRET signify the shy, fragrant strawberry.

RINGS OF STONE: THE PREHISTORIC STONE CIRCLES OF BRITAIN AND IRELAND, by Aubrey Burl, with photographs by Edward Piper. Ticknor & Fields, New Haven and New York ($19.95). From Land's End past John o'Groat's House to the Orkneys and across the sea to Ireland the scattered dots fill the map to mark more than 900 stone rings. These range from world-famous Stonehenge to obscure little Kealkil, a setting of five stones in an oval the size of a small room, not far from Bantry Bay. Fifty such sites are here described, arranged by region, each with a page or so of description, and a set of photographs evocative of the monuments on the quiet land. These are sites meant to be visited; they are plotted on highway maps, and the tenor of the entire volume is to encourage quiet and thoughtful attendance. The text itself, no less than the photographs, all made by a single well-traveled artist, takes the personal stance of a fascinated and knowing traveler.

This is not at all to say that the volume is the work of tyro or dilettante. Aubrey Burl of the Hull College of Higher Education is as scholarly an archaeologist as these ruins know. He has carried out detailed work in the field at many sites; one recent task was supervision of the total excavation of the recumbent ring of Berrybrae, north of Aberdeen on highway A92. (During three seasons "of over five hundred hours spent at the site we lost only two because of rain," redeeming the Scottish summer.) What his informal introduction of 90 pages does is to offer, instead of single-minded inquiry into the geometry of the sites, a rounded reflection of how the people of those times can have made and used the sites, as far as all the evidence we can gather can tell.

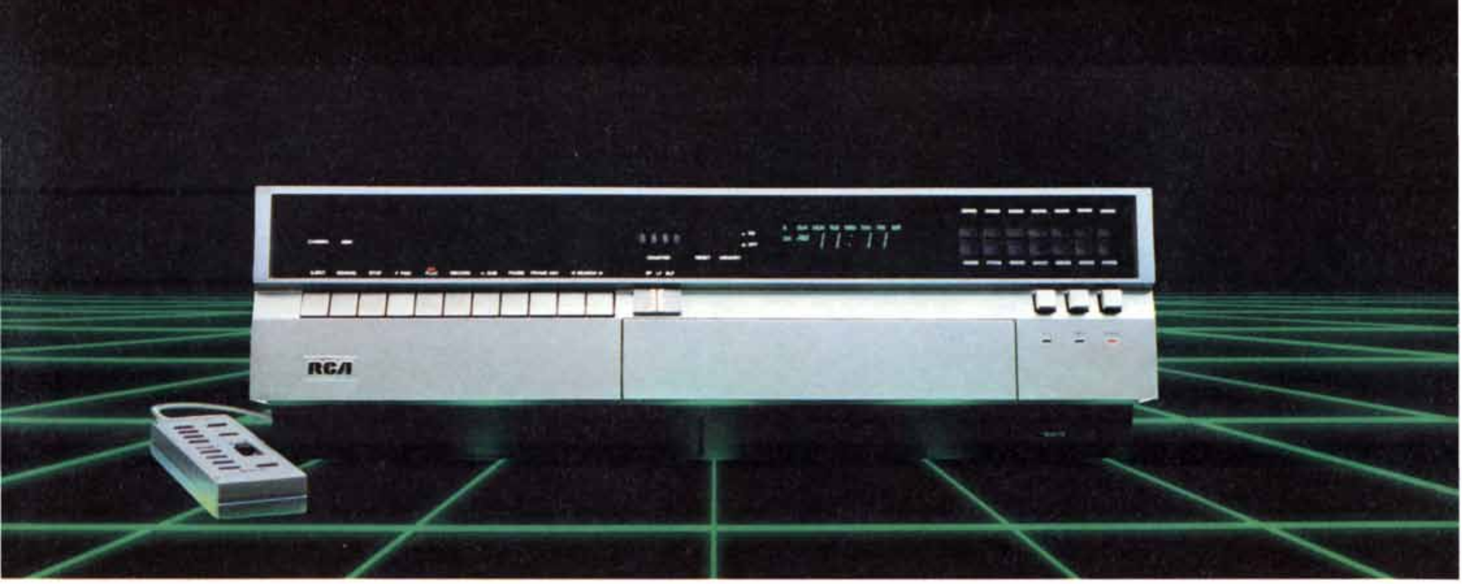
The dates are pretty reliable, and getting more so; the Scottish circles Burl studied are Early Bronze Age work, just short of 4,000 years old. The youngest circle known in the British Isles was built in about 1500 B.C., and the oldest, amid the carved passage graves of the Neolithic farmers in the green hills of Ireland along the River Boyne, go back more than 5,000 years.

The great stone rings were chieftains' works, communal in use and manufacture. The smaller ones are plainly the work of families, who had to gather their neighbors to haul together the one great recumbent stone that gave meaning to the ring. No single use is likely, as no single people made all the marks and all the settings. Conflict no less than unity is found in the works. Burl is direct in his criticism of the stronger claims so loud today. He dispatches the various lunacies of the esoteric, the seekers of ley lines, the dowsers, the electrifiers and the rest in a few witty and sensible pages. Some mysterymongers look at "occult Britain" to find antigravity liquids and priests who could raise huge stones by sound. "One would imagine from this that Callanish was composed of prodigious blocks, high as trees, thick as elephants. It is not." The heaviest stone there could have been raised by 14 adults "with patience and rawhide."

Burl tells us as well of the troubles of the more rational, the archaeoastronomers and the searchers for modular statistics. Here it is no simple rout, although some important lines of sight turn out to be blocked once you look in the field, and others refer to much-disturbed positions. The "delightful ring of Blakely Raise" is so close to the road that picnickers drive their cars into it. That ring is "exactly twenty Megalithic Yards in diameter" and a line can be taken across it right to the setting direction of the maximum moon. "This is strange. The stones of Blakely Raise are set in concrete ... put up by Dr. Quine in 1925." The measurements there are lucky coincidences.

That too is not the entire story. True north-south alignments are frequent; the recumbent stones Burl himself dug are wonderfully leveled, and they are well oriented to serve as ritual horizons for the midsummer full moon at the top of its arc. Burl found these directions himself, even in rings made by men and women in small groups, struggling farmers in a poor land, at work in skill and piety to fulfill their traditions, not to tabulate the sky. Yet these farmers were journeymen of fundamental geometry and simple, effective measurements. You can level a big stone with a water trough if you shape the bottom so that levers and chock stones under it can finally bring it to the right tilt to adjust the broad flat surface above. Serious attention of that kind is what we need to impute to the people of the past; they were not astronomers, but they were manifestly aware of level, line, season and sky, probably more aware than an editor of *Antiquity* today. And they left enigmatic evidence of their awareness. The elegant statistics may often be wishful; it is the optional hypotheses—there are plenty of lines to seek—that make uncertain not so much the data as the true statistical significance of noisy data.

Folklore remains; it has meaning too, amid the noise of merged and misunderstood traditions and the attenuation down languages and generations. "Dancing, water, children, healing, the stones"—all have bearing. The rings are not cemeteries, but they often hold ritual remains, "little and pathetic," a savagely split skull of a child or mere burned ashes. That they were also places to dance is a good guess, intuitive for the sites, ethnographically supported, enshrined in lore. The dead children and the offerings, or the sacrifices, are hard evidence. Often there is water nearby, and the stones still speak mutely of the events of the sky. The sun and the moon were not to be ignored. "In their lifetimes the people ... went to the circles where the powerful bones lay," circles that were constructed rightly. We go to the circles still, without knowing our ancestors. "We shall never see them, ... darker figures against the moon, the dancers, the images of death." The sym-



No other video cassette recorder gives you more than this one.

RCA SELECTAVISION 650.

6-hour recording time, 14-day memory, slow motion, fast motion, frame-by-frame advance, stop action, Automatic Rewind, Remote Control and ***Picture Search***—all important performance features. And all available with the RCA SelectaVision 650.

6-Hour Recording Time—The Maximum Available Today.

This is the big time. The new RCA SelectaVision 650 can record up to 6 hours on a single cassette.* That means fewer interruptions because you change cassettes less often. It also means saving money on tape because you can put more shows on a single cassette.

An Unsurpassed 14-Day Memory.

Just program your selections into the timer up to two weeks in advance. Your RCA SelectaVision 650 will automatically turn itself on and off—and even change

channels—up to eight separate times. So now you can be missing from home without having missed your favorite programs.

RCA Remote Control Special Effects—More Than Ever Before.

You can expect a VCR to let you control *when* and *what* you watch on television. RCA SelectaVision 650 lets you control *how* you watch it, too—without having to leave your chair.

Thanks to an advanced headwheel system, our new remote control unit allows you to view cassettes in ***slow motion*** at 1/30 to 1/4 normal speed. Or ***fast motion*** at twice the normal speed. You can also advance the picture ***frame by frame*** and ***freeze*** it at the moment of your choice.

But that's not all. This year we've added a new ***Picture Search*** mode that enables you to review cassettes in forward or reverse motion at high speed so you can locate footage you want without having the picture go dark.

There are even Remote Pause and Channel Control features for editing out unwanted programming and changing channels—even if your television set isn't equipped with remote control.

Etcetera. Etcetera. Etcetera.

You'll also find a new ***Automatic Rewind*** that resets the cassette when the tape reaches the end of either the recording or playback sequence. Illuminated soft-touch controls. And exclusive new styling that proves a VCR this smart can also be beautiful.

We'll let your RCA Dealer tell you the rest. Just ask to

Simulated TV picture.

see the new SelectaVision 650. You can ask for something less. But you can't ask for anything more.

*RCA Tape is made to our own high standards to insure picture quality and long life. Ask your RCA Dealer for VK-125 (3 hr.) and VK-250 (6 hr.) video cassettes.

RCA

Let RCA turn your television into

SelectaVision®

bols are nonetheless there; the fears are constants. The world in which human life flickers, an enduring world of cycle and time, is also present. Like death, like propitiation, like fear, like reverence, reason too is in the human mind, then as now.

THE PARTICLE ATLAS. VOLUME V: LIGHT MICROSCOPY ATLAS AND TECHNIQUES, by Walter C. McCrone, John Gustav Delly and Samuel James Palenik. VOLUME VI: ELECTRON OPTICAL ATLAS AND TECHNIQUES, Walter C. McCrone, editor, John A. Brown and Ian M. Stewart. Ann Arbor Science Publishers, Inc. (each volume $90). A dozen years ago the first edition of this guide to small-particle identification appeared, one volume of optical micrographs. Six years ago the second edition brought four big volumes; the Chicago center had grown apace in size and power and had acquired a battery of new electron-optical techniques, with modern computerization. These two volumes keep the work up to date.

The clever scanning beams are even more versatile today, but the eye at the lens retains its full importance. At this laboratory 11 microscopists specialize in visible light and only eight use electron optics. The power of light microscopy is plain; here it is exploited with a variety of clever dispersive and interference techniques, although almost always with polarization. "When, oh when, will more biologists learn to use crossed polars?... The beauty of the microscope is not the small samples it deals with, but the shortness of time in which it deals with a sample."

The aim of the volumes is to present a review of techniques and a catalogue raisonné for those who would identify the nature and source of the ubiquitous tiny samples that can be collected from every surface where contact or the flow of air or water deposits them. We see dusts, ashes, pigments, fibers, drugs, pollens and much more. For these analysts a tenth of a nanogram is a generous portion. In Volume V there are about 400 new samples, each shown magnified fifty- or a hundredfold in the subtle dispersive colors of circularly polarized light, artfully augmented by a second flash exposure in plane-polarized light and often by a third flash of unpolarized top light. In this way each photograph carries the most information to the analyst, who will of course be able to manipulate real samples personally.

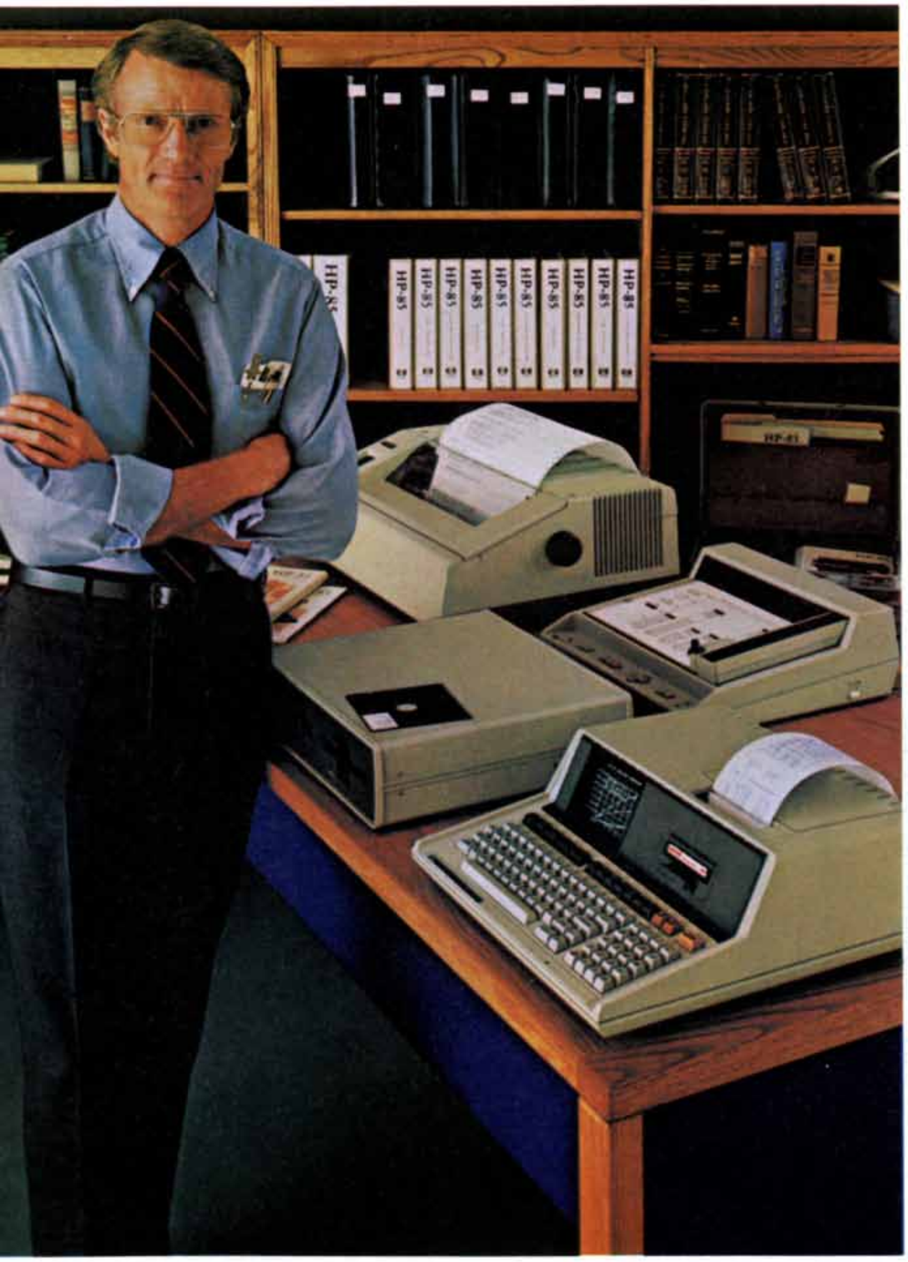
Volume VI is instead in the subtle gray tones of the electron world. The same particles are seen generally at three magnifications (100X, 1,000X, 10,000X) under the scanning electron beam, together with many graphs of atomic compositions provided by energy-dispersive analysis of the X-ray fluorescence under the scanning beam. Some samples are also described by special-area electron-diffraction photographs. All these results can be obtained with a single versatile instrument based on the high spatial resolution of the transmission electron microscope but capable of analyzing the X rays emitted under its fine beam.

Electron-energy-loss spectroscopy is now coming into use, complementing the X-ray techniques poorly suited to the lightest atoms. Scanning with the primary beam of the transmission microscope is now practical as well, offering chemical resolution down to the nanometer range. Even the organic particle, important to many a pollution study, is coming under molecular analysis. It is not much use to study such a fragment with electrons or even with ion beams; its atomic composition is often simple and stereotyped.

A laser-excited Raman microspectroscopy is newly available to these analysts. The instrument can analyze the weak Raman-scattered light to map the particle according to the distribution of many important chemical linkages, if not yet the overall molecular configuration. When heavier trace impurities are present, a valuable aid is offered by the ion microprobe. Sputtering away for each data point a few hundred monolayers of atoms into the mass spectrometer, an ion beam can map a human hair shaft, say, to yield 100 data points, each one analyzed for 40 elements, along the half millimeter of normal hair growth during a single day. The effects of a single meal can be seen! Such techniques should someday allow the sure attribution of a hair to a single individual, even to fix the date of its loss.

Right now such work is too taxing for practical use. Today's pictures nonetheless make plain how hairs differ species by species. A mink hair has remarkable lanceolate scales, and an elk hair seems to have hairs of the second order—whiskers of its own. Nowadays the explosives manufacturers mix little flakes of color-coded layered titanium and aluminum oxides into their daily output. These refractory flakes survive detonation, in micron sizes. One outside surface of the sandwich is made magnetic, the other fluorescent. It is no trick for these microscopists to spot the fragments by their glow, to collect them with a magnet and to read out the color layers under the light microscope. Thus they learn the provenance of any sample of high explosive, down to the lot number. Even such a store of information is clumsy, of course, in the face of the marvelous specificity of living matter from the molecular level upward. Pollen grains are patently artful microsculpture; a microtaggant in explosive is merely a clever shipping label.

Think system. Grow systematically. HP-85.



A COMPLETE COMPUTING SYSTEM

Encapsulated in a mere 20 lb. package, is a complete computing system. Keyboard, tape unit, CRT with alphanumerics and integrated graphics, printer and read/write memory—all ready for instant use. Just turn it on. That's all. Its language is powerful HP BASIC. The HP-85 is simple to use. Extremely effective. Highly productive. In brief, very professional.

"INSTANT INTERFACE" FOR EASY GROWTH

Now the HP-85 can be configured even further to accommodate your increasing demands. Instantly. Just plug in the HP interface bus and handle up to 14 peripherals or instruments. There's no need to write special operating programs. HP's peripheral ROM's do it for you. Options available now include flexible disc drives, additional read/write memory, graphics plotter, full-width printer, many software Application Pacs. More options are under development. Indeed, Hewlett-Packard is committed to providing solutions that meet your varied and changing needs.

ONLY FROM HEWLETT-PACKARD

There is no comparable computing system in the marketplace that offers you this combination of power, sophisticated growth potential, software, lightweight portability, convenience and ease-of-use. In addition, HP maintains service facilities throughout the world. On-site service, if you wish. For details and address of nearest dealer, CALL TOLL-FREE 800-547-3400, Department 236U; except Hawaii/Alaska. In Oregon, call 758-1010. Or write Hewlett-Packard, 1000 N.E. Circle Blvd., Corvallis, OR 97330, Dept: 236U.

610/16

The Personal Computer For Professionals: HP-85 specifications: Memory—16K RAM expands to 32K, 32K ROM expands to 80K; CRT—32 chars./line, 16 lines; graphics—256 X 192 dots; thermal printer—2 lines/sec.; magnetic cartridge capacity —200K; I/O—HP-IB (IEEE 488), RS-232, 16-bit GP-IO, BCD; Application Pacs include Statistics & Regression Analysis, Finance, Math, Linear Programming, Text Editing, Waveform Analysis, Circuit Analysis, BASIC Training, & Games.

HEWLETT PACKARD

Established 1845 SCIENTIFIC AMERICAN October 1980 Volume 243 Number 4

Health Maintenance Organizations

A decade ago the Federal Government endorsed the concept of prepaid medical care. Today the number of HMO's is less than was envisioned then, but it is nevertheless increasing

by Ernest W. Saward and Scott Fleming

Nearly 10 years have passed since the Nixon Administration, concerned about the rapid rise in health-care expenditures in the U.S. following the adoption of the Medicare and Medicaid programs in 1965, announced a plan to stimulate the development of Health Maintenance Organizations (HMO's) as the leading feature of a "national health strategy" aimed at constraining the rising costs. The HMO, based on the principle of prepaid medical care, was seen as a competitive alternative to the traditional fee-for-service system. Although the HMO program has fallen considerably short of the ambitious goals set by the Administration in 1971, it has progressed significantly and now appears to be expanding steadily. Our purpose here is to review the concept, its present standing and its prospects.

The name Health Maintenance Organization is a promotional term first employed in 1970 by Paul M. Ellwood, Jr. (the head of InterStudy, a "think tank" specializing in health policy), in advocating the concept that the public interest would be better served by competition among alternative health-care systems than by increased Government regulation of the field. Although many organizations are known as HMO's, not all of them fit the definition of the term contained in Federal laws and regulations. Even those restrictive provisions, however, accommodate a diversity of organizations, ranging from individual-practice associations, which operate on a fee-for-service basis, to what can be called the classic prepaid-group-practice plans such as the Kaiser-Permanente Medical Care Program (operating in several states) and the Group Health Cooperative of Puget Sound. (Such programs inspired the Nixon Administration's endorsement of the HMO concept.) Notwithstanding this diversity a number of characteristics distinguish HMO's and similar organizations from the traditional systems of financing and delivering health care. They include (1) a defined population of enrolled members, (2) payments determined in advance for a specific period of time and made periodically, (3) services provided to the patient by HMO physicians for essentially all medical needs, with referrals to outside specialists being controlled by the HMO, and (4) voluntary enrollment by each individual or family.

Because HMO's and similar organizations constitute an alternative to traditional systems, they are inherently in competition with those systems and can succeed only by attracting a substantial amount of patronage from them. The competitive factors include not only relative costs but also the concern of consumers with the quality of care, the accessibility and availability of service and the convenience of the place where the service is provided.

The term health maintenance organization reflects rather inexactly what a prepaid medical service does, since patients still tend to seek care mainly when they are sick. It does, however, reflect a crucial difference in economic outlook between the traditional fee-for-service physician and the physician member of an individual practice association or an HMO organized on the classic Kaiser-Permanente model. In the prepaid group practice the physician is compensated by a payment fixed in the organization's annual budget, with his prospects for an increase in earnings, through a bonus or a salary increase, largely dependent on whether the year's operations meet, improve on or fail to meet the budget. A patient who requires elaborate tertiary procedures can add large sums to the expense side of the budget. The HMO physician therefore has an incentive to urge his patients to have periodic physical examinations, to report symptoms of illness promptly and to cure habits that might lead to illness. He also monitors carefully the treatments a sick patient receives in order to avoid inappropriate procedures. Self-interest and peer pressure combine to establish an environment in which efficiency in professional practice and the appropriate use of hospital facilities and other health-care resources become the goal of both professionals and administrators.

The fee-for-service physician, on the other hand, is little concerned with expense, which nowadays is in any case mostly borne by third parties such as insurance companies and the Federal Government. Other factors contribute to the environment in which costs under the traditional system are open-ended, with few rewards for efficiency and economy. Hospitals, competing for the patronage of physicians, find it necessary to provide more (and usually more elaborate) services. And although the physician receives slightly less than 20 cents from each dollar spent on health care, he effectively controls an additional 60 cents through his specifying of hospital service, diagnostic procedures, drugs and diverse therapies.

Because of the contrast in economic outlook between the two modes of medical care HMO's are now acclaimed by many policymakers in the Government and by other people concerned with the escalation of health-care costs, all of whom cite the cost-saving capability that studies of many HMO's and HMO prototypes have demonstrated. This attitude overlooks other factors that gave rise to the predecessors of HMO's long ago and are still embodied in the modern HMO.

The predecessors of the HMO include organizations that mutualized costs and provided direct health-care service for Venetian seamen in the 13th century, for British seamen in the 16th century and for merchant seamen of the U.S. beginning in 1798 (the origin of the hospital system now operated by the U.S. Public Health Service). Analogous systems were formed late in the 19th century and early in this one by employers, particularly those with work forces in isolated places, and by mutual-benefit societies with an ethnic or work-related basis. The contemporary HMO began to emerge with the formation of such organizations as the Ross-Loos Medical Group, which began serving employees of the Los Angeles Department of Water and Power in 1929, and the Group Health Association of Washington, D.C., which was organized by employees of the Federal Home Loan Bank Board as a consumer cooperative in 1937.

Kaiser-Permanente, the major prototype of the HMO, also traces its origins to that time; it developed from a system serving workers building the Los Angeles Aqueduct in 1933, construction workers and their families at the Grand Coulee Dam project in eastern Washington in 1938 and, on a larger scale, Kaiser shipyard employees in and around Portland, Ore., and San Francisco and employees of the Kaiser steel plant at Fontana (near Los Angeles) during World War II. These moves established the system that was opened to public enrollment in 1945. Organized consumers formed the Group Health Cooperative of Puget Sound in 1946, and the Health Insurance Plan of Greater New York was organized in 1947.

A look at the history of these organizations reveals several motives and claimed advantages. They include the assured availability of and access to services through integrated group-practice facilities; the mutualization and budgeting of the costs of illness through prepayment; budgeting for the provision of services; the creation of incentives for health maintenance and preventive health services, and the assurance of a good quality of care through the group practice of physicians offering multiple specialties. The main issues of health care today—rising costs and concern about the overutilization of services—were notably absent.

Developments during World War II strongly influenced the evolution of health insurance in the U.S. In 1940 less than 10 percent of the population had health insurance. Federal measures to relieve the pressure on the wage-control system toward the end of the war, particularly the treatment of employers' contributions toward health-care coverage for employees and their families as a nontaxable element of compensation exempt from wage controls, enormously accelerated the development of today's fringe-benefit economy.

Another boost came from a Supreme Court ruling that health care is an appropriate subject for collective bargaining under the National Labor Relations Act. These events greatly stimulated the growth of group-health coverage by Blue Cross and Blue Shield plans, commercial insurance carriers and prototype HMO's. By the early 1960's most economically self-sufficient people in the U.S. had a significant amount of health-care coverage. The standard thus established gave rise to the concept that health care is a right, and this concept in turn brought forth tax-supported funds to ensure that economically dependent people would have access to health-care services. The Government's role was reinforced and amplified by the enactment of the Medicare (for the aged) and Medicaid (for the indigent) programs.

Since Medicare is a social-insurance program whose beneficiaries are entitled to receive specified care regardless of the cost, it offers the same open-ended funding as private health insurance. It has therefore become a major uncontrollable cost in the Federal budget. Medicaid is also an entitlement program and so is similarly open-ended, although access to the program is subject to a means test. The fiscal consequences of the Government's commitment to these two major programs underlie Federal and state concern over the rising costs of health care.

By the 1970's most of the country's major employers, including governmental bodies, had established fringe-benefit programs with the provision of health care for employees and their dependents. Except in the few parts of the country where prototype HMO's were well established the health-insurance business was held and closely guarded by the commercial insurance companies and by Blue Cross–Blue Shield. Moreover, the idea that employers could help to contain health-care costs by promoting competition through dual-choice programs (an employee can choose either conventional care or membership in a prepaid-care group) had gained little acceptance.

It was in this setting that the Nixon Administration endorsed the HMO concept as a cost-constraining alternative to the fee-for-service system. The President presented the plan to Congress in February, 1971, and three months later Elliot Richardson, the Secretary of Health, Education, and Welfare, announced that "the goals of the Administration are to develop 450 HMO's by the end of fiscal year 1973," to have 1,700 by the end of fiscal year 1976 and by the

OPERATING HMO'S are depicted on this map as they existed in June. A prepaid group practice consists of several physicians working together on a fixed annual budget. A network embraces two or more group practices

INDIANA
- Fort Wayne ●
- Indianapolis ●

KENTUCKY
- Lexington ●
- Louisville ●
- Madisonville ●
- Middlesboro ○

LOUISIANA
- Baton Rouge ●●
- Lafayette ●

MAINE
- Farmington ○
- Rockland ○

MARYLAND
- Baltimore ●●●●●●●●●○

MASSACHUSETTS
- Boston ●●●○
- Springfield ●⊙
- Worcester ●○
- Athol ○
- Brockton ●

MICHIGAN
- Detroit ●●●●○⊙
- Flint ○
- Kalamazoo ●
- Lansing ●
- Saginaw ●

MINNESOTA
- Minneapolis–Saint Paul ●●●●●○○
- Saint Cloud ●
- Virginia ●
- Two Harbors ●

MISSOURI
- Kansas City ●
- Saint Louis ●●●○

NEBRASKA
- Lincoln ●
- Omaha ○
- McCook ○

NEW HAMPSHIRE
- Nashua ●

NEW JERSEY
- Atlantic City ○
- Newark ●●○○○
- Trenton ●
- Vineland ●

NEW MEXICO
- Albuquerque ●○

NEW YORK
- Buffalo ●○
- Nassau-Suffolk ●●
- Albany ●
- New York ●●●
- Rochester ●○⊙
- Syracuse ●

NORTH CAROLINA
- Winston-Salem ●

NORTH DAKOTA
- Hettinger ⊙

OHIO
- Cincinnati ●○
- Cleveland ●●⊙
- Columbus ●○
- Dayton ●○
- Toledo ●
- Bellaire ●
- Marion ○

OREGON
- Eugene ○
- Portland ●●○○○
- Salem ○

PENNSYLVANIA
- Allentown ○
- Philadelphia ●●●○○
- Pittsburgh ●●○
- Reading ●
- Philipsburg ●
- Danville ●
- Harrisburg ⊙

RHODE ISLAND
- Providence ●●●○

SOUTH CAROLINA
- Greenville ●

TEXAS
- Dallas ●○○
- El Paso ●
- Houston ●○
- San Antonio ●○

UTAH
- Salt Lake City ●

WASHINGTON
- Seattle ●●○○
- Spokane ●●
- Bremerton ○

WEST VIRGINIA
- Wheeling ○
- Beckley ●○

WISCONSIN
- Eau Claire ●○○
- Green Bay ○
- Madison ●○
- Milwaukee ●●●○
- Marshfield ●
- Stevens Point ○
- Wausau ○
- La Crosse ●
- Wisconsin Rapids ○

that have joined to form an HMO. In individual-practice associations the physicians generally work on a fee-for-service basis but the patient (as in all HMO's) pays in advance for his medical care for a year or some other period. By June the 240 HMO's in the U.S. had a total membership of about nine million, which is approximately 4 percent of the U.S. population. The data for the map were supplied by InterStudy, a "think tank" specializing in health policy. Between a survey made by InterStudy in 1978 and the review the organization carried out this year the number of HMO's rose by 40 (20 percent) and the membership rose by approximately 1.7 million (22 percent).

end of the decade "to have a sufficient number of HMO's to enroll 90 percent of the population if they [desire] to enroll." Under pressure from elements in organized medicine and the medical-insurance industry, however, the Administration soon lost its enthusiasm for HMO's.

Nevertheless, the HMO concept, having acquired a momentum and a constituency of its own, received an impetus in the form of the Health Maintenance Organization Act, which was passed by Congress in 1973 and approved by President Nixon even though his administration had by then retreated from its strong advocacy of the concept. The law provided several kinds of financial assistance to developing programs. Nonprofit sponsors could obtain grants to study the feasibility of establishing an HMO. Grants for planning and early development were also available. Newly operating nonprofit plans could obtain loans to cover the deficits of the start-up period and profit-oriented programs could obtain Federal guarantees of loans. These financial programs aided greatly the development of HMO's and were absolutely vital to the genesis of many new plans.

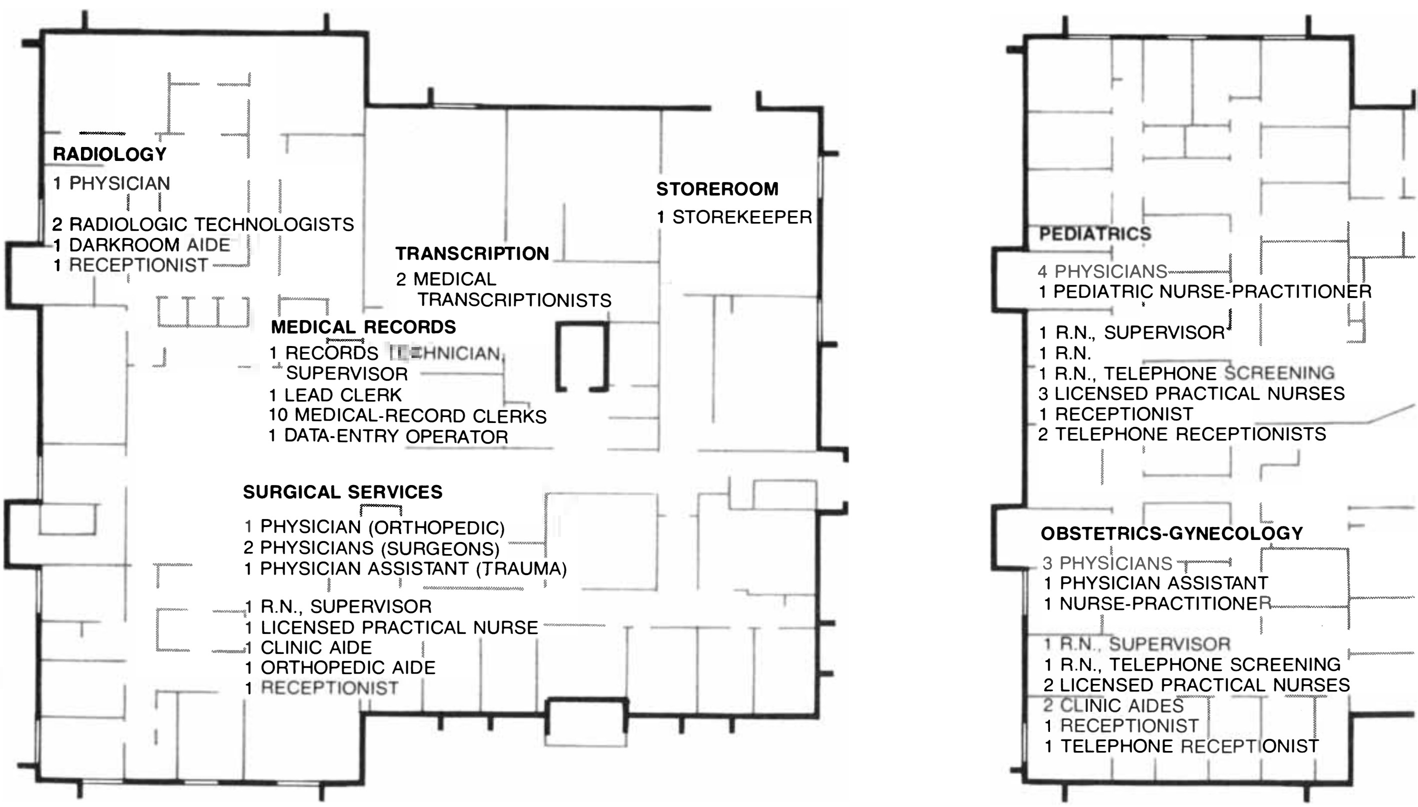
A crucial feature of the law was the requirement that an employer with more than 25 workers at a particular location had to offer them (if he provided health-care coverage as a fringe benefit) a choice between conventional coverage and an HMO if the choice was available in the area. Although this stipulation has been criticized in some quarters as heavy-handed Government interference, it was an essential step toward breaking the grip held by traditional carriers on health-care coverage.

Certain other features of the HMO law presented competitive difficulties for HMO's. For example, a prepaid-care group seeking to qualify as an HMO under the law was required to offer a broader range of benefits than the prototype HMO's had offered and one that cost more than most competing health-insurance programs. The HMO was also required to conduct an annual open enrollment, accepting all applicants up to the capacity of the facility, whereas conventional carriers need not offer individual coverage and can reject individual applicants who seem likely to be costly risks. The law has since been amended to moderate the open-enrollment qualification.

Against this background it is appropriate to ask how well HMO's have performed. One answer was provided two years ago by Harold S. Luft, a specialist in health economics who was then at Stanford University. "The available studies [since 1950]," he said, "support the claims and expectations that medical-care expenditures are lower for HMO enrollees than for people with conventional insurance coverage. The most convincing evidence pertains to enrollees in the California Kaiser plans, whose total costs were 10 to 40 percent below those with conventional coverage.... There is no documentary evidence that the costs for enrollees in individual-practice associations are lower than those for people with conventional insurance."

The most pertinent factor in making HMO costs lower is that members of HMO's spend far less time in the hospital than people with other forms of coverage. Excluding maternity cases, the hospitalization rate per 1,000 members per year may be as low as half that of people in Blue Cross–Blue Shield plans and those with commercial insurance. Less elective surgery and more use of one-day surgery without hospitalization account for a significant part of the difference, but hospitalization for nonsurgical reasons is also generally lower in group-practice prepayment plans.

Critics of the HMO concept say several explanations other than efficiency in the delivery of health care can be offered for such statistics. The explanations usually advanced by the critics are (1) HMO's systematically select young, healthy people as members, (2) they skimp on necessary services and do not provide care of the highest quality and (3) the savings achieved by HMO's are overstated because members obtain a substantial amount of medical service outside the HMO.



SERVICES AVAILABLE in a typical fairly large HMO, the three-story Westminster–North Denver facility of the Kaiser Foundation Health Plan of Colorado, are indicated on a series of floor plans beginning at the left with the ground floor. Kaiser has four medical-

Both experience and the results of major studies indicate that these criticisms are not well founded. It is true that enrollees of HMO's constitute a selected population in which the aged and the disadvantaged are underrepresented, but the same thing is true of the rolls of Blue Cross, Blue Shield and commercial health-insurance companies. In each case the selectivity reflects the fact that health-care coverage in the U.S. is based primarily on employment.

A thoroughly objective evaluation of the quality of medical care and the appropriateness of medical services is difficult in any health-care system. Nevertheless, the published evidence indicates that plans of the HMO type maintain a quality of care that is equal to and possibly better than the care available elsewhere in the same communities. Evidence is supplied by studies published in 1959 by the Commission on Medical Care Plans of the American Medical Association, the report of the National Advisory Commission on Health Manpower in 1967 and a report by John W. Williamson and Frances Cunningham of the Johns Hopkins School of Public Health covering 1958–79.

Several structural features of a group-practice prepayment system tend to ensure that the care is of good quality. The physicians are aware that substandard care or the withholding of necessary service for any patient could give rise to a need for more and costlier care in the future. Consultation is readily available without charge to the patient, and it does not carry the risk of diminished future income for the physician who recommends it, since it is provided by a specialist who is either in the same HMO or has a working agreement with it. Finally, physicians in prepaid-care organizations are constrained by the budget to think twice about performing any service that is unnecessary, inappropriate or of doubtful value.

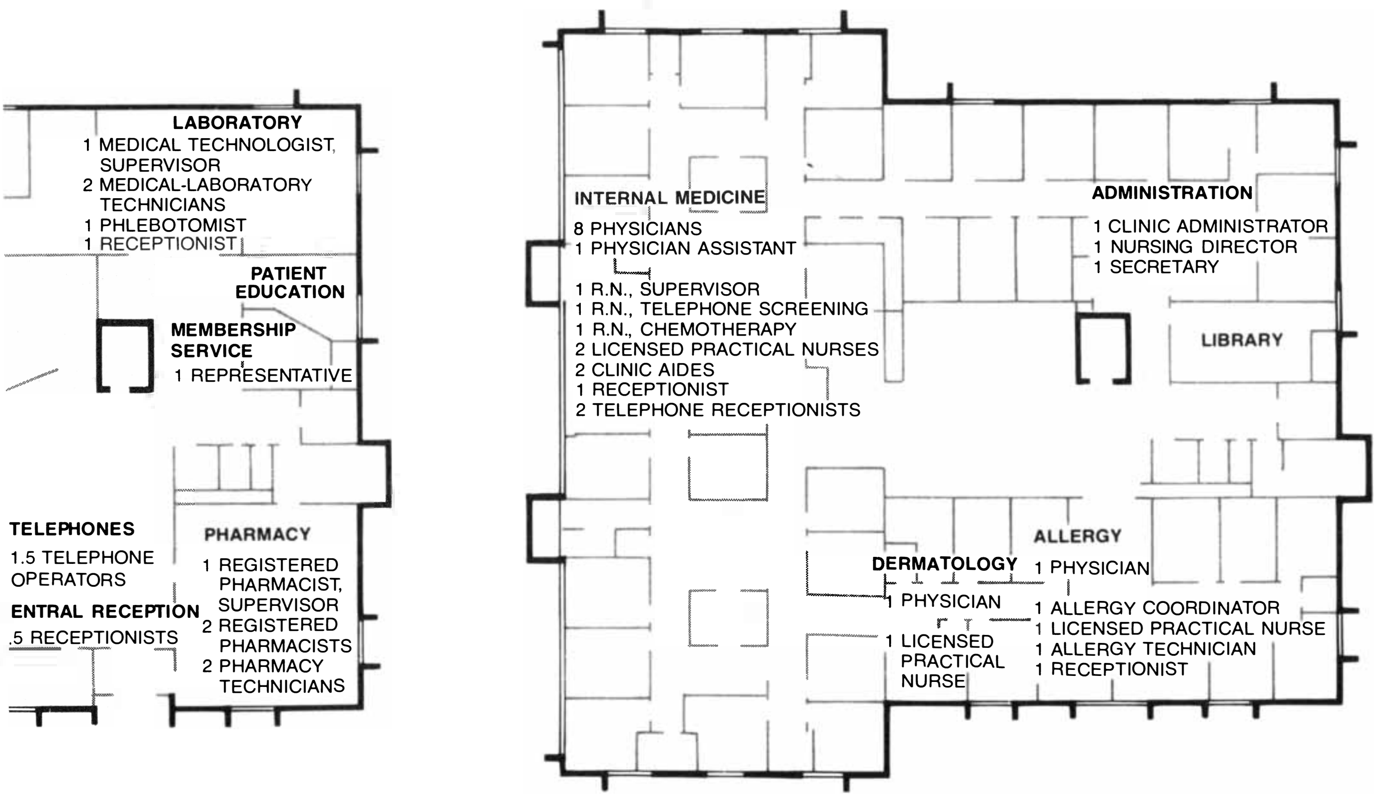
It is certainly true that for a variety of reasons members of HMO's sometimes go outside the organization for medical service. The study by Luft concluded, however, that out-of-plan use is generally not substantial.

A census of HMO programs made in 1978 by InterStudy showed that the concept had become well established, although it was not yet widely available. The census found 13 HMO's that had been founded before 1950; they accounted for about 4.5 million (62 percent) of the total HMO membership. Another 16 programs, started in the 1950's and 1960's, included about 11 percent of the total membership, and about 170 programs begun since 1970 served the remaining 27 percent. A survey by InterStudy this June found 240 HMO's in operation. Their enrollment was about nine million, or approximately 4 percent of the population.

A most significant development in the HMO movement is the sponsorship of HMO's by medical schools that deliver a large percentage of the country's total medical care. In the Boston metropolitan region the Harvard Community Health Plan is now approaching an enrollment of 100,000, and Rush Medical College in Chicago is supporting the growth of satellite HMO ventures.

Most of the new programs remain relatively small, however, with fewer than 30,000 members each. (Kaiser-Permanente, the giant of the field, had more than 3.8 million members as of June, 1980.) It usually takes quite a long time for a new HMO to develop a substantial growth in membership. Part of the reason is that the HMO Act still has burdensome requirements that do not apply to competing plans. Medical economists frequently point out that HMO's have not yet been given a fair test in the market. Another reason for the ordinarily slow growth of new HMO's is that most people are generally satisfied with their physician and must perceive rather convincing advantages before they will switch to an HMO.

Nevertheless, the broad appeal of the HMO is demonstrated by the large share of the market that such organizations have taken on the West Coast. In the Los Angeles area nearly two in seven people belong to an HMO. The ratio in the San Francisco, Portland and Seattle areas is about one in four. In these communities the enrollment in the Kaiser-



office locations (for 110,000 members) in the Colorado region; the Westminster–North Denver one serves about 22,000 people. The services available to them extend beyond this facility, since referrals for certain specialties not available there are made to other facilities.

DATE	PATIENT A	PATIENT B	PATIENT C
JAN. 3	ARTERIOSCLEROTIC HEART DISEASE WITH ANGINA PECTORIS.		
FEB. 19			CHEST TIGHTNESS, CONGESTION.
MARCH 18			SYNOVITIS (BURSITIS AND TENOSYNOVITIS), SHOULDER.
MARCH 31		REFRACTION, WITH REFRACTIVE ERROR FOUND AND GLASSES GIVEN. HEADACHE.	
APRIL 9		HEADACHE.	
APRIL 10		HEADACHE.	
APRIL 22		HEADACHE.	
APRIL 26			FOREIGN BODY IN EYE.
APRIL 29	ARTERIOSCLEROTIC HEART DISEASE WITH ANGINA PECTORIS.		
APRIL 30		WELL-CHILD CARE. HEADACHE.	
MAY 10		(DID NOT KEEP SCHEDULED APPOINTMENT WITH OPHTHALMOLOGIST.)	
MAY 29	PHARYNGITIS, BACTERIAL.		
JUNE 13		OBESITY, NOT SPECIFIED AS ENDOCRINE ORIGIN.	
JUNE 18	ARTERIOSCLEROTIC HEART DISEASE WITH ANGINA PECTORIS.		
JULY 3		OBESITY, NOT SPECIFIED AS ENDOCRINE ORIGIN.	
JULY 10	ARTERIOSCLEROTIC HEART DISEASE WITH ANGINA PECTORIS. PHARYNGITIS, BACTERIAL.		
JULY 11	DEBILITY AND UNDUE FATIGUE. ANXIETY REACTION.		
JULY 26	PAIN IN CHEST. HEALED CORONARY OCCLUSION. ARTERIOSCLEROTIC HEART DISEASE WITH ANGINA PECTORIS.		
JULY 26	MYOCARDIAL INFARCTION.		
AUG. 4	ANXIETY REACTION. ARTERIOSCLEROTIC HEART DISEASE WITH ANGINA PECTORIS.		
AUG. 18	ARTERIOSCLEROTIC HEART DISEASE WITH ANGINA PECTORIS.		DERMATOPHYTOSIS OF FOOT. HERPES FIBRILIS.
OCT. 1	ARTERIOSCLEROTIC HEART DISEASE WITH ANGINA PECTORIS. ASTHMA DUE TO OTHER AND UNSPECIFIED CAUSES.		
OCT. 5	LARYNGITIS AND TRACHEITIS, NOT SPECIFIED; NO ANTIBIOTIC GIVEN.		
NOV. 24	ARTERIOSCLEROTIC HEART DISEASE WITH ANGINA PECTORIS.		
DEC. 11	IMMUNIZATION.		
DEC. 30			HEADACHE. DROOPING FACE.

USE OF AN HMO over the course of a year is charted for three types of member: a relatively extensive user, Patient *A*, a 55-year-old man; a moderate user, Patient *B*, a boy of 14, and a fairly light user, Patient *C*, a man aged 41. The records are based on computer printouts reflecting the utilization of prepaid-health-care facilities in Oregon in 1975 by members of the Kaiser Foundation Health Plan there.

Permanente service has generally been increasing as fast as the organization can mobilize physicians and facilities.

Even these figures substantially understate the potential attractiveness of HMO's in the market because large segments of the population have not had an opportunity to enroll in an HMO. Many people do not have regular employment in fields that offer health-care coverage as a fringe benefit, and many employers do not offer the HMO alternative even where it is available. A number of the more successful HMO's have had to suspend new enrollment at times because limits on the size of the staff or the facility prevented them from serving any more members. The potential of the HMO is suggested by the fact that in certain large employment-based groups where the option of joining an HMO has been available for years the HMO's cover from 30 to 60 percent of the market.

The trend of public policy in the face of soaring costs for health care has been toward stronger regulatory intervention. Examples include a mounting emphasis on reviewing the utilization of services offered by Medicare and Medicaid, reviews of the appropriateness and quality of care through the Government-sponsored medical groups known as professional-standards review organizations and the control of capital expenditures on health-care facilities through the National Health Planning and Resources Development Act of 1974. The regulation of hospital rates by the equivalent of public-utility commissions is now operative in many states. Only recently have the Congressional policymakers on health perceived the inconsistency between the regulatory approach and trying to stimulate the development of HMO's as a competitive alternative to conventional care. Even so, it is likely that Congress will continue to straddle, simultaneously pursuing both a market-oriented approach and pervasive regulation. Notwithstanding these ambiguities, the HMO concept is continuing to spread at an impressive rate. Industrial leaders are awakening to the cost-saving potential of competition among alternative systems. Organized labor has generally advocated the HMO concept and many locals urge their members to join available plans.

Throughout the past decade personal expenditures for health care have increased much faster than the productivity of society as measured by the gross national product. The trend cannot continue indefinitely. The process by which society brings the rise in health-care costs into line with the growth of the economy will be as important to the people of the country as it will be to health-care organizations, of which HMO's are only a part.

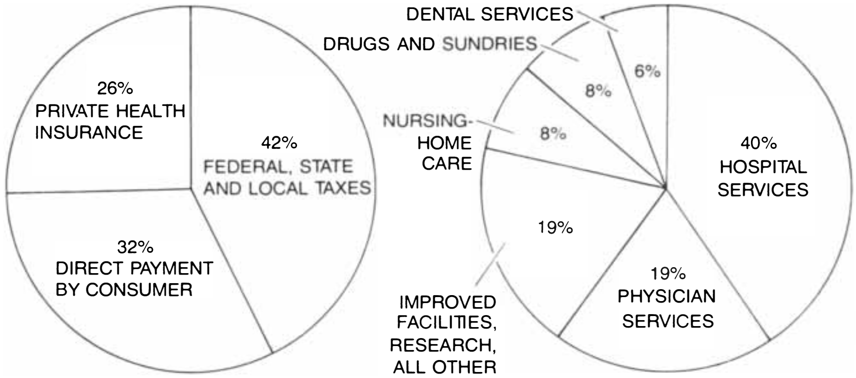
Several possibilities can be envisioned. If the U.S. adopts a national health service along the lines of the British system, which at present seems quite unlikely, HMO's would become irrelevant or nearly so. If the Government sought to control costs by extending a system such as Medicare to the general population and combined it with strict limitations on physicians' fees and hospital costs (also unlikely), the development of HMO's might be severely curtailed, since the system would reduce or eliminate the financial incentives for funding agencies to consider costs in choosing health-care providers.

Another possibility, which some observers see as the only likely one for the near future, is "catastrophe" national health insurance (Federally established protection that would apply only after an individual or a family had incurred substantial health-care costs during a year). HMO's, which emphasize comprehensive coverage and primary care, would probably have difficulty integrating their benefits with such a system.

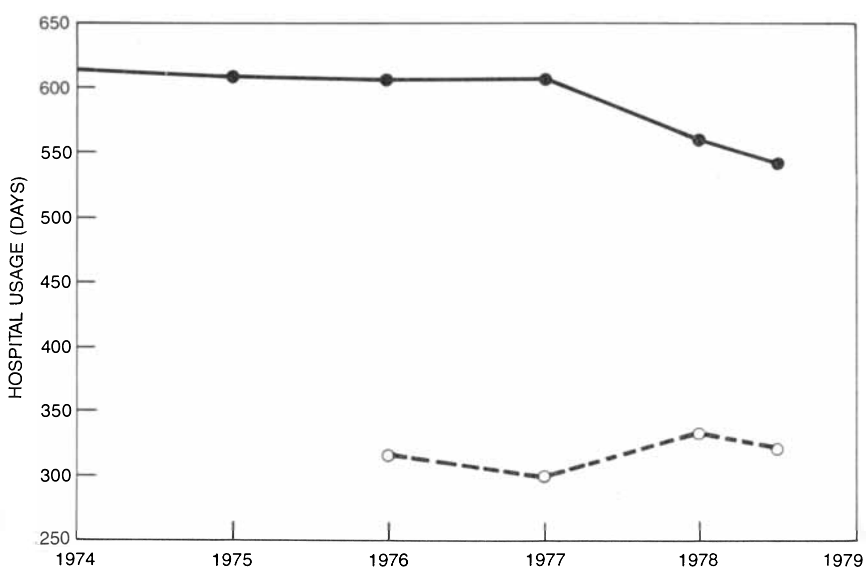
Effective HMO's should thrive, however, under any plan seeking to foster competition among providers of health care. As long as the service rendered by HMO's continues to cost less than conventional care, HMO's will grow at the expense of less efficient providers.

It would be unrealistic to expect anything like that to happen soon. The HMO initiative has suffered severely from grandiose and impractical expectations. Years of effort, a substantial amount of capital and an astute management are required to build up an effective system for delivering health care, particularly an HMO.

The HMO should be extolled for what it is: a creative and effective rationalization of one aspect of health planning. For HMO's to become a stronger alternative in the U.S. medical system they must become more numerous and widespread than they are now. Only by having a large and satisfied constituency will they acquire the political strength to ensure their future.



ECONOMICS OF HEALTH CARE in the U.S. in 1977 are charted. The diagram on the left reflects the sources of funding for health care, the one on the right the classes of expenditure.



"RIPPLE EFFECT" often observed when an HMO is established in a community is reflected in this graph of hospital use in Rochester, N.Y. The top curve represents inpatient medical or surgical days per 1,000 members of Blue Cross, the bottom curve the same thing for HMO members. The HMO members had a much lower hospital-utilization rate, and through the ripple effect the number of days spent in the hospital by Blue Cross members soon declined.

The Bizarre Spectrum of SS 433

Displaced emission lines found in the spectrum of this extraordinary stellar object suggest that it is spewing two narrow high-speed jets of matter in opposite directions. How can such behavior be explained?

by Bruce Margon

With 100 billion stars in the Milky Way there is not much room for individuality. Virtually every phenomenon in the galaxy has a chance of happening more than once. Hence "unique" is not a term often applicable in stellar astronomy. Even some highly unusual astronomical objects have been found to have analogues; for example, the Crab Nebula, the spectacular remnant of a supernova explosion that was observed in A.D. 1054, apparently has two or three counterparts elsewhere in the galaxy. Over the past two years, however, observations have accumulated on a stellar object that probably is deserving of the term unique. The object, known as SS 433, appears to be ejecting two very narrow jets of matter in opposite directions at incredibly high velocities. No other star has ever been observed to behave in such a manner.

SS 433 is a fascinating object for many reasons, not the least of which is the fact that it was discovered independently several times in the past two decades, although each time its discoverers failed to recognize its truly exotic character. The object emits an exceptional pattern of radiation in several different parts of the electromagnetic spectrum (at wavelengths corresponding to visible light, radio waves and X rays), and each of these peculiarities had been individually noted over the years. It was not until quite recently, however, that a comprehensive picture of SS 433 began to emerge. It is instructive to trace the early history of the observations, in order to see how close the earlier investigators came to detecting the unique properties of the object.

The first pertinent clues were obtained in the early 1960's when astronomers at Case Western Reserve University conducted a systematic survey aimed at discovering certain types of faint stars near the central plane of our galaxy. The particular telescopic technique they employed incorporated an objective prism, a light-dispersing optical element that smears stellar images on a photographic plate into tiny trails representing the visible spectrum of the starlight. Unlike the more conventional techniques of astronomical spectroscopy, in which the light from a single star is passed through a narrow slit near the focus of the telescope and then dispersed over a considerable part of a photographic plate, the objective-prism technique separates the starlight by wavelength only slightly. It does so simultaneously for every visible stellar image in the field of view, however, rather than for one star at a time. In this way very crude spectrograms, suitable mainly for classifying different types of objects, can be obtained for large numbers of stars with only a modest investment of precious telescope time.

The Case survey was specifically designed to find new stars with emission lines in their spectra. The existence of such lines forms the basis of a versatile diagnostic technique common to many areas of physics and chemistry. When the electrons bound to the atoms in any diffuse gas are excited to higher "orbital" energy states, either through collisions with other particles or through radiative interactions, their subsequent de-excitation creates new photons, or light quanta, at certain discrete wavelengths that are determined by the energy difference between the orbital states and hence by the electronic structure of the particular element. The appearance in a spectrogram of these lines (so named because they manifest themselves as a linear darkening at a particular wavelength on the exposed photographic negative) can then serve as a highly specific probe of the physical conditions in the emitting gas. The precise wavelength of such a line, for example, is an unambiguous indicator of the original light-emitting chemical element, since it bears the memory of the atom's electronic structure.

Objective-prism studies such as those carried out by the Case workers have shown that perhaps as many as 10 percent of all stars produce emission lines characteristic of a hot, excited gas; such a gas is thought to exist in a diffuse outer layer surrounding an otherwise normal star. There are many possible reasons for this abnormal situation, and they vary from star to star. Emission lines are often found, for example, in association with very young stars and very old stars. On occasion they are also identified with a stable middle-aged star.

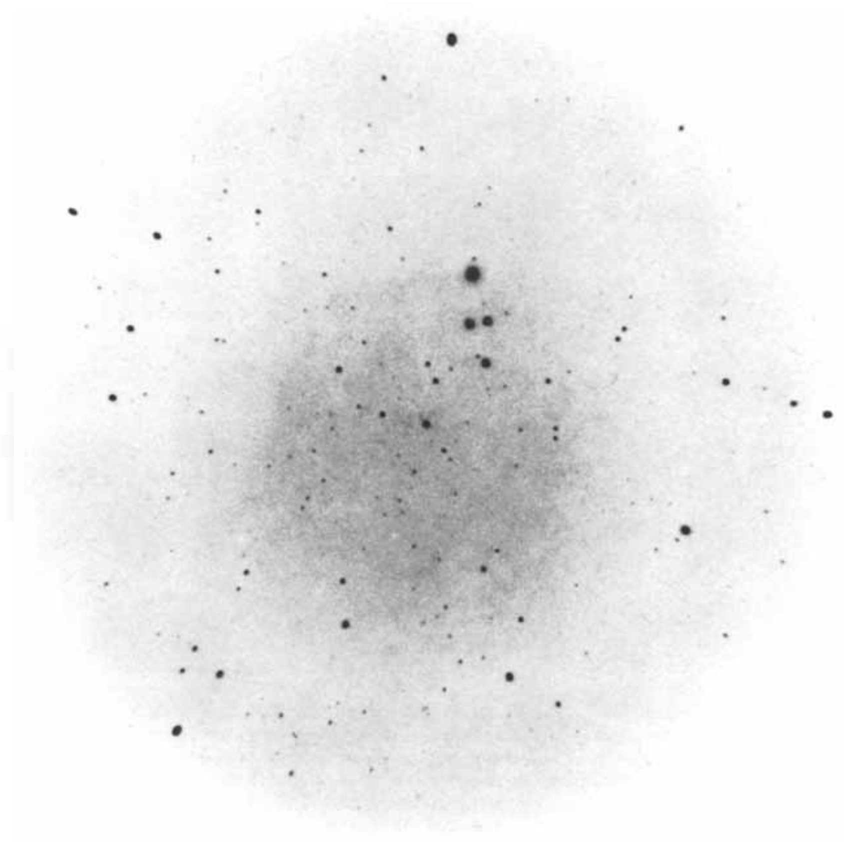
One of the plates exposed by the Case workers was centered on the constellation Aquila in the midst of the Milky Way; many strong emission-line objects were visible on the plate. In 1977 C. Bruce Stephenson and Nicholas Sanduleak published a list of the emission-line objects in the area; the 433rd entry on their list was the star now known as SS 433. Stephenson and Sanduleak had no reason for singling out SS 433 as being particularly different from the hundreds of other emission-line stars identified in their survey; spectroscopic data more detailed than those that could be obtained by the objective-prism technique would be needed for the purpose. Indeed, one of the primary goals of surveys such as the Case effort is to stimulate future detailed observations of interesting-looking objects. As it turned out, with SS 433 Stephenson and Sanduleak succeeded in this respect beyond their fondest dreams.

With an apparent brightness equivalent to that of a 14th-magnitude star, SS 433 is more than 1,000 times too faint to be visible to the unaided eye. It is an easy target for spectroscopic observations, however, even with a rather small telescope. The follow-up observations might well have been stimulated as early as 1975, when Lawrence Krumenaker, another Case astronomer, published a short paper that appeared before the publication of the Stephenson-Sanduleak catalogue. The paper mentioned a small subset of emission-line stars from the full survey, including the then-anonymous SS 433, and it gave the celestial coordinates and charts needed to enable other astronomers to locate these objects. Unfortunately the coordinates listed for SS 433 were incorrect, probably because of a simple transcription error. The object could not have been found again even if another observer's curiosity had been piqued.

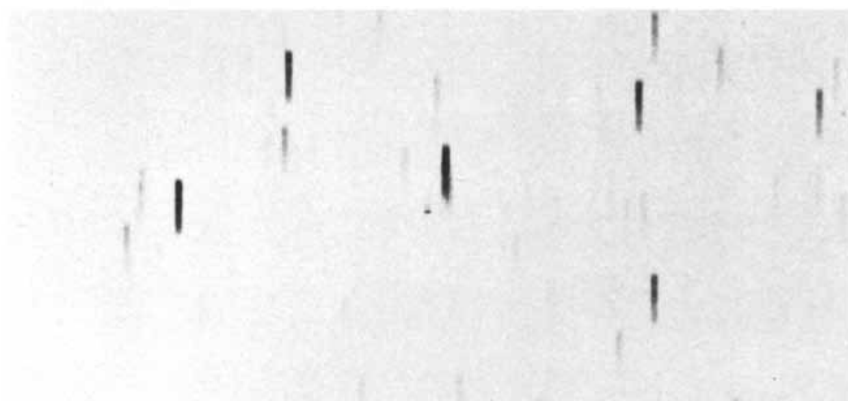
At roughly the same time as the Case survey was in progress an entirely unrelated observational program also (unwittingly) discovered SS 433. Radio astronomers at the University of Cambridge were compiling one of their several comprehensive lists of celestial radio sources, in this case the list published as the Fourth Cambridge Catalogue. Most of the sources of radio waves included in the catalogue turn out on detailed examination of optical photographs to be associated with distant extragalactic objects such as galaxies and quasars. The radio emission from normal stars in our galaxy is simply too feeble to detect in such surveys, even for objects close to the solar system. In the course of the survey a rather bright radio source was noted by the Cambridge workers in Aquila, and it was designated 4C 04.66.

If anyone at this point had noted the coincidence in the position of the radio source with the position of a fairly bright visible star, further optical observations would have almost surely been undertaken at that point, owing to the scarcity of visually detectable stellar radio sources. Again, however, the published celestial coordinates for 4C 04.66 did not agree with the actual position of SS 433. Here the probable reason for the discrepancy was instrumental; that particular region of the sky is a confused and patchy jumble of foreground and background radio sources, and the instruments used in the Cambridge survey were incapable of accurately fixing the position of even a comparatively strong source against such confusion. Once again a chance to notice the special nature of SS 433 was missed.

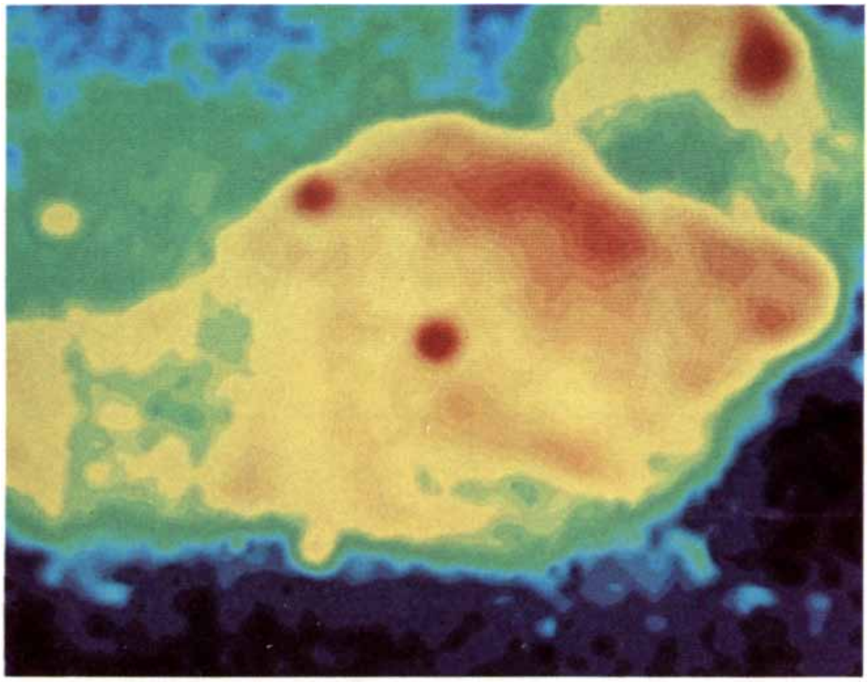
Actually part of the reason for the confusion in this case is directly relevant to the story of SS 433. Centered near that region of sky is an intense, extended patch of radio emission covering an area twice the angular size of the full moon. The radio feature is also a previously catalogued and largely forgotten object, known as W50. (The designation is derived from the source's appearance in a catalogue of radio features compiled in the 1950's by the radio astronomer Gart Westerhout of the University of Maryland.) The radio structure and radio spectrum of W50 have led most workers to suggest that it is an old supernova remnant: the diffuse, expanding remains of an ancient stellar explosion, similar to the Crab Nebula. Since there is no historical record indicating the observation of a supernova in that part of the sky, however, it has been difficult to obtain absolute proof of this assertion of the nature of W50. If indeed it is a supernova remnant, its size and radio brightness, when compared with analogous data for better-studied remnants, suggest that



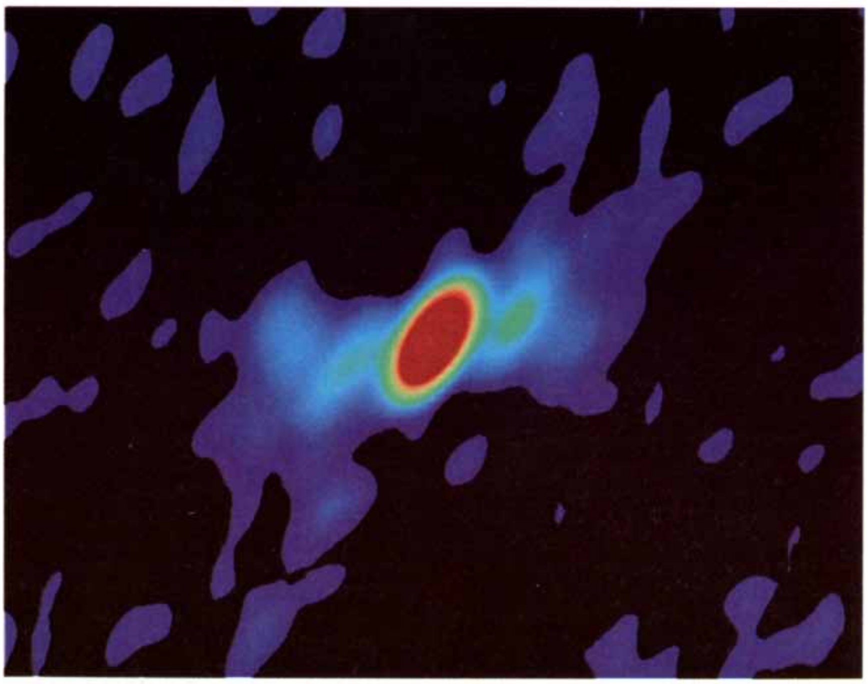
SS 433 IS QUITE INCONSPICUOUS in conventional astronomical photographs, such as this one made by Eugene A. Harlan with the 36-inch refracting telescope at the Lick Observatory of the University of California. The image of SS 433 is the black dot at the precise center of this circular photographic negative, which encompasses a field of view approximately a tenth of a degree across in the constellation Aquila. Classified in terms of its apparent brightness as a 14th-magnitude object, SS 433 is 1,000 times too faint to be seen with the unaided eye. The obvious lack of any visible features distinguishing the object from its many neighboring stars in this densely populated region of the Milky Way helps to explain why it was long overlooked.



FIRST CLUE to the unusual nature of SS 433 was discovered on this comparatively wide-field spectroscopic plate obtained more than 20 years ago in the course of a special stellar survey conducted by astronomers at the Warner and Swasey Observatory of Case-Western Reserve University. The plate was made by means of the objective-prism technique, which has the effect of smearing stellar images into short trails representing the visible spectrum of the starlight. In this case the exposure was limited by the use of appropriate filters to the red part of the spectrum, including wavelengths in the range between 6,000 and 6,800 angstrom units. In the part of the sky covered by this particular plate (an area measuring roughly half a degree across, or about the diameter of the full moon) only one spectral image shows up as a line rather than a smear; it is the inconspicuous dash near the center of the plate. Most of the red light from this object is concentrated in the extremely strong hydrogen-alpha emission line, rather than being spread continuously across all the wavelengths in this range, as in the case of the neighboring stars. In 1977 C. Bruce Stephenson and Nicholas Sanduleak of Case Western Reserve published a list of such bright emission-line objects in this part of the sky. The object at the center of the plate was the 433rd entry on the Stephenson-Sanduleak list, and hence it is designated SS 433.



RADIO MAP of a large area surrounding SS 433 reveals that it is embedded in an extended source of radio emission, known as W50, that is believed to be the remnant of an ancient supernova explosion. The map, which covers a region of the sky approximately two degrees across, is color-coded so that red represents the most intense radio emission and blue the least intense. The bright red spot near the center coincides with SS 433. The similar spot to the northeast (*upper left*) is probably an extragalactic compact radio source, seen through the fringe of W50. The extended radio source to the northwest (*upper right*) is an unrelated cloud of ionized hydrogen between W50 and the solar system. The data for the map were recorded at a wavelength of 11 centimeters by B. J. Geldzahler, T. Pauls and C. J. Salter, working with the 100-meter radio telescope of Max Planck Institute for Radio Astronomy at Effelsberg in West Germany.



HIGH-RESOLUTION RADIO MAP of the area in the immediate vicinity of SS 433 was made at a wavelength of six centimeters with the Very Large Array, a complex of radio telescopes near Socorro, N.Mex. The two elongated structures emanating from the central point source in the map are aligned with the bulges in the surrounding radio source, W50. The map was produced in a collaborative study by John T. Stocke of the University of Arizona and Ernest R. Seaquist and William S. Gilmore of the University of Toronto. The color representation of the map was prepared by Eric W. Greisen of the National Radio Astronomy Observatory. The ellipticity of central source in this representation is an artifact of the imaging process.

the explosion occurred some 10,000 years ago. Although such an age is clearly too great for the explosion to have been documented by human observers, it would still qualify the event that created W50 as one of the more recent supernovas.

A series of radio observations designed to map the detailed structure of W50 were published in 1975 by David H. Clark and his colleagues at the University of Sydney. The map also shows the bright, pointlike radio source now known to be SS 433, surrounded by the diffuse radio emission from W50. Now yet another chance occurrence intervened to inhibit further study of this curious configuration. Although the actual radio structure of W50 is quite symmetrical, with SS 433 close to the center, the map published by Clark and his colleagues showed only the northern half of the remnant, thereby somewhat downplaying the prominent central location of the mysterious pointlike source. If other observers had appreciated the striking symmetry of the remnant and therefore had recognized that the unidentified point radio source was quite precisely centered, further observation would almost surely have been stimulated.

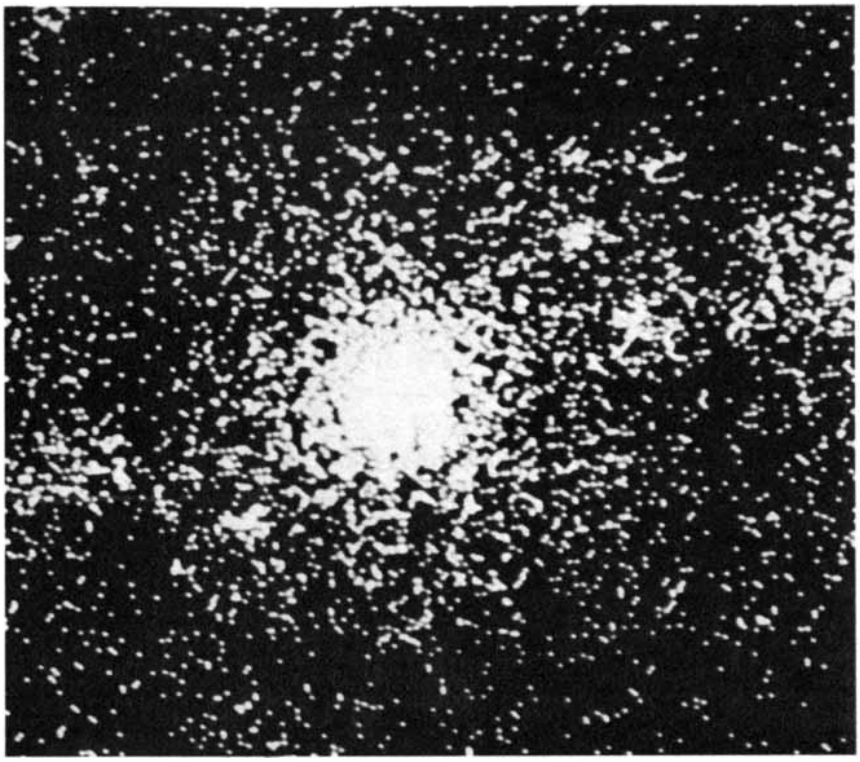
Although supernovas are widely believed to often (perhaps always) leave behind a collapsed, exotic stellar remnant such as a neutron star or a black hole, there are only two unambiguous cases where a collapsed star has actually been found in a supernova remnant. The radio and visible remnants of both the Crab event and one in the constellation Vela also harbor a pulsar, pointing to the existence of a rotating neutron star. The fact that only two such coincidences are known, in spite of the existence of dozens of radio-emitting and light-emitting supernova remnants and hundreds of radio pulsars, is annoying, although probably not profound. The opportunity to explore another potential coincidence of this type, if it had been recognized, would surely not have been missed.

The final preface to the recognition of the strange properties of SS 433 involves its X-ray emission. As is the case with radio emission, very few normal stars are a detectable source of X rays, although here again many supernova remnants are found to be X-ray sources. In the early 1970's two earth-orbiting satellites independently recorded X-ray emission from the vicinity of Aquila, and once again the source of these emissions, now known to be SS 433, was duly named and catalogued. Observers working with data from the American satellite *Uhuru* designated the X-ray source 4U1908 + 05 (from the fourth *Uhuru* catalogue, with the rough celestial coordinates of the position in the sky). Meanwhile a largely British group (led by a visiting American, Frederick D. Seward), working with data from the British satellite *Ariel V,* designated the object A1909 + 04.

The poor spatial resolution of the experiments prevented either group from perceiving the coincidence of the X-ray source and the bright star. Seward and his co-workers did note, however, that the X-ray intensity seemed to be changing with time, a characteristic not seen in the X-ray emission from supernova remnants because of their comparatively slow evolution after the initial explosion. Seward and his colleagues presciently commented in their 1976 publication that the X-ray source was probably not simply the remnant W50 but perhaps something more exotic related to it.

A synthesis of these numerous clues to the unusual characteristics of SS 433 finally emerged in the summer of 1978, as the result of the contemporaneous but largely independent efforts of three separate research groups. A group of Canadian radio astronomers led by Ernest R. Seaquist of the University of Toronto searched for emissions from young stars by conducting a new survey of objects in the Stephenson-Sanduleak catalogue. Although this led to still another rediscovery of the radio source, Seaquist and his colleagues correctly realized the source was associated with the bright stellar object SS 433. Meanwhile a group of British and Australian radio astronomers led by Sir Martin Ryle of Cambridge were tackling the quite separate problem of the observed scarcity of pointlike radio sources inside extended supernova remnants. In a sensitive survey aimed at finding and refining the positions of such objects, they again found the radio source in W50, and they also noted the coincidence with the 14th-magnitude optical object.

Finally, Clark and his colleague Paul Murdin decided to obtain spectrograms of visible stars whose position was close to the rather poorly located radio source Clark had previously noted in his map of W50. Working with the Anglo-Australian Telescope in Australia, they recorded the spectrum of SS 433 in June, 1978, the first reported spectroscopic observations since the crude objective-prism plates in the Stephenson-Sanduleak catalogue had been made. Clark and Murdin did not immediately realize that they were observing a previously catalogued object. Nevertheless, the spectrogram they obtained left little doubt that they had properly located the visible counterpart of the strange radio source. Their observations showed emission lines with an intensity found only in the most unusual stars. An extremely accurate radio position provided by Ryle's group made the identification of SS 433 with the radio source conclusive; the positions of the radio and the optical objects coincide precisely. A final, irrevocable link with the X-ray source was forged in a recent series of extremely precise X-ray measurements made by Seaquist and his group with the earth-orbiting satellite *HEAO-2,* also called the Einstein X-ray Observatory; the X-ray position is now also known to agree perfectly with the radio and optical ones.



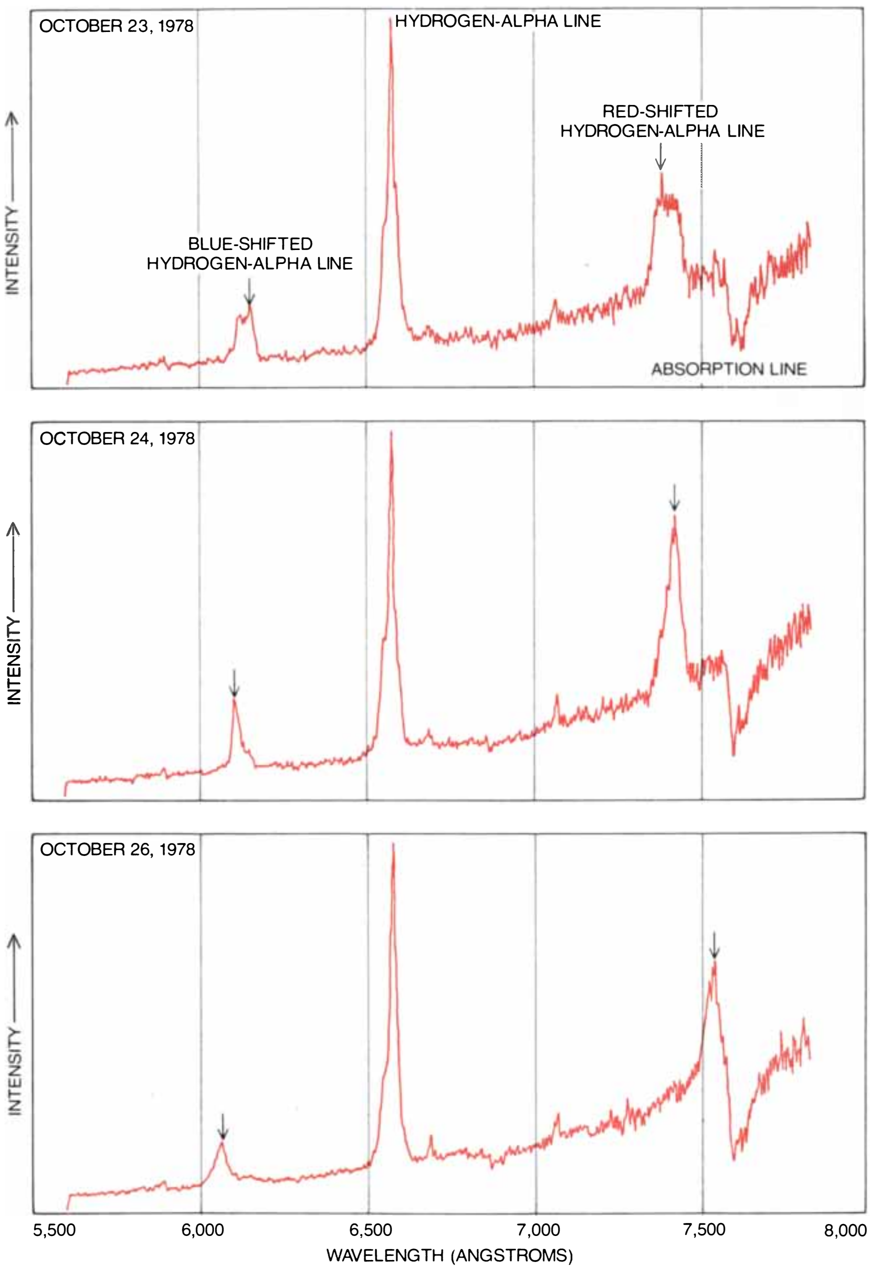
STRONG X-RAY SOURCE coincident with the position of SS 433 appears at the center of this computer-generated picture made with data obtained by an X-ray telescope aboard the satellite *HEAO-2,* also known as the Einstein X-ray Observatory. The faint evidence of X-ray emission extending outward from the central source on both sides suggests the presence of jets of hot gas streaming toward the east and west extremities of W50. This X-ray picture, the product of an exposure of five hours, was made by Seaquist and Gilmore with Jonathan E. Grindlay and Frederick D. Seward of the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory. Field of view is approximately one degree.

Clark and Murdin published their observations in a brief note in *Nature* in the autumn of 1978. They identified the prominent emission lines as having wavelengths appropriate for excited hydrogen and helium atoms (which was not in itself surprising, since these are invariably the most abundant chemical elements in stars). They also alluded indirectly to weaker emission lines of uncertain origin in the spectrum. (These briefly mentioned features later proved to hold the key to the entire mystery of SS 433.) Finally they stressed the striking triple coincidence of a visible emission-line star, an X-ray source and a radio source, all centered in a supernova remnant, and they suggested that SS 433 and W50 might be causally related.

At this point my own involvement with SS 433 began. I had long been interested in the optical characteristics of the faint visible counterparts of celestial X-ray sources. Most of the X-ray stars in our galaxy prove to be in binary systems consisting of a compact object such as a white dwarf or a neutron star bound in a close orbit with a comparatively normal star, often not too different from the sun. The gravitational attraction exerted by the compact star on its companion frequently causes that otherwise normal star to lose some of its mass by transferring it to the compact object. In the transfer process the streaming matter often reaches temperatures and densities sufficient to cause copious X-ray emission. Spectroscopic studies of the normal star in such systems often yield valuable data on the characteristics of both objects in them. The enormously high densities of matter in compact objects such as neutron stars cannot be attained in laboratories on the earth, and so astrophysical data provide virtually the only direct information on the behavior of matter in these exotic states.

I acquired my first spectrogram of SS 433 in September, 1978, working with the three-meter Shane telescope at the Lick Observatory of the University of California. My primary intent was simply to confirm the results of Clark and Murdin. The electronically recorded data revealed the strong emission lines of hydrogen and helium noted by those workers and also by Stephenson and Sanduleak. To my surprise, however, also present in the spectrum were very prominent emission lines not familiar to me. This was a disquieting state of affairs. Which spectral lines appear in a stellar spectrum depends on the abundance, the temperature and the density of the individual elements in the star. In astrophysical situations there is some variety in these parameters, but it is not infinite; accordingly the stellar spectroscopist gets accustomed to the appearance of certain familiar spectral lines. To encounter spectral emission lines at completely miscellaneous wavelengths is an experience somewhat akin to a driver's suddenly finding that his familiar homeward-bound freeway has all new exit ramps.

The strengths of the unidentified emission lines in the spectrum of SS 433 were particularly interesting. These lines are only slightly less prominent than the familiar hydrogen lines that also appear in the spectrum. One might therefore attribute them to a chemical element with a cosmic abundance comparable to the abundance of hydrogen. Since 90 percent of all the atoms in most stars are thought to be those of hydrogen, however, there is no such comparably abundant element. Could the unidentified lines also be due to hydrogen, and could they for some reason be displaced from the normal wavelengths seen in all other laboratory and astrophysical situations?

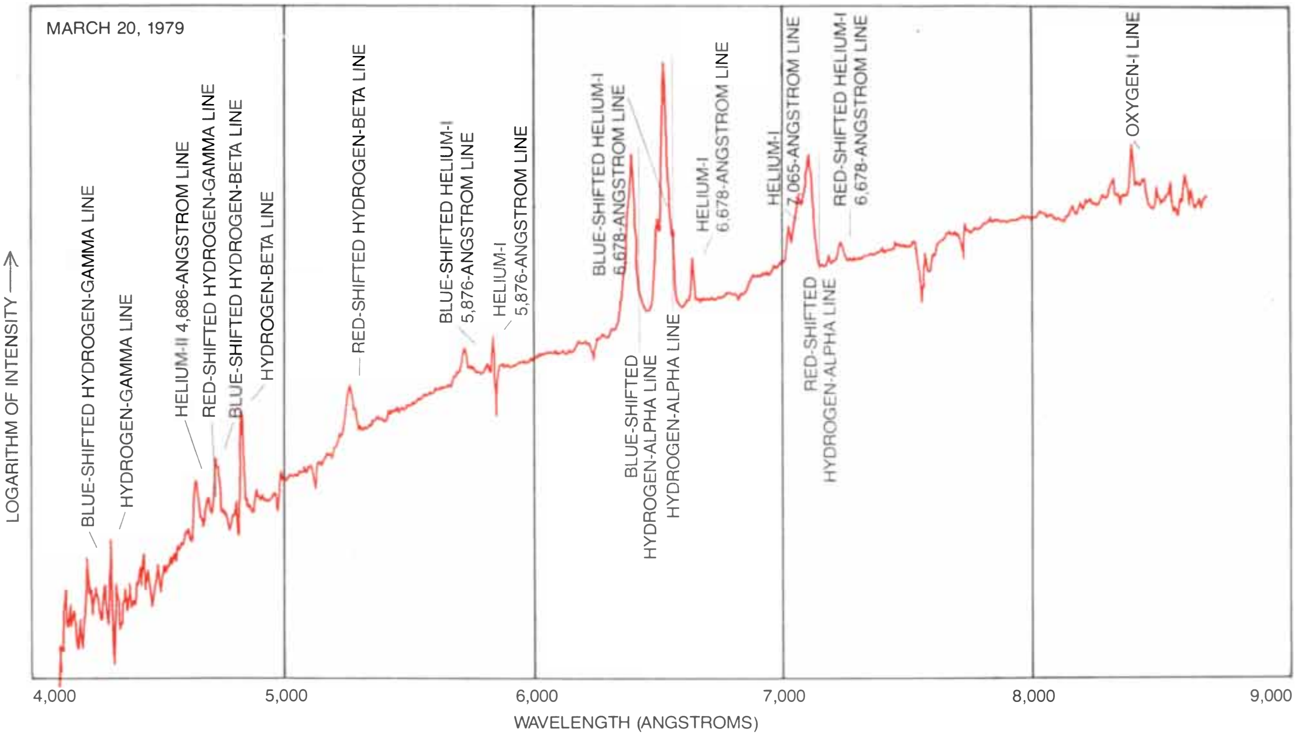


SPECTRUM OF SS 433 was recorded on three different nights in a four-day period by Remington P. S. Stone, working with the 24-inch reflecting telescope at the Lick Observatory. The most prominent feature, the peak at a wavelength of 6,563 angstroms, corresponds to the extremely strong hydrogen-alpha emission line. Much weaker emission lines attributable to helium are detectable at 5,876, 6,678 and 7,065 angstroms. Two very strong emission features, now known to be Doppler-shifted versions of the central hydrogen-alpha line, can be seen flanking the central line, one toward the shorter-wavelength (blue) end of the spectrum (*left*) and the other toward the longer-wavelength (red) end (*right*). In the course of the three nights the red-shifted line obviously moved farther toward the red and the blue-shifted line moved farther toward the blue. The large dip in the curves near a wavelength of 7,600 angstroms is unrelated to SS 433; it is an absorption line caused by molecules in the earth's atmosphere.

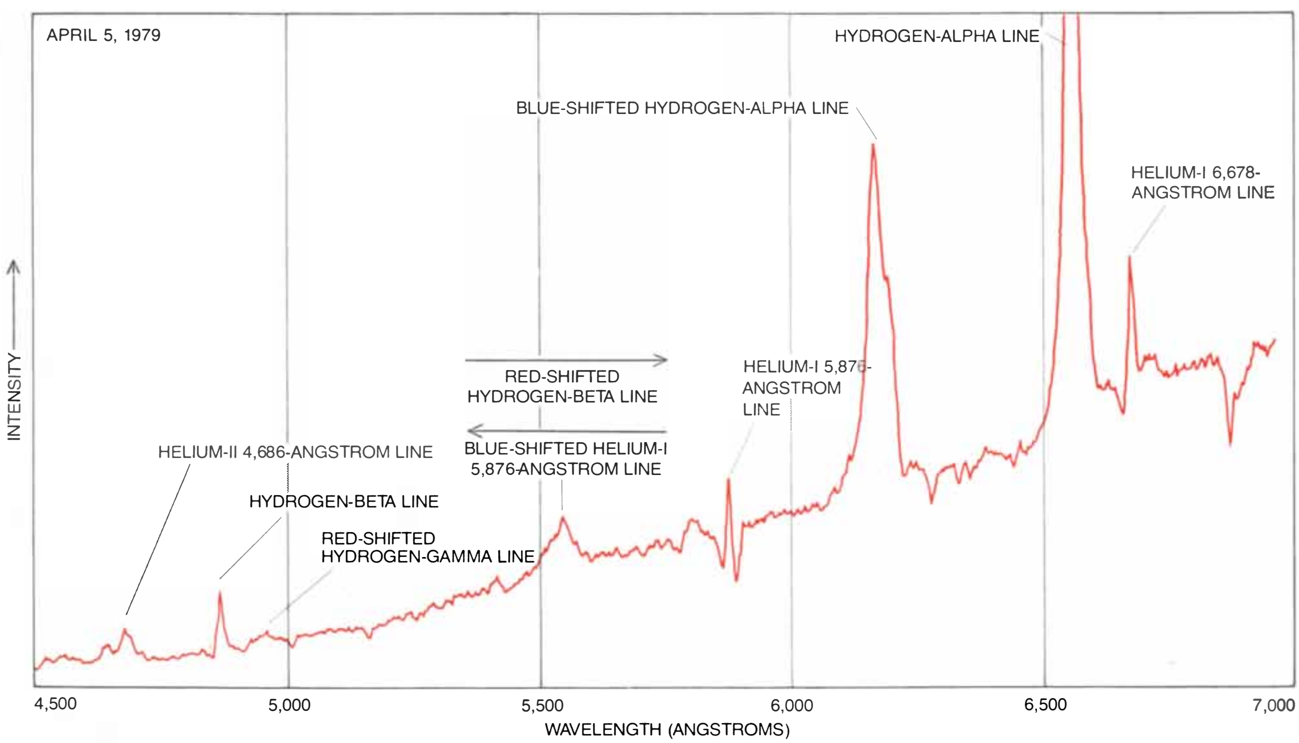
There is one such displacement mechanism familiar to the spectroscopist, namely the Doppler effect. Relative motion of the source with respect to the observer is known to slightly displace the perceived wavelength of any wave phenomenon, such as sound or light. Furthermore, the magnitude of the displacement conveniently reflects the velocity of the motion, and the sense of the displacement (toward longer or shorter wavelengths) indicates whether the object is receding or approaching. Most stars have a random motion of a few tens of kilometers per second with respect to the sun, causing a spectral-line displacement of about .01 percent from the standard values. With some binary stars the periodic orbital motion can be 10 times greater than that, causing a proportionately larger Doppler displacement.

The unidentified lines in SS 433, however, were at wavelengths not at all close to any hydrogen lines; thus if they were Doppler-displaced hydrogen emissions, the velocities implied would have to be prodigious. For the most prominent unidentified line in these spectrograms, located in the red part of the spectrum near a wavelength of 7,400 angstrom units, the velocity needed to give rise to the necessary displacement from the nearest hydrogen line (at 6,563 angstroms) is about 40,000 kilometers per second, or more than 10 percent of the speed of light! Because the escape velocity from the galaxy is only a few hundred kilometers per second stellar velocities higher than that are never encountered; any such object would quickly (on an astronomical time scale) leave the galaxy entirely. In short, Doppler-shifted hydrogen emission seemed a poor explanation of the observations.

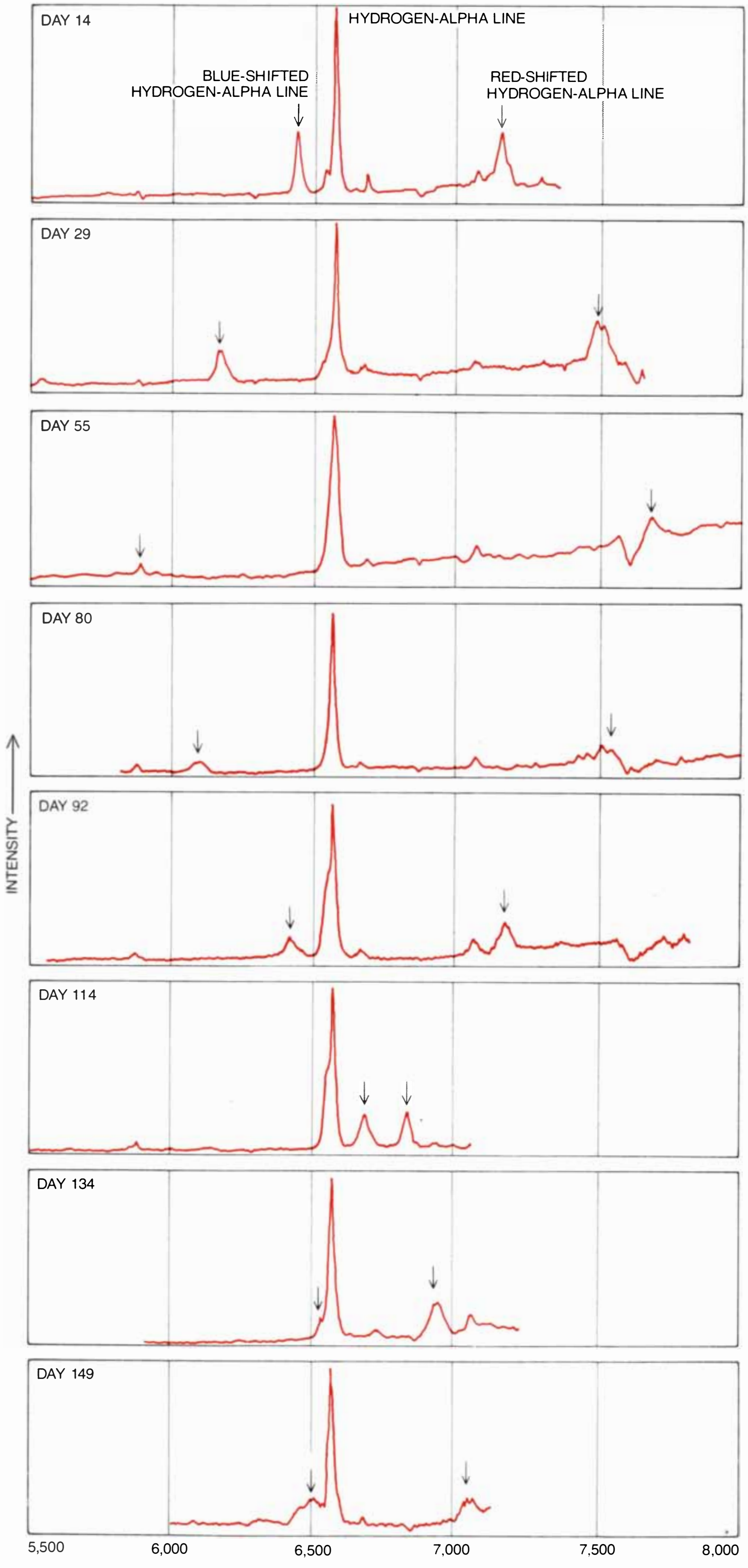
An even more bizarre characteristic of the spectrum of SS 433 became apparent after several nights of repeated study of the object. The unidentified features at unfamiliar wavelengths were seen to change wavelength, and by very substantial amounts. For example, in one four-night period the strong feature in the red part of the spectrum increased its wavelength by more than 1 percent. If the Doppler effect were responsible,



TRIPLED EMISSION LINES in this particular spectrogram of SS 433 are what convinced the author and his colleagues at the University of California at Los Angeles that the "moving" emission lines observed in various spectra of this object must be attributable to the Doppler effect. The spectrogram was obtained by Steven A. Grandi of U.C.L.A. on the night of March 20, 1979, with the three-meter (120-inch) Shane reflector at the Lick Observatory. A number of prominent hydrogen and helium emission lines in the spectrum are each represented three times: by a strong central component and by two flanking Doppler-shifted components, one of which is red-shifted and the other blue-shifted. The strong absorption features that appear in the spectrum are again unrelated to the stellar object.



CONFUSING COINCIDENCE is sometimes presented by the large number of moving Doppler-shifted emission lines found in spectra of SS 433. In this case, for example, several prominent stationary and moving emission lines are identified, among them the comparatively broad peak at the center (near 5,500 angstroms), which consists equally of a blue-shifted helium line moving to the left and a red-shifted hydrogen line moving to the right. On the night this spectrum was obtained (April 5, 1979) the two lines just happened to be passing each other in opposite directions. The spectrum was recorded at the Lick Observatory by Lawrence H. Aller and Charles D. Keyes of U.C.L.A.



this seemingly small discrepancy would need to be explained by a change in velocity of nearly 5,000 kilometers per second in those four days. Furthermore, the wavelength changes of the unidentified lines were not even consistent: some of the features moved to longer wavelengths and others simultaneously moved toward shorter wavelengths.

At this point it became clear that the spectral behavior of SS 433 was considerably more exotic than the previous observations had implied. My colleagues and I at the University of California at Los Angeles therefore initiated a program to obtain at least a brief spectral observation of SS 433 on every possible night. In spite of the severe shortage of observing time on large research telescopes, this was feasible because of the comparative brightness of the object. For example, the sensitive computer-controlled spectroscopic instrumentation of the Lick three-meter reflector is designed for the observation of very distant galaxies and quasars, objects hundreds of times fainter than SS 433. A good-quality spectrogram of this star can be had in about 10 minutes from telescopes in this class, and spectrograms can therefore be made frequently without disrupting previously planned observing programs.

Observers from all four University of California campuses where research in optical astronomy is conducted—Los Angeles, Berkeley, Santa Cruz and San Diego—participated in the observations. More than a dozen astronomers, whose primary research interests ranged from distant galaxies to nearby normal stars, generously gave of their scarce observing time to help monitor the object. Deserving of special mention are three observers who obtained as many spectrograms as I did: Steven A. Grandi and Holland C. Ford of U.C.L.A. and Remington P.S. Stone of the Lick staff.

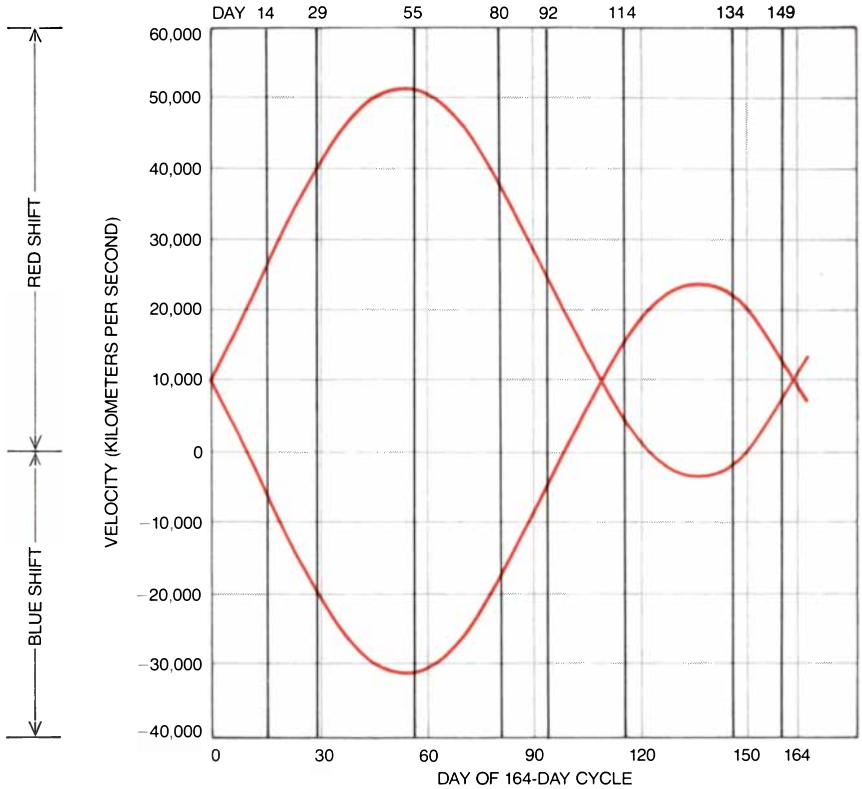
Our consortium had little time to lose, because from December through February of each year the line of sight to SS 433 is too close to the sun for nighttime observations. By the end of the 1978 observing season we had watched the "moving" spectral lines, as we came to call them, traverse a staggering range of wavelengths. For example, the reddest emission feature changed its wavelength by about 700 angstroms in 30 days, which, if it was interpreted as a gradually increasing Doppler shift, would im-

SELECTED SPECTRA of SS 433, obtained by the author and his colleagues over a period of 164 days, cover the entire cycle of the motion of the Doppler-shifted emission lines. The days the observations were made are given in relation to a model of the 164-day cycle in the illustration on the opposite page.

ply a steady increase in velocity from 20,000 kilometers per second to 50,000.

In our initial reports of the data to the scientific community we tried to remain skeptical about the Doppler-shift interpretation of the moving lines for a variety of reasons that all seemed valid at the time. The most obvious objections were those I have already mentioned. Both the velocities implied and the changes in the velocities were larger by a factor of 100 than comparable values found in any other stellar object. There were also subtler difficulties with the Doppler-shift explanation. Some of the moving lines shifted toward longer wavelengths and others shifted toward shorter ones, defying the simple interpretation of a single cloud of gas either approaching or receding.

Furthermore, the identification of the moving emission lines as Doppler-shifted hydrogen lines would imply the existence of a gas at a rather modest temperature, about 20,000 degrees Kelvin or less; at temperatures higher than that most of the hydrogen would be ionized, and spectral lines due to the transitions of an electron bound to an atomic nucleus could not appear. Yet almost any mechanism one can imagine that could accelerate a gas to the enormous velocities implied (a substantial fraction of the speed of light) would heat the gas to a far higher temperature. Another way to put this difficulty is to note that if the thermal and the kinetic energy of the hydrogen nuclei in the emitting gas are roughly equal, as is often the case in a variety of physical systems, the observed velocities imply temperatures of more than 30 billion degrees K., far higher than the temperature inferred for the gas observed in SS 433.



THEORETICAL CURVES trace out one complete 164-day cycle in the predicted pattern of red-shifted and blue-shifted spectral emission lines for SS 433, on the assumption that the light-emitting gas is concentrated in two oppositely directed rotating jets, each with an ejection velocity of 78,000 kilometers per second, or about a fourth the speed of light. The black lines indicate the days on which the selected spectra that appear on the opposite page were obtained. The emission features in the spectra match the predicted red shifts and blue shifts quite closely. The curves exhibit a constant average red shift equivalent to an ejection velocity of 12,000 kilometers per second, which results from the effect of special relativity called time dilation.

Our concerns did not inhibit a host of imaginative theorists. Andrew Fabian and Martin Rees of Cambridge pointed out that the existence of a gas Doppler-shifted in both directions could be understood if a central object were ejecting two jets of gas in roughly opposite directions. Then some gas would be approaching the observer and other gas would be receding. Fabian and Rees noted the existence of similar double-lobed structures seen on a vastly larger scale in radio galaxies, where the radio emission is often found to be confined to two opposed jets. A similar scheme was proposed independently by Mordechai Milgrom of the Weizmann Institute of Science in Israel, who went even further to make a guess that later proved to be spectacularly successful. On the basis of only a handful of our data points Milgrom speculated that the line motions might be periodic, with a repetition time of about a few months.

Meanwhile we continued to worry about the entire basic idea. Were the unidentified moving emission lines actually Doppler-shifted hydrogen lines? The answer became apparent in March of last year, when SS 433 was again far enough from the sun for spectroscopy. Our first spectrogram of the new observing season clearly showed all of the principal emission lines to be tripled: one component at the laboratory (undisplaced) wavelength, one at a wavelength displaced toward the red end of the spectrum (toward the longer wavelengths) and one displaced toward the blue end (toward the shorter wavelengths). Both hydrogen and helium emission lines, about half a dozen different features in all, exhibited this peculiar triple pattern. Furthermore, each of the red-shifted lines independently implied an identical velocity of recession, about 27,000 kilometers per second, and each of the blue-shifted features similarly implied an identical velocity of approach, about 6,000 kilometers per second. This multiple set of coincidences could be explained only by the Doppler shift. A similar conclusion was reached independently at about the same time by a group of astronomers at the University of Arizona directed by James W. Liebert.

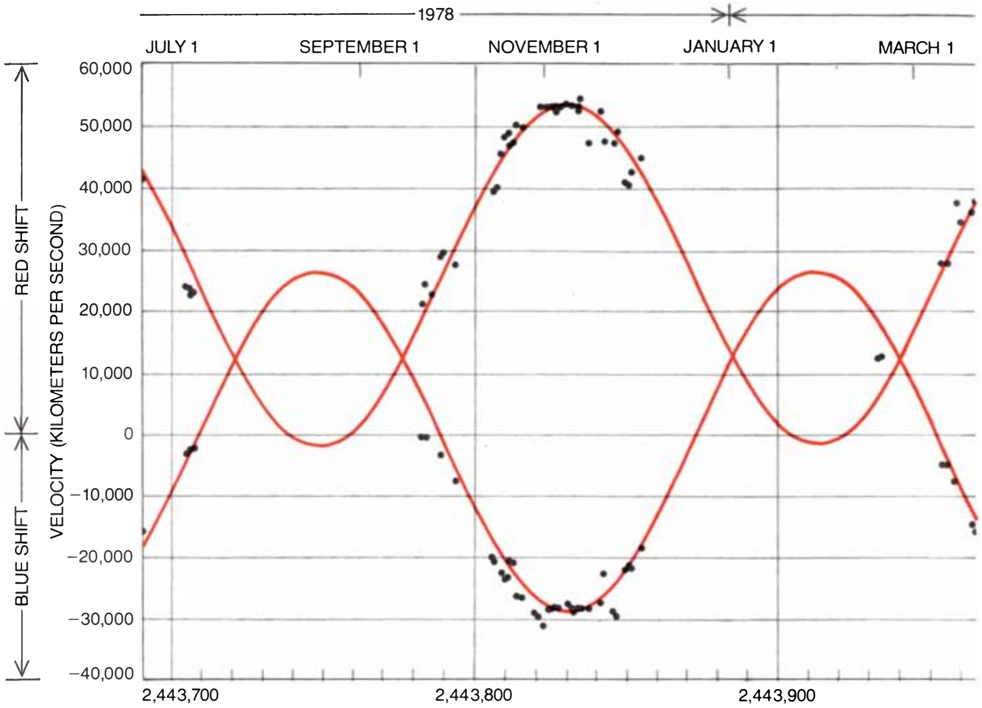
In a curious way the difficulty in interpreting the moving lines in SS 433 was an eerie repetition of the sudden recognition of the meaning of quasar spectra by Maarten Schmidt two decades ago. The huge red shifts of the quasars, caused by the expansion of the universe, also displaced the emission lines so much from their normal wavelengths that it was hard to identify the atoms that gave rise to them, in spite of the fact that the lines turned out to be associated with the most abundant and familiar elements in the universe. Even though the history of the discovery of quasars was well known to all of us, SS 433 presented an initial interpretive problem because of two unprecedented complications: first, such enormous red shifts had never been associated with stars within our galaxy, and second, the object was evidently showing blue shifts as well as red shifts.

It is clear that since the wavelength of the Doppler-shifted features changes rapidly with time, the velocity of both the approaching gas and the receding gas is changing. An obvious question then is: Is there any regular pattern to these changes? In our two years of observations we have accumulated some 150 separate measurements of both red-shifted and blue-shifted emission lines; a graph of the values of the implied velocities as a function of time reveals several fascinating features in the strange behavior of the moving lines [*see illustration on these two pages*]. One sees immediately that the velocities are truly enormous: the red-shifted (receding) gas reaches values up to 50,000 kilometers per second (16 percent of the speed of light) on several different occasions, and the blue-shifted (approaching) gas reaches velocities of up to 30,000 kilometers per second. Because the velocities of stars in our galaxy (either approaching or receding) never exceed a few hundred kilometers per second, and because all extragalactic objects beyond the immediate vicinity of our galaxy show only red shifts, attributable to the expansion of the universe, SS 433 has the distinction of exhibiting the largest blue shift (by a factor of 100) of any known celestial object, galactic or extragalactic.

More pieces of the puzzle fell into place after an analysis of the pattern of velocity changes. The approaching and receding volumes of radiating gas are certainly not coincident in space; if they were, they would separate rapidly at these enormous oppositely directed velocities. Yet in spite of their physical separation the two clouds of gas are definitely related; the variations in the red-shifted and blue-shifted systems reach their extremes of velocity at identical times. Moreover, the average value of the two velocities on any given night is roughly constant but very large: about 12,000 kilometers per second.

The constancy of the mean of the two Doppler-shifted velocities, in spite of the huge change in their individual values on a time scale of days, could perhaps be understood if one central object were responsible for ejecting both radiating clouds; the average velocity would then be that of the central star. Once again, however, we were confronted with the problem that this value exceeds by a wide margin the escape velocity from our galaxy; the object would depart from the galaxy forever in a tiny fraction of the age of the stars in the neighborhood of the sun (at least 10 billion years). Is it reasonable to believe that one just happens to be alive and doing astronomy during this brief interval? As it turns out, a considerably less strained explanation is available.

I have yet to address perhaps the most startling characteristic of the variation in the velocities of the material associated with SS 433. In spite of the large and continuous variations in velocity, and the considerable gaps in the observations, it can be seen from an examination of the illustration on these two pages that both the red-shifted gas and the blue-shifted gas repeatedly return to the same velocity values approximately once every six months; in other words, the variations in velocity are periodic. A simple analysis of the current



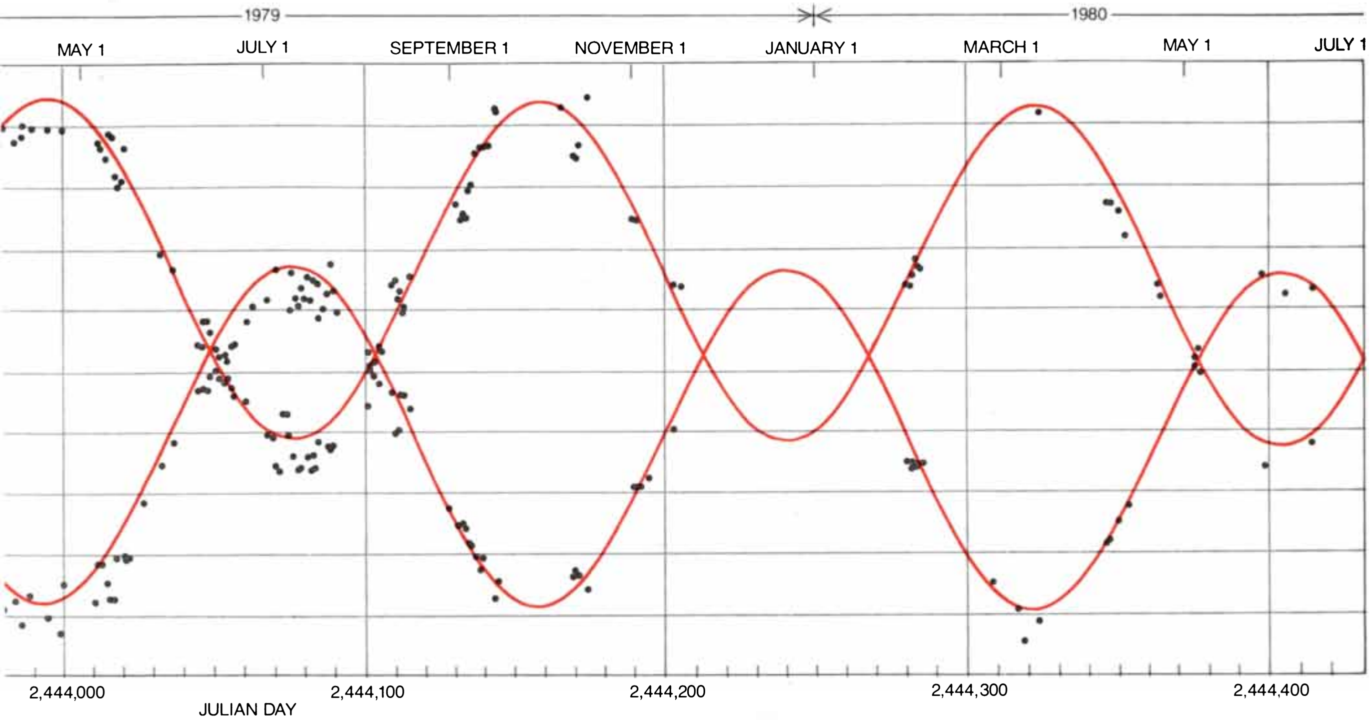
TWO YEARS OF OBSERVATIONS are summarized in this illustration, which plots the values of the red- and blue-shifted emission lines observed in the spectrum of SS 433 from mid-1978 to mid-1980 in terms of the equivalent velocity of the ejected gas. The large gaps in the

data shows the exact period to be 164 days, with an uncertainty of about half a day. The pattern of the variations is quite distinctive. Twice each 164 days the two emitting systems merge to the identical velocity and then change places, that is, the blue-shifted gas becomes red-shifted and vice versa. Observationally these events are seen in the spectrum as a gradual merging of the moving lines, which then pass through each other, drawing apart in opposite directions. We have observed this "crossover" event several times, although we have not yet been lucky enough to obtain a spectrogram during the precise (presumably brief) interval when the moving lines are exactly superposed.

There is a familiar precedent for this type of periodic spectral variability, albeit one on a greatly reduced scale. In certain binary star systems if the brightness of the two stars is comparable, two independent sets of spectral lines are sometimes visible. Because at any instant one star has some component of velocity toward the solar system and the other star has some component of velocity away from it, the Doppler shifts of the same spectral line in each star are slightly different; hence the lines appear as a resolvable pair at slightly different wavelengths. As the two stars revolve around their common center of mass the component of motion of each star toward or away from the earth changes smoothly and periodically; accordingly the wavelengths of each of the two spectral lines evince a periodic change between a short-wavelength limit and a long-wavelength one. In binary star systems, however, the amplitude of the change in velocity never exceeds a few hundred kilometers per second, whereas in SS 433 the amplitude is more than 100 times greater.

Could SS 433 represent some enormously scaled-up version of this phenomenon, consisting simply of two mutually orbiting objects? If this were the case, there would be a fascinating consequence. One corollary of the simplicity of the gravitational force is that all orbiting bodies, regardless of their nature, obey a basic relation between the orbital period, the orbital velocity and the total mass in the system. Therefore if one assumes that the dramatic wavelength variations of SS 433 are due to orbital motion, then with only the observed amplitude and period of the variation in velocity the total amount of matter in SS 433 can be directly calculated. The result is a total mass equivalent to a billion times the mass of the sun! Since the most massive stars known have a mass of less than 100 solar masses, this result is startling. In fact, the mass inferred in this way for SS 433 is about 1 percent of the total mass of the 100 billion stars in the entire galaxy.

Is it credible that an object with such an extraordinary mass has been overlooked until now? The answer is no. There are several different arguments that can be relied on to show that the strange spectral-line variations cannot be due to orbital motion. For example,



data are correlated with times when the line of sight to the object was too close to the sun for nighttime observation; the smaller gaps are attributable to the proximity of the moon, which makes observing difficult, or to the lack of observing time on a suitable telescope. Curves show the predicted behavior, based on the assumption that the gas responsible for Doppler-shifted lines is in two oppositely directed jets.

the putative orbital parameters are such that the diameter of the orbit would be so large that it would take light some two weeks to travel between the orbiting objects. Yet the peaks and valleys in the plot of the system's red shifts and blue shifts remain synchronized with each other to an accuracy of within a day or so. According to the special theory of relativity no information, regardless of its method of transmission, can propagate faster than the speed of light. Therefore it would be impossible for the two objects to remain synchronized; they cannot, so to speak, "tell" each other where they are.

A second item of evidence arguing strongly against orbital motion is the stationary emission system in the SS 433 spectrum, that is, the set of spectral lines that are constantly close to their laboratory wavelengths. If the gas emitting these lines were in the vicinity of an object with a mass of a billion solar masses, it would rapidly feel this enormous gravitational attraction and fall in toward the larger mass. Yet the velocities indicated by these lines are quite small.

For all of these reasons, together with subtler arguments, it is not feasible to invoke two mutually orbiting bodies to explain the rapid wavelength changes in the spectrum of SS 433. Most of the workers interested in the problem have turned instead to modifications of the concept in which one central object ejects two jets of matter, one jet directed approximately toward the earth, thereby giving rise to blue-shifted emission-line radiation, and one jet directed approximately away from the earth, thereby giving rise to red-shifted radiation. If in addition it is postulated that the imaginary line joining the two jets rotates at a rate such that a complete turn is made every 164 days, the periodic modulations of the observed velocity values are also explained. That is because at each point in the 164-day cycle the angle of each jet with respect to the line of sight to the earth varies. When the jets are closest to pointing directly toward (or away from) the earth, the largest velocities of approach (and recession) will be observed. On the other hand, when the jets are directed across the line of sight, there is no gas moving either toward or away from the earth, and one would expect the velocities to be at a minimum.

It is easy to quantify this idea with a simple set of equations. Surprisingly there are only five unknowns in the equations; in other words, five characteristic parameters should be enough to completely describe the behavior of the object. To begin with there are two unknown geometric angles. One is the inclination of the axis of rotation of the object to the line of sight. For example, an extraterrestrial observer able to perceive the earth's rotation would not necessarily have to be stationed exactly above the Equator; he would see the rotation from any latitude. We do not know the "latitude" from which we are observing SS 433. The second unknown angle is the inclination of the axis of the jets to the rotation axis. Our angle with respect to the jets need not be the same as this first angle, just as the axes of the earth's geographic and magnetic poles are slightly inclined with respect to each other. A third unknown is the velocity with which the jets are ejected; for the sake of simplicity it is convenient to assume that the two jets are ejected in opposite directions but with the same speed. (If this assumption is wrong, there is no solution to the equations.) Finally, the precise values of the period and the phase of the 164-day cycle of SS 433 are both unknowns that must be determined from the data.

Our spectroscopic observations can be fitted to this theoretical model to determine if any values of the five parameters can be found that seem to agree with the data. It turns out that this simple concept, perhaps surprisingly, fits the data quite well. The two angles prove to be respectively about 80 and 20 degrees, and the jet velocity implied by the observations is 78,000 kilometers per second, or 26 percent of the speed of light. The reason we never directly observe a red shift or a blue shift quite as large as this value is simply that the two geometric angles turn out not to be right angles. Thus the jets do not point directly toward or away from the earth. (Indeed, it would be suspicious if they did.) The maximum velocity component of the gas moving toward or away from the earth therefore cannot reach the actual velocity inferred for the jets. An observer fortuitously located along the projected path of the jets would see this maximum and minimum velocity.

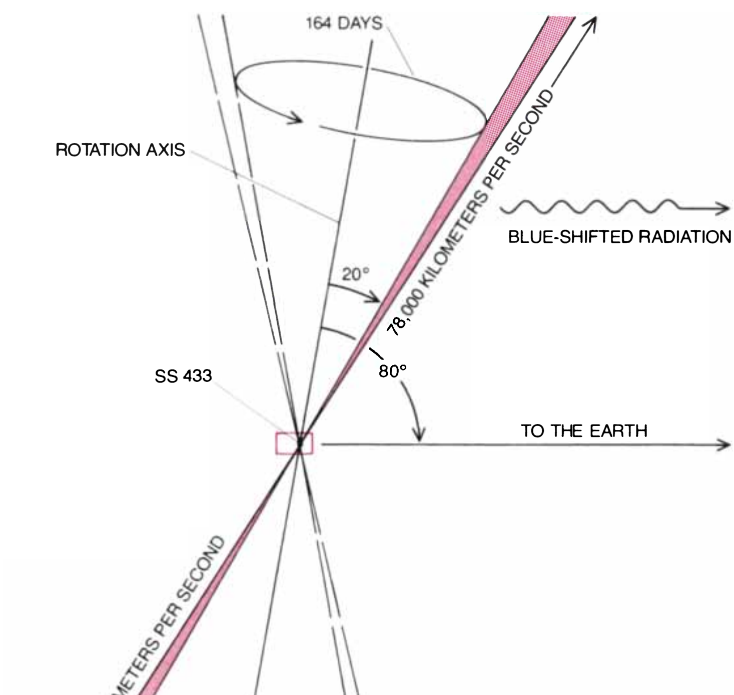
For matter moving at a significant fraction of the speed of light, effects predicted by the special theory of relativity become important. In fact, it is just such an effect that explains one of the more puzzling aspects of SS 433: the very large and constant average velocity (12,000 kilometers per second) of the two jets. The relevant effect has several different technical names, such as the second-order or transverse Doppler shift, but it is often referred to simply as time dilation. An imaginary observer watching someone moving at high velocity and carrying a clock would perceive the clock to be running slow; the higher the velocity, the slower the clock. The person carrying the clock, on the other hand, would perceive the clock to be keeping perfect time. Indeed, since all motion is relative, he might accuse the "stationary" observer of faulty time perception.

What does this have to do with SS 433? Each atom in the ejected gas has the equivalent of a clock, since it must keep track of the frequency (or equivalently of the wavelength) in order to emit light at the proper wavelength whenever its electrons are de-excited. Since time dilation always slows the clock, and since the frequency and the wavelength of light are inversely proportional, the decrease in frequency will appear as an increase in wavelength, in other words as a red shift. The amount of the time-dilation red shift depends only on the velocity of the clock, and it is easy to calculate that at a velocity of about a quarter of the speed of light the effect of time dilation is 4 percent.

This may seem a small discrepancy, but remember that a red shift of 4 percent of the speed of light (300,000 kilometers per second) is 12,000 kilometers per second. This, of course, is exactly the observed average value of the two beams of SS 433. The time-dilation red shift is always present in the jets, regardless of their angle with respect to the line of sight. Therefore the 164-day, periodically varying Doppler shift is superposed on (adds to and subtracts from) the time-dilation red shift. This interpretation nicely explains not only why 12,000 kilometers per second is the average velocity actually observed but also why the emission lines associated with the two jets merge twice every 164 days at that large value rather than at zero velocity. The crossover events occur when the two jets are pointing at right angles to the line of sight. Even though at those times there is no approach or recession of the emitting gas, and thus no red shift or blue shift other than the time-dilation one, the ever present effect of time dilation gives both beams a red shift equivalent to 12,000 kilometers per second.

What is the mysterious central object that emits the jets? It is hard to say, but we do have one important clue, namely the observed velocity: 26 percent of the speed of light. Why is it that value and not some other one? The answer may be that that velocity is quite close to the escape velocity of matter from the surface of a neutron star. Theoretical calculations show that there is a limited range of parameters over which these exotic stars can support their own mass and thus be stable. A neutron star with a mass equal to that of the sun would have a radius of only about 10 kilometers. The velocity needed to escape from the surface of a neutron star turns out to be similar to the one observed in the jets of SS 433. Perhaps this is a coincidence, but if it is not, it suggests the possibility of a self-regulating mechanism of expulsion. The gas may be accelerated up to whatever velocity is necessary to expel it permanently; then the acceleration mechanism, having done its job, need work no harder. One can also reverse the argument. If the star is not that compact, why should it generate and maintain the observed enormous expulsion velocity, if instead it could economically be rid of the material forever by imparting to it a much lower velocity?

If SS 433 does harbor a neutron star, the 164-day clock (that is, the mechanism that rotates the jet axis) is probably not simply the rotation of the star every 164 days. That is because the very small size of the star implies that at that low rate of rotation the object is not very stiff. (Technically one would say that its moment of inertia is not very large.) On the other hand, the ejected material carries away a tremendous amount of energy at a fantastic velocity. The resulting recoil given to the star would very quickly disrupt the periodic behavior unless the two masses of ejected material are exactly matched in velocity and alignment, so that their effects on the star perfectly cancel each other. It seems more likely that the 164-day period is instead a precession effect, a slow wobble of the rotation axis. The actual rotation period of the star could then be rapid, as it is with radio pulsars, where the periods are on the order of a few seconds or less. An apt analogue is a toy gyroscope, whose wheel can spin very



GEOMETRY OF THE ROTATING-JET MODEL of SS 433 is laid out in this schematic diagram. The rotation axis of the object is inclined to the line of sight to the solar system by an angle of about 80 degrees. The jets themselves are in turn inclined by about 20 degrees to the rotation axis. The period of the rotation is 164 days. The velocity of the jets works out in this model to be about 78,000 kilometers per second. The component of this velocity projected along the line of sight varies periodically as the axis of the jets rotates; hence the Doppler-shifted emission lines in the spectrum of the object fluctuate between maximum and minimum positions, depending on the inclination of the jets to the line of sight. In the position shown here, for example, both the red shifts and the blue shifts would be at a maximum; minimum red shifts and blue shifts would be observed when the jets are at right angles to the line of sight.

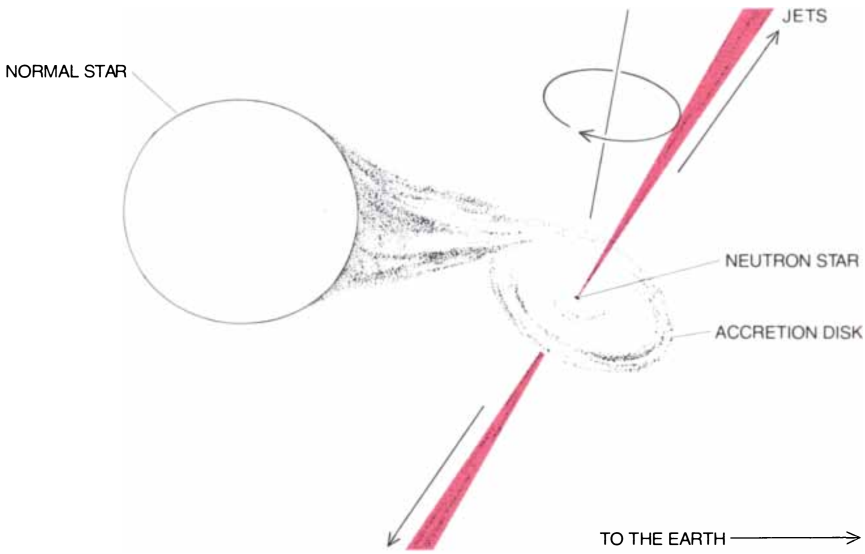
rapidly while the entire assembly turns slowly in a circle.

What is the source of the ejected material? Since the emission lines are from hydrogen and helium, elements typically found in objects far less evolved than neutron stars, the material probably does not come from such a star. The interstellar medium is largely composed of hydrogen and helium, but it is far too tenuous to supply enough gas. One is therefore led to suspect the presence of a second, less evolved star.

This line of thought led David Crampton, Anne P. Cowley and John B. Hutchings of the Dominion Astrophysical Observatory in Canada to obtain a series of spectrograms of SS 433 that might be sensitive to such a companion star. They soon discovered that the "stationary" emission lines, the emissions of hydrogen and helium at their laboratory wavelengths, are also cyclically shifting in their wavelength. The amplitude of the shift is very small, corresponding to 70 kilometers per second (about .1 percent of the amplitude of the shift of the moving lines), and therefore it could not be detected by the equipment my colleagues and I were using for our observations. The period of the minor variation is 13 days. Again the observed period and amplitude of the velocity variation lead to estimates of the masses of the two stars. A consistent, although not unique, solution is obtained if both stars are assumed to have a mass comparable to that of the sun. The parameters of the orbital solution suggest that the companion of the neutron star is probably a normal star, with characteristics not very different from those of the sun. Unfortunately such an object would be too faint against the background of the radiation from the jets to be observed directly.

Lest it seem that all the mysteries of SS 433 are now understood, I should review the host of perplexing problems remaining to be solved. Most of them center around the physical conditions in the ejected material. It is possible to calculate some of the parameters in the emitting gas; to do so, however, one must make some guess at the distance of SS 433, so that the observed intensity of the jets' radiation can be used to calculate the intrinsic luminosity. Like most estimates of astronomical distances, this one is grossly uncertain. On the basis of arguments such as the strength of spectral absorptions due to the intervening interstellar gas, I estimate the distance to be about 4,000 parsecs. (One parsec is 3.258 light-years.) Our galaxy is about 30,000 parsecs across, and so SS 433, although not an immediate neighbor of the sun, is not extremely distant.

It is now possible to calculate the luminosity of the jets. The few moving emission lines prove to have more energy, by about a factor of 10, than the sun radiates at all wavelengths. The length of the visibly radiating material is about 10 billion kilometers, that is, about 100 times the distance from the earth to the sun, or twice the size of the entire solar system. The most startling parameter, however, is the amount of kinetic energy inferred to be necessary to accelerate the considerable mass of material to the extraordinary velocity observed. Although the figure depends somewhat on uncertain assumptions, it is on the order of 10^{39} ergs per second, or a million times the energy radiated at all wavelengths every second by the sun. What is the source of this fantastic energy output? We are not at all certain.



HYPOTHETICAL CENTRAL OBJECTS postulated by the author and his colleagues to help explain the observed characteristics of the spectrum of SS 433 according to the rotating-jet model are depicted in this diagram, which can be seen as an enlargement of the area inside the small colored box in the illustration on the opposite page. The object responsible for ejecting the jets is thought to be part of a binary star system, consisting of a comparatively normal star (*left*), not unlike the sun, bound in a close orbit with a compact neutron star (*right*), which is in the process of pulling material away from its normal companion by virtue of its strong gravitational field. (Both stars are assumed to be about as massive as the sun.) The gas streaming from the normal star forms a rotating accretion disk around the neutron star, and it is from the faces of this disk that the two jets are ejected in opposite directions. Precession of the plane of the accretion disk around the neutron star is presumably what causes the axis of the jets to rotate.

The more detailed questions we ask about the jets, the more our ignorance is revealed. What mechanism harnesses this energy to accelerate the gas, yielding just one accurately controlled and unchanging velocity? Why is the gas so cool compared with the temperatures expected at these velocities? What process collimates and directs the material into a jet? The last question is particularly vexing. We now have an observational measure of the width of the jets based on the width of the moving spectral lines. If the jets were broad, at any given instant such a wide swath would present a variety of angles as well as different velocities and different Doppler shifts. We would then expect the moving lines to be broad, spanning this range of Doppler velocities. Actually we observe the opposite: the moving lines are quite narrow compared with the huge velocity of the jets. The inference is that the jets are less than a few degrees wide, almost like two sharp needles.

Finally, and perhaps most intriguing, is the question: Where are the other objects like SS 433? Why do we observe only one such object in a galaxy of 100 billion stars? A possible answer is that the lifetime of this bizarre event may be very short by astronomical standards, perhaps only on the order of 10,000 years. Many stars may pass through this phase but only for an astronomical instant. There may be only one such object active at any one time.

It is surely foolish to speculate what might be said about the general significance of SS 433 to astrophysics five or 10 years from now. Nevertheless, it is intriguing to guess. The most interesting possibility is that the resemblance of the twin-jet structure of SS 433 to the double-lobed radio emission from giant galaxies and quasars is not a coincidence. If the same basic mechanism underlies both phenomena, and if the answers to the above questions of energetics, acceleration and collimation are related in both, it would be tremendously exciting. We would then be privileged to have a revealing closeup view of a comparatively nearby object within the galaxy that could serve as a prototype for gaining an understanding of the violent extragalactic events that are among the greatest mysteries in astronomy.

Monoclonal Antibodies

Cells that secrete antibodies can be made immortal by fusing them with tumor cells and cloning the hybrids. Each clone is a long-term source of substantial quantities of a single highly specific antibody

by Cesar Milstein

When a foreign substance enters the body of a vertebrate animal or is injected into it, one aspect of the immune response is the secretion by plasma cells of antibodies: immunoglobulin molecules with combining sites that recognize the shape of particular determinants on the surface of the foreign substance, or antigen, and bind to them. The combination of antibody with antigen sets in train processes that can neutralize and eliminate the foreign substance. Quite apart from the natural function of antibodies in the immune response they have long been an important tool for investigators, who capitalize on their specificity to identify or label particular molecules or cells and to separate them from a mixture.

The antibody response to a typical antigen is highly heterogeneous. There are perhaps a million different lines of *B* lymphocytes, the precursors of plasma cells, in the spleen of a mouse or a man. All are derived from a common stem cell, but each line develops an independent capacity to make an antibody that recognizes a different antigenic determinant. When an animal is injected with an immunizing agent, it responds by making diverse antibodies directed against different antigen molecules on the injected substance and different determinants on a single antigen, and even different antibodies that fit, more or less well, a single determinant. It is next to impossible to separate the various antibodies, and so conventional antiserums contain mixtures of antibodies, and the mixtures vary from animal to animal.

Each antibody is made, however, by a different line of lymphocytes and their derived plasma cells. What if one could pluck out one such cell making a single specific antibody and grow it in culture? The single cell's progeny, or clone, would be a source of large amounts of identical antibody against a single antigenic determinant: a monoclonal antibody. Unfortunately antibody-secreting cells cannot be maintained in a culture medium.

There are malignant tumors of the immune system called myelomas, however, whose rapidly proliferating cells produce large amounts of abnormal immunoglobulins called myeloma proteins. A tumor is itself an immortal clone of cells descended from a single progenitor, and so myeloma cells can be cultured indefinitely, and all the immunoglobulins they secrete are identical in chemical structure. They are in effect monoclonal antibodies, but there is no way to know what antigen they are directed against, nor can one induce myelomas that produce antibody to a specific antigen.

In 1975 my colleagues and I learned how to fuse mouse myeloma cells with lymphocytes from the spleen of mice immunized with a particular antigen. The resulting hybrid-myeloma, or "hybridoma," cells express both the lymphocyte's property of specific-antibody production and the immortal character of the myeloma cells. Such hybrid cells can be manipulated by the techniques applicable to animal cells in permanent culture. Individual hybrid cells can be cloned, and each clone produces large amounts of identical antibody to a single antigenic determinant. The individual clones can be maintained indefinitely, and at any time samples can be grown in culture or injected into animals for the large-scale production of monoclonal antibody. Highly specific monoclonal antibodies produced by this general method have proved to be a remarkably versatile tool in many areas of biological research and clinical medicine.

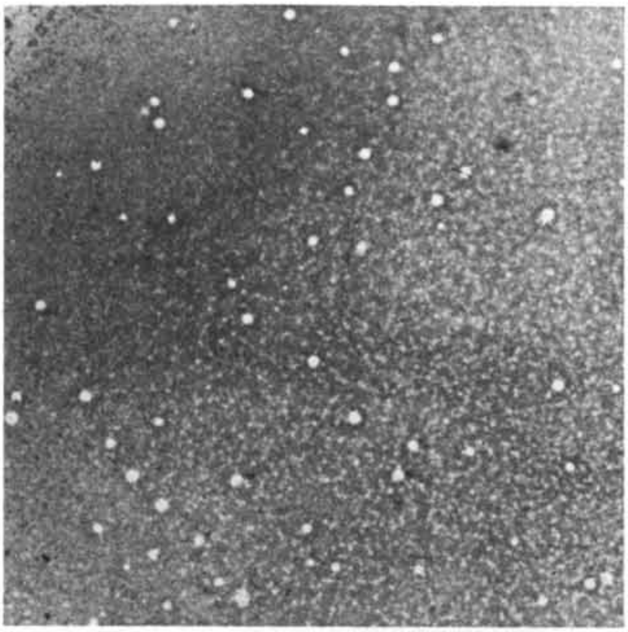
Human myelomas have been known to physicians for a long time, but it was not until the early 1960's that the precise nature of myeloma proteins was elucidated by immunologists. Michael Potter of the National Cancer Institute then induced myelomas in mice, and these too produced large amounts of monoclonal immunoglobulins. In spite of much effort, however, it was not possible to induce tumors that could synthesize antibodies to an injected antigen. Leo Sachs, Kenko Horibata, Edwin S. Lennox and Melvin Cohn did succeed in establishing a line of mouse myeloma cells in tissue culture at the Salk Institute for Biological Studies, but the line was then lost. Eventually Horibata and A. W. Harris were able to establish a number of lines, which they distributed to other laboratories. My group at the Medical Research Council Laboratory of Molecular Biology in Cambridge subjected a line derived from one of Potter's tumors to intensive study.

At that time we were not thinking about monoclonal antibodies. We were studying how somatic (body) cells diversify in culture and how mutations modify the combining specificity of antibodies, and the mouse myeloma line was for us simply another appropriate tissue-culture line. By 1973 Richard G. Cotton, David S. Secher and I were able for the first time to produce structural mutants of a mouse myeloma protein secreted by a cultured cell line. That work and parallel investigations by Matthew D. Scharff of the Albert Einstein College of Medicine in New York demonstrated spontaneous mutations in cultured cells that affected the structure of the proteins they manufactured, and also told something about the molecular nature of the mutations and their frequency. The search for mutants was laborious, however, because the proteins made by the parental cells lacked recognizable antibody activity, changes in which would be the most effective indication of slight differences caused by mutations. Clearly what was needed was a cell line that secreted an immunoglobulin exhibiting antibody activity that could be easily assayed. No such line existed.

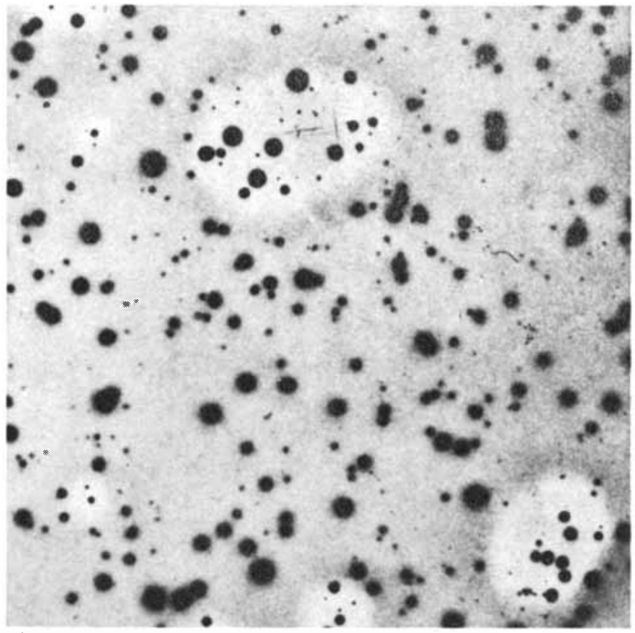
At that point a lucky circumstance led us to the hybrid-myeloma technique. While we were working on somatic mutations Georges Köhler and I were also following a quite different line of research in an attempt to learn more about the genetic control of the synthesis of antibodies. The synthesis of antibodies is controlled by two sets of genes. One set encodes the "variable" region of the antibody molecule's light and heavy chains, the region that controls antibody specificity; the other set encodes the "constant" region of the chains, the region that is responsible for such effector functions as the binding of complement (a complex of blood-plasma proteins implicated in the immune response), the transport of the antibody molecule across membranes and the binding of the molecule to membranes. Each lymphocyte synthesizes an antibody encoded by a single pair of *V* (for variable) and *C* (for constant) genes out of a large repertory of such genes in the cell, and when there are different alleles, or variant forms, of a *V* or a *C* gene on each of the cell's two chromosomes, only the allele on one of the chromosomes is active; the other is excluded.

In 1973 Cotton and I did an experiment to find out if allelic exclusion could be broken and, if it could, what the molecular consequences would be. We fused two myeloma cells, one from a mouse line and one from a rat line. Analysis of the hybrid cells showed that they secreted hybrid molecules consisting of various combinations of the chains synthesized by the parental cells but never a combination of a *V* region from one animal and a *C* region from the other. That meant the genes for the *V* and the *C* regions must be on the same chromosome. It is now known that the DNA sequences coding for the *V* and the *C* regions are separated by introns, or intervening sequences of DNA. The entire stretch of DNA is transcribed

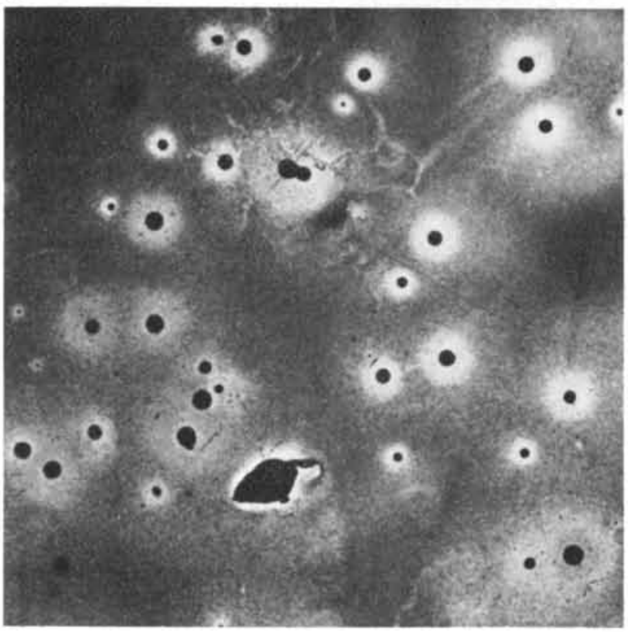
a



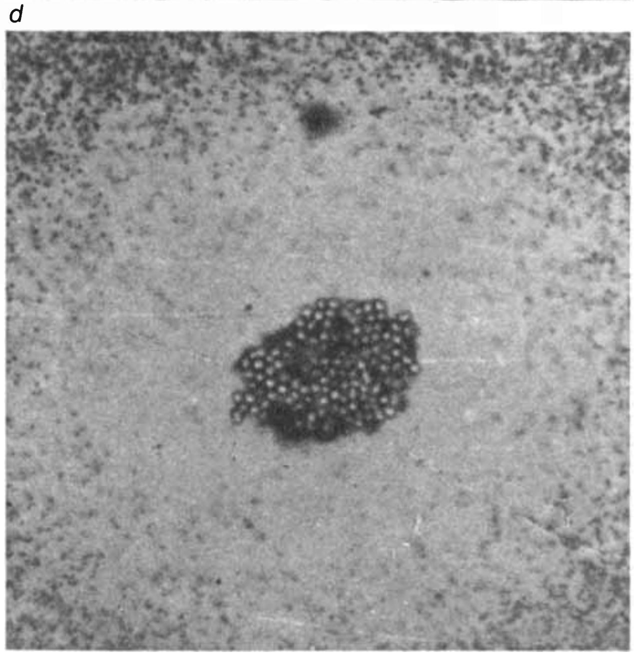
b



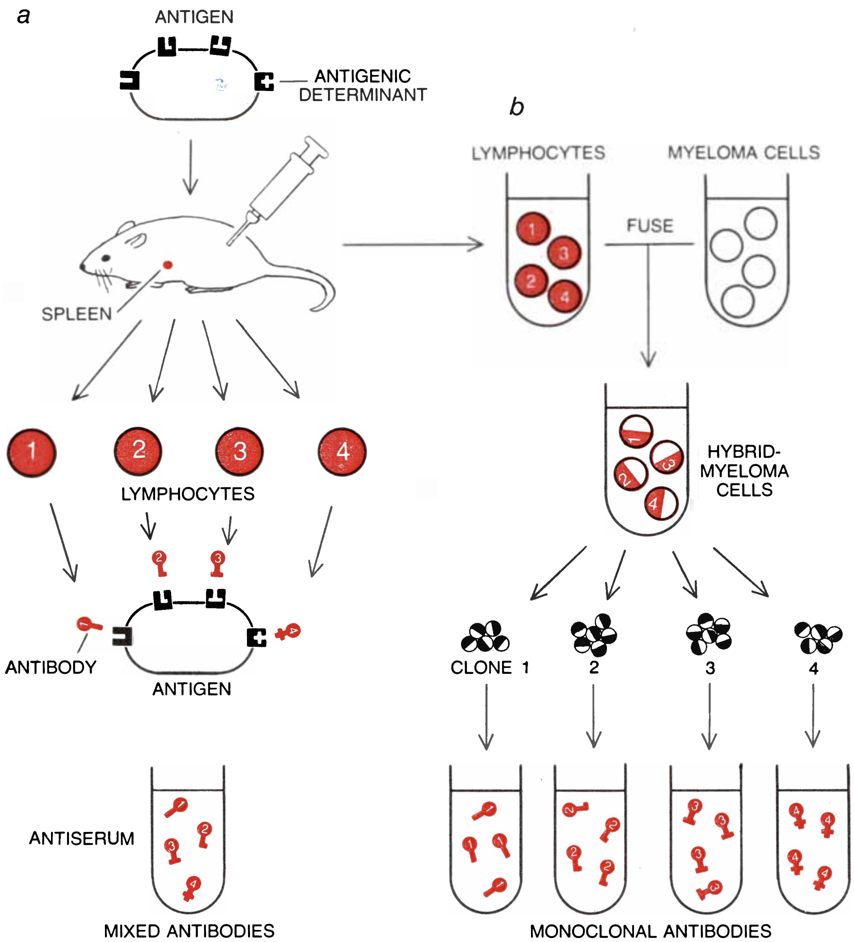
c



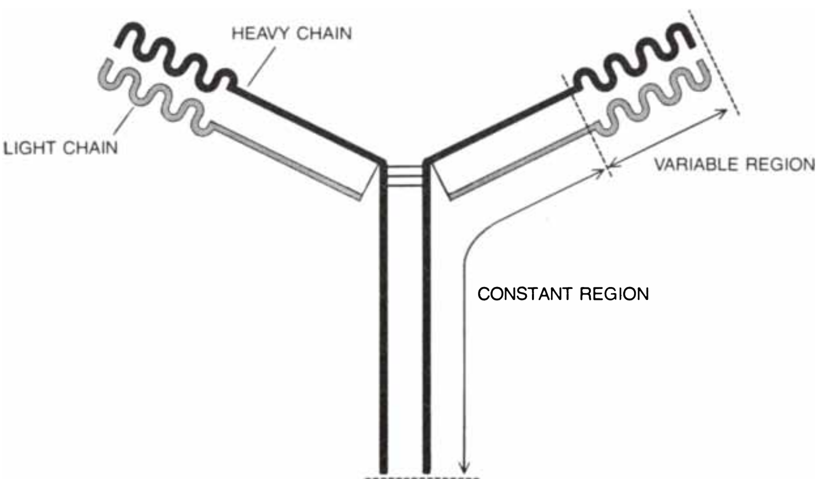
d



ANTIBODY-SECRETING CLONES of hybrid-myeloma cells were first detected by a test for antibodies to sheep red blood cells. In the standard test (*top left*) the red cells and antibody-secreting cells are incubated on agar, and complement (a protein complex from blood plasma) is added. Antibody diffusing from each secreting cell binds to antigens on nearby blood cells, initiating a complement reaction that kills blood cells, forming a plaque: a clear area (*white spots*) around each secreting cell. The author and Georges Köhler fused mouse cells secreting antibody to the blood cells with tumor (myeloma) cells and plated the hybrids (*top right*). Hybrid-cell colonies developed (*black spots*). When a layer of sheep red cells was added along with complement, a few hybrid colonies gave rise to plaques (*white areas around colonies*), indicating that they were secreting specific antibody. Individual cells were picked from a colony of antibody-secreting cells and plated thinly (*bottom left*); most of the clones derived from them proved to be secretors of the anti-red-cell antibody. A photomicrograph of a single secreting clone (*bottom right*) shows the individual cells of the clone and the area of dead cells around it.



IMMUNE RESPONSE is initiated (*a*) when an antigen molecule carrying several different antigenic determinants enters the body of an animal. The immune system responds: lines of *B* lymphocytes proliferate, each secreting an immunoglobulin molecule that fits a single antigenic determinant (or a part of it). A conventional antiserum contains a mixture of these antibodies. Monoclonal antibodies are derived by fusing lymphocytes from the spleen with malignant myeloma cells (*b*). Individual hybrid cells are cloned, and each of the clones secretes a monoclonal antibody that specifically fits a single antigenic determinant on the antibody molecule.



ANTIBODIES, shown in the illustration at the top of the page as stylized shapes, belong to the family of proteins called immunoglobulins. The basic shape of an immunoglobulin molecule is that of a molecule of the class immunoglobulin G (IgG), a heterogeneous population of molecules sharing a *Y*-shaped structure composed of two kinds of molecular chain, heavy and light, linked by disulfide bonds. The number and precise position of the disulfide bonds differ and are characteristic of the IgG subclass. Each chain has two regions. In the variable region amino acid sequences that differ from antibody to antibody provide differently shaped combining sites that bind specifically to different antigens. Constant region of the chains, with the same amino acid sequence in all antibodies of a given subclass, is responsible for other functions.

into a complementary strand of nuclear RNA. The RNA is processed by cellular enzymes—the intron is excised and the sequences coding for the *V* and the *C* regions are spliced together—to make the messenger RNA that is thereupon translated into the protein of the immunoglobulin chain. The hybrid-myeloma experiment, in other words, showed that the splicing of the *V* and the *C* sequences takes place within a single molecule of RNA.

This experiment also showed that there was no allelic exclusion in the hybrids, because the information from both parents was "codominantly" expressed by the fused cells. That finding suggested to Köhler and me a possible answer to our need for an antibody-producing cell in the mutation experiment. It occurred to us that it might be possible to fuse a normal lymphocyte or plasma cell with a myeloma cell and thus to immortalize the expression of the plasma cell's specific-antibody secretion. We would be applying the well-established cell-fusion technique to a new purpose, namely to fix in a permanent cell line a function that is normally expressed only in a "terminal" cell: the plasma cell derived from a *B* lymphocyte stimulated by an antigen.

For our first attempt we chose sheep red blood cells as the immunogen because antibody against such cells is easily detected by an assay developed by Niels Kaj Jerne and Albert Nordin in 1963. We mixed mouse myeloma cells with spleen cells from immunized mice in the presence of an agent that promotes cell fusion, identified successfully fused cells in a selective medium and found that they secreted immunoglobulins from both parents. Some of them secreted antibody against the red blood cells [*see illustration on preceding page*]. We were able to isolate clones that secreted single molecular species of that antibody, and the clones could be maintained in culture. We had for the first time developed continuous cultures of fused cells secreting a monoclonal antibody of predefined specificity.

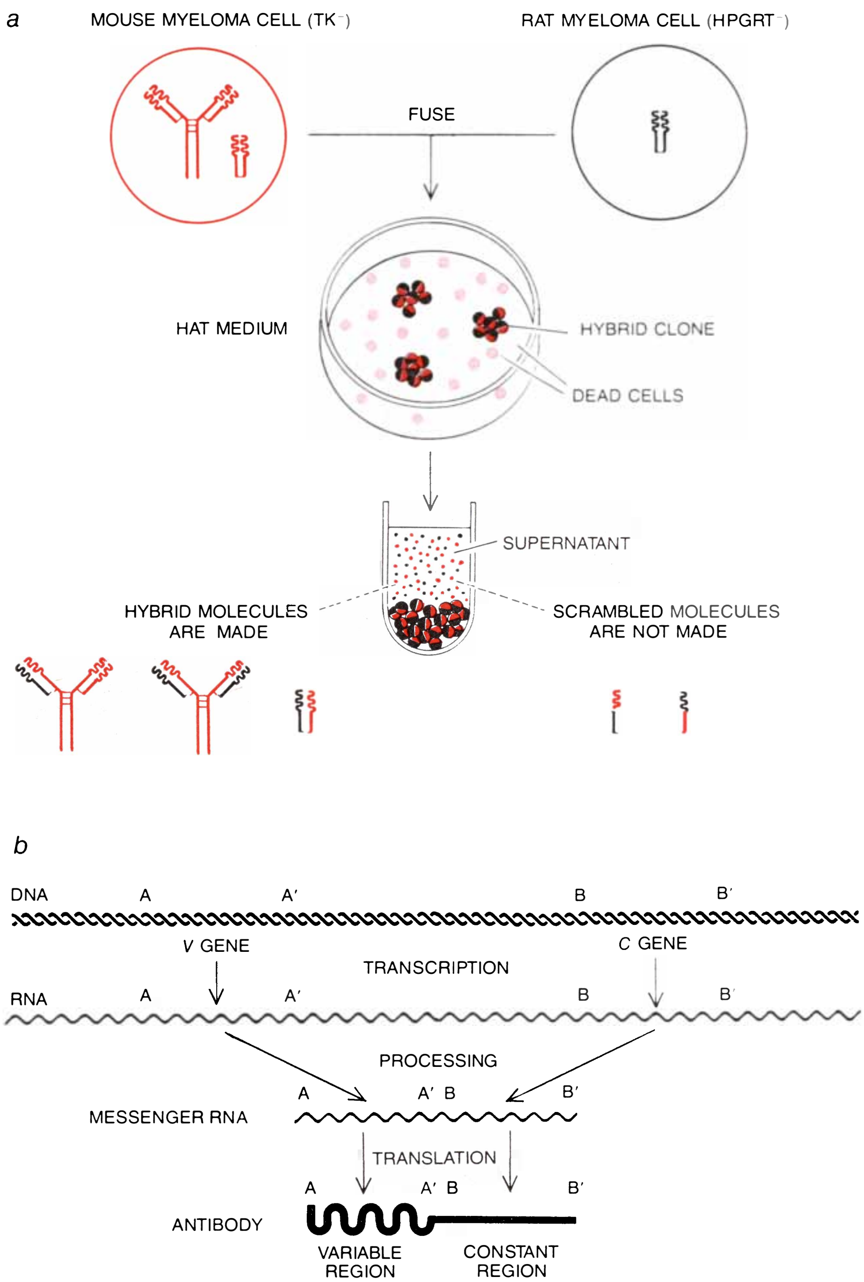
After our initial success in 1975 we ran into trouble, and for almost six months the experiments in our laboratory went badly or not at all; Köhler, who had moved to the Basel Institute for Immunology, reported difficulties too. Then Giovanni Galfré came to work in our laboratory and tried various modifications, in particular the use of polyethylene glycol as the fusing agent. It was a depressing period, but it did allow us to optimize the conditions for each step. Eventually the basic problem was discovered: a toxic batch of one of the reagents. Once that fault was remedied, with Jonathan C. Howard and Geoffrey W. Butcher of the Agricultural Research Council Institute of Animal Physiology in Cambridge we achieved a spectacularly successful fusion that produced a series of monoclonal antibodies to rat histocompatibility antigens: the cell-surface markers that establish individual identity and are responsible for the rejection of grafts. Other results began to come in at a rapid pace as we established a standard protocol for the experiments and developed new methods for assaying antibody secretion.

The success of these experiments was enhanced by an unexpected feature. Of the spleen cells we were fusing only perhaps one in 100 was an actively antibody-secreting plasma cell, and yet about one in 10 of our hybrid clones turned out to secrete antibody. That is, we had 10 times as many positive, immortal hybrids as one would expect if immortality were randomly transferred to the heterogeneous spleen-cell population; apparently we were achieving selectivity along with immortality. The explanation for this selectivity is not completely established, but according to recent evidence it probably has two components. On the one hand, secretion seems to be amplified, with lymphocytes that synthesize antibody but do not normally secrete it giving rise to hybrids that both synthesize and secrete antibody. Probably the myeloma parent provides the secreting machinery some antibody producers lack. On the other hand, the conditions under which the fusion takes place apparently make it unlikely that spleen cells other than *B* lymphocytes will give rise to long-lived hybrids.

When a clone of fused cells has been established, by definition all of the antibody it secretes is genetically derived from a single cell. It is not yet necessarily a monoclonal antibody in the immunological sense of the word, however, because each cell of the hybrid clone has some chromosomes from the myeloma-cell parent and some from the spleen-cell parent and is expressing both sets of chromosomes. Potentially a hybrid cell, instead of producing only the two components of a true monoclonal antibody (one kind of heavy chain and one kind of light chain), can produce two heavy chains and two light ones. We refer to such a cell as HLGK because it secretes the heavy and the light chains of the spleen-cell parent and the corresponding gamma and kappa chains of the myeloma-cell parent. It is in the nature of hybrid cells, particularly in the early stages of proliferation, to lose chromosomes rapidly. In this case the detectable loss is not random: heavy chains (H or G) are usually lost first, and then one or the other of the light chains (L or K) is lost. The HLGK hybrid therefore gives rise to variants whose secretion pattern is HLK or GLK, and these in turn to such variants as HL, HK, LK, L and K.

It is the clone of HL cells, expressing only the heavy chain and the light chain of the specifically immune spleen cell, that one is looking for (although there are reasons for also preserving other variants, notably HK). As one clones it is therefore necessary not only to assay for the specific antibody but also to analyze the immunoglobulin for its type of chain and select a strongly secreting HL (or HK) clone. The selection process can be simplified by choosing a mutant myeloma line that expresses only the light chain (K) and therefore yields HLK hybrids or one that does not express any immunoglobulin and therefore yields an immunologically monoclonal HL hybrid at the outset.

Once the desired clone is selected it can be frozen for long-term storage. At any time a sample of the clone can be



FIRST SUCCESSFUL FUSION of myeloma cells was between a mouse line that secreted complete IgG molecules and excess light chains and a rat line that secreted only light chains. Mutants of each line, respectively deficient in the enzymes TK and HPGRT, were cultured together in a selective medium, HAT, in which both enzymes are required for cell survival; only fused cells survived, forming hybrid clones. The clones secreted into the supernatant various hybrid molecules consisting of mouse heavy chains and rat light chains (*left*) but did not produce scrambled molecules combining variable regions from one animal with constant regions from the other animal (*right*). These results indicated that the variable and the constant regions are transcribed from DNA into RNA, and the RNA is processed and translated into protein, as is indicated (*b*), and also that the genetic information of both parental cells is "codominantly" expressed by the hybrids. More recent studies on DNA show that the *V* and *C* genes themselves are interrupted by intervening sequences that are excised in the course of processing.

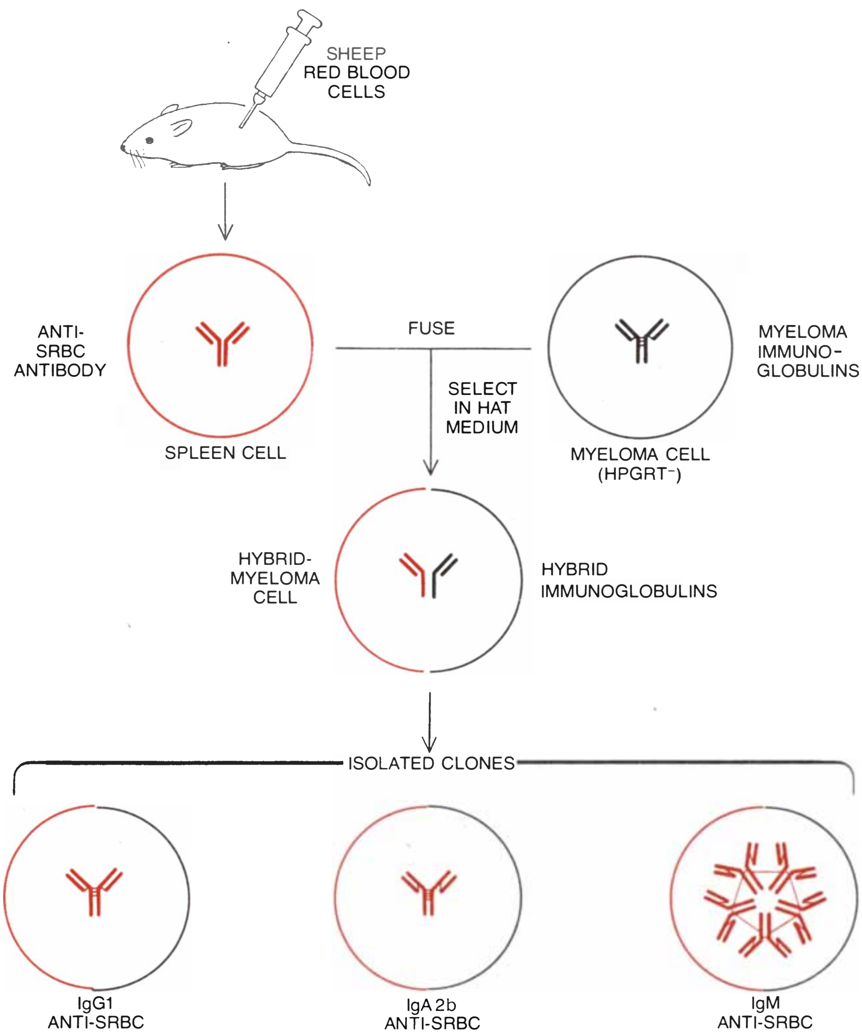
injected into animals of the same strain as those that provided the original cells for fusing. The animals develop tumors secreting the specific monoclonal antibody produced by the clone, and the antibody is present in their serum in extremely high concentrations: usually more than 10 milligrams of antibody per milliliter of serum, which in some examples is equivalent to a titer of perhaps a million. Alternatively a sample of the clone can be grown in a mass culture and the antibody can be harvested from the medium.

We have exploited methods similar to the one I have described to produce antibodies to a broad range of substances: the small antigens called haptens; proteins (including enzymes), carbohydrates and glycolipids; cell-surface components and viruses. The methods seem to be generally applicable. If the animal makes a certain antibody, one should be able to immortalize the antibody as a hybrid myeloma. The degree of difficulty in deriving a specific hybrid myeloma seems to vary with the immune response of the animal. When the response is very weak, the search for a hybrid clone secreting the specific antibody among the many clones secreting nonspecific immunoglobulins can require special procedures. We have developed several ways of screening large numbers of cells or clones for antibody production; we have also learned how to preselect spleens that are enriched in cells producing the specific antibody. There are no miracles, however. If the animal does not make an antibody, there is no way to immortalize that antibody.

A monoclonal antibody is a well-defined chemical reagent that can be reproduced at will, in contrast to a conventional antiserum, which is a variable mixture of reagents and can never be reproduced once the original supply has been exhausted. As monoclonals become available they are therefore likely to supersede conventional antibodies in many investigative and clinical laboratories. An example is provided by the standard test for the blood groups *A, B, AB* and *O*. Reagents for the test (antibodies to the *A* and the *B* red-cell antigens) are conventionally obtained from human serum. The best results are obtained by hyperimmunizing the donor of the serum by injecting red cells of the appropriate group (or better, a purified antigen), a potentially hazardous procedure; where such hyperimmunization is not done, as in the U.K., the reagents tend to be of lower quality. Moreover, the donated serum must be carefully screened for the presence of unwanted antibodies whose activity might obscure the anti-*A* or the anti-*B* reaction. (That is why reagents for the test cannot be obtained by immunizing laboratory animals. Antibodies in the animal serum would recognize the human character of the red cells being tested, completely eliminating the distinction between the *A* and the *B* groups.)

We have been able to establish that a monoclonal antibody to the individual specificity of a blood group need not be of human origin. In collaboration with Lennox, Steven Sacks and others a reagent is being produced at the Medical Research Council Laboratory of Molecular Biology from mass cultures of hybrid-myeloma cells that specifically recognizes Group *A* antigen. The reagent has been tested in comparison with the best available commercial reagents by Douglas Voak and Jack Darnborough of the Cambridge Regional Blood Transfusion Center and has been found to be equally effective.

The ability to derive antibodies to a single component of a "dirty" mixture opens up a new approach to the purification of natural products. As we became convinced of this premise we felt it needed to be put to the most stringent test we could think of. The substance of choice was interferon, which is notoriously difficult to purify and to obtain in significant quantity. When Secher and Derek C. Burke of the University of Warwick set out to purify interferon by immunoadsorption, the best interferon preparations to which they had access were about 1 percent pure. They immunized mice with a preparation of this crude interferon, fused spleen cells from the immunized mice with mouse myeloma cells and then tested the hybrid clones for their production of anti-interferon antibody, supplementing the unreliable and laborious biological assay for anti-interferon activity with tests for immunoglobulin secretion. With much difficulty they were able to select a positive clone, and by injecting its cells into mice they induced tumors that secret-



FIRST HYBRID-MYELOMA CELLS were prepared by immunizing mice with sheep red blood cells (SRBC). The cells involved in the experiment and their immunoglobulins are depicted here schematically. Cells from the spleen of immunized mice (which die in any tissue culture) were fused with mouse myeloma cells deficient in HPGRT (which die in the selective medium HAT). Hybrid cells survived in HAT and secreted immunoglobulins derived from both parents; some secreted active anti-SRBC antibody (*see illustration on page 67*). Clones of hybrid cells were isolated that secreted only a single species of antibody to SRBC: IgG1, IgA2b and IgM (a pentamer). In other words, each clone secreted an antibody to the sheep red cells.

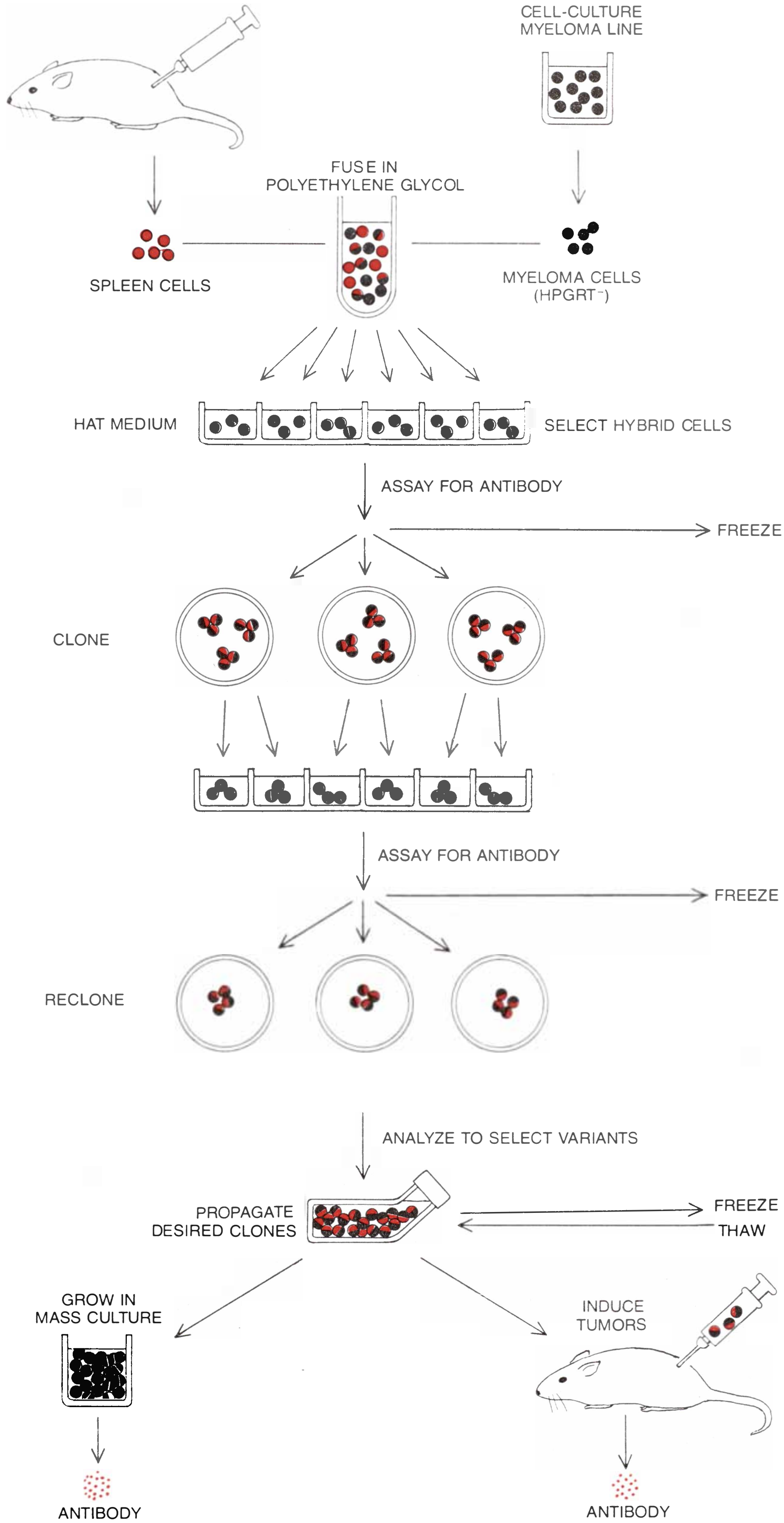
ed the antibody in large amounts. The antibody was attached to carbohydrate beads to prepare an immunoadsorbent column. Passing a totally crude interferon preparation through the column purified it 5,000-fold in a single step. Purification on an industrial scale is now being explored.

Monoclonal antibodies can be prepared that are specific for individual components of any complex mixture, and unlimited amounts of each antibody can be produced for immunoadsorbent columns. This makes it possible to dissect a mixture of completely unknown substances into its components. Animals are immunized with the mixture to be analyzed; hybrid-myeloma clones are derived and the antibodies from each clone serve to remove the components from the mixture one by one in cascade fashion.

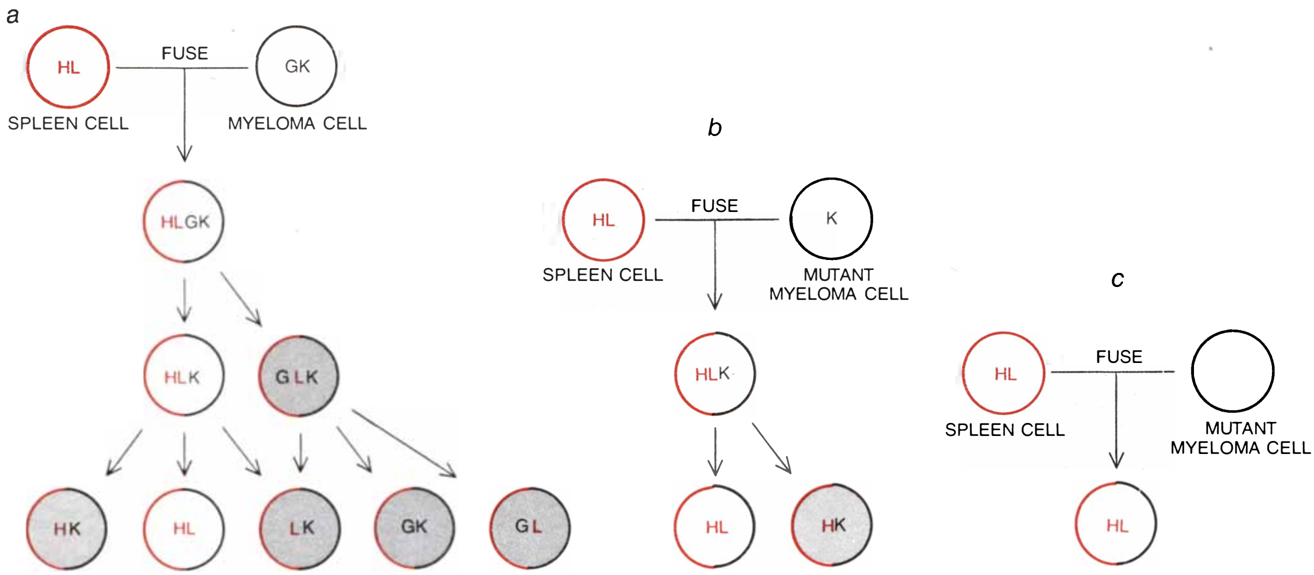
One of the most exciting investigative applications of monoclonal antibodies is in the area of membrane biology. Membrane proteins are hard to purify. They are present in cells in small amounts; often they have no easily measured biological activity, or else their activity is destroyed when the membranes are solubilized for analysis. One way to overcome these problems is to characterize cell-surface molecules by immunological methods, an approach that has been fruitful in the recognition of surface antigens that characterize particular cell types at different stages of tissue differentiation. Conventional antibodies to surface antigens are usually complex, however, and do not recognize single molecules; elaborate procedures have been required to circumvent this complexity.

In 1977 Galfré and I, along with Alan F. Williams of the University of Oxford, showed how the hybrid-myeloma technique could identify individual differentiation antigens. We immunized a mouse with cell membranes from rat thymus. After immortalizing the mouse cells that secreted antibody to thymus lymphocytes we were able to isolate clones producing different specific antibodies. In that one attempt we defined three new antigenic specificities, a task that might have required years of sophisticated immunology by conventional methods. Since then a number of other monoclonal antibodies to differentiation antigens of mouse, rat and human cells have been prepared.

The antigenic structure of a cell's surface establishes the cell's lineage and defines subsets of cells. For example, *B* lymphocytes can be distinguished from *T* lymphocytes (which take part in cellular immune reactions rather than in the secretion of antibody) largely because the former have immunoglobulin molecules on their surface and the latter have characteristic markers such as the Thy-1 antigen. Most surface markers, how-



STANDARD PROCEDURE for deriving monoclonal antibodies begins with the fusion, mediated by polyethylene glycol, of spleen cells from an immunized mouse (or rat) with mouse (or rat) myeloma cells. Hybrids are selected in HAT. The medium is assayed for antibody secretion, and a portion of each positive culture is frozen as a precaution. Positive cultures are cloned and the clones are assayed. The positive ones are then frozen, recloned and assayed for the presence of immunoglobulin variants (*see illustration on the next page*). The clones finally selected can be stored frozen. When the samples are thawed, they can be either grown in culture to produce the antibody or injected into animals to induce myelomas that secrete the antibody.



HYBRID-MYELOMA CELL produced by the usual fusion process (*a*) synthesizes not only the heavy (*H*) and light (*L*) chains of its spleen-cell parent but also the gamma (*G*) heavy and the kappa (*K*) light chains of its myeloma parent: it is thus designated HLGK. Hybrids tend to lose chromosomes, so that as the clone grows some chains are lost as is indicated here. Cultures are assayed to select desired variants (*unshaded cells*) until a stable HL variant secreting only specific antibody (or for some purposes an HK variant) is isolated. Fusion with mutant myeloma cells that make only a K chain (*b*) or no immunoglobulin at all (*c*) makes it easier to derive the desired HL clone.

ever, are not specific for a single subset of cells. Even the *B* cell's characteristic immunoglobulin may be present in several functionally different members of the *B*-cell lineage, from the so-called memory cells (which respond to reimmunization with an antigen the organism has previously been exposed to) to plasma cells (which secrete antibody). What characterizes a particular differentiated state is the presence of a particular ensemble of surface antigens and their quantitative expression.

To establish a unique antigenic profile for each of many cell types will require a vast collection of monoclonal antibodies and will take a long time. A good start has been made with the few reagents already available and with the help of cytofluorometers and fluorescence-activated cell sorters: instruments that can quickly measure both the size and the fluorescent intensity of large numbers of cells to which monoclonal antibodies, tagged with a fluorescent dye, have been attached. A large cell population can thus be fractionated into subpopulations on the basis of their size and surface-antigen pattern, and then the function of each subpopulation can be studied. Monoclonal antibodies, in other words, are standard reagents that can identify new surface molecules and at the same time distinguish among cell populations. So far the best results have been reported with various blood-forming and lymphoid cells; one directly practical application has been the differential diagnosis of various leukemias and related disorders.

The pattern of reactivity of monoclonal antibodies against subpopulations of cells is sometimes consistent with a given cell line's pattern of maturation, but not always. One monoclonal antibody seems to recognize an antigen characteristic of certain bone-marrow cells in the rat, whereas in the peripheral lymphoid organs it recognizes lymphocytes and in nervous tissue it recognizes some component that is as yet unidentified. Among the peripheral lymphocytes the antigen is present on *T* cells but not on *B* cells; yet it reappears on plasma cells, which are derived from *B* cells. We say this monoclonal antibody recognizes a kind of "jumping" antigenic determinant.

The monoclonal approach to characterizing differentiation antigens thus makes it possible to probe for the particular stage at which an antigen is expressed as well as for the line of cells that expresses it. The cascade purification method I described above can be applied not only to characterize the antigenic complexity of the cell surface but also to dissect functional as well as structural components of other biological materials such as cell organelles and pharmacologically active cell extracts.

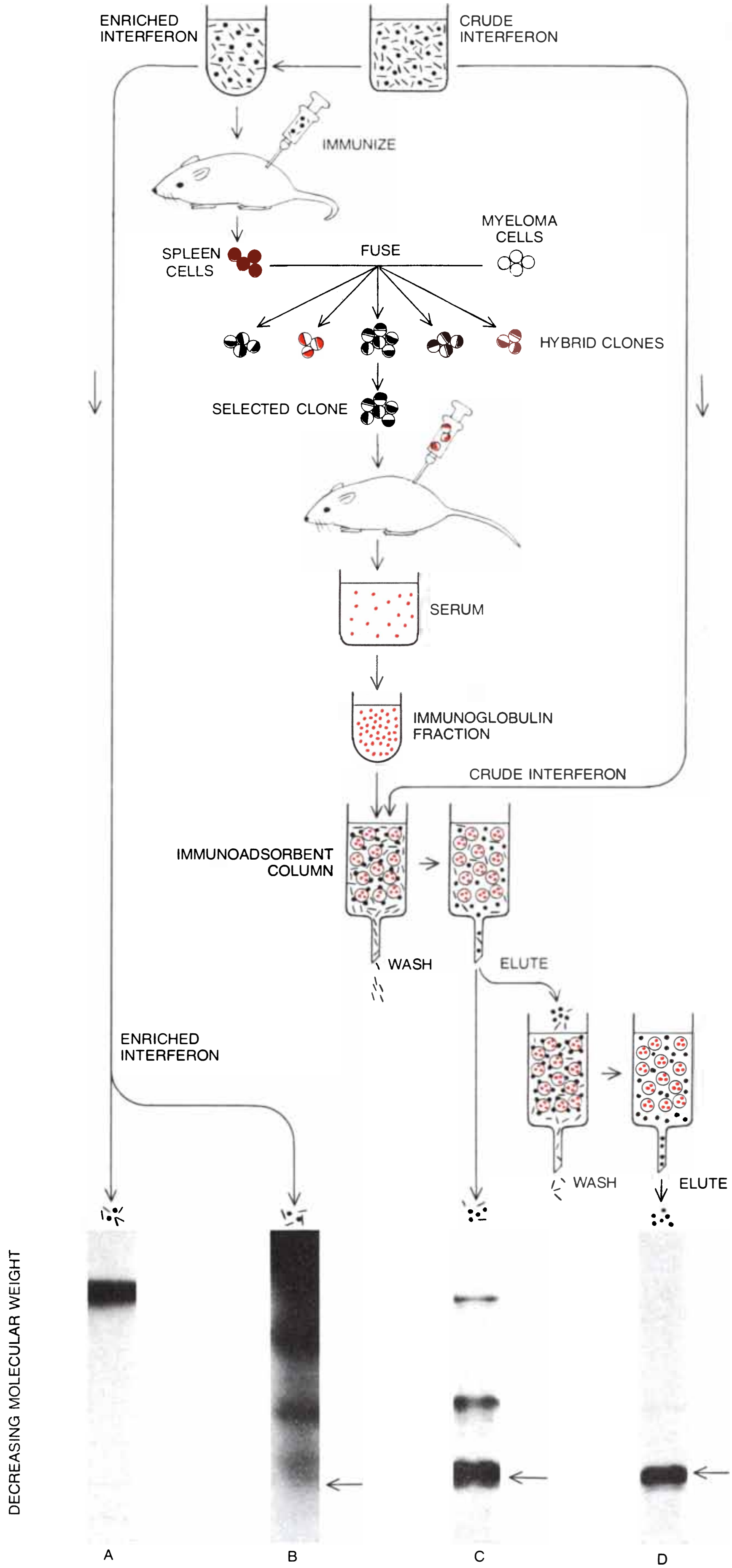
The "monospecificity" of antibodies from hybrid-myeloma clones has thrown new light on some well-known phenomena of antigen-antibody reactions. One indication of the binding of antigen to conventional antibodies in a test tube, for example, is the formation of a precipitate. The effect is not generally observed when the antibody is monoclonal. This is perhaps the first formal proof of the theory, advanced more than 40 years ago, that the precipitate is a three-dimensional lattice of antigens and antibodies. A monoclonal antibody binds only to a single antigenic determinant on an antibody molecule, so that no such lattice can be formed by a monoclonal antibody and most antigens. It can be formed only if the antigen is a polymer composed of repeated identical structural elements.

Monospecificity has also revealed some hitherto unsuspected phenomena that call for new interpretations of antigen-antibody reactions. To take just one example, it appears that the binding of different antibodies to neighboring sites on the same antigen is an important factor in the rupture of a cell membrane by complement. This synergistic effect was discovered as we were isolating the rat antibodies to histocompatibility antigens. We assayed for the presence of antibody-secreting hybrid myelomas by measuring the cytotoxic, or cell-killing, activity of their culture mediums. The supernatants of the uncloned cultures were consistently cytotoxic, but once we had cloned individual cells their supernatants showed no such activity. It occurred to Howard to measure the activity of a mixture of the supernatants of these apparently negative clones. To our delight the mixture was active, and then it was easy to purify two complementary components.

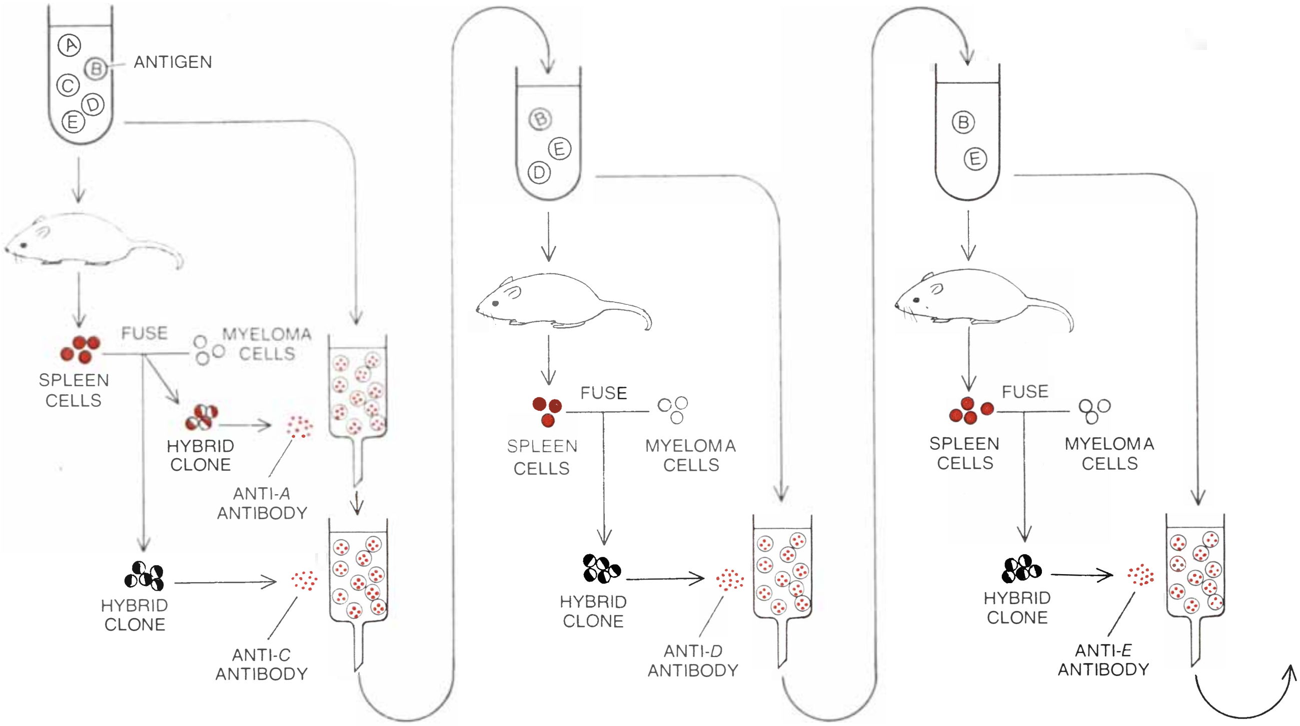
Once the synergistic effect was understood the "silent" activity of the isolated components could be exploited in a special way. Test cells could be "sensitized" by exposure to one monoclonal antibody and then exposed to the antibodies from other clones, thus revealing entire repertories of antibodies that act synergistically. Clearly there are cases where mixtures of monoclonal antibodies will be essential to produce a desired effect. In each case a decision will have to be made whether the advantages of blending monoclonals in specific proportions (rather than relying on the uncontrollable mixtures present in ordinary antiserums) will justify the effort involved in deriving the monoclonals.

Monoclonal antibodies are slowly beginning to replace conventional antiserums in standard kits for such procedures as the radioimmunoassay; many commercial companies are marketing them. Because they can be produced in large quantities they will make possible widespread use of kits of diagnostic reagents that until now were either not available at all or were considered too highly specialized for general application; one example is an antibody to the neurotransmitter called Substance *P*, derived recently by A. Claudio Cuello and me. The impact of monoclonals in virology, parasitology and bacteriology is only beginning to be felt. Great hopes are placed on their application to organ transplantation, just one aspect of which should be the worldwide standardization of tissue typing. In basic research the possibilities are even wider, with applications already reported in embryology and pharmacology and in the study of receptors for hormones and neurotransmitters.

Possible roles for monoclonal antibodies in direct therapy are under serious investigation. The most obvious role is in passive immunization (the injection of an antibody into a patient, as opposed to active immunization with an antigen that stimulates the patient's own antibody response). Given the impurity of conventional antibodies, passive immunization is not a common method of treatment, but it may prove to be effective when a purified antibody can be administered. In tumor therapy two kinds of role are foreseen for monoclon-



CRUDE INTERFERON was purified by immunoadsorption. Spleen cells of mice immunized with a somewhat enriched preparation of interferon were fused with myeloma cells. A hybrid-myeloma clone selected for anti-interferon activity and immunoglobulin secretion was injected to induce myelomas: purified antibody from the serum of tumor-bearing mice was attached to carbohydrate beads to prepare immunoadsorbent columns. When crude interferon was passed through such a column, it bound to the antibody and was retained when other components of the crude mixture were washed out; then the interferon was eluted. One passage through a column increased the preparation's interferon activity about 5,000-fold. Electrophoresis of the proteins (radioactively labeled to increase their visibility) made the results visible (*bottom*). The partially enriched preparation showed a single band (*A*), apparently albumin; overexposure of the same gel (*B*) showed minor bands, one of them (*arrow*) at a position corresponding to the molecular weight of interferon. After the passage of crude interferon through a column (*C*) there was less contaminant and a strong band at the interferon position (*arrow*). A second passage (*D*) produced material with a single strong interferon band.

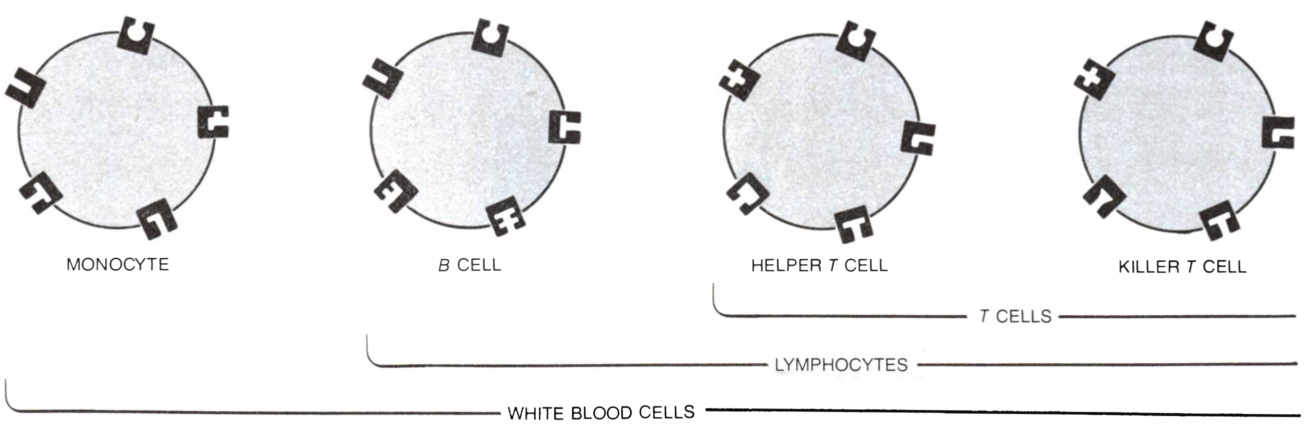


UNKNOWN MIXTURE can in effect be dissected by the monoclonal antibodies it engenders. A random set of antibodies, derived by immunizing animals with the unknown mixture, are applied to immunoadsorbent columns, which remove the corresponding antigens. The depleted mixture, now enriched in the remaining antigens, is injected to produce more antibody, and so on in cascade fashion. Hybrid-myeloma cells therefore provide a tool for characterizing the components and at the same time for separating and purifying them.

al antibodies. One role is the targeting of toxic drugs: antibodies to the tissues of a particular organ or to specific tumor antigens could be attached to drug molecules to concentrate the drug's effect. Alternatively it may be possible to produce antitumor antibodies that will themselves find and attack tumor cells.

For therapeutic applications antibodies derived from human lymphocytes rather than from the mouse or the rat would be desirable. Contrary to early hopes, this has proved to be difficult; attempts to immortalize antibody-producing human cells by fusing them with mouse or rat myeloma cells have so far been disappointing. The problem is that when human cells are fused with animal cells, there is a rapid preferential loss of human chromosomes from the resulting interspecific hybrid cells. And so far the search for a suitable human myeloma line that can be cultured and fused to make an intraspecific hybrid has not borne fruit.

In this overview of the uses of hybrid-myeloma antibodies I have referred only superficially to their obvious applications in basic immunological research. I have preferred to emphasize the fact that, although the technique originated in our effort to understand the genetic organization and expression of immunoglobulins, there has already been an impressive "spin-off" into many other areas. It is always hard to define the boundary between basic and applied research, but to experience personally the transition from one to the other has made a deep impression on me. I cannot think that if my research aim five or six years ago had been the production of monoclonal antibodies, I would ever have stumbled on the idea of attempting simultaneously to derive mutant antibody-secreting cells in one corner of the laboratory and to fuse two myeloma cells in another corner. Yet that was the combination that led to the initial production of monoclonal antibodies against sheep red blood cells.



DIFFERENTIATION ANTIGENS are cell-surface antigenic determinants that are either specific for individual cell types or common to sets or subsets of cell types. Here hypothetical antigens are shown for four white blood cells. The best definition of cell lineage and subsets is the pattern of expression of such markers. Monoclonal antibodies are an ideal tool for establishing such patterns of expression.


12 YEARS OLD WORLDWIDE • 86 PROOF • GENERAL WINE & SPIRITS CO. N.Y. N.Y.
12
REGAL
FOUNDED
1801
BLENDED
SCOTCH
WHISKY
PRODUCE
OF SCOTLAND
86
PROOF
BLENDED & BOTTLED IN SCOTLAND BY
BROTHERS LTD.
ABERDEEN, SCOTLAND
Truth in advertising.

SCIENCE AND THE CITIZEN

Shortfall

The most recent projections by the U.S. electric-power industry predict that the demand for electricity at the summer peak will grow by 4.3 percent per year in the coming decade, compared with a rate of 4.5 percent per year in the decade just ended. Capacity to meet the expected demand, however, will grow by only 3.6 percent per year even if all generating facilities now planned are completed on schedule. The projected increase in summer capacity is 233,500 megawatts, of which some 94,000 megawatts would be supplied by 86 nuclear power plants now under construction. (Today 74 nuclear plants with a capacity of 51,250 megawatts supply about 10.5 percent of the nation's electric power.) On the basis of these predictions the reserve margin of capacity over demand in the summer of 1989 will fall to 27.4 percent, compared with a reserve of 35.8 percent in 1979.

The foregoing estimates were compiled by the electric industry's Regional Electric Reliability Councils; the estimates have recently been subjected to a critical analysis by the Department of Energy. Although the DOE does not carry its own projection of power consumption beyond 1983, it believes the industry's estimates of both supply and demand are substantially too high. Whereas the industry expects an increase in demand of 4.54 percent per year for the period 1980–83, the DOE foresees an increase of only 2.06 percent per year. The DOE points out that the demand for power grew only 1.9 percent in 1979 and in the first five months of 1980 consumption was 1.2 percent less than it was in the corresponding period of 1979. The slackening demand is attributed to the economic slowdown, to energy conservation and to the restraining effect of rising electric rates. In support of its lower estimate of demand the DOE notes that each year since 1976 the regional councils' own projections have fallen steadily. In 1976, for example, the councils predicted a 1985 demand of 3,486,700 gigawatt-hours; this year the estimate has been reduced almost 17 percent to 2,909,000 gigawatt-hours.

The estimates by the DOE of slower growth in demand are more than matched by a lower expected rate of new power-plant construction. In particular, the DOE believes the completion dates listed for the majority of the nuclear units "are highly optimistic." Its own appraisal is that 53 of the nuclear plants due to be completed by the end of 1985 "are not likely to be operational on the dates [scheduled]. Failure of these units to be available as scheduled will degrade the projected reserve margins and will result in increased consumption of oil and gas." Many large coal-fired units are also behind schedule. If none of the 53 nuclear units is operating in the summer of 1985, the nation's reserve margin will be only 17 percent.

If licensing and other delays prevent all 53 nuclear units from operating in 1985, the U.S. will "lose" 57,000 megawatts of capacity, which would have supplied almost 10 percent of the industry's projected 1985 demand of 2.9 million gigawatt-hours. The DOE report identifies three regions "that will most likely not be able to completely make up the loss of planned nuclear generation, even with [power] imports from other regions." The areas are the Gulf States Utilities Group, the Texas Interconnected Systems Group and the Northern California–Nevada Group. A fourth area, the Northwest Power Pool, will probably run short of power even if the five nuclear units being built in the region are completed on schedule.

The DOE has estimated the amounts of coal, oil and gas that may be needed by 1985 to replace the lost nuclear power. Coal consumption may have to rise by some 150,000 tons per day, or 55 million tons per year, an increase of 7 percent over the 743 million tons otherwise projected for 1985. Oil use may have to increase 33.5 percent, from 2.1 million barrels per day to 2.8 million. (This year utilities are burning some 1.2 million barrels of oil per day.) Without the added nuclear capacity, estimated demand for natural gas will be higher by 932 million cubic feet per day, an increase of 22 percent.

These estimates of additional coal, oil and gas needs are based on the regional councils' projections of demand for 1985, which the DOE, as noted above, believes are too high. If the growth in demand conforms to the DOE's lower estimates, there will be no need for additional coal or gas to replace lost nuclear power; additional oil, however, will still be needed: about 237,000 barrels per day. The DOE report concludes that "failure of nuclear and coal-fired generating units to meet projected schedules for commercial operation could have a serious negative effect on power supply reliability and adequacy."

Waiting for Decay

The digging of a pit 60 by 70 by 80 feet is now complete some 1,950 feet underground in the Morton Salt Mine east of Cleveland, Ohio. In a few months the pit will be lined on all sides with 2,000 photomultiplier tubes and filled with water so clear that light can traverse the longest diagonal of the pool and still retain more than a third of its intensity. When flashes of light are detected, most of them will be caused by the arrival of high-energy particles from space. Some, however, may signal the decay of a proton or of a neutron bound in a nucleus. Either event would be evidence supporting the efforts of theoretical physicists to establish a mathematical formalism that unifies three of the four basic forces in nature: the strong force, the weak force and the electromagnetic force. (The stubborn holdout is gravity.) It would also be the first observation of a process that slowly robs the universe of its atoms.

By failing to find an instance of such a decay, a number of earlier searches have already established that the proton is prodigiously stable. Its mean lifetime is at least 10^{30} years, a time that exceeds the present age of the universe by 20 orders of magnitude. A neutron bound in a stable nucleus is thought to last equally long, although a neutron not in a nucleus has a lifetime measured in minutes. Until recently it was conjectured that protons and bound neutrons might be absolutely stable. No law of nature (or at least no law accorded fundamental status) forbade their decay; on the other hand, no mechanism was known that could convert a proton into particles of lower mass.

The recent unifying theories incorporate such a mechanism. To begin with, the proton and the neutron are thought to be composite particles, made up of quarks. There is one other family of elementary material particles: the leptons, a group that includes the electron, the muon and the neutrino. Quarks and leptons can be distinguished by their differing responses to the fundamental forces. Quarks are subject to all four forces, whereas leptons do not "feel" the strong force. The hypothesis that the proton can never decay is equivalent to a statement that quarks can never be converted entirely into leptons. Conversely, the unifying theories establish a connection between the strong force and the weak and electromagnetic forces and thereby posit a mechanism for transmuting quarks into leptons.

The mechanism is as follows. The proton consists of three quarks. In one hypothetical mode of decay the proton gives rise to a positron (a positively charged electron) and a neutral pion. The pion itself then decays, most often into two photons, or quanta of electromagnetic radiation. Thus three quarks disappear, one lepton (the positron) is created and some excess energy is liberated, to be carried off by the photons.

An intermediary in the decay process is a hypothetical particle designated *X*, which is expected to have a mass equivalent to an energy of 10^{14} gigaelectron

volts. For a subnuclear particle this is an enormous mass, comparable to the mass of a virus particle. One consequence of the X particle's large mass is that it has only a fleeting existence and cannot be detected directly even in principle. Another consequence is that an X particle is unlikely to be created in the first place. Soon after the big bang, energy was available to make X particles in profusion, but in the universe today the spontaneous appearance of an X is an extremely rare event. That is why protons so rarely decay. From present estimates of the X's mass the unifying theories predict a proton lifetime of between 10^{31} and 10^{33} years.

The experimental searches for the decay of the proton will attempt to detect evidence of the more enduring products of the decay, such as the positron and the pion. The way this is done is to assemble a multitude of protons and neutrons (which are collectively called nucleons) in a place well protected from any outside influence, such as cosmic-ray particles. The places chosen so far are all deep in the ground. The next step in the search is to wait for a nucleon to decay.

In the U.S. the largest detector (and the largest one contemplated anywhere in the world) will be the instrumented pit now under construction in the salt mine east of Cleveland. There the multitude of nucleons will be constituents of the atoms in 10,000 tons of water. The consortium building the apparatus includes investigators from the University of California at Irvine, the University of Michigan and the Brookhaven National Laboratory.

A second detector consisting of 1,000 tons of water is soon to be constructed in a silver mine near Salt Lake City, Utah, by a consortium from Harvard University, the University of Wisconsin and Purdue University. In both of these experiments the detection of a decay event will rely on the circumstance that the speed of light in water is less than it is in a vacuum. Indeed, the speed of light in water is less than the speed of the particles that arise from the decay. Under these conditions the particles give off the light known as Čerenkov radiation. (In much the same way an airplane traveling faster than the speed of sound gives rise to a sonic boom.)

The two experiments differ in a number of details. For one thing the smaller detector will have photomultiplier tubes distributed throughout the volume of water to make the collection of light more efficient. Another difference is that the smaller detector is surrounded by a shield of concrete and of particle counters. The physicists doing the smaller experiment are facing the problem of outside influences with the dual strategy of making the apparatus compact and of shielding it massively. Still, the assumption of both of the U.S. groups is that the outside influences can be compensated for. Two European groups, somewhat in contrast, intend to place detectors deeper in the earth. An Italian consortium is preparing a small prototype detector for a shaft in the Mont Blanc tunnel; a French group has similar plans for the Fréjus tunnel 50 miles to the south.

Meanwhile the smaller pioneering experiments continue. In one of them 200 tons of nucleons in the Homestake Mine in South Dakota are being monitored by investigators from the University of Pennsylvania. In another 150 tons of nucleons will be monitored in the Kolar Gold Field of southern India by workers from the Tata Institute of Fundamental Research, Osaka City University and the University of Tokyo. A third search (now ended) monitored only 25 tons, but they lay at a depth of almost two miles. They were placed in a South African gold mine by investigators from the University of California at Irvine, Case Western Reserve University and the University of the Witwatersrand.

Some calculations for the largest detector suggest how the efforts may fare. Roughly three times per second, or 10^8 times per year, a cosmic-ray muon will pass through the experiment, causing intense Čerenkov radiation. The high intensity of the light should distinguish the event from the decay of a proton or of a neutron bound in a nucleus. The difficulty is that some of the muons will interact with the rock or the salt near the detector, sending various particles into the volume of water. The subsequent collisions with atoms in the water can yield reaction products that resemble those of nucleon decay. For this reason the investigators plan to disregard all events in the two-meter thickness of water nearest the walls.

A second problem ultimately has no solution. Neutrinos are notorious for passing through matter without interacting with anything, but 1,000 times per year a neutrino that enters the usable volume of the detector will collide with a nucleon and give rise to a positron and a pion. This is precisely the pair of particles that the decay of a nucleon is most likely to produce. The positron and the pion usually will leave the collision with an angle between them of less than 120 degrees. Each will generate Čerenkov radiation. For the decay of an isolated proton (the nucleus of a hydrogen atom in the water) the expected angle between the two particles will be 180 degrees. For a nucleon in a nucleus the angle will be greater than 130 degrees. The investigators hope that of 1,000 neutrino interactions per year, 999 can be dismissed, in part by measuring the angle between the two flashes of light. This leaves one neutrino event per year.

It is this one event that sets a limit on the sensitivity of the experiment. In the usable volume of the detector there are 2.5×10^{33} nucleons. If the mean life of a proton or a bound neutron were 10^{33} years, the passage of a year would entail 2.5 decays. It is hypothesized that two-fifths of the decays will be of the form that yields a positron and a pion, and that is the decay mode the experiment is best suited to detect. Hence if the experiment were entirely free of error, the observation of one decay in a year would imply a nucleon lifetime of 10^{33} years. A similar calculation links the observation of 10 decays per year to a lifetime of 10^{32} years, and 100 decays per year to a lifetime of 10^{31} years.

From these calculations it would appear that any proton lifetime of 10^{33} years or less should be detected by the experiment. Actually, because of the unsuppressible background of neutrinos, the detector is not quite that sensitive. Even in the absence of proton decays one event per year can be expected from neutrino interactions. In practice, therefore, the observation of one apparent decay per year over the course of two or three years would only set a floor on the proton's lifetime: it could only be said to be more than 10^{33} years.

A larger detector would then be called for, but as the detector is enlarged the background of neutrino events grows. The consequence is plain: If it takes 10^4 tons of nucleons to establish a minimum of 10^{33} years, then it may take 10^5 or even 10^6 tons to seek a lifetime 10 times longer. The search might then be moved from deep underground to deep underwater. One proposal is to outfit the hold of a supertanker with the appropriate instruments and simply sink the ship.

Persistent Poliomyelitis

The control of poliomyelitis is one of the success stories of 20th-century medicine. In the developed countries of the world the paralytic disease has become a rarity in the years since vaccination programs were begun in the 1950's and 1960's. It may therefore come as a surprise that some 35,590 cases of poliomyelitis were reported to the World Health Organization in 1978. Almost all of them were in developing countries that do not have an effective vaccination program.

The worldwide status of poliomyelitis is reviewed in a recent report by the WHO, covering 150 countries and more than 90 percent of the world's population. The 35,590 cases are only those that came to the attention of health officials. The actual total was surely higher, not only because of countries not included in the survey but also because of underreporting in the records of those countries that were included.

In the WHO's American region 99 percent of the cases were recorded in 13 of the 41 nations reporting, but those 13 nations have almost half of the area's population. Most of the cases were re-

What if you chose HP products to

For example: The HP-85 computer can now double as an I/O controller, with interfacing power disproportionate to its modest price.

Since its introduction earlier this year, the HP-85 has proved itself as an easy-to-use, fully integrated computer, with keyboard, display, CPU, magnetic tape data storage, and thermal printer in one package small enough to fit on a corner of your desk or lab bench. Add to that BASIC language, 16K bytes of memory (expandable to 32K bytes), and full graphics capability, and you begin to see just how versatile the HP-85 is. And at 20 pounds (9 kilograms) it can be moved around without strain.

Now, an input-output ROM and Hewlett-Packard Interface Bus (HP-IB) card are available which virtually double the HP-85's versatility, turning it into a low-cost controller for just about any HP-IB - compatible instrument. Other interface protocols are also available, including RS-232C, GP-IO (parallel), and BCD.

More than 120 HP-IB - compatible, "designed-for-systems" instruments are available from HP, as well as optional peripherals to enhance the HP-85's performance: a high-speed, full-width line printer, a full-sized plotter, and flexible disc drives for data storage.

The I/O ROM provides straightforward I/O commands and statements to configure, control, pass data to and from, and check the status of devices in the system. With the I/O ROM's input and output capabilities, a programmer need key in only a few standard BASIC commands.

The HP-IB is the cornerstone in building computed measurement systems you can spend your time using rather than connecting and testing. It incorporates 10 years of interfacing experience from which the IEEE-488 interface standard was derived. As many as 14 instruments can be interconnected and controlled by the HP-85 via a simple, one-bus system.

Price of the HP-85 system including I/O ROM, ROM drawer, and HP-IB card, starts at $3985*.

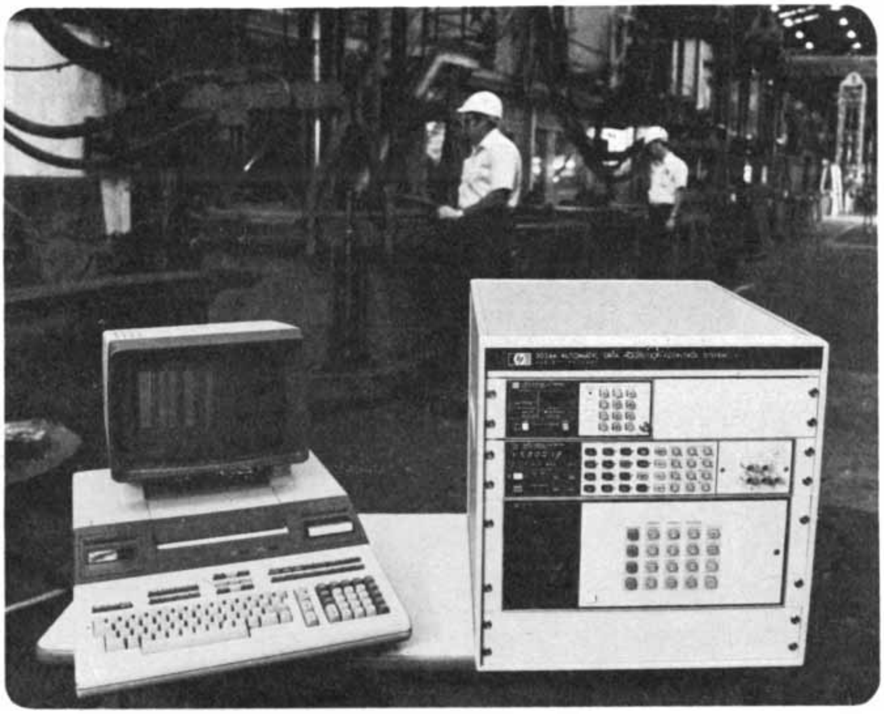
enhance your productivity?

For example: a system that can be precisely configured to acquire, analyze, and present your crucial data—automatically.

With conventional measuring instruments, getting answers to complex measurement questions requires three separate operations: making the measurement to obtain data, analyzing the data, and synthesizing it to achieve the answers. This takes time and considerable technique, and occupies skilled people in repetitive tasks.

Now, through a single process which we term *computed measurement*, HP instruments make the measurements, perform the analyses, synthesize the data, and deliver fast, precise, direct answers. In many instances, they provide answers previously unobtainable or economically unfeasible.

A case in point is HP's new computer-based data acquisition and control system, the HP 3054. It brings capability and economy to applications in electronic testing, process control development, transducer measurements, production testing, R&D, and signal analysis.



In its fullest configuration, the HP 3054 comprises a data acquisition and control unit and two voltmeters interfaced via HP-IB to one of several HP computers, depending on the nature of the application.

For instance, a product test application where variables are known and measurements are made against design standards usually requires different computer characteristics than a measurement application where evaluation is made without benefit of a known model, or a process control application where measurements serve primarily to regulate a sequence of events.

HP 3054 system computers include the HP-85 as described on the opposite page, or one of several desktop computers in the HP 9800 series, or a large mainframe computer. Powerful system software includes instrument driver routines, data analysis and presentation programs, instrument verification routines, and applications programs.

Prices for the HP 3054 system range upward from $9000* depending on the number of options and the computer selected.

For more information, mail the coupon to Hewlett-Packard, 1503 Page Mill Road, Palo Alto, CA 94304. Or call the HP regional office nearest you: East (201) 265-5000, West (213) 970-7500, Midwest (312) 255-9800, South (404) 955-1500, Canada (416) 678-9430.

These are only two of many thousand product reasons for choosing HP.

hp HEWLETT PACKARD

Mail to: Hewlett-Packard, 1503 Page Mill Road, Palo Alto, CA 94304
Please send me further information on

( ) HP-85 I/O controller
( ) HP 3954 data acquisition and control system

Name ________
Company ________
Address ________
City ________ State ________ Zip ________

*Domestic U.S. prices only.

00045

ported from Brazil, Colombia and Mexico. Countries in other regions with a relatively high incidence of the disease were Turkey, Kenya, Afghanistan, Iraq, Vietnam, Pakistan and Egypt.

The situation in Egypt reflects the difficulty that some developing countries have in establishing an effective program of vaccination. Immunization has been compulsory in Egypt since 1965, and a mass-vaccination campaign was carried out in 1976–77. Nevertheless, there were more cases of poliomyelitis in Egypt in 1978 than there were in 1977. A survey showed that 80 percent of the patients had received fewer than the recommended three doses of oral vaccine, and some of them had received none. In tropical countries climate compounds the difficulty of eliminating the disease because heat tends to inactivate the vaccine.

Even in the developed countries there have been occasional outbreaks of poliomyelitis. The Netherlands had such an outbreak in 1978 among a group of people who for religious reasons refuse all immunizations. The disease has also appeared among similar groups in the U.S. and Canada.

The WHO report inspired an editorial in *British Medical Journal,* which described the outbreak in the Netherlands as "a reminder of the epidemic potential of pockets of susceptible individuals." For this reason, the *BMJ* said, it is important for the developed countries to maintain their immunization programs and for people traveling abroad to make sure they have been fully immunized.

Small Talk

In mastering language a child learns first to understand spoken words and then to speak; only later are reading and writing introduced. Curiously, the linguistic capabilities of the electronic digital computer have emerged in exactly the reverse sequence. Writing (or at least typing) is a trivial task for a computer, and reading is only a little more difficult if the text is carefully prepared. Spoken language presents much greater challenges. Electronic devices that synthesize speech are just now becoming economically practical, and the automatic recognition of natural spoken language remains in a primitive state of development.

In speech synthesis some refinements have been made recently in the algorithms, or formal procedures, by which artificial vocal sounds are created and assembled into words. The most important advances, however, have been technological rather than conceptual. The synthesis of intelligible speech once required the dedicated services of a large, high-speed digital computer, along with specialized devices for signal processing. Now all the microelectronic circuitry needed for speech synthesis fits on two or three chips of silicon, which can be made in large quantities at a cost of a few dollars each.

The most straightforward method of speech synthesis resembles a digital recording of the human voice. The waveform of each word or phrase to be spoken is sampled at short intervals, perhaps 10,000 times per second, and the amplitude during each interval is encoded as a binary number. The complete array of numbers that defines the waveform is then stored in the random-access memory of a computer. A word is reproduced by converting the numerical values into voltages, which are amplified and supplied to a loudspeaker. Such a recording differs from one on magnetic tape or on a phonograph record in that the stored data can be recovered in any sequence. Words and parts of words can therefore be recombined to create new sentences.

The digital encoding of waveforms is the brute-force approach to speech synthesis. The technique yields artificial speech of the highest quality, but it makes prodigious demands on computer resources. Storage of the encoded waveform requires about 50,000 bits, or units of information capacity, for each second of speech. Furthermore, the rate at which this information must flow through the central processor strains the capabilities of some computers. For a practical low-cost system the data rate must be reduced.

A set of chips recently introduced by the National Semiconductor Corporation operates by encoding the speech waveform, but the data are greatly compressed before they are stored. For example, segments of the waveform that are repeated several times in succession are stored only once, and all segments are converted into a mirror-symmetrical form, so that only half of each segment has to be recorded. Low-amplitude sounds are deleted entirely. Through these techniques and others the data rate is reduced to about 1,000 bits per second of speech, with only a slight loss of quality. Not only is the speech intelligible but also the voice of the speaker who made the recording remains identifiable.

Another technique, called formant synthesis, achieves a still lower data rate by assembling words from phonemes, or elementary units of spoken sound. The human vocal tract is capable of articulating more than 100 distinct phonemes, but any one language draws on only a subset of them. A single-chip synthesizer made by the Votrax Division of the Federal Screw Works reproduces some 45 phonemes, including all those required for English. Each phoneme is generated by a particular setting of various tone generators, noise generators and acoustic filters, all of which are included on the chip. In order to specify a word one merely lists its component phonemes. The data rate can be less


CONCORD

ONE TOUGH AMERICAN ECONOMY CAR.

23/34
EPA EST MPG/EST HWY*

What makes the American Motors Concord DL different from every other good-looking, high-mileage car? Of the world's most comfortable cars, only Concord gives you galvanized steel in every exterior body panel as part of its exclusive Ziebart® Factory Rust Protection, a full 5-year No Rust-Thru Warranty,™ and the famous Buyer Protection Plan® So for all its style, Concord is built to last. If you're the tough customer we think you are, Concord is your kind of car.

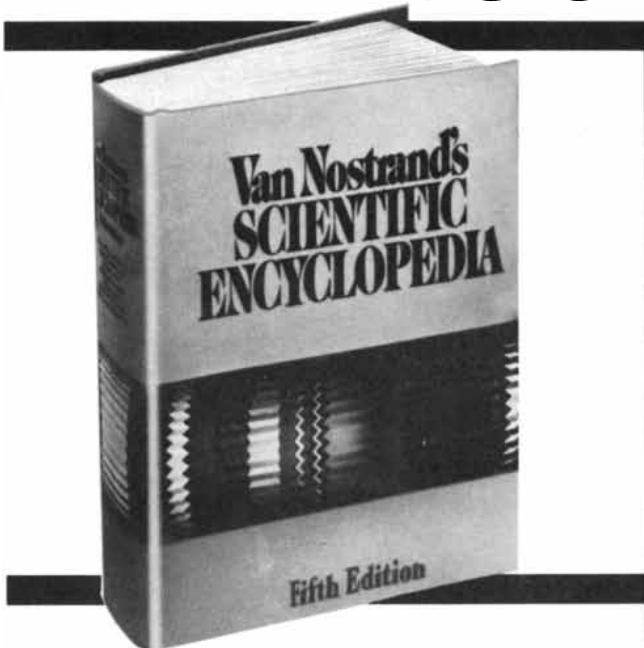
*Use EPA est MPG for comparison. Your mileage may vary with speed, weather, trip length. Actual highway will probably be less. Calif. ests. may vary.
Ziebart is a registered trademark of Ziebart International Corporation.
See your American Motors dealer for warranty and rust program details. Aluminized exhaust warranted for 12 months/12,000 miles.

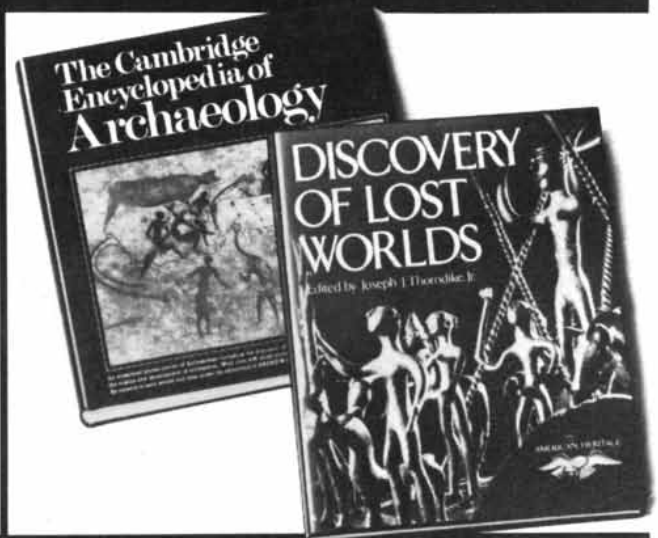
BUILT TO LAST **AMERICAN MOTORS**

YOU ARE LOOKING AT THE ONLY CAR BUILT IN AMERICA WITH 100% EXTERIOR BODY PANELS OF GALVANIZED STEEL.

5 good reasons for joining the

Natural Science Book Club



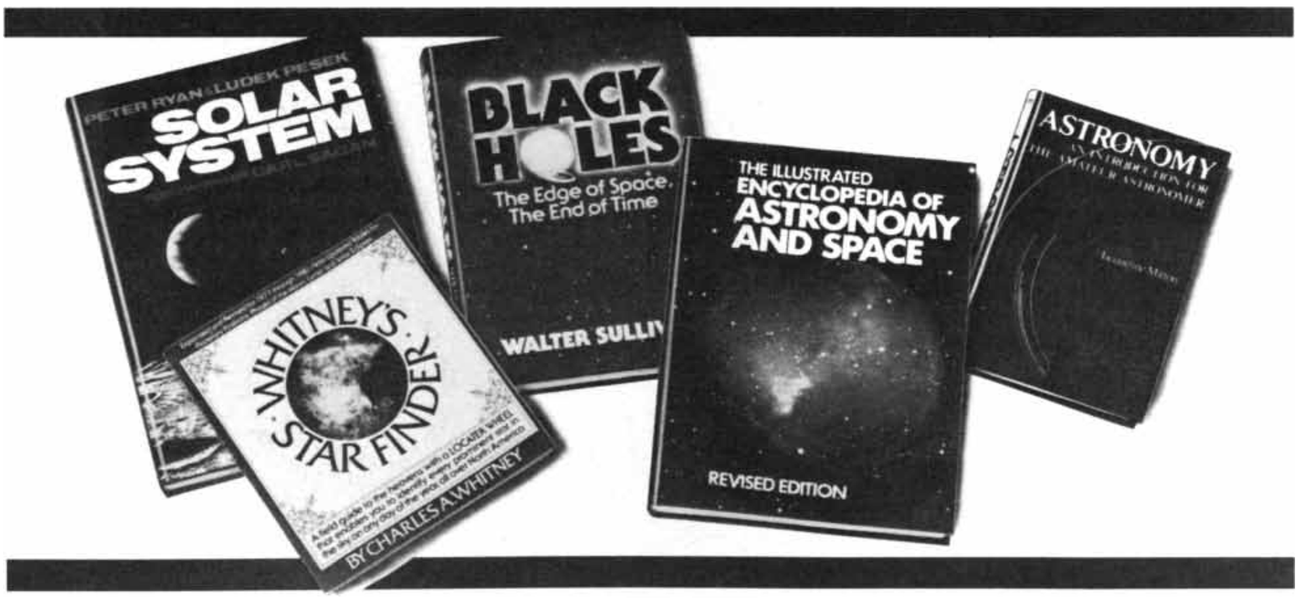


1. Van Nostrand's Scientific Encyclopedia

Edited by Douglas M. Considine. Nearly 200 experts have contributed to this thoroughly revised and greatly expanded fifth edition of the most authoritative single-volume source of scientific knowledge ever assembled. Enormous 9¼″ × 12″ volume contains 2.2 million words, 2382 pages, 2500 photographs, drawings, and charts, and 500 tables. 7200 articles cover everything from the earth and space sciences to energy technology, from the life sciences to mathematics, from the information sciences to physics and chemistry. ". . . an amazing book . . . for both the general and scientific reader."—*The New York Times.* **Publisher's Price $69.50**

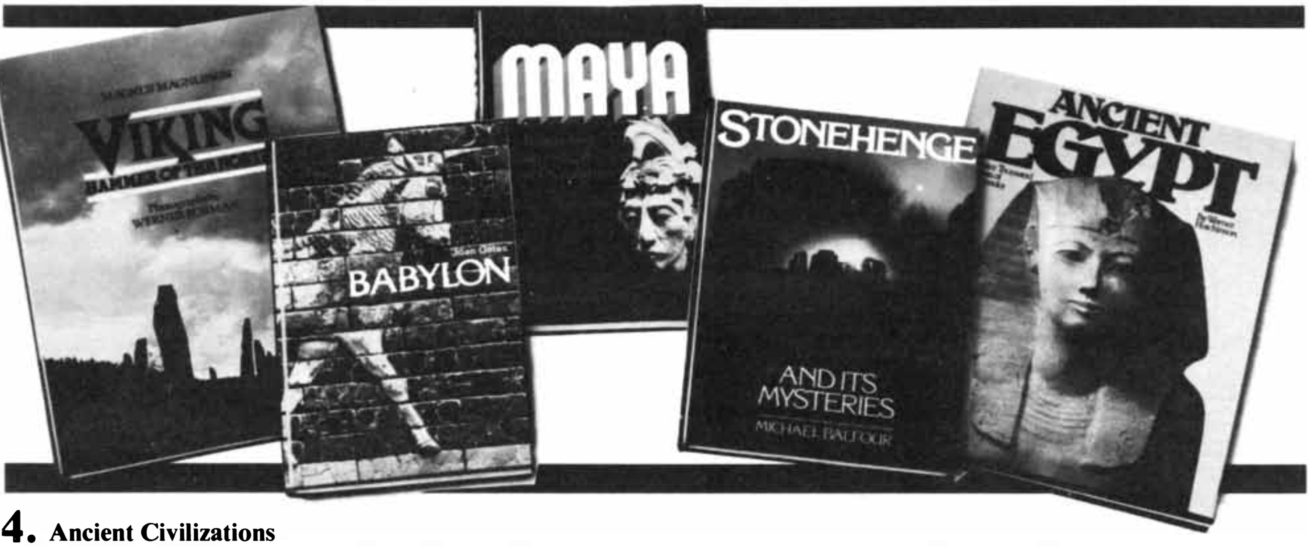
2. Archaeology Library

Two extensively illustrated volumes provide a panoramic overview of ancient cultures. ***The Cambridge Encyclopedia of Archaeology***—an over-sized reference with more than 400 photos, drawings, maps, and plans. Includes writings by 55 experts. ***Discovery of Lost Worlds***—the major archaeological finds of the past 300 years recreated in the writings of well-known scholars. Beautifully illustrated. **Publishers' Price $69.90**



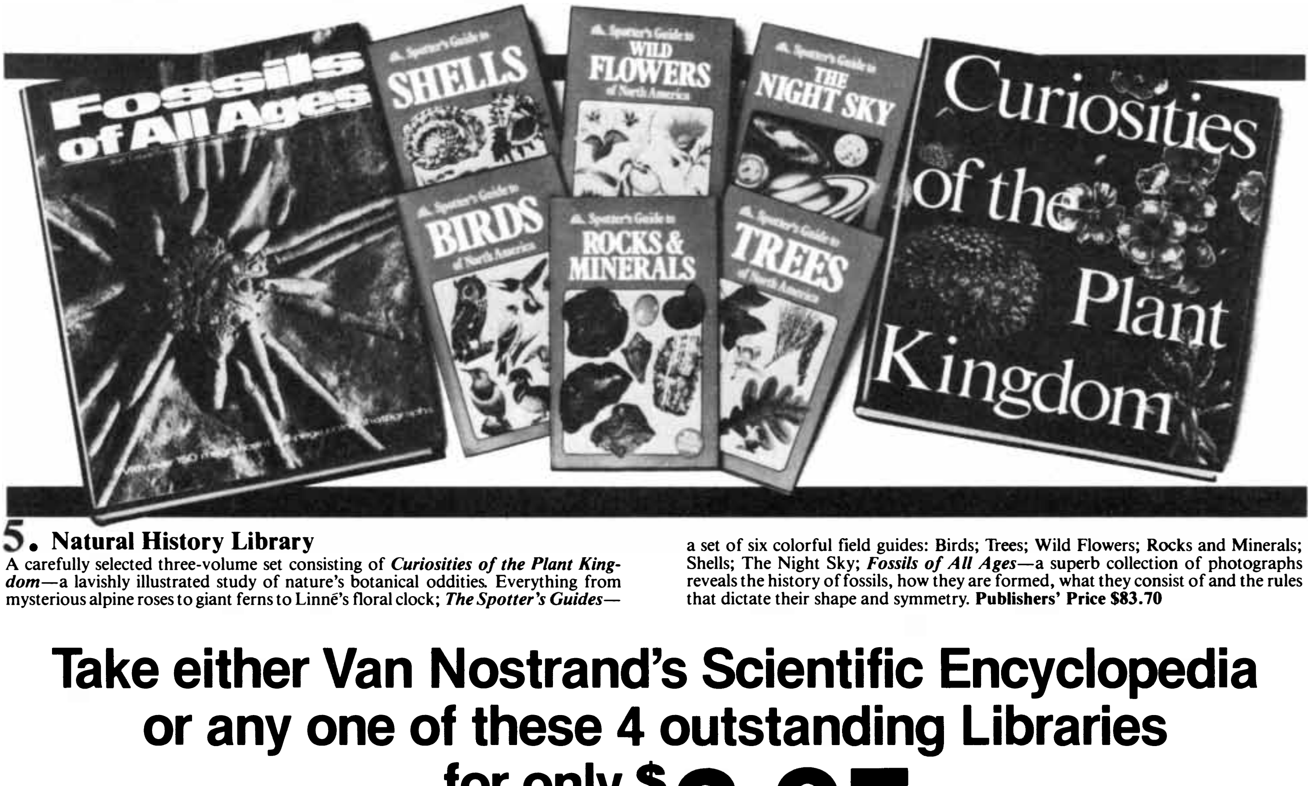
3. Basic Astronomy Library

Five books that will provide you with a basic, well-rounded library. Start with ***Astronomy: An Introduction for the Amateur Astronomer. Solar System*** contains paintings, charts, diagrams and over 100 photographs taken in space. ***Black Holes: The Edge of Space, The End of Time. Whitney's Star Finder*** enables you to identify every major star in the sky over North America. ***The Illustrated Encyclopedia of Astronomy and Space*** contains over 1000 entries and 150 photographs and illustrations. **Publishers' Price $70.75**



4. Ancient Civilizations

A five-volume set including ***Viking: Hammer of the North***—an illustrated study of the Age of the Vikings, AD 800 to 1100; ***Maya: The Riddle and Rediscovery of a Lost Civilization; Ancient Egypt***—a fascinating survey of Egyptian civilization, contains 125 photographs; ***Stonehenge and Its Mysteries***—over 200 maps, diagrams and photographs and ***Babylon***—beautifully illustrated, reveals the lifestyle, art, architecture and more of this fabulous city of antiquity. **Publishers' Price $79.75**



5. Natural History Library

A carefully selected three-volume set consisting of ***Curiosities of the Plant Kingdom***—a lavishly illustrated study of nature's botanical oddities. Everything from mysterious alpine roses to giant ferns to Linné's floral clock; ***The Spotter's Guides***—a set of six colorful field guides: Birds; Trees; Wild Flowers; Rocks and Minerals; Shells; The Night Sky; ***Fossils of All Ages***—a superb collection of photographs reveals the history of fossils, how they are formed, what they consist of and the rules that dictate their shape and symmetry. **Publishers' Price $83.70**

Take either Van Nostrand's Scientific Encyclopedia or any one of these 4 outstanding Libraries for only $2.97

(values to $83.70)

if you will join now for a trial period and agree to take 3 more books—at handsome discounts—over the next 12 months.

As a member, you'll be kept informed of the most interesting, most important science books being published. You'll get to choose from a dazzling array of books—from fascinating accounts of the discovery of lost civilizations to speculations on the origin of our universe, from sumptuous word-and-picture books on the world of nature to stimulating books on where man came from and where he is going.

And you get these books at handsome savings. In addition to receiving up to $83.70 worth of books for less than three dollars when you join, you keep on saving substantially—up to 30% and occasionally more. Your total savings as a trial member, for example, can *easily* be over 50% if you include your introductory offer. And these are real savings, because our books are always of equal quality to the publishers' editions, *never* "economy" editions.

Moreover, you will *immediately* become eligible for our Bonus Book Plan, with savings of at least 70% off the publishers' prices. And you can choose your Bonus Books from *any* of the titles we offer.

Why not start enjoying the many benefits of Club membership? Join the **Natural Science Book Club** today.

If the reply card has been removed, please write to **Natural Science Book Club, Dept. 4-AC6, Riverside, N.J. 08075** to obtain membership information and an application.

than 100 bits per second, although the quality of the speech is inferior to that obtained by waveform encoding.

A similar but more sophisticated method, called linear predictive coding, is employed in a set of two chips made by Texas Instruments, Inc. Again various sound sources and filters are adjusted to simulate a range of phonemes, but the quality of the resulting speech is enhanced by a more realistic mathematical model of the vocal tract. In human speech physical constraints on the movement of the tongue and of other organs of articulation alter intonation and influence the transition from one phoneme to the next. Linear predictive coding incorporates some of these constraints into the algorithms that generate an utterance, modifying phonemes according to their context. The data rate is somewhat more than 1,000 bits per second of speech.

The Texas Instruments chips were the basis of the first microelectronic voice encoder. They were employed two years ago in a Texas Instruments product: a toy called Speak & Spell, which gives spelling drills to children. With the wider availability of chip sets for voice synthesis, talking toys, games and appliances may become commonplace. A chess-playing machine that reports its moves aloud has been manufactured for some time by Fidelity Electronics Ltd.; a clock that announces the time and a talking calculator were recently introduced by the Sharp Corporation of Japan. An application that draws on a much larger vocabulary is an automated language laboratory; such a system is being installed at Stanford University.

Accurate recognition of natural human speech is inherently more difficult than accurate imitation of it. Distinctive acoustic features of a word, such as its frequency spectrum, can readily be extracted; the problem lies in matching the resulting pattern with a prerecorded "template" in order to identify the word. Pronunciation varies from speaker to speaker, and it changes from time to time even for the same speaker. The boundaries between words are often entirely unmarked, and pauses within a word can be longer than those between words. Background noise, which the human ear and brain somehow segregate from speech, cannot easily be suppressed in an electronic system. Whereas a person listens, the computer can at best only hear.

Because of these complications most devices for speech recognition respond to only a limited number of words, and even those words are recognized only when they are isolated by pauses from the preceding and following words. Moreover, most of the devices can be operated by only one speaker at a time, who must "train" the machine to recognize his voice. In general the recognition systems have not yet been reduced to microelectronic form, although one company, Threshold Technology, Inc., has announced plans to do so.

Even a limited capability for voice recognition is useful enough for such devices to be put to the test of commercial practice. For example, they have been employed for the control of production-line machinery where a worker must keep both hands free for other tasks. An airline has tested electronic recognition of spoken commands for baggage handling, and a parcel-delivery service has modified a sorting machine to accept voice entry of destinations. Major cities are designated by the names of their football teams. An experimental system now being tested by the Federal Aviation Administration would allow an air-traffic controller to communicate with a computer without glancing away from the radar display.

The more ambitious goal of recognizing continous speech is being pursued in the laboratory. At the International Business Machines Corporation a large computer has been programmed to recognize about 1,000 words drawn from patent applications for lasers. In sentences pronounced by a trained speaker the program identifies these words with an accuracy of 91 percent, but the processing time is quite long: more than three hours for each minute of speech. One aim of the IBM project is an automated dictation system, which would accept voice input and return a typed transcript.

Several other potential applications of both voice synthesis and speech recognition are to be found in the telephone system. Certain kinds of messages, such as changes of telephone numbers, are already given by an artificial voice, and in Japan telephone banking services are provided by a device that recognizes a few words. At Bell Laboratories a prototype of an automated directory-assistance service is under development. Names and telephone numbers of some 18,000 laboratory employees are stored in the system. Information can be retrieved in a few seconds by spelling a name aloud; when the inquiry service is completed it will respond by saying the corresponding number in a synthetic voice. Thus the user will be engaged in a dialogue with the computer.

Fish Story

The coelacanth, a primitive fish once thought to have died out with the last of the dinosaurs, was found some 40 years ago to be alive and well and living in the Indian Ocean. Still, biologists had no opportunity to observe a living coelacanth, or even to examine well-preserved tissues, until 1972, when a coelacanth was captured alive by a joint British, French and U.S. expedition. Two more specimens were caught in 1975 during an expedition sponsored by the California Academy of Sciences. The findings of the two groups are reviewed in a recent issue of *Proceedings of the Royal Society* and in a volume published by the California Academy: *The Biology and Physiology of the Living Coelacanth.* The results may call for a drastic revision of the taxonomic classification of the ancient fish.

The first modern coelacanth was brought up in 1938 by a trawler near the mouth of the Chalumna River on the east coast of South Africa. The carcass of the unlikely looking fish eventually reached J. L. B. Smith, a British chemist and amateur ichthyologist, who identified it as a coelacanth from its hollow spine and stumpy, paddle-shaped fins. ("Coelacanth" comes from the Greek for "hollow spine.") Coelacanths had been seen previously only in the fossil record and were thought to have become extinct 70 million years ago.

Smith and others searched the waters of the Indian Ocean for additional specimens, but 14 years passed before another was found. In 1952 a fisherman in the Comoro Islands, near Madagascar 1,000 miles to the north, caught a second coelacanth. Since then the catch has generally been three or four fish per year, all from waters near the Comoro Islands.

The fish are typically between four feet and five and a half feet in length and between 70 and 180 pounds in weight. They have thick, scaly skin and lobed fins that resemble degenerate limbs. Ironically, the biologists interested in coelacanths have proved unable to catch them, whereas the Comoro fishermen catch them accidentally but seem to take no interest in them. All the specimens except the first have been brought in by the island fishermen. Investigators must await the chance arrival of a coelacanth in the coastal villages of the islands. Until 1972 no facilities had been set up for quick examination of the specimens; as a result they were badly decomposed and unfit for study.

The expedition of 1972 was conducted by the British Royal Society, the French National Museum of Natural History and the U.S. National Academy of Sciences. Two coelacanths were caught during the expedition, again by local fishermen, but only one was alive when it was brought to the surface. It survived several hours, then died from the effects of decreased pressure and increased temperature. The fish was dissected and its tissues were frozen or chemically fixed; the preserved specimens were then distributed according to a prearranged plan for biochemical analysis, electron microscopy and other kinds of study.

On the basis of fossil evidence the coelacanth has long been considered a close relative of the rhipidistian fishes, the group that eventually gave rise to all terrestrial vertebrates. This classifica-


LOOK AT IT THIS WAY:
EVERYONE THOUGHT YOUR PAPER ON "LASER SATELLITES" WAS PIE-IN-THE-SKY. TODAY YOU CAN'T MAKE ENOUGH REPRINTS.
IT'S TIME YOU STOPPED DRINKING ORDINARY SCOTCH.
Pinch
BY HAIG & HAIG
Old Blended Scotch Whisky
750 ml (25.4 fl. oz.)
Pinch
Twelve Years Old
86° PROOF
PRODUCT
SCOTLAND
Pinch 12 year old Scotch
EXTRAORDINARY TASTE BY HAIG & HAIG
86 PROOF BLENDED SCOTCH WHISKY—RENFIELD IMPORTERS, LTD. N.Y.

tion is based on certain anatomical similarities: for example, both groups have a bony skeleton and lobelike fins. The availability of fresh coelacanth tissue has made possible the first physiological tests of this hypothesis. Surprisingly, the recent findings suggest closer affinities with another group, the sharks and rays. These fishes are members of the ray-finned line, from which almost all surviving fishes are descended. The lobe-finned fishes proved adaptable to land, but they have not thrived in the water; the only survivors of the group today (if the coelacanth is excluded) are the lungfishes, which are freshwater species, and some primitive marine forms.

The evidence supporting a proposed reclassification of the coelacanth with the sharks and rays is both anatomical and biochemical. The shape of the pancreas and pituitary glands and the presence of a rectal gland in both groups suggest a close relation. The similarity that many investigators consider the most compelling is a mechanism of blood regulation shared by coelacanths and sharks and rays but found in almost no other group. All saltwater fishes have blood with a lower concentration of salts than the surrounding seawater; as a result osmotic pressure tends to force water out of the body fluids and to raise the internal salt concentration. Coelacanths, sharks and rays are almost unique in resisting this pressure through a buildup of urea, an end product of protein metabolism.

The revised taxonomy of the coelacanth has not been universally accepted. Defenders of the earlier classification point out that nothing in the recent findings has altered the most obvious characteristics of the fish: it has a bony skeleton and lobed fins, like the rhipidistians and unlike the sharks. Both sides can explain contrary evidence by arguing that related organisms may well diverge in the course of 70 million years, while unrelated organisms may develop similar features independently.

The question is likely to remain unsettled for some time to come because fresh specimens are no longer available. Since 1975 the Comoro Islands have been in political upheaval, and coelacanth expeditions are no longer welcome. On the other hand, specimens may turn up elsewhere in the Indian Ocean; indeed, it is still unexplained why the first coelacanth was discovered 1,000 miles away from the rest.

Steamed Up

Conventional wisdom traces the Industrial Revolution in Britain to three 18th-century developments in steam technology: James Watt's improvement, in the 1770's, on Thomas Newcomen's 1710 atmospheric pumping engine; the use of Watt engines to keep the island's coal mines pumped dry, and Watt's near-monopoly in steam-engine construction until his patent expired in 1800. A census of steam engines built in Britain during the 18th century has now cast doubt on the significance of all three points. Two English scholars, John Kanefsky and John Robey, found that at least 2,191 steam engines were built during the century. Almost half were of the Newcomen type and fewer than a fourth were of the Watt type. Although about 1,100 were pumping engines, they were by no means employed exclusively as mine pumps. Moreover, many 18th-century engines were pirated versions of the Watt design, made in defiance of his patent.

Writing in *Technology and Culture,* Kanefsky and Robey point out that the number of 18th-century steam engines in Britain has long been underestimated, and the variety of their applications has been ignored. In their survey they divide the century into two unequal parts. From 1701 to 1780 all but a few steam engines produced reciprocating motion; the only exceptions were 33 atmospheric pumps made by Thomas Savery. In the last 20 years of the century Newcomen, Watt and other engines that directly produced rotary motion were introduced. In the earlier period, however, rotary motion had often been achieved indirectly. The engine was employed to pump water, which then powered a water wheel. Water wheels driven in this way were often installed at coal mines to hoist coal from the pits.

Summarizing their county-by-county inventory, Kanefsky and Robey report that by 1780 the mining industry had installed 482 steam engines, 376 of them at coal mines and the remainder at lead, tin, copper and salt mines. By 1800 the number of mining engines had risen to 1,064, of which 828 were at coal mines. Up to 1780 only one steam engine had been installed in the textile industry. By 1800 there were 469, more than half of them in cotton mills. Metalworkers had installed 51 steam engines by 1780, most of them at iron foundries, where they supplied water to water wheels, which in turn pumped the blast bellows. By 1800 there were 263 engines in the metals industry, including 217 at iron foundries. Only five food processors, such as flour millers and brewers, had steam engines working by 1780; the number was 112 by 1800. By then 44 canals also had steam engines, as did 36 waterworks, 18 oil mills, 13 paper mills and 12 potteries. The county with the most steam engines was West Riding, which had 271; seven other counties had more than 100.

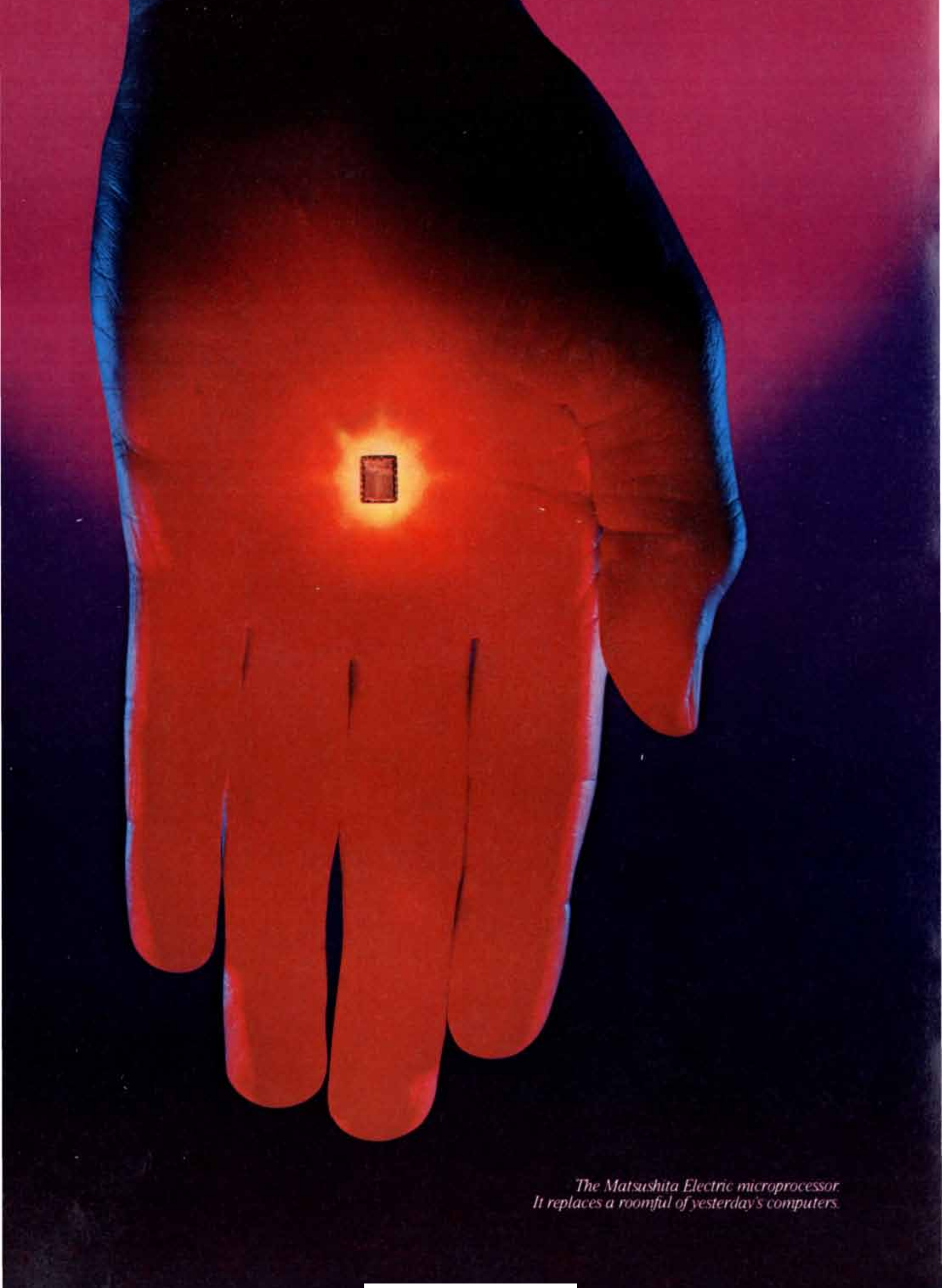
All the early exploitation of steam technology notwithstanding, conventional water wheels continued to provide most of Britain's power for another 100 years. The mechanization of British manufacturing, Kanefsky and Robey point out, was largely a development of the period after 1870.

JAPANESE TECHNOLOGY TODAY

The Japanese draw on a long tradition of craftsmanship as they serve the needs of today's technological world. In this comprehensive report you'll learn how the worlds of research, production, technology, finance, government, and culture have been woven together to make the words "Made in Japan" synonymous with some of today's most advanced and sophisticated products.

SCIENTIFIC AMERICAN

Advertisement


The Matsushita Electric microprocessor.
It replaces a roomful of yesterday's computers.

JAPANESE TECHNOLOGY TODAY

THE AUTHORS

This report is the result of the combined efforts of two of the world's leading experts on Japan's economic and technological worlds—Dr. James C. Abegglen and Mr. Akio Etori.

Dr. Abegglen is vice president of the Boston Consulting Group and a leading expert in the world of international marketing. He has been a consultant to multinational corporations in the chemical, petroleum, food and other industries and has worked with consulting firms in Japan for over 15 years.

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

Their report is based on firsthand knowledge as well as interviews with top executives in Japanese industry.

THE COVER

Symbolically resting on a tray used in the ancient Japanese tea ceremony are two of today's most sophisticated electronic components. The 4-bit one-chip microprocessor (upper right), a product of Matsushita Electric, Japan's largest consumer electronic company, is used in 6-hour VHS video recorders, microwave ovens and Technics high-fidelity components. The IC/Flexible Printed Circuit (lower left), consists of 6 IC chips in three separate packages and was designed and developed by Nikon engineers for use in their most advanced camera, the Nikon F-3, the first camera to offer an LCD viewfinder readout.

Cover photo by Phil Marco.

"Following the war, our nation actively imported technology from the Western advanced economies, devised improvements and also put into effect our own development capability. As a result, the technological level of our nation's industry, while still showing a gap in certain leading technologies, has in most sectors advanced to the top level of the world. Our nation's products have achieved a high evaluation internationally based on a high level of product technology reflected in high quality and reliability, and in precise delivery and effective after-service. These special qualities are to be maintained in the future and the general level of technology heightened.

"Still, in order to deal with various problems of the 1980's, it is not merely a matter of refining existing technology or improving the competitiveness of products. We must advance individual, creative technological development. There is some concern about this capability. However, our nation holds the largest share of patents issued by the United States to foreign nationals; since 1972, technological exports have exceeded imports in new agreements entered into; considering the current high level of technology and the considerable increase in the number of science graduates, we conclude that the potential for technological development is quite high."[1]

HOW GOOD IS JAPANESE TECHNOLOGY

This is the conclusion reached regarding Japan's technology today by the Industry Structure Council of the Ministry of International Trade and Industry in its April, 1980, report, "Vision of Industrial Policy in the 1980's." This report makes an effort to quantify Japan's relative technological position. Using an index of several factors related to technological performance, with the United States as 100, Japan is rated as only 22 in the late 1960's—about equal to France and the United Kingdom, but well behind West Germany's index of 40. Applying the same factors a decade later, Japan is rated at 50, well ahead of France and the United Kingdom, and about equal to Germany.

GENERAL INDEX OF TECHNOLOGICAL LEVEL

	Late 1960s	Late 1970s
United States	100	200
West Germany	40	56
France	24	38
United Kingdom	25	26
Japan	22	50

Index: Using U.S. as 100 the average of index positions in patents registered, technological trade balance, value-added in manufacturing, and proportion of exports technology intensive.
Source: Vision of Industrial Policy in the 1980s, p. 277.

The Industry Structure Council also reports on a survey of views of Japanese businessmen, asking whether in 71 industry categories Japan is superior to, equal to or inferior to the United States in technology. In only two categories, both related to steel, is Japan given a superior rating, while Japan is rated inferior to the United States in 26 industries. However, Japan is seen by these businessmen as technologically inferior to Western Europe in only eight industry sectors.

BUSINESSMEN'S EVALUATION OF JAPAN'S TECHNOLOGY (71 industrial sectors)

	Technological Level (number of sectors)	
	U.S.	W. Eur.
Japan superior to	2	2
Japan equal to	41	55
Japan inferior to	26	8
Not rated	2	6

Source: Vision of Industrial Policy in the 1980s, pp. 279-80.

Like the MITI advisory group, the businessmen see Japan's capability for development as lagging behind current technological levels, although they feel Japan is about at Western European levels in development as well as technology.

Whatever the concerns Japanese might express about the future, the accomplishment is extraordinary. If one accepts these calculations, in only one decade, Japan moved from a technical level one-fifth that of the United States to a level half that of the United States, and in the same decade from one-half the level of West Germany to approximate parity. Other calculations give Japan an even more substantial position in technology. In a recent econometric analysis, it was concluded that Japan had reached U. S. levels of technology around 1972 or 1973, and it was further concluded that a continuation of high investment by Japan relative to the United States would mean a steady shift to Japan's advantage in technology. Since the differential in investment has remained very greatly to Japan's advantage, this leads to the view that Japan now has a measurable advantage in overall technological level over the United States—and, by extension, an even greater advantage over other economies.[2]

[1] *Hachiju Nendai no Tsusan Seisaku Bijion.* Sangyo Kozo Shingikai, Tsusho Sangyo Sho, 1980, pp. 83-4. Translated by writer.

[2] *U.S. and Japanese Economic Growth, 1952-1973: An International Comparison.* Dale W. Jorgenson and Mieko Nishimizu. Discussion Paper Number 566. Harvard Institute of Economic Research, Harvard University, Cambridge, Massachusetts, August, 1977.

JAPANESE TECHNOLOGY TODAY

MORE RESEARCHERS THAN GERMANY, FRANCE AND THE U.K.

There is no question that Japan's current research effort is substantial. It is not, however, especially large in research expenditure. As a percent of gross national product, Japan's research expenditures lag behind those of France and Germany and well behind those of the United States. The effort is greater when seen in terms of numbers of researchers deployed. Japan reports more research personnel than Germany, France and the UK combined—a total of 272,000, or about half that of the United States. This suggests, of course, a rather lower level of expenditure per researcher, but in any event does not indicate an exceptional effort or capability overall.

	Researchers (000) late 1960's	late 1970's	Percent increase
Japan	158	272	+72
U.S.	550	571	+ 4
W. Ger.	72	94	+31
France	55	62	+13
U.K.	57	78	+37

Source: Vision of Industrial Policy in the 1980s, p. 278.

Further, the bottom line measure of technological level—the balance of payments on technology—is very much to Japan's disadvantage. While there has been some slight improvement over the past decade or so, Japan remains very heavily in deficit on the balance between payments for technology and income from technology. For the United States, the payments balance remains overwhelmingly in surplus. No doubt some lag effects are here, especially from early contracts with long contract life. Nonetheless, the figures hardly suggest that Japan has closed the gap in technological development.

The seeming contradiction between the measures of general technological level and the measures of research effort and output makes a point especially important in the case of Japan, but not confined to Japan. Basic research and discovery are important, but not critical to an economy that is seeking to catch up with world levels of output. Nor is invention necessary in a world in which ideas move freely, and in which purchase of technology is possible. The move of Japan at extraordinary speed to parity in technology level needs to be seen as a result of strengths in the economic system and in the system of employment in Japan.

"YANKEE INGENUITY" JAPANESE STYLE

This is somewhat analogous to the experience of the United States. The dominance of the United States in awards of Nobel prizes in science has come after, rather than before, the period in which the United States moved to world leadership in technology. Indeed, for a rather long time the phrase "Yankee ingenuity" referred to the ability of the American engineer to take developments from other countries and apply them to commercial needs and purposes. In other words, the determinants of the technological position of the United States earlier, and Japan today, lie not so much in science, nor yet even in invention necessarily, but rather in characteristics of the economy as a whole. There is a suggestion in this that high levels of investment in research may well be a luxury, not to be afforded by economies bent on maximizing the use of limited resources.

The issue of discovery and invention arises for Japan now, with some urgency, since Japan has attained the heights of currently available technology. MITI's Industry Structure Council properly expressed concern over this issue. However, the explanation of Japan's current success lies in the careful and wide-ranging selection and purchase of world technology, in the speed and extent of its adoption, and in the ability to refine and build on purchased technology. In all of these respects, Japan has performed with unique and unprecedented capability. And it is this phenomenon that requires examination.

ROLE OF GOVERNMENT

There is a widespread tendency to attribute Japan's successes to the exceptional powers and abilities supposed to be exercised by the government of Japan. The field of technology is another case of this misplaced, or egregiously oversimplified, argument. Much, for example, has been made of the subsidy provided by MITI to the electronics industry for the development of very large-scale integrated circuitry. There was, indeed, a subsidy. But it totaled only $100 million over a seven-year period, divided among several producers. By the standards of R&D expenditure of that industry, and the subsidies provided by other governments to their semiconductor industries, the subsidy is trivial.

Indeed, the direct financial support of the Japanese government to science and technology is conspicuously—and perhaps even dangerously—limited. For a very long time, two-thirds of total research expenditure in Japan has been funded by the private sector. That is, while the total research expenditure by Japan is not great, the proportion provided by government is smaller than in any of the other major economies. This is consistent with the general pattern in Japan, since the total direct intervention by government in the economy is less in Japan than in the countries of the West (e.g., total taxes as percent of GNP are only about 20 percent in Japan, compared with over 30 percent in the United States and over 40 percent for most of Western Europe).

The Japanese government summarized its position recently as follows: *"... Japan's ratio of research expenditure to GNP is low at 2.1 percent (as of fiscal 1977) compared with 2.5 percent in the United States and 2.6 percent in the Federal Republic of Germany (both as of 1977). Meanwhile, in Japan the governmental contribution to the total national research expenditure also stands at only 27.4 percent (as of fiscal 1977) as against 50.5 percent in the United States and 48.5 percent in the Federal Republic of Germany (both in 1977) ... As a matter of principle, technological innovation should be the job of innovators—namely, businesses themselves."*[3]

COOPERATION, NOT ANTAGONISM

Perhaps the conclusion should be that government R&D expenditure is largely wasted, and that Japan has benefited precisely because of the small amounts of government funding available for technology, support, and the pressure on the business firm to fund developments.

This too would be an oversimplification. The Japanese government has played an important role in Japan's moving to technological parity with the world. That role is more subtle and indirect than is generally seen, but no less important.

The role of the Japanese government in support of technological advancement can be seen in at least three aspects: 1/ the formulation of industrial policy, 2/ measures in support of that industrial policy, especially in indirect ways through tax and depreciation schedules and industry rationalization measures and 3/ in the earlier period through the 1960's, direct control over technological imports. Most important perhaps, the government's role in relation to technology needs to be seen in the more general context of government-business relations in Japan: a context of cooperation and mutual support rather than antagonism.

"JAPAN INC." MISUNDERSTOOD

The Japanese government-business relationship has by now been much commented on, sometimes in a distorted fashion in

[3]*Economic Survey of Japan 1978/1979.* Economic Planning Agency, Japanese Government. The Japan Times, 1980, p. 137.

In simple terms, why every I

Have you noticed how more and more car manufacturers are switching to front-wheel drive these days?

We're not surprised. We knew the advantages of front-wheel drive a long, long time ago.

In fact, only Honda has sold all its cars in the U.S. with front-wheel drive and a transverse-mounted engine. (Not to mention having sold more than anyone else too.)

So just what *are* the advantages?

For a start, we've designed a car that's smaller on the outside but with more than enough room inside for both driver and passengers.

This is partly due to the fact that there's no driveshaft running through the passenger compartment. So your passengers have more foot and legroom.

Eliminating the driveshaft also reduces another problem. Because the driveshaft can be a source of noise

which the two sectors are seen as conspiring together. This view—"Japan, Inc." taken to an extreme—takes no account, either from distance or from ignorance, of the diversity of views and the frequent conflicts of interest that must arise in so large and complex an economy. The relevant aspect is that the business community of Japan tends on the whole to see the bureaucracy of Japan as competent and well-intentioned and as working toward the national interest. (Rather less respect is paid elected government officials, but they play little part in these issues of technology.)

For their part, the bureaucrats concerned with industry, trade and finance are generally active in support of Japan's businesses as the necessary and appropriate instruments of economic growth and well-being. Further, the government tends to be pragmatic rather than doctrinaire in policy, this, too, suiting the business relationship well. For example, anti-trust enforcement in Japan is not cast in a moral mode as in the United States. Monopoly is seen as hazardous and costly, and warranted only under special conditions. However, cooperation between business firms under specified conditions is not necessarily seen as immoral. The implications for industrial policy in general, and technological development in particular, are important: anti-trust and other policies are a means to an economic end, and not part of a higher morality.

Thus, for example, government-sponsored programs in pursuit of a new technology can be made more effective by cooperation between otherwise competing firms. The LSI program referred to earlier was an example. Although the total government funding was small, it was the more effective by being channeled through an intra-industry committee which allocated funds and research tasks, and arranged the sharing of the results. The LSI development program is seen as important to long-term industrial development, and an anti-trust policy was adapted to meet the national requirement.

This directing and coordinating role of government ministries is informed and given justification by Japan's overall industrial policy. That policy, formed through a long and complex process of study and discussion adjusted at frequent intervals, is given the support of the community generally. The policy is broad and general, as good policy must be. It is not a detailed plan, and has, even in its planning aspects, no force of law or other discipline. It is rather an explicit and rational view of how the Japanese standard of living can be improved.

INDUSTRY'S SOPHISTICATED POLICY

Briefly, Japan's industrial policy notes that a nation's income is a result of the productivity of the nation's resources of capital and labor. It is through the improvement of the technological level that these resources are made more productive. This leads to the conclusion that the nation must steadily shift its resources of capital and labor to higher levels of value-added, to more sophisticated sectors technologically, if steady improvement in living standards is to be achieved.

Like any policy, however well-founded, Japan's industrial policy in execution often yields to some extent to political and strategic considerations. Agriculture is an example. While Japan's yields per land unit are among the highest in the world, the productivity of agricultural workers is low and product cost high. Yet the familiar political problems of conservative governments trying to deal rationally with farmers handicap Japan no less than other countries. Also, abandoning domestic food production in favor of more cost-efficient imports poses a strategic problem like that facing Japan in terms of supplies of energy and most raw materials. Thus Japan's agricultural policy is not simply economically rational, but is a complex outcome of conflicting requirements.

Yet, Japan's industrial policy is applied, usually effectively. The shipbuilding crisis of the last few years is a good example. Japan supplied about half of the world's shipbuilding market, with very efficient yards and leading technology. However, much of shipbuilding is labor-intensive and steel-intensive. With low-cost steel, and with a relatively low-cost and very flexible labor force, Japan led in such low value-added sectors as the building of tankers. This is not the kind of industry in which the resources of a highly developed economy should be invested.

HOW TO LET A DYING INDUSTRY DIE

With the collapse of the world shipbuilding market after the 1973 oil crisis, Japan was left with enormous surplus capacity. The response, painful but entirely rational economically, has been to close down 40 percent of total capacity. Government funds were made available to assist in buying back and scrapping excess capacity, but in fact only about half of these funds have been used while the target reduction of capacity was achieved. A government-approved cartel was established to allocate production, banks worked together to provide emergency funding, and the industry organized itself to plan and oversee the capacity reduction. All yards deemed obsolete have been closed, and the remaining yards are the newest, specializing in more sophisticated ship production, with the highest efficiency. The remaining shipbuilding industry in Japan should be notably productive and efficient as the market recovers.

Simple enough, one might say. Yet this is hardly the response of the other major shipbuilding economies, where funds were devoted to support dying yards rather than to closing them. It is the wide recognition and understanding of economic forces, the ability to elicit cooperation, and the ability to move toward the economically rational solution that distinguishes the policy of the Japanese government.

Note the implications for technology. In Japan, only yards less than a decade or so old will continue in production. All will be the most efficient of Japan's efficient industry. Several of the smaller companies have combined, with the resulting potential for larger-scale, more efficient operations. This raises the level of technology in the industry, but without invention, development or significant technological change.

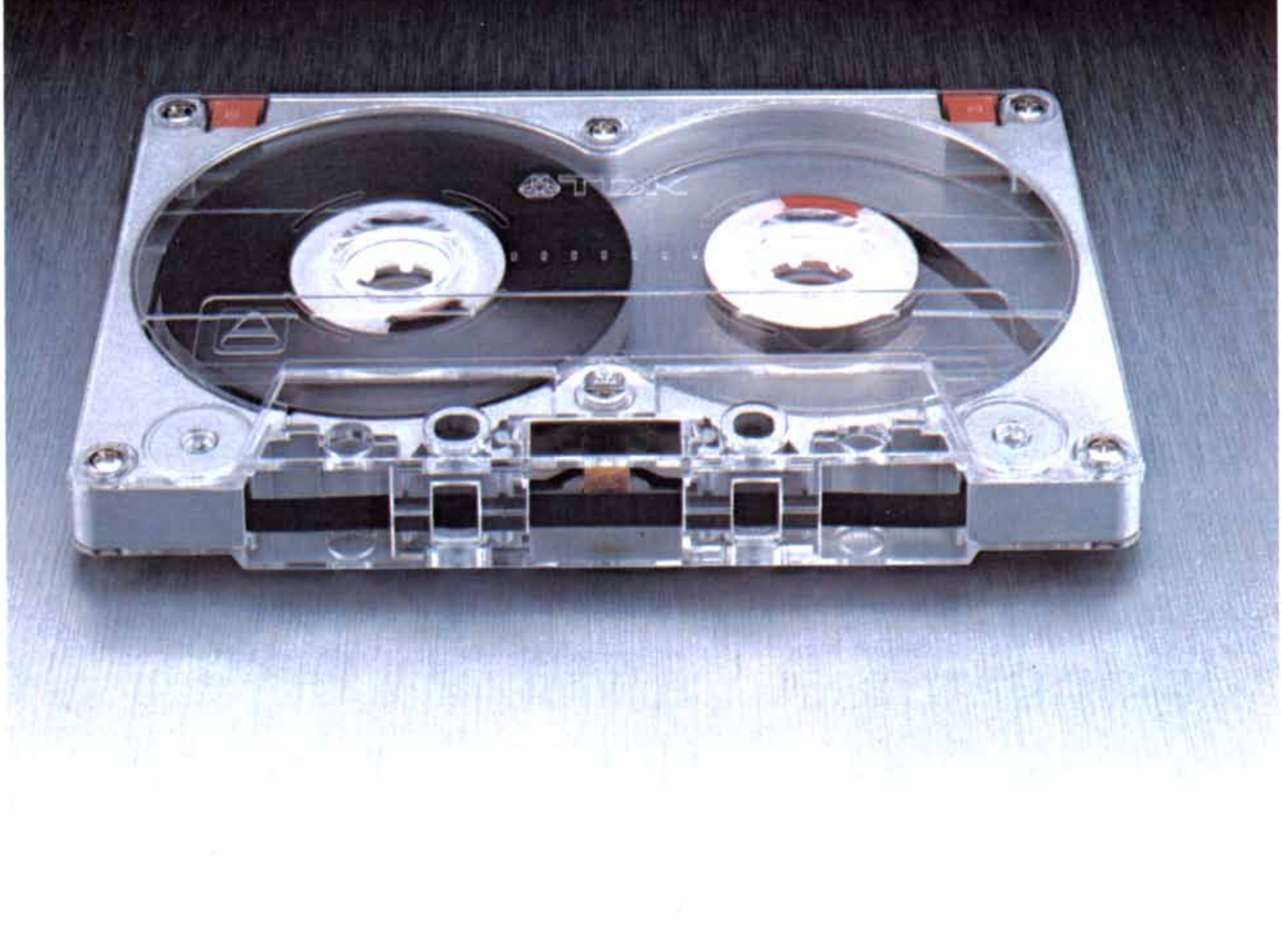
Trade policy is another sector that is informed by overall industrial policy, and that has direct relevance to technology. The contrast with the United States is especially marked. U.S. trade policy has as its central concern the protection of industries in trouble. One need only remember the enormous political efforts made by the United States to protect the domestic textile and shoe industries to appreciate the U.S. position. Again, as the U.S. steel industry began to encounter the inevitable result of its long failure to invest to improve its productivity, so-called orderly marketing agreements—an exercise in hypocritical protectionism—were forced on U.S. overseas suppliers. More recently, a laggard television industry was granted the same protection. Autos are now being proposed for a similar kind of shelter.

ZERO TARIFFS ON AUTOMOBILES

Japanese protectionism, leaving aside agriculture, has had an opposite thrust. Protection has been provided those industries that are in need of protection because of their newness and their fragility as emerging industries. Thus protection is negotiated for the semiconductor and computer industries, and telecommunications. Textiles are a mature industry and, in the large, not appropriate for Japan's level of development. Therefore, Japan's textile tariffs are among the lowest in the world, and textile imports increased more than 40 percent annually through much of the 1970's. Similarly, the auto industry is mature and doing well. Japanese auto tariffs are now zero, and imported cars are, in the bargain, exempt from Japan's stiff emission control requirements.

The impact on technology level is again direct. Sectors of high value-added, and

We've got the future on tape.



TDK has been shaping the future of recording tape since 1952, when we introduced our first tape. Over the years, TDK breakthroughs have changed the way people see and hear.

In 1968, TDK SD became the world's first high fidelity cassette. Its outstanding performance made it NASA's choice to take to the moon.

In 1973, TDK introduced Super Avilyn. This revolutionary magnetic particle created a new set of standards for audio and video tape. Today the research continues, raising recording standards even higher. From TDK SA, the high bias reference standard cassette, to SA-X, which surpasses it. To the most recent Super Avilyn advance, High Grade home videotape, which made brilliant 6-hour pictures possible. TDK also explores the far reaches of metal sound. The classic MA-R cassette is at the forefront of metal recording.

These technological achievements are part of TDK's philosophy. A way of seeing the future as a challenge to be confronted now. It's evident in the standard-setting performance of our audio and video tapes. It was evident in the initial pioneering which established TDK as the leader in ferrite technology in 1935. This scope of technological growth has led to an impressive increase in market share. Today TDK is an acknowledged leader in the quality magnetic recording tape market.

Tomorrow has already begun.

TDK®

The future in sound and pictures.

high technology, with high growth potential, are afforded as much trade protection as can be arranged. This allows a nurturing of technology in the domestic market until competitive scale and sophistication are achieved. Conversely, low-value-added, low-technology, and low-growth sectors are exposed to trade competition to help force a shift of capital and labor out of those sectors and into higher technology sectors.

All of this is, however, rather far from the image of a government that possesses some special insight or foresight in its planning for technology. There is a view in the West that somehow the Japanese government has been especially clever in predicting the future of technology. This is simply not the case. MITI's so-called national research projects have not provided notable output—witness the long efforts to develop an electric car, for example. Indeed the Japanese government has been distinctly in error in its forecasts rather often, in much the same fashion as governments everywhere. The familiar story about the resistance to Sony's initial efforts to license transistor technology is an example. The government's total failure to foresee IBM's third-generation computer technology is another.

It is, in fact, not the affirmative actions of the Japanese government in support of specific technologies that make its role valuable. It is rather the general policies pursued to encourage growth and encourage the move to higher levels of value-added that make the government effective. We have again an instance of the adage that it is not in the power of governments to do great good, but rather they have the power to prevent harm. Tax policies that encourage new products are an example. There was an interesting case a few years ago when the government allowed a cutback on the high commodity tax on color television if the set employed solid-state technology. Clearly, the industry moved fast to employ semiconductors.

Depreciation schedules are a better example. Fast writeoffs on new investment do not require that the government make a judgement on the merits of a particular investment. It is only required that the government understand that new investment will, by definition, tend to incorporate more advanced technology. Therefore, any tax or depreciation schedules in support of new investment in general will have the effect of raising the overall level of technology with the judgement as to the usefulness of the technology left to the entrepreneur and the marketplace rather than to a government decision. Japan's success is a powerful argument in favor of less government rather than more, in favor of the competitive marketplace rather than central planning, and in favor of policies to encourage vigorous investment rather than to stimulate consumption.

This emphasis on the indirect role of the Japanese government in support of technological advance is not to say that the role has always been so indirect. In the very earliest stages of Japanese industrialization, in the 19th century, the Japanese government took the initiative in establishing industries in a number of cases, turning them over to private investors after the initial period. (Indeed, this history no doubt explains in part the positive attitude toward government on the part of the business community today.)

More to the point of this review, during the immediate postwar period the government played a very direct role in supervising and regulating the inflow of foreign technology to Japan. The context needs brief review. Japan went to war in the mid-1930's. The long focus on military production, and the long isolation from the advanced economies, added to wartime devastation of plant and equipment, meant that in the mid-1940's, Japan was virtually bankrupt of modern technology. The military situation influenced Japan's industrial decision-making process.

THE GREAT "MAKE-OR-BUY" DECISION

Under these conditions, a critical initial decision was the classic make-or-buy decision. That was rather easily resolved; the need was so great and the availability quite sufficient to lead to a buy decision. Indeed, the need was so great that Japanese companies, out of desperation, were prepared to pay any price for foreign technology. At this point the government intervened, to regulate both the price paid and the length of the agreement—another form of price. All import of technology to Japan required specific government approval, with the government an active participant in the negotiations. This state of affairs continued essentially until the end of the 1960's, when it was judged that the marketplace could work without assistance and without too great a cost to Japan.

In retrospect, two aspects of this period and of the nature of government control are of special interest. The first is that the government did not attempt to tell individual companies which technology should be purchased. Japanese companies studied the world, sending teams scurrying through plants and research centers everywhere attempting to determine which technologies were best. The government intervened in terms of price, but did not presume to judge the appropriate technology. Again, the view that the Japanese government exercises a special prescience in respect of technology seemed unwarranted.

AVOIDING THE TECHNOLOGY MONOPOLY

Another aspect of government policy during this period is worth noting. There was an explicit effort to try to avoid monopoly. That is, when one Japanese company sought a technical agreement that might give it market advantage in the domestic market over competitors, the competitors were encouraged to seek foreign technology for their own operations. An example is computers. When Hitachi indicated that it was entering into an agreement with RCA for computer technology, the government approached a major competitor, Nippon Electric, and indicated that it would hold the Hitachi-RCA agreement up until Nippon Electric decided on a foreign licensor. Nippon Electric chose Honeywell as their technology source. But in the final event, all were wrong, since RCA got out of computers and Nippon Electric is unwinding its connections with Honeywell. In any case, here we see the government both leaving the technology decision to the private sector, and also seeking to avoid monopoly advantage through foreign licensing.

This effort to avoid special technological advantage can also be seen in two famous cases of foreign technology and investment in Japan. In the late 1950's, IBM, which had long had a wholly-owned operation in Japan, sought to import technology to enter into computer production in Japan. The Japanese government made as a condition for that import, the availability at fair market price to all interested Japanese firms of the key patents held by IBM. IBM finally agreed, and is today Japan's largest computer company with $2 billion in sales. One might infer that the Japanese government wanted IBM in Japan, while wanting a competent domestic industry as well. The patent access was useful to both ends.

The Texas Instruments case was a rather different example. TI applied for a 100 percent investment in Japan at a time when the Japanese government required that foreign investors hold no more than 50 percent equity in a manufacturing firm. This was in the mid-1960's, and as a bargaining ploy TI held an absolutely basic patent in semiconductors, one that Japanese companies had to have access to if they were to export semiconductor-using products to the United States. After a long delay on the part of the Japanese and then after the U.S. government intervention at the highest level, TI was allowed 50 percent investment, provided patent access was available to the domestic industry, and provided the partner was one of six companies specified by the Japanese government. The final outcome was a joint venture with Sony with an understanding that the ownership would revert


Look at television the way we do.
Television should be more than an electronic tranquilizer.
At its best it is art, and like any art form it should involve you, touch you, *move* you.
Experience the art of television at its best, on the one television that is itself a work of art: the Emmy Award-winning Sony Trinitron.
From the world's smallest color screen to the largest available in the U.S., there are 23 different models, and every one is every inch a Trinitron.
Experience television at its finest.
Look at it our way.
SONY
THE ONE AND ONLY
SONY
Model shown: KV-1945R
© 1980 Sony Corporation of America.
Sony and Trinitron are trademarks of Sony Corporation.

totally to TI after three years, but with patent access a condition of the agreement.

THE PRIVATE VS. PUBLIC TECHNICAL DECISION

The conclusion must be twofold. The technology import decisions were basically made by private firms, with the government providing an important, and from the companies' point of view, a very helpful role as price controller. Further, the government consistently sought to avoid technological monopoly, and rather aimed to make the technology available to domestic competitors.

The role of government bureaucracy in Japan's technological development has been, and is, important. But the importance derives not from the simplistic role often assigned, of perceiving sectors of importance and pouring funds in support of them. Rather, the role of the government has been a broadly facilitating one, leaving the specific decision about technology to the private sector and supporting private sector initiatives with the full array of public sector instruments—tax, depreciation and trade policy; joint and cooperative programs within and with industry; surveillance of industry decisions in terms of cost and in terms of maximizing domestic competition.

The explanation, therefore, of the success of the Japanese in moving to ever higher levels of technology must be sought elsewhere. It is not to be found in terms of a simple answer. Rather, the explanations lie in the dynamics of the Japanese economic system, and in the support structure in aid of new technology that results from the employment, labor relations and trade union system of Japan.

THE BASIS FOR TECHNOLOGICAL PROGRESS

Japan's technological progress is not the result of government programs and funding. Rather, it is the result of a highly competitive market economy, in which companies are required to invest heavily to maintain market position. This high level of investment is made possible by a high savings rate, both individual and corporate. As will be noted later, this process of competitive investment is greatly aided by a system of labor relations that greatly facilitates the investment process. An understanding of the Japanese achievement and potential in the area of technology must, however, begin with an appreciation of the very high investment levels that have characterized modern Japanese history.

Investment is, of course, the correlate of savings. It seems that there is no very clear or definitive explanation as to why a given country has had for very long periods a characteristic level of savings and investment. One might speculate that the keen Japanese sense of being a poor nation with few resources and, thereby, a need to husband available resources, explains the Japanese pattern. Thus the Americans, with a long-held sense of infinite space and resources, have a more profligate pattern of low savings and investment. In this notion, the basic attitudes eventually become embedded in tax and other policies that perpetuate the underlying attitudes.

Or again the Japanese sense of being besieged in an unfriendly world, of being isolated and subject to imperial predators, may play a part in determining savings behavior. It has also been noted that Japanese traditional, pre-industrial patterns of food, dress and housing continued well into the modern period. Thus the productivity returns from industrialization were available for savings and investment from the lag in modernization of consumption habits.

Whatever the explanation—and the matter is no doubt multi-determined—there can be no question of the results. For a long time, Japan has had the highest levels of capital formation, and of private investment in plant and equipment, of any of the developed economies. For Japan, as for other economies, the post-1973 depression brought some reduction in the overall level of investment. But Japan's distinctive relative position remains as before.

JAPANESE INVESTMENT LEVELS REMAIN HIGH
Gross Fixed Capital Formation as Percent of GNP

	Average 1970-1974	1975	1976	1977	1978
Japan	34.9	32.2	31.0	30.1	30.2
West Germany	24.9	20.7	20.6	20.8	21.5
France	23.8	23.2	23.2	22.3	21.4
United Kingdom	19.1	19.6	18.9	18.2	18.0
Italy	20.9	20.6	20.1	19.7	18.8
United States	18.8	16.3	16.4	17.2	18.0

Source: Union Bank of Switzerland
IMF

$3 OUT OF EVERY $10 REINVESTED

The Japanese reinvest $3 of every $10 produced. The United States invests less than $2 and, in fact, invests less than any other major economy in proportion to its output—even less than the United Kingdom. It is Japan's high level of investment from savings, of course, that accounts for her more rapid economic growth. It is the investment that makes possible the great advances in productivity, in product quality, in product design and in product innovation that has marked Japanese output in recent years.

BANK SAVINGS ARE AT A HIGH LEVEL
Bank Savings Deposits

	Total Bank Savings, 1977 ($ Billion)	Bank Savings per Capita, 1977 ($)
Japan	1,085.2	9,531
West Germany	235.2	3,830
France	180.2	3,394
United Kingdom	139.3	2,494
United States	943.9	4,354

(Savings deposits with local savings banks, post office and central savings banks, commercial banks, and other financial institutions, converted to U.S. dollars at year-end rates.)

Source: Union Bank of Switzerland "Business Facts and Figures"

The effects of this investment rate can be seen throughout Japanese industry. The current guess is that of the total of industrial robots deployed in world manufacturing, one-half are in Japan, with one-quarter in the United States and one-quarter in Western Europe. The story of steel is well-known, with Japan having world-scale facilities, oxygen furnaces and continuous casting in a proportion far greater than any other nation's steel industry. The investment in redesign of television sets reduced the number of parts in Japanese sets well below the numbers in competitive sets, with great benefits in terms of both manufacturing cost and reliability.

As noted, this high level of investment, which translates under competitive conditions to a steady improvement in technological level, must rest on a high savings rate. A principal source of these savings is the Japanese individual. In sharp contrast to the United States, the Japanese save a higher proportion of their after-tax income than most other people: about 20 per cent for each man, woman and child. In the

JAPANESE TECHNOLOGY TODAY

event that the savings rate notion seems a bit abstract, it is helpful to note some data compiled by the Union Bank of Switzerland for 1977. Total bank savings in that year in Japan were $1,085 billion, compared to $944 billion for the United States with twice the population. That is, on a per capita basis in 1977, the average Japanese had bank savings of $9,531 compared to the average American's $4,354.

These savings, as well as those of corporations (and, in some years, government), flow to the private sector for investment. (It has already been observed that the government's share of cash flow through tax collection is unusually low in Japan.)

It is, of course, entirely possible to have high savings and investment, and yet relatively low growth and relatively low levels of technology. Most of the socialist economies are testimony to this alternative, with essentially forced savings through government intervention, and inefficient central government investment of the savings.

JAPAN, A FIERCELY COMPETITIVE MARKET

Thus it is essential to complete the view of Japan to appreciate that the Japanese economy is a market economy, and fiercely competitive. Again, like misperceptions of the role of the Japanese government, this fact is frequently not appreciated by commentators on Japan. Perhaps distance blurs corporate distinctions. In any event, the competition for market share in Japan is inexorable. This is especially true in a market economy under conditions of high growth, since it is with high growth that share positions can most readily change. The firm that fails to add capacity under conditions of high growth quickly finds itself out of the race. As growth slows, the final successful survivor is the firm that has invested most heavily and efficiently. Since growth in Japan has been especially high, it is no surprise that competitive investment has been heavy.

This sets in motion a virtuous cycle. The investments in plant and equipment, in cost improvement and product improvement, that make for success in the highly competitive domestic market, are the very factors that make export competitiveness possible. The numerically controlled lathe is an example. The high investment rate of Japanese companies, directed in part to reducing labor input and labor costs, provided a wide market for the NC lathe. The exploitation of this domestic market led to a cost and quality position in NC lathes that has given rise to an explosive growth in exports over the decade. This allows further investment and product innovation. Just as the U.S. cycle of low savings, low investment and declining export performance leads to inflation, currency crises and declining real incomes, so the inverse of the cycle is enormously rewarding.

THE INDIVIDUAL TAKES ON THE GIANT

It might be noted that the leader in this NC lathe field is not a large and internationally known company, but is Mori Seiki, a firm that was until recently small and is still privately held. The competitive, innovative thrust in Japan has tended not to come from the companies bearing the great and famous names of history. Seiko, leading the world in watch production, is a family firm. Honda, leading in motorcycle production and making an impressive run at the auto industry, is a post-war, independent success. The highly individual business philosophy of giant Matsushita Electric's founder helped it become Japan's largest consumer electronics company and is one reason its *National, Panasonic, Quasar,* and *Technics* products have worldwide respect. The list of *non-saibatsu* firms of world position is a long one—Shiseido, Pioneer, Sony, Fuji Film, Canon, Toyota, Casio. These names alone should be sufficient evidence of the degree of competition that prevails in those industries in Japan that have achieved greatest international success in technological innovation. These were not firms that began with special advantage, either with government or banks, and these were not firms that were nurtured by group affiliations. In the fast-moving sectors of high technology and consumer marketing, the seemingly favored firms of Japan have generally not prospered.

The search for the key to Japan's economic and technological success is all too often focused on the policies and plans of the Japanese government. It is a competent government, at least in its bureaucratic structure, and deserves credit. It has not earned position as the principal cause of success, however, and is best seen as facilitator and moderator. In an ultimate sense, the real cause lies in whatever it is that has led the Japanese people to be savers and investors, and to compete vigorously for economic advantage. Without the competitive thrust, investment languishes and savings lie fallow.

QUALITY—A PRODUCT OF LABOR OR AUTOMATION

It is this investment level that explains the high quality of Japanese products. There is a quite understandable tendency of many Japanese to attribute the high quality of Japanese products to the diligence of the Japanese work force. As will be noted, the work force is indeed a key factor in the move of Japan to higher technology. But the explanation of high quality lies in the investment in design, and in the investment made to take labor out of the product. That the quality of labor is high in the bargain is a great advantage. However, it is investment in automation that explains the consistent quality. Matsushita Electric's Panasert, for example, is not just a machine, but a patented process that automatically selects, inserts and affixes components so that color TV printed circuit boards—the heart of the chassis—can be assembled from start to finish without human hands.

A startling piece of evidence on Japanese quality levels was offered recently by a U.S. businessman. He reported the experience of his company in purchasing semiconductors of a particular type from three Japanese and from three U.S. suppliers.

PRODUCT QUALITY DIFFERENCES: 4K AND 16K RAM'S

	Failed Test on Arrival (Percent)	Field Failure per 1000 Hours (Percent)	Hewlett-Packard Quality Index
Japanese Supplier			
A	0	0.01	89.9
B	0	0.019	87.2
C	0	0.012	87.2
U.S. Supplier			
X	0.19	0.09	86.1
Y	0.11	0.059	63.3
Z	0.19	0.267	48.1

Source: Richard Anderson, Computer Division, Hewlett-Packard as reported in *The Economist,* April 26, 1980.

The quality of the Japanese product was high. This is of special interest because semiconductors are an instance of the U.S. producers taking the labor-intensive part of semiconductor production—the wiring of the chip—off-shore to low labor rate countries to achieve cost savings. The Japanese, handicapped by their inability to fire their labor force and by their relative lack of experience in managing foreign labor, invested heavily to automate what had been the labor-intensive part of semiconductor production. The quality differences reported by Hewlett-Packard are the result of that heavy investment.

LABOR RELATIONS AND TECHNOLOGY

High savings, leading to high investment, in a competitive economy. That is the message of Japan regarding technological advancement. To complete the pattern, however, it must also be noted that the system of employment and labor relations of Japan provides exceptional support for the intro-


Creating, with brush and ink, the characters of written Japanese,
Kanji, is more than writing. It is also a meditation and an art form.

duction and diffusion of technology. Indeed, it is perhaps not too much to say that none of the Western systems of employee relations, with the possible exception of Germany's, would allow the rate and degree of technological change that has taken place in Japan in the post-war period.

The Japanese labor force has many advantages in terms of technological change and advance. It is, first of all, a highly educated labor force.

THE EDUCATION LEVEL IS VERY HIGH

	Full-Time School Enrollment, 1977 (Percent Of 15-19 Age Group)
Japan	71
West Germany	42
France	55
United Kingdom	45
United States	72

Source: OECD, "Observer March, 1979"

Indeed, given the intensive nature of Japanese primary and secondary education, it may well be at the shop floor level the best educated work force in the world. Dropouts from the educational system are few, and the U.S. concern with functional illiteracy seems not to be a concern in Japan. The educational system also seems well-designed for technological matters. Earlier studies by Prof. B. Bloom at the University of Chicago indicated that Japanese secondary-level students score well above the youth of other nations in mathematics, when standard tests are administered.

The labor force is homogeneous. There are no divisions in race, ethnic background, religion and the like that can make management difficult. Further, the value system associated with the Confucian ethic—especially with regard to respect for elders and superiors—that is pervasive in Japan is very useful indeed in the context of a large corporate structure. Finally, the labor force of Japan is still fairly young, although this factor changes as the age structure of Japan shifts sharply upward over the next decades.

The employment system of Japan has patterns somewhat special to the country, patterns that in the past were often thought to be economically dysfunctional, but that are increasingly seen to be positive to growth and technical change. The general pattern for male employees of large corporations is to enter the company directly from school. The larger, more prestigious and faster-growing companies can usually be quite selective in hiring, and employ batteries of tests and interviews to gauge the intelligence, personality and values of the young candidate.

It is important to note that the employee is not selected on the basis of a particular skill. Candidates are broadly divided between those with technical and those with non-technical education. Within these groupings, however, the worker is hired for general potential and is expected to acquire specific skills on the job. The company invests large amounts of both money and time in the training of personnel because the employee expects to be with the company for his entire career, without layoff or discharge, and the company expects that the employee will not seek employment elsewhere.

HOW IMPORTANT IS THE LOYALTY FACTOR

This employment pattern creates an unusually close tie between the individual and the firm. It is this pattern that causes much discussion of the unusual loyalty of the Japanese employee. The term is perhaps misleading. The tie with the company is entirely explainable in economic terms. It is very hard for a man to change jobs in Japan between Japanese companies, since in doing so his quality is immediately suspect. It is very hard to take an employee later in his career, both because of resistance by the informal structure and because the system of compensation and reward is closely tied to seniority in the organization. And, by the same token, a man leaving a company is forgoing a good deal in potential income and benefits. Finally, of course, job security has a real value that can be traded off against possible short-term income gains.

This close identification with the company is reinforced by a compensation system that is heavily weighted to seniority. Wages rise with length of service basically, with additional components of "merit" and rank. In addition, semiannual bonuses are an important component of compensation, ranging from perhaps a quarter of annual income up to half or more. This has the effect of tying the long-term well-being and prosperity of the company directly to the well-being and prosperity of the employee and his dependents.

The linkage between employee and employer, far from being attenuated by the widespread presence of trade unions, is strengthened. The Japanese employee is not naive; he has a labor union representing his interests in wage and bonus negotiations. However, his trade union is made up of all the non-executive employees of his company—and only the employees of his company. The existence of the trade union depends on the continuing existence of the company. Further, Japanese trade unions—like most unions everywhere—are highly conservative in their strong support of job security and seniority privileges. Therefore, the union tends to reinforce the employment system.

PRODUCTIVITY IS SUPPORTED BY PEACEFUL LABOR RELATIONS
Man Days Lost To Strikes Per 1,000 Employees

	1970	1977
Japan	120	40
West Germany	4	190
France	110	130
Italy	2,230	430
United Kingdom	480	410
United States	940	450

Source: Japan Productivity Center

Note that the Japanese worker is not organized into skill unions. The single-skill, or occupational union in Japan is the seaman's union. (Leave to one side the unions of doctors and the like that go by other names and are not industrial.) There is nothing in the union structure that prevents job reassignments or reclassification. Nor is there any such provision in the union contract, and that contract does not deal with manufacturing technique issues. Thus the potential economic disadvantage arising from the lack of inter-company worker mobility in Japan is largely balanced by the unusual amount of intra-company mobility that the employment and trade union systems make possible. The union is concerned with maintaining overall employment, not with defending specific skill categories.

WHY NEW TECHNOLOGY IS NO THREAT

In terms of technological change, this labor force and its mode of organization is exceptionally supportive. In the Western system of employee relations, new technology is a real threat. By definition, it threatens job security and compensation. In the Japanese case, there is no threat. Job security is assured, and compensation is only slightly linked to a particular task. Thus, to the extent that new technology, whether process or product, will make the company more prosperous, the worker has every reason to encourage its introduction. Further, the worker has a sufficient level of education to be able to deal with the new technology. And the company, obliged to continue the worker's employment, has every incentive to train him to do so.

It is this system of employment and compensation that makes possible the worker's concern with product quality that has so impressed foreign visitors to Japan. There

JAPANESE TECHNOLOGY TODAY

has recently been much interest in the so-called Quality Control Circles of Japan, where workers meet to improve product reliability and production efficiency. Attention to a particular device like the QC circle may be a case of mistaking form for substance. The important fact is that the basic system of employment in Japan is more secure and less conflict-ridden than in most countries. To an unusual degree, the potential for confrontation between management and worker has been minimized, and a considerable identity of interest has been established. On that base, QC circles and other devices can prosper. Their effectiveness in many U.S. companies might be less.

(Nearly twenty percent of the presidents of Japan's largest companies were previously the heads of the unions in those companies. A young man who is chosen by his peers to head the union moves up in rank and out of the union, but has been marked as a comer. Not surprisingly, discussions of co-determination and other devices to share management power with union officials have met with little response in Japan.)

MANAGEMENT'S SUPPORT OF TECHNOLOGY

The Japanese workplace, therefore, is highly receptive to new technology, and is entirely capable of absorbing it. Factors supportive of technological change are by no means confined to the shop floor; management shares these interests. There are additional aspects of management that would appear to be supportive of technological developments in the company. First, nearly all Japanese executives are university graduates, in contrast to many of those in Western Europe. Their interest and appreciation of new technology is thereby that much greater.

The continuity of management would also seem important. Programs of research and development can be carried through and are not subject to jarring discontinuity owing to sudden changes in senior management. Nor are programs subject to plunder by a senior research or management official leaving to set up his own firm. Assuming that continuity does not stifle initiative—and for most Japanese companies it seems not to do so—this continuity must be helpful.

Two additional management issues noted are first, the Japanese company is under extraordinary pressure to grow, not simply owing to the wishes of shareholders primarily. Every person in the company sees his future prosperity as dependent on the growth and success of the firm. The firm cannot cut back the labor force, or sell a part of its business, to deal with economic downturns. It must fight its way back by internal growth—again, the theme of high investment rates, with the notion that each product group in the company must seek improvement and innovation.

And the growth must be internal. It is a corollary of the employment system that companies in Japan, or even parts of companies, are not bought and sold. Since the employee is an integral part of the company, it cannot be dealt with as a simple aggregate of assets and liabilities. Further, combining two work forces under these employment conditions is exceptionally difficult. On the whole, one might conclude that this inability to diversify the acquisition, or to cut losses by sale of a business, is supportive of solid technological change and improvement. Growth must be generic; programs must be pursued to a conclusion; innovation must come from within.

With this, Japanese management is not subject to the kind of short-term earnings-per-share pressure that characterizes much of U.S. business. The stock market tends to value growth—that is, capital gains potential—more highly than short-term changes in profitability. Further, the executive's compensation is not related directly to share price since stock options and the like are not available. All of this tends to make for a longer time horizon, and a willingness to undertake investments that offer no near term advantage but will pay out handsomely in the longer term. Thus there is both executive continuity, and a need and willingness to make long-term investments.

It is in this highly supportive context of labor and management characteristics that competitive investment is made. The potential for technological change is high from the investment level, and the opportunity for introduction is high from the nature of the relationships within the company. The result is Japan's leap to technical parity, and even leadership.

FREEDOM FROM THE "NOT-INVENTED-HERE" SYNDROME

There doesn't seem to be an explanation for Japan's freedom from the NIH syndrome—the rejection of things Not Invented Here. That Japan has, from its early history, periodically taken in from abroad a great range of things is well-known. Why so proud a people should also be so willing to learn from abroad and adapt and improve on foreign learning is something of a mystery. Whatever the reasons, Japan's businesses seem immune to the NIH problem.

"Japan's technological progress has been achieved so far primarily through the introduction of foreign technologies. This has been inevitable, it may be said, because Japan made a late start and therefore had to catch up to advanced nations in a short period. After the oil crisis, cases of foreign technology introduction diminished in a similar development as experienced by private capital spending, but still remained at a high level. In fiscal 1978, Japan's earnings from exports of technologies were equal to only 22 percent of its payments for foreign technology introduction. Nonetheless, the low ratio itself does not necessarily give cause for worry; the ratios of Britain and France stood at 100 percent (1975) and 120 percent (1976), respectively, compared with 41 percent for the Federal Republic of Germany (1977). A brisk introduction of foreign technologies may be taken as an indication that a country has a great capacity to assimilate them." [4]

BUYING THE TECHNOLOGY OF THE WORLD

The acquisition of technology from abroad by the Japanese has been discussed. What needs to be appreciated is the massive scale of that acquisition. From 1950 through 1978, Japanese entities entered into a total of some 32,000 contracts for the input of new technology to Japan. As noted, the government played a part in controlling the cost to Japan of this importation of technology. It seems to have done so with some skill. The cumulative cost of the purchase of essentially all of the technology in the world has been only $9 billion over the period. Current U.S. research expenditure is variously estimated at about $50 billion. In other words, for a fraction of the U.S. annual expenditure, Japan closed the technology gap.

It is through these figures that one can appreciate the extent to which the workplace of Japan provides a supportive setting for innovation. One might argue that this massive flood of new technology could not have been introduced effectively into the workshops of the United States or Great Britain. The system of employment and of union relations would simply not have allowed it, even assuming a labor force of sufficient adaptability to incorporate so great a series of changes.

From the point of view of the seller of the technology, much of this sale to Japan represents disaster. At the time of sale, the prospect of windfall income from R&D write-offs was attractive. In the early period at least, Japan seemed remote and harmless enough, not worth the effort to enter its marketplace directly, and sale of technology seemed a good move. In actuality, it created competitors of competence, competing in markets throughout the world. Further, the sale of the technology meant the loss of leverage that might have been used to establish a direct position in Japan. (Does this suggest some caution in the sale

[4] *Economic Survey of Japan 1978/1979.* Economic Planning Agency, Japanese Government. The Japan Times, 1980, p. 135.


TO TRULY APPRECIATE CREATIVE GENIUS, YOU NEED RECREATIVE GENIUS.
There are many sophisticated recordings today that capture a performer's true artistry.
But unfortunately, most people never get to hear it.
Because many of today's so-called high fidelity components give you something other than the genius of the artist. They give you distortion. Noise that can dramatically interfere with the music you're listening to. Noise that can make a surging musical passage seem lifeless. And a quiet moment altogether disappear.
But at Pioneer, we design all our turntables, receivers, speakers and tapedecks to make sure you get everything out of a piece of music that a musician puts into it. From the soprano voice of a piccolo. To the deepest tones of a contrabassoon.
It's no wonder Pioneer High Fidelity Components are the first choice of people who care about music.
So if you're going to pay a big price for one of today's better recordings, why hear only a small part of it?
PIONEER
We bring it back alive.

JAPANESE TECHNOLOGY TODAY

of technology to the Soviet Union and People's Republic of China?)

In the 1950's, both Renault and British Austin were licensed by Japan for their technology. Within a generation, Honda introduced its models into British Leyland to raise that company's technical level. Corning Glass licensed its technology for TV picture tube glass to Asahi Glass. Corning is now a member of the COMPACT group, vigorously protesting Japanese television exports to the United States. In this context, the battles of IBM and Texas Instruments to enter Japan directly take on added significance.

It is all too easy to dismiss this phenomenon as mere copying. It is not. First, some of the technology contracts were entered into to gain patent access on products and processes that the Japanese had independently developed but that had prior foreign patent coverage. Thus, the then Toyo Rayon, now Toray Industries, paid a fee equal to its entire equity capitalization to duPont for access to the Nylon patent. Since duPont provided no know-how, Toray developed its technology independently.

Second, as the Economic Planning Agency notes, the process of industrializing and catching up technologically necessarily involves use of foreign technology, for any nation. The technology is there for anyone. The critical issue is the ability of an economy to absorb the technology, and that is in fact not copying but a creative act in itself.

Third, as nearly every Western company can attest, the Japanese deployment of these technologies was not a passive one. In virtually every case, improvements were made rather quickly by the Japanese licensee, often with the result that the original seller found it to his advantage to incorporate in his own product or process the Japanese developments.

In any event, the period of massive input of foreign basic technology lasted through the 1960's. In the 1970's the focus has been on improvements by the Japanese, and on production technology. There has been less available and of interest for purchase as Japanese industry has caught up. The purchase of technology has shifted to luxury-goods designs and fashion, and to software, rather than continuing the earlier focus on hardware.

As Japan moves into the 1980's, there is widespread recognition that this is the era in which Japan must move to its own independent technological development. There are many questions as to whether Japan can succeed, and the issue needs close examination. However, it is well to note that if patents are a measure of competence, Japan seems to be doing well. The Industry Structure Council of MITI examined the pattern of the registration in the United States of patents by foreign nationals. The results are startling.

A 600% INCREASE IN PATENTS

Patents registered in the United States to Japanese have increased nearly six times in the decade from 1966 to 1976.

Japanese nationals are now the largest single foreign national source of U.S. patents. In addition, U.S. patents registered by foreigners have grown from 20 percent of total patents to nearly 40 percent of total patents in the same period, while the number of patents issued has not increased. Is this a measure of U.S. loss of research leadership? What is the potential for Japan in technological development?

RESEARCH EXECUTIVES' VIEW OF JAPAN'S POSITION

The men who manage the Research and Development efforts of Japan's major companies appear to have a realistic view of the present state of their technology. Discussions indicate a good deal of self-confidence in the present level of competitive development.

Matsushita Electric's Senior Managing Director Dr. S. Kisaka, for example, noting that *"We'd like to keep up the unique position of Matsushita as the top electronics manufacturer,"* goes on to explain that, *"We at Matsushita have distinguished ourselves by 'vertical integration' in which we produce from very basic electronic components and parts to the finished products. By producing high-quality components and parts for our own use, we can always maintain the highest possible quality in the finished products. The outstanding characteristic of our firm is that all research at Matsushita Electric has been conducted with the firm conviction that 'All research is for the happiness of mankind.' It's because of this philosophy that we've come up with the VTR, Video Camera, theater-sized Projection TV, high-resolution TV/VTR system and other products."*

Managing Director Mr. M. Masujima of TDK Electronics Co. states a general view:

"We are certainly strong in production technology. The principal reason is the high quality of our workers. If you look at factories in the United States, Taiwan and Korea, you find no workplaces that can assemble a work force of the high quality that we have in Japan. Our engineers also are first-class, and this is one factor in our level of technology. But beyond that is the overall contribution made by the quality of our workers, of our human resources. Given the same equipment, a work force in Taiwan twice our size, or one in America 40 percent greater cannot reach our levels of output. Company scale is a factor. But without adequate motivation, such things as inspection systems will not provide product quality. Even with integrated circuits, because our Japanese work force is the most diligent, we can achieve high production levels with high quality."

OVERTAKING THE WEST

Again, asked about the basis for the overtaking of the West by the Japanese auto industry, Senior Managing Director Mr. J. Tanaka of Nissan Motors concludes that *"The greatest factor is the diligence of the Japanese. This is true not only of the way in which they work, but in the content of the work as well. It is this factor that has been our support."*

The recurring view is that the homogeneity of the work force, the high educational level of the work force, and the system of relatively strife-free work relations has been a principal factor in these companies' technical achievements to date. Managing Director Mr. T. Inoue of Victor Company of Japan takes this view:

"The Japanese production system involving career employment is a truly unique feature, isn't it. The fact of being an island nation; with that, the homogeneity of the population—because of these factors, human relations are not simply based on rights and duties. Our companies are groups, de-

PATENTS REGISTERED IN THE UNITED STATES BY FOREIGN NATIONALS

	1966		1970		1976	
	Number	Percent	Number	Percent	Number	Percent
Japan	1,122	8.1	2,625	15.1	6,542	25.1
W. Germany	3,981	28.9	4,434	25.6	6,243	23.9
United Kingdom	2,677	19.4	2,952	17.0	3,013	11.6
France	1,436	10.4	1,732	10.0	2,417	9.3
Other	4,506	33.2	5,611	32.3	8,215	30.1
Total	13,772	100.0	17,354	100.0	26,074	100.0
Total US Patents	68,406		64,427		70,236	
Foreign Patents as a percent of the U.S. total	20.1%		26.9%		37.1%	

Source: Vision of Industrial Policy in the 1980's.

All we ask is that you take the wheel.

Experience the feeling of uncompromising luxury.

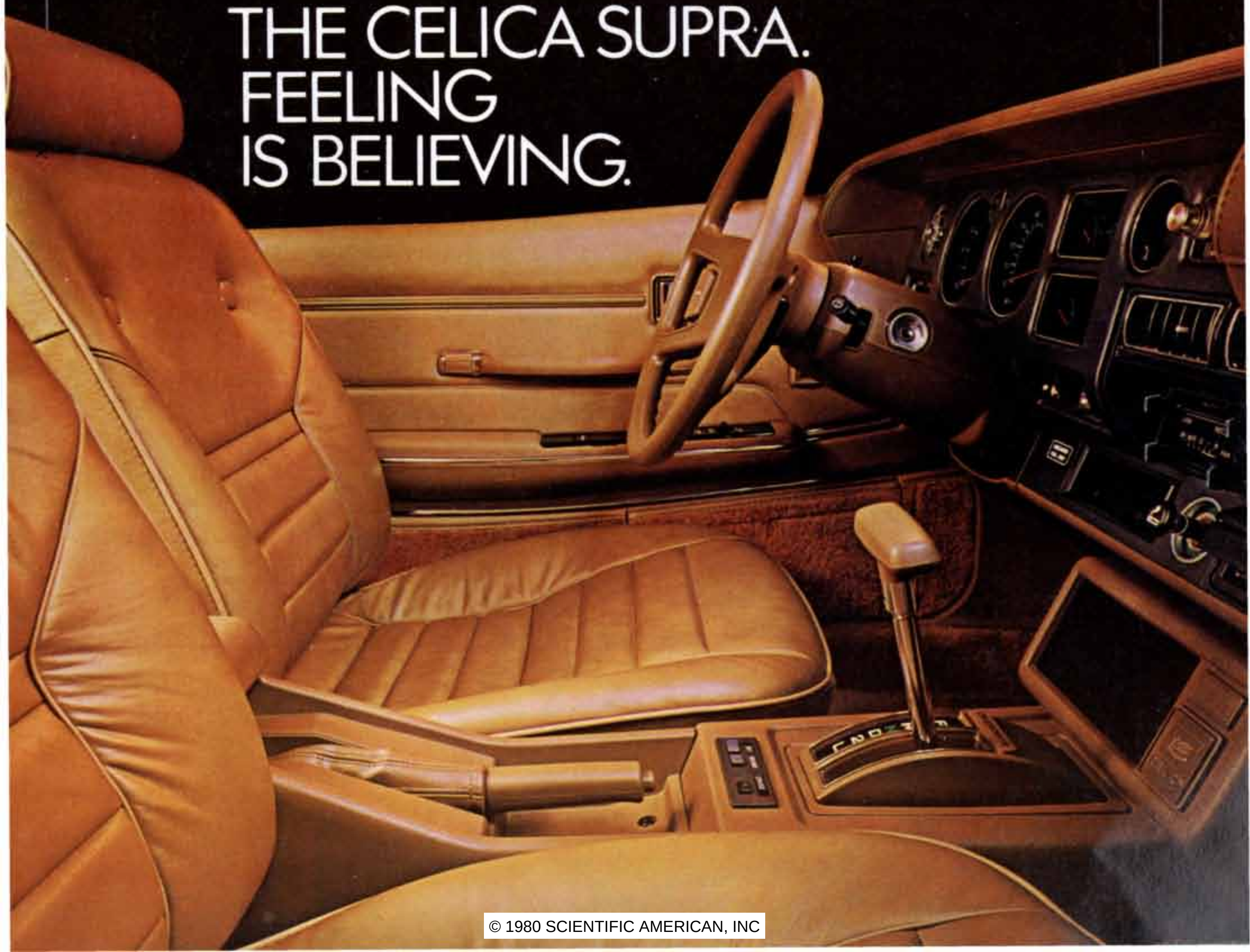
Then drive the car one mile.

We guarantee that you too will become a believer in this remarkable automotive achievement that is called Celica Supra.

Supra thrills you with a six-cylinder electronically fuel-injected engine. Soothes you with air conditioning, six-position tilt wheel and four speaker AM/FM Stereo radio —all standard. And cushions you with options like genuine leather upholstery (shown).

Performance and luxury in one automobile? Certainly. Plus an unexpected surprise: remarkable fuel economy. Thanks to fuel injection, and an optional 4-speed automatic overdrive transmission, Supra is rated at 26 EPA Estimated Highway MPG. (21) EPA Estimated MPG. Remember: Compare this estimate to the EPA "Estimated MPG" of other cars. You may get different mileage depending on how fast you drive, weather conditions and trip length. Actual highway mileage will probably be less than the EPA "Highway Estimate."

The Celica Supra. Believe it.



JAPANESE TECHNOLOGY TODAY

rived from this background. There is a context of loyalty. This has provided a positive environment for technological development."

Some question of the durability of these factors in the face of increasing size and complexity of companies is expressed. Senior Managing Director Mr. K. Yamamoto of Toyo Kogyo notes:

"In Japan, human relations and organization are in a single tune. However, as the organization becomes larger, and as it comes to include a number of separate sections, consensus becomes more difficult to achieve. Sectionalism begins to appear. Radical ideas, and ideas rich in originality, are more difficult to express."

While Mr. Yamamoto expresses confidence that Toyo Kogyo is dealing successfully with these problems, he raises an important issue for the future of Japan's work system.

THE COMPETITIVE AUTO INDUSTRY

In considering their situation and contrasting it with other countries', there is also a theme in the remarks of these executives that is critical of United States Management in particular. Mr. J. Tanaka of Nissan, looking to future competition, touches on this issue:

"Our future competitors are General Motors and the German auto industry. Looking at their plans, it appears that they are following closely the approach of the Japanese industry, and aiming at the Japanese industry. Since they are large-scale companies, their financial power is considerable, their technical capability adequate, and as they plan to roll us back, we will have some problems. However, in the American case, perhaps because of the structure of the companies, each single year's company results become very important. Thus, for example, when setting up a long-term plan for five years forward, company results are seen to drop because of capital requirements for the plan. And the long-term plans are not carried out. In my view, this is the major problem for America. For the sake of short-term profits, old equipment is used far too long. This is true of all U.S. industry. Even looking at General Motors plants, the assembly lines are very much out of date. They are using the methods Nissan used in 1965."

Similar themes are expressed by Dr. A. Ouchi, Senior Managing Director of Nippon Electric Company:

"Using integrated circuits as an example, America was the innovator, and in the 1960s manufactured the chip domestically, carrying out assembly in the low-wage countries of Southeast Asia. If I were an American engineer, I would probably have done the same. However, fortunately or not, the Japanese were some years behind and trying to catch up. Because we were behind in using low-wage labor, we instead sought to automate. With automation, we were able to achieve very high quality and at the same time, with appropriate machinery, achieve very low cost levels. For example, in our Kyushu plant, with complete automation we can produce at ten times the level of hand assembly, and leaving depreciation aside, have much lower costs with complete product quality.

"Labor unions are a factor in this. In the West, automation means firing people. In the case of Japan, the company gives priority to maintenance of work and therefore if automation improves the profits of the company, there is basic agreement to it. Again, in America, because long-range plans cannot be established, short-term profitability becomes the target. Thus, as with ICs, they are quick to take machinery to places where wages are low."

Rather clearly, Japan's R&D managers credit their success to good management, good labor, and a willingness to invest. There is little suggestion in their remarks of a special brilliance in planning, nor in creativity. There is a strong suggestion of self-confidence, however, as expressed by Senior Managing Director Dr. T. Sasaki of Sharp Corporation:

"In the process which proceeds from scientific concept to engineering and actual product production, I feel that in the stage of production technology, Japan generally leads the United States. Beginning with Germany and compared with Europe generally, Japan is well in the lead. However, our definitive lack is in ideas and concepts. Japan takes these from the United States, Britain and France who excel in this area, and engineers them into high levels of output."

This view of Japan as adapter of imported concepts is rather widely held by Japan's research directors. Dr. T. Kitsuregawa, Managing Director of Mitsubishi Electric says: *"In Japan there have been no breakthrough inventions. Rather, the Japanese are clever at refinements; instead of invention, they take pride in carrying out improvements."*

"WE NEVER LOST OUR SENSE OF CONFIDENCE"

Again, Honda's Vice President, Mr. H. Sugiura feels that *"In terms of developments from basic research such as basic science, new products, inventions and new processes, Japan perhaps cannot yet be said to be even. Whether or not Japan can be said to be even in that regard, we seem not yet to be realizing results. Because we never lost our sense of confidence, we have been able fully to catch up. Still, it is another matter to say we are ahead."*

There is some disagreement on this issue of new invention and creativity. Mr. H. Ohsawa, Division Manager in Toshiba Electric says, *"Unfortunately, the general results until now in terms of invention and innovation lag behind the United States, West Germany and the United Kingdom."*

Perhaps reflecting the differences in recent experience of the two companies, Japan Victor has a different view according to Managing Director T. Inoue. Asked whether Japanese technology can be said to have emphasized production technology and product reliability rather than invention, he states:

"Currently I do not think that is so. Japanese technology also has the characteristic of being creative. The video disk is a good example. It is creative technology. I believe it is necessary to have management that is challenged to produce new technology."

SEIKO VS SWITZERLAND

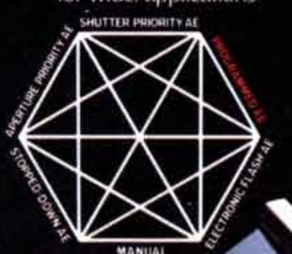
These views rather naturally reflect company experience, and one of the successes of Japan, the move of the Seiko watch group owned by K. Hattori and Co. to leading world share in the watch market, is described in terms of Japanese innovation. The success is not permanent, however, as the technology continues to move forward. Director Mr. K. Kubota of Seiko describes the history, and the issue, as follows:

"Ten years ago Seiko introduced the quartz watch into the marketplace. Before that, in terms of manufacturers, Switzerland was the kingdom of the watch industry. The post-war reconstruction of the watch industry began in the mid-1950's. Our stimulus and objective were the Swiss. We produced a number of original ideas and drew even with the Swiss even in accuracy and precision. Since the 1969 quartz watch introduction, we have left the Swiss behind. Now we are in the electronics era, and our competitors are the major electronics companies, not the watch makers. Electronics makers are the biggest threat."

THE CAMERA VIEW

Another sector, in which Japan has had entire success, now also faces a similar challenge from electronics. Mr. S. Fukuoka, Vice President of Nippon Kogaku, traces Nikon's growth in cameras:

"Before the war, German cameras had all our respect and admiration. After the war, as we stopped production of precision machines and began work on cameras, we had some confidence we could accomplish our goals. Even then though, our policy was that we wouldn't sell cameras unless they were better than those of our



Programmed 350 9.5

'seniors,' so to speak. We were trying out one experimental lens after another, never to our satisfaction. When the Korean War started, our product was introduced to the world in a manner that has since become legendary. (Editor's Note: Mr. Fukuoka is referring to the incident when the prominent photographer, David Douglas Duncan, went to the battlefields of the Korean War to take photographs and said publicly that he preferred the Nikon camera to the Leica.) The fruit of our efforts was recognized and our products began to sell. Now, leaving out toy-like cameras and special cameras such as Polaroid, there is no camera industry in the world except in Japan. Our development effort now concerns new features in new products and that is the meaning of innovation for all of the companies in the industry. But this means electronics. The trend today is to smaller cameras, and to such electronic developments as shutter speed control and autofocus."

To deal with these requirements, Mr. T. Watanabe, Managing Director, Olympus Optical Co., reports that they are spending 6 to 7 percent of annual sales on R&D against an industry average of 4 to 5 percent. In fact, he believes that the total figure is nearer 10 percent of sales if all development-related costs are included. The Japanese camera producers clearly are not content with their world-dominant position.

Still another camera maker has extended its technology to a new product area and seems to be considering a more diversified approach. In view of the apparent maturity of the camera industry, Senior Managing Director Mr. T. Tada of Minolta Camera says:

"We have the word 'camera' in our company name. Originally we developed our copy machine from our camera technology. Thus we have focused very much on camera-related developments. In addition, with some fear of moving into an unrelated field, we are considering seriously a so-called third division."

Another executive, Dr. A. Ouchi of Nippon Electric, seems also to have greater confidence in Japanese innovative abilities. He reports:

"NEC has been working on fiber optics for more than fifteen years. We have developed the world's top technology for systems applications. In fact, Japan as a whole is at a rather high level relative to the world (in this key technology). Corning Glass held the original patents and there was a patent problem. However, with the Nippon Telephone and Telegraph Corporation as the focus, we carried out a joint development program to develop Japanese optical fiber technology. According to recent data, ours is now the finest technology in the world."

A PEEK AT THE FUTURE

Whatever their view of Japanese basic research capability—and the general view remains that it is the sector that requires further and rapid development—many Japanese companies have clear and ambitious notions of where they intend to seek innovation over the next decade. Some are ambitious and broad in scope. Dr. T. Kitsuregawa says that *"In looking to the future, companies in the electrical industry have a similar view, not limited to Mitsubishi Electric — the three E's: electronics, energy, exotic materials."* This view partly reflects the broad interests of Mitsubishi Electric, ranging from atomic power generators to household appliances.

Dr. Kitsuregawa's list for the future is broad and includes such research sectors as audio warning systems for autos and elevators, single-frequency semi-conductor lasers, pattern information systems for computers, antennae for satellite signal reception, office automation equipment, solar and other alternative energy sources, VTR developments and the like.

No less ambitious in certain regards are the plans of Nippon Electric. According to Dr. A. Ouchi:

"The theme for this company for the 1980's is termed 'C and C.' This is a fusion of what until now have been distinct activities, computers and communications. For this fusion, we will need systems and approaches not heretofore available. NEC will stand against IBM in computers and against Western Electric in communcations. Because these two companies are developing in parallel, and with this fusion will be matched against each other, in view of all their power we are now attempting throughout the entire company to strengthen ourselves."

Still another major company in the electronics sector, Matsushita Electric, looks at the 80's as "The Age of Information/the Age of Energy," according to Senior Managing Director Dr. S. Kisaka. He explains:

"The coming age of information may be looked at from four viewpoints: 1) Global: With expansion of worldwide communications networks, including, among others, submarine cables, optical fibers, and communications satellites; 2) Social: CATV, multiplex TV broadcasting, office automation, and electronic transaction of business; 3) Home: Information interfaced with society through the system listed above; and 4) Personal: Pocket computer/data bank/translator, and pocket telephone. On the other hand, the coming Age of Energy will be characterized by new energy technologies, especially solar energy, and energy conservation with energy saving through more refined electronic control."

Given the general Western view of Japan, these research executives put remarkably little emphasis on government support in meeting their research and development goals. The general view seems to make a quite clear distinction between basic research and corporate research activity, and seems to indicate that government support should not be directed toward the business community. In a somewhat critical vein, Mr. H. Ohsawa of Toshiba states:

"Japanese companies, including my own, have insufficient concern with basic research. Their attitude is that the government, through the national universities and MITI's research facilities, should focus on this. In their view, manufacturers have the task of applying the results of the basic research."

This view is in fact generally accurate. Mr. M. Masujima of TDK was asked what he thinks about government aid to business in the research sector. He replied:

"In the case of our company there is absolutely none. Further, we do not need it. Instead of aid to companies, there should be more aid to universities and to national research institutions. It is there that basic research should be strengthened. From there, projects can be handed on to companies. It is only in that way that major results can be achieved."

Dr. T. Sasaki of Sharp, whose positive views of the level of Japanese production technology have been noted, seems to summarize the general view about the needs of the future:

"I think that each sector — government, the universities, and industry — should put more stress on encouraging science rather than encouraging scientific technology. Industry can amply develop technology. However, for the encouragement of science, scientific research budgets are insufficient. A cause of limited government aid is the fact that there is no military funding. To balance the lack of funding, I think it is necessary to have industry-university cooperation. For example, required facilities might be jointly established at the companies. University staff can make good use of the facilities, and at the same time, more effort will be needed in support of basic and general concepts for which equipment is not necessary."

These views are those of executives in companies in which the directions and potential for further technological development can be fairly clearly identified, in the near term at least. But not all industries have a promising future. It is useful to recall the general view in Japan that the nation grows not only by pursuing promising technologies and growth, but

JAPANESE TECHNOLOGY TODAY

by concurrently allowing the marketplace to diminish the role of industries that have reached technological maturity. Steel is just such an industry. Despite the fact that Japan is the acknowledged leader in technology and in production cost position, steel imports into Japan are growing rapidly, and new sources of steel products are clearly emerging as effective competitors.

Technological change and economic growth are the shift of resources from mature industries to growth sectors with higher technological levels and therefore higher levels of value-added. The Japanese response to pressure on the steel industry is therefore of special interest. Dr. T. Ikeshima, Executive Vice President of Sumitomo Metal Industries offers a point of view:

"Thinking only of the steel industry, the advance of the developing countries as competitors is not pleasant to contemplate. However, from a different point of view, a widening of the GNP gap is certainly not a plus in terms of the international environment. As countries raise their living standards, and increase their use of steel, the increase in demand will lead to a growth of their steel industry. At the same time, there will also be a requirement for higher value items, and the supply of these will be a plus for Japan. We should view the issue in this manner; we Japanese have been trained to see the issue in this way.

"Earlier we studied America, and we can now be considered to have surpassed America. However, that is because of America's own stagnation. I feel that when humans are more or less hungry, they should be pursuing some goal. As they gradually begin to reach their objective, they strive to pull ahead as leaders. In that view, we intend to support the technical efforts of other countries as much as possible, and in as many places as possible. The most eager are Taiwan and South Korea. However, if leaders and workers do not really share common interests, great difficulties result. Both countries depend on imported raw materials, in the same way as Japan. We hope they will expand and develop."

THE ISSUE OF INNOVATION

In technology, as in so much else related to their economy, the Japanese have caught up. By all indications, their technology is at the highest levels. The achievement is extraordinary, and historic. In 1936, the Mitsubishi Economic Research Bureau stated:

"In the early years of the Meiji era, Japanese industry was in the handicraft stage, and only typical Japanese trades such as porcelain, lacquer ware, paper manufacturing, silk reeling, weaving, flower-mat making showed more or less development. The industrial development has now reached a stage hardly inferior to that prevailing in the West, and the former imports of ordinary manufactured articles are now replaced by domestic products of good quality."[5]

From flower-mat making, to import substitution, to industrial competence of the highest sort. The first non-Western society to have traveled the entire distance from handicrafts to technical eminence. It will be some time before the world, and the Western world in particular, entirely adjusts to this change.

Japan too is having some difficulty adjusting to the change. Trade and investment policy, appropriately protectionist earlier, have changed only in the past decade. Perhaps the most difficult adjustment though is in the selection of further goals. Having caught up in so many areas, what targets are next?

"... (A) major requirement for the stable growth of the Japanese economy in the future is the acceleration of technological innovation. The chances for our success in overcoming the energy problem, for one, crucially depend on it. There are more reasons why we must accelerate the development of new technology. First, the level of Japanese technical accomplishments has generally caught up with those of the major industrial countries of the West. Also, the level of income in this country... has rapidly improved so that it is close to parity with the Western countries.

"If only in the interest of the international division of labor, therefore, Japan is faced with a pressing need for developing a more knowledge-intensive type structure of industries in a short time. We also need a technology of our own if we are to improve the living environment, given the "high-density society" we have in this country. All these considerations ... underscore the importance of technological innovation, especially in the development of technology of indigenous origin."[6]

There are substantial questions regarding Japan's ability to move in the next decade to a high level of innovation. First, Japan's success in many industrial sectors should not be allowed to hide the fact that in industries like chemicals Japanese industry has shown no special distinction technically. Again, by general view, Japanese companies lag in the development of software and systems technology for all of the progress in hardware.

To date, the technical accomplishments of Japanese industry have been focused on what might be called the middle-range of technology—watches, cameras, autos, and home entertainment products. Japan has a minuscule aerospace industry, only a modest position in atomic energy developments and a distinctly secondary position in computers and computer-related areas. Japan's successes have been greatest in mass-produced items, in process engineering and in quality control. The successes have been considerable and should not be minimized in any way, but are in a sense a measure of the task of moving to new technology in terms of innovation.

In considering whether Japanese industry can, in fact, move to a new level of technological output, the initial questions usually relate to whether the pattern of high savings and investment, of supportive labor relations and work force quality, and of stable and competent government can be maintained. These are perhaps not the critical questions. The crises of the 1970's provided a severe test of the durability of these aspects of Japanese society and industry. They have survived the test very well indeed. It is reasonable to expect their continued survival well into the future, barring some external catastrophe.

A question that must be asked concerns the ability of the Japanese to innovate in the industrial sector, in the sense of creating new products and systems. There can be no present answer to the question, since it only now arises as an appropriate issue. That the Japanese are creative is clear enough. The existence of a distinct and unique culture, with special art forms and living styles, is surely testimony to creativity. Individuals like Tange in architecture, Munakata in art, Kenzo in high fashion, and Mishima and Kawabata in literature are creative by any measure and in every sense.

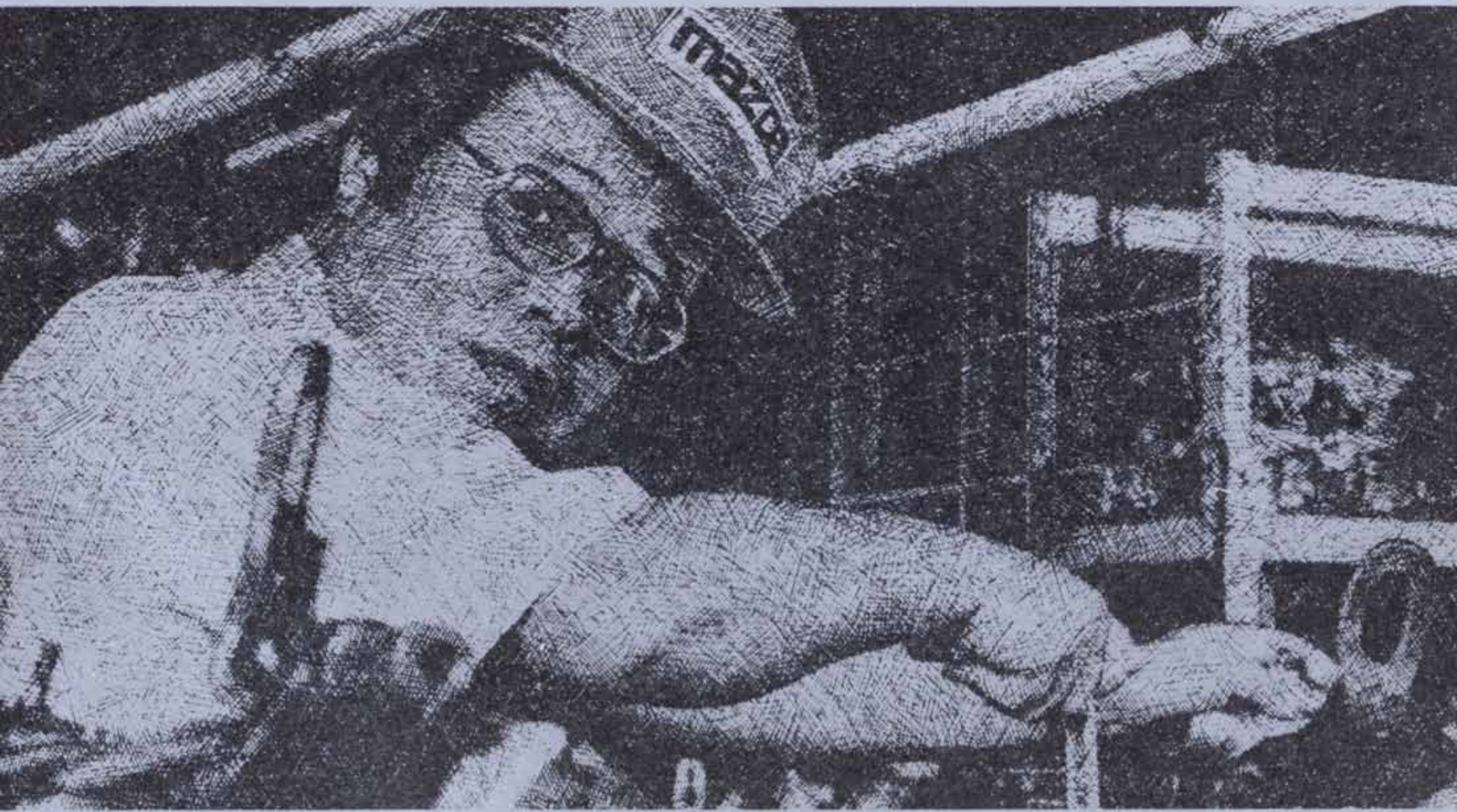
The question remains, however, as to whether in a corporate context, with emphasis on group values and group-centered behaviors, the same degree of creativity can be fostered. Japanese cliches like *"The nail that sticks out gets hammered down"* are folk-wisdom testimony to the pressures to conform to group expectations. These pressures must be oppressive to the development of innovation.

Yet it can also be argued that a supportive group context allows creative risk-taking because the costs of failure are less extreme than in the more punishing climate of work relations in the West. It can also be held that the processes of

[5]*Japanese Trade and Industry, Present and Future.* Mitsubishi Economic Research Bureau. Macmillan and Co., Ltd. London 1936, p. 203.

[6]*Economic Survey of Japan 1978/1979.* Economic Planning Agency, Japanese Government. The Japan Times, 1980, p. 135

What car company increased its productivity 42% in just two years?



It was Mazda.

In 1977, we built 27.1 Mazda cars and trucks per employee. In 1978, we built 30.6. In 1979, our productivity rose to an incredible 38.6.

This dramatic increase was made necessary because of a quantum leap in demand for our products in 110 markets worldwide. A demand that propelled us into an elite corps of 10 top world carmakers last year, when we produced about a million vehicles.

But while we've increased our production, we also made certain we didn't compromise quality to get there. All 26,809 Mazda employees are fiercely dedicated to keeping our quality level high. They know that Mazda's future is also *their* future. And they're given countless opportunities to actually *participate* and *contribute* in the company.

Throughout Mazda, small employee groups called Mazda Circles meet informally on their own time to create ways to improve the level of efficiency and quality in their group. Last year alone, we adopted tens of thousands of these suggestions. And these workers now feel more compelled than ever to make them work.

Workers also have the authority to stop the entire mile-long assembly line if they see anything that doesn't measure up. That, too, puts a tremendous responsibility in the hands of the workers.

This may all sound a little unusual. But it's all part of a new spirit that's taken over Mazda. And it's working.

Keep a close eye on Mazda. Because the more you look at what we're all about, the more you like the kind of company we stand for.

mazda

research and discovery in modern industry are in their nature more a group activity than an individual activity. The model of a Thomas Edison tinkering alone in a lab is hardly a suitable model for development work in aerospace or next-generation semi-conductors.

A related and often discussed question is whether the Japanese suffer some special handicap in terms of the development of software and of systems technology. As with the issue of industrial creativity, the answer will have to await the outcomes of the next years. Certainly Japanese competence in mathematics is generally high. Clearly, Japanese investment in this area has been relatively limited from the absence of large-scale government programs, as in the United States. And, in a sense, the need for development in this area arises only after the hardware has been put in place, and is therefore in any event the next stage of concern for the Japanese.

Another major question about Japan's ability to deal with the issue of developing unique technology is set out by the Economic White Paper. *"... What matters is creativity in the private sector. But the government also has as large a role to play not only in sharing research and development expenditure with private industry but in improving the educational and research institutions at large."*[7]

Just as there has been general agreement that the quality of primary and secondary education in Japan is very high, so is there general agreement that the financing and facilities for higher education in Japan are at a low level. Further, graduate training is largely for academic careers, and the system of compensation of the major companies offers little incentive to undertake graduate study before entering the work force. Japan's companies have made up for these deficiencies by in-house training programs for the most part. The question remains as to whether Japan will produce a sufficient supply of highly trained specialists. The current efforts to establish a research-centered university and community near Tokyo are a recognition of the need. The slowness of development of that complex is a measure of the problem.

Apart from government support to education, there is the more general question of government support to research. Substantial military and space budgets have provided a subsidy base in the United States for the development of jet engines, aircraft, electronics, special metals and other leading-edge technologies. There is little prospect of Japanese industry getting this kind of base support. No doubt tax concessions and very fast depreciation schedules will be employed, and semi-governmental units like the Nippon Telephone and Telegraph Corporation and the utilities companies will provide indirect financing. But for very large-scale projects more direct government financing is likely to be needed. Japan has not yet established a system for such support.

Thus there are substantial and unresolved questions about Japan's ability to become a major source of new technology for the world. Yet perhaps the factor that makes a positive outcome most likely is the clear recognition by the Japanese government and business community that there is an overriding need for innovation, and a wide agreement that the national interest requires that major efforts be concentrated in this area. Along with this agreement on goal, there is a generally high level of self-confidence about the ability of the Japanese economy to meet its goals.

The period since the 1973 oil crisis has been for most of Japanese industry a period of consolidation and regrouping. The success of the effort, measured by relative economic growth currently, international competitiveness and improved corporate results, has led to a degree of self-assurance not previously visible on the part of the Japanese economic community. The next goal is innovation. It would be an imprudent gambler who would bet against their reaching the goal. And, in fact, a shift to significant innovation in science and technology would be simply a natural and logical extension of Japan's economic development until now. For the world, the prospect of Japan as a major source of capital and as a major source of technology as well over future decades, should be an encouraging one.

[7] *Economic Survey of Japan 1978/1979.* Economic Planning Agency, Japanese Government. The Japan Times, p. 203.

The publisher, authors, Ted Bates & Company, coordinators of "Japanese Technology Today," would like to acknowledge the important contribution of the members of the international business community whose messages appear on these pages:

American Honda Motor Co., Inc.
Canon USA, Inc.
Fuji Photo Film USA, Inc.
Japan Air Lines Co., Ltd.
Matsushita Electric Industrial Co., Ltd.
Mazda Motors of America, Inc.
Minolta Corporation
Nikon Inc.
Nissan Motor Corporation in U.S.A.
Olympus Camera Corporation
U.S. Pioneer Electronics Corporation
Sony Corporation of America
TDK Electronics Corporation
Toshiba America, Inc.
Toyota Motor Sales U.S.A., Inc.
US JVC Corp.

In addition, we acknowledge the significant contribution to this section by the distinguished scientific and technical leaders who were interviewed:

Mr. S. Fukuoka
Nippon Kogaku

Dr. T. Ikeshima
Sumitomo Metal Industries

Mr. T. Inoue
Victor Company of Japan

Dr. S. Kisaka
Matsushita Electric

Dr. T. Kitsuregawa
Mitsubishi Electric Corp.

Mr. K. Kubota
K. Hattori and Co.

Mr. M. Masujima
TDK Electronics Corp.

Mr. H. Ohsawa
Toshiba Corp.

Dr. A. Ouchi
Nippon Electric Co.

Dr. T. Sasaki
Sharp Corporation

Mr. H. Sugiura
Honda Motor Co.

Mr. T. Tada
Minolta Camera Co.

Mr. J. Tanaka
Nissan Motor Co.

Mr. T. Watanabe
Olympus Optical Co.

Mr. K. Yamamoto
Toyo Kogyo

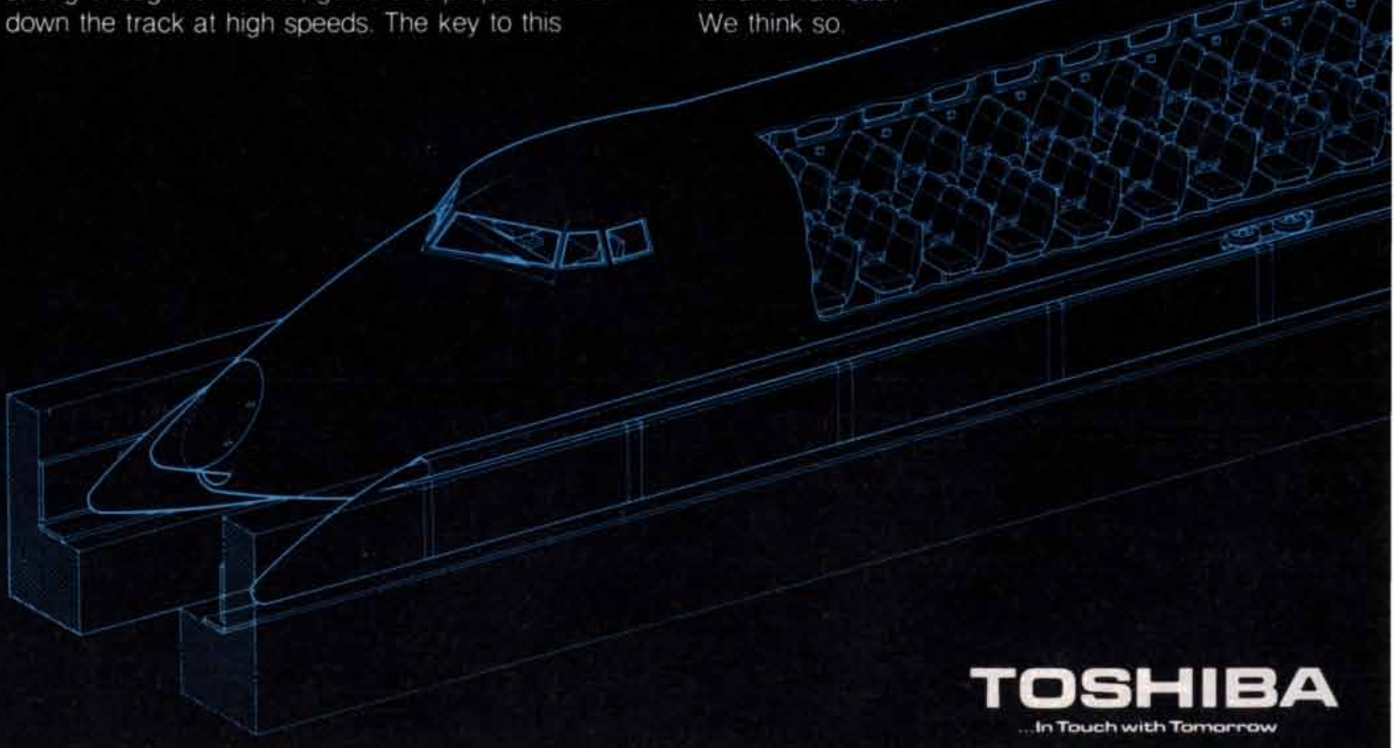
Cold Logic.

A magnet cooled to −268.8°C will provide the driving force behind tomorrow's high-speed transportation system.

No more clickety-clack. Toshiba is a member of the group working with the Japanese National Railways on the train of the future. At a test track in Miyazaki, Southern Japan, Toshiba engineers from many different disciplines are working together to develop an advanced transportation system that will float 10cm above a guideway and travel over 500km/hr. Last December a test car attained a speed of 517km/hr.

The heart of the train's new linear motor is a unique superconducting electromagnet operating at a temperature close to absolute zero. Cooled by helium refrigeration equipment on board, this electromagnet requires no electrical contact to generate a force strong enough to levitate, guide and propel the train down the track at high speeds. The key to this process is a special system developed by Toshiba to recover the cooled helium.

The benefits, of course, are enormous: high speeds, no friction or wear on the running equipment, and greatly improved passenger comfort and capacity. The technical innovations in all fields needed to bring a train like this into being are enormous. That's why Toshiba, with its experience in heavy engineering, electrical engineering, electronics, and transportation was chosen. We're working for the day when passengers will be riding at twice the speed of the fastest trains operating today. Will that be any way to run a railroad? We think so.



TOSHIBA

...In Touch with Tomorrow

The Causes of Color

They are diverse, but they all stem from the same root: It is the electrons in matter, through their varied responses to different wavelengths of light, that make the world a many-colored place

by Kurt Nassau

What makes the ruby red? Why is the emerald green? On the most superficial level these questions can be given simple answers. When white light passes through a ruby, it emerges with a disproportionate share of longer wavelengths, which the eye recognizes as red. Light passing through an emerald acquires a different distribution of wavelengths, which are perceived as green. This explanation of color is correct as far as it goes, but it is hardly satisfying. What is missing is some understanding of how matter alters the composition of the light it transmits or reflects. Ruby and emerald both derive their color from the same impurity element: Why then do they differ so dramatically in color? What gives rise to the fine gradations in spectral emphasis that constitute the colors of materials?

It turns out that the ultimate causes of color are remarkably diverse. An informal classification I shall adopt here has some 14 categories of causes, and some of the categories embrace several related phenomena. With one exception, however, the mechanisms have an element in common: the colors come about through the interaction of light waves with electrons. Such interactions have been a central preoccupation of physics in the 20th century, and so it is no surprise that explanations of color invoke a number of fundamental physical theories. Indeed, color is a visible (and even conspicuous) manifestation of some of the subtle effects that determine the structure of matter.

The Energy Ladder

The perception of color is a subjective experience, in which physiological and psychological factors have an important part; these matters will not be taken up in detail here. It seems reasonable to assume, however, that perceived color is merely the eye's measure and the brain's interpretation of the dominant wavelength or frequency or energy of a light wave. The meaning of this assumption is clear in the case of monochromatic light, which has a single, well-defined wavelength. The interpretation of light that is a mixture of many wavelengths is more complicated, but it is still the relative contributions of the various wavelengths that determine the color.

Wavelength, frequency and energy are alternative means of characterizing a light wave. Energy is directly proportional to frequency; both energy and frequency are inversely proportional to wavelength. In other words, high frequencies and high energies correspond to short wavelengths, as at the violet end of the visible spectrum. A common unit of measure for light wavelengths is the nanometer, which is equal to a billionth of a meter. The energy of light is conveniently measured in electron volts, one electron volt being the energy gained by an electron when it accelerates through a potential difference of one volt. In terms of wavelength human vision extends from about 700 nanometers, where red light grades into infrared radiation, down to about 400 nanometers, at the boundary between violet light and ultraviolet radiation. The same range in energy units runs from 1.77 electron volts to 3.1 electron volts.

An important constraint on all interactions of electromagnetic radiation with matter is the quantum-mechanical rule that says atoms can have only certain discrete states, each with a precisely defined energy; intermediate energies are forbidden. Each atom has a lowest-possible energy, called the ground state, and a range of excited states of higher energy. The allowed energy states can be likened to the rungs of a ladder, although their spacing is highly irregular. Light or other radiation can be absorbed only if it carries precisely the right amount of energy to promote an atom from one rung to a higher rung. Similarly, when an atom falls from an excited state to a lower-lying one, it must emit radiation that will carry off the difference in energy between the two levels. The energy appears as a photon, or quantum of light, whose frequency and wavelength are determined by the energy difference.

The states that are of the greatest interest in the analysis of color represent various possible energy levels of electrons. In atoms, ions and molecules each electron must occupy an orbital, which describes a particular geometric distribution of the electron's charge around the atomic nucleus. The orbitals in turn are organized in shells. A further constraint on the possible states of the atom is that each rung on the energy ladder can be occupied by only a limited number of electrons. In general, when proceeding from the smallest atoms to the largest ones, electrons are added in sequence from the bottom rung up. Two electrons fill the first shell; each of the next two shells holds eight electrons. The electrons in any filled or closed shell form pairs, and they have a notably stable configuration.

A comparatively large quantity of energy is needed to promote one of the paired electrons from a closed shell to the next vacant position on the ladder. The energy required for such a transition can usually be supplied only by radiation in the ultraviolet or even in the X-ray region of the spectrum; as a result closed shells have no direct influence on the colors of materials. Instead color usually results from transitions of unpaired electrons, which are most often the outermost ones. They are the valence electrons, the ones that participate in chemical bonds.

Atomic Transitions

Consider a vapor of the element sodium in which the density is low enough for each atom to act independently of its neighbors. The sodium atom has 11 electrons, but 10 of them lie in closed shells, and it is only the single valence electron that takes a direct part in the interactions of the atom with light. When the sodium atom is in the ground state the outermost electron occupies an orbital designated $3S_{1/2}$. The next-highest energy levels (the next rungs on the ladder) are labeled $3P_{1/2}$ and $3P_{3/2}$, and they lie at energies 2.103 and 2.105 electron volts above the ground state. These are the smallest quantities of energy a

CHAMELEONLIKE GEMSTONE seems to adapt its color to the spectrum of the light with which it is illuminated. The stone is alexandrite, which appears red in the red- and yellow-rich light of a candle flame or an incandescent lamp but turns blue-green in sunlight or in the light from a fluorescent lamp. Alexandrite is a beryllium aluminate, $BeAl_2O_4$, but the color is generated by chromium ions present in the crystal as impurities. Transitions between various energy levels of unpaired electrons in the chromium ions are responsible for the colors of several gemstones. The exact color produced depends on the chemical environment of the ions. For example, chromium ions give rise to the red of ruby and to the green of emerald under slightly different conditions. The environment of the chromium ions in alexandrite is intermediate between that of ruby and that of emerald, and so the color is changeable. Natural alexandrites are exceedingly rare. This one is a synthetic crystal grown by Creative Crystals, Inc., of Concord, Calif. The photographs, which were made by Fritz Goro, do not show the exact colors of the crystal because the response of photographic film is somewhat different from that of the human eye.

Physician, did you miss any of these significant developments in medical science?

- The isolation of coxsackievirus B4 from the damaged pancreatic cells of a 10-year-old boy who died of diabetic ketoacidosis adds to the growing evidence that implicates viruses in some forms of insulin-dependent diabetes mellitus (IDDM).

- *Campylobacter fetus* subsp. *jejuni* causes at least as much acute gastroenteritis as does *Shigella* or *Salmonella;* the culturing and antimicrobial therapy for *C. fetus* subsp. *jejuni* differs from that for the other two bacteria.

- Miconazole shows great promise in patients with disseminated and meningeal coccidioidomycosis who do not respond to or cannot tolerate amphotericin B.

- Pittsburgh pneumonia agent (PPA) and atypical *Legionella*-like organisms (ALLO) are particularly dangerous in immunosuppressed patients.

- Lithium batteries for pacemakers have made mercury-zinc cells virtually obsolete; they can power a unit for 8 to 20 years. Microcircuitry has revolutionized the therapeutic and diagnostic potential of these devices.

IF THESE ITEMS are familiar you must be a prodigiously energetic or prodigiously lucky reader. With 2,000 or more journals published each year, information that significantly affects patient management all too easily slips by. Textbooks are out-of-date before they are published.

SCIENTIFIC AMERICAN *Medicine* is the busy clinician's answer to this problem.

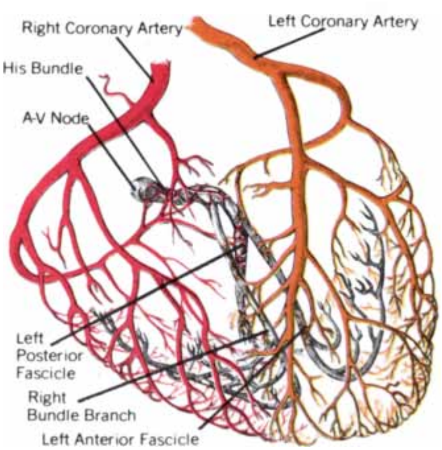
Because its authors update SCIENTIFIC AMERICAN *Medicine* every month, it is always current. Because the new information appears in a single source, it is there when you need it.

This 2,000-page, innovative union of publishing and electronic technology is the work of leading scholar-practitioners from Harvard and Stanford. The editors are Edward Rubenstein, M.D., F.A.C.P., and Daniel Federman, M.D., F.A.C.P.

Each month as authors update their contributions, revisions are entered on the magnetic tape on which the text and index are stored. The tape drives high-speed phototypesetting equipment so that subscribers receive about eight new chapters and a new index every four weeks; a bulletin highlights new developments.

New material replaces old material in the living text, so that the information is there – up-to-date, at your fingertips.

A CME program of eight patient management problems offered over a 12-month period is available at no extra cost. As an organization accredited for continuing medical education, the Stanford University School of Medicine designates this continuing medical education activity as meeting the criteria for 32 credit hours in Category 1 for Educational Materials for the Physician's Recognition Award of the American Medical Association, provided it has been completed according to instructions. This program is approved by the American Academy of Family Physicians for 32 Elective Hours of CME credit.



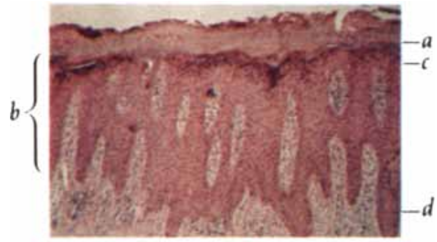
Branches of the right and left coronary arteries supply blood to the AV node and intraventricular conduction system.

SCIENTIFIC AMERICAN *Medicine* is lucidly illustrated with drawings and photographs. Some examples are seen here and on the facing page.

Trial Offer

We invite you to try SCIENTIFIC AMERICAN *Medicine* — for two months at no cost. Send us the coupon and you will receive the two-volume text and two monthly updates. You may also take a CME test for credit. At the end of 60 days, if you decide to continue the subscription, we will bill you for $185 for the full 12 months (renewal is $160); otherwise just return the volumes to us.

Please mail the coupon today and let us take the hassle out of keeping up.



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., Harvard Medical School and Massachusetts General Hospital

E. William Hancock, M.D., F.A.C.P., Stanford University School of Medicine

Roman W. DeSanctis, M.D., Harvard Medical School and Massachusetts General Hospital

Adolph M. Hutter, Jr., M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Eve Elizabeth Slater, M.D., Harvard Medical School and Massachusetts General Hospital

2. Dermatology

Eugene M. Farber, M.D., Stanford University School of Medicine

Elizabeth A. Abel, M.D., Stanford University School of Medicine

3. Endocrinology

Daniel D. Federman, M.D., F.A.C.P., Harvard Medical School

4. Gastroenterology

Gary M. Gray, M.D., Stanford University School of Medicine

Peter B. Gregory, M.D., Stanford University School of Medicine

John Austin Collins, M.D., Stanford University School of Medicine

5. Hematology

Stanley L. Schrier, M.D., Stanford University School of Medicine

6. Immunology

John David, M.D., Harvard Medical School and Robert B. Brigham Hospital

7. Infectious Disease

Thomas C. Merigan, M.D., Stanford University School of Medicine

Morton N. Swartz, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Cyrus C. Hopkins, M.D., Harvard Medical School and Massachusetts General Hospital

Adolf W. Karchmer, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Robert H. Rubin, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Harvey B. Simon, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Peter F. Weller, M.D., Harvard Medical School

8. Intensive and Emergency Care

Edward Rubenstein, M.D., F.A.C.P., Stanford University School of Medicine

9. Metabolism

George F. Cahill, Jr., M.D., Harvard Medical School, Howard Hughes Medical Institute, and Peter Bent Brigham Hospital

10. Nephrology

Roy H. Maffly, M.D., Stanford University School of Medicine and Palo Alto Veterans Administration Medical Center

11. Neurology

Robert W. P. Cutler, M.D., Stanford University School of Medicine

12. Oncology

Saul A. Rosenberg, M.D., F.A.C.P., Stanford University School of Medicine

13. Psychiatry

Ned H. Cassem, M.D., Harvard Medical School and Massachusetts General Hospital

14. Respiratory Medicine

Eugene D. Robin, M.D., F.A.C.P., Stanford University School of Medicine

15. Rheumatology

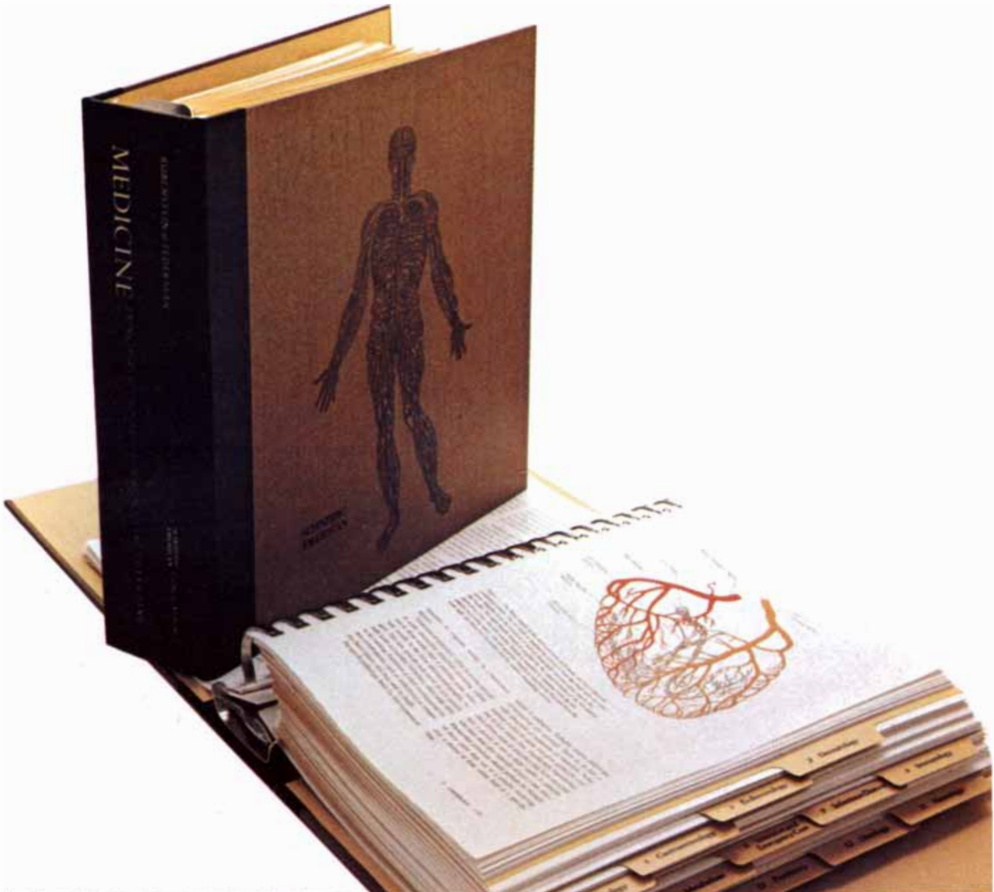
Stephen M. Krane, M.D., Harvard Medical School and Massachusetts General Hospital

Dwight R. Robinson, M.D., Harvard Medical School and Massachusetts General Hospital

Andrei Calin, M.D., M.A., M.R.C.P., Stanford University School of Medicine and Palo Alto Veterans Administration Medical Center

Order by Phone

You can order SCIENTIFIC AMERICAN *Medicine* by telephone. **Please call this toll-free number: 1-800-227-6222 (in California call 800-632-2122);** you will be billed after your subscription begins. Toll-free calls are acceptable only for orders placed in the continental United States.



SCIENTIFIC AMERICAN MEDICINE

415 Madison Avenue, New York, N.Y. 10017

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. I understand that the price of $185 for the first year of service is tax deductible, as is the renewal price of $160. If I am not entirely satisfied, I may cancel at any time during the first 60 days, returning all materials for a *complete refund*.

☐ 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 $185*
☐ Please bill me

* Please add sales tax for California, Illinois, Massachusetts, Michigan, Ohio and New York

Name ______________________

MD Specialty ______________________

Address ______________________

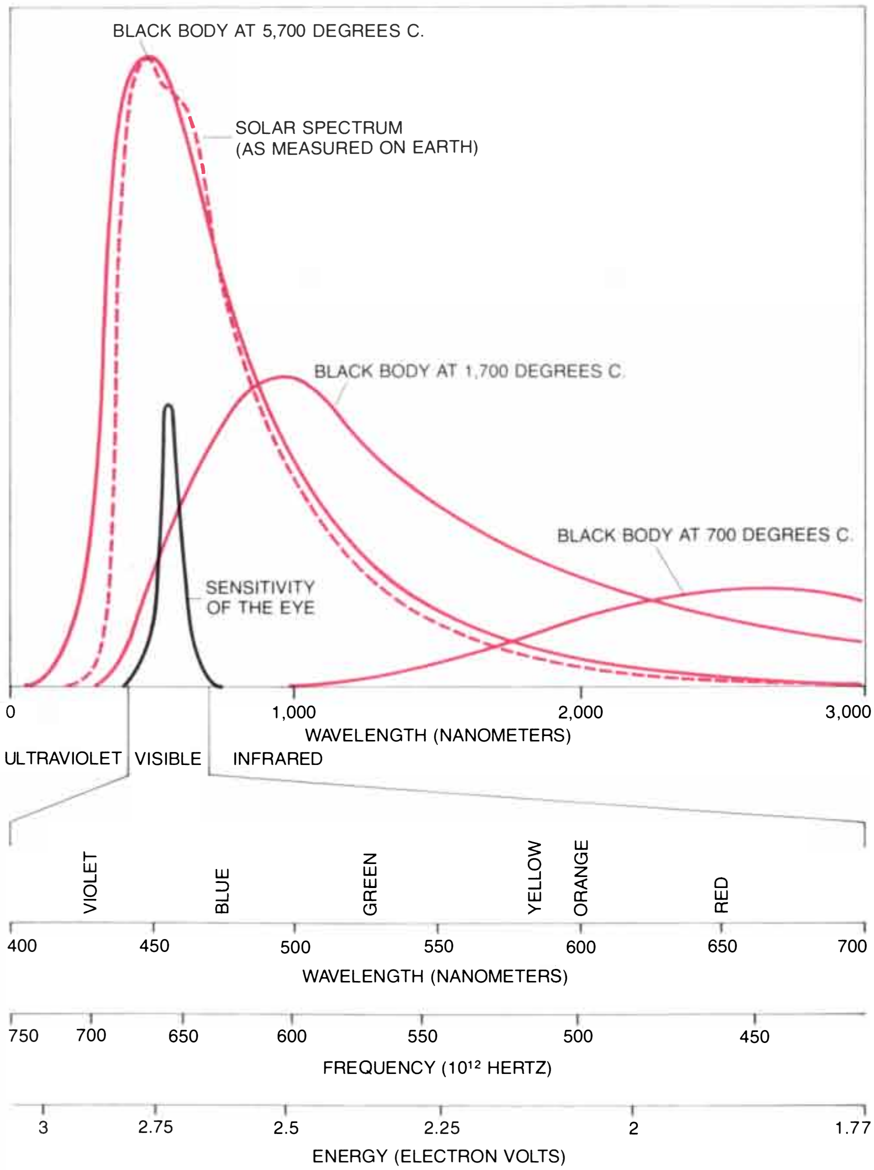
City ____________ State ____________ Zip ________

Signature ______________________

Subscribers outside of the U.S. and possessions will be charged extra for shipping and handling; updates will be sent by surface routes unless airmail delivery is requested. Please allow 6-8 weeks for delivery. 1R sodium atom in the ground state can absorb. They correspond to wavelengths of 589.6 and 589.1 nanometers, in the yellow part of the spectrum.

Above the 3*P* orbitals are a multitude of other excited states, where the electron has a greater average distance from the nucleus and a higher average energy. The number of such states is infinite, but the interval between levels becomes smaller as the energy increases, so that the series converges on a finite limit. For sodium the limit comes at 5.12 electron volts, where the outermost electron is no longer merely excited but is torn loose from the atom entirely; in other words the atom is ionized.

Suppose a sodium atom is ionized, perhaps by a quantum of ultraviolet radiation, and the free electron and the ion then recombine. Initially the electron may occupy one of the higher orbitals, but it quickly falls to a lower energy level. If the descent were made in a single step, from the ionization limit to the ground state, the atom would emit a single ultraviolet photon with an energy of 5.12 electron volts. A much likelier route would pass through several intermediate states, accompanied by the emission of a lower-energy quantum at each stage. Not all such cascades are possible; "selection rules" determine which ones are allowed. Most of the allowed pathways proceed through one of the 3*P* orbitals and thence to the ground



SPECTRUM OF SUNLIGHT closely matches the sensitivity of the eye; as a result light appears white if its spectrum resembles the solar one, and other colors can be described by how they depart from the solar spectrum. The sun's radiation is approximately that of a black body with a temperature of 5,700 degrees Celsius. The shape of a black-body spectrum is determined entirely by the temperature, becoming steeper and shifting to shorter wavelengths as the temperature increases. Thus as an object is heated its color changes from black (no emission) to red to yellow to white and finally to pale blue. The spectrum can be measured in units of wavelength, frequency or energy, which are merely alternative means of describing a light wave.

state. As a result quanta of yellow light with energies of 2.103 and 2.105 electron volts are among those emitted. Indeed, these two lines are by far the brightest in the spectrum of atomic sodium and a vapor of excited or ionized sodium glows bright yellow.

The characteristic yellow radiance of atomic sodium can be observed when a salt of sodium is heated in a flame hot enough to vaporize some of the atoms. In analytic chemistry this property serves as the basis of the flame test for the presence of sodium. The doublet of yellow lines is also prominent in the spectrum of a sodium-vapor lamp, where the sodium atoms are ionized by a high-voltage discharge.

Other atoms also yield distinctive emission lines when they are excited or ionized and then allowed to return to the ground state; in each element, however, the spacing of the energy levels is different, and so the color of the emitted light also differs. In neon the strongest lines are in the red part of the spectrum, which accounts for the red glow of neon lights and signs. The mercury atom has prominent lines in the green and the violet regions of the visible spectrum, and consequently a mercury-vapor lamp gives off blue-tinged light that is deficient in red and yellow. Lasers whose working medium is a monatomic gas exploit emission lines of the same kind. Lightning and electric arcs also derive their color from electronic excitations of the atoms in gases.

Black-body Radiation

Sharply defined emission and absorption lines are typical of gases. The spectrum of light emitted by a solid or a liquid is usually quite different, in that it extends over a continuous range of wavelengths.

A universal form of radiation from condensed matter is black-body radiation, which has a continuous spectrum with a distinctive shape. Here "black body" refers simply to an idealized material that absorbs all wavelengths without favor and is also a perfect emitter of all wavelengths. (Real materials all have lower emissivity, but many approach the black-body spectrum at high temperature.) Such radiation has an important place in the history of physics, since it was through an analysis of the black-body spectrum that Max Planck deduced the quantum principle in 1900. He found he could explain the shape of the spectrum only by assuming the quantization of energy.

In ideal black-body radiation the spectrum is independent of the chemical composition of the emitter and is determined by a single parameter: temperature. At absolute zero all the atoms occupy the lowest energy level available, and no radiation is emitted. As the temperature rises some atoms are promoted

THE MARK™ T/A®

Introducing the high performance Radial T/A that adds elegance and superb handling to magnificent machines.

From BFGoodrich,® The Mark™ T/A® radial. The newest member of America's first family in high performance radials. The ones more people buy than any other.

With The Mark T/A, you possess more than an elegant radial. Its appearance complements any quality machine. You ride with the assurance of superb handling, tough durability, straight tracking, and a tread designed to offer excellent wet traction.

The DuroGard® folded belt system wraps The Mark T/A radial in a network of strength, while the dual compound tread resists heat build-up and provides a commanding grip on the road.

Ride T/A radials. Ride number one.

USA ♥ T/A

BFGoodrich

#1 IN HIGH P.ERFORMANCE RADIALS.

to excited states, but the process is a random or statistical one, and the atoms are distributed over a broad range of energies. At any finite temperature the number of occupied states increases gradually with energy up to some maximum value; then it declines again. Thus the shape of the spectrum is somewhat like the profile of an ocean wave about to break. The steepness of the wave and the position of the crest depend on the temperature of the body.

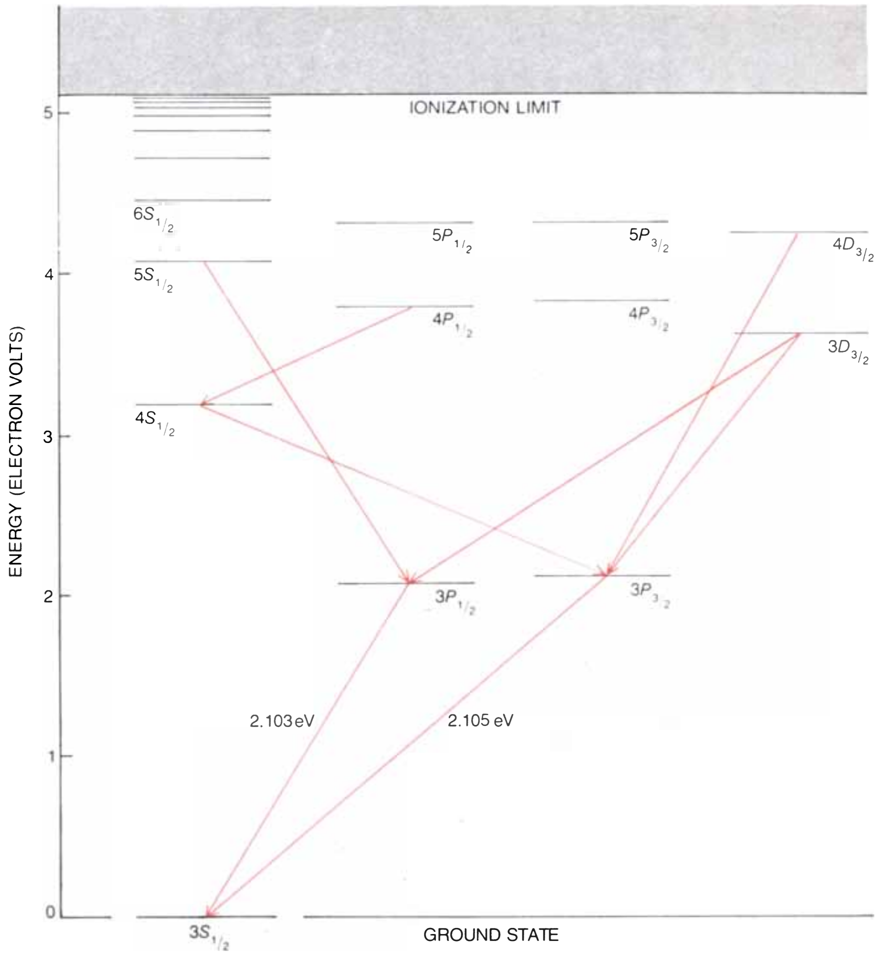
At room temperature the thermal excitations are confined to small energies, and radiation is emitted only in the infrared. When the temperature reaches about 700 degrees Celsius the maximum emissions are still in the infrared, but a little visible light begins to appear; it is perceptible as a dull red glow. As the temperature rises further the peak of the emission curve shifts to higher energies and shorter wavelengths, so that the object glows brighter and its color changes. The sequence of colors runs from red to orange to yellow to white to pale blue, in accord with the colloquial descriptions "red hot," "white hot" and so on.

In a log fire or a candle flame incandescent particles of carbon give off radiation with an effective black-body temperature of at most 1,500 degrees C., where the light ranges from red to yellow. The tungsten filament of an incandescent light bulb has a temperature of about 2,200 degrees and yields a warm yellow-white. A flash bulb, which can reach a temperature of 4,000 degrees, yields a somewhat more accurate version of white.

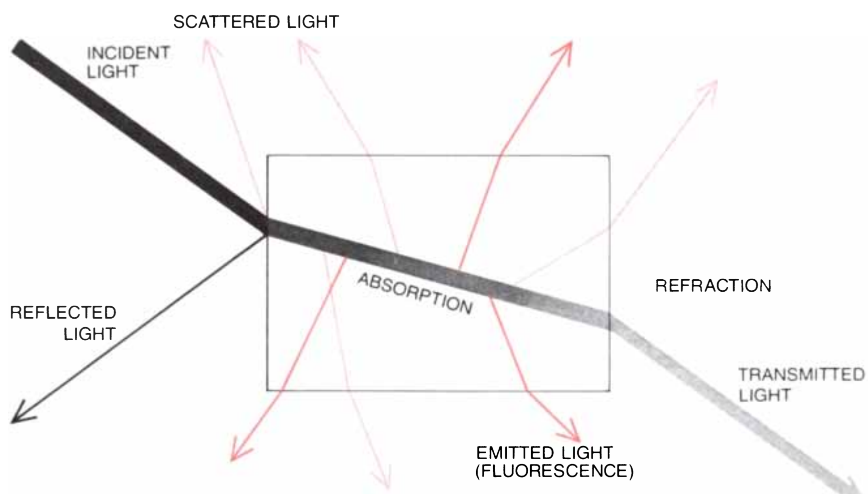
The solar spectrum has the approximate form of a black-body curve; its shape is determined by the temperature at the surface of the sun, about 5,700 degrees C. The spectrum has a broad peak centered near 2.2 electron volts, or 560 nanometers, a yellow-green wavelength. The eye is most sensitive to just this wavelength. Indeed, the concept of white seems to be conditioned largely by the spectrum of daylight, which is dominated by solar radiation. Roughly speaking, light is perceived as being white if its spectrum resembles that of sunlight; other colors can be defined according to how they depart from the solar spectrum.

Crystal-field Colors

When atoms combine to form a molecule or condense to form a liquid or a solid, new modes of excitation are introduced. Among them are mechanical vibrations and rotations that are not possible in an isolated atom. For example, the atoms of a diatomic molecule can oscillate as if they were connected by a spring, and they can rotate about their common center of mass. Such motions can occasionally influence the color of a material. In water, for example, a complex bending of the molecules absorbs a

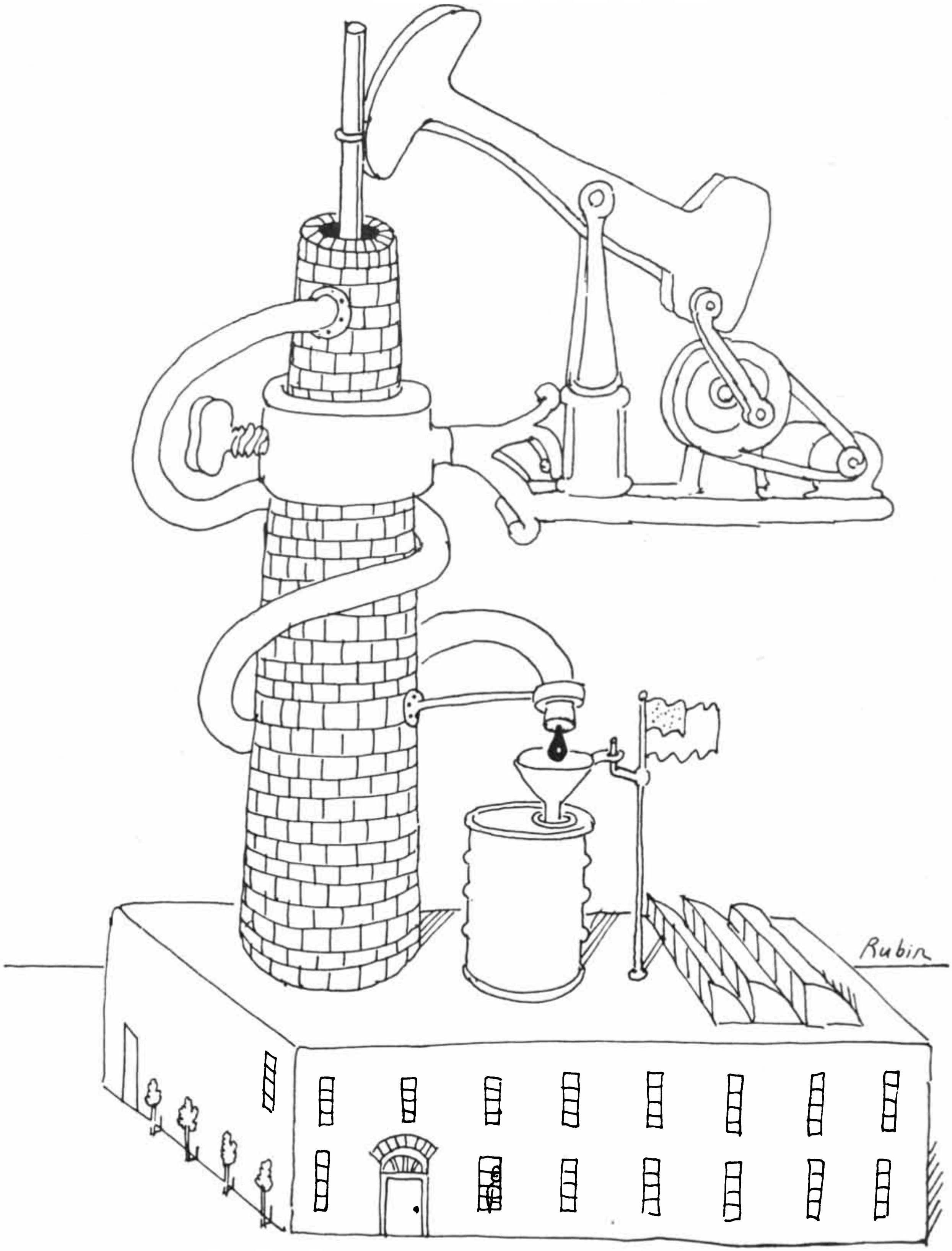


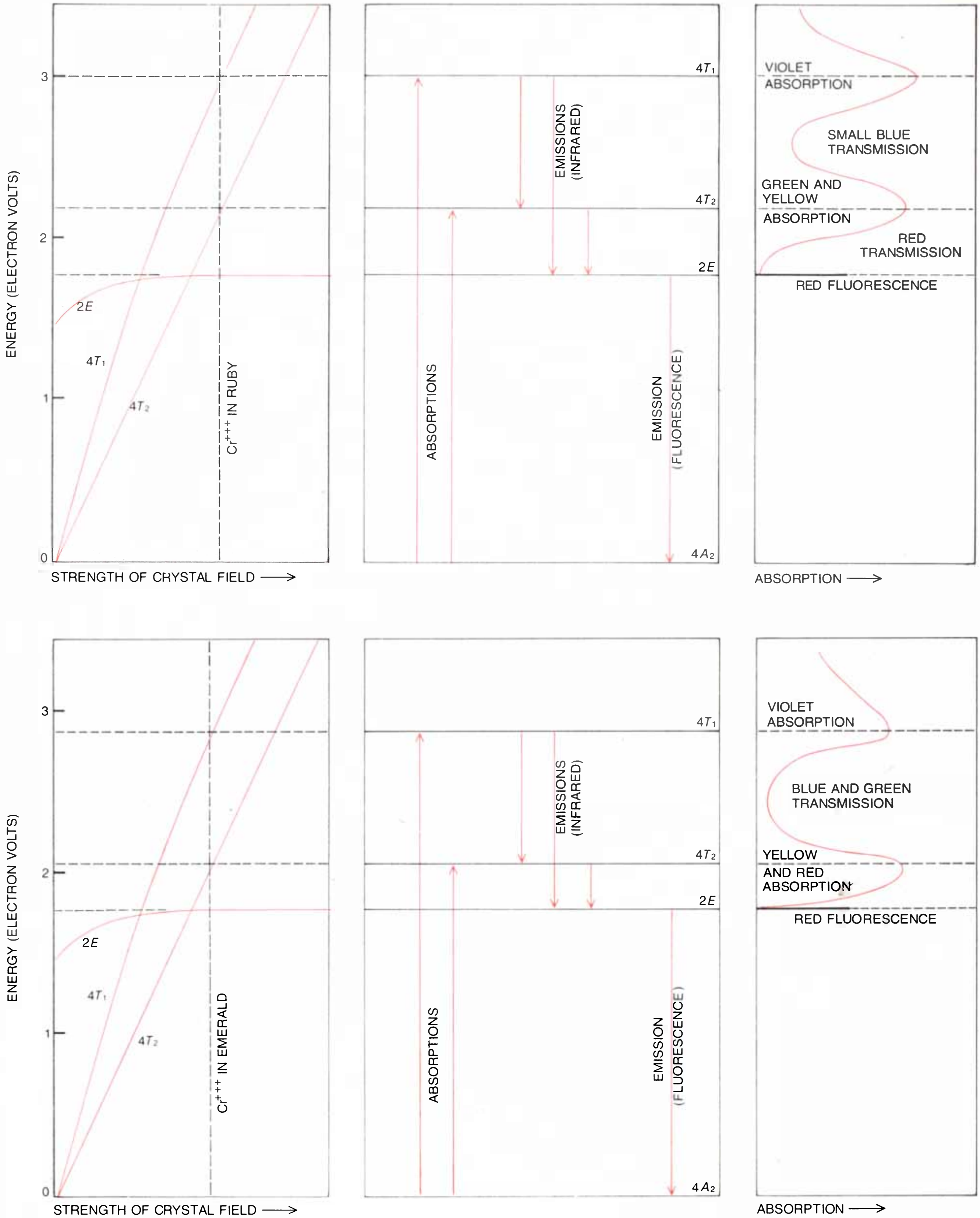
LADDER DIAGRAM for the sodium atom defines a spectrum of discrete wavelengths, which are the only ones the atom can emit or absorb. In order to climb to a higher rung the atom must absorb a quantum of radiation whose energy corresponds exactly to the difference in energy between the initial and the final states. On falling to a lower rung the atom emits a quantum with the same energy. Most downward transitions pass through the levels designated $3P_{1/2}$ and $3P_{3/2}$ to the lowest level, or ground state, labeled $3S_{1/2}$. In these transitions quanta are emitted with energies of 2.103 and 2.105 electron volts, in the yellow part of the spectrum, and so a vapor of excited sodium atoms glows bright yellow. In the ladder diagram only the vertical dimension has meaning, but the various series of levels are separated horizontally for clarity.



INTERACTIONS OF LIGHT with condensed matter include reflection, refraction, scattering and absorption; some absorbed light can also be reemitted (usually at a longer wavelength) as fluorescence. The effects of each of these processes can vary with wavelength and so can give rise to color. For example, the preferential absorption of short wavelengths and reflection or transmission of long ones makes an object appear yellow, orange or red. In general condensed matter absorbs broad and essentially continuous bands of wavelengths rather than discrete lines.

Imagine a smokestack that turns heat into oil.





ELECTRIC FIELD IN A CRYSTAL can influence color by altering the state of atoms or ions within the crystal structure. In ruby and emerald color results from the absorption of selected wavelengths by unpaired electrons in chromium ions. The transitions that cause the absorption are the same in both cases: ions are promoted from the ground state, $4A_2$, to the excited levels $4T_2$ and $4T_1$. The energies at which these states lie are set by the magnitude of the crystal field. In ruby, the absorption bands block violet light and green and yellow light; red light is transmitted and so is a little blue, which gives the ruby its deep red color with a slight purple cast. In emerald the crystal field is weaker, which depresses both absorption bands. As a result red transmission is eliminated and green and blue are enhanced. In both materials an excited ion returns to the ground state through an intermediate level designated $2E$, whose energy is little affected by the crystal field. Transitions from the $2E$ level to the ground state give rise to red fluorescence that is almost identical in ruby and emerald.

little energy at the red end of the spectrum and gives pure water and ice a pale blue cast. For the most part, however, the energy of vibrational and rotational excitations is small, and it is dissipated as infrared radiation, or heat.

Another consequence of the binding together of atoms is a change in the state of the valence electrons. In an isolated atom the valence electrons are unpaired and are the primary cause of color. In a molecule and in many solids, on the other hand, the valence electrons of one atom form pairs with the valence electrons of adjacent atoms; it is these pairs that constitute the chemical bonds that hold the atoms together. As a result of this pair formation the absorption bands of the valence electrons are displaced to ultraviolet wavelengths, and they are no longer available for the production of color. Only electrons in exceptional states remain to give rise to coloration. It is evident, however, that such exceptional states cannot be too rare; if they were, most molecules and solids would be transparent to visible light.

One set of unusual electronic states appears in the transition-metal elements, such as iron, chromium and copper and in the rare-earth elements. The atoms of metals in the transition series have inner shells that remain only partly filled. These unfilled inner shells hold unpaired electrons, which have excited states that often fall in the visible spectrum. They are responsible for a wide range of intense colors. For example, both ruby and emerald derive their color from trace amounts of chromium.

The basic material of ruby is corundum, an oxide of aluminum with the formula Al_2O_3. Pure corundum is colorless, but in ruby a brilliant color results from the substitution of chromium ions (Cr^{+++}) for a few percent of the aluminum ions. Each chromium ion has three unpaired electrons, whose lowest-possible energy is a ground state designated $4A_2$; there is also a complicated spectrum of excited states. All the excited states are broadened to form bands, and they are also modified in another way by the presence of the crystal matrix. Although the identity of the states is defined by the electronic configuration of the chromium ion, the absolute position of each level in the energy spectrum is determined by the electric field in which the ion is immersed. The symmetry and strength of the field are determined in turn by the nature of the ions surrounding the chromium and by their arrangement.

In ruby each chromium ion lies at the center of a distorted octahedron formed by six oxygen ions. The interatomic bonds in ruby are about .19 nanometer long, and they have about 63 percent ionic character, which means that the electron pairs that make up the bonds spend more of the time near the oxygen ions than they do near the aluminum or

chromium ions. This distribution of the electronic charge gives rise to a comparatively strong electric field, which is called the crystal field or the ligand field. When a chromium ion is immersed in this field, three excited states of its unpaired electrons have energies in the visible range.

The three excited states are designated $2E$, $4T_2$ and $4T_1$. Selection rules forbid a direct transition from the ground state to the $2E$ level, but both of the $4T$ levels can be entered from the ground state. The energies associated with these transitions correspond to wavelengths in the violet and in the yellow-green regions of the spectrum. Because the levels are not sharp lines but broad bands a range of wavelengths can be absorbed. Hence when white light passes through a ruby it emerges depleted of its violet and yellow-green components. Essentially all red is transmitted, along with some blue, giving the ruby its deep red color with a slight purple cast.

Because of the selection rules electrons can return from the excited $4T$ levels to the $4A_2$ ground state only through the intermediate $2E$ level. The initial transitions from $4T$ to $2E$ release small amounts of energy corresponding to infrared wavelengths, but the drop from $2E$ to the ground state gives rise to strong emission of red light. It should be noted that this red light, unlike the transmitted bands, is not present in the beam originally incident on the crystal; it is generated within by the process of fluorescence. Indeed, the red fluorescence of ruby can be observed most clearly when the crystal is illuminated with green or violet light or with ultraviolet radiation. On the other hand, the fluorescence can be quenched by iron impurities, which are often present in natural rubies. The light of a ruby laser derives from red fluorescence in synthetic rubies, which are free of iron.

Ruby and Emerald

The subtlety of crystal-field colors can be made apparent through a comparison of ruby and emerald. The color-generating impurity in emerald is again the Cr^{+++} ion, and it again replaces aluminum in small amounts. The similarity of the two substances extends further: in emerald too the chromium ions are surrounded by six oxygen ions in an octahedral configuration, and the bond length is again about .19 nanometer. In emerald, however, the fundamental crystal lattice is that of a beryllium aluminum silicate, $Be_3Al_2Si_6O_{18}$. And the most significant difference is in the nature of the chemical bonds, which are less ionic by a few percent, so that the magnitude of the electric field surrounding a chromium ion is somewhat reduced. As a result the two $4T$ levels lie at slightly lower energies; the position of the $2E$ band is essentially unaltered. The major effect of these changes is to shift the absorption band that in ruby blocks green and yellow light downward into the yellow and red part of the spectrum. The emerald therefore absorbs most of the red light, but the transmission of blue and green is greatly enhanced.

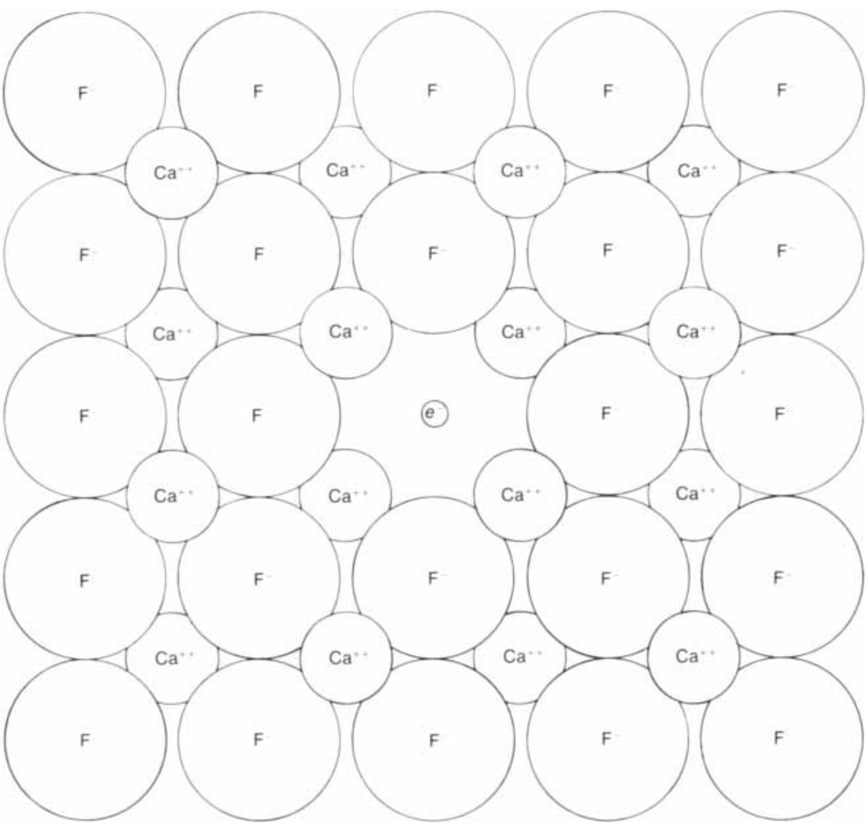
Curiously fluorescence in emerald is almost identical with that in ruby. The reason is that the energy of the $2E$ level is scarcely altered by the reduction in the crystal field. The similarity of the fluorescent emissions shows it is merely a coincidence that ruby has a red color and also red fluorescence.

Intermediate in spectral composition between ruby and emerald is the rare (and therefore costly) gemstone alexandrite. Again the color arises from chromium ions that replace aluminum, but in this case the underlying crystal structure is that of a beryllium aluminate, $BeAl_2O_4$. The crystal field that sets the energy scale of the chromium ions is stronger than that in emerald but weaker than that in ruby, with the result that the red and green transmission bands are quite evenly balanced. The near equality of the two bands has an extraordinary consequence: in blue-rich sunlight the gemstone appears blue-green, but in the redder light of a candle flame or an incandescent lamp it appears red.

Crystal-field colors can arise whenever there are ions bearing unpaired electrons in a solid. Aquamarine, jade and citrine quartz get their colors from a mechanism similar to that in ruby and emerald, but the transition-metal impurity is iron instead of chromium. Many compounds and crystals in which the transition metals appear as major constituents rather than impurities are also strongly colored. Among natural minerals in this category are the blue or green azurite, turquoise and malachite, in which the color is produced by copper, and the red garnets, which owe their color to iron. Most pigments in paints are transition-metal compounds.

Color Centers

The physical mechanism responsible for crystal-field colors is not confined to electrons in transition-metal ions; indeed, the electrons need not be an intrinsic component of any atom. An excess electron unattached to any single atom will suffice if the electron can be trapped at some structural defect, such as a miss-



COLOR CENTER can form in a crystal when an electron takes the place of a dislodged ion. In fluorite, or calcium fluoride, the electron fills the vacancy created when a fluorine ion is removed. The electron has a spectrum of excited states that extends into the range of visible wavelengths. The color centers in fluorite, which are called *F* centers, give it a purple hue. The original crystal structure can be restored by heating, whereupon the color fades. There are also electron and "hole" color centers, where a single electron rather than an entire ion is displaced.

The GE Computer Radio.
At 6:00 A.M. it's smarter than you are.

6:00 — TIME

The Great Awakening from General Electric. For starters, it's smart enough to let you set the time directly...no flipping around the clock.

6:15 — WAKE-UP 1

You can program it to change stations for you. So it will rock you to sleep with Strauss, switch to your news station, and wake you at 6:15.

7:53 — WAKE-UP 2

Then it comes back on to wake up your better half to Beethoven at 7:53. All with push-button ease.

OFF — ALARM OFF

When you forget to set the alarm...The Great Awakening remembers to remind you.

1410 — RADIO AM

You can scan all the AM or FM stations by pressing a button or, to tune in one station, just punch in the frequency of your choice on the keyboard.

102.7 — RADIO FM

You can also program up to six stations into the memory. And recall any one with the touch of a finger.

15 — SNOOZ-ALARM

For a little extra sleep, press the Snooz-Alarm.® It lets you sleep an extra minute or an extra hour. You tell the memory how long.

E — ERROR

The Great Awakening is so smart it even tells you when you've made an error. But it's easy to correct...just press a button.

Model 7-4880

We bring good things to life.

GENERAL ELECTRIC

If you don't have millions in R&D behind you, your microcomputer should.

When it comes to choosing a high performance microcomputer company, you only have two choices.

A semiconductor company that's taking its first stab at the world of computers.

Or Digital.

No one has more time or money behind their 16-bit micros than Digital. Our LSI-11/2, LSI-11/23, PDP-11/03, and PDP-11/23 are the result of tens of millions of dollars and hundreds of man-years of development. They're direct descendants of our incredibly successful PDP-11 family of minicomputers. And they're every bit as successful in their own right. With more than 100,000 sold into the market now — more than our next three competitors combined.

It all means that with Digital's high performance micros, you get a head start no semiconductor company can give you.

Like a broad range of microcomputer boards, boxes, systems, and development tools to work with. High performance software that's literally years ahead of the competition. All backed by technical consultation and training from over 13,000 support people worldwide.

And best of all, Digital's microcomputers are software-compatible. Which means you won't ever have to worry about losing your software investment.

So if you don't have millions to spend on research and development, maybe you really don't have two choices at all.

Maybe you have just one.

Digital Equipment Corporation,
MR2-2/M65, One Iron Way,
Marlboro, MA 01752.

It took the minicomputer company to make micros this easy.

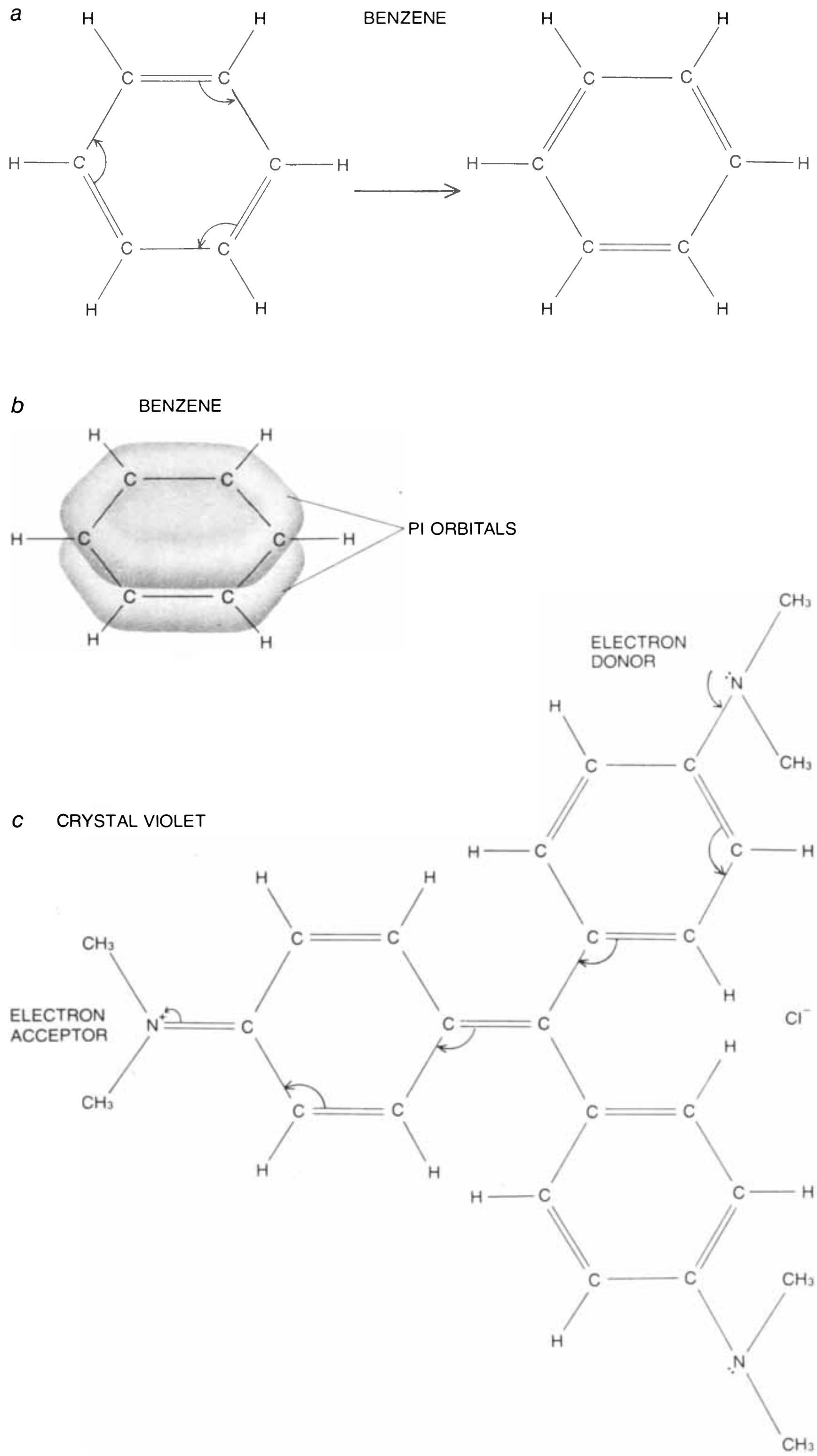
digital

ing ion or an impurity. A "hole," the absence of one electron from a pair, can have the same effect. Anomalies of this kind are called color centers or *F* centers (from the German *Farbe,* color).

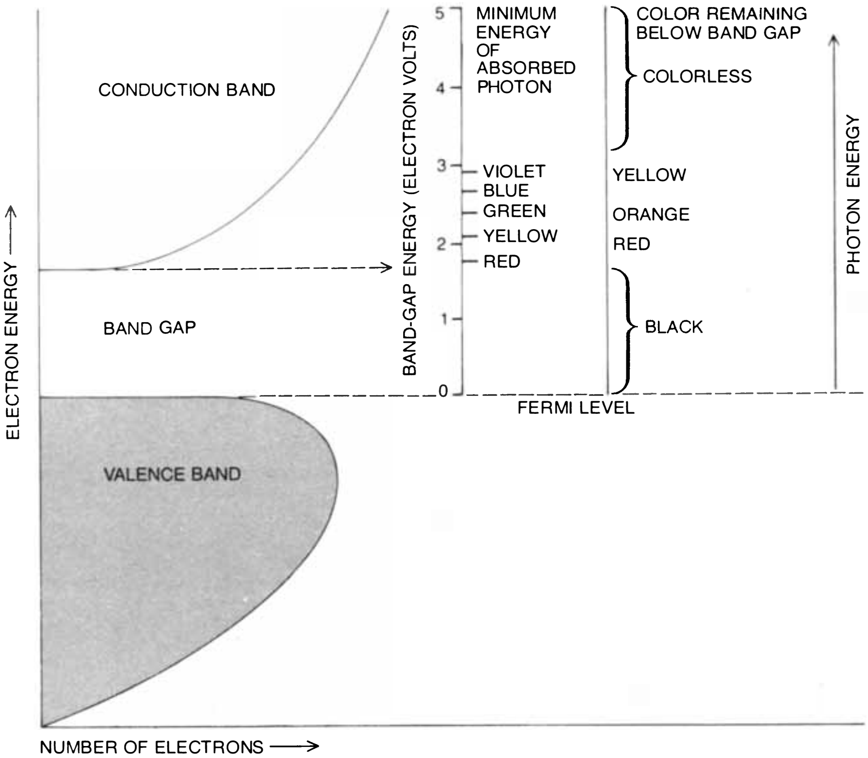
Although many color centers are known, the mechanism of color-production is understood in only a few. One of these is the purple *F* center of fluorite, a mineral that exhibits a great variety of color-center activities. Fluorite is calcium fluoride, or CaF_2, in which each calcium ion is normally surrounded by eight fluorine ions. An *F* center forms when a fluorine ion is missing from its usual position. The loss of the fluorine can come about in several ways: by growing the crystal in the presence of excess calcium, by exposing the crystal to high-energy radiation (which can displace an ion from its usual position) or by applying a strong electric field (which removes fluorine by the process of electrolysis). In order to preserve the electrical neutrality of the crystal some other negatively charged entity must take up the position in the lattice left vacant by the absent fluorine ion. When the charge is supplied by an electron, an *F* center is created. The electron is bound in place not by a central nucleus, as in an atom or an ion, but by the crystal field of all the surrounding ions. Within the field it can occupy a ground state and various excited states similar to those in the transition metals. The movement of electrons between these states gives rise both to color and to fluorescence.

The color of smoky quartz is attributed to a hole color center. The basic lattice of quartz is silicon dioxide (SiO_2), but a prerequisite to the formation of the color center is the presence of aluminum impurities replacing a few silicon ions. Because aluminum has a valence of +3 and silicon a valence of +4, an alkali metal ion or a hydrogen ion must be nearby to maintain electrical neutrality. Quartz is almost always contaminated with traces of aluminum, but that alone will not give rise to color because there are no unpaired electrons. The color center is created when the quartz is exposed for a few minutes to intense X rays or gamma rays, or when it is exposed to low levels of such radiation over geologic periods. The radiation expels one electron from a pair of electrons in an oxygen atom adjacent to an aluminum impurity, thereby leaving an unpaired electron in the orbital. The absent electron is called a hole, and the remaining unpaired electron has a set of excited states much like that of an excess electron.

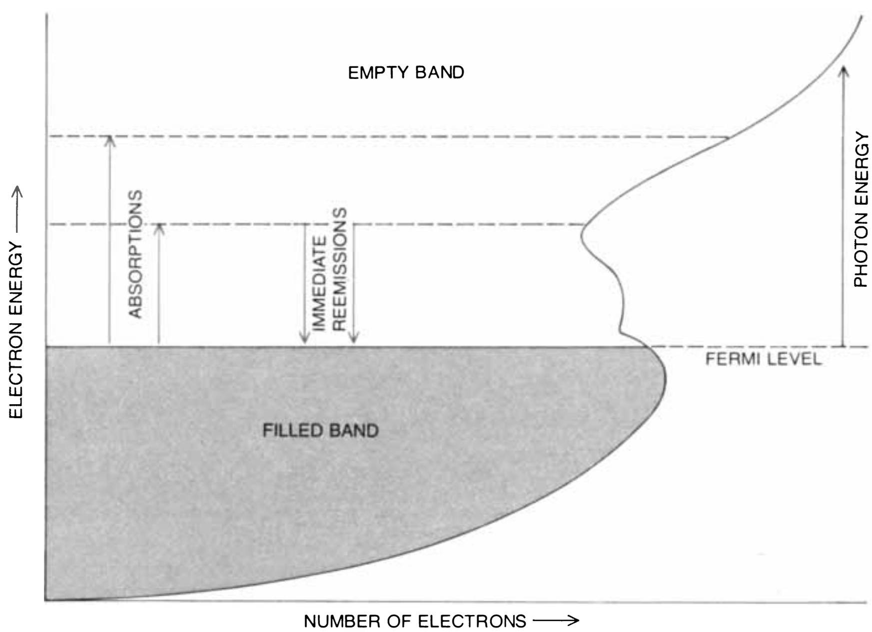
Amethyst is another form of quartz that derives its color from a hole color center, but the impurity ion is iron rather than aluminum. Certain old bottles were made of a glass containing iron and manganese; after many years of exposure to intense sunlight the glass turns purple through the development of col-



MOLECULAR ORBITALS describe paired electrons distributed over several atoms or over many atoms; they often absorb strongly at visible wavelengths. The pair of electrons that forms a chemical bond ordinarily absorbs only in the ultraviolet, but when alternative configurations of the bonds spread the pair over a number of atoms, the excitation energy of the pair is reduced. The commonest molecular orbitals that cause color are those associated with systems of conjugated bonds (alternating single and double bonds) in organic compounds such as benzene (*a*). By shifting three pairs of electrons in benzene the sequence of bonds is reversed. A better way of representing the structure of the molecule (*b*) shows single bonds between all the carbon atoms, with the extra three pairs of electrons distributed throughout the molecule in "pi" orbitals. It is the pi orbitals that are active in producing color. In benzene the excited states of the pi orbitals still lie in the ultraviolet, and so benzene is colorless, but in molecules with larger conjugated systems, such as the dye crystal violet (*c*), the bands are at visible wavelengths. The color is enhanced by chemical groups called auxochromes that donate and accept electrons.



ELECTRONIC STRUCTURE OF A METAL is distinguished by an essentially continuous band of allowed energy levels. The band is filled from the ground state up to an energy called the Fermi level; all higher energy states are empty and can therefore accept excited electrons. A consequence of this electronic configuration is that all wavelengths of radiation can be absorbed, from the infrared through the visible to the ultraviolet and beyond. A material that absorbs light of all colors might be expected to appear black; metals are not black because an excited electron can immediately return to its original state by reemitting a quantum with the same wavelength as the absorbed one. The metallic surface is therefore highly reflective.



BAND STRUCTURE OF A SEMICONDUCTOR is similar to that of a metal, except that a gap of forbidden energies separates the filled valence band from the empty conduction band. There is therefore a minimum energy that a quantum of radiation must have in order to be absorbed, namely the energy needed to promote an electron from the top of the valence band to the bottom of the conduction band. The color of a pure semiconductor is determined by the magnitude of the band gap. If it lies in the infrared, all visible wavelengths are absorbed and the material is black. A gap energy in the visible region allows some colors to be transmitted, so that the semiconductor takes on a color ranging from red to yellow. When the gap energy is in the ultraviolet, all visible wavelengths are transmitted and the material is colorless.

or centers and is called desert-amethyst glass. The same effect is achieved by a 10-minute exposure to the intense gamma rays emitted by cobalt 60.

Most color centers are stable if the material is not heated excessively. In fluorite raising the temperature makes the displaced fluorine ion mobile, so that it eventually resumes its original position; the color center and the color are then abolished. Amethyst when heated changes color, becoming either citrine quartz or a rare greened quartz. Both these colors represent the influence of iron without the amethyst color center. In some materials even the energy of sunlight can cause a color center to fade. For example, some colorless topaz can be irradiated to induce a deep orange-brown tint; the color fades, however, after only a few days in sunlight. Natural topaz of the same color is quite stable.

Molecular Orbitals

It was pointed out above that in molecules and solids the valence electrons are paired in chemical bonds and as a result their excited states are shifted into the ultraviolet. Actually this is true only when the paired electrons remain confined to a particular bond between two atoms. In many cases the electrons can move over longer distances; they can range throughout a molecule or even throughout a macroscopic solid. They are then bound less tightly, and the energy needed to create an excited state is reduced. The electrons are said to occupy molecular orbitals (in contrast to atomic orbitals), and they are responsible for a varied class of colors in nature.

One mechanism by which molecular orbitals can contribute color is the transfer of electric charge from one ion to another; blue sapphire provides an example of the process. Like ruby, sapphire is based on corundum, but it has two significant impurities, iron and titanium, both appearing in positions normally filled by aluminum. In the lowest energy state the iron has a formal charge of +2 and the titanium has a formal charge of +4. An excited state is formed when an electron is transferred from the iron to the titanium, so that both ions have a formal charge of +3. An energy of about two electron volts is needed to drive the charge transfer. Such transitions create a broad absorption band that extends from the yellow through the red and leaves the sapphire with a deep blue color.

In a number of materials iron is present in both its common valences, Fe^{++} and Fe^{+++}. Charge transfers between these forms give rise to colors ranging from deep blue to black, as in the black iron ore magnetite.

Molecular-orbital theory also applies to the colors observed in many organic substances in which carbon atoms (and

HOW TO BUILD A BOEING.

Boeing is now building an airplane that'll save millions of gallons of fuel: the new generation 757.

How much will it save?

A 757 flying 1.3 million miles a year will save 1.8 million gallons of fuel annually over the airplane it will replace.

Fuel economy has always been a concern to an airplane designer. Now it's more critical than ever.

Heading up one engineering group is Doug Miller. His specialty is preliminary design and development.

Doug and the design team were involved in the development of an airplane wing that has exceptionally efficient aerodynamics. It

would never have been possible if Boeing had not been testing and perfecting components of lightweight carbon-fiber and high-strength pure aluminum alloys.

Result: The 757 wing will get the airplane off the ground and flying with less fuel consumption than any airplane its size.

New aerodynamics, new materials, and engineering inventiveness along with a long-term commitment to constantly reach out and explore the unknown have made the 757 a reality.

Boeing believes almost anything is possible. Doug and the team are but one part of the immensely complex process that has helped create a future world of efficient commercial aviation — a step that can keep air travel one of the best values in the world of rising inflation.

THE BOEING FAMILY

Getting people together.

sometimes nitrogen atoms) are joined by a system of alternating single and double bonds, which are called conjugated bonds. The best-known example of such a conjugated system is the six-carbon benzene ring, but there are many others. Because each bond represents a pair of shared electrons, moving a pair of electrons from each double bond to the adjacent single bond reverses the entire sequence of bonds. The two structures defined in this way are equivalent, and there is no basis for choosing between them. Actually the best representation of the structure shows all the atoms connected only by single bonds; the remaining pairs of bonding electrons are distributed over the entire structure in molecular orbitals, which in this instance are called pi orbitals.

The extended nature of the pi orbitals in a system of conjugated bonds tends to diminish the excitation energy of the electron pairs. In benzene the energy of the lowest excited state is still in the ultraviolet, and so benzene is colorless, but in larger molecules, and notably in those built up of multiple rings, absorption can extend into the visible region. Such colored organic molecules are called chromophores (meaning color-bearers). Chemical side groups that donate or accept electrons can be attached to the conjugated system to enhance the color; they are called auxochromes (color-increasers).

A number of biological pigments owe their color to extended systems of pi orbitals. Among them are the green chlorophyll of plants and the red hemoglobin of blood. Organic dyes employ the same mechanism. The *Color Index* of the Society of Dyers and Colourists lists 8,000 such substances.

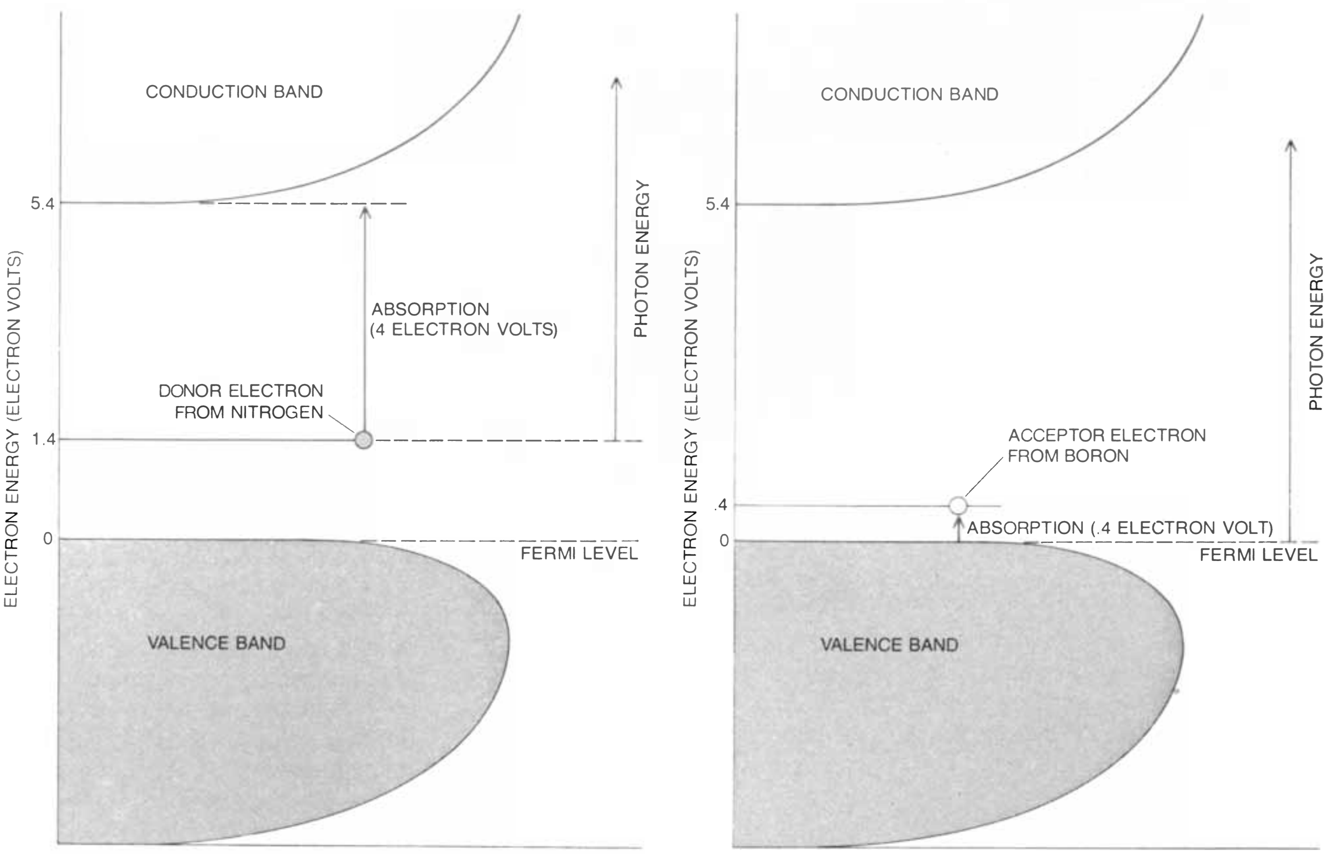
Some chemical species with extended molecular orbitals are also capable of fluorescence. As in ruby and emerald, the fluorescent light is emitted when an excited state decays through an intermediate level and the energy of at least one of the intermediate transitions corresponds to a visible wavelength. The fabric brighteners added to some detergents achieve their effect by absorbing ultraviolet radiation in daylight and reemitting a part of the energy as blue light. Lasers that have a dye as their active medium operate on the same principle.

Fluorescence is also observed in some systems of molecular orbitals where the energy to create the initial excited state comes from a source other than radiation. The bioluminescence of fireflies and of some deep-sea fishes is driven by a sequence of chemical reactions that culminates in the formation of an excited state of a molecule with extended pi orbitals. Manmade "cold light" devices imitate the biological process.

Metals and Semiconductors

The spatial extent of electron orbitals reaches its maximum possible value in metals and semiconductors. Here the electrons are released entirely from attachments to particular atoms or ions and can move freely throughout a macroscopic volume. Their range is limited only by the dimensions of the material. The enormous numbers of mobile electrons (on the order of 10^{23} per cubic centimeter) give metals and semiconductors optical and electrical properties unlike those of any other material.

In a metal all the valence electrons are essentially equivalent since they can freely exchange places. One might suppose, therefore, that they would all have the same energy, but a rule of quantum



DOPING OF A SEMICONDUCTOR with impurities creates allowed energy levels within the band gap. In diamond, which is a semiconductor with a band gap of 5.4 electron volts, doping with nitrogen introduces a level of filled states 1.4 electron volts above the Fermi level. One of these electrons can be excited to the conduction band by absorbing radiation with an energy of four electron volts. Because the level is broadened somewhat, there is some absorption of violet and blue light, and the diamond is colored yellow. Adding boron to a diamond creates holes, or empty states, in a level centered .4 electron volt above the Fermi level. Light of the lowest visible energies can promote an electron from the valence band into this broadened level, so that red and yellow light is absorbed and the diamond becomes blue.

mechanics forbids that. In a solid as in an isolated atom only a limited number of electrons can occupy the same rung of the energy ladder. Accordingly there must be many energy levels in the metal, and the levels are necessarily spaced very closely. In effect they form a continuum from the ground state up.

At zero temperature this continuous band of states is filled from the lowest level up to an energy designated the Fermi level; all states above the Fermi level are vacant. Any input of energy, no matter how small, propels an electron into one of the empty states of higher energy.

Since a metal effectively has a continuum of excited states, it can absorb radiation of any wavelength. A surface with this property of absorbing all colors might be expected to appear black, but metals are not an absorptive black. The reason again has to do with the agility of the metallic electrons: when an electron in a metal absorbs a photon and jumps to an excited state, it can immediately reemit a photon of the same energy and return to its original level. Because of the rapid and efficient reradiation the surface appears reflective rather than absorbent; it has the luster characteristic of metals. If the surface is smooth enough, the reflection can be specular, as in a mirror.

The variations in the color of metallic surfaces, which, for example, distinguish gold from silver, result from differences in the number of states available at particular energies above the Fermi level. Because the density of states is not uniform some wavelengths are absorbed and reemitted more efficiently than others. Besides absorption and reemission, transmission is also possible in some metals, although it is always weak and can be observed only in thin layers. The transmission varies with wavelength, and so it gives rise to color. Beaten gold leaf, in which the gold structure is highly distorted, transmits only green light. Colloidal gold in ruby glass is not under strain, and the light transmitted is purple-red.

Band-gap Colors

Wide, continuous bands of electronic states are a feature of another class of materials: the semiconductors. The definitive characteristic of these materials is that the average number of bonding electrons per atom is exactly four. Included are crystalline forms of some elements in Group IV of the periodic table, such as silicon, germanium and the diamond phase of carbon; there are also many compound semiconductors, such as gallium arsenide.

What distinguishes a semiconductor from a metal is a splitting of the band of energy levels into two parts. All the lower energy levels form a valence band, which in the ground state is completely

WHAT'S BETTER THAN SPEED READING?

SPEED LEARNING

(SPEED PLUS COMPREHENSION)

Speed Learning is replacing speed reading. It's easy to learn...lasts a lifetime...applies to everything you read...and is the only accredited course with the option of college or continuing education credits.

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

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

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

The new *Speed Learning Program* shows you step-by-proven-step how to increase your reading skill and speed, so you understand more, remember more and use more of everything you read. The typical remark made by the 75,000 slow readers who completed the *Speed Learning Program* was: "Why didn't someone teach me this a long time ago?" They were no longer held back by the lack of skills and poor reading habits. They could read almost as fast as they could think.

What makes Speed Learning so successful?

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

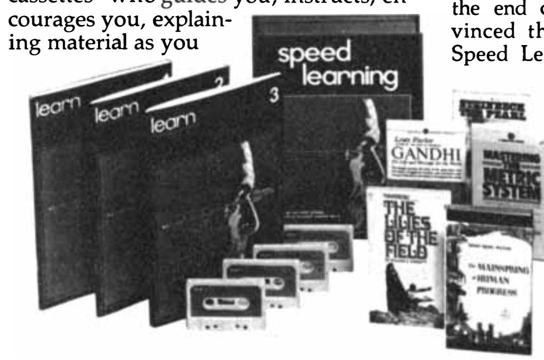
In just a few spare minutes a day of easy reading and exciting listening, you discover an entirely new way to read and think — a radical departure from anything you have ever seen or heard about. Research shows that reading is 95% *thinking* and only 5% eye movement. Yet most of today's speed reading programs spend their time teaching you rapid eye movement (5% of the problem) and ignore the most important part (95%) *thinking*. In brief, *Speed Learning* gives you what speed reading *can't*.

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

Listen-and-learn at your own pace

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

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



Examine Speed Learning FREE for 15 days

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

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

Note: Many companies and government agencies have tuition assistance plans for employees providing full or partial payment for college credit programs.

In most cases, the entire cost of your Speed Learning Program is Tax Deductible.

COLLEGE CREDITS

You may obtain *2 full semester hour credits* for course completion, wherever you reside. Credits offered through Whittier College (California). Details included in your program.

CONTINUING EDUCATION UNITS

National Management Association, the world's largest association of professional managers, awards 3.0 CEU's for course completion. CEU's can be applied toward the certificate in Management Studies.

PROFESSIONAL SOCIETIES

Speed Learning is offered internationally to members of professional associations such as: American Chemical Society, Foundation for Accounting Education, Institute of Electrical and Electronics Engineers and dozens more. Consult your Education Director for information.

BUSINESS, INDUSTRY, GOVERNMENT

Many companies and government agencies offer Speed Learning as a wholly-paid or tuition reimbursement program. Consult your Training or Personnel Director for details.

31SAY

learn INCORPORATED 113 Gaither Drive, Mt. Laurel, NJ 08054

YES! Please rush me the materials checked below:

- ☐ Please send the Speed Learning program @ $89.95 plus $4 postage and handling.
- ☐ Please send the Speed Learning Medical Edition @ $99.95 plus $4 postage and handling.
- ☐ Please send the Junior Speed Learning program (ages 11 to 16) @ $79.95 plus $4 postage and handling.

NJ residents add 5% sales tax.

Check method of payment below:

- ☐ Check or money order enclosed
- ☐ Charge my credit card under the regular payment terms

☐ Visa ☐ Master Card Interbank No.________ ☐ American Express

Card No.________ Exp. Date________

I understand that if after 15 days I am not delighted in every way, that I may return the materials in their original condition for a full refund. No questions asked.

Name ________

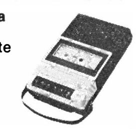
Address ________

City________ State________ Zip________

Signature________

If you don't already own a cassette player, you may order this Deluxe Cassette Recorder for only $49.95. (Includes handling and delivery.)

Check here to order ☐



Outside USA add $10 per item plus $4 surface mail—Airmail extra

A 1½" Panasonic TV with AM/FM makes it amazing. Playing almost anywhere in the world makes it TravelVision.™

It's a whole new breed of portable television. TravelVision from Panasonic.

It's TravelVision because with this slim 1½" TV (meas diag), you can watch programs in more than 100 countries. From Alexandria, Virginia to Alexandria, Egypt and just about anywhere in between. In North, South and Central America. Europe. Australia. Asia and Africa.

It's TravelVision because you can use it in airports, cars and train stations waiting to get to all those places. Not to mention your own backyard or bedroom, or just about anywhere else.

And it's TravelVision because it's so sleek and slim it fits into your attaché case or pocketbook, so you can take it almost anywhere. And it can play almost anywhere because it comes with both a universal AC adapter and a built-in rechargeable battery. There's even an optional car-adapter cord available.

Snap on the magnifying lens and you'll enjoy an even larger picture. Or, if you get tired of TV, enjoy superb AM or FM radio.

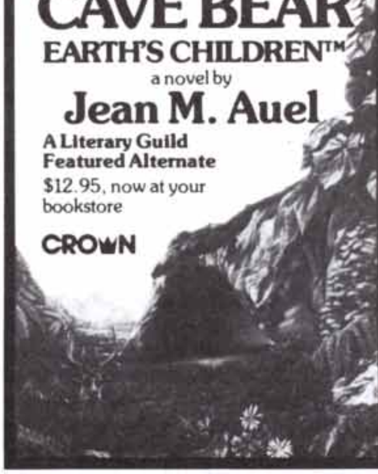
TravelVision. Electronic tuning. Automatic voltage regulator. And more. Which all makes it amazing by any standard, even ours. TV picture simulated

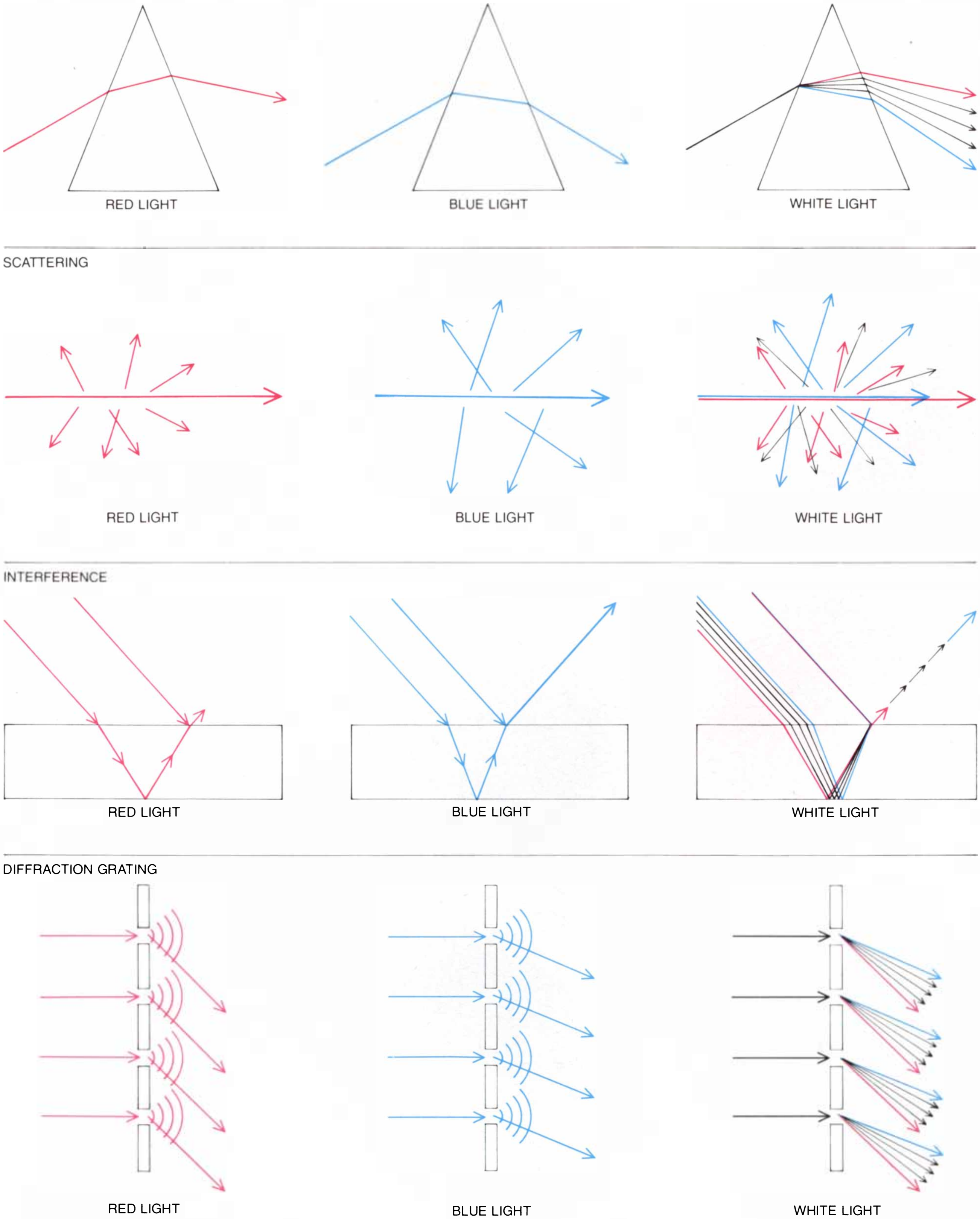
Panasonic.
just slightly ahead of our time.

filled. All excited states lie in a separate conduction band, which in the ground state is entirely empty. Separating the two bands is a gap of forbidden energies. The division of the energy band has profound effects on the optical properties of a semiconductor. No longer can the electrons absorb radiation of arbitrarily low energy; the minimum energy is the energy needed to lift an electron from the top of the valence band to the bottom of the conduction band.

The color of a pure semiconductor depends only on the magnitude of the energy gap. If the gap is smaller than the lowest energy of visible light, all visible wavelengths are absorbed. Small-gap semiconductors in which reemission is efficient and rapid have a metal-like luster, as silicon does, but other small-gap

BAND-GAP COLORS are observed in six semiconducting compounds. Pure cadmium sulfide (*bottom left*) has a band-gap energy of 2.6 electron volts, with the result that only violet light can be absorbed; the material has the color complementary to violet, namely yellow. In pure cadmium selenide (*bottom right*) the band-gap energy is 1.6 electron volts, so that all visible wavelengths are absorbed and the crystals are black. The other four materials are mixtures of various proportions of the two compounds; they show a gradient in absorption, with colors from orange through red. The compounds are employed as pigments under the name cadmium red.





PHYSICAL OPTICS provides the most convenient interpretation of colors generated by several mechanisms that involve a change in the direction of light. The dispersion of white light into its component colors by a transparent prism comes about because short wavelengths are refracted through a larger angle than long wavelengths. In a similar way the scattering of light by small particles is more effective at short wavelengths, so that more blue light is scattered than red. Interference is observed when a light wave is split into two parts that are then brought together again. If the waves are in phase when they recombine, the intensity is enhanced; if they are out of phase, the waves cancel each other. Since the phase difference between the two beams can depend on wavelength, interference can enhance some colors and suppress others. In a diffraction grating light is scattered by many uniformly spaced centers and the resulting multiple wave fronts interfere with one another. Each wavelength is reinforced in one set of directions and canceled in all others, so that white light is dispersed.

semiconductors are black. At the opposite extreme the band gap can be greater than the highest energy of visible light; in that case no visible wavelengths can be absorbed and the material is colorless. Diamond, with a band gap of 5.4 electron volts, is one such large-gap semiconductor and is transparent both to visible light and to a limited range of ultraviolet radiation.

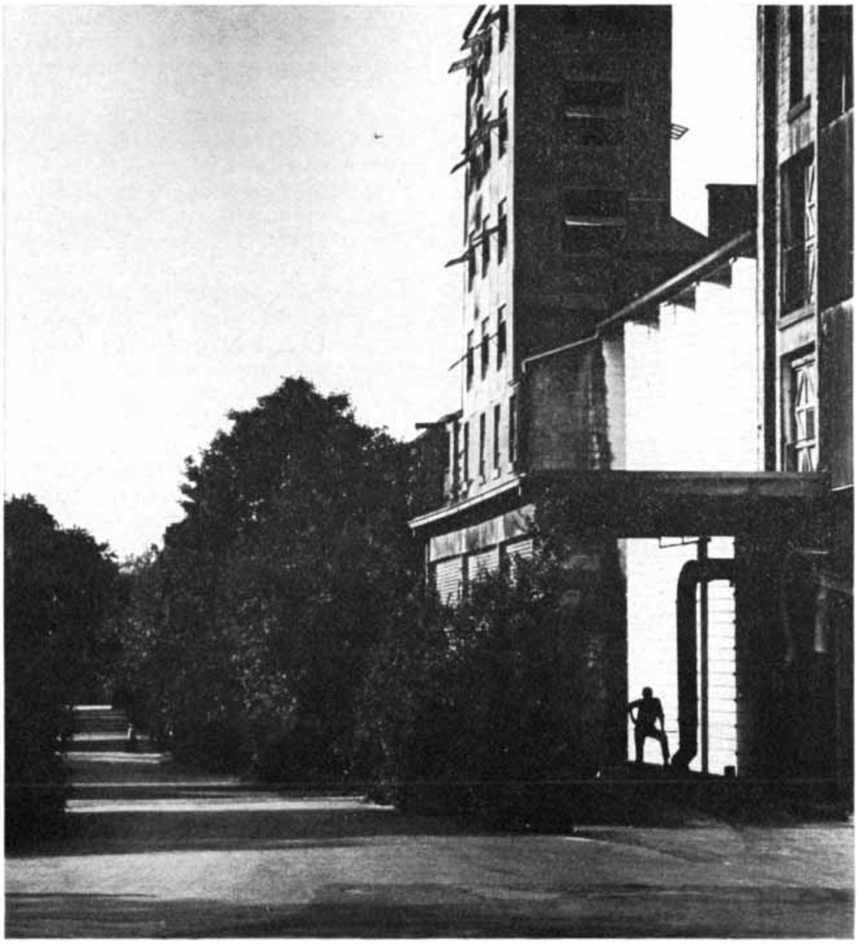
Where the band-gap energy falls in the visible range, the semiconductor has a definite color. The mercury ore cinnabar (which has the formula HgS and is also known as the pigment vermilion) has a band-gap energy of 2.1 electron volts. All photons with energies higher than this level are absorbed, and only the longest visible wavelengths are transmitted; as a result cinnabar appears red. The pigment cadmium yellow, CdS, has a band gap of 2.6 electron volts and absorbs only blue and violet light; after these wavelengths are subtracted from white light the color remaining is yellow. The sequence of band-gap colors, from the smallest gap to the largest, is black, red, orange, yellow, colorless.

Although large-gap semiconductors are colorless when they are pure, they can take on color when they are "doped" with traces of an impurity. The impurity can be either a donor of electrons or an acceptor of them; in either case it introduces a set of energy levels in the gap between the valence band and the conduction band. Transitions to or from the new levels require a smaller quantity of energy than transitions across the entire gap.

In diamond the replacement of a few carbon atoms with nitrogen creates a donor band about 1.4 electron volts above the valence band. Nitrogen has one electron more than carbon, and it is these extra electrons that form the donor band. The nominal energy needed to promote one of them to the conduction band is four electron volts, which is still in the ultraviolet; the donor level is sufficiently wide, however, for some violet light to be absorbed. At a concentration of one nitrogen atom per 100,000 carbon atoms diamond is yellow; with more nitrogen it becomes green.

Boron has one electron fewer than carbon and gives rise to acceptor, or hole, levels in the band gap of diamond. An electron can be excited from the valence band to occupy one of these holes. The impurity level is centered .4 electron volt above the valence band but extends far enough to absorb some of the longer visible wavelengths. The boron-doped diamond therefore appears blue.

It is well known that doping also alters the electrical properties of semiconductors, making possible all the devices of solid-state electronics. Among these devices are light-emitting diodes and semiconductor lasers, in which an elec-

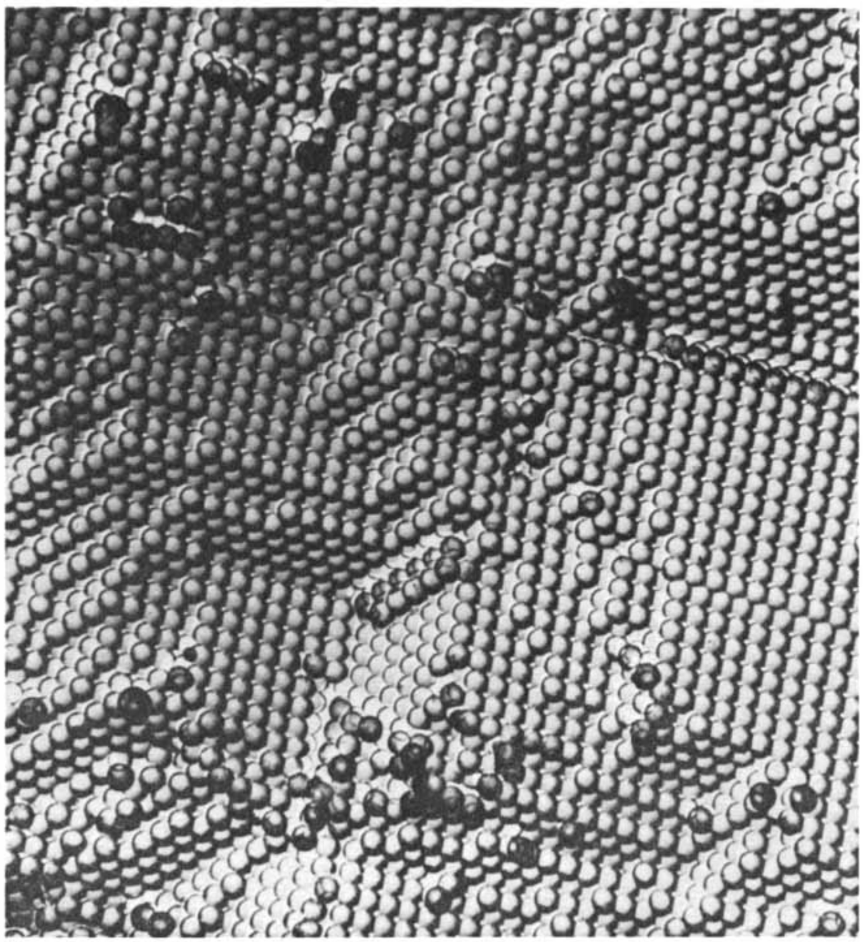


tric current populates excited states and the electrons emit radiation in returning to the ground state. Doped semiconductors can also function as phosphors, which are materials that give off light with high efficiency when they are stimulated electrically or by some other means. Phosphors are the luminous sources in a fluorescent lamp and in the picture tube of a television set. In a color picture tube three phosphors are distributed over the surface, with emissions at red, green and blue wavelengths. Certain compounds of the transition metals also act as phosphors.

Geometrical Optics

In the phenomena I have described so far color results either from the direct emission of colored light or from the selective absorption of some wavelengths and transmission of others. In a last group of color-producing phenomena interactions of light with matter change the direction of the light. The change in direction is the primary cause of color in refraction and diffraction, where the magnitude of the deflection can vary with wavelength. In the scattering of light from small particles the deflection of any given ray is not a deterministic function of wavelength, but the average intensity of the scattered light does depend on wavelength. Interference gives rise to color through an interaction of light with light, but a change in direction is needed to enable one beam of light to interfere with another.

At the most fundamental level these processes can be understood in terms of electronic excitations in matter. Refraction, for example, results from a change in speed when light passes from one medium into another; the speed in each material is determined by the interaction of the electromagnetic field of the radiation with the electric charges of the electrons. An analysis of this kind is always possible, but it is often too cumbersome to be very informative. What is needed is a "higher level" analysis; it is provid-



NATURAL DIFFRACTION GRATING is formed by closely packed spheres in a synthetic opal. The spheres consist of silicon dioxide and a little water, embedded in a transparent matrix that has a similar composition but a slightly different index of refraction. The spheres have a diameter of about 250 nanometers. Dispersion of white light by the three-dimensional diffraction grating gives rise to spectrally pure colors that glint from within an opal. The synthetic opal was made by Pierre Gilson in France, who also made this scanning electron micrograph.

ed by the methods of geometrical and physical optics.

A light ray is refracted, or bent, by a transparent prism because light is slowed more by traveling in a solid than it is by traveling in air. (The highest possible speed of light, and the speed that relativity theory says can never be exceeded, is attained only in a vacuum; in any material medium the speed is lower.) Why a change in speed should lead to a change in direction can be understood by imagining the light ray as a series of plane wave fronts. When a wave front strikes a transparent surface obliquely, one edge enters and is slowed before the opposite edge reaches the surface; the edge that is slowed first thus falls behind the rest of the wave front, and the ray is deflected toward the perpendicular. On leaving the solid the ray is bent away from the perpendicular.

The magnitude of the change in direction when a light ray is refracted depends on the angle at which it strikes the surface. For any given angle of incidence it also depends on the ratio of the speeds of light in the two materials. It turns out that this ratio, which is called the refractive index, is generally not the same for all wavelengths; that is why refraction can give rise to color. The mechanisms that retard light in a transparent medium have a greater effect on high frequencies and short wavelengths than they do on low frequencies and long wavelengths. As a result violet light is refracted through a larger angle than red light, and a beam of white light on passing through a prism is separated into its component colors.

This dispersion of white light according to wavelength was discovered by Isaac Newton in 1704, or at least it was Newton who first recognized its significance. The dispersion of sunlight by refraction in water droplets or ice crystals is responsible for the colors of a rainbow and of other colored halos occasionally seen around the sun and the moon. Dispersion also causes the "fire," or flashes of color, characteristic of diamond and, to a lesser extent, of other faceted transparent gemstones. Dispersion is not always a welcome phenomenon: in telescopes, cameras and other optical systems it gives rise to chromatic aberration, or the misregistration of images in different colors.

Interference was also first investigated in detail by Newton. It can be observed in a system of waves where two waves combine to yield a single new wave whose amplitude at each point is simply the sum of the amplitudes of the original waves. Thus a monochromatic light wave can be split into two components, which follow different paths and then merge again. Where the two components are in phase, so that the peaks and valleys coincide, the waves reinforce each other and the light is bright. Where the components are out of phase the waves cancel and appear dark.

Interference is often observed in thin transparent films, where part of the light is reflected by the first surface and part by the second. Whether the beams reinforce or cancel depends on the nature and thickness of the film, on the angle of reflection and on the wavelength of the light. If the film has a uniform thickness, different wavelengths emerge at different angles. If the layer varies in thickness, then at any given viewing angle different colors appear at different positions. Interference in a thin, transparent cuticle gives rise to the colors of some beetles and butterflies. A thin layer of oil on water swirls with color as a result of the same mechanism.

The Diffraction Grating

Diffraction is the bending or spreading of a light wave at the edge of an opaque obstacle. When the obstacle is a macroscopic object, the effect is generally a minor one, but it becomes important when the dimensions of the obstacle are comparable to the wavelength of the light. As might be expected, the magnitude of the effect then becomes strongly dependent on wavelength.

Diffraction is an important cause of color in the scattering of light by small particles. Lord Rayleigh showed that the intensity of the scattered light is inversely proportional to the fourth power of the wavelength; accordingly blue light is scattered about four times as much as red light. The disparity in scattering effectiveness is apparent in the daytime sky. Molecules, dust and fluctuations in density in the atmosphere preferentially scatter blue light, so that the sky appears blue. The direct light of the sun, on the other hand, is depleted of blue and therefore appears reddish, particularly at sunrise and sunset when the light must traverse a greater depth of the atmosphere.

The same process can be demonstrated on a smaller scale by passing white light through diluted unhomogenized milk, where the scattering centers are particles of fat. The fluid appears bluish, whereas the directly transmitted beam is reddish. The blue sheen of moonstone almost certainly comes from a similar process. Scattering from much larger inclusions, which do not alter color, accounts for the patterns in star ruby and star sapphire and in tiger's-eye quartz.

Interference and diffractive scattering act in combination in a final mechanism for generating color: the diffraction grating. Such a grating is an array of many equally spaced lines or points in which the spacing is not too large in comparison with light wavelengths. Light passing through the grating is scattered in all directions at each opening, and so waves from adjacent openings interfere. For any one wavelength there are angles at which the interference is constructive; at all other angles the waves cancel one another. The actual value of these angles depends only on the wavelength and on the spacing between lines or points on the grating. When white light passes through a diffraction grating or is reflected from it, each wavelength is enhanced in a different set of directions. Hence the light in any given direction is spectrally pure and monochromatic, but as the grating is rotated a series of spectra comes into view.

A diffraction grating can be made by inscribing fine rulings on a glass plate, and a few natural systems also have the

ELECTRONIC TRANSITIONS IN FREE ATOMS AND IONS; VIBRATIONAL TRANSITIONS IN MOLECULES	ELECTRONIC EXCITATIONS	INCANDESCENCE, FLAMES, ARCS, SPARKS, LIGHTNING, GAS DISCHARGES, SOME LASERS.
	VIBRATIONS	BLUE-GREEN TINT OF PURE WATER AND ICE.
CRYSTAL-FIELD COLORS	TRANSITION-METAL COMPOUNDS	TURQUOISE, MOST PIGMENTS, SOME LASERS, SOME PHOSPHORS, SOME FLUORESCENT MATERIALS.
	TRANSITION-METAL IMPURITIES	RUBY, EMERALD, RED SANDSTONE, SOME LASERS, SOME FLUORESCENCE.
	COLOR CENTERS	AMETHYST, SMOKY QUARTZ, DESERT-AMETHYST GLASS, SOME FLUORESCENCE.
TRANSITIONS BETWEEN MOLECULAR ORBITALS	CHARGE TRANSFER	BLUE SAPPHIRE, MAGNETITE.
	CONJUGATED BONDS	ORGANIC DYES, MOST PLANT AND ANIMAL COLORS, LAPIS LAZULI, FIREFLIES, DYE LASERS, SOME FLUORESCENCE.
TRANSITIONS IN MATERIALS HAVING ENERGY BANDS	METALLIC CONDUCTORS	COPPER, SILVER, GOLD, IRON, BRASS.
	PURE SEMICONDUCTORS	SILICON, GALENA, CINNABAR, DIAMOND.
	DOPED SEMICONDUCTORS	BLUE DIAMOND, YELLOW DIAMOND, LIGHT-EMITTING DIODES, SEMICONDUCTOR LASERS, SOME PHOSPHORS.
GEOMETRICAL AND PHYSICAL OPTICS	DISPERSIVE REFRACTION	THE RAINBOW, "FIRE" IN GEMSTONES, CHROMATIC ABERRATION.
	SCATTERING	BLUE OF THE SKY, RED OF SUNSETS, MOONSTONE, STAR SAPPHIRE.
	INTERFERENCE	OIL FILM ON WATER, LENS COATINGS, SOME INSECT COLORS.
	DIFFRACTION GRATING	OPAL, LIQUID CRYSTALS, SOME INSECT COLORS

CAUSES OF COLOR are classified in 14 categories of five broad types. All but one of the color-causing mechanisms (vibrations of the atoms in molecules) can be traced to changes in the state of the electrons in matter. Electronic transitions are the most important causes of color because the energy needed to excite an electron commonly falls in the range that corresponds to visible wavelengths of light.

dimensions and the high degree of order needed for an optical grating. For example, in the materials called liquid crystals molecules are stacked with sufficient regularity to act as a diffraction grating. Because the spacing between molecules depends on the temperature so does the color of the material, and it can be employed as a temperature indicator. Similar ordered arrays of molecules produce diffraction-grating colors in some butterflies and beetles. Spectral colors arising from diffraction also appear when a distant streetlight is seen through the fabric of an umbrella or when the surface of a phonograph record is viewed at a glancing angle.

The preeminent natural diffraction grating is the opal. In this gemstone spheres made up of silicon dioxide and a little water are closely packed in a three-dimensional array with a spacing of about 250 nanometers. The transparent or translucent matrix that fills the space between spheres has a similar composition but a slightly different index of refraction. When white light enters this three-dimensional diffraction grating, pure colors appear inside the stone; the colors change as the eye or the stone is moved. The mechanism of color production in opal came to be understood only in 1964, when the structure was first resolved with the electron microscope. Synthetic opals were created not long thereafter.

The Visible Spectrum

In this catalogue of colors, which is surely not complete, I have described more than a dozen mechanisms. They can be put into five broad categories: excitations of free atoms and ions and vibrations in molecules, crystal-field effects, transitions between states of molecular orbitals, transitions in the energy bands of solids and effects interpreted through physical optics. It may seem an extraordinary coincidence that such a diversity of phenomena is encompassed in a band of wavelengths that is not even a full octave wide; it may seem still more remarkable that this narrow band happens to be just the one to which the human eye is sensitive.

Actually it may not be a coincidence at all. So much of interest happens in this narrow region of the electromagnetic spectrum because these are the wavelengths where interactions of light with electrons first become important. Waves of lower energy mainly stimulate the motions of atoms and molecules, and so they are usually sensed as heat. Radiation of higher energy can ionize atoms and permanently damage molecules, so that its effects seem largely destructive. Only in the narrow transition zone between these extremes is the energy of light well tuned to the electronic structure of matter.

The Southern Appalachians and the Growth of Continents

A large seismic-reflection survey suggests that for at least half of the history of the earth continents have evolved by the stacking and shuffling of relatively thin sheets of material at their margins

by Frederick A. Cook, Larry D. Brown and Jack E. Oliver

The theory of plate tectonics, which regards the crust of the earth as the upper part of a set of interacting rigid plates, has done much to describe the evolution of the ocean floor but has left much unexplained about the formation and structure of the continents. This state of affairs is beginning to change with the intensive application of the petroleum industry's exploration technique of seismic-reflection profiling to the study of the deep crust and the underlying upper mantle. The technique, which depends on the reflection of sound waves by discontinuities in the density of the rock and in the velocity of sound through it, has mapped new details of the geological structure of the continental basement.

The application of the technique in the southern Appalachian area of the U.S. reveals how the margins of continents change as ocean basins close and continents collide at subduction zones: the deep oceanic trenches near continental margins where oceanic crust plunges into the mantle. It has long been suspected that the continents grow by the accretion of crustal material at the subduction zones. The seismic-reflection profiles of the southern Appalachians and related geological studies show how this actually happens: continents evolve out of the stacking and shuffling of thin horizontal sheets of crustal material.

The seismic profiles were made by the Consortium for Continental Reflection Profiling (COCORP), a group of geologists and geophysicists from universities, industry and Government that is headed by Sidney Kaufman of Cornell University and two of us (Brown and Oliver). COCORP collected its first data in Hardeman County, Tex., in 1975. Since then it has made profiles at 11 sites. By probing the crystalline rocks below sedimentary strata to depths of as much as 50 kilometers, the study has focused on the continental basement. The effort has been fruitful. COCORP has mapped a mid-crustal body of magma (liquid rock) in central New Mexico, explored an ancient buried rift valley in Michigan and traced a major thrust fault deep into the crust in southwestern Wyoming. Before the survey in Wyoming was done the depth and attitude of the fault, which bounds an important feature of the Rockies, had been the subject of much controversy.

The most spectacular finding was made in the southern Appalachians. The profiles revealed that the mountains are underlain to a depth of at least 18 kilometers by horizontal layers of material that is sedimentary or once was. Sedimentary rocks, of course, are one of the three broad types of rock, the other two being igneous and metamorphic. Sedimentary rocks originate with sediments deposited at the surface by water and wind. Igneous rocks are formed by the congealing of magma. Metamorphic rocks are formed when sedimentary or igneous rocks are subjected to heat and pressure over long periods of time. Metamorphic rocks can be described as high-grade or low-grade, depending on the degree to which they have been recrystallized in the process of metamorphism.

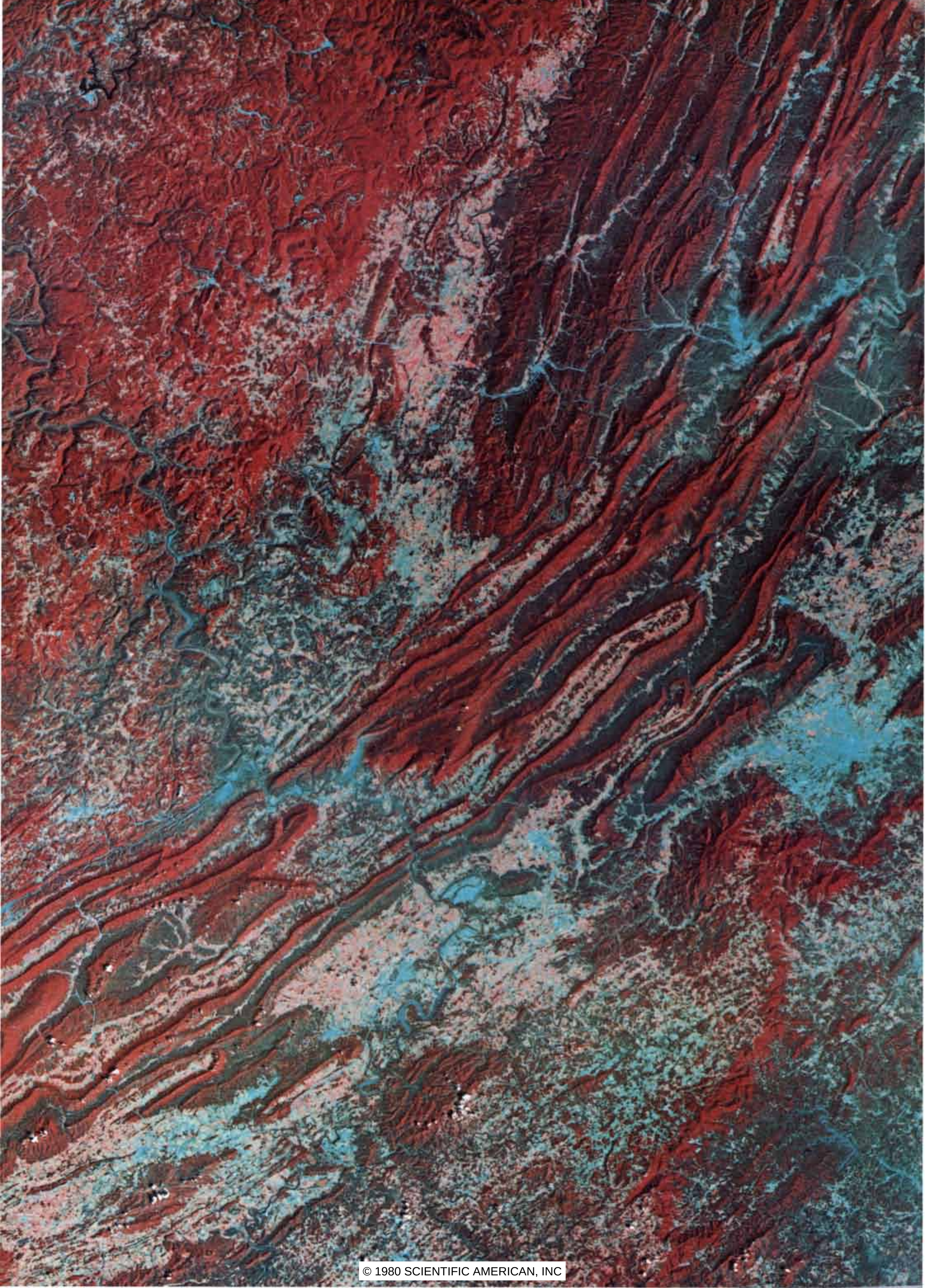
Most of the rocks at the surface in the southern Appalachians are highly deformed metamorphic ones. Furthermore, they are older than or contemporaneous with the horizontal sedimentary strata that were discovered under them. This fact suggests that roughly 475 million years ago the surface rocks began to be transported as a thin sheet for at least 260 kilometers over the eastern continental margin of the land mass that was to become North America. The discovery confirms the hypothesis that horizontal or near-horizontal thrusting can carry large volumes of crustal material great distances, and it suggests that a continent could grow and evolve by the emplacement of thin horizontal slices of material at the continental margins.

The seismic-reflection profiles may also have a more immediate kind of import. They have led geologists to speculate that there may be undiscovered deposits of oil or gas in the sedimentary rocks under part of the overthrust sheets. Sedimentary rocks are notably the best target for oil and gas exploration, because unlike igneous rocks and most metamorphic ones they have not been subjected to the pressures and temperatures that would destroy or expel the hydrocarbons.

The first seismic-reflection studies of sedimentary basins were made in the 1920's. Since then the technique has been developed intensively by the petroleum industry as part of its effort to locate economic oil- and gas-bearing formations. The theory of the technique is quite simple. Sound waves generated by an explosion or some other source of acoustic energy at the surface radiate downward to impinge on discontinuities in subsurface layers of rock. The waves are reflected wherever there is an abrupt change in the rock's density or in the waves' velocity. The reflected energy is detected at the surface by an array of geophones, or vibration sensors.

The time it takes the wave to travel from the source to the reflecting discontinuity and then back to the surface depends on the depth of the discontinuity.

SATELLITE IMAGE OF THE APPALACHIANS to the west of Roanoke, Va., is a false-color one made in the fall. Roanoke is the light green area to the right below the middle. The red areas are vegetation, the bright bluish green areas rivers and lakes. Here the mountains, which reach heights of between 500 and 1,000 meters, shift from a northeasterly trend to a more northerly one. The linear trends constitute the sedimentary material of the Valley and Ridge province. Seismic-reflection survey by the Consortium for Continental Reflection Profiling (COCORP) covered part of this area and terrains of crystalline rocks to the southeast.



The more you travel, the more you entertain, the more you need the American Express® Card.

Because the American Express Card is good for charging almost anything.

But the American Express Card can do far more than charge things.

For example, no other card has anything like the more than 1,000 Travel Service Offices of American Express, its subsidiaries and Representatives worldwide. For travel help at home and abroad. For help replacing a lost Card fast, or even with emergency money.

And unlike bank cards, there is no fixed limit imposed on your spending with the Card. Purchases are approved based on your ability to pay, as demonstrated by your spending, payment patterns and personal resources.

So if you want the world's most respected card for travel and entertaining, one that will help you manage your money responsibly, here's a perfect opportunity to apply.

If the application is missing, just phone 800-528-8000 toll-free and we'll send you another one.

AMERICAN EXPRESS

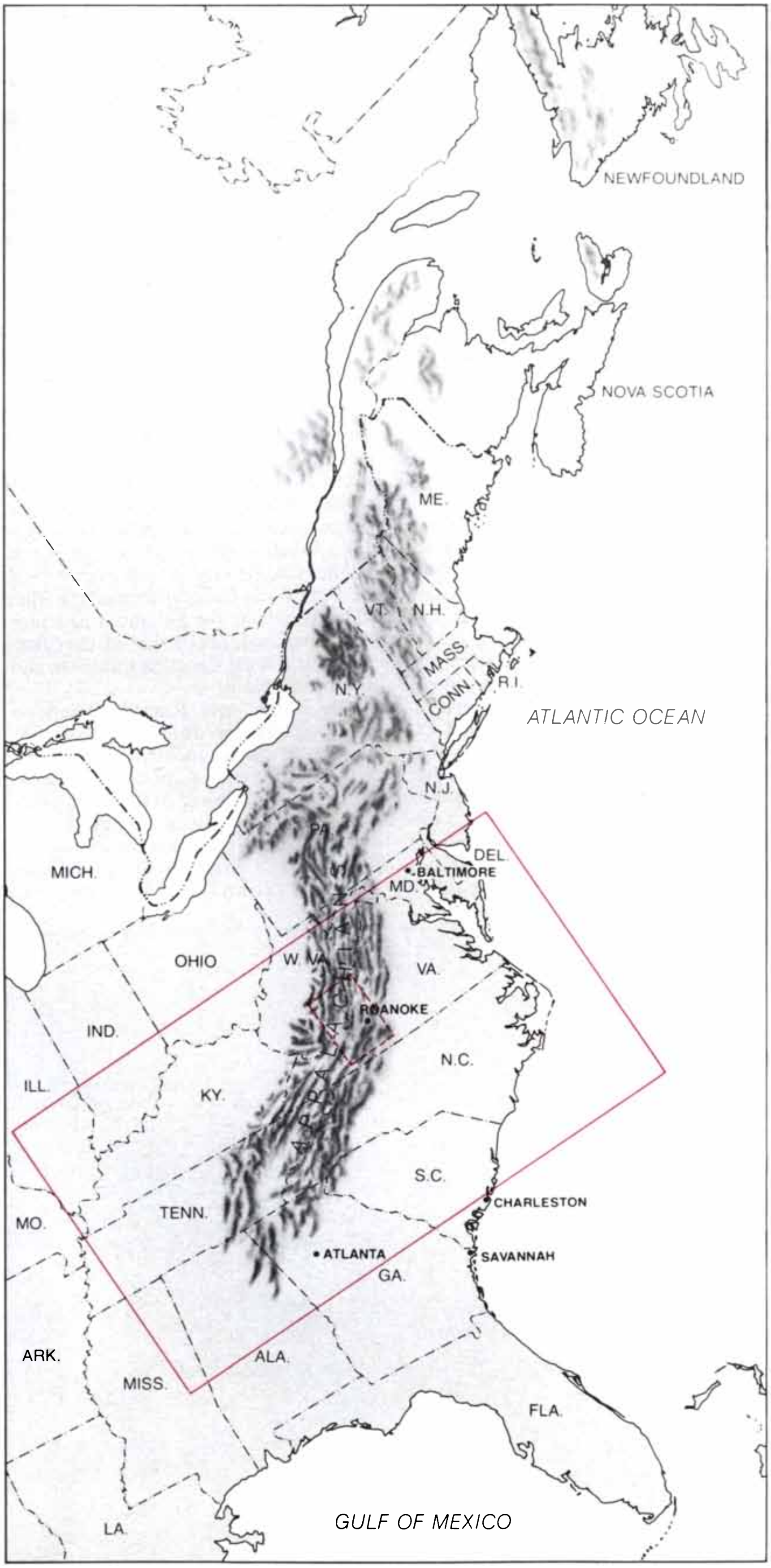
The American Express Card. Don't leave home without it. SM

Additional information about the discontinuity can be deduced from the recordings gathered by an array of geophones that respond to waves whose paths are almost vertical. The seismic-reflection profiles that are used to locate oil and gas respond to waves whose travel times are less than four or five seconds, corresponding to depths of between eight and 10 kilometers. Waves with longer travel times are mainly bouncing off rocks that are too deep to be reached by current techniques for recovering oil. The COCORP survey worked with waves whose travel times were as much as 20 seconds; such times correspond to depths of 60 or 70 kilometers.

COCORP did a multichannel reflection survey in which as many as 2,304 geophones (arranged in 96 groups of 24) were laid out over a distance of more than 6.5 kilometers. The entire spread of detectors was connected by cable to a truck in which the collected data were digitally stored on magnetic tape for the subsequent processing by computers that would construct a seismic image of the subsurface geology. If the wave sources and the geophones are positioned properly and the data are processed correctly, undesirable noise will be suppressed and the reflected signals will be enhanced.

Although explosives still sometimes serve as the source of acoustic energy in explorations for oil and gas, COCORP and much of the petroleum industry have turned to a nonexplosive, environmentally acceptable source named Vibroseis, which was invented at the Continental Oil Company in the 1950's. In a typical COCORP survey four or five truck-mounted vibrators direct into the ground a signal of between eight and 32 hertz. Over the sweep period of 30 seconds the frequency is made to vary linearly to form a "chirp." After each sweep the vibrators are moved forward a few meters and another sweep is made. After 16 sweeps covering a distance of 122 meters the process is repeated at the next source station 134 meters away.

The purpose of all of this is to record many reflections from the same subsurface point. The redundancy of the data enables the computer analysis to suppress noise, to enhance the strength of the reflected signal and to estimate the velocity of the subsurface vibrational waves. A seismic image of the structure of the continental basement is then constructed from the readings taken from many different configurations of sources and sensors. The process is somewhat like the medical technique of computed axial tomography (the "CAT scan"), in which X-ray readings taken from many different angles are combined to yield a representation of the internal structure of the living body in cross section.



MAP OF THE EASTERN U.S. AND SOUTHERN CANADA shows how the Appalachians extend from central Alabama to Newfoundland. The broken rectangle on the map marks the Roanoke area that is shown in the satellite image on page 157. The colored rectangle outlines the area of the southern Appalachians, which is shown in the map on the next page.

In seismic-reflection profiling it often takes a few months to gather and process the data for a particular site. The sources and detectors are moved across the surface at a rate of between one and four kilometers a day. To reconstruct the subsurface geological structure requires the processing of enormous quantities of data. For example, the analysis of a major northwest-to-southeast traverse in Georgia and Tennessee called for the manipulation of roughly two billion items of information: 3,843 vibration points multiplied by 96 channels multiplied by a recording time of 50 seconds divided by a sampling rate of .008 second.

The computer analysis of the COCORP data included the demultiplexing of the multichannel field recordings, the elimination of particularly noisy data, the calculation of wave velocities, the collection of all reflection signals from each common depth point, the compensation for differences in the distance between sources and receivers, the adjustment for differences in topography and near-surface geology and the stacking, or superposition, of the several coherent signals for each common reflection point. The result of this intensive processing is a seismic cross section that resembles a geological cross section except for the important differences that depth is represented not by distance but by travel time and that lateral variations in the velocity of sound in the rocks can distort the geometry of the reflections.

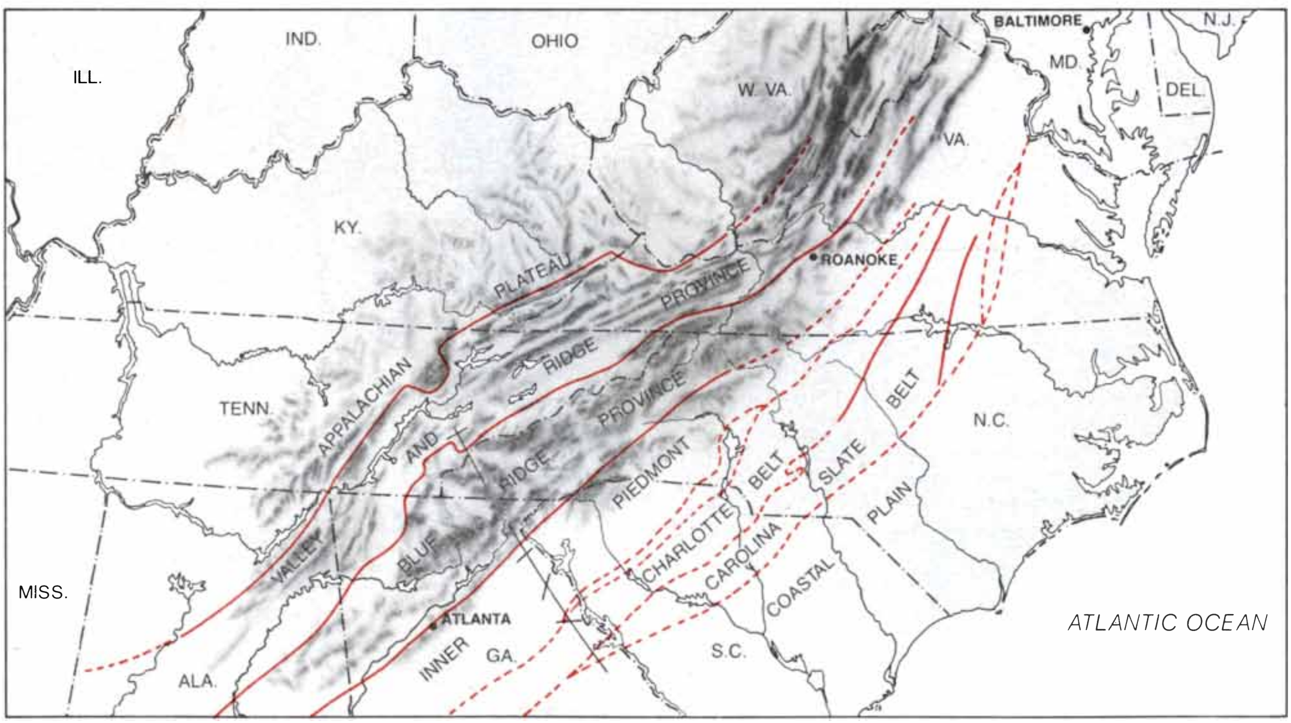
To infer an accurate geological cross section from the seismic section requires substantial skill and experience. For example, dipping interfaces are represented on the seismic cross section by reflections that seem out of position; the steeper the dip, the greater the "misplacement." Reflections often must be "migrated" back to give an accurate picture of the actual subsurface geology. (The COCORP profile of the southern Appalachians did not show many steep dips, so that migration was not much needed.)

Such was the exploration technique that was applied to the southern Appalachians. The Appalachians extend more than 3,000 kilometers from Newfoundland to central Alabama. The southern Appalachians consist of a series of distinct geological provinces and belts trending from northeast to southwest. (A province is a general term for any region whose rocks have a similar history; a belt is a long, linear feature whose rocks have a similar composition.) From northwest to southeast the southern Appalachians are made up of the Valley and Ridge province, the Blue Ridge province, the Piedmont province (including the Inner Piedmont, the Charlotte belt and the Carolina slate belt) and the coastal plain.

The Valley and Ridge province is characterized by folded and thrust-faulted strata of mostly unmetamorphosed sedimentary rocks formed between 600 million and 300 million years ago. The thrust faults and folds indicate that the rocks were much compressed in the horizontal direction. For a long time it was not known whether the deformation in the Valley and Ridge province involved the crystalline rocks forming the basement under the sedimentary strata (in which case the process would be called thick-skinned tectonics) or whether the deformation was confined chiefly to the sedimentary strata over the basement (thin-skinned tectonics). Now it is known from seismic-reflection profiles and explorations for oil and gas that the deformation is predominantly thin-skinned. Like a wrinkled rug on a floor the sedimentary strata seem to have ridden westward on top of large, horizontal detachment zones over the crystalline basement.

Southeast of the Valley and Ridge province is the Blue Ridge province. The main transition between the two is a large thrust fault that dips to the southeast. Unlike the sedimentary rocks of the Valley and Ridge province, the rocks of the Blue Ridge province have generally undergone much metamorphism. In both provinces however, the rocks have been deformed in three orogenies, or episodes of mountain building: the Taconic, the Acadian and the Alleghenian.

The basement of the Blue Ridge province includes Precambrian rocks that are at least a billion years old. Many geologists cited this fact as evidence that the Blue Ridge province is rooted in place, making it part of the main body of basement rock that forms the crust there. New data show, however, that although unmetamorphosed sedimentary rocks are rare in the province, they are occasionally found in "windows": areas



SOUTHERN APPALACHIANS are made up of four major topographic features trending from northeast to southwest: the Valley and Ridge province, the Blue Ridge province, the Piedmont province (including the Inner Piedmont, the Charlotte belt and the Carolina slate belt) and the coastal plain. COCORP studied the continental basement of the area along black line. Red lines mark thrust faults.

where the metamorphic rocks have eroded away to expose the underlying material. The windows are mostly in the western Blue Ridge, and so it is now known that at least there the crystalline basement rocks are underlain by layers of unmetamorphosed sedimentary rocks. Moreover, one of the largest windows, the Grandfather Mountain window of North Carolina, is in the eastern Blue Ridge. This area, which exposes Cambrian carbonate rocks and shales, demonstrates that sedimentary rocks that have undergone little metamorphism are present under the eastern Blue Ridge as well as under the western part.

At the eastern edge of the Blue Ridge province is a major topographic feature that extends from Alabama to Virginia. This feature, called the Brevard zone, is a narrow belt of multiply deformed rocks that marks the boundary between the Blue Ridge and the Piedmont. The history and the nature of the deformation in the Brevard zone have been the subject of much debate. In fact, COCORP did its first survey of the area in the hope of getting new data so that the debate could be resolved. One of the most important geological findings in the Brevard zone was made by Robert D. Hatcher, Jr., of the University of South Carolina. He interpreted sedimentary rocks of an unusually low metamorphic grade as evidence of sedimentary strata deep under the Brevard zone, with the rocks having been brought to the surface by faulting. This was additional evidence that sedimentary material may underlie the Blue Ridge province at least as far east as the Brevard zone.

Southeast of the Brevard zone is the Inner Piedmont, which consists primarily of high-grade metamorphic rocks intruded by bodies of igneous rock such as the Stone Mountain granite and the Elberton granite. The Piedmont has traditionally been thought to be the metamorphic "core" of the southern Appalachians. Because much of the metamorphic material seems to be sedimentary rocks that were extensively deformed, geologists believed the Piedmont had been vertically uplifted and the deformation had propagated westward into the Blue Ridge province and the Valley and Ridge province. The new seismic data have completely changed this view.

The Inner Piedmont is flanked on the southeast by the Kings Mountain belt, a narrow band of metamorphosed sedimentary rocks and volcanic rocks, some of which may be the remnants of a closed ocean or a geological basin at the continental margin. The major folds of the belt and the folds of the area to the southeast are aligned in a northeasterly trend, similar to the folds of the Inner Piedmont and of the Blue Ridge. The belt was scarred by at least two periods of deformation, one probably about 450 million years ago and the other about 350 million years ago.

The southeastern Piedmont is made up of the Charlotte belt, which consists of metamorphosed sedimentary material, and the Carolina slate belt, which consists primarily of metamorphosed volcanic material. The Carolina slate belt seems to be the remnant of a volcanic arc like the ones that are active today in the western Pacific. Volcanism in the belt probably started in the late Precambrian between 700 million and 650 million years ago and continued into the Cambrian until about 500 million years ago.

The coastal plain southeast of the Piedmont consists of a sequence of young sediments (less than 200 million years old) overlying a crystalline basement. Wells penetrating the basement have unearthed metamorphosed sedimentary rocks like those of the Inner Piedmont and metamorphosed volcanic rocks like those of the Carolina slate belt. Perhaps the metamorphosed sedimentary and volcanic basement of the Piedmont and the Carolina slate belt extends below the coastal plain.

According to the historical picture developed by the theory of plate tectonics, the continents that now border the Atlantic were joined 200 million years ago like the pieces of a jigsaw puzzle to form one huge expanse of land. At that time North America began to separate from Europe, Africa and South America. Since the continents were once connected, geological data from the eastern side of the North Atlantic may help in the interpretation of the structure and formation of the Appalachians. Although most plate-tectonic models of the Carboniferous period before the Atlantic started to open put western Africa adjacent to southern North America, a recent interpretation of paleomagnetic data by Edward Irving of the Department of Energy, Mines and Resources of Canada suggests that it was northern South America that was then adjacent to eastern North America. In any event western Africa and northern South America both have belts of folding and thrusting that were probably created in the same Carboniferous orogeny.

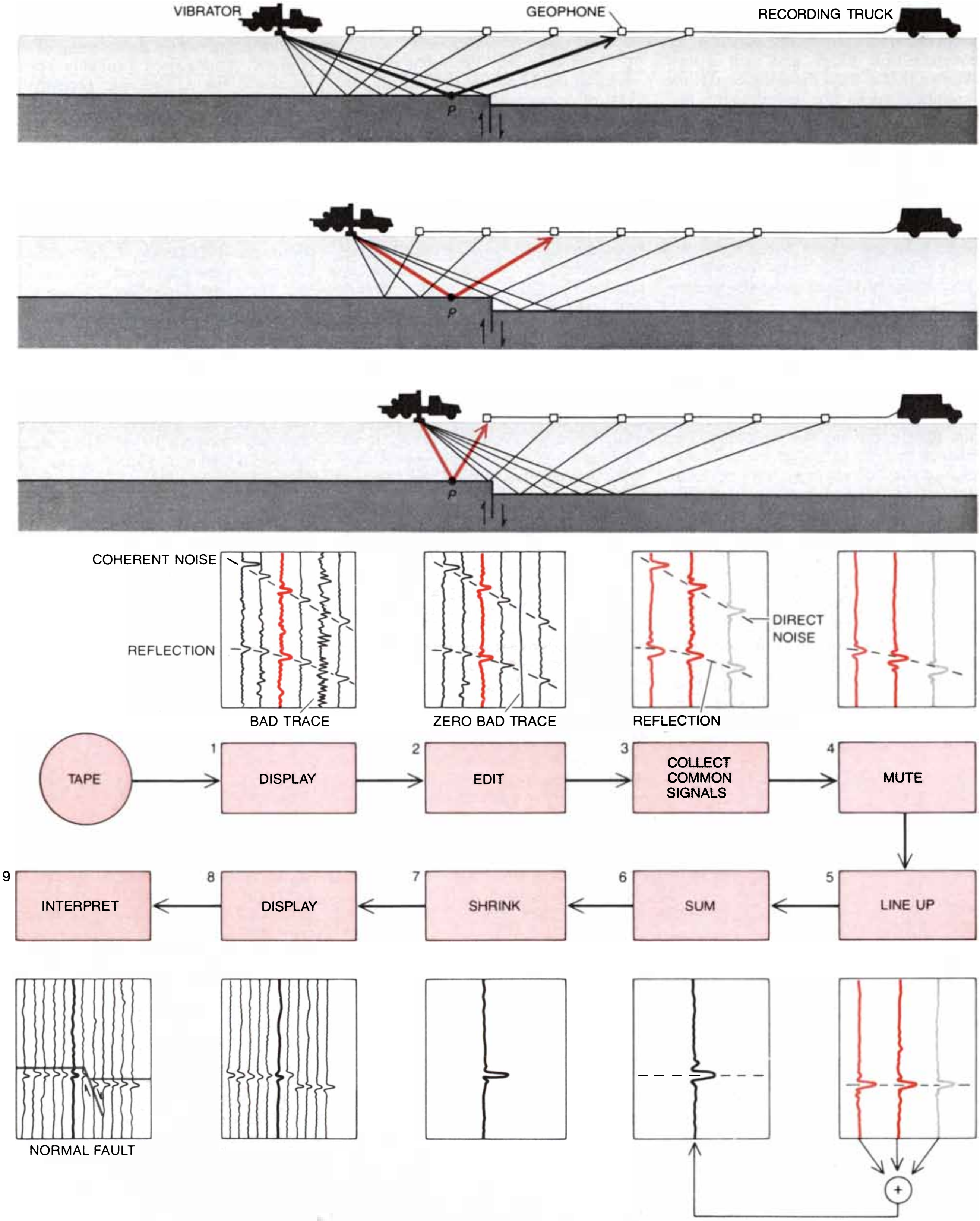
The Mauritanide mountain chain of western Africa is characterized from east to west by a series of belts that are similar in some ways to the Appalachian belts. The eastern Mauritanides are made up of unmetamorphosed sedimentary strata partially covered by metamorphic rocks that have overridden the sediments from the west along thrust faults. To the west are older high-grade metamorphic rocks that resemble those of the southern Appalachian Piedmont. A coastal plain of horizontal younger rocks covers the rest of the orogen. Geological mapping of the area suggests that a mild episode of deformation about 550 million years ago was followed by a period of metamorphism and thrusting before the opening of the Atlantic. This period probably corresponds to the Alleghenian orogeny in the Appalachians. In a broad sense the Mauritanides of western Africa are a mirror image of the Appalachians.

COCORP investigated the Appalachians in Tennessee, North Carolina and Georgia. The initial survey covered an area from about 100 kilometers southwest of Knoxville, Tenn., to about 100 kilometers northwest of Augusta, Ga. The most prominent finding is a southeast-dipping layer of reflections extending from the Valley and Ridge province under the Blue Ridge and the Inner Piedmont between four and 10 kilometers below the surface.

Although several types and configurations of rock could give rise to the reflections, we interpret the reflecting materials as layered sedimentary strata. The interpretation is based on four items of evidence. First, the reflections under the Blue Ridge and the Inner Piedmont can be traced to and correlated with similar units of the Valley and Ridge province in which the rocks are known to be sedimentary. In fact, parts of the seismic-reflection profile of the Blue Ridge and the Piedmont are quite similar to the seismic-reflection profile of the Valley and Ridge. Second, the presence of sedimentary rocks in the windows of the Blue Ridge indicates that the crystalline rocks there overlie sedimentary material. Third, the discovery of unusual carbonate rocks in the Brevard zone suggests that they were scraped from underlying sedimentary layers by activity along the edges of the fault. Last, the COCORP results resemble those from seismic-reflection surveys of the current continental margins, which of course consist of sedimentary material.

H. Clark and his fellow workers at the Virginia Polytechnic Institute have made limited seismic-reflection studies of the area 100 kilometers north of the COCORP-survey area in North Carolina. The studies show that layered, horizontal strata are also present under that area of the Blue Ridge. Therefore the strata probably underlie much of the southern Appalachians. Leonard Harris and Ken Bayer of the U.S. Geological Survey have analyzed recently collected seismic-reflection data from North Carolina that again show layered strata extending from the Valley and Ridge province under the Blue Ridge into the Piedmont.

The only explanation for the buried strata is that the overlying crystalline rocks were emplaced along a major subhorizontal thrust fault (a horizontal fault below the surface). The COCORP data also indicate that the Brevard fault is a thrust that splayed, or broke off, from the major subhorizontal thrust.



NINE-STEP PROCESSING OF DATA is diagrammed in a highly schematic form. Truck-mounted vibrators (*top*) direct into the ground a low-frequency signal that will be partially reflected by discontinuities in the rock. The reflected energy is detected at the surface by an array of geophones, or sound sensors, that are connected to a truck in which the collected data are stored for subsequent processing that will reveal the subsurface geology. After each episode of vibration the truck-mounted vibrators are moved forward a few meters. The thick black and colored lines show the path of a signal reflected by a point *P* for three different positions of a vibrator. The data are processed in nine steps (*bottom*). The recorded signals are displayed (*1*) and noisy data are edited out of them (*2*). Next reflection signals from a common depth point are collected (*3*). The high-amplitude noise is then muted (*4*) and the common signals are lined up (*5*) so that they can be summed in phase (*6*). The summed pulse is shrunk (*7*) and the summed signals are displayed (*8*). Then data are interpreted (*9*).

Splaying may also have given rise to other faults in the area.

The COCORP study traces the horizontally layered reflections to under the Inner Piedmont and the Charlotte belt. Since these reflections are similar to and laterally continuous with the reflections under the Blue Ridge, we think the sedimentary layers are essentially continuous from the Valley and Ridge province to the Charlotte belt. Nevertheless, there is a substantial change in the character of the reflections at a distance of about 250 to 300 kilometers from the northwestern end of the profile. A series of eastward-dipping reflections, similar to those that often characterize the deposits on a continental slope, distinguish this part of the seismic section from the thin, layered reflection band to the west.

At the surface over the dipping reflections is the Elberton granite: a large body of igneous rock intruded into the crust. Under the folded and metamorphosed rocks of the eastern part of the Charlotte belt is another discrete, multilayered and horizontal sequence of reflections. The wedge of dipping reflections under both the Elberton granite and the Charlotte belt, together with horizontal reflections from under the belt, suggests there is a thick layered sequence of rocks between 12 and 18 kilometers below the surface.

Other significant changes in the reflection response of the deep crust are found at the southeastern end of the surveyed area. There anomalous horizontal reflections of between 10.5 and 11 seconds (equivalent to depths of between 30 and 33 kilometers) may correspond to the transition between the crust and the mantle. This transition is known as the Mohorovičić discontinuity. Reflections from the Mohorovičić discontinuity, and many other deep reflections as well, are not observed on the northwestern two-thirds of the profile. Other seismic data suggest that the crust under the Inner Piedmont and the Blue Ridge may deepen to between 40 and 45 kilometers. Over the Mohorovičić discontinuity at the southern end are intriguing westward-dipping reflections, which will be difficult to interpret until the profile is extended to the east. They could represent fault zones, parts of an ancient subduction-zone complex or perhaps even an earlier geological structure unrelated to the Appalachians.

It is still not known what the reflection changes at the transition between the Inner Piedmont and the Charlotte belt represent. One interpretation has the eastward-dipping reflections under the Charlotte belt marking the points where the major thrust fault steepens and plunges deep into the crust. Because this interpretation cannot easily account for the horizontal reflections to the southeast under the Carolina slate belt we prefer a different one. We think the horizontal reflections and the thick sequence of eastward-dipping reflections under the Charlotte belt are caused by sedimentary rocks that were part of a continental margin and continental-shelf edge now buried under the overthrust sheet. These sedimentary rocks are now quite deep, and so they probably have been metamorphosed.

This interpretation can also explain the decrease in depth of the Mohorovičić discontinuity from northwest to southeast along the survey: it is the result of a transition from continental crust on the west to former oceanic crust or thinned continental crust on the east. This suggests that the deformed rocks of the Charlotte belt and the Carolina slate belt, like the deformed rocks of the Blue Ridge and the Inner Piedmont, were thrust over a pile of sediments. According to this interpretation, the thrusting must continue even farther to the east. Seismic studies now in progress should reveal whether or not it actually does.

If the reflections under the Charlotte and Carolina slate belts are indeed from sedimentary or metamorphosed-sedimentary layers, then before thrusting began the highly metamorphosed and deformed rocks of the Blue Ridge must have been a great distance away from their current position with respect to the continental interior. They were probably east of the current position of the Carolina slate belt. The distance of about 260 kilometers from the southeastern edge of the profiled area to the western edge of the Blue Ridge is the minimum lateral distance the rocks of the Blue Ridge were transported.

What does seismic-reflection profiling reveal about the tectonic history of the Appalachians? According to the theory of plate tectonics, the rigid plates of the lithosphere, the solid surface of the earth extending downward about 100 kilometers, float on the asthenosphere: a mobile and moderately fluid layer of the mantle a few hundred kilometers thick. Each plate consists of an upper layer of crust (between 35 and 40 kilometers thick for a continental plate and between five and 10 kilometers thick for an oceanic one) and a lower layer of solid and strong mantle. Oceanic lithosphere is created at the mid-ocean ridges, where magma continually wells up, cools and hardens to form the trailing edge of the moving plate. Such lithosphere is ultimately reabsorbed into the mantle in the subduction zones, the deep oceanic trenches created where two plates collide and one plunges under the other.

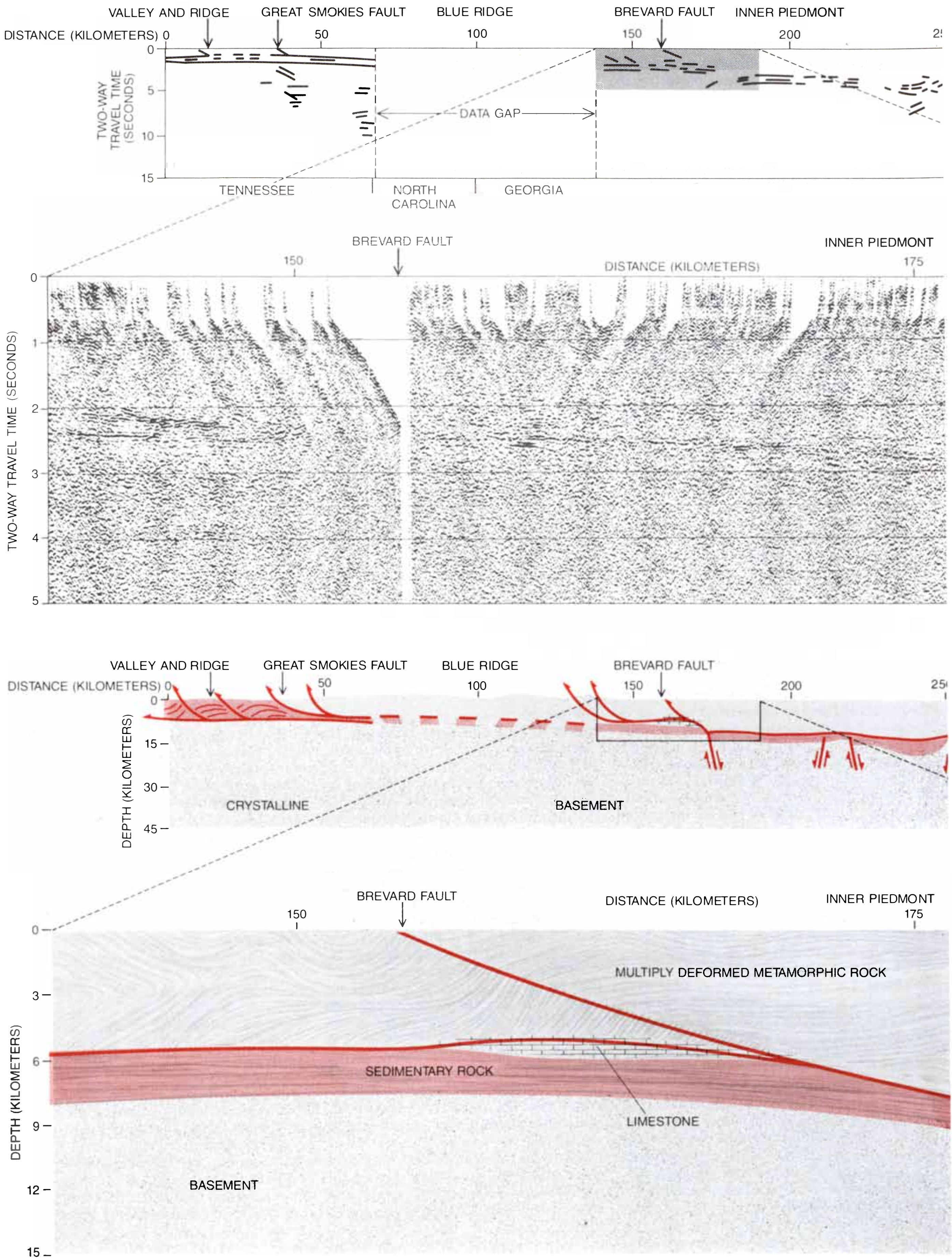
A subduction zone continues to consume lithosphere until a continent or an arc of oceanic islands impinges on it. Because the material of the continental crust is much lighter than that of the mantle most geologists think a continent can scarcely be subducted as a unit. If it indeed cannot, then the collision of a continent or an island arc with a subduction-zone complex will drastically change the nature of the subduction or even arrest it. The deformation of lithosphere at the collision site may give rise to a mountain range.

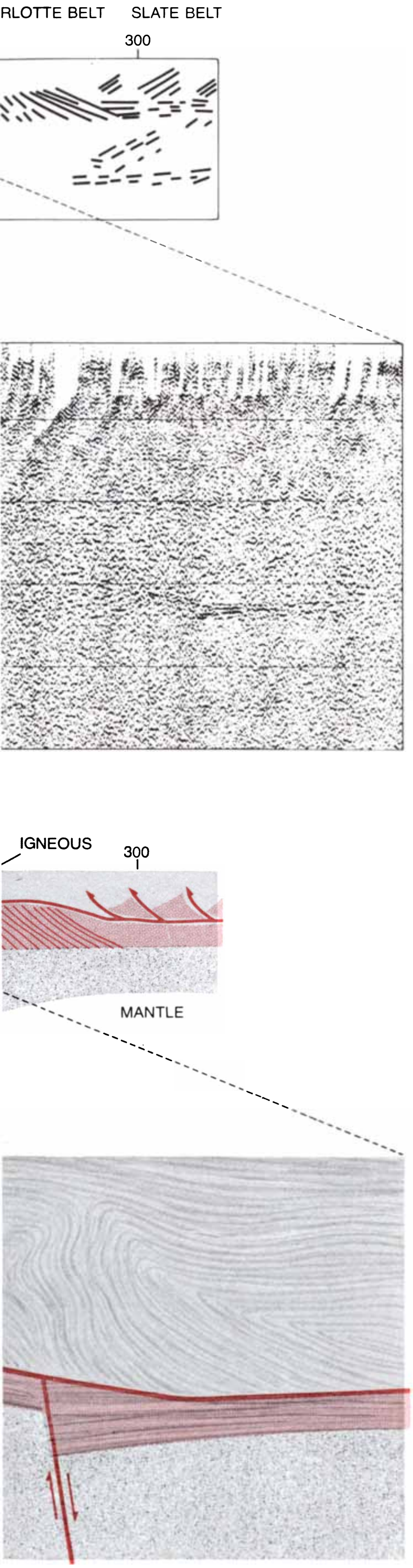
The southern Appalachians have evolved in a series of collisions of fragments of continental or island-arc material at the eastern edge of North America in the Taconic, the Acadian and the Alleghenian orogenies. Several models of the evolutionary process have been developed, and the COCORP data put major new constraints on them. We shall discuss one promising model that was developed by Hatcher to fit the surface geological data. We have modified the model to take into account the COCORP subsurface information. The dates of the orogenic episodes have been determined from the deposition of sediments in the Appalachian basin and from the radioactive dating of metamorphic and igneous rocks. The overall composition of the colliding masses has been inferred from the types of rock in the various provinces and belts.

About 750 million years ago magma rising from the deep interior of the earth split a megacontinental expanse into at least two large continents (Laurentia, or proto–North America, and Gondwana, or proto-Africa) and at least two continental fragments that included the Inner Piedmont–Blue Ridge fragment and the Carolina slate belt fragment. In the period of early rifting the rocks that are now the volcanic and the metamorphosed sedimentary material of the Blue Ridge were deposited in the basin separating proto–North America from the Inner Piedmont–Blue Ridge fragment.

There is no indisputable evidence that the Piedmont and the Carolina slate belt fragments came from the same megacontinent that North America did. Nevertheless, the fact that some rocks in the Blue Ridge and the Piedmont have the same radioactive-dating age (about a billion years) as the basement rocks of eastern North America suggests that these crustal pieces underwent metamorphism at about the same time in the Precambrian. They may therefore have been part of the same continental block. The structure of the Appalachians is the result not of a single crustal block's colliding with North America but more likely of collisions of small continental fragments or island arcs in a way that resembles the numerous collisions occurring today in the southwestern Pacific as Australia moves toward Asia with a group of island arcs and continental fragments trapped between them.

Volcanism started in the island arc of the Carolina slate belt fragment some 650 million years ago. This means that subduction, which gave rise to the volcanic activity, also began at about that time. Some 500 million years ago the basin between proto–North America


VALLEY AND RIDGE
GREAT SMOKIES FAULT
BLUE RIDGE
BREVARD FAULT
INNER PIEDMONT
DISTANCE (KILOMETERS) 0
50
100
150
200
TWO-WAY TRAVEL TIME (SECONDS)
0
5
10
15
DATA GAP
TENNESSEE
NORTH CAROLINA
GEORGIA
BREVARD FAULT
INNER PIEDMONT
150
DISTANCE (KILOMETERS)
175
TWO-WAY TRAVEL TIME (SECONDS)
0
1
2
3
4
5
VALLEY AND RIDGE
GREAT SMOKIES FAULT
BLUE RIDGE
BREVARD FAULT
DISTANCE (KILOMETERS) 0
50
100
150
200
DEPTH (KILOMETERS)
0
15
30
45
CRYSTALLINE
BASEMENT
BREVARD FAULT
DISTANCE (KILOMETERS)
INNER PIEDMONT
150
175
DEPTH (KILOMETERS)
0
3
6
9
12
15
MULTIPLY DEFORMED METAMORPHIC ROCK
SEDIMENTARY ROCK
LIMESTONE
BASEMENT



and the Inner Piedmont–Blue Ridge fragment began to close as a result of subduction. The very existence of sedimentary layers under the Blue Ridge and the Piedmont implies that the subduction zone dipped eastward; westward subduction at that time would probably have destroyed the sediments when volcanism took place. Before 450 million to 500 million years ago the sediments that gave rise to the sandstones and shales now found in the Valley and Ridge province came from proto–North America. For the next 250 million years sediments were derived chiefly from land in the opposite (easterly) direction. This shift in sedimentary history coincides with the first episode of deformation and metamorphism (the Taconic orogeny from 500 million to 450 million years ago) and can be attributed to the result of the closing of the basin between the Inner Piedmont–Blue Ridge fragment and proto–North America and the subsequent collision of these land masses. The sedimentary rocks of the east came from the thrust sheet that had started to move westward onto the continent.

It is not at all clear how a sheet between 10 and 20 kilometers thick was detached from the lower crust of the Inner Piedmont and the Blue Ridge and was then thrust over the continental shelf. Why did it split where it did and what became of the remaining 80 kilometers or so of underlying lithosphere? One hypothesis suggests that as the upper crust became detached the remaining lower crust and upper mantle continued to be subducted eastward and collided with the island arc of the Carolina slate belt. Another similar hypothesis proposes that the polarity of the subduction in the basin reversed, so that a "flake" of the upper crust was pushed toward the continent as the lower crust of the Inner Piedmont and the Blue Ridge was subducted westward. (The flaking hypothesis had previously been adopted by E. R. Oxburgh of the University of Cambridge to explain some features of the Alps.)

The second episode of mountain building was the Acadian orogeny, from 400 million to 350 million years ago. Characterized by extensive metamorphism and deformation, the orogeny was triggered by the closing of the ocean basin between the Inner Piedmont–Blue Ridge fragment and the Carolina slate belt fragment. By that time the Inner Piedmont–Blue Ridge fragment was probably accreted to the proto–North American continent by overthrusting. Today the Kings Mountain belt may be the surface remnant of the ancient collision zone between the Carolina slate belt island arc and the Inner Piedmont. Although the size of the Carolina slate belt fragment is not known, the fragment could well have been quite wide because metamorphic rocks that are similar to the rocks of the Carolina slate belt extend under the coastal plain.

After the Acadian orogeny the next (and last) major compressional event in the southern Appalachians was the Alleghenian orogeny from 300 million to 250 million years ago. This mountain-building episode can be attributed to the collision of proto–North America and proto-Africa (or perhaps South America) to form the supercontinent of Pangaea. Although the Alleghenian orogeny was not as prominent in the northern Appalachians as the other two orogenies were, it gave rise in the southern Appalachians to large-scale overthrusting and extensive igneous activity.

At that time the Brevard zone broke through the surface, transporting carbonate sedimentary rocks from below. Radioactive dating indicates that many of the igneous bodies in the Piedmont were emplaced between 300 million and 250 million years ago. Large-scale overthrusting in this period has also been mapped in western Africa, although the western limit of the thrust faults has not yet been determined. We speculate that a segment of the African (or South American) continental shelf underthrust the eastern margin of the Carolina slate belt fragment, resulting in a fold-and-thrust belt that went in the opposite direction.

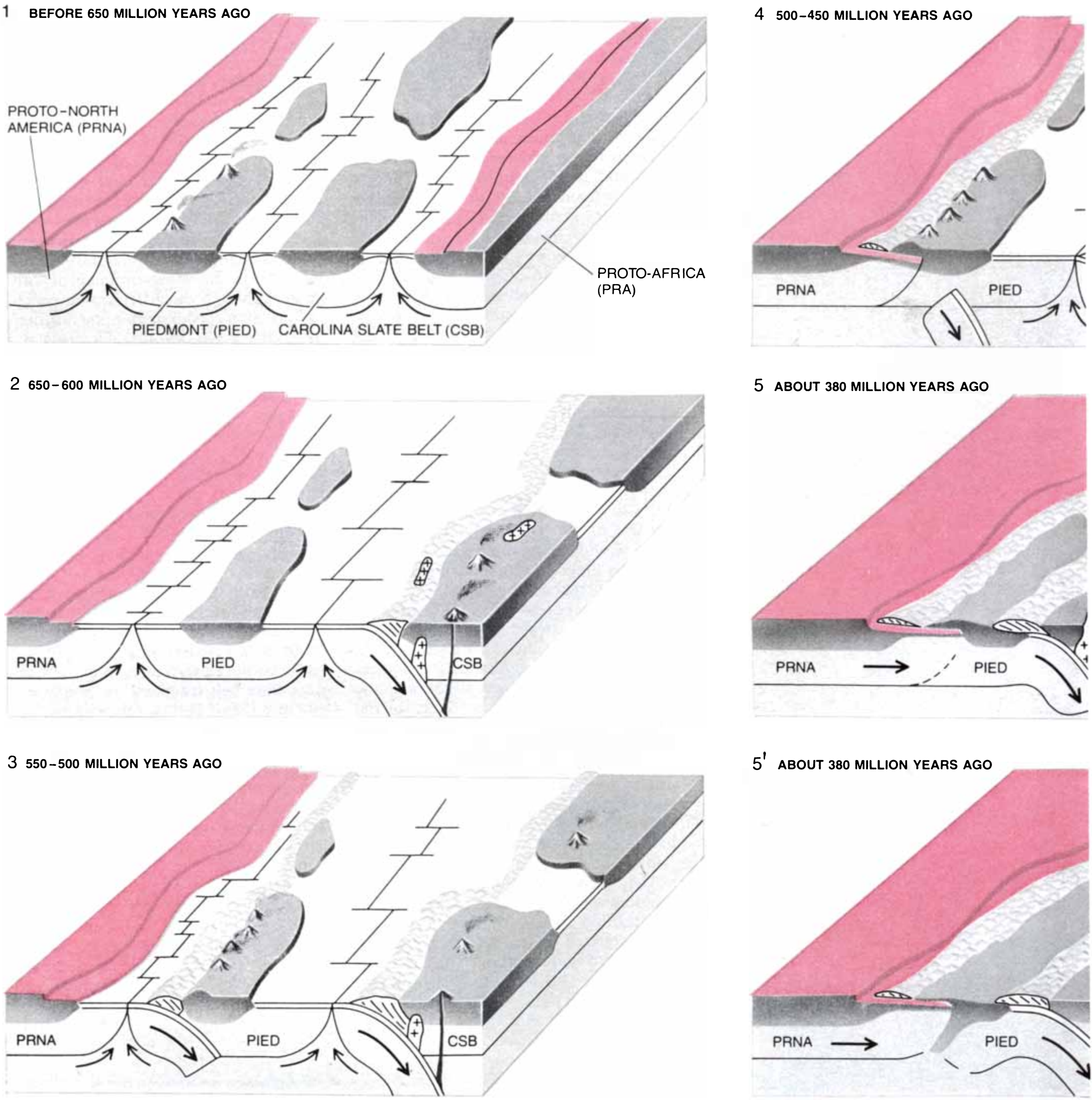
After the Alleghenian orogeny extensional tectonism again started to break a megacontinent (Pangaea) into smaller continents between 250 million and 200 million years ago. As the continents drifted apart the Atlantic Ocean was left in their wake. The Triassic basins of the eastern coast of the U.S., such as the grabens (troughs with near-vertical sides) in New Jersey and Connecticut, formed at that time. As the Atlantic grew, the current continental shelf was built up off the eastern coast of North America (and off the western coast of

SEISMIC-REFLECTION PROFILE of the southern Appalachians is shown in a highly schematic form at the top. The vertical axis indicates the two-way travel times of the reflected waves. The horizontal axis locates the major geological features and indicates the distance along the surface from the northwestern end of the beginning of the COCORP-survey area. The state boundaries are also marked. Below the schematic profile is a photograph of the actual profile across the Brevard fault, a narrow belt of multiply deformed rocks that marks the boundary between the Blue Ridge province and the Piedmont province. Next a possible geological interpretation of the profiles is shown. Crystalline rocks of chiefly continental origin are gray. Folded metamorphic rocks near the surface are wavy gray lines. Sedimentary strata are a light color. The inferred faults are lines with arrowheads that show the relative motion.

Africa and the northern coast of South America).

Is it possible that large-scale thin thrust sheets are a general result of continental collisions? There are several indications that they are. We have discussed the fold-and-thrust belts of western Africa and northern South America, which have a geological structure quite similar to that of the Appalachian belts. David Gee of the Geological Survey of Sweden has mapped major horizontal thrust faults in the Caledonide mountains of Scandinavia, some of which seem to have been displaced by hundreds of kilometers.

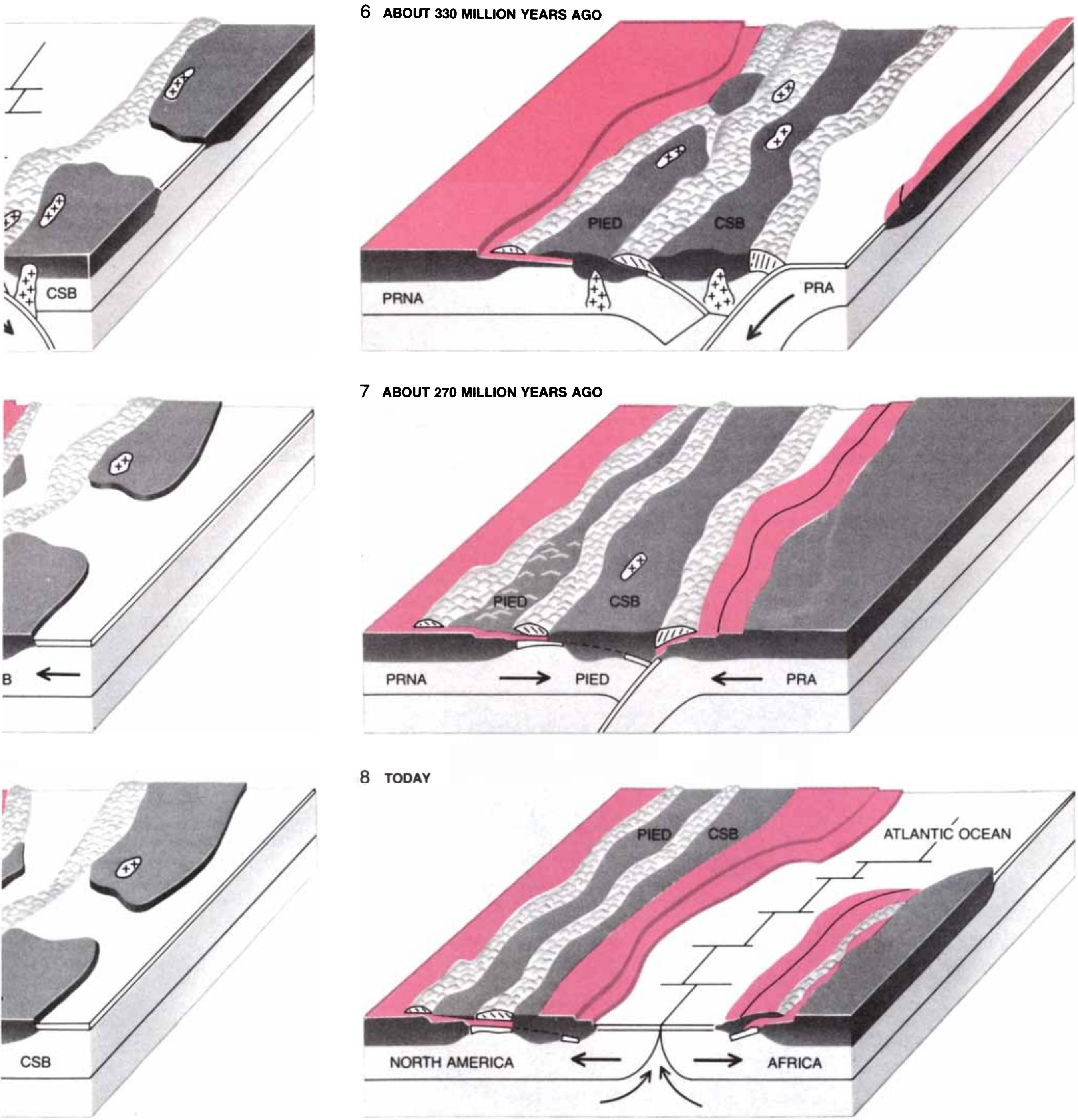
The kind of thin-skinned thrusting found in the Valley and Ridge province can also be seen in the fold-and-thrust belts of the Montana and Alberta cordillera, which is part of the Rockies. This area is a future site for COCORP studies that will try to determine the western limit of the thrusting. Thin-



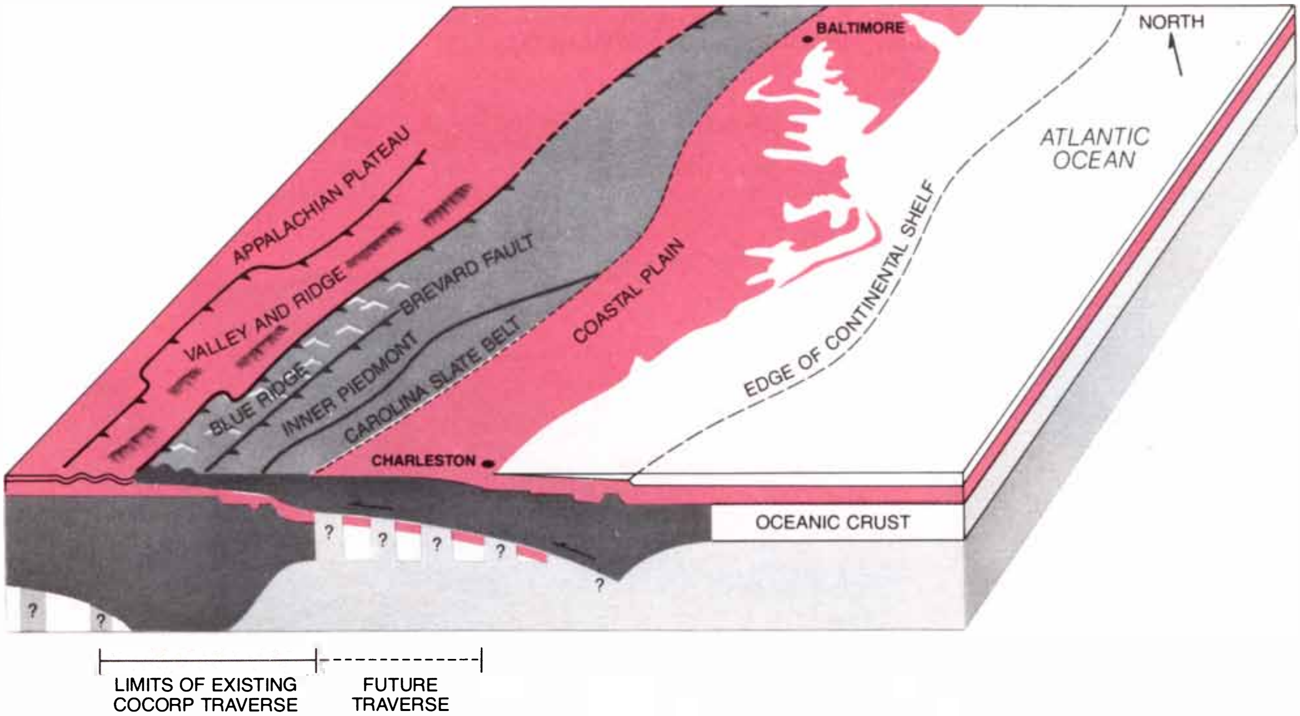
TECTONIC HISTORY of the southern Appalachians is shown in a series of nine block diagrams. Continental crust is dark gray, oceanic crust is white and the sedimentary rocks of the continental shelf are a light color. The solid surface of the earth extending downward about 100 kilometers is a set of interacting rigid plates that can be absorbed into the mantle at subduction zones: deep oceanic trenches where two plates collide and one plunges under the other. The collision of a continent or an arc of oceanic islands with a subduction-zone complex causes the surface rock to buckle. The buckling gave rise to three orogenies, or episodes of mountain building, in the southern Appalachians. Some 650 million years ago in the Precambrian (*1*) a mega-continental expanse split into two continental plates (Laurentia, or proto–North America, and Gondwana, or proto-Africa and proto–South America) and several smaller fragments including the Piedmont and the Carolina slate belt. Subduction gave rise to volcanism in the Carolina slate belt, probably about 625 million years ago (*2*).

skinned thrusting may also have been dominant in other mountain ranges, including the Alps, the Himalayas and the Zagros of the Middle East. Some geologists believe the kind of overthrusting evident in the Appalachians is even now taking place in the Timor region of the southwestern Pacific, where the Australian continental shelf has underthrust the Timor island arc to the northwest for perhaps 150 kilometers.

The entrapment of sedimentary rocks under large metamorphic sheets has major consequences not only for the growth of continents but also for the recovery of gas and oil. The discovery of sedimentary material under the metamorphic rocks of the Blue Ridge and Piedmont provinces calls for reevaluating the possibility of extracting hydrocarbons from that part of the Appalachians. Perhaps the metamorphism of the sediments has not been intense



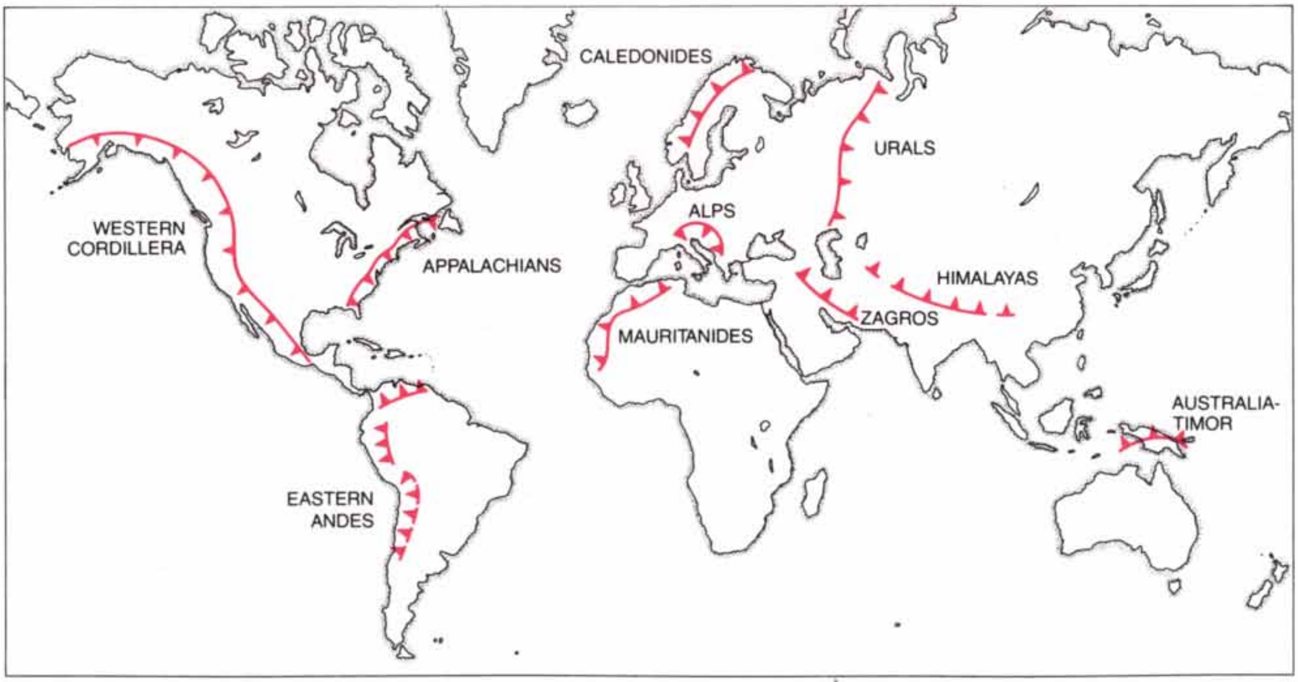
The volcanic activity lasted until about 500 million years ago (*4*). The closing of the basin between proto–North America and the Piedmont (*3, 4*) may have been responsible for the Taconic orogeny from 500 million to 450 million years ago. The thin-skinned thrusting that began at that time could have been caused by two possible processes. The first process (*4, 5*) involves the continued eastward subduction of the bottom part of the Piedmont as the top 10 or 15 kilometers is peeled off. The other process (*5′*) involves the westward subduction of the bottom part of the Inner Piedmont and the Blue Ridge as the upper part starts to form the thrust sheet. The closing of the ocean basin between the Inner Piedmont and the Carolina slate belt (*5, 6*) resulted in the Acadian orogeny from about 400 million to 350 million years ago and the continued thrusting of the Piedmont and the North American continent. The last orogeny was the Alleghenian, which put proto-Africa against North America (*6, 7*). The opening of the Atlantic about 200 million years ago pushed the continents apart (*8*).



SUBSURFACE GEOLOGY OF THE EASTERN U.S. was inferred from the seismic-reflection profiles of the existing COCORP survey. A planned extension of the surveyed area is marked. The sedimentary strata are colored, the crystalline rocks of chiefly continental origin are dark gray and the basaltic oceanic crust is white. Between 500 million and 250 million years ago the continental shelf was overthrust by a thin near-horizontal sheet of crystalline rock. A new continental shelf has been forming off the Atlantic coast for 200 million years.

enough to remove all the hydrocarbon deposits from the rocks under the thrust sheet.

The discovery of the layer of sediments under the core of the Appalachians has implications that go far beyond this particular mountain range. The mechanism of thin-skinned thrusting involving basement rocks may have been responsible for the formation of mountain belts for as long as plate-tectonic processes have been operating. In that case the evolution of continents is characterized not by lateral accretion along vertical boundaries but by the juggling and stacking of thrust sheets. Much work remains to be done, but we suspect that for at least half of the earth's history such thrusting could well have been a chief mechanism in the evolution and growth of continents.



MOUNTAIN RANGES AROUND THE WORLD (the western Cordillera, the eastern Andes, the Mauritanides, the Alps, the Urals, the Caledonides, the Zagros, the Himalayas and Australia-Timor) may consist of thin thrust sheets as the southern Appalachians do.

Buy now and get FREE Solid State Software™ Libraries.

Choose from these.

TI Programmables lead the field in performance, quality and value. When you're choosing a programmable calculator consider: Power. Total software support. Flexibility. And price/performance. You'll discover a clear-cut answer. A TI Programmable 58C or 59.

The TI-59 has up to 960 program steps or up to 100 memories. Magnetic card read/write capability lets you record your own custom programs or those received from PPX (Professional Program Exchange).

The TI-58C features up to 480 program steps or 60 memories. And it has TI's Constant Memory™ feature that retains data and program information even when the calculator is turned off.

And now free modules give you that added productivity you need. From August 15 to October 31, 1980 is your special opportunity to purchase one of the world's most advanced programmable calculators. And get a minimum of $40 worth of free software modules with a TI-58C. Or, when you buy a TI-59, get a minimum of $80 worth of software modules and an $18 one-year membership in PPX. This will allow you to select up to 3 programs (from over 2,500) written by professionals in your field.

Visit your TI retailer for more information, and let him help you select the TI Programmable and free software that's right for you.

☐ I've bought a TI-58C, send me my free module. Here is my first choice and an alternate.
☐ I've bought a TI-59, send me my two free modules and my membership (which entitles me to select three programs from the source catalog at no charge). Here are my module choices and an alternate
1. ________ 2. ________
3. ________

Send to: TI Library Offer, P.O. Box 1984, Lubbock, TX 79408.
Return this coupon: (1) with customer information card (packed in box), (2) a dated copy of proof of purchase, between Aug 15 and Oct 31, 1980 — items must be postmarked by Nov 7, 1980.

Name ________
Address ________
City ________ State ________ Zip ________ SA
Calculator Serial Number (from back of unit) ________
Please allow 30 days for delivery. Offer void where prohibited. Offer good in U.S. only.
TI reserves the right to substitute modules.

‡U.S. suggested retail for all Libraries is $40, except Farming, $55, and Pool Water Analysis, $45.
**For use with TI-59 only

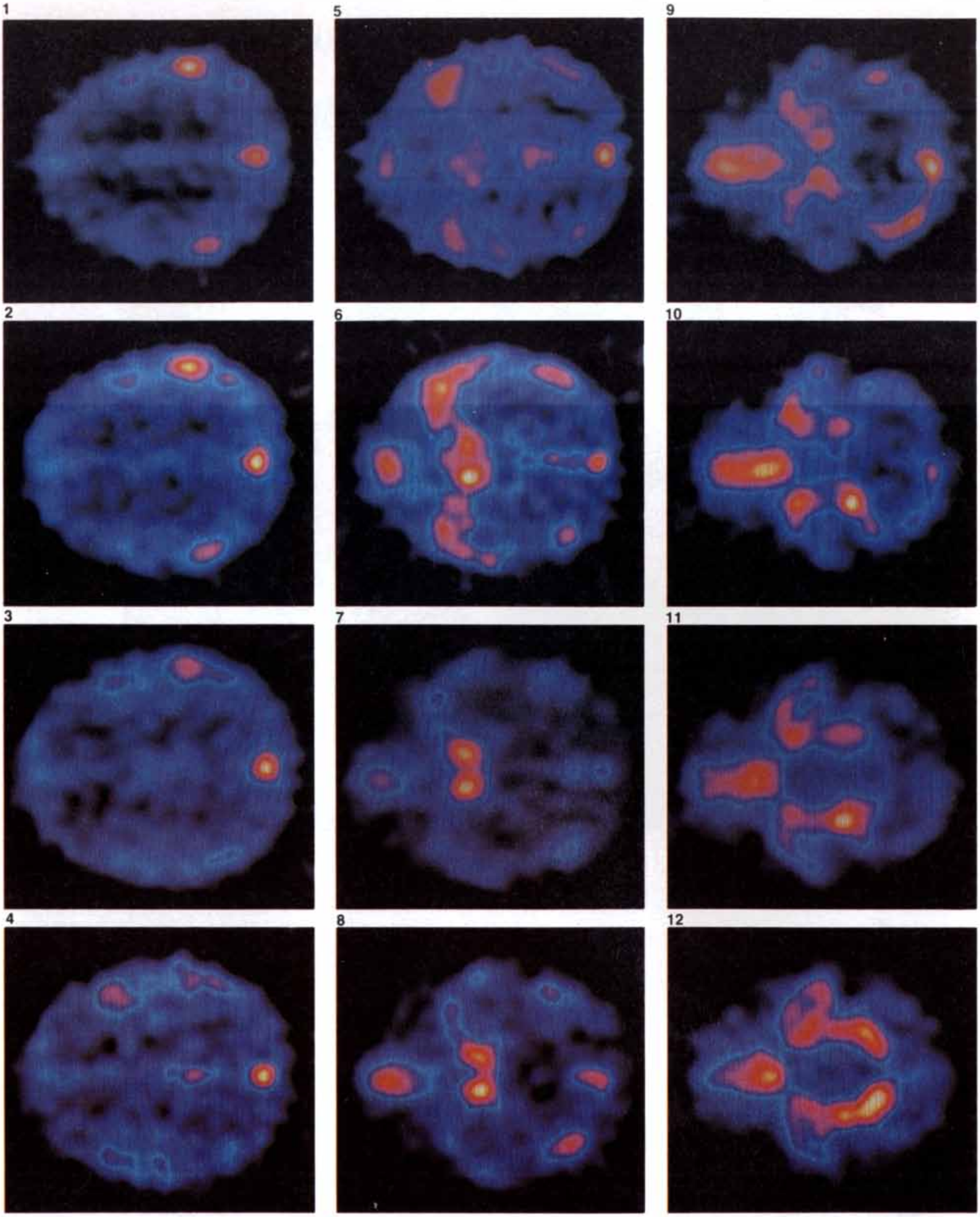
Fifty Years of Innovation

Texas Instruments technology — bringing affordable electronics to your fingertips.

TEXAS INSTRUMENTS
INCORPORATED

*US suggested retail price.

45735



CEREBRAL BLOOD VOLUME in a normal man is shown in images that have been made at 12 different levels in the head by positron-emission tomography (PET). The drawing on the opposite page identifies the levels corresponding to the numbered PET images. The multicolored regions in the images disclose the distribution in the head and the brain of the radioactive tracer carbon 11 (^{11}C), which has a half-life of only 20 minutes. The subject inhales the tracer in the form of ^{11}C–carbon monoxide, which attaches itself to the hemoglobin in the red blood cells and is transported rapidly throughout the body. When an atom of carbon 11 decays, it releases a high-energy positron, or positive electron. Almost immediately the positron interacts with an electron, and the two particles annihilate each other with the release of two gamma rays, which fly off in nearly opposite directions. The PET images are reconstructed from measurements of the radiation. In these images the regions of high blood volume appear in shades of red, yellow and white and delineate the large venous collecting system that courses over the surface of the brain. The shades of blue generally depict the actual blood volume within tissues of the brain. These PET images and the others in this article were made by the authors at the Washington University School of Medicine.

Positron-Emission Tomography

This technique displays the concentration of radioactively labeled substances in the living body with improved clarity. It holds much promise for both biological investigations and clinical diagnosis

by Michel M. Ter-Pogossian, Marcus E. Raichle and Burton E. Sobel

A central premise of medicine is that all biological activity is the result of biochemical reactions and that for every pathology there is an underlying biochemical defect. A continuing goal of medicine is therefore to identify the abnormal biochemical activity associated with a given pathology and to observe the abnormality directly and as early as possible in the afflicted organism. One useful technique, developed with the availability of artificial radioactive isotopes, is to administer a radioactively labeled substance to a patient and to follow its fate within the body by means of an instrument that detects the decay of the isotope. The conventional imaging devices of nuclear medicine provide somewhat distorted images of the distribution of a radioactive isotope because their field of view varies as a function of depth and also because they fail to distinguish between the region of interest and radioactive regions in front of it or behind it. Essentially such devices compress three dimensions into two. Nevertheless, for a number of purposes the images of conventional nuclear medicine are quite suitable for clinical applications.

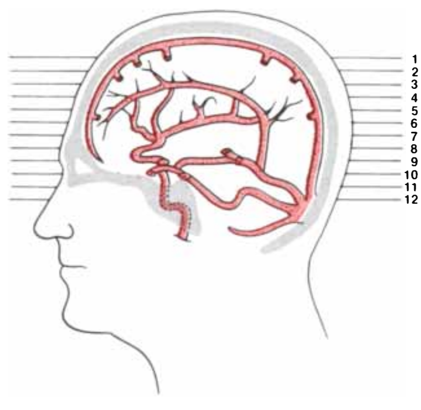
Within the past few years a more precise localization of the emitted radiation has been made possible by the development of a technique analogous to that of computerized axial tomography (CAT), which constructs an image by rotating a source of X rays completely around a human subject. A detector on the other side of the subject from the X-ray source rotates synchronously with the source to record variations in the absorption of the radiation as it passes through the intervening tissues. A computer reconstructs the image in the form of an axial cross section. It is evident that by the time a pathological condition shows up either in a conventional X-ray image or in a CAT scan the underlying biochemical abnormality is usually well advanced.

The technique that combines early biochemical assessment of pathology achieved by nuclear medicine with the precise localization achieved by computerized image reconstruction is positron-emission tomography (PET). In this technique a chemical compound with the desired biological activity is labeled with a radioactive isotope that decays by emitting a positron, or positive electron. The emitted positron almost immediately combines with an electron, and the two are mutually annihilated with the emission of two gamma rays. The two gamma rays fly off in very nearly opposite directions, penetrate the surrounding tissue and are recorded outside the subject by a circular array of detectors. A mathematical algorithm applied by computer rapidly reconstructs the spatial distribution of the radioactivity within the subject for a selected plane and displays the resulting image on a cathode-ray screen. The images are recorded at intervals after the administration of the labeled compound and can be color-coded to show differences in the level of activity from point to point. With suitable interpretation PET images can provide a noninvasive, regional assessment of many biochemical processes that are essential to the functioning of the organ that is being visualized.

So far PET has been extensively used with experimental animals, primarily dogs and monkeys, and to a more limited extent with human subjects. It has already thrown new light on many physiological processes that cannot be studied as effectively in any other way. Only a few years ago PET was still regarded as an exotic and costly technique, limited to a few laboratories. Today it is on the threshold of becoming a tool of fundamental importance in diagnostic medicine, with implementation in some 40 major medical centers throughout the world.

In addition to providing precise localization of the labeled substance administered to the subject PET has a second important advantage over the conventional techniques of nuclear medicine. PET offers a more useful choice of biologically significant chemical elements for labeling. Biological systems consist mainly of compounds of carbon, nitrogen, oxygen and hydrogen. The first three elements have short-lived radioactive isotopes that decay through the emission of positrons and therefore are suitable for positron-emission tomography. For PET studies three isotopes of particular value are oxygen 15, nitrogen 13 and carbon 11, with half-lives respectively of roughly two, 10 and 20 minutes. Although there is no positron-emitting isotope of hydrogen (tritium, or hydrogen 3, emits only low-energy electrons), water can be labeled for PET studies with oxygen 15. Before the development of PET carbon 11, oxygen 15 and nitrogen 13 had seen limited application to biochemical studies in conjunction with conventional detectors to record the gamma rays created by the annihilation of positrons.

The development of PET was delayed in large measure by the short half-lives of the positron-emitting isotopes of the elements most useful for labeling. Since their existence is so fleeting, carbon 11, oxygen 15 and nitrogen 13 must be created close to the point of detection, and fast chemical techniques must be devised to introduce the labels into useful compounds. Finally, PET studies with



PLANES OF THE PET IMAGES that appear on the opposite page are identified in this cross section. The principal venous structures of the brain, which show up prominently in the PET images, are depicted here in color.

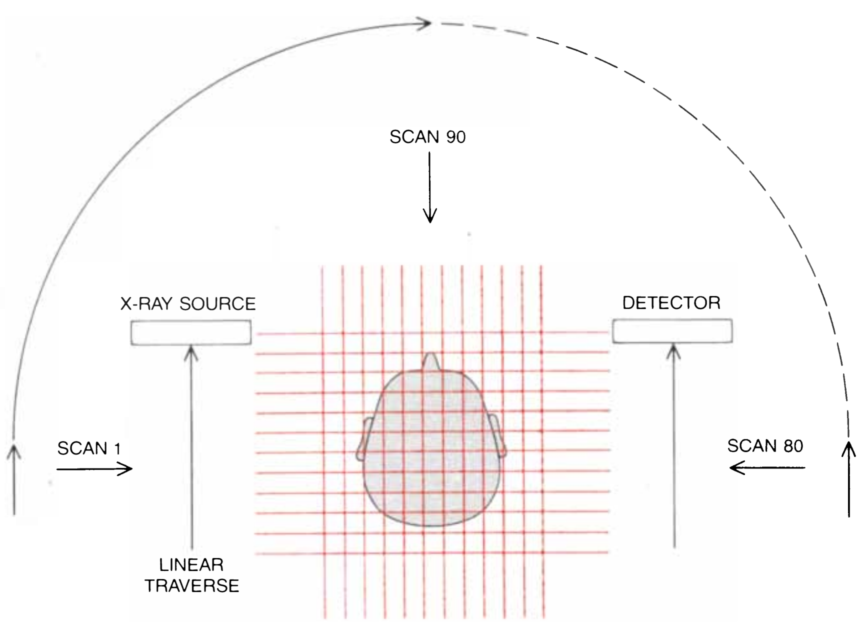
the labeled compounds are limited to biochemical processes with rapid rates of turnover.

Within the past 20 years the difficulties associated with producing and handling such short-lived isotopes have been largely resolved. Several centers have now been equipped with cyclotrons expressly designed for creating a variety of positron-emitting isotopes. Cyclotrons for this purpose typically accelerate protons (hydrogen nuclei) up to an energy of 15 million electron volts (MeV) and deuterons (heavy-hydrogen nuclei) up to half that energy. Such an instrument weighs some 20 tons and is shielded by several feet of concrete.

The target materials subjected to bombardment in the cyclotron vary depending on the radioactive isotope desired and on the chemical form in which it emerges from the target. Often the target is a gas to facilitate transporting the active material from the cyclotron to the area where it is to be used. Even when the target is a solid the newly formed isotope is frequently released as a gas. Carbon 11 is usually prepared by irradiating nitrogen or boron with protons, oxygen 15 by irradiating nitrogen with deuterons and nitrogen 13 either by irradiating oxygen with protons or carbon with deuterons. The reactions are selected so that the radioactive isotopes are synthesized relatively free of contamination by stable isotopes of the same element. Such "carrier-free" isotopes are highly desirable for many labeling procedures.

In some instances the radioactive isotope can be used directly in its elemental form. Thus oxygen 15 serves for metabolic studies and nitrogen 13 for studies of lung ventilation. A number of simple compounds are easily labeled: water with oxygen 15, carbon monoxide or carbon dioxide with either carbon 11 or oxygen 15 and ammonia with nitrogen 13. In other instances the labeling of a complex compound may stump the most imaginative chemist for months or years before he can devise a labeling procedure fast enough to be compatible with the half-life of the isotope. For example, the labeling of glucose with carbon 11 proved to be notably difficult even though labeling with carbon 14 (which has a half-life of 5,730 years) was well established. Illuminated leaves of the Swiss chard plant are exposed to carbon dioxide labeled with carbon 11. Through photosynthesis the carbon dioxide is incorporated into the sugar glucose, which is then extracted in expeditious chemical steps and purified by chromatography.

Difficulties are sometimes encountered even after a fast labeling process has been devised. For example, the labeled molecule may not be readily soluble in body fluids. This problem arose with the salt of palmitic acid, a fatty acid used in studying the integrity of the heart muscle. The salt was made soluble by binding it to the protein serum albumin. Sometimes the problem lies in achieving an extremely high specific activity. For example, compounds suitable for studying nerve-cell receptors must have a specific activity of thousands of curies per millimole. This means that most or all of the molecules in a given sample must be radioactively



COMPUTERIZED AXIAL TOMOGRAPHY, or CAT, is based on programs similar to those used by PET. In a CAT scan the attenuation of X rays by tissue along many axes is recorded as a radiation source is rotated around the subject. In a typical procedure the X-ray source makes a linear traverse at each of 180 positions one degree apart. Collected measurements are sufficient for a computer to reconstruct the relative opacity of tissue in a two-dimensional section.

labeled. In spite of various difficulties methods have now been worked out for labeling several hundred species of molecules with positron-emitting isotopes, exceeding the number used in conventional nuclear medicine.

An isotope of particular usefulness in PET studies is fluorine 18, which has a half-life of 110 minutes. Although it has the obvious potential of clarifying the role of the trace element fluorine in animal physiology, fluorine 18 has an additional value: it can be incorporated into molecules that are analogues of natural substrates of metabolism. Within the organism the analogue follows the metabolic pathway of the normal molecule up to a certain point, at which the reaction sequence comes to an abrupt halt because the analogue molecule is no longer accepted by the cell's biochemical machinery. A glucose analogue in which fluorine 18 is substituted for one of the oxygen atoms has proved useful in studying the metabolism of glucose.

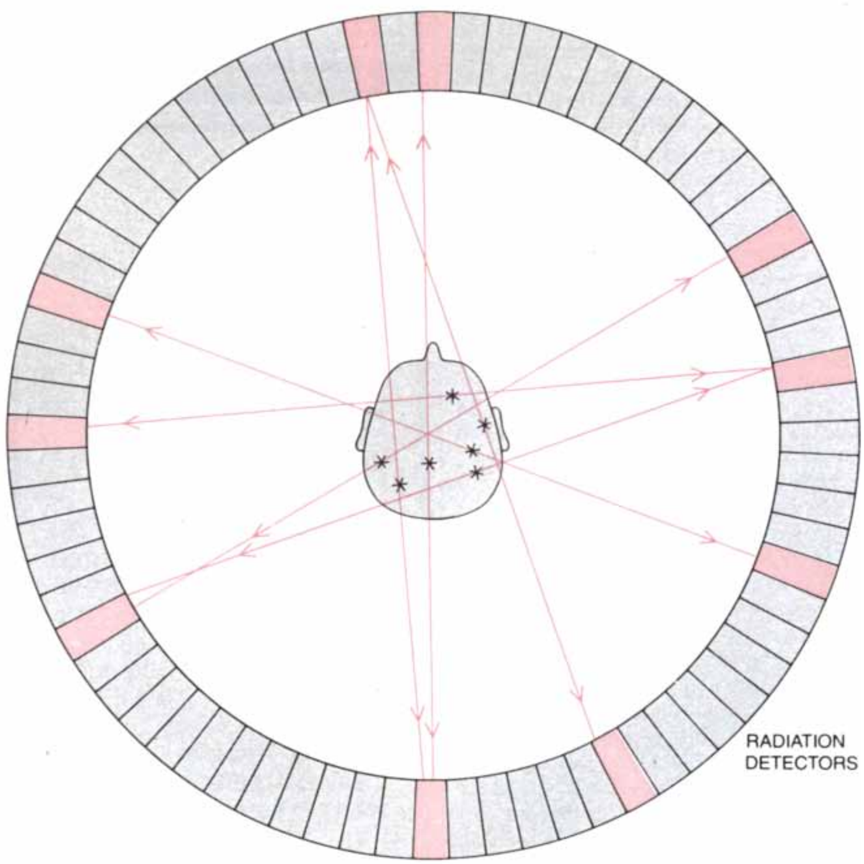
The image-forming process in positron-emission tomography has three principal steps: the detection of the gamma rays emitted in the positron-annihilation process, the identification of the radiation's direction of travel and the reconstruction of the distribution of radiation into an accurate geometric image. A tomographic image can be reconstructed with the aid of isotopes that emit gamma rays directly on decay, but the number of suitable isotopes is limited. Moreover, the reconstruction is simplified and the image more faithfully represents the distribution of the labeled compound if the isotope decays through positron emission. The emitted positron loses most of its kinetic energy after traveling only a few millimeters in living tissue. It is then highly susceptible to interaction with an electron, an event that annihilates both particles. The mass of the two particles is converted into 1.02 million electron volts of energy, divided equally between two gamma rays, or high-energy photons.

Since the two gamma rays are emitted simultaneously and travel in almost exactly opposite directions, their source can be established with high accuracy. This is achieved by linking up the radiation detectors in pairs, with each member of a pair 180 degrees away from the other; the two detectors register a signal only if both sense high-energy photons coincidentally. In this way a circumferential array of detectors can establish the sources of all coincident pairs of gamma rays that originate within a volume defined by straight lines joining paired detectors. Annihilation events that occur outside the defined volume are not detected. The spatial resolution that can be achieved depends on the aperture of the detectors. The resolution is slightly degraded at the site of radioactive decay, however, by the range of positrons in matter and by the fact that the paired annihilation photons are not emitted precisely 180 degrees from each other.

The events recorded by the coincidence detectors are the raw data from which the PET images are reconstructed. The principle of image reconstruction by computerized axial tomography is that an object can be accurately reproduced from a set of its projections taken at different angles. The faithfulness of such a reconstruction is proportional to the number of projections. Given an infinite set of projections an object can be reconstructed uniquely. In a typical PET system between 100 and 300 projections will yield a spatial resolution of a few millimeters.

By measuring coincident photons the detector array in a PET system identifies a series of straight lines, all in a single plane. This information is sorted into a series of closely spaced parallel lines that represent the radioactivity in the subject viewed at different angles. The groups of parallel lines, recorded as profiles, are modified by the mathematical process called convolution, which endows the profiles with negative values designed to remove artifactual structures in the image. The convolved profiles are "back-projected" to yield the final image of the object. Other algorithms that do not rely on convolution have been successfully applied to the same end, but the convolution, or "filtered back-projection," technique is the one now most widely used.

A complete PET unit consists of a data-acquisition system and a computer. The acquisition system consists of the radiation detectors, their associated circuitry and in most designs a mechanical system that imparts a small motion to the detectors in order to get better sampling. In some alternative designs the number of detectors is large enough for movement to be unnecessary. Data from the acquisition system are fed into a fast computer, usually one with a large memory. The computer's peripheral equipment includes magnetic disks and tapes, a line printer (which supplies numerical values of the distribution of radioactivity in the image), a display system for immediate viewing and recording of the image, and often some interactive capabilities for the analysis of the image. Some PET systems incorporate electronic circuitry capable of carrying out parts of the reconstruction



DISTINGUISHING FEATURE OF A PET SCAN is that the penetrating radiation originates within the subject rather than from the outside. The source of radiation is carried to the tissue of interest by a biologically active compound that has been labeled with an isotope that decays by emitting a positron. The positron and an electron annihilate each other within a few millimeters of the decay site. The detectors are arranged so that only simultaneous events 180 degrees apart are recorded. A computer program reconstructs the distribution of the decaying isotopes.

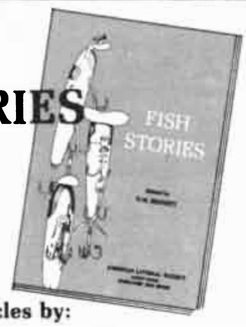
process that other systems leave entirely to the computer.

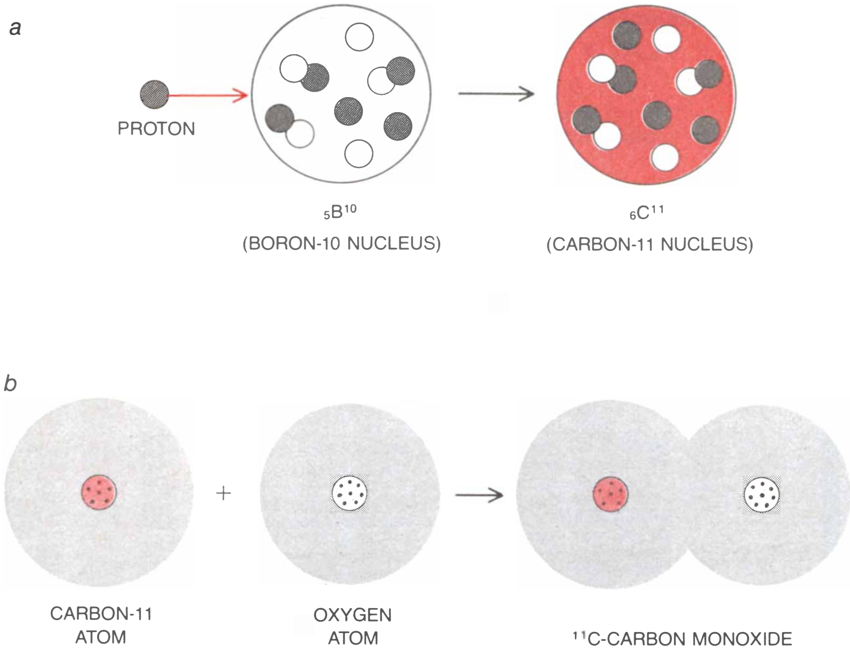
The latest PET systems can record up to seven tomographic images of the human head or torso simultaneously. They achieve a resolution of about a centimeter and assess the distribution of radioactivity within a few percent. Successive images can be obtained at intervals of less than a minute. The dose of radioactive material given to the subject produces a radiation exposure comparable to typical exposures in other kinds of diagnostic nuclear medicine. Several companies already offer PET systems at prices ranging from $600,000 to more than $1 million.

Let us turn now to the kinds of physiological study that can be conducted with positron-emission tomography. One of the first measurements accomplished with PET was the determination of regional blood volume in transverse sections of the human brain. The most satisfactory results are obtained by having the subject inhale air that contains trace amounts of carbon monoxide labeled with carbon 11. The ^{11}C–carbon monoxide avidly binds with hemoglobin to form ^{11}C-carboxyhemoglobin, thereby effectively labeling the entire blood pool of the body. The regional blood volume in the brain is readily determined by comparing the amount of ^{11}C-carboxyhemoglobin observed in tomographic images of the brain with the amount in the total blood pool as determined from samples of venous blood taken concurrently with the emission scan. Measurements of the regional blood volume in the brain by PET not only reveal the expected regional differences resulting from the fact that the blood vessels in gray matter of the cerebral cortex are more densely packed than those in the underlying white matter but also delineate the major blood vessels, primarily veins, surrounding the brain.

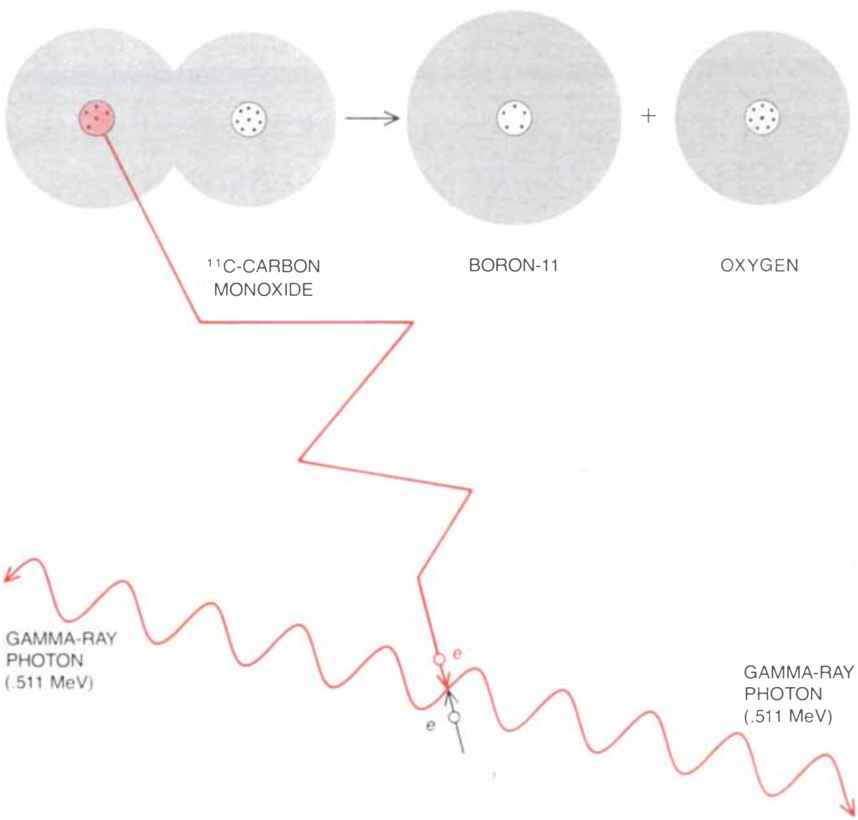
Although there are several other well-proved techniques for measuring cerebral blood volume, the success of the PET method was important for various reasons. First, it showed that the regional perfusion of an organ such as the brain can be assessed by a simple and accurate method. Second, the PET technique for measuring cerebral blood volume can be extended to other organs and tissues by labeling things other than carbon monoxide. For example, by labeling blood platelets it is now possible for the first time to study the process of atherosclerosis in the blood vessels of a living animal. Such information should assist greatly in the diagnosis and treatment of human patients with various forms of vascular disease, including the arterial blockages that cause stroke and heart attack.

Third, the PET technique of measuring blood volume can be adapted to de-





POSITRON-EMITTING ISOTOPES are created by bombarding nonradioactive elements with particles such as protons, deuterons or helium nuclei. For example, carbon 11 is produced when boron 10, whose nucleus consists of five neutrons and five protons, captures another proton (*a*). In many studies carbon 11 is combined with oxygen to form ^{11}C–carbon monoxide (*b*).



FATE OF A LABELED COMPOUND in an organism is revealed when the radioactive isotope incorporated as a tracer decays. For example, ^{11}C–carbon monoxide acts as a tracer for hemoglobin by forming ^{11}C-carboxyhemoglobin. When a carbon-11 nucleus decays, it releases a positron and becomes boron 11, whose nucleus consists of five protons and six neutrons. As a result the carbon monoxide is dissociated into an atom of boron and one of oxygen. When the positron (e^+) encounters an electron (e^-), each of the two gamma rays resulting from their annihilation carries off a little more than half a million electron volts (MeV) of energy.

termine the concentration of virtually any suitably labeled compound, including drugs and pharmaceuticals, in the vascular compartment of a target organ compared with its concentration in the extravascular compartment. The ability to make this distinction is unique to PET and is of much importance in the measurement of metabolism, of the chemical composition of tissue and of blood flow.

The measurement of regional metabolism is somewhat more complex than the measurement of regional blood volume. Quantitative studies of brain metabolism have now been conducted with oxygen labeled with oxygen 15, with glucose labeled with carbon 11 and with the fluorine-18 analogue of glucose (^{18}F-2-deoxy-D-glucose). The metabolism of the heart muscle has also been studied with palmitate, the salt of palmitic acid and the primary physiological substrate of the heart, labeled with carbon 11.

Several promising techniques have recently been developed for noninvasive diagnosis in cardiology. They include the assessment of the dimensions of the chambers of the heart and the motion of the valves with ultrasound (echocardiography), the evaluation of the pumping function of the heart with radioactive isotopes administered intravenously (radioventriculography) and the definition of heart anatomy with CAT scans. These methods and other available ones, however, do not have the inherent characteristics needed for a definitive quantification of the regional perfusion and metabolism of heart muscle. The commonest cardiac problem encountered in adults is ischemic heart disease: a condition caused by the inadequate flow of blood to regions of the heart. Symptoms range from episodic chest pain (angina pectoris) to circulatory insufficiency (congestive heart failure) to myocardial infarction ("heart attack").

Coronary artery disease, which is responsible for these disorders, leads not only to symptoms but also to derangements in regional metabolism within the heart. With the PET technique tracers such as carbon 11 can be localized precisely and accurately within the heart muscle. Furthermore, when the positron-emitting tracers are incorporated in appropriate substances such as palmitate, they make it possible to characterize regional metabolism with a substance whose chemical behavior is identical with that of the physiological substance being traced.

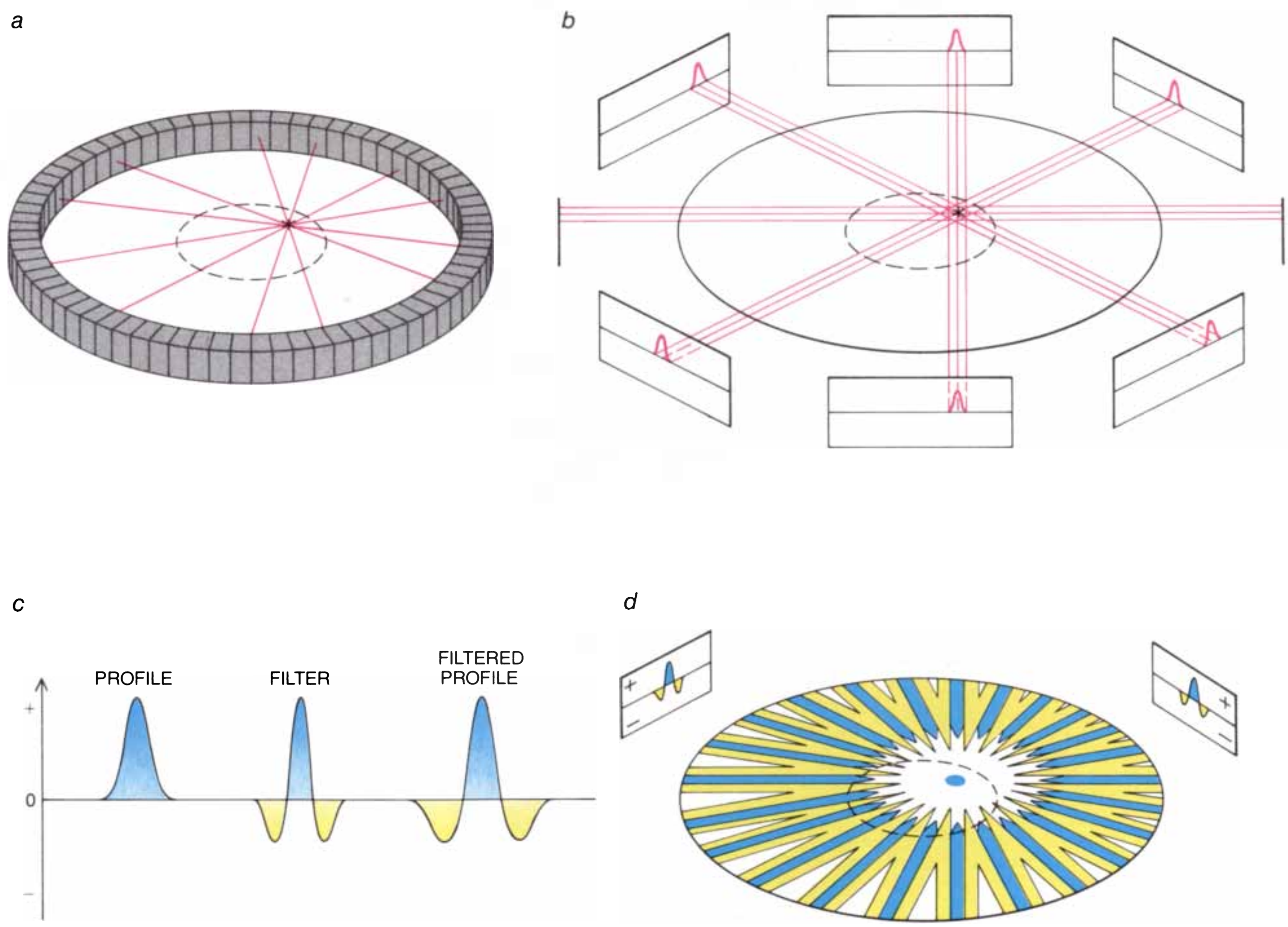
In order to define the externally detectable nature of the important biochemical processes in the heart that are amenable to assessment by PET, studies have been done in hearts isolated from experimental animals and also in the hearts of living human subjects. The accumulation of any tracer in a region of the heart depends on several factors, including the rate of delivery of the tracer to the tissue (a function of regional perfusion), the residence time of the tracer, the fraction of the tracer that is extracted on a single pass through the coronary circulation and the rates of metabolism and washout of the tracer initially extracted by the heart muscle. The extraction fraction and the metabolism are critically dependent not only on the nature of the traced substance but also on the status of regional heart-muscle metabolism at the time of its evaluation.

This critical relation was established in studies with hearts isolated from animals and perfused with a nutrient fluid. Within seconds after the injection of palmitate labeled with carbon 11 into the perfusion fluid the PET image records a peak of activity that is proportional to the amount of tracer injected. The peak activity declines rapidly as a sizable fraction of the tracer is carried away by the coronary circulation. This decline is followed by a slower one that represents primarily the exchange and washout of tracer from the fluid between the blood vessels. The third and final phase of the decline in activity follows a slow exponential curve that reflects primarily the fate of ^{11}C-palmitate incorporated into lipids within the heart-muscle cells.

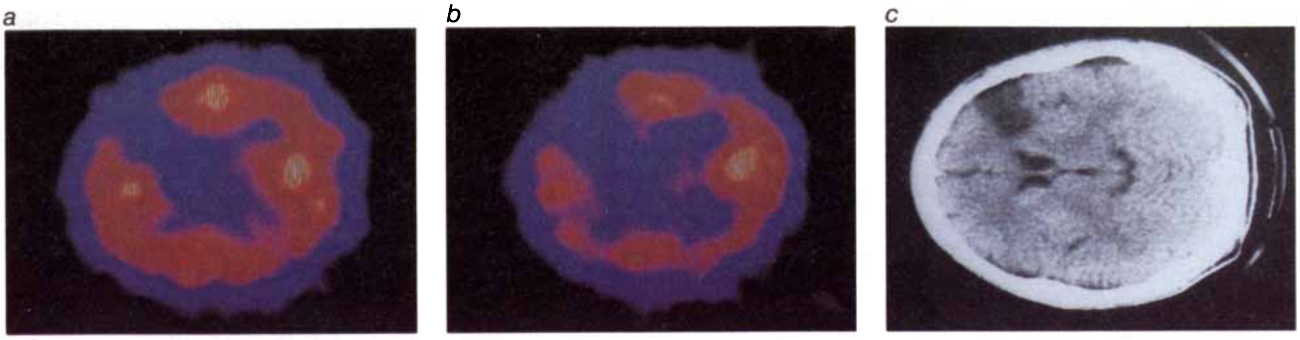
By comparing this final phase with the initial upstroke of the curve of time v. radioactivity one can obtain an index of the net extraction fraction of ^{11}C-palmitate into neutral lipids. In studies with perfused animal hearts it is possible to control the flow requirements and the metabolic requirements of the heart independently. Under conditions in which flow is kept constant but metabolic demands are modified the rate of decline of radioactivity in the heart toward the end of the curve of time v. radioactivity quantitatively reflects changes in the metabolism of the heart muscle. Such observations suggested that decreased uptake of free fatty acid induced by ischemia should be readily detectable by PET in experimental animals and human patients.

When ^{11}C-palmitate is administered intravenously to experimental animals, the heart muscle accumulates the labeled material homogeneously. When heart muscle is made ischemic, however, PET images show a marked decrease in the uptake of ^{11}C-palmitate. The distribution of the palmitate and the extent of defects in its uptake seen in the images correspond closely to histological and enzymatic criteria of irreversible injury of the heart muscle observed in postmortem examinations.

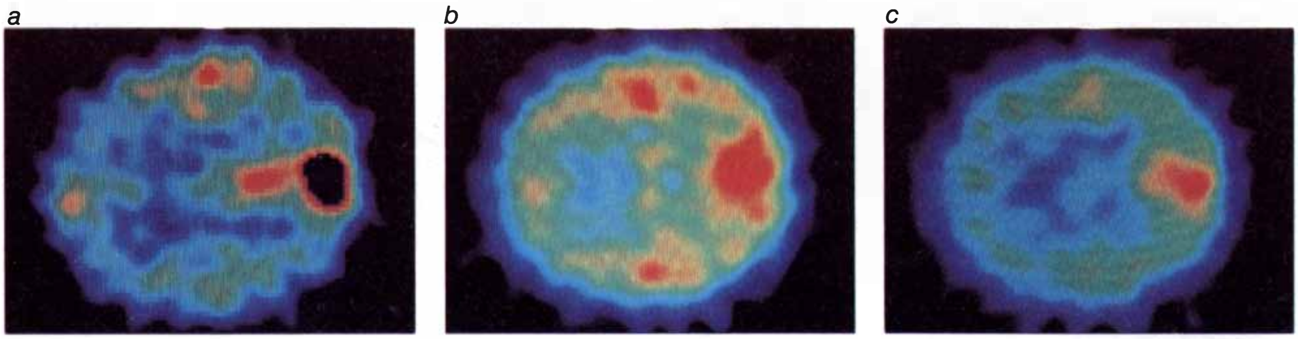
In addition to revealing irreversible injury of the heart PET studies with ^{11}C-palmitate can delineate depressed heart muscle where the flow in the coronary arteries has been subjected to transitory occlusion. Sequential tomographic studies make it possible to differentiate between reversible and irreversible



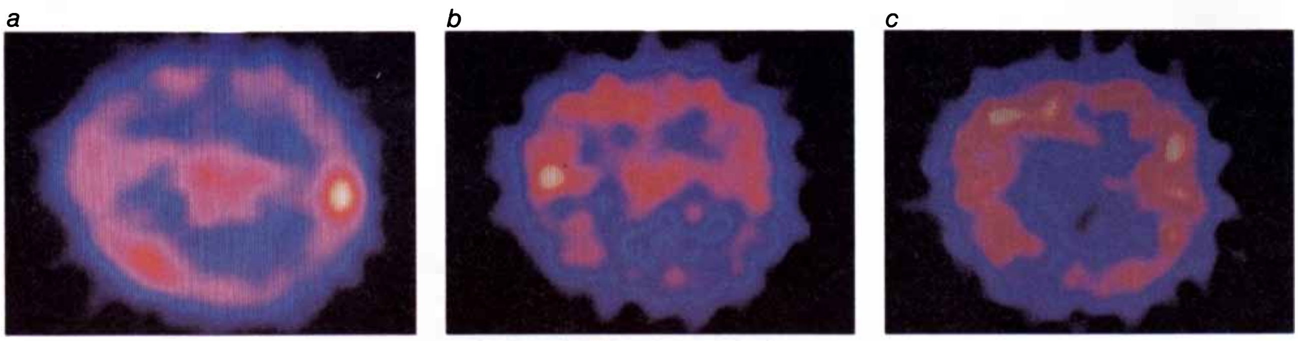
PET PROCESS has four basic steps. In the first step (*a*) the gamma-ray photons that mark the decay sites of a positron-emitting compound are collected by a circular array of detectors. The measurements represent a series of straight lines, all in a single plane. The information is sorted into a series of parallel lines (*b*) that represent a projection at different angles of the radioactivity within a selected region of the subject. These projections, initially recorded as profiles, are passed mathematically through a filter, a process termed convolution that provides a means of removing artifactual structures in the image (*c*). The convolved profiles are then "back-projected" to reconstruct an image of how the labeled compound was distributed (*d*). Manipulation of the large amount of collected data calls for a fast computer.



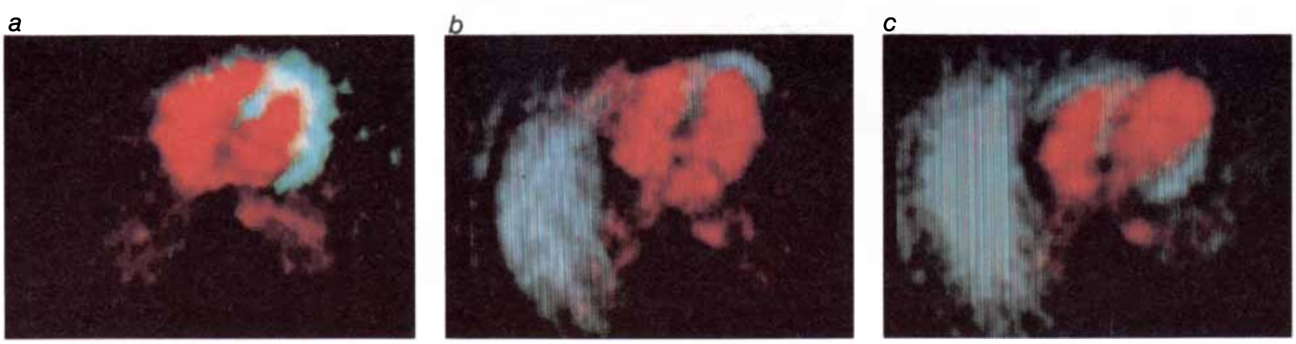
IMAGES OF A BRAIN INFARCT in stroke are shown in PET scans (*a and b*) and a CAT scan (*c*). The infarcted region is in the upper left quadrant (with the patient's face to the left). Such infarcts show up immediately in PET scans. A CAT scan reveals only changes in tissue density, which develop slowly. The first PET image (*a*), made after the patient had received a sample of blood labeled with ^{15}O-water, shows cerebral blood flow. Second PET image (*b*) shows oxygen consumption after patient had inhaled oxygen including ^{15}O.



SITE OF A FOCAL SEIZURE in an epileptic appears in PET images, which show cerebral blood volume (*a*), blood flow (*b*) and oxygen consumption (*c*). In image *a* the tiny red dot appearing at 12 o'clock depicts increased blood volume at the site of the focal seizure.



SITE OF A BRAIN TUMOR, a large meningioma, can be identified in these PET images, which also show cerebral blood volume (*a*), blood flow (*b*) and oxygen consumption (*c*). Meningiomas are highly vascularized tumors in the brain's covering membrane. This one is revealed by increased blood volume in the center. There is also a local increase in blood flow but a decrease in the oxygen consumption.



NORMAL AND DAMAGED HEARTS exhibit striking differences in these PET images, which depict the response of the heart to two tracer compounds. The green areas show the distribution of carbon 11 injected in the form of ^{11}C-palmitate, the salt of palmitic acid, one of the readily metabolized fatty acids. The red areas show the distribution of carbon 11 inhaled as ^{11}C–carbon monoxide. In a normal heart (*a*) the ventricles are bright green. The green areas are truncated in the PET image of a patient with a lateral myocardial infarction (*b*) and in a patient with an anterior myocardial infarction (*c*). Green structure that appears to the left of the heart in *b* and *c* is the liver.

injury: the reversibly injured heart muscle initially fails to accumulate ^{11}C-palmitate but subsequently accumulates the material in a normal fashion when the flow of blood has been restored.

Among subjects without overt or suspected coronary artery disease PET images show that the distribution of ^{11}C-palmitate administered intravenously is consistently homogeneous throughout the ventricle of the heart. On the other hand, among patients who have suffered a myocardial infarction at some time in the past (an episode resulting in the death of some heart-muscle cells) PET images disclose regions of diminished uptake of ^{11}C-palmitate in places that correspond to those of the initial infarction (as defined by electrocardiography). Moreover, the extent of infarction sustained in these patients and detected by PET corresponds closely to the extent of infarction estimated by biochemical measurements.

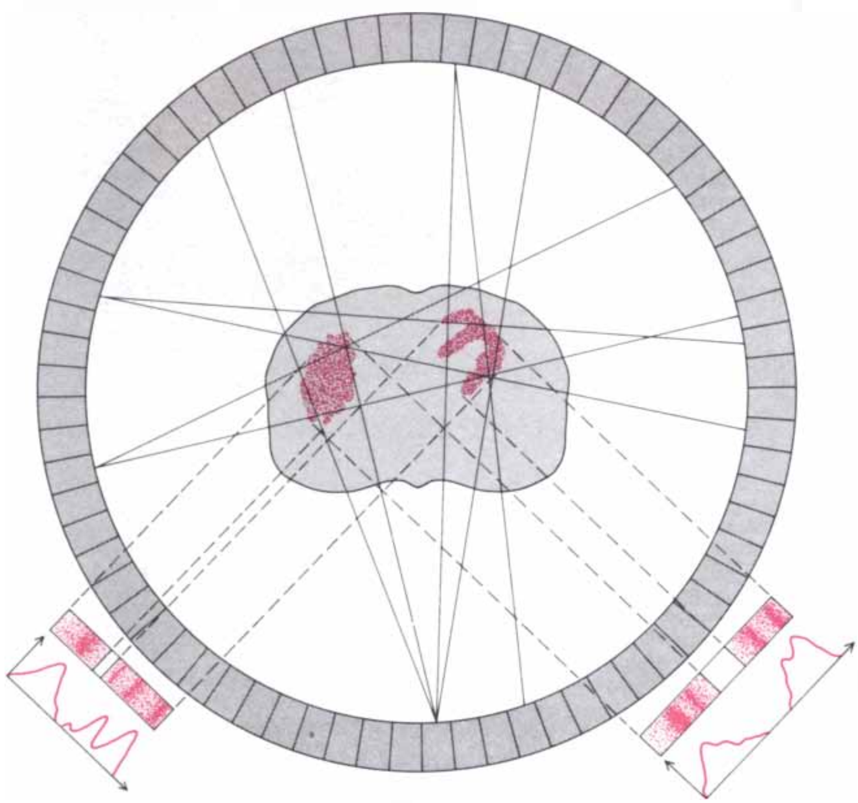
PET with ^{11}C-palmitate has also helped to clarify the mechanism responsible for a global, or generalized, impairment of the pumping function of the heart in patients with acute myocardial infarction. According to one view regional abnormalities in the motion of the heart wall may be due primarily to the delayed conduction of the electrical impulses that excite the heart muscle. Another hypothesis is that the abnormalities reflect altered biochemical and mechanical functions of the heart muscle itself. In the patients we have studied the global impairment of ventricular function and the localization of abnormalities in heart-wall motion are both closely correlated with impaired metabolism revealed by PET in the ventricular muscle. It is clear that in these patients at least the wall-motion abnormalities are attributable primarily to impaired regional metabolism in the heart muscle induced by decreased perfusion rather than to abnormalities in conduction.

The experience with ^{11}C-palmitate provides one example of the potential of PET for the delineation of regional metabolism in the heart muscle. Other agents such as deoxyglucose labeled with fluorine 18 have been used to obtain images that qualitatively resemble those obtained with ^{11}C-palmitate. It is somewhat difficult, however, to interpret fluorine-18 images unequivocally because fluorinated compounds frequently exhibit metabolic properties different from those exhibited by naturally occurring metabolites.

Because the tracers in PET studies of metabolism exhibit a complex behavior in the organism one must often resort to fairly elaborate mathematical models and data-acquisition schemes in order to extract metabolic rates from the quantitative data supplied by PET images in conjunction with the radioactivity measured in samples of arterial or venous blood taken from the subject's arm or leg. There are several important requirements in such measurements. First, one assumes that the labeled substance is transported and metabolized in the same manner and at the same rate as the unlabeled substance one seeks to trace. The requirement is clearly met when a substance such as glucose is labeled with carbon 11. It is not met, however, with analogues in which fluorine 18, for example, replaces oxygen; suitable correction factors must be applied.

A second requirement is that it is necessary for the metabolized tracer to be retained within the organ of interest during the measurement period. With a substrate such as glucose labeled with carbon 11 one can expect the substance to be retained in the brain, for example, for the length of time needed to perform a scan. Immediately on entering brain cells ^{11}C-glucose is phosphorylated, a reaction that yields ^{11}C-glucose-6-phosphate. Because the brain is poor in enzymes capable of reversing this reaction and is impermeable to glucose-6-phosphate the carbon 11 is retained until the ^{11}C-glucose-6-phosphate is transformed into a series of degradation products, which ultimately leave the brain in the venous blood flow. As a result carbon 11 remains effectively trapped in the organ long enough (usually five minutes) for the metabolism to be measured by the recording of PET images. When analogues such as glucose labeled with fluorine 18 are used, the time requirement is much relaxed because metabolism stops with the phosphorylation step. The fluorine-labeled phosphate remains irreversibly trapped in the cells of the nervous system.

A third requirement in metabolic studies with the PET technique is that the amount of tracer not metabolized, and hence remaining as free tracer in the blood and the extracellular fluid, be either negligible or accurately accounted for at the time of measurement. This requirement is most easily met when the tracers are radioactive analogues. Since such tracers are irreversibly trapped in the organ being studied (and in cells elsewhere), one can delay the PET mea-



BACK-PROJECTION OF CONVOLVED PROFILES reconstructs the distribution of radioactivity within the subject at a selected section of the body. Here the selected section passes through the heart and the liver; the subject's spinal column is at the bottom. By PET convention the view is from the feet, so that the *W*-shaped structure at the right is the heart, with the liver at the left. The distribution is typical of the kind obtained when the patient receives an injection of ^{11}C-palmitate. Palmitate is carried to the heart and liver, where it supplies energy.

surement until the free tracer in blood and tissue has fallen to insignificant levels. When the tracers are glucose analogues, a delay of between 30 minutes and an hour is enough. When the tracers are not analogues but true metabolic substrates, such a delay is not possible because the labeled metabolites begin leaving the tissue five minutes or so after the time of the injection. The experimental design must then explicitly account for the free tracer in blood and tissue. This is accomplished in part by measuring the regional blood volume in conjunction with measuring the regional metabolism. By knowing the regional blood volume one can calculate the free tracer not only in the blood but also in the tissue itself.

One alternative to accounting for the amount of free tracer in blood and tissue is to follow the egress of metabolized tracer from the organ under study by repeating the measurement of metabolism several times in the first five minutes after the injection of the tracer. The calculated metabolic rate from successive measurements will obviously decrease with time, and from the rate of decline one can derive the true metabolic rate. In preliminary studies such a procedure seems quite workable. We have tested the approach with both ^{11}C-glucose and ^{15}O-oxygen. The latter test is particularly stringent because of the speed with which oxygen 15 is metabolized into water and leaves the tissue. In both cases we obtained excellent values for the true rate at which the labeled substances are utilized. In early experiments we were limited to PET images of a single "slice" of tissue. Now even better results can be obtained from advanced PET systems that are capable of rapidly scanning several slices simultaneously.

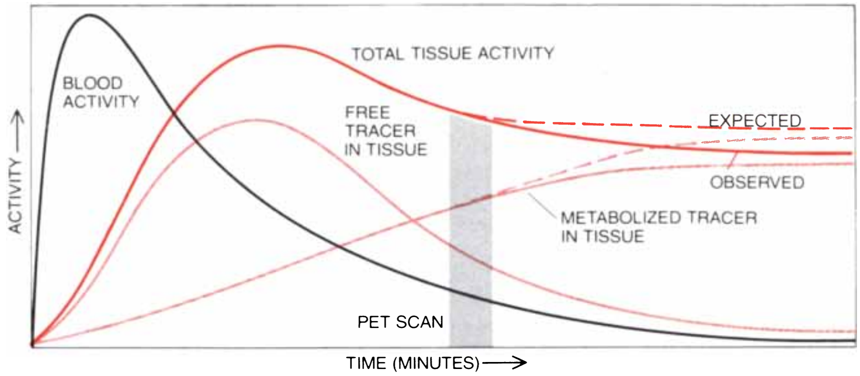
With fast multiple-slice PET systems it is also easier to establish the amount of tracer in the tissue that remains unmetabolized. The necessary data can be derived from sequential PET images of regional tissue activity as a function of time in the course of a slow intravenous infusion of tracer. If such an approach works as it is expected to, one should be able to get more than just the local utilization rate of a substrate such as glucose or palmitate. One should also be able to establish the substrate's concentration in the organ of interest, its spatial distribution and its duration of residence. In human patients such determinations should be extremely valuable in the ultimate understanding of disease processes as well as in their diagnosis and management.

Although metabolism studies with radioactively labeled substrate analogues rather than with the true substrates have the advantage that the analogues are irreversibly trapped once they are metabolized, the analogue approach has disadvantages. Since the analogues are not biochemically identical to the natural substrates, departures from the normal rate of metabolism must be determined experimentally. Moreover, there are important differences among animal species in the metabolism of analogues. Additional differences can be expected if the organ under study is diseased. Other difficulties have to do with time and with radiation-dose levels. With substrate analogues a period of 30 minutes to an hour must elapse before the PET measurement can be made, in order to allow time for unmetabolized tracer to clear from the blood and the tissue.

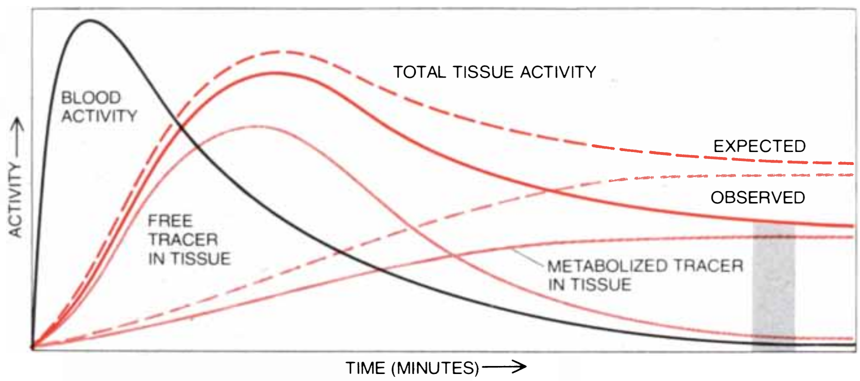
This time requirement makes it hard to study transient phenomena such as an epileptic seizure that may last only a few minutes. It also necessitates tracers with long half-lives or, if the tracer can only have a short half-life, calls for the administration of a fairly large dose of radiation to the patient. Finally, substrate analogues can disrupt normal cell metabolism if they are administered in excessive amounts. One must therefore take care that the radioactively labeled analogue be relatively free of unlabeled material.

In addition to determining the rates of such continuous processes as metabolism the PET technique can quantitatively measure the chemical composition of body tissues with a suitable choice of labeled compounds that rapidly distribute themselves between the blood and the organ of interest. For example, the carbon dioxide content of the brain of the rhesus monkey has been measured by having the animal inhale carbon dioxide labeled with carbon 11. The measurements were made with various concentrations of carbon dioxide in the blood and showed the expected relation between the level of carbon dioxide in the arterial blood supply and the level in the brain. The experiments also illustrate the unique potential of PET for quantitatively determining the acidity of living brain tissue. Such measurements of acidity levels should be of great value in evaluating brain damage resulting from stroke or physical injury.

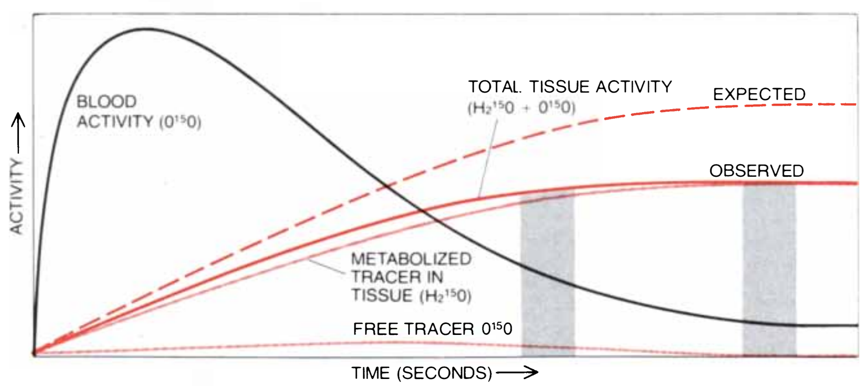
The general scheme of measuring carbon dioxide in the brain can be extended to explore the response of virtually any organ to any pharmaceutical that can be labeled with a positron-emitting isotope. An obvious application is to measure drug levels in various organs. PET studies should help to define, for example, the therapeutic efficacy of anticonvulsants in the brain or digitalis in the heart. For the first time it should be possible to map in human subjects the response of specific brain receptors and transmitters to drugs with specific binding characteristics. Such investigations have already been conducted in animals. There is reason to hope that similar investigations in human beings will clarify the action



INTERPRETATION OF PET MEASUREMENTS depends critically on the use of mathematical models that relate the concentration of particular trace compounds observed by PET at specific tissue sites to a particular metabolic process. Two simultaneous measurements are required: the level of tracer activity in the blood, determined by sampling a peripheral artery or vein, and the activity in the target tissue, determined by PET. The tissue activity consists of both metabolized tracer and unmetabolized tracer. Because the PET technique cannot distinguish between the two a mathematical model becomes critical. For example, when metabolism is measured with a labeled substrate such as ^{11}C-glucose, tissue activity must be measured early to avoid a deviation from the expected values due to the loss of metabolized tracer from the tissue. Much of the measured activity is unmetabolized tracer, even when the blood activity in the region of interest has been accounted for. A good mathematical model enables the user of PET to distinguish between metabolized and free tracer and establish the local metabolism.



METABOLISM OF THE SUBSTRATE ANALOGUE ^{18}F-2-deoxy-D-glucose differs in two important ways from that of normal glucose: It is not taken up as avidly by tissue cells, and once it is taken up it is not released in any form. Because of the latter property the PET measurement can be delayed, thereby minimizing the radioactive contribution of free tracer in the blood and tissue. Because the analogue is not accepted by the cells as readily as the normal metabolite, however, the observed tissue activity departs substantially from the activity expected.



MEASUREMENT OF OXYGEN CONSUMPTION calls for yet another mathematical model. Because actively metabolizing tissue, such as the brain, stores almost no oxygen, radioactive oxygen (^{15}O) is rapidly metabolized to water ($H_2{}^{15}O$). Therefore the PET measurement, when it is corrected for the presence of blood activity, covers only the metabolized tracer. Since water leaves the tissue rapidly, however, the observed activity increasingly falls below the level expected. One way to improve the accuracy of the measurement of oxygen consumption is to make several sequential PET scans at intervals that are spaced only a few seconds apart.

of psychotropic agents on certain conditions for which satisfactory animal models do not exist, such as schizophrenia and Parkinson's disease.

The development of a completely satisfactory measurement of blood flow by means of PET has lagged behind other PET applications for several reasons. First, the extraction fraction of many tracers is influenced by the metabolic status of the tissue as well as by the residence time, impairing accurate measurement of regional perfusion. Second, the incomplete trapping of tracer initially extracted by the tissue of interest and the recirculation of tracer obscure the relation between accumulation and regional perfusion. Third, the uptake of tracer into fat-soluble constituents of the tissue may differ from that taken up into water-soluble constituents, contributing to a variability in the relation between regional perfusion and tracer accumulated within a given region of the organ.

In addition to these biological complexities, the quantification of regional perfusion of organs such as the heart and the brain with radioactive isotopes has been limited by difficulties encountered in the accurate determination of the amount of tracer in specific regions of the organ. PET offers the promise of overcoming such difficulties, and so regional perfusion of the heart has been assessed by intravenous injection of selected positron-emitting isotopes and detection of their distribution by tomography. One tracer employed for this purpose is ^{13}N-ammonia ($^{13}NH_3$), a compound believed to behave like potassium and other singly charged positive ions. Another promising tracer is rubidium 81, a closer analogue to potassium. Ammonia is metabolized actively and is incorporated into glutamine by heart muscle, so that its extraction fraction is not constant but depends on metabolic factors.

In perfused isolated rabbit hearts, the retention of ^{13}N-ammonia is influenced markedly by the metabolic status of the heart under conditions where flow can be rigorously controlled. In spite of such limitations it has been proposed that the reduced accumulation of ^{13}N-ammonia in heart muscle provides a quantitative estimate of diminished regional perfusion. The net accumulation of tracer may be the result, however, of several competing phenomena: a decrease in the extraction fraction accompanying regional ischemia and caused by an altered metabolism of the heart muscle, possible increases in the extraction fraction reflecting a prolonged residence time in regions of low blood flow, possible changes in permeability in ischemic zones influencing the extraction fraction and a decreased delivery of the tracer reflecting the low blood flow itself. The complex interplay of these phenomena and related ones may limit the utility of ^{13}N-ammonia for defining regional perfusion quantitatively even though the distribution of the tracer can be detected by PET.

In particular, the measurement of blood flow in the brain has been beyond the capability of most PET systems currently in operation because they are too slow to collect enough data for an adequate reconstruction of a tissue slice in the brief time available, which is measured in seconds. As a result alternative approaches have been explored. One approach is to construct an equilibrium image of water labeled with oxygen 15, obtained while the subject is inhaling ^{15}O–carbon dioxide, and to deduce the blood flow from it. Although the amount of oxygen-labeled water in the brain is clearly proportional to the blood flow, the relation is not linear. Moreover, for reasons not yet

PET APPARATUS is the latest model, PET VI, designed and built at the Washington University School of Medicine. The structures in a circle are the photon detectors that record passage of the gamma rays created when a positron and an electron annihilate each other.

understood the method seems to give erratic results.

Perhaps the most promising approach at the moment is a modification of one devised originally for measuring blood flow in mice and other small laboratory animals. A freely diffusible labeled tracer is infused intravenously for one minute, after which the animal is decapitated. The amount of tracer in actual slices of brain is determined from autoradiograms and is related to the time v. activity curve of tracer in the animal's arterial blood up to the time of its death. From the two sets of data the blood flow can be calculated. It should not be difficult to substitute PET images for the autoradiograms and to arrive at the blood flow in living subjects.

The technology of positron-emission tomography is well developed, and both its capabilities and limitations are increasingly understood by biological investigators and to a lesser degree by clinical workers. Many PET studies have now been carried out with results that cannot be obtained by any other technique. They include the determination of regional metabolism in the brain and the heart, studies of the permeability of tissues, the measurement of the size of infarcts left in the heart by coronary attack and promising investigations in the physiology of psychoses. The great promise of PET technology for medicine is that it can provide glimpses into pathological conditions that are difficult or impossible to get by any other means. PET offers a particularly powerful tool for assessing the effects of drugs on tissues that are diseased or malfunctioning. A striking example is the demonstrated ability of PET to disclose regional metabolic changes apparently reversed by drug therapy in schizophrenic patients.

In clinical oncology PET offers the possibility of regarding malignant tissue as a metabolic entity and of measuring the effect of treatment—whether it is radiation therapy, hormone therapy or chemotherapy—by the changes in the malignant tissue and by the biochemical reactions of the normal tissues around it. In current practice the size of a malignant tumor and its response to therapy are based almost entirely on morphological examination: direct observation, including palpation, of accessible tumors or X-ray imaging of inaccessible ones.

LABEL	HALF-LIFE	LABELED SUBSTANCE	USE
OXYGEN 15	~2 MINUTES	OXYGEN	METABOLISM
		WATER	BLOOD FLOW
		CARBON MONOXIDE	BLOOD VOLUME
		CARBON DIOXIDE	BLOOD FLOW
NITROGEN 13	~10 MINUTES	AMMONIA	ORGAN PERFUSION METABOLISM
		VARIOUS AMINO ACIDS	METABOLISM PERMEABILITY (BRAIN)
		NITROUS OXIDE	BLOOD FLOW
		1, 3 BIS (2-CHLOROETHYL) 1-NITROSOUREA (BCNU)	TUMOR DRUG LEVELS
CARBON 11	~20 MINUTES	CARBON MONOXIDE	BLOOD VOLUME
		CARBON DIOXIDE	TISSUE *p*H
		VARIOUS ALCOHOLS	PERMEABILITY (BRAIN) BLOOD FLOW
		VARIOUS ETHERS	BLOOD FLOW TISSUE LIPID CONTENT
		ACETATE	METABOLISM (HEART)
		PALMITATE	METABOLISM (HEART)
		METHYLALBUMIN	TISSUE HEMATOCRIT (RATIO OF CELLS TO PLASMA)
		OCTYLAMINE	MAPPING OF MONOAMINE OXIDASE RECEPTORS (LUNG)
		GLUCOSE	METABOLISM
		2-DEOXY-D-GLUCOSE	METABOLISM
		PHENYTOIN	TISSUE DRUG LEVELS (EPILEPSY)
		THYMIDINE	METABOLISM (TUMORS)
		DOPAMINE NOREPINEPHRINE	NEUROTRANSMITTER PHARMACOLOGY (BRAIN)
		FLUNITRAZEPAM ETORPHINE PIMOZIDE	NEURORECEPTOR PHARMACOLOGY (BRAIN)
FLUORINE 18	~110 MINUTES	2-DEOXY-D-GLUCOSE	METABOLISM
		3-DEOXY-D-GLUCOSE	METABOLISM
		HALOPERIDOL SPIROPERIDOL	NEURORECEPTOR PHARMACOLOGY (BRAIN)

POSITRON-EMITTING SUBSTANCES synthesized for PET studies provide probes for a wide variety of biological processes. The short-lived isotopes of oxygen, nitrogen and carbon are particularly useful for tracing the metabolism of their normal counterparts but can also be incorporated in compounds of pharmacological interest. The long-lived positron-emitting isotope of fluorine, ^{18}F, is substituted for oxygen in analogues of normal metabolites or drugs.

The price of providing PET technology is high. It includes the presence of a cyclotron in the medical center, appropriate chemical facilities and one or more arrays of PET imaging and computing equipment, together with physicists, chemists, mathematicians, physiologists and physicians able and willing to work together as a group. Such groups have been assembled, and they are actively at work in some 40 institutions throughout the world. To a large degree the development of PET can be credited to the Atomic Energy Commission (now superseded by the Department of Energy), which supported some of the pioneer PET efforts, to the National Heart and Lung Institute, which appreciated PET's potential early in the investigation of brain and heart diseases, and more recently to the National Institute of Neurological and Communicative Disorders and Stroke, which took the initiative in subsidizing PET investigations in several medical centers. Recognizing the potential market, a number of companies are making the dedicated cyclotrons and complete PET imaging-and-computing systems.

At present the resolution achieved by the latest commercial instruments is about one centimeter in exposures of one minute. It seems reasonable to hope that the next generation of instruments will improve the resolution by a factor of two with exposures of a few seconds. The incorporation of photon time-of-flight information in the reconstruction process, which calls for the accurate measurement of the time differential between the arrival of two annihilation photons at the coincidence detectors, promises to enhance the quality of PET images considerably by improving the overall signal-to-noise ratio. PET has reached a stage where its future in biology and medicine seems assured.

Oil Mining

Conventional drilling and pumping removes less than half of the oil in an average petroleum reservoir. As the price of oil rises, mining it underground or at the surface becomes economically more attractive

by Richard A. Dick and Sheldon P. Wimpfen

The conventional method of tapping a petroleum reservoir is notoriously inefficient. The portion of the resource that can be withdrawn from an underground oil pool by pumping drilled wells can be less than a tenth of the total and averages considerably less than half. In the days of cheap and plentiful oil inefficient extraction of this kind drew little notice. Today, with domestic petroleum reserves shrinking steadily and the price of imported oil 1,500 percent higher than it was only a decade ago, the U.S. should no longer neglect a major asset at a time of need. Much of the oil left in the ground after pumping can be mined.

The concept of oil mining may appear radical. The fact is that in the U.S. mining for oil is nearly as old as drilling for oil. Moreover, oil has been and is still being successfully mined abroad. In Ohio and in California more than a century ago the mining approach was applied to the extraction of oil in two different ways. At Macksburg, Ohio, in 1865, when a conventional well yielded little oil and much water, drilling was halted and a large-diameter shaft was sunk in place of the small drill hole. The oil was then skimmed from the top of the water. Unfortunately the shaft yielded little more oil than the well had.

The next year, attracted by oil seeps on Sulphur Mountain, north of Ventura, Calif., Josiah Stanford, a pioneer of the oil industry in the West, excavated a series of tunnels into the mountainside, all slanted slightly upward. Before Stanford was through 30 tunnels had been dug and each of them was yielding between one and 20 barrels of oil per day, delivered by gravity to the tunnel mouths. These tunnels and an equal number of deeper ones dug into the Sulphur Mountain oil reservoir in the early 1890's continued to yield oil for several decades.

Oil mining in Europe is an even older art. At Merkwiller-Péchelbronn in the department of Bas-Rhin in eastern France mine galleries were first dug into an oil-bearing formation in 1735. Initially the oil simply seeped into ditches cut into the tunnel floor, but in the 19th century gravity collection was supplemented by drilling. By the 1930's the network of oil-mine tunnels extended over an underground area of three square miles. At Wietze, between Hamburg and Hannover in West Germany, a less ambitious oil-mining operation late in the 19th century seems to have paralleled the Ohio shaft-sinking effort. As the demand for oil in Germany rose during World War I the effort at Wietze was increased. In 1917 a shaft 13 feet in diameter was sunk to a depth of 495 feet; in the year 1918 the shaft was skimmed of 127,000 barrels of oil.

Today the world's major oil-mining efforts are in Canada and in the U.S.S.R. The Canadian enterprise is actually the exploitation of "tar sand" rather than an oil reservoir, but the end product is oil and the operations are instructive. The tar sand (a sandstone permeated with soft bitumen) lies between a limestone bedrock and a soil overburden. The material is mined by the open-pit method. A swath of overburden some 50 feet wide is stripped off to expose the tar-sand stratum, which is some 130 feet deep. The soft rock is gathered into long windrows and transferred mechanically to conveyor belts that carry it away to the processing plant. To produce 45,000 barrels of synthetic crude oil per day from the bitumen in the rock requires the delivery of 117,000 tons of tar sand to the processing plant. (The crude is called synthetic because it is obtained by distilling the bitumen.) The 45,000-barrel-per-day production target has already been exceeded, and a second facility, financed by an international consortium of oil companies, should soon be producing 125,000 additional barrels per day.

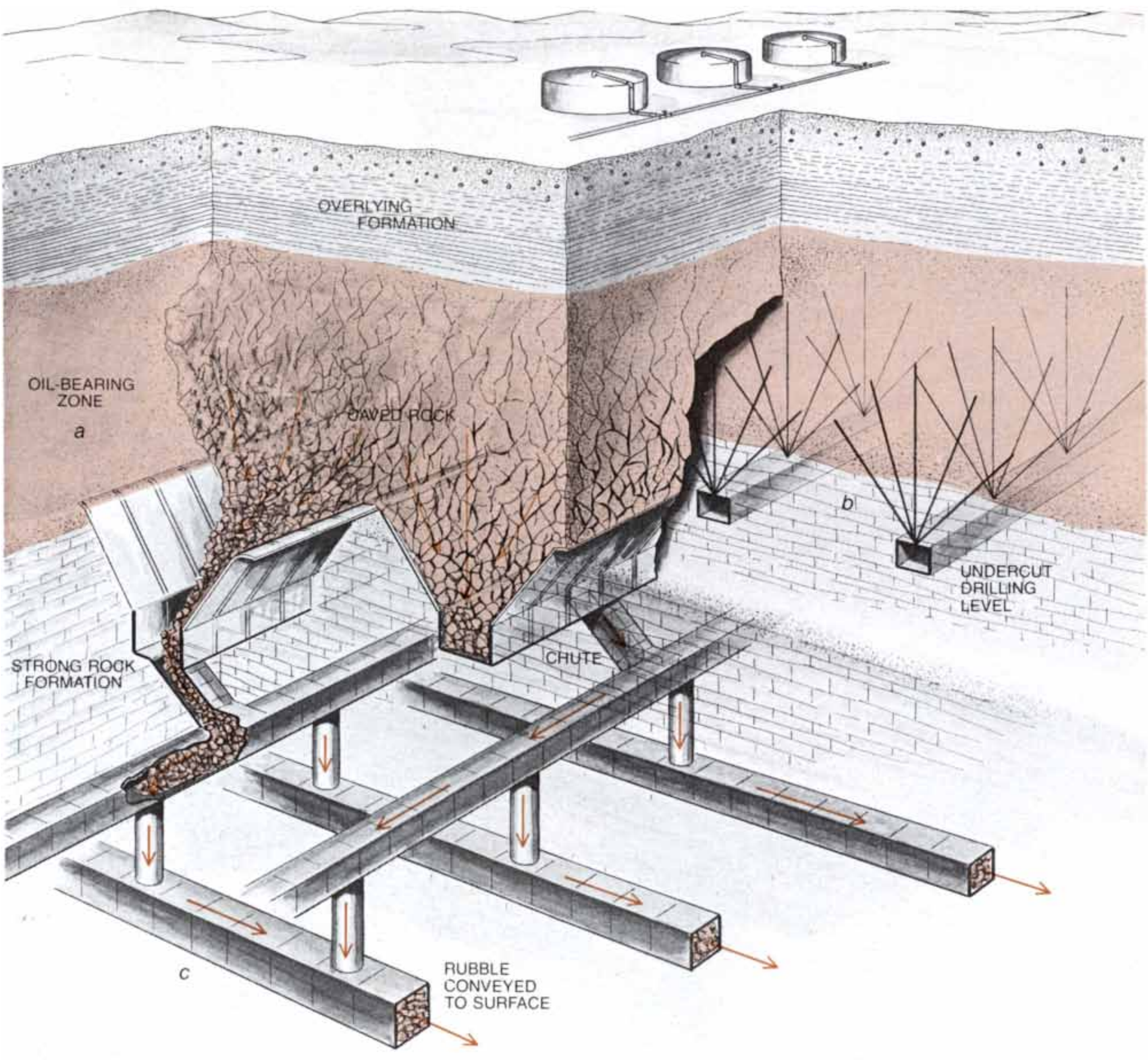
The Russian effort was a response to low yields from drilling at the Yarega oil field, west of the Urals near Ukhta in the Komi Autonomous S.S.R. The reservoir, some 650 feet below the surface, consists largely of "heavy oil," that is, oil too viscous to be extracted by conventional drilling and pumping. Indeed, it was estimated that surface pumping had exploited only 2 percent of the reservoir. To improve the situation shafts were sunk to a depth of more than 500 feet and from each shaft a network of horizontal galleries was excavated. From these work areas inclined passages were driven down into the oil-bearing formation and a production gallery, 80 feet in diameter, was excavated at the bottom of each passage. Drilling then produced a series of horizontal and slightly uptilted holes that radiated from each production gallery for a distance of as much as 800 feet.

In order to promote the flow of the heavy oil to these collection holes steam was injected into the reservoir rock through vertical holes drilled down from the horizontal galleries overhead. The warmed crude was heated further at a central underground terminal before being pumped to the surface. The Russians estimate that the combination of lateral drilling from the underground production galleries and the injection of steam will make it possible to recover 50 to 60 percent of the oil in the Yarega reservoir at the rate of more than six million barrels per year. Meanwhile they have begun similar mining operations at a depleted reservoir north of the Baku oil fields in the Azerbaijan S.S.R. and are considering doing the same at more than 300 known reservoirs of heavy oil elsewhere in the U.S.S.R.

How much oil might become available through oil mining in the U.S.? It has been estimated that there now remain in depleted U.S. oil fields and will eventually remain in fields currently producing or in reserve a total of more than 300 billion barrels of oil. The size of this potential resource can best be judged by comparing it with the estimated existing domestic oil reserve recoverable by conventional methods: less than 30 billion barrels. Of course, not all of the 300 billion barrels remaining in U.S. fields can be recovered by mining. The 300-billion-barrel figure, however, does not take into account still other large potential hydrocarbon resources. If one adds up known U.S. reservoirs of heavy oil similar to the Yarega crude, certain hydrocarbon-impregnated diatomaceous deposits in California and the tar sands of Utah (which alone are believed to contain a potential 30 billion barrels of synthetic crude), one finds that an additional 200 billion barrels of oil are potentially available. Even this figure excludes the potential for synthetic crude-oil production from the Green River oil-shale formation in Utah, Colorado and Wyoming. This formation is estimated to represent a reserve of 1.8 trillion barrels. Between 400 billion and 600 billion barrels of the reserve are believed to be recoverable. We shall not, however, address ourselves here to the mining of oil shale.

Not long ago the Bureau of Mines, an agency of the U.S. Department of the Interior, commissioned two feasibility studies in order to make an up-to-date assessment of the potential for oil mining in the U.S. The last such assessment had been conducted by the Bureau more than 40 years ago, when oil produced by conventional methods rarely cost more than $2 per barrel and U.S. reserves of petroleum seemed inexhaustible. To present the findings of the recent studies in their proper context we must first sketch out certain aspects of mining technology and then discuss three broad criteria of feasibility.

The methods of mineral extraction can be classified under three headings: surface mining, in situ mining and underground mining. The tar-sand operations in Alberta are typical of surface mining; conventional drilling (not only for petroleum but also for sulfur and other soluble mineral resources) is an example of in situ mining. Underground mining can take many forms, but so far as its potential for oil mining is concerned suitable methods are limited. For example, to tunnel underground and then transport oil-bearing rock to the



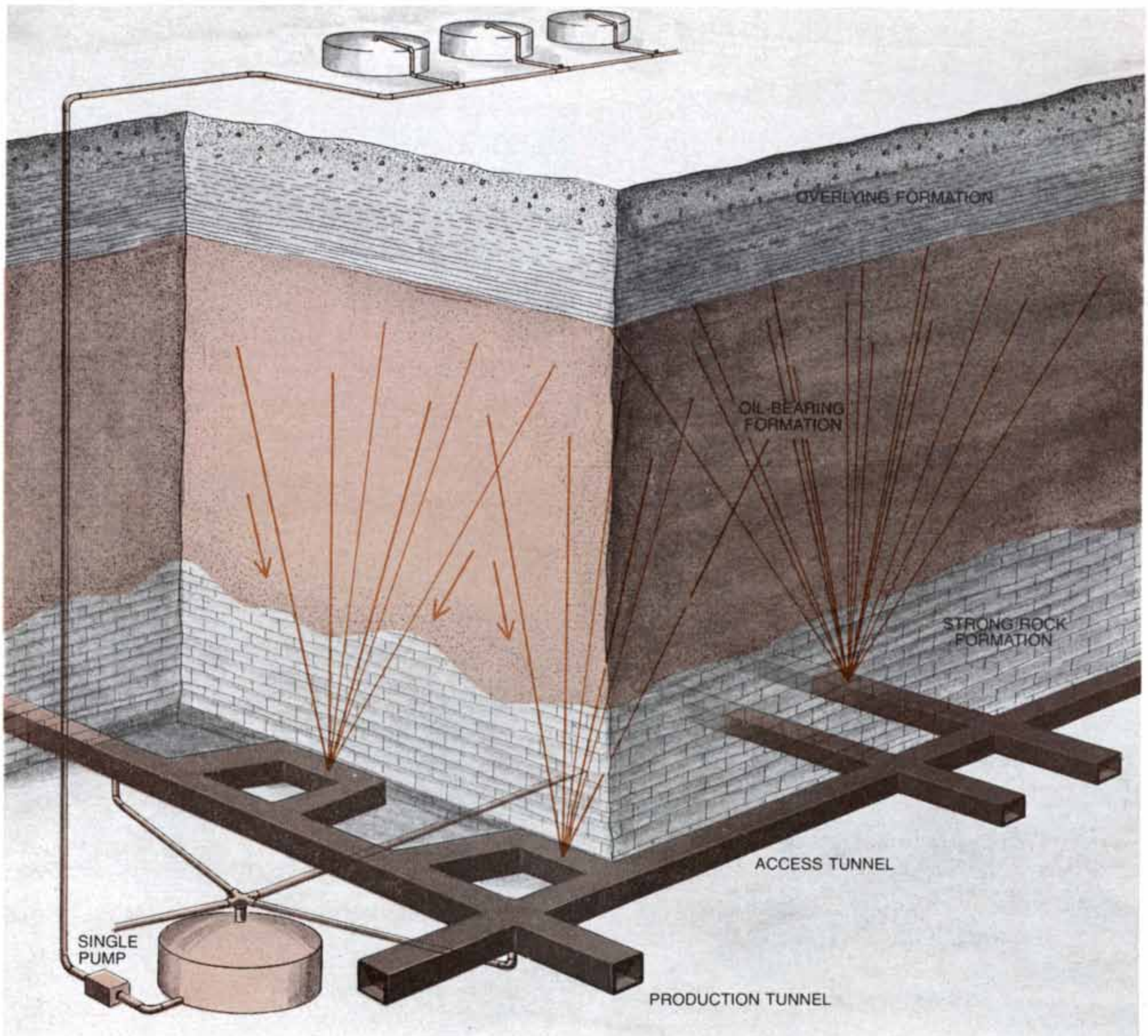
OIL IN WEAK ROCK could be tapped by underground mining. The first step would be the excavation of a series of connected tunnels in a strong rock formation underlying the oil-bearing formation of weak rock. The upper grid of tunnels is linked to chutes that rise to the bottom of the weak rock formation (*a*). Drilling, blasting and undercut excavation (*b*) cause the weak rock to collapse. The rubble is then drawn through the chutes into the upper tunnels. It next passes to a second and lower grid of tunnels (*c*) for haulage to the surface. A beneficiation plant at the surface would separate the crude oil from the rock rubble. This method of oil mining is called block caving.

surface, much as bituminous coal is mined, is not practical. Those oil reservoirs that exist in rock formations strong enough to be safe for tunneling are low in their proportion of oil to rock. Conversely, when the proportion of oil to rock is high enough to make mining the reservoir rock economically attractive, the rock itself is dangerously weak.

A more promising approach combines the underground and in situ methods of mining, as in the Russian oil mine at Yarega. A prerequisite for such "modified" in situ operations is the existence of a structurally strong rock stratum close to the oil reservoir, preferably under the reservoir but not more than 100 feet under it. The first step in such a mining operation would be to drive a shaft down into the strong rock under the reservoir; from the bottom of the shaft a series of tunnels would be extended horizontally until a streetlike grid underlay the entire reservoir area.

A typical oil reservoir contains, in ascending order, water, oil and gas. Working in the underlying tunnels the miners would drill upward into the reservoir; solid casings would extend through the water zone but the casing entering the oil zone would be perforated so that gravity could drain the oil into a collecting system in the tunnels below. From there the oil would be pumped up the mine shaft to the surface. Because the drill holes would be short ones, no longer than 100 to 150 feet, it would be economically feasible to drive many closely spaced holes up into the reservoir, thereby promoting maximum drainage. With proper controls the reservoir could be drawn down uniformly so that a high percentage of oil recovery would result. As at Yarega, steam stimulation could be used, if necessary, to ensure that the oil would flow to the drill holes.

One of our commissioned studies, taking the Wheeler Ridge reservoir near Bakersfield, Calif., as a model, estimates that the average cost of oil obtained by such a combination of underground mining and gravity drainage would be $14 per barrel. (The cost, cal-



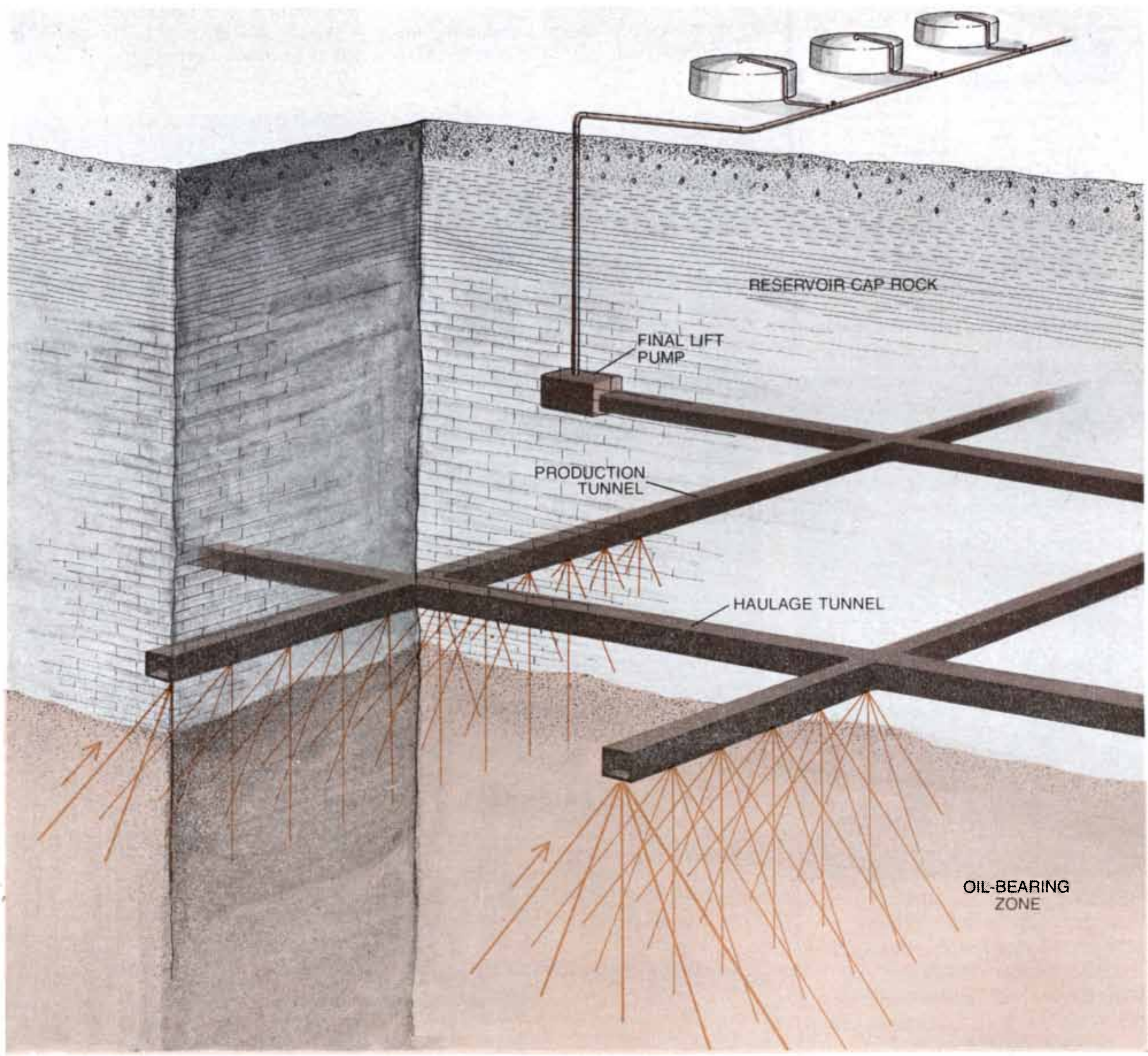
GRAVITY DRAINAGE is an oil-mining method that also depends on the existence of a strong rock formation under the oil reservoir. A grid of tunnels (*a*) is excavated in the strong rock, as close to the bottom of the reservoir as possible. A large number of short collecting holes are drilled up into the reservoir rock. The oil moves by gravity through a network of pipes to a collection tank and is then pumped to the surface for refining. The single pumping system makes the gravity-drainage method of oil mining economically advantageous.

culated in 1977 dollars, includes an allowance for royalty and profit.)

In petroleum exploration, however, it is the custom to stop drilling once the drill hole enters the oil-bearing stratum. As a result no one knows how many oil reservoirs in the U.S. are underlain by rock strata strong enough to allow the excavation of the tunnel system that gravity mines would require. Before any such mines could be planned an ambitious program of exploratory drilling would be needed to determine just which U.S. oil reservoirs are underlain by strong rock strata sufficiently close to the bottom of the reservoir.

Conversely, normal petroleum-exploration practices already provide a record of many U.S. oil reservoirs that are overlain by rock strata strong enough to allow a modified in situ system of extraction more closely resembling the Yarega operation. Here the tunnel system would be excavated above the reservoir and the holes would be drilled down into it. In lieu of gravity drainage each drill hole would be equipped with a pump. This combination of mining and pumping would necessarily be more complex and expensive than the gravity-drainage system but need not be prohibitively so. For example, another of our studies, taking the Irma reservoir in southern Arkansas as a model, estimates the inclusive cost of oil from such a pumping mine at $16 per barrel. (This estimate and all those that follow are also in 1977 dollars.) If steam injection should be required, as it is at Yarega, the estimated price would be increased to $24 per barrel. With the 1980 cost of imported oil in excess of $30 per barrel, pumping mines begin to appear economically attractive.

Economic considerations, although paramount, still constitute only one of three basic criteria of feasibility for oil mining. The other two are the degree of safety afforded to the miners and the extent of the disturbance to the natural environment. Obviously no mining system is acceptable that unduly jeopardizes the safety or health of the miners or unduly degrades the environment or



PUMPED DRAINAGE is an alternative method of exploiting oil reservoirs that are incompletely drained but do not overlie a strong rock formation. Here the access tunnels are excavated above the reservoir and many short holes are drilled down into the oil-bearing rock. Because each drill hole requires its own pump the method is less economical than gravity drainage. In the U.S.S.R. oil mines of this kind, utilizing steam injection to lower viscosity, recover heavy crude oil that cannot be obtained by conventional surface pumping.

both. (The word unduly is appropriate because no mining system exists that does not expose miners to some degree of risk and does not to some degree alter the environment.) In these two respects modified in situ systems of oil mining score from average to excellent.

With respect to miner safety, any underground work in proximity to hydrocarbon deposits involves the risks associated with fires and explosions. In modified in situ systems, however, the tunnel networks would be at some distance from the actual oil reservoir and would be contained in strong rock. Moreover, the oil taken from the reservoir would travel to the surface in a sealed system. The potential for gas leakage would nonetheless exist. Such mines would certainly require safety regulations fully as stringent as those imposed in underground coal mines.

With respect to environmental impact, underground mining in general has the potential for considerable degradation in terms of both mine waste and geological subsidence (the sinking of the ground after mining). The modified in situ system of oil mining presents a minimum threat to the environment on both counts. The amount of excavated material brought to the surface and later disposed of is small compared with that brought up by conventional mineral extraction. The prospect of subsidence is also slight. Most of the material extracted from such a mine is the oil itself. For example, the amounts of rock and water produced would be small, and the water could be reinjected into the oil reservoir, further minimizing the likelihood of subsidence. In summary, modified in situ oil mining would be a remarkably clean process.

Most U.S. mineral extraction is done not underground but by one or another of three techniques of surface

TAR SANDS OF ALBERTA are exploited by means of surface-mining techniques. Seen here are two of the trenches excavated in a tar-sand deposit by Syncrude Canada Ltd. Draglines pile up the tar-bearing soft rock in windrows; two can be seen at work close to the near ends of the trenches. The rock is then transferred to conveyor belts by means of bucket-wheel reclaimers; one can be seen at work just above the left dragline. The long conveyor belts deliver the rock to the processing plant (not seen), where the tar is converted into synthetic crude oil.


The man, the wi

…ng and the Apple.

If you could talk to Orville Wright, he'd tell you the problems he faced as a turn-of-the-century engineer. You could tell him all about the technological solutions available to today's engineer and scientist… particularly a 20th century phenomenon that tests assumptions and defines models before a project gets off the ground. The Apple personal computer.

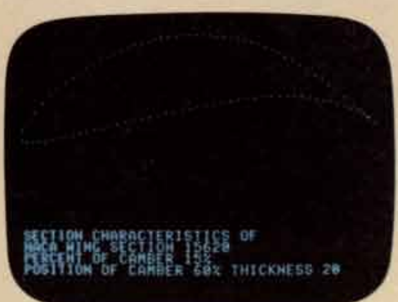
Computation, calculation, analysis…the power to pilot your projects.

With a highly-integrated system from the extensive Apple personal computer family, Orville and brother Wilbur would have increased their productivity. Perhaps even launched the Kitty Hawk Flyer well before 1903.

An Apple in their hangar would have freed them from the time and tedium of crunching numbers by hand.

An Apple in your lab or office will give you the problem-solving capabilities you demand from a big computer…without the time-consuming problems typical of remote processing.

But the Apple system solution doesn't stop there. It keeps on soaring with proven performance, power and expandability



Apple's existing software library includes a program that plots the shape of an airfoil, given its parameters.

that's unparalleled for analyzing alternative paths of design and modeling a wide variety of physical processes.

Want more memory? Depending on your choice of system, Apple has memory expandable to 64K bytes or 128K bytes. Prefer wide displays? Choose 40 or 80 characters. Need to control instruments in the lab? Get on the IEEE 488 bus. Over 100 companies also supply peripherals for Apple because Apple is the most popular personal computer with the least complicated interface.

Want an efficient system of data storage and access? Apple's 5 ¼″ disk drive not only offers you increased application versatility, but high density (143K bytes), high speed and low cost. You can even add up to four or more drives to your Apple system. With proven reliability, no wonder it's the most popular drive on the market today.

Wilbur determined that birds didn't have to constantly flap their wings to fly. With an Apple, he could've determined the fixed-wing design of the Kitty Hawk Flyer much faster.

FORTRAN that helped to design a 20th century flying machine.

Fluent in the same language that helped to design the 747, Apple FORTRAN lets you tackle differential equations at the touch of a key. And since more the 170 companies also offer software for the Apple family, you can have one of the most impressive program libraries ever…including vast subroutine libraries for math, science, engineering and statistics. When you write your own programs, the Apple also speaks in languages other than FORTRAN: Pascal, BASIC, PILOT and 6502 assembly language.

Where to learn more about Apple, the small-yet-serious solution.

Let your imagination soar with Apple.

Discover the 20th century tool versatile enough to monitor quality controls and manufacturing schedules, orchestrate tolerance tests and determine alternative parts selection. Learn why Apple emerges as the technological leader of reliable personal computer products that increase your productivity.

Let the Apple dealer show you how, by putting the system of your choice through its paces. He'll tell you about our extended warranty, support and service. And he'll prove that a personal computer is not just a flight of fancy but a serious solution. Don't let history pass you by. Visit your nearest Apple dealer, or call 800-538-9696. In California, 800-662-9238.

apple computer

Behavioral Science Book Service
1st Place

to look for the essential and exciting books in every area of the behavioral sciences

54405. HYPNOTIC REALITIES: The Induction of Clinical Hypnosis and Forms of Indirect Suggestion. *Milton H. Erickson et al.* Includes book plus one-hour cassette recording of actual hypnotic session with author's commentary. **$24.95**

48260. THE FAMILY LIFE CYCLE: A Framework for Family Therapy. *Carter and McGoldrick.* **$25.00**

39205-2. COGNITIVE THERAPY OF DEPRESSION. *Aaron Beck et al.* The long-awaited new book by Beck and his associates plus an exciting audio cassette demonstrating an actual clinical session. *Counts as 2 of your 3 books.* **$33.00**

36660-2. THE BORDERLINE SYNDROMES: Constitution, Personality and Adaptation. *Michael Stone.* This outstanding new reference is a "unique, fundamental book for the 1980's."—*Otto F. Kernberg, M.D. Counts as 2 of your 3 books.* **$27.50**

60595. MARRIAGE AND MARITAL THERAPY. *Paolino and McCrady.* An outstanding collection of original articles. Contributors include Meissner, Gurman, Sluzki, Nadelson, O'Leary. **$25.00**

36880. THE BROKEN CONNECTION. *Robert Jay Lifton.* NBA winner for his *Death in Life,* Lifton offers a new psychology for our culture. **$15.95**

41960. DEPRESSION: Comparative Studies of Normal, Neurotic and Psychotic Conditions. *Edith Jacobson.* **$22.50**

40245. CONSCIOUSNESS: The Brain, States of Awareness and Alternate Realities. *Edited by Goleman and Davidson.* **$22.50**

53535. HOMOSEXUAL BEHAVIOR. *Edited by Judd Marmor.* The most recent research findings from the biological and social sciences. **$22.50**

50335. GOING IT ALONE: The Family Life and Social Situation of the Single Parent. *Robert S. Weiss.* **$13.95**

52241-2. THE HANDBOOK OF PRIVATE PRACTICE IN PSYCHOLOGY/FINANCIAL PLANNING FOR THE INDEPENDENT PROFESSIONAL. *Counts as 2 of your 3 books.* **$33.95**

54750. IDENTITY AND THE LIFE CYCLE. *Erik H. Erikson.* Original observations by the world-famous psychoanalyst. **$10.95**

72880. THE PSYCHOPATHOLOGY OF WOMEN. *Ihsan Al-Issa.* **$15.95**

Take any 3 books for only $1 each (values to $74.95)

if you will join now for a trial period and agree to take 3 more books—at handsome discounts—over the next 12 months.

(Publishers' Prices shown)

60495. THE MANY FACES OF SUICIDE. *Edited by Norman L. Farberow.* "An important and outstanding collection of original essays dealing with unconscious self-destructive patterns that can be manifested in almost every aspect of contemporary life."—*Judd Marmor.* **$18.95**

87418. WOMEN DISCOVER ORGASM. *Lonnie Garfield Barbach.* A low-cost approach which has had a remarkable success rate. **$20.00**

70840-2. PSYCHIATRIC DICTIONARY. Fourth Edition. *Hinsie and Campbell.* The best and most comprehensive one-volume reference work in the field. *Counts as 2 of your 3 books.* **$26.50**

36163-2. BIAS IN MENTAL TESTING. *Arthur R. Jensen.* The well-known author examines standardized intelligence tests. *Counts as 2 of your 3 books.* **$29.95**

52300-2. HANDBOOK OF GENERAL PSYCHOLOGY. *Benjamin B. Wolman.* Comprehensive compilation of up-to-date facts and theories in all areas of psychology. *Counts as 2 of your 3 books.* **$49.95**

32238. ADOLESCENT PASSAGE: Developmental Issues. *Peter Blos.* **$24.95**

35738. BEFORE THE BEST INTERESTS OF THE CHILD. *Joseph Goldstein et al.* **$12.95**

62668. MID-LIFE: Developmental and Clinical Issues. *Edited by Norman and Scaramella.* Adult life phases, sexual problems, changing roles and more. **$15.00**

57915. LAW AND PSYCHOLOGICAL PRACTICE. *Schwitzgebel and Schwitzgebel.* An up-to-date, comprehensive survey of law relevant to practice and research in mental health. **$14.95**

78145. SEXUAL DISORDERS: Treatment, Therapy, and Research. *Tollison and Adams.* **$22.50**

67115. PARENT AND FAMILY THERAPY: An Integrative Approach to Family Interventions. *Alexander J. Tymchuk.* Theoretical background, assessment, practice. **$14.95**

53385. HOLISTIC MEDICINE: From Stress to Optimum Health. *Kenneth Pelletier.* Prevention and health maintenance approach, of special importance in relieving psychosocial and environmental stress. **$10.00**

83995. THERAPLAY. *Ann M. Jernberg.* "The most significant work on play therapy to appear in the past twenty years."—*Charles E. Schaefer.* **$17.95**

41963. DEPRESSION IN CHILDREN AND ADOLESCENTS. *Edited by French and Berlin.* Clinical reports and theoretical contributions. **$18.95**

42370. DISORDERS OF SEXUAL DESIRE. *Helen Singer Kaplan.* "Very simply the best book available today on the treatment of sexual problems."—*Maj-Britt Rosenbaum, M.D.* **$17.50**

87343. THE WOMAN PATIENT. *Edited by Notman and Nadelson.* A critically acclaimed volume presenting fresh new knowledge and insights on the crucial issues involved in modern female health care. Covers such vital topics as adolescent sexuality and pregnancy, surgical treatment and menopause. **$19.50**

37495. CHANGING LIVES THROUGH REDECISION THERAPY. *Mary and Robert L. Goulding.* A unique combination of theoretical understanding and clinical practicality. **$15.00**

42193. DICTIONARY OF BEHAVIORAL SCIENCE. *Edited by Benjamin Wolman.* Nearly 20,000 entries. **$24.50**

55790. INTELLIGENCE. *David W. Pyle.* **$14.95**

If reply card is missing, please write to **Behavioral Science Book Service, Dept. 1-AE2, Riverside, N.J. 08075** to obtain membership information and an application.

mining: strip mining, open-pit mining and terrace mining. These same techniques are applicable to oil mining. All three have one characteristic in common. They call for the movement of earth and rock on a very large scale, and therefore they require a major investment in heavy equipment, if unit costs are to remain economical.

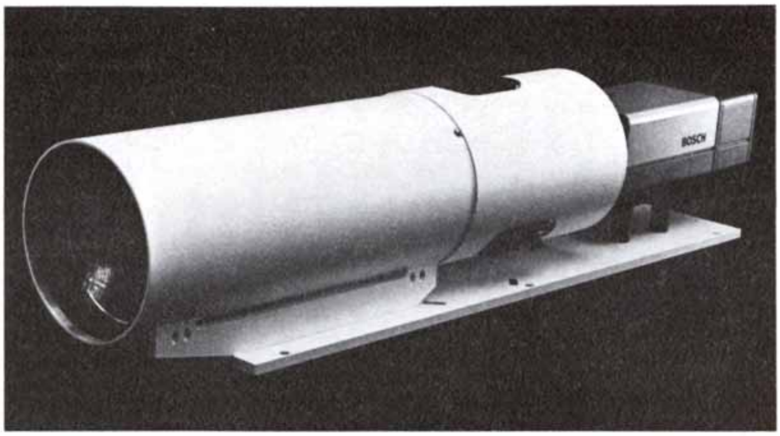
Just as modified in situ mining systems are appropriate for the exploitation of oil reservoirs 500 feet or more below the surface, the techniques of surface mining are applicable to shallower deposits. We shall describe the three methods in order. Strip mining, a technique frequently used in U.S. coal fields of the Midwest and West, begins with the removal of overburden by excavating a long, narrow trench with a dragline or a large stripping shovel. Usually the trench is 50 to 100 feet wide. It may be more than a mile long and is deep enough to expose the mineral deposit. The mineral is then removed and transported to a processing plant. Next a second trench is excavated parallel to the first. The overburden from this trench is dumped into the first trench and the mineral is removed. The sequence is repeated as the mining progresses.

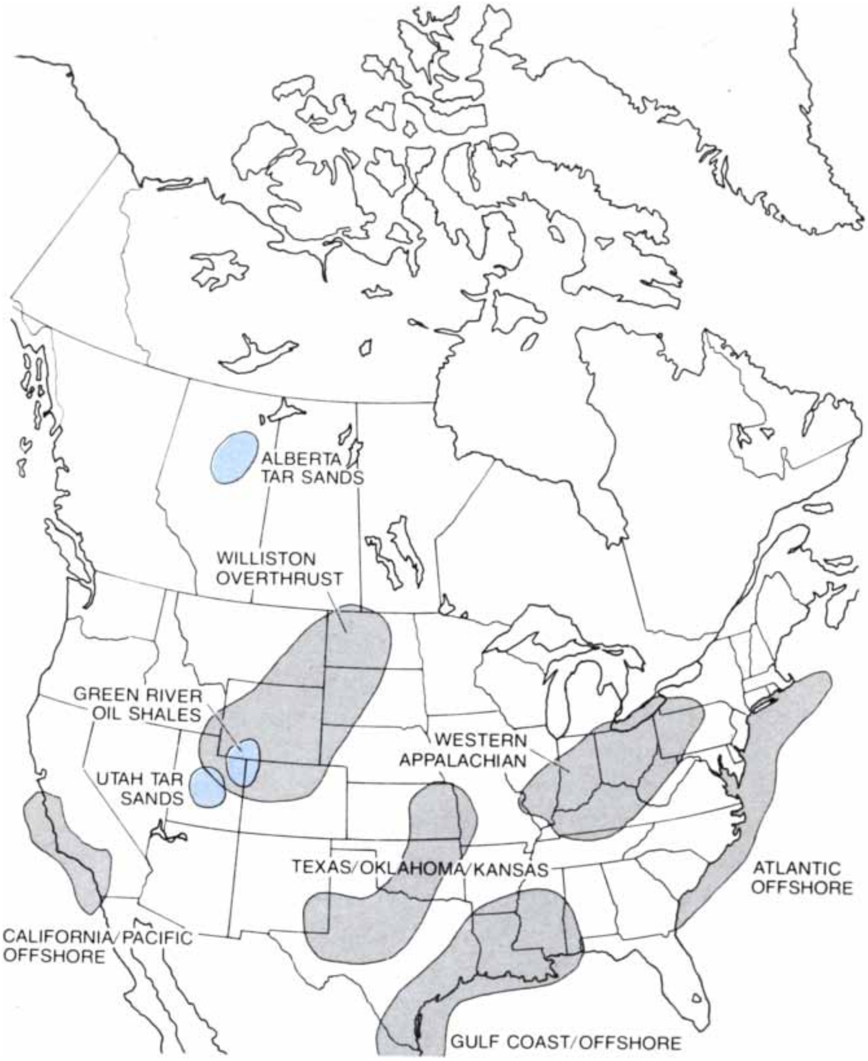
Strip mining is an efficient system for deposits at depths down to 180 feet in generally flat terrain. The reclamation of strip-mined land involves the relatively simple processes of flattening the piles of overburden, replacing the topsoil and replanting it. In many places in the Midwest strip-mined land is now routinely being returned to farming.

Open-pit mining is suited to mineral deposits with irregular surface topography and also to those that are deeper than 180 feet. Both the overburden and the mineral deposit are removed with large shovels and are usually trucked out of the pit along a system of haulage roads. Conveyor-belt systems may also be used. The mineral loads go to the processing plant and the overburden is dumped some distance from the pit. It is generally impractical to use the overburden to backfill the pit.

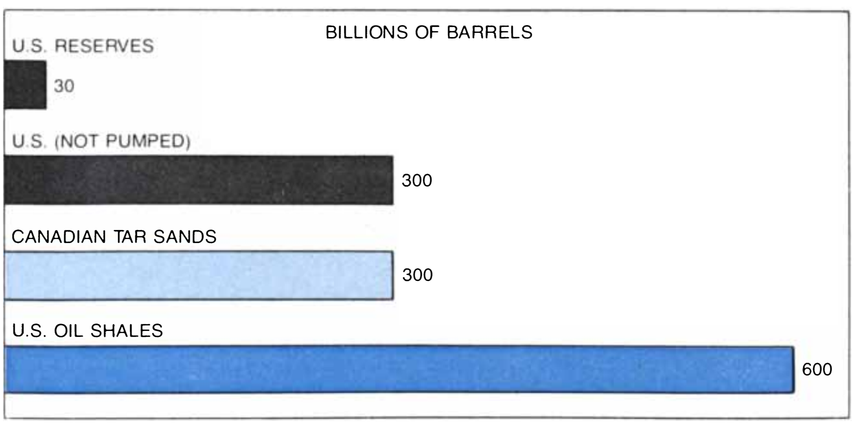
The third surface-mining technique, terrace mining, is a variation of open-pit mining practiced when the mineral deposit covers an extended area but is relatively shallow. It may be regarded as a kind of outwardly spiraling strip-mining system except that the overburden is trucked away and stored at least temporarily rather than being successively dumped back into the pit. Because a terrace mine eventually includes a very large worked-out area, however, some backfilling with overburden is usually practical.

Estimates indicate that the production of oil by surface mining in the U.S. is economically feasible. One study, with the Kern River oil reservoir in California as a model and terrace mining as the technique, suggests that the inclusive





FIVE MAJOR REGIONS of petroleum production in the contiguous 48 states of the U.S. are shown along with the still unproductive Atlantic offshore area. Also shown (*color*) are three unconventional hydrocarbon resources, the tar sands of Alberta and Utah and the oil shales of the Green River formation in Utah, Colorado and Wyoming. Each of the five regions has depleted oil reservoirs where petroleum might be mined. The most promising is California, where in addition to normal oil deposits a vast bed of diatomaceous sediments holds hydrocarbons.



ESTIMATED OIL RESERVES in the U.S., some 30 billion barrels, are compared in this bar chart with the estimated amount of oil that now remains or will be left behind in depleted oil reservoirs: 300 billion barrels. This is equal to the estimated total amount of synthetic petroleum available in the tar sands of Alberta (*light color*) and about half the amount available from the Green River oil shales that can yield more than 15 gallons of oil per ton (*dark color*).

cost of the oil would be $11 per barrel. Another study, based on open-pit mining as the technique, projects the cost at $15 per barrel with the Edna oil reservoir in California as a model. With the Santa Cruz oil reservoir in California as a model and strip-mining as the technique, the projected cost is $18 per barrel. When a modified terrace-mining system was projected for the tar-sand deposit near Sunnyside, Utah, the cost estimate came to about $21 per barrel, which is not a great deal higher than the cost per barrel of synthetic crude oil from the Canadian tar sands.

Although surface mining is somewhat safer than modified in situ mining, its environmental impact is greater. Gas, dust and noxious odors can be expected near the mines. Both the overburden and the tailings from the processing plants would present substantial disposal problems. It is, of course, possible at a price to reduce in scope or even eliminate these unfavorable consequences of surface mining, but the question of exactly what environmental restrictions may be placed on the surface mining of oil remains unanswered at present. Meanwhile the Bureau of Mines is undertaking a study to identify and quantify such environmental problems and to suggest solutions to them.

Why is no oil now being regularly mined in the U.S.? An answer that was valid in the past is that the nation could get along without oil mining. As the cost of imported oil has increased this answer has become less and less logical. Indeed, some pilot oil-mining efforts have already begun, including some oil-shale extraction. A 17-member consortium that includes many prominent U.S. oil companies recently completed a 30-month demonstration of the practicality of deriving synthetic crude oil from the oil-shale formation near Rifle, Colo. By the end of the demonstration period, late in 1978, the oil-shale retorts had produced nearly 100,000 barrels of crude oil at a rate of 25 gallons of crude per ton of shale. At another Colorado site research is in progress on oil shale that has been reduced to rubble by blasting.

Meanwhile one petroleum producer is planning to establish an open-pit mine to exploit the hydrocarbons in a 400-foot-deep formation of diatomaceous sediments in the McKittrick field in California. Another producer, working in cooperation with the Department of Energy, is investigating the prospects for recovering hydrocarbons from a tar-sand formation in Utah, utilizing in situ extraction methods. It seems safe to forecast that the future will see an accelerating exploitation by oil mining both of the nation's unconventional hydrocarbon deposits and of the substantial supplies of petroleum still present in conventionally pumped-out reservoirs.

The Buoyancy of the Chambered Nautilus

The animal trims its weight and gains mobility by dividing its shell into compartments and removing their watery content. It succeeds in spite of the ocean's pressure, which tends to drive water back in

by Peter Ward, Lewis Greenwald and Olive E. Greenwald

Before the end of the Cretaceous period some 65 million years ago life in the oceans was quite different from that of today. Among the most important large inhabitants of the seas were chambered cephalopods: mollusks that dwell in the outermost compartment of a shell they divide into compartments by secreting a sequence of septa, or walls, inside it. What gave them their dominance was their development of the ability to achieve neutral buoyancy: an overall density, or weight per unit volume, virtually equal to that of the seawater around them. In particular the chambered cephalopods evolved an organ capable of removing water from the inner compartments of their shell. With the advent of this ability (and also the evolution of a directional water jet) the chambered cephalopods were freed from the confines of their bottom-crawling ancestors and became the first of the large free-swimming carnivores in the sea.

The chambered cephalopods reached their greatest diversity during the Triassic, Jurassic and Cretaceous periods, from 225 to 65 million years ago. In the latter part of the Cretaceous their numbers began to diminish, perhaps in response to increasing numbers of a new type of mobile predator, the modern bony fishes, which maintain their neutral buoyancy in an entirely different way: they have an inflatable bladder and a gland that can fill it with gas. Whatever the reason, the extinction of the chambered cephalopods was almost complete by the end of the Cretaceous. Only the genus *Nautilus* now remains. The animals of the genus provide an opportunity to study the ancient buoyancy system. The precision of the system in the nautilus is remarkable: the difference between the weight of a mature nautilus (as much as 1,400 grams, or more than three pounds) and that of an equal volume of seawater can be as little as a gram.

The surviving species of *Nautilus* are found in the sea just outside the coral reefs that surround the islands of the tropical western Pacific. They are rarely seen near the surface, but they have been trapped at depths as great as 600 meters. Since they live below the typical range of human divers, little is known about their behavior and their ecology. Nevertheless, some observations have been made. For one thing, the stomach of a dissected nautilus usually contains pieces of bottom-dwelling crustaceans. Moreover, on the rare occasions when we ourselves have observed the nautilus in the sea the animal has been on the bottom or close to it. Finally, the nautilus is a slow swimmer; a diver can easily keep up with one. The impression is strong that the nautilus feeds by slowly foraging on the slopes of the fore reef.

The living soft parts of the nautilus consist of two principal sections, the head (more properly the head-foot) and the body. The head is covered by the hood, a tough, fleshy tissue that acts as a shield. The nautilus can retract into its shell, leaving only the hood exposed.

The most notable features of the head are its tentacles, which number more than 90, far more than there are in any other living cephalopod. Each tentacle is lodged in a sheath, into which it can be retracted and from which it can be protruded. The surface of the tentacle lacks the suckers found on the tentacles of other cephalopods. Instead it is covered with a sticky substance that aids it in holding prey.

The tentacles surround a massive pair of jaw parts that resemble the beak of a large parrot. The jaws are heavily calcified, allowing the nautilus to break up even the most heavily armored exoskeleton of a crustacean. Unlike the saliva of the octopus, a fellow cephalopod, the saliva of the nautilus contains no toxins to incapacitate a struggling prey animal. The jaws of the nautilus are its only offensive weapon. Below the tentacles and the jaws is a fold of tissue called the hyponome, which is used for locomotion: it expels a jet of water. The hyponome is actually a pair of muscular flaps that curl around each other to form a highly flexible funnel.

One other feature of the head deserves mention. Like other cephalopods the nautilus has prominent eyes, one on each side of the head, but unlike the eyes of the others, the eyes of the nautilus are poorly developed. For example, they have no lens. A tiny opening admits light, and presumably images form the way they do in a pinhole camera. Our own encounters with the nautilus suggest that its eyes may serve solely to detect changes in the intensity of light. On the other hand, the tentacles seem to bear cells that are sensitive to the presence of chemical substances. These cells may serve the animal in place of accurate vision.

The other main section of the living nautilus, the body, includes a large sac that contains the animal's organ systems. Much of the space is occupied by the systems for digestion and reproduction. The sac itself is enveloped by the mantle, a sheet of tissue that secretes the shell. The posterior mantle secretes the septa that divide the shell into compartments. The space between the mantle and the sac is open under the body, creating a large cavity that communicates with the hyponome. The cavity contains four large gills, in contrast to the single pair of gills found in all other cephalopods. The cavity also receives the exit pores for the digestive and reproductive systems.

The sexes are distinct in the nautilus. The male is slightly larger than the female because of the presence of a large organ, the spadix, which during copulation introduces a packet of sperm into the mantle cavity of the female. The females of many cephalopod species produce thousands of eggs in a year; an average female nautilus produces no more than 10 large eggs.

The shell in which the nautilus lives

is approximately a hollow cone wound tightly around itself. The shell is made of layers of aragonite, a crystalline form of calcium carbonate, in alternation with layers of a proteinaceous substance having a chemical composition resembling that of fingernail.

The internal division of the shell into chambers immediately suggests the possibility that the nautilus might change its buoyancy as a submarine does. As early as 1696 Robert Hooke proposed that "the animal has a power to fill or empty each of [the chambers] with water, as shall suffice to poise and trim the posture of his vessel, or shell, fitteth for that navigation or voyage he is to make; or if he be to rise, then he can empty these cavities of water, or fill them with air." We ourselves, however, have never found a nautilus in shallow water to refill a chamber with water; it can only empty the chambers. Moreover, it can empty them only slowly; the maintenance of buoyancy is in fact a lifelong effort. As the animal grows in its shell it builds a septum behind itself, thereby sealing off a new chamber. Each new chamber is filled at first with a watery fluid called the cameral liquid, but the liquid is slowly removed. This provides the buoyancy the animal needs to counteract the growing weight both of its living parts and of its shell.

The removal of the liquid from the chambers is accomplished by the siphuncle, a strand of living tissue enclosed in a calcareous tube that spirals from the posterior mantle of the animal through all the chambers of the shell, including the earliest chambers. At the center of the strand is a network of blood vessels. The surrounding seawater exerts a hydrostatic pressure on the body of a nautilus, and this pressure is transmitted to the blood that circulates in it. Hence the pressure of the blood in the siphuncle is equal to the blood pressure generated by the heart plus the pressure of the seawater, which increases by one atmosphere for every 10 meters of depth. At 400 meters, which is a typical depth for the nautilus, the inside of the siphuncle would have a pressure of more than 40 atmospheres, or nearly 600 pounds per square inch. In an emptied chamber, therefore, where the pressure is always less than one atmosphere, the calcareous tube surrounding the tissue of the siphuncle must keep the tissue from bursting.

One aspect of our research has been to

CHAMBERED NAUTILUS was photographed at night at a depth of 30 feet near the coral reef surrounding the South Pacific island of New Caledonia. The animal hovers almost effortlessly because its weight is virtually equal to that of the seawater it displaces. A few feet from where the photograph was taken the reef begins to drop off to a depth of several hundred feet. The nautilus may have migrated upward during the night (the time when it is active), or it may have hidden near the top of the reef during the day. Because living nautiluses are seldom seen, little is known about their behavior. The diver at the right is Pierre Laboute, a participant in the authors' research.

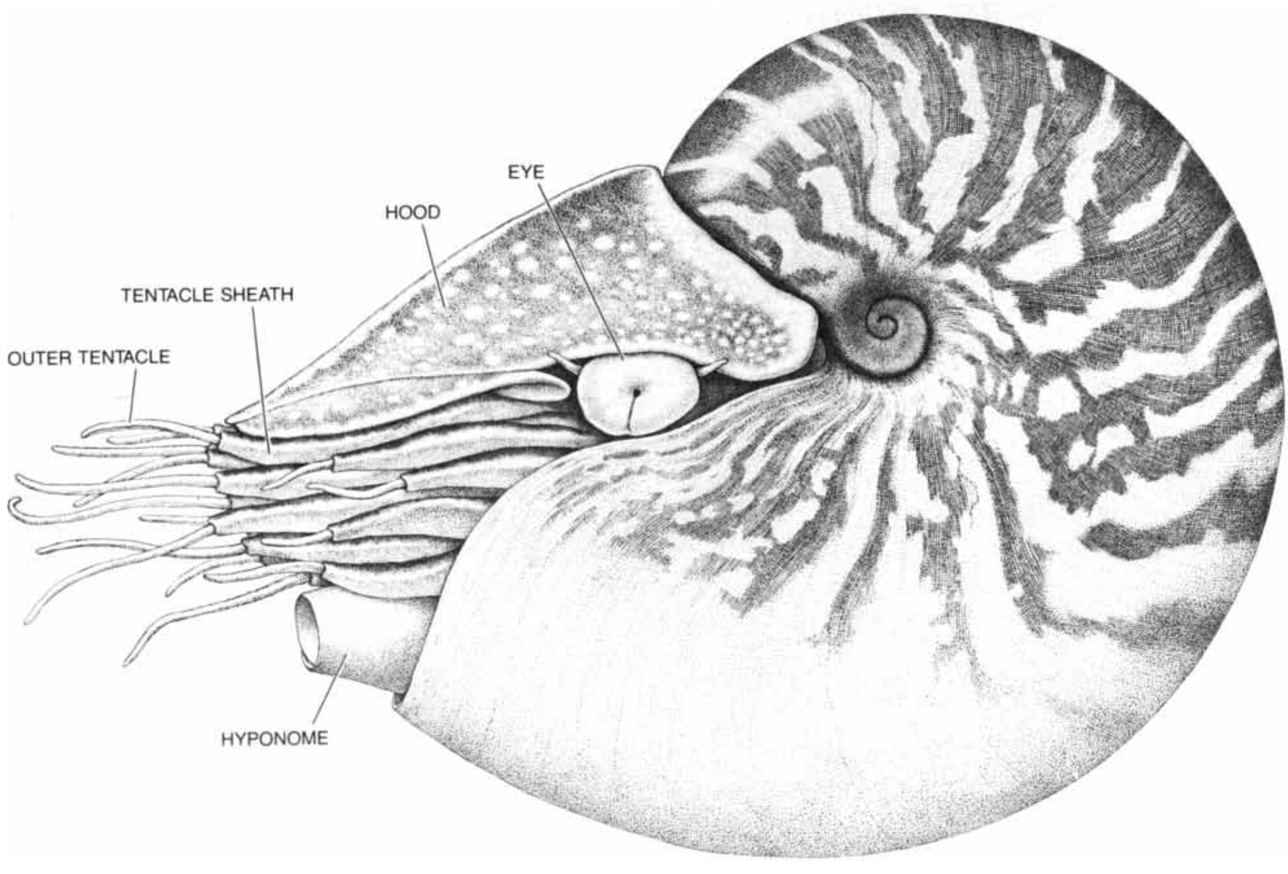
examine the mechanism by which the siphuncle removes the cameral liquid that initially fills each chamber. The liquid closely resembles nautilus blood, and also seawater, in that the concentrations of potassium, sodium and chloride ions are virtually the same in all three. It is therefore impossible to say with certainty whether the liquid is a filtrate of blood, a secretion of the siphuncle or of the mantle or perhaps seawater the animal has modified. Nevertheless, if siphuncular cells were to transport some of the ions from the liquid into the blood vessels at the center of the siphuncle, the concentration of ions in the liquid would fall below that in the blood. As a result the water in the liquid would flow by osmosis from the chamber into the blood. It could then be removed from the blood by the animal's kidneys.

The trouble with this hypothesis is that the hydrostatic pressure deep in the ocean favors the tendency for water to pass from the siphuncle back into the chamber. Indeed, even if the siphuncle could reduce the concentration of ions in the cameral liquid to zero and thereby maximize the osmosis of water out of the chamber, a calculation shows that at a depth of just over 240 meters the hydrostatic pressure becomes great enough to drive water in the opposite direction. Since nautiluses are found with completely empty chambers at depths as great as 600 meters, we were prepared to rule out the simple osmotic mechanism.

We also wanted to rule it out experimentally. Experiments in the open ocean were not practicable. Instead we drilled a small hole through the shell of the nautilus, piercing a nearly empty chamber. We removed any liquid that was present and then added five milliliters of a solution whose concentration of ions was greater than that of nautilus blood. If the movement of water across the siphuncle were governed by simple osmotic forces, water would diffuse from the siphuncular blood into the artificial cameral liquid. Actually we observed the opposite. The siphuncle was capable of removing the liquid from the chamber even when the concentration of ions in the liquid was almost twice that of the blood.

At this point we turned to the work of Jared M. Diamond. In the 1960's Diamond and his colleagues at Harvard University demonstrated that the gall bladder of the rabbit could transport water against both osmotic and hydrostatic gradients. The water passed from the lumen, or hollow, of the bladder into the blood vessels in the bladder wall. The investigators then showed that the cells forming the wall of the bladder abutted one another at the surface facing the lumen, but that there were spaces between the cells at the places in the wall of the bladder where the cells were closest to blood vessels.

The Harvard workers proposed that the cell membrane bounding each space might include enzymes that transport ions across the membrane. These molecular pumps could build up a high but local concentration of ions in the spaces between the cells. A mathematical anal-



GROSS ANATOMY of *Nautilus macromphalus*, the species studied by the authors, is displayed in these two views. The view at the left shows the features that are visible externally. The head of the animal protrudes from the shell. It has more than 90 tentacles that are arrayed in two whorls on each side of the mouth. The eyes are prominent but the vision they provide is not acute. The head is protected by a tough fleshy hood. The view at the right shows the internal structure of both the shell and the living animal. The outermost living part is the mantle, a tissue whose secretions enlarge the shell. The posterior mantle secretes the succession of septa, or walls, that divide the shell

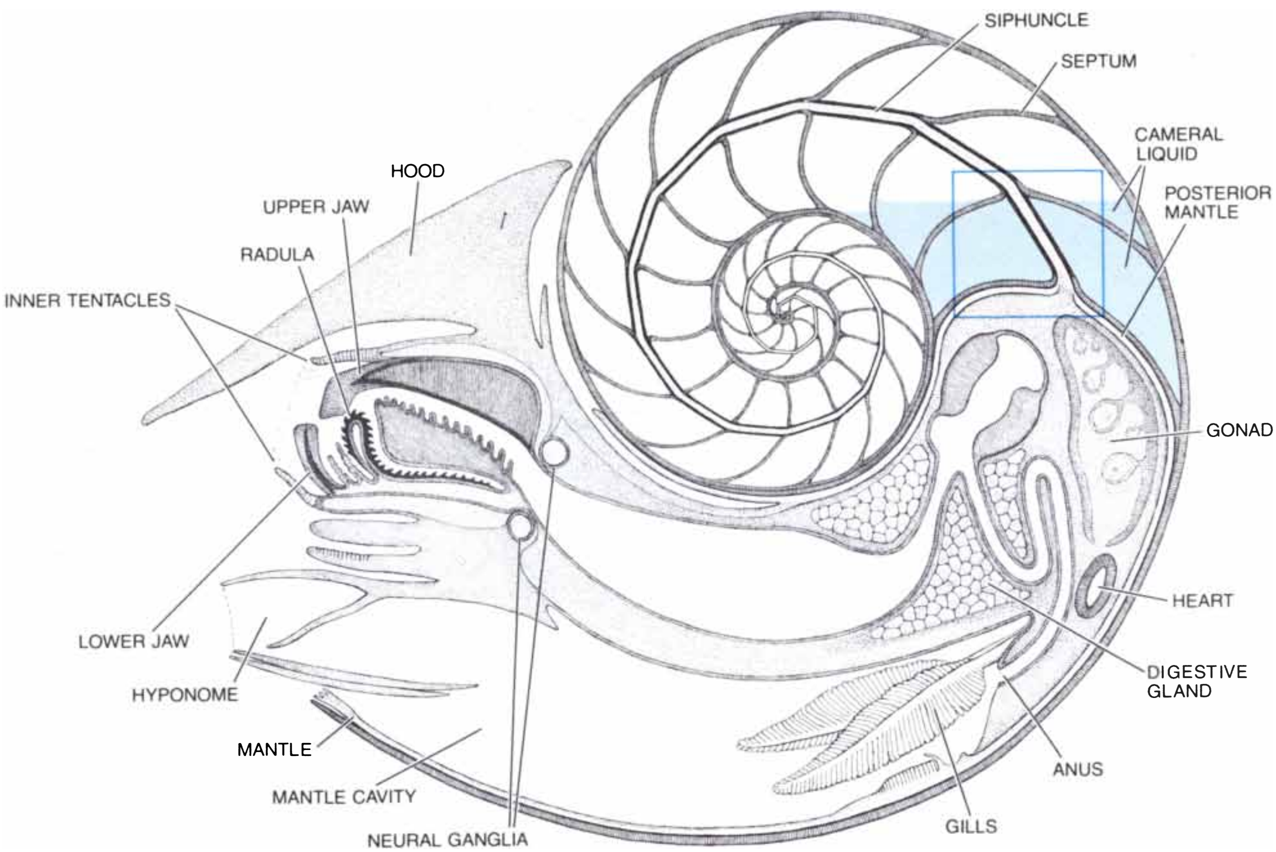
ysis showed that under these conditions water would diffuse from the lumen of the bladder into the cells forming the wall of the bladder and from there into the regions of high ion concentration in the spaces between the cells. If the ion pumps were in the cell membrane along the blind ends of the intercellular spaces, the water that entered the spaces would flow toward the open ends of the spaces against either osmotic or hydrostatic pressure. Ultimately it would enter the blood vessels.

Because this hypothesis involves osmosis into small regions in a cell layer it is called a local-osmosis model. The significance of the model with regard to the nautilus is that it could account for the ability of the siphuncle to transport cameral liquid against either osmotic gradients (such as we produced in the laboratory) or the hydrostatic gradients an animal faces deep in the ocean. The question is whether the siphuncle has a cellular geometry consistent with local osmosis.

Evidently it does. In 1966 Eric J. Denton and John B. Gilpin-Brown of the Marine Biological Association of the United Kingdom at Plymouth published drawings of the siphuncular tissue that bore a striking resemblance to Diamond's low-power electron micrographs of gall-bladder tissue. Denton and Gilpin-Brown later suggested that local osmosis might function in the siphuncle. In our own examination of siphuncular tissue we too have noted that segments of siphuncle taken from chambers that were emptying or had already emptied have prominent intercellular spaces. Our electron micrographs suggest to us, however, that the large intercellular spaces may serve merely as drainage channels for fluid from a much finer set of canals leading into the larger spaces. Our reason for proposing that the cameral liquid is transported first into the finer set of canals is that they are lined (in the cytoplasm surrounding the canals) by numerous mitochondria, the intracellular organelles supplying energy to a cell. We hypothesize that in this case the mitochondria provide energy for the transport of ions or some other solute.

Tissue taken from the part of a siphuncle that traverses a chamber whose septum is still being built lacks both large and small intercellular spaces. On the other hand, the part of the siphuncle in chambers that have been emptied maintains its network of spaces. Hence the siphuncle appears to be capable of bailing out those chambers, which might otherwise be filled by water forced inward by the substantial hydrostatic pressure at depths greater than 240 meters.

In the emptied chambers the liquid has been replaced by a gas that is essentially air from which most of the oxygen has been removed and into which some extra carbon dioxide has been secreted, presumably by the siphuncular cells. Denton and Gilpin-Brown found that when a chamber is being emptied, the gas pressure inside it can be as low as .1 atmosphere. This led them to speculate that the gas is not transported into the chamber by a process requiring en-



into chambers. The siphuncle, a strand of living tissue encased in a calcareous tube, extends from the posterior mantle in a spiral that takes it through every chamber. The animal's two pairs of gills are deployed in a cavity between the mantle and the viscera. The cavity connects with the hyponome, a flexible funnel through which water is expelled in a jet for locomotion. When a nautilus of the species *N. macromphalus* is fully grown, it is several inches across and weighs nearly a kilogram. At that stage the animal's shell has 30 or more chambers. Colored rectangle in the view at the right marks the part of the shell that appears in the illustration at the top of page 195.

tech talk:

Quartz phase lock circuitry

explained.

When you get right down to it, the most important function a turntable performs is turning. Turning your records at a specific and constant rate of speed.

Simple, right?

Wrong.

The complicated part lies in the word *constant.*

Inconsistencies in platter rotation cause distortion. One form of distortion is caused by variations in platter speed. If the variations occur at a slow speed, the distortion is called "wow." If they occur at a rapid rate, it's called "flutter."

The solution to this problem brings us to our tech talk and the incredible MCS Series® 6720 Quartz lock direct-drive turntable.

The key to both the tech talk and the 6720 is quartz. An oscillating quartz crystal to be exact. And when it comes to timing, that's about as exact as you can get.

This crystal is part of a circuit called a phase lock loop or P.L.L. Part of this circuit is called a divide by. As its name implies, it divides. It divides the frequency of the crystal into a frequency that matches the speed at which the platter should revolve. Another portion of this circuit, called the comparator, continuously measures and compares the speed of the turntable to the divided frequency of the crystal. At the slightest sign of deviation the P.L.L. instantly and automatically makes the necessary corrections.

The result: incredible platter speed accuracy and amazingly low wow and flutter.

To add to this avant-garde technology, the 6720 also offers a brushless DC servo motor, a front control panel for operation without lifting the dust cover, a damped pushbutton cueing control and an Audio-Technica cartridge.

But judge for yourself. Come to your nearest JCPenney and see the incredible MCS Series® 6720 Quartz lock turntable. We think you'll agree that record-playing technology just took a turn for the better.

The MCS Series® 6720 Quartz lock direct-drive turntable. Only $219.95.* Only at JCPenney.

Full 5-Year Warranty MCS® Speaker. Full 3-Year Warranty MCS® Components. Within ten days of purchase of MCS® speakers, single or multiple play turntables, receiver, tuner, amplifier, tape deck, headphone, or graphic equalizer we will at our option, repair or replace any item of this modular component system that is defective in material or workmanship. Just return it to a JCPenney store with the original packing, carton, and accessories.

For the balance of the 5-year warranty on MCS® speakers and the 3-year warranty on MCS® components we will repair any item that is defective in material or workmanship. Parts and labor are included. Just return it to a JCPenney service center for service.

*Prices higher in Alaska, Hawaii and Puerto Rico.

MCS® Series

IT MAKES EVERYTHING CLEAR.

Sold exclusively at JCPenney.

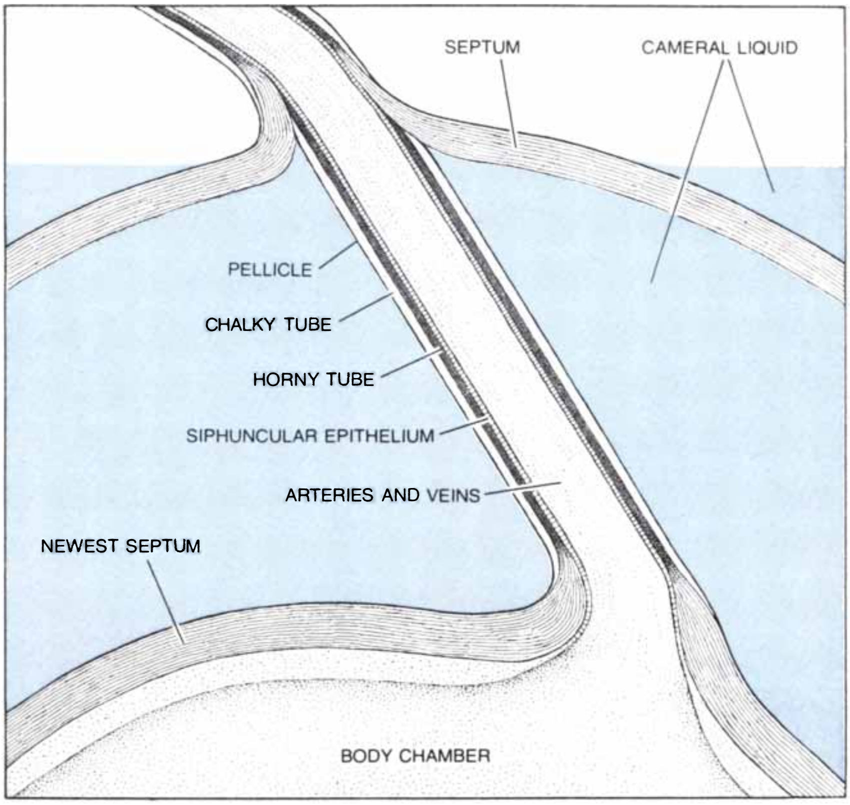
ergy. Perhaps it simply leaves the siphuncular blood by diffusion.

Because the seawater in which the nautilus swims is in chemical equilibrium with the atmospheric gases at the surface of the ocean, and since nautilus blood is in chemical equilibrium with seawater, it is not surprising that the pressures of nitrogen and argon in the chambers filled with gas are the same as the pressures of those gases in sea-level air. In any case the gas does not contribute to the buoyancy of the animal. Actually it slightly decreases the buoyancy, because the gas after all has weight whereas a vacuum weighs nothing. The point of putting gas in a balloon is to establish the balloon's size and shape. Those functions are served in the nautilus by the rigid shell itself.

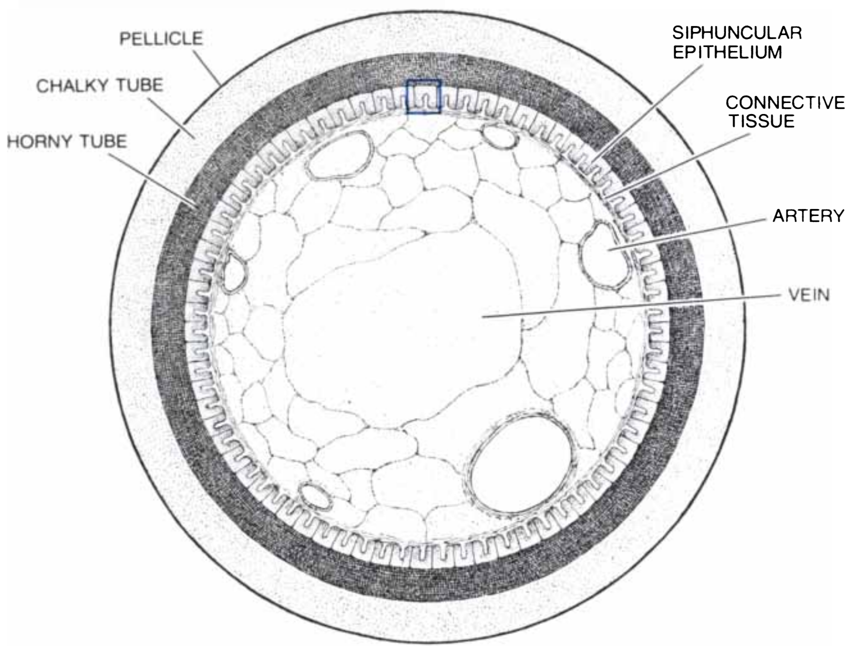
A nautilus weighs from one gram to five grams in seawater, regardless of its size. Hence the animal maintains its neutral buoyancy as it grows. It does so by keeping two rates in balance: the rate at which its growing shell and soft parts add weight and the rate at which it empties the shell's chambers of their liquid. To understand how these rates are related Desmond H. Collins of the Royal Ontario Museum, G. E. G. Westermann of McMaster University, Arthur W. Martin, Jr., of the University of Washington and we, in a variety of studies, have considered various aspects of shell growth in the nautilus.

X-ray images of a growing nautilus show that the growth involves a sequence of processes. First the part of the mantle investing the forward part of the body at the open end of the shell secretes new shell material. The open end thus grows both longer and wider, thereby enlarging the body chamber for the growing soft parts of the animal. When the open end of the shell has got long enough, the animal moves forward in the shell, leaving a volume of cameral liquid behind. The mantle forms a seal between the animal and the walls of the body chamber. A thin layer of a slimy substance may complete the seal.

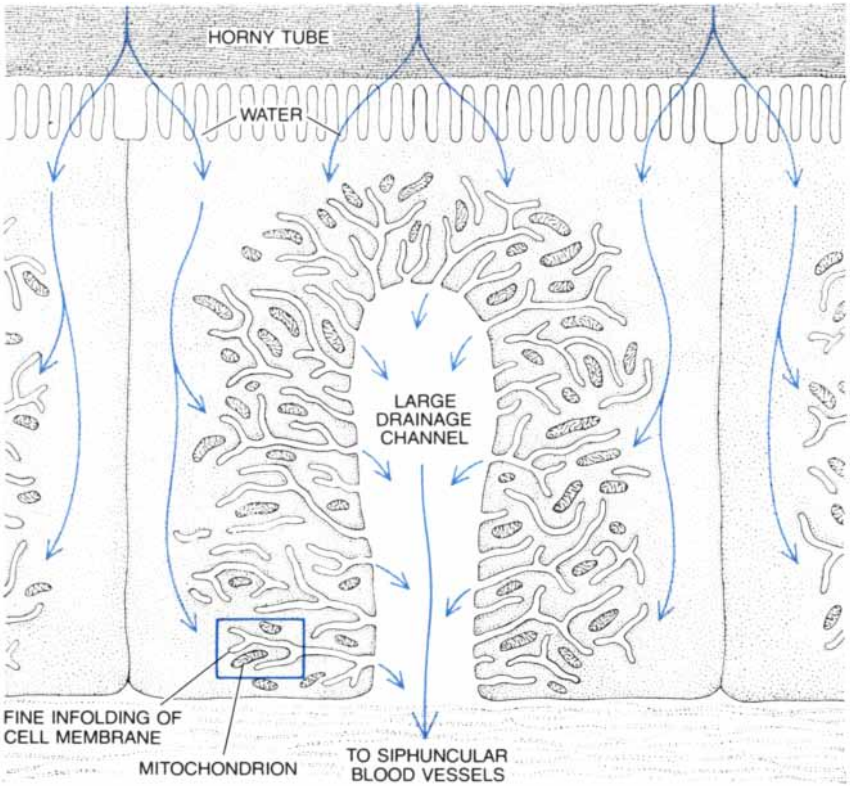
Next the posterior mantle secretes a new calcareous septum, starting at the edges of the shell and progressing inward toward the centrally located siphuncle. At the same time the siphuncle secretes its own calcareous covering. The cellular geometry of the siphuncle in the newly formed chamber is not yet of the type we hypothesize is appropriate for the emptying of a chamber by local osmosis. At this stage, however, the newly forming septum is not yet fully thickened, and the liquid in the newly formed chamber is acting as a brace against the pressure of the surrounding seawater, which is transmitted to the septum through the body of the nautilus. If the liquid were removed be-



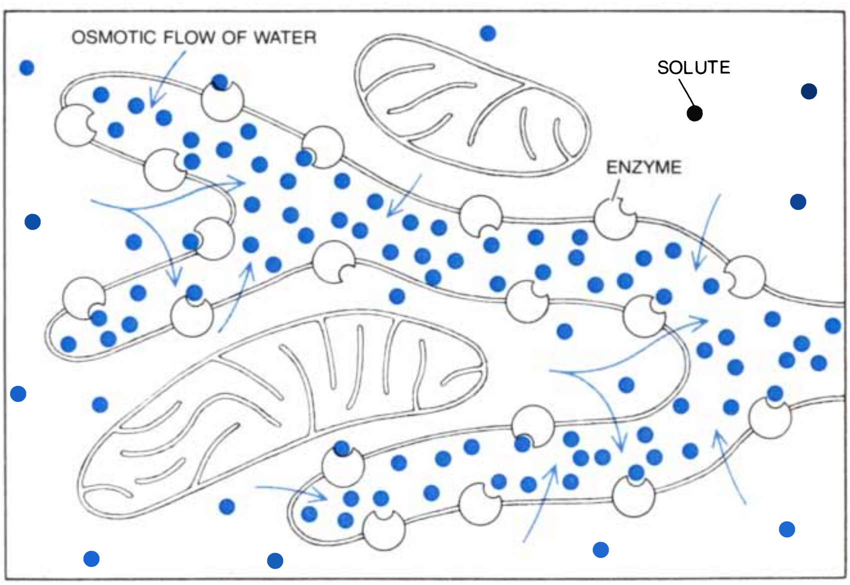
NEWEST CHAMBER in the shell is the one that lies behind the living animal and whose floor is formed by the newest septum. At first the chamber is filled with a watery fluid called the cameral liquid, but the liquid is slowly removed. Hence as the animal grows it adds buoyancy to counter the increasing weight both of its living parts and of its shell. At the time the newest chamber begins to empty, the next-newest chamber still contains a few milliliters of liquid.



SIPHUNCLE, the organ that spirals through the chambers of the shell and removes the liquid from them, is shown in cross section. The outer layers of the organ are porous and calcareous. They encase living tissue: an epithelial layer (that is, a layer of closely packed cells) inside of which is a core consisting of arteries and veins. The pressure of the blood in the siphuncle corresponds to the pressure of the sea around the animal. In an emptied chamber, therefore, the outer layers of the siphuncle must keep the inner tissues from exploding. The rectangle in color shows the part of the siphuncle that appears in the illustration at the top of the next page.



EMPTYING OF A CHAMBER requires that water molecules pass through the porous outer layers of the siphuncle and then through the cells of the siphuncular epithelium. The membrane of each such epithelial cell forms a network of fine infoldings. The infoldings communicate in turn with a larger extracellular space that may serve as a drainage channel, allowing the water to enter the blood vessels at the center of the siphuncle. The water can later be removed from the blood by the kidneys of the animal, which then excrete the water into the mantle cavity. The rectangle in color shows the part of an epithelial cell that appears in the illustration below.



DETAILS OF THE MECHANISM by which the authors propose that the chamber is emptied appear in this schematic drawing of the network of fine infoldings in the membrane of a cell of the siphuncular epithelium. The emptying begins when substances dissolvea in the cameral liquid (*colored dots*) diffuse into the siphuncle. Enzymes on the membrane of the siphuncular cells then pump the solute into the infoldings. The energy for the pumping is supplied by mitochondria, the intracellular organelles that make energy available to cells. When the concentration of the solute in the network of infoldings is sufficiently great, water enters (*arrows*) by osmosis in spite of the pressure deep in the ocean, which tends to drive water back out of the siphuncle.

fore the septum were sufficiently thick, the new chamber would collapse. Moreover, the siphuncle would explode from the pressure of the blood inside it if the cameral liquid were absent while the siphuncle lacked its calcareous sheath.

When the septum is complete, the structure of the siphuncle has become the one we think is appropriate for local osmosis, and the chamber begins to empty. When the chamber is full, it contains as much as 30 milliliters, or about an ounce, of liquid. Meanwhile the next-newest chamber, and sometimes the one or two before that, contains a few milliliters of liquid. The older chambers are empty. In chambers where the level of the liquid has fallen below the siphuncle the cameral liquid can nonetheless be removed. It can reach the siphuncle by capillary action along the inner surface of the shell.

The removal of liquid from the newest chamber and from the adjacent chambers adds buoyancy to the nautilus at a rate that counters the decrease in buoyancy due to the addition of weight to both the shell and the animal's tissues. This relation is apparent when the length of the body chamber is graphed against the volume of cameral liquid in the chamber just behind it. Immediately after the chamber is formed the body chamber is relatively short and the volume of cameral liquid is maximal. As the body chamber lengthens, the volume of cameral liquid decreases. The cycle begins whenever a new chamber is formed.

The nautilus does not, however, grow indefinitely large. Like many animals it reaches a certain final size, coincident with the attainment of sexual maturity. In the nautilus the approach of maturity is marked by a reduction in the spacing between the final two or three septa. The reduction is quite variable from one animal to another, and it may reflect an ability of the nautilus to make final adjustments to its buoyancy. In any case the crowding of septa can be seen not only in the nautilus but also in the remains of many fossil chambered cephalopods. It follows that many or even all chambered cephalopods had definite limits in size and were not ever-growing.

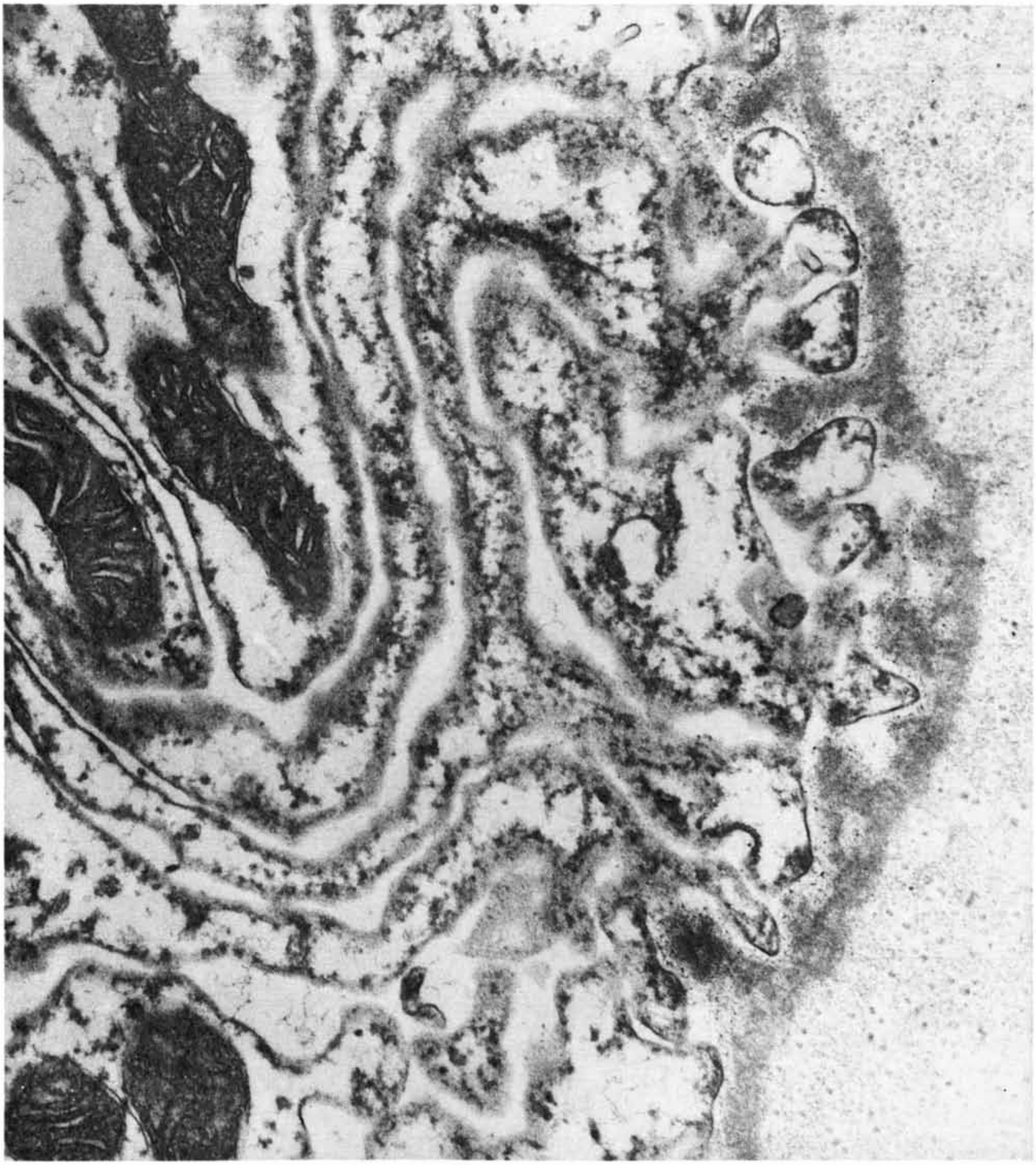
When the final septum is complete, the nautilus enters on the final processes that act to trim its buoyancy. The open end of the shell again is enlarged, and the body of the animal fills the space. Meanwhile the last of the cameral liquid is being removed from the newest of the chambers. Growth has now ended. The animal with its shell is several inches across. The shell has 30 or more compartments. The animal is thought to be at least three years old. The enlarging of the shell and the secretion of septa in it has led up to this point: the attainment of neutral buoyancy in the fully grown

SCIENCE/SCOPE

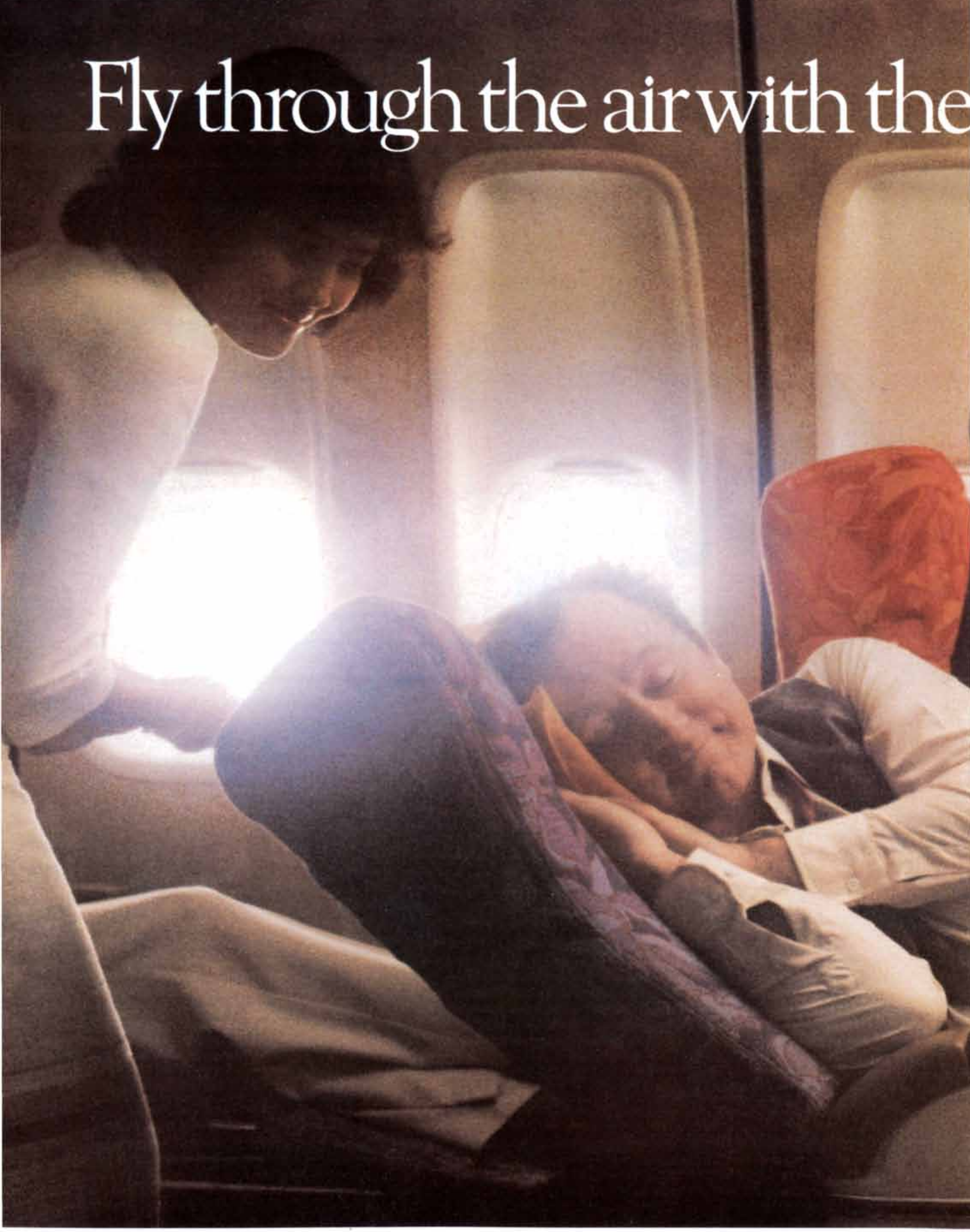
nautilus. The buoyancy will remain neutral if the siphuncle continues to bail out the shell against the hydrostatic pressure of the ocean. How long the animal lives is not known.

We have not yet learned whether there is a biological feedback mechanism in the nautilus that keeps the growth of the shell and the emptying of the chambers in phase. If there is, we will find, for example, that the animal is inhibited from starting to make a new chamber when we prevent the previous chamber from emptying. We have already noted that when the level of the cameral liquid in the newest chamber falls below the level of the siphuncle, the secretion of a new septum begins.

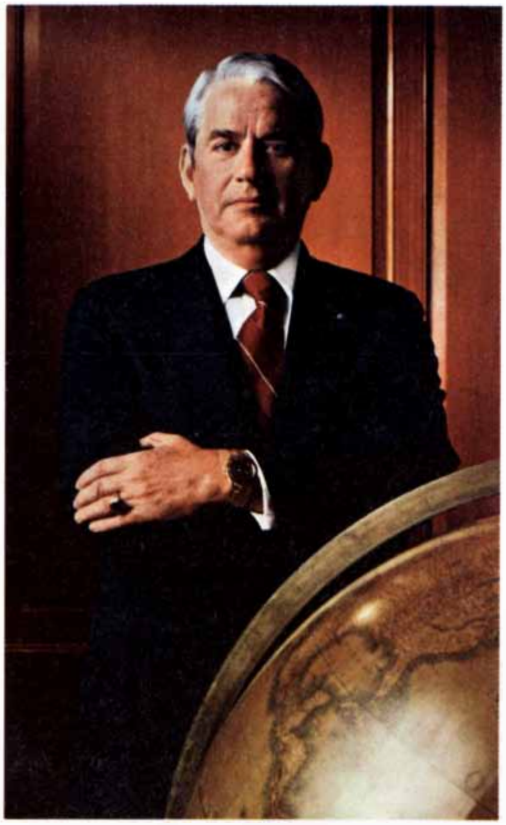
Moreover, we have shown that the nautilus can compensate for artificial changes in its buoyancy. In our earliest experiments we had been confronted with the problem of rates of emptying so low that they could barely be



ELECTRON MICROGRAPH of a siphuncular cell confirms that mitochondria congregate near the fine infoldings. The mitochondria are the darkest objects; five appear in whole or in part in the left half of the image. The larger extracellular drainage channel with which the infoldings communicate runs from top to bottom at the far right. The magnification of the micrograph is 40,000 diameters.


Fly through the air with the

His business is world travel. His watch is Rolex.



From bomber squadron duty in the flak-filled skies of World War II to the challenging office of Chairman and Chief Executive Officer of Pan Am, William T. Seawell's firm hands at the controls have proved over and over again his qualities as a leader of men.

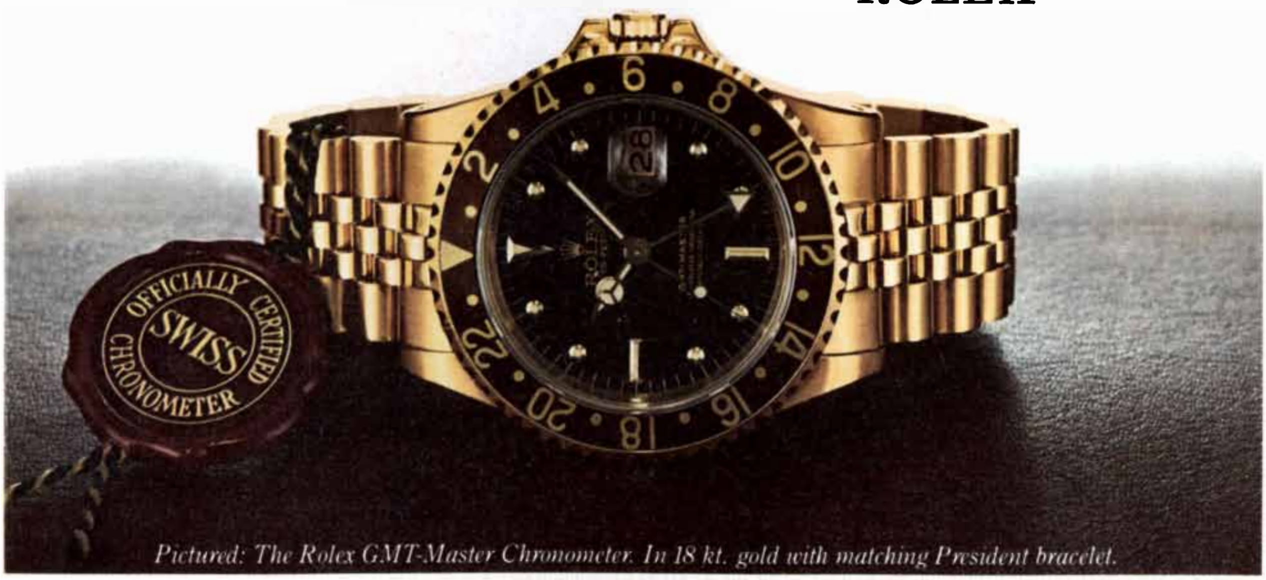
Leadership. It's what it took to accomplish the largest airline merger in history. Leadership. It's what sets Bill Seawell's sights on horizons more distant than most can see. His personal attention has meant billions of dollars in expanding tourism and international commerce.

His calm, take-charge style prevails, whether as an Air Force general or an airline executive. His manner is precise, exact...like the Rolex he chooses to wear.

Precision and exactness are Rolex guidelines. What you might expect from a watch that is handcrafted in 192 individual steps from a solid block of gold or stainless steel. The unique Oyster case is world famous for its unparalleled durability. The handsome movement world renowned for its reliability.

Pan Am: An airline that goes to the world's most prestigious places. Rolex: A watch that goes on the world's most prestigious wrists.

ROLEX

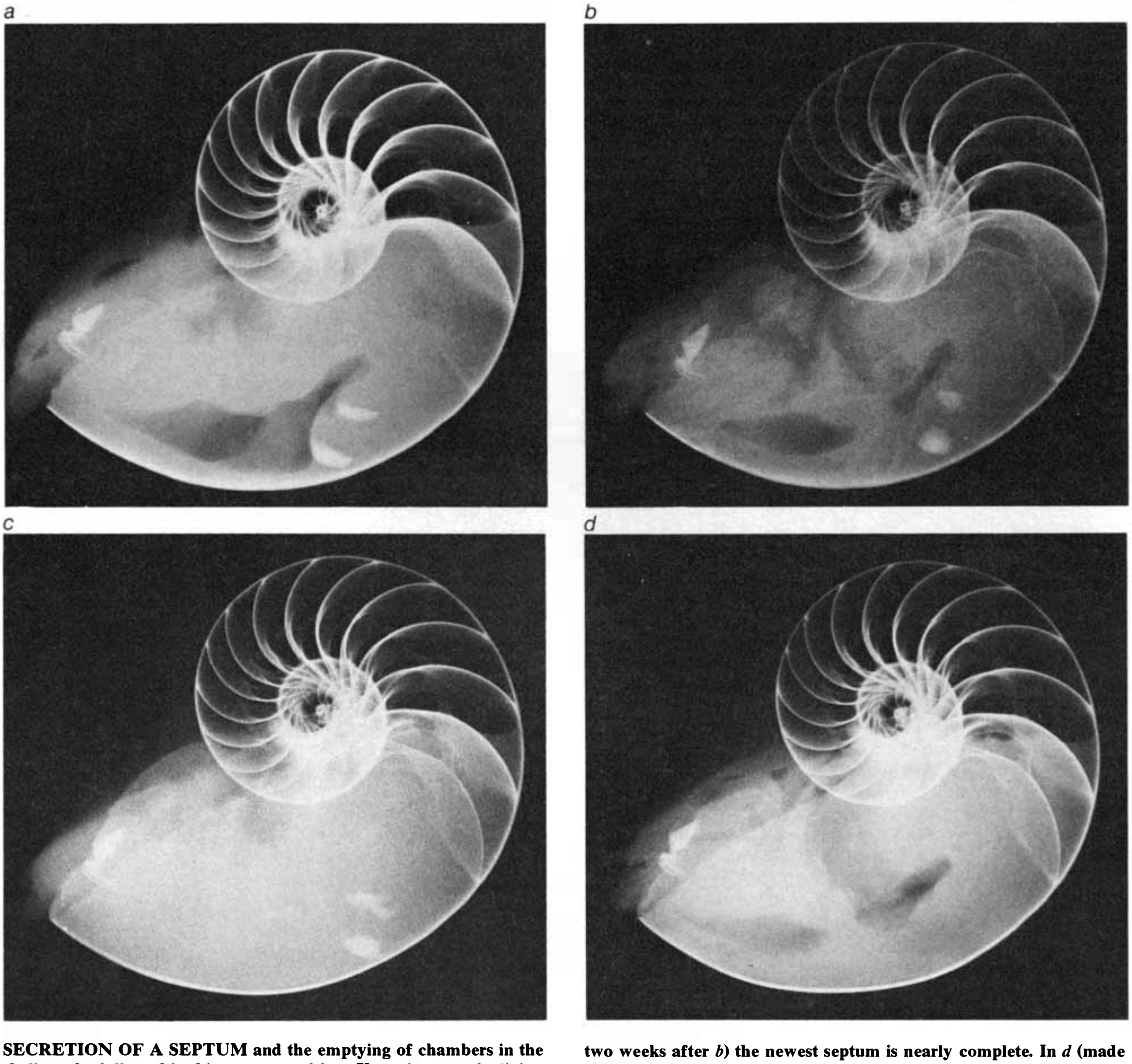


Pictured: The Rolex GMT-Master Chronometer. In 18 kt. gold with matching President bracelet.

Write for brochure. Rolex Watch, U.S.A., Inc., Dept. 115, Rolex Building, 665 Fifth Avenue, New York, NY 10022. World headquarters in Geneva. Other offices in Toronto and major cities around the world.

measured. We therefore gave our animals an "incentive" to empty the experimental chamber (the chamber in which we made measurements) by adding liquid to a second, empty chamber, thereby increasing the animal's weight in seawater. In all the animals made heavy by flooding a chamber we observed a much more rapid emptying of the experimental chamber. When the same animals were later made abnormally buoyant by emptying the flooded chamber, we observed a depression in the emptying rate. The most rapid emptying we observed in the laboratory was less than a milliliter per day, a value consistent with the estimate that the animal needs at least a month to empty a chamber.

Since in all our observations of the nautilus in shallow water we have never noted an animal adding liquid to a previously emptied chamber, we conclude that it has only the option of increasing or decreasing the rate of chamber emptying (and perhaps the rate of shell formation). Why then does the animal maintain its living siphuncle back to the very first chamber? Its life would be more efficient if it had developed instead a mechanism for permanently making each chamber waterproof after it has been emptied. Perhaps the part of the siphuncle that extends to the early chambers is the vestige of an organ that could flood and empty chambers for the ancestral chambered cephalopods. Those creatures, now extinct, would then have been far more mobile than the ones into which they evolved.



SECRETION OF A SEPTUM and the emptying of chambers in the shell can be followed in this sequence of four X-ray images of a living nautilus. In *a* the level of the cameral liquid in the newest chamber has fallen below the base of the siphuncle in that chamber. A much smaller amount of liquid remains in the chamber above it. All the other chambers are empty. In *b* (made three weeks after *a*) a new septum is faintly apparent. The chamber it defines is filled with cameral liquid. Meanwhile the level of the liquid in what had been the newest chamber has continued to fall. Although the liquid in that chamber no longer makes contact with the siphuncle directly, it can reach the siphuncle by wicking along the inner surface of the shell. In *c* (made two weeks after *b*) the newest septum is nearly complete. In *d* (made two weeks after *c* and seven weeks after *a*) the newest septum is complete. Two dark regions at the upper left of the newly formed chamber are gas bubbles. Their presence shows that the newest chamber is now being emptied. Presumably the newest septum has become strong enough to withstand the pressure of the ocean without the hydrostatic brace the cameral liquid provides. In each image the living nautilus occupies the body chamber at the bottom of the shell. Two whitish spots at the right of the body in *a* are kidney stones. They have disappeared in *d*. The white lines at the left in each image are the jaws of the animal. The dark structures are part of the mantle cavity.

THE AMATEUR SCIENTIST

A homemade mercury-vapor ion laser that emits both green and red-orange

by Jearl Walker

Dean Morelli of Rye, N.Y., whose homemade spectrophotometer I described in this department in January, has devised a way for an amateur to build a mercury-vapor ion laser. Morelli's laser emits bright pulses of green light (wavelength 567.7 nanometers) and dimmer pulses of red-orange light (615 nanometers). This type of laser is distinctive in several respects. It was the first gaseous-ion laser. Its gain and the width of its beam are unusually large for a laser emitting visible light. (Indeed, it was with this type of gas laser that high gain and high power were first demonstrated in the visible range.) The laser's emissions near 615 nanometers are almost purely monochromatic, that is, its range of frequencies and wavelengths in that region is very narrow, so that it could serve as a standard of frequency. Construction of Morelli's laser is in many respects easier than building a helium-neon laser.

The gain of a laser is a measure of how much the intensity of light is amplified in the laser (through stimulated emission) compared with the losses (through scattering, diffraction and absorption). In the green emissions Morelli's laser probably has a gain of about 50 percent per meter along the length of the laser tube. Between the mirrors of the laser the intensity of the light is quite high. Several experiments in spectroscopy can be done by putting the object of study inside the cavity of the laser, where it intercepts part of the intense light. (The part of the beam that emerges from the laser can still be utilized in other laser-beam experiments.)

The tube of Morelli's laser contains mercury vapor. The laser is electrically excited by a high-voltage discharge through the tube. When an electron in the discharge collides with an atom of mercury, an electron is removed from the atom and the result is a singly ionized mercury ion that is at an excited energy level. The mercury ions are excited into a variety of energy levels, from which they spontaneously drop to lower levels, emitting light as they do so. This light is the glow—not the beam—from the laser tube. Some of the ions will be at a level of energy that is important to the laser process (the level labeled $5f^2F^o{}_{7/2}$ in spectroscopic notation). An ion at this level can emit a photon at a wavelength of 567.7 nanometers if it drops to the lower energy level $6d^2D_{5/2}$.

A few of the photons thus emitted go down the axis of the laser tube. As such a photon passes other mercury ions at the $5f^2F^o{}_{7/2}$ level the electric fields of the photon can stimulate those ions to make the same downward transition. With each transition another photon is emitted; it is identical with the passing photon in phase, associated wavelength and direction of travel. By the time the initial photon reaches the end of the tube it has stimulated the emission of a large number of identical photons.

The picture of stimulated emissions is usually made clearer at this point by switching from a description of light in terms of photons to one in terms of waves. The light waves generated by the stimulated emission are said to be coherent because they travel in the same direction, have the same wavelength and are in step. They therefore interfere constructively, thereby augmenting one another, and so they yield a wave of large amplitude, which means that the light is bright.

The waves are reflected back and forth through the tube of mercury vapor by the mirrors built into the laser, stimulating still more emissions at the wavelength of 567.7 nanometers. Part of the light leaks through the mirrors to form the external laser beam. What happens to the rest of the light emitted by the mercury atoms in the tube, both the unionized ones and the singly ionized ones that do not participate in stimulated emission? Each of those atoms either is at the wrong energy level or emits light spontaneously before being stimulated by a passing photon. The light from such atoms goes outward from the tube in all directions and merely contributes to the glow of the tube. This light is incoherent, that is, the emitted waves are not all at the same wavelength, do not go in the same direction and do not proceed in step.

Most of the energy levels attained by the singly ionized atoms of mercury do not participate in the laser action because on the average the ions stay too briefly at those levels. If a particular transition between energy levels is to contribute to laser action, more atoms must be in the upper level than are in the lower one. This situation is called a population inversion.

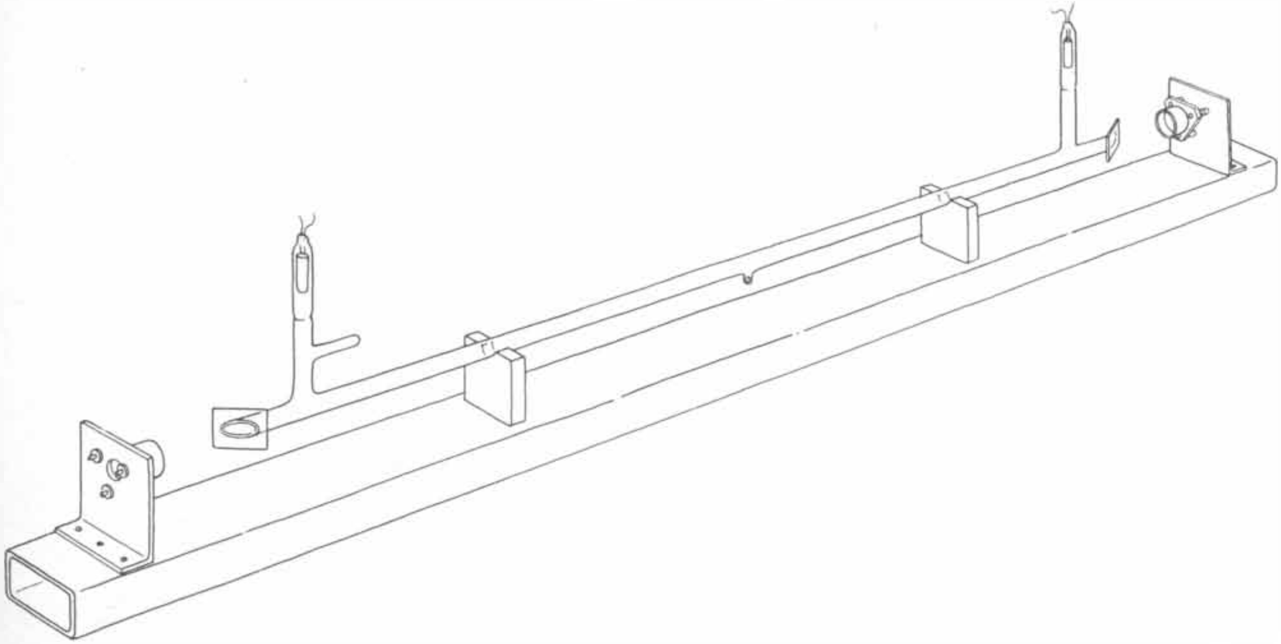
Suppose population inversion is absent, so that there are more atoms in the lower level of a pair of levels than there are in the upper one. A fortuitous downward transition by an atom could still send a photon down the laser tube, and the photon could cause a stimulated emission if it passes another mercury ion in the upper energy level. Without inversion, however, the stimulated photons are likelier to encounter lower-level atoms than upper-level ones. A passing photon meeting a lower-level atom will be absorbed by it. Absorption dominates, and the result is an absence of laser action.

Population inversion is possible if there is an energy level at which an atom can stay for a long time compared with its lifetime at a lower level to which it can make a transition. The atoms at the lower level drop to still lower levels quite quickly, leaving the level they vacated empty and the topmost level relatively full; the result is a population inversion.

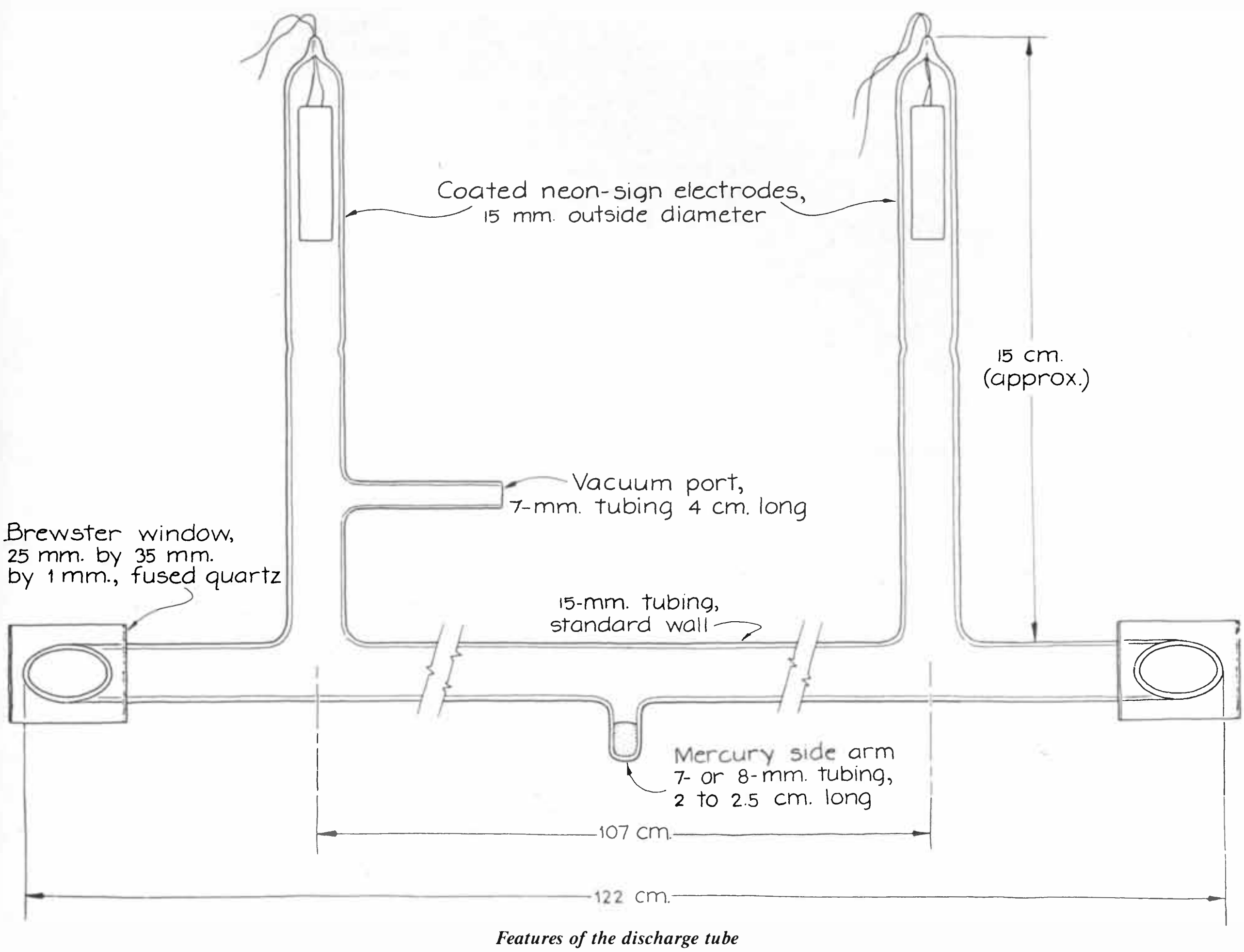
The energy level of the singly ionized mercury ion that can give rise to laser light at 615 nanometers is the one labeled $7p^2P^o{}_{3/2}$. The transition is to the level $7s^2S_{1/2}$. Part of the stimulated light leaks through a mirror to form another laser beam, which is in the red-orange region of the visible spectrum.

In the right conditions Morelli's laser can emit both wavelengths (567.7 and 615 nanometers) simultaneously. The two colors can be separated by passing the output of the laser through a prism. The red-orange output is relatively weak, however, and so Morelli has designed his laser to enhance the green output.

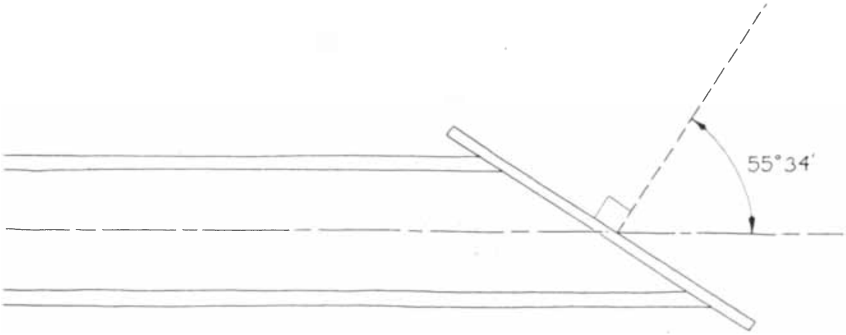
Morelli made his tube from a standard length (122 centimeters, or four feet) of Pyrex glass tubing that had a standard wall and an outside diameter of 15 millimeters. He cut another piece of the same kind of tubing to a length of 7.5 centimeters and sealed it at right angles to the long tube 7.5 centimeters from one end, employing a gas-oxygen torch to make the seal. A neon-sign electrode 15 millimeters in diameter was fused to the open end of the shorter



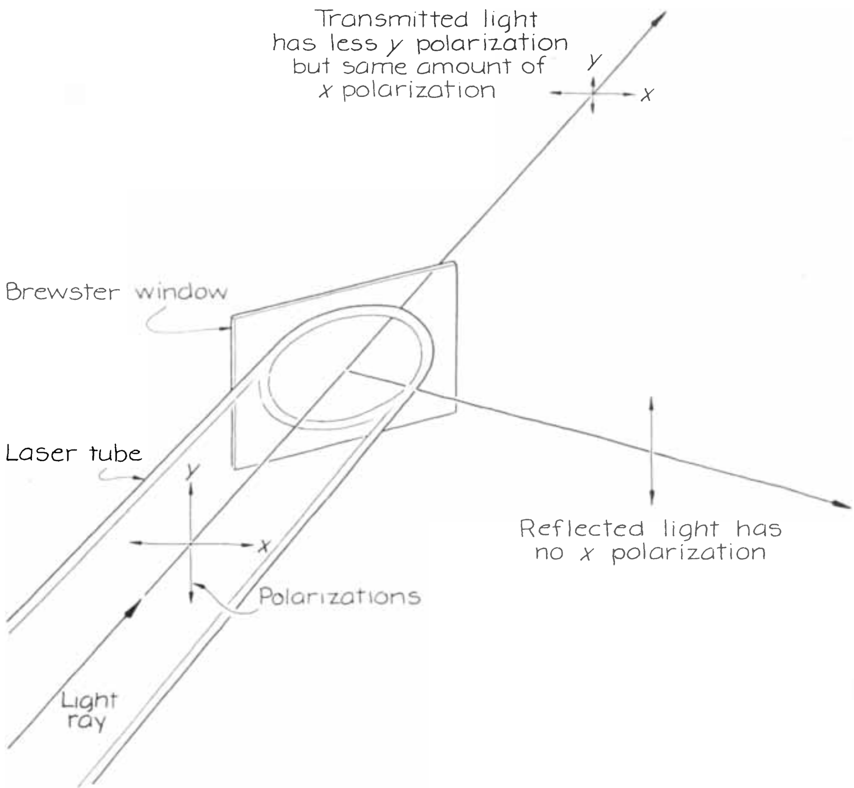
The mercury-vapor ion laser built by Dean Morelli



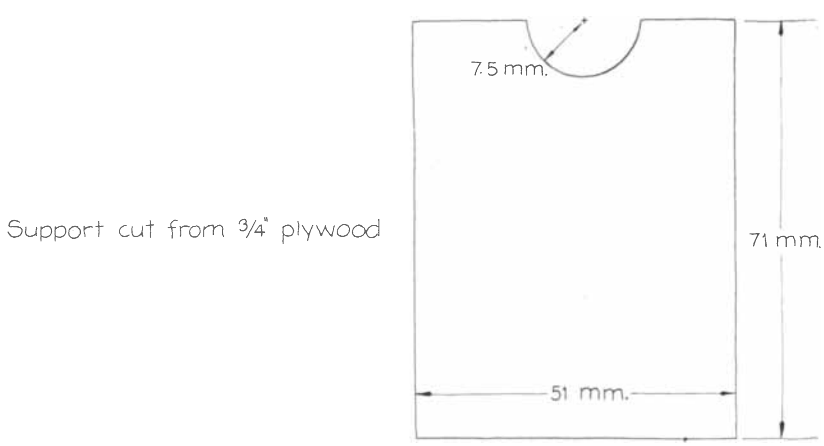
Features of the discharge tube



The angle of the Brewster window



The role of the Brewster window in polarization



The wood support for the discharge tube

tube. An identical electrode assembly was mounted at the other end of the long tube.

Another section of tubing, four centimeters long with an outside diameter of seven millimeters, was joined to one of the electrode tubes to serve as a vacuum port. A second section of the seven-millimeter tubing was fused to the center of the long tube on the side opposite to the electrode arms. This section, 2.5 centimeters long, served as the source of the mercury vapor.

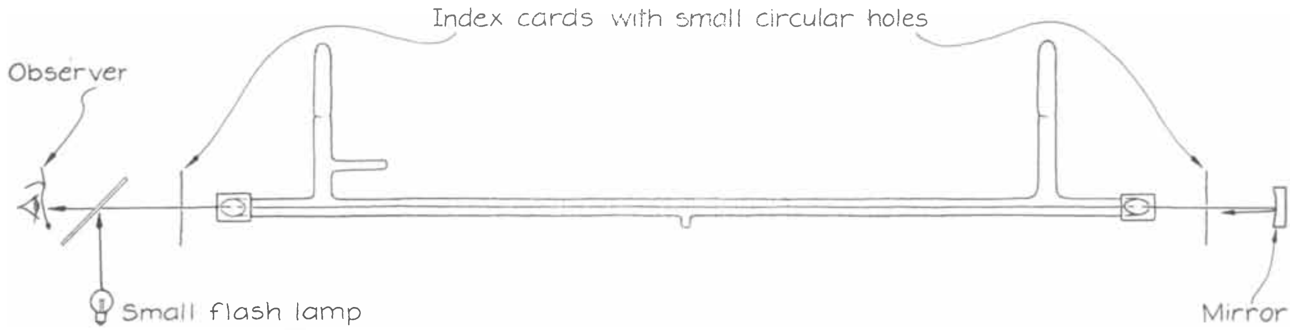
If you do not have a gas-oxygen torch, you can make your tubing of soda-lime glass that can be fused with a propane torch of the kind available in hardware stores. Two of the books cited in the bibliography for this month [*see page 210*] are particularly helpful: *Creative Glass Blowing* for the work with glass and *Light and Its Uses* for a number of pointers on the construction of lasers.

Morelli sealed the ends of the long tube with Brewster windows, which are made of flat glass of optical quality. Each piece is mounted on the tube at a precise angle chosen to minimize the loss of light reflecting from the windows. The angle is measured between the axis of the tube and a line perpendicular to the surface of the window.

This Brewster angle, as it is called, is equal to the arc tangent of the window's index of refraction. The angle has a bearing on the polarization of the light. Early in the laser process the light incident on the window is unpolarized. Because of the Brewster angle, however, one type of polarization is enhanced, and it quickly comes to dominate the laser process.

Suppose a ray of unpolarized light is incident on a Brewster window. The electric fields of the light oscillate along all possible axes in a plane perpendicular to the ray. It is helpful to imagine that the oscillations are along two perpendicular axes. One axis, labeled *x*, is in the same plane as the incident ray and the reflected ray. The other axis, labeled *y*, is perpendicular to *x*. Because of the particular geometry of the Brewster window the light reflected from the surface of the window is entirely polarized in the *y* direction. This light is reflected out of the laser tube and no longer contributes to the laser process. None of the *x*-polarized light is reflected, and so it goes through the window and is sent back through the tube by one of the laser mirrors. In this way the *x*-polarized light comes to dominate the laser process. Thereafter the light created by stimulated emission is almost entirely polarized to avoid reflection at the windows and little light is lost by them.

Morelli employed quartz windows for which the index of refraction was 1.459 for light at the wavelength of 567.7 nanometers. The arc tangent of 1.459 is 55 degrees 34 minutes, which therefore was the angle of mounting of



A system for aligning the mirrors

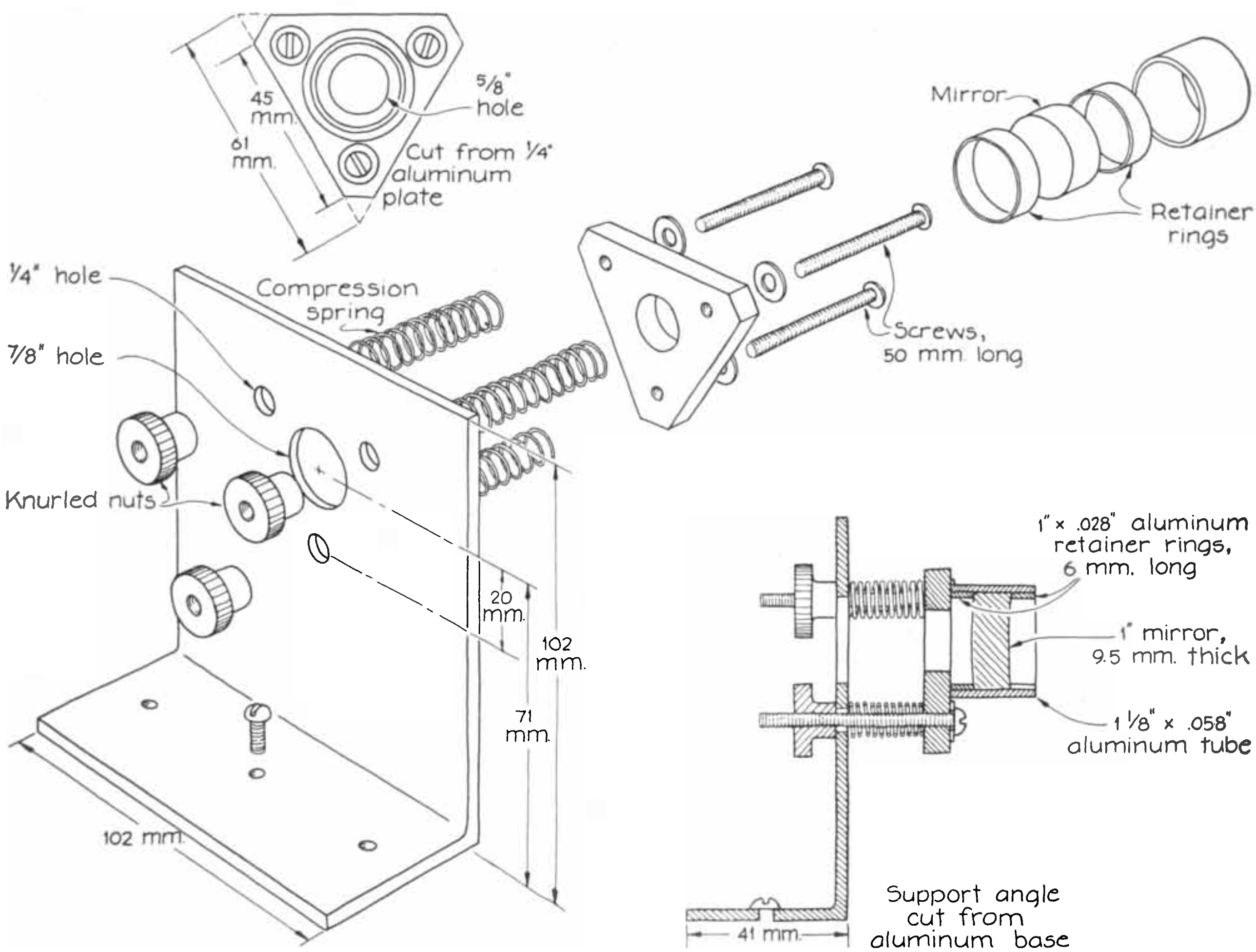
the Brewster windows. (The emission at 615 nanometers does not reach the windows at exactly the proper angle, because the index of refraction is slightly lower for this longer wavelength. The Brewster angle, however, is almost the same for the two wavelengths and so the laser will work at 615 nanometers.)

Morelli prepared the ends of the tube for the windows by sawing them at the complement of the Brewster angle, 34 degrees 26 minutes. He made the cut with a hacksaw that had a brass blade .025 inch thick, holding the tube and a pair of saw guides in a miter box. A mixture of No. 120 abrasive and water served as the lubricant. Morelli points out that the saw must be moved slowly and carefully to avoid chipping the edges of the glass. Chipped edges give rise to a loss of gas as the laser runs.

When the sawing is half completed, the tube is turned over for the rest of the cut. Then the ends are polished flat by grinding them against a flat glass tool supplied with progressively finer abrasive (the last and finest being No. 600). The cuts should not deviate from the complement of the Brewster angle by more than half a degree.

Having finished the cutting and polishing, Morelli cleaned the long tube and mounted it on its support. With an eyedropper he put a few drops of mercury into the side arm of the tube by tilting the tube slightly and carefully letting the drops run down the inside. This mercury must be very clean. Morelli worked with mercury that had been distilled three times.

Once the mercury is in place the side arm is immersed in a beaker containing dry ice and ethanol until the mercury is frozen. Now the long tube is connected to a vacuum pump at the vacuum



Details of a mirror cell

port. With the pump running, the windows are carefully held against the ends of the long tube. (You may need a helper here.) If the ends of the tube are sufficiently flat, the suction from the pump should hold the windows firmly. The slightest leak will call for more polishing to make the ends of the tube flatter. If the contact is airtight, the windows can be glued to the tube. Morelli did his gluing with a disposable hypodermic syringe that had a needle of large bore. With the syringe he applied a coat of epoxy around the connections between the tube and the windows. The suction should be maintained until the epoxy hardens.

The base for the tube and the mirrors was a rectangular piece of aluminum tubing 72 inches long, with a cross section of four inches by 1¾ inches and walls 1/8 inch thick. It is the kind of tubing sold for the making of metal doors and windows. Two supports cut from 3/4-inch plywood were attached to the base to support the laser tube. Epoxy was applied between the support and the base and the support and the laser tube. (In order to apply epoxy to aluminum, Morelli prepared the aluminum by scrubbing it with a toothbrush dipped in an abrasive cleaning powder. He kept at it until water on the metal formed a thin layer with no tendency to bead.)

Each mirror was mounted in a cell that could be adjusted with three knurled nuts. Morelli aligned his mirrors with the laser tube by means of a standard technique involving two index cards, a small flashlight and a microscope slide that served as a beam splitter. A circular hole was punched in each card. A card was attached to each end of the laser tube so that the hole was aligned with the tube. The beam splitter was put between a mirror and one end of the tube at an angle of about 45 degrees with respect to the axis of the tube. The cell holding the nearest mirror was removed so that the observer could look down the axis of the tube and see the mirror at the far end.

The flashlight, which should have a bulb with a small filament, was positioned near the slide so that its light was reflected from the slide and entered the laser tube through the hole in the nearby index card. The screws on the distant mirror were adjusted (again by a helper if one was available) until light was reflected back to the observer sighting through the tube. The alignment is correct when the light from the far mirror is brightest.

The support cell was removed from the base without disturbing the adjusting screws. That cell would remain in adjustment when it was reattached to the base. Meanwhile Morelli aligned the second mirror in the same way.

The mirrors were spherical, had equal radii and were separated from each other by slightly less than one radius (an arrangement called a confocal cavity). Their radius of curvature was 147 centimeters; they were 25.4 millimeters in diameter and 9.5 millimeters thick. The back of each mirror was ground to be convex. The separation between the mirrors was about 140 centimeters. (The mirrors could be flat on the outside surface if an external lens was employed to collimate the external beam. They could even be flat on the inside surface, but a spherical-mirror system is easier to align than a flat-mirror one.)

Each mirror was mounted in an aluminum tube 29 millimeters (1⅛ inches) wide and 21.5 millimeters long. The wall of the tube was 1.5 millimeters (.058 inch) thick. With epoxy the tube was fastened to a triangular plate that was attached to a support plate by three screws. An aluminum retainer ring 25.4 millimeters (one inch) wide and six millimeters long was slipped into the tube and glued into place. The mirror and another ring were then slipped in. Morelli held the ring in place with transparent tape. It would not do to glue the ring, he notes, because the mirrors must be removed at times for cleaning.

The mirrors ordinarily installed in lasers that emit visible light have several layers of dielectric thin films designed to increase the reflectivity of light at a certain wavelength. Morelli had his mirrors designed to reflect best for a wavelength of 567.7 nanometers. The thin film next to the glass had a refractive index higher than that of the glass. Its optical thickness was one-fourth of the designated wavelength.

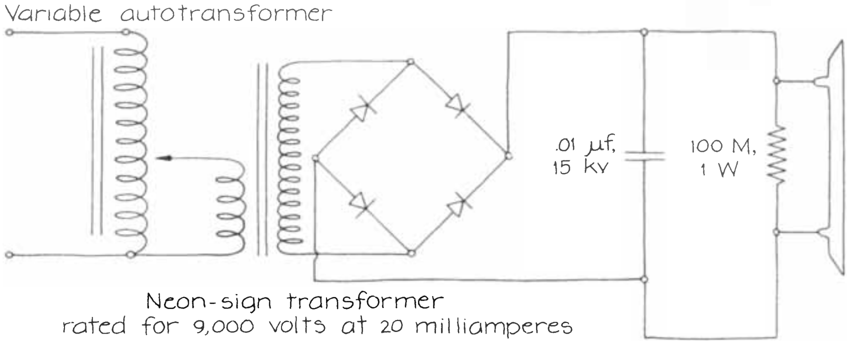
On top of this film is another layer with a lower refractive index. More layers are added, alternating between high and low refractive indexes. The outermost layer of each pair has the higher value. Each layer reflects part of the light reaching the mirror. With this design all the reflected rays interfere constructively, yielding a bright reflection. A mirror usually reflects about 99.8 percent of the light reaching it. Without the dielectric coatings the reflectivity would be considerably lower.

The range of wavelengths over which a mirror will reflect depends on the ratio of the refractive indexes of the dielectric-coating films. Morelli's mirrors reflect about 99.5 percent of the light with a wavelength of 567.7 nanometers. The range of reflection extends over about 100 nanometers and includes the 615-nanometer emission that can also be obtained with his laser. He notes, however, that the reflection is relatively poor at this wavelength, with the result that the emission may be dim or absent.

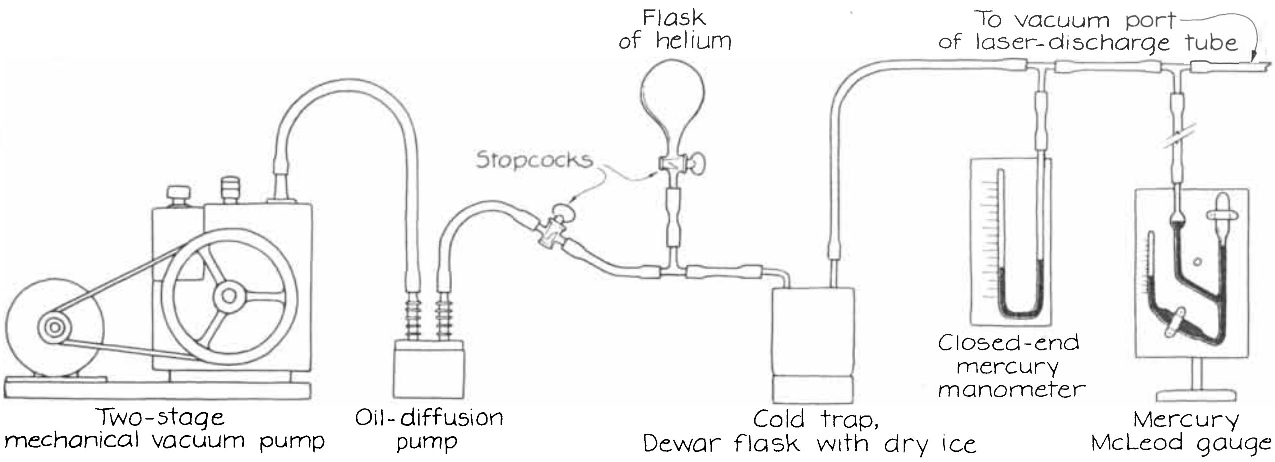
The laser is excited by a high-voltage discharge through the laser tube. A variable autotransformer feeds current to the primary winding of a neon-sign transformer rated at 9,000 volts (at 20 milliamperes). The output from the transformer is rectified by a bridge of four diodes and then fed to a capacitor connected across the laser tube. When the voltage difference between the electrodes in the laser tube reaches about 5,000 volts, the gas in the tube becomes ionized and so can conduct electricity, thereby enabling the capacitor to discharge its current through the tube. The high-current arc lasts for about a microsecond. Depending on the setting of the autotransformer, the discharge can be repeated as often as several hundred times per second. At such a rapid repetition the laser output appears to be a steady beam.

The capacitor should have a capacitance of about .01 microfarad and a breakdown voltage of 15 kilovolts. Morelli makes it from a piece of double-sided printed-circuit (PC) board. The required area of board (in square-meter units) is determined by the formula $a = ct/d\epsilon_o$, where ϵ_o is 8.85 × 10^{-12} farad per meter, c is the capacitance in farads, t is the thickness of the dielectric in meters and d is the dielectric constant of the insulating material. The epoxy-glass material in the PC board has a dielectric constant of 4.8 and a breakdown voltage of 95 kilovolts per millimeter of thickness.

Morelli etched away one centimeter of the copper along the perimeter of the board on both sides to prevent arcing in that region. Then he experimented to find the best area of the board for the laser. The larger the area, the larger the optimum capacitance. The capacitance



The circuit of the power supply



The vacuum system

is smaller for the brightest output at 567.7 nanometers than it is for the 615-nanometer beam.

The two electrodes must be protected from the bombardment of electrons they receive as the electric arcs course through the tube. Morelli obtained commercial electrodes that were coated with a mixture of barium and strontium carbonate suspended in a binder of nitrocellulose. After he had sealed the laser tube he heated the electrodes in order to convert the coating into a protective layer. He did the heating with a coil of wire slipped over the electrode assembly and excited by a high-frequency oscillator. The oscillating current in the coil induced an oscillating current in the electrode. He connected a variable capacitor in parallel with the coil so that he could generate a resonance between the oscillator and the capacitor-coil system. The oscillator should be rated for at least 50 watts.

The heating of the electrodes decomposed the nitrocellulose, liberating hydrocarbon gas. The carbonates were converted into oxides and released carbon dioxide. The oxides protected the electrodes from the arcing current. In addition the reactions gave rise to metallic barium and strontium that acted as "getters," removing active but unwanted gases from the tube.

Morelli's vacuum system consisted of a two-stage mechanical pump, an oil-diffusion pump, a closed-end mercury manometer, a mercury McLeod gauge, a cold trap, a flask of helium and two stopcocks. He says the diffusion pump is helpful but not crucial. Plans for making the gauges are given in *Creative Glass Blowing*. Additional instructions for the gauges and other elements of the vacuum system appear in *Light and Its Uses* (see the chapters on helium-neon lasers and argon lasers).

Morelli's first step toward operating the vacuum system was to freeze the mercury in the central side arm. Then he turned on the pumps and began heating the laser tube with a Bunsen burner to remove molecules of gas adhering to the walls of the tube. He took care to avoid heating the Brewster windows and the mercury in the side arm.

Having prepared the electrodes with oxide coating, Morelli connected them to the alternating-current output of the neon-sign transformer. Helium was let into the tube under a pressure of a few torr. The glow discharge that built up while the helium was in the tube cleaned the interior of the tube. The helium was pumped out and fresh helium was admitted, and the procedure was repeated until the discharge through the tube generated a pink glow. Then the helium pressure was adjusted to be between .5 and one torr. The side arm of mercury was put in a beaker of water at a temperature of 40 degrees Celsius and the electrodes were connected to the pulse-power supply.

Once mercury has diffused into the tube adjust the mirrors until the laser begins to operate, at least in the green region. Several adjustments should be made to maximize the output. The cells holding the mirrors may need further adjustment. The operator can also vary the gas pressure, the temperature of the side arm holding the mercury and the capacitor of the power supply.

Morelli suggests that all the leads on the electrodes be reversed periodically so that the electrodes exchange roles as anode and cathode. This procedure prevents one of the electrodes from receiving the entire bombardment of electrons. It also keeps the mercury ions from being pumped toward only one end of the tube. The gas pressure will gradually drop as some of the helium is buried under sputtered electrode material. The helium must be replenished after the laser has been operating for several hours.

I cannot overemphasize the potential danger of a laser. Most people realize that a laser beam entering the eye can permanently damage the retina, but the danger extends to stray reflections of the beam from objects in the room. You must also beware of the high currents developed by the power supply. This danger is particularly acute if you work with the laser in partial darkness that prevents the circuit from being fully visible. Insulate all the exposed leads. Finally, work carefully with the mercury, which evaporates not only at high temperature but also at room temperature. The vapor is poisonous. Make sure the room is well ventilated when you load the mercury into the side arm.

If you want to buy parts for this mercury-vapor laser or other lasers that have been described in this department, check the advertisements in two publications: *Laser Focus* and *EOSD* (Electro-Optical Systems Design). They specialize in lasers and publish an annual directory of suppliers of laser equipment. A library is not likely to have either publication; you may have to borrow a copy from someone in the physics or engineering department of a university or in a local company that works with lasers.

Morelli bought his mirrors, windows, glass, electrodes and accessories from North Country Scientific, R.F.D. 1, Plymouth, N.H. 03264. The abrasives came from the A. Jaegers Optical Corporation, 691 South Merrick Road, Lynbrook, N.Y. 11563. The glass-blowing equipment and supplies were from the Wale Apparatus Company, 400 Front Street, Hellertown, Pa. 18055. Helium in glass flasks is available from the EGL Co., 730 South 13th Street, Newark, N.J. 07103. The neon-sign transformer was manufactured by the Jefferson Electric Company, 840 25th Avenue, Bellwood, Ill. 60104. The vacuum pumps can be obtained from the Sargent-Welch Scientific Company, 7300 North Linder Avenue, Skokie, Ill. 60077, or from Edwards High Vacuum, 3279 Grand Island Boulevard, Grand Island, N.Y. 14072. Morelli's vacuum gauges were from Ace Glass, Inc., P.O. Box 688, Vineland, N.J. 08360.

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

THEORY OF VOTING. Robin Farquharson. Yale University Press, 1969.

THE PRESIDENTIAL ELECTION GAME. Steven J. Brams. Yale University Press, 1978.

HEALTH MAINTENANCE ORGANIZATIONS

TOWARDS A COMPREHENSIVE HEALTH POLICY FOR THE 1970'S: A WHITE PAPER. Department of Health, Education, and Welfare. U.S. Government Printing Office, 1971.

THE ORGANIZATION OF MEDICAL CARE. Ernest W. Saward in *Scientific American,* Vol. 229, No. 3, pages 169–175; September, 1973.

HMO GROWTH: 1977–1978. InterStudy. Health Systems Demographic Center, 1978.

HOW DO HEALTH-MAINTENANCE ORGANIZATIONS ACHIEVE THEIR SAVINGS? RHETORIC AND EVIDENCE. Harold S. Luft in *The New England Journal of Medicine,* Vol. 298, No. 24, pages 1336–1343; June 15, 1978.

THE BIZARRE SPECTRUM OF SS 433

THE BIZARRE SPECTRUM OF SS 433. Bruce Margon, Holland C. Ford, Jonathan I. Katz, Karen B. Kwitter, Roger K. Ulrich, Remington P. S. Stone and Arnold Klemola in *Astrophysical Journal Letters,* Vol. 230, No. 1, Part 2, pages L41–L45; May 15, 1979.

A KINEMATIC MODEL FOR SS 433. George O. Abell and Bruce Margon in *Nature,* Vol. 279, No. 5715, pages 701–703; June 21, 1979.

ENORMOUS PERIODIC DOPPLER SHIFTS IN SS 433. Bruce Margon, Holland C. Ford, Steven A. Grandi and Remington P. S. Stone in *Astrophysical Journal Letters,* Vol. 233, No. 2, Part 2, pages L63–L68; October 15, 1979.

DOES ANYONE UNDERSTAND SS 433? Dennis Overbye in *Sky and Telescope,* Vol. 58, No. 6, pages 510–516; December, 1979.

MONOCLONAL ANTIBODIES

DERIVATION OF SPECIFIC ANTIBODY-PRODUCING TISSUE CULTURE AND TUMOR LINES BY CELL FUSION. G. Köhler and C. Milstein in *European Journal of Immunology,* Vol. 6, pages 511–519; 1976.

LYMPHOCYTE HYBRIDOMAS. Edited by F. Melchers, Basel M. Potter and N. Warner. *Current Topics in Microbiology and Immunology,* Vol. 81, 1978.

HYBRID MYELOMA MONOCLONAL ANTIBODIES AGAINST MHC PRODUCTS. Edited by Göran Möller. *Immunological Reviews,* Vol. 47, 1979.

MONOCLONAL ANTIBODIES AND CELL SURFACE ANTIGENS. C. Milstein, G. Galfré, D. S. Secher and T. Springer in *Ciba Foundation Symposia,* No. 66, pages 251–276; 1979.

THE CAUSES OF COLOR

THE ORIGINS OF COLOR IN MINERALS AND GEMS. K. Nassau in *Lapidary Journal,* Vol. 29, pages 920–928, August; pages 1060–1070, September; pages 1250–1258, October and page 1521, November, 1975.

THE ORIGINS OF COLOR IN MINERALS. Kurt Nassau in *The American Mineralogist,* Vol. 63, No. 3/4, pages 219–229; March/April, 1978.

THE SOUTHERN APPALACHIANS AND THE GROWTH OF CONTINENTS

WEST AFRICAN FOLD BELT. J. Sougy in *Geological Society of America Bulletin,* Vol. 73, No. 7, pages 871–876; July, 1962.

LITHOSPHERE PLATE: CONTINENTAL MARGIN TECTONICS AND THE EVOLUTION OF THE APPALACHIAN OROGEN. John M. Bird and John F. Dewey in *Geological Society of America Bulletin,* Vol. 81, No. 4, pages 1031–1060; April, 1970.

TECTONICS OF THE WESTERN PIEDMONT AND BLUE RIDGE, SOUTHERN APPALACHIANS: REVIEW AND SPECULATION. Robert D. Hatcher, Jr., in *American Journal of Science,* Vol. 278, No. 3, pages 276–304; March, 1978.

THIN-SKINNED TECTONICS IN THE CRYSTALLINE SOUTHERN APPALACHIANS: COCORP SEISMIC-REFLECTION PROFILING OF THE BLUE RIDGE AND PIEDMONT. Frederick A. Cook, Dennis S. Albaugh, Larry D. Brown, Sidney Kaufman, Jack E. Oliver and R. D. Hatcher, Jr., in *Geology,* Vol. 7, No. 12, pages 563–567; December, 1979.

POSITRON-EMISSION TOMOGRAPHY

EXTERNAL QUANTIFICATION OF MYOCARDIAL ISCHEMIA AND INFARCTION WITH POSITRON-EMITTING RADIONUCLIDES. Burton E. Sobel in *Advances in Cardiology,* Vol. 22, pages 29–47; 1978.

FATTY ACID UPTAKE AND "METABOLIC IMAGING" OF THE HEART. Milton S. Klein and Burton E. Sobel in *Cardiovascular Clinics,* Vol. 10, No. 2, pages 165–176; 1979.

QUANTITATIVE IN VIVO AUTORADIOGRAPHY WITH POSITRON-EMISSION TOMOGRAPHY. Marcus E. Raichle in *Brain Research Reviews,* Vol. 1, No. 1, pages 47–68; July, 1979.

REGIONAL ASSESSMENT OF MYOCARDIAL METABOLIC INTEGRITY IN VIVO BY POSITRON-EMISSION TOMOGRAPHY WITH ^{11}C-LABELED PALMITATE. Michel M. Ter-Pogossian, Milton S. Klein, Joanne Markham, Robert Roberts and Burton E. Sobel in *Circulation,* Vol. 61, No. 2, pages 242–255; February, 1980.

OIL MINING

MINING FOR PETROLEUM BY UNDERGROUND METHODS. George S. Rice. U.S. Bureau of Mines Bulletin 351, U.S. Department of Commerce, 1932.

OIL MINING. A TECHNICAL AND ECONOMIC FEASIBILITY STUDY OF OIL PRODUCTION BY MINING METHODS. Anthony Edey, Bruce A. Kennedy and Leigh A. Readdy in U.S. Bureau of Mines Open File Report 55–79, PB 297 134/AS, Golden Associates, 1978.

THE BUOYANCY OF THE CHAMBERED NAUTILUS

FLOTATION MECHANISMS IN MODERN AND FOSSIL CEPHALOPODS. E. J. Denton and J. B. Gilpin-Brown in *Advances in Marine Biology,* Vol. 11, pages 197–268; 1973.

CAMERAL LIQUID IN *NAUTILUS* AND AMMONITES. Peter D. Ward in *Paleobiology,* Vol. 5, No. 1, pages 40–49; Winter, 1979.

FUNCTION OF CAMERAL WATER IN *NAUTILUS.* Desmond Collins, Peter D. Ward and Gerd E. G. Westermann in *Paleobiology,* Vol. 6, No. 2, pages 168–172; Spring, 1980.

CAMERAL LIQUID TRANSPORT AND BUOYANCY CONTROL IN THE CHAMBERED NAUTILUS (*NAUTILUS MACROMPHALUS*). Lewis Greenwald, Peter D. Ward and O. E. Greenwald in *Nature,* Vol. 286, No. 5768, pages 55–56; July 3, 1980.

THE AMATEUR SCIENTIST

VISIBLE LASER TRANSITIONS IN Hg^+. W. E. Bell in *Applied Physics Letters,* Vol. 4, No. 2, pages 34–35; January 15, 1964.

CREATIVE GLASS BLOWING. James E. Hammesfahr and Clair L. Stong. W. H. Freeman and Company, 1968.

LIGHT AND ITS USES: MAKING AND USING LASERS, HOLOGRAMS, INTERFEROMETERS AND INSTRUMENTS OF DISPERSION. T. D. Walker. W. H. Freeman and Company, 1980.

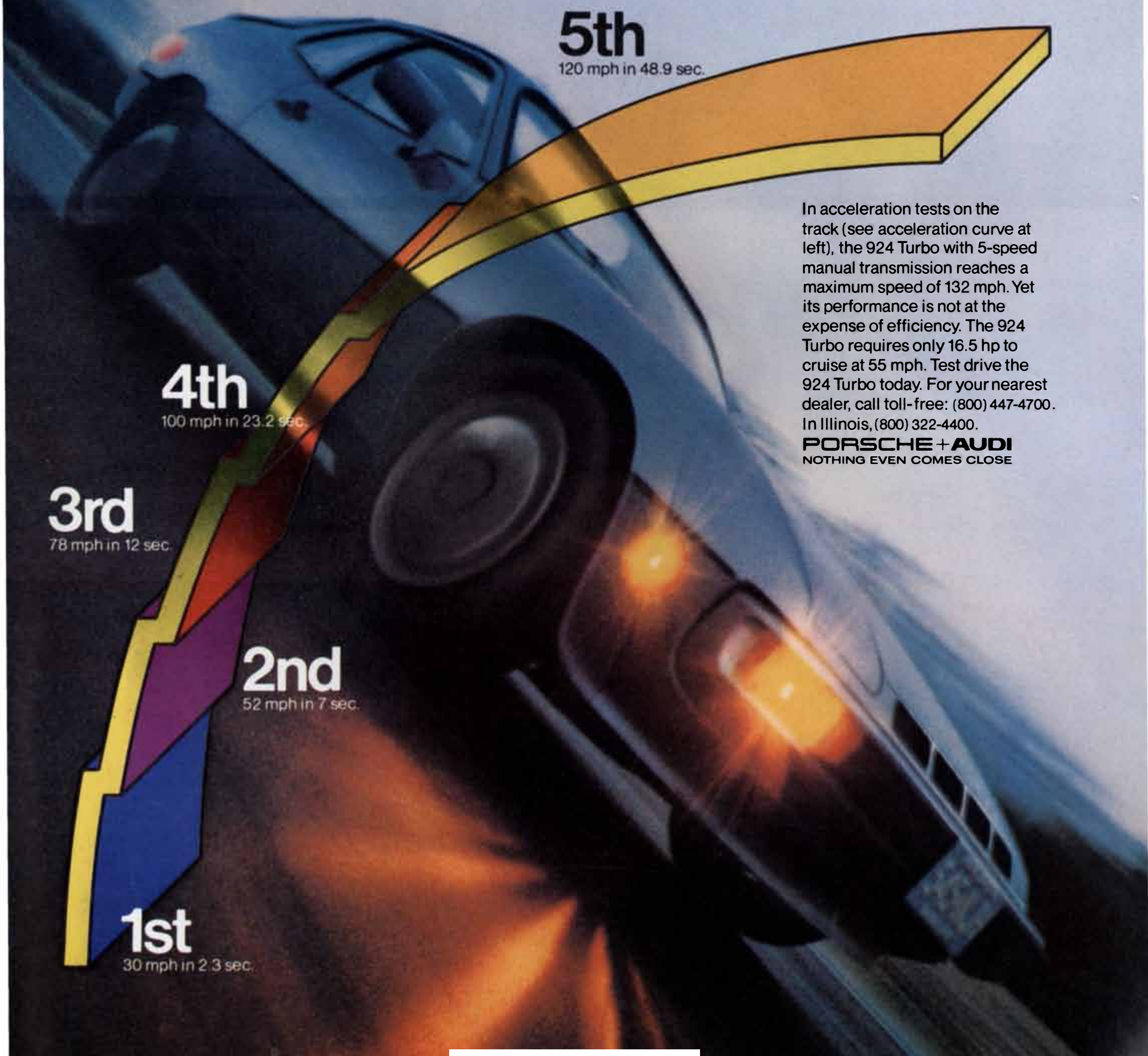
Consumer Orientation
No. 8 in a Series
Subject: Maximizing Volumetric Efficiency

8 Porsche 924 Turbo

At Weissach, where Porsche's Research and Development Center (Entwicklungszentrum) is located, the price of gasoline is $2.55 per gallon. Yet throughout Germany, you can drive the Autobahn which has virtually no speed limits. To meet the diverse goals of performance and economy*—Porsche developed the 924 Turbo.

Turbocharging maximizes the volumetric efficiency of an engine. Instead of increasing the engine's displacement, it increases the density of the charge supplied to the engine. A naturally-aspirated engine's air supply is limited by normal atmospheric pressure. But on demand, the 924's turbocharger forces air into its engine at up to one-and-a-half times normal atmospheric pressure. The result: more efficient combustion and a 30% increase in horsepower and torque. The 924 Turbo's engine produces 143 bhp at 5500 rpm and generates 147 ft-lbs of torque at 3500 rpm. The power that drives the turbocharger comes from normally-wasted engine exhaust gases. Thus, in effect, the 924 Turbo creates energy from waste.

*EPA estimated (19) mpg. 31 mpg estimated highway. Use the "estimated mpg" for comparison. Mpg varies with speed, trip length, weather. Actual highway mpg will probably be less.



In acceleration tests on the track (see acceleration curve at left), the 924 Turbo with 5-speed manual transmission reaches a maximum speed of 132 mph. Yet its performance is not at the expense of efficiency. The 924 Turbo requires only 16.5 hp to cruise at 55 mph. Test drive the 924 Turbo today. For your nearest dealer, call toll-free: (800) 447-4700. In Illinois, (800) 322-4400.

PORSCHE+AUDI
NOTHING EVEN COMES CLOSE

transmit up to 500 million bits of information per second. There are projects in Aerospace. Digital systems. And, of course, home entertainment, as exemplified by color television and the new "SelectaVision" VideoDisc System, developed entirely by RCA.

Dedicated research has been a tradition at RCA for over sixty years. Today, many years after Nipper started listening to "His Master's Voice," research and development of electronic products are funded at the highest level in our company's history. Our commitment is, as it has always been, to the advancement of technology—to the creation of products that expand the human horizon.

RCA

A tradition on the move