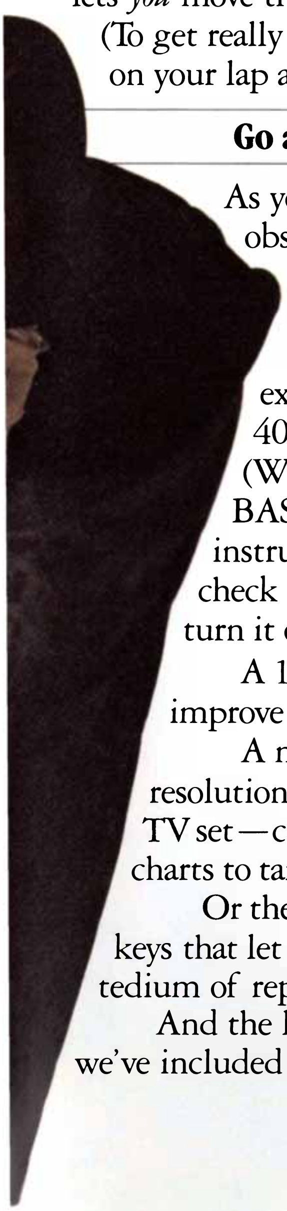
SCIENTIFIC AMERICAN

ULTRALIGHT AIRPLANES

$2.00

July 1982

Right away, you can see a difference.



Stop in a store near you. Take a look.

You'll be instantly taken with some of the features that make the IBM Personal Computer so different.

Like the non-glare screen—easy on the eyes during those number-crunching tasks like payroll and general ledger.

80 characters a line—with upper and lower case letters for a quick and easy read.

And the flexibility of a system that lets *you* move the components around at will. (To get really comfortable, try the keyboard on your lap and put your feet up.)

Go ahead, compare.

As you progress from casual observer to comparison shopper, you'll want the inside story of the IBM Personal Computer.

Like user memory expandable up to 256KB. And 40KB of permanent memory. (Which not only includes the BASIC language, but diagnostic instructions that automatically check the system every time you turn it on.)

A 16-bit microprocessor that can improve speed and productivity.

A mix of crisp text and high-resolution color graphics on your own TV set—clearly helpful for creating charts to target forecasts and trends.

Or the 10 programmable function keys that let you bid goodbye to the tedium of repetitious tasks.

And the list goes on. Which is why we've included a box (at right) that tells all.

There's more than meets the eye.

Some of the best things about the IBM Personal Computer aren't part of the computer.

Like the instruction manuals that help you set up your system and teach you to use it with the greatest of ease.

IBM PERSONAL COMPUTER SPECIFICATIONS

User Memory
16K-256K bytes*

Microprocessor
16-bit, 8088*

Auxiliary Memory
2 optional internal diskette drives, 5¼", 160K bytes or 320K bytes per diskette

Keyboard
83 keys, 6 ft. cord attaches to system unit*
10 function keys*
10-key numeric pad
Tactile feedback*

Diagnostics
Power-on self testing*
Parity checking*

Display Screen
High-resolution*
80 characters x 25 lines
Upper and lower case
Green phosphor screen*

Operating Systems
DOS, UCSD p-System, CP/M-86†

Languages
BASIC, Pascal, FORTRAN, MACRO Assembler, COBOL

Printer
Bidirectional*
80 characters/second
12 character styles, up to 132 characters/line*
9 x 9 character matrix*

Permanent Memory
(ROM) 40K bytes*

Color/Graphics
Text mode:
16 colors*
256 characters and symbols in ROM*
Graphics mode:
4-color resolution: 320h x 200v*
Black & white resolution: 640h x 200v*
Simultaneous graphics & text capability*

Communications
RS-232-C interface
Asynchronous (start/stop) protocol
Up to 9600 bits per second

*ADVANCED FEATURES FOR PERSONAL COMPUTERS

And an expanding library of software programs that meet IBM's demanding specifications.

Programs for business. Education. The lab and the home. Programs that make the IBM Personal Computer *your* tool for modern times.

See for yourself.

The quality, power and performance of the IBM Personal Computer are what you'd expect from IBM. The price isn't.

So stop in and take a look.

Visit an authorized IBM Personal Computer dealer. For a store near you, (or for information from IBM about quantity purchases) call (800) 447-4700. In Illinois, (800) 322-4400. In Alaska or Hawaii, (800) 447-0890. IBM®

The IBM Personal Computer
A tool for modern times

†UCSD p-System is a trademark of the Regents of the University of California. CP/M-86 is a trademark of Digital Research, Inc.

WHOEVER YOU ARE, WHEREVER YOU GO,

YOU HAVE A WORLD OF CHOICES WITH VISA.

THE VISA EMBLEM — SYMBOL OF WORLDWIDE ACCEPTANCE.

Introducing a new Visa card, with preferred banking and travel services. Financial institutions around the world are introducing a new card, to give you preferred banking and travel services. Each card will feature the unique design of the financial institution and will display the Visa emblem in Gold, White and Gold—your assurance of recognition, acceptance and convenience. Worldwide.

Visa Travelers Cheques: now with an expanded refund system that reports the loss of your cards, too. The worldwide refund service protects your Visa Travelers Cheques if they are lost. If your payment cards are lost too, Visa will notify the proper institutions, at no charge, and arrange a cablegram or mailgram to any person you need to contact in the emergency.

The Blue, White and Gold card. The Blue, White and Gold Visa is the most widely used payment card in the world, accepted by over 3 million stores, hotels, airlines, car rentals and restaurants in 155 countries. It's good for cash at over 100,000 financial offices, too.

YOU CAN HAVE IT THE WAY YOU WANT IT WITH VISA.

Each Visa service available at the option of participating financial institutions

Established 1845 **SCIENTIFIC AMERICAN** July 1982 Volume 247 Number 1

ARTICLES

DEPARTMENTS

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

THE COVER

The painting on the cover shows an ultralight airplane, the Quicksilver MX, a design considerably advanced over the first ultralight craft, which was made seven years ago by bolting a 12-horsepower Go-Kart engine to a biplane hang glider (see "Ultralight Airplanes," by Michael A. Markowski, page 62). The Quicksilver MX, manufactured by Eipper-Formance, Inc., has a 30-horsepower Cuyuna engine designed primarily for ultralight flight. The craft takes off and lands at about 20 miles per hour, climbs at 28 m.p.h. and cruises at 43 m.p.h. The designation MX (for multiaxis) refers to the airplane's control system, which acts independently on three axes: the pilot operates controls that move the rudder and surfaces on the wing and the tail to affect pitch, yaw and roll. He operates the control stick with his right hand; the control device visible in the painting is the throttle, which he operates with his left hand. The Quicksilver has a wingspan of 32 feet and weighs 200 pounds. Many types of ultralight airplane are available at prices ranging from $2,800 to $7,000.

THE ILLUSTRATIONS

Cover painting by Ted Lodigensky

Page	Source
16–31	Jerome Kuhl
41–42	Perkin-Elmer Corporation
43–49	George V. Kelvin
50–51	John L. Tonry, Institute for Advanced Study
53–58	Gabor Kiss
59	Courtesy of French National Center for Scientific Research (C.N.R.S.)
60	Gabor Kiss
63–67	Walken Graphics
68	Michael A. Markowski
83	David F. Malin, Anglo-Australian Observatory and Photolabs, Royal Observatory, Edinburgh
86	© National Geographic Society–Palomar Observatory Sky Survey
87	J. T. Williams, Multiple Mirror Telescope Observatory
88–89	Ilil Arbel
90	© National Geographic Society–Palomar Observatory Sky Survey (*left*), Ilil Arbel (*right*)
91	© National Geographic Society–Palomar Observatory Sky Survey
92–93	Ilil Arbel
95	James C. Wang, Harvard University
96	Ian Worpole
97	Ian Worpole (*top*); James C. Wang, Harvard University (*bottom*)
100–108	Ian Worpole
111–120	Carol Donner
123	Kaj Johansen, University of Washington School of Medicine (*top left and bottom*); Howard Ricketts, University of Washington School of Medicine (*top right*)
124	Howard Ricketts, University of Washington School of Medicine (*top left and bottom left*); Carol Donner (*top right*); Laurence D. Cromwell, University of Washington School of Medicine (*bottom right*)
127–129	Patricia J. Wynne
130–134	Ilil Arbel
137–138	Judith Farr Tormey, Temple University
139–141	Karl Hartig
142	Gabinetto Fotografico–Sopr. Gallerie, Uffizi Gallery, Florence
147	R. F. Bonifield (*top*), William M. Strouse (*bottom*)
150–153	Michael Goodman

SCIENCE/SCOPE

Computers are being called upon to help create the "super chips" that will give military electronics system a tenfold increase in data processing capability. Hughes is using computer-aided design programs to develop Very High Speed Integrated Circuits (VHSIC) and the systems in which these chips will be used. Computer help is essential because of the tremendous amount of circuitry per unit area. VHSIC chips are as complex as 100 Los Angeles street maps printed on a thumb tack, and they themselves are mere components of larger, more complex systems. Computer programs will help engineers design, lay out, and test a chip. They describe an entire system (a signal processor, for example) at many different levels of detail simultaneously to predict the system's performance under various operating conditions.

Intelsat VI will become the world's most sophisticated commercial communications satellite upon launch in 1986. The drum-shaped, spin-stabilized satellite will have twice the capacity of Intelsat V. It will be able to carry 33,000 telephone calls and four TV channels simultaneously. It will weigh more than 8200 pounds at launch, measure 12 feet in diameter, and deploy to 39 feet in height. Hughes heads an international team building Intelsat VI spacecraft for the International Telecommunications Satellite Organization.

A new software system can translate naval tactical messages into understandable form. Messages within a command, control, and communications (C^3) system are typically hard to understand because they are transmitted in telegram form and often omit subjects, direct objects, articles, prepositions, and punctuation. If grammatical errors creep in, messages can be rendered unintelligible. While conventional computer techniques can't make sense of a garbled message, a Hughes message understanding system called GRACIE can. Using artificial intelligence techniques, GRACIE understands general descriptions of flights of aircraft over ships, of attacks, and of encounters with hostile ships. It constructs grammatical sentences based on what it expects messages to be, referring when necessary to a "rule book" of examples. It can be adapted for other than naval use.

An improved process of growing silicon for laser detectors provides the highest-purity silicon available in the world. With the float zone crystal process, a polycrystalline silicon rod is heated in a vacuum chamber, producing a zone of molten silicon that passes from the bottom to the top of the rod several times, thus removing impurities and providing for growth of a highly pure crystal. The silicon can be used for laser seekers, designators, trackers, and range finders. Hughes developed the process for the U.S. Army and Air Force.

Hughes Radar Systems Group has career opportunities for engineers, scientists, and programmers. We design and build many of today's most complex airborne and spaceborne radar electronics systems, including data links, electronic warfare systems, and display systems. We need scientific programmers, systems analysts, microwave specialists (antenna, receivers, transmitters, data processors), circuit designers (analog, digital, RF/IF), mechanical designers, systems and test engineers. Send resume to Engineering Employment, Dept. SE, Hughes Radar Systems Group, P.O. Box 92426, Los Angeles, CA 90009. Equal opportunity employer.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY
CULVER CITY CALIFORNIA 90230

(213) 670-1515 EXTENSION 5964

LETTERS

Sirs:

Douglas R. Hofstadter's discussion of innumeracy ["Metamagical Themas," SCIENTIFIC AMERICAN, May] struck a responsive chord with me. I would nonetheless suggest that he is far too generous in his appraisal of the general level of numeracy when he focuses on the ability to comprehend large numbers.

My point can best be illustrated in terms of an incident that occurred while I served on the White House science staff in 1970–71. A member of the staff was charged with arranging for prototype or pilot technology-assessment studies as models for future "scientific" assessments of where new developments might lead. (President Nixon had become enamored of Alvin Toffler's *Future Shock*.) I was asked to nominate subjects for pilot technological assessments and had, inter alia, nominated mariculture. A prestigious corporation was placed under contract to prepare several assessments, including the one on mariculture. After three months of work the draft assessment was brought to me for a cursory review. Without spending more than a few minutes making rough order-of-magnitude calculations, I convinced myself that the conclusions of the report were in error by six orders of magnitude.

The corporation team of three scientists headed by a senior Ph.D. reported back that they had redone all their calculations and that the conclusions were correct. I replied that I did not care whether they were correct; they simply did not make sense. The annual per capita availability of protein from a mariculture program of reasonable size should be measured in kilograms, not milligrams. It was not until I had led the leader of the team step by step through the calculations that they discovered they had neglected a factor of 10^6 at the head of one column.

The phenomenon illustrated by this incident is commonplace. People in engineering and the sciences have a strong tendency to calculate "madly off in all directions" without stopping to think whether their results make numerical sense, even at the commonplace level of milligrams or kilograms of food.

CARL H. SAVIT

Senior Vice-President, Technology
Western Geophysical
Houston, Tex.

Sirs:

By mentioning fractals in his column ["Metamagical Themas," SCIENTIFIC AMERICAN, November, 1981] Douglas R. Hofstadter has added to my mailman's burden. I keep being asked in what sense it is true that strange attractors are "closely related to the 'fractal' curves described by Benoit B. Mandelbrot in his book *Fractals: Form, Chance, and Dimension*." The response is that this relationship is as intimate as it could be: for every known strange attractor the fractal (Hausdorff-Besicovitch) dimension *D* is a fraction, which suffices for the attractors to be fractal sets; the value of *D* is related to the Lyapunov exponents, to which Hofstadter also alludes. Since "strange" has (mercifully) not been defined, many people now identify "strange" with "fractal."

Furthermore, the fact that attractors are fractal is by no means a mere matter of vocabulary. The themes that permeate the study of fractals, such as self-similarity and asymptotic self-similarity, are also central to the study of attractors. For this reason and diverse others I am delighted to welcome the fractal attractors into my fold. My new book, *The Fractal Geometry of Nature* (in press), devotes some space to this topic.

Incidentally, the wonderful effects of the iteration of functions were first explored not by contemporary mathematicians but by Pierre Fatou (1878–1929), mostly in 1906 and 1919, and by Gaston Julia (1893–1978), mostly in 1918. Fatou in particular focused on the quadratic mapping in the complex plane.

BENOIT B. MANDELBROT

Thomas J. Watson Research Center
Yorktown Heights, N.Y.

Sirs:

Dr. Keely ["Illegal Migration," by Charles B. Keely; SCIENTIFIC AMERICAN, March] cites two authors as stating that net illegal migration to the U.S. is running about 500,000 per year and notes that neither author indicates how this figure was arrived at. He therefore concludes that this estimate lacks "empirical grounding." It is far too common in the social sciences for its practitioners to put forward important and basic numbers without feeling constrained to say how they arrived at them. Nevertheless, a little further searching by Dr. Keely of the immigration literature would have shown him that the figure does have an empirical basis.

Dr. Keely himself has estimated net legal migration to be about 250,000 per year, at least until the very large recent increase in this number (*Demography*, Vol. 15, page 272; *The New York Times*, August 18, 1971, page 19). I have estimated net migration by air (legal plus illegal) to have been between 400,000 and 500,000 per year between 1969 and 1978 (*Research in Population Economics*, Vol. 4, in press). J. Gregory Robinson has estimated net illegal migration from Mexico to be about 200,000 per year for the period 1970–75 (*Demography*, Vol. 17, page 175). A reasonable assumption is that virtually all net illegal migration from Mexico is by land and that somewhat less than a fourth of total net legal migration is also by land. Approximately a fourth of our legal in-migration is across a land border, but some of it is by air (*Immigration and Naturalization Service 1976 Annual Report*, Table 5). Net illegal migration by sea and across the Canadian border is unknown but is certainly positive. This gives us a lower bound on total net migration of (400,000 to 500,000 by air) + (200,000 illegal by Mexican border) + (50,000 legal by land), which comes to a total of 650,000 to 750,000. Net illegal migration is simply total net migration *minus* net legal migration, that is, (650,000 to 750,000) − 250,000 or 400,000 to 500,000. Note that this is a lower bound that does not include net illegal migration by sea and by the Canadian border. All the above estimates were available to Dr. Keely, and all have empirical grounding. He may disagree with me as to how firm this grounding is, but the sense one gets from his article is not that the estimate may be in error but rather that it lacks any empirical basis. This is an error on Dr. Keely's part.

DANIEL R. VINING, JR.

Regional Science Department
University of Pennsylvania
Philadelphia

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 75 cents.

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 5969, 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 ______

______

______


Renault Fuego. From the first people to turbocharge Grand Prix racing and win on three continents. Wind-smooth styling that simmers with the performance of a fuel injected 1.6 litre engine. EPA's that give you 24 est. MPG, 36 est. hwy.*
Renault Fuego. An optional Turbo for performance that's more responsive yet offers EPA's of 26 est. MPG, 39 est. hwy.*
Renault Fuego. Aerodynamics more slippery than the Porsche 928. Agile front-wheel drive. Five forward speeds or optional three-speed automatic. And standard Michelin radials. All give
New. Rac
RENAUL


•ICE COLD
SODAS
•SHAKES
•MALTS.
HOT

Why not?

It's true we've been making acquisitions and expanding into new markets.

So why not fast foods? Or breakfast cereal? Or beach balls?

We'll tell you why not.

We don't think consumer products are where we should be.

Our business is providing products and services to industry. We know how to design products to meet our customers' specific needs. That's where our experience is. That's where we think we should stay.

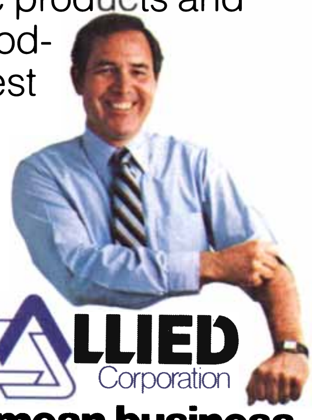
We're only interested in expanding into high-growth, technology-oriented businesses where our skills can help put us out front.

For years all our efforts were in chemicals, oil and gas, fibers and plastics.

Today we're also expanding into electrical and electronic products and health and scientific products. Two of industry's fastest growing markets.

We plan to grow in all these fields. By acquiring new businesses and developing new products through our rapidly expanding research and development program.

With all that to keep us growing, who has time for hot dogs?



ALLIED
Corporation
We mean business.

Our holographic camera makes it easy to see how the turbine blades in a jet engine react to stress.

Left: turbine blade. Right: hologram showing stress pattern of blade vibrating at resonance.

Just push a button, and get non-destructive test results in 10 seconds.

For some time now, holography has been used to test critical parts. However, conventional film and film processing make the task tedious, time-consuming, and costly. Wet processing, and the need to reposition ordinary photographic plates with ½-micron accuracy, has limited the use of holography as a design and a quality control tool.

In order to be truly practical, holograms of products like turbine blades, TV tubes or automotive and aircraft tires must not add appreciably to the cost of the item, or the time taken to produce it. Recently a team of research scientists from Honeywell's Corporate Materials Science Center and Test Instruments Division introduced a holographic camera and thermoplastic film. This system reduces development time from 30 minutes to 10 seconds. And the typical material cost per hologram from $5 to less than a dollar.

Good vibrations.

To examine a turbine blade, a technique called time-average holography is used. A time exposure is made while the blade vibrates at its resonating frequency. Because blades typically resonate at anywhere from several hundred to several thousand Hz, exposure times of as little as 1/10th of a second contain enough cycles to provide a significant average. The vibration patterns captured by the hologram reveal imbalances which may, in actual use, cause metal fatigue.

The thermoplastic film we developed for our system is the key to the system's speed and economy. It's real time, in-place write/develop/read capabilities eliminate conventional processing and repositioning problems. To record an image, the thermoplastic film is electrically charged, exposed, recharged and then heated to the softening point—allowing the electrostatic forces to deform its surface according to the distribution of illumination on the surface. After the image has been read, the image can be erased by simply reheating the thermoplastic film.

The latest development.

Unlike conventional photographic films, our thermoplastic film can be developed *in situ,* without the use of chemicals, in less then 10 seconds. In fact, an entire write/develop/read/erase cycle can be completed in well under a minute. And because a single film plate can be reused up to 300 times, manufacturers are able to test products at a material cost of less than a dollar a shot.

Our holocamera and thermoplastic film plates are available today. But our research in electro-optics doesn't end there. Investigation into alternative sources of laser light and improved layering techniques are already under way.

If you have an advanced degree and are interested in a career in solid state electronics, sensors, optics, or material sciences, please write to Dr. W. T. Sackett, Vice President. Honeywell Corporate Technology Center, 10701 Lyndale Avenue South, Minneapolis, MN 55420.

For more information on the holocamera and its application, write to Ms. Dorothy Saxum, Honeywell, Honeywell Plaza 4118, Minneapolis, MN 55408.

Honeywell

Although Honeywell engineering is world-wide, the bulk of research and applied research is done in Minneapolis. The most recent Quality of Life Study conducted by Midwest Research Institute shows Minnesota to be one of the best places in the country to live and work considering cultural, social, economic, educational and political factors.

This ideal research environment is further enhanced by Honeywell's affiliations with universities. We have an ongoing program of seminars with Berkeley, MIT, Stanford, the University of Kiel, the University of Illinois, Cornell, Purdue and the University of Minnesota.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JULY, 1932: "Just what may happen if physicists succeed in making available the incomparable energy within the atom—enough of it in a dime to drive a steamship across the Atlantic and back—is anybody's guess. In the famous Cavendish Laboratory at Cambridge University, Drs. Cockcroft and Walton have split the atom, releasing on a minute scale, but nevertheless releasing, far more energy than was put into the process. Although science cherishes and treasures every fragment of new knowledge of nature it can gain, in the present case there is unfortunately a possible by-product that might prove to be a doubtful asset to humanity. Sir Oliver Lodge thinks the young human race is still too primitive for this dangerous present. Perhaps humanity would not handle it judiciously. The control of such stupendous forces might also fall into the ruthless hands of the soldier-minded, and what fun it would provide for them while the human race and civilization lasted!"

"Archie M. Palmer, associate secretary of the Association of American Colleges, has described the means in which various universities throughout the country have met the question of what they shall do with valuable and patentable discoveries made by faculty members. Perhaps the outstanding patent now held by any university is that by which the University of Toronto controls the manufacture of insulin. 'So successful,' writes Mr. Palmer, 'has been the administration of the patent rights with respect to this product that the income of the university for one year was $500,000.' In treating the legal problems arising from the financing of research by industry the method employed by the University of Michigan is one workable solution. 'There the Department of Engineering Research performs considerable experimental and research work in its laboratories under contractual arrangements with industrial concerns. However, the University reserves the right to publish for the benefit of science such results as are in the nature of fundamental principles.'"

"The spectra of the four outer planets show broad and diffuse bands in the orange and red. Rupert Wildt of the University of Göttingen, working with the new plates sensitive to the deep red, has found several more bands beyond the limit of previous work. Band spectra come from compounds, and such wide and hazy bands as these suggest that they arise in a considerable thickness of thoroughly dense gas. There is good reason to suppose the substance or substances that produce these bands are compounds of familiar elements that remain gaseous, or at least condense only partially at very low temperatures. Hydrogen compounds afford the 'best bet,' because the low densities of the larger planets suggest that, unlike the earth, they may have retained a great deal of the original solar hydrogen. Ammonia gas shows three bands within the wide ones newly discovered in the infra-red. Wildt has no hesitation in stating that the presence of ammonia in Jupiter's atmosphere is proved."

SCIENTIFIC AMERICAN

JULY, 1882: "The demonstration of the intimate relation of bacteria to certain fevers and other diseases would seem at first to greatly simplify the work of the physician in searching for efficient remedies. Put in plain English the problem is: Find some element or compound that is fatal to bacteria and administer it in the way best calculated to reach the mischievous fungi in the patient's blood. The problem is easier stated than solved. The lower forms of life that appear to cause the trouble are able to live and thrive under the widest possible range of conditions, so that, as far as is known, any reagent able to kill them would be much more quickly fatal to the patients. It is barely possible that these vicious organisms may be reached and killed by some drug in doses the human system can tolerate, but the prospect is certainly not bright."

"The factory of the Edison Lamp Company has been moved from Menlo Park, N.J., to East Newark, N.J. The manufacturing of lamps was begun in the new factory on June 1, and 150 men are now employed. The tools and power in the factory are adequate for making 1,200 lamps a day, but the factory has an ultimate capacity of 40,000 lamps a day, which will require from 3,000 to 4,000 hands. The lamp factory has always been managed with unusual skill and intelligence, and all visitors have united in praising the perfection of the system and the economy and precision of the work."

"Men of high authority and experience declare that opium is as much a necessity for the natives of the East as wine, spirits and beer are for the natives of the West, and that the evils arising from its use are altogether less than those caused by the use or abuse of alcohol. For three years past the Indian opium crop has been short, and coupled with this deficiency there has been an increase in the production of Persian and Chinese opium. If the government monopoly of opium were abandoned, not only would India lose a revenue that would have to be made up by some other tax but also the extent of the poppy cultivation would almost certainly be largely increased in the hands of private growers. If the government went further than this and altogether forbade the poppy cultivation, they could not stop the Chinese demand, which would then be supplied by inferior qualities of Persian and native Chinese growth."

"Liebig regarded the production of muscular energy as an expenditure of the 'vital force' of the muscle substance itself, involving its death and chemical transformation, and requiring *nitrogenous* or tissue food for its regeneration. Mayer attributed the production of that energy to the oxidation of a portion of the *non-nitrogenous* or respiratory food, regarding the muscular apparatus as a mere instrument by which that oxidation is made to produce motion in place of heat. By experimental inquiries Liebig's doctrine has been disproved and that of Mayer has been established. The *mechanical working* of the body of a living animal is as directly dependent as its *heating* upon the oxidation of the hydrocarbons of its food, and these may be most economically supplied by non-nitrogenous substances. On the other hand, the mechanism can only be kept in working order by the continual renovation of its substance, and for this renovation a supply of *proteids* is essential."

"The history of the discoveries in the Pennsylvania oil fields has been one of a series of disappointments to the producers. From 1866 to 1872 the price per barrel averaged from $4 to $5, and the producers were making money rapidly. Then the field in Butler County was struck, and from that day to this the production has been greater than the consumption. Then came the Bullion pool with its 2,000- and 3,000-barrel wells, which forced the price down to $1.50. This field was soon exhausted, and better times for the producers were at hand when the Bradford field, the largest in extent ever known, was opened. Then Bradford began to decline and again a silver lining was seen, but again disappointment came. In May of last year the first well was struck in Allegany County, New York, and a new field was opened that soon more than made up for the decline. Then was the great '646' well struck, and with it followed disaster to the owners of wells generally, and lower-priced oil than since the summer of 1874, when for a time it sold for 45 cents a barrel. Where the next field will be is only a matter of conjecture."

THE AUTHORS

JOHN N. BAHCALL and LYMAN SPITZER, JR. ("The Space Telescope"), are astrophysicists who have worked on the planning for the Space Telescope since early in that effort. Bahcall is professor at the Institute for Advanced Study and visiting lecturer in astronomy with the rank of professor at Princeton University. He was graduated from the University of California at Berkeley with a B.A. in 1956. His M.A. (1957) is from the University of Chicago. His Ph.D. in physics (1961) is from Harvard University. From 1960 to 1962 he was research fellow in physics at Indiana University. In 1962 he moved to the California Institute of Technology, where he served as research fellow, assistant professor and associate professor. In 1968 Bahcall joined the faculty of the Institute for Advanced Study. Spitzer is retiring this month as Charles A. Young Professor of Astronomy at Princeton. He obtained his bachelor's degree at Yale University in 1935 and his Ph.D. from Princeton University in 1938. From 1939 to 1942 he was an instructor at Yale. During World War II he worked on several military projects, including one on sonar. In 1946 he returned to Yale as associate professor. In 1947 he moved to Princeton as professor and chairman of the department of astrophysical sciences and director of the university observatory. He held the last two posts until 1979. He was an early advocate of an orbiting platform for astronomical observation.

KLAUS BECHGAARD and DENIS JÉROME ("Organic Superconductors") are respectively a physical chemist and a physicist who have collaborated since 1975 in work on the subject of the current article. Bechgaard is lecturer in chemistry at the University of Copenhagen. He received his Ph.D. in 1973 from the University of Copenhagen. After a year of postdoctoral research at Johns Hopkins University he joined the faculty at Copenhagen, where he started his investigation of organic superconductors. Bechgaard is currently on sabbatical leave at the Université de Paris-Sud in Orsay. Jérome is at the French National Center for Scientific Research (C.N.R.S.) laboratory at the Université de Paris-Sud. His Ph.D., awarded in 1965, is from the Université de Paris-Sud. He moved to the C.N.R.S. laboratory in 1967 after a year of postdoctoral work at the University of California at San Diego. Jérome has worked on magnetism and the transitions between the phases of matter in addition to superconductivity.

MICHAEL A. MARKOWSKI ("Ultralight Airplanes") is a writer on small, light flying machines and a consultant to companies that manufacture them. He was graduated from Pennsylvania State University with a B.S. in 1968. After graduation he worked as an aeronautical engineer. In 1968 he was with the McDonnell Douglas Corporation. From 1969 to 1971 he worked for Sikorsky Aircraft. From 1973 to 1975 he worked for Atkins & Merrill Inc., a company that specializes in testing aircraft designs. During the last period he was a partner in a firm that manufactured hang gliders. Since 1975 Markowski's main interest has been ultralight aircraft.

CHARLES J. LADA ("Energetic Outflows from Young Stars") is assistant professor of astronomy at the University of Arizona. His bachelor's degree was given by Boston University in 1971. His master's degree (1973) and his doctorate (1975) are from Harvard University. From 1975 to 1977 he was a postdoctoral fellow at the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory. In 1977 and 1978 he was research fellow at Harvard. In 1978 he moved to Arizona, becoming assistant professor and assistant astronomer at the Steward Observatory in 1980. Lada is the winner of the 1982 Bart J. Bok Prize for astrophysical research, awarded annually by Harvard.

JAMES C. WANG ("DNA Topoisomerases") is professor of biochemistry and molecular biology at Harvard University. Born in mainland China, he moved with his family to Taiwan in 1949. He got his B.S. at the National Taiwan University in 1959 and was assistant instructor there in 1959 and 1960. In 1960 he moved to the U.S. to continue his education. His M.A. (1961) is from the University of South Dakota. His Ph.D. in chemistry (1964) is from the University of Missouri. From 1964 to 1966 he was research fellow in chemistry at the California Institute of Technology. In 1966 he moved to the University of California at Berkeley, becoming professor in 1974. In 1977 he joined the faculty at Harvard.

KAJ JOHANSEN ("Aneurysms") is associate professor of surgery at the University of Washington School of Medicine. He was graduated from the University of Washington in 1966 with a B.S., and he went on to obtain his M.D. from the University of Washington School of Medicine in 1970. After serving a residency in surgery he earned his Ph.D. in physiology and biophysics from the University of California at San Diego in 1977. In 1978 he returned to the University of Washington School of Medicine, where he serves as director of the residency program in surgery as well as associate professor. Johansen writes: "I am an avid long-distance runner, having run about 35 marathons. I was American 50-kilometer road-running champion in 1974, and my time of 2 : 57 : 00 stood as the American record until 1979."

J. DAVID LIGON and SANDRA H. LIGON ("The Cooperative Breeding Behavior of the Green Woodhoopoe") are biologists who have been jointly studying the green woodhoopoe in Kenya intermittently since 1975. David Ligon is professor of biology at the University of New Mexico. His degrees are from the University of Oklahoma (B.S., 1961), the University of Florida (M.S., 1963) and the University of Michigan (Ph.D., 1967). In 1967 and 1968 he was assistant professor of biology at Idaho State University. He joined the faculty of the University of New Mexico in 1968, becoming professor in 1977. His main research interest is the adaptive advantages conferred on birds by social behavior. Sandra Ligon is lecturer in biology at the University of New Mexico. She was graduated from Knox College in 1971 with a B.S. and went on to obtain an M.S. from the University of New Mexico in 1973. After receiving her master's degree she worked for two years for the U.S. National Fish and Wildlife Laboratory. Her work at the Fish and Wildlife laboratory was on the sirenians, an order of herbivorous aquatic mammals that includes the manatees and the dugongs.

ALAN TORMEY and JUDITH FARR TORMEY ("Renaissance Intarsia: The Art of Geometry") are teachers of philosophy who share an interest in the art of the Renaissance. Alan Tormey is professor of philosophy at the University of Maryland, Baltimore County. He attended the Eastman School of Music and the Juilliard School of Music, studying the flute. His B.A. (1962) is from New York University. His Ph.D. (1968) is from Columbia University. From 1968 to 1972 he was at Rutgers University, first as assistant professor, then as associate professor. He left in 1972 to take up his current job. His main interests are the philosophy of mind and the philosophy of art. Judith Tormey is associate professor of philosophy at Temple University. She was graduated from Barnard College with a B.A. in 1961, going on to obtain a Ph.D. from Columbia in 1970. Her main interests are medical ethics, the philosophy of psychiatry and the philosophy of art. In 1976 and 1977 she was assistant professor in the department of psychiatry at the University of Maryland School of Medicine under the university's Humanities in Medicine program.

This is Thomas. One day, he might win the Nobel Prize. Or perhaps he'll even get a job. (Anything's possible, after all.)

But for two hours last Autumn, all his glorious futures hung in the balance.

He contracted an infection. His temperature soared and he fell into convulsions.

Before we got together with British Telecom, four hours might elapse between the development of serious symptoms and a patient's admission to hospital.

Now, thanks to the national radio paging system, a doctor can be alerted to an emergency by a 'bleep,' carried about his person, enabling him to save vital minutes in the treatment of serious illness.

So it was with Thomas. His GP was at his side within an hour, and a potentially-fatal situation defused.

One day, he'll know enough words to thank his lucky stars.

Meanwhile, it's hard to resist raising an eyebrow at those who depict the microchip as the harbinger of a new Dark Age, in which honest flesh and blood will be surplus to requirements.

For there's nothing inherently sinister about a silicon chip. It is, after all, merely a slave. It does what it's told.

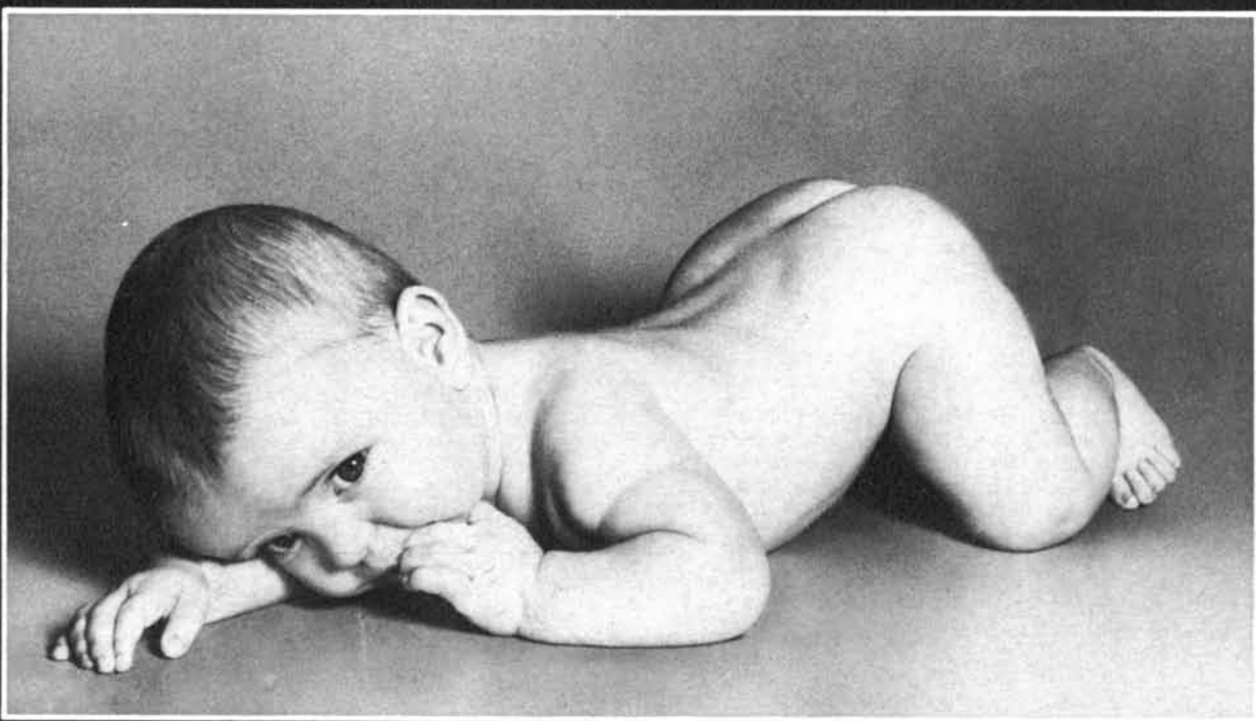
We tell ours, among other things, to train pilots in our flight simulators; to entertain us through our television and recorded music systems; and to take the drudgery out of office work through our advanced, desktop Teleputer terminals.

In fact, Thomas will bump into us in all sorts of unexpected places as he grows up.

And we believe that, thanks to us, his world will be rather better than the one he almost left, last September.

REDIFFUSION

A MICROCHIP GAVE THIS MAN A FUTURE.



WITH A TURNOVER OF £250M, REDIFFUSION IS A LEADING NAME IN BRITISH ELECTRONICS. IF YOU WOULD LIKE TO KNOW MOREABOUT US WRITE FOR A BROCHURE TO:— GROUP PUBLIC RELATIONS EXECUTIVE, REDIFFUSION LTD., CARLTON HOUSE, LOWER REGENT ST., LONDON SW1Y 4LS.

METAMAGICAL THEMAS

Beyond Rubik's Cube: spheres, pyramids, dodecahedrons and God knows what else

by Douglas R. Hofstadter

O cursed spite,
that ever I was born to set it right!
—*Hamlet,* Act I, scene 5

These days just "the Cube" will suffice; no one needs to say "Rubik's Cube" to be understood as referring to that mighty puzzle object. In fact, I have a cube in the shape of a sphere, which I sometimes call "the round cube" but equally often refer to as "that cube over there." It has been sliced up in the proper way, with rotating "sides" and an inner mechanism the same as Rubik's. What is even more marvelous, I have what poses as a cube but is most definitely not a cube: a cubical object sliced in a strange diagonal way, which scrambles in a devilishly skew manner [*see illustrations below*]. The sphere is, of course, a cube, and the cubical object is an impostor in cube's clothing.

This proliferation of varieties of "cube" (hereinafter a generic term for these scrambling-by-rotation puzzles) is really an astonishing phenomenon. Ernő Rubik (and his somewhat eclipsed Japanese counterpart Terutoshi Ishige) began it, but then it just took off like a prairie fire. Suddenly there were variations on the Cube (Rubik's 3 × 3 × 3, that is) turning up all over: little ones, teeny-weeny ones, prettily decorated ones and so forth. In some sense, however, none of these puzzles was genuinely different from the Cube itself.

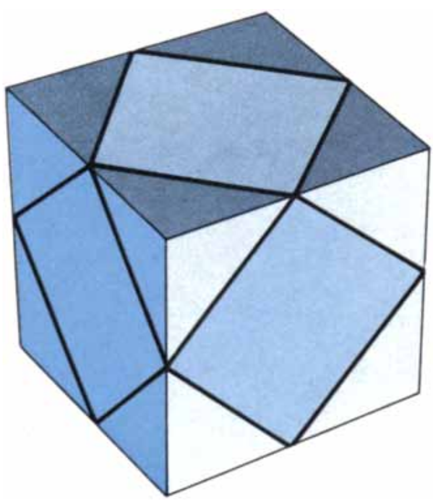
The first genuinely different cubes I saw came from Japan. They were 2 × 2 × 2's! One was magnetic, with eight metal cubies sliding around a central magnetic sphere. The other was plastic and had an intricate mechanism similar to the Rubik-Ishige 3 × 3 × 3 but not identical with it. It could not be identical, since the keystones of the 3 × 3 × 3 mechanism are the six face centers and in a 2 × 2 × 2 there aren't any centers! Later I found out that this mechanism is also Rubik's, based on the 3 × 3 × 3 mechanism. The 2 × 2 × 2 [*see illustration at top left on opposite page*] is a wonderful, inevitable object, in some ways even more beguiling than the 3 × 3 × 3. Indeed, I wonder why they are not available all over. The 2 × 2 × 2 (Twobik's Cube?) seems to me an ideal steppingstone from total novicehood to an intermediate level of cubistry, since it involves solving only the *corners* of a 3 × 3 × 3.

I should not say that the 2 × 2 × 2 was the first essentially different cube I encountered. I had seen a "Magic Domino," another Rubik invention, much earlier. The Domino is like two of the three layers of a 3 × 3 × 3 cube [*see illustration at top right on opposite page*]. Its square top and bottom layers can both turn 90 degrees, but its four rectangular sides must turn 180 degrees to allow further moves. Another early variant was the "Octagonal Prism," a cube four of whose edges had been shaved and which, when twisted, gave rise to some rather grotesque shapes [*see middle illustrations on opposite page*]. In this version some of the information about edge parities is lost (you cannot tell whether the "shaved" edge pieces are in forward or backward), so that it has some quirks that make solving it slightly different from solving the full Cube. On the full Cube flipped edges always come in pairs. Here the same is true except that since you cannot see whether a shaved edge piece is flipped or not, sometimes you will wind up with what appears to be a solved cube but with only a single flipped edge. The first time it can be quite confusing if you are used to the full Cube.

The next variation I encountered was one devised by a young German named Kersten Meier, who was then a graduate student in operations research at Stanford. He had built a rough working prototype of a "magic pyramid." (Actually it is not a pyramid with a square base, like the Egyptian monuments, but a regular tetrahedron whose four sides are equilateral triangles.) Meier's prototype was so rough, in fact, that it often fell apart as one twisted its sides. Nevertheless, it was clearly an innovative step, and it deserved to be marketed. I later found out that at nearly the same time Ben Halpern, a mathematician at Indiana University, had come up with the same concept. Both Meier and Halpern had generalized the Rubik-Ishige 3 × 3 × 3 cube mechanism and had seen how to make a dodecahedral puzzle on the same principles. Halpern built working prototypes of both the pyramid and the dodecahedron [*see bottom illustrations on opposite page*].

As it turns out, Uwe Mèffert, another German-born inventor, beat both Meier and Halpern to the pyramidal punch, although in a different way. In 1972 Mèffert had been interested in pyramids and their pleasing qualities when they are held in the hand. Somehow he hit on the notion of a pyramid with twisting sides and made a few [*see illustration on page 18*]. He found them soothing to play with and helpful for meditation, but after a while he stored them away and more or less forgot about them. Then along came Rubik's Cube. Seeing its phenomenal success, Mèffert realized that his old invention might have some commercial value. He quickly patented his design, made arrangements to have

A cube



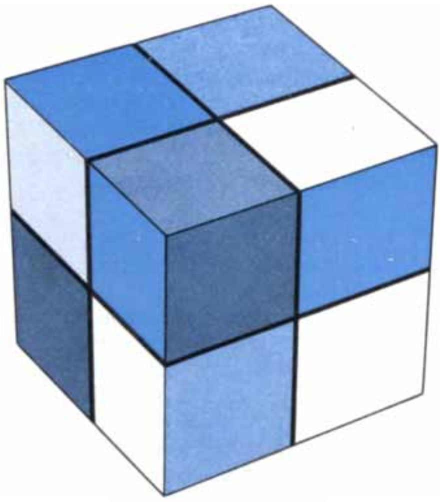
Not a cube

the device manufactured in quantity and got in touch with a toy company for the marketing. The end result was the world success of the Pyraminx, a pyramidal cube (in the generic sense) that operates quite differently from the Meier-Halpern pyramid.

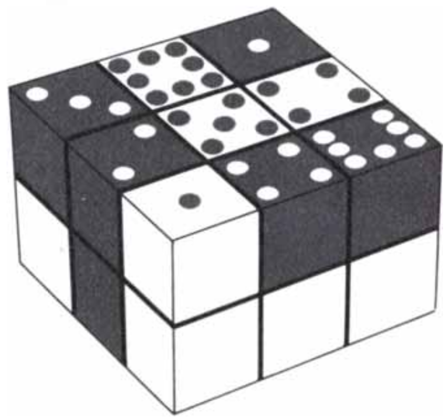
Mèffert, who now lives in Hong Kong, became deeply involved in the production and marketing of his Pyraminx and began traveling a lot. He came in contact with other inventors around the world and decided it would be a good idea to market various toys of the cube family worldwide. Among these inventors were Meier and Halpern, and as a result their pyramid too will soon be available to puzzle lovers. It will be known as the Pyraminx Magic Tetrahedron. The dodecahedron will also be available, under the name Pyraminx Magic Dodecahedron. (I would have preferred King Tet.) For a catalogue of Mèffert's complete 1982 range, write to him in care of Pricewell (Far East), Ltd., P.O. Box 31008, Causeway Bay, Hong Kong. Mèffert welcomes all ideas for new puzzles; please send them to him rather than to me. Mèffert also wants to develop a Puzzler's Club, in which members would subscribe at a yearly flat rate and receive in return six or more new puzzles a year. These would be limited editions of particularly complex or esoteric forms of "cubic" puzzles. He would like to hear from prospective members.

Ronald Turner-Smith, a friend of Mèffert's in the mathematics department at the Chinese University of Hong Kong, has written a charming little book on the patterns and the mathematics of the Pyraminx called *The Amazing Pyraminx,* which is available in paperback through Mèffert. In it Turner-Smith does for the Pyraminx what David Singmaster did for the Cube in his *Notes on Rubik's "Magic Cube."* (Incidentally, Singmaster is continuing in his role as world clearinghouse for cubology. He now puts out a quarterly newsletter amusingly titled *Cubic Circular.* You can subscribe for $5 per year by writing to David Singmaster, Ltd., 66 Mount View Road, London N4 4JR, England. Again, please send all ideas and inquiries about the Cube to Singmaster, not to me. I should also mention that a quarterly magazine called *Rubik's* will be coming out of Hungary this summer. It is available for $8 per year through Post Office Box 223, Budapest 1906, Hungary.) Like Singmaster, Turner-Smith develops a notation and uses it to convey some of the group theory connected with it that affords one a deeper appreciation of the object than mere mechanical solving does.

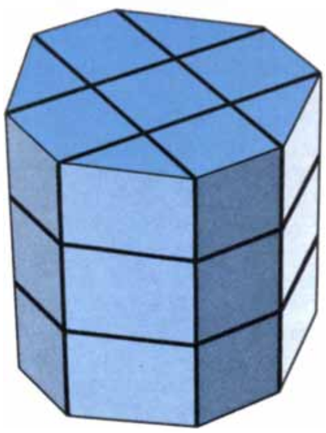
It is interesting that there are two distinct ways of manipulating and describing the action of the Pyraminx. You can rotate either a face or a small pyramid. The two views are equivalent but complementary, since a face and its opposing small pyramid make up the entire object. Turner-Smith sees the small pyramids as being movable and the faces as being stationary. I shall adopt this view for the moment and then return to the complementary one. Let us, then, name the four possible moves [*see illustration at top left on page 19*]. Each move rotates a small pyramid, either at the Top (T), the Back (B), the Left (L) or the Right (R). The letters T, B, L, and R stand for clockwise 120-degree turns and T', B', L' and R' stand for counterclockwise 120-degree turns (as



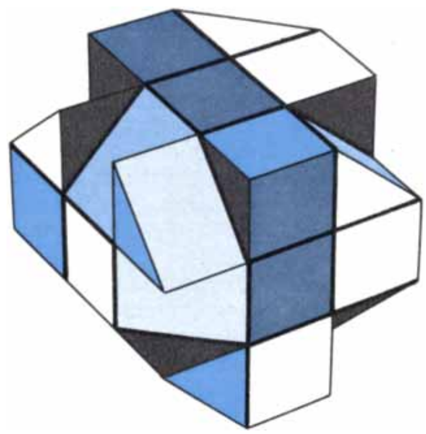
A 2 × 2 × 2 cube



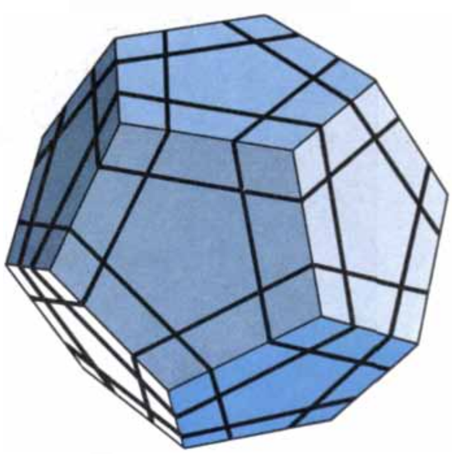
The Magic Domino



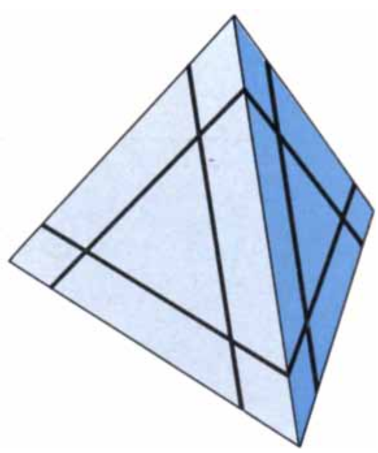
The Octagonal Prism in its pristine state



The Octagonal Prism in a scrambled state



The Pyraminx Magic Dodecahedron



The Pyraminx Magic Tetrahedron

seen when looking at the rotating tip along the axis of rotation). Notice that a move leaves all the vertex tips in place (although twisted). Therefore one can consider the four tips as stationary reference points, much like the six face centers of the Cube. In fact, at the very start of the solving process the tips can quickly be twisted to agree, and from then on they provide an identifying color for each face. Thus one can consider the four tip pyramids either as decorative ornaments or as useful signposts.

With the Cube the elementary objects that change location are usually called cubies or cubelets. What are the corresponding elementary objects here? They are certainly not just small pyramids. As with the cube it turns out that there are three types: edge blocks, middle blocks and the above-mentioned tips [*see bottom illustrations on opposite page*]. As you can see, at each vertex there is a middle block with three "trianglets" of different colors, the same colors as those of the tip perched on top of it. Like a tip, a middle block never leaves its home location but only twists. As a result the tips can be considered "trivially solvable" parts of the Pyraminx and the middle blocks "easily solvable" parts.

This leaves six edge blocks, each having two colors, that can travel and flip, just like the edge cubies on the Cube. As a matter of fact, it turns out that the constraints on flipping and swapping edges are exactly analogous to those applying to the edge cubies on the Cube: two edges must flip at once, and only *even* permutations of edge locations—permutations where an even number of edge swaps have taken place—are allowed.

This means one can quickly enumerate the number of different ways edges can be distributed about the Pyraminx. Without the constraints the edges could be dropped into place in 6!, or 720, different ways: the first edge into six slots, the second into five and so on. The requirement that the permutation be even, however, divides this by 2, to give 360. Moreover, if each edge is unconstrained, it could be in either of its two orientations, thus giving 2^6, or 64, different possibilities, but once again we must divide by 2 because of the flipping constraint, thus getting 32 distinct flip states. Multiplying these two figures together, we come up with 11,520 "interestingly different" states of the Pyraminx. Of course, if you want to take into account the middle blocks and the tips, each of them has 3^4, or 81, ways of twisting, and they are quite unconstrained, so that you can inflate the figure up to 75,582,720 distinct scramblings! Perhaps the most realistic figure discounts the tip orientations but counts the middle blocks. In that case one has $81 \times 11{,}520$, or 933,120, "nontrivially distinct" states of the Pyraminx.

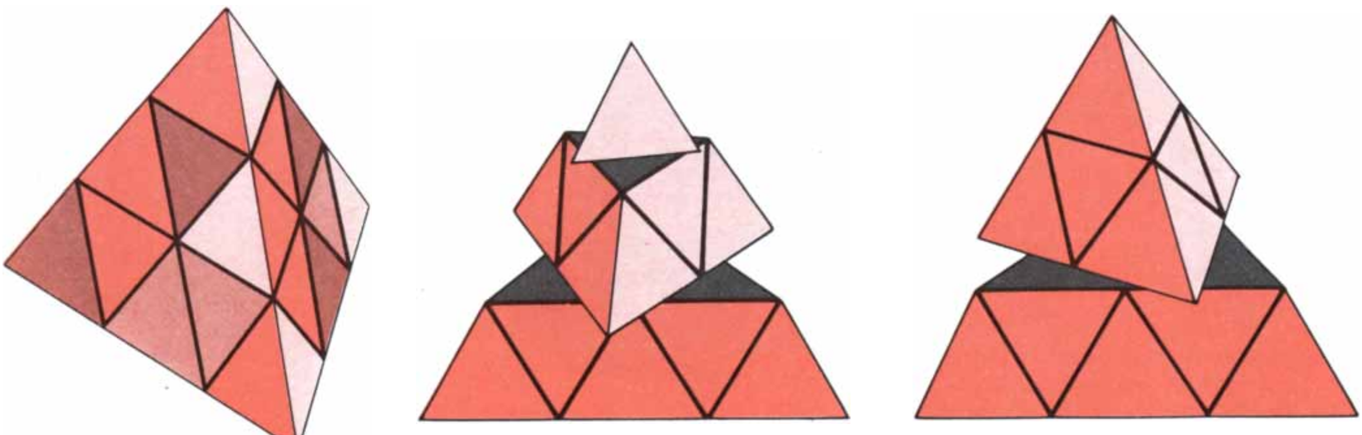
The shortest solving algorithm now known, discovered with the aid of a computer, takes 21 twists. It is easy to prove that from some positions one needs at least 12 twists to get back to Start, but the nature of "God's algorithm" (which always chooses the shortest possible route home) and the maximum number of twists it requires are unknown, as they are on the Cube.

When Mèffert designed the Pyraminx, he was quite aware that there were other ways to slice it up internally, even while keeping the same surface appearance, with nine trianglets per face. Therefore he figured out some alternate internal mechanisms that allow richer modes of twisting. The object I have just described is called the Popular Pyraminx. The Master Pyraminx is a different version and is slated to become available early next year. On it, above and beyond all the movements of the Popular Pyraminx, each edge can swivel about its midpoint by 180 degrees, thereby allowing the exchange of any two tips along with the flipping of a single edge piece [*see illustration at top right on opposite page*]. The flexibility requires each middle block to break up into several pieces as well, some of which can travel all around the pyramid.

Thus one has a much more complicated puzzle. The mechanism is exceedingly tricky, because in the course of such swiveling each of the two moving tips is in contact with the rest of the Pyraminx through a little invisible piece inside the now broken-up middle block. That little piece does not know to which edge it owes allegiance. As a result the invisible piece and the tip would together fall off if it were not for a clever bit of engineering that allows each tip to "lock" its little piece to the appropriate edge before the swiveling starts and then to "unlock" it after the swiveling is done. Turner-Smith cites the number of scrambled states of the Master Pyraminx as being in excess of 446 trillion.

Once bitten by the cube bug, Mèffert did not stop here but moved further into the world of regular polyhedrons. His next step was to design an eight-colored octahedron each of whose triangular faces is again divided into nine trianglets. How does it twist? Just as with the Pyraminx, Mèffert perceived the possibility of various modes of twisting. It is interesting that the two equivalent ways of describing the twists of the Popular Pyraminx become inequivalent when they are applied to the octahedron. Remember that they involved twisting either faces or small pyramids. The reason they were essentially equivalent is that the rotation of a face is complementary to the rotation of a small pyramid. On an octahedron, however, rotating a face 120 degrees is obviously not complementary to spinning a small pyramid (centered on a vertex) 90 degrees [*see top illustration on page 20*]. Realizing this extra degree of freedom, Mèffert designed a mechanism for each of the two ways of turning.

The octahedron that will soon be marketed (under the unsurprising name Pyraminx Magic Octahedron) is the one in which the six small pyramids can spin. Hence there are three orthogonal axes of rotation, just as in the Cube. This seemingly trivial connection with the Cube actually embodies much more than a grain of truth. In fact, the Mèffert octahedron and the Cube amount to two surface manifestations of one deep abstract idea. To see how this comes about, note that a cube and an octahedron are *dual* to each other: the face centers of one shape form the vertexes of the other shape. Thus the six face cen-



The Pyraminx and how it rotates

ters of a cube define an octahedron and the eight face centers of an octahedron define a cube.

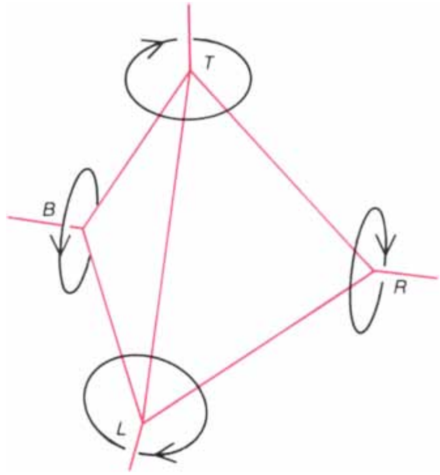
Imagine a Cube and, sitting inside it, the octahedron its face centers define [*see illustration at bottom left on next page*]. Each twist of a face of the Cube induces a twist on the corresponding pyramid of the octahedron. Each scrambled position of the Cube seems then to correspond to a scrambled position of the octahedron. However, this is not quite true. To see what is correct one needs to see what maps onto what in the Cube-octahedron correspondence. Like the Popular Pyraminx, the octahedron has tips, middle pieces and edges. As before, the tips are largely ornamental and the middle pieces rotate as wholes. Hence a middle piece on the octahedron (together with its decorative tip) maps onto a face center on the Cube. This leaves only edge pieces on the octahedron, and it is apparent that these, having two "facelets," must map onto edge pieces on the Cube. Where does this leave the Cube's corners? Nowhere. They have no analogue on the octahedron, which is a simplification.

In order to visualize the Cube-octahedron correspondence properly you must color one of the puzzles in an alternate manner. Since the Cube is the more familiar of the two, let us see how it has to be altered to "become" a Magic Octahedron. The proper coloring is corner-centered rather than face-centered [*see illustration at bottom right on next page*]. Stan Isaacs, a computer scientist and puzzlist par excellence, has made up one of his dozens of cubes to simulate a Mèffert octahedron.

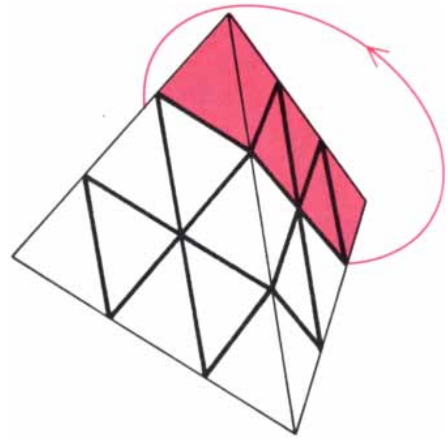
People fluent in solving the ordinarily colored $3 \times 3 \times 3$ Cube will therefore find that their expertise does not suffice to handle Isaacs' strangely colored cube, because now the orientation of the face centers matters. On the other hand, there is a corresponding simplification as well: on this cube "quarks" no longer exist. That is, there is no such thing as a twisted corner or even a misplaced one, simply because all the corner cubelets are white on all sides. On the standard Cube, quarks (corners whose twist is $+1/3$, or 120 degrees clockwise) and antiquarks (with $-1/3$ twist, or 120 degrees counterclockwise) always add up to make a total twist that is integral.

All you need to solve this cube (or the octahedron) is the ability to restore the edges and face centers (with the added novelty of orientations). Of course, not all magic octahedrons will be equivalent to simple recolorings of the $3 \times 3 \times 3$ cube, since they may not turn about those three axes. In particular, Mèffert's alternate twisting mode for the octahedron (where faces twist 120 degrees) is quite unrelated to the Cube.

Mèffert, in his 1982 catalogue, shows a picture of an icosahedron (guess what



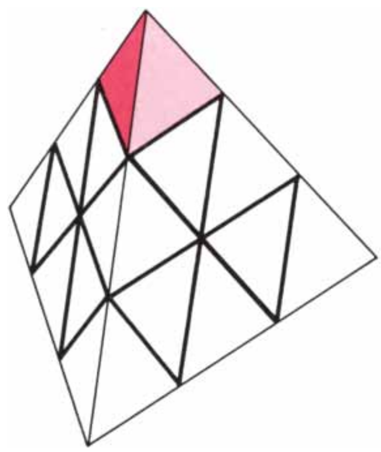
Twist notation for the Pyraminx



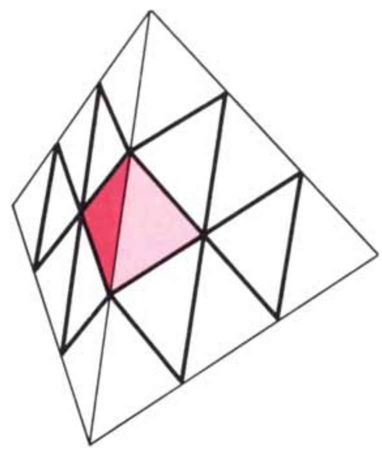
Second twisting mode for Senior Pyraminx

its name is) whose 20 triangular faces are not subdivided at all; they move five at a time, swirling about any of the 12 vertexes [*see illustration at left on page 23*]. Since the movement is vertex-centered rather than face-centered, it should make you think of its dual solid, the dodecahedron. The dual puzzle would have face-centered movement, just as the puzzle dual to the octahedron, with its vertex-centered movement, is the Cube, with its face-centered movement. (Incidentally, what would be the puzzle dual to the Pyraminx?)

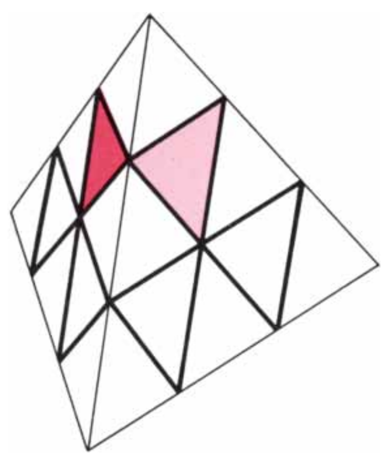
Actually in Mèffert's catalogue there are two other dodecahedral puzzles.



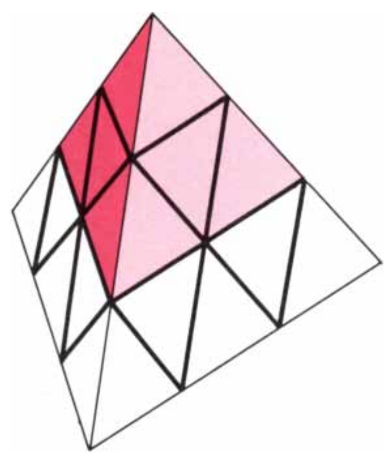
A tip on the Pyraminx



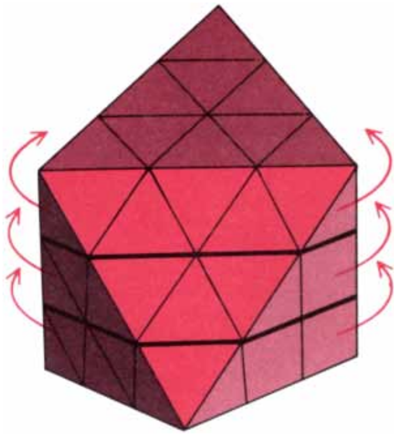
An edge piece on the Pyraminx



A middle block on the Pyraminx



A small pyramid on the Pyraminx



Mèffert's Octahedron **(left)** *and an alternate twisting mode for it* **(right)**

The less complicated one is called the Pyraminx Ball and the beautifully criss-crossed one is called the Pyraminx Crystal [*see middle and right illustrations on page 23*]. The Ball has four axes of rotation, like the Pyraminx; the Crystal has six. They should be hitting the market by the time this issue appears.

At this point you might well be wondering whether there could be a cube—I mean a genuine, six-sided, square-faced cube—with a vertex-centered twisting mechanism. Ah! No sooner said than done. In fact, that is the skew-twisting cubical object shown both in the illustration at the left on page 16 and in the illustration at the top left on page 24. To see it in the act of twisting, look at the illustration at the top right on page 24. Tony Durham, a British journalist, was the first to think of the idea. He showed his design to Mèffert, who developed it into a marketable product incorporating mechanical features that had proved useful in the Pyraminx. I like to call this one the Skewb, although Mèffert gives it the more prosaic name Pyraminx Cube.

Each of the Skewb's four cuts slices the whole into two equal halves. Each cut perpendicularly bisects one of the four spatial diagonals of the cube. If you think about it, you will see that the shape traced out by each cut as you run around the cube's surface is a perfect hexagon. Each cut crosses all six faces, so that every turn affects all the faces at once. In this respect the Skewb is more vicious than the Cube, where on each turn two faces are exempt from change. In spite of the simplicity of this object, it is quite hard to get used to its skew twist. To be sure, that is part of its charm.

Durham offers some perceptive commentary on his invention in a remarkable set of notes he has written titled "Four-Axis Puzzles." I should like to quote a few paragraphs from them:

"The symmetry group generated by four threefold axes is the rotation group of the tetrahedron, and has order 12. Almost all the well-known polyhedra, regular as well as semiregular, possess this tetrahedral symmetry, though their own symmetry may be much richer. So a four-axis mechanism may be put inside a polyhedral puzzle of any regular or semiregular shape, and the puzzle will keep its shape during play. The Pyraminx Ball may look odd at first glance, but it illustrates the way in which tetrahedral symmetry is buried in the richer symmetry of the dodecahedron.

"Rubik's cube mechanism does not have this property. It uses fourfold rotation axes, which are generally found only in the cube/octahedron family of solids. Thus it is possible to 'build out' a Rubik cube into the shape of a dodecahedron. But to preserve that shape you must restrict yourself to half-turns. Quarter-turns invoke a symmetry the dodecahedron does not possess.

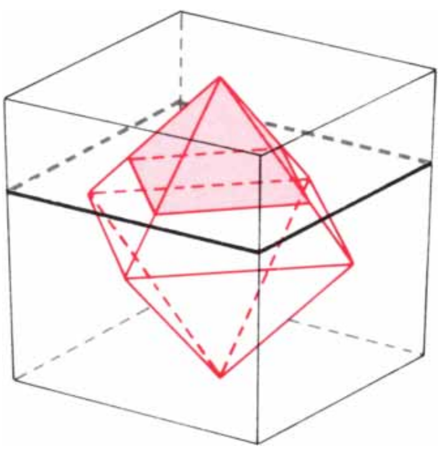
"All four-axis puzzles have a central ball or spindle. Four pieces (usually corners) are pinned directly to the ball. The standard Pyraminx has six free-floating edge pieces with 'wings' that hook under the corner pieces. The analogous free-floating pieces on the Pyraminx Cube are the square face-centers. The four-faceted pieces on the dodecahedral Pyraminx Ball play the same role.

"The Pyraminx Cube and Ball have four more free-floating pieces, which again are corners. These pieces have their own 'wings' which, in the Start position, hook under the first set of free-floating pieces. Thus there is a three-level hierarchy of interlocking pieces, conceptually similar to Rubik's, but geometrically very different.

"All eight corners of the Pyraminx Cube look alike. At first sight one might think that any two corners could be made to change places. In fact, four of the corners are free-floating and four are rigidly fixed to the central ball. The two types can never change place. The square shape of the face center pieces is deceptive, too. Inside, the mechanical parts of the square pieces are not so symmetrical. Such a piece can never return to its starting position (relative to the rigid set of four corners) rotated by 90 degrees. Only half-turns are possible.

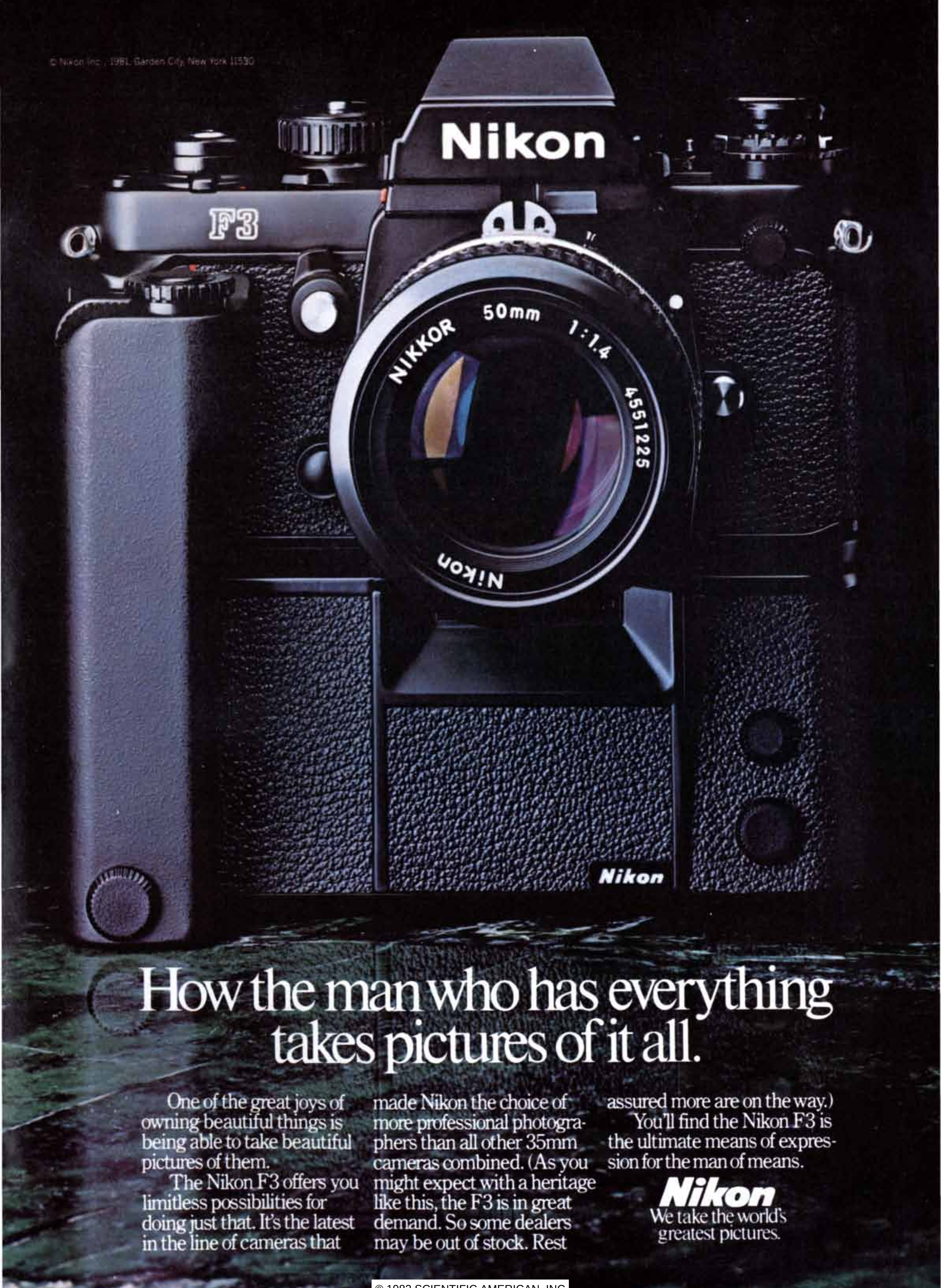
"The standard Pyraminx has obvious fixed points—the four corners. Confronted with a Pyraminx Cube and knowing that four corners are fixed and four are free, one naturally wonders which are which. Actually it makes no difference. The four free corners move independently of the fixed ones, but they always move together as if physically linked."

Durham proceeds to give Turner-Smith's *TBLR* notation for the Pyraminx, and he mentions that it can be adapted to any four-axis puzzle (such as his Skewb) simply by letting *TBLR* name four of the centers of rotation. (On the Pyraminx this could mean either the four tips or the four face centers. On the Skewb it would be four of the tips, leaving four others unnamed.) Then any move can be transcribed. If it is centered on one of the named spots, just use the proper notation. If it is centered on one of the four unnamed spots, use the name for the complementary move, since it



Cube and Octahedron rotations are related

Isaacs' cube simulates Mèffert's Octahedron


© Nikon Inc., 1981, Garden City, New York 11530
Nikon
F3
NIKKOR 50mm 1:1.4 4551225
Nikon
Nikon
How the man who has everything takes pictures of it all.
One of the great joys of owning beautiful things is being able to take beautiful pictures of them.
The Nikon F3 offers you limitless possibilities for doing just that. It's the latest in the line of cameras that made Nikon the choice of more professional photographers than all other 35mm cameras combined. (As you might expect with a heritage like this, the F3 is in great demand. So some dealers may be out of stock. Rest assured more are on the way.)
You'll find the Nikon F3 is the ultimate means of expression for the man of means.
Nikon
We take the world's greatest pictures.

does not matter which half of the puzzle twists. (You may want to think about that for a moment. Actually it is obvious, but it sounds like a tricky point.) Durham points out that it is sometimes useful to have names for the four remaining spots and for twists around them. He lets *t, b, l* and *r* fulfill that purpose. Thus *T* and *t* accomplish the same thing internally to the puzzle, but they leave it hovering in space in a different overall orientation. Although Durham concedes that a mixed notation may become confusing, he advocates using it on occasion.

"Sometimes you need to mix the notations to see what is going on. *TbT'b'* is one of the useful class of moves called 'commutators' (two moves followed by their inverses—thus of the form *xyx'y'*), though you would never guess so from its description in regular coordinates (*TBL'B'*) or alternate coordinates (*tlt'b'*).

"The Pyraminx Cube and Ball may be described as 'deep-cut' puzzles in contrast to 'shallow-cut' puzzles such as Rubik's Cube. In the latter, the cuts are made near to the surface. In deep-cut puzzles they slash close to the puzzle's heart. The bulk of a shallow-cut puzzle remains stationary while you turn a small part of it. A deep-cut puzzle, however, raises serious doubt as to which part has been turned and which has remained stationary. This is why alternate sets of coordinates have to be taken seriously on deep-cut puzzles.

"Deep-cut puzzles also dictate a 'global' approach to solution. It is peculiarly difficult to work on one area of the puzzle without affecting the rest. However, as solution proceeds, this very fact comes to your aid. Pairs of corners magically untwist in synchrony. The last flip, the last swap is done for you automatically. As you close in for the kill, billions of pathways down which the puzzle might escape are closed off to it. Parity constraints are at work, and when every move activates five or eight interlocked permutation cycles—as it does in a deep-cut puzzle—parity constraints are powerful."

In the section of Durham's notes having to do with parity constraints he includes the following apology: "Please forgive the loose use of the term 'parity' to include tests for divisibility by 3 (not only 2) or even more distant concepts. We shall use the term 'parity restriction' for any constraint on imaginable transformations of the puzzle that prevents their accomplishment in normal operation of the puzzle. The list does not, for example, include the rule: 'Thou shalt not swap a face piece with a corner piece.' It is just too far-fetched. One might as well try to imagine a move that transformed the entire puzzle into depleted uranium or Gorgonzola cheese."

Then Durham lists all the Skewb's "parity" constraints, in his generalized sense of the term.

"1. The four (fixed) corners *TBLR* may be permuted among themselves, as may the remaining four corners *tblr*, but mixing the two sets is prohibited.

"2. *TBLR* themselves move as a rigid tetrahedral unit. This constraint applies to their positions in space only (not to their orientations).

"2*a*. For exactly the same reasons, the remaining four (free) corners *tblr* move as a tetrahedral unit. They move independently of *TBLR*. In fact, any of the 12 possible relative positions of *tblr* and *TBLR* can be reached in at most two puzzle moves.

"Although *TBLR* are fixed and *tblr* are free-floating, mathematically speaking, 2 and 2*a* have exactly the same status. Writers on the Rubik Cube have generally regarded the transposition of two face centers as an 'unimaginable' transformation, while the swapping of two edge pieces is 'prohibited but imaginable.' By analogy with this convention, 2*a* counts as a parity restriction while 2 does not! This is plainly unsatisfactory, and a better and more precise definition of 'parity' is badly needed. Is it a question of geometry? Of mechanics? Of topology? Note that the problem is in enumerating *impossible* positions. *Possible* positions are readily counted.

"3. The sum of the twists of corners *TBLR* is always equal, modulo 3, to the twistedness of the puzzle, taken as a whole.

"(Here, 'twist' applies to corners, and is either 0, +1 or −1. A corner's twist is measured relative to the rigid tetrahedron to which it belongs. Thus the twist of *T* is measured relative to *TBLR*. A clockwise rotation of a corner counts as +1, counterclockwise as −1. By contrast, the 'twistedness' of the puzzle as a whole is a function only of the *positions* of the corners, not of their orientations. If the relative positions of *TBLR* and *tblr* are as in the Start position, then the twistedness is zero. If they can be restored to Start by one clockwise puzzle move, the twistedness is −1, and if by one counterclockwise move, then +1. If it takes one of each type, then the twistedness is again zero.)

"3*a*. Same as 3, only with *tblr*.

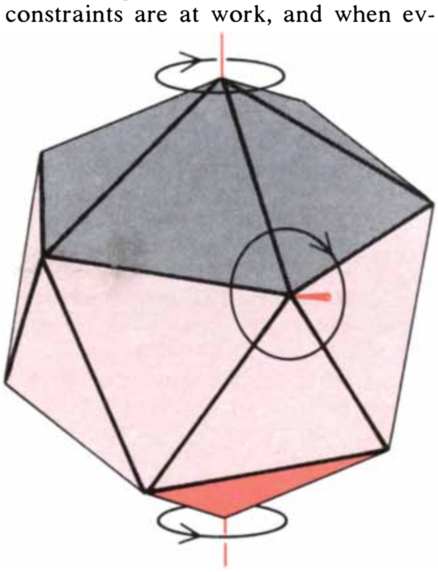
"From 3 and 3*a*, it follows that the total twist of *TBLR* always equals the total twist of *tblr*. Also, it follows that it is impossible to turn a single corner by 120 degrees (i.e., to create an isolated quark). One might paraphrase 3 and 3*a* by saying that the puzzle 'knows,' in three distinct ways, how many turns it is away from start (modulo 3).

"4. It is impossible to transpose exactly two face pieces.

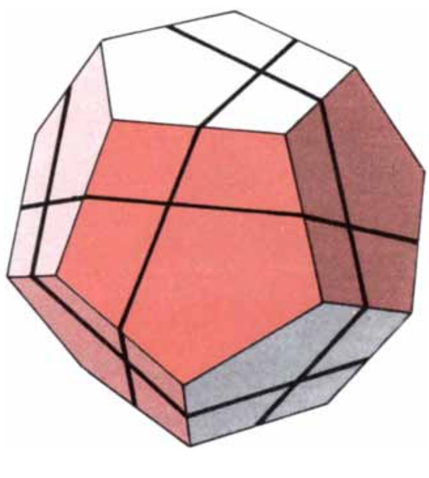
"5. It is impossible for any face piece to turn in place by 90 degrees.

"6. It is impossible to flip a single face piece through 180 degrees."

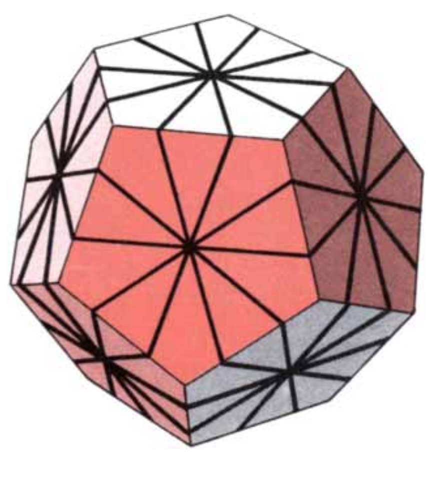
Durham offers proofs of these interesting facts, but because they are for the most part analogous to those on the Cube I shall omit them here. By combining all these constraints Durham comes



The Pyraminx Magic Icosahedron



The Pyraminx Ball



The Pyraminx Magic Crystal

up with the total number of scrambled states of his Skewb, which is 100,776,960. This assumes, however, that you have a way of telling the orientation of a face center, which (unless you mark it up) you don't. Hence the number of *visually distinguishable* states is reduced by five factors of two, to 3,149,280. It is a number rather smaller than the one for the Cube (4×10^{19}), but certainly the difficulty does not scale down proportionately with the number of states. (Could you even imagine what a puzzle 10 trillion times easier than Rubik's Cube would be like?)

Durham's final observations carry Solomon W. Golomb's beautiful analogy between cubological phenomena and those of particle physics to even greater heights. Golomb pointed out that many fundamental particles have their counterparts on the $3 \times 3 \times 3$ Cube. They include the quarks (q), antiquarks ($\bar{q}$), mesons ($q\bar{q}$ pairs), baryons and antibaryons (qqq and $\bar{q}\bar{q}\bar{q}$ trios). Durham extends the analogy as follows:

"The definition of 'twist' must be modified for the purpose of particle physics. A clockwise twist of one of the corners *TBLR* is now given the value of $+1/3$, as is a *counterclockwise* twist of any of the corners *tblr*. Either of these is a quark. Its opposite is an antiquark with value $-1/3$. It will be seen that twist corresponds to *baryon number*. The total twist of all corners is always an integer. A single puzzle move is always a meson.

"Quarks at the corners *TBLR* will be regarded as 'up' or *u* quarks; those at *tblr* will be 'down' or *d* quarks. Both quarks have isotopic spin 1/2. They are distinguished by the orientation of the isospin vector in its abstract space. The projection of the isospin, I_z, has the value $+1/2$ for the *u* quark and $-1/2$ for the *d* quark. In the absence of strangeness, charm, etc., the electric charge Q of a particle is given by $Q = I_z + B/2$, where B is the baryon number. So *u* quarks have charge 2/3, while *d* quarks have charge $-1/3$. (All the quantum numbers are multiplied by -1 for the antiquarks.) Again the puzzle models an important feature of observed reality: all particles have integral electric charge.

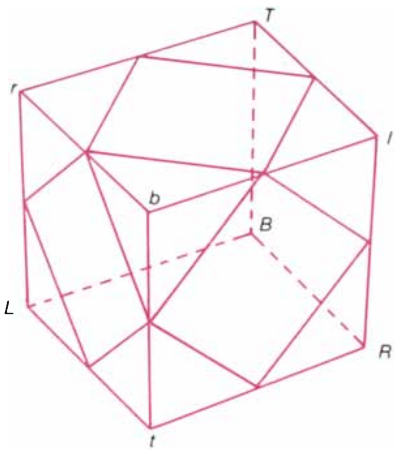
"The relevant quantum numbers for our two quarks are as follows:

	u	*d*
B	1/3	1/3
I	1/2	1/2
I_z	1/2	−1/2
Q	2/3	−1/3

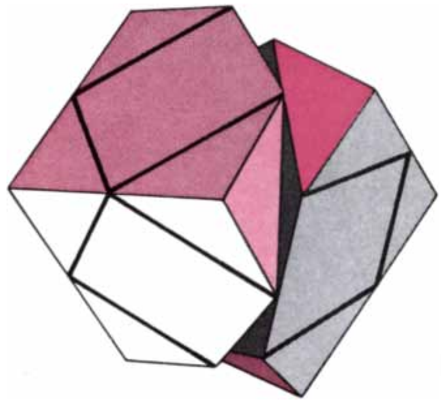
"We can now assemble various 'hadrons,' or strongly interacting particles [*see top illustration on opposite page*]. Each particle is represented by two rows having four symbols each. The four places in the top row represent the twists on the *TBLR* corners; in the bottom row the same is done for the *tblr* corners. A quark is denoted by '+,' an antiquark by '−.'

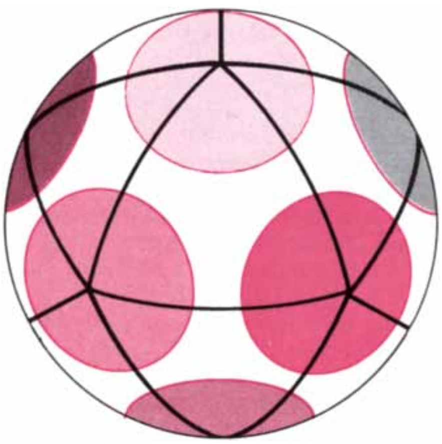
"Isotopic symmetry is a global symmetry, and the strong (nuclear) force is invariant under transformations that rotate the isotopic spin vector by the same amount for all particles. Such a transformation would, in a continuous fashion, transform all *u* quarks into *d* quarks and vice versa. Protons and neutrons would swap roles. The analogous process for the puzzle is the *continuous rotation of the whole puzzle* in space. It can indeed bring the *TBLR* corners to the former position of the *tblr* corners, so that an up quark becomes a down quark.

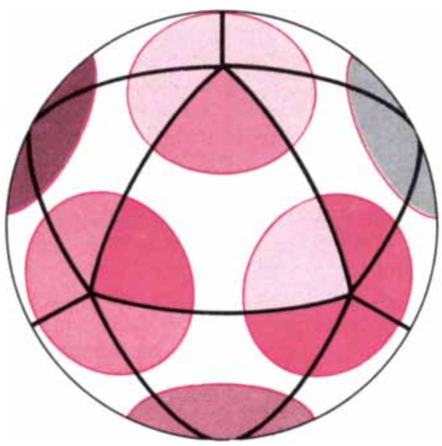
"This makes no difference to the 'strong interaction' (i.e., the normal operation of the puzzle). The *TBLR* and *tblr* corners are functionally identical. But it matters, if you try to dismantle the puzzle: you will find that one set of corners is fixed to the core, and one is



Names for the Skewb's eight corners


The Skewb in mid-twist


The IncrediBall in its pristine state


The IncrediBall with one quark visible


The IncrediBall in a "bumpy twist"


The IncrediBall with one icosalet removed

not. Such dismantling operations can be thought of as weak or electromagnetic interactions, which can break the conservation rules obeyed by the strong interaction. Actually they break the rules rather too well, since they allow the creation of single free quarks."

Durham points out that the analogy still has weaknesses, such as the fact that neither charge nor baryon number is conserved, that there is no analogue to spin, that only two "flavors" of quark are represented (up and down) and that quark "color" is not modeled. Golomb in the meantime has been actively trying to find a way of modeling quark color in the 3 × 3 × 3 Cube analogy.

Whatever the failings of this analogy, I find it one of the most provocative analogies I have ever encountered anywhere, and I shall be astonished if it is purely coincidental. I somehow cannot help but believe that the fascinating patterns shared by these macroscopic puzzles and the microscopic particles reveal some underlying order and set of principles common to both. Indeed, I have faith that, if they are looked at in the proper way, the group-theoretical principles governing these parity constraints on "cubes" can be transferred to the domain of particle physics and can yield fresh insights into the reasons for the symmetries among particles. There! If *that* doesn't prod some particle physicist into looking into this, I don't know what will.

+ 0 0 0 − 0 0 0	π^+ meson ($u\bar{d}$)	+ + 0 0 + 0 0 0	proton (uud)	+ 0 0 0 + + 0 0	neutron (udd)
+ + + 0 0 0 0 0	Δ^{++} (uuu)	0 0 0 0 + + + 0	Δ^- (ddd)	− − 0 0 − 0 0 0	antiproton ($\bar{u}\bar{u}\bar{d}$)
− 0 0 0 + 0 0 0	π^- meson ($\bar{u}d$)	+ − 0 0 0 0 0 0	η^0 meson ($u\bar{u}$)	0 0 0 0 + − 0 0	π^0 meson ($d\bar{d}$)

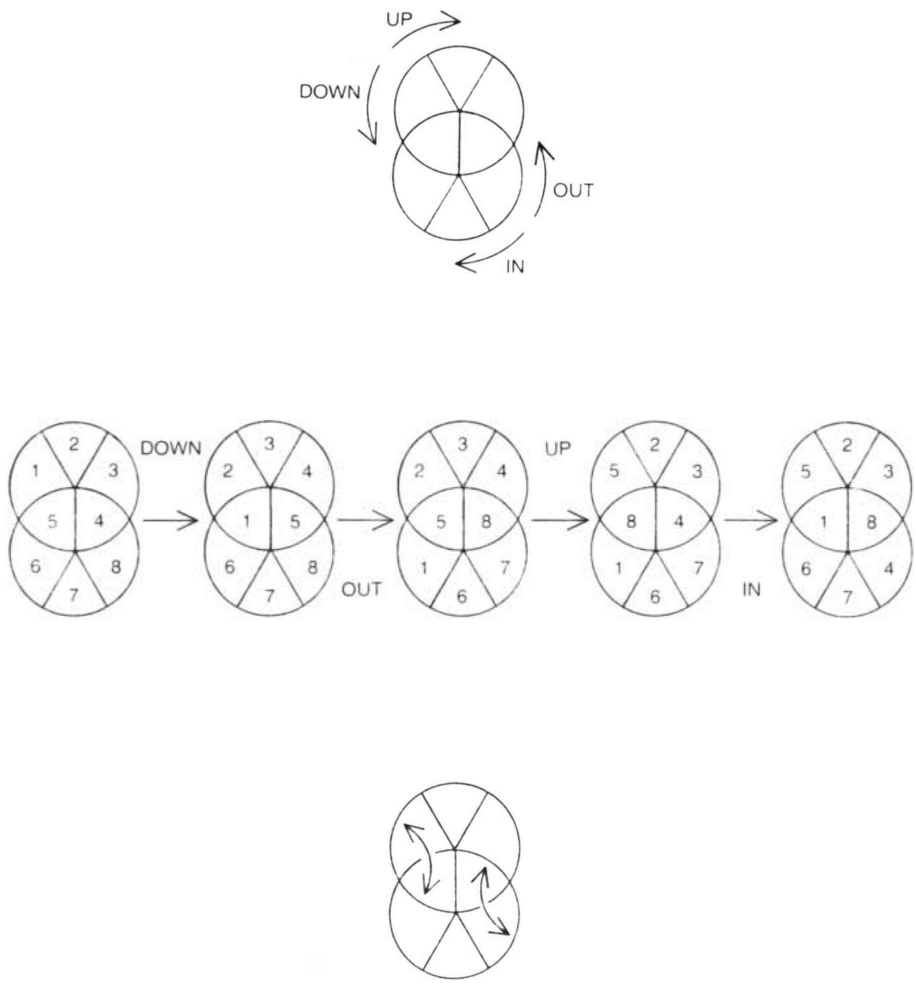
"Hadrons" are assembled from Cube quarks

Perhaps my favorite "cube" is the one I dubbed the IncrediBall [*see illustration at left in middle of opposite page*]. It was devised by a German educator from Dortmund named Wolfgang Küppers, and it is in Mèffert's catalogue. As of this writing, I may be the world's fastest IncrediBall solver (or at least the fastest on my block), with an average time of about five minutes. I am sure, however, that my glory will not last long once this puzzle is marketed by the Milton Bradley Company sometime this summer. Their trade name for it will be Impossi*Ball.

This I-ball is basically a rounded-off dodecahedron each of whose 12 faces (dodecalets, I'll call them) has been subdivided into five elementary trianglets. Thus there are 60 such trianglets. If instead of seeing them in groups of five, you take them *three* at a time, you will find they define a rounded-off icosahedron (the dual of the dodecahedron). Such a group of three trianglets I call an icosalet, and there are 20 of them, each with a unique arrangement of three colors. The icosalets are the elementary, unbreakable units out of which the IncrediBall is constructed; they correspond to the cubelets on the Cube or the elementary blocks of the Pyraminx. Whereas on the Cube there are three kinds of cubelets (edges, faces and corners), here all the icosalets are of a single type. For this reason the ball is less forbidding than it might appear at first. Its pristine state is one in which each dodecalet is all one color. Mèffert has used only six colors rather than 12, each color being used in two antipodal dodecalets, but this does not change the difficulty of the puzzle in any way.

The way the I-ball turns is a little surprising. Any group of five icosalets that meet at a point (the center of a dodecalet) form what I call a "circle." It will rotate as a unit, twisting 72 degrees to the left or the right. Hence five such twists return that group to its starting position. A "circle" is analogous to a "layer" (a face together with a fringe) on the Cube. The "circle" defined by the boundary of five icosalets is not, however, truly circular, and if the icosalets were rigidly held at a fixed distance from the center, it simply would not be possible to rotate such a group. Mèffert's mechanism ingeniously gets around the problem by having the icosalets lift up slightly as they go over "bumps," so that the solid flexes noticeably. As a result twisting the I-ball has a delightful "organic" feel to it.

The constraints here are the same old thing: all permutations are even, which means you cannot swap two icosa-



Stepping through the commutator "down-out-up-in" (doui). The result is at the bottom lets. The best you can do is cycle three of them or swap two pairs simultaneously, and of course quarks and antiquarks must add up to a total twist that is integral. Taking into account these constraints, I calculate that the total number of IncrediBall scramblings is 23,563,902,142,421,896,679,424,000, or 24×10^{24}—about 24 trillion trillion. This is not quite a million times larger than the figure for Rubik's Cube.

How hard is it to solve this puzzle? Is it harder than the Cube? I found it easier, but that is hardly fair, since I had already done the Cube. In Durham's terms, however, the IncrediBall is decidedly a "shallow cut" puzzle, which means that a more or less local approach will work. I found that when I loosened my conceptual grip on the exact qualities of my hard-won operators for the Cube and took them more metaphorically, I could transfer some of my expertise from Cube to I-ball. Not everything transferred, needless to say. What pleased me most was when I discovered that my "quarkscrew" and "antiquarkscrew" were directly exportable. Of course, it took a while to determine what such an export would consist of. What is the essence of a move? What aspects of it are provincial and shedable? How can one learn to tell easily? These are difficult questions for which I do not have the answers.

I gradually learned my way around the IncrediBall by realizing that a powerful class of moves consists of turning only two overlapping "circles" in a commutator pattern (*xyx'y'*). I therefore studied such two-circle commutators on paper until I found ones that filled all my objectives [*see illustration on preceding page*]. They included quarkscrews, double swaps and 3-cycles, which form the basis of a complete solution. In doing so I came up with just barely enough notation to cover my needs, but I did not develop a complete notation for the IncrediBall. This, it seems to me, would be most useful: a standard universal notation, psychologically as well as mathematically satisfying, for all cubelike puzzles. It is, however, a very ambitious project, given that you would have to anticipate all conceivable variations on this fertile theme, which is hardly a trivial undertaking.

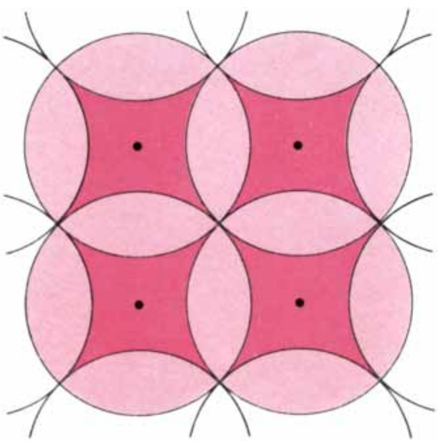
It is interesting that my diagrams of overlapping circles turn out to be closely connected with another lovely family of generalizations of the Cube, arrived at by a Spanish physicist named Gabriel Lorente. His puzzles are mostly planar, and they consist precisely of networks of overlapping circles. He calls the planar puzzles the Grill and the Trebol [*see illustrations on this page*]. In each of them circles can be given partial twists and pieces of them are thereby shuffled and redistributed. Extending this notion to a spherical surface, Lorente came up with an elegant IncrediBall-like puzzle that he calls the Florid Sphere.

When you look closely at Lorente's puzzles, the IncrediBall and even the Cube, you begin to see that the essence of all of them seems to lie in such overlapping orbits. In fact, one could even maintain that the three-dimensionality of these puzzles is irrelevant; their interest essentially lies only in the properties of intricately overlapping closed orbits in a two-dimensional space, possibly curved like a sphere.

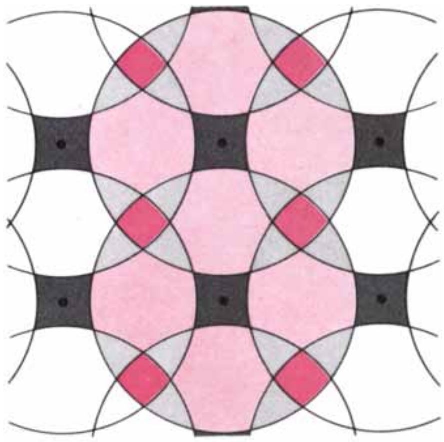
The quintessential planar overlapping-circle puzzle was invented, as it turns out, back in the 1890's, although recently it has been repeatedly discovered in the wake of the Cube. All such puzzles basically involve two circles of marbles that intersect at various points [*see top illustration on page 30*]. You can choose to cycle either circle, and the marbles at the intersections will thus be absorbed into whichever circle is moving.

While we are considering two-dimensionality, it is worthwhile pointing out that the IncrediBall's internal construction allows it to be transformed rather amazingly into what I call the 19 puzzle: a two-dimensional curved-space version of Sam Loyd's famous 15 puzzle (the 4×4 square puzzle with one "squarelet" removed, enabling you to rearrange the remaining 15 squarelets by shifting the hole). This was first observed by both Halpern and Mèffert. They found that if you remove a single icosalet (which is possible, one of the beauties of the IncrediBall being that its mechanism readily allows disassembly and reassembly), the hole it leaves can now wander all over the sphere, just like the square hole in the 15 puzzle [*see illustration at bottom right on page 24*]. This again seems to bring out the two-dimensional nature of these puzzles.

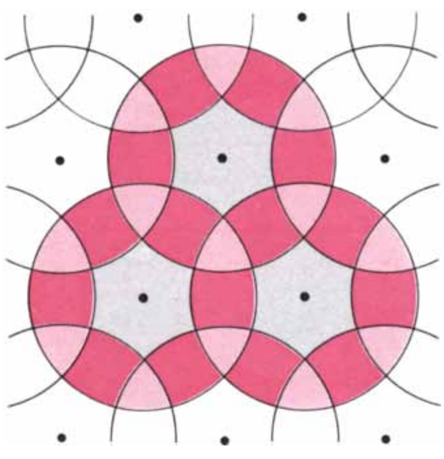
The view that these puzzles are two-dimensional arises from the fact that only pieces on their surfaces move; there is no exchange between the interior and the exterior. For an extreme case imagine the earth as a giant puzzle, its entire surface covered with trillions of overlapping circles of marbles. With a hundred million turns you could ship a marble from New York to San Francisco. Clearly this would in essence be a two-dimensional puzzle. The small-



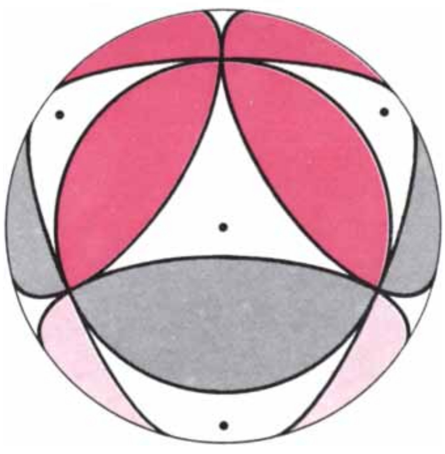
Lorente's Grill puzzle



Another Grill puzzle by Lorente



Lorente's Trebol puzzle



Lorente's Florid Sphere

Finally you can afford to satisfy your lust for power.


sinclair
ZX81
EDIT AND THEN TO GRAPHICS DELETE
1 2 3 4 5 6 7 8 9 Ø
PLOT UNPLOT REM RUN RAND RETURN IF INPUT POKE PRINT
"" OR STEP <= <> >= $ ( ) "
Q W E R T Y U I O P
SIN COS TAN INT RND STR$ CHR$ CODE PEEK TAB
NEW SAVE DIM FOR GOTO GOSUB LOAD LIST LET
STOP LPRINT SLOW FAST LLIST ** − + = FUNCTION
A S D F G H J K L ENTER
ARCSIN ARCCOS ARCTAN SGN ABS SQR VAL LEN USR
COPY CLEAR CONT CLS SCROLL NEXT PAUSE BREAK
SHIFT
: ; ? / * < > , £
Z X C V B N M . SPACE
LN EXP AT IN KEY$ NOT π

For $99.95 you can have a full powered personal computer.

Most people know by now that the ZX81 from Sinclair Research is the lowest priced personal computer in the world.

But serious programmers are looking for more than a low price. They're looking for true computer power. And that's where the ZX81 surprises a lot of people.

Just look at the keyboard and you'll get some idea of the ZX81 's power. It has more than 60 BASIC commands, 20 graphic symbols, and complete mathematical functions. And there's even more power that you can't see.

Sinclair programs are available

A breakthrough in personal computers. The ZX81 offers features found only on computers costing two or three times as much.

Just look at what you get:

- Continuous display, including moving graphics
- Multi-dimensional string and numerical arrays
- Mathematical and scientific functions accurate to 8 decimal places
- Unique one-touch entry of key words like PRINT, RUN and LIST
- Automatic syntax error detection and easy editing
- Randomize function useful for both games and serious applications
- Built-in interface for ZX Printer
- 1K of memory expandable to 16K
- A comprehensive programming guide and operating manual

16K Memory Module

The ZX81 is also very convenient to use. It hooks up to any television set to produce a clear 32-column by 24-line display. And you can use a regular cassette recorder to store and recall programs by name.

What you get. When you order your ZX81, you get everything you need to start programming.

It comes with connectors for your TV and cassette recorder, an AC adaptor, and a free programming guide and operating manual that completely documents the capabilities of the ZX81.

Options and add-ons. Like any full-powered computer, the ZX81 can be expanded and upgraded.

Its 1K memory can be expanded to over 16K just by plugging the Sinclair Memory Module onto the back of the unit. The cost is only $49.95.

Sinclair has also published pre-recorded programs on cassettes for your ZX81. We're constantly coming out with new programs, so we'll send you our latest software catalog when you order your computer.

How did we do it? The question most often asked about the ZX81 is, "How can so much computer power cost so little money?"

The answer is that Sinclair Research simply took a different approach. Our only goal was to make programming power as affordable as possible. So we developed a radical new design that cuts costs dramatically without cutting computer power. For example, our unique Master Chip replaces as many as 18 chips used in other personal computers.

The success of the ZX81 speaks for itself. It is now the fastest-selling personal computer in the world. And we stand behind our product. If anything goes wrong in the first 90 days, we'll repair or replace your unit free of charge. Even after that, you can take advantage of our national service-by-mail facilities for a minimum fee.

Order now and try it out for 10 days. Simply send the coupon along with a check or money order. For faster delivery, call our toll-free number and use your MasterCard or VISA.

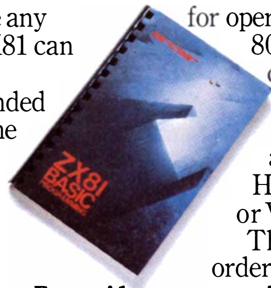
You have 10 days to try out the ZX81. If it isn't all we say it is, just send it back and we'll refund your money.

Why wait any longer? With the Sinclair ZX81, you can finally afford to have the computer power you've always wanted.

Call toll free 800-543-3000. Ask for operator #509. In Ohio call: 800-582-1364; in Canada call: 513-729-4300. Ask for operator #509.

Phones open 24 hours a day, 7 days a week. Have your MasterCard or VISA ready.

These numbers are for orders only. If you just want information, please write: Sinclair Research Ltd., 2 Sinclair Plaza, Nashua, NH 03061.



Free guide to programming

To order call toll free: 800-543-3000.

AD CODE 07SA

MAIL TO: Sinclair Research Ltd., One Sinclair Plaza, Nashua, NH 03061.

	PRICE*	QTY.	AMOUNT
ZX81	$99.95		
16K Memory Module	$49.95		
Shipping and Handling	$4.95		$4.95
*U.S. dollars		TOTAL	

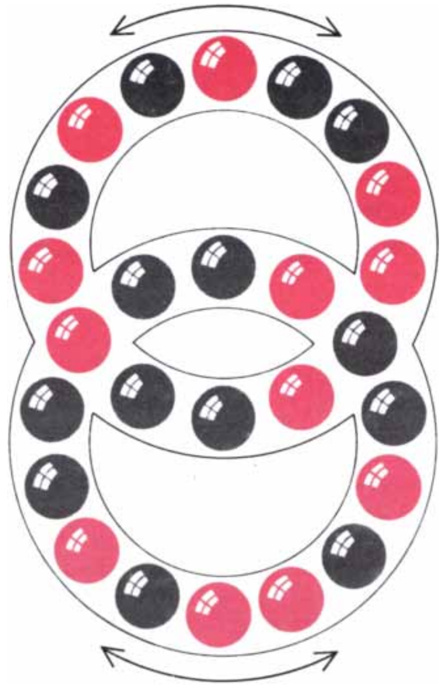
Name

Address

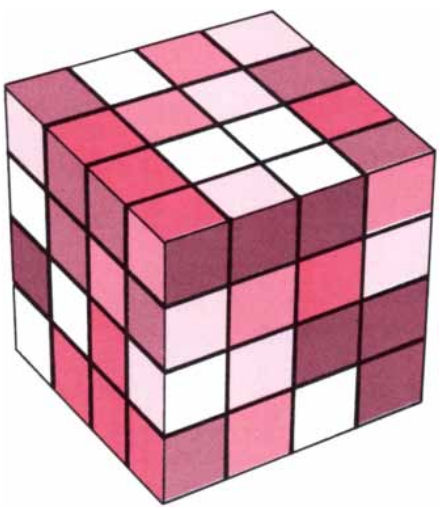
City State Zip

Sinclair technology is also available in Timex/Sinclair computers under a license from Sinclair Research Ltd.

sinclair



Puzzle with overlapping circles of marbles



The 4 × 4 × 4 Cube in a scrambled state

The Pyraminx Ultimate

ness of the circles with respect to the size of the earth makes this obvious. (One surely would not even want to think about *solving* such a puzzle.)

In contrast, consider two objects that will be released soon: Ideal's 4 × 4 × 4 cube, unimaginatively called Rubik's Revenge, and Mèffert's Pyraminx Ultimate, a 5 × 5 × 5 with shaved corners and edges [*see middle illustration on this page*]. (Incidentally, for impatient readers with home computers, a program called Quadcube that simulates a 4 × 4 × 4 Cube in color graphics is available for the Texas Instruments 99/4 and 99/4A computers and for the IBM Personal Computer, from Linear Aesthetic Systems, P.O. Box 23, West Cornwall, Conn. 06796, for $14.95 per cassette.) In both of these objects there are circles on a much more global scale. The 4 × 4 × 4 has an Arctic Circle, a Tropic of Cancer, a Tropic of Capricorn and an Antarctic Circle. The 5 × 5 × 5 has an Equator as well.

On the 3 × 3 × 3 cube one could get away with ignoring the Equator by describing equatorial twists in terms of their complements, like rotating the slices of bread in a sandwich without rotating the meat. Singmaster's notational choice for the 3 × 3 × 3 cube reflects his propensity for regarding face centers as stationary. Thus for him the bread slices move while the meat stays put. In principle this is fine but in practice people just do not hold their sandwiches—pardon me, their Cubes—in one fixed orientation. Moreover, when you go to higher orders, this view will not suffice. Imagine a multilayer club sandwich with three slices of bread and two kinds of meat. For this you simply must expand your notational horizons.

An elegant set of names for the six meat-slice, or equatorial, moves on the 3 × 3 × 3 Cube has been suggested by John H. Conway, Elwyn R. Berlekamp and Richard K. Guy in their book *Winning Ways* (Academic Press, 1981). They employ Greek letters with clever mnemonic justifications [*see top illustration on opposite page*]. With some modification the letters could be adapted to slices on higher-order cubes.

Slice moves of this more global kind are like giant circles of marbles stretching around the Equator or the Tropic of Capricorn; their radii are of the same order of magnitude as the radius of the underlying three-dimensional object. The topology of linkage of the circles becomes much more complicated than it is in the case where the circles are small and every connection is quite local. To describe the linkage economically one would be forced to talk about the way the circles are embedded in 3-space. In this sense the higher-order cubes can truly be said to be intrinsically three-dimensional puzzles.

There are, it seems, endless new spin-offs of the Cube being created. It is a very fertile idea. I have been sent a beautiful paper by H. J. Kamack and T. R. Keane, who have simulated a four-dimensional 3 × 3 × 3 × 3 cube on a computer. They call it Rubik's Tesseract, and the number of states it has is nearly a googol (10^{100}). (If Ideal were marketing it, they might say "over three trillion combinations.") Kamack and Keane have made many provocative discoveries, which I have no space to report on here. Interested readers can find out more by writing to Kamack at 610 Lehigh Road (Apartment 10), Newark, Del. 19711.

I was also sent a fascinating paper by György Marx and Eva Gajzágó, physicists at Roland Eötvös University in Budapest. In it they give a definition of "entropy" on the Cube and describe some statistical results computed by a grammar school student named Victor Zámbó. These are matters I should like to go into at some future time.

I should like to close by discussing the astonishing popularity of the Cube. In *The New York Times* paperback best-seller list of November 15, 1981, three cube booklets figured on the list. The positions they occupied? Nos. 1, 2 and 5. People often ask, "Why is the Cube so popular? Will it last or is it just some sort of fad?" My personal opinion is that it will last. I think the Cube has some kind of basic, instinctual, "primordial" appeal. Its conceptual pizzazz comes from the fact that it fits into a niche in the human mind that connects with many general notions about the world. The following is an attempt to characterize that quality.

To begin with, the Cube is small and colorful. It fits snugly in the hand and has a pleasing feel. Twisting is a fundamental and intriguing motion the hand performs naturally. The object itself has overall symmetry, so that it can be rotated as a whole without its feel changing. (This is in contrast to many puzzles that have at most one axis of symmetry.) Quite surprisingly, there are not many puzzles or toys that give the mind and fingers a genuine three-dimensional workout.

Although the object gets scrambled, it still stays whole. (This is in contrast to many Humpty-Dumptyish puzzles that come apart into scads of pieces that can get scattered all over the floor.) That the object manages to stay in one piece when it has so many independent ways of being twisted is initially amazing and remains mysterious even after you have seen its innards.

The object is a miniature incarnation of that subtle blend of order and chaos that our world is. Most of the time one cannot predict what repercussions even simple actions will have: they simply have too many side effects. A few tiny actions can have vast, interlocking consequences and become practically un-

undoable. One can easily become paralyzed by fear, not wanting to make any move at all, sensing that with no trouble at all one can get totally, irretrievably, hopelessly lost.

There are plenty of patterns, some attainable, some unattainable. Sometimes they are simple to generate but one cannot see how it happens. Sometimes they are hard to generate and one clearly understands how they arise.

There are many routes to any state, and the shortest one is nearly always completely unknowable. The solution to a difficult situation is hardly ever to back out the way you came in but to find an alternate and completely different escape route. One feels a little like someone trapped in a cave with no light, unable to sense the entire space, able only to feel about very locally, wondering whether it is even humanly possible to have such an overview.

The Cube is a rich source of metaphors. It furnishes analogies to particle physics (quarks, say), to biology (a move sequence as a "genotype" and the pattern it codes for as a "phenotype"), to problem solving in everyday life (breaking a problem down into parts, solving it stage by stage), to entropy and pathfinding. It even touches theology (God's algorithm).

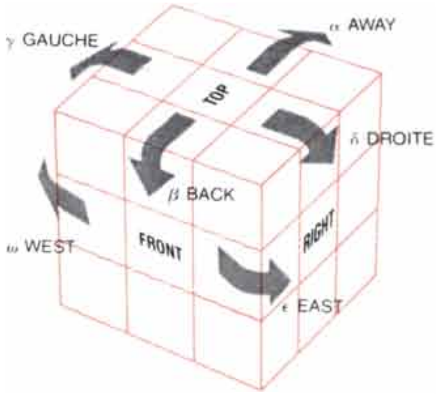
There are different approaches to understanding the Cube. In particular there is a strong contrast between the "algebraic" approach and the "geometric" one. In the algebraic, or mathematical, approach long sequences of operations are compounded out of shorter sequences, so that after a while one has no idea why one is doing the various individual twists; one just relies on the sequences to work as wholes. That is efficient but risky. In the geometric, or commonsense, approach the eye and the mind combine to choose twist after twist, each twist having a clear reason as part of a carefully charted pathway. That is inefficient but reliable. The two approaches can indeed serve as metaphors for styles of attacking problems.

The Cube's universe has a strange population. Its varieties of cubelets and modes of twisting aside, there are such ephemeral entities as "flippedness" or "twistedness," which one quite literally moves about on the Cube just as one moves the tangible cubelets. Similarly, the word "here" can designate a "place" that moves to and fro in a sequence of twists. The interlocking and nested frames of reference one jumps between in trying to restore order to the Cube vividly exemplify the layered way in which we conceive of space, indeed the layered way in which thought itself works.

Among the Cube's less intellectual charms are the magic of motion too swift for the eye, the thrills of speed, competition and grace, the varying levels of knowledge one can gain, the enjoyment of exchanging information and insight. And even after playing with the Cube for a long time one can still be amazed at how such a tiny, innocent object can conceal such a vast universe of potential.

Finally, consider the metaphor the Cube offers for the state of the world. (It is a metaphor that has in fact been exploited in numerous current political cartoons.) The globe is in a mess [*see illustration below*] and the leaders of various nations want it to be "fixed." They are unwilling, however, to relinquish any tiny bit of order they achieve. They cling to old, useless achievements because they are too fearful of letting go and temporarily abandoning what partial order they have in the pursuit of greater order and harmony. Can it be they lack a mature, global view, one recognizing that a willingness to make sacrifices in the short run can result in much greater gains in the long run?

I am confident that the Cube, and "cubes" in general, will flourish. I expect new varieties to appear for a long time to come, and to enrich our lives in many ways. It is gratifying a toy that so challenges the mind has found such worldwide success. I hear it is now popular in China. Perhaps it will even penetrate into the U.S.S.R., to my knowledge the last bastion of the Cube-free world.



Notation for slice moves on 3 × 3 × 3 cube

The sad state of the globe

BOOKS

Handedness, plant geography, tree rings and great spectacles of the solar system

by Philip Morrison

LATERAL PREFERENCES AND HUMAN BEHAVIOR, by Clare Porac and Stanley Coren. Springer-Verlag ($23.90). Many animals, from mice to gorillas, show consistent paw preference. Each individual is either right- or left-pawed, yet the overall incidence of right-pawedness approaches 50 percent for the population. Not so for human beings: in every age and every place for which we have data most people are right-handed. They are also right-footed, right-eyed and right-eared. Right-handedness runs to 85 or 90 percent, but right-earedness stands only at about 60 percent. Gross bodily asymmetry is not found; once it was a popular research goal, but nowadays the seekers study the brain, or perhaps development rates overall. The left hemisphere of the brain controls the motor movements of the right hand and the right foot, and it may first receive the signal of the right ear. Once it was realized that the same left hemisphere was the site of the primary speech centers, the link between human handedness and human speech was seductive. Left cerebral dominance became the mode; in pop psychology the left hand is now plainly the dreamer.

The stubborn facts are not so simple. These two British Columbia psychologist authors have examined some 20,000 subjects and thousands of papers, to offer it all up in a meticulous and appealing monograph. For instance, they sought out some 10,000 pieces of art from all places and times; nearly 1,200 of them showed recognizable handedness in the depicted use of tools and weapons. From 3000 B.C. to 1950 every sample of these representations gave the usual edge to right-handedness, without any trend over time. The result is nearly the same when the art is analyzed by place; there is a suggestion that the extremes (88 percent in the Americas and 96 percent in the Middle East) might differ beyond sampling error. The fact remains that in all the cross-cultural studies no clear pattern emerges from the noise.

It is the exceptions that bring down the naive view. Fewer than half of the cases in a large sample show complete right-sidedness in all four indexes: hand, foot, eye and ear. Nor is the brain architecture so simple. The right side of the brain shows a greater blood volume and pressure in right-handers than does the speech-fostering left; in left-handers it is the reverse. Both right- and left-handers have more pyramidal motor fibers on the right side. The pace of development is not one-sided either: the earliest gradients of growth favor the left side. Single-egg twins with identical genotypes do not always agree in sidedness; their concordance differs little from siblings of similar age or from chance expectation.

There is a dextral trend in some but not all preferences from early childhood to adulthood. Cultural pressure is plain, both overt and covert; forcible efforts to make children right-handed are notorious in certain times and places, and the tools of the world are subtly right-handed. But the right-handed world does not always prevail; right-ear preference actually declines with age. If there is any genetic factor, it is associated not with the right side but with the display of consistent rather than mixed preferences across the four choices.

Birth stress and deficiencies of many kinds have excited transient explanations of handedness. Sinistrality is a "soft sign" for damage to many a neurologist, but normal lefties show no clear neural differences at all; the clinical relation is itself one-sided. Of course, we are not really one-handed at all. The two hands work in mutual aid; handedness refers to the active partner in many two-handed forms of behavior. Similarly, eyedness is by no means simple. Preferences all show a kind of *U*-shaped distribution; not many people stand ambivalently in the middle. Proficiency with each of the four choices tends to a single-peak distribution, although it is one that is right-biased. The simple decision conceals the intricacy of the tasks; let the reader beware: "Reading the fine print in the methodology section may be as important as reading the main text." The discussions of reading, sports performance, cognitive abilities and sensory preferences bear out the point; these behaviors are complex, and the right-left choices are bins for many parameters of function.

A conclusion is not yet at hand. Certainly the most ingenious proposal is the one based on the fact that 80 percent of all mothers, both right- and left-handed ones, carry their newborn infants on the left side. This is supposed to derive from the calming effect of the heartbeat, so steadfast in utero. With the rise of human tool use right-handed mothers acquired a selective advantage, since they could work deftly with a free right hand. The genetic transmission of handedness, however, is at most debatable; the asserted advantage does not appear to be heritable.

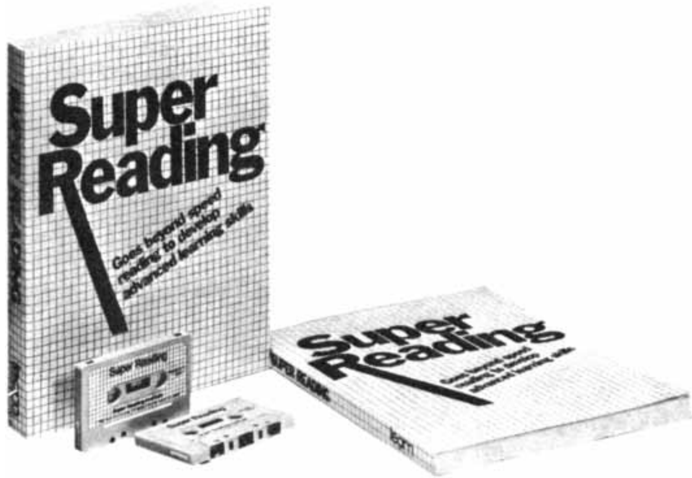
A newer speculative view, ascribed to a two-page paper in 1975 by a neurologist, P. Reynolds, clearly appeals to the authors, who remain models of good sense and objectivity to the last word. Since practice certainly does make perfect, we have a case for increased skill to follow any consistent early preference. Survival will then favor individuals with strong lateral preferences, a tendency that is the only heritable pattern now visible. Primates show lateral preference, although the side chosen does not matter. The weak relation between cerebral asymmetry and handedness is not at all incompatible with this view.

For human beings alone some early general bias toward the use of the right side could turn this impartial mechanism for a learned side specialization into an overall but not universal dextrality. Perhaps the faster maturation of the left cerebral hemisphere causes the infant hand to grasp for a few moments longer whenever it is in the right palm that the grasping reflex is excited. That hand will become preferred. The sensory choices will act rather differently, as indeed they do. On this view the tendency to left-sidedness (or at least to mixed laterality) associated with neuropathology is plain enough. Learning then is slower. Normal lefties are simply those who by chance experienced a few left-handed stimuli first and thus progressed to skillful sinistrality.

This work ends with Pooh looking thoughtfully at his two front paws: "He knew that one of them was the right, and he knew that when you had decided which one of them was the right, then the other one was the left, but he never could remember how to begin."

HISTORICAL PLANT GEOGRAPHY: AN INTRODUCTION, by Philip Stott. George Allen & Unwin, Inc., 9 Winchester Terrace, Winchester, Mass. 01890 ($30). For all the centuries of modern science botanists have been happily collecting and describing the 300,000 known species of flowering plants. The standardized herbarium sheets worldwide hold specimens by the tens of millions; a shelf of flora catalogues plants from almost every land. The first yield of this treasury is certainly knowledge

Speed Reading Breakthrough



The Super Reading System

A remarkable innovation that will launch you on a new way to read and think. You will double your reading speed and ability to learn and understand more of everything.

How many times have you stared at the pile of reading you must do for work, school, or just to keep current and wondered, "How will I ever be able to get to the bottom of the stack?" Sure, you could enroll in one of those 'speed reading' courses. Some cost as much as $400. But, you don't have the time . . . let alone the money. Even courses offered by mail can cost over $100. There must be a better way.

GOOD NEWS

We went to one of America's leading publishers of self-study educational programs and the man who is regarded as one of today's foremost authorities on reading and learning. Our instructions were to create a program that would make it easy for everyone to learn to read faster, with greater understanding. The program he developed also had to be economical so that it would be within almost everyone's reach.

After two years of development he finally achieved the goal . . . we call it THE SUPER READING SYSTEM.

WHAT, EXACTLY, IS SUPER READING?

First, the man who authored this home-study program is Russell Stauffer, Ph. D., Professor Emeritus at the University of Delaware. Dr. Stauffer has authored major textbooks and is considered to be one of the top researchers on how and why we learn. He is one of the few people admitted for membership in the Reading Hall of Fame. Second, the program is published by the same organization that publishes communication skills programs that are used by many of this country's leading businesses, government agencies and professional societies. In fact, many of the skills taught in THE SUPER READING SYSTEM are the very same as those taught in their reading programs that sell for three to four times the price of SUPER READING. Third, there are no gadgets or devices with SUPER READING. Authorities agree that these do not make you a better reader.

What you will get is a practical program that shows you *a new way to think.* Using a comprehensive instruction manual—over 280 pages—a teacher-on-cassette guiding and motivating you through every step of the program, you will discover a totally new way to read and learn . . . incredibly fast. *You will learn how to use your brain more efficiently.*

NEW READING

After completing the course, you will zip through all kinds of reading: text books, novels, correspondence, reports, technical journals, magazines, newspapers. You name it, you'll read it faster . . . get a heck-of-a-lot more out of it and remember what you've read. *The benefits last a lifetime.* Just like riding a bicycle, the more you use the SUPER READING skills the better and faster you become. The value of reading more efficiently means you will also have time for all the other things you have or want to do. No longer will you be bogged down by poor reading habits.

WHO CAN BENEFIT FROM THE SUPER READING SYSTEM?

Executives, students, professional people, men, women, managers, technicians . . . anyone who reads for career or pleasure will benefit from SUPER READING. There's even a special edition for children ages 10-16. Just think of how much time you now spend reading. Then consider what your life would be like if you could *read twice as efficiently.*

LISTEN AND LEARN AT YOUR OWN PACE

No matter how slow a reader you are now, THE SUPER READING SYSTEM is designed to get you started on this new way to read quickly and painlessly. In just a few days you'll begin to see an improvement. The teacher-on-cassette will be your guide through this remarkable program. You'll practice on materials you read everyday. *Your living room becomes your classroom.* You should complete the course in about 20-25 hours. But, you set the pace . . . you create your own schedule. Practice at home, the office . . . wherever or whenever it's convenient.

College Credits You may obtain college credits for successfully completing Super Reading. Full details and an application form are included with 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.

EXAMINE THE COMPLETE SUPER READING SYSTEM WITHOUT OBLIGATION

We've said a lot of things in this advertisement that may be difficult to believe. So we want you to see SUPER READING for yourself. Examine it for 15 days. Listen to the teacher-on-cassette. Then decide. If you feel that SUPER READING will not benefit you, simply return it for a full, no questions asked, refund. You have nothing to lose.

Super Reading Institute
113 Gaither Drive, Mt. Laurel, NJ 08054

☐ **YES! Please send me the Super Reading System at $39.95 plus $3 postage and insured delivery. If I'm not completely satisfied within 15 days, I may return it for a full refund.**
☐ **Send the Junior Edition of Super Reading.**
☐ **Check or money order enclosed.**
☐ **Charge my credit card under regular terms.**
☐ Visa ☐ Mastercard (Interbank #________)
☐ American Express
Card No. ________
Exp. Date ________
Name: ________
Address: ________
City: ________ State ____ Zip ____
Signature: ________
DZ-19

of the variety of plants themselves, evident extension of our long symbiosis with a number of them, in particular the grasses. The second theme is where the flowers grow: on some imaginary map there is a dot for every water lily and daisy. From such ideal maps (the real ones, of course, are mere approximations beset with problems) flows our knowledge of world vegetation in its intricate ecosystems. This brief and agreeable survey is meant instead to introduce the student and the amateur naturalist to the less flourishing outcome of taxonomic botany: the elucidation of how each group of plants came over time to flower where we find it. This is not ecology but geography, an effort to reconstruct a process from present and sometimes fossil evidence.

The topic was pre-Darwinian, Alexander von Humboldt and Alphonse de Candolle were its early paladins; with the *Origin,* particularly under Darwin's close friend Joseph Hooker, it became evolutionary. All but strangled under the luxuriant growth of environmental ecology, the geographic tradition was restored by the big treatise of E. V. Wulff, a scholar killed in the catastrophic siege of Leningrad, his last volume incomplete. Stott's small book is no such grand compendium; rather it is a lively introduction, not to the voluminous data themselves but to the methods and problems of the science.

The data are rich indeed; in the 1950's British field botanists surveyed and reported on the flowers of 3,500 10-kilometer squares; with an average of more than 400 species recorded per square, they generated a heap of sorted punch cards. Certain small European areas are particularly well studied. (A map of sweet violets in Staffordshire is one sample here.) At stake, however, are larger issues. The tall common reed of the world's wetlands is the most widely distributed of all flowering species; mapped in a generalized way, its worldwide distribution silhouettes the domains of desert and ice. Such results, taken with a modicum of paleontological data, for a long time suggested transoceanic land bridges; now we see in them the plain mark of continental drift.

Disjoined distributions are puzzles; witch hazels are found only in the eastern U.S., Japan and southern China. They share that pattern with no fewer than 74 other genera. Climate and sea-level changes are available to the theorist, and even rare dispersal over long distances by ocean currents and migratory birds is plausible enough. Darwin himself studied how long seeds could remain viable in seawater.

Two special cases are striking. It is surely unlikely that one species could evolve twice: a genome never steps twice into the same river. Yet two disjunct areas in Malaya support a single species of tree. One area has it as a conspicuous wayside tree of the mountains, the other as a tree of the swamp. To the botanist the trees hardly differ, except for "an indefinable difference in the velvety texture of the living leaves, which disappears on drying." It is maintained that in this instance the same form has evolved twice out of a large complex of related species, one for each habitat! Instances of closely similar plants with differing chromosome numbers, the result of polyploidy, are easier to accept than this subtle example, whose chromosomes show equal counts.

A more widespread phenomenon is exemplified by a pair of related plant species (kin to the source of a well-known laxative seed), one found in the California desert and one in a band running across North Africa to the deserts of western India. These two plants are called vicariants; they are recognized as distinct species, although they are easily hybridized. They must have evolved from a single ancestor that spread from the Old World to the New over a land bridge 20 million years ago, when the climate of the Bering Strait might have allowed it.

Fascinating too are the endemic groups, plants that are unique to, or at least strongly characteristic of, some particular region. There are living fossils, such as the dawn redwood, which was known as a Tertiary fossil until it was discovered to be a living tree in Szechwan. Such relict species can sometimes burst out, to give rise to an entire new local flora. That is seen in the Canary Islands for several forms; whether they are vicariant species, with common parents but evolved in place, or whether they first became distinct and then spread is unclear. Islands and tropical mountains (which are cool islands in the sky) are the usual scene of such tangled evolutionary dramas, more frequently in the Southern Hemisphere. The story feathers out into the larger topic of speciation in isolation, about where Darwin entered the theater of the Galápagos.

THE SURFACE OF MARS, by Michael H. Carr. Yale University Press ($45). THE GRAND TOUR: A TRAVELER'S GUIDE TO THE SOLAR SYSTEM, by Ron Miller and William K. Hartmann. Workman Publishing Company, Inc. ($19.95). Ample remote sensing (close to 60,000 sharp photographs from orbit) and some scarce ground truth to support it now exist for rust red Mars. Michael Carr is a senior geologist of the U.S. Geological Survey (seconded to the neighbor planet since the planning days of the Viking missions) and chief of the Orbiter imaging team. This big illustrated book presents our present overall understanding of Mars, a complex, changing mineral ball as study, reflection and a lively scientific literature have formed it up to about 1980. Now is a good time for such a perspective, with the last word not spoken but the early judgments well tested. One chapter by Harold P. Klein sums up the search for life on Mars: all organics are rapidly decomposed by that dry, weathered, iron-rich clay. This fits ground truth too, at least for two points on the windswept planet.

The treatment is comprehensive. Chapters survey the history of our growing knowledge, the atmosphere and climate of Mars, impact craters in all their variety, volcanoes, the complex relief and what it means, the surface chemistry, wind, ice and the two little moons. Three themes sound the differences between Mars and the earth. The chief of minerals—corrosive water—is there hidden away, whereas here a thick layer of the stuff rolls over most of the surface. There are still dry places on the watery earth; as a result Viking landers look out on a familiar enough desert scene.

The shifting continental plates of the earth are not found on smaller Mars; a single rift valley and the Tharsis bulge merely hint at the tectonic vigor of our own world. The big Martian volcanoes mark no plate collisions but arise from hot spots in the Tharsis bulge, a domed plateau of an area and height well exceeding Tibet's. Impact craters and the ancient enigmatic networks of natural channels sculpture the smaller surface relief, plainly etched away here and blanketed there on the scale of a few meters by the howling, dust-laden winds.

This second geology is already too rich for easy summary. The treatment of impact craters is particularly interesting; it is comparative in approach, contrasting the moon scars with those on Mars, drawing both on the tiny experimental craters of the laboratory simulators and a few full-scale craters made on the alluvial prairie of northern Alberta by test detonations of 500-ton hemispheres of TNT. The four concentric rims of the Prairie Flat crater are particularly striking.

This topic shades into the leitmotif of the volume: Where is the Martian water? Arguments for its presence take many forms, including theoretical outgassing models. Strange fluidized lobes cover the areas around the big craters on Mars, very different from the bouldery rays that stretch out across the surface of the moon. The smaller channels suggest running water; the larger ones imply watery eruption to carve away land either as bottom-lubricated glacier flow or as torrents of the fluid. Liquid water cannot persist on Mars today: it rapidly freezes or evaporates. Many arguments, uncertain but seductive, suggest that the now bone-dry Martian surface hides water. Perhaps it forms a layer 100 meters thick, compared with the 30-to-40-


TAKE YOUR SEAT.
THE PERFORMANCE IS ABOUT TO BEGIN.
DATSUN 280-ZX.
Slip into something comfortable: a multi-adjustable bucket seat you fine-tune to the perfect driving position.
A quick systems checkout assures you all is in readiness: computer sensors report on lights and vital fluids; a vocalized system alerts you if doors are open, fuel level low, parking brake on.
Remote-control mirrors? Check. Four-speaker stereo/cassette? Check. Power windows? Check.
Now, unleash the fuel-injected, computer-controlled powerplant. As energy surges effortlessly, a tingle travels down your spine. No wonder Z-cars have won national racing championships 11 of the past 12 years.
The 1982 Datsun 280-ZX and Turbo-ZX. Now both models are available in 2-seater and 2+2.
Come and see the four newest faces of "awesome."
©1982 NMC-USA
DATSUN
WE ARE DRIVEN
NISSAN
Product of NISSAN

times-deeper oceans of the earth. If it does, it must be stored mostly in a buried permafrost, perhaps sealing artesian waters kilometers deep, sometimes to be tapped by big impacts and to flood out onto the lowlands as both water and ice.

The big square pages of Carr's survey are all black, white and gray, illustrations and maps no less than the careful argument and the scrupulous comparison of evidence. Tentative to a degree, it is nonetheless strictly a cognitive exercise, a science newborn. *The Grand Tour,* the second book reviewed here, crosses that limit. It is literally a colorful, not to say a romantic, look at the remarkable variety of our solar system, a tour of some 40 worlds large and small, ordered by size from Jupiter down to a sun-grazing comet and its meteor-shower debris. Page after page is filled with new color paintings, each the well-controlled evocation of a spectacular scene, sometimes distant and orbital, sometimes as if done right on the spot. Besides the paintings there are plenty of photographs and straight maps to gloss the larger, more striking confections. What except admiration can one express for a daring view of a molten black sulfur lake filling a purple-walled caldera on Io, a blood red volcanic cone of sulfur in the distance, the blue glows of sunlit sulfur dioxide smoking out of the hot fluid? There are many more: the knife-sharp bright ring around sideways Uranus parting the night hemisphere from the sunlit one, as it might be seen from icy Titania; ring-shadowed Saturn a backdrop like a giant "golden waterfall" from closeup rocky Mimas; a total solar eclipse by Tethys seen from Dione!

The two imaginative painters of these wild scenes are knowing and serious. Ron Miller is an artist who for years has specialized in space art for the National Air and Space Museum, and William Hartmann is a planetary astronomer, with the energy and talent for his own paintings of views no human being has seen. Their joint work is nothing less than grand, as its title proclaims, surely tempting to every space-alert reader. The text is simple but up-to-date and quite reliable; such novelties as the double asteroids, the companion of Pluto and the two satellites of Saturn that dance in a common orbit are included, always in simple language. One could wish for a single page of references to allow a reader to go beyond informed assertions about the evidence.

We do find a couple of pages of the key statistics behind these leaps of the imagination. The presence of lightning in many of the scenes shown evokes the reflection that the thunderbolt, like liquid water, must be a rare phenomenon in the universe. Most matter is ionized plasma, which cannot support strong static electric fields; perhaps only in the inner insulating layers of planetary atmospheres, protected alike from stellar heat and ultraviolet and from galactic cosmic rays, can gas persist that is neutral enough to allow the buildup of slow-moving accumulations of electric charge. There is plenty of lightning here, but only one painting shows that still rarer spark, life, here represented by a patch of cabbagy stromatolites, ancient microbial dwellers on terrestrial shores.

TREE-RING DATING AND ARCHAEOLOGY, by M. G. L. Baillie. University of Chicago Press ($25). Hearts of oak are known in the folklore of the British Isles, but trusses and rafters and beams of oak are there in solid reality. The common oak has been a building material in Britain and Ireland back to time before record; no other wood comes close to oak in ubiquity. This small volume recounts the progress of a Belfast archaeologist toward the establishment of an ambitious chronology for the islands since Roman times. The dates are held in tree rings with an accuracy approaching a decade or so, giving hope of a bridge across the Channel to link someday into "one 7,500-year European chronology."

Baillie began his work in 1968; before

him there were a number of pioneering efforts. The first sustained one was made before World War I by the founder of the method, A. E. Douglass of Flagstaff, Ariz. His chronology relied on the yellow pine of the arid U.S. Southwest, which he extended to the older California sequoias. The southwestern trees showed such clear signs of frequent aridity that the method could be applied with comparative simplicity. The bristlecone pines of the Sierras were next; there the slow-growing timberline trees may disclose a pattern four millenniums long in a single individual. In spite of the difficulty of analyzing the bristlecone's very narrow rings, a fully confirming sequence over about five millenniums has been obtained for the species in a quite distinct mountain locality. It is hard to doubt that the entire digital bristlecone sequence is correct over some 8,200 years, accurate enough for its ring samples to calibrate the more variable analogue data of worldwide carbon-14 radioactivity.

A 1,000-year medieval chronology has been built on German oak rings by Bruno Huber, and one notable Russian investigation was able to date a sequence of 28 successive re-layings of the corduroy pathways of Velikaya Street in Novgorod the Great from the 10th century to the 15th. There tree rings sufficed for a floating chronology of considerable relative accuracy; old but historically well-documented standing wood churches anchored the tree rings in time. The wood pavements showed close agreement with the chronicles, which of course did not record the new sidewalks but did on at least 10 occasions describe the conflagrations that had made them necessary.

Among European timbers oak is well ringed. Between March and May, before the great crown of leaves is well formed, there grow right under the bark a set of large, thin-walled water-carrying vessels. These make the directly visible springwood ring, followed once the tree is great with leaf by a wider, denser, darker ring of mixed fibrous growth and small summerwood vessels. Springwood is derived from the reserves of the previous year; it is narrow and hence less varied. In the summer growth is written the season's nourishment, but the springwood mainly notes the occurrence of the invariable thaw of spring. One photograph here shows abnormal rings from the "years without a summer," 1816 and 1817, years of little summer growth in Ireland too.

Step by step as the problems unfold in practice they are described, a method that turns this quite technical study into a well-plotted narrative of logical progress. In the Temperate Zone world the simple signature of Douglass—patterns of ring alternation that leap to the eye and can be recorded by mere tally marks—becomes a width curve carefully measured under the binocular microscope and handled with computer statistical techniques of cross-correlation. Features still stand out from time to time; a few such unmistakably matchable pieces of wood are shown in the photographs. Huber was delighted with his "German Saw," a regular two-year cycle of wide and narrow rings between 1530 and 1540. Long floating sequences are thus fitted together.

Then comes the search for an anchor in time. An elegant new corer serves to sample living trees a foot deep. A dozen or more woodlands, at least 10 oaks in each, were fairly well correlated across both Britain and Ireland, the deviations showing no trend with distance. The conclusion, perhaps optimistic, is that the sites support one another; their differences are simple noise. A general signal is present over all these places. It turns out that the great oaks are much younger than their admirers think; big-crowned, freestanding oaks are often younger than less imposing specimens. Several ancient woodlands are said to go back 800 years and more, but no living British oaks are known to be older than about 400 years. Bicentennial celebrants in the U.S. had a similar disappointment. Every tree that could be shown to have been standing in 1776 was promised a suitable plaque. Hundreds of trees were cored; most of them were not even 100 years old, let alone 200. King William III tied his horse to a certain tree south of Belfast; that tradition is firm. But the tree itself blew down long ago; its large neighbor inherited the story. When it too died, the next large tree carried on by propinquity not in time but in space.

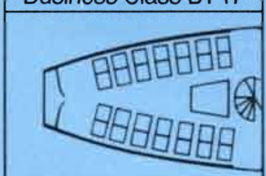
The hardest problem of a widespread chronology suited for reliable archaeology is sampling without gaps. The gaps arise from the shortage of big oaks over certain periods of the past. What the builders could not buy the dater cannot tabulate. There remain gaps without any cross dating. Nevertheless, there are wonderful old horizontal mills and lowly dugout canoes that can be dated to about 1,500 years ago. In their time or thereafter a cross-check exists with some dated sample, in particular one piece of wood excavated from Tudor Street in London.

A useful section treats of the new results in radiocarbon dating. Routine radiocarbon counts are now supplemented by high-precision methods: long counting times can offer statistical accuracy only if the stability of the counter itself is kept high enough. Such specially stabilized counting devices are now in service; they confirm the manifold wiggles in the natural radiocarbon-activity curve, and they may in the end yield radiocarbon dates several times more precise and much more reliable than those usually cited. The techniques are not cheap, nor are they always unambiguous, but they now cover nearly 7,000 years with high precision, with all the dates calibrated against tree rings to within two decades or so. The most substantial errors left in tree-ring dates, apart from the gaps and the variability with location still present, are those due to the moist sapwood that forms the outermost layers of the tree. These either are usually lacking, are damaged by insects or are removed from the log when it is put to use; a careful account of the clues to such errors suggests that they amount to at worst a couple of decades.

Seasonal and climatic rhythms are plain in oak and pine, our near contemporaries in the span of life. It is good to read a story so tightly argued it would be of value these days to see a summary of work on those still older "rings" left in the bottom of bays and lagoons by the seasonal runoff of silt and sand. The immediacy of those subtly modulated layers in clay and shale somehow makes them more sharply convincing of the depth of time than the valid but imperceptible radioactive-decay signal to the physicist's elegant electronics.

On international flights you'll find Alitalia's new Business Class right up front, where other airlines usually place first class. Why? Because our Business Class now offers all the comforts of first class at a cost only slightly higher than economy fare. For more room, Alitalia has reduced the number of seats. That's two seats per row, not three across. And our new reclinable seats invite total relaxation. For your enjoyment, there are free feature films and stereo music for good listening. Our menu offers a wide selection of authentic Italian cuisine, along with your choice of cocktails or excellent Italian wines. Our distinctively Italian ambience is enhanced by superior service and the kind of personal attention that is so rare, yet so Alitalia. Our Business Class is such a pleasure, you just might forget all about business.

Business Class B747



Alitalia

Established 1845 SCIENTIFIC AMERICAN July 1982 Volume 247 Number 1

The Space Telescope

The largest astronomical telescope designed to operate beyond the interfering effects of the earth's atmosphere is scheduled to be transported into orbit by the U.S. space shuttle in 1985

by John N. Bahcall and Lyman Spitzer, Jr.

The earth's atmosphere is an imperfect window on the universe. Electromagnetic waves in the optical part of the spectrum (that is, waves longer than X rays and shorter than radio waves) penetrate to the surface of the earth only in a few narrow spectral bands. The widest of the transmitted bands corresponds roughly to the colors of visible light; waves in the flanking ultraviolet and infrared regions of the optical spectrum are almost totally absorbed by the atmosphere. In addition, atmospheric turbulence blurs the images of celestial objects, even when they are viewed through the most powerful ground-based telescopes.

Accordingly the advantages of making astronomical observations from outside the atmosphere have long been recognized. In the past few decades considerable experience has been gained in the remote operation of telescopes that have been carried above most or all of the atmosphere by suborbital rockets, high-altitude balloons and artificial earth satellites. Significant findings have come from these efforts, altering theories of the structure and evolution of the universe.

The next stage in this program of exploration is the Space Telescope, which is scheduled to be put into orbit around the earth by the U.S. space shuttle in 1985. The Space Telescope will be a conventional reflecting telescope with unconventional capabilities. It will be the largest astronomical telescope ever orbited. It will also be the first long-term international scientific facility in space.

The Space Telescope, which is now under construction, is designed as a multipurpose astronomical observatory. It will have a 2.4-meter (94-inch) primary mirror capable of concentrating electromagnetic radiation in the entire optical part of the spectrum. It will be equipped initially with an assortment of scientific instruments for recording extraordinarily high-resolution astronomical images, for detecting extremely faint objects, for collecting various kinds of spectrographic data and for making very precise measurements of the position of radiant sources in the sky. The observations will be made from an altitude of some 500 kilometers (300 miles), well above the obscuring layers of the atmosphere.

The plans for the Space Telescope have been developed by a large number of scientists and engineers, working for almost a decade under the supervision of the National Aeronautics and Space Administration (NASA). The prime contractors charged with the actual construction are the Perkin-Elmer Corporation (responsible for the telescope itself) and the Lockheed Missiles and Space Company, Inc. (responsible for the supporting spacecraft system and for the integration of the components into a working whole). The total cost of the project is currently estimated at $750 million.

The projected lifetime of the Space Telescope is 15 years, although in principle there is no reason it could not be operated for many decades. An essential element in ensuring such a long lifetime (and in keeping costs within reasonable limits) is the availability of the space shuttle, which not only will deploy the telescope but also will service it on a regular basis. Astronauts from the shuttle will visit the Space Telescope whenever the instruments on board the observatory need maintenance, repair or replacement. At longer intervals (perhaps every five years) the entire Space Telescope will be returned to the earth by the shuttle for refurbishment of the mirror and other major components. The telescope will then be returned to orbit.

With suitable instrumentation the Space Telescope should be able to respond to electromagnetic waves ranging in length from about 115 nanometers (billionths of a meter) in the far-ultraviolet region of the spectrum to about a million nanometers (or one millimeter) in the far-infrared. Thus the spectral band accessible to the telescope could extend over a range of wavelengths that differ by a factor of 10,000. In contrast, ground-based telescopes have a clear view of colors that range in wavelength from about 300 to 1,000 nanometers, a span of less than a factor of 10.

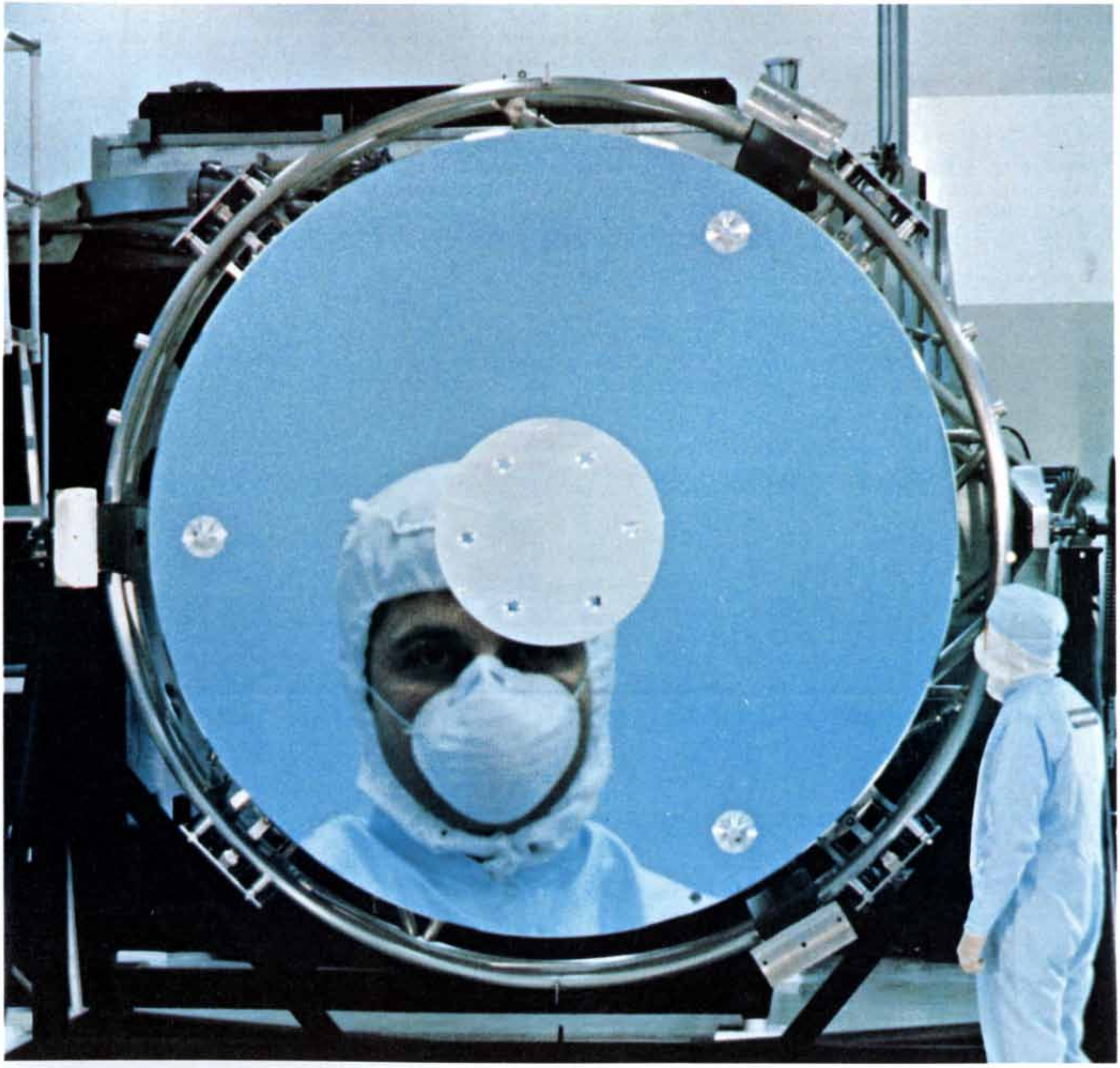
Because the Space Telescope will be immune to the blurring effects of atmospheric turbulence it will be able to obtain much sharper images of celestial objects than ground-based telescopes can, even at the same wavelengths that are observable from the ground. The maximum spatial resolution attainable with the Space Telescope will be on the order of a tenth of an arc-second; most astronomical images made with ground-based instruments have a resolution not much better than an arc-second. The tenfold improvement in resolution will make possible more detailed observations of extended objects. It is also expected to enable astronomers to see stars some seven times farther from the solar system than is now possible.

The observing program for the Space Telescope will be administered for NASA by the Association of Universities for Research in Astronomy (AURA), a consortium of 17 universities organized originally to operate several facilities for the National Science Foundation, including the national astronomical observatories at Kitt Peak in

Arizona and at Cerro Tololo in Chile. The center for the initial processing and analysis of data from the telescope will be the Space Telescope Science Institute, a new facility that is being established by AURA on the campus of Johns Hopkins University. The first director of the institute is Riccardo Giacconi, who led the scientific teams for the highly successful *Uhuru* and *Einstein* X-ray satellites. The operation of the Space Telescope will be the joint responsibility of the institute and of NASA's Goddard Space Flight Center in Greenbelt, Md. The Goddard center will have direct control of the satellite and will serve as the collection point for the data transmitted back to the earth.

The European Space Agency (ESA) is covering approximately 15 percent of the cost of the Space Telescope and will have an independent data-analysis center at the headquarters of the European Southern Observatory in Munich. The ESA is supplying the solar panels for powering the observatory, a high-resolution faint-object camera for the instrument section and a number of scientists and technicians for the staff of the Space Telescope Science Institute. In return European astronomers will be entitled to 15 percent of the observing time. Astronomers from other parts of the world will also work with the telescope, making it a truly international observatory.

The first astronomical observations



PRIMARY MIRROR for the Space Telescope was photographed at the Wilton, Conn., plant of the Perkin-Elmer Corporation just after its front surface had been coated with a reflective film of aluminum three millionths of an inch thick, followed by a protective layer of magnesium fluoride a millionth of an inch thick. The mirror, which is made of fused silica glass with an extremely low coefficient of thermal expansion, is 2.4 meters (94 inches) in diameter and weighs about 1,800 pounds. It consists of a lightweight cellular core approximately 10 inches thick sandwiched between two endplates, each about an inch thick. Some 200 pounds of material were removed from the front plate in the course of the 28 months of grinding and polishing required to give the surface its proper figure, which is that of a concave hyperboloid. The masked man seen enlarged in reflection is standing next to the photographer some 60 feet from the mirror. Another technician, also wearing a mask and a special suit to maintain the cleanliness of the mirror's surface, is at the right. A metal plate temporarily covers the hole in the center of the mirror through which light from the telescope's secondary mirror will pass.

from space were made in the late 1940's with captured German V-2 rockets. Some of these early liquid-fuel rockets were brought to the U.S. after World War II and were used to send various scientific instruments far above the atmosphere for several minutes of observation. Smaller solid-fuel rockets were later developed specifically for scientific research; they typically lifted a payload of about 100 pounds to a maximum altitude of 100 miles, giving an observation time of a few minutes above the obscuring layers of the atmosphere. The subsequent development of lightweight, solid-state electronic devices has made it possible to build increasingly complex and capable scientific instruments for such missions without prohibitive increases in the power needed to lift them.

The first application of the high-altitude technology was to the study of the sun. In 1946 a rocket-launched spectrometer developed by workers at the U.S. Naval Observatory obtained the first ultraviolet solar spectrogram, revealing absorption features not previously detected in the radiation from any celestial object. It was not until 1957 that ultraviolet radiation from a star was recorded. The spectrographic resolution of this early measurement was quite coarse, with a measuring bandwidth of several tens of nanometers. The early rockets could not be aimed accurately; they rotated freely in space and so could not give the long exposure needed for a precise measurement of the faint radiation from a distant star. In the 1960's techniques were developed for pointing rocket-borne instruments at a star, utilizing small gyroscopes to provide an inertial reference system. As a result stellar spectrograms were recorded with a measuring bandwidth of about a tenth of a nanometer. This achievement marked the beginning of active research on many aspects of stellar atmospheres and interstellar matter.

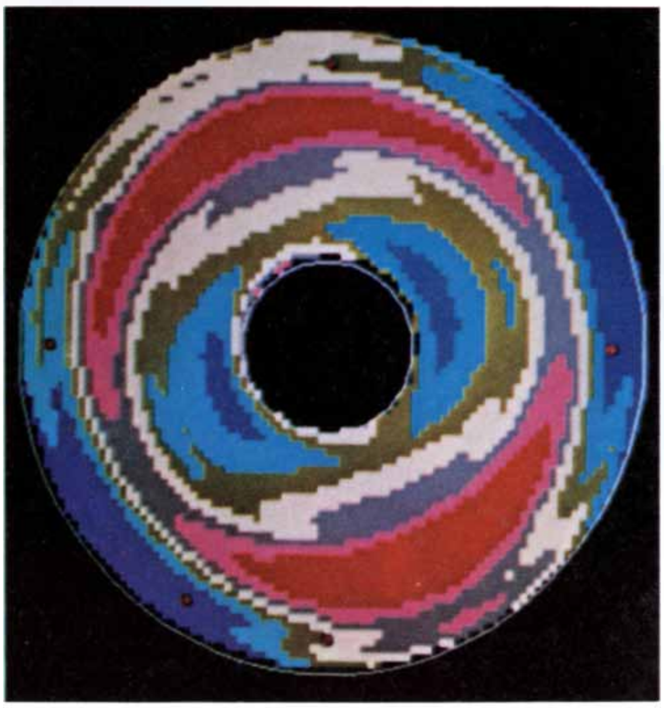
Meanwhile another group of astronomers employed balloons to lift optical telescopes to altitudes of about 20 miles, above the densest part of the atmosphere. In the late 1950's a 12-inch telescope of this type, named *Stratoscope I*, obtained extraordinarily sharp pictures of the sun. In the following decade its successor, the 36-inch *Stratoscope II*, made several photographs of planets and star systems with a resolution close to a tenth of an arc-second.

An artificial earth satellite, which can operate in orbit for years, offers a much better platform for mounting an optical telescope than either a suborbital rocket or a balloon. As aerospace technology has progressed satellites have become the primary vehicles for extraterrestrial astronomy.

With satellites as with the earlier rockets and balloons the first observations made were of the sun. The process of finding an object in the sky and pointing a telescope at it is much easier with the sun than with a more distant star. Beginning in the 1960's NASA built and operated a series of Orbiting Solar Observatories, equipped with various instruments for studying the solar atmosphere.

The first NASA satellites designed for stellar observations were named the Orbiting Astronomical Observatories. Two satellites of this type were operated successfully, one from 1968 to 1973 and the other from 1972 to 1981. Both of them were used mainly for analyzing ultraviolet radiation from stars. The first one had a fairly low spectrographic resolution: its measuring bandwidth was 1.2 nanometers. The second, named *Copernicus*, was far superior in this respect: its measuring bandwidth was .005 nanometer. The development of precise guidance systems for such satellites was a major technological achievement. The *Copernicus* telescope, which had a mirror 32 inches in diameter, could stay pointed toward a star for several minutes with a maximum deviation of about .02 arc-second.

The two Orbiting Astronomical Observatories yielded a wealth of data. For example, observations made with the *Copernicus* satellite showed that much of the hydrogen in interstellar clouds is in the form of molecules rather than indi-



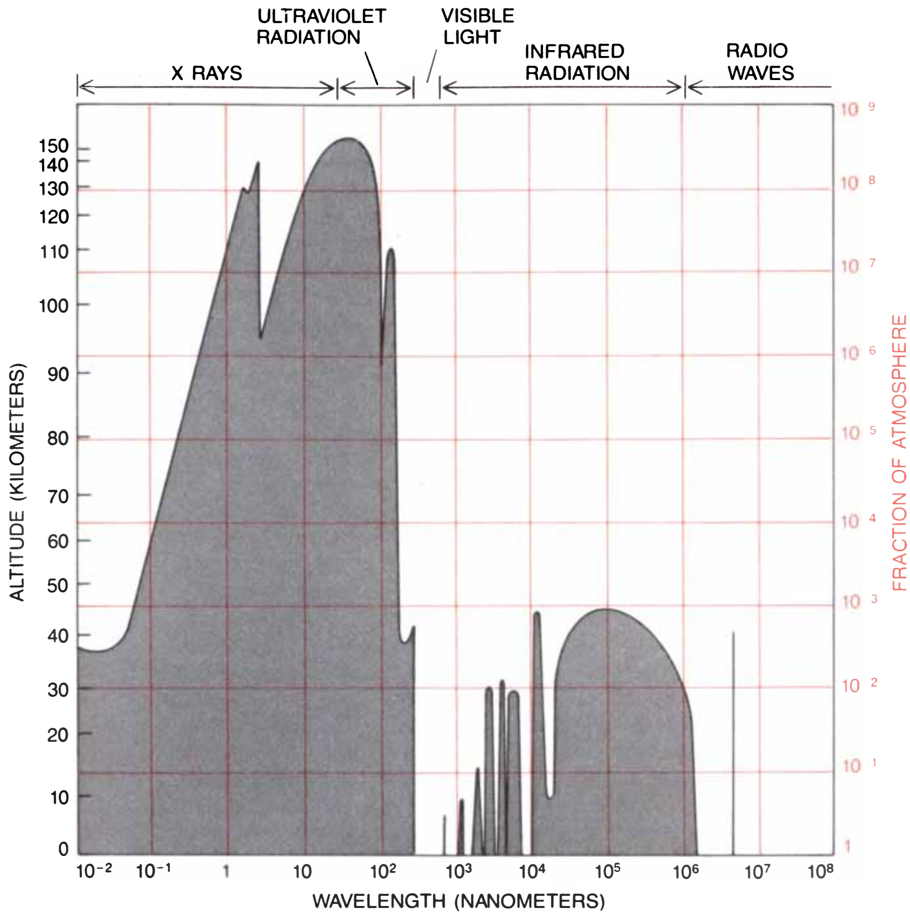
COLOR-CODED TOPOGRAPHIC MAPS of the surface of the primary mirror were made on the screen of a computer-graphics terminal as an aid in determining the corrective action needed for each of the 24 cycles in the final, eight-month polishing process. The maps were based on precise interferometric measurements of the shape of the surface. The two maps shown were made at the start and the finish of the computer-controlled polishing process. The white areas represent the average surface of the mirror; the dark blue and dark red areas correspond respectively to highs and lows. At the start deviations from the prescribed shape were as great as 100 millionths of an inch; at the finish the maximum deviation was less than a millionth of an inch over most of the surface. The finished primary is the finest large astronomical mirror ever made. According to Perkin-Elmer, "so nearly perfect is the surface that if the mirror were scaled up to the width of the continental United States, no hill or valley would depart from the mean surface by more than about 2½ inches."

vidual atoms. Moreover, many oxygen atoms in the regions between the clouds were found to be highly ionized, indicating that the gas between the clouds is very hot: on the order of a million degrees Kelvin. The satellite data also showed that the cosmic ratio of atoms of deuterium, or heavy hydrogen, to atoms of ordinary hydrogen is about one to 100,000. According to certain cosmological theories, this measurement supports the view that the universe will continue to expand forever.

The most recent optical telescope in space is the International Ultraviolet Explorer, a satellite developed jointly by NASA, the ESA and the British Science Research Council; it has been measuring the ultraviolet spectrum of comparatively faint objects since 1978. Although the performance of this instrument is limited by the size of its mirror (which is 18 inches in diameter), it has been particularly effective in obtaining ultraviolet spectrograms of galactic nuclei and in analyzing the interstellar gas in remote parts of our galaxy.

The concept of a much larger space telescope has evolved slowly over the past two decades. The first official notice of such a project appeared in 1962 in the report of a group of scientists organized for NASA by the National Academy of Sciences to study the future of space science. The group recommended the development of a large space telescope as a logical long-range goal of the U.S. space-science program. The recommendation was repeated by a similar study group in 1965. Soon afterward the National Academy established a committee chaired by one of us (Spitzer) to define the scientific objectives of a proposed space telescope with an aperture of approximately three meters. The report of this group was issued in 1969. In spite of the many advantages cited for such a large space telescope, most astronomers were simply too busy at the time to take an active part in promoting its development. Ground-based astronomy had entered an exciting "golden era" with the discovery of phenomena such as quasars, the cosmic microwave background radiation and pulsating neutron stars, and few people were prepared to devote the many years of effort needed to develop a facility as complex and costly as a large space telescope.

In 1972 another committee of the National Academy of Sciences, chaired by Jesse L. Greenstein of the California Institute of Technology, reviewed the needs and priorities of astronomy in the 1970's and again drew attention to the capabilities of a large space telescope. Although the nature and cost of such a device were then only partially defined, it was viewed as a realistic and desirable long-range goal.



ATMOSPHERIC ABSORPTION of electromagnetic radiation limits ground-based optical astronomy primarily to the narrow spectral band corresponding to visible light. Radiation in the flanking ultraviolet and infrared regions is almost totally blocked. The upper edge of the gray areas indicates the boundary where the intensity of the radiation at each wavelength is reduced to half its original value. A nanometer is a billionth of a meter, or 10 angstrom units.

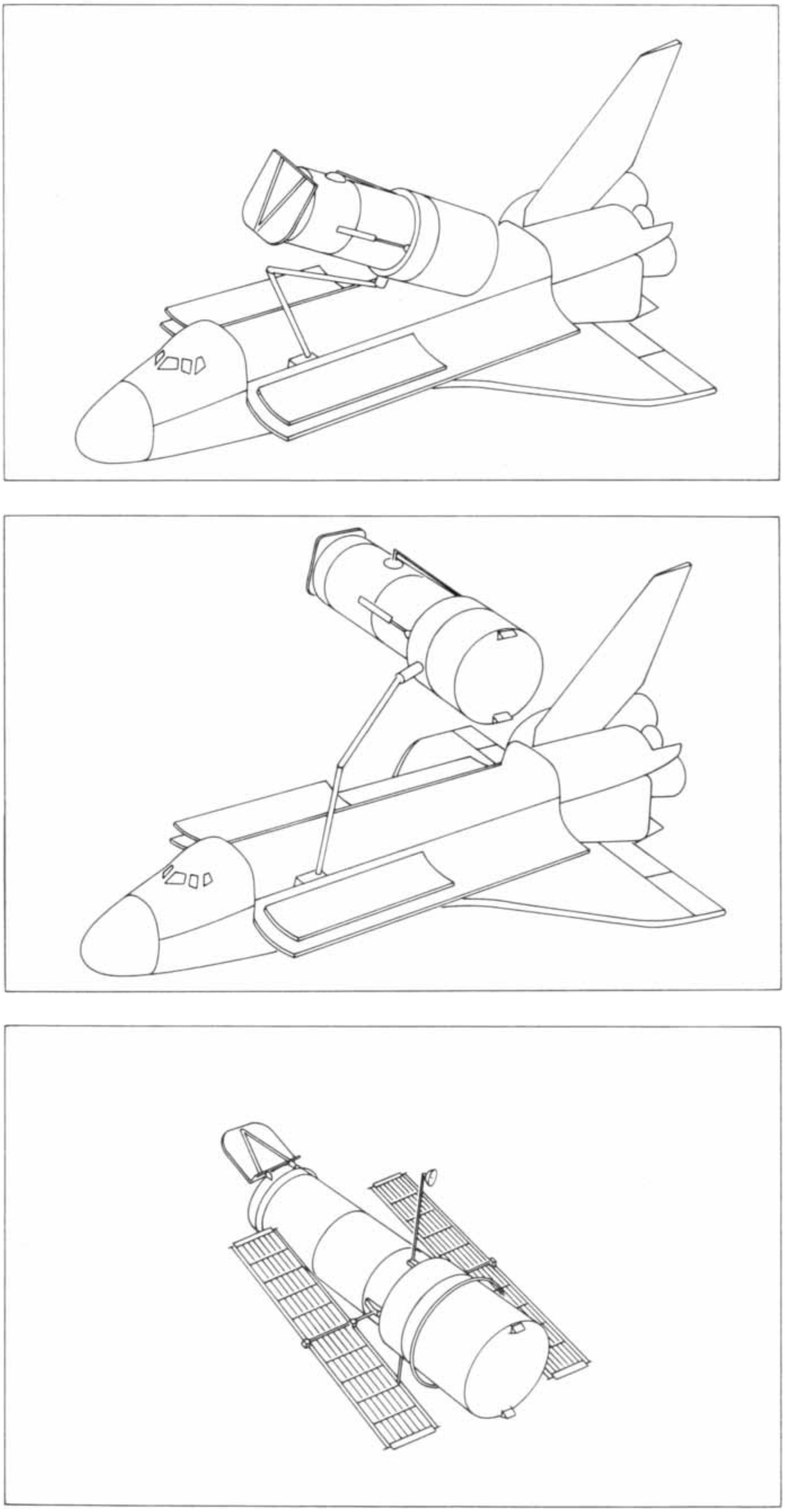
Meanwhile NASA had assembled a small group of astronomers under the direction of Nancy G. Roman to provide scientific guidance for the space-telescope feasibility studies then being done at Goddard and at the George C. Marshall Space Flight Center in Huntsville, Ala. Representatives of academic institutions, NASA research centers and industrial contractors assisted in the initial effort.

In 1973 NASA selected a group of scientists from several academic institutions to help establish the basic design of the telescope and its instruments. The group worked with NASA scientists and engineers to determine what objectives for the telescope were feasible and which of them should be given priority. The main scientific guidance was provided by a 12-member working group (on which both of us served) chaired by C. R. O'Dell of the University of Chicago. In order to head the scientific effort for the still unfunded Space Telescope project O'Dell left his positions as professor and chairman of the astronomy department at Chicago and as director of the Yerkes Observatory.

In 1977 NASA selected a new group of 60 scientists from 38 institutions to participate in the design and development of the proposed observatory. The scientific direction of this effort is again guided by a science working group headed by O'Dell; the current membership of the working group includes key NASA employees, the principal investigators responsible for the initial scientific instruments, several interdisciplinary scientists (including Bahcall) and specialists in data handling, spacecraft operations and telescope optics.

The Space Telescope program almost didn't happen. Between 1974 and 1978 the project was repeatedly in danger of being canceled or postponed indefinitely as a result of congressional and executive budgetary reviews. After an intensive lobbying effort, joined not only by hundreds of astronomers but also by many interested scientists in other fields, construction was finally authorized in 1977. The program survived its first appropriations test in Congress in 1978, and since then it has consistently met with a sympathetic and informed response on Capitol Hill.

By the time the Space Telescope was formally approved detailed NASA studies had led to a comprehensive de-



SPACE SHUTTLE will carry the Space Telescope to an altitude of approximately 500 kilometers (300 miles) and then release it into orbit with the aid of a mechanical arm. The solar-power panels, communications antennas and aperture door, which will be stowed while the satellite is being carried in the shuttle's cargo bay, will be deployed by the satellite after its release. The telescope will be visited by the shuttle for maintenance, repair and replacement of parts. Every five years or so the entire satellite will be returned to the earth for refurbishment.

sign, which is being followed for the most part in the actual construction of the observatory. The telescope itself consists of two hyperboloidal reflecting surfaces: the 94-inch concave primary mirror and a much smaller convex secondary mirror mounted about 16 feet in front of the primary. Light striking the primary mirror is reflected to the secondary, where it is directed through a hole in the center of the primary; the image comes to a focus several feet behind the primary. The telescope is described as a Ritchey-Chrétien type of Cassegrain optical system.

The scientific instruments that detect and measure the radiation concentrated in the focal plane are installed in an array of boxes mounted behind the primary mirror. Four of the boxes are aligned parallel to the optical axis of the telescope and four are arranged radially around the axis. Of the four radial boxes three house the telescope's fine-guidance system. The tube of the telescope extends more than 10 feet in front of the secondary in order to shield the optical system from stray light, most of which is direct light from the sun and scattered sunlight from the earth and the moon. A system of internal baffles provides additional shielding. Electronic equipment and other devices are housed in a toroidal section surrounding the telescope tube at its base. Two panels of solar cells for powering the equipment and two dish-shaped radio antennas for communicating with the earth extend from the midsection. The cylindrical body of the satellite is about 42 feet long and 14 feet in diameter.

The most remarkable feature of the Space Telescope will be the unprecedented quality of the images formed at its focal plane. The optical surfaces will be as nearly perfect as modern technology can make them: the average deviation of the two reflecting surfaces from their ideal contour will not exceed 10 nanometers. To avoid thermal distortions the mirrors are made of fused silica glass with an extremely low coefficient of thermal expansion. In addition they will be maintained thermostatically at a nearly constant temperature while they are in space. The position of the two mirrors with respect to each other and to the focal surface will be adjustable by remote control to yield the sharpest images possible. The fine-guidance system, which will take a fix on stellar images in the outer part of the telescope's field of view, is expected to be able to hold the optical axis steady to within .01 arc-second for as long as 10 hours. (Internal reaction wheels will serve to aim the telescope and hold it steady; commanding such a wheel to rotate faster in one direction will cause the entire telescope to turn in the opposite direction.)

Six major scientific instruments are

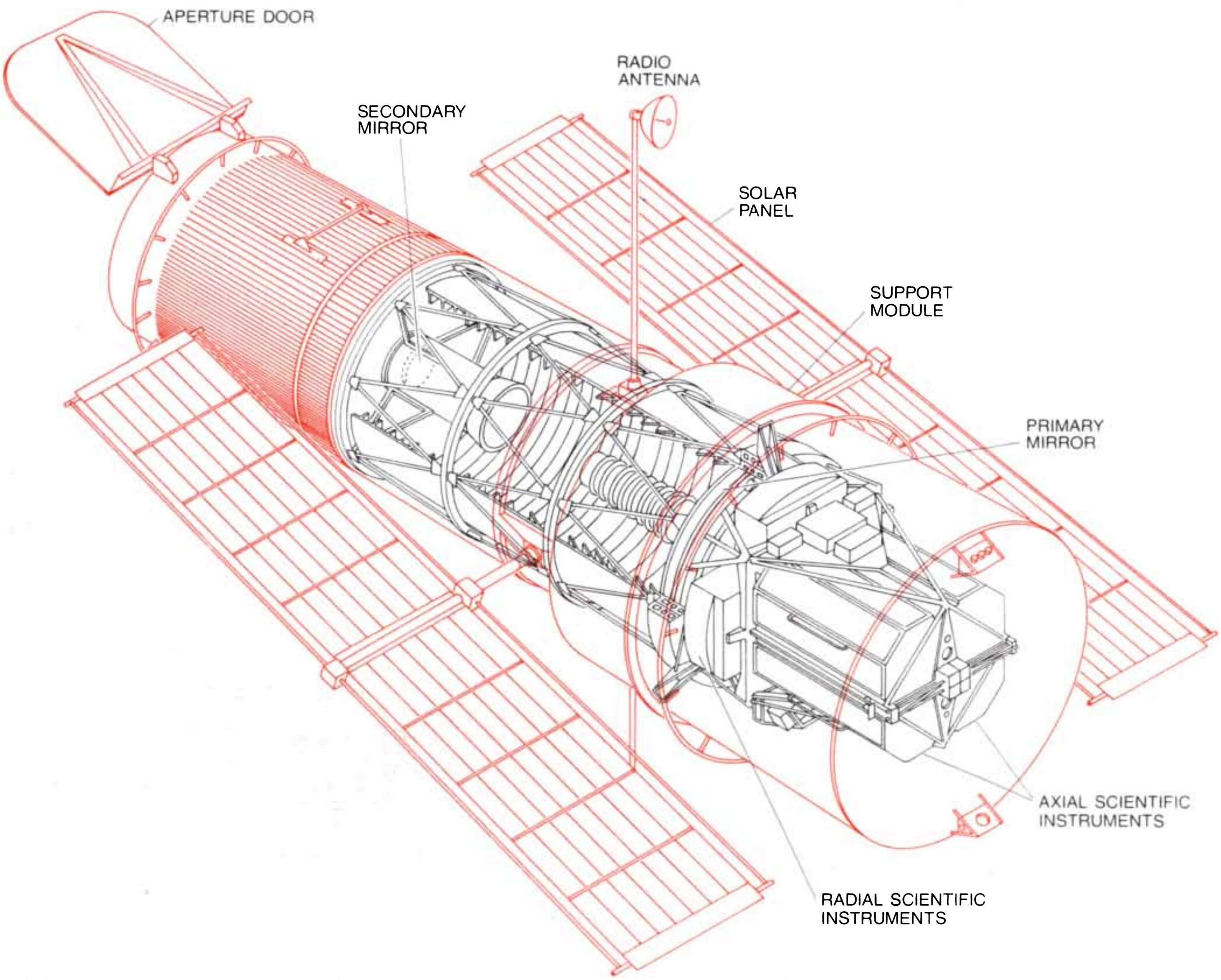
scheduled to be included in the Space Telescope's instrument section from the time it is launched through its first few years of operation. The first five are called the wide-field/planetary camera, the faint-object camera, the faint-object spectrograph, the high-resolution spectrograph and the high-speed photometer. In addition the fine-guidance system will give the telescope an astrometric capability, that is, an ability to measure the exact position of stars. Although the two mirrors will have a high reflection efficiency for radiation at all wavelengths in the optical region of the spectrum, no infrared-sensitive instrument will be included in the initial stage. Nevertheless, all aspects of the observatory are planned to be consistent with the possible future inclusion of an instrument sensitive to radiation with wavelengths as long as a millimeter.

The entrance apertures of the four axially mounted instruments are at the focal plane of the telescope. There the total field of view, which measures 28 arc-minutes in angular units, is almost half a meter in linear diameter; the resulting scale of the image at the focal plane is 3.58 arc-seconds per millimeter. With suitable pointing commands the image of any object in the field of view can be directed toward any one of the four axial instruments or toward the fifth, radially mounted one. Each instrument is designed so that it can be removed in orbit and a new instrument installed in its place by a space-suited astronaut operating from the space shuttle.

An on-board computer, external to the scientific instruments, will control the operation of the observatory and handle the flow of data. The computer will be reprogrammable, making it possible to modify the procedures as experience is gained with the instruments. Astronomers and spacecraft controllers will communicate with the Space Telescope by means of the NASA Tracking and Data Relay Satellite System. All data will be relayed back to the earth through this system also, for delivery ultimately to the Space Telescope Science Institute.

The principal investigators responsible for developing the initial set of instruments were chosen after intense competition. By the time the satellite is launched each of these investigators and his colleagues will have spent more than eight years building a general-purpose instrument for the potential use of all astronomers. In recognition of this effort each principal investigator and his team will be awarded more than a month of observing time.

The principal investigator for the wide-field/planetary camera is James A. Westphal of Cal Tech. This instrument, as its name suggests, can be operated in either of two modes: as a wide-field camera or as a higher-resolution camera suitable for, among other things, planetary observations. In each mode the detection system consists of four charge-coupled devices (CCD's): microelectronic silicon "chips" that con-



INTERNAL COMPONENTS are drawn in black and external components in color in this overall perspective view of the Space Telescope in its deployed configuration. The cylindrical body of the satellite is approximately 42 feet long and 14 feet in diameter. The scientific instruments are designed so that they can be replaced in orbit by a space-suited astronaut operating from the space shuttle.

vert a pattern of incident light into a sequence of electrical signals. Each chip is a square measuring almost half an inch on a side and is subdivided into an array of pixels, or individual picture elements, with 800 pixels on a side. A single chip therefore has a total of 640,000 pixels, and the four-part mosaic image formed by a set of four CCD's has more than 2.5 million pixels. Each pixel yields an electrical signal proportional to the number of photons, or quanta of electromagnetic radiation, reaching it during an exposure.

The wide-field/planetary camera is mounted on the side of the telescope that will generally be kept away from the sun. Incoming light passing along the optical axis of the telescope is directed outward at a right angle by means of a flat "pick-off" mirror held by a rigid arm at a 45-degree angle to the optical axis. The diagonal mirror diverts only the central part of the incoming beam; the rest of the light passes around the mirror to the other instruments.

In the wide-field mode the camera has a square field of view 2.67 arc-minutes on a side, the largest field of any of the instruments. Each pixel in this mode subtends an angle of .1 arc-second. In a sense the wide-field camera compromises the angular resolution of the telescope in order to provide a field of view large enough for the study of extended sources such as planetary nebulas, galaxies and clusters of galaxies. Even so, the field of view is much smaller than the field that can be recorded on a photographic plate by a ground-based telescope. In the Space Telescope the field is limited by the size of the microelectronic detectors available for remotely acquiring, storing and digitizing pictures. The CCD's for the wide-field/planetary camera, which are being supplied by Texas Instruments, Inc., have more pixels than any other CCD's used for astronomical purposes.

In the planetary mode the square field of view of the camera covers an area of the sky about a fifth as large as it does in the wide-field mode; the field in the planetary mode measures 68.7 arc-seconds on a side, and an individual pixel subtends an angle of .043 arc-second. The planetary camera takes advantage of almost the full resolution of the optical system while providing a field of view that is more than adequate for full-disk images of the planets. The high sensitivity of the CCD detection system makes possible the short exposure time required for certain planetary observations. The planetary mode will also be employed by many observers for high-resolution studies of extended galactic and extragalactic objects.

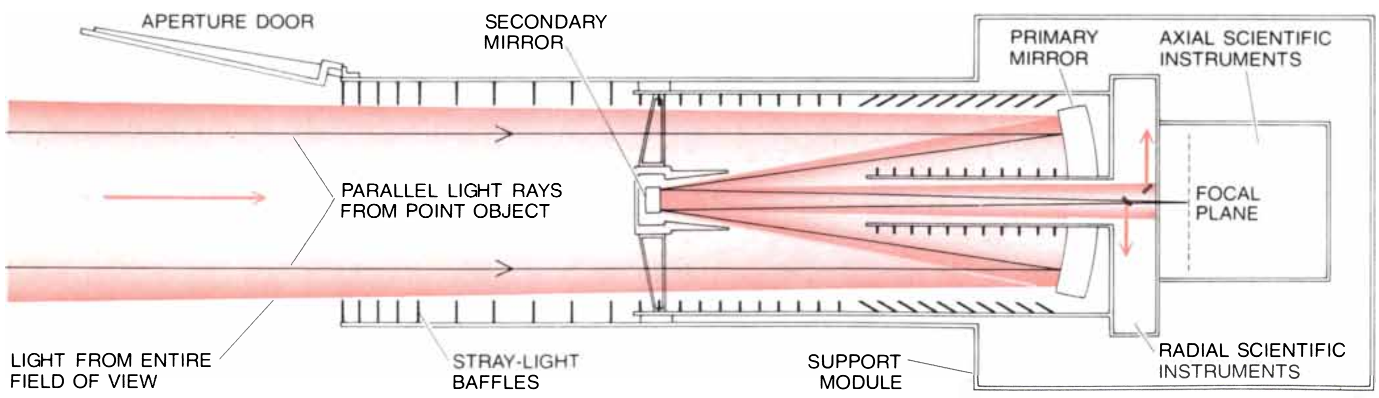
The wide-field/planetary camera is unique among the Space Telescope's instruments in several respects. It will gather by far the greatest number of bits of information: more than 30 million bits per picture. The spectral response of the detector will also be the widest available with any of the telescope's instruments: the camera will be sensitive to wavelengths ranging from 115 nanometers in the far-ultraviolet region to 1,100 nanometers in the near-infrared. The wide spectral coverage is made possible by coating the CCD's with an organic phosphor, called Coronene, that converts photons of ultraviolet radiation into photons of visible light, which the silicon sensors can detect. The excellent response at the red end of the visible band is attributable to the natural sensitivity of the CCD's.

The CCD's used in both the wide-field mode and the planetary mode have a low level of background electrical "noise" and hence are well suited for making pictures of faint sources. Part of the noise in such a device is thermal, and it will be reduced by cooling the detector elements thermoelectrically to about −95 degrees Celsius. The heat generated by the cooling system will be dissipated by a radiator that will form part of the outside surface of the satellite.

The incoming light to the instrument can be directed onto either the four CCD's of the wide-field camera or the four CCD's of the planetary camera by means of a pyramidal mirror that can be rotated by 45 degrees about its axis, thereby allowing two essentially independent optical systems to be housed in one instrument compartment. Any of 48 filters can be inserted into the optical path. Thus the wide-field/planetary camera is an extremely versatile instrument that will serve a broad range of astronomical purposes. We shall mention here just two of the many investigations that will be undertaken with this instrument.

The camera will be employed in both modes to make a series of images of certain nearby stars to see if they have planetary companions. The 10 or so stars selected for the study have been chosen because they all have a large proper motion (that is, motion across the sky). If any of the stars does have a planetary system, it may be possible, given the extraordinary resolution and accurate guidance of the Space Telescope, to detect periodic "wobbles" in the path of the star caused by the gravitational attraction of an unseen companion. The measurements are difficult ones, but the Space Telescope may finally resolve the long-standing question of whether there are planetary systems similar to the solar system among the nearby stars.

Quasars are the most distant and the most energetic objects known in the universe. Each of these compact sources emits on the order of 100 times as much energy as a bright galaxy made up of 10 billion stars. Several competing theories have been put forward to explain how a quasar produces such an enormous amount of energy in such a small space, but some crucial observational tests required to settle the matter are not feasible with ground-based instruments. Some of the theories are based on the idea that quasars are "sick" galaxies; in



OPTICAL PATH in the Space Telescope is said to be folded: light from the concave primary mirror is reflected from the convex secondary mirror and passes through a hole in the center of the primary before coming to a focus at the image plane in the instrument section several feet behind the primary. Technically the telescope is described as a Ritchey-Chrétien type of Cassegrain optical system.

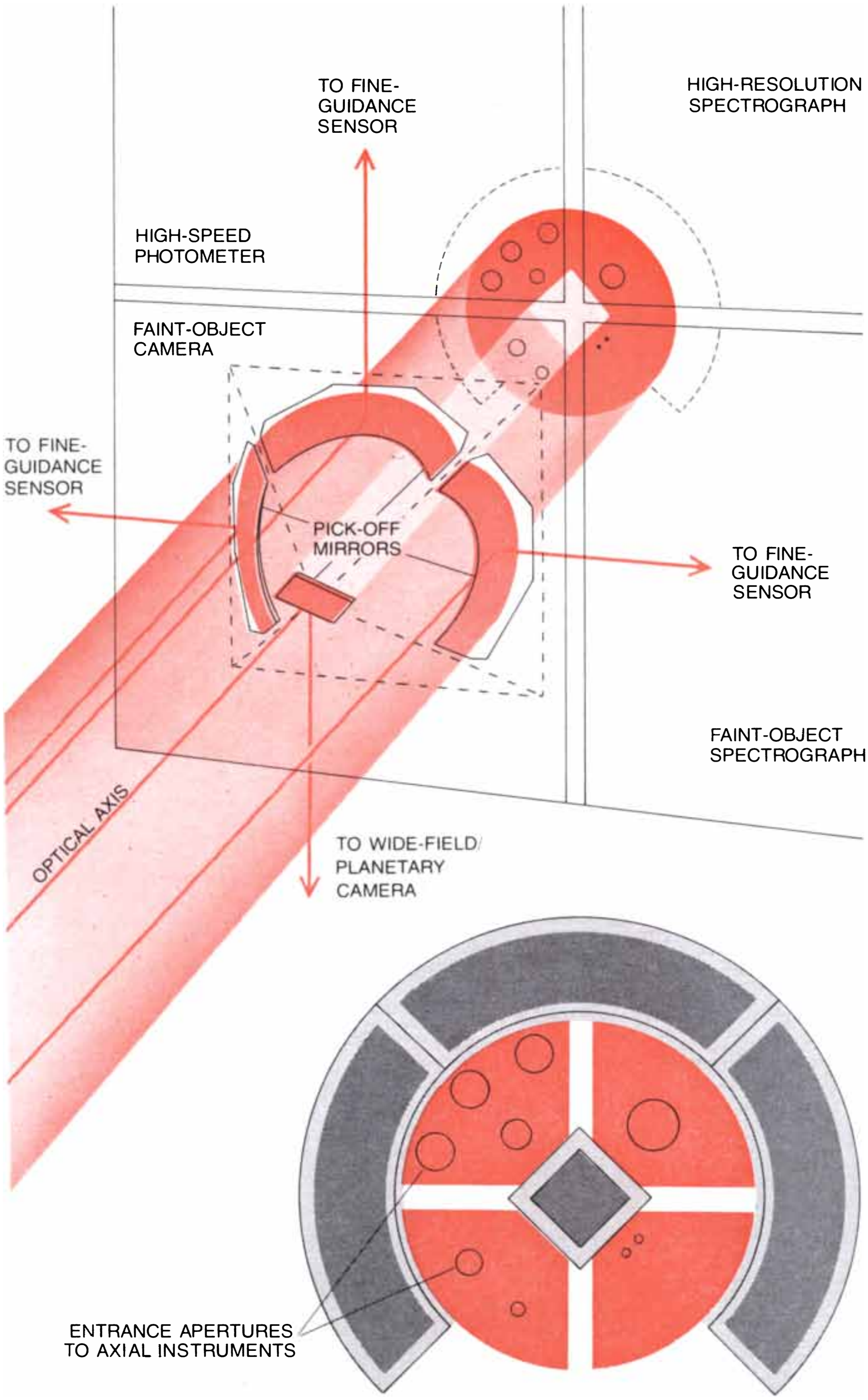
other words, the quasars are supposed to represent a transient, diseaselike stage in the evolution of an otherwise normal galaxy. To test these theories high-resolution images of quasars will be obtained with the wide-field/planetary camera to determine whether the bright objects that appear as point sources from the earth are surrounded by the fainter, more diffuse light of a galaxy. It should even be possible to tell whether the quasar stage is a disease of young galaxies or of old ones. This fundamental question is currently unanswerable because of the fuzziness of the images obtained with ground-based instruments.

The faint-object camera that will be supplied by the ESA is one of the four axially mounted instruments. The primary purpose of this second camera is to exploit the full optical power of the Space Telescope. It will detect the faintest objects visible with the telescope and will record images having the highest angular resolution attainable with the optical system. The project scientist for the faint-object camera is F. Macchetto of the ESA.

The faint-object camera is complementary in several ways to the wide-field/planetary camera. The faint-object camera will have a higher spatial resolution, whereas the wide-field/planetary camera will have a larger field of view. In the spectral region between 120 and 400 nanometers the faint-object camera will acquire an image more rapidly than the wide-field/planetary camera will. In the longer-wavelength, redward regions of the spectrum, however, the wide-field/planetary camera will be faster. In addition to forming images the faint-object camera will be able to determine the polarization of the detected radiation and to make spectroscopic measurements of both point objects and extended objects. The two cameras are not redundant, but they are designed to be sufficiently similar in function to ensure that an operable camera of some kind will be among the initial instruments even if a camera were to fail in orbit.

In the faint-object camera two similar but independent optical systems are provided to form an image of a point source. One system has a very small, square field of view, measuring 11 arc-seconds on a side; it has a pixel size of only .022 arc-second. The other system has a square field of view 22 arc-seconds on a side and a pixel size of .044 arc-second. In each system the detector consists of an image-intensifying device similar to the light-sensitive cathode-ray tube in a television camera. Unlike the CCD's in the wide-field/planetary camera, a detector of this kind counts individual photons.

The faint-object camera is designed so that each point-source image produced by the telescope is sampled by several



INCOMING LIGHT is routed in different directions by an array of small "pick-off" mirrors positioned near the center of the Space Telescope's scientific-instrument section behind the primary mirror. The diamond-shaped flat mirror mounted diagonally on the optical axis directs light outward to the radially mounted wide-field/planetary camera. The three arc-shaped flat mirrors arranged around the outside of the incoming beam send light to the three fine-guidance sensors, which are also radially mounted. The light that bypasses these four mirrors comes to a focus at an image plane at the entrance apertures near at the front of the four axially mounted instrument boxes. The projections of the pick-off mirrors on this focal plane are shown in dark gray in the plan view at the bottom. Because the incoming beam is interrupted by the pick-off mirrors well in advance of the focal plane the areas blocked by the mirrors are slightly enlarged; the additional vignetted zones are represented by the light gray bands outlining the projected mirror zones. At the focal plane the field of view is 28 arc-minutes in angular diameter. The wide-field/planetary camera views a square region about three arc-minutes on a side in the center of the field. The remainder of the field out to a radius of about nine arc-minutes is divided into quadrants, each of which is viewed by one of the four axially mounted instruments. The outermost part of the field, roughly between nine and 14 arc-minutes from the optical axis, is sampled by the fine-guidance system, which is designed not only to point the telescope but also to make precise measurements of the position of stars.

pixels. Hence it will be the instrument of choice when the highest possible resolution and the maximum contrast against the background sky are required. The camera will also be able to carry out spectroscopic and polarimetric studies of comparatively faint objects. In addition the camera will be able to view extremely narrow fields with an even smaller pixel size (approximately .007 arc-second).

The scientific tasks of the wide-field/planetary camera and the faint-object camera are expected to overlap. Depending on the specific resolution, field of view and spectral region required, an observer may choose to work with one camera or the other. We shall mention here only one type of observation for which the faint-object camera should be particularly suited.

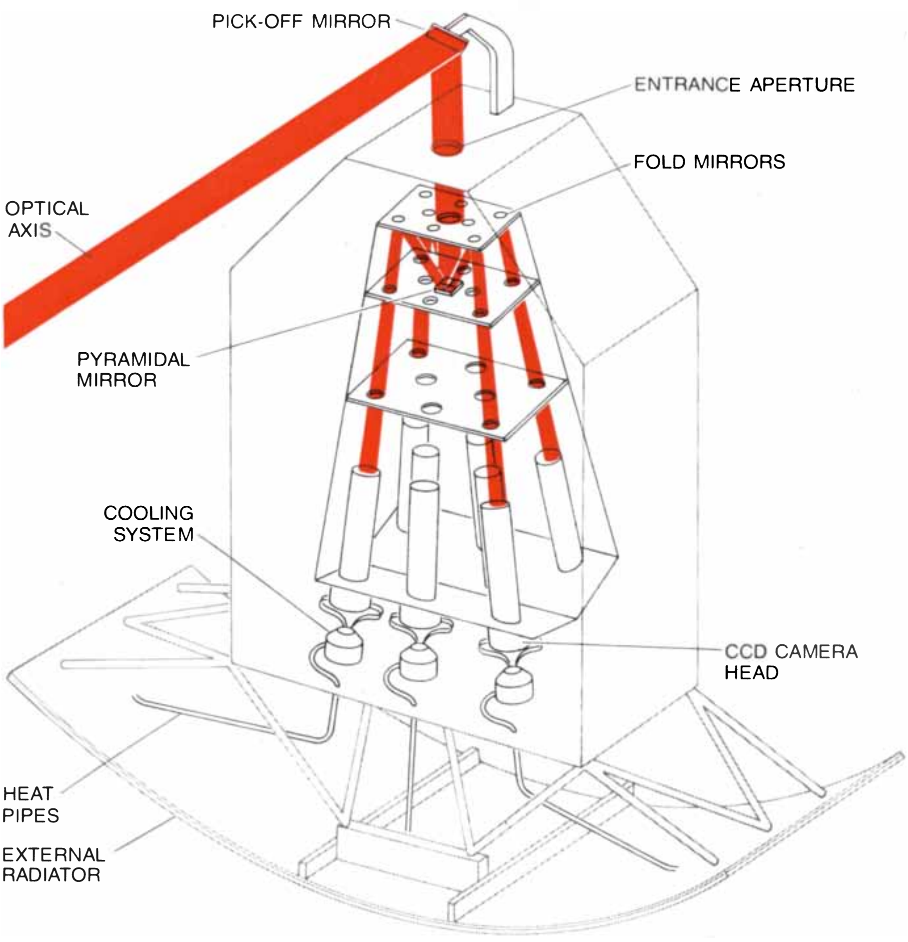
Globular clusters are spherical collections of millions of stars that can be seen from the ground on a clear night with a small telescope or even with binoculars. Because all the stars in a cluster are at approximately the same distance from the solar system one can test theoretical models of stellar evolution simply by counting the stars of various types in a cluster. The standard theory predicts that each globular cluster should include between about 10,000 and 100,000 of the stars called white dwarfs. These compact objects represent the last stage in the evolution of stars that have exhausted their nuclear fuels, cooled and collapsed. Because white dwarfs are very faint they cannot be seen at the great distances of the globular clusters with ground-based instruments. The Space Telescope's faint-object camera, however, should be able to detect many white dwarfs in globular clusters. By studying their properties it will be possible to learn much more about the evolution of stars.

The Space Telescope will have two spectrographs: optical devices that divide the incoming light from an astronomical source into separate beams according to wavelength. In spectroscopy resolution is usually defined as the ratio of the wavelength of the incoming light to the smallest separation that can be measured between two wavelengths. One of the two spectrographs on board the observatory, the faint-object spectrograph, will be able to observe faint stellar objects with a spectrographic resolution of 1,000 (equivalent to a measuring bandwidth of 1/1,000th of the wavelength). The principal investigator for this instrument is Richard J. Harms of the University of California at San Diego.

The faint-object spectrograph will be equipped with two systems of detectors. Both detectors are devices called Digicons; one is sensitive to red light and the other to blue light and ultraviolet radiation. A Digicon sensor operates on the basis of the photoelectric effect. The incoming light is spread out according to wavelength by a diffraction grating and strikes the surface of a thin photocathode layer deposited on a transparent plate. Light of a particular wavelength reaches a particular position along the photocathode, producing a spray of free electrons known as photoelectrons. A magnetic field focuses the photoelectrons at a point whose position depends on where they emerge from the photocathode and hence on the wavelength of the incident light. The photoelectrons are collected by a linear array of 512 diodes, each of which records the intensity of the incident light at a particular wavelength.

The faint-object spectrograph will be sensitive to radiation ranging in wavelength from about 115 to 800 nanometers. In addition the instrument will have two special features: it will be able to measure the polarization of the incoming light and to detect extremely fast variations (perhaps as brief as a few milliseconds) in the spectrum of radiation emitted by bright sources. Because the investigation of many astronomical problems depends on the spectral analysis of the radiation from extremely faint objects, this instrument is expected to be one of the busiest on the Space Telescope. By measuring the spectra of very distant quasars, for example, it should be possible to study the properties of the universe more than 10 billion years ago, perhaps 85 percent of the way back to the beginning of time (if, as the standard big-bang model of cosmology assumes, time actually had a beginning). Spectrograms of the most distant quasars are expected to indicate the chemical constitution of matter at that early stage in the evolution of the universe.

The investigation of some astronomical questions requires a higher spectrographic resolution than can be attained with the faint-object spectrograph, because the width of many emission and absorption features is narrower than the



WIDE-FIELD/PLANETARY CAMERA is one of the instruments scheduled to be included in the Space Telescope during its first few years of operation. The camera is designed to operate in either of two modes. In each mode the detection system consists of a rectangular array of four light-sensitive silicon "chips" called charge-coupled devices (CCD's). The incoming light reflected into the radially mounted instrument compartment by the diagonal pick-off mirror can be directed onto either the four CCD's of the wide-field camera or the four CCD's of the higher-resolution planetary camera by means of a pyramidal mirror that can be rotated by 45 degrees about its axis. Any of 48 filters can be inserted into the optical path. The external radiator serves to dissipate the heat generated by the cooling system associated with the detectors.

measuring bandwidth of the instrument. The high-resolution spectrograph will meet this need. Under normal operating conditions it will have a spectrographic resolution of 20,000. Narrow spectral features that might not even be detected with the lower-resolution faint-object spectrograph will be accurately measured, yielding information about the physical conditions under which the radiation was emitted. The high-resolution spectrograph will also have an ultrahigh-resolution mode of operation in which the spectrographic resolution will be improved by an additional factor of five to about 100,000. The principal investigator for the high-resolution spectrograph is John C. Brandt of Goddard.

Of course, there is a price to be paid for the higher resolution of this second spectrograph. Dividing the spectrum into a much larger number of bands in order to measure the flux of photons separately in each band has the effect of decreasing the number of photons detected per band. Thus higher resolution results in lower sensitivity, and the larger quantity of information provided by the high-resolution spectrograph can be obtained only for stars that are some 60 times brighter than those that can be studied with the faint-object spectrograph. This difference amounts to about 4.5 stellar magnitudes. For the ultrahigh-resolution mode the difference in brightness is a factor of more than 300, or the equivalent of about six stellar magnitudes.

The high-resolution spectrograph has six interchangeable diffraction gratings, each of which disperses light of different wavelengths in different directions. A camera mirror or grating then forms an image of the spectrum on the photoelectron-emitting surface of a Digicon sensor. By rotating the carousel on which the gratings are mounted, any one of them can be brought into the optical path of the instrument, making it possible to obtain a spectrographic reading at any wavelength between 110 and 320 nanometers.

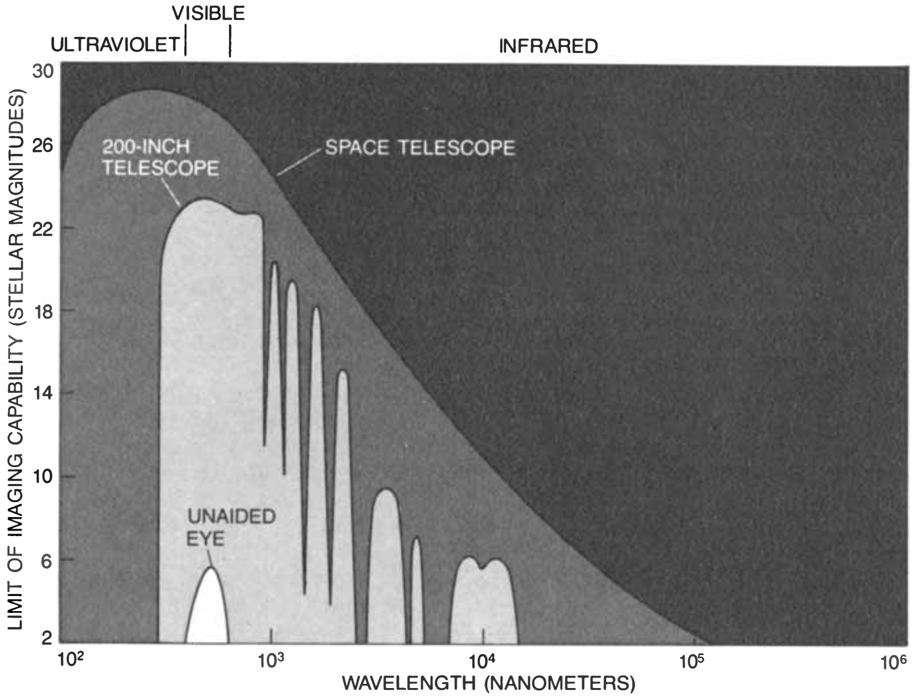
This spectrograph with its normal resolution should be able to observe stars as faint as the 13th magnitude, or about six stellar magnitudes fainter than those observed by the *Copernicus* telescope. The gain in sensitivity over the spectrograph on the International Ultraviolet Explorer is not as great—about four magnitudes—but the spectrographic resolution and the photometric accuracy will be significantly better for the instrument on the Space Telescope.

The power of this instrument should open up a number of interesting new lines of inquiry. For example, the high-resolution spectrograph will make possible the study of interstellar gas at places in our galaxy and other galaxies where it cannot now be observed. Preliminary measurements by the International Ultraviolet Explorer have shown that the gas in the galactic "halo" between the earth and the nearest neighboring galaxy (one of the two Magellanic clouds) includes carbon atoms that have been stripped of three electrons, indicating that the temperature in this region is about 100,000 degrees Kelvin. With the high-resolution spectrograph much more accurate data will be obtainable, perhaps revealing the relation between this gas and the even hotter material detected by *Copernicus*. Measurements of the way in which the properties of our galaxy vary from place to place will provide much-needed clues to the evolution of the system as a whole.

The high-resolution spectrograph will also be applied to the study of interstellar clouds. Ground-based observations of such clouds are able to detect only a few dark lines in the spectrum, created when the gas of the cloud absorbs radiation from background stars. In many instances each absorption line is split into multiple subfeatures, which can be attributed to separate clouds along the same line of sight. The clouds are moving with somewhat different speeds toward the solar system or away from it, altering the characteristic wavelengths at which they absorb radiation. The splitting of the absorption lines makes it possible to study each cloud separately, provided the spectrographic resolution is high enough. With the high-resolution spectrograph it will be possible to analyze a wide range of ultraviolet absorption features from various atoms and molecules and to determine the physical conditions in each cloud. Our understanding of how such interstellar clouds come together and contract to form stars may depend critically on the results of such studies.

The high-speed photometer, which is being developed by Robert C. Bless and his colleagues at the University of Wisconsin at Madison, is designed to make highly accurate measurements, with an extraordinary temporal resolution, of the intensity of the light from astronomical sources over a wide range of wavelengths. The photometer will be capable of distinguishing events separated in time by only 10 microseconds. Observations of sources that vary over time scales this short are difficult or impossible with ground-based instruments because of fluctuations in the atmosphere.

The high-speed photometer is the simplest of the instruments in the initial group on board the Space Telescope. It has no moving parts and relies entirely on the fine pointing of the spacecraft to direct the light from an astronomical target onto one of its 100 or so combinations of spectral filters and entrance apertures. The photometer has four independent, magnetically focused detectors, called image dissectors; they



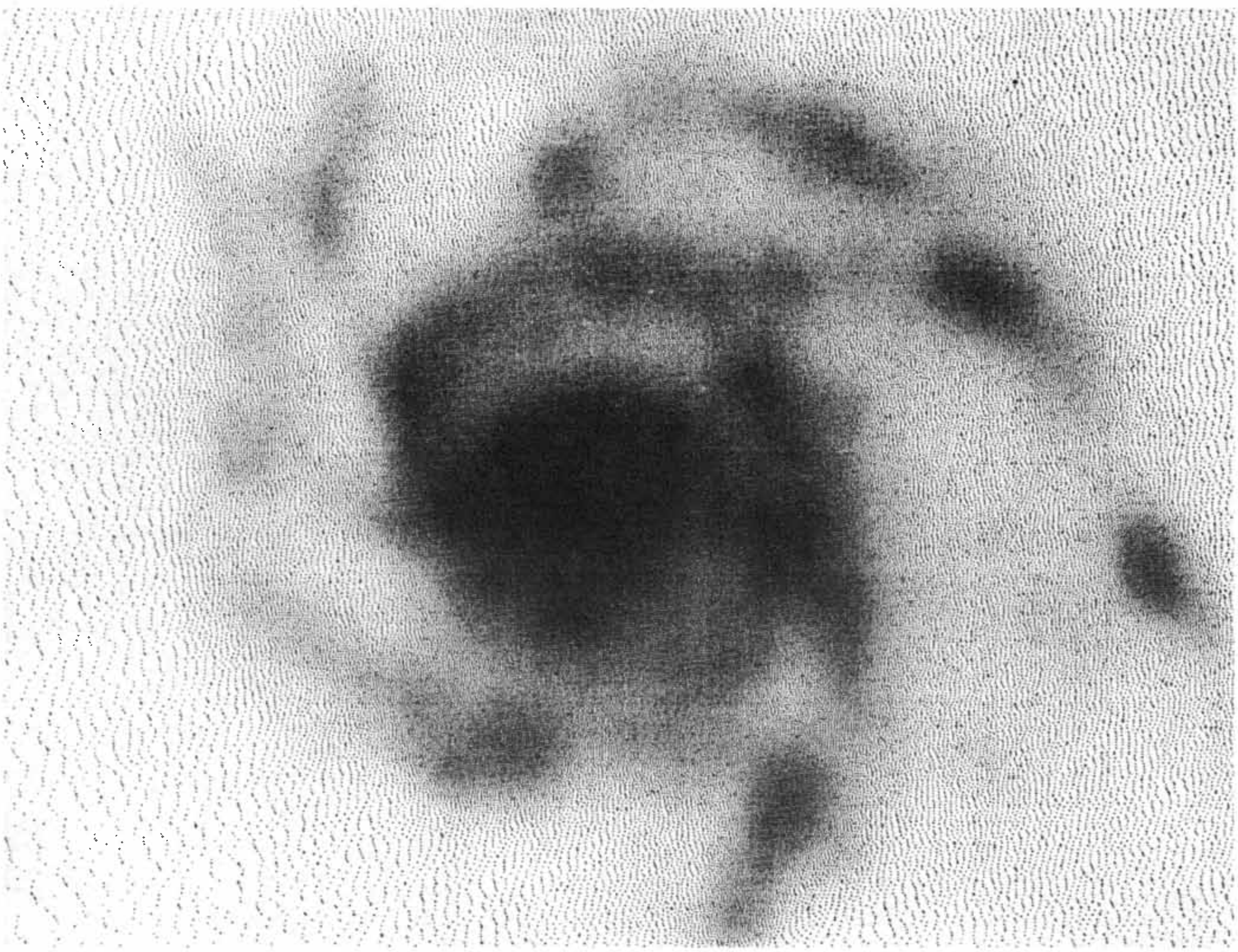
RANGE OF WAVELENGTHS potentially accessible to the Space Telescope extends from the far-ultraviolet part of the spectrum (*left*) to the far-infrared (*right*). For comparison the spectral bands that can be observed with the unaided human eye and with a large ground-based telescope (in this case the 200-inch Hale telescope on Palomar Mountain) under normal observing conditions are also indicated. The vertical scale gives the relative brightness (in terms of stellar magnitude) of the faintest celestial object that can be imaged; an increase of one unit in stellar magnitude corresponds to a decrease in apparent brightness by a factor of about 2.5.

resemble photomultiplier tubes in operation, except that they can be made to respond only to photoelectrons coming from the small region of the photocathode on which the light is focused. Each image dissector is mounted behind a plate that holds an assortment of filters and entrance apertures.

The overall spectral response of the image dissectors extends from about 115 to 650 nanometers. The instrument is also equipped with a red-sensitive photomultiplier tube and a system for measuring the polarization of ultraviolet radiation with the aid of one of the image dissectors.

The high-speed photometer will be capable in principle of detecting the smallest objects observable with any of the instruments on the Space Telescope. The ability to distinguish events that are separated in time by only 10 microseconds implies (according to the special theory of relativity) that variations in the light output of a star as small as three kilometers across could be detected. This is an extraordinarily small linear dimension for a star; indeed, it is very close to the diameter the sun would have if it were compressed to such a high density that it formed a black hole. Accordingly, one program scheduled for the high-speed photometer is to search for extremely fast variations in astronomical systems that are suspected of harboring a black hole, in the hope of finding further evidence of these elusive entities. The high-speed photometer will also be used for less exotic observations, including an attempt to identify optically faint objects observed mainly at radio or X-ray wavelengths.

Under the best observing conditions ground-based measurements of the position of any star are limited by the size of the star's blurry "seeing disk," which is generally at least one arc-second in diameter. In determining the angular distance between two stars an uncertainty of about .1 arc-second, or a tenth of the diameter of the stellar image, is typical for a single observation. By averaging many exposures the uncertainty can be reduced to about .01 arc-second. Random errors result in corresponding uncertainties in the determination of a star's parallax. (Parallax is the average angular change in the apparent position of a star resulting from the revolution of the earth about the sun.) The determination of distance beyond the solar system is based largely on measurements of the parallax of comparatively nearby stars. Since the measurement of a stellar image with the Space Telescope will be accurate to within about .002 arc-second, the determination of stellar position, and hence of stellar parallax, should be about five times better than it is with ground-based telescopes. The fivefold improvement in the accuracy of stellar-parallax measurements is of fundamental importance to all of stellar astronomy. For example, knowing the precise distance of certain comparatively young star clusters in our galaxy will enable astronomers to determine the absolute brightness of the stars in the clusters. This knowledge in turn will make it possible to extend the calibrated distance scale, which is based on the com-



TENFOLD IMPROVEMENT in spatial resolution expected with the Space Telescope will enable astronomers to make more detailed observations of extended objects. In this simulation the picture at the left represents the image of a distant spiral galaxy obtained with the 200-inch Hale telescope and the picture at the right represents the corresponding image obtained with the Space Telescope. Actual-

parison of apparent brightness and absolute brightness, to stars that are much farther away.

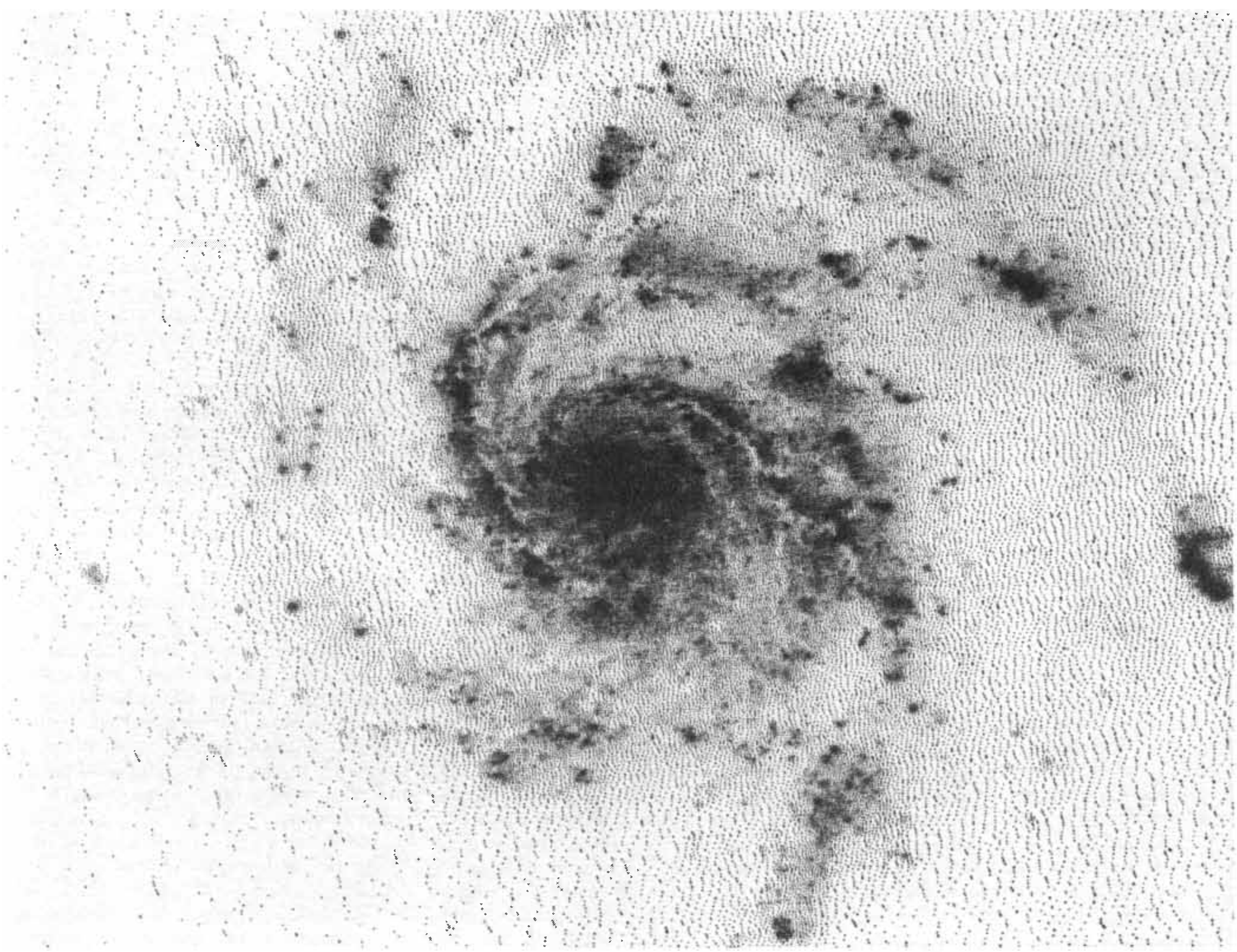
The Space Telescope has not been equipped with a separate instrument for astrometry because the fine-guidance system will be accurate enough to make the necessary measurements of the angular distance between stars. The leader of the team for astrometry is William H. Jefferys of the University of Texas at Austin.

Observing time on the Space Telescope will be allocated to astronomers from all parts of the world by the Space Telescope Science Institute, which will be responsible for facilitating the most effective scientific use of the powerful new observatory. To provide visiting astronomers with the most efficient operating systems, to assist and advise observers on the optimum use of the various instruments and to help create a stimulating atmosphere for research with the Space Telescope outstanding astronomers from the U.S. and abroad are being recruited to serve on the institute's staff. It is expected that half of their time will be devoted to the diverse tasks of the institute, with the other half available for their own research programs. The new institute will also make recommendations to NASA on broad policy matters pertaining to the Space Telescope. The involvement of outside astronomers in determining the policies of the institute is being ensured through a number of external committees.

The institute will solicit outside proposals for specific observing programs for the Space Telescope. With the aid of peer-review groups the institute will evaluate the proposals and select the most promising programs for inclusion in the telescope's schedule. In many cases the programs selected will be combined with those submitted by the original scientific-instrument teams, by other members of the Space Telescope working group and by the European groups. The final scheduling and the preparation of a complete list of commands for the operating computer will be done by NASA, which will retain responsibility for the day-to-day operation of the observatory.

Astronomers on the staff of the institute will advise outside astronomers on the formulation of observing plans. Other staff astronomers will be responsible for maintaining the calibration of the instruments and for the initial processing of data. Computer specialists will help to develop suitable programs for use by the astronomers in analyzing the data. Finally, the Space Telescope Science Institute will assist astronomers in communicating the results of their studies to other scientists, to NASA, to Congress and to the public.

The Space Telescope will help to solve many outstanding astronomical puzzles. The greatest excitement, however, will come when the pictures returned from the satellite reveal things no one in this generation of astronomers has dreamed of, phenomena that only the next generation will be privileged to understand.



ly the picture at the right is a digitized version of a photograph of a nearby galaxy made with the 200-inch telescope and the picture at the left is a blurred version of the same image made by defocusing the original by an amount proportional to the difference in the effective resolution obtainable with the two instruments. The simulation was prepared by John L. Tonry of the Institute for Advanced Study.

Organic Superconductors

The ability to carry an electric current without resistance had been known only in metals. Now superconductivity is seen in organic crystals made up of flat molecules in zigzag stacks

by Klaus Bechgaard and Denis Jérome

The diversity of synthetic organic compounds in everyday life is a testament to the architectural resources of the carbon atom. By assembling chains or rings of carbon atoms chemists have fabricated more than two million organic substances, which serve as fibers, coatings, adhesives, pharmaceuticals, synthetic rubbers, structural plastics and so on. The collective experience with carbon-based systems offers the molecular designer a vast set of raw materials and a battery of proved experimental techniques with which to create new materials. It seems an organic molecule can be constructed to have almost any physical property.

Recently we and our colleagues have created a series of organic materials with a rare and extraordinary property: superconductivity, or the ability to conduct an electric current without resistance. Superconductivity had been observed previously only in metals and metallic alloys. Even in those materials, most of which are good ordinary conductors, it is a remarkable phenomenon. The resistance of a superconductor is not merely small but exactly zero; a current set up in a superconductor is not dissipated as heat (as it is in an ordinary conductor) but instead continues to flow forever. Establishing superconductivity in an organic solid seems still more remarkable because the great majority of synthetic organic materials are electrical insulators.

The conditions under which we first observed the superconducting state in an organic compound were extreme. The temperature was .9 degree Kelvin (or about −272 degrees Celsius) and the pressure was 12,000 atmospheres (or 12,000 times the atmospheric pressure at the surface of the earth). In less than a year, however, five other synthetic organic compounds were found to be superconducting; one of them is superconducting at normal atmospheric pressure, although a low temperature is still required for all the materials. There is good reason to believe that superconductivity at higher temperatures, perhaps as high as 20 degrees K., can be achieved. Indeed, now that superconductivity has been added to the list of properties available in organic substances, the full resources of synthetic chemistry can be brought to bear on relaxing the conditions under which superconductivity can be maintained.

Ever since the discovery of superconductivity in 1911 by Heike Kamerlingh Onnes of the University of Leiden, it has been considered a low-temperature phenomenon. Kamerlingh Onnes' discovery was the outcome of a remarkable technological achievement: the liquefaction of helium. At atmospheric pressure helium condenses at 4.2 degrees K. Kamerlingh Onnes employed the new liquid as a refrigerant, and he proceeded to study the electrical resistance of various metals at low temperature. Typically in a metal the resistance decreases with the temperature, and it was predicted even by pre-quantum-mechanical theories that if absolute zero could be achieved in a metal whose interatomic structure was perfectly regular, the electrical resistance would be zero. Kamerlingh Onnes found, however, that the resistance of a mercury wire suddenly fell to zero at 4.2 degrees K., a low temperature but still infinitely far above absolute zero. Soon after the initial discovery investigators at Leiden observed superconductivity in other metals: in lead below 7.2 degrees K. and in tin below 3.7 degrees.

The temperature at which the transition to superconductivity takes place is called the critical temperature. As new superconductors were discovered the greatest measured value of the critical temperature increased slowly at an average rate of .3 degree per year. Relatively high critical temperatures have been found for metallic alloys such as vanadium-silicon (17 degrees), niobium-tin (18 degrees) and niobium-aluminum (18.7 degrees). Nevertheless, the steady rise of the maximum critical temperature since 1911 now shows signs of leveling off. The most recent increase was recorded in 1973 for an alloy of niobium and germanium (Nb_3Ge); the critical temperature of the material is 23.2 degrees.

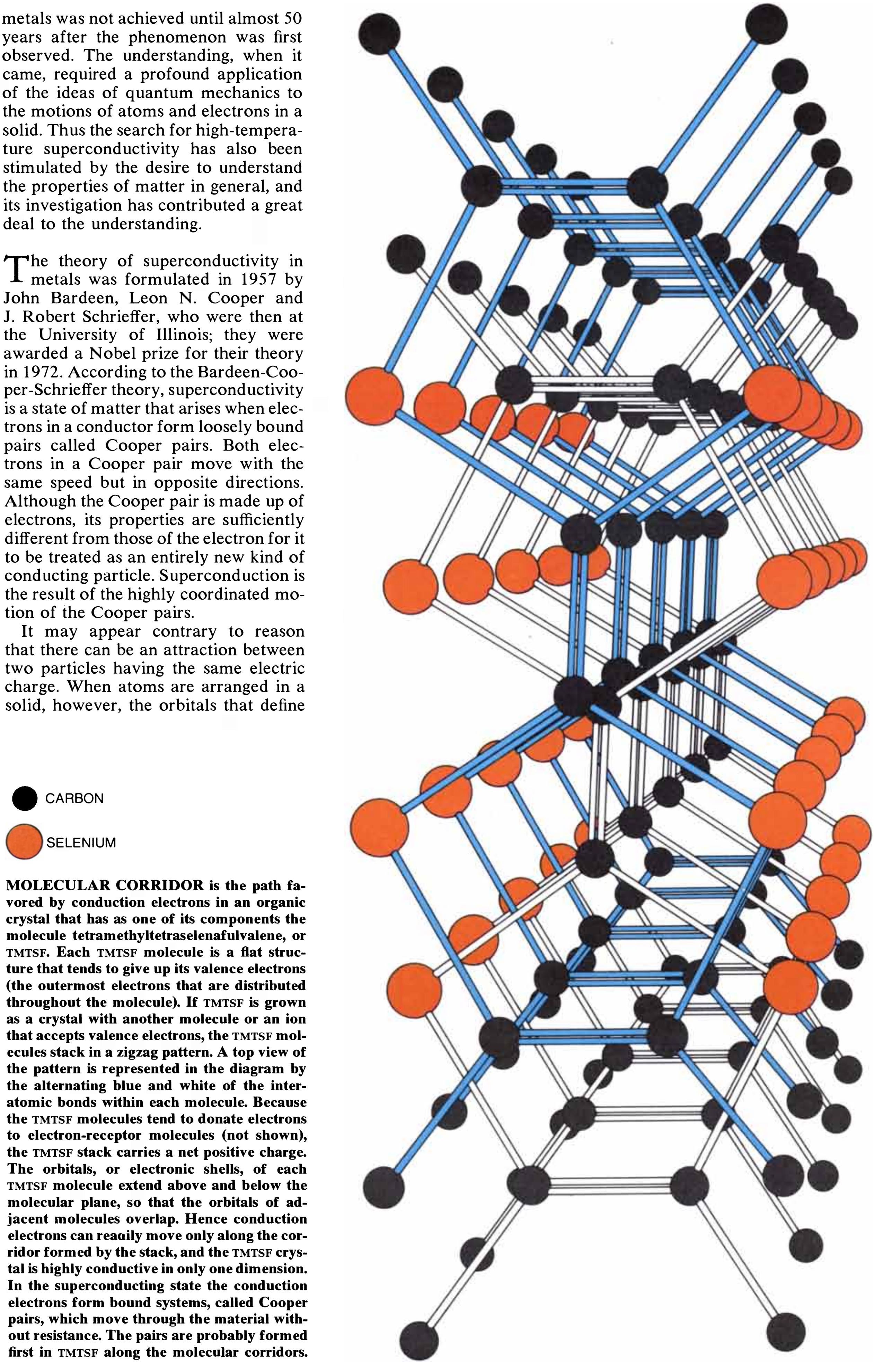
It has been apparent since its discovery that if superconductivity could be developed in materials at less severe temperatures, its technological potential would be almost unbounded. Lossless power transmission, more efficient electric motors and enormously powerful electromagnets could become generally available. Supercomputers have been proposed that exploit the small heat output of superconducting circuit elements. The speed of a computer is fundamentally limited by its size, which determines the average distance a signal must travel within the computer. The size of a computer is limited in turn by the ability of its circuit elements to dissipate heat. If the density of circuit elements is too great, the heat they generate will melt the computer. Superconducting elements could largely circumvent the problem. Moreover, such elements can respond to a signal much faster than semiconducting devices.

Superconductivity is also characterized by a magnetic effect, now called the Meissner effect, first observed in 1933 by the German investigators K. W. Meissner and R. Ochsenfeld. A superconductor placed in a magnetic field generates its own internal field that tends to expel the external one. The lines of force of the external magnetic field bend around the surface of the superconductor instead of passing through it. (In a certain class of superconductors the magnetic lines of force may penetrate the surface of the material to a shallow depth.) The Meissner effect also suggests a technological application: a roadway made of a superconducting material would repel a magnetic field generated by a vehicle. Such a vehicle would float above the roadway on a magnetic cushion and move virtually without friction.

The search for high-temperature superconductivity has therefore been one of the driving forces in solid-state physics. Yet in spite of the technological and economic incentive, a fundamental understanding of superconductivity in metals was not achieved until almost 50 years after the phenomenon was first observed. The understanding, when it came, required a profound application of the ideas of quantum mechanics to the motions of atoms and electrons in a solid. Thus the search for high-temperature superconductivity has also been stimulated by the desire to understand the properties of matter in general, and its investigation has contributed a great deal to the understanding.

The theory of superconductivity in metals was formulated in 1957 by John Bardeen, Leon N. Cooper and J. Robert Schrieffer, who were then at the University of Illinois; they were awarded a Nobel prize for their theory in 1972. According to the Bardeen-Cooper-Schrieffer theory, superconductivity is a state of matter that arises when electrons in a conductor form loosely bound pairs called Cooper pairs. Both electrons in a Cooper pair move with the same speed but in opposite directions. Although the Cooper pair is made up of electrons, its properties are sufficiently different from those of the electron for it to be treated as an entirely new kind of conducting particle. Superconduction is the result of the highly coordinated motion of the Cooper pairs.

It may appear contrary to reason that there can be an attraction between two particles having the same electric charge. When atoms are arranged in a solid, however, the orbitals that define



MOLECULAR CORRIDOR is the path favored by conduction electrons in an organic crystal that has as one of its components the molecule tetramethyltetraselenafulvalene, or TMTSF. Each TMTSF molecule is a flat structure that tends to give up its valence electrons (the outermost electrons that are distributed throughout the molecule). If TMTSF is grown as a crystal with another molecule or an ion that accepts valence electrons, the TMTSF molecules stack in a zigzag pattern. A top view of the pattern is represented in the diagram by the alternating blue and white of the interatomic bonds within each molecule. Because the TMTSF molecules tend to donate electrons to electron-receptor molecules (not shown), the TMTSF stack carries a net positive charge. The orbitals, or electronic shells, of each TMTSF molecule extend above and below the molecular plane, so that the orbitals of adjacent molecules overlap. Hence conduction electrons can reaoily move only along the corridor formed by the stack, and the TMTSF crystal is highly conductive in only one dimension. In the superconducting state the conduction electrons form bound systems, called Cooper pairs, which move through the material without resistance. The pairs are probably formed first in TMTSF along the molecular corridors.

the spatial distribution of their outer electrons overlap. Such electrons are called valence electrons; they are the least tightly bound electrons in an atom, and they tend to wander throughout the overlapping orbitals in the solid. The atoms that remain fixed in the lattice are in effect positively charged ions, since they have been stripped of at least one valence electron. The attraction between valence electrons that make up a pair develops through the mediation of the positive ions in the lattice.

As an electron moves through the lattice the electrostatic attraction between the electron and the positive ions generates a ripple or distortion in the lattice. The passage of the negatively charged electron pulls the positive ions together in its wake; interactions among the ions give rise to restorative forces that cause the lattice to vibrate. The vibratory motions of the lattice are much slower than the motions of the electron. Hence for a relatively long period after the passage of the electron the positive ions remain close together. They create a region of net positive charge in the lattice to which another electron can be attracted, and a Cooper pair is formed.

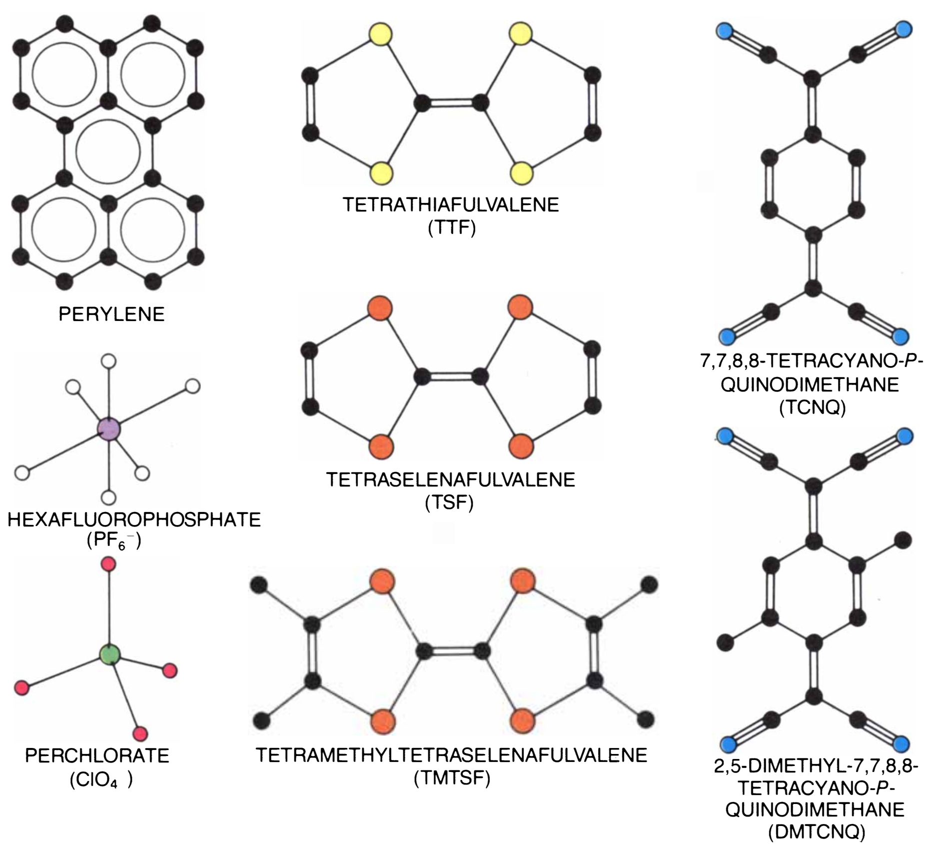
As the ions move apart under the influence of the restorative forces a region is created in which the density of positive charge is less than it is in the surrounding parts of the lattice. The two electrons tend to avoid this region. At high temperature the effect is overwhelmed by the random, thermal vibrations of the lattice, but it can emerge below the critical temperature and cause the electrons that make up each Cooper pair to act as if they were connected by a spring. The electrons that form the pair vibrate back and forth in response to the variations in charge density set up by their own motions in the atomic lattice.

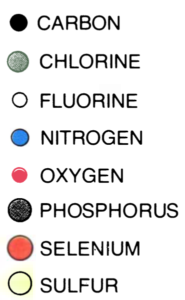
Electrical conduction in a superconductor is caused by the net motion of the centers of mass of the Cooper pairs, whereas electrical conduction in an ordinary metal is caused by the net motion of individual valence electrons. In order to explain the motion of the Cooper pairs in the lattice, however, one must draw on the principles of quantum mechanics. When the electrons enter the bound state, the total energy of the system is minimized if all the Cooper pairs have the same momentum. If no current is flowing through the material, the momentum of each Cooper pair is zero: the electrons in each pair are thereby constrained to move in opposite directions with respect to the lattice, and the velocity of the center of mass of each pair must be exactly zero. When there is a current in the material, the electrons in each pair must move in such a way that the centers of mass of all the Cooper pairs have the same constant momentum. Like the hands of a tightrope walker on his balance bar, every motion of one electron that threatens to break up the Cooper pair must be immediately compensated for by its partner so that equilibrium is always maintained.

Why does a material that resists the flow of individual electrons not also resist the flow of Cooper pairs? Ordinary electrical resistance is a result of the scattering of moving electrons by imperfections in the atomic lattice and by the thermal vibrations of the lattice at any temperature above absolute zero. The current or net motion of the electrons through the lattice is impeded by the scattering, and the energy of the electronic system is correspondingly reduced by an extremely small amount. The scattering of each electron and the resulting decrease in the electric current are accomplished by a small change in the vibrational energy stored in the lattice at a finite temperature. The resistivity of a perfectly regular metallic lattice would theoretically be zero at absolute zero because at that temperature there is no vibrational energy in the lattice available for scattering the electrons.

In a superconductor, on the other hand, the electrical resistivity can be zero at a finite temperature because once the Cooper pairs are put into motion by an electric field they acquire a net momentum. The net momentum cannot decay through scattering: unlike the scattering of an ordinary electron, which costs little energy, the scattering of a Cooper pair requires much more energy than the binding energy that holds the electrons together. Furthermore, in the highly organized superconducting state a change in the momentum of one pair requires a change in the momenta of all the other pairs as well. Hence the energy needed to redistribute the momenta of the Cooper pairs and so to create electrical resistance is much larger than the vibrational energy available in the lattice at low temperature.

The binding of the electrons does not





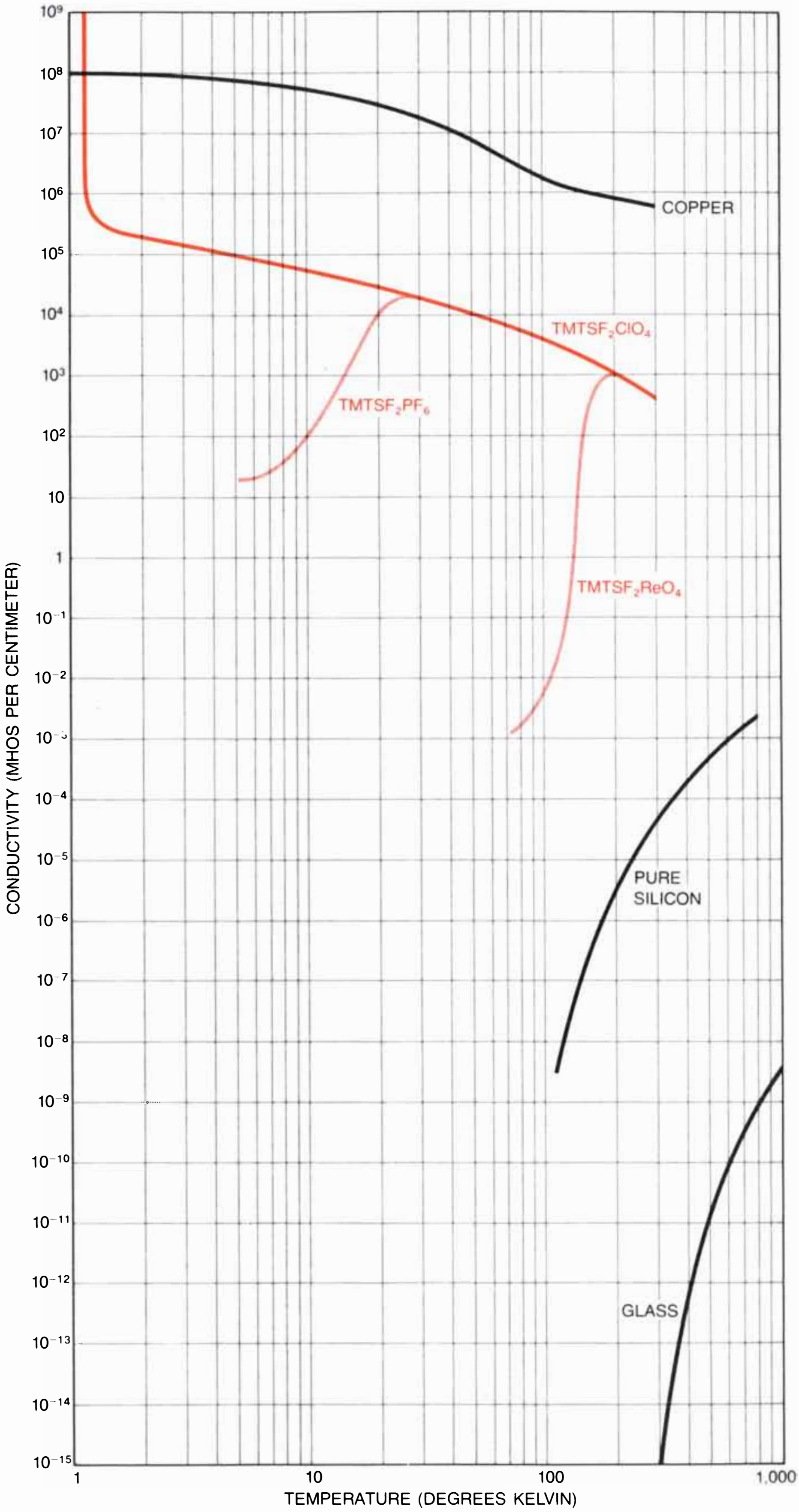
STRUCTURAL BUILDING BLOCKS of the organic materials in which superconductivity has been detected fit together in long, quasi-one-dimensional structures. Typically two kinds of molecule are combined in one crystal. Because the two molecules differ in their affinity for the valence electrons in the crystal, one kind of molecule tends to accept valence electrons from the other. The resulting structure is a network of stacks of molecules, in which the charge alternates from stack to stack. Perylene bromide was the first highly conductive (but not superconductive) organic substance to be discovered. TTF, TSF and TMTSF are molecules closely related to one another that donate electrons; TCNQ and DMTCNQ accept electrons. PF_6^- and ClO_4^- have been combined with TMTSF to create the first organic superconductors.

take place above the critical temperature because at such higher temperatures the thermal agitation of the electrons tends to break up the pairs. In other words, if there were enough energy available in the lattice to redistribute the momenta of the pairs, the energy would be sufficient to break them up as well. Below the critical temperature the pairs can persist in their lockstep motion through the lattice for years without detectable change. Above the critical temperature the electrons are scattered as they are in an ordinary conductor; the resistance of the material jumps from zero to some finite value.

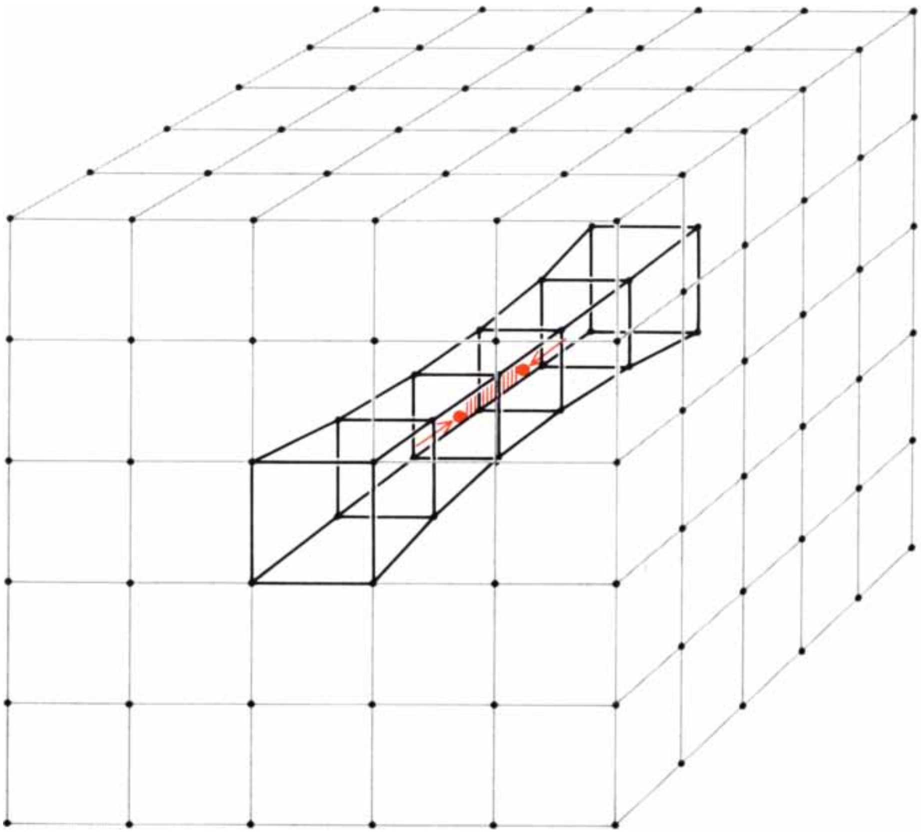
Since the condensation of electrons into Cooper pairs depends on deformations of the atomic lattice, the critical temperature for superconductivity is related to the stiffness of the lattice: its resistance to deformation by a passing electron. If the atoms of the lattice have small masses and the interatomic bonds are relatively "soft," the deformations are large compared with the deformations caused by thermal vibrations and the critical temperature is high. More precisely, the critical temperature varies inversely with the square root of the mass of the ions in the lattice.

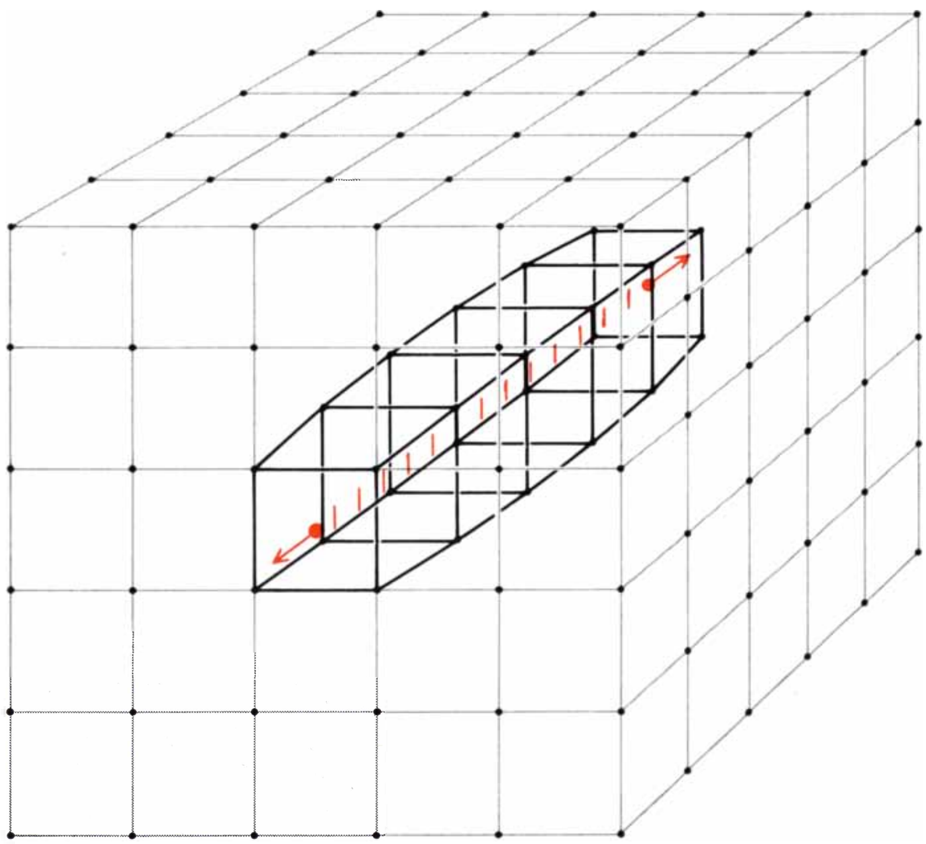
William L. McMillan of Bell Laboratories has argued, however, that the softening of the lattice may establish an upper limit to the critical temperature for superconductivity. When the lattice becomes softer, its vibration frequency decreases. The lower frequency tends to inhibit the formation of Cooper pairs, an effect McMillan predicts will become dominant between 25 and 30 degrees K. Thus McMillan's theory may account for the failure to find any substance with a critical temperature higher than that of niobium-germanium.

In 1964 W. A. Little of Stanford University proposed an alternative mechanism for the formation of Cooper pairs, which does not depend on lattice distortions [see "Superconductivity at Room Temperature," by W. A. Little; SCIENTIFIC AMERICAN, February, 1965]. According to Little, a superconductor could be fabricated from a chain of organic molecules. A spine of carbon atoms would act as a conducting path for Cooper pairs, and organic molecules attached to the spine would carry out the mediating role of the deforming lattice. Little proposed that each organic molecule be a hydrocarbon (specifically, a dye) with loosely bound valence electrons. Such a molecule can easily be polarized, so that one part acquires a positive charge and another part a negative charge. An electron traveling along the spine of carbon atoms would repel the outer electrons of the hydrocarbon molecules, thereby creating a region of net positive charge on the regions of the hy-



ELECTRICAL CONDUCTIVITY is one of the most widely varying quantities in nature. Materials can differ in conductivity by a factor of more than 10^{20}. (If superconductivity is considered, the factor is infinite.) Metallic conductors, such as copper, become more conductive as the temperature decreases. In organic conductors, however, structural changes or the onset of magnetic properties can cause the material to become a semiconductor or an insulator at low temperatures. The dependence of conductivity on temperature is plotted for various materials.





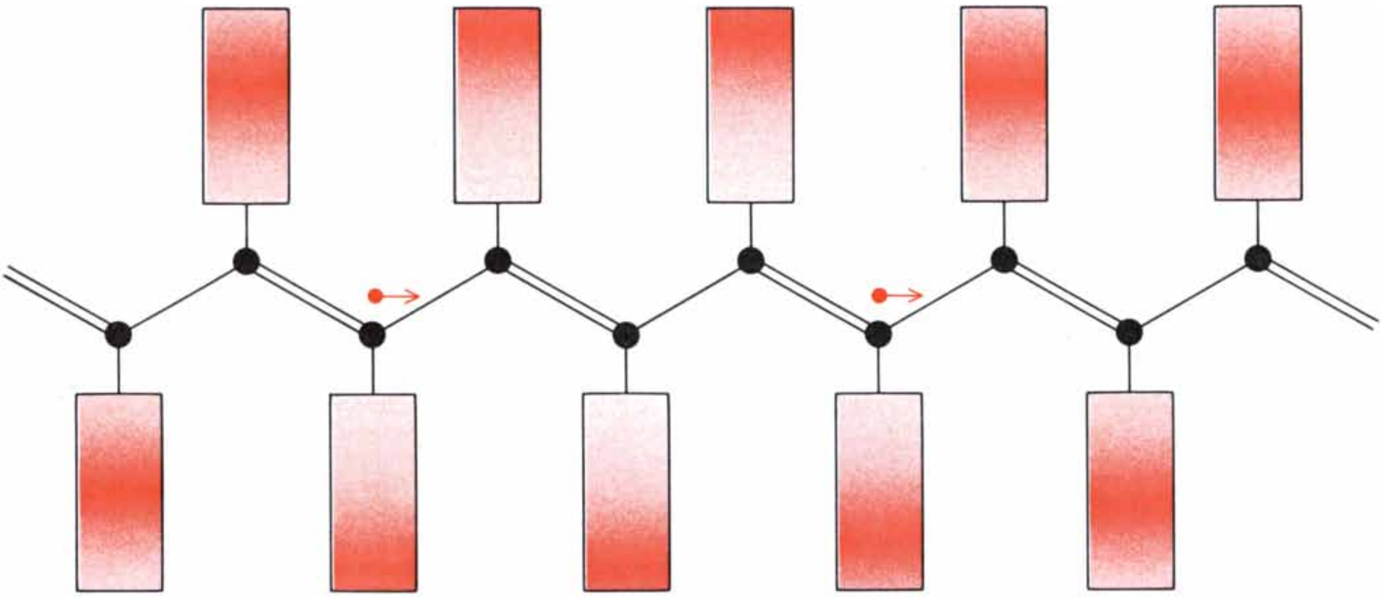
SUPERCONDUCTIVITY in a metal lattice arises through the interaction of conduction electrons with the lattice. A negatively charged electron moving through the metal attracts the fixed, positively charged ions of the lattice, which are therefore drawn together in the wake of the electron. The region of net positive charge formed in this way tends to attract a second electron, and so the motion of the second electron can become correlated with the motion of the first one. The two electrons form a Cooper pair, acting somewhat as if they were connected by a spring. They move at the same speed but in opposite directions. In regions of excess positive charge the electrons move toward each other (*upper illustration*). Restorative forces on the lattice also act like springs. The positive ions overshoot their equilibrium position and begin to oscillate, briefly creating a region where the density of positive charge is less than it is in the surrounding lattice. The two bound electrons avoid this region and move apart (*lower illustration*).

drocarbon molecules nearest the carbon spine. A second electron moving along the spine would be attracted by the relatively dense positive charge and so would be attracted indirectly to the first electron.

Little argued that in such a model one could envision superconductivity at room temperature; indeed, he suggested that in principle superconductivity could be established at temperatures as high as 2,000 degrees K. Such a high critical temperature could come about, he maintained, because the region of enhanced positive charge in the wake of the first electron is caused by the displacement of an electron in the hydrocarbon molecule instead of by the displacement of an atom in a lattice. Because the electron is about 100,000 times less massive than a typical positive ion in an ordinary metal the tendency to form Cooper pairs would be increased by a factor of the square root of 100,000, or roughly 300. Although Little's idea continues to be highly controversial, it has generated strong interest in the development of organic superconductors.

No one has yet succeeded in building a molecule even approximating the properties Little envisioned. The first organic superconductors have instead been an outgrowth of attempts to build ordinary electrical conductors from organic materials. The effort is historically linked to the development of electronic components made of silicon and other semiconductors. Organic chemists and solid-state physicists sought to match the range of electrical properties exhibited by the chemical elements and so create from organic substances what Little called a "shadow periodic table." Because most synthetic organic materials such as vinyl or polyethylene are electrically insulating the search for organic conductors and superconductors has been a difficult one.

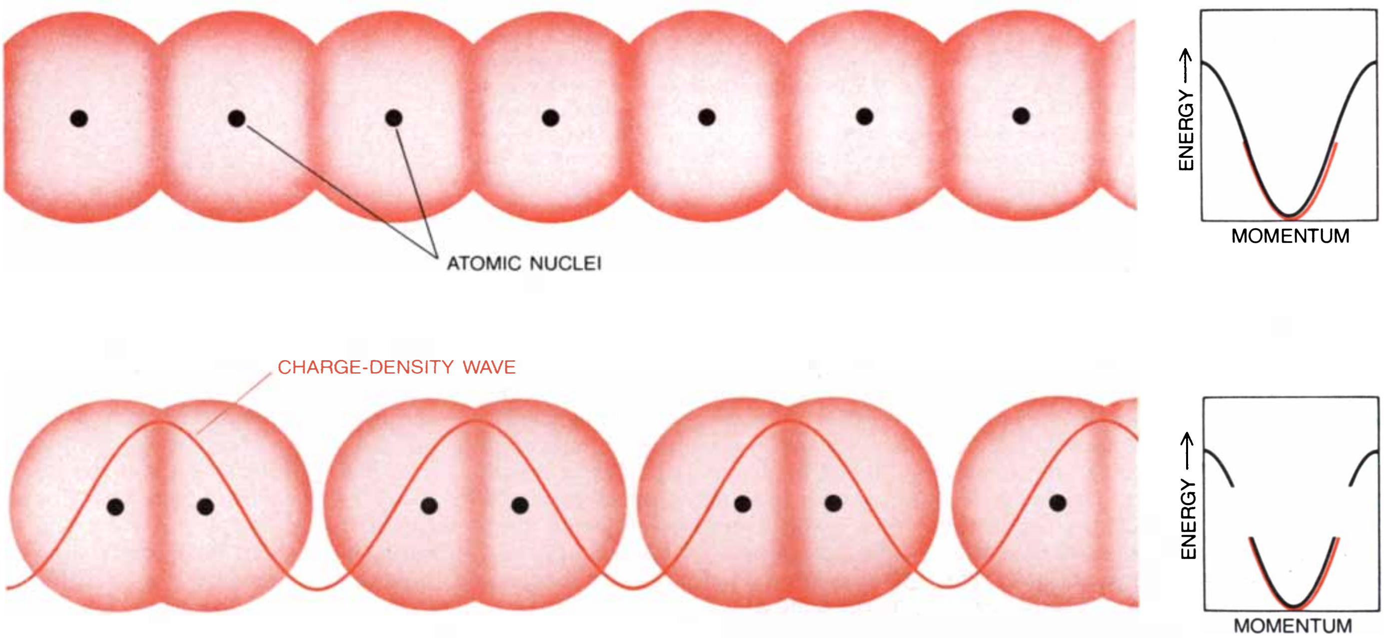
Electrical conductivity is one of the most widely varying properties in nature: the conductivity of a good conductor such as copper is greater than 10^{20} times the conductivity of a good insulator such as glass. As we have mentioned, ordinary conduction depends on a net flow, or drift, of the valence electrons through a material. Under the influence of an electromagnetic field the essentially chaotic motions of the valence electrons are altered slightly to favor a single direction. In an isolated atom each valence electron occupies a discrete orbital, which represents a state with a particular energy. In a molecule or in a solid, however, the principles of quantum mechanics do not allow more than two electrons to have precisely the same energy and momentum. As a result each discrete energy level of the atomic valence electrons is partitioned into many slightly different levels that are spread



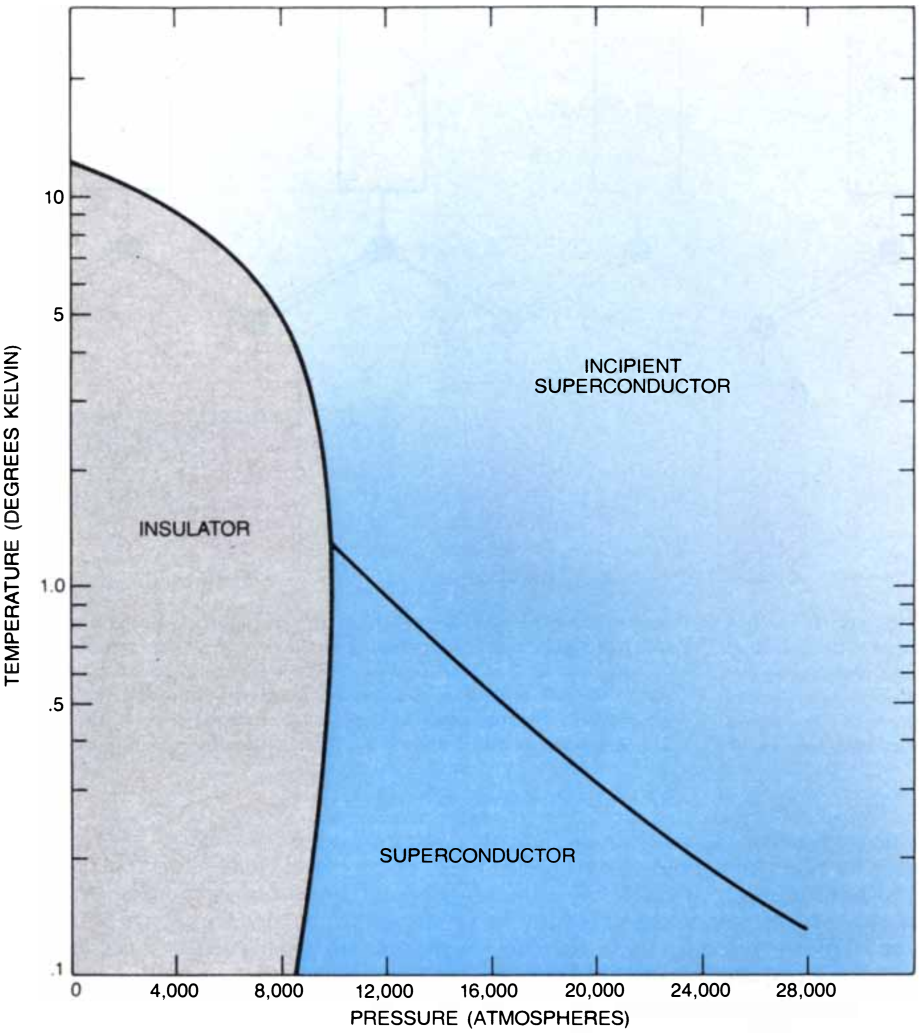
COOPER PAIRS in an organic superconductor may also form by a different mechanism, one based on the polarization of side molecules attached to a spine of carbon atoms. As conduction electrons move along the spine the valence electrons of the side molecules are repelled and tend to concentrate their charge away from the spine. An excess of positive charge is thereby created on the region of each side molecule nearest the conducting spine, so that another electron moving along the spine can become bound to the first one. Once two electrons become bound they oscillate as if they were connected by a spring, just as they do in a metal. The arrows indicate the motion of the electrons during the formation of the bond, before they begin to oscillate. This mechanism was proposed by W. A. Little of Stanford University.

throughout a nearly continuous band.

The drift of electrons through a material corresponds to an increase in the total electronic energy of the material; it can readily take place only if the allowed energy states in the valence energy band are not entirely occupied. The valence electrons in such a material can drift through the regions of the lattice in which the partially filled atomic orbitals overlap. If the valence band is fully occupied, on the other hand, the total electronic energy can be increased only by raising the energy of some of the valence electrons across an energy gap: a range of energy levels that is forbidden by the laws of quantum mechanics. Any net motion of the valence electrons in such a material is inhibited because there is no place for them to go within the overlapping atomic orbitals of the material.



ONE-DIMENSIONAL CONDUCTION can be understood by considering a simple system in which it might be observed: a linear array of sodium atoms whose valence-electron orbitals overlap. Each sodium atom has one valence electron, and when the orbitals overlap, the energy states of the valence electrons form a nearly continuous band of energy levels. The band is only half filled with electrons; if an external electric field is applied, some of the electrons can easily move to unoccupied energy states within the band. The shift in the energy distribution of the electrons corresponds to a net motion of electrons along the overlapping atomic orbitals (*upper illustration*). Because the system is one-dimensional, however, its energy can be reduced if the elements of the linear array are not uniformly distributed. The charge density of the valence electrons in such an array may also vary with a periodicity equal to twice the interatomic spacing. The distortions in the structure and in the charge density split the energy band in the middle, leaving a full band and an empty band separated by a gap (*lower illustration*). Such a configuration cannot readily conduct a current because the required shift in the energy levels is not easy to make. Enough energy to promote valence electrons across the energy gap must be supplied before the array can become conductive.



PHASE DIAGRAM of the organic compound TMTSF$_2$PF$_6$ illustrates the typical response of an organic superconductor to variations in temperature and pressure. At low pressure and temperature (*gray region*) the material is magnetic and becomes an insulator. The effect of magnetic ordering is similar to the effect of the onset of a charge-density wave: an energy gap is created between the filled and the empty parts of the valence-electron energy band. At much higher pressure the material becomes superconductive (*solid colored region*). The authors have suggested that outside these two regions and up to about 30 degrees Kelvin superconductivity is incipient: Cooper pairs form for short periods over short distances along the stacks of molecules. The incipient superconductivity is manifested by a low but nonzero electrical resistance.

Thus a material readily conducts a current only if some of the energy states in the valence-electron energy band are not occupied by electrons.

In many organic materials partially filled valence orbitals make it energetically favorable for the material to combine chemically with other substances. Hence the requirement that the valence orbitals be partially filled in a conductor can lead to the preparation of a substance that is chemically unstable. The first highly conducting organic compound, for example, was reported by Hideo Akamatu, Hiroo Inokuchi and Yoshia Matsunaga of the University of Tokyo in 1954. The compound was made by treating the organic molecule perylene with bromine, but it broke down almost immediately into a nonconducting substance through continued reaction with the bromine. Because of such chemical instability relatively few molecules suitable for constructing organic conductors have been found.

The structure of any successful organic conductor is probably dictated by two requirements. First, its molecular building blocks must fit closely together, so that the conduction electrons can move easily from one molecule to another. Second, the energy cost of partially filling or partially opening a valence energy band must be small. All the molecules investigated so far have a large planar skeleton made up of carbon and hydrogen atoms and diverse atoms such as nitrogen, sulfur and selenium. The valence electrons of such a molecule are typically found above and below the planar framework. Hence the molecules can stack close together like pancakes and the valence electrons can propagate a current primarily from plane to plane along the stack.

One of the first molecules of this kind to be synthesized is 7,7,8,8-tetracyano-*p*-quinodimethane, which is usually abbreviated TCNQ. It was prepared in 1960 by workers at E. I. du Pont de Nemours & Company. Little energy is needed to introduce an extra electron into the TCNQ molecule and so to begin filling a new valence energy band in a stack of such molecules. The negatively charged structure that results is chemically stable. TCNQ by itself, however, is not able to conduct an electric current: all the TCNQ molecules have the same affinity for electrons, and so in an electrically neutral system there is no tendency for electrons to move from one TCNQ molecule to another. If the TCNQ molecules can obtain electrons from other atoms or molecules in their environment that readily give up electrons, the material can become conductive. In the cesium salt of TCNQ (Cs_2TCNQ_3), for example, each cesium atom gives up an average of two-thirds of an electron to each TCNQ molecule. Thus two out of every three TCNQ molecules become negatively charged, and if an electric field is imposed, the extra electrons can move from the charged TCNQ molecules to the neutral ones.

In the early 1970's Fred Wudl of the State University of New York at Buffalo prepared an organic molecule called tetrathiafulvalene, or TTF, which incorporates four sulfur atoms. TTF is a counterpart to TCNQ in that it readily gives up an electron to form a stable, positively charged structure. When Wudl combined TTF with chlorine atoms, the chlorine accepted some of the electrons given up by the TTF and the resulting material was highly conductive.

In 1973, shortly after TTF was first synthesized, Dwaine O. Cowan and John P. Ferraris of Johns Hopkins University and Alan J. Heeger and Anthony F. Garito of the University of Pennsylvania independently combined the two organic molecules to form TTF-TCNQ. In the solid form of this substance the TTF and the TCNQ molecules stack in separate columns, and electrons are donated from the TTF stack to the TCNQ stack. X-ray experiments by Robert Comes and his co-workers at the Université de Paris-Sud at Orsay have shown that .59 electron per molecule is transferred from one stack to the other. Because of the electron transfer there can be a net motion of electrons along both stacks; hence the material is conductive.

The architecture of the TTF-TCNQ crystal gives rise to a striking electrical property: the material is highly conductive in one direction only. The most favorable direction is 500 times as conductive as the least favorable direction.

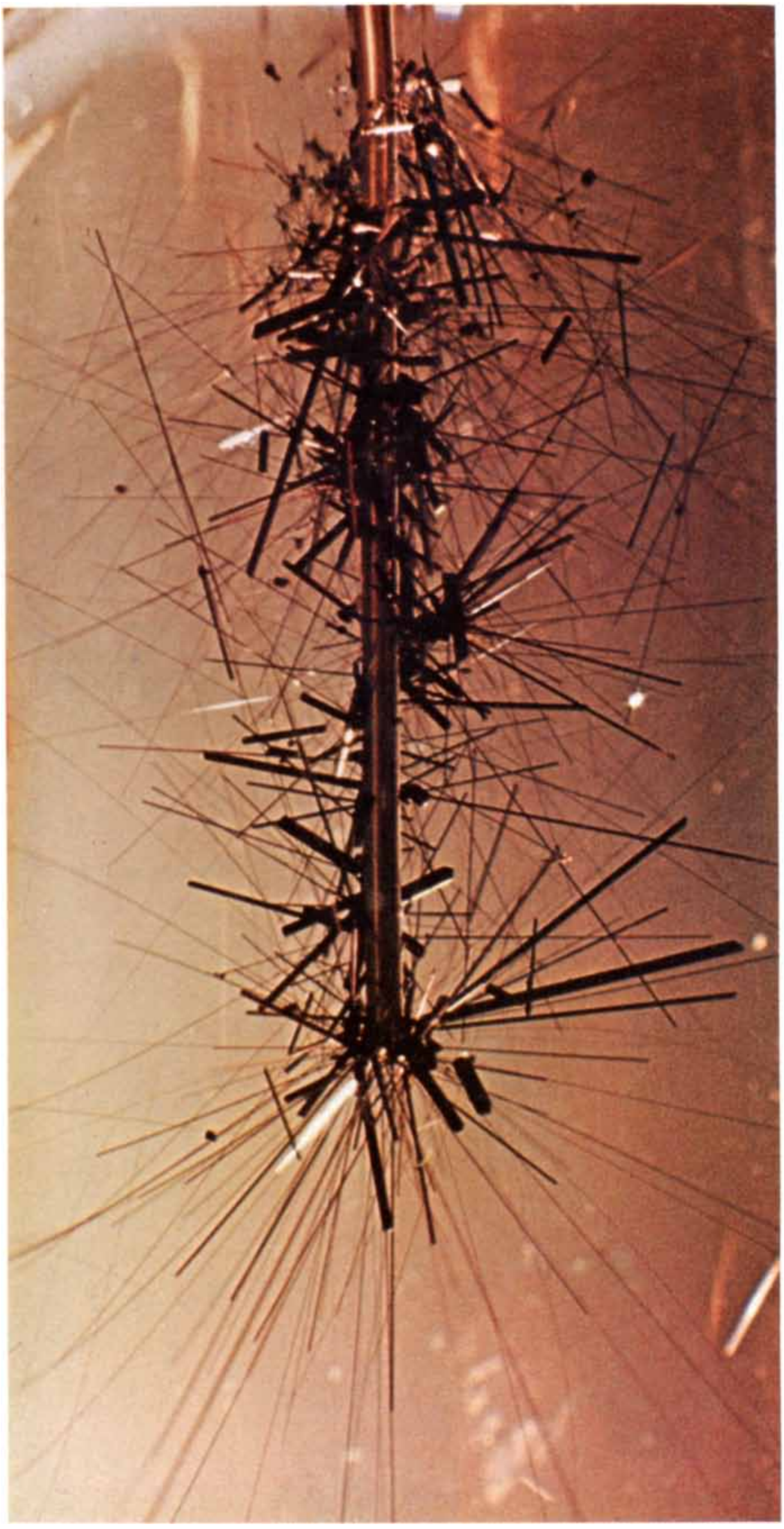
The reason for the anisotropic conductivity is that all the constituent molecules of the crystal are stacked in parallel planes. The orbitals of the molecules interact mainly along the stacks.

In view of Little's suggestions about organic superconductivity, the discovery of such a one-dimensional organic conductor is highly suggestive. Could TTF-TCNQ, perhaps under extreme conditions of temperature and pressure or perhaps with slight modifications to its molecular structure, be made superconducting according to the mechanism Little proposed? It is known from theoretical calculations that it is impossible to sustain the formation of Cooper pairs in a strictly one-dimensional conductor at any temperature above absolute zero. Thus the one-dimensional superconductor, as proposed by Little, cannot exist. (The formation of Cooper pairs through the mediation of polarizing electrons, however, which is the heart of Little's theory, is not thereby disproved.) Moreover, the structure of the stacks of TTF-TCNQ is actually quite different from the linear conducting chain Little described.

In spite of such differences the analogy with Little's hypothetical superconductor is not without merit. In any real molecular system the conductors are only approximately one-dimensional; there is always some interaction between stacks of molecules that would allow a Cooper pair to jump from one stack to another. Hence the theoretical obstacle to superconductivity in one-dimensional systems is not a practical one. Moreover, investigations of the properties of TTF-TCNQ have led to the development of related compounds that are quasi-one-dimensional and superconductive. It is not yet known whether the mechanism for Cooper pairing in such systems is the one Little envisioned.

In order to develop superconductivity in organic materials investigators had to study and overcome problems unique to one-dimensional systems. One of the most important properties of such systems was noted by Rudolf Peierls of the University of Birmingham as early as 1954. Peierls pointed out that the uniform separation between adjacent planes of molecules, characteristic of substances such as TTF-TCNQ at room temperature, is not energetically favored at all temperatures. As the temperature is lowered the total energy of the solid can be reduced if the interplanar spacing and the distribution of the charge density of the valence electrons are modified.

The charge density of the electron undergoes a periodic concentration and rarefaction that is called a charge-density wave. When the charge-density wave forms and the lattice is distorted, the partially filled valence-electron energy



CRYSTALS OF TMTSF$_2$PF$_6$ are formed by passing an electric current through a solution in which molecules of TMTSF and negative PF$_6^-$ ions are dissolved. Valence electrons from the TMTSF molecules are drawn out through the positive electrode, and crystals of high purity collect there. The needlelike crystals in the photograph are from five to 10 millimeters long.

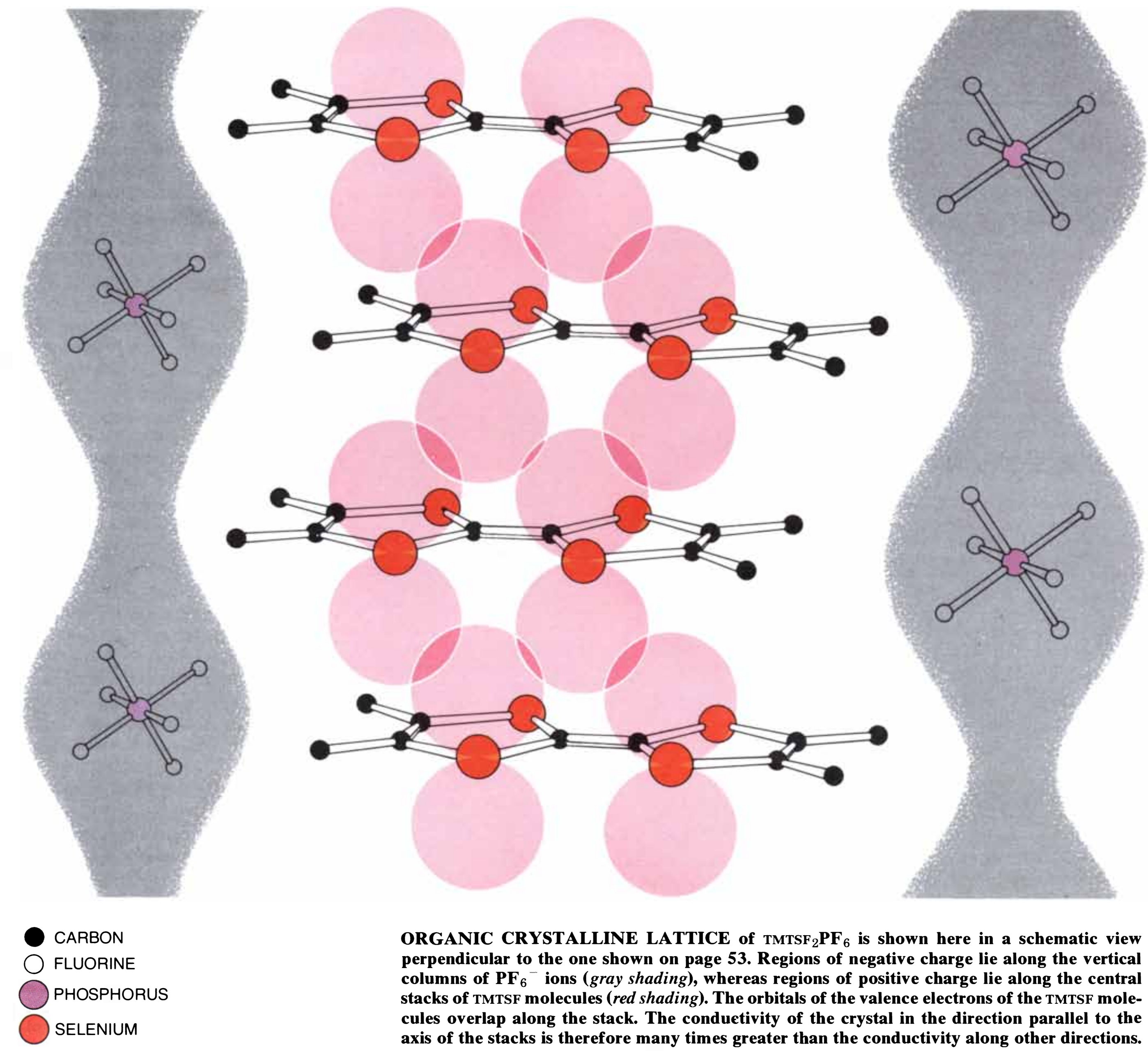
band splits into two bands, the lower one completely filled and the upper one completely empty. After such a split conduction can occur only if electrons are promoted across the energy gap by an external source of thermal or electromagnetic energy.

The onset of the charge-density wave can thus cause a material to change abruptly from a conductor into a semiconductor or an insulator. Typically the transition takes place between 50 and 100 degrees K.; the mechanism causes TTF-TCNQ to become a semiconductor below 53 degrees.

To modify the electron-donating properties of TTF and similar molecules investigators substituted heavier selenium atoms for the sulfur atoms in TTF. Edward M. Engler of the International Business Machines Corporation prepared tetraselenafulvalene-TCNQ, or TSF-TCNQ, and one of us (Bechgaard) prepared tetramethyl-TSF, or TMTSF, while working with a group at Johns Hopkins. It is the latter molecule that was finally employed to create superconducting organic molecular crystals.

At first TMTSF looked like a real loser. When it was reacted with TCNQ, red, insulating crystals were obtained. TMTSF-TCNQ is quite stable in this state; with luck and elaborate experimental techniques black, conducting crystals representing a different phase, or structural configuration, can be prepared. Below 60 degrees K. a charge-density wave splits the valence energy band and the material becomes an insulator again.

In collaboration with Jan R. Andersen of the Risø National Laboratory in Roskilde, Denmark, we reacted TMTSF with a substance closely related to TCNQ called 2,5-dimethyl-TCNQ, or DMTCNQ. The resulting crystals of TMTSF-DMTCNQ conduct current at room temperature along the parallel stacks of molecules in a way similar to that observed in TTF-TCNQ, but the material becomes an insulator below 41 degrees K. At Orsay we had already begun to study the effects of high pressure on the electronic properties of TTF-TCNQ and similar systems. In 1973 Meir Weger of the Hebrew University of Jerusalem had suggested that the insulating phase of such materials, caused by charge-density distortions, could be inhibited by squeezing the chains of molecules together under high pressure. We had therefore hoped that high pressure would induce superconductivity in TTF-TCNQ, but we found instead that high pressure helps to stabi-



ORGANIC CRYSTALLINE LATTICE of $TMTSF_2PF_6$ is shown here in a schematic view perpendicular to the one shown on page 53. Regions of negative charge lie along the vertical columns of PF_6^- ions (*gray shading*), whereas regions of positive charge lie along the central stacks of TMTSF molecules (*red shading*). The orbitals of the valence electrons of the TMTSF molecules overlap along the stack. The conductivity of the crystal in the direction parallel to the axis of the stacks is therefore many times greater than the conductivity along other directions.

lize the insulating phase of the material.

If pressure is applied to TMTSF-DMTCNQ at room temperature, however, a new phase of the crystal is generated. If the material is then cooled under pressure, it does not become an insulator but instead remains highly conducting in the temperature range of liquid helium, that is, below about four degrees K.

Although the phase transition is not yet fully understood, the stabilization of a highly conducting state of TMTSF-DMTCNQ was a milestone in the quest for an organic superconductor. At low temperature and at a pressure of 10,000 atmospheres the conductivity of the material exceeds 200,000 mhos per centimeter, which exceeds the maximum conductivity of TTF-TCNQ by a factor of 10. (The mho, a unit of conductance, is the reciprocal of the ohm, the basic unit of resistance.)

Although the conductivity of the best conducting metals such as copper and silver can be as great as 100 times the conductivity of TMTSF-DMTCNQ, the mobility of the conducting electrons is much greater along the stacks of an organic molecule than it is in a metal. An electron in the stack of an organic molecule is scattered once every 1,000 molecules; an electron in a metal such as copper is scattered once every 50 to 100 lattice spacings, unless the metal is exceptionally pure. (The greater conductivity of a metallic conductor is a result of its far more numerous conduction electrons.) The highly conducting state of TMTSF-DMTCNQ can be suppressed by applying a magnetic field perpendicular to the stacks or by exposing the material to small amounts of radiation.

Superconductivity has not been observed in TMTSF-DMTCNQ, perhaps because of small amounts of impurities in the samples. Its intriguing properties, however, have led us to propose that its high conductivity is a signature of an incipient, transient superconductivity that arises along the one-dimensional stacks of the crystal below about 30 degrees K. Although superconductivity cannot be sustained in a one-dimensional conductor, transient and fluctuating Cooper pairing can take place along a one-dimensional chain of molecules. The transient Cooper pairs can contribute significantly to electrical conduction well above the critical temperature of about 1.2 degrees K. for bulk, or three-dimensional, superconductivity.

The critical temperature for the formation of Cooper pairs in one dimension depends only on the strength of the pairing interactions within the chain. The critical temperature for three-dimensional superconductivity also depends on the strength of interactions between chains. If the strength of the former interactions is much greater than the strength of the latter, incipient, one-dimensional superconductivity can set in at temperatures well above the three-dimensional critical temperature. The formation of the one-dimensional Cooper pairs is expected to take place gradually as the temperature is decreased.

Our detailed investigations of TMTSF-DMTCNQ indicated that the properties of the material are determined primarily by the stack of TMTSF molecules. We therefore decided to prepare a group of materials in which the properties of the TMTSF stacks are enhanced. To make crystals of high purity we dissolved neutral TMTSF molecules in an organic solvent and added a salt that would release negatively charged ions. We then placed two platinum electrodes in the solution and passed an electric current through it. At the positive electrode electrons were removed from the TMTSF molecules to form positively charged molecules. As soon as a positively charged molecule formed it combined spontaneously with a neutral TMTSF molecule and with one of the negatively charged ions and separated as an insoluble salt at the positive electrode. In this way many so-called $TMTSF_2X$ salts have been prepared, where X represents a negatively charged ion such as hexafluorophosphate (PF_6^-) or perchlorate (ClO_4^-).

The $TMTSF_2X$ salts have quite unusual properties. Unlike other stacks of organic molecules such as TTF-TCNQ, most of the $TMTSF_2X$ salts remain highly conducting at temperatures well below 20 degrees K. The transfer of electrons from TMTSF molecules to the other molecules in the solid and the periodic arrangement of the negatively charged ions in the crystal are such that the material is not susceptible to the distortions introduced by the charge-density wave. It turns out that if an organic conductor is not susceptible to such distortions, the material becomes a superconductor instead of an insulator at low temperature. Ordinary conductivity does not persist as the material is cooled.

In December, 1979, we subjected the salt $TMTSF_2PF_6$ to conditions similar to the ones that resulted in highly conducting TMTSF-DMTCNQ. In an elaborate experiment carried out by Michel Ribault at Orsay, crystals of $TMTSF_2PF_6$ were first pressurized to 12,000 atmospheres and then cooled slowly at constant pressure over a 12-hour period to below one degree K. As the material cooled we measured its conductivity and its susceptibility to a magnetic field. As the temperature approached one degree K. the electrical resistance began to decrease, and at .9 degree it dropped abruptly to zero. The magnetic measurements also showed that at .9 degree a weak magnetic field was expelled from the sample, which is the sign of the Meissner effect in metallic superconductors. The material had become superconductive.

Since then we have found superconductivity in crystals of $TMTSF_2X$ made from hexafluoroarsenate (AsF_6^-), hexafluoroantimonate (SbF_6^-), hexafluorotantallate (TaF_6^-), perrhenate (ReO_4^-) and perchlorate. The properties of the crystals depend somewhat on the properties of the negatively charged ions, although the ions do not take part directly in conduction. At atmospheric pressure the PF_6^- salt becomes magnetic and insulating at 12 degrees K., and the ReO_4^- salt gives way to charge-density-wave distortions and becomes insulating at 180 degrees. High pressures are needed to make the two kinds of crystal superconductive. The crystals of $TMTSF_2ClO_4$, on the other hand, become superconductive at 1.2 degree K. at normal atmospheric pressure.

There is now considerable experimental evidence to support our contention that superconductivity is incipient along the one-dimensional molecular stacks, at temperatures well above the onset of bulk superconductivity. Radiation at far-infrared frequencies is absorbed by the materials above a threshold frequency that is related to the bonding energy for Cooper pairs within a single chain. Experiments demonstrating the movement of electrons from the superconductor to a thin layer of semiconductor applied on the surface of the superconductor also suggest that Cooper pairs form at relatively high temperatures within each chain. Finally, the high conductivity of the materials at temperatures well above the critical temperature can be sharply reduced by placing them in a magnetic field.

These results and others all point to a fairly consistent value for the temperature below which electrons in an isolated chain form Cooper pairs, namely about 30 degrees K. Thus the comparatively low critical temperatures found so far in organic superconductors are not upper limits; on the contrary, the experiments done so far strongly indicate a transient, one-dimensional superconductivity at temperatures higher than those found for any other superconductor.

There is much still to be learned about superconductivity in organic materials. It is not known, for example, whether the Cooper pairs form because of interactions mediated by the lattice, as they do in an ordinary superconducting metal, or whether they form through some mechanism of electronic mediation, as Little proposed. As one-dimensional systems come to be better understood, synthetic chemistry may open the way to the development of superconducting materials that have practical technological applications.

Ultralight Airplanes

The conjunction of the hang glider and the small engine has brought into being the "air recreational vehicle." A typical craft carries 200 pounds and cruises at 50 miles per hour

by Michael A. Markowski

Since the earliest days of aviation a widely sought objective has been an inexpensive airplane that is not unduly difficult to fly. From the time of the Wright brothers, however, the trend has been almost entirely toward bigger and more sophisticated airplanes. Only within the past few years has the conjunction of the hang glider and the small engine (Go-Kart or snowmobile) brought the long-sought objective into being as the ultralight airplane. The numerous models available today range in price from $2,800 to $7,000. For a pilot who has received the proper instruction and is acutely sensitive to the vagaries of the wind the craft are not hard to fly.

Until the ultralight aircraft came on the market most airplanes were designed and built principally to serve a commercial or military purpose. Even light aircraft cost so much to buy and maintain that few people could afford to have one solely for recreation. The ultralight airplane is the first to have been successfully developed and marketed as an "air recreational vehicle" (often abbreviated ARV). Last year more than 10,000 such aircraft were sold, surpassing the sales of general-aviation craft (airplanes employed for commercial purposes other than scheduled passenger service). The ultralights are now even finding commercial application in agricultural surveying, crop dusting and aerial photography. A few have been adapted for military observation service because they have the advantage of presenting virtually no radar image.

Several things besides affordability account for the popularity of the ultralights. One is the sheer pleasure of flying in this way. The pilot of an ultralight is not encapsulated in a cockpit but instead is able to be part of the wind. With the engine off the craft can be soared like a hang glider. At the end of the day the airplane can be folded up and stored at home.

In addition most ultralight airplanes do not have to be registered with the Federal Aviation Administration, nor is the operator required to have a pilot's license. Until recently these exemptions were granted only for craft that could be launched by a person on foot, as a hang glider is. As ultralights evolved, however, their larger engines and higher unloaded weight called for wheeled landing gear. Last fall the FAA set in motion a proceeding intended to establish clearer rules on what type of plane will be exempt. If the recommendations of manufacturers and other leaders in the industry are accepted, the primary criteria will be a maximum empty weight of 220 pounds and a maximum wing loading of three pounds per square foot. With these constraints the exempt craft would have sufficient structural strength and would land at a speed of less than 30 miles per hour, a necessary condition for easy handling.

Most of the models now available have an empty weight of about 200 pounds. The wingspan is 30 feet or more, the cruising speed about 50 m.p.h. and the stalling speed about 25. The glide ratio (the amount of forward movement for each foot of descent) averages 9 : 1 and the rate of climb is 500 feet per minute or more. Most of the craft can lift more than their empty weight, meaning that the pilot and the fuel can have a combined weight of more than 200 pounds.

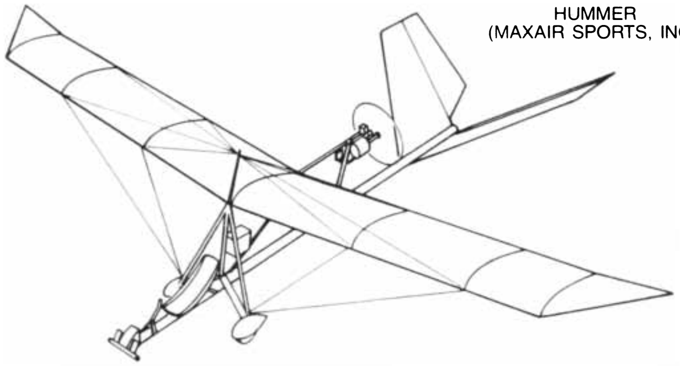
In retrospect it can be seen that the ultralight airplane has its origins in the first attempts of people to fly. Otto Lilienthal of Germany made more than 2,000 flights during the 1890's in what were essentially hang gliders. Lawrence Hargrave of Australia designed and flew model airplanes powered variously by rubber bands and compressed-air and steam engines. He also invented the box kite, which served as the basis for all the externally braced biplanes made later.

In the U.S., Octave Chanute, who in 1894 had written the aeronautical classic *Progress in Flying Machines,* designed a hang glider based on the box kite and incorporating a Pratt truss, which had been patented in 1844 as a method of bracing railroad bridges. The two wings of Chanute's glider were connected by vertical posts braced by crisscrossing wires in both longitudinal and lateral planes, forming a rigid but lightweight structure that became the choice of the Wrights and all future designers of biplanes. The glider also included an aft tail assembly for stability and a curved-section airfoil for improved lift.

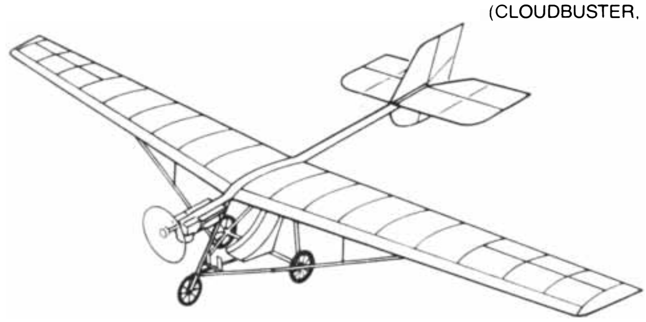
Augustus M. Herring, who had been an assistant to Chanute, was probably the first to fly a powered ultralight aircraft. He built a heavier version of the Chanute biplane hang glider (a glider he had helped Chanute with and had flown for him), mounting a two-cylinder, compressed-air motor ahead of the lower wing to drive two five-foot propellers mounted in tandem, one ahead of the wings and one behind them. On October 11, 1898, Herring made his first powered flight at St. Joseph, Mich., traveling about 267 feet in the air against a head wind approximating 25 m.p.h. He could control the glider only by shifting his weight. A few days later he made another flight. Satisfied that he had proved the feasibility of powered flight, he went on to design a steam engine and a larger aircraft, all of which were destroyed by fire in 1899. Later he organized the Herring-Curtiss Company, the first firm to manufacture airplanes in the U.S.

When the Wright brothers succeeded in piloting their "Flyer" on December 17, 1903, they achieved the first powered flight in which the pilot could control the craft by moving aerodynamic surfaces rather than by shifting his weight. The Flyer was a direct ancestor of today's ultralight airplanes. In succeeding years several other forerunners appeared: the Demoiselle of Alberto Santos-Dumont in 1909, the first truly professional ultralight of D. W. Huntington after World War I, the White Monoplane in 1920, the English Electric Wren in 1923 and the French Pou-du-Ciel in 1935. None of them achieved wide commercial success.

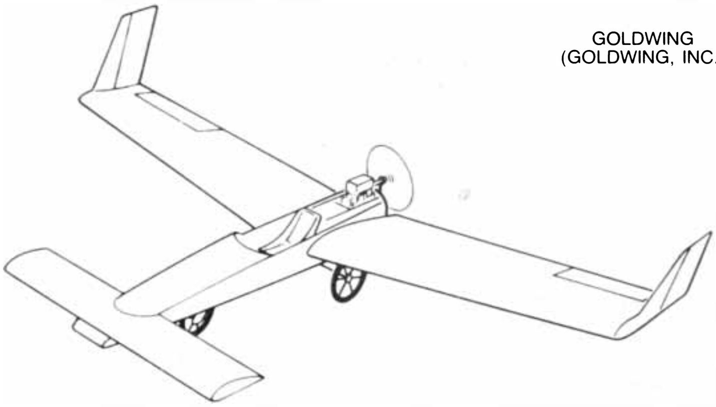
LARGE VARIETY of ultralight airplanes now available is suggested by the 10 versions on the opposite page. Each weighs about 200 pounds and has a wingspan of about 32 feet.



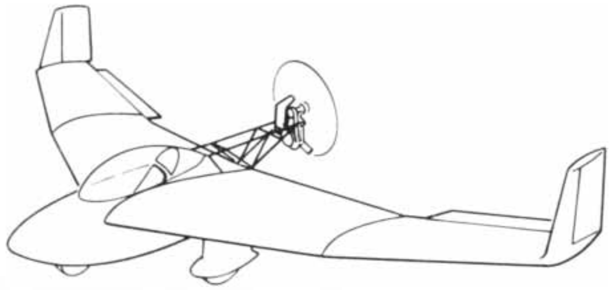
HUMMER
(MAXAIR SPORTS, INC.)



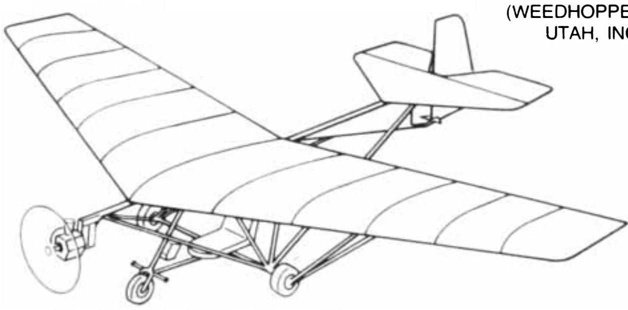
CLOUDBUSTER
(CLOUDBUSTER, INC.)



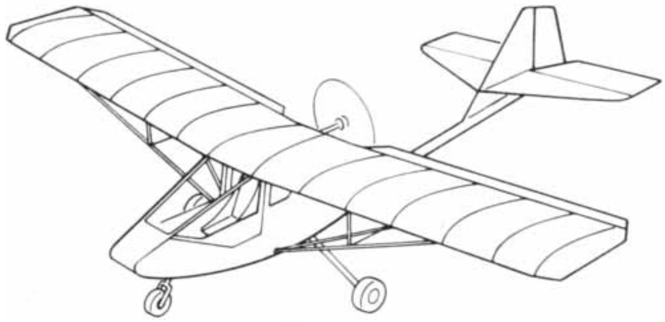
GOLDWING
(GOLDWING, INC.)



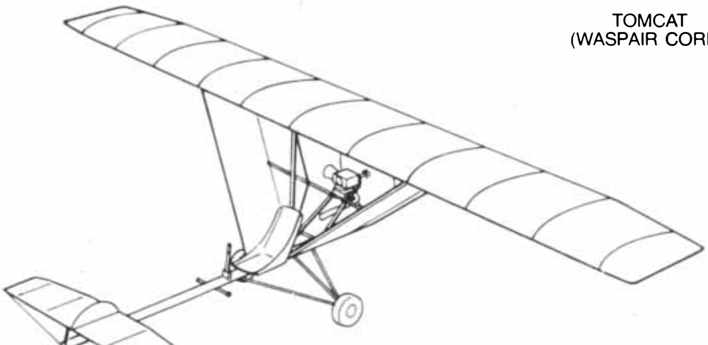
MITCHELL U-2
(MITCHELL AIRCRAFT CORP.)



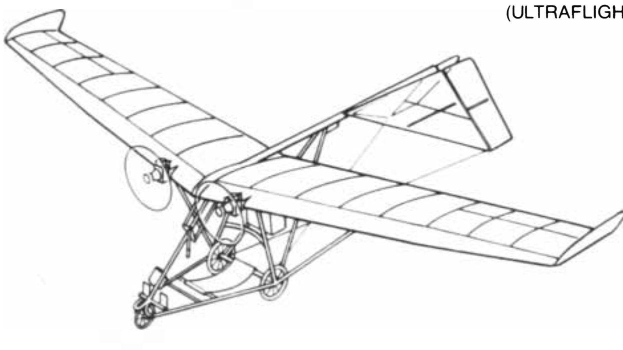
WEEDHOPPER
(WEEDHOPPER OF UTAH, INC.)



CGS HAWK
(CGS AVIATION, INC.)

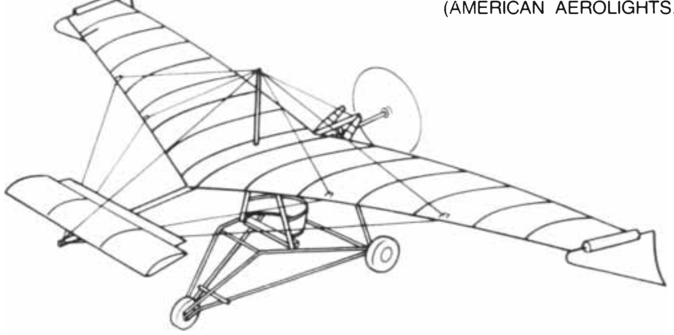


TOMCAT
(WASPAIR CORP.)



LAZAIR
(ULTRAFLIGHT, INC.)

MITCHELL WING
(MITCHELL AIRCRAFT CORP.)



EAGLE CANARD
(AMERICAN AEROLIGHTS, INC.)

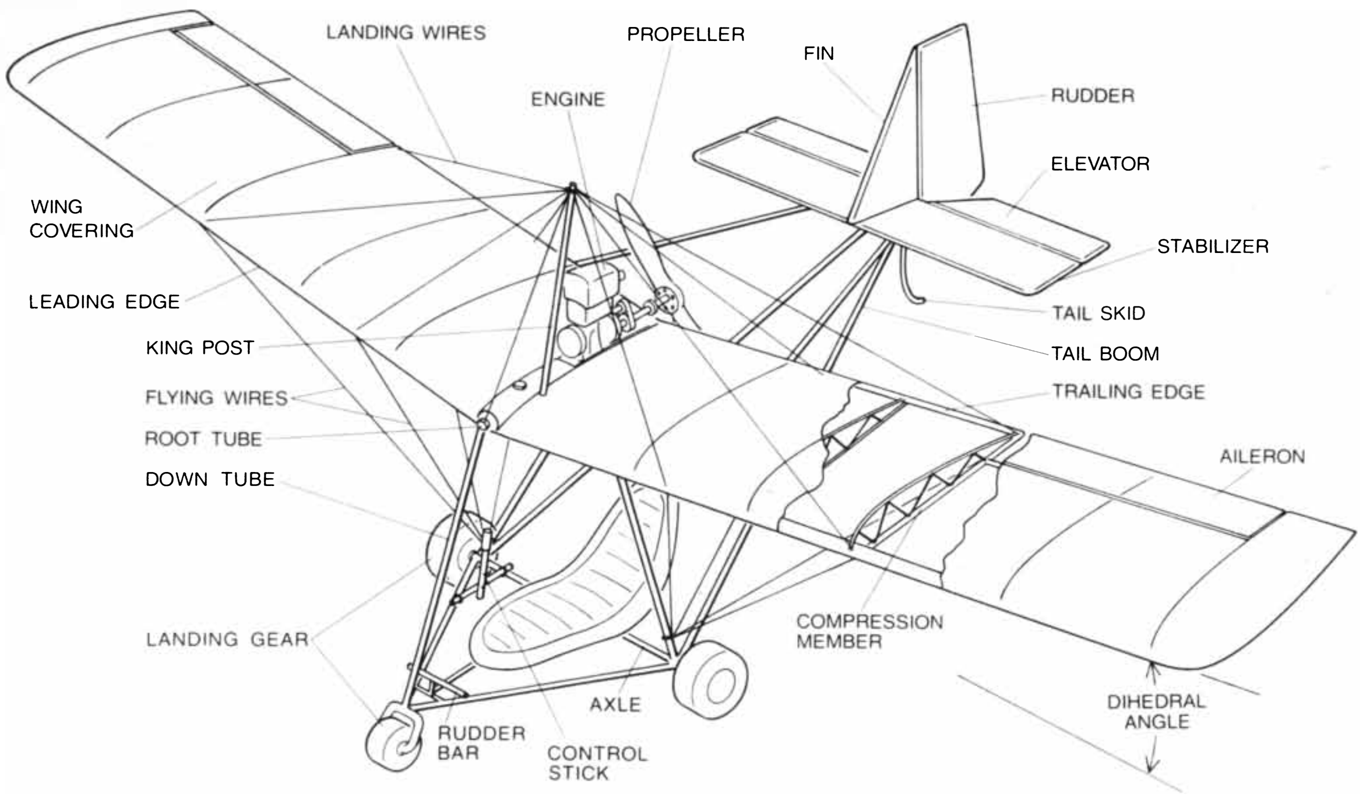
The first modern ultralight airplane was made in Wisconsin in the winter of 1974–75 by John Moody, an electrical engineer and a hang-glider pilot. He mounted a 12-horsepower Go-Kart engine on his Icarus II biplane hang glider. His aim was only to be able to climb to altitudes where he could fly the craft as a hang glider, soaring and gliding with the engine turned off until he needed to climb again. It was not long, however, before a good many of those who followed his lead began to realize that the addition of an engine to a hang glider created a new kind of craft: a small airplane that could be flown independently of the natural lift required for a hang glider alone.

Other hang-glider designs were soon adapted to power. One adventure of this type demonstrated that there is more to making a successful ultralight airplane than just bolting a small engine to a hang glider. It involved the Rogallo wing, a hang glider adapted from a triangular kite patented by Francis M. Rogallo and his wife in 1951. At first the engine was bolted to the king post on Rogallo wings being adapted for power. The high thrust line (the direction of propulsion) of that location proved to be disastrous at times when the force of gravity was not acting on the pilot. In this zero-*g* situation the high thrust line created a nose-down pitching moment the pilot could not counteract. (In normal flight, when the wing is under a positive *g* loading, a shift of the pilot's weight is effective as a control, but in zero *g* the pilot in effect has no weight.) With the nose down the sail began to luff, or flutter, and the glider was forced into either a dive that could not be stopped or an inverted loop that was likely to cause structural failure. Fortunately this dangerous flaw in design was corrected by a lower thrust line before the powered Rogallo wing was put on the market.

The most popular bolt-on engine installation for the Rogallo wing positioned a pusher propeller at the aft end of the keel, which is the center tube of the wing. The rig consisted of a Go-Kart engine mounted above the pilot's head and a long shaft running parallel to the keel. The same unit could be bolted to almost any flexible-wing hang glider. A glider thus fitted out was launched from level ground on foot, that is, by a pilot who held the rig over his head and ran into the wind. Since the thrust line was still higher than the center of gravity, flight in turbulent conditions was not recommended because the craft tended to become longitudinally unstable with the engine running. In calm conditions, however, the arrangement was workable. The rig was employed primarily to gain altitude for soaring rather than to cruise under power.

Until 1977 most ultralight airplanes received their thrust from a Go-Kart engine turning a propeller bolted directly to the engine shaft. The arrangement was simple and mechanically trouble-free, but the engine ran at such a high number of revolutions per minute that the propeller had to be quite small (about 28 inches in diameter) in order to keep the propeller tips from exceeding the speed of sound. As it was, the propeller turned at more than 9,000 r.p.m., quite close to sonic speed, with the result that the propeller noise was extremely loud and the propulsive efficiency was only about 50 percent. The performance of the early powered hang glider was therefore quite marginal. The cruising speed was only slightly higher than the stalling speed, and the rate of climb was a dangerously low 100 feet per minute. Furthermore, the engines themselves were short-lived because they had to run at such high speeds in order to deliver enough power. It seemed clear that the ultralight aircraft would not gain wide acceptance until the level of performance was improved.

An enterprising experimenter, Charles Slusarczyk, approached the thrust problem in a scientific way. Aware that a propeller is most efficient when its tips are moving at a rate well below sonic speed, he devised a reduction-drive system for powered hang gliders; it was patented last year. His idea was to move a large volume of air as slowly as possible through the disk formed by



COMPONENT PARTS of a generalized ultralight airplane are identified. The craft represents the most recent stage of design in that it incorporates an independent three-axis control system: pitch (the up and down movement of the nose) is regulated by moving the control stick back and forth, roll by moving the stick from side to side (making one aileron go up and the other down) and yaw by operating the rudder bar with the feet. An ultralight airplane with this control system is flown in much the same way as a heavier airplane.

the turning propeller while minimizing the drag from compressibility that absorbs a great deal of power near sonic velocity. The result is a vast improvement in thrust and efficiency and a substantial reduction in propeller noise. Nearly all contemporary ultralight aircraft incorporate a reduction-drive propulsion system that gets superior performance out of a small engine. A typical figure for today's reduction-drive transmission is 10 pounds of thrust per horsepower.

Before long the Go-Kart engine was largely replaced by the snowmobile engine, which has a larger displacement and runs at a lower r.p.m. The tuning of the engine is modified from its snowmobile specifications, thereby enhancing its thrust, improving its reliability and increasing its expected service lifetime. Typically the carburetion is reduced or the compression is lowered; sometimes both are done.

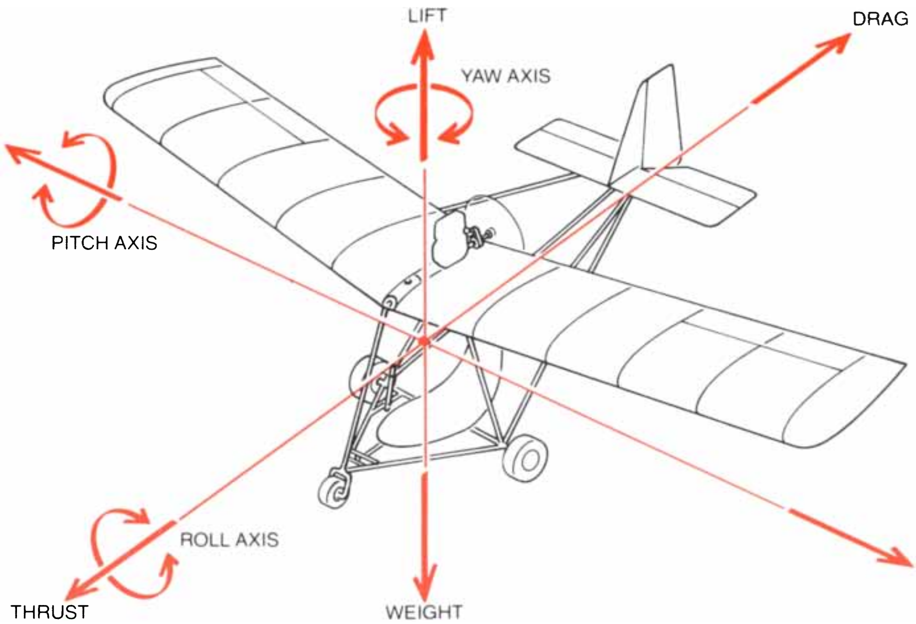
A change resulting from the switch to snowmobile engines was the wheeled landing gear. A hang glider with a direct-drive Go-Kart engine had an empty weight of about 100 pounds and could be launched on foot. The snowmobile engine, the associated reduction-drive components and a strengthening of the frame brought the empty weight to something over 160 pounds and made launching on foot dangerous.

The incorporation of landing gear made for a complete ultralight airplane. Once landing gear was accepted it was possible for engineers to design an ultralight craft around the propulsion system instead of simply bolting an engine to a hang glider. The result was a new generation of designs, most of which moved the new aviation away from its origins in the powered hang glider and toward the "little airplane."

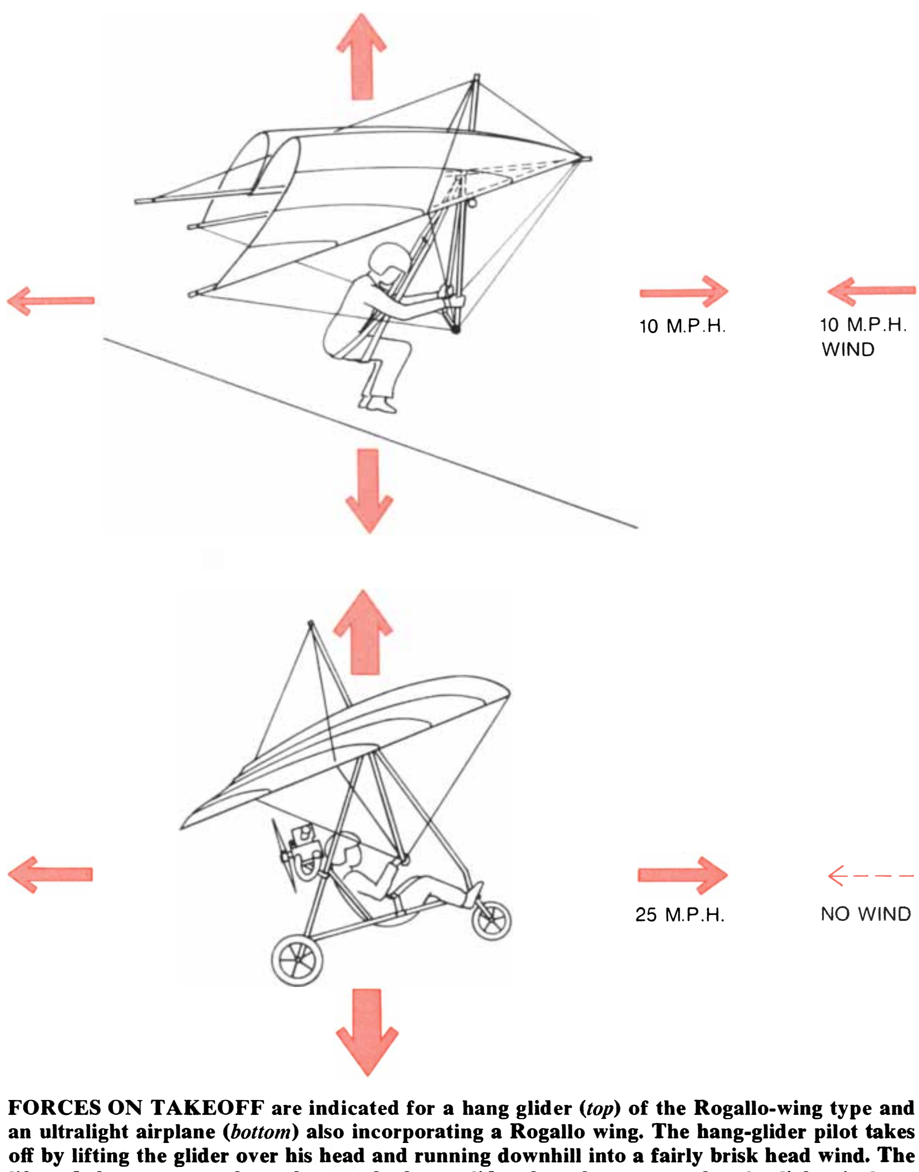
The ultralight aircraft of today can be grouped in four categories, which reflect the basic construction embodied in each design: Rogallo, cable-braced, strut-braced and cantilevered. The order of the grouping also approximates the relative level of performance of each type, particularly with respect to the top speed.

The Rogallo types include not only the foot-launched versions with a bolted-on engine but also the craft known as a trike. A trike (from tricycle) is a pyramidal tubular frame that holds the engine, the pilot's seat and the landing gear. The entire unit is bolted onto a Rogallo-wing hang glider. As a result of this arrangement the owner has in effect two aircraft: a hang glider and an ultralight airplane. In either form the method of flying is the same: the pilot pushes and pulls on a control bar for pitch and shifts his weight to one side or the other to make the craft bank.

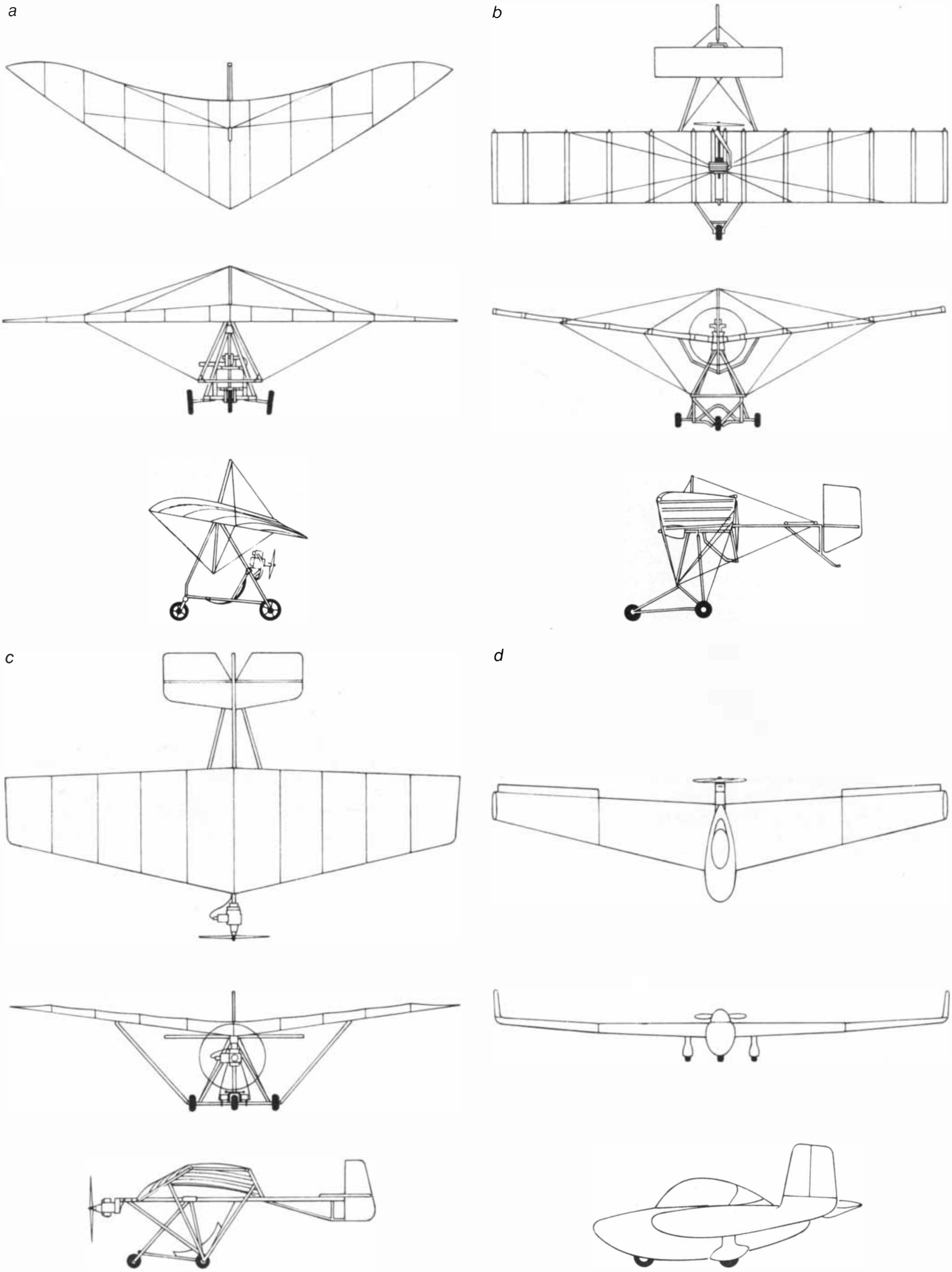
One current ultralight airplane, the Eagle, incorporates a hybrid Rogallo wing plus a canard. The main wing is a



FOUR FORCES that act on an aircraft in flight are lift, weight, thrust and drag. Lift is provided by the pressure difference of the air around the wings developed by the flow of air over them. Weight is a reflection of gravity. Thrust is provided by the propeller. Drag results from a variety of forces that tend to hold the airplane back. Pitch, roll and yaw axes are also shown.



FORCES ON TAKEOFF are indicated for a hang glider (*top*) of the Rogallo-wing type and an ultralight airplane (*bottom*) also incorporating a Rogallo wing. The hang-glider pilot takes off by lifting the glider over his head and running downhill into a fairly brisk head wind. The lift and thrust are much weaker on the hang glider than they are on the ultralight airplane.



FOUR BASIC TYPES of ultralight airplane are the Rogallo (*a*), represented by the Jet Wing; the cable-braced (*b*), represented by the Quicksilver, which is also the airplane shown on the cover of this issue; the strut-braced (*c*), here the Weedhopper, and the cantilevered (*d*), which has a wing on each side of the fuselage supported only by its attachment to the fuselage. The cantilevered craft shown here is a Mitchell U-2. The level of performance of the planes, particularly their top speed, is approximately reflected by the grouping.

Rogallo wing that has been modified by the insertion of ribs into the sail and the addition of drooped "tip draggers," or rudders, at the wing tips. The canard is an inflexible wing with an elevator (a movable flap) that serves to control pitch. Pitch control is augmented by the ability of the pilot to shift his body fore and aft.

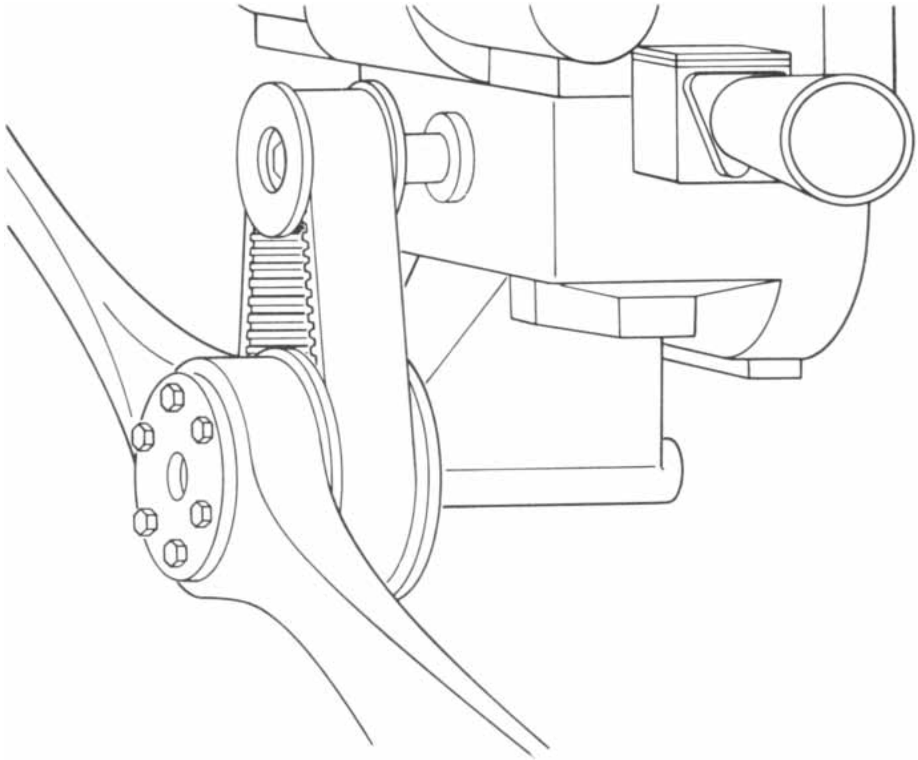
Most contemporary ultralight airplanes are in the cable-braced category. A typical arrangement consists of a pyramidal tubular frame surrounding the pilot and a tricycle landing gear attached to the frame. A ladder-frame wing is connected by a pin to the top of the pyramid and braced by cable to the bottom of the pyramid and to a king post above it. Another tube arrangement runs aft of the wing and holds the tail, which is also braced by cable to the pyramid, the king post and the wing.

The engine is usually mounted in the center section of the wing. A reduction-drive transmission conveys the power to a pusher propeller. Envelopes of presewn Dacron slip on over the framework of the wing to create the surface that provides lift. On the upper surface of the wing preformed aluminum ribs are inserted into pockets to create an airfoil of curved shape, which enhances lift. The lower surface is flat or slightly cambered to compensate for the fact that in flight the air tends to push the surface upward. On some designs only the upper side of the wing has a Dacron "skin." Such a craft is slower than one with a double-surface wing.

The strut-braced construction is gaining in popularity as designers move to improve the aerodynamics of the ultralight airplane. The group includes a fairly wide variety of wing structures. In the simplest designs a basic ladder-frame wing is covered with a presewn Dacron envelope and ribs are slipped into the upper surface. The most advanced designs incorporate a main spar, made of aluminum tubing with a *D*-shaped cross section, that also serves as the leading edge of the wing. The ribs are made of either aluminum or a composite consisting of a foam core and a strong shell of aluminum or fiberglass. Fuselages range from simple tubular pyramids to partly enclosed structures made of foam-and-fiberglass components.

The planes of the fourth group are made in much the same way but have separately cantilevered wings, that is, a wing on each side of the fuselage supported only by its attachment to the fuselage. The cantilever design results in the most streamlined of the ultralight craft and also in the closest resemblance to a conventional light airplane. Another of the virtues of a cantilevered ultralight is that it takes the least time to assemble once it is at its takeoff site.

The controls of ultralight airplanes have also evolved significantly in recent years. When such a craft was made by simply bolting an engine onto a hang glider, control often consisted of nothing more than shifts of the pilot's weight. A few craft had in addition some rudimentary aerodynamic controls, that is, surfaces the pilot could move to affect pitch, yaw and roll. Hang-glider pilots found these familiar arrangements adequate, but to people trained to fly standard airplanes they were strange and confusing. The designers of ultralight aircraft soon recognized the problem and began developing suitable aerodynamic controls. Today it is unusual to find an ultralight airplane that depends in any way on the shifting of weight by the pilot for control.



REDUCTION DRIVE, patented last year by Charles Slusarczyk, greatly improved the thrust and efficiency of propellers on ultralight airplanes. A toothed belt transmits power from a small pulley on the drive shaft to a larger pulley on the propeller shaft. A typical reduction-drive transmission of this type provides 10 pounds of thrust per horsepower. With direct drive the propeller tips ran at almost the speed of sound, which was inefficient and extremely noisy.

The first modern ultralight airplane, the powered Icarus II hang glider, had a hybrid control system because that was how the glider was controlled before Moody thought of adding an engine. The pilot, who was suspended prone in a harness, controlled the pitch and therefore the speed by moving forward or backward. He could also move the tip draggers to make the craft yaw and thereby induce a roll. The wings on the inside of a turn lost speed and therefore lift, whereas the outer wings gained speed and lift, generating a net rolling moment and a turn. The simultaneous deflection of both tip draggers increased the aircraft's drag and so served as a means of controlling the glide path.

Another hybrid system was employed in the Quicksilver, which had appeared originally as a hang glider in the early 1970's. The pilot was suspended in a swing and could therefore shift his weight fore and aft and from side to side. To increase the turning capability of the aircraft lines were connected from the harness to the rudder, so that a sideward shift by the pilot deflected the rudder, causing a yaw and thus a roll. The system was workable, but it was still a weight-shift scheme and licensed pilots did not accept it.

The first major advance in control systems appeared in the first true ultralight aircraft (as distinguished from the powered hang gliders). The new idea was a two-axis control system. The pilot is held in place by a seat belt and cannot effectively shift his weight. Instead he grips a control stick that is connected to the rudder and its associated elevator. The wings have no movable surfaces. The control stick works partly in the conventional way: forward movement pitches the nose down, backward movement pulls it up. Sideward movement of the stick, however, moves the rudder in the same direction, generating yaw and hence roll. (In a conventional stick-controlled airplane a sideward movement of the stick actuates ailerons on the wings to achieve roll.)

In the two-axis system a deflection of the rudder causes the aircraft to yaw and skid. The velocity of the outside wing increases, generating a rolling moment and its attendant bank. The wings must therefore have sufficient dihedral (the angle at which they meet) to stop the skid as it develops. In a properly designed aircraft the resulting motion is a fairly well-coordinated turn. The major drawback of the two-axis system is that the craft becomes difficult to handle in a crosswind of more than 5 m.p.h. because the wing cannot be raised quickly enough to forestall a skid.

The most recent models incorporate the independent three-axis control system, which was the key to the success of the Wright brothers. One version of it, which is essentially the same as the conventional arrangement, encompasses control of the elevator and the ailerons with the stick and of the rudder with foot pedals. A modified version employs spoilers in place of ailerons. A spoiler is on the upper surface of each wing; deflecting one of the spoilers disrupts the airflow above the corresponding wing, reducing the lift and causing the wing to drop. The deflected spoiler also increases the drag, enhancing yaw and inducing additional lift on the outer wing because of its increased speed. By deflecting both spoilers simultaneously the pilot can control the glide path.

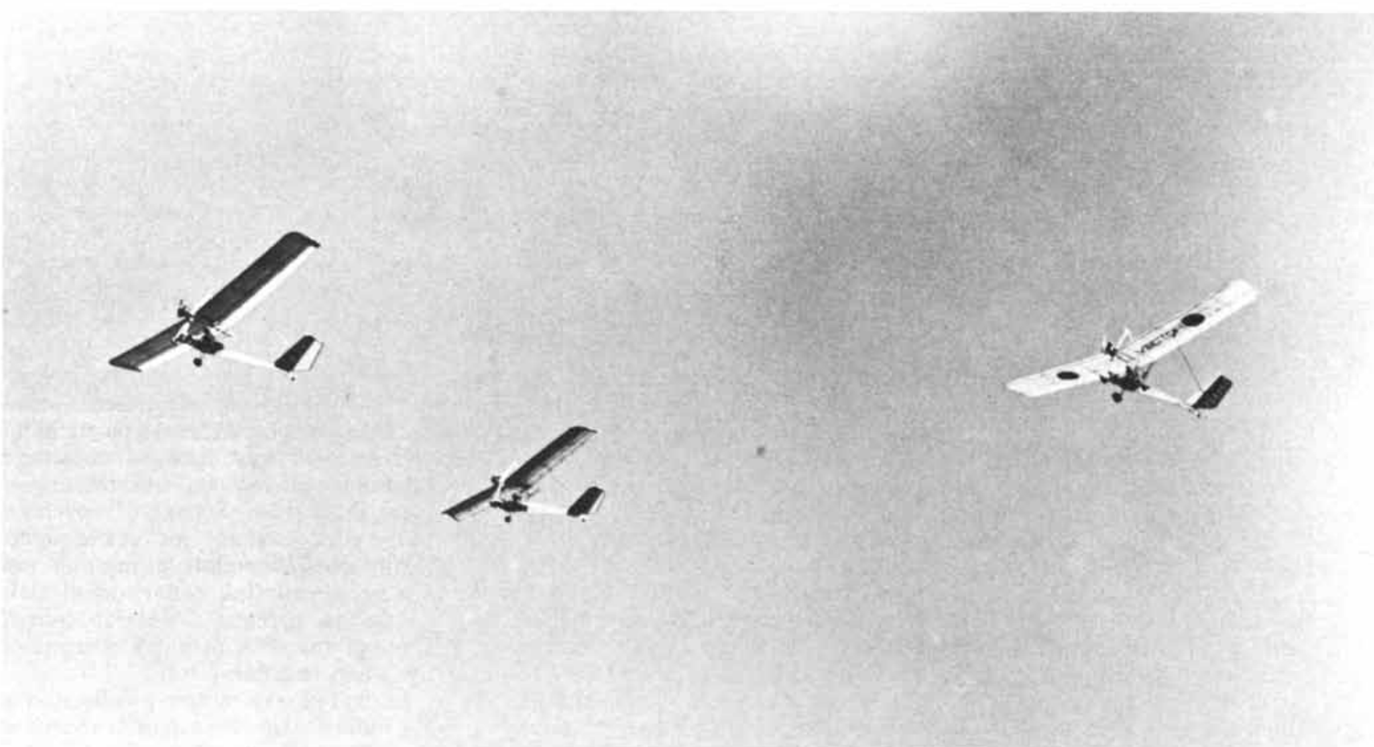
Anyone who is of a mind to try flying an ultralight airplane should have a grasp of basic aerodynamics. The four forces that act on the craft are lift, weight, thrust and drag. Lift is provided by the pressure of air under and the flow of air over the wings. Weight (of the airplane, the pilot and the fuel) is an effect of gravity. Thrust is the force supplied by the propeller. Drag is the combination of forces tending to hold the airplane back. Induced drag is a by-product of lift, and it increases with increasing angle of attack (the angle between the chord of the wing and the relative wind, or more loosely between the wing and the horizontal). Parasitic drag is generated by the parts of the aircraft that do not contribute to the lift, profile drag by the airfoil (the cross-sectional shape of the wing).

The pilot's major concern is the angle of attack. When it is high, the speed of flight is low, and when it is low, the speed of flight is high. As the angle of attack is increased, a point is reached at which the wing will stall, losing its lift and increasing its drag because the airflow becomes quite turbulent and separates from the upper surface of the wing. At a point near stall the pilot's ability to control the craft is diminished because the ailerons become virtually useless. Only a decrease in the angle of attack will enable the wing to regain lift, which is to say that the pilot must push the control stick forward so that the nose goes down and the airplane gains airspeed by losing altitude. A stall can develop into a spin, particularly during a turn. If one wing stalls first, it drops, and the airplane begins to rotate about a vertical axis as it falls. It is not always possible to recover from a spin, which is why it is important for the pilot (particularly for a novice) to know the conditions that can bring on a stall, to avoid them if possible and to apply the proper corrective steps quickly if a stall develops or seems to be imminent.

Although ultralight airplanes are basically easier to handle than heavier aircraft and no license is required to fly most of them, the FAA does require anyone who flies to be familiar with the Federal Air Regulations. All pilots of ultralight aircraft must pay particular attention to the applicable visual-flight rules. Flight near controlled airports should be avoided.

It would be foolhardy for anyone to try to fly an ultralight airplane without first acquiring a sound knowledge of aerodynamics and taking some formal flight training. An ultralight is a real airplane, not a toy, and a good deal of skill and understanding is needed to fly one safely. Many flight schools require their students to first take dual instruction in a conventional airplane and to be able to make a solo flight in such a craft. Since an ultralight is a single-person craft, the dual-instruction approach is certainly wise for anyone who is not already a pilot. You must know the basics of flight before you venture aloft alone.

I mentioned at the outset that the pilot of an ultralight airplane must be acutely sensitive to the vagaries of the wind. The necessity is reflected in the fact that licensed pilots, who make up about half of the people flying ultralight craft, often have more trouble than nonpilots in adapting to the conditions. Being accustomed to heavier aircraft, they seem not to respect the wind as much as they should and are often taken by surprise when an ultralight airplane displays its extreme sensitivity to gusts. The prudent pilot flies an ultralight airplane only when the weather is fair and the wind is gentle and steady.



FLYING CIRCUS at a recent show of ultralight airplanes in Lakeland, Fla., consists of three Vector 610 ultralight aircraft. The 610 is a cable-braced airplane with a pusher propeller and an upright *V* tail. The pilot is seated in a cage under the craft's 34-foot wing.

Important stuff.

What this elixir makes possible would have gotten people in serious trouble for witchcraft in times gone by.

Actually it is only a polymer which has precipitated in hexane. You dry the precipitate under vacuum. You dissolve it in something like 2-ethoxyethyl acetate. You spin-cast the solution onto silicon to leave a coating a wee bit less than a micrometre thick. You bake at 90°C for 30 minutes. You draw a pattern on the baked coating. (Not just any pattern.)

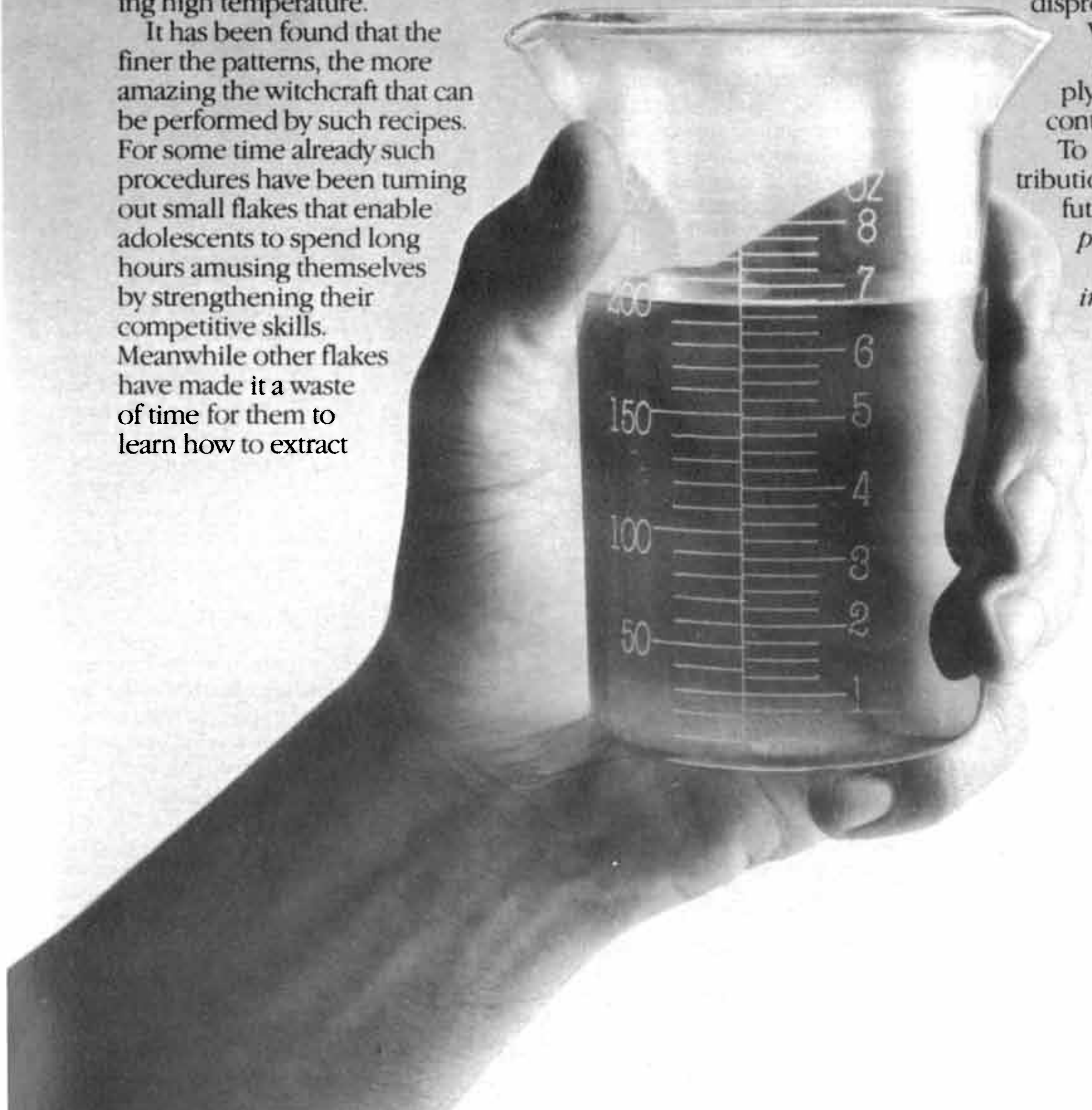
The pattern has lines and spaces as narrow as half a micrometre. That is only about the wavelength of visible light. If light is used to write the pattern, the pattern gets smeared out somewhat by the very wave nature of light. A 10 keV electron beam has a shorter effective wavelength. With it you draw one of the layers of a large-scale integrated circuit (LSI). A dip in common solvent then washes away the coating where electrons did not hit. This exposes the underlying substrate so that other substances can be deposited on it by some method involving high temperature.

It has been found that the finer the patterns, the more amazing the witchcraft that can be performed by such recipes. For some time already such procedures have been turning out small flakes that enable adolescents to spend long hours amusing themselves by strengthening their competitive skills. Meanwhile other flakes have made it a waste of time for them to learn how to extract square roots. Still others permit inanimate devices to explain to a distant device that is also inanimate what ails them, and get cured. At a still higher level, other little chips work together to enable the animate mind to relate spots on a sheet of film to sequences of nucleotides that are presumed to account ultimately for the existence of the animate mind itself.

What more can be expected as the patterns on the chips get down to submicron scale may not be as obvious to all as it is to some people. Laying down submicron patterns with x-rays through masks that are themselves prepared with electron beams is simpler and cheaper. The composition of the polymer determines whether the elixir (more prosaically termed a photolithographic resist) is for electrons or x-rays. It has been thought that high sensitivity to x-ray and electron-beam exposure is incompatible with high resistance to the hot plasma that treats the uncovered substrate. With those particular compositions we have disproved that.

We like to think of formulations that we are now supplying to Silicon Valley as a contribution to civilization.

To learn more about our contribution and commitment to the future of microelectronics, *please send for our brochure (B3-190) "...Resources to innovate, commitment to serve...."* Write Eastman Kodak Company, Dept. SA-L-GB, 343 State Street, Rochester, NY 14650.



Kodak

SCIENCE AND THE CITIZEN

Public Enterprise

A number of companies engaged in exploiting new biological technologies are eager to invest in basic research done in universities. Discussions of the propriety of such arrangements tend to be overshadowed by a pragmatic concern: Even if the influence of industrial funding is not wholly salutary, it is argued, universities cannot afford to be too fastidious at a time when governmental support for fundamental research is dwindling. A contrary but equally pragmatic view has lately been expressed, namely that even if industrial participation is not altogether harmful, it cannot possibly replace public financial resources.

At the meeting of the American Physical Society in January, George Pake of the Xerox Corporation reviewed the funding of research and development in the U.S. in all disciplines. In 1980 some $8.1 billion was spent for basic research, 60 percent of it in the universities. Of the university share the Federal Government supplied 65.5 percent and industry only 3.7 percent. Thus industrial contributions would have to be multiplied several times before commercial interests had a significant stake in the basic research done in universities.

At a meeting in March sponsored by Beckman Instruments, Arthur Kornberg of Stanford University pointed out that the biological processes now finding commercial applications are the product of three decades of fundamental research paid for almost entirely by Federal agencies (chiefly the National Institutes of Health and the National Science Foundation). Although the recent flowering of industrial biochemistry has proved stimulating, private and corporate philanthropy cannot supply a major part of the continuing investment needed. "We're talking about billions of dollars a year," Kornberg said. "There just aren't such resources available nor are there equitable mechanisms for obtaining and distributing such sums except through Federal taxation and administration.... There is a widespread feeling that the biotechnology companies, having obtained their science and scientists from the universities, should share their fortunes with their academic parents and benefactors.... To put it bluntly, there aren't that many rich biotechnology companies; few of the successful enterprises are able or inclined to give their money away, and if such a rare [company] exists, it will surely not pour millions of dollars into an abstruse project with a 20-year gestation period."

Kornberg suggested that collaboration between industry and the universities would be most effective if it took a form other than direct support. Since both parties have an interest in the well-being of basic research, they should "mobilize to preserve and enlarge the Federal programs that support it."

Funny Valentine

At 1:53 P.M. on February 14—Valentine's Day—an event without precedent took place in an unoccupied laboratory of the physics department at Stanford University. The event, which was not entirely unexpected, triggered a carefully laid trap: its passing was signaled by an abrupt increase in a persistent electric current that had been set in motion months earlier in a small coil of superconducting niobium wire. The coil was shielded from outside electromagnetic fields by a thin cylinder of superconducting lead, and the entire assembly was immersed in a pool of liquid helium maintained at a temperature of 4.2 degrees Kelvin, below the temperature at which both lead and niobium lose all resistance to an electric current.

In the guarded words of Blas Cabrera, the physicist who prepared the trap, the recorded event "is consistent with the passage of a single Dirac charge." In somewhat more familiar terms, the signal may have been caused by the fleeting presence of a magnetic monopole, the long-sought particle of isolated magnetic charge whose existence was predicted more than 50 years ago by the British theoretical physicist P. A. M. Dirac.

In a brief account of the Valentine's Day event published in *Physical Review Letters* Cabrera gives some of the reasoning behind his tentative identification of the observed current jump as evidence for the detection of a free cosmic-ray monopole. Recent efforts to construct a grand unified theory of the forces of nature suggest that superheavy magnetic monopoles would have been created in the first 10^{-35} second of the big bang that presumably gave birth to the universe. Monopoles of this type would be much too massive and slow-moving to be recorded in conventional particle detectors that rely on the ionization trail left behind by a charged particle moving at a speed close to the speed of light. That would explain why no such monopole has ever been detected before, in spite of numerous attempts to find one [see "Superheavy Magnetic Monopoles," by Richard A. Carrigan, Jr., and W. Peter Trower; SCIENTIFIC AMERICAN, April].

The extraordinary method resorted to by Cabrera to search for a superheavy monopole is based solely on the long-range electromagnetic interaction of a moving magnetic charge with the macroscopic quantum state represented by the superconducting ring. According to Cabrera, his detector, which incorporates a current-monitoring system called a SQUID (for superconducting quantum interference device), is insensitive to the speed, direction, mass and electric charge of a particle passing through it. The detector also would show no response to a magnetic dipole, made up of both a north and a south pole. The passage of a single magnetic charge through the ring, however, would result in a sharp, well-defined increase in the circulating current.

Cabrera reports that his detector has been in operation more or less continuously for a total of more than 200 days. In that period only one event, the one of February 14, was large enough to have been caused by the passage of a magnetic monopole. This result sets an upper limit on the number of magnetically charged particles moving through the earth's surface. The implied limit in turn is remarkably close to an estimate based on the assumption that superheavy magnetic monopoles account for all the "missing mass" in our galaxy. The latter assumption suggests that an average of 1.5 events per year should be observed in a detector the size of Cabrera's ring. The first order of business, therefore, is to continue the search with the present detector, in the hope of finding another monopole in the next few months. In the meantime, Cabrera notes, "two new systems of larger sensing area are being built."

The implications of Cabrera's discovery, if it is borne out by further observations, are far-reaching. In the view of W. Peter Trower, a veteran monopole hunter at the Virginia Polytechnic Institute and State University, "Cabrera's experiment is an elegant one, and many alternative explanations of his solitary event have been ruled out. Although it is rational at this stage to reserve judgment, we are all very excited."

Ursprache of the Cell

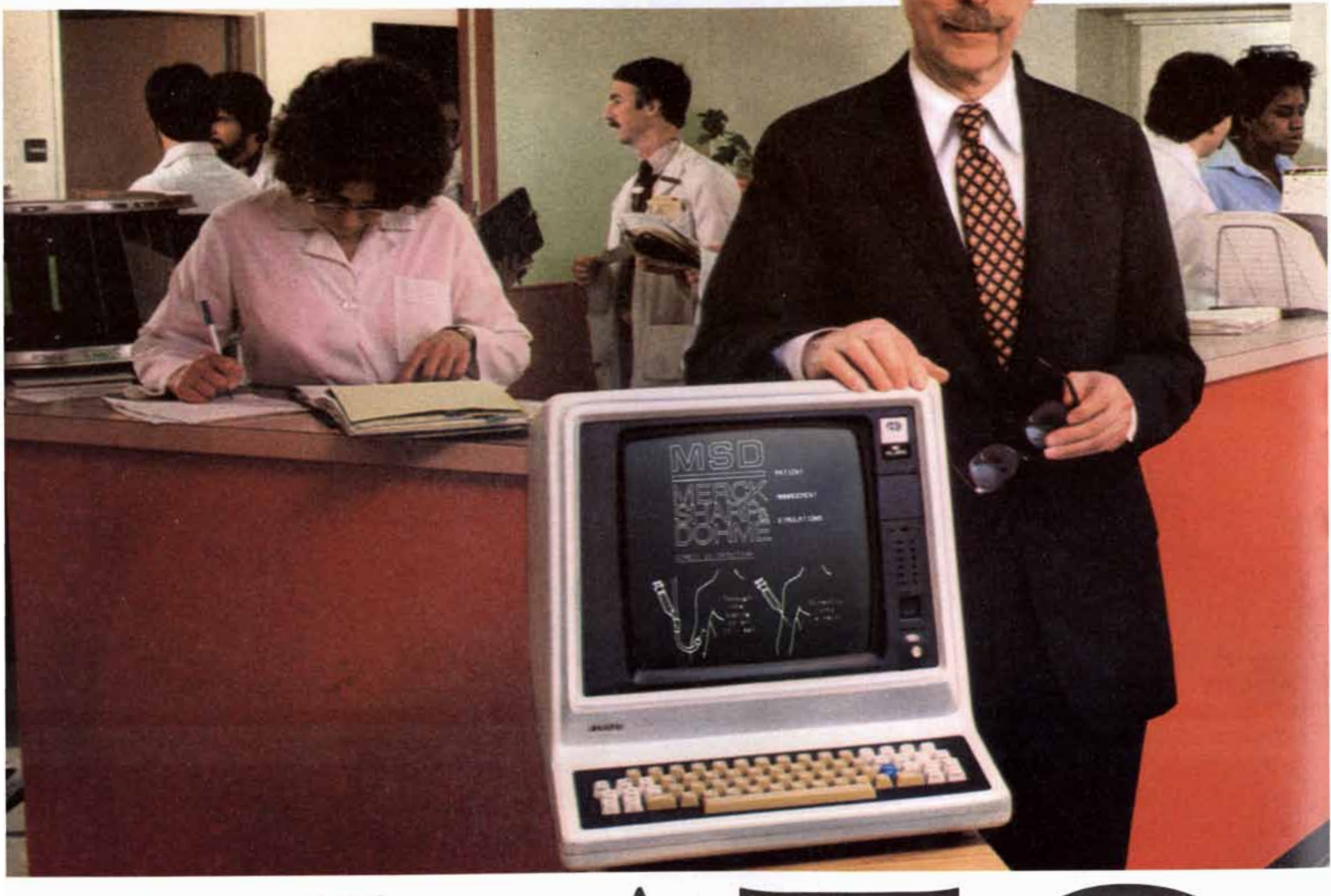
There was a time, not so long ago, when hormones were hormones and neurotransmitters were neurotransmitters. A hormone was a substance secreted by a particular kind of cell in an endocrine gland and released into the bloodstream to affect specific receptor cells elsewhere in the body of a vertebrate animal. A neurotransmitter was a chemical messenger carrying signals from a nerve cell to other nerve cells or to muscle cells. Whereas neurotransmitters were known to be present even in simple multicellular animals, hormones were thought to have arisen only with the evolution of the vertebrates.

The lines have become blurred. It has

Solve your field training problems the way Merck Sharp & Dohme does.

"It's vital that pharmaceutical representatives be fully qualified to relate complex product information to the medical community. At Merck Sharp & Dohme, PLATO computer-based training adds depth, motivation and consistency to a program that has developed one of the largest and most professional field organizations in the prescription drug industry."

Elliot J. Margolis, MD
Executive Director, Field Administration
Merck Sharp & Dohme



PLATO

CHANGING HOW THE WORLD LEARNS

Merck Sharp & Dohme has expanded its field organization more than 30 percent in the past five years. To accommodate this growth, America's leading prescription drug company has increased the quality and productivity of its training with the help of PLATO* computer-based training.

With PLATO, each learner is a class of one. At the touch of the screen or keyboard, the terminal turns into a tutor. PLATO explains, reviews, tests, grades and re-tests when necessary. It encourages and counsels. It makes complex technology easy to understand. So people at all levels learn more, better and faster than by traditional methods.

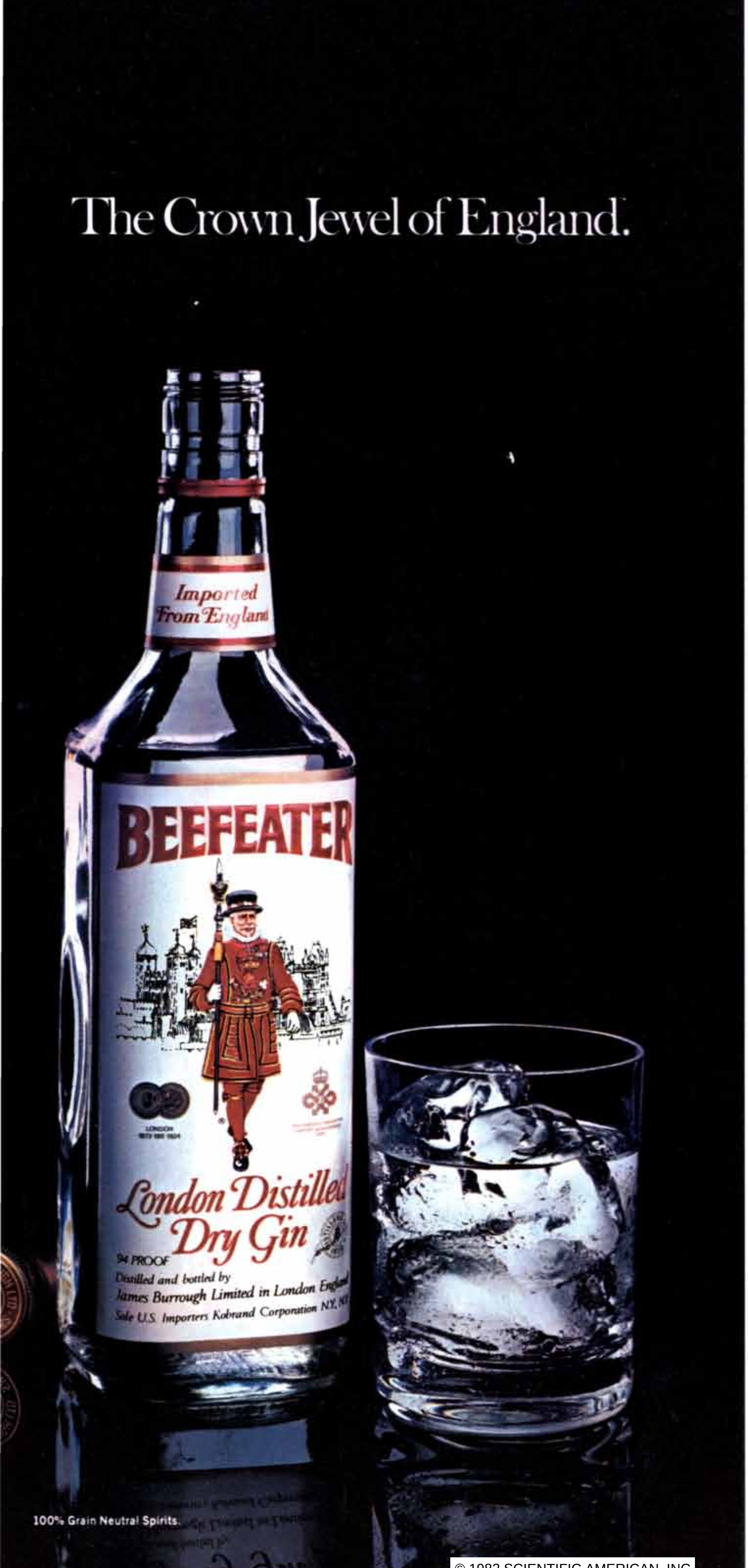
There is no better way to train, whether you have a PLATO system installed at your location or use any of over 100 Control Data Learning Centers. There are more than 7000 hours of standardized PLATO courses. Or we can help you develop a course for your unique requirements.

Ask for an Executive Briefing. Learn how PLATO can improve the value of your company's training. *Call 800-328-1109 or write Control Data, Dept. HQV003-4, P.O. Box 0, Minneapolis, MN 55440.

*In Minnesota, call collect 612-853-7518.

CONTROL DATA

Addressing society's major unmet needs as profitable business opportunities



been found that hormones are synthesized not only in endocrine glands but also by cells in other tissues, including many cancer cells. Some hormones are made by nerve cells and serve as neurotransmitters; some neurotransmitters are released by glands and serve as hormones. It has developed that both hormones and neurotransmitters also serve as "tissue factors" mediating local interactions among cells. Some vertebrate hormones have been found in insects and even in sponges. And now both hormones and neurotransmitters have been discovered in unicellular organisms, and certain hormones have been found even in bacteria.

A new theory is taking shape that relates the hormones, neurotransmitters and other intercellular messengers of vertebrate animals to one another and traces their ancestry to primitive messenger molecules in unicellular organisms. Both the theory and much of the recent data bearing on it come from the laboratory of Jesse Roth at the National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases. The data, the theory and its ability to explain what is known about intercellular communication are summarized in *The New England Journal of Medicine* by Roth and his colleagues Derek LeRoith, Joseph Shiloach, James L. Rosenzweig, Maxine A. Lesniak and Jana Havrankova.

Roth and his associates have concentrated on a search for peptide hormones in unicellular organisms. A peptide is a fairly short chain of amino acids (the constituents of proteins) and is synthesized directly from genetic information. The presence of a peptide hormone in an organism therefore means that the organism has inherited and is expressing the DNA encoding the hormone. (Other kinds of hormones are assembled from molecules present in many living cells, and their presence might conceivably be accidental.) The investigators grow organisms in a simple culture medium, treat them with agents calculated to extract a particular peptide and detect the peptide with a radioactively labeled antibody that binds to it. The suspected hormone is purified and its identity is confirmed by testing its biological activity: its ability to do what the hormone does in a vertebrate animal.

Roth's group found material indistinguishable from the peptide hormone insulin in a protozoan, in two species of fungi and in three strains of the bacterium *Escherichia coli*. In protozoans they also found somatostatin, a peptide made in the brain and pancreas of mammals that inhibits the synthesis of insulin and a number of other regulatory substances, and the hormone ACTH, or corticotropin. Other investigators had earlier reported the presence of several mammalian neurotransmitters in protozoans and of human chorionic gonadotropin, a peptide hormone, in bacteria.

The function of these various messenger molecules in unicellular organisms is not known, but there are indications some of them bind to receptor molecules of the same kind as those they bind to in vertebrates.

According to Roth and his colleagues, the data suggest that the biochemical agents of the vertebrate endocrine and nervous systems, as well as other intercellular messengers, originated in unicellular organisms. In the course of evolution the anatomical elements—the cells that make the messengers, the ducts and vessels that carry them and the cells that receive them—have differentiated, specialized and become far more complex. The biochemistry of the messenger agents, however, has largely been conserved.

Such an approach to the history of messenger systems helps to account for many recent observations about the vertebrate systems, the investigators maintain. It explains the overlap of the endocrine and the nervous systems, both of which are held to have evolved from a common ancestral system. It explains the overlap, both as to cell type and as to secretions, of the endocrine glands and the exocrine glands: those that discharge their products not into the bloodstream but directly onto a surface such as the lining of the stomach or of the intestine. It explains the presence in vertebrates of substances such as growth factors, prostaglandins and interferon that resemble the endocrine and brain peptides but are neither hormones nor neurotransmitters. In this view all messenger molecules were once local tissue factors acting on the cells that secreted them or on their neighbors, and it was a subset of those molecules that became hormones and neurotransmitters. The theory even casts a new light on pheromones, which are chemical messengers released by one animal that affect the behavior of another individual of the same species. Now pheromones begin to seem more like hormones. Indeed, when the only organisms were unicellular ones, all intercellular messengers were by definition pheromones.

Primal Screen

Mathematicians who scout the number system for primes have gained a few distant promontories, but the exploration of the vast territory between them has turned out to be a difficult undertaking. The trouble is logistic rather than conceptual: given unlimited time the primality of any number could be checked by trial division. (If a number is prime, only 1 and the prime number itself divide it without remainder; if a number has other divisors, it is composite.) For large numbers, however, the constraints of time are decisive. Except for numbers having a special form, no method was known until recently for testing the primality of a given 100-digit number in less than 100 years, even with the aid of a high-speed computer.

In 1980 a new test for primality was devised by Leonard M. Adleman of the Massachusetts Institute of Technology and Robert S. Rumely of the University of Georgia. Carl Pomerance, also of Georgia, and Andrew Odlyzko of Bell Laboratories then showed that the time required to carry out the test should theoretically grow quite slowly as the numbers being tested become larger. Soon thereafter a modified version of the Adleman-Rumely test was developed into a computer program by Hendrik W. Lenstra of the University of Amsterdam and Henri Cohen of the University of Bordeaux. The new test has now proved its mettle: the program, running on a large computer, has demonstrated the primality of an arbitrary 97-digit number in 78 seconds of computer time.

Unlike the most straightforward methods of testing primality, such as trial division, the Adleman-Rumely test does not factor a number in order to determine whether it is prime or composite. Instead the test is based on a generalization of a theorem first proved by the French mathematician Pierre de Fermat in the 17th century. Fermat showed that if an integer p is prime and if n is some number between 1 and $p-1$, then $n^{p-1}-1$ is divisible by p. For example, suppose p is the prime number 7 and n is 2; then $2^{7-1}-1$ is equal to 63, which is divisible by 7.

According to Pomerance, Chinese mathematicians of the fifth century B.C., who knew of the theorem for the special case when n is equal to 2, believed its converse as well. The converse states that if $2^{p-1}-1$ is divisible by p, then p is a prime number. The converse proposition is not true in all instances, however. For example, the number 341 passes the test (that is, 341 evenly divides $2^{340}-1$, a fact that can be demonstrated without calculating the value of $2^{340}-1$), but 341 is a composite number; it is the product of 11 and 31. A composite number such as 341 that satisfies the converse of Fermat's theorem is called a pseudoprime to the base n.

Most numbers that satisfy the converse of Fermat's test to any given base turn out to be genuine primes and not pseudoprimes. Adleman and Rumely reasoned, therefore, that if such a test were carried out for several bases, the pseudoprimes might be weeded out. Although the strategy is basically sound, Adleman and Rumely were well aware that it would not work without being modified: the composite number 561, for example, is a pseudoprime to every base from 1 to 560. They proposed instead a method whereby a sequence of tests analogous to the pseudoprime test is carried out, providing additional information about the number p. The extra information converges to fix the status of the number p so that its primality can be efficiently determined.

The most immediate practical significance of the new primality test may be for cryptography. The security of many coded messages is now dependent on the impracticality of factoring a composite number that is the product of two large primes. The primality test may be helpful in ensuring that the large factors of the composite number are indeed primes. On the other hand, Pomerance remarks, the problem of factoring is closely related to the problem of testing a number for primality, although no one currently sees any link between the Adleman-Rumely test and the factoring problem. Nevertheless, the long-term security of encryption systems based on factoring may now be open to question. Whereas the product of two 100-digit primes was once thought sufficiently secure, work is now under way at the Sandia Laboratories in Albuquerque, N.M., to build encoding and decoding devices capable of manipulating primes having several hundred digits.

Mesozoic Mishap

The hypothesis that a catastrophe at the end of the Cretaceous period, some 63 million years ago, was responsible for the extinction of many plants and animals gained support in 1979 from the discovery of an unusual mineral enrichment in sediments at the boundary layer between the Cretaceous and the Tertiary periods (a boundary that also separates the Mesozoic era from the Cenozoic era). The most distinctive feature of the layer is an overabundance of the element iridium, which has been attributed to the introduction of extraterrestrial material by the impact of a comet or a large meteorite. Evidence of the anomaly has been found in Italy, Spain, Denmark and New Zealand and in cores extracted from the floor of the Pacific and the Atlantic oceans. Now detailed analyses of several sea-floor cores from the South Atlantic have given further support to the catastrophe hypothesis and have suggested how the impact of a comet or a meteorite may have led to the extinctions.

The cores were obtained in May, 1980, as part of the Deep Sea Drilling Project. They come from an area called Site 524, adjacent to the submarine mountain range called the Walvis Ridge near the prime meridian at 30 degrees south latitude. The examination of the cores, which is still in progress, is being done by 20 individuals representing 13 universities and other institutions. They have presented a preliminary report in *Science.*

Unlike other sea-floor cores, which are poorly defined at the Cretaceous-Tertiary boundary, the cores from Site 524 include more than 30 centimeters of alternating layers of sediments, with a three-centimeter layer of clay defining the boundary itself. The clay is unusually enriched in iridium.

Another telltale substance is calcium carbonate, which makes up about 40 percent by weight of the sediments above and below the boundary layer but diminishes to only 2 percent within the layer. This can be attributed to the mass disappearance of plankton that normally shower the sea bottom with their calcium-rich exoskeletons.

An analysis of oxygen isotopes indicates that the surface waters in the South Atlantic were warmed by as much as 10 degrees Celsius at the end of the Cretaceous. The rise in temperature is attributed to the "greenhouse effect" caused by excess carbon dioxide. Ordinarily the carbon dioxide would have been taken up by photosynthetic marine plants; the extinction of many plants may have allowed the gas to pass instead into the atmosphere.

Addressing the catastrophe hypothesis directly, the investigators conclude that the oceanic extinctions can logically be attributed to the impact of a large extraterrestrial body. The consequent elevation of the earth's temperature, lasting for a period of perhaps 50,000 years, may have been responsible for the extinction of certain land animals, such as the dinosaurs.

Bits per Inch

The workhorse of computer memory systems for the bulk storage of information is magnetic recording, mostly on spinning disks bearing a thin-film magnetic medium. The disks have been facing increasing competition from semiconductor memories, which have no moving parts and offer faster recording and retrieval. As the price of semiconductor memories has declined and their capacity has increased, the obsolescence of the magnetic-disk memory has been predicted repeatedly. That obsolescence may now be postponed again by a new geometric mode of magnetic recording.

On a conventional disk the recording is done longitudinally, that is, the magnetic domains embodying the stored information are aligned parallel to the surface of the film and parallel to the direction of disk motion. The new technique is vertical recording, in which the magnetic domains are perpendicular to the film. If vertical recording can be perfected, it will lead to a large increase in the amount of information that can be stored in a given area of the recording medium.

In all forms of magnetic recording the surface coated with magnetic particles moves with respect to a recording head, which is an electromagnet. Initially all the particles can be oriented in one direction. The magnetization of a given particle moving past the head is reversed if the magnetic field generated by the head exceeds a certain threshold strength. The strength of the field in turn varies according to the strength of the signal reaching the head, and so the number of particles left in a reversed state constitutes a record of the information in the signal.

In longitudinal recording the head consists of two magnetic poles pressed close to the recording surface and arranged so that the field is parallel to the motion of the surface. The limits of the technique are approached when the density of recorded information is about 10,000 bits, or binary digits, per inch of recording track. At this density stray fields in the medium cause areas of demagnetization, which increase with density If the magnetic domains in the recording surface are oriented perpendicular to the surface, however, the areas of demagnetization shrink as the information density increases. The density that might be achieved ultimately is limited only by the width of the "wall" separating two oppositely magnetized domains. In principle the density could be greater than 445,000 bits per inch of track, with as many as 2,000 tracks per inch of width in the recording medium.

A number of companies in the U.S., Europe and Japan are working on high-density memory systems based on vertical recording. In most instances the recording medium is a cobalt-chromium film on a thin substrate The major difference from longitudinal recording is in the recording head, which has one pole on each side of the medium. In laboratory experiments densities as high as 100,000 bits per inch have been demonstrated. If the technique can be brought to the commercial stage, the initial densities will probably be from 20,000 to 40,000 bits per inch.

The Great Sickness

One of the unfortunate results of the conquest of the New World by Europeans was the decimation of the native population by disease. European settlers introduced into many parts of North and South America diseases that had been unknown there. The mortality rate in the ensuing epidemics was much higher than it would have been among Europeans exposed to the same pathogens. In early studies of the epidemics it was assumed that the death rate among the native peoples was exceptionally high because they lacked the capacity to make antibodies to these agents of disease. Recent work in immunology suggests that on the contrary the capacity to make antibodies to all antigens is latent in every human being. Some mechanism other than immunologic incompetence must therefore underlie the high mortality. In *Proceedings of the American Philosophical Society* Robert J. Wolfe, an anthropologist employed by the Alaska Department of Fish and Game, argues that the severity of an epidemic in a previously unexposed population is determined by the interaction of several social and biological factors; among the most important is the degree of social disorganization caused by the disease.

Wolfe has analyzed records of simultaneous epidemics of measles and influenza that devastated the native population of western Alaska in the summer of 1900. Three groups of Alaskans were affected: the Eskimos, the Aleuts and the Athapascans. They called the episode the Great Sickness.

Measles was unknown in the population before 1900, when it was apparently transmitted to Alaska from Siberia along trading routes. The measles epidemic began in June on St. Lawrence Island, between Siberia and Alaska, and then spread south and east. There had been earlier experience with influenza, but the influenza virus of the Great Sickness may have been a strain new to the area. It appears to have been carried by travelers on ships from Seattle or San Francisco. The influenza virus spread north from the Pribilof Islands, a common port for U.S. ships. The areas affected by the two epidemics intersected on the mainland in July.

The overall death rate in the three native groups was 22 percent in the seven communities for which reliable data are available. The death rate varied greatly, however, from one community to another. The lowest rate among the seven villages was 13 percent; the highest was 74 percent. Of the Caucasians in Alaska very few contracted either disease, even among those who lived in close contact with the indigenous population.

Wolfe compares conditions at St. Paul's Island in the Pribilofs, where the death rate was 13 percent, with those on the mainland. Almost all the 200 Aleuts on St. Paul's were employees of the Alaska Commercial Company, a U.S. company that traded in seal pelts. Measles appeared on St. Paul's in June, near the beginning of the annual seal drive. In late August, when most of the population had recovered, the influenza epidemic began. Both measles and influenza affected almost every native resident, but none of the Caucasian employees of the company became sick. Measles was the deadlier disease on St. Paul's: 20 died of it, compared with seven who died of influenza.

Wolfe notes that living conditions in the company town were comparatively good. Each Aleut family had a wood house with a stove and an outhouse. There was a hospital, and free medical care was provided. Perhaps more important, the social organization of the town was not disrupted by the epidemics. On September 9, when all but two Aleuts were bedridden, the company was delivering food and fuel to each household, and company physicians continued to treat the symptoms of the diseases.

Along the Yukon and Kuskokwim rivers the native population of Eskimos and Athapascans had maintained a culture based primarily on salmon fishing. In June, when the food stock was at its lowest point and the salmon run had begun, families dispersed to small fishing camps along the river.

Measles and influenza struck the fishing camps almost simultaneously. Conditions were made worse by a cold rain that fell during the acute stage of the epidemic and by the low reserves of food. In some villages every inhabitant was sick. Joseph Romig, a physician who canoed down the Kuskokwim providing medical treatment, wrote: "The story is just about the same for every village. There are not enough well ones to care for the sick, and in many cases the sick are in serious need of care." Aleck Sipary, a Jesuit from the mission at Holy Cross on the Yukon, wrote that the residents of a nearby village "were entirely depressed by their awful condition and could not make a move to help themselves."

In the interior the overall death rate was about 25 percent. Even along the rivers, however, the death rate seems to have been sensitive to the presence of healthy outsiders. At a mission school near Holy Cross only one out of 20 boys died; at Holy Cross itself the death rate was 13 percent. In remote areas the toll was heavy. In Ugavig, a large settlement on the Kuskokwim, the death rate was 46 percent; in Dog Fish Village it was 74 percent; in some small camps every inhabitant died.

Tests of measles vaccines have shown that there are only slight differences among human populations in the capacity to make antibodies to the measles virus. What may have made measles so deadly in the Great Sickness was not the lack of an immune response but the age of those affected. In populations where the disease is common immunity is generally acquired in childhood; many of the deaths in 1900 appear to have been those of older people. An effect of genetics, however, cannot be ruled out entirely; some workers have suggested that in historically isolated populations the capacity to make antibodies may be limited. Nevertheless, Wolfe concludes that the major factors were the simultaneous presence of two epidemics and the loss of social cohesion that accompanied them.

War Games

For more than 20 years planners of U.S. nuclear strategy have been playing a game of simulated nuclear warfare developed by workers at the Rand Corporation, the Air Force Academy, Harvard University, the Massachusetts Institute of Technology and the Naval Postgraduate School. The rule book for a version of the game called Half-SAFE has recently been made available to the public. It is described by Susan Dahlberg and Tom McNevin in *The Bulletin of the Atomic Scientists.*

Under the rules of Half-SAFE the world is divided into two opposing forces, represented by teams of from four to 10 players. A game director acts as referee, messenger and occasionally as mischief-maker. The game begins with three planning periods, in which each team invests its resources in developing, building and deploying weapons and in gathering information about the other side. After every planning period each team prepares a "strategy statement," explaining how it would respond to various possible situations. If the strategy calls for using nuclear weapons or threatening to use them, the team must also prepare a "megatonnage statement," listing the targets to be attacked and the megatonnage to be allocated to each target.

At the completion of planning, a crisis begins, defined by messages from the game director. The moment at which one team orders either a "premeditated attack" or a "last-minute preemptive strike" is designated Zero-Hour. A few minutes later, at H-Hour, the actual launch begins. A little later still comes X-Hour, when the defending team launches its retaliatory weapons. The final stage of the game is aptly called the postmortem.

Although Half-SAFE is meant to train the commanders of American nuclear forces, Dahlberg and McNevin point out that the game is curiously inconsistent with certain elements of U.S. policy. In particular the team representing the U.S. is not constrained by the doctrine that nuclear weapons should serve primarily as a deterrent to war and would be put to use only in retaliation after an attack. In the game either team can launch a first strike or can "launch on warning," having detected what it takes to be evidence of an impending attack. Moreover, military leaders can initiate a nuclear exchange without an order from the president. The emphasis on winning a war rather than averting it suggests the game is not so much practice in thinking the unthinkable as an exercise in planning the inevitable.

Finland's Technological Advance: Specialization and Innovation

by Michael Frenchman

Innovation has been one of the keys to Finland's use of applied technology for industrial development and production. In an ever-competitive world the 4.7 million Finns have had little chance to match the wide-ranging production and volume of manufactured goods from the major industrialized countries of Europe and other parts of the world.

Traditionally its economy has been centered on the so-called forest-based industries—pulp and papermaking, and wood processing—and it has achieved a fine degree of expertise in these areas especially in the design and manufacture of machinery for these industries. But during the last decade there has been a concentration of effort to capitalize on the country's rather limited R&D facilities for product development across a wide cross section of industrial applications.

Its industries now produce many highly technical specialized products and skills from the construction of ice-breakers and Arctic oil-drilling ships, new developments in thin-film technology to a variety of original ideas for the conservation of energy.

Since the oil crisis of the mid-70's, the Finns, like everyone else, had a preoccupation with saving energy which had a catalytic effect on R&D in this field as the country was totally dependent on imported oil and gas, mostly from the Soviet Union.

The oil crisis stimulated more efficient ways of producing heat and power including using the country's substantial peat resources equivalent to 5,000 million tonnes of oil. More research has gone into the co-generation of fuels and the development of fluidized bed combustion chambers by companies like Strömberg and Ahlström. District heating plants have been expanded in many areas of the country and the Finns are trying to promote their expertise in these and other energy saving fields abroad. New ideas have been put forward for "living energy" houses and offices which have technically proved to be both environmentally and economically successful.

Mr. Kaarlo Kirvela, a vice-president of Ekono, believes there are more possibilities for conserving energy in pulp and paper production by co-generation of electrical power from mixed fuels such as wood chips and peat. More industrial waste heat should also be recycled.

Ekono, which has its headquarters in the technical research center of Otaniemi just outside Helsinki, has designed and built its "energy" house which is a five-story block entirely heated by self-generated energy backed up by only a minimum amount of applied electrical power at night. This has been achieved by a high degree of efficiency in design and function of the building at comparatively low cost. Using a modular system of construction with hollow concrete slab forms, air can circulate through floors and ceilings and through specially designed air extracting windows with quadruple glazing. Solar heat source is derived from the windows and from collecting panels on the roof. Other heat energy is generated by body temperature of the room occupants and from artificial lighting and office equipment. All this is basically "free" heat and the entire building acts as a kind of heat reservoir with additional storage in the bedrock beneath the structure. Another Ekono expert, Mr. Antero Punntila claims there is only a five-year payback period for the extra capital cost.

Fluidized bed combustion (FBC) was first developed in the late 1920's and almost any fuel such as wood chips, peat, oil and sewage sludge can be burnt. Air is blown through a bed of sand in the bottom of the combustion chamber and after primary ignition with oil or gas extremely efficient combustion takes place in the swirling air stream which resembles a mini-cyclone burning at a low temperature.

One of the most successful developments has been Ahlström's fast circulating FBC PYROFLOW™ chamber which originated from the need to lower fuel costs at a board mill. Tests showed that the FBC developed 98 percent and even 100 percent combustion efficiency. A retrofitted furnace using coal and peat at the Pihlava mill, which has been in operation for a year, produces 65 MWt process heat from a 14.3 MWe back-pressure steam turbine. Ahlström has now joined with the San Diego Pyropower Corporation to use an advanced FBC steam generator for oil recovery from heavy crude and tar sand deposits.

Other innovations on the industrial front have been the direct result of trying to produce something that was not only different but also competitive. These have included such developments as Outokumpu's flash smelting and continuous casting systems in use worldwide; Rautaruukki's special steel plates for building ice-going vessels to be found under construction in many Finnish shipyards and even specialized shipbuilding techniques themselves. Then there are the traditional based industries of pulp and papermaking for which highly advanced machinery has been designed and manufactured by companies like Valmet, Wärtsilä and Tampella.

The shipbuilders are among the most buoyant in the world at present with full order books for the next three or four years. If there is a case for out and out versatility and innovation it is in this field. Because they have been unable to compete with the rest of the world the Finns have concentrated on specialization and prompt delivery which has earned them international recognition.

The yards of Valmet, Wärtsilä, Rauma-Repola, and Hollming have never gone in for conventional ships. Instead they have concentrated on ice-breakers [*see page* 7], ice-cargo vessels, drilling ships for the Soviet Arctic, oil platforms, floating "hostel" ships, tugs and specialized cargo carriers. In addition Wärtsilä, which has made a name for itself with its ship-surgery operations (cut and lengthen), has just captured the cream of the cruise liner market with an $80 million order from P & O in the face of some of the stiffest competition seen by the industry. Delivery on time, particularly as far as the Soviet Union (Finland's major shipping customer) is concerned, and quality have been key factors in their success.

Closely allied with the construction of ice-breakers (more than two-thirds of all those built in the world come from Finland) has been the production of a special steel called RAEX Polar from Rautaruukki's foundries at Raahe. Research showed that the greatest weakness in polar shipbuilding design was the welds between the steel plates which were liable to corrosion and possible failure. Rautaruukki's experts eventually came up with a special micro-alloyed steel with restricted manganese, silicon, and sulfur content which, after welding, proved to be resistant to low temperature seawater conditions—in fact down to −60 degrees C.

Cold has played a vital part in many aspects of Finnish basic research, not least because of the country's geographical position and the fact that a third of the land area lies above the Arctic Circle. Much of the country's grass-roots research as a whole comes from the VTT (technical research center) and the Helsinki University of Technology's laboratories at Otaniemi, the red-brick science campus designed by one of Finland's best known architects, Alvar Aalto. Low temperature physics has become a major research area attracting much international interest.

It is not surprising that one of the coldest temperatures ever reached—50 nK—was obtained with one of the Low Temperature Laboratory's cascade nuclear refrig-


If your last flight to Europe made you feel like an intruder instead of a guest,
it's time for the Finnish experience.
Beneath that smart mink hat stands someone who cares. Someone of competence as well as charm. Yes, she'll serve your champagne graciously. And snug you into your First Class sleeper seat. She'll also converse intelligently in four languages. And tend to your needs throughout the flight with quiet efficiency.
Such spirit and pride among all our people make flying with Finnair an exceptional experience. And it's yours. To Finland, Scandinavia, Europe and beyond. Talk to your travel consultant or call Finnair, the national airline of Finland.
FINNAIR
We bring the Finnish experience to flying, all over the world.
FINNAIR

erators. But from a more practical point of view the modus vivendi for much of the work at the nearby Arctic Research Laboratory involves the study of ice-fracture mechanics at the comparatively warm temperature of −60 degrees C. Professor Pauli Jumppanen, who is busy organizing next year's polar engineering conference which is being held in Finland, believes that much work needs to be done on the study of the effects of cold on men and machines in the Arctic regions. This is because of the growing commercial interests in these areas as the frontiers of exploitation of natural resources are pushed further back.

Not so far removed perhaps from the magnetic anomalies of the Pole is the magnetically shielded room (MSR) which has been constructed as a joint experiment with the VTT instrument laboratory. The MSR consists of three mumetal and six aluminum layers for eddy current shielding plus a fluxgate coil system. The entire room weighs seven tonnes and the degree of magnetism inside the MSR is 10,000 times less than outside. Its purpose has been to develop sophisticated magnetometric instruments which can detect minute magnetic signals such as might be emitted from the human brain or other parts of the body. These are measured by a modified SQUID magnetometer. Similar work on clinical analysis is being done in the field of proton measurement using nuclear magnetic resonance (NMR) techniques. This has enabled researchers to open a "window" into the heart of a cell or other soft tissue which could lead to the development of a new family of body scanner equipment.

The role of individuals in R&D has always been important and many of Finland's younger innovators had done their early research on the VTT campus. One of them, Tuomo Suntola, was a student in the mid-70's.

Today Suntola is vice-president of Lohja's electronics division and has successfully developed the world's first practical electroluminescent display based on new materials technology. His method is known as "atomic layer epitaxy" (ALE). It is a totally new approach to thin-film fabrication involving controlled growth of the deposits on the film surface. Now being marketed as the Finlux display, it basically consists of a glass sandwich with a fine transparent matrix of electrodes. The panels can be used for a wide variety of instrument displays, moving logos, or public information boards.

Market considerations have often been the stimulus for R&D. One example in Finland has been the production of advanced radiosonde and weather measuring devices by an old established company, Vaisala. This company, which today produces 25 percent of all the weather balloons in the world, was one of the first in the country to work on silicon chip technology.

Innovation in electronics has been a key factor from Salora's low energy consuming television sets to Euroka's new family of computer speech synthesizers and "bubble" memory chips. Bigger companies like Nokia have edged into the hardware market with a new kind of 32-bit computer, the MPS 10 with advanced ADA program support.

Euroka is a relatively small company that began life making shop fittings. It now has a market lead in computer component modules and was the first European company to commercially develop the 1 M-bit bubble memory which has a capacity of 128 K which it calls the "bubblebug" memory. Mr. Jussi Liesiö, one of Euroka's engineers believes that his "bubblebug" modules have many advantages over conventional dynamic 64 K RAM's which are subject to alpha particle error and need sustained battery back-up circuits.

Another major area of research has been into voice synthesizers which can be preprogrammed and have unlimited capacity. Euroka has already developed a prototype intelligent interactive communications system using voice synthesizers called Syntelcom which Mr. Eero Viljanen believes is further advanced than any U.S. concept.

Mr. Liesiö, commenting on the future for microcomputers in Finland, believes that there is plenty of scope for the expansion of industrial process control systems and for more computerized services. This is a view shared by other companies such as Nokia, which began life in the rubber and electrical cable manufacture industry. Electronics was yet another example of how a Finnish company has had to diversify and specialize in order to compete. Nokia has been producing both its own hardware and software for computerized banking services, industrial automation and advanced communication systems. A recent innovation just launched is the Mikko Processing System MPS 10 32-bit computer which offers real-time multi-user systems supported by the little-known American ADA programming language. This was originally conceived by the U.S. Department of Defense and Kurt Wiksted, president of Nokia Electronics, and his R&D team are convinced that ADA may replace many contemporary computer languages particularly for business and financial applications.

The MPS 10 is a new concept from Nokia and is essentially a true stack machine with no "visible" registers employing a single level address system. The main objective has been to reduce costs and yet improve protection and reliability of both hardware and software.

Most successful innovations are often the result of bringing together different aspects of technology, the application of R&D to solve a variety of problems, particularly in industry. One aspect of this has been the difficulty of controlling the speed of electric motors. A leading role in this area has been taken by one of Finland's oldest and largest companies, Strömberg, whose interest in regulatory and control systems for such motors coincided with the early development of semiconductors in the early 1960's. R&D has played an important role in the company's development, a point stressed by its president, Mr. Antti Potila, who is convinced that more needs to be done in the country as a whole. "The solution to our future lies in innovation within very narrow gaps of development," says Potila. His own company has pioneered the AC-motor drive, or inverted frequency converter drive, often used for traction drive systems in trains and trolley buses as well as for industrial applications. The man behind this innovation and chief designer of the frequency controller system called SAMI is Mr. Martti Harmoinen, manager of the company's electronics division. In 1981 he was made Finnish Engineer of the Year. SAMI is a device which converts a fixed frequency into a variable one which can alter the speed of an AC induction motor with excellent energy saving capabilities. It also provides stepless speed controls where variable speeds are needed such as in pumps or fans.

One of the principal applications for the SAMI control is for traction power in locomotives and trolley buses. Recently Strömberg agreed to develop a 15,000-h.p. traction unit for a 12-axle prototype Soviet locomotive which will be one of the largest ever built. The company is also providing frequency control equipment for 33 articulated trams in Rotterdam and 11 trolley buses in Switzerland. It has now concluded a licensing agreement with the Los Angeles Garrett Corporation for the production of AC motors.

The frequency converter is based on pulse-width modulation (PWM) because of its particular suitability for traction drives. Thyristor switches are programmed in such a way that symmetric three-phase voltage with a rectangular form is generated. By changing switching speed the three-phase voltage is varied. The pulse widths are controlled in such a way as to produce a smooth current which overcomes any harmful torque. Because of this particular feature Strömberg controls are fitted to nearly all the diesel electric drives on ice-breakers which have to overcome major thrust torques. Another industrial application has been to fit one of the first frequency controllers to papermaking machines at one of United Paper's mills in Finland.

The traditional backbone of the economy has always been the forest industries of which pulp and papermaking have been the main contributing sectors. Providing design and manufacturing capacity on the machine side of the industry has been developed in the face of keen international competition from Finland's Nordic neighbors as well as France and the U.S. The first completely locally-made papermaking machine was not produced in Finland until Valmet entered the scene in the early 1950's. Since then the company has gone from strength to strength in this highly specialized field. Valmet, which makes everything from sporting rifles to container-handling equipment, had a record year in 1981 with 20 paper-machine deliveries and rebuilds worth Fmk. 1.3 billion.

A growing market for the company is the U.S., which the Finns consider to be wide open because of energy-wasting and environmentally unsuitable mills. Valmet is one of the three members of the TVW paper machine manufacturers' group. The others are Tampella and Wärtsilä. Mr. Otto Freund, managing director of the group, says that it was formed to combat technical competition and improve research. All three members are in fact complementary to each other specializing in different areas of production as pulp and papermaking machines are complex hybrid machines. TVW is in fact an innovation in marketing for the Finns but its operations are not just confined to the sales side. Technically speaking all three partners have achieved major developments. Valmet's research director Professor Antti Lehtinen, who was recently made a Fellow of TAPPI, has pioneered many papermaking innovations, particularly on the instrumentation and process control side. He now holds 13 patents in these areas, the most notable of which is Valmet's "Symconcept"—a method of producing paper with symmetric qualities by controlled means. Tampella has been producing stronger pulp with less energy consumption by the pressure groundwood (PGW) system in response to the development of thermo-mechanical pulp (TMP). Wärtsilä's forte has been to create an automatic wrapping innovation for the bulky awkward-to-handle paper rolls. It reduces manpower to a minimum.

Mr. Freund says that there is still room for much further technical development in the industry. In a final word, he adds: "We need more paper from less pulp with less energy . . . the key thing is to maintain our level of R&D and to find more innovations; we need more creative people who can maintain a very high standard of research. The name of the game today is not only cost but reliability and delivery on time. This is what our customers want and what we can provide."

One of the names known best internationally for its approach to innovation is Outokumpu, which started in 1910 as a copper mining company. Again in the search for energy economy and the need to compete with other metal producers, Outokumpu's researchers developed a new technique for producing copper after the war. This became known as "flash smelting." It later applied the same technique to smelting sulfidic nickel concentrates. The basic idea is to replace conventional roasting and smelting by a combined process which also partly converts the concentrates in a single flash smelting action. The dried, fine-grain concentrates are fed into the oxygen-rich furnace where they are held in suspension. Heat generated by the rapid exothermic reaction in the suspension is used for the actual smelting process and only a small amount of additional energy is required because the suspended concentrates generate their own energy and feed on themselves.

Now, according to Mr. Asko Parviainen, the company's metallurgical manager, they have made a breakthrough by extending the flash smelting process to the production of lead. This has considerable environmental advantages as well as being energy saving. He explains, "Flash smelting has developed from using latent energy in mineral concentrates and this has been the nucleus, if you like, of our innovation. It has also created these environmental conservation advantages." For instance today more than 86 percent of all the world's lead production is from conventional sinter-plant blast furnaces in which lead is mixed with coke before being roasted and blasted. As a result there is considerable atmospheric pollution from sulfur and lead emissions. Mr. Parviainen believes that flash smelting of lead will fulfill a much-needed development for the industry, particularly in the U.S. where the traditional reverberating smelters will have to be replaced for environmental reasons.

Undoubtedly innovation and specialization have been the keys to Finland's technical and economic success. Although there may be traditional rivalry between the metal men and the wood men it has not stood in the way of progress. The present decade will see new advances in biochemistry and pharmaceuticals, two fast growing sectors. And where but in Finland, where all alcohol is controlled by a state monopoly, Alko, would you expect to find research into the structure and manufacture of enzymes? Innovation has also spread to some of the heavier sectors, such as the construction industry. Both Lohja and Partek have introduced new methods of building concrete elements, including transportable site factories—some even working in the Middle East and the Soviet Union.

Even the sacred sauna has undergone a change, but primarily for export only. Instead of the traditional woodburning stoves to heat the water and produce searing steam, it is all done by electricity—an innovation which most Finns would rather not know about.

If you're in the paper business you should be able to spot the difference.

Or should you?

As a professional you would certainly spot the difference at once – if there was a difference. But in this case there is none, even though we're talking about uncoated paper. This paper doesn't have the usual A and B sides. It just has two A sides – identical, symmetrical and of equal quality. Symmetrical paper offers significant advantages in terms of quality, production and marketing.

In 1971 we started an extensive and thorough paper-machine product development project at Valmet. The targets we set ourselves were to produce a superior paper quality and also to raise the level of automation and increase production speed. As a result of intensive and relentless work we are now one of the world's leading manufacturers of paper machines producing symmetrical paper. The product of this work is known as the Valmet SYM-concept®.

Valmet has made important contributions to Finland's industrial and technological know-how, not only as a manufacturer of paper machines but also in its other engineering sectors.
We have the resources and we are ready to design and implement technology for the benefit of today's industrial and business ventures.

We sell technology that sells

VALMET

Head Office P.O.B. 155, SF-00131 Helsinki 13, Finland, telephone +358 0 171441, telex 124427 valp sf
Valmet Information Office, 7 Westchester Plaza, Elmsford, N.Y. 10523, USA, telephone (914) 347-4440, telex 646179 valmet emfd

HOW TO BREAK THE ICE: PUTTING DESIGN TO THE TEST

by Eero Makinen*

Since the 1950's the extension of winter navigation in the Baltic and in the Soviet Arctic seas and Siberian rivers has been the stimulus for greater research into icebreaking design and construction in Finland. Breaking evenly formed and level icefields is the fundamental requirement of designing all icebreakers but it is only one of several factors which has to be taken into account. The thickness, variation in form, coverage, and natural characteristics of the ice itself must be considered too.

After a ship has broken the ice by loosening pieces from the solid field it must then penetrate the broken floes. Usually a ship cracks the ice with a downward movement of the bows, using her own headway and weight after first rising up over the ice. This method has proved to be the most efficient way of forcing a passage through ice of all kinds. There is naturally also a relationship between the amount of energy needed by the ship, not only for breaking the ice, but for thrusting the ship through the sea and overcoming friction from the broken ice and water contacting the hull. In addition, the ship must overcome resistance caused by friction between the broken ice blocks themselves as they are pushed aside by the vessel.

It has taken twenty years to perfect a satisfactory downward-bearing hull form which can achieve maximum breaking efficiency. The performance of the ship through breaking ice can still be improved by several methods. One is to heel the ship from side to side by transferring fuel or water ballasts from one tank to another; with the fastest methods in use today this can be done in under half a minute. The other is to use an air bubbling system known as WABS which has been devised by Wärtsilä. Air is blown into the water through holes in the ship's bilges. As the bubbles rise alongside the ship's side they decrease friction between the hull and the ice reducing resistance by forming a "lubricating layer." This can lower ice resistance from between 20–60 percent. WABS has been installed in more than 50 different vessels, half of them icebreakers. Both heeling and WABS greatly reduce the risk of vessels becoming stuck in the ice.

So far, the downward-bearing hull form has been the best solution. But, as in other research fields, icebreaking has attracted its fair share of unconventional ideas, most of them unrealistic. These have included sawing, melting, using banks of multiple drills, and even lasers. One alternative method, which has proved successful within certain limitations on level ice, has been wave movement induced by air-cushion vehicles.

Wärtsilä shipyards have built more than 60 percent of all the world's icebreakers. These include the towering 36,000 shaft horse power (s.h.p.) polar icebreakers like the Admiral Makarov to shallow draft ones for use in Siberian Arctic rivers which can punch their way through more than a meter of ice with ease. With the focus of attention now turning to exploitation of offshore resources in the Arctic, Finland is in a unique position to play a leading role in the development and construction of specialized vessels from towed air-cushion vehicles for transferring cargoes across unbroken ice to new types of offshore platforms. Wärtsilä is now doing a joint study with the Soviet Union for the design of a new generation of nuclear powered icebreakers which will help to open up the Arctic oil fields. Wärtsilä shipyards currently have more than 19 icebreakers and ice breaking cargo vessels on order. The company has also been taking an interest in the Antarctic and has delivered a special passenger and cargo carrying 16,200 s.h.p. icebreaker for Argentina.

Two critical elements governing the performance of the icebreaking ship are its propellers and machinery. These have to be designed so that the ship can move at a desired speed in both ice-free waters and through icefields of a particular thickness. Screw propellers driven by diesel engines, steam turbines, or electric motors have proved to be the most efficient. The propellers frequently hit the ice blocks which usually cause a sudden decrease in the thrust bringing additional loading to shafts and engines which have to be designed to much stronger specifications than for use in conventional ships. In addition, propeller bosses, rudders, and bilge keels must be considerably stronger and heavier. Furthermore, ice damage and collision with the hull itself must be taken into account—the basic requirement for the naval architect as only the ice should be crushed, not the ship's hull!

Meeting these demands, and yet producing an efficient design, all adds to the capital cost. Increasing the weight and use of heavier materials to produce adequate structural requirements also proportionately reduces the vessel's payload, especially in respect of cargo vessels with icebreaking capability.

This makes it all the more necessary for the designer to know very accurately what iceloads can be expected under varying conditions and what effect these will have on different parts of the vessel. Apart from the traditional voice of experience from ships' captains used to operating in icefield conditions, the modern designer has three basic research aids: theoretical calculations, laboratory model tests, and sea trials in icefields. All three are essential and complementary. Although from between five and ten ships, from polar icebreakers to harbor tugs, can sometimes be used for sea trials in one winter's season, this is often impractical as it is very costly in terms of time and money. Also the results obtained vary greatly because of different environmental conditions and changing parameters. Many solutions for design problems can be arrived at from theoretical calculations which are now enhanced by computer aided design (CAD).

But perhaps the most useful facility is the icebreaking model basin for testing scale design models in simulated icefields. In 1969 Wärtsilä built the world's first icebreaking model basin in which more than 100 models have now been tested. An essential factor is the proper correlation between model tests and full-scale trials. In order to obtain the correct parameters considerable research has had to be undertaken but it is still not always possible to simulate faithfully all physical aspects and ice conditions in the test tank. Some of the natural mechanical properties of ice, such as elasticity and flexibility, cannot so far as is known be scaled down simultaneously. Many factors can be studied in the basin—notably the ship's performance under different conditions of level ice, ridge formations, pressure ice, and broken channels.

All icebreakers and ice-class cargo vessels are now tank-tested at an early stage of design. This year a new, much larger model test tank measuring 78m × 6.5m will be completed. This will be the largest test tank of its kind in the world.

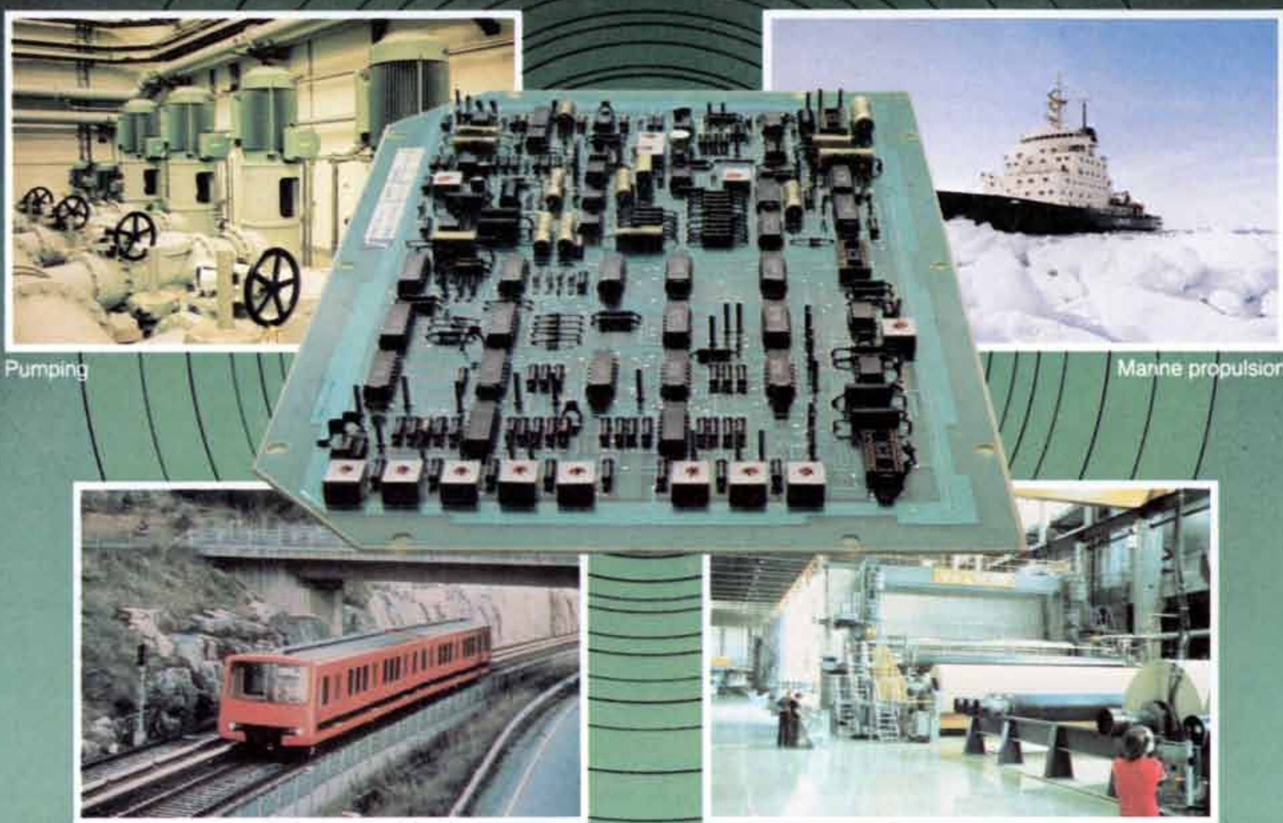
All activities in the icebreaking field are carried out by the company's Arctic Design and Marketing Department (WADAM) which carries out a wide range of operational studies and design investigations. These are not just limited to the Baltic, Siberia, and the Sea of Azov, but also include the Canadian Arctic and the Antarctic oceans of the southern hemisphere.

Wärtsilä is the only company in the world specializing in the building of icebreakers. The need for icebreaking technology and the capability of operating vessels in ice waters will increase dramatically in the next decade as development of the Arctic oceans becomes a reality. Already exploration is being intensified and actual production has begun in some areas.

*The author is assistant managing director, Oy Wärtsilä Ab, Helsinki Shipyard in charge of Arctic Design and Marketing (WADAM).

STROMBERG

CONTROLLED REVOLUTION



Pumping

Marine propulsion

Traction

Line drives – paper

The breakthrough, allowing the speed of a standard squirrel cage motor to be continuously varied, is the most significant advance in variable speed drive technology in recent years. Stromberg's Sami Frequency Converter, combines all the benefits of variable speed with the advantages of using A.C. motors.

As one of the first companies to introduce solid state technology to their controllers, Stromberg have now truly revolutionised operating efficiency in the transport, marine, papermaking and process industries.

For further information, contact:

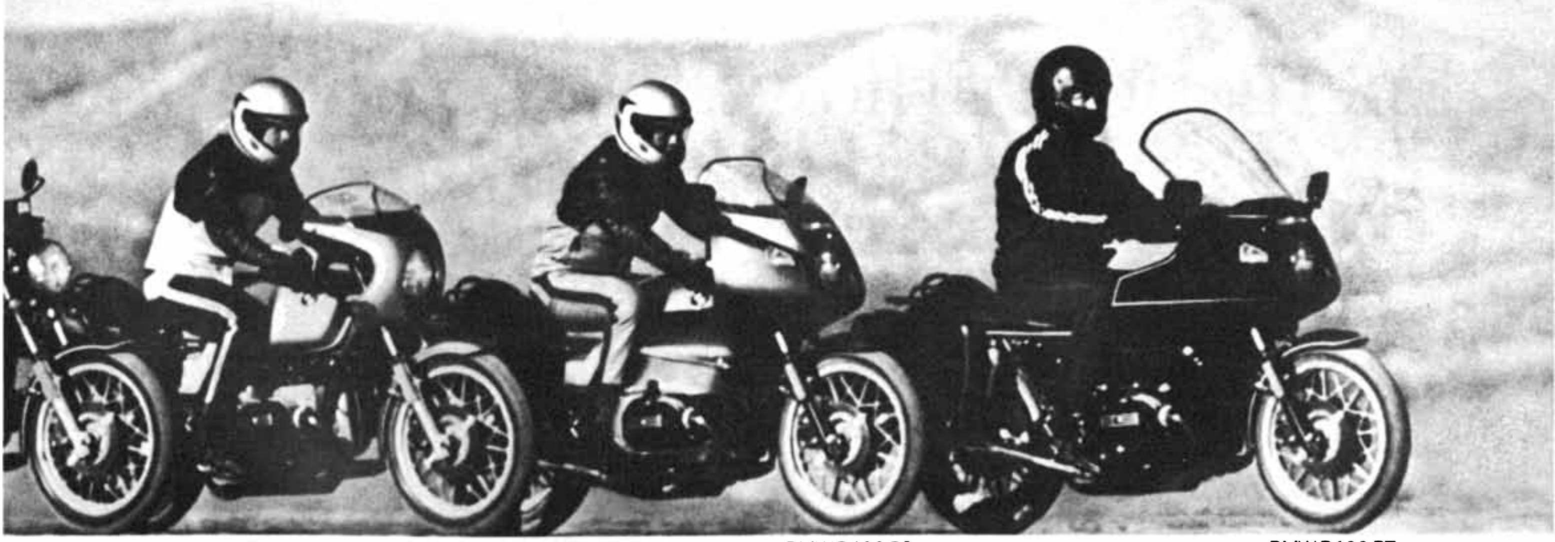
OY Strömberg Ab
PO Box 118 SF-00101 Helsinki 10

Strömberg

Tel. 550045
Telex 12405

BMW R65 LS BMW R65 BMW R80 G/S

THERE MAY BE ANOTHER 200,000-MILE MOTORCYCLE CLUB. BUT WE KNOW OF ONLY ONE WITH ANY MEMBERS.



BMW R 100 BMW R 100 BMW R 100 RS BMW R 100 RT

In September of 1981, Clyde Rees eased his BMW R90S out of the driveway of his home in Shelbyville, Tennessee.

Though he had no particular destination in mind, by mid-October Clyde was in the northern-most reaches of Nova Scotia.

A long haul, to be sure, but nothing out of the ordinary for Clyde. Over the past 8 years he has logged over 225,000 miles.

All on the same BMW.

The odometer on Carl Goldsby's BMW R75/5, meanwhile, reads 268,000 miles. Says Carl, of La Grande, Oregon, "I'm just as attached to my wife as I am to my bike. We've traveled almost every mile two-up."

Greg Beemer of San Jose, California, has put over 210,000 miles on his BMW. Miguel A'Llerio of Chicago, Illinois, has logged 245,000. Elwin Russell of Los Angeles actually has 342,000.

And the list goes on to form the steadily growing membership of an elite riding corps known as the BMW 200,000-Mile Motorcycle Club.

An organization whose very existence becomes all the more remarkable when you consider that today the average motorcycle will actually be scrapped after logging fewer than 14,000 miles.

(Obviously, no motorcycle manufacturer can guarantee that a motorcycle will last for 200,000 miles, since longevity depends largely upon individual riding habits, maintenance and other variables.)

THE ALTERNATIVE TO DISPOSABLE MOTORCYCLES.

"In typical BMW fashion," observed Cycle Guide recently, "the BMW engineers have brought their motor into the Eighties without resorting to needless complexity."

Indeed, the design of the BMW engine has grown not increasingly complicated with the years but

ingeniously simple, the hallmark of a brilliantly engineered machine.

It is the simple genius of the BMW engine that helps it to perform with what Cycle Guide refers to as "doorstop-like reliability."

And it is precisely this kind of reliability that is becoming more and more appreciated as the price of a motorcycle mechanic, now hovering at approximately $24 an hour, continues to rise.

In the end, a BMW does what we believe all motorcycles should do. But which few accomplish.

It places life's complexities behind you. Not beneath you.

HOW TO PARTICIPATE IN THE BMW MOTORCYCLE CLUBS.

Should you be approaching 200,000 miles on your BMW, your local BMW dealer will gladly arrange for you to receive a medallion like the one pictured here.

Should you be nearing the halfway point, you'll be happy to know that there's also a BMW 100,000-Mile Motorcycle Club.

Both clubs are informal, with no mandatory meetings and no dues to pay.

The only price of admission is the predictably high cost of a BMW, which ranges between $3,390 and $6,890.*

And should you ever decide to part company with your BMW, you'll find even that to be a rewarding experience.

Indeed, according to the NADA's Used-Motorcycle Handbook,** BMW's hold a higher percentage of their original purchase price than any other line of motorcycles on the road today.

THE LEGENDARY MOTORCYCLES OF GERMANY.

*Manufacturer's suggested retail price. Actual price will depend upon dealer. Price excludes state and local taxes, dealer prep, destination and handling charges. **Average retail selling price based on January 1982 NADA Used-Motorcycle Guide. Your selling price may vary depending upon the condition of your motorcycle and whether you sell it privately or to a dealer. © 1982 BMW of North America, Inc. The BMW trademark and logo are registered trademarks of Bayerische Motoren Werke, A.G.

Energetic Outflows from Young Stars

The radiation emitted by carbon monoxide molecules in clouds of gas and dust where new stars are forming discloses masses of gas moving outward at high velocity in opposed directions

by Charles J. Lada

In our galaxy stars are born in clouds of gas, bodies so cold that they do not emit any radiation at the visible wavelengths of the electromagnetic spectrum. Moreover, the clouds are permeated by cosmic dust, so that the visible radiation emitted by the new stars is absorbed. These conditions long presented a seemingly impenetrable barrier to astronomers seeking to understand how stars form. Radiation at the wavelengths of infrared waves and the shortest radio waves, however, can pass through these clouds, and thanks to the development of new telescopes and equipment for the detection of radiation at those wavelengths, astronomers are now able to explore the dark and dusty clouds where stars are born.

A prime probe of the conditions within star-forming clouds has turned out to be the molecule of carbon monoxide (CO), a gas first detected in space only a dozen years ago. In interstellar molecular clouds carbon monoxide molecules emit radiation at a wavelength of 2.6 millimeters. The study of such radiation has recently revealed a new and intriguing astrophysical phenomenon closely associated with the birth and early evolution of stars. When certain stars are in the earliest stages of their life, they appear to be associated with violent outflows of mass. In a number of instances molecular gas is found to be flowing outward from around newly formed stars in two supersonic streams 180 degrees apart. The gas in these bipolar streams typically has a mass many times that of the sun and accordingly carries enormous amounts of kinetic energy. The origin and nature of these energetic outflows is a mystery. It is nonetheless clear that they represent a significant stage in the evolution of young stars.

The new glimpses of the birth and early evolution of stars are being provided by half a dozen new instruments specially designed to detect radiation in the near-infrared region of the spectrum (wavelengths from two to 20 micrometers), in the far-infrared region (from 30 to 300 micrometers) and in the region of the shortest radio waves (from less than a millimeter to a few millimeters). The instruments include the three-meter Infrared Telescope Facility Project on Mauna Kea in Hawaii and the five-meter millimeter-wave telescope on Mount Locke in Texas. The gold-coated surface of the five-meter telescope is the most precisely figured millimeter-wave radio reflector in the world. Three years ago this instrument was the first to detect a bipolar high-velocity outflow in a molecular cloud.

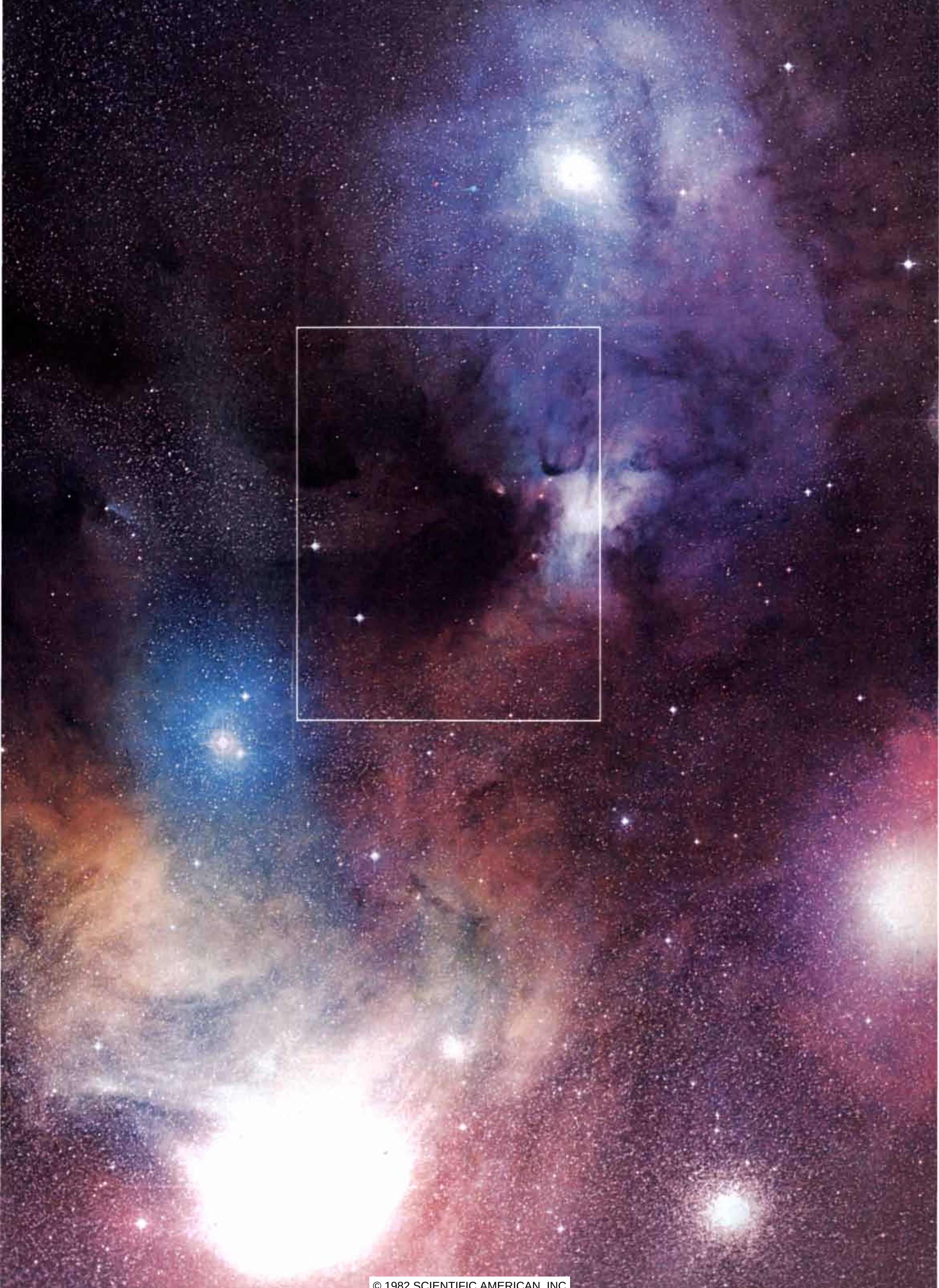
Perhaps the most remarkable of the new family of instruments is the Multiple Mirror Telescope (MMT) on Mount Hopkins near Tucson, Ariz., operated jointly by the University of Arizona and the Smithsonian Astrophysical Observatory. This unique instrument consists of six 1.8-meter mirrors mounted within a seven-meter circle. When it is operated so that the radiation striking the six mirrors is combined coherently, that is, in phase, it becomes the equivalent of a single seven-meter reflector. The MMT was recently operated in such a coherent mode to examine emissions from molecular clouds at wavelengths of less than a millimeter, a region of the spectrum in which it has been difficult to make high-resolution observations.

Studies of emissions of specific lines in the millimeter-wave region of the spectrum have enabled astronomers to probe the physical, chemical and dynamical nature of the dark clouds from which new stars must be forming. Far-infrared observations disclose the physical state of the dust that permeates the dark clouds and shrouds the newborn stars. Near-infrared observations have revealed the newborn stars themselves. For example, in the dark cloud known as Rho Ophiuchi infrared observations have revealed an entire cluster of 20 very recently formed stars buried too deep in gas and dust ever to have been detected at visible wavelengths. The synthesis of data from these various invisible-wavelength regions of the spectrum has added immeasurably to knowledge of star birth and early stellar evolution. In due course such work should yield a general theory of star formation.

Sites of Stellar Birth

It was not long ago that the subject of star formation and the genesis of planetary systems was limited to theoretical speculation. Relevant observational data simply did not exist. Some 40 years ago the situation changed abruptly with developments in nuclear physics and stellar dynamics. It became clear that the power supply of stars is nuclear fusion. With this recognition it became possible to predict how long stars could live and what their ultimate fates would be. It was apparent that the very luminous stars designated *O* and *B* are burning their supply of hydrogen nuclear fuel from 10,000 to 100,000 times faster than the sun is and therefore are doomed to exhaust their fuel in 10 million years or less. Such stars do not survive even one rotation of our galaxy

TYPICAL BIRTHPLACE OF STARS is the Rho Ophiuchi dark cloud, named for the bright star surrounded by a nebulosity of blue reflected light at the top center in this photograph. The brilliant star at the lower left is Antares in the constellation Scorpius. To the right of Antares is the globular cluster M4. The dark cloud between Rho Ophiuchi and Antares holds molecular hydrogen gas equivalent in mass to about 1,000 suns. Heavy concentrations of dust within the molecular cloud absorb more than 99.99 percent of the visible radiation emitted by a cluster of 20 new stars recently discovered in a survey made at infrared wavelengths by Bruce A. Wilking of the University of Texas at Austin and the author. Infrared radiation is 10 times more effective than visible radiation in penetrating the dust. The cluster of new stars is buried near the center of the outlined area. A photograph of the outlined area showing the location of the infrared objects is on page 86. The photograph was made by David F. Malin from three black-and-white plates exposed to different regions of the spectrum with the 1.2-meter Schmidt telescope on Siding Spring Mountain in Australia, operated by the Royal Observatory, Edinburgh.



What if you chose as a technical

"At AMP, Inc., an HP 1000 computer cuts product test costs by 96 percent."

AMP's Communication Products Division, in Winston-Salem, N.C. manufactures electrical and electronic connectors. Last year, thousands of products passed through an extensive battery of environmental tests, all controlled and processed by an HP 1000 computer system.

Group Systems Manager Anthony L. Chitea says, "With the HP system, we traded manual probing and a costly time-share program for automatic test, real-time data, and a 30-fold increase in productivity. Testing procedures which previously took three hours to complete are performed on the HP 1000 in less than 15 minutes. Based on this savings alone, our system easily paid for itself in less than a year.

"By enhancing productivity, we're able to increase testing volume 60 times over, while reducing costs per test by 96 percent."

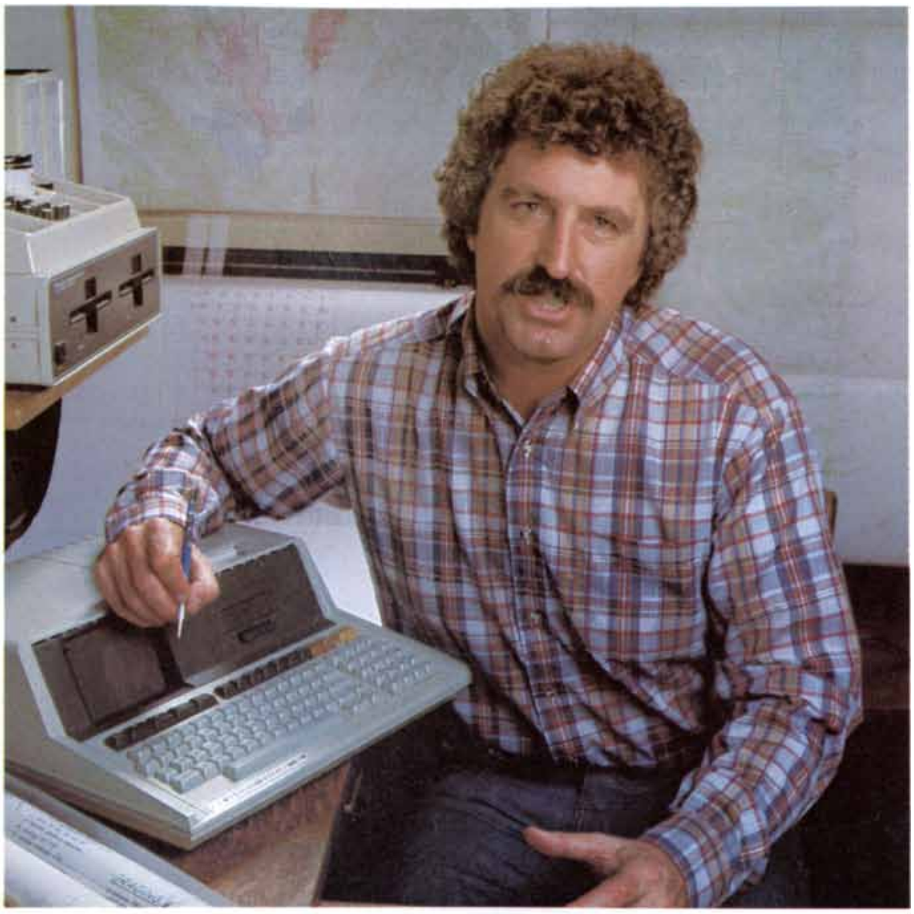
HP can be your personal computer partner, too!

A growing family of personal computers is available through HP's network of qualified dealers, or through HP's computer sales organization.

For example:

The new HP-87, with enhanced BASIC, up to 544K bytes of user memory, and dynamic graphics capabilities, can handle large computing tasks and present the data in easily understood charts and graphs. Extensive software offerings are availablefrom HP, as well as a CP/M® system module which opens the HP-87 to an extensive repertoire of software solutions. A wide range of peripherals gives your personal computing needs unlimited expression: a new color plotter; printers; and versatile mass storage including a new Winchester hard disc drive and flexible disc drives. Altogether, the HP-87 is a new, high-performance, highly affordable personal computer system.

Hewlett-Packard computer partner?



"At Ag Surveys, Inc. an HP personal computer has increased our productivity by 35 percent."

Ag Surveys, Inc. in Salinas, Ca., is a small business that provides consultation and a range of services to the agricultural business community.

The company's president, Clark Whittle, says, "A significant part of our business involves resurveying property lines that have become vague over the years. With HP-85 personal computer, our surveyors can evaluate amounts of data that would be impractical to attempt manually, and get instantaneous results. Using HP's land surveying and engineering software package and enhanced BASIC language, our people can work with the computer without lengthy training.

"In our first year with the HP-85 we've improved the precision of our services, increased the orders we can handle by 35 percent, and reduced data processing costs by 40 percent."

The HP 125, specifically designed for the business office, offers versatile management software, graphics, and impressive word processing capabilities to users who are not computer specialists. Through instructions displayed on its screen, it guides the user through procedures step by step. As part of a distributed systems network, it can also communicate with large computers. Through its CP/M operating system, the HP 125 offers access to hundreds of software programs.

CP/M is a registered trademark of Digital Research, Inc.

More good reasons to consider a working partnership with HP.

For more information on these personal computer systems, write to A.P. Oliverio, Vice-President, Marketing, Dept. 242, Hewlett-Packard Co., 3000 Hanover St., Palo Alto, CA 94304.

When performance must be measured by results

HEWLETT PACKARD

00242

(100 million years) and have a life expectancy barely a thousandth that of the sun. *O* and *B* stars are so young that they cannot have got far from their places of birth in clouds of gas and dust.

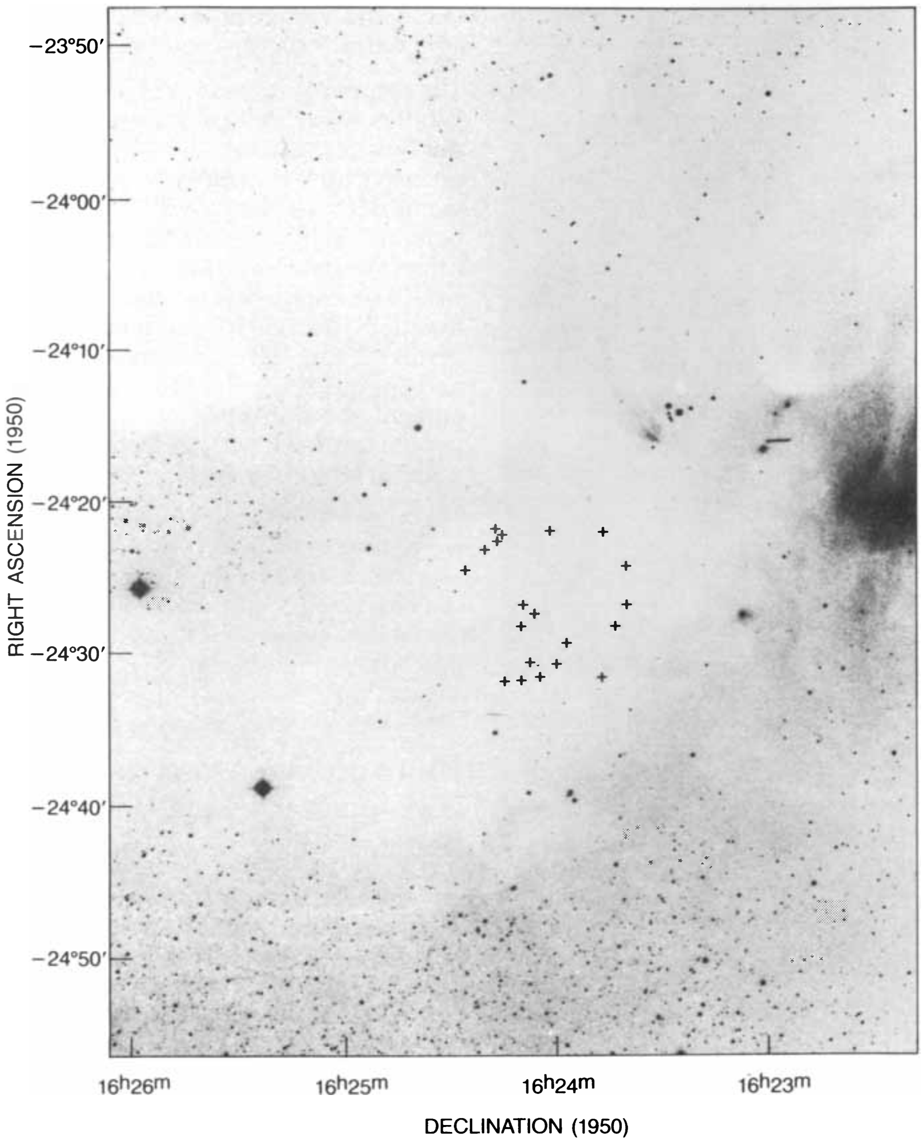
Studies of the distribution and motion of *O* and *B* stars have shown that many of them form groups, which have been named *OB* associations. Instead of being gravitationally bound like the stars in a star cluster the members of an *OB* association are moving outward, away from one another. The rate of expansion indicates the associations are no more than 10 million years old. These two independent items of information—*O* and *B* stars' high rate of fuel consumption and their outward motion—demonstrate that in the present epoch of our galaxy's history stars are still being born. Since the life cycle of *O* and *B* stars from birth to death is so short, their distribution and orientation in a small region of space can be regarded as a fossil record of the distribution and orientation of the protostars that were their precursors and presumably of the gas and dust that came before the protostars.

The advent of observations in the infrared and millimeter regions of the spectrum has revealed that *OB* associations are formed in giant molecular clouds some 300 light-years in diameter consisting of perhaps 100,000 solar masses of material in a gas more tenuous than the best laboratory vacuum. Such gaseous aggregations may be the most massive objects in the galaxy. Twelve years ago their existence was unknown [see "Giant Molecular-Cloud Complexes in the Galaxy," by Leo Blitz; SCIENTIFIC AMERICAN, April]. It is now recognized, primarily as a result of infrared and millimeter-wave observations, that there may be several ways in which stars can actually form from giant molecular clouds. Exactly how many mechanisms are there? How are they related? What is their relative contribution to all star formation? Such questions, the focus of much current astronomical research, could not have been addressed a decade ago.

The Detection of Molecular Gas

The clouds that give birth to stars seem to consist almost entirely of hydrogen in molecular form (H_2). Because the clouds are cold (between 10 and 50 degrees Kelvin, or degrees Celsius above absolute zero) the hydrogen molecules cannot be detected directly. Molecules in space can be detected only if they emit photons, as they do when they drop from a higher energy state to a lower one. The transitions from one state to another give rise to spectral lines, which, depending on the energy lost in making the transition, can fall anywhere in the electromagnetic spectrum. In molecular gases at low temperatures the principal transitions are from one state of rotation of the molecule to another state, transitions in which the radiation emitted is at radio frequencies. According to the rules of quantum mechanics, for a symmetrical molecule such as hydrogen to make transitions between adjacent rotational energy levels is forbidden. The first transition above the ground (lowest) level allowed for the hydrogen molecule therefore requires that the molecule gain enough energy to move up two rotational levels. At the temperatures prevailing in molecular clouds collisions between molecules can rarely supply enough energy for such a transition.

Fortunately molecular clouds also contain nonsymmetrical molecules, such as carbon monoxide, for which transitions between adjacent rotational levels are allowed. In even the coldest clouds collisions between molecules readily raise carbon monoxide molecules to their first excited rotational state. When the molecule drops back to its ground state, it emits a photon with a wavelength of 2.6 millimeters. Although carbon monoxide in its most abundant isotopic form (carbon 12 combined with oxygen 16) is nearly 10,000 times less abundant than molecular hydrogen, it is one of the principal trace molecules in interstellar clouds and therefore serves as an excellent surrogate for studying the unobservable molecules of hydrogen. The 2.6-millimeter spectral line of carbon monoxide was first detected from interstellar space in 1970 by Robert W. Wilson, Keith B. Jefferts and Arno A. Penzias of Bell Laboratories with the 11-meter radio telescope of the National Radio Astronomy Observatory on Kitt Peak in Arizona. The spectral line of carbon monoxide has since been observed throughout



NEW STARS IN RHO OPHIUCHI DARK CLOUD are identified by crosses on a negative print of the area outlined in white on page 83. Wilking and the author probed the darkest regions of the cloud at a wavelength of two micrometers with the 61-inch infrared telescope of the University of Arizona. It is estimated that the 20 stars have been formed within the past million years. The photograph, made with the 48-inch Schmidt telescope on Palomar Mountain, is from the National Geographic Society–Palomar Observatory Sky Survey.

the galaxy and in other galaxies as well.

From the intensity of molecular spectral lines astronomers are able to derive the temperature, density and molecular composition of the gas in star-forming clouds. In addition, because of the Doppler effect (which raises the frequency of radiating sources that are approaching the observer and lowers the frequency of sources that are receding) astronomers can determine the motion of a molecular cloud with respect to the solar system. Approaching sources are said to be blue-shifted, receding sources red-shifted. The Doppler effect also discloses the range of velocities of molecules in a cloud; the greater the range of velocities, the broader the spectral line. The observation of shifted and broadened molecular-line profiles led to the discovery of unusual gas flows in molecular clouds.

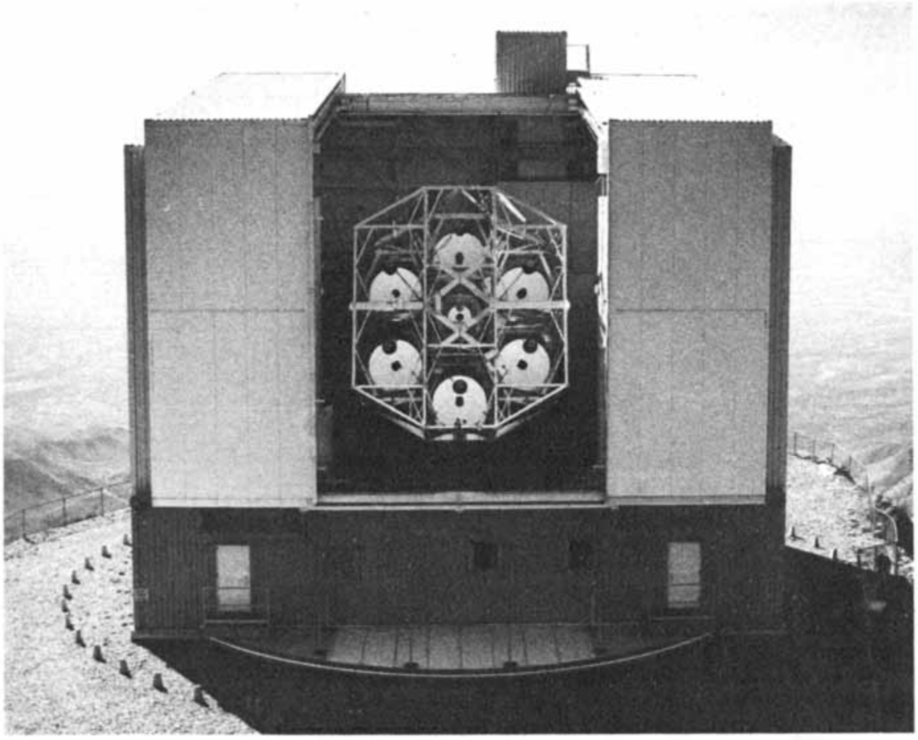
The High-Velocity Gas in Orion

If thermal motions were the only source of Doppler line broadening, interstellar molecular clouds would emit lines with widths equivalent to only a few tenths of a kilometer per second. The widths of lines are commonly expressed in units of velocity rather than wavelength because it is the motions of the gas cloud and of its constituents that are of primary interest. The carbon monoxide lines from molecular clouds are almost always much wider than the lines that would be produced by thermal motions alone.

With giant molecular clouds such as those near *OB* associations line widths in the range between one kilometer per second and three kilometers per second are observed over most of the cloud. In regions where star formation is most intense the line widths increase to between four and 10 kilometers per second. The source of such motions is a puzzle.

The most extreme instance of molecular-line broadening was discovered about five years ago in a tiny region of a giant molecular cloud 200 to 300 light-years across that is intimately associated with the lovely Orion Nebula. In a region less than half a light-year in extent carbon monoxide emission exhibits a velocity spread that is slightly greater than 100 kilometers per second. Studies of this small region at many wavelengths demonstrate that the high velocity of its carbon monoxide molecules is the result of an energetic outflow of gas from around a young object, possibly a protostar.

The existence of such an energetic and perhaps violent outflow of gaseous material from the vicinity of a young object did not fit in with the accepted picture of the formation and early evolution of stars. At first it was thought the outflow was associated with a very rare and possibly unique event, such as the explosion of a supernova in the Orion molecular cloud only about 1,000 years earlier. It has since become clear from millimeter-wave observations of other molecular clouds that such energetic outflows of gas are common and may even be a consequence of star birth.



MULTIPLE MIRROR TELESCOPE (MMT), recently completed on Mount Hopkins near Tucson, Ariz., is the first of its kind. In its normal mode of operation it brings to a common focus the radiation from six 1.8-meter mirrors mounted in a circle seven meters in diameter, thereby equaling the light-gathering power of six 1.8-meter telescopes or of a single 4.4-meter telescope. Because the light is combined incoherently, however, the resolving power of the MMT is still only that of a 1.8-meter instrument. In order to achieve the full resolving power of the MMT its light must be combined coherently, or in phase. When that is done, the MMT becomes the equivalent of a telescope seven meters in diameter. A phase-coherent system at submillimeter wavelengths was used for the first time with the MMT by Neal Erickson, Paul F. Goldsmith, G. Richard Huguenin and Ronald L. Snell of the University of Massachusetts at Amherst in collaboration with Bobby L. Ulich of the MMT observatory and the author.

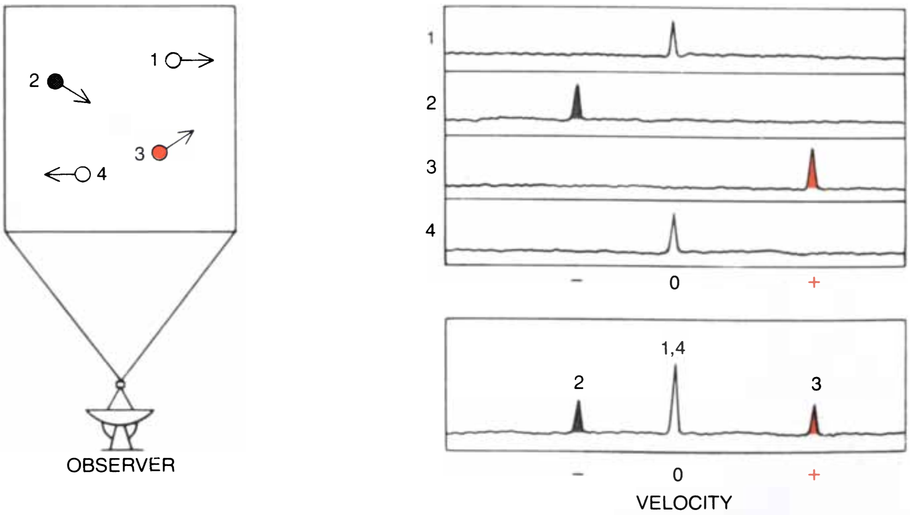
The Cloud around AFGL 490

One set of such observations was made two years ago by Paul M. Harvey of the University of Texas at Austin and me. The subject of our study was a molecular cloud around an infrared source suspected of being a protostellar object in the constellation Camelopardalis. It is known only by its designations in two infrared catalogues: either as U of A 1 (object No. 1 in the University of Arizona catalogue) or AFGL 490 (object No. 490 in the catalogue of the Air Force Geophysics Laboratory).

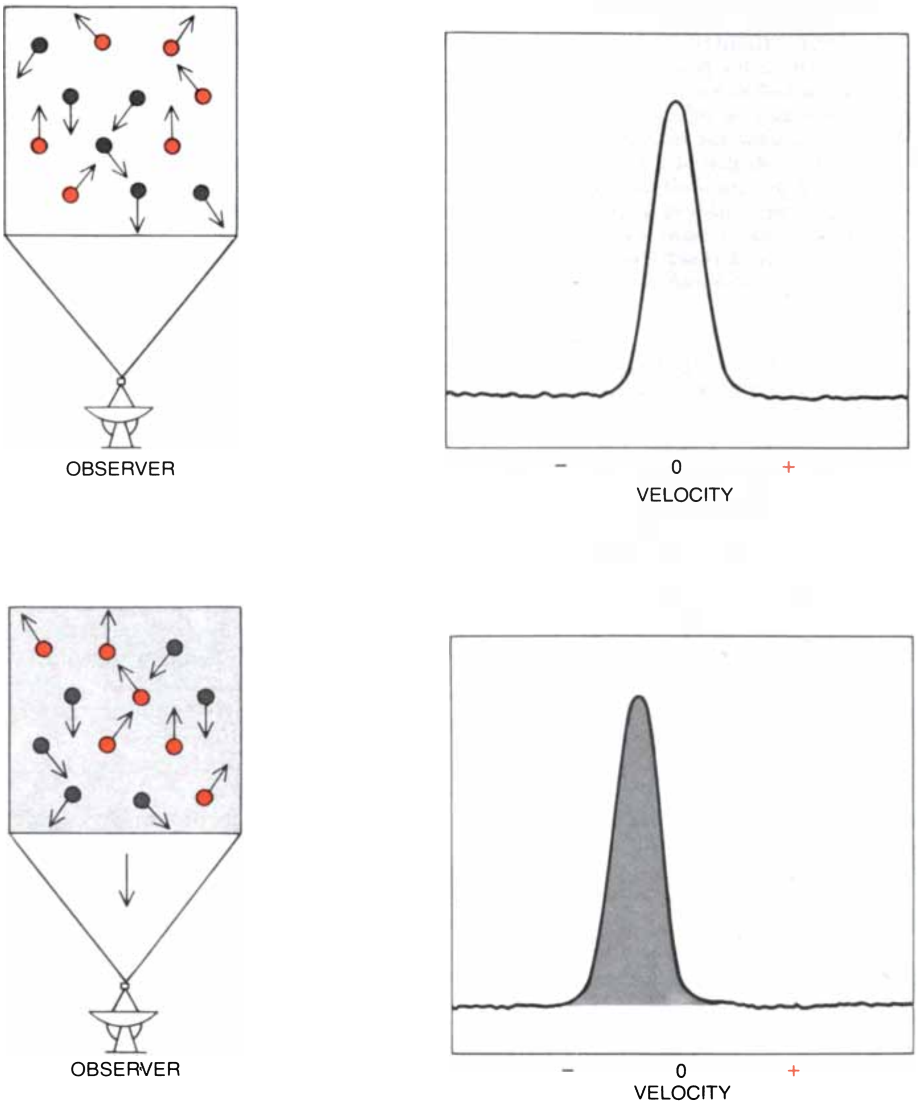
We made our first observations with the gold-coated five-meter reflector of the McDonald Observatory of the University of Texas at Austin on Mount Locke. When we aimed the telescope directly at the embedded infrared source, we were surprised to find that the carbon monoxide profile had broad velocity "wings." Instead of having the expected line width of two or three kilometers per second the carbon monoxide emission could be detected over a range of some 60 kilometers per second. At positions in the cloud other than the position of the infrared source the broad velocity wings disappeared and the lines had the narrow widths we had originally expected.

Such an unusual finding merited further investigation with a larger and more sensitive telescope. We therefore transferred our study to the 11-meter instrument on Kitt Peak, which has a resolution twice that of the five-meter telescope. This improvement in resolution led to another unexpected finding. As we scanned the telescope from north to south across the infrared source the shape of the carbon monoxide profile changed in a systematic manner. Slightly north of the source we observed only the red-shifted wing of the profile. Slightly south of the source we observed only the blue-shifted wing. When the telescope was centered directly on the infrared source, the carbon monoxide spectrum exhibited high-velocity components shifted both to the blue and to the red, hardly what one would expect from a gas cloud that was collapsing gravitationally to form a new star.

The gas flow extends for only about one light-year, or .3 parsec, in any direction, so that it must be a very young



RELATIVE MOTION OF MOLECULES IN SPACE alters the wavelength at which their emission is observed because of the Doppler effect. If the molecule is at rest or is moving transversely (*1, 4*), its "line profile" will not be affected. If the molecule is approaching the observer (*2*), its line profile will be shifted to a "bluer," or shorter, wavelength. If the molecule is receding (*3*), its profile will be shifted to a "redder" wavelength. Shifts reveal the velocity. The intensity at a particular Doppler velocity is related to number of molecules emitting at that velocity.



ACTUAL OBSERVATION OF A MOLECULAR CLOUD sums the emission from molecules moving with a range of velocities. If the cloud is neither approaching nor receding (*top*), spread of velocities due to internal motions will yield a broad line profile centered on the rest, or normal, wavelength. If the cloud is approaching (*bottom*), profile is blue-shifted, described as a negative velocity. A receding cloud will show a red-shifted profile, or a positive velocity.

dynamical event. (One parsec equals 3.26 light-years.) It would take only 10,000 years for the fastest-moving gas in this flow to travel one light-year. Molecular clouds themselves have an estimated lifetime of 10 million to 100 million years. By observing the emission from a rare isotopic form of carbon monoxide (in which the carbon was carbon 13 instead of carbon 12) we were able to determine that the mass of hydrogen in the flow is about 30 times the mass of the sun. A simple calculation (half the mass multiplied by its velocity squared) shows that the gas is carrying about 2×10^{47} ergs of kinetic energy, equivalent to the total radiant output of the sun for 1.6 million years. Obviously a very energetic event has recently occurred in the vicinity of AFGL 490. How might one account for the observed activity?

Conceivably the blue- and red-shifted flows might represent the approaching and receding edges of a rotating disk of gas viewed edge on. To be stable against centripetal disruption, however, the disk would have to consist of at least 20,000 solar masses of material within the .6-parsec diameter of the high-velocity emission. Our observations indicate that within this region the total amount of material (including both the gas in the high-velocity flow and the ambient low-velocity gas) has an upper limit of 100 solar masses. In addition infrared observations of the embedded central object, that is, AFGL 490 itself, suggest that its mass is no more than 15 times that of the sun. The most plausible explanation for the blue- and red-shifted high-velocity flows is that they are not gravitationally confined. Instead they are somehow being driven outward by the central object in two streams, one stream aimed roughly in our direction and the other aimed away from us. Such a bipolar outflow would not require a large concentration of mass in or around the central object.

The energetic gas-flow phenomena observed in the Orion molecular cloud and AFGL 490 have since been detected elsewhere. At this writing 24 other examples have been identified. Of these about a dozen have been mapped and nearly all have a bipolar structure.

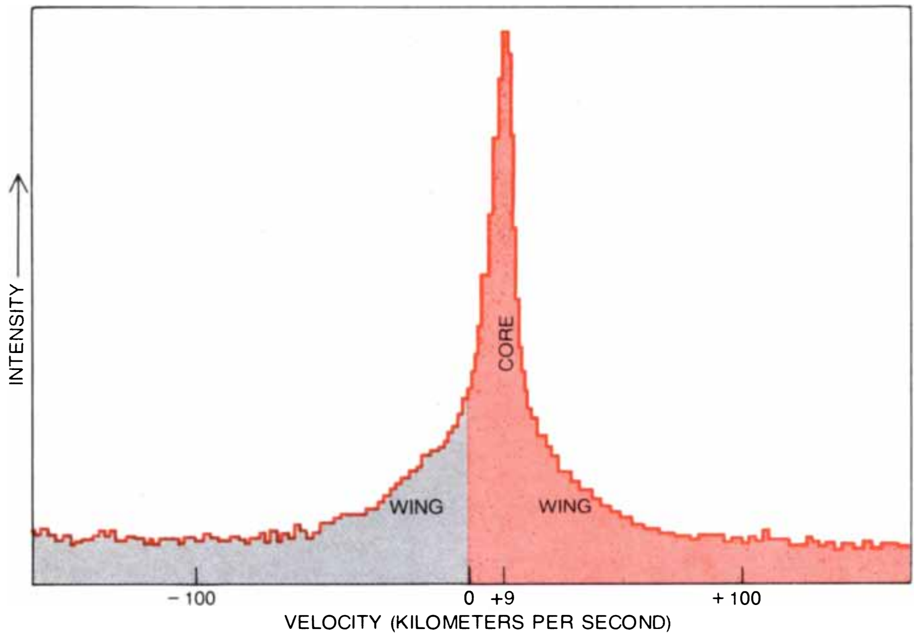
The Molecular Jets of L1551

Perhaps the most remarkable example of bipolar molecular outflow is associated with an infrared source in the dark molecular cloud L1551 in the constellation Taurus. Early observations suggested something peculiar about the carbon monoxide profiles in the neighborhood of the source. With the aid of the five-meter telescope on Mount Locke, Ronald L. Snell of the University of Massachusetts at Amherst, Robert B. Loren of the University of Texas at Austin and Richard L. Plambeck of the University of California at Berkeley examined the source more closely. They were the first to discover molecular gas flowing outward from a young star in a bipolar pattern.

The pattern they observed is unusual in two respects. First, it is the only molecular-cloud flow system so far that has been completely resolved by the observations. The lengths of the flows of high-velocity gas are longer than their widths in the ratio of about three to one, suggesting that the flows are narrow and jetlike. Second, the jet that is blue-shifted coincides with three visible patches of nebulous emission believed to be Herbig-Haro objects. Such objects (named for their discoverers, George H. Herbig of the University of California at Santa Cruz and Guillermo Haro of the National Institute of Astrophysics, Optics and Electronics in Mexico) appear to be created by shock waves that arise when strong outflowing stellar winds interact with a surrounding cloud of molecular gas.

Optical observations of emission lines from the Herbig-Haro objects in L1551 made by Stephen E. Strom and Karen M. Strom of the Kitt Peak National Observatory and Gary L. Grasdalen of the University of Wyoming have shown that the lines exhibit a blue shift consonant with the blue shift of the molecular jet in which the objects seem to be embedded. In fact, the Herbig-Haro objects are found to be approaching along the line of sight at a velocity 50 kilometers per second greater than that of the gas in the molecular jet. In separate observations W. J. Luyten of the University of Minnesota has observed two Herbig-Haro objects, H-H 28 and H-H 29, to be moving so fast in a direction transverse to the line of sight that they occupy different positions on photographic plates taken only a few years apart. When Kyle M. Cudworth of the Yerkes Observatory of the University of Chicago and Herbig analyzed plates of the L1551 region made over a 30-year period, they found that H-H 28 and H-H 29 have transverse velocities of some 150 kilometers per second. They have also shown that if the motions of the two objects are extrapolated backward in time, their paths would intersect only 3,000 years ago at a point near the origin of the bipolar outflow of molecular gas.

Presumably the reason Herbig-Haro objects are observed in association with the blue-shifted jet but not with the red-shifted one is that the blue-shifted gas is closer to the near surface of the molecular cloud, generating a shock wave that gives rise to the luminous Herbig-Haro nebulosities as the gas begins to rupture the surface. The red-shifted flow is directed away from us and is much deeper in the cloud, and so any Herbig-Haro objects that may be created in the outflow are completely obscured by intervening molecular gas and dust.



MOLECULAR CLOUD IN ORION overlying a cluster of newly formed stars exhibited this profile of carbon monoxide emission when it was observed by the author and his colleagues with the Multiple Mirror Telescope. The radiation, which has a rest wavelength of .87 millimeter, is emitted when carbon monoxide molecules decay from their third excited rotational state to their second excited rotational state. Overall the cloud is receding at nine kilometers per second. The broad "wings" on the profile show, however, that many molecules within the cloud are moving at supersonic velocities of more than 50 kilometers per second, some approaching, others receding. The high-velocity emission, confined to a region only half a light-year in extent, evidently represents a highly energetic flow of gas from deeply embedded young stars.

Perhaps the most remarkable aspect of the violent activity observed in L1551 is the nature of the object that is presumably responsible. An infrared source near the center of the outflowing gas was originally discovered in 1975 by Stephen Strom and Karen Strom, working in collaboration with Frederick J. Vrba of the Lowell Observatory in Flagstaff, Ariz. It differs markedly, however, from the infrared sources previously discovered in the Orion molecular cloud and in AFGL 490. The sources driving the gas outflows at those two sites appear to be young stars or protostars, probably stars of spectral type *B*, which are at least 15 times as massive as the sun and at least 1,000 times as luminous. In view of their mass and luminosity it is perhaps not surprising that energetic activity is observed in their vicinity.

In contrast, far-infrared and near-infrared observations of the source in L1551 carried out from the ground by Charles Beichman of the California Institute of Technology and with a balloon-borne telescope by a group of European astronomers (led by C. V. Fridlund of the Stockholm Observatory) show that the source is no more than 25 times as luminous as the sun. The central object resembles a star of the T Tauri type, a young stellar object that may be no more than twice as massive as the sun. From this one must conclude that unusually massive and luminous stars are not the only stellar objects capable of energizing a strong bipolar outflow of gas in the early stages of their evolution. Stars no larger than the sun may be capable of inciting the same phenomena early in their lifetime. Of the 26 high-velocity molecular outflows now known, four appear to be associated with stars of relatively low mass.

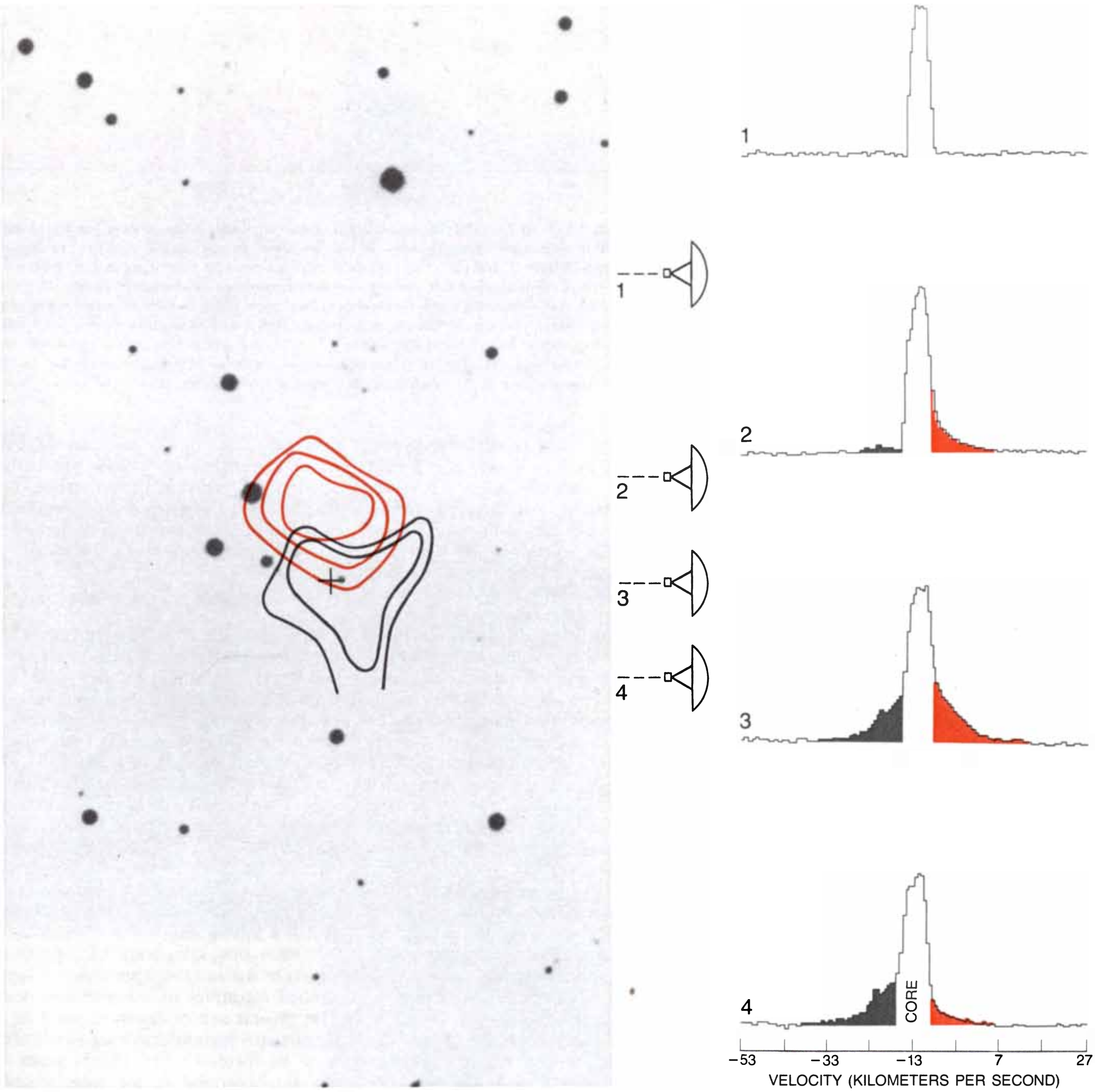
The Frequency of Molecular Jets

The energetic activity exemplified by the Orion molecular cloud, AFGL 490 and L1551 represents a new and extraordinary stellar phenomenon marked by three distinguishing characteristics and often four. The primary feature is a high-velocity outflow of molecular gas. Second, the dynamical activity has a short evolutionary time scale: 1,000 to 10,000 years. Third, large kinetic energies are involved: from 10^{45} to 10^{48} ergs. And fourth, in most instances the flowing gas exhibits a bipolar pattern centered on an embedded infrared source that is a young star.

Within one kiloparsec (3,260 light-years) of the sun there are now 10 well-studied examples of this phenomenon. The objects are evidently created at a remarkably high rate: at least 10 of them must be formed every 10,000 years in the neighborhood of the solar system alone. In fact, the rate of formation of such star-centered molecular jets appears roughly equal to the rate of formation of all stars with masses in excess of three solar masses. Because a systematic search for molecular jets has not yet been made the above estimate of their formation rate is probably a lower limit. A tentative conclusion is that all stars a few times more massive than the sun go through an early evolutionary stage in which they give rise to energetic and massive outflows of gas. Whether or not smaller, sunlike stars go through the same kind of stage is not yet known.

The possible existence of such an evolutionary stage for stars like the sun may throw light on the still poorly understood process of planet formation. The Jovian planets of the solar system (Jupiter, Saturn, Uranus and Neptune) retain their original atmospheres but the terrestrial planets (Mercury, Venus, the earth and Mars) do not. The present atmospheres of Venus, the earth and Mars (Mercury has no atmosphere) have been built up primarily by volcanic venting over the eons of the planets' history. The original atmospheres consisted chiefly of hydrogen and helium, with smaller amounts of methane, ammonia, nitrogen and neon; the present atmospheres are quite different. (The earth's atmosphere was uniquely enriched in oxygen by the metabolism of living organisms.)

What became of the original atmospheres of the terrestrial planets? The prevailing hypothesis is that the primordial atmospheres were violently swept away by a gale of matter flowing outward from the sun. The nature of the wind that cleansed the inner solar system has never been clearly understood. It now seems possible that the terrestrial

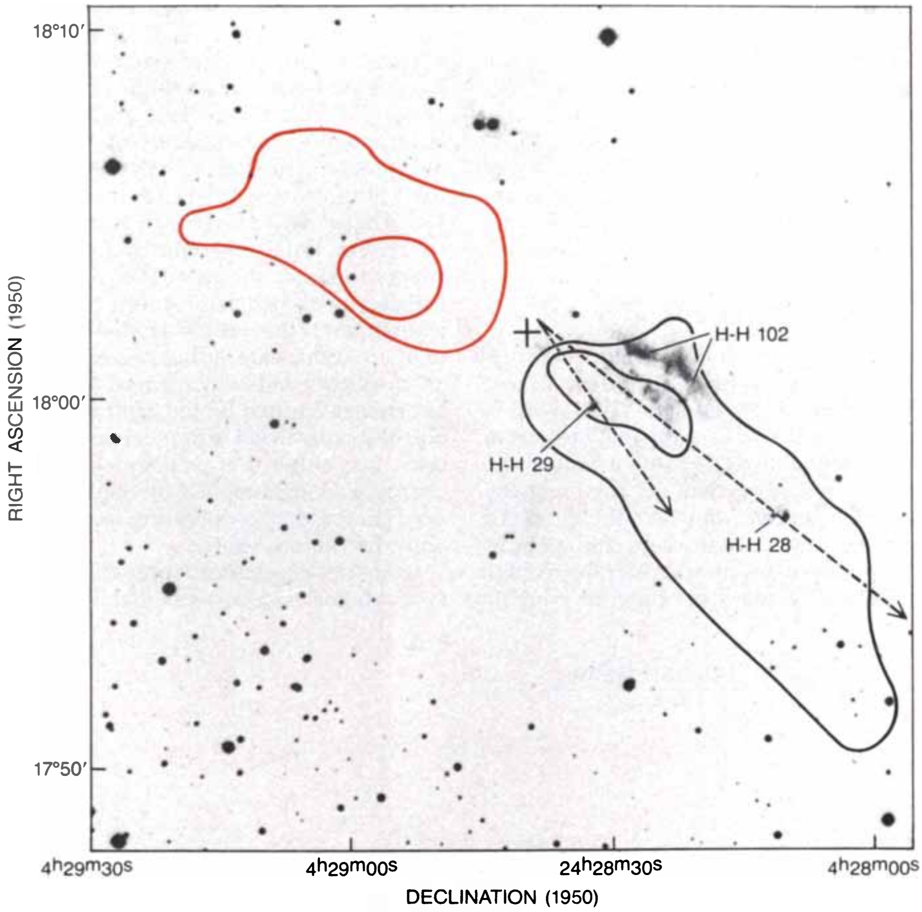


DOUBLE-LOBED FLOW OF CARBON MONOXIDE is observed in the vicinity of the infrared source AFGL 490, which is embedded in a giant molecular cloud some 300 light-years across in the constellation Camelopardalis. The core of the gas that is red-shifted, or receding (*contours in color*), seems to be offset from the core of the gas that is blue-shifted, or approaching (*black contours*). This suggests that the gas is being ejected in two oppositely directed streams from around AFGL 490, indicated by the cross in this photograph made with the 48-inch Palomar Schmidt telescope. The region of double-lobed carbon monoxide emission is about one light-year across. The four profiles at the right show the emission of carbon monoxide at four locations in or near the bipolar molecular flow. The highest gas velocities are found at location No. 3, near the position of AFGL 490 itself. Observations were made by Paul M. Harvey of the University of Texas at Austin and the author with the 11-meter radio telescope of National Radio Astronomy Observatory on Kitt Peak in Arizona.

planets were subjected to the same kind of violent outflow of gas that has been observed in the Orion molecular cloud, AFGL 490 and L1551.

The energetic outflows may also help to explain the puzzling longevity of giant molecular clouds. When spectral line widths of one kilometer to 10 kilometers per second were first observed over large regions in molecular clouds, it was thought the velocities, too high to be thermal motions, could be explained if the clouds were collapsing under the influence of gravitation, which seemed plausible because the clouds are clearly active sites of star formation. The trouble with this idea is that the clouds would have collapsed completely in about a million years, and there are strong grounds for believing the clouds have lifetimes ranging from 10 million to 100 million years. Moreover, if molecular clouds were all in a state of collapse, the rate of star formation in the galaxy would be considerably higher than the rate that is actually observed. Something has been preventing molecular clouds from collapsing and in so doing has poured enough mechanical energy into the gas to maintain line widths of several kilometers per second over the entire cloud.

With the discovery of high-velocity molecular outflows around young stars in dark clouds the mystery may be solved. If every *B* star gives rise to the kind of molecular jets observed in the Orion molecular cloud and AFGL 490 in the course of its early evolution, the observed rate of *B*-star formation in subgroups of *OB* associations may be adequate to account for the mechanical energy that keeps 100,000 solar masses of molecular gas moving at average velocities of two kilometers per second for as long as 10 million years.



OPPOSITELY DIRECTED JETS OF GAS are cleanly separated in the vicinity of the embedded infrared source L1551 in the constellation Taurus, marked by a cross near the center of this photograph made with the 48-inch Palomar Schmidt. As in AFGL 490, the streams evidently arise from the infrared object, which is presumed to be a young star. The blue-shifted stream (*black contours*) seems to include three nebulous patches of visible emission designated as Herbig-Haro objects: H-H 28, H-H 29 and H-H 102. The first two are moving so rapidly (150 kilometers per second) that their position has changed in photographs made only a few years apart. The direction of their motion is shown by the arrows. By extrapolating the motions backward one finds that the objects were ejected from a region close to L1551 only 3,000 years ago. Map of carbon monoxide emission was made by Snell, Robert B. Loren of the University of Texas at Austin and Richard L. Plambeck of the University of California at Berkeley.

The Role of Circumstellar Disks

Although the discovery of the high-velocity outflows in molecular clouds seems to solve some old problems, it raises new ones just as puzzling. The extraordinary energy of the outflows and their high frequency of occurrence are not easily reconciled with current views of the formation and early evolution of stars. The new evidence that energetic physical processes accompany the creation of stars more massive than the sun (and perhaps in addition stars no more massive than the sun) is neither predicted nor explained by current theories. What mechanism could be responsible for the high-velocity outflows around young stars?

Any explanation must be able to account for two important features: the enormous energy embodied in the outflows and, in several significant examples, the collimation, or focusing, of the flows into bipolar jets. It is not necessary to assume that the bulk of the flowing material has been ejected from the central source. In all probability it has not been ejected. More likely the bulk of the outflow consists of molecular-cloud gases swept up by a small volume of material ejected from the young star.

The bipolar, or double-lobed, distribution of the flow may be easier to explain than the energy source. If, for example, the outflow were initially spherical, it could be channeled into a bipolar flow if the central object were embedded in a disk of gas. The outflowing gas would meet much less resistance going through the poles of the disk than it would going through the equator, and as a result the flow would be channeled in two opposite directions.

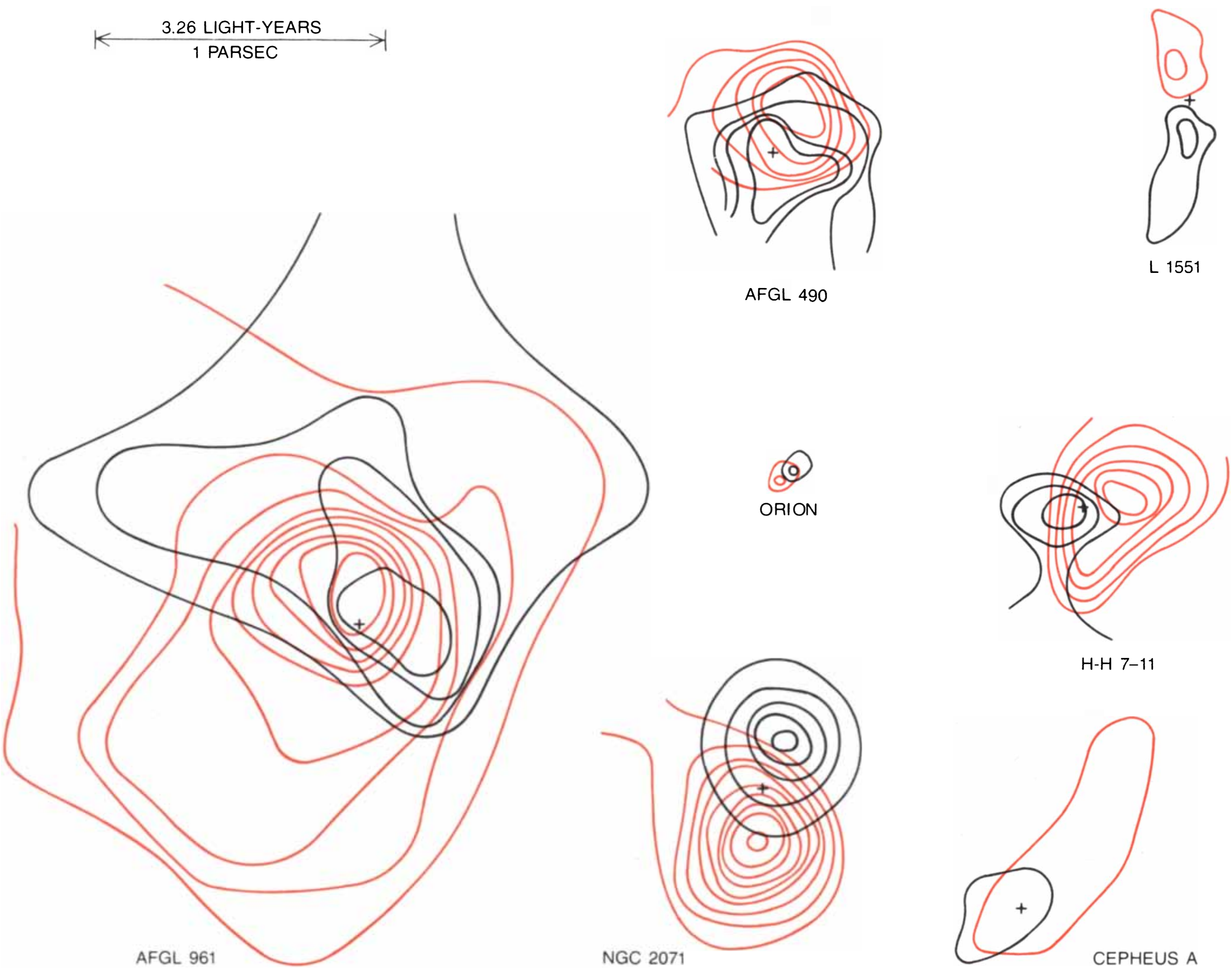
Is there any evidence to suggest that such disks might form around young stars and protostars? There is, in fact, considerable indirect evidence. Our own sun is an example. According to present views, the planets and the sun were formed together out of the same collapsing protostellar cloud. Evidently the collapsing cloud took the form of a thin disk, so that when the planets ultimately condensed, all ended up more or less in the same plane very close to the equatorial plane of the sun, where they remain in orbit today. On an immensely larger scale, the collapsing cloud that formed the galaxy itself took the form of a thin disk (as is indicated by the narrowness of the Milky Way when it is seen on a clear dark night). Indeed, theoretical calculations show that when a rotating sphere of gas collapses, it inevitably forms a rotating disk.

Observational evidence for a disk around a star with a peculiar infrared spectrum, the star MWC 349 in the constellation Cygnus, has recently been reported. Rodger I. Thompson and Peter A. Strittmatter of the University of Arizona and Edwin F. Erickson, Fred C. Witteborn and Donald W. Strecker of the Ames Research Center of the National Aeronautics and Space Administration have proposed that the star's spectrum is most plausibly produced by a circumstellar disk. Comprehensive spectral-line studies in the infrared have yet to be made of the hidden young stars associated with any of the molecular jets, but data that are available for two infrared sources, AFGL 490 and the closely similar AFGL 961, suggest the presence of a disk around the central object. The results are far from conclusive at this stage, but the belief is growing that most recently formed stars are girdled by a disk of gas and dust.

Nature of the Energy Source

The source of energy sufficient to support the high-velocity outflows is more perplexing. In the case of AFGL 490, 30 solar masses of outflowing molecular gas is somehow driven by a central object of no more than 15 solar masses. In other words, whatever the energetic process may be, it musters enough power to sweep up and accelerate to high velocities a mass of material roughly double the central mass. On the basis of far-infrared observations made at an altitude of 40,000 feet from NASA's Kuiper Airborne Observatory, Harvey, Murray F. Campbell and William F. Hoffmann of the University of Arizona have calculated that AFGL 490 has a total luminosity 1,400 times that of the sun. Therefore it would take 1,000 years for AFGL 490 to emit enough luminous energy to match the amount of mechanical energy carried in the outflowing molecular gas. If the outflow has been going on for 10,000 years, as has been estimated, it follows that 10 percent of the radiant energy emitted by the central object has been converted into mechanical energy. It is either that or the mechanical energy is being supplied by another energy source that is operating in parallel with the radiant source.

Other possible energy sources fall into two categories: a process that involves one or more explosive ejections of gas from the central object or a process that yields a steady outflow of gaseous material from the central object. If a single explosive event gave rise to the jets, the original energy of the explosion would have to be much greater than the kinetic energy now observed in the high-velocity molecular gas. The reason is that most of the initial energy of the blast would have been radiated away in shock waves generated by the ejected material. Only a fraction of the energy could have been directly deposited in the form of bulk mechanical energy. Nevertheless, a supernova explosion buried in a molecular cloud could easily deliver enough energy to explain the outflow. The weakness in this hypothesis is that the birth rate of supernovas in the vicinity of the solar system is some 20 times lower than the birth rate of bipolar molecular outflows.



SEVEN HIGH-VELOCITY JET SYSTEMS around seven young stellar objects are mapped to the same scale. The contours in color show red-shifted carbon monoxide emissions; the contours in black show blue-shifted emissions. The carbon monoxide emission from each source originates in two lobes centered on an infrared object (*cross*). The map of Orion was made by Erickson, Goldsmith, Huguenin, Snell, Ulich and the author with the Multiple Mirror Telescope. The maps of AFGL 490 and AFGL 961 were made by Harvey, T. N. Gautier of the University of Arizona and the author. The map of NGC 2071 was made by John Bally of Bell Laboratories. The map of H-H 7–11 was made by Snell and Suzan Edwards of Smith College. The map of Cepheus A was made by Luis F. Rodriguez of the Mexican Institute of Astronomy, Paul T. P. Ho of Berkeley and James M. Moran of the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory. The workers who made the map of L1551 are listed with illustration on preceding page.

Observational evidence provides a more direct argument against the single-explosion hypothesis. With the 11-meter radio telescope on Kitt Peak I have made sensitive observations of the outflowing jets in AFGL 490 and found that the gas of highest velocity covers only a small fraction of the total area of the gas in motion. Evidently the high-velocity flow is decelerating strongly as it moves away from the source. One would also expect material ejected in a single explosion to decelerate as it collides with the surrounding cloud; in that case one would not observe a range of velocities across the jet because all the material swept up in a single explosion would decelerate at the same rate and would have the same velocity at present. Another objection to the single-explosion hypothesis is that a single great burst of energy might very well disrupt the circumstellar disk necessary to keep the flow bipolar.

Continuous Energy Sources

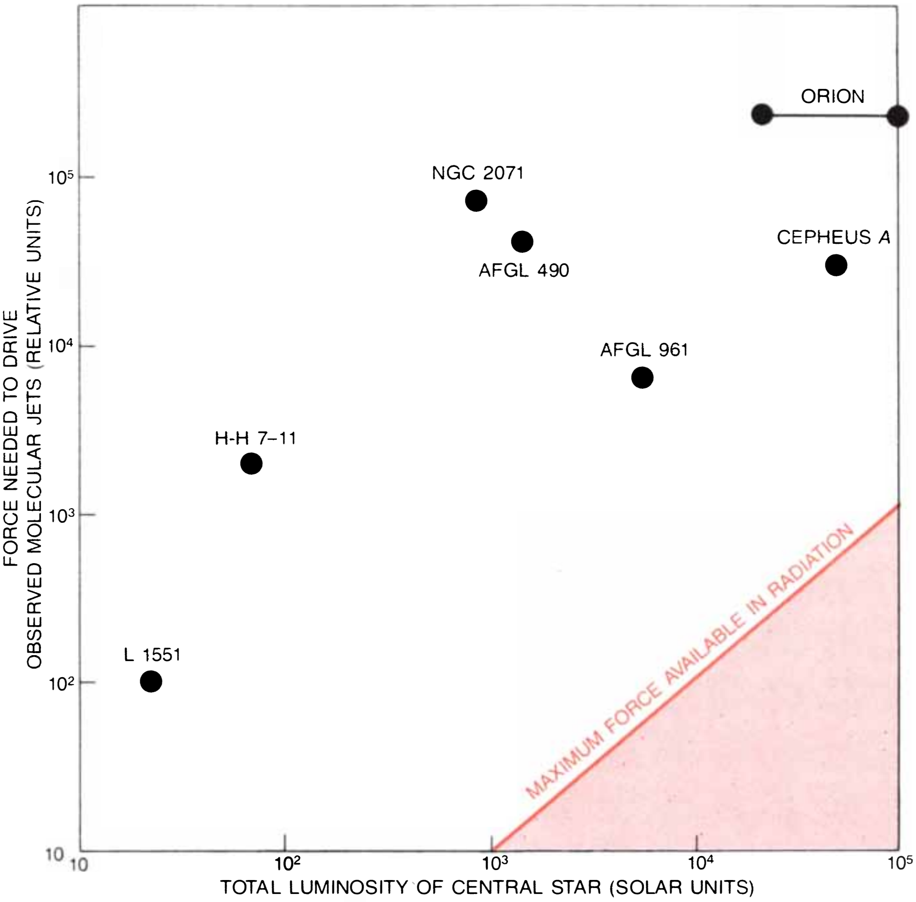
For these reasons either a steady outflow or a series of explosive bursts provides a better explanation for the observations. Although a succession of small bursts cannot be ruled out, a steady outflow seems more probable. There are precedents in astrophysics for such a phenomenon. Most stars, including the sun, are known to be losing mass in the form of a stellar wind. For the sun at present the rate of loss is very small, about 10^{-14} of the sun's total mass per year. The momentum and energy in such a gentle wind are far too small to drive the outflow observed in molecular jets. It seems probable, however, that when the sun was young, its rate of mass loss was much higher, perhaps as much as 10^{-7} of its mass per year. T Tauri stars, objects thought to be progenitors of stars like the sun, are observed to be losing mass at that rate today. Even so, the momentum transferred by such a wind into the surrounding molecular gas would be a factor of 10 short of the momentum needed to explain the outflow around L1551 and a factor of 10,000 short of being able to account for the outflow around AFGL 490 and the Orion molecular cloud.

One final source of energy remains to be considered: the outward pressure of stellar radiation itself. We have already seen that the total kinetic energy carried by molecular jets is equal to no more than 10 percent of the luminous energy emitted by the central stars over the dynamical age of the flows. Could radiation, then, be the driving force of the outflows?

In order to evaluate the effectiveness of radiation as a driving force we must calculate the pressure in a radiation field: the actual amount of momentum carried by the photons. In a collision between a photon and a particle of gas or dust momentum is transferred from the photon to the particle. According to the special theory of relativity, the momentum, p, carried by a photon is related to the photon's energy, E, by the equation $p = E/c$, where c is the velocity of light. The total luminosity, L, of a star is equal to the amount of energy emitted in photons per second. Therefore if every photon is absorbed at least once, the total amount of momentum the star can transfer to surrounding gas and dust per second is just L/c. For example, in the case of hot, giant *O* stars the amount of momentum that can be transferred by photons exceeds the momentum carried by the particles in the stellar winds. Would such intense radiation pressure be sufficient to account for the molecular jets?

John Bally of Bell Laboratories and I recently investigated this possibility by comparing the radiation-pressure force with the force needed to drive the outflow for the best-studied molecular jets. In every case we found that the momentum in the outflowing gas was from 100 to 1,000 times greater than the momentum available from radiation pressure if every photon is absorbed exactly once. Hence radiation pressure does not seem capable of driving the observed outflows of molecular gas.

If neither explosive release of energy nor radiation pressure can account for the molecular jets observed around young stars, what is the mechanism at work? This pivotal question is bound to keep astrophysical theorists busy for some time. Whatever the answer, it may illuminate not only our understanding of the formation and early evolution of stars but perhaps also our understanding of the processes that give rise to the enigmatic jets found near quasars and radio galaxies [see "Cosmic Jets," by Roger D. Blandford, Mitchell C. Begelman and Martin J. Rees; SCIENTIFIC AMERICAN, May]. Infrared and millimeter-wave observations have placed another small piece in the jigsaw puzzle of star formation and evolution. What other pieces remain to be found? How many more pieces will be needed before the puzzle is finally solved? These are exciting problems of the present and future, problems that will be confronted by continued progress on both the observational and the theoretical fronts of astrophysics.



INADEQUACY OF RADIATION PRESSURE to accelerate the seven molecular jet systems mapped in the illustration on the opposite page can be seen in this diagram. The mass of molecular gas in a bipolar flow can often exceed that of 10 suns. The force required to accelerate such a large mass to the observed high velocities of molecular jets is enormous. The diagonal line indicates the maximum force that can be generated by radiation pressure available from the young stellar objects at the center of the flows. Even for stars that emit 10,000 times as much radiation as the sun, or 4×10^{37} ergs per second, radiation pressure falls short of supplying the force by many orders of magnitude. Diagram is based on data compiled by Bally and the author.

DNA Topoisomerases

They are enzymes that convert rings of DNA from one topological form to another, linking them and even tying them in knots. The enzymes may take part in genetic replication and transcription

by James C. Wang

The information content of a DNA molecule is embodied in its sequence of paired nucleotide bases and is independent of how the molecule is twisted, tangled or knotted. In the past decade, however, it has become apparent that the topological form of a DNA molecule has an important influence on how it functions in the cell. Enzymes called DNA topoisomerases, which convert DNA from one topological form to another, appear to have a profound role in the central genetic events of DNA replication, transcription and recombination.

A molecule of DNA is usually thought of as two linear strands intertwined to form a double helix with a linear axis. In many cases, however, DNA has the form of a ring consisting either of a single strand or of two strands wound in a double helix. Moreover, both kinds of DNA ring can be converted to other topological configurations. A single-strand ring can be tied in knots. Two single-strand rings can be linked and twisted to form a double-strand ring. The double helix of a double-strand ring can be unwound a few turns if one strand is cut. The unwinding puts a strain on the molecule; as a result the axis of the double helix itself twists into a helix. In this form the DNA is said to be supercoiled.

It would appear that quite different manipulations of the strands are needed to accomplish such topological conversions. Actually all the transformations are based on the same general mechanism. In each case the ring must be broken temporarily and a segment of DNA passed through the break before the cut ends are rejoined. In 1971 it was discovered that a single topoisomerase molecule can carry out the complete operation of breaking, passage and resealing. It is now known that there are two kinds of topoisomerase. One kind cuts a single strand of DNA; the other kind cuts two strands simultaneously.

Since 1971 many topoisomerases of both kinds have been found in organisms as disparate as bacteria and man. In spite of their ubiquity the topoisomerases might have remained curiosities—biological molecules that do mathematical work—if their manipulation of the DNA molecule had not been found to coincide with important moments in the genetic cycle of the cell. In order for the genetic information of a DNA molecule to be replicated, transcribed into RNA or recombined with that of other genes, a variety of enzymes must gain access to the nucleotide bases. In some topological configurations the base pairs are accessible to such enzymes; in other configurations they are not. For example, supercoiling makes it much easier for the double helix of a double-strand ring to unwind. Unwinding exposes the bases at the core of the helix to the action of enzymes. It seems that in some organisms unwinding is a precondition for DNA replication and for the transcription of DNA into RNA. One topoisomerase is capable of putting a DNA ring into the supercoiled form; another can later uncoil it. Hence it is possible that by regulating the degree of supercoiling the topoisomerases help to control the rate of replication and transcription.

That the topological form of DNA can influence its structure and function began to be suspected only after the discovery of circular DNA molecules. In the early 1960's Robert L. Sinsheimer and Walter Fiers of the California Institute of Technology found that the genetic material of the virus ϕX174, which infects bacteria, is in the form of a ring made up of a single strand of DNA. The development of efficient methods of determining the sequence of bases in a DNA molecule has made it possible to examine every part of the ring; by this means it has been confirmed that the ring is uninterrupted.

Double-strand DNA rings were first observed in 1963 by John Cairns of the Australian National University. It is now known that the DNA of many organisms has the form of a double-strand ring. Such rings are common in bacteria and viruses. They are also found in the mitochondrion, the energy-producing organelle in the cells of higher organisms, and in the chloroplast, where photosynthesis is carried out in plant cells.

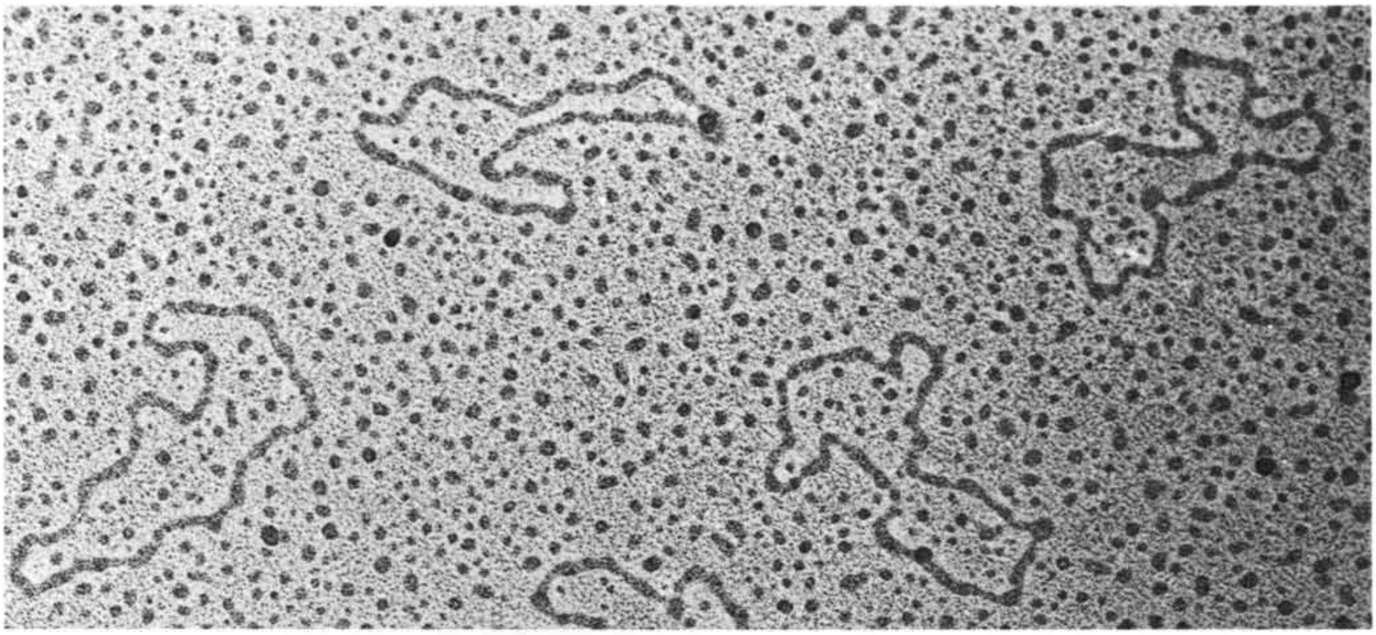
A further level of structure of both single-strand and double-strand rings is defined by their topological configuration; the configurations assumed by a double-strand ring in which both strands are intact are particularly intriguing. A quantity that has an important influence on the configuration of a double-strand ring is the number of times one strand goes completely around the other strand. The quantity is known as the linking number. [For a more precise mathematical definition of the linking number see "Supercoiled DNA," by William R. Bauer, F. H. C. Crick and James H. White; SCIENTIFIC AMERICAN, July, 1980.]

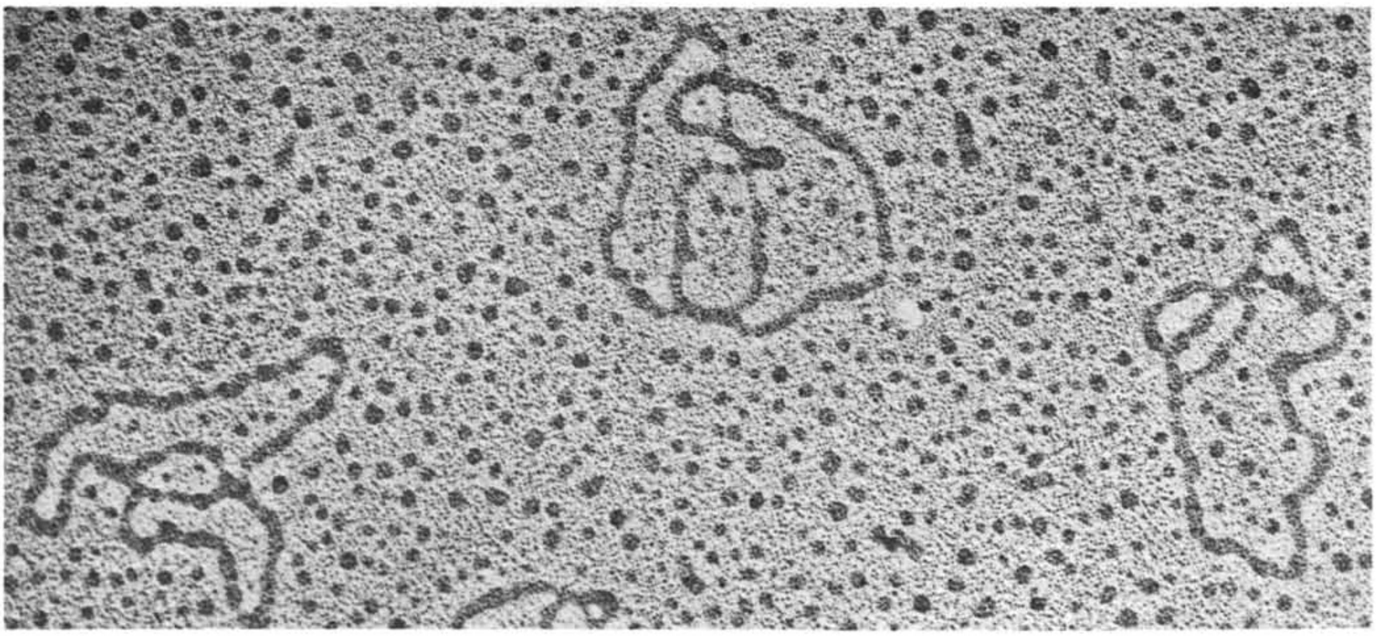
The linking number is a topological property of a double-strand ring. Topological properties are dependent on the spatial arrangement of a structure; they are quite different, however, from geometric properties. Indeed, the essential feature of a topological property is that it remains unaltered when the shape of the structure is changed. No matter how the double-strand ring is pulled or twisted, the linking number does not change as long as the two strands remain unbroken.

In order to understand how the linking number of a ring influences its configuration and physical properties, it is necessary to appreciate the structure of the DNA molecule. Biological molecules have several levels of structure, beginning with the level of the atoms and the covalent chemical bonds between them. (A covalent bond is formed by the sharing of electrons between adjacent atoms; it is the strongest kind of chemical bond.) The atoms of the DNA molecule are arranged in three kinds of groups: five-carbon sugar units, phosphate groups and nucleotide bases. Alternating sugar and phosphate groups linked by covalent bonds make up the backbone of each strand. Each sugar unit in the chain is attached to the phosphate group on one side by the carbon atom designated 5′ and to the phosphate group on the other side by the carbon atom designated 3′. The sugar-phosphate backbone thus has directionality: its length can be traced by proceeding from the 5′ carbon to the 3′ carbon on each sugar molecule or in the reverse direction.

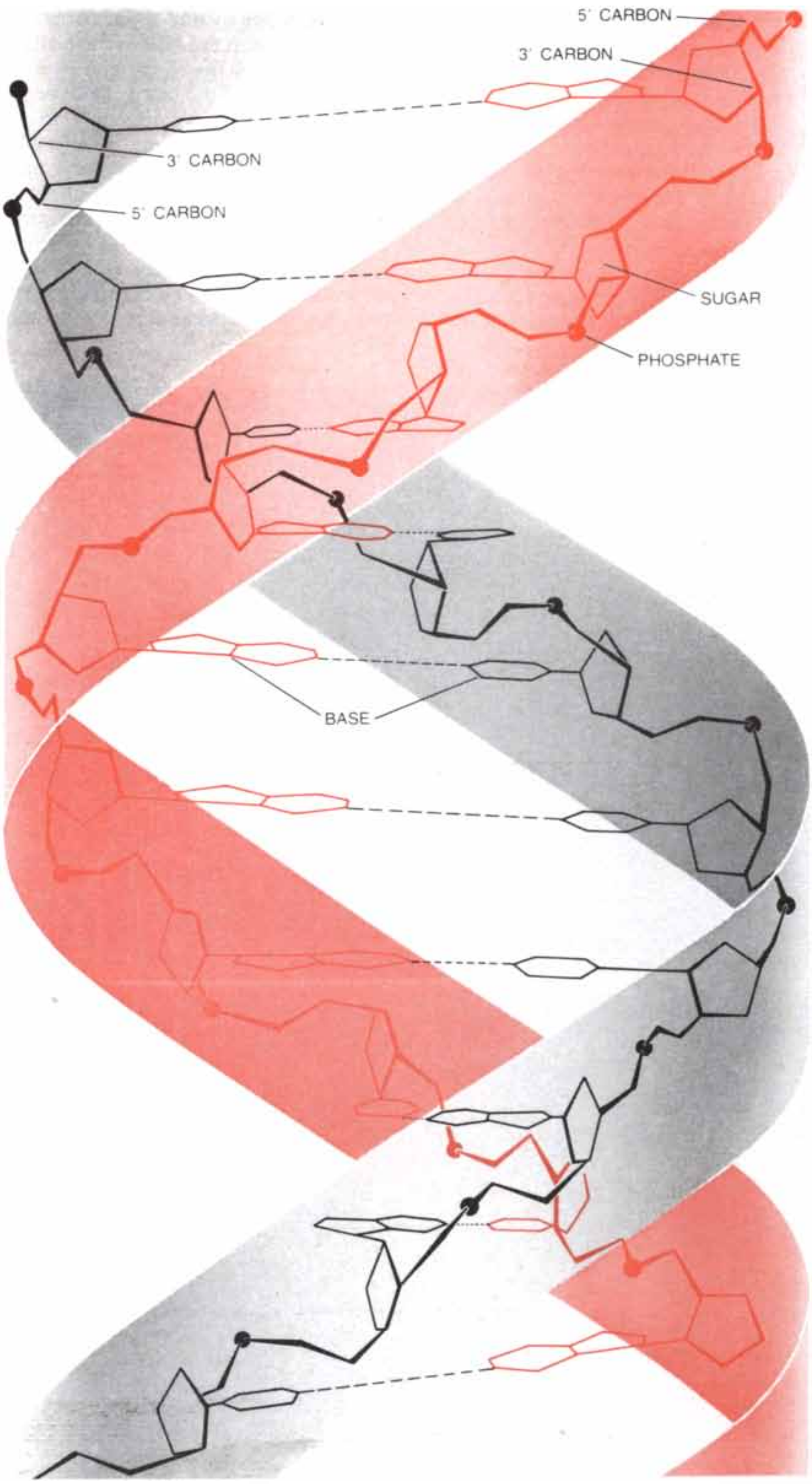
Attached to each sugar molecule is one of the bases adenine, thymine, cytosine and guanine, abbreviated *A, T, C* and *G*. The bases extend from the backbone roughly at right angles. The second level of structure in the DNA molecule is the hydrogen bonding of the bases on two sugar-phosphate chains to form a double helix. (Hydrogen bonding is an attraction between a hydrogen atom attached to an oxygen or nitrogen atom and another oxygen or nitrogen atom.)

If two sugar-phosphate backbones with their attached bases are brought together, the bases of one chain form hydrogen bonds with the bases of the other chain. Because of the shape and chemical composition of the bases *A* can bond only to *T* and *C* can bond only to *G*. Hence wherever *A* appears on the first chain *T* must appear on the second, and wherever *T* appears on the first chain *A* must appear on the second. The same relations hold for *C* and *G*. Strands whose nucleotide bases are able to pair in this way are said to be complementary. In the double-strand molecule the





TYING OF KNOTS in rings of DNA is one of the capabilities of the enzymes known as topoisomerases. The genetic material of many organisms has the form of a ring made up either of one strand of DNA or of two strands twisted in a double helix. The ring can assume a number of topological configurations. The conversion of the DNA ring from one configuration to another is catalyzed by the topoisomerases. The upper electron micrograph shows single-strand DNA rings from a virus known as bacteriophage *fd*, which infects bacteria. The lower micrograph shows the rings after they were exposed to a topoisomerase from the bacterium *Escherichia coli*. By cutting the DNA strand, passing a segment of the ring through the break and rejoining the cut ends the enzyme has tied a knot in each ring; such knotting has been seen only in the laboratory. The process of breaking, passage and resealing is essential to the action of all the topoisomerases. Some of the enzymes, designated type I, cut a single strand of DNA; others, designated type II, cut both strands of a double helix.



DNA DOUBLE HELIX consists of two strands twisted about each other in a right-hand sense. On the outside of the helix are the backbones of the strands. Each backbone is a chain of alternating sugar and phosphate units linked by the strong chemical bonds called covalent bonds. Each five-carbon sugar molecule is attached to the phosphate group on one side by the carbon atom designated 5′ and to the phosphate group on the other side by the carbon atom designated 3′. The backbone is thus directional: its length can be traced from the 5′ carbon to the 3′ carbon on each sugar or in the reverse direction. The strands run in opposite directions. Attached to each sugar and extending into the interior of the helix is one of four bases: adenine (*A*), thymine (*T*), cytosine (*C*) or guanine (*G*). Bases on opposite strands are held together by hydrogen bonds, which are weaker than covalent ones. *A* can form bonds only with *T*, and *C* can form bonds only with *G*. The sequence of base pairs constitutes the genetic information carried by the molecule. The shape and chemical composition of the subunits of DNA determine that in the living cell the helix is most stable when the strands revolve completely about each other once every 10.5 base pairs. The structure shown is the one deduced from observations of crystalline fibers of DNA; it is most stable when the strands revolve once every 10 base pairs.

complementary strands run in opposite directions: one from 5′ to 3′ and one from 3′ to 5′.

The geometry and the chemical composition of the subunits of DNA determine that when hydrogen bonds are formed between bases, the strands assume the form of a double helix with the sugar-phosphate backbones on the outside and the base pairs in the middle. The helix is right-handed: it has the same direction of twist as an ordinary wood screw. The structure of the double helix was first analyzed in crystalline fibers, where it was found that the strands revolve about each other completely once every 10 base pairs; this structure is designated the Watson-Crick *B* helix. Recent results from my group at Harvard University and from that of Aaron Klug at the British Medical Research Council Laboratory of Molecular Biology in Cambridge show that in DNA molecules in solution the strands of the double helix make a full turn every 10.5 base pairs.

As I noted above, the linking number of a double-strand ring is equal to the number of times one strand goes around the other. If a linear DNA molecule in which the strands revolve every 10.5 base pairs is formed into a ring, the linking number is the total number of base pairs divided by 10.5. The DNA of even a small virus has thousands of base pairs; the DNA of a bacterium consists of millions of base pairs. The linking number of a double-strand ring in such organisms therefore ranges from several hundred to several hundred thousand. The alteration of the linking number of a double-strand ring is among the topological operations now known to be carried out by the topoisomerases.

Two molecules that have the same chemical composition but differ in structure are known as isomers. Therefore two double-strand DNA rings with the same sequence of base pairs but different linking numbers are isomers. Because they differ in a topological quantity they are called topoisomers. Hence the name topoisomerase for the enzyme that changes the topological form.

Reducing the linking number of a double-strand ring induces twisting and coiling of the DNA double helix itself. In this form the ring is said to be supercoiled. Supercoiled rings were first noted by Jerome Vinograd of Cal Tech. Vinograd and his colleagues observed that the DNA of the polyoma virus, which infects animals, had physical properties somewhat different from those that would have been expected on the basis of its size alone. The viral DNA is a double-strand ring of about 5,000 base pairs. Vinograd showed that the properties of the ring were the result of supercoiling caused by a reduced linking number.

That the linking number of a DNA

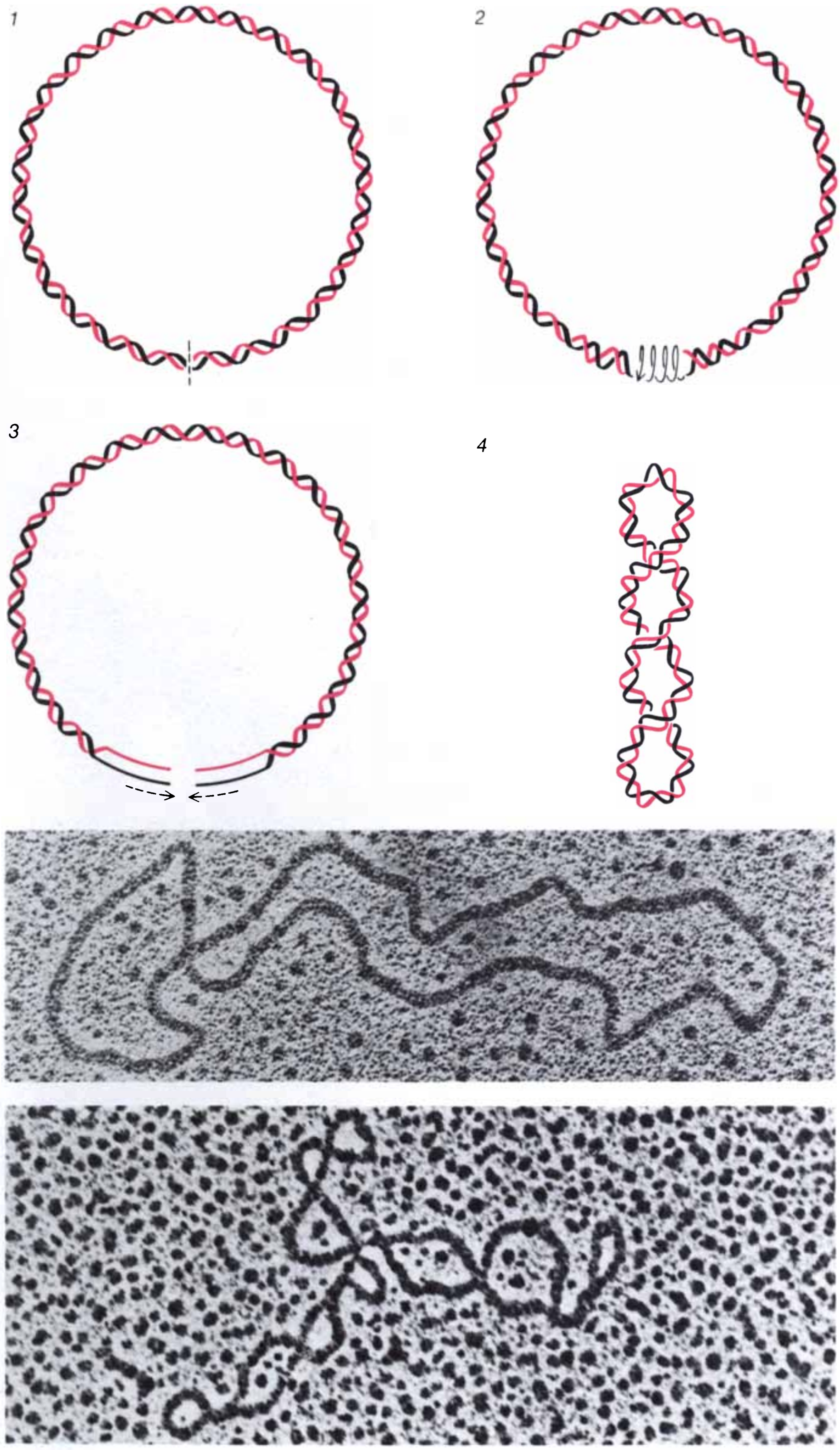
ring, and so its degree of supercoiling, have a strong influence on the physical properties of the ring can be demonstrated by preparing a family of rings that differ only in linking number. A method of preparing such a set of molecules was known even before the discovery of the topoisomerases. One begins with a solution of double-strand rings of known length and any linking number. An enzyme called pancreatic DNase I is employed to cut one strand of each ring. DNase breaks the covalent bond between a sugar group and a phosphate group. (A water molecule is split at the same time; one part of the water molecule is attached to the sugar and the other part is attached to the phosphate group.)

With one strand broken there is no constraint on the linking number of the ring; the cut strand can revolve freely about the intact strand and raise or lower the linking number. In the linear molecule the strands revolve every 10.5 base pairs because that configuration puts the least strain on the double helix; it is the most stable configuration of the molecule. When one strand is cut, the double helix gradually reverts to this form. When the linking number equals the number of base pairs divided by 10.5, the ring is said to be relaxed.

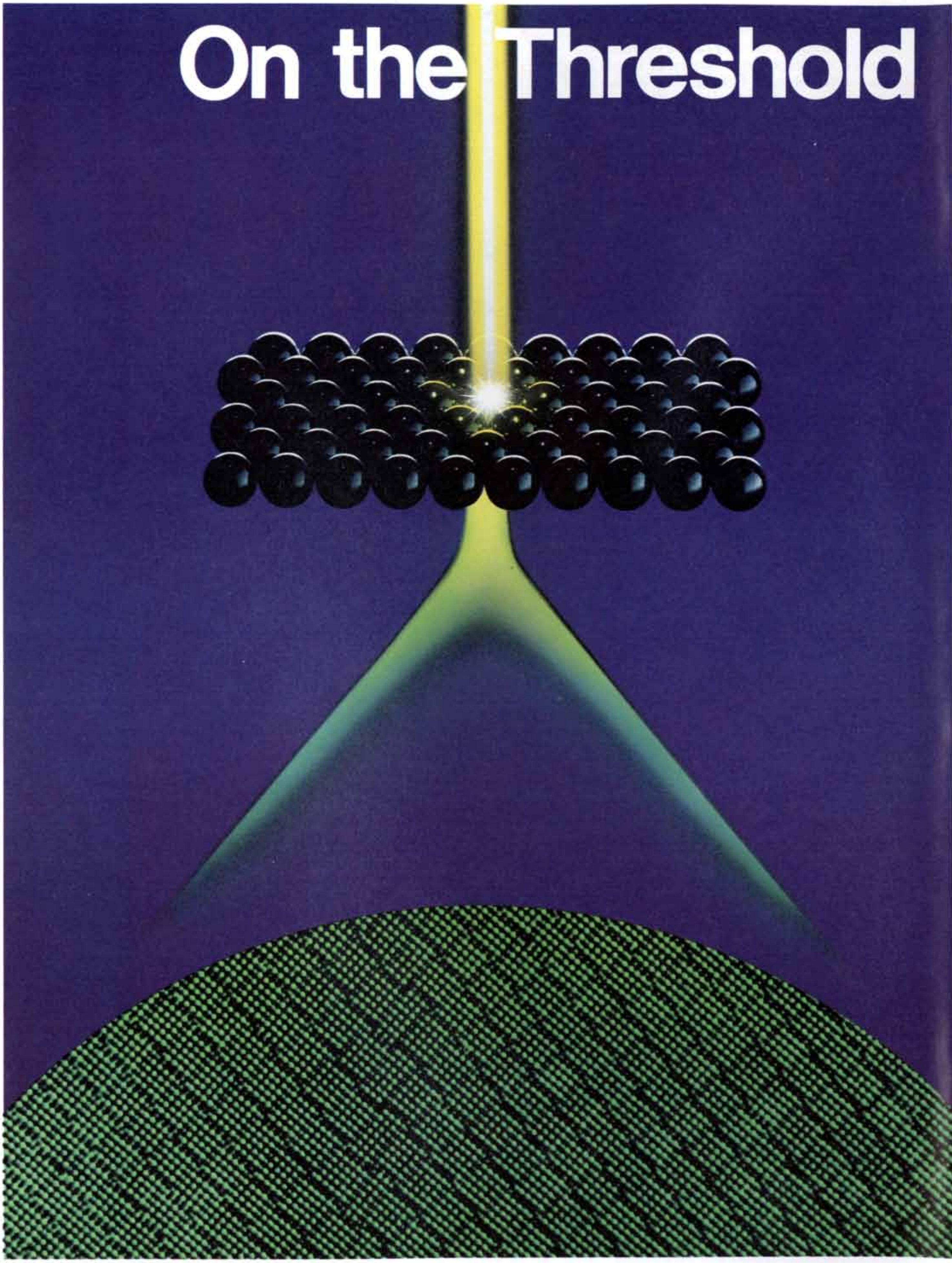
The linking number can be reduced below that of the relaxed state by employing ethidium, a planar molecule that slides between two base pairs when it binds to the double helix. Each molecule of bound ethidium unwinds the double helix by about 26 degrees. Thus 14 bound molecules unwind the double helix one complete revolution, reducing the linking number by one. How much the linking number is reduced depends on the quantity of bound ethidium, which depends in turn on the concentration of ethidium in the solution.

By controlling the concentration of ethidium the linking number of a ring can be reduced by a specified amount. After the ethidium has been bound the enzyme DNA ligase is employed to rejoin the ends of the broken strand. DNA ligase reverses the action of DNase. It reunites the separated portions of the water molecule that was inserted into the sugar-phosphate bond and removes the water molecule, thereby restoring the continuity of the strand. With both strands intact the reduced linking number is sealed into the ring. When the ethidium is removed, the ring supercoils. Supercoiled rings observed by means of the electron microscope appear twisted and more compact than relaxed rings. The more the linking number has been reduced, the more twisted and compact the ring is.

Supercoiling that results from reducing the linking number is called negative supercoiling; raising the linking number causes positive supercoiling. Negatively



SUPERCOILING of a double-strand DNA ring deforms the ring into a more twisted and compact shape. The shape of a DNA ring is strongly affected by the number of times one strand goes around the other; the quantity is called the linking number. It is a topological quantity: it cannot be altered while the strands are intact regardless of how the ring is pulled or twisted. If the strands are cut, however (*1*), and then rotated about each other in the direction opposite to that of the twist of the helix (*2*), the helix unwinds. When the cut ends are rejoined (*3*), the linking number is decreased by the number of rotations that have been made. (The illustration shows the principle of reducing the linking number; the topoisomerases do not necessarily work by the same mechanism.) The strands of DNA in a linear molecule revolve once every 10.5 base pairs because that configuration puts the least strain on the double helix. A ring in which the ratio of base pairs to linking number is 10.5 is said to be relaxed. Increasing or decreasing the ratio strains the double helix, which responds by supercoiling (*4*). Reducing the linking number causes negative supercoiling; raising the linking number leads to positive supercoiling. The upper electron micrograph shows a relaxed DNA ring from a bacterial virus called *PM2*. The lower micrograph shows a negatively supercoiled DNA ring from the same virus.


On the Threshold

# of Atomic Vision

Electron optical lens systems of the past wouldn't do

When demands arose for instruments capable of probing down to the molecular level, scientists devised powerful electron microscopes. But even these have their drawbacks. Imaging techniques limit the clarity and degree of magnification possible. Manual focusing is still a tedious process. And elemental analysis is virtually impossible for anyone but an expert. Without ongoing innovations in microscopy, visual access to the worlds within would be greatly restricted.

Computerized TEM as an answer

Hitachi has brought microscopy to the threshold of atomic vision by developing a computer-controlled electron microscope with a resolving power of just 1.02 Å (about one ten millionth of a millimeter). This incredible resolution, highest in the world for commercial devices, was made possible by applying transmission electron microscope (TEM) techniques: an electron beam, shot from an ultrafine-tipped electron gun and condensed by lenses, penetrates the specimen and emerges to be magnified for projection on a fluorescent display screen. Increasing the gun's accelerating voltage to 200 kV enables observation of thicker specimens, which can be viewed in magnifications of 200 to 1,000,000X. And fully computerized focusing has improved microscopic image clarity, while reducing operator procedures to the mere touch of a button.

Balance of power and precision

The greatest difficulty to overcome in creating the computerized TEM was how to maintain computer control and accuracy while increasing the power supplied. Computers do not ordinarily perform well in combination with high voltage equipment. Their electronic circuits are highly susceptible to interference caused by the generation of electrical noise. But it was possible to achieve this delicate balance of power and precision by drawing on the know-how and experience gained in developing Hitachi's highly advanced computer technology. A capability that has also been applied to the scanning electron microscope (SEM). Today, Hitachi's computerized SEMs and TEMs can be used for a wide range of applications, from genetic engineering and aerospace materials research to discovering clues for criminal investigations.

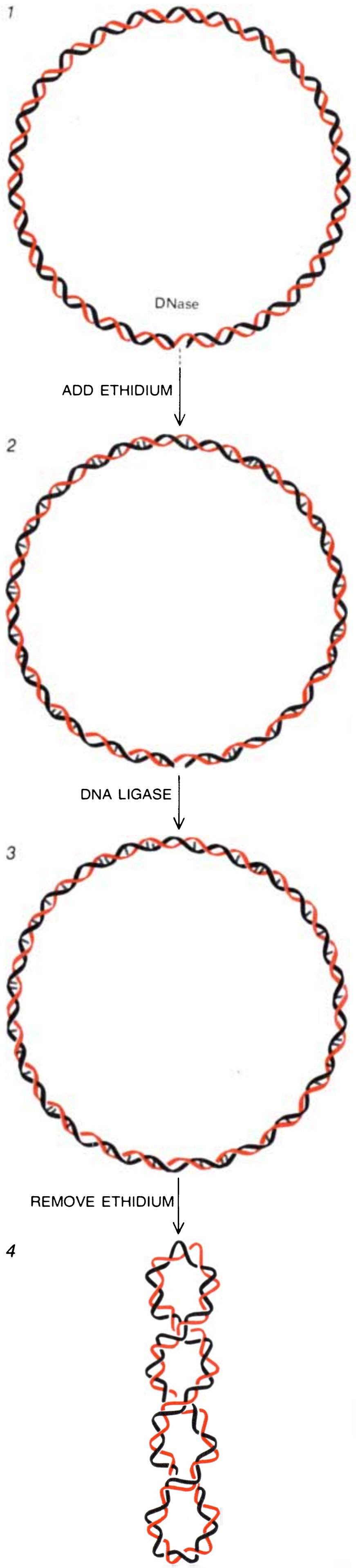
Technical excellence is embodied in all Hitachi products

The development of computerized electron microscopes with greater resolving power is just one case demonstrating Hitachi's technological strength. You'll find other examples in all of Hitachi's products, from spectrophotometers and chromatographs to a whole host of high-quality electronic appliances. Our comprehensive technological expertise is your guarantee of convenience, easy operation and high reliability in every product that bears the Hitachi brand.

The H-800 high resolution analytical TEM system

HITACHI

A World Leader in Technology



supercoiled DNA rings are known to be common in intact cells, but positively supercoiled rings have been made only in the laboratory. It should be noted that if one strand of a supercoiled ring of either kind is cut, the molecule returns to the relaxed state.

Although the supercoiling of a DNA ring may seem implausible, it has analogues in everyday life. If the ends of a piece of rubber tubing are twisted several times and then joined, the tubing can twist on itself to form a helix with untwisted loops extending from both ends. If the receiver of a telephone with a helical cord is lifted and rotated several times before being replaced, the cord can twist into a larger helix.

One technique often used to study supercoiled DNA rings is gel electrophoresis. When DNA molecules are placed on a gel and an electric current is passed through the gel, the molecules move toward the positive electrode because they have a negative charge. The velocity depends on the shape of the molecule as well as on its charge. Because supercoiling makes a double-strand ring more compact the speed of the ring in a gel depends on its linking number. When a current has been passed through the gel for a certain period, the position of the DNA rings can be revealed by staining the gel with a fluorescent dye. When the gel is placed over a source of ultraviolet radiation, the rings appear as bright red bands on a dark background. The technique allows rings with linking numbers differing by one to be readily distinguished.

UNTWISTING OF THE HELIX can be induced by ethidium, a molecule that slides between adjacent base pairs when it binds to the DNA. It can be used to prepare a family of double-strand DNA rings that differ in linking number and therefore have different degrees of supercoiling. The enzyme known as DNase cuts one strand of the ring (*1*). With the cut strand free to rotate about the intact strand the ring reverts to the least stressful state: the relaxed state. Ethidium molecules are inserted into the relaxed helix (*2*). Each bound molecule untwists the double helix by 26 degrees, and so 14 molecules reduce the linking number by one. After the ethidium is bound the enzyme DNA ligase rejoins the cut ends, thus sealing the reduced linking number into the ring (*3*). When the ethidium is removed, the ring becomes negatively supercoiled because of the decrease in linking number (*4*). The degree of supercoiling can be controlled by varying the amount of ethidium in the solution. Double-strand DNA rings that have the same chemical composition but differ in linking number have different shapes and physical properties; they are called topoisomers. Hence the name topoisomerase for an enzyme that changes the linking number and so alters the configuration of the ring.

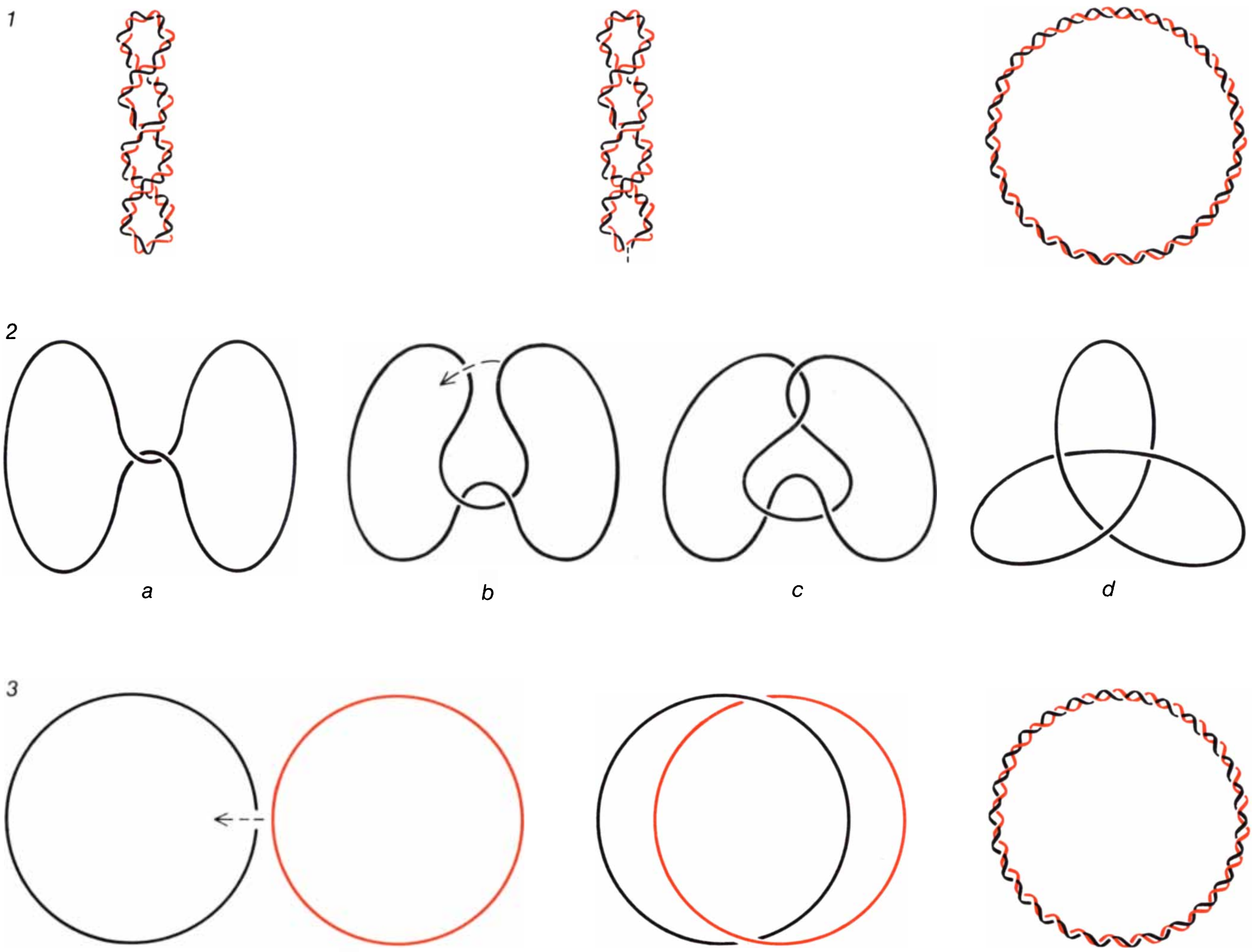
The supercoiling of DNA and the techniques for inducing it in the laboratory had been known for a few years when the first enzyme capable of carrying out such a topological transformation was discovered. The enzyme was found by me in the much-studied bacterial species *Escherichia coli.* It is a type I topoisomerase, one of the family of enzymes that cut one strand of DNA at a time.

The first action of *E. coli* type I topoisomerase to be noted was the relaxation of a supercoiled double-strand DNA ring. The enzyme relaxes negatively supercoiled rings with far greater ease than positively supercoiled rings. Relaxing a supercoiled ring is among the simplest operations of the topoisomerases, requiring only that one strand of the ring be cut and resealed. A few years later it was found that the *E. coli* type I topoisomerase is capable of other topological conversions. It can tie a knot in a single-strand ring and undo the knot. It can also catalyze the interlocking of two complementary single-strand rings to form a double-strand ring.

The knotting of single-strand rings has been investigated by Leroy F. Liu, who was then working in my laboratory; his findings serve to illustrate the techniques by which the action of the topoisomerases is being elucidated. The single-strand rings of DNA employed in Liu's experiments come from the virus called bacteriophage *fd,* which infects *E. coli.* When the viral DNA is exposed to the *E. coli* type I topoisomerase in the presence of a small quantity of magnesium ions, which the operation of the enzyme requires, an intriguing transformation is observed: after exposure to the enzyme the single-strand DNA ring is twisted and compact.

Experimental work suggests that knots have been tied in the ring. Analysis of the untreated viral DNA by gel electrophoresis yields two bands. One band corresponds to the single-strand ring and the other band to the linear molecule that results from accidental breakage of the ring. When the enzyme-treated DNA is placed on the gel, an additional group of bands is seen. The new bands correspond to molecules that migrate through the gel with slightly different speeds. All the new molecules, however, move faster than the original simple rings.

Certain possible explanations of these observations can be eliminated from consideration. The medium in which the electrophoresis was done was quite alkaline. A high degree of alkalinity prevents pairing between the bases attached to the DNA backbone. The new bands in the gel therefore cannot represent structures created by base pairing between segments of the ring. Electron microscopy reveals that the new structures



TYPE I TOPOISOMERASES, which cut one strand at a time, can carry out several topological operations. The first topoisomerase to be found, a type I enzyme of *E. coli,* was discovered by the author in 1971. Topoisomerases are now known to exist in many species, including man. By cutting one strand of a supercoiled DNA ring the type I enzyme can put the ring in the relaxed state (*1*). It can tie a single-strand ring in a knot (*2*). The knot is tied when the single-strand ring crosses over itself. If the two loops formed in this way are pulled together, the enzyme can cut one loop and pass the other loop through the opening. When the break is sealed, the ring is tied in a knot. Structures *2c* and *2d* are topologically equivalent: each form can be made from the other without cutting the strand. The type I enzyme can also interlock two single-strand rings (*3*). If the rings have complementary base sequences, a double helix results. Although the operations seem quite different, each requires that a strand be broken, a segment of DNA be passed through the break and the break be resealed.

are single molecules, and so the differences in speed cannot be a result of the aggregation of several DNA molecules.

One plausible hypothesis is that the new bands correspond to rings that have been tied in knots of varying degrees of complexity. There is a simple way to test this hypothesis. When a ring has been tied in a knot, a single cut anywhere in the ring yields a linear strand regardless of the complexity of the knot. If the new bands created by treating bacteriophage *fd* DNA with the topoisomerase are the result of knotting, cutting the rings once should yield a single band in the position occupied by the linear molecules in the original gel. When the DNA in the new bands is treated very lightly with pancreatic DNase I, a single band is indeed seen in that position.

What is the topological and chemical mechanism by means of which the *E. coli* type I topoisomerase can tie knots in a single-strand ring? It appears that complementary segments of the single-strand ring play an important role in the knot-tying reaction. Such complementary segments could form a double helix in one part of the ring. If the two loops extending from the ends of the helix were pulled together, the enzyme could cut one loop and pass the other loop through the opening before the break was closed. The complexity of the resulting knot would depend in part on the length of the helical region. If more than one helix had been formed in the ring, the complexity of the knot would also depend on the ways the various loops had become linked.

It is known that several topoisomerase molecules can work simultaneously on a single ring. If several cuts are made in the ring at the same time, however, the loose ends could diffuse away from each other. The ends could then be rejoined incorrectly, so that the sequence of bases in the ring would be changed. Whether the ring is rearranged can be tested directly by determining the sequence of bases in the ring before and after the enzyme has worked on it. Such a determination demonstrates that the sequence of bases remains unchanged even when many enzyme molecules act on the ring at the same time. The finding implies that when the ring is cut, the cut ends of the DNA strands are not free to drift apart.

It is possible that the breaking of a strand is always done in a part of the ring where complementary bases have formed a double helix; the unbroken complementary strand could serve to keep the cut ends together. It is also possible that the enzyme itself somehow holds the ends together.

Until more is known about the structure of the topoisomerases it will be difficult to resolve detailed questions of mechanism. Like other biological molecules, the topoisomerases have several levels of structure. So far only the chemical composition has been determined, and that only for *E. coli* type I topoisomerase. The enzyme is a chain of about 895 amino acids. The sequence of amino acids has been determined by finding the sequence of bases in the bacterial DNA that codes for the enzyme. Almost nothing is known, however, about the three-dimensional form the enzyme assumes in the bacterium.

As I mentioned above, all the reactions catalyzed by the type I topoisomerases appear to require the same sequence of actions: the enzyme must break the DNA backbone, a segment of DNA must pass through the break and the enzyme must then reseal the break. There are a number of ways the three operations might be accomplished. One possibility is that a topoisomerase incorporates into a single molecule the capabilities of the two enzymes DNase I and ligase: a topoisomerase might break the covalent bond between a sugar group and a phosphate group by the insertion of a water molecule and restore the bond by removing the water molecule.

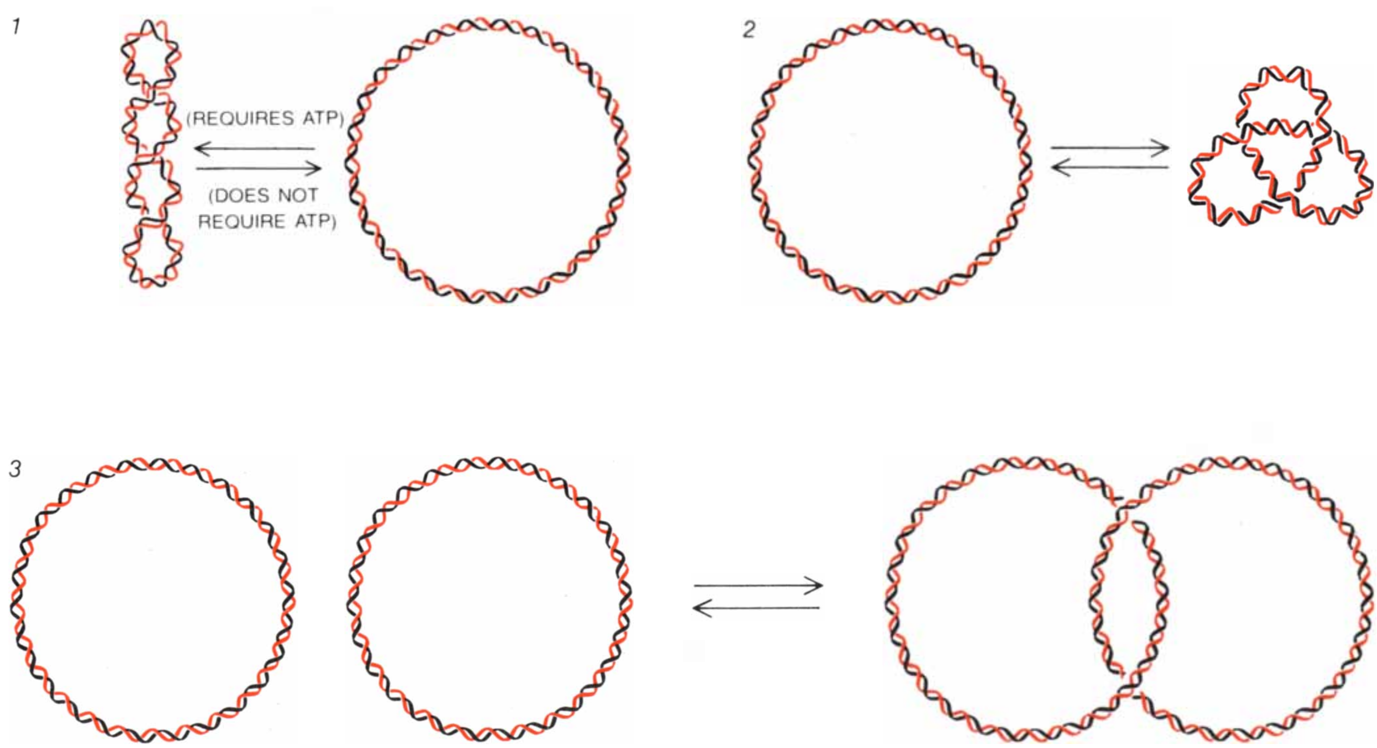
The energetics of the combined mechanism of DNase and ligase, however, suggest that this is not how the topoisomerases work. The making and breaking of chemical bonds can either absorb or release energy. A particular reaction may proceed in either direction, depending on its energetic balance. An enzyme does not determine the direction of a reaction; it merely accelerates it in the favored direction.

The breaking of the sugar-phosphate bond by DNase is favored energetically. The reaction can take place spontaneously; DNase merely accelerates it. The re-forming of the sugar-phosphate bond by ligase, however, has an unfavorable energy balance and does not occur spontaneously. DNA ligase can re-form the bond only by means of a subsidiary reaction with a molecule called a high-energy cofactor. The splitting of the cofactor provides the energy needed to reseal the sugar-phosphate bond. DNA ligases in bacteria utilize nicotinamide adenine dinucleotide (NAD) as a cofactor. Ligases from other organisms utilize adenosine triphosphate (ATP).

Type I topoisomerases do not require a cofactor to work. It is therefore unlikely that the resealing of the DNA backbone by a topoisomerase is done by the mechanism of ligase. The absence of a cofactor led me to hypothesize that a type I topoisomerase breaks and then rejoins the sugar-phosphate DNA backbone by a quite different mechanism. Instead of adding parts of a water molecule to the sugar and the phosphate groups the topoisomerase could itself be split, with a small part of the enzyme molecule, a hydrogen atom, becoming attached to the sugar and the remainder of the enzyme molecule becoming covalently bound to the phosphate group.

If this is the mechanism of the type I topoisomerases, the transitory break in the DNA strand must be accompanied by the formation of a covalent complex of DNA and enzyme. For several years attempts to find such a complex were unsuccessful. Recently, however, it has been shown that adding a substance that disrupts the folding of proteins to a mixture of DNA and a topoisomerase often yields such covalent complexes. The nature of the bond between the enzyme and the strand of DNA has been elucidated for three bacterial topoisomerases. In all of them a phosphate group at the free 5′ end of the strand is linked by a covalent bond to a unit of the amino acid tyrosine in the enzyme.

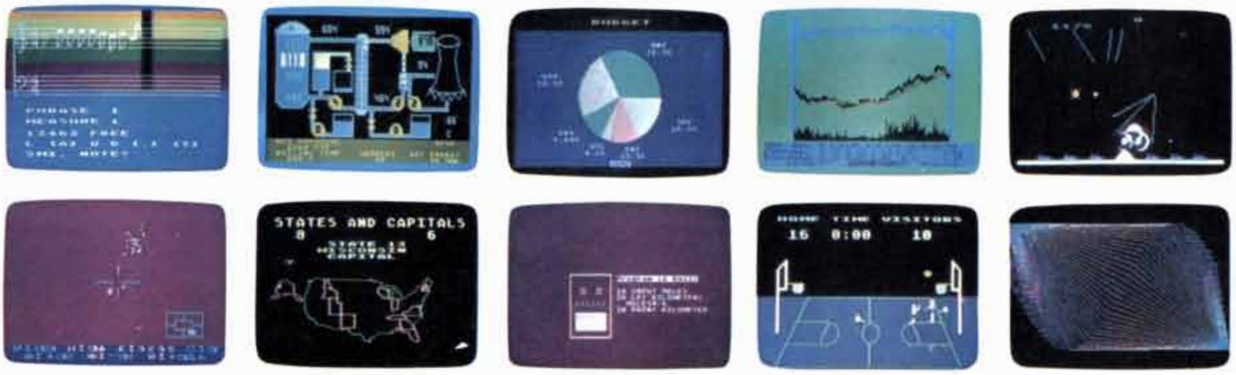
It was not until several years after the discovery of *E. coli* type I topoisomerase that the first type II topoisomerase was found; the discovery was made by Mar-



TYPE II TOPOISOMERASES, which cut two DNA strands at once, act only on a double-strand ring. The first type II enzyme to be found was the one called DNA gyrase; it was discovered in *E. coli*. DNA gyrase can negatively supercoil a relaxed double-strand ring in the presence of ATP (*1*). It can relax a negatively supercoiled ring without ATP, but the reaction proceeds slowly. The type II enzyme can tie and untie knots in a double-strand ring, as the type I enzyme does in a single-strand ring (*2*). Many topoisomerase molecules can work on a ring at once. If the cut ends were free to drift apart during the knot-tying process, the base sequence of the ring might be rearranged when the breaks were sealed. No rearrangement has been seen. When the ring is cut, the enzyme forms a covalent bond to the DNA strand; the bond prevents the cut ends from separating. DNA gyrase can also interlock two double-strand rings and separate interlocked rings (*3*).

THE GRAPHIC DIFFERENCE

BETWEEN ATARI® COMPUTERS AND ALL OTHERS.

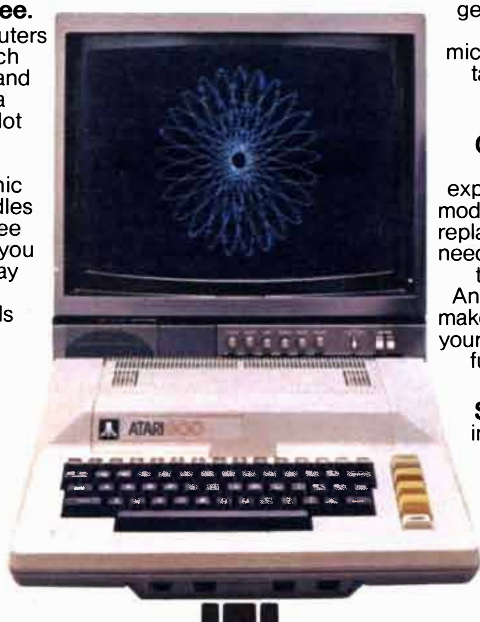


3.7 million reasons why the ATARI Home Computer is something to see.
The display screen used with our computers is composed of 192 horizontal lines, each containing 320 dots. Delivering color and luminosity instructions to each dot for a second requires 3.7 million cycles . . . a lot of work for the normal 6502 processor.

That's why the ATARI computer has equipped its 6502 with its own electronic assistant. It's called ANTIC, and it handles all the display work, leaving the 6502 free to handle the rest. What this means to you is uncompromisingly spectacular display capabilities without loss of computer power needed to carry out the demands of your program.

That's a quality you just don't find in ordinary personal computers. And it's one of the reasons some computer experts say that ATARI computers are so far ahead of their time.

There's more . . . which is what you'd expect from ATARI.
Language. The ATARI Personal Computer uses several programming languages to give the user maximum control of its extraordinary capabilities. PILOT, Microsoft BASIC,* and ATARI BASIC are understood and spoken by the ATARI computer. You'll also find our Assembler Editor cartridge indispensable for machine language programming.



Sound. An ATARI computer has four sound generators, or voices, activated by a separate microchip. This leaves the principal microprocessor chips free to perform other tasks. And you can take full advantage of this capability which is designed for easy programming.

Change. ATARI Home Computers have been designed to make change and expansion easy. The ATARI computer has a modular operating system* that can be easily replaced as new technology develops. If you need it, memory expansion requires no more than inserting additional RAM modules.* And the ATARI ROM cartridge system also makes it easy to change languages. In short, your ATARI computer won't be obsoleted by future developments . . . *beause it already incorporates the future.*

Sharing. To learn more about the amazing capabilities of ATARI computers, visit your local computer store for a demonstration. Or send for our Technical User's Notes, intended for the serious programmer. They are only $27 and contain a lot more information about our computers' special capabilities than most companies could tell. See your ATARI dealer, or send $30 ($27 plus $3 postage and handling), payable to ATARI, to Technical User's Notes, c/o ATARI Customer Service, 1340 Bordeaux Avenue, Sunnyvale, CA 94086.

*ATARI 800™ computer only.

ATARI®

A Warner Communications Company

DIESEL CARS SHOULD BE SEEN AND NOT HEARD.

If you buy an ordinary diesel car, you could end up with a headache.

Because the diesel's reputation for dependability and getting a lot of miles per gallon, is only exceeded by the noise they make along the way.

Which is the beauty of the new Isuzu.

The Isuzu diesel is not only among the highest mileage* diesel cars in the country, it's also engineered to be a quiet diesel car in the city.

With Isuzu's ingenious rubber drive belt, instead of conventional steel. With special Isuzu noise reducing engine mounts. And extensive soundproofing.

So while an Isuzu can give you all the advantages of owning an economical diesel, it gives you all the comforts of owning a luxurious car.

Like fast acceleration. Smooth, even shifts. Responsive handling. But most of all, a nice, quiet ride.

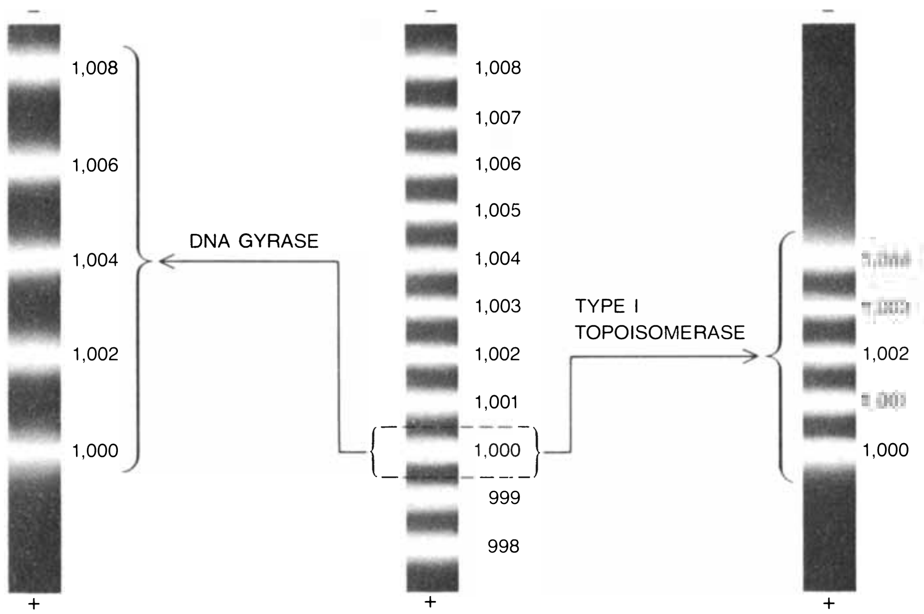
The new I-Mark diesel from Isuzu.

It's the one way you can keep up with the Joneses. Without waking up the Reillys.

ISUZU

*ISUZU DIESEL COUPE [41] EST. MPG, 51 MPG HIGHWAY. USE ESTIMATED MPG FIGURES FOR COMPARISON. YOUR OWN MILEAGE MAY VARY DUE TO SPEED, WEATHER, TRIP LENGTH AND ROAD CONDITIONS tin Gellert of the National Institutes of Health in 1976. The discovery was made in a study of the virus infection of the bacterial cell. The DNA of the virus called bacteriophage λ, which infects *E. coli,* can form a double-strand ring after entering the bacterium. Under certain conditions the viral DNA can then insert itself into the DNA of the host. This reaction is catalyzed by an enzyme called λ integrase that is coded for by the viral DNA. Integrase cuts both the viral DNA ring and the bacterial DNA and joins the cut ends, so that the viral DNA becomes a segment of the bacterial chromosome.

In 1975 Howard A. Nash of the National Institute of Mental Health and Kiyoshi Mizuuchi of the National Institutes of Health accomplished the insertion in the laboratory with extracts of cells that had been infected with phage λ. They observed that if the viral DNA ring is initially relaxed, ATP is required for the insertion. If the ring is negatively supercoiled, however, no ATP is needed. Nash, Mizuuchi and their colleagues Gellert and Mary H. O'Dea hypothesized that the ring must be supercoiled for insertion to occur and that ATP is required as a cofactor by the supercoiling enzyme. These ideas have since been confirmed.



DNA GYRASE lowers the linking number of a double-strand ring by two or a multiple of two. DNA molecules can be separated according to linking number by the method of gel electrophoresis, in which the molecules migrate through a gel toward a positive electrode. The more the linking number of a ring differs from that of the relaxed state, the faster the ring moves. The reason is that the more supercoiled a ring is, the more twisted and compact it is. By a photographic process the molecules can be made to appear as light bands on a dark background. The position of a band indicates the linking number of the ring. Ethidium can be employed to prepare a family of topoisomers with different linking numbers. If the topoisomers are put on a gel, they form a group of narrowly spaced bands (*center*). Cutting out one band isolates rings with the same linking number, for example 1,000. If the slice of gel is treated with *E. coli* type I topoisomerase, the enzyme relaxes the rings, raising the linking number in steps of one (*right*). If the slice is treated with gyrase in the absence of ATP, more widely spaced bands appear, which correspond to rings whose linking numbers have been raised by even numbers (*left*).

In 1976 Gellert and his colleagues isolated the supercoiling enzyme from *E. coli.* It is designated DNA gyrase, and it can convert double-strand rings to a number of topological forms. Gyrases from bacteria are the only topoisomerases known to be able to supercoil a relaxed DNA ring. ATP is required as a cofactor in the supercoiling; the splitting of ATP by means of the insertion and splitting of a water molecule, which releases energy, provides the driving force for the energy-absorbing process of supercoiling. In the absence of ATP gyrase can relax a negatively supercoiled ring, albeit quite slowly.

When gyrase was discovered, it was thought to cut the double helix one strand at a time, as all the known topoisomerases did. In 1977 Gellert's group and that of Nicholas R. Cozzarelli of the University of Chicago began treating solutions of DNA and gyrase with a substance that disrupts the folding of proteins. When the substance was applied, the rings were converted to a linear form. Thus the enzyme clearly breaks both DNA strands at once. Moreover, a part of the enzyme molecule was found to be attached by a covalent bond to the free 5′ end of each strand.

The significance of the cleavage of both strands was not appreciated fully until two further reactions catalyzed by gyrase were discovered. In 1980 Leroy Liu, Chung-Cheng Liu and Bruce M. Alberts of the University of California at San Francisco found that a topoisomerase coded for by the DNA of the bacterial virus T4 can tie knots in a double-strand ring; the enzyme can also untie the knots (as gyrase can). At about the same time Cozzarelli and Kenneth N. Kreuzer of the University of Chicago showed that gyrase can interlock two double-strand rings and separate interlocked rings.

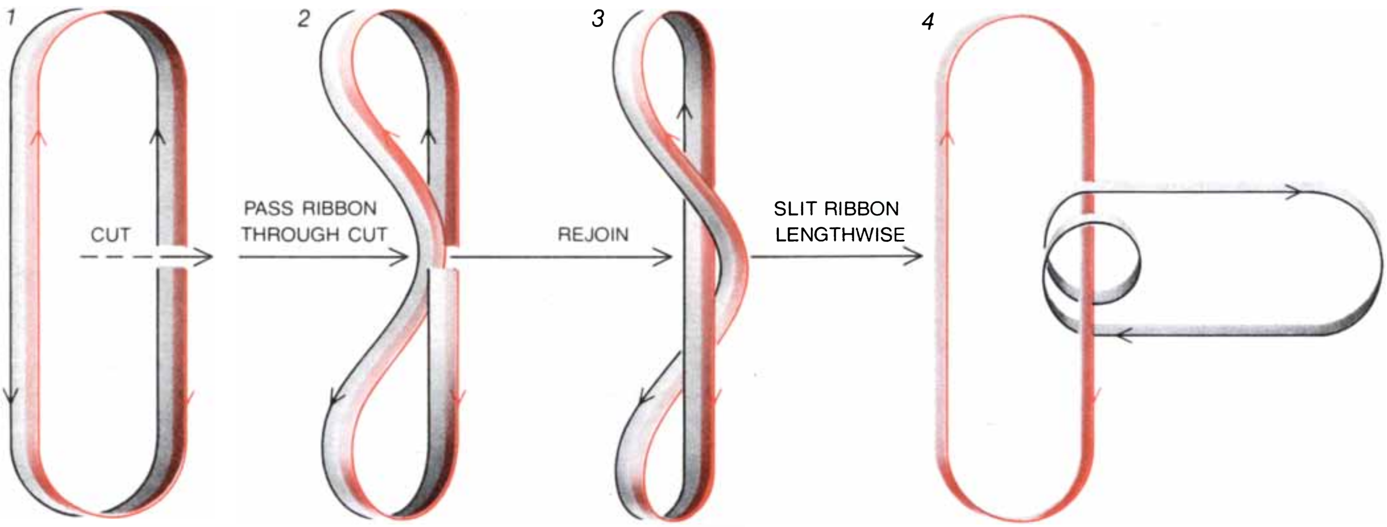
To undo a knot in a double-strand ring it is necessary to break both strands and to pass a segment of DNA through the break. Separating linked rings also requires cutting both strands of one ring. The enzyme might not, however, sever both strands at once. Gyrase could break a single strand, pass the segment into the core of the helix and reseal the cut strand before the second strand is broken. Such a mechanism seems clumsy, however, compared with a transitory double-strand break.

Some of the most convincing evidence that DNA gyrase breaks both strands at once comes from examining the way the enzyme alters the linking number. If negatively supercoiled rings with a linking number of 1,000 are exposed to a type I topoisomerase, the linking number gradually increases to that of the relaxed state in steps of one. If the reaction is interrupted before the relaxed state is reached, a family of topoisomers with a range of linking numbers greater than 1,000 are observed by means of electrophoresis. Such an experiment was first carried out in 1975 by Vinograd and his colleagues. The rings were found to have linking numbers that could differ from the original value by any integer amount and that could differ from one another by as little as one.

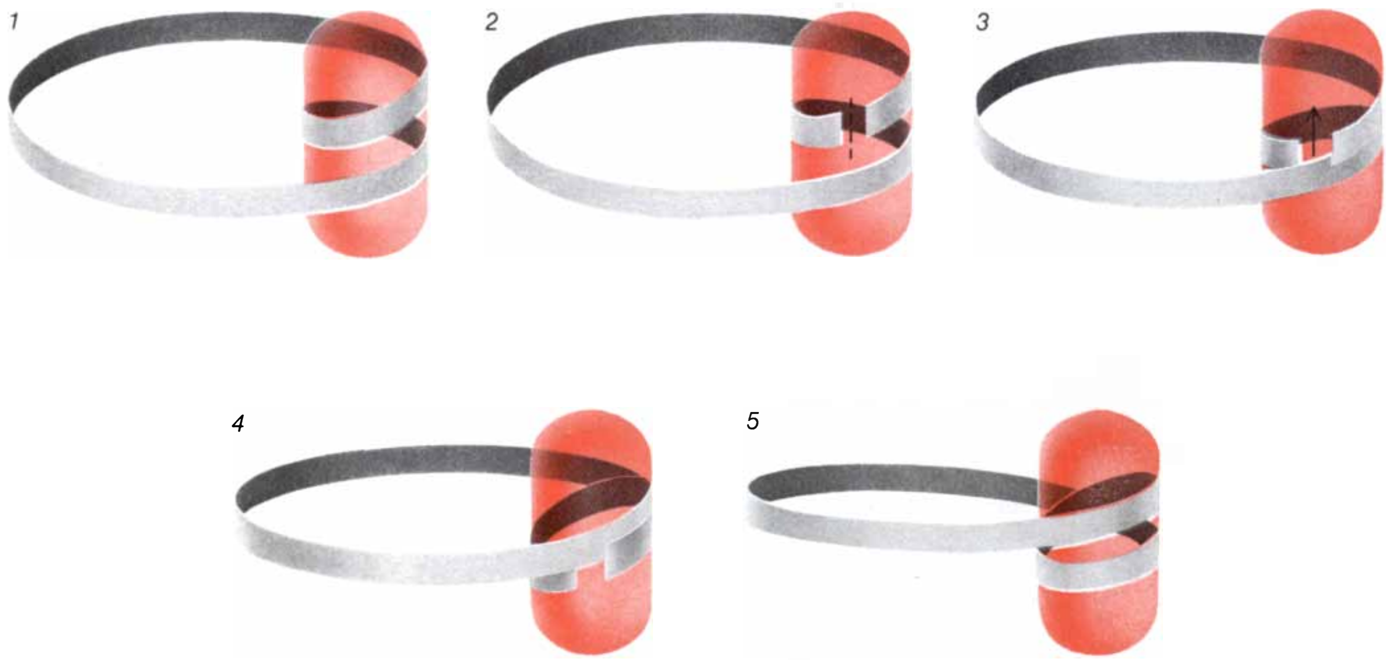
When double-strand rings were exposed to a type II topoisomerase, however, only topoisomers with linking numbers greater by two or by multiples of two than the linking number of the original rings were observed. When the reaction was carried out in the presence of ATP, the enzyme supercoiled the ring rather than relaxing it; again, however, the change in the linking number took place in steps of two.

A simple exercise can serve to illuminate how gyrase could reduce the linking number of a double-strand ring by two or a multiple of two. Consider a linear piece of ribbon. If the ends of the piece are joined to form a circle, the edges take the form of closed circular loops. The edges can stand for complementary strands of DNA, one running from 5′ to 3′ and the other from 3′ to 5′. If the ends of the ribbon are not twisted before being joined, the linking number of the loops traced by the edges is zero. That the number is zero can readily be shown by slitting the ribbon along its longitudinal axis: the halves, each with one original edge, will separate.

Suppose that the circular ribbon is cut transversely rather than longitudinally and the two ends are held close together without being rotated. If a segment of ribbon opposite the cut is passed through the break and the ribbon is resealed, the linking number of the edges changes by two: from zero to two. The increase is confirmed by slitting the ribbon along its longitudinal axis. Doing so yields linked rings, one of which is seen to go twice around the other [*see upper illustration below*]. Because the linking number of the edges was initially zero the exercise is a special case. It can be shown, however, that the linking number changes by two regardless of the initial quantity. This method of causing even-numbered changes in linking number was first noted by the math-



RIBBON-CUTTING EXERCISE can serve as a simplified model of the way DNA gyrase changes the linking number of a double-strand ring by multiples of two. The ends of a ribbon are joined without being twisted to form a simple loop. The edges of the loop represent strands of DNA running in opposite directions: one from 5′ to 3′ and one from 3′ to 5′. Since the edges do not cross each other, the linking number is zero. The ribbon is slit across its width (*1*). The ends are held together without being rotated while a segment of the loop is passed through the cut (*2*). The cut ends are then rejoined (*3*). If the ribbon is slit along its length, one edge is seen to make two complete revolutions about the other: the linking number has been raised from zero to two (*4*). It can be shown that the operation always alters the linking number by two; whether the linking number is raised or lowered depends on the exact way in which the passage is done. If one cut end of the ribbon is rotated 360 degrees, however, the linking number changes by one because one revolution has been added. Since DNA gyrase always changes the linking number by an even number, the enzyme must somehow prevent the cut ends from rotating.



PROPOSED MECHANISM for DNA gyrase begins with a DNA ring wrapped around the enzyme molecule in a right-hand coil (*1*). The enzyme binds to a segment of the DNA about 150 base pairs long. The enzyme cuts the ring by breaking the covalent bond between a sugar group and a phosphate group in each strand of the DNA backbone (*2*). A segment of the ring is passed through the gap (*3, 4*). When the break is sealed, the linking number has been changed by two (*5*). In addition the handedness of the DNA coil is reversed by the passage through the break. It is thought that the coil of DNA is somehow returned to its original handedness before the reaction is repeated.

ematician Brock Fuller of Cal Tech before the work on the type II topoisomerases was done.

Work on DNA gyrase tends to confirm the idea that the enzyme alters the linking number of the double-strand ring by a mechanism with the same general form as the ribbon-cutting exercise. It should be noted that a condition of the exercise is that the cut ends of the loop not be rotated. If one end is rotated by 360 degrees about the longitudinal axis of the ribbon, the linking number changes by one, an odd number. Thus DNA gyrase must somehow prevent the cut ends from rotating while the reaction proceeds.

Work in my laboratory and in others has shown that DNA gyrase is a protein with a molecular weight of about 400,000. The protein consists of two identical pairs of subunits, designated *A* and *B*. If the enzyme is denatured while cutting the ring, it is the *A* subunits that are attached to the cut 5′ ends. The result suggests why the ends of the ring do not separate during the breaking of the strands. The 5′ ends of the cut strands are linked to the *A* subunits. The *A* subunits of the enzyme are in turn held together by the *B* subunits. Further work has shown that the enzyme binds to a piece of DNA about 150 base pairs long. Most of the bound segment of DNA appears to be wrapped around the enzyme molecule.

Several mechanisms have been proposed to explain how gyrase alters the linking number. In one mechanism, which has been suggested by Cozzarelli and Patrick O. Brown, a piece of the DNA ring is wrapped around the enzyme as a right-hand coil. The binding of a molecule of ATP to the complex made up of the DNA ring and the enzyme causes a piece of the ring to move through a break in the segment of the DNA that lies on the surface of the gyrase molecule. The break is then resealed. In addition to changing the linking number by two such a passage would reverse the direction of the coil of DNA around the enzyme [*see lower illustration on opposite page*]. Following the splitting of the ATP molecule the coil would somehow revert to its original right-handed form.

In a second proposed mechanism the DNA molecule also wraps around the enzyme, but the handedness of the coil does not change during the reaction. Instead a segment of DNA is brought close to the one lying across the face of the enzyme. The breaking of the double-strand ring is done near the middle of the segment lying across the enzyme. The intact upright segment is then driven through the break. The binding of a molecule of ATP changes the shape of the enzyme to allow the intact segment of DNA to be driven into the interior of the enzyme molecule. The splitting of the ATP molecule leads to the passage of the strand out of the far side of the enzyme [*see upper illustration on next page*]. Either mechanism can accommodate the topological transformations that DNA gyrase is known to do; a considerable amount of work will be needed to determine whether one of them is correct.

While the action of the topoisomerases was being investigated at the molecular level the enzymes were being found in many species, including man. It now appears that all prokaryotic organisms and all eukaryotic organisms have both type I and type II topoisomerases. (Prokaryotes are organisms such as bacteria whose genetic material is not enclosed in a nucleus. The cells of eukaryotic organisms have a nucleus.)

The topoisomerases of eukaryotes apparently differ somewhat from those of prokaryotes. Soon after the discovery of the type I topoisomerase in *E. coli* a type I enzyme was found in the nucleus of mouse cells by James J. Champoux and Renato Dulbecco of the Salk Institute for Biological Studies. The mouse topoisomerase can readily relax both positively supercoiled and negatively supercoiled double-strand DNA rings, in contrast to the bacterial enzyme, which more easily relaxes negatively supercoiled rings. Moreover, when the mouse enzyme cuts the DNA backbone, it forms a covalent bond with the cut 3′ end of the DNA strand rather than with the 5′ end. The type I topoisomerases of other eukaryotes appear to share these properties.

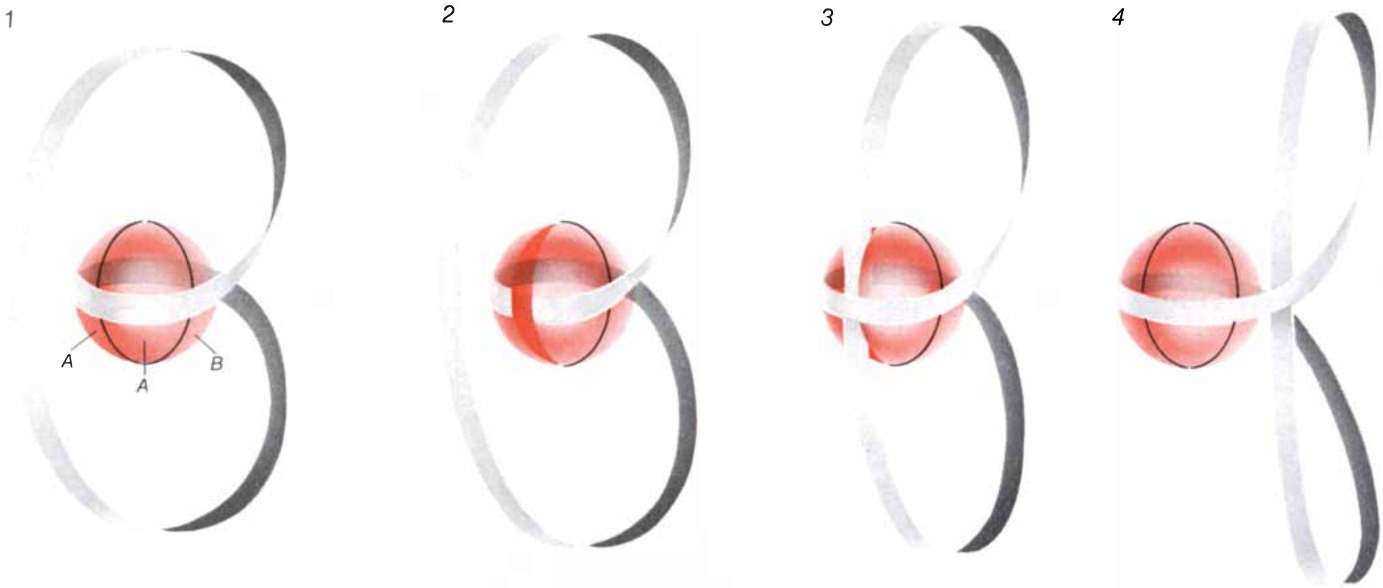
Type II topoisomerases have also been found in many species. Only the bacterial enzyme DNA gyrase, however, can supercoil a DNA ring. On the other hand, the type II topoisomerases of viruses and eukaryotes can carry out the other operations done by DNA gyrase, such as knotting or linking double-strand rings.

The wide distribution of the topoisomerases leads naturally to questions about their biological function. The information available to answer such questions comes mainly from studies of bacteria and the viruses that infect them. Even the limited data now available suggest that the enzymes have a role in the replication of DNA, in the transcription of genetic information and in certain types of genetic recombination.

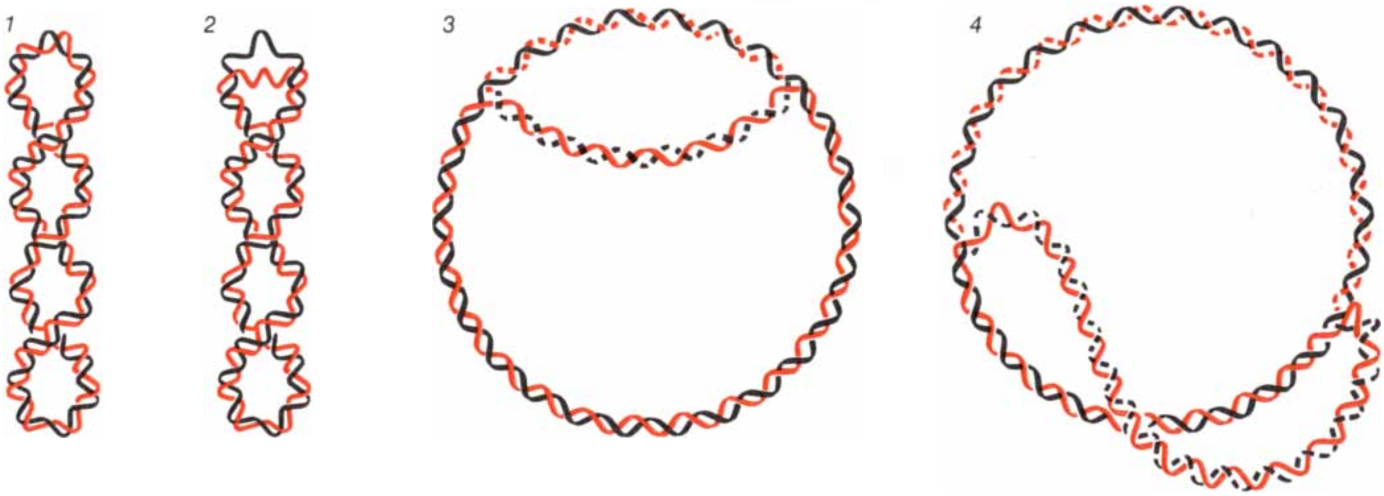
It seems the DNA of most bacterial species is ordinarily negatively supercoiled. The degree of supercoiling is controlled by the topoisomerases. Gyrase supercoils the DNA; a type I topoisomerase can relax it to the required degree. The balanced action of the two enzymes may have significant implications. Both the replication of genetic information and its transcription into RNA require the strands of the double helix to be separated by breaking the hydrogen bonds between base pairs. The separation exposes the bases of each strand to the action of enzymes. Negative supercoiling makes it easier for the strands of the ring to unwind and hence separate. The reason is that unwinding is one way the double helix can approximate the relaxed state.

When part of the double helix unwinds, leaving a segment of each strand free, the total linking number of the ring remains unchanged because its strands are intact. The untwisting, however, redistributes the revolutions of the two strands about each other: the revolutions are primarily confined to the part of the ring that retains the helical configuration. Because the same number of revolutions is distributed over a diminished length the ratio of revolutions to base pairs in the helical segment can approach that of the relaxed ring: one revolution per 10.5 base pairs.

By making it easier for the strands of the double helix to come apart, negative supercoiling can facilitate both replication and transcription. For example, an enzyme coded for by a gene called the *A* gene in the DNA ring of the virus ϕX174 plays an important role in the life cycle of the virus. By cleaving the DNA ring the enzyme initiates the repli-



MOLECULAR GATE has an important role in a second mechanism proposed for DNA gyrase. Little is known of the exact three-dimensional structure of the enzyme. It is known, however, that the gyrase molecule is made up of two pairs of identical subunits, designated *A* and *B*. A segment of about 150 base pairs of the DNA ring is wrapped around the enzyme in a right-hand coil (*1*). The two *A* subunits make up the surface of the enzyme against which the middle part of the segment of 150 base pairs lies. The two *A* units pull apart and break the ring (*2*). The vertical segment of the ring is driven between the two *A* units, which act as an entrance gate, and into the interior of the enzyme (*3*). The break in the helix is then resealed. When the vertical segment has passed halfway into the enzyme, the entrance gate closes. The *B* units open as the entrance gate closes; they act as an exit gate through which the vertical segment is expelled from the enzyme (*4*).



REPLICATION of the information in the sequence of base pairs is necessary for a cell to divide. The replication of a double-strand DNA ring may depend on the capacity of a topoisomerase to reduce the linking number. One of the common mechanisms for the duplication of a double-strand ring is the manufacture of a new complementary strand for each original strand. For the new strands to be assembled the original strands must separate. In bacteria gyrase negatively supercoils the ring (*1*). In the negatively supercoiled form the unwinding of the double helix is made easier. Unwinding is probably required for replication to begin in some organisms (*2*). Assembly of the new strands is begun using the original strands as templates (*3*). In order for replication to proceed the double helix must be progressively unwound. The unwinding may be accomplished by one or more topoisomerase molecules, which repeatedly reduce the linking number (*4*). When the linking number is zero, the two double-strand rings separate; each ring is made up of one original strand and one new strand.

cation of the virus particle in the host cell. In the laboratory the gene-*A* enzyme cuts the viral DNA only if the DNA is negatively supercoiled; the supercoiling may also be necessary inside the bacterial cell.

The replication of the genetic information of both bacteria and eukaryotic cells prior to cell division also requires that the strands separate. As they move apart a new complementary strand is formed for each original strand. In some organisms a high degree of supercoiling may be needed for the process of replication to begin. In addition, as replication proceeds the strands of DNA must be progressively untwisted, thereby reducing the linking number [*see lower illustration on opposite page*]. In order for the new double-strand rings to separate at the end of replication the linking number must be zero. If the number is not exactly zero, the rings remain interlocked. Such interlocking has been seen in viral DNA at the end of the process of replication; presumably a topoisomerase does the final unlinking.

The function of the topoisomerases in transcription may similarly call on their ability to supercoil the double helix. Transcription is initiated by the binding of the enzyme RNA polymerase to the DNA molecule. The enzyme separates about 10 base pairs at a time to allow the sequence of bases to be transcribed. The separation of the strands is facilitated by supercoiling. The binding of RNA polymerase to some DNA sequences is affected by the binding of other proteins to adjacent sites; in many cases the binding of the other proteins is influenced by the degree of supercoiling. Thus adjusting the degree of supercoiling could help to regulate the rate of transcription, which is a crucial element of cellular metabolism.

The breaking and rejoining of DNA strands is also an essential step in gene recombination and rearrangement. The integration of the viral DNA into the host DNA catalyzed by λ integrase is one of the best-studied examples. The enzyme cuts the viral DNA and the host DNA, and the cut ends of the viral DNA are then joined to the cut ends of the host DNA and vice versa. Indeed, Nash and Yoshiko Kikuchi have shown that the λ integrase is a DNA topoisomerase: the enzyme can relax negatively supercoiled DNA, and it forms a covalent bond with the DNA when it cuts the strands.

Biological processes as complicated as those carried out on DNA by the topoisomerases can be expected to have vital functions. When more work has been done on the topoisomerases, they will probably be found to be even more deeply implicated in the genetic operations of the cell.

Aneurysms

Local perturbations of blood flow interact with the repetitive hydraulic stress generated 100,000 times a day by the heart to give rise to these balloonlike dilatations of an artery wall

by Kaj Johansen

Every beat of the heart pumps some 70 milliliters of blood into the root of the aorta under a peak pressure of at least 120 millimeters of mercury, or almost two and a half pounds per square inch. Following each heartbeat a pulse wave is propagated along the arteries, the ramifying system of muscular, elastic conduits carrying oxygen and nutrients to the tissues. In the aorta the wall tension generated by each beat is approximately 170,000 dynes per centimeter. At an average heart rate of 70 beats per minute this stress on the arteries is repeated 100,000 times per day, or 2.5 billion times during a lifetime.

Generally the structural integrity of the arteries enables them to resist the destructive effects of this repetitive hydraulic stress. Sometimes, however, the wall of an artery gives way. A segment of the artery expands to form a balloonlike dilatation: an aneurysm. Once initiated, the dilatation tends to progress in accordance with well-established hydraulic principles. Untreated, the aneurysm may burst under the insistent internal pressure, causing death or grave disability. Even an unruptured aneurysm can lead to damage by interrupting the flow of blood or by impinging on and in some cases eroding nearby blood vessels, organs or bone.

The incidence of aneurysm rises linearly with age, and so in an aging population these lesions can be expected to appear with increasing frequency. On the other hand, better diagnostic techniques are making it possible to detect aneurysms before they do harm. More is known about their causes and natural history, and improved techniques have been developed for dealing with them.

The arteries include blood vessels with a great range of diameters. The aorta is about the size of a garden hose; where it issues from the left ventricle of the heart it has an internal diameter of about 2.5 centimeters and a wall about two millimeters thick. The smallest arteries—the arterioles that supply the capillaries—are about 60 micrometers, or .06 millimeter, in diameter. The detailed structure of the arteries varies with their size, but in all of them the structure seems well suited to resisting the stress of repetitive pulsatile flow.

The wall of an artery has three layers: the intima, the media and the adventitia. The innermost component of the intima, lining the lumen (bore) of the artery, is a continuous tessellation of thin endothelial cells, which offer little resistance to outward pressure but are resistant to the shearing force of the flowing blood. Depending on the size of the artery, the endothelial lining is overlain by more or less fibrous and connective tissue. The intima is separated from the media by the internal elastic membrane, a sheathing composed largely of the protein elastin.

The media, which is responsible for most of the strength and elasticity of the arterial wall, is composed primarily of elastin and smooth-muscle cells, the proportions of which vary in different arteries. In the aorta and other large "elastic" arteries the media is built up of concentric wrappings of elastic membrane, interspersed with smooth-muscle cells and some fibrous material. The elastic walls stretch to accommodate the volume of blood ejected from the heart with each heartbeat; then they recoil, acting as a second, subsidiary pump impelling the column of blood through the arterial tree. The media of the smaller, "muscular" arteries has less elastic membrane and a greater proportion of muscle cells; these arteries contract and dilate in response to signals from the sympathetic and the parasympathetic nervous systems and in response to blood-borne chemical signals.

The adventitia, the outermost layer of the arteries, is a fibrous sheath composed largely of collagen, a connective-tissue protein. Because the adventitia has little elasticity and few smooth-muscle cells it lacks the dynamic-recoil property of the media, but it accounts for much of the static strength of the arterial wall.

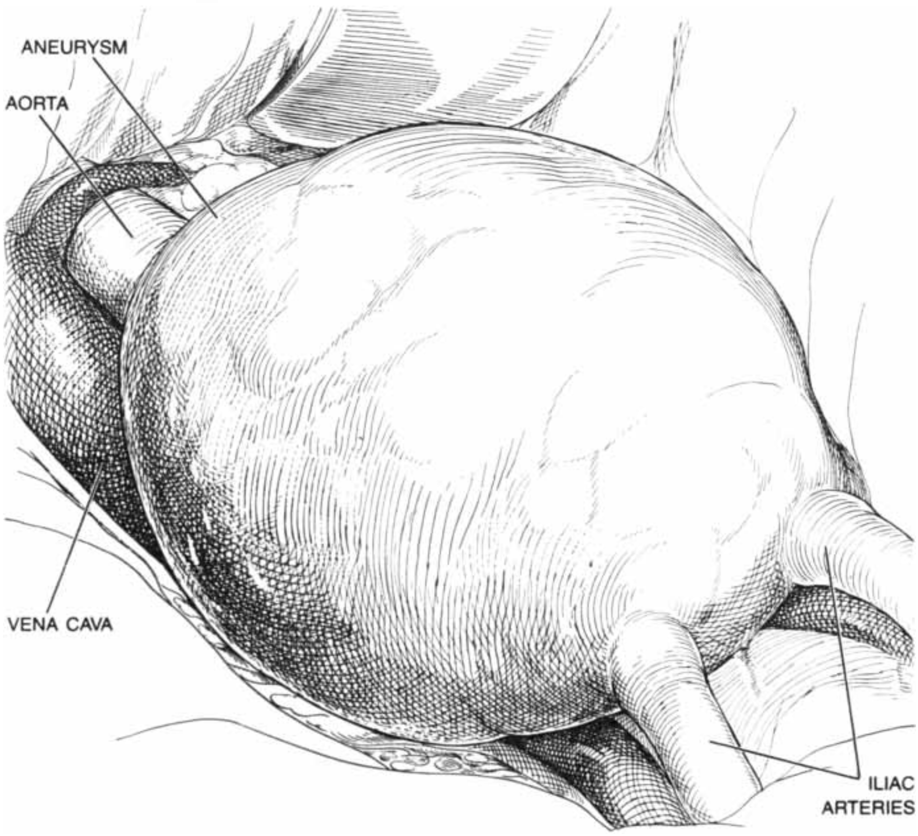
True aneurysms are seen most often in larger arteries, the structure of which may provide a clue to the origin of aneurysms. The leaflike elastic elements and the muscle cells of the media constitute musculoelastic lamellae, which are wrapped around the vessel in a direction roughly perpendicular to its axis. In this configuration they are optimally oriented to resist the distending force of the pulsatile intraluminal pressure. The arteries of individuals who develop an aneurysm often manifest ectasia, a marked elongation of the arteries that tends to come with age. In itself the condition is not pathological, but it may be that ectasia disrupts the orientation of the media's lamellae. Since the maintenance of vessel structure depends on a balance between intraluminal pressure and the wall's ability to resist pressure, a loss of optimal orientation could make it more likely that the wall will weaken. Even in the absence of ectasia, increasing age is accompanied by a reduction in the elastin content of the vessel wall and an increase in the smooth-muscle and fibrous content; old arteries are stiffer than young ones.

Given that aneurysms develop primarily in older people and primarily in elastic arteries that have undergone structural change, one must still ask why some people develop aneurysms whereas others do not and why aneurysms are most commonly seen at certain sites in certain arteries. The assumption has been that the lesions are generally a manifestation of arterial disease, notably atherosclerosis. It is not hard to understand the assumption. In atherosclerosis the smooth-muscle cells of the media proliferate and invade the intima. Lipids and scar tissue are deposited in the intima, forming a "plaque" that begins to narrow the lumen. So-called complicated plaques lead to calcification, ulceration and degeneration of the vessel wall. Observing the frayed and deteriorated lining of an atherosclerotic vessel, one might easily conclude that atherosclerosis can lead to a progressive weakening of the arterial wall and eventually to an aneurysm. Certainly patients with aneurysms often have coexisting symptoms and other indications of atherosclerosis.

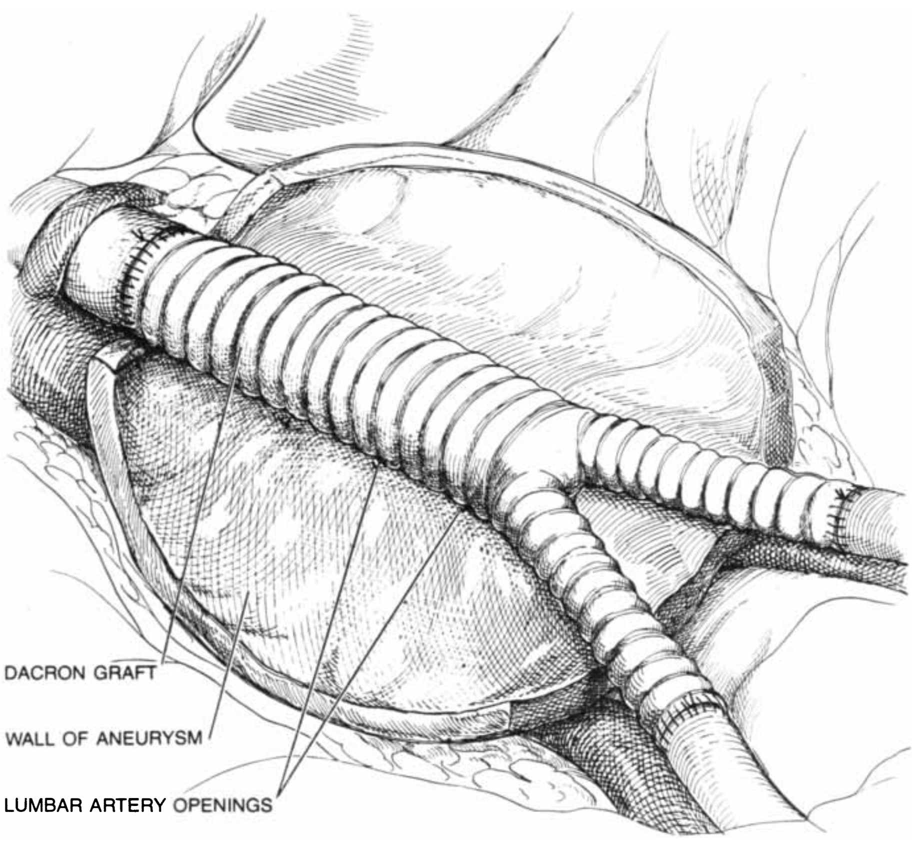
The connection between atherosclerosis and aneurysm may be casual rather than causal, however. The commonest site for aneurysm is the abdominal aorta, below the arteries leading to the kidneys and above the bifurcation of the aorta to form the two iliac arteries. The same aortic segment is a common site of occlusive atherosclerosis. If aneurysm were simply a late manifestation of atherosclerosis, many individuals with aorto-iliac occlusive disease should develop an abdominal aortic aneurysm. They do not seem to do so, however. The incidence of occlusive disease of the terminal aorta peaks at the age of about 55, whereas the peak incidence of abdominal aneurysm comes at 70 or more; the gap is rather too long to suggest that the disease causes the aneurysm. Moreover, atherosclerosis is a systemic disease that manifests itself in many arteries. Yet individuals with aneurysms show little more evidence of generalized atherosclerosis than an age-matched group of people without aneurysms. Any vascular surgeon can attest to the fact that aneurysms are often seen in patients with no evidence of atherosclerosis. Finally, the sites where one sees the most concentrated arterial deposition of lipids in atherosclerosis (in the femoral artery in the lower thigh, for example, or at the bifurcation of the carotid artery in the neck) are only rarely the site of an aneurysm.

Inherited biochemical or structural defects having nothing to do with atherosclerosis may influence the genesis of aneurysms. David W. Rowe, Ermona B. McGoodwin and their colleagues at the Argonne National Laboratory have shown that aortic aneurysms develop spontaneously in "blotchy" mice, a strain with a particular mutation on the *X* chromosome leading to a high incidence of arterial-wall degeneration. The mutation seems to inhibit the formation of a molecular bridge between elastin and collagen by lysyl oxidase, a copper-containing enzyme. Clearly a failure of structural cross-linking could lead to weakening of the arterial wall. In human beings a rare condition known as Menkes' kinky-hair syndrome, characterized by collagen abnormalities and malabsorption of copper, also seems to be inherited by way of the *X* chromosome; arterial aneurysms are sometimes seen in infants with this condition.

Studies done by Ronald W. Busuttil of the University of California at Los Angeles School of Medicine and Philip B.



LARGE ANEURYSM of the abdominal aorta is depicted as it appears to the vascular surgeon about to repair it. The aneurysm, a dilatation of the artery wall, is just above the bifurcation of the aorta, which divides to form the two iliac arteries. The arteries are clamped above and below the aneurysm, the dilated sac is opened and a graft is sutured in place from within.



PROSTHETIC GRAFT, a *Y*-shaped tube of knitted Dacron, is shown in place, sutured to undamaged regions of the aorta and the two iliac arteries. Blood flow has been reestablished.

THE LEADING EDGE

#3 in a series of reports on new technology from Xerox

Xerox introduced the first xerographic copier, Model 914, in 1959. The 914 and xerography were great surprises to the world. They revealed and then satisfied an immense latent demand for plain paper copying. And they surprised scientists and engineers by demonstrating practical applications of physical phenomena that had not been thought to have much practical value. They used large page-sized sheets of semiconducting selenium to capture entire images by the phenomenon of photoconductivity. What's more surprising, they found a practical use for "static electrification," which is the transfer of electric charge that occurs when, for example, one rubs a cat on a dry day.

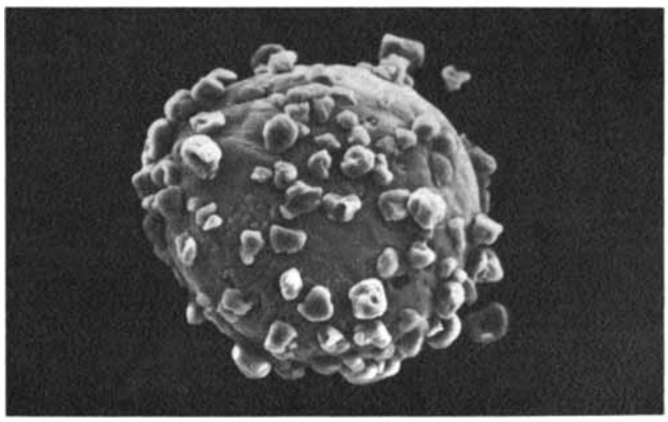
This fundamental part of the xerographic process was not at all understood scientifically. Early technology work was by trial and error.

Some years ago a few of us in Xerox Research became aware that as the demands on xerography increased, trial and error methods of xerographic design would become increasingly risky. We undertook a small, deliberate research program on the fundamentals of the xerographic process.

THE XEROGRAPHIC PROCESS

A xerographic image—a copy—is made of about 100 million carefully arranged bits of pigmented plastic called toner particles. The whole xerographic process is simply the means of arranging toner particles into faithful patterns, then transferring the patterns to paper and making them permanent by melting.

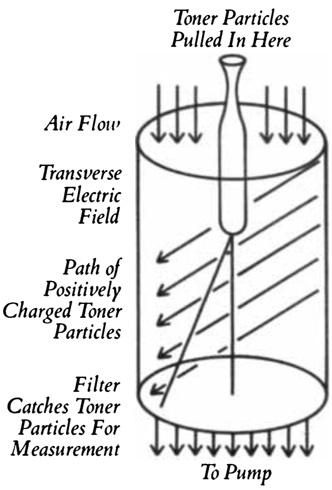
The process works as follows: First, a photoconductor, a paper-thin layer of glassy selenium coated on aluminum, is given a uniform electric charge on its top surface. Usually this is done with a spray of positive ions from a wire or point at high voltage. In darkness selenium is a good electrical insulator so the electric charges are stable. Next, the image of a document to be copied is projected onto the photoconductor discharging it in light-struck areas and leaving charge patterns corresponding to the black letters and marks of the original document. Finally, this latent electrostatic image is developed by attracting oppositely charged toner particles to it. Here is where static electrification becomes crucial. Toner particles get their charges by repeated rubbing against the surfaces of larger particles called carrier beads. A carrier bead with toner is shown in the picture. When the process is working properly, each toner particle carries the electric charge of about 20,000 electrons.



The photo shows a single carrier bead with toner clinging by electrostatic force. A xerographic copier might hold about three kilograms of carrier which itself binds 50 to 100 grams of oppositely charged toner. The combination is a xerographic developer.

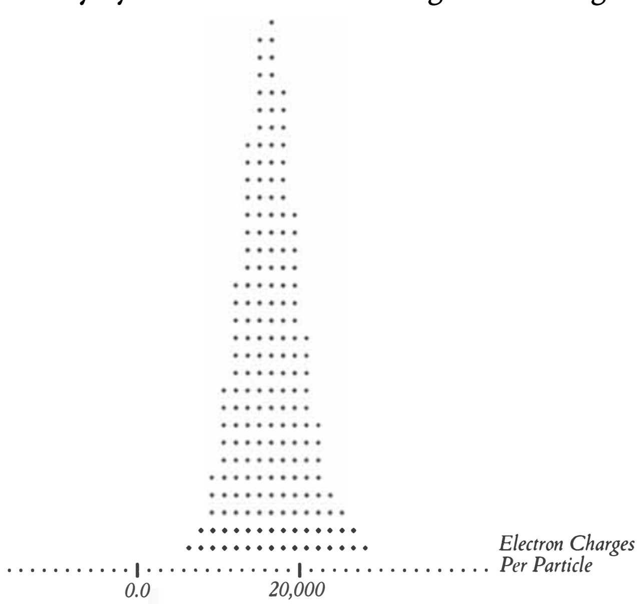
THE CHARGE SPECTROGRAPH

The trouble was that we had no way to make certain that a xerographic developer was in fact working properly. We could not measure the charge on individual toner particles to be sure that all of them had the right charge. We suspected that wrong-charge particles might often be present and responsible for poor copy quality and for at least some of the dirt contamination of xerographic equipment. A reversed-sign particle will go in exactly the wrong place—the white areas—and only a small percentage of these in the whole toner population can produce a visible flaw. A particle with no charge is uncontrollable. By measuring charge distributions on toner we could characterize xerographic developers independently of copier hardware. We hoped also that charge distributions would eventually provide insight into the physical mechanisms responsible for static electrification.

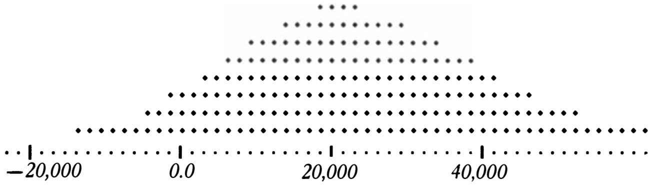


With these things in mind, E.W. Connors, R.F. Koehler and I built the charge spectrograph sketched here.

It's a kind of low-speed wind tunnel with a uniform electric field across the direction of air flow. Toner particles are stripped from carrier beads with a tiny air jet and introduced on the axis of the instrument. As they move downward with the air flow, they are pulled laterally by the electric field according to their charge and size and finally caught on an exit filter for examination. Afterwards, the filter is removed and the drift distance and size of each toner particle is read through a microscope by an automated analysis system.



Example 1. *The charge distribution on the 10 micron size class of toner particles. This is a fairly sharp distribution.*



Example 2. *Charge on the 10 micron toner particles in a worn-out developer. Note the reverse-sign and uncharged toner.*

Experimental results are in the form of distributions or histograms, which are plots of the fractions of the toner particles having various charges. Here we show only a particular size category, the toner particles with measured diameter of 10±1 micron. The first example is free of low-charged or reverse-charged toner particles. This xerographic developer performs well in equipment.

The second distribution is from a developer that has been used to make too many copies. This developer is worn out; the charge-exchanging character of its carrier surfaces has changed.

And in cases that we haven't shown by example, the charge on each toner particle is proportional to its surface area; and in some others, it is proportional to toner diameter. These and other regularities, once seen, beg for explanation. These explanations lead to completely new developer designs which work as predicted.

The most satisfying outcome is that we are beginning to understand the underlying physics well enough to formulate rules for the design of xerographic materials.

Over many years we have examined hundreds of developers under many conditions of use. A few have already gone to market and others are on the way.

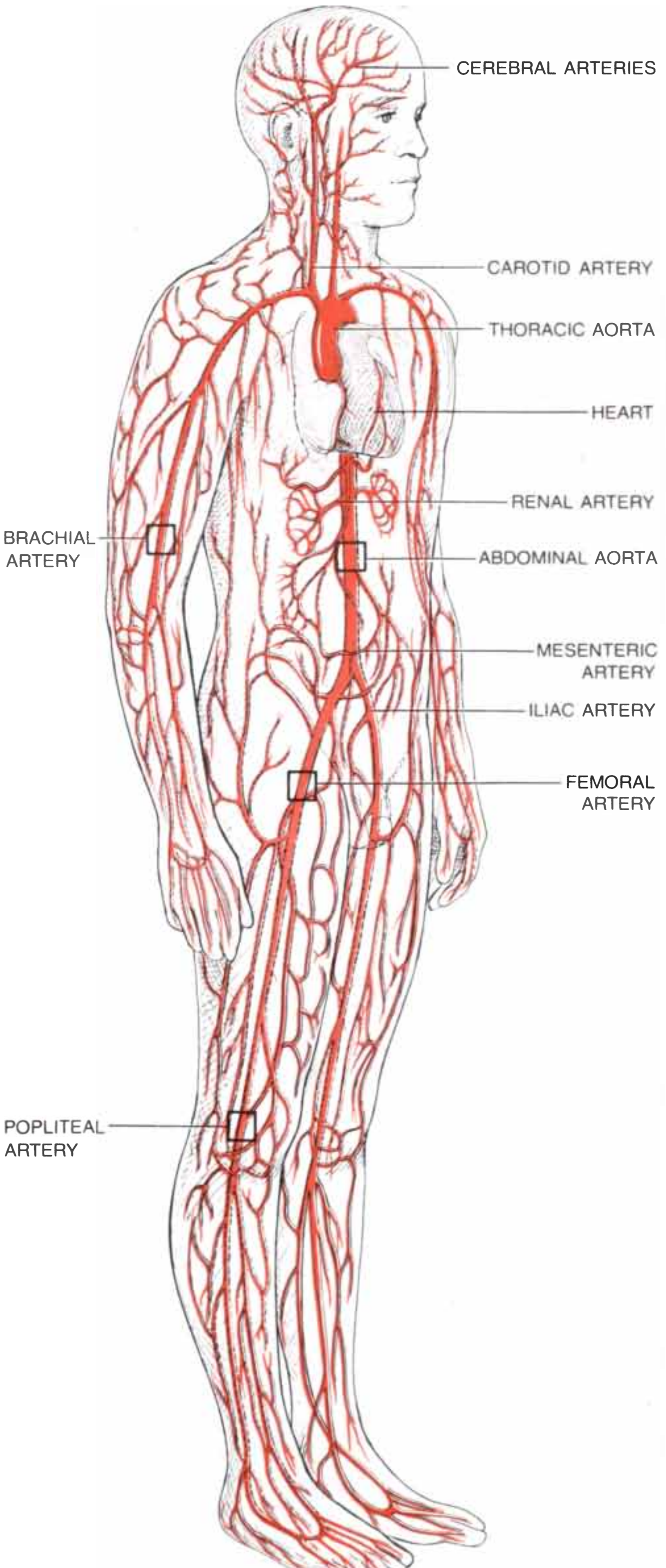
The charge spectrograph is only one example of how we can make xerography work even better in future generations of copiers and printers. During the past ten years people have sometimes said that xerography was a mature technology. We never believed it.

ABOUT THE AUTHOR

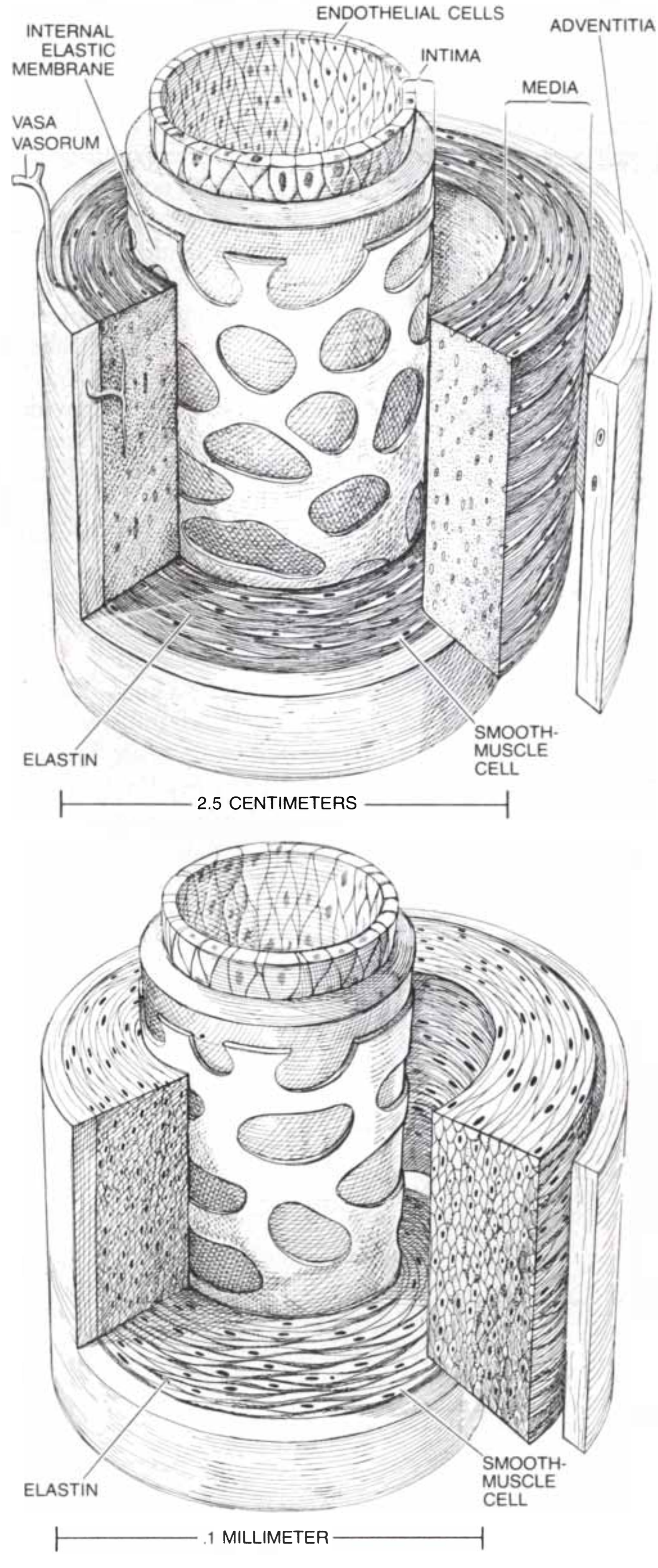
R.B. Lewis is a member of Xerox Corporate Staff specializing in electronic printing strategy. The work described here was done when he was manager of the Imaging Physics and Materials Laboratory at the Xerox Webster Research Center. He holds a B.A. degree from Yale and a PhD degree in physics from Princeton.

Dobrin of the Loyola University Stritch School of Medicine suggest that enzymatic dissolution of structural subunits in the artery wall may have a role in weakening the wall. In patients with an aneurysm the enzymes that break down collagen and elastin may be present in unusually large quantity or in abnormal proportions.

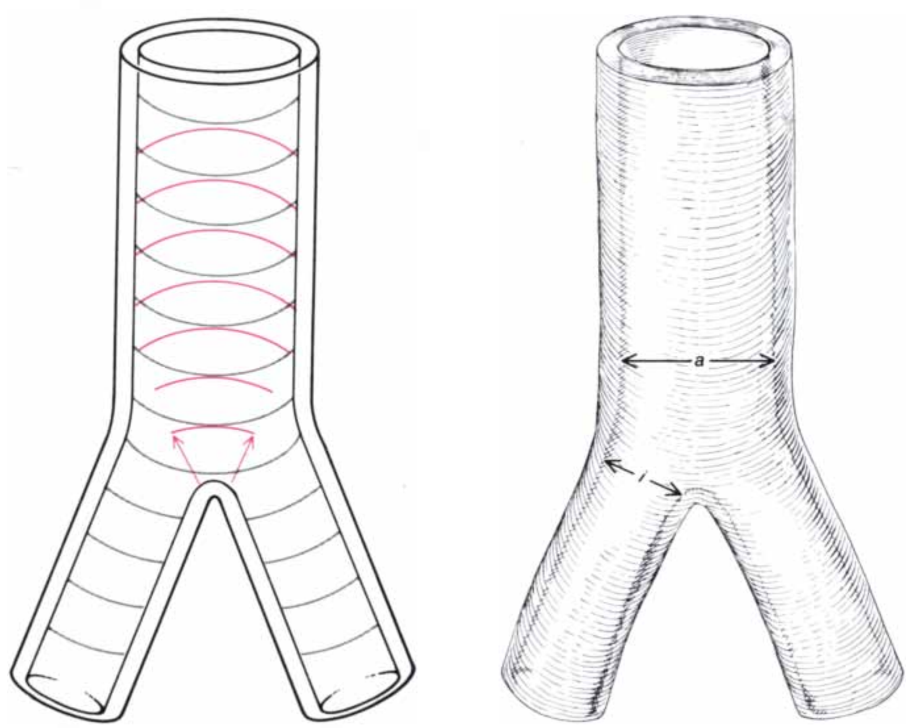
M. David Tilson of the Yale University School of Medicine has reported a correlation between arteriomegaly, or inherently large arteries, and the development of aneurysms. Our group at the Harborview Medical Center of the University of Washington School of Medicine is studying the families of patients who have aortic aneurysms. Our preliminary impression is that the incidence of the condition in close relatives of a patient is higher than could be ascribed to



MAJOR SITES OF ANEURYSMS are indicated (*black squares*) on this diagram of the arterial tree. The commonest sites are just upstream from where an artery bifurcates or gives off a major branch.



STRUCTURE OF ARTERY WALL varies with the size of the artery and its function. The drawings show (at very different scales) the wall of the aorta (*top*) and of a small muscular artery (*bottom*).



WAVE OF PRESSURE moving through fluid in a tube is partially reflected by a bifurcation (*left*), putting lateral stress on the wall just upstream from the bifurcation. Such reflections may be partly responsible for aneurysms just above the bifurcation of the aorta (*right*). The magnitude of the reflected wave is inversely related to the ratio between the area of the lumen (bore) of the two iliac arteries and the area of the aorta, or $2i/a$. The area ratio tends to become smaller with age as aorta becomes more flaccid, and so reflected wave increases in magnitude.

chance, implying there may indeed be a heritable trait (expressed late in life) that can lead to the deterioration and eventual dilatation of the arterial wall.

It may be that neither arterial disease nor any heritable defect is primarily implicated in the genesis of an aneurysm. The lesions may simply be generated at particularly vulnerable sites by long years of pulsatile hydraulic stress. The commonest aneurysm, as I have mentioned, is found near the lower end of the abdominal aorta, just above its bifurcation into the iliac arteries. Experimental work at several centers suggests that the incidence of artery-wall deterioration at this site can be linked specifically to local perturbations of the blood flow.

A wave conducted down a tube is partially reflected from a downstream bifurcation or branch point. Ray G. Gosling and his colleagues at Guy's Hospital Medical School in London have shown in roosters and dogs and in postmortem studies of human arteries that the nature and magnitude of the wave reflected from the aortic bifurcation can be related mathematically to the ratio of the cross-sectional area of the two iliac arteries to the cross-sectional area of the aorta. The area ratio becomes progressively smaller with age as the aorta becomes more flaccid, and the smaller the ratio, the larger the wave reflected back up the aorta. Any initial aortic dilatation (due perhaps to arteriomegaly, ectasia or the normal age-related reduction in arterial-wall elasticity) would further reduce the area ratio, thereby augmenting the wave reflectance and the resulting stress on the aortic wall. In 1954 Arthur B. Voorhees, Jr., and Arthur H. Blakemore of the Columbia University College of Physicians and Surgeons suggested that a "water hammer" effect, generated when the reflected wave collides with the pressure pulse of the next heartbeat, results in additional lateral stress on the aortic wall.

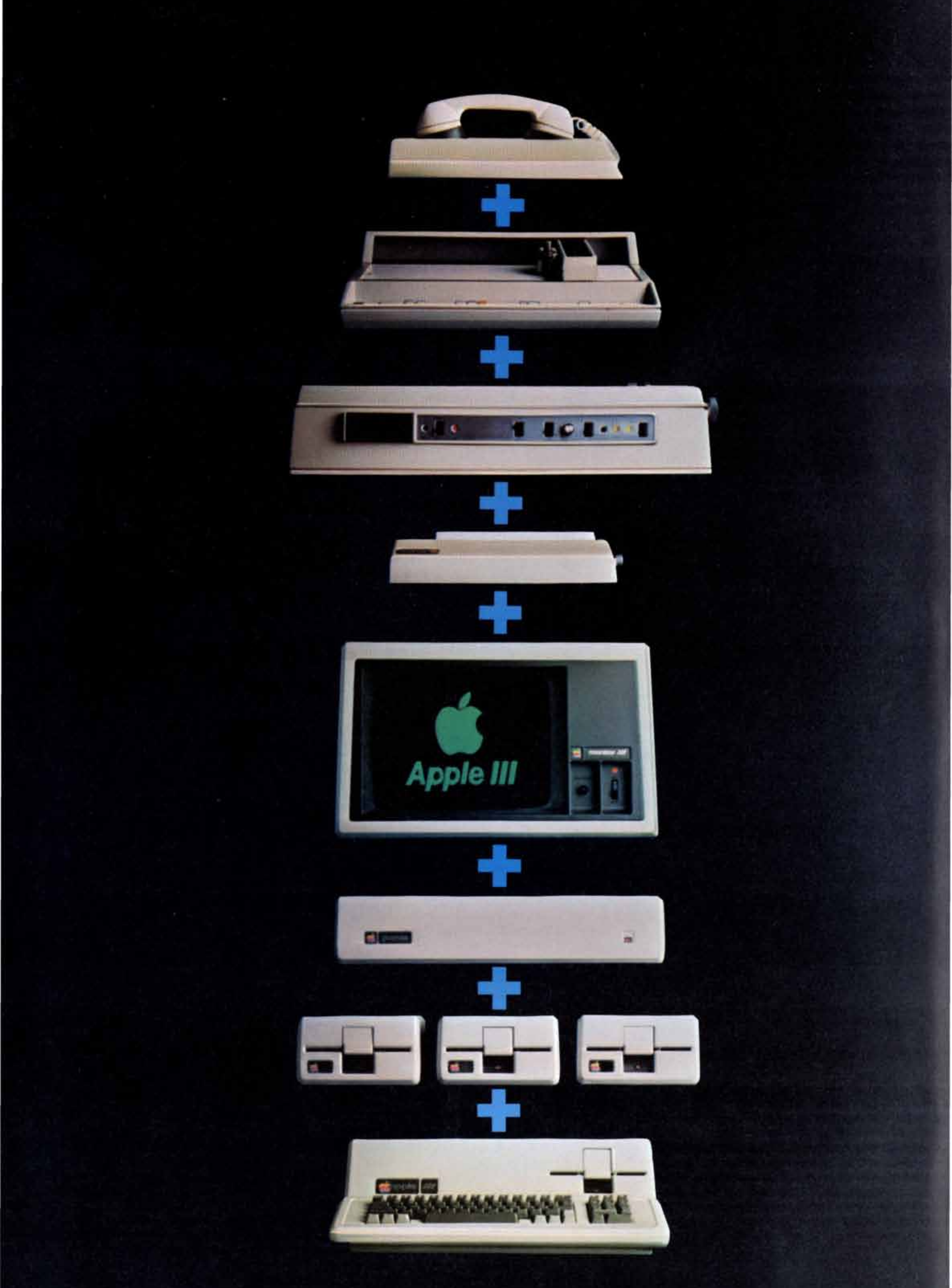
Another hydraulic principle may affect the progress of an aneurysm. When the bore of a tube flares gradually (as the aortic lumen does in an incipient aneurysm), the fluid near the wall slows down, generating turbulent flow. The turbulence can make the wall vibrate. Margot R. Roach of the University of Western Ontario Faculty of Medicine has shown that arterial-wall vibration at a low frequency (up to 200 cycles per second in the human iliac artery) can be highly destructive to structural components of the wall. Anatomically the importance of these various hydraulic mechanisms is supported by the fact that a common site of aneurysms is just upstream from a bifurcation or a branch point. Susceptible sites of this kind are found not only in the aorta but also in the iliac and femoral arteries of the pelvis and upper thigh and in the popliteal artery behind the knee.

Evidence has accumulated that the abdominal aorta may be particularly prone to structural deterioration because it is inadequately supplied with oxygen and nutrients and also is structurally deficient. The terminal segment of the aorta has fewer vasa vasorum (arterioles that bring nutrients to the outer layers of a blood vessel) than other arteries have. Investigations by Seymour Glagov and his colleagues at the University of Chicago Pritzker School of Medicine indicate that the abdominal segment of the human aorta may be deficient in its musculoelastic wrapping: between the renal arteries and the bifurcation only 28 lamellae can be identified in the media, whereas 35 would be predicted on the basis of comparative animal studies. The result is that under normal arterial pressure the tension per musculoelastic lamella is more than twice as high in the human abdominal aorta as it is in experimental animals; the tension per lamella is also twice as high in the human abdominal aorta (where aneurysms are common) as it is in the human thoracic aorta (where aneurysms are rare).

Some specific vascular conditions are associated with the formation of aneurysms. Fibromuscular dysplasia, a condition that seems to affect the media, can make an artery look like a string of beads, with narrowed areas between aneurysmal dilatations. Such lesions appear most often in the renal arteries and occasionally in the carotid arteries, and they can sometimes be a cause of hypertension and stroke.

Small "berry" aneurysms in brain arteries arise at sites (often branch points) where the media is deficient or absent. They can burst, causing bleeding within or around the brain that results in neurological damage. Kawasaki disease, an affliction of newborn infants that may be viral in origin, is sometimes characterized by aneurysms of the coronary arteries. Laborers who work with vibrating tools such as jackhammers or chain saws may develop aneurysms of small arteries in the hand. By and large, however, the causative chain leading to aneurysmal degeneration seems to begin with a local deficiency of structure, perhaps associated with inadequate blood supply. Hydraulic anomalies that arise from turbulent flow and pulse-wave reflection increase the wall tension inordinately. The result is a dilatation that, once initiated, progresses inexorably to the stage of aneurysm.

A true aneurysm is characterized by the dilatation of all three layers of the artery wall. A pseudoaneurysm, which is a pulsatile mass with many of the clinical features of a true aneurysm, can

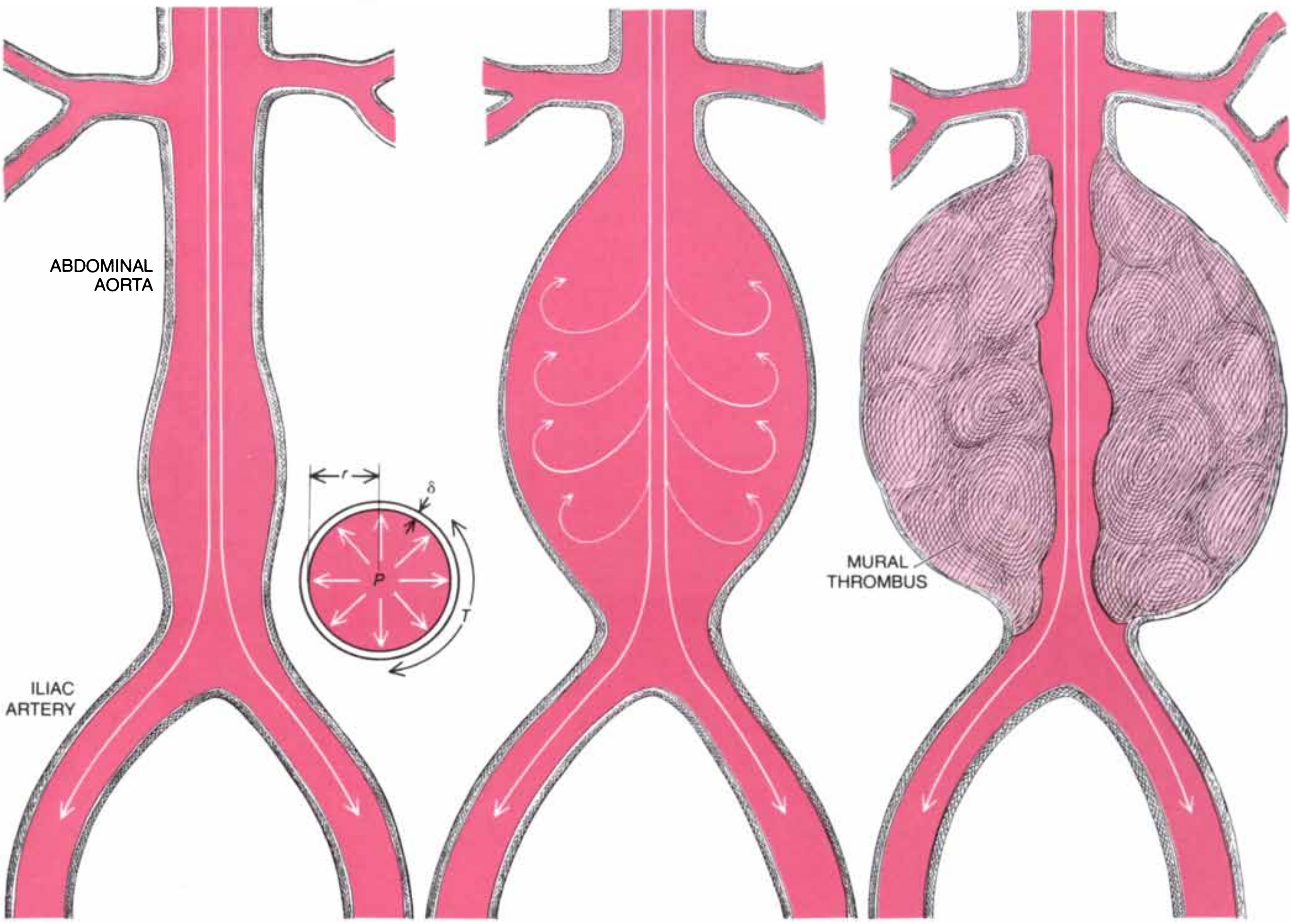

Apple III

arise when blood under pressure breaks through one or two layers of the artery wall, or through the entire wall to form a hematoma. The commonest cause of a pseudoaneurysm is a penetrating injury: a stab or gunshot wound, damage from a needle or catheter introduced into an artery in a diagnostic or therapeutic procedure, damage from infection or a defect in the suturing of an artery to an artificial-bypass graft.

Once an aneurysm has developed, the tension on its wall increases in accordance with a fundamental hydraulic principle enunciated some 200 years ago by the French astronomer and mathematician Pierre Simon de Laplace. The principle relates wall tension to pressure and to vessel size. It explains, among other things, the bursting of a balloon or a soap bubble. Laplace's law states that the tension (T) in the wall of a spherical container is proportional to the product of the intraluminal pressure (P) and the radius (r) of the sphere: T is proportional to Pr. The law applies strictly only to a sphere with an infinitely thin wall. D. Eugene Strandness, Jr., and David S. Sumner of the University of Washington School of Medicine have pointed out that wall thickness (δ) must play a role in the distention and ultimate rupture of an aneurysm. (A thick-walled balloon resists bursting better than a thin-walled balloon.) The tension in the arterial wall, then, varies directly with the product of the pressure and the radius and inversely with the wall thickness: T is proportional to Pr/δ.

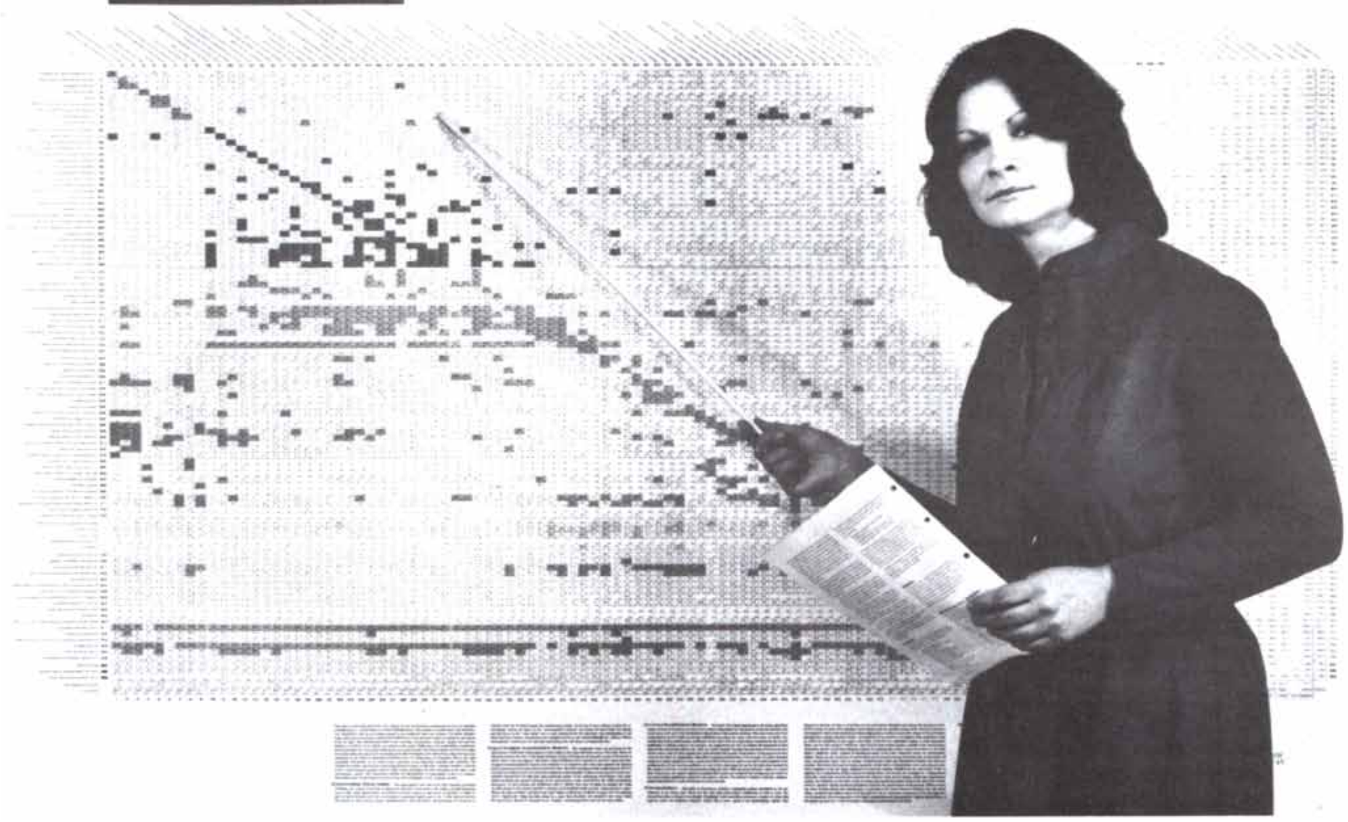
Because the volume of the arterial wall presumably remains constant the wall thickness must decrease as the aneurysm enlarges. The value of $(Pr)/\delta$ therefore increases very fast: the wall tension in an aneurysm six centimeters in diameter is 12 times the tension in the wall of the normal abdominal aorta with an internal diameter of two centimeters. Increased tension makes the wall thinner as it enlarges the aneurysm, which further increases the tension, and so on. This vicious positive feedback, which is often exacerbated clinically by high blood pressure (an increase in P), can lead eventually to the rupture of the aneurysm.

As an aneurysm develops, its wall becomes lined with a mural thrombus: an aggregate of blood cells, platelets, blood proteins and cellular debris. The thrombus grows as material is deposited from the flowing blood, so that in time the channel left open through the aneurysm may not exceed the diameter of the intact lumen on each side of the aneurysm. It has been argued (perhaps teleologically) that the thrombus represents an attempt by the body to increase the wall thickness, thereby reducing wall tension and the likelihood of rupture. The thrombus is actually a gelatinous mass with many crevices, however, and because pressure is equilibrated at all points in a hydraulic system the wall



ABDOMINAL AORTIC ANEURYSM, once initiated, tends to enlarge inexorably. The tension (*T*) on the wall varies directly with the product of the intraluminal pressure (*P*) and the radius (*r*) of the lumen and inversely with the thickness (δ) of the wall: *T* is proportional to *Pr*/δ. Any slight dilatation (*left*) increases the radius and decreases the thickness, thereby increasing the tension and enlarging the aneurysm. As the artery dilates, blood near the wall flows more slowly, creating turbulence (*middle*). Turbulent flow causes the wall to vibrate, further weakening it. Both turbulence and irregularities on the lining of the damaged wall promote formation of a mural thrombus: an aggregation of blood platelets, blood cells and cellular debris knitted together by the clotting protein fibrin. The thrombus may significantly reduce the diameter of the lumen (*right*). It can also fragment, forming emboli that can close off a smaller artery downstream.

THE INPUT/OUTPUT STRUCTURE OF THE UNITED STATES ECONOMY



WHAT MAKES THE U.S. ECONOMY TICK?

The editors of SCIENTIFIC AMERICAN have prepared a wall chart displaying for the 1980's the Input/Output Structure of the U.S. Economy based on the latest interindustry study from the U.S. Department of Commerce.

The SCIENTIFIC AMERICAN Input/Output wall chart does for economics what the table of elements does for chemistry. It answers at a glance questions about the linkage between the microeconomics of the firm and the macroeconomics of the system; about the web of technological interdependencies that tie industry to industry; about the industry-by-industry direct and indirect consequences of swings in public and private spending; about the impact of change in technology, and about any other topic you can think of. You are rewarded by surprise as well as by confirmation of your hunches. For teaching and practical and theoretical studies, here is a powerful, graphic tool.

In the familiar format of the SCIENTIFIC AMERICAN Input/Output wall charts for the 1960's and 1970's, the wall chart for the 1980's measures 65″ × 52″ and is printed in eight colors. Each of the nearly 10,000 cells in the 97-sector interindustry matrix shows (1) the interindustry commodity flow, (2) the direct input/output coefficient and (3) the "inverse" coefficient. Where the direct input/output coefficient exceeds .01, the cell is tinted in the color code of the industrial bloc from which the input comes. This device, combined with triangulation of the matrix, brings the structure of interindustry transactions into graphic visibility.

A supplementary table displays, industry by industry, the capital stock employed; the employment of managerial, technical-professional, white-collar and blue-collar personnel; the energy consumption by major categories of fuel, and environmental stress measured by tons of pollutants.

The editors of SCIENTIFIC AMERICAN are happy to acknowledge the collaboration, in the preparation of this wall chart, of Wassily Leontief, originator of input/output analysis—for which contribution to the intellectual apparatus of economics he received the 1973 Nobel prize—and director of the Institute for Economic Analysis at New York University.

Packaged with the chart is an index showing the BEA and SIC code industries aggregated in each of the 97 sectors.

SCIENTIFIC AMERICAN
415 Madison Avenue, Dept. I/O, New York, N.Y. 10017

YES, send me the new SCIENTIFIC AMERICAN 1980's INPUT/OUTPUT wall chart—printed in eight colors on heavy paper measuring 65″ × 52″.

I understand my chart(s) will be shipped prepaid, uncreased*, flat-rolled and in a sturdy mailing tube. (Allow 4-6 weeks for delivery)

I enclose a check for $_____ for _____ chart(s) at $35 each, saving me $5 each in handling charges.

Bill me for $_____ for _____ chart(s) at $40 each.

Name ____________________

Company ____________________

Street Address ____________________

City, State, Zip Code ____________________

*Charts shipped international require one fold. 7/82

Make check payable to SCIENTIFIC AMERICAN. Please add sales tax for California, Illinois, Massachusetts, Michigan and New York.

tension at the bottom of a crevice is the same as it would be if no thrombus were present. Accordingly the mural thrombus can be only marginally helpful in resisting rupture.

Aneurysms are not rare. In a postmortem study done in Sweden aneurysms of the abdominal aorta were found in almost 2 percent of the autopsies. Aneurysms of all kinds, in other arteries as well as the aorta, are seen in up to 10 percent of all postmortem examinations.

In some cases an aneurysm of the abdominal aorta is discovered during a routine physical examination, when palpation of the abdomen reveals a pulsatile mass. Alternatively one of several complications may betray the presence of a previously undiagnosed aneurysm. Bits of mural thrombus may break off and be carried downstream; these emboli can become lodged in a smaller vessel and interrupt the flow of blood. The growing mural thrombus may encroach on the lumen of the artery until finally it occludes blood flow through the aneurysm. Thrombosis of an aneurysm is particularly serious in the popliteal artery behind the knee. There are few collateral vessels that can substitute for the occluded artery, and so limb loss is common following thrombosis of a popliteal aneurysm. In the aorta or the iliac arteries an aneurysm may enlarge to the extent that it obstructs nearby hollow organs such as the ureter; it may erode the spine or a segment of the intestine.

Of course the gravest manifestation of aneurysmal disease is rupture, which is commonest in large aneurysms of the abdominal aorta. Several studies, including one by Eugene F. Bernstein and George R. Leopold of the University of California at San Diego School of Medicine, have shown that most aneurysms enlarge slowly, with a mean annual dilatation of .4 centimeter. They generally keep growing, however, as is predicted by Laplace's law, and the larger the aneurysm grows the more likely it is to burst. Lesions as small as four centimeters in diameter may be subject to rupture. A significant incidence of rupture can be expected once the diameter exceeds five centimeters. D. Emerick Szilagyi of the Henry Ford Hospital in Detroit has shown that the likelihood of rupture for lesions more than six centimeters in diameter is greater than 50 percent.

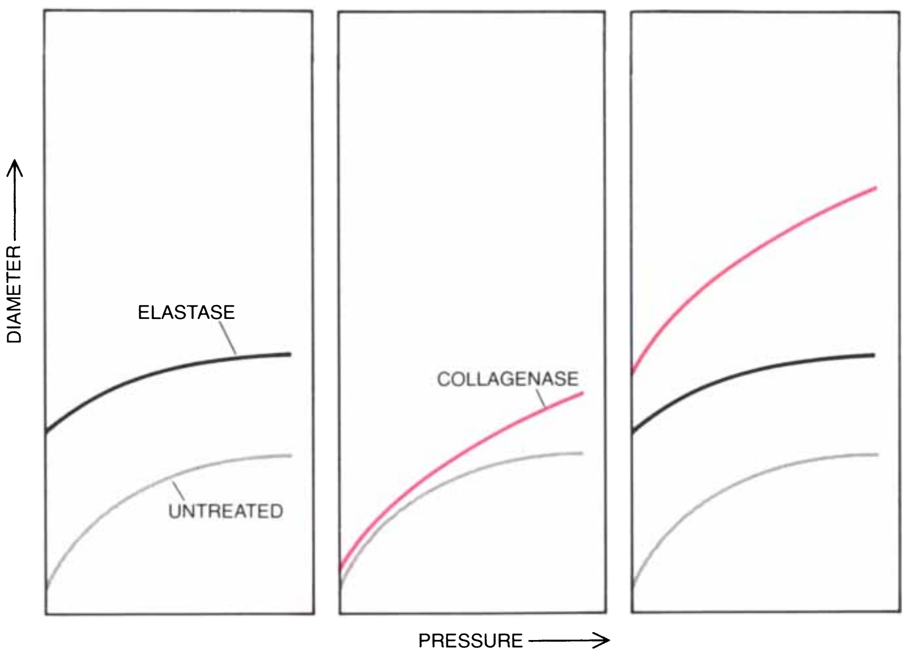
If blood from a bursting aneurysm breaks through the peritoneum (the membrane enclosing the visceral organs), the patient may quickly lose large amounts of blood into the abdominal cavity; irreversible shock, cardiac arrest and death can follow within minutes. More commonly the break in the vessel is contained transiently within the retroperitoneum, behind the abdominal cavity, where resistance and pressure provided by surrounding muscles and other soft tissues may temporarily check the flow of blood from the breached vessel. In this situation patients either seek medical attention quickly because of severe pain in the back or the groin or are brought to the hospital because they go into shock.

The lost blood must be replaced immediately while an emergency operation is undertaken in which the aorta is clamped just above the aneurysm. If this can be done before the patient has lost so much blood that vital organs have been irreversibly damaged, survival may be possible. Most patients with a ruptured aneurysm are elderly, however, and they may have other medical problems, notably heart disease; if they survive the emergency, they may nonetheless succumb to myocardial infarction, a stroke or kidney failure. A study by our group at the Harborview Medical Center suggests that a third of the survivors of aneurysmal rupture may show evidence of gangrene of the large intestine, presumably as a result of the loss of blood. The mortality rate for rupture of an aortic aneurysm approaches 50 percent and in some reported series has exceeded 90 percent for patients who were in shock.

Since the late 1940's it has been possible to demonstrate the presence of a suspected aneurysm and to assess its size by means of arteriography: a technique in which the arteries are made visible in an X-ray photograph by injecting a dyeopaque to X rays into the artery with a needle or a catheter. The technique calls for puncturing the artery; moreover, the presence of a mural thrombus can keep the dye from showing accurately the size of the aneurysm.

More recently two important noninvasive techniques have come into use: ultrasound imaging and computerized axial tomography (CAT). An ultrasonogram is formed by the differential reflection of high-frequency sound waves from different tissues. A CAT scan is a cross section of a plane of the body, revealing soft-tissue structures as well as bone, that is constructed by computer processing of signals generated as an X-ray source is rotated about the body. With these new methods the size of a small, asymptomatic aneurysm can be measured accurately; it is also possible to make serial examinations and thereby trace the development of the lesion.

The mortality rate for elective surgical repair of an aortic aneurysm has fallen as low as 1 or 2 percent in some series, even when a number of patients in their eighties have been included. The incidence of serious complications is also low, and individuals who have undergone the operation have about the same life expectancy as an age-matched group of people who have never developed an aneurysm. These results, together with the high mortality attending aneurysmal rupture and the fact that there is no effective nonsurgical treat-



EFFECT OF ENZYMES on carotid arteries from dogs was studied by Philip B. Dobrin of the Loyola University Stritch School of Medicine. He treated the arteries with elastase and collagenase, enzymes that break down elastin and collagen respectively, and measured the vessels' diameter as the internal pressure was increased. Elastase caused an artery to dilate (*left*) but did not lead to rupture. Collagenase caused only slight dilatation (*middle*) but led to rupture. Treatment with elastase and then collagenase led to massive dilatation, then rupture (*right*).

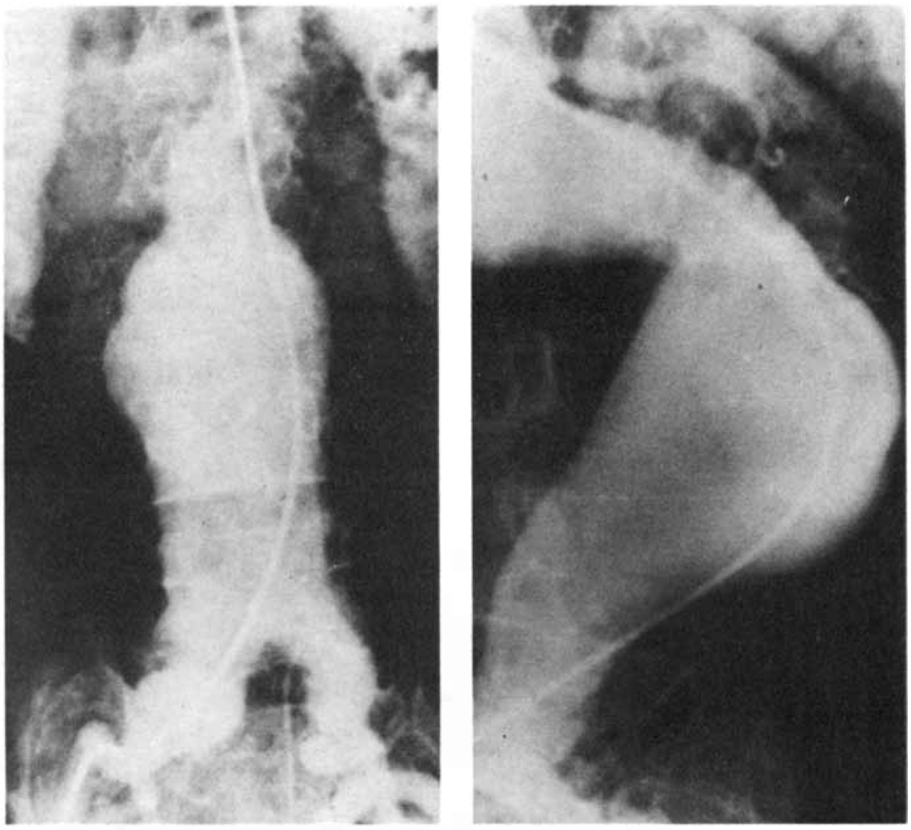
ment, argue in favor of elective repair. The elective procedure should certainly be performed for any aneurysm that is more than five centimeters in diameter, is tender to the touch or is giving rise to symptoms such as back pain. Even small, asymptomatic aneurysms should probably be repaired unless the patient has another serious disease that makes him a very poor candidate for surgery.

Aneurysms have been recognized as lethal lesions for a long time. Until recent decades the commonest cause was probably not mechanical stress on aging arteries but infection (due either to syphilis or to the lodging of bacteria from an infected heart valve) or arterial damage, often as a result of the widespread practice of bloodletting. Surgical repair of aneurysms in the popliteal and brachial arteries was attempted as early as the 18th century. The artery was tied off above and below the aneurysm, almost inevitably at the expense of subsequent loss of the limb.

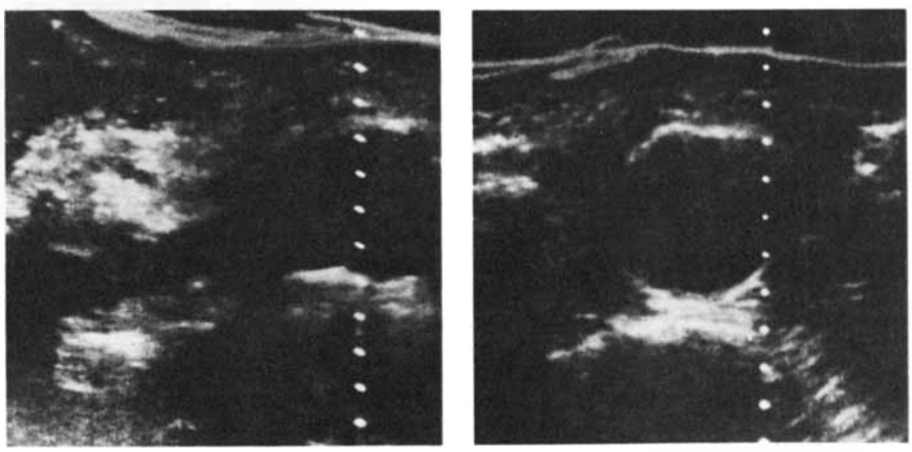
In the 19th century attempts were made to inhibit the enlargement of an aneurysm by inducing clotting within the vessel, either by inserting needles or by applying a compressive dressing around a limb or the abdomen. In 1864 an English surgeon was reported to have introduced 75 feet of fine iron wire into an aneurysm to stimulate thrombosis. Electrocoagulation of an aneurysm was first tried in 1879 by the Italian physician Corradi, who passed an electric current through 42 centimeters of wire inserted into an aneurysm. The technique was applied widely, with sporadic success, for many years.

True repair of an aneurysm became possible only when several technological developments emboldened vascular surgeons to replace or bypass a diseased blood vessel. The new developments included the clinical application of heparin to prevent clotting in a clamped vessel, the availability of arteriography and better understanding of several principles of vascular surgery. The first replacement of the terminal aorta (it was for occlusive atherosclerosis rather than for aneurysm) was done by Jacques Oudot at the Hôpital Broussais in Paris in 1951. Later that year and at the same institution Charles Dubost, Michel Allary and Nicolas Oeconomos removed an abdominal aortic aneurysm in its entirety and replaced it with a length of artery taken from a cadaver. Michael E. DeBakey and Denton A. Cooley, who were then working at the Baylor University College of Medicine, repaired an aneurysm of the thoracic aorta in 1953. A year later Charles G. Rob of St. Mary's Hospital in London replaced an abdominal aneurysm with the first synthetic graft: a tube of shirt linen fashioned on the family sewing machine. Soon prosthetic grafts, usually made out of knitted Dacron, became available, and by the mid-1960's they were being implanted in a procedure that had become fairly routine.

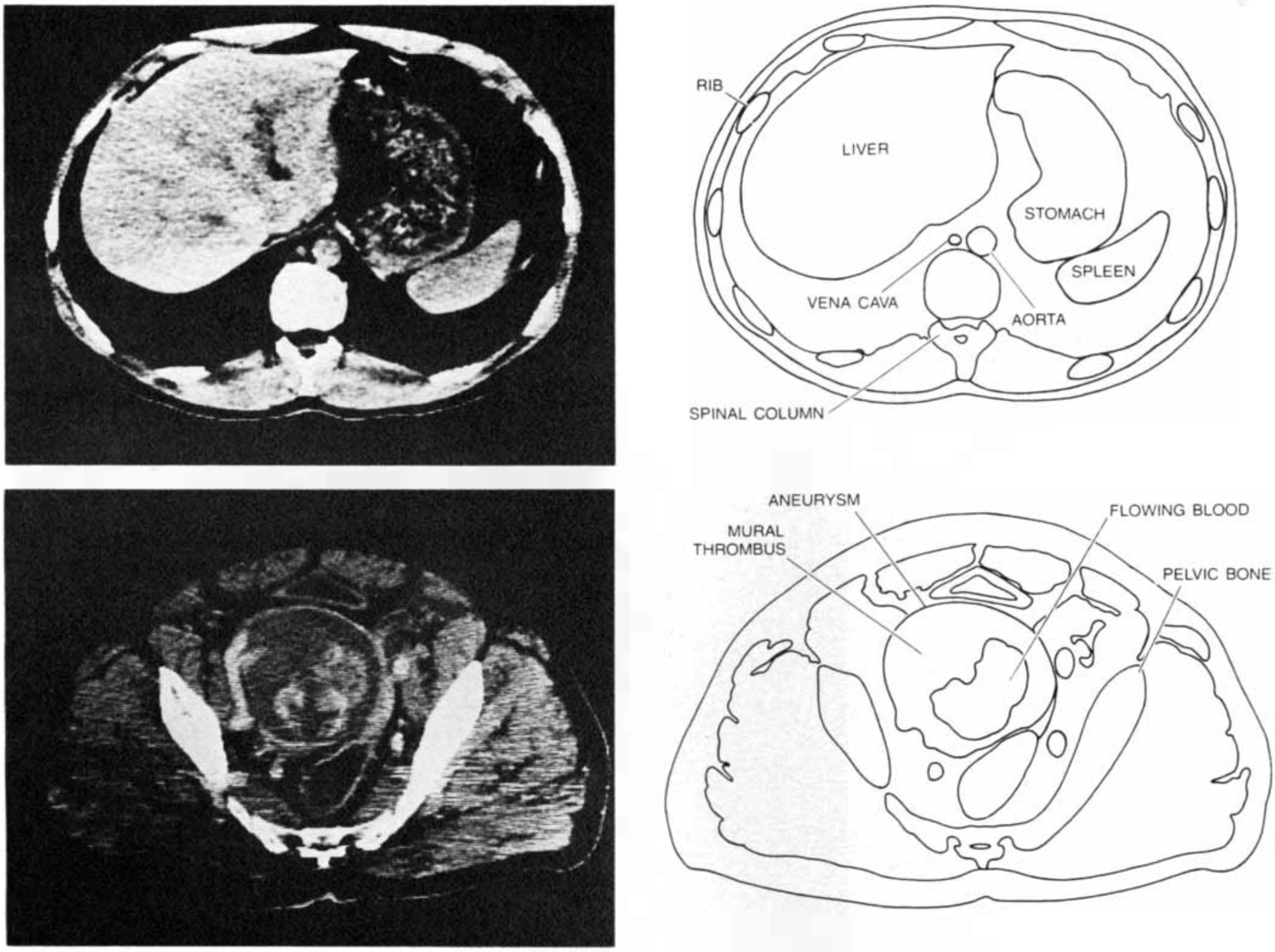
The mortality rate was unacceptably high, however. The degenerating tissue of an aortic aneurysm often becomes entangled with and even fused to neighboring structures. In many cases it was difficult to cut out the aneurysm without damaging such structures, in particular the inferior vena cava, the great vein that collects blood from the lower extremities and the trunk and returns it to the heart. In the mid-1960's many vas-



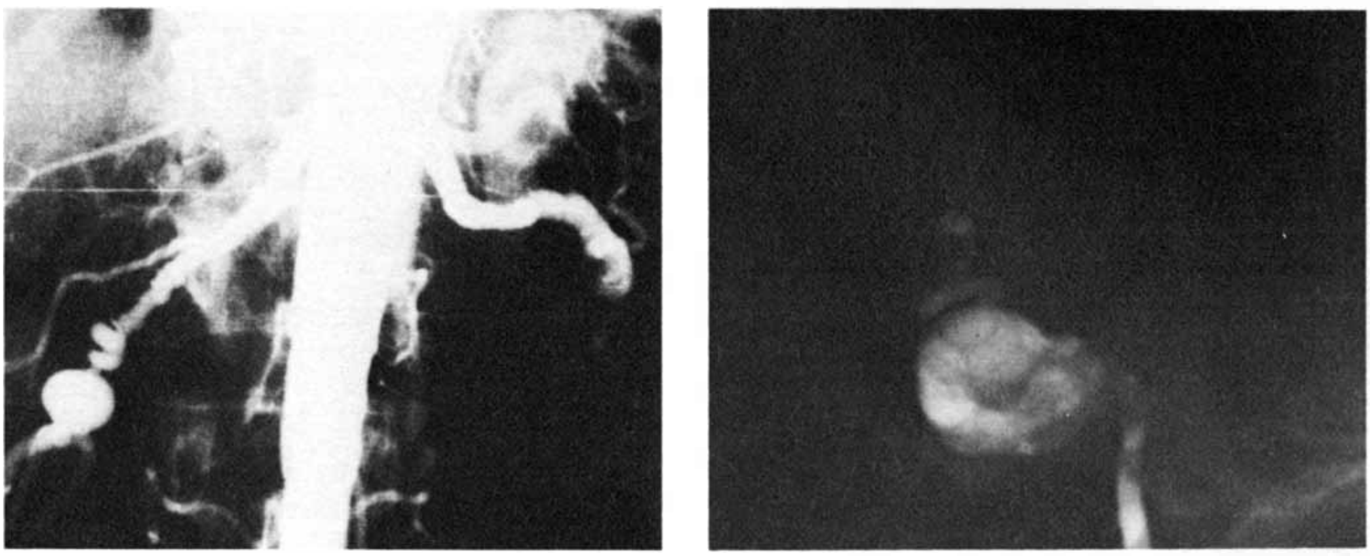
ARTERIOGRAMS delineate two abdominal aortic aneurysms, one of them (*right*) a particularly large lesion in an ectatic, or elongated, aorta. The X-ray photographs were made after the introduction into the aorta of a contrast medium (a substance opaque to X rays). In both photographs the catheter through which the medium was introduced is visible as a thin white line. The catheter was inserted into a femoral artery and was advanced through the iliac artery and the aorta to a site upstream from the aneurysm, where the contrast medium was released.



ULTRASONOGRAMS make it possible to assess the shape and size of an aneurysm accurately without having to inject a contrast medium. Ultrahigh-frequency sound waves are directed into the body and the intensity of their echo is recorded; tissues are distinguished by their differing capacities for reflecting the sound waves. These are longitudinal (*left*) and transverse (*right*) images of an abdominal aortic aneurysm, made respectively by moving the sound source down the midline and from side to side across the abdomen. The dots are one centimeter apart.



CROSS-SECTIONAL VIEWS of a normal aorta (*top*) and a massive pelvic aneurysm (*bottom*) were made by computerized axial tomography, or CAT scan. The structures seen in the scans are identified in the drawings. CAT scanning constructs an image of a cross section of the body by recording the degree of attenuation of low-intensity X rays as a source of X rays is rotated about the subject. The two scans were done at different levels: the normal aorta is seen high in the abdomen, whereas the section revealing the aneurysm is in the pelvis. Scans are displayed as transverse sections of body seen from below. Note that technique can distinguish a thrombus from flowing blood.



ANEURYSMS MAY RESULT from specific vascular conditions. Fibromuscular dysplasia (*left*) is a disorder of the media (the layer responsible for most of the artery wall's strength). The disorder gives rise to a string of small aneurysms separated by narrowed areas. The arteriogram shows such aneurysms in the renal arteries, which branch from the aorta to the kidneys. A "berry" aneurysm (*right*) can arise in a brain artery where the media is thin or absent. Unusually large one seen in arteriogram is in internal carotid artery at base of brain.

cular surgeons therefore adopted an "intrasaccular" approach, wherein the artery is temporarily clamped above and below the aneurysm and the dilated sac is opened. Then a prosthetic graft is sutured from within to undilated, healthy segments of the artery; extensive dissection outside the aneurysm is not needed, with the result that damage to neighboring structures is avoided.

Aneurysms in smaller arteries of the abdomen and in arteries supplying the head and extremities have a lower risk of rupturing than those of the aorta, but they seem to have a greater risk of thrombosis or embolization, and they too should generally be repaired when diagnosed. If the site of the aneurysm is one that has enough collateral vessels parallel to the diseased one, the aneurysm can simply be ligated, or tied off. In other cases a bypass graft is required; usually the graft is either a segment of the saphenous vein taken from the patient's thigh or a piece of plastic tubing. Aneurysms that arise from infection create special difficulties. They often result from penetrating damage, as when an artery is punctured for illicit drug injection or in diagnostic procedures such as arteriography. The area of infection must first be widely excised and drained and then blood flow beyond the aneurysm must be reestablished with a bypass graft tunneled through healthy tissue.

Cerebral berry aneurysms can be extremely hard to treat both because they are difficult to reach and because many patients already have a severe neurological deficit as a result of sudden, catastrophic intracranial bleeding. A direct surgical approach—usually an attempt to ligate the base of the aneurysm—is the appropriate treatment in most instances. Indirect approaches are also tried. One method diminishes the intraluminal pressure in the aneurysm by gradual occlusion of the carotid artery in the neck. A tiny balloon, inserted by means of a catheter and then maneuvered into the aneurysm, may stimulate thrombosis; in some cases a quick-setting glue injected into the aneurysm may work as well. John F. Alksne of the University of California at San Diego School of Medicine has investigated a technique in which the aneurysm is pierced with a needle introduced through an opening in the skull; iron filings are injected through the needle and a magnetic field is applied, again in an effort to stimulate clotting.

Along with improved diagnostic technology and surgical techniques, better understanding of the factors predisposing an individual to the development of aneurysms should continue to enhance the ability of physicians and surgeons to deal with these dangerous lesions.

The Cooperative Breeding Behavior of the Green Woodhoopoe

Among these East African birds one pair in each flock breed while the nonbreeding adults assist in the raising of the breeders' chicks. How did this pattern of seemingly altruistic animal behavior evolve?

by J. David Ligon and Sandra H. Ligon

Over the past few years investigators of animal behavior have devoted much attention to the evolution of complex social organizations. One aspect of social behavior is cooperation: exchange between individuals of some kind of resource or assistance. Cooperation is common in human societies and in general is mutually beneficial. The giving of assistance can, however, be unequal and even one-sided (as in life-sacrificing human heroism). The commonest instances of unequal exchange among the lower animals are found in the social insects. In these animal societies workers are sterile and spend their entire adult lives engaged in activities that benefit the reproduction of the queen. Often they give their lives to defend the nest. Such behavior seems to go beyond the bounds of cooperation and to meet the biologist's definition of altruism: "behavior that decreases or potentially decreases the lifetime reproductive output of the altruist to the benefit of another individual or individuals." This definition does not, of course, imply any conscious reasoning or forethought.

Although the social insects provide the most striking examples of apparent altruism, complex levels of cooperation, including apparently unequal aid giving, are also found among higher animals other than man. A vivid case in point is a species of insect-eating bird of Africa: the green woodhoopoe (*Phoeniculus purpureus*). The birds' social unit, the flock, may have as many as 16 members but only one breeding pair. The other sexually mature but nonbreeding flock members serve both as "nest helpers," sharing the burden of bringing food to the incubating female and later to the nestlings, and as "guards," defending the nestlings and later the fledglings against predators and participating throughout the year in such flock activities as protecting the home territory.

More than 100 species of birds share with the green woodhoopoe this pattern of cooperative breeding, but no single set of factors can explain why one species displays the behavior pattern and an apparently similar species does not. To understand the selective pressures that favor the evolution of cooperative breeding one must take into account the influence of the environment on a given social system. Among the environmental factors are the climate, the distribution of food both in time (with the seasons) and in space (within the utilized area), the availability and quality of roost sites and nest sites and the kinds and numbers of predators. Another important component of natural selection is what may be called the social environment: the effect on the individual of interaction with other members of its own species.

We began our work with the woodhoopoes in 1975. The birds were well suited for study. The flocks we observed inhabit acacia woodland near Lake Naivasha in Kenya, where they roost at night inside cavities in the trunks of the acacias, some of them naturally formed and some unoccupied woodpecker holes. This makes individuals easy to catch and mark; one simply plugs the mouth of the cavity after dark and then lets the bird emerge into a transparent bag the next morning. Both the openness of the woodland and the sparseness of the acacia foliage made observing the birds fairly easy; so did their generally calm disregard of human beings and the birds' comparatively large size. (The males are about 36 centimeters long and the females about 30.)

During our first period in the field, from July, 1975, through May, 1976, we color-banded 151 birds, a total that included nearly all the members of 25 flocks. We subsequently banded an additional 218 birds, most of them offspring of the first flocks we had banded. As a result our field notes include the sex, the age and the flock affiliation of the birds and also the parentage of nearly all of the second group.

The birds in our study area were highly territorial. It was unusual when a suitable habitat, for example a patch of woods that included roosting cavities, was not "owned" by one flock or another. Each territory was vigorously defended both against intrusion by birds from neighboring flocks and against individual "outsider" woodhoopoes that occasionally wandered into the area. This pattern of behavior made the establishment of new territories rare. The groups varied in size, some consisting only of the breeding pair. Generally, however, the flocks were larger, and the birds that "helped" the breeding pair were usually, although not invariably, the siblings of one breeder or the offspring of one or both birds in the breeding pair.

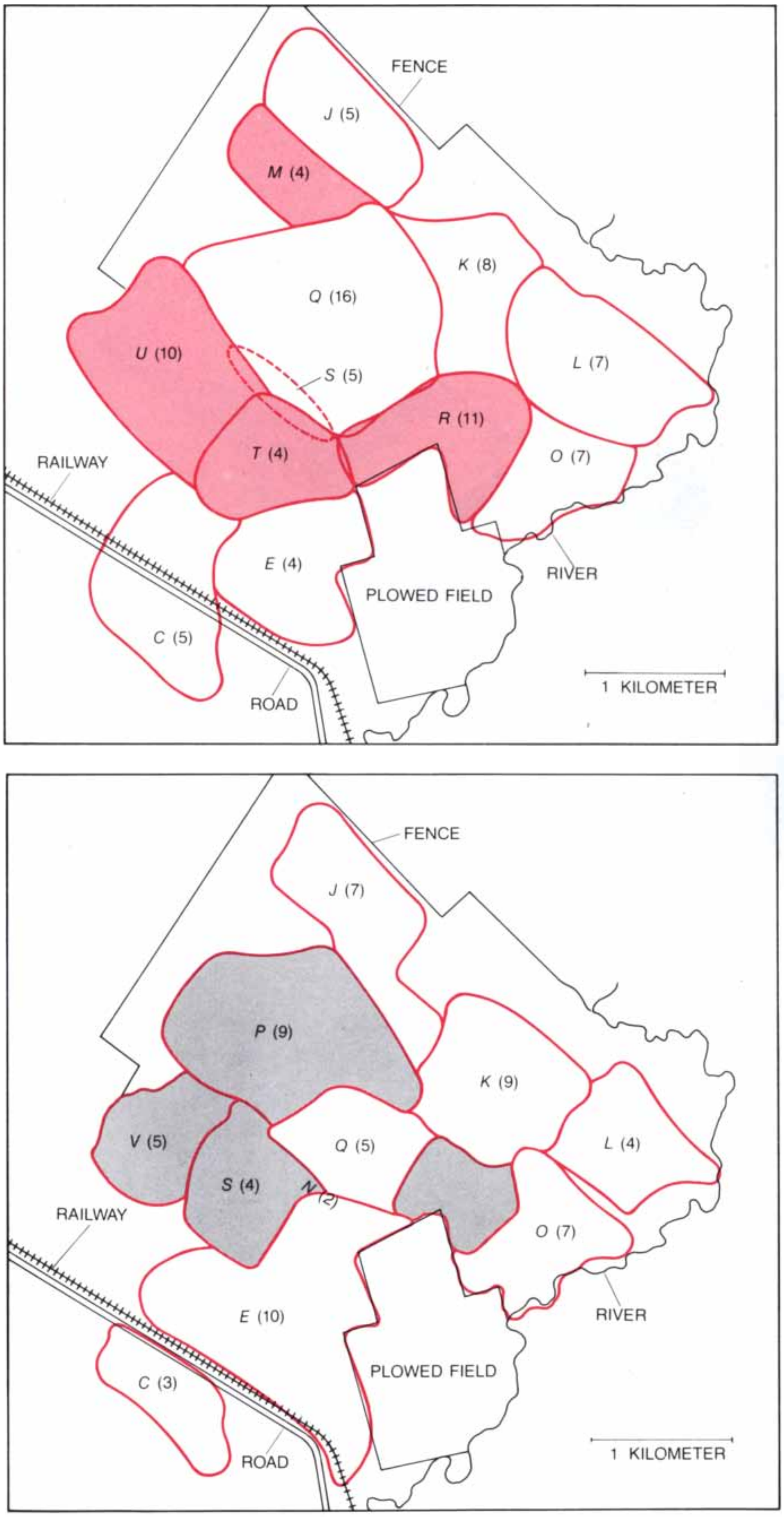
The woodhoopoes in our study area did not on the average live long. Each year between 30 and 40 percent of the population died. The distribution of the death rate in the population was uneven. Breeding males were the most frequent casualties, their deaths outnumbering those of the other three categories: breeding females, female helpers and male helpers. Most of the deaths were caused by predation. The primary predators appeared to be genets (catlike members of the mongoose family) and driver ants (the African counterpart of the New World army ants). Both predators raid the woodhoopoes' roosts at night. On several occasions we found the remains of dead birds still in the roost holes or on the ground below the nest sites.

Two environmental factors in addition to predation also have a great effect on the woodhoopoe population. The first is the timing and amount of rainfall. The rains in this region of Kenya are highly variable. The "normal" annual pattern includes a dry season that extends from December through February, followed by the "long rains," which begin in March or April. Over the seven years of our study the woodhoopoes began their breeding activity in May or June, a few weeks after the onset of the long rains.

Abnormal rainfall can sharply limit the woodhoopoes' main food supply: the larvae of some 2,000 species of moths that inhabit the woodland. If the dry season is truly dry, the pupating moths thrive in their buried cocoons. As the rains begin the moths emerge to breed and lay the eggs that give rise to the next generation. If, however, there is rain in the dry season, the pupae may rot or be destroyed by fungi and other microorganisms. With fewer pupae to become moths and lay eggs there are fewer caterpillars.

We witnessed such an unfavorable state of affairs in 1979. That year nearly 30 centimeters of rain fell in January and February, compared with a 34-year average of less than six centimeters. The woodhoopoes started nesting in June,

NONBREEDING "HELPERS" from a flock of green woodhoopoes surround a fledgling (*center*) in this painting. Two have brought the main local foodstuff, caterpillars, to the juvenile woodhoopoe and one is grooming it. The three remaining helpers, their bills agape, are vocalizing in the direction of the fledgling. A woodhoopoe flock may include as many as 16 birds, but only one adult pair breed. The nonbreeding adults in the flock take part in the upbringing of the breeding pair's offspring and also assist in the defense of the flock's territory.



OPEN ACACIA WOODLAND on a large farm in the Rift Valley of Kenya was the area selected for the woodhoopoe study. In late 1975 (*top*) the 2,000-acre site was inhabited by 11 established flocks and a 12th flock in the process of establishment (*broken boundary line to the left of center*). The numerals indicate the number of birds in that flock at the end of the year. The total was 86. The four territories shown in color no longer existed by late 1981 (*bottom*), but four new territories, shown in gray, had been established. The number of birds then was 65.

but only two of the 11 flocks in our study area managed to raise any young. Exactly the opposite happened in 1981. Only 2.1 centimeters of rain fell in January and February, and the moth pupae thrived. The long rains of March through May were unusually heavy: 36 centimeters. All 19 of the woodhoopoe flocks in our area that year, including four pairs with no helpers, nested in May and June. Reproductive success was unusually high, with all but two groups producing young. It would have been higher still if predators had not eaten two broods of nestlings. This large annual variation in food supply (and reproductive success) represents one critical environmental variable for the woodhoopoe.

Another variable, perhaps equally critical, consists of the number of roost holes available to the woodhoopoes. The birds compete for the cavities with birds of other species and also with honeybees and small mammals. The presence or absence of roost holes in a particular woodland area determines both the distribution of woodhoopoe flocks and the reproductive success of the flocks over the years. Suitable roosts are often so scarce that most or all members of each sex roost together. Such "dormitory" behavior may be a byproduct of sexual dimorphism: the difference in size between the males and the females. The females, being smaller, can enter cavities the larger males cannot. In territories where suitable roost holes were rare we have found as many as eight females roosting together in a single cavity.

This segregation by sex has an important territorial implication. A nocturnal predator can by chance eliminate most or all of the males or females in a flock in one night. Such an event leaves the flock's territory open to colonization by birds of the sex that has been eliminated; the colonizers may be either solitary individuals or members of flocks in adjacent territories.

Our years of observation yielded numerous examples of the adaptive and evolutionary significance of the behavior of the helper woodhoopoes. For example, consider the birds' high mortality rate. For a newly fledged bird to leave its parents' territory and search for a territorial vacancy elsewhere without a known roost for shelter at night is quite risky. It is also indirectly risky for the young bird's parents: in an evolutionary sense producing mature offspring is their only reason for existence. For the young bird to remain in its parents' territory in the role of a nonbreeding helper is far less risky. Since helpers are usually the offspring of one or both of the members of the flock's breeding pair, their remaining with the flock for an extended period is also indirectly beneficial in an evolutionary sense to one or both of the parents. Indeed, we view this pattern of continuing affiliation with the flock as constituting a form of extended parental care.

An observed fact supporting this interpretation of continuing affiliation by the helper birds is that there is no positive correlation between the number of helpers in a flock and the number of offspring produced by the breeding pair in the course of a season. This means the protection that continuing affiliation with the flock gives the helpers does not necessarily increase the reproductive output of the breeding pair and may sometimes even decrease it.

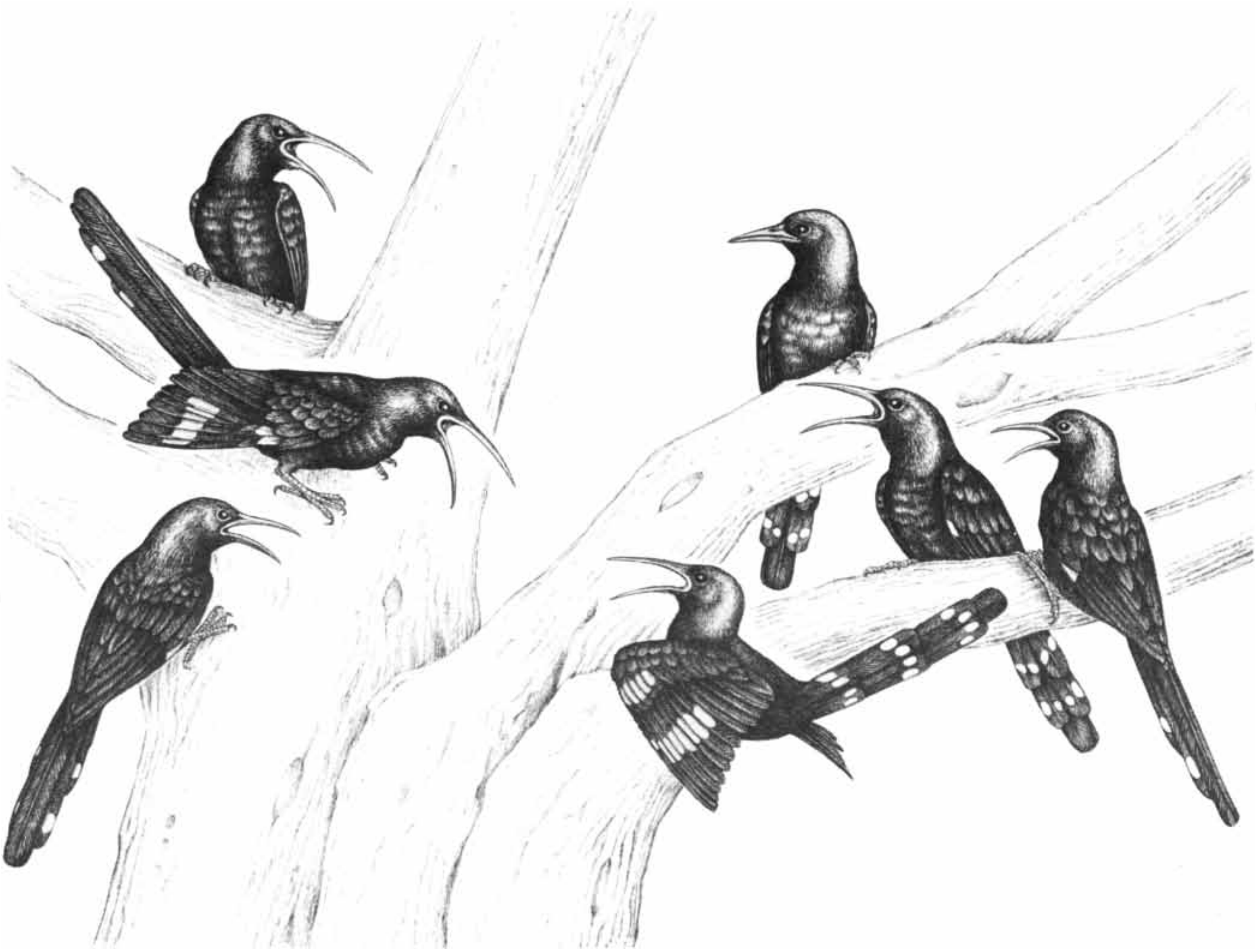
A further aspect of high mortality is the likelihood that one or another of the breeding pair in a flock will die and be replaced within a short span of time. As an example, in 1981, a year of exceptionally high breeding success, the 12 flocks in our study area included only 25 helpers. Of this total five helpers (20 percent) had the same parents as the nestlings the helpers were feeding. The degree of relatedness between the other 20 helpers and the nestlings they attended ranged from zero (three birds known not to share close relatives) to 37.5 percent (six birds with one parent in common with the nestlings and the other parents siblings). In other words, high adult mortality leads to lower levels of relatedness between helpers and nestlings than kinship theory might predict.

If one asks whether helpers have the option of being able to colonize new territory, the answer is no, at least most of the time. For example, the number of woodhoopoes in our main study area fluctuated widely between August, 1975, and August, 1981, ranging from a low of 46 birds to a high of 94. The number of territories containing a breeding pair at the onset of each breeding season, however, fluctuated far less, ranging from a low of 12 territories to a high of 14. Between 1975 and 1981 four flocks ceased to exist; their former territories remained empty. In the same period four new territories were developed.

Considering the large variation in the number of individuals over the seven-year period, the relative stability in the number of breeders and territories suggests that the opportunity to colonize a new territory is rare. A pattern of wandering—in effect searching out a new territory—with its attendant dangers is evidently uncommon because it is usually disadvantageous. These considerations help to explain why young woodhoopoes stay with their parents' flock. They do not, however, account for the birds' helping behavior.

That is not to say emigration is always disadvantageous. The adult woodhoopoes' high mortality rate means that neighboring territories have frequent breeding vacancies. As a result the commonest way for a mature nonbreeder of either sex to attain breeding status is to emigrate to a territory where a vacancy has been created by the death of a breeder of the same sex (and of mature nonbreeders of that sex, if any).

Emigrants seldom made these shifts alone. We observed that a successful move into a strange territory usually involved a team of two or more birds of the same sex. Such teams were generally composed of flock mates but they never included nest mates; antagonism between former nest mates is strong and



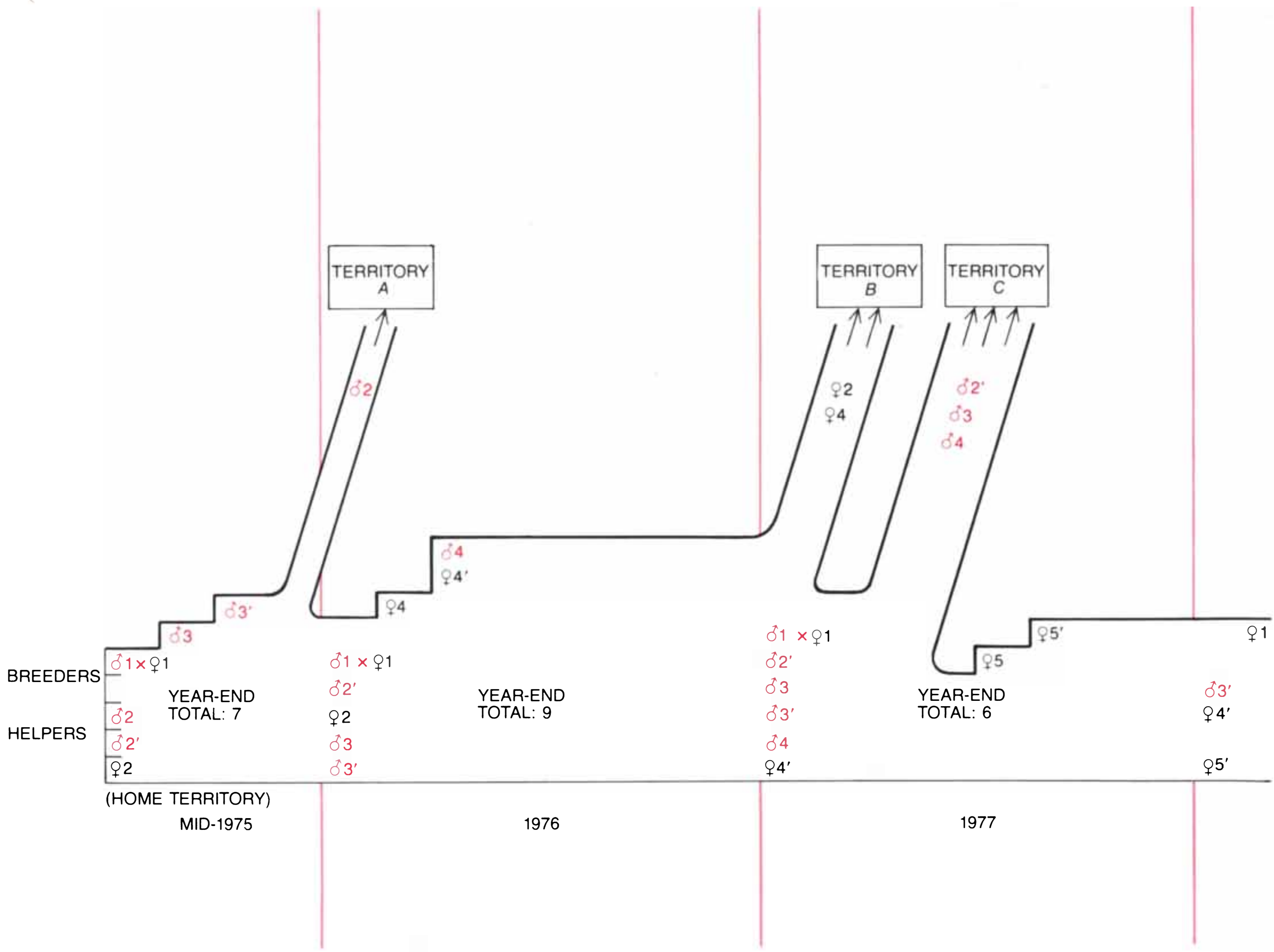
TERRITORIAL DISPUTE depicted in this drawing based on observations in the field brings three birds from one flock to a confrontation with four from a neighboring flock. The dispute is chiefly vocal. One bird in each group, however, stands with wings loosely held.

they do not emigrate together. As a result the emigrant teams consisted of an older nonbreeder (we labeled them alpha birds) and a younger one or two (beta and gamma birds). A subtle dominance hierarchy, positively correlated with greater age, was apparent. We found that the alpha emigrant became the breeding replacement in the adopting flock. If the alpha emigrant then died, the beta emigrant inherited the alpha's breeding status.

The observed facts of team emigration, with the dominant older partner a helper that had been involved in rearing the subordinate younger partner or partners, gave us an insight into the evolution of helper behavior. For a nonbreeder to achieve breeder status by emigration—one of the two roads to survival in the genetic sense—it was first necessary for it to acquire one or more subordinate allies. Could this be, at least in part, why helpers help? Nothing in this evolutionary process need involve the exercise of reasoning: the genes of unhelpful and therefore unallied nonbreeders are at risk of being eliminated from the gene pool of the species within a single generation.

How is the bond established between the helper and the nestling it helps? We found that helpers not only brought the nestlings food but also stole caterpillars from other helpers in order to deliver them to the nestlings they tended. Helpers interacted with nestlings in other ways. For example, they groomed them. They also perched nearby and vocalized with their attention directed to the nestlings. Neither these activities nor the fetching of food seems to represent a costly investment in time and energy compared with the high genetic return that may result: achievement of breeding status.

The fact that older nonbreeders may benefit in this way still leaves a key question unanswered: Why do younger birds accompany their elders when the opportunity to emigrate arises? Not only do the betas help the alphas to become established in the new territory but also they take up the burdens of a helper when the offspring of the alphas are hatched. The answer again seems to lie in the pattern of woodhoopoe mortality. Without implying any reasoned behavior, an anthropomorphic description of the young birds' actions would not be



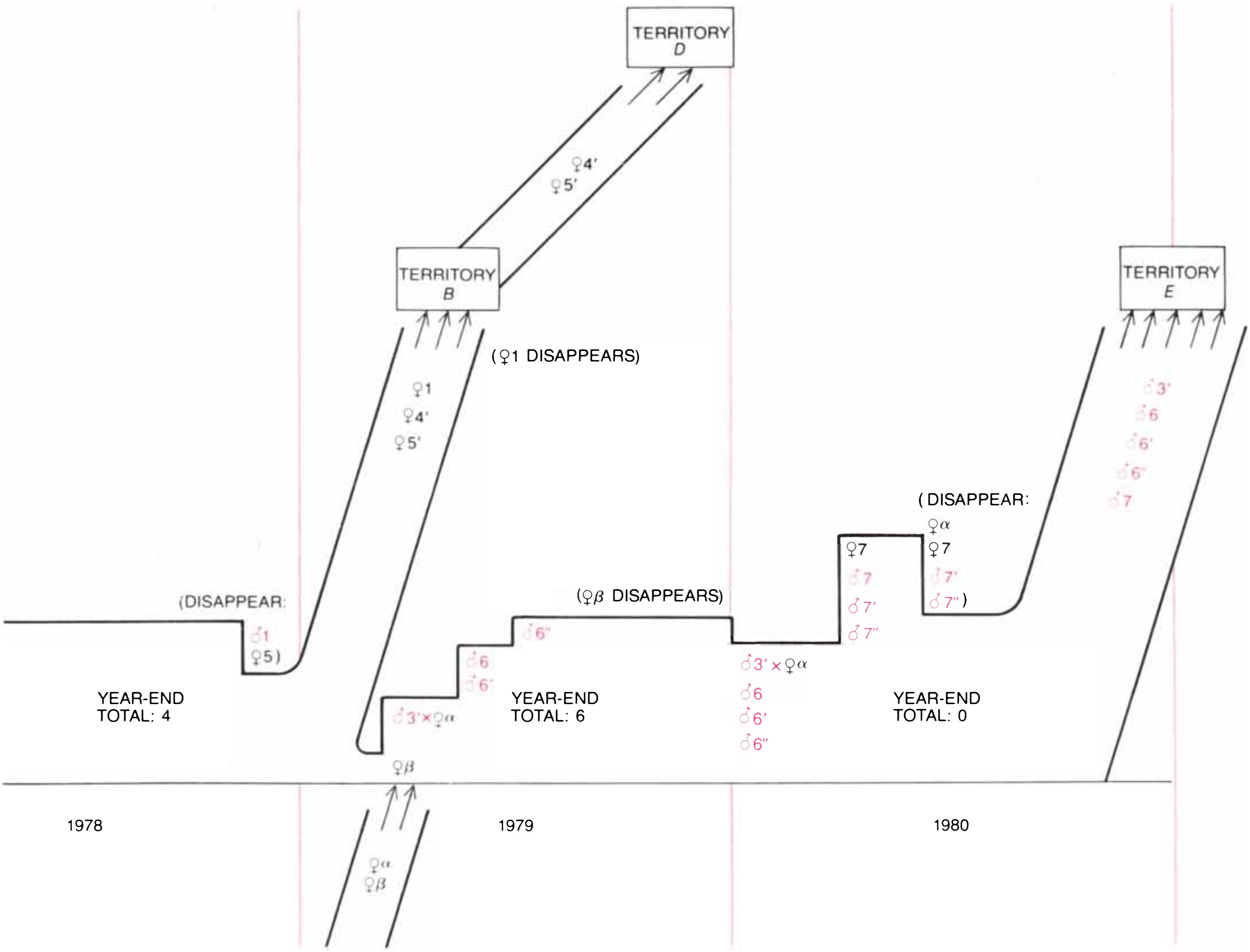
SHIFTING MAKEUP of one woodhoopoe flock over a 66-month interval is shown in this diagram of the individual birds' breeding successes, emigrations, immigrations and disappearances. When the birds were first banded, in a territory adjacent to the main study area, the flock included a breeding pair (*male 1 and female 1; the numbers identify them as the first observed generation*) and three other mature birds of a second generation (*males 2 and 2' and female 2*), all probably offspring of the breeding pair. In the 1975 breeding season two more birds (*males 3 and 3'*) were added to the flock. Early in 1976 male 2 emigrated alone to Territory *A*. In the 1976 breeding season three offspring (*female 4 and males 4 and 4'*) were added to the flock. Early in 1977 two birds (*female 2 and female 4*) emigrated to Territory *B*. This accorded with the normal pattern of emigration: an older bird being accompanied by a younger subordinate of the same sex. Later in 1977 three more emigrants left the flock for Territory *C* (*male 2', male 3 and male 4*). This left behind the breeding pair, a

that they were "paying back" the older birds for their earlier help but rather that they were "playing the odds" in expectation, as it were, of their own genetic reward.

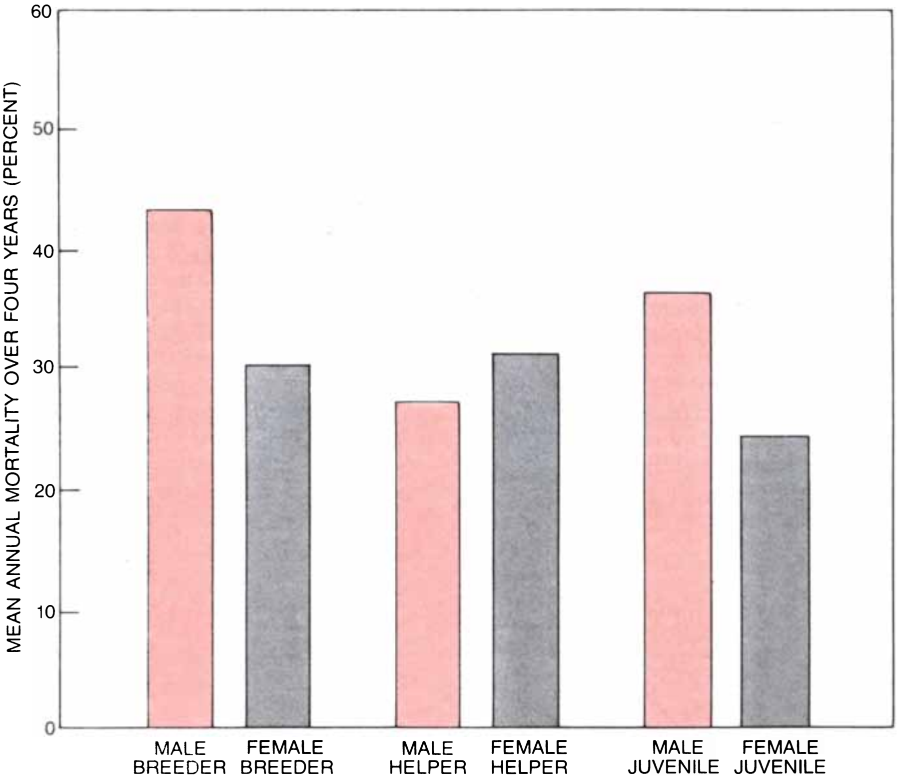
For example, male alphas, it will be remembered, have a significantly higher mortality rate than male betas. As a result subordinate males will frequently outlive their dominant allies and then themselves inherit breeding status. We observed that among groups of two to four emigrant males, and among newly formed flocks composed of only two adult males and one or two adult females, in 10 out of 13 instances the alpha male died before the younger males. We also found that over their lifetime once-subordinate emigrant males left as many surviving offspring as the originally dominant males. This observation bears crucially on the evolutionary concept known as individual fitness. With regard to the beta birds' behavior, as with regard to the alphas', there is no reason to suppose that when they emigrate with older flock mates they are in any way acting contrary to their own long-term reproductive interests.

In this connection we also observed that some female helpers, unlike male helpers, voluntarily left their original territory unaccompanied by subordinates and wandered over a large area apparently searching for a territorial vacancy. This behavior may arise from the fact that the mortality rate for female breeders is significantly lower than that for male breeders. Therefore the likelihood of a female helper's achieving breeder status at an early age in the home flock is remote, and seeking breeder status somewhere else, although it is risky for a solitary bird, may be genetically rewarded.

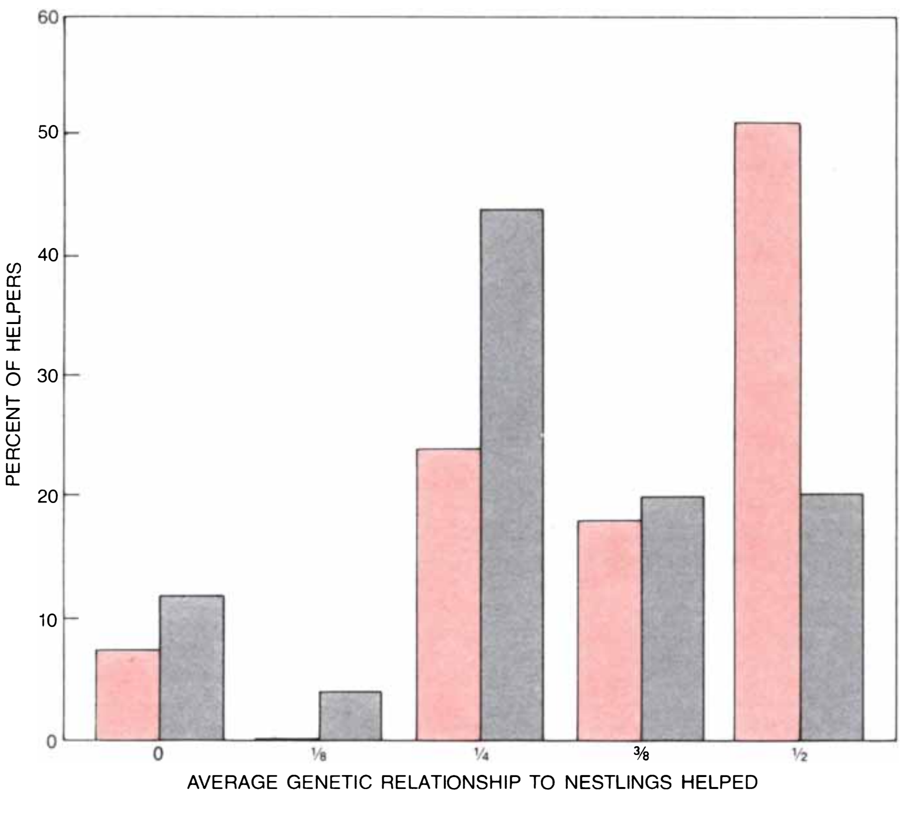
Since any individual bird's future is uncertain, it is genetically advantageous for breeding woodhoopoes to produce behaviorally variable young. The point can be illustrated by observations we began to make in 1977. That year three out of four male-sibling helpers in a flock composed of the usual breeding pair and five nonbreeders emigrated to a neighboring territory. There they drove out the lone male occupant and joined three females. The fourth male helper did not emigrate. Two years later the male parent died and the



third-generation male and a fourth-generation female. Late in the year the breeding pair nested twice, raising two fifth-generation females. This was the last season for male 1; it disappeared in 1978, as did one of the young females: female 5. No offspring were produced that year. In mid-1979 the breeding female, female 4′ and female 5′ emigrated to Territory *B*, leaving male 3′ as the sole occupant of the home territory. The emigrant mother soon disappeared and the two younger females moved on to Territory *D*. Two female emigrants from an adjacent flock (*females alpha and beta*) now joined male 3′. At the very end of the year male 3′ and female alpha became the second breeding pair in the home territory. They produced the sixth generation: males 6, 6′ and 6″. In 1980 they produced four more offspring: female 7 and males 7, 7′ and 7″. Both female alpha and female beta disappeared that year, as did female 7, male 7′ and male 7′. The breeding male and its offspring (*males 6, 6′, 6″ and 7*) then emigrated to Territory *E*, leaving the home territory deserted.



HIGH MORTALITY RATE among woodhoopoes over a four-year period (January, 1976, through December, 1979) was not evenly distributed either by sex or by social status. Breeding males and juvenile males led in fatalities; juvenile females and male helpers trailed. The rates for breeding females and female helpers were much alike and fell between those of the others.



TIES OF KINSHIP within individual flocks were increasingly diluted over time. This graph compares the relationship between helpers and the nestlings they cared for in 1978 (*color*) and 1981 (*gray*). In the earlier year 14 flocks included a total of 55 helpers; in the latter year nine flocks included a total of 25 helpers. More than 50 percent of the helpers in 1978 shared common parentage with the nestlings. By 1981 only 20 percent of the helpers did so. Fewer than 25 percent of the helpers in 1978 were the equivalent of half-siblings of the nestlings and none were the equivalent of first cousins. By 1981 these relationships were close to 45 and 5 percent.

fourth male helper's mother and two sisters subsequently emigrated, leaving the male helper as the sole inheritor of his parents' territory.

This male soon mated with one of two female emigrants from a neighboring flock. In late 1979 and in 1980 the new breeding pair raised a total of seven offspring. In contrast to this record of reproductive success two of the male's three emigrant siblings had died by the end of 1979 without ever having produced young. The youngest of the three was still alive in early 1982, but it has fledged only one offspring. If the fourth male sibling had not stayed at home, its contribution to the woodhoopoe gene pool might easily have been nil.

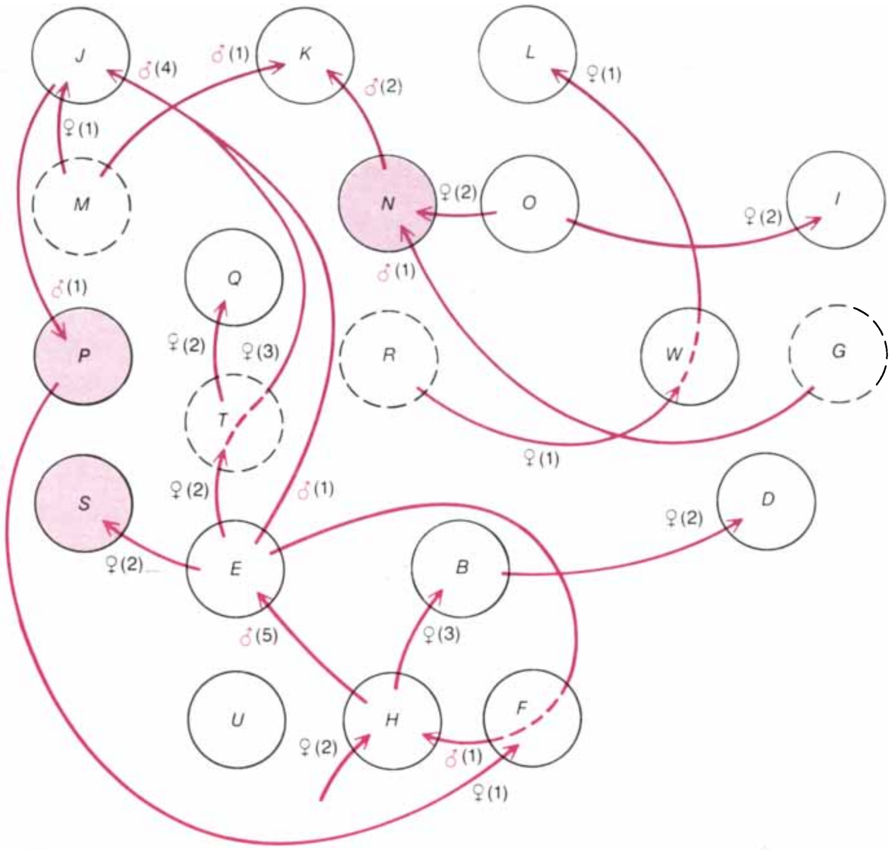
Although many emigrant groups consist of siblings, we also observed unrelated woodhoopoes of the same sex merging to provide the nucleus of a new flock. Typically this happens when an older, dominant bird allows a younger, subordinate individual to join it. We observed such behavior only when an older male needed an ally to defend and hold its territory. Two females would occasionally allow a third unrelated female to join them under the same circumstances. Of 40 observed instances of the formation of a new flock (or the replacement of all the members of one sex or all but one member of one sex) 17 of the groups included nonrelatives of the same sex and 17 included known or probable relatives of the same sex. In six instances kinship ties, if there were any, were not known to us.

Let us now suggest how cooperative breeding behavior may have originated among the green woodhoopoes and how this behavior in turn may have led to reciprocal assistance among unrelated individuals of the same sex. First, as we have seen, three features of the woodhoopoes' environment and life cycle appear to have set the stage for the development of cooperative breeding by favoring those breeding pairs that allowed at least some of their offspring to remain indefinitely in the parental territory. The three features are a high adult death rate, an unpredictable birth rate due to a fluctuating food supply and the uncertain availability of roost sites.

The extended term of residence for adult offspring confers two benefits on the breeding pair. The first of them is the opportunity for territorial expansion available to a flock that consists of more than two adults. In this uniform acacia-woodland environment a larger territory means an increased foraging area. Similarly, the presence of more than two adult birds means that territorial defense weighs less heavily on the individual adults.

The second benefit of an extended term of residence is whatever assistance the young adults may offer the breeding pair in feeding and protecting their nestlings. The presence of helpers probably reduces the cost to the breeding pair of annual nesting, thereby increasing the pair's potential lifetime reproduction record. Then when opportunities to occupy new territory arise for the adult nonbreeders, these emigrants find their chances of success enhanced by the availability of allies from among their younger flock mates of the same sex.

Once this interdependence for territorial defense and expansion, for reproduction and later for the acquisition of new territories became established among related individuals, any unaided woodhoopoe pairs or single birds were placed at an overwhelming competitive disadvantage. If most such birds are ever to gain territory for themselves and eventually to breed, the best option open to them once a mate is found is to obtain unrelated allies of their own sex. This they do by allowing unaffiliated younger and subordinate birds to join them.



EMIGRANTS' MOVEMENTS among 20 woodhoopoe territories during a four-year period (1978–81) are shown in this schematic diagram. Where the territories are the same as those in the illustrations on pages 128 and 130–131 the identifying letters correspond. Thirty-one individual migrations are shown, involving flights by 14 males and 17 females. Birds that enter and leave one territory en route to another (for example from *E* to *J* by way of *T*) are counted only once. The same birds, however, may move more than once. For example, two of the three females that went from *H* to *B* subsequently moved from *B* to *D*. The four territories with broken perimeters were abandoned and three shown in color were colonized during this period.

How well do the facts of the woodhoopoe social system correspond to the various theories that deal with cooperative behavior? Three theories are particularly relevant: Darwinian individual selection, kin selection and reciprocity. The theory of individual selection, as set forth by Charles Darwin, is based on the premise that animals behave in a way that maximizes opportunities for them to produce their own offspring. In the context of cooperative breeding among birds, a nest helper of either sex might eventually gain valuable resources such as territory, for example by first serving as an apprentice to the older, socially dominant breeders.

Kin selection, an extension of individual selection, is a theory developed (primarily by W. D. Hamilton of the University of Michigan at Ann Arbor) to account for what seemed to be altruistic behavior. In order to explain the apparent altruism seen in nature Hamilton proposed that if the unselfish behavior shown by an individual was directed toward a relative, it would serve to promote genes shared by the altruist and the recipient of the altruist's aid, so that the costs and benefits of such behavior should be correlated with the degree of genetic relatedness between the two interacting individuals.

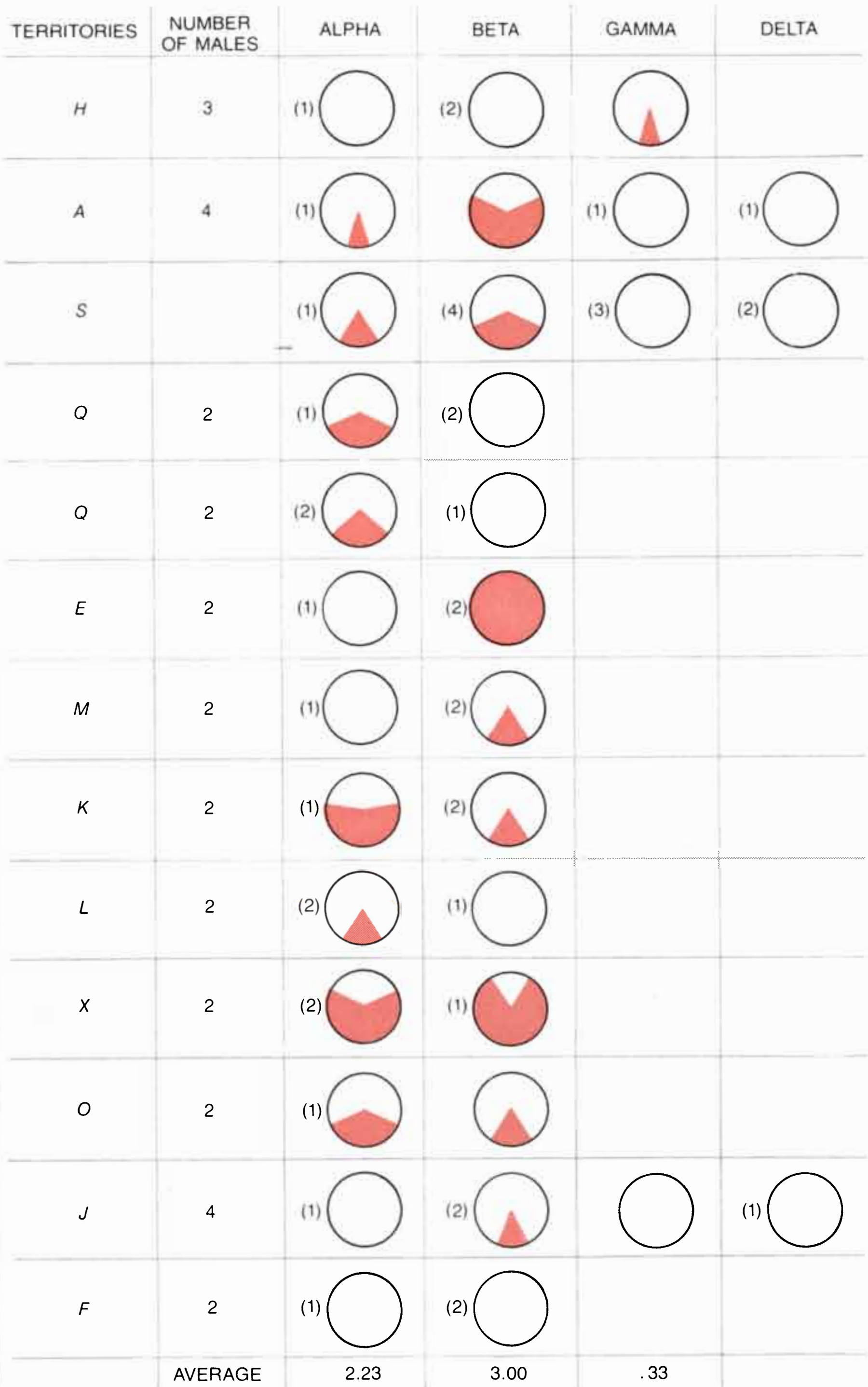
The idea behind the third theory, reciprocity, is that an individual provides aid to another individual with the expectation that it will be repaid (at a value equal to or greater than that of the aid but not necessarily in the same form). Reciprocity is a specialized means of obtaining benefits for the individual and thus does not differ fundamentally from Darwinian individual selection.

Now, because the majority of the birds that interact within the woodhoopoes' closed social unit are relatives, it could be argued that what has been crucial to the evolution of the system is kin selection: natural selection where animals preferentially aid relatives, perhaps at a cost to themselves. Indeed, one kinship bond has clearly been fundamental to the evolution of the woodhoopoe social system: the bond between parent and offspring. It is not clear, however, whether other kinship ties are necessary to the maintenance of the system. For example, in three to four flocks each year since at least 1977 we have observed helpers that were not related to the nestlings they were tending. Some of these nonkin not only fed the nestlings but also fed the female member of the breeding pair while she was still incubating the eggs. Moreover, the merger of unrelated individuals either to increase the size of the flock or to form new flocks suggests that genetic relationships are not all-important to the woodhoopoes' forms of cooperation.

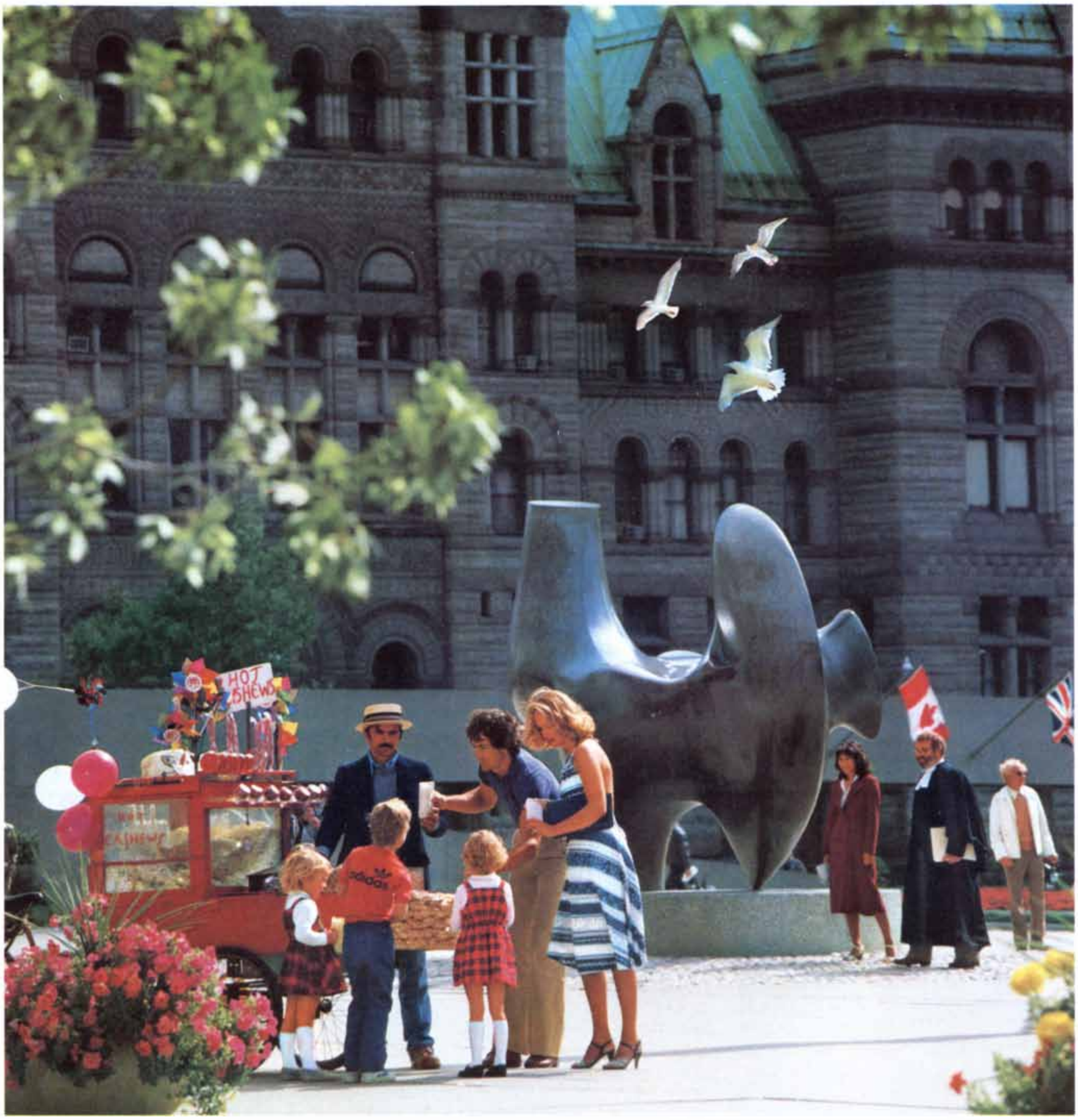
Reciprocity of various kinds occurs among flock mates, whether they are related or not, and it seems to be an important component of the woodhoopoe social system. As we have seen, nonbreeding members of the flock both serve as helpers and aid in territorial defense. In return these individuals can benefit in two ways. First, if the helper outlives the breeder of its own sex in the flock, it may inherit breeder status. Second, as a result of interaction with the developing young the helper wins potentially valuable allies in the flock in the event of emigration.

Moreover, when an older helper emigrates to take advantage of a territorial breeding vacancy, it is usually accompanied by one or more younger allies. By helping to establish the older bird's position in the new flock the younger birds can be thought of as repaying the older bird for the services they received in the nest. The younger birds will also tend the older bird's offspring, thereby establishing a kind of cross-generational reciprocity because the fledglings the older bird feeds may eventually help the helper to advance its own breeding program. The merger of unrelated adult birds is also suggestive of reciprocity; because of the high death rate this form of cooperation may often lead to an eventual breeding status for all involved.

Finally, what appears to be the primary underlying basis for the woodhoopoes' complex behavior is Darwinian individual selection: the maximizing of individual reproductive opportunities with respect to the other members of the population. The various woodhoopoe strategies I have described are most parsimoniously explained as behavior that has evolved to maximize these opportunities under an unusual and unpredictable set of circumstances.



REPRODUCTIVE SUCCESS of emigrating male woodhoopoes is plotted. Nine pairs of males and four larger male groups emigrated from the territories identified at the far left. An empty pie chart in any of the four columns indicates lack of breeding success for the bird in question. (Alpha is the oldest male, beta the second-oldest and so on.) A full pie chart indicates the maximum breeding success observed: 11 sired by the beta bird from Territory *E*. The extent to which the pie charts are filled elsewhere is proportional to the breeding success of that bird. The numeral beside each pie chart indicates the order of death in the group; for example, of the four emigrants from Territory *A* three died simultaneously. The beta male in that group was still alive when observations were concluded, and so its pie chart is unnumbered. Although the alpha males were the first to breed, the beta males' average reproductive success was greater.



Nathan Phillips Square, Toronto, Ontario/Canada

Toronto à la carte!

Savour the old world flavours, the green and gracious public spaces, and the gleaming new world sights of our capital city. Take a clip-clop calèche or a trolley ride or a water slide on our island playgrounds. Come for the concerts and cabarets, the creature comforts, the urban delights and cosmopolitan pleasures. The options are endless, and yours to discover! Call TOLL FREE 1-800-828-8585 or from New York State 1-800-462-8404 or write: Ontario Travel, Queen's Park, Toronto M7A 2E5, Ontario/Canada.

ONTARIO yours to discover!

Renaissance Intarsia: The Art of Geometry

Intarsia is wood inlay. It flourished in the Renaissance, when it notably embodied the development of a mathematical procedure for representing the three-dimensional world in two dimensions

by Alan Tormey and Judith Farr Tormey

In the *Zibaldone,* a remarkable commentary on art and life in 15th-century Florence, Giovanni Rucellai notes that of all the arts being practiced with distinction, one art in particular is worthy of special attention: the art of intarsia. Indeed, Rucellai lists intarsia ahead of painting among the glories of his city and his age. In 1478 there were 84 workshops in Florence alone where intarsia was the principal occupation, and in the Florentine archives of the period the practitioners of intarsia, the *intarsiatori,* were listed as "masters of perspective."

Intarsia, or wood inlay, was a central artistic phenomenon of the latter half of the 15th century and the first quarter of the 16th. The art, or craft, of intarsia (the distinction was controversial at the time) was transformed in the mid-15th century from the peripheral status of a decorative embellishment of architectural borders into what may justifiably be considered the geometrical art par excellence. Intarsia panels frequently represent street scenes and architectural complexes (real or imaginary) in perspective, as if they were seen through an open window. Many panels convincingly depict the contents of a cupboard whose doors are ajar. Virtually every panel is an illusion of a three-dimensional prospect. The illusion is created by the application of geometrical laws. The sudden flourishing and the subsequent fortunes of intarsia coincided with the Renaissance effort to give art a mathematical basis, and the story of its development dramatically exemplifies the Renaissance fusion of art, mathematics and philosophy.

The practice of intarsia required the selection and cutting of a large number of variously colored and textured pieces of wood such as ebony, cypress, boxwood, walnut and fruit woods, in accordance with a cartoon, or preliminary design. The resulting precisely shaped fragments were embedded in or placed on a backboard of walnut, glued in place and varnished. Giorgio Vasari remarks in *Lives of the Most Excellent Architects, Painters and Sculptors* (first published in 1550) that the work begins with designs in perspective, because the forms of these designs meet at plane angles. The pieces of wood joined together show the contours of the figures, and the intarsia panel appears to be one flat piece, although it is made up of more than 1,000 pieces. In many of these works the natural qualities of the wood were modified by staining or burning to enhance the sense of contrast and the illusion of depth. It is worth noting that the procedures for preparing and varnishing wood perfected by the *intarsiatori* were later adopted by the master violin makers of northern Italy.

It was not by coincidence that the artistic activity of the *intarsiatori* was most apparent in and around Florence, because the most decisive influence on the development of intarsia was the emergence of the theory of linear perspective, a means of producing two-dimensional representations of the three-dimensional visual world. The principal architects of the theory were the Florentines Filippo Brunelleschi (1377–1446) and Leon Battista Alberti (1404–72). Alberti's *Della pittura* (1436) contains the first published account of the theory. He writes, "I draw a rectangle, which I regard as an open window through which the subject to be painted is seen" [*see illustrations on page 139*]. Next he places a standing human figure in the foreground. The height of the figure he divides into three parts; then he marks off the resulting unit of measure, a *braccio,* along the baseline of the rectangle. He chooses a "centric point" (we would call it a vanishing point) level with the top of the figure's head and connects it to each point marked off on the baseline. He also draws through the centric point what we would call the horizon line.

The subsequent steps are obscure in Alberti's writings, but a succession of art historians have formulated plausible reconstructions. Evidently Alberti extended the horizon line on one side of the rectangle and chose a "distance point" on it. With diagonal lines he connected the distance point to the points marked off on the baseline. The intersections of these diagonals with a side of the rectangle determine a set of horizontal lines. A grid (interpreted, for example, as a checkerboard floor that recedes into the distance) is now available to receive the pictorial representations of various three-dimensional forms.

The origin of Alberti's theory (and more broadly of Renaissance theorizing about the visual arts) was in large part medieval optics. From antiquity through the Renaissance there had been a debate over the origin of the rays by which objects are perceived. Defenders of the "extramission" theory held that rays issue conically from the observer's eye; the "intromissionists" argued that "visual rays" proceed from the object to the eye. Medieval theorists on each side of the debate agreed, however, that the "centric visual ray" at the middle of the network was the ray by which the world was perceived most clearly. The centric ray became a symbol of the moral power of God, and it was thought knowledge of one would reveal essential aspects of the other.

In effect the Renaissance artist-theoreticians transformed that medieval theory of vision into a theory of pictorial representation. The geometrical link between the two was supplied by the "pyramid of sight." The pyramid was formed by the rays that entered or left the eye. (It made no difference to Alberti, since the geometry of the pyramid was the same.) Thus the pyramid had its apex at the eye. Its base coincided with what the eye saw. Along the axis of the pyramid passed the centric visual ray, which Alberti called "the prince of rays." In Alberti's fully formulated method it connected the vanishing point to the eye.

The intent of Alberti's method was to determine the projection of what the eye sees at the base of the pyramid of sight onto a plane that passes vertically through the pyramid. Hence the problem of representation was formulated as a problem in plane geometry. With the formulation in place the construction of visual works of art was given a theoretical basis comparable to the long-envied mathematical foundation of music devised by the Pythagoreans, who had discovered in about 550 B.C. the numerical ratios underlying musical intervals. The importance of mathematics was institutionalized in the medieval division of the seven liberal arts into a lower three, or trivium, consisting of grammar, rhetoric and dialectic, and a higher four, or quadrivium, consisting of arithmetic, astronomy, geometry and music.

For Alberti and his contemporaries, including the artists Paolo Uccello and Piero della Francesca, depicting the seen world was a means of investigating the natural laws underlying visual perception and of expressing their knowledge of those laws. It is not surprising that in this atmosphere the concept of pictorial fidelity was enshrined as the dominant artistic paradigm. The work in intarsia from this period is a perfect reflection of that commitment.

The story of the development of the theory of linear perspective and its manifestations in Renaissance art has been frequently if not always consistently told. The major monuments of painting and architecture that exemplify the theory are traditionally cited. It is nonetheless clear that intarsia was the art form of the Renaissance whose fortunes were most intimately linked to the concept of a mathematically based representation of the world. From the outset the practice of intarsia had required a sophisticated understanding of applied geometry and precise means of measurement for the cutting and fitting of the pieces of wood that made up the final composition and gave it its characteristic depth, shading and texture. It was only with the development and codification of the theory of linear perspective, however, that the *intarsiatori* were able to develop their interest in illusionistic representation and to produce the most persuasively illusionist and symbolically suggestive art of the period. It appears, in this instance at least, that a theory was needed in order to elevate a craft to the status of an art.

Because the *intarsiatori* were artists, practical geometers and carpenters, it is understandable that pride in their skills is frequently reflected in the portrayal of geometrical instruments and the tools of inlay. The depiction of the tools would

OPEN CUPBOARD in an intarsia panel made in about 1519 by Fra Giovanni da Verona displays the illusion of three dimensions that is typical of intarsia. With the exception of the raised border of the frame the panel is a flat assembly of inlaid wood. Uppermost in the cupboard is a 72-sided polyhedron, which was a symbol of architecture. Hanging in the center of the cupboard are instruments employed by geometers and by the practitioners of intarsia: the compass, the ruler, the set square and the right-angled triangle. The Greek inscription (which is also inlaid) reads: "These are the instruments of inlay." At the bottom of the cupboard is a *mazzocchio*, which was both a headdress for Florentine men and a symbol of solid geometry.

ANOTHER OPEN CUPBOARD was also executed in intarsia by Fra Giovanni. At the top is a vase containing hazelnuts and apples. It is in essence an early still life. Below it is a dodecahedron rendered so that the figure is a lattice and each of its faces is the base of a pyramid. By these means Fra Giovanni displayed his mastery of perspectival rendering. This panel and the one on the preceding page are now in the Monastery of Monte Olivetto Maggiore near Siena.

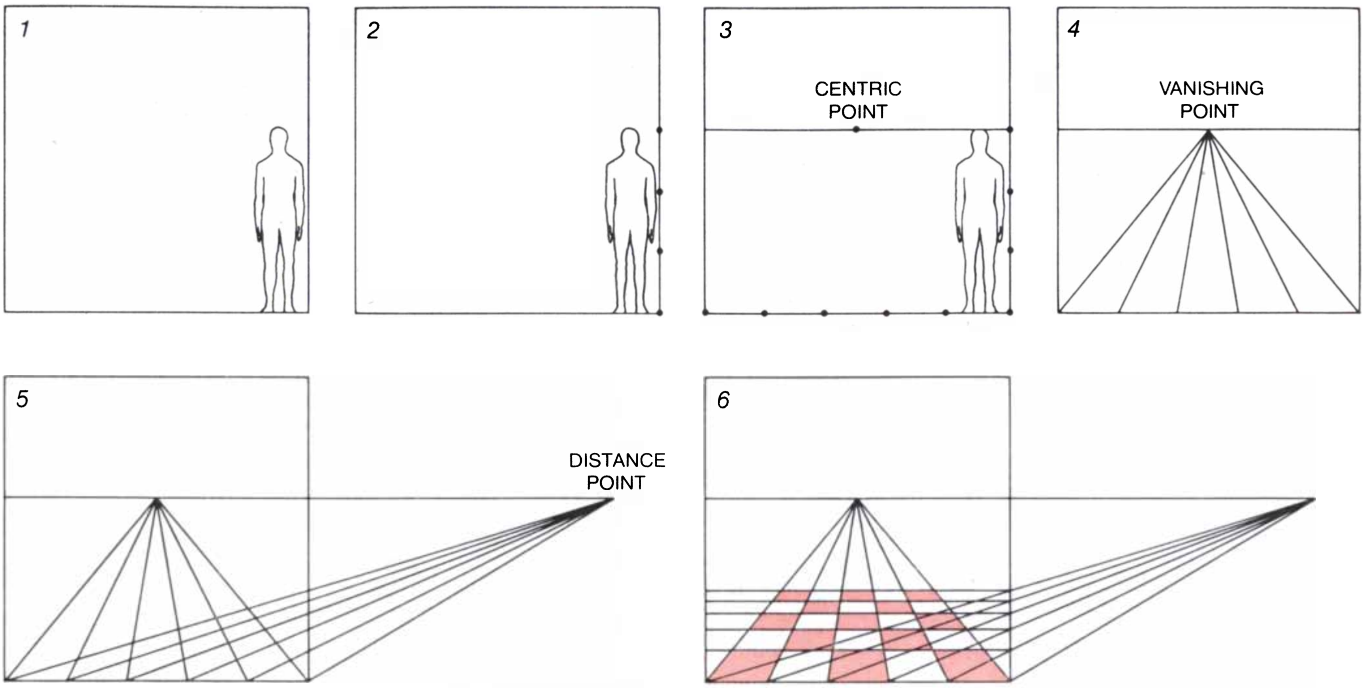
require their use in their own representation. Thus in a number of the most impressive and interesting works of intarsia, geometry is not only employed in conception and execution but also internalized in representation. Intarsia is distinguished as an intriguing instance of systematic self-reference, a property that is of particular interest in the philosophies of art, language and mathematics.

One of the greatest practitioners of intarsia, the Olivettan monk Fra Giovanni da Verona, constructed panels that incorporate geometry explicitly. He represented geometrical objects, in particular various polyhedrons. Of special importance for an understanding of the intense interest in polyhedrons during the Renaissance is the influence of the Platonic Dialogues, particularly the *Timaeus.* Although many of Plato's works had been translated into Latin and were known to Renaissance scholars, it was the *Timaeus* that most affected the emerging theory that there was an isomorphic relation between the mathematical foundation of pictorial representation and the mathematical structure of reality. Commentaries supplied by translators of the *Timaeus* contained arguments for this intellectualized conception of art.

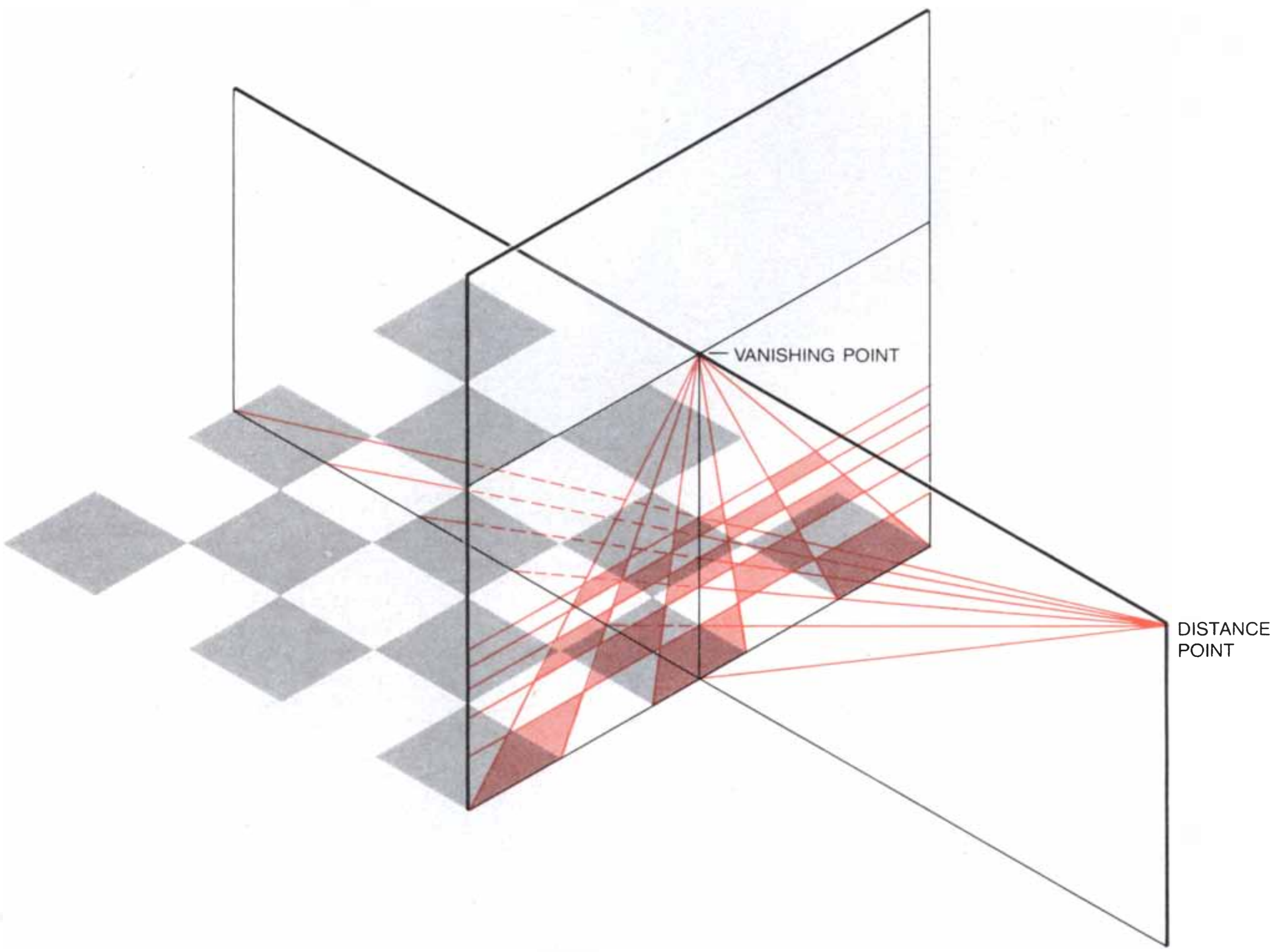
In the *Timaeus* four of the five regular solids (which have been called the Platonic solids) are linked to four fundamental metaphysical elements: the cube is the form of earth, the octahedron of air, the tetrahedron of fire and the icosahedron of water. The fifth regular solid, the dodecahedron, is said to be the form of the universe. The Renaissance mathematician Luca Pacioli (ca. 1445–1514) echoed this Platonic theme when he made the same associations in his *De Divina Proportione* (Milan, 1498; Venice, 1509).

Under the influence of the *Timaeus* Piero della Francesca sought to reduce all visual appearances to constructions consisting of the five regular solids. After all, the Platonic solids reflected the design for the construction of the universe that was employed by Plato's "master craftsman" (the *Demiurgos*); thus they became symbols of divine perfection and order. Both Alberti and the philosopher-mathematician Nicholas of Cusa (1401–64) maintained that mathematical symbols were the best way to represent God, since mathematical forms were permanent, universal and indestructible.

Some of the most striking panels in intarsia contain, then, representations of geometrical solids. These were done by Fra Giovanni sometime after the publication of *De Divina Proportione.* The sequence is suggestive. The solids discussed in Pacioli's work were elegantly realized in woodcuts by Leonardo da Vinci, whose close association with Pa-



THEORY OF LINEAR PERSPECTIVE was codified by Leon Battista Alberti (1404–72). First Alberti drew a rectangle and placed a standing human figure in the foreground (*1*); then he divided the figure into three parts and marked off the resulting unit of measure on the baseline (*2*). He also chose a "centric point" level with the top of the figure's head (*3*). He connected the centric point to the points on the baseline (*4*). Then, choosing a "distance point" on an extension of the construction's horizon line, he drew diagonals to the baseline (*5*). The diagonals' intersections with the rectangle determine a set of horizontal lines (*6*). Horizontals complete a floor grid in perspective.

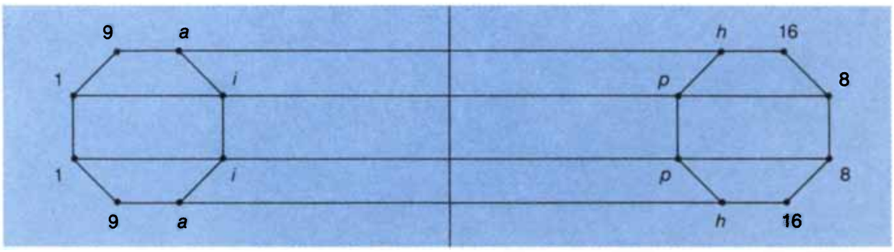


REASON THE THEORY WORKS is evident when the diagonals from the distance point are placed at a right angle to the rectangle. The diagonals then intersect the grid lines on the floor behind the rectangle. Alberti considered a painting to be in effect a window.

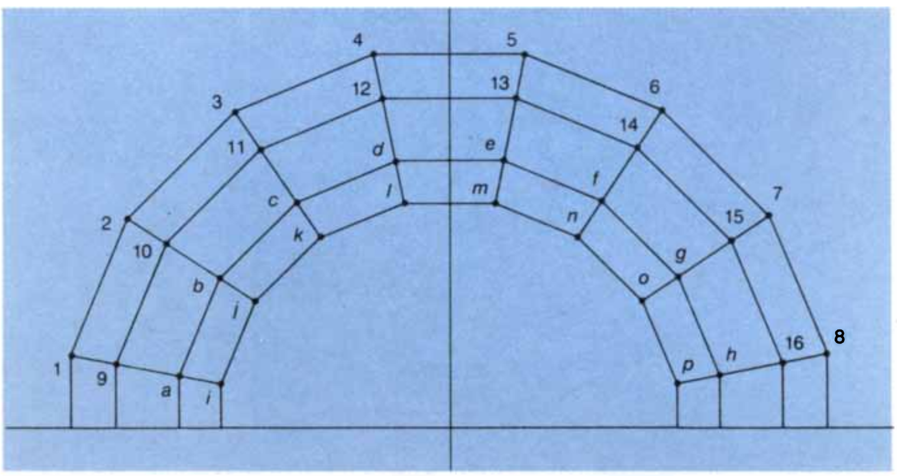
cioli in Milan during the writing of *De Divina Proportione* and later in Venice is well documented. The solids represented in Fra Giovanni's intarsia panels are transcriptions of Leonardo's illustrations, but their orientation is in many cases different from that of the originals, for the sake, one imagines, of an aesthetic harmony in the overall design.

Although Pacioli is said to have kept the original Leonardos in his possession, the accuracy with which they were replicated in intarsia suggests that Fra Giovanni had access to the drawings while designing his work in Verona and Monte Olivetto Maggiore. Their replication in intarsia was at once a statement of a metaphysical view of the world and a demonstration of the beauty of the visual expression of the metaphysics through the mathematical theories that linked the world of appearance to its underlying reality.

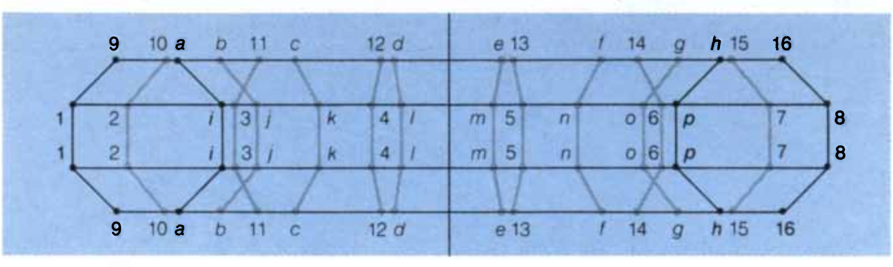
A famous inscription over the entrance to Plato's Academy read: "Let no one unskilled in geometry enter [here]." This too had echoes during the Renaissance. Alberti counseled artists not to attempt their work unless they were familiar with geometry, and Leonardo admonished that no one unfamiliar with mathematics should read his own advice to artists. Both are reflections of Plato, yet ironically they foreshadow a rejoinder to Plato's criticism of the arts. In Book 10 of *The Republic* Plato contrasts the falsity of art with the truth of mathematics. For example, he observes that in order to be true to the look of things the artist must be false to their essential properties, notably their various measurements (their metrical properties). By virtue of the reciprocal relation between image and object in perspectival projections, however, it is possible to determine the essential metrical properties of a geometrical object and to construct or reconstruct three-dimensional models from their two-dimensional depictions. Thus the way is cleared toward resolving Plato's quarrel with representational artists. That the Platonic solids themselves became models for geometrically precise representation is a delightful footnote to Plato.



CONSTRUCTION OF A MAZZOCCHIO in perspective as it might have been done by a Renaissance artist is shown in a sequence of illustrations. The sequence is based in part on *La Pratica della perspettiva* (1569), by Daniel Barbaro, and in part on the work of G. I. Kern, a German historian of mathematics. Here the construction begins. Two equilateral octagons are marked off at equal distances from a vertical line. They are connected by four horizontal lines that represent four planes of the *mazzocchio.* The vertexes of the octagons are labeled.



SECOND STAGE in the construction requires that the lines *1–8, 9–16, a–h* and *i–p* be made the diameters of four concentric semicircles. The semicircles are divided into 12 equal arcs; then the endpoints of the arcs are labeled and are connected by line segments. The result (four concentric polygons) is a plane view of a *mazzocchio* as it would be seen from above.

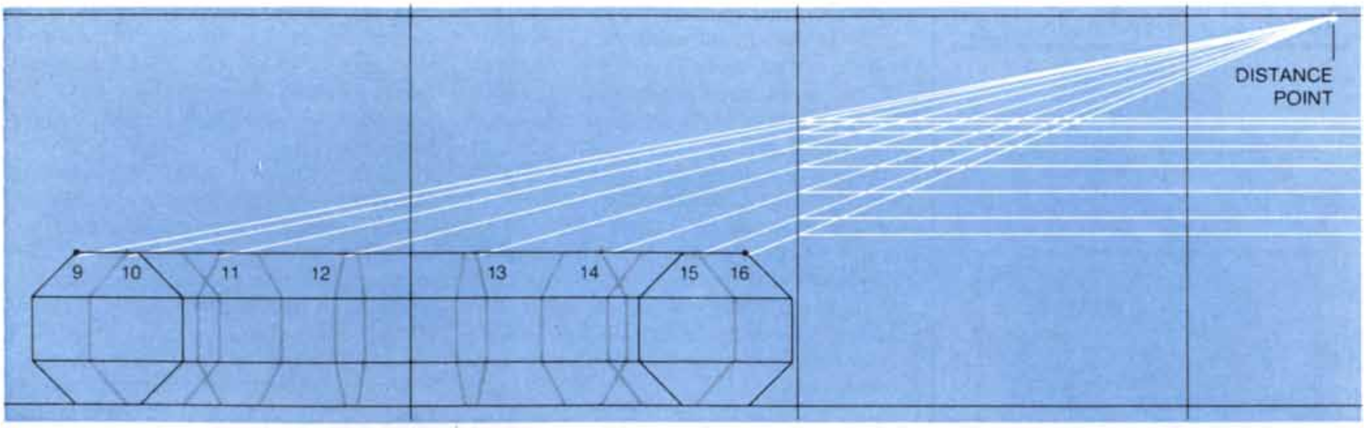


THIRD STAGE requires that the various points determined in the second stage be mapped onto the construction made in the first stage. The simplest way to accomplish the mapping is to drop a vertical line from each such point until the line intersects the two appropriate horizontal lines in the initial construction. The points *9–16* and *a–h* must all be mapped, for example, onto the line representing the top of the *mazzocchio* and the line representing the bottom of it.

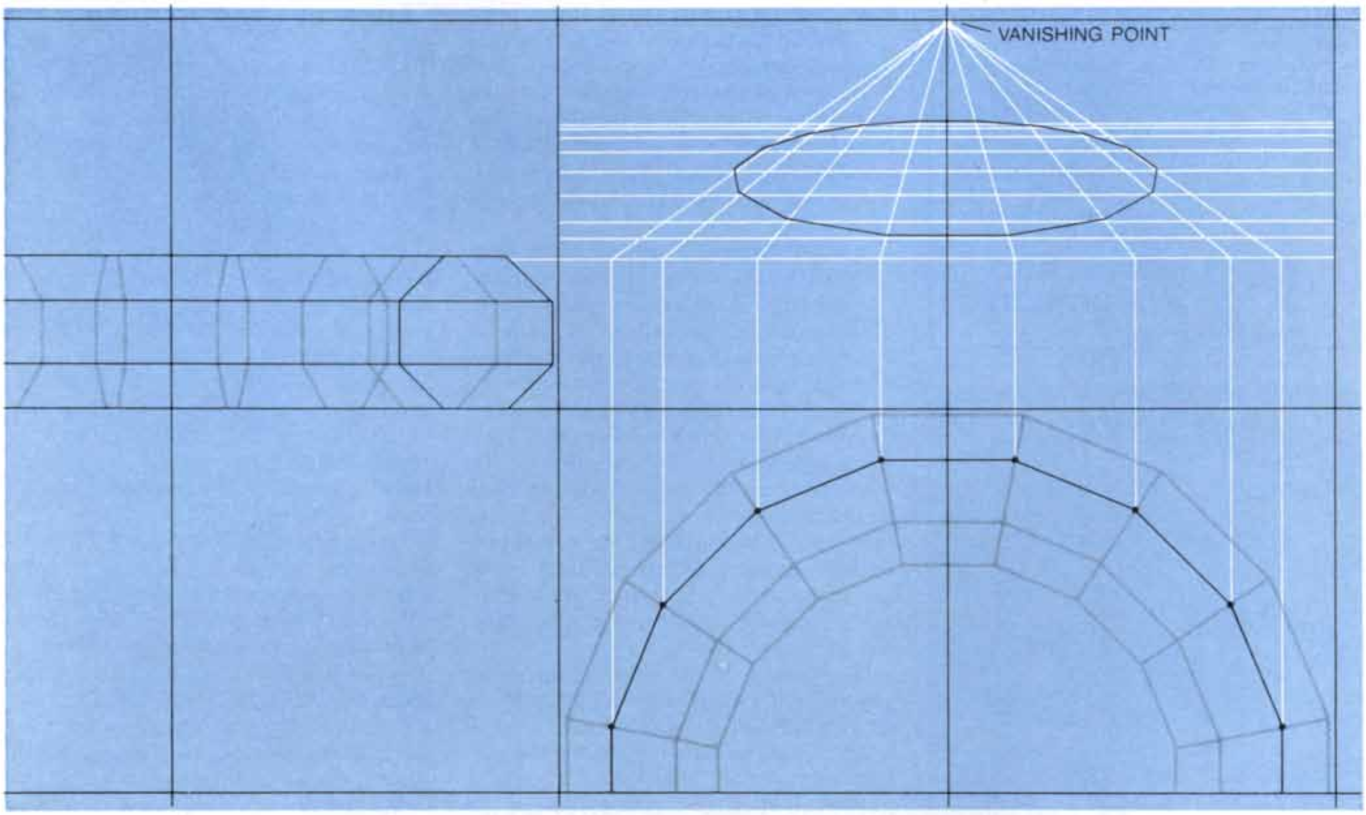
The intimate linkage among mathematics, art theory, philosophy and intarsia can be documented at a personal level as well as a theoretical level. In 1471 Alberti (then in his sixties) was Pacioli's patron when Pacioli made his first visit to Rome. Lorenzo da Lendinara, widely respected as a major *intarsiatore,* was Pacioli's godfather and a teacher of Piero della Francesca. Piero in turn influenced Pacioli. In fact, there has been a lengthy debate over the accusation that Pacioli was guilty of plagiarism in including in *De Divina Proportione* some of Piero's mathematical writings.

Pacioli was a friend of Leonardo's and of Pope Julius II's, who commissioned Fra Giovanni to decorate with intarsia panels the lower part of the Stanza della Segnatura, the room in the Vatican that contains Raphael's *The School of Athens.* The intarsia was later removed except for four panels on the doors leading into the Stanza. The panels are rarely seen, because the doors are usually open to admit the stream of visitors anxious to admire Raphael's work. *The School of Athens* includes depictions of philosophers (Aristotle and Plato, who carries a copy of the *Timaeus*), mathematicians (Pythagoras and Euclid) and a self-portrait. Juxtaposed with the intarsia panels it must have been one of the strongest possible statements of the influence of mathematics and philosophy on the theory of pictorial representation.

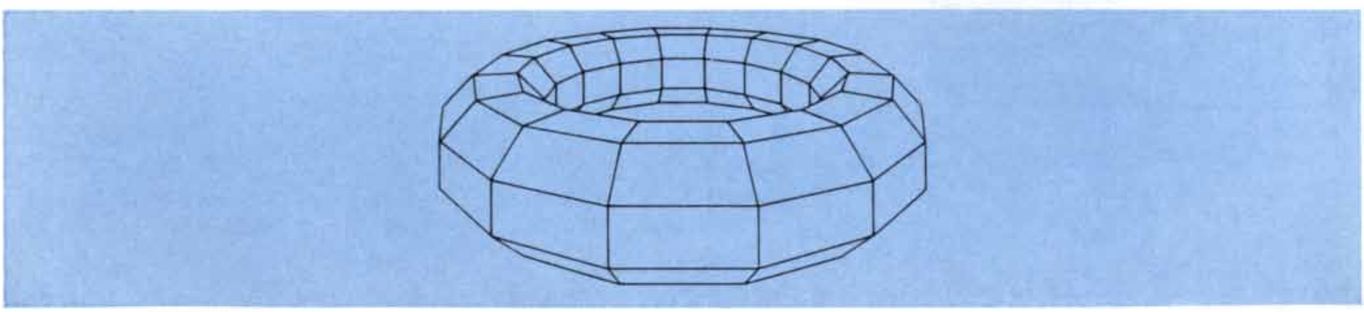
A unique and complicated geometrical form, the *mazzocchio,* came to symbolize the linkage between mathematics and pictorial representation. This solid, decidedly not Platonic, was much studied and discussed during the Renaissance. The *mazzocchio* was a wood or wicker headdress worn by well-to-



FOURTH STAGE is in essence an elaboration of Alberti's method. Each of the four horizontal planes includes two concentric polygons, and each polygon is determined in the construction by eight points. Thus diagonal lines are extended from a set of eight points (points *9–16* on the top of the *mazzocchio*) to a distance point. Horizontal lines are drawn from points where the diagonals intersect a vertical.



FIFTH STAGE is not clear in Renaissance texts, but the method shown here is plausible. Vertical lines rise to intersect what amounts to the forward edge of a horizontal plane that includes the top of the *mazzocchio*. Then from the points of intersection lines extend toward the vanishing point on horizon line. Intersections of latter lines with horizontals in fourth stage give vertexes of polygon in perspective.



COMPLETED MAZZOCCHIO is produced by seven further iterations of the fourth and fifth stages. Each iteration yields (in perspective) another horizontal polygon. The mastery of the *mazzocchio* was believed by Barbaro to illuminate the study of linear perspective.

do Florentine men of the 15th century. It was also fashionable in Burgundy, where it may have originated. Because of its complicated geometrical form (it is a ring made of equilateral polygons) its representation presented a difficult constructional problem that many perspectival theorists sought to solve. Some of the best-known efforts are by Uccello, whose passion for the study of geometrical forms and their perspectival representation was noted by Vasari. Vasari also remarked that some of Uccello's designs were intended for use by *intarsiatori.*

The perspectival representation of the *mazzocchio* was considered a sign of mastery of perspectival theory. Daniel Barbaro notes in *La Pratica della perspettiva,* an influential book on the theory of perspective that was published during the latter half of the 16th century: "The construction of the *mazzocchio* was still considered very difficult even at the end of the 15th century. In spite of the difficulties, the construction was used because it allowed the creation of beautiful figures and the mastery of it aided in the resolution of problems in perspective."

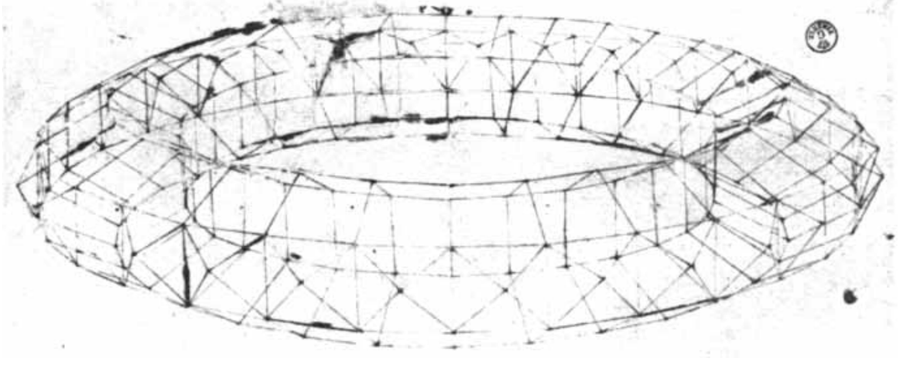
It is not surprising, then, that the *mazzocchio* became a symbol of solid geometry and that it appears in major works of intarsia in the cathedrals of Pisa and Modena and in two famous secular sites: the intarsia executed for Federigo da Montefeltro's studies in Urbino and Gubbio. (The Gubbio intarsia is now being restored for reinstallation at the Metropolitan Museum of Art in New York.) For the *intarsiatori* the *mazzocchio* would undoubtedly have symbolized the connection between their work, the concurrent renaissance of mathematics and their right to the title "masters of perspective."

In addition to its interest on other grounds intarsia prefigured painting in its treatment of still-life motifs. That is, intarsia must be counted among the earliest instances of the depiction of objects for their own sake. The depiction of the *mazzocchio* is evidence for this claim. The representational content of intarsia panels, such as musical instruments, instruments for navigation and instruments for the measurement of space and time, links the iconography of intarsia directly with the philosophical, mathematical and scientific concerns of the age.

Renaissance interest in the quantification of space was paralleled by interest in the measurement of time. Brunelleschi, who initiated the study of linear perspective, also designed and built clocks. The earliest representations of a domestic clock—the monastic alarm clock—are in intarsia. The depiction of such instruments of measurement in intarsia might well have been predicted from the fact that many of the most notable *intarsiatori* were monks, who displayed an animated interest in the accurate depiction of objects that entered the regulated routines of their daily lives.

If the reasons for the precocious development of intarsia in the 15th century are remarkable, the reasons for its precipitous decline and subsequent neglect are no less so. We have first of all Vasari to thank for prejudicing the attitudes of both contemporaneous artists and later scholars by his remark that intarsia had always been practiced by those having more patience than skill and by the inaccurate judgment that it does no more than counterfeit painting and must necessarily be transitory because of the constant threat of damage by worms and fire. (In all fairness we should mention that Vasari also thought Uccello's studies in perspective were a waste of time.)

More important, the essential link between intarsia and the Renaissance commitment to a geometrically anchored conception of perspective helps to explain why intarsia flourished only briefly as a significant representational art form. By the middle of the 16th century the artistic climate had changed radically. Michelangelo, for one, protested against the formulation of rules for pictorial composition; he complained that all the reasonings of geometry and arith-



DRAWING BY PAOLO UCCELLO is evidently a study for the construction of a *mazzocchio.* In the actual drawing one can see holes where a compass was placed. One also sees networks of lines incised in the paper. Drawing may have led to a cartoon for an intarsia panel.

metic and all the proofs of perspective were of no use to a man "without the eye." Art theorists began to criticize the attempts to put artistic representation on a mathematical basis.

Soon mathematics, which had been honored by the preceding century as the firmest foundation for pictorial art, was disparaged with great intensity. In his *Idea of Sculptors, Painters and Architects* (1607) Federico Zuccari, a defender of the mannerist view, denied the relevance of mathematics altogether. "I say [he asserted] that the art of painting does not derive from the mathematical sciences, nor has it any need to resort to them to learn rules of means for its own art, nor even in order to reason abstractly about this art: for painting is not the daughter of mathematics but of Nature and of Drawing."

With the erosion of interest in perspective and the geometrical studies so closely allied to it, intarsia reached the end of its greatest artistic potential. Unlike painting, it could not transcend its roots in perspectival theory, except perhaps by a leap into abstraction not encouraged by the ascendancy of mannerism and the Baroque. In fact, the "cubistic" quality of the best intarsia panels suggests a closer aesthetic affinity with the 20th century than with intervening centuries.

Work in intarsia did not, of course, simply stop in the face of shifting fashions of taste and revisions of aesthetic doctrine. What happened was really worse: Vasari's accusation that intarsia could do no more than counterfeit painting turned into pathetic prophecy. The greatest part of the work done in intarsia after 1525 was devoted to the mimicry of cluttered, chaotic, "painterly" designs, so that it lost its singular character. The effort to accommodate the capacities of the material and technique of intarsia to the demands of an incongruous style is a vivid instance of the inherent limits of any artistic medium.

More recently Renaissance intarsia has suffered neglect for economic reasons. Since most of the extant work is in choir stalls and sacristies scattered across northern Italy, it is commercially inaccessible. It cannot be bought, sold, traded or toured, and in the international art market that is nearly enough to ensure oblivion.

History, however, has a way of redressing prolonged neglect. And if, finally (as some historians have argued), the most significant accomplishment of the Renaissance was neither the invention of printing nor the discovery of America nor the fall of Constantinople but rather the emergence of the ideas that led to the "rationalization of sight," then the achievements of the *intarsiatori* truly deserve, as the Florentine Rucellai maintained, to be counted among the glories of his age.

GIVE AMERICAN WORKERS THE RIGHT TOOLS AND THEY CAN OUTPERFORM ANYBODY IN THE WORLD.

At Motorola, that's a long-held belief. More than that, it's a long-term business principle.

We've been operating on this principle for at least fifteen years And we've coined a name for it. We call it Tool Management Culture (TMC).

TMC means supplying every employee with the best equipment to do the job. From pencils to computers. TMC also means investing the resources to make it work. And invest we do. Of 1981's capital expenditures of $317 million, nearly 70% was put behind TMC.

But it also takes employee understanding and motivation to make TMC work. Because employees are the prime movers in recommending new tools, methods and techniques.

That's where our Participative Management Program has been a tremendous help. Because this program not only makes every employee a member of the management team, it allows them to make recommendations to improve their own productivity.

Motorola's R2001. In one "in-the-field" test unit: a signal generator, a frequency, modulation, and RF power meter, a spectrum analyzer, a duplex generator, RF memory tables, a code synthesizer, an audio frequency counter, a digital voltmeter, an oscilloscope, and a sweep generator.

As participants in management, Motorola people know that better tools produce better quality, higher productivity, increased demand in the marketplace. And that helps reduce costs, preserve jobs, and make the company's products more competitive.

Take our Communications Sector for example. With thousands of people in the field servicing land mobile communications products with dozens of separate pieces of test equipment, our people saw a need for something better. So they invented the R2001, a single in-the-field test unit that could handle 12 functions in one test unit.

There isn't another piece of service equipment like it. Amazingly, it not only services all Motorola's land mobile communications products, it can service anyone else's As a result, the R2001 has expanded the test equipment market for us, in addition to providing a better tool for our own people.

Clearly, it's for reasons like this we're a world leader in land mobile communications. And it's also an example that proves our point.

Giving people the right tools can help a company work a lot smarter.

But to give them the right tools, you've also got to give them their heads.

And once you do that, they'll help you compete with anybody at home or abroad.

MOTOROLA A World Leader in Electronics.

Quality and productivity through employee participation in management.

THE AMATEUR SCIENTIST

In which a Lifesaver lights up in the mouth and light takes funny bounces through a lens

by Jearl Walker

If you stand in front of a mirror in a dark room long enough for your eyes to adapt to the darkness and then start chewing on a piece of hard candy, you may see in the reflection of your open mouth a faint light like the glow from the dial of a luminescent watch. The glow from the candy is triboluminescence (from the Greek *tribein,* to rub) and can be demonstrated with most hard candies that are made with granulated sugar. It is the sugar, or more precisely the molecules of gaseous nitrogen around it, that accounts for the glow.

I first noticed the phenomenon when I was standing in front of a mirror in a dark room while chewing on a wintergreen Lifesaver. As I crunched the confection I could see in the reflection that the inside of my mouth was aglow with a blue green light. The phenomenon has been known since at least the 17th century, when the Accadèmia del Cimento in Florence published a description that in a contemporary translation reads: "Of Bodies affording Light.... Besides Firestones there are other Conservatories of Light; for by striking them together or breaking them in the Dark, they Sparkle. Such are White Sugar, Loaf Sugar and Sal Gemme [rock salt] in the Stone; all which being broken in a Mortar, give forth so great a Light as distinctly to discern the sides of the Mortar and the shape of the Pestle thereby." At about the same time Robert Boyle noted that "hard sugar being nimbly scraped with a knife would afford a sparkling light." Many investigators since then have explored the phenomenon, but the nature of triboluminescence was not understood until this century. Even now the details have not been fully worked out.

Until quite recently the study of triboluminescence focused on determining what substances glow, examining the spectrum of the light and exploring how the glow might be correlated with the physical properties of the substance. Much of the modern research has been done by Jeffrey I. Zink and his associates at the University of California at Los Angeles. They and others have ascertained the source of the light in sugar and certain other substances. An explanation of the relation between the glow and the pressure on the substance has remained elusive.

A good many materials display triboluminescence, not always in the same form as ordinary sugar (sucrose) does. Zink and others have found that the sugars D-glucose, lactose, maltose and L-rhamnose emit the glow but the sugars fructose, cellobiose, fucose, galactose and mannose do not.

Early in this century the spectrum of the glow from sucrose was shown to be the same as the spectrum of gaseous nitrogen. It was clear that the glow did not originate in the molecules of the sugar itself, because they would yield an entirely different spectrum. The spectrograph shows that the triboluminescence of sugar has several bright peaks near the ultraviolet region of the spectrum. Strong emissions appear at 330 and 360 nanometers, weaker ones near 315 and 380. The visible glow is fainter, lying between 380 and 435 nanometers (the violet and blue regions of the spectrum). If the eye could see farther into the ultraviolet, the glow would be brighter because the strong emission peaks would then be visible.

The crushing of sugar crystals somehow excites the molecular nitrogen associated with the sugar. An analysis of the spectrum indicates that the nitrogen is excited from its lowest molecular energy state to the $^3\pi_u$ state. It quickly drops to the $^3\pi_g$ state, emitting light in a certain assortment of wavelengths. It is these wavelengths that can be identified as the assortment unique to molecular nitrogen.

The role of nitrogen can also be demonstrated by removing from the sugar as much nitrogen as possible. When sugar is crushed in a vacuum, the glow is dimmer. It is still there because even in a vacuum some molecules of nitrogen are adsorbed on or absorbed by the sugar. When nitrogen is entirely eliminated, as Zink has managed to do, the glow disappears.

The importance of the ambient gas is similarly revealed when other gases are substituted for nitrogen. When the sugar is placed in an atmosphere of neon at low pressure, the glow from crushing the sugar is red. The neon is somehow excited by the pressure; it then deexcites much as it does in neon signs. Neon's emission is strong in the red region of the spectrum.

Certain hard candies display additional colors. Zink and his associates have found, for example, that wintergreen Lifesavers give off more of a blue green color than other flavors, which have thc weak blue emission of sucrose. The reason is that the ultraviolet radiation of the nitrogen is absorbed by the methyl salicylate (oil of wintergreen) with which the candy is flavored. When the methyl salicylate deexcites, it emits in a broad range from about 400 nanometers to about 540. The chewer sees a composite of the spectrum from molecular nitrogen and methyl salicylate.

I demonstrated triboluminescence by grinding table sugar and various sugar products in a beaker after giving my eyes time to adapt to darkness. What you grind with does not seem to matter. I had good results with either the blunt end of a metal table knife or a wood dowel. Nor does it matter whether the beaker is made of glass, plastic or metal. (Be careful with glass. Since you are working in the dark, you could break a glass beaker by applying pressure in the wrong way.)

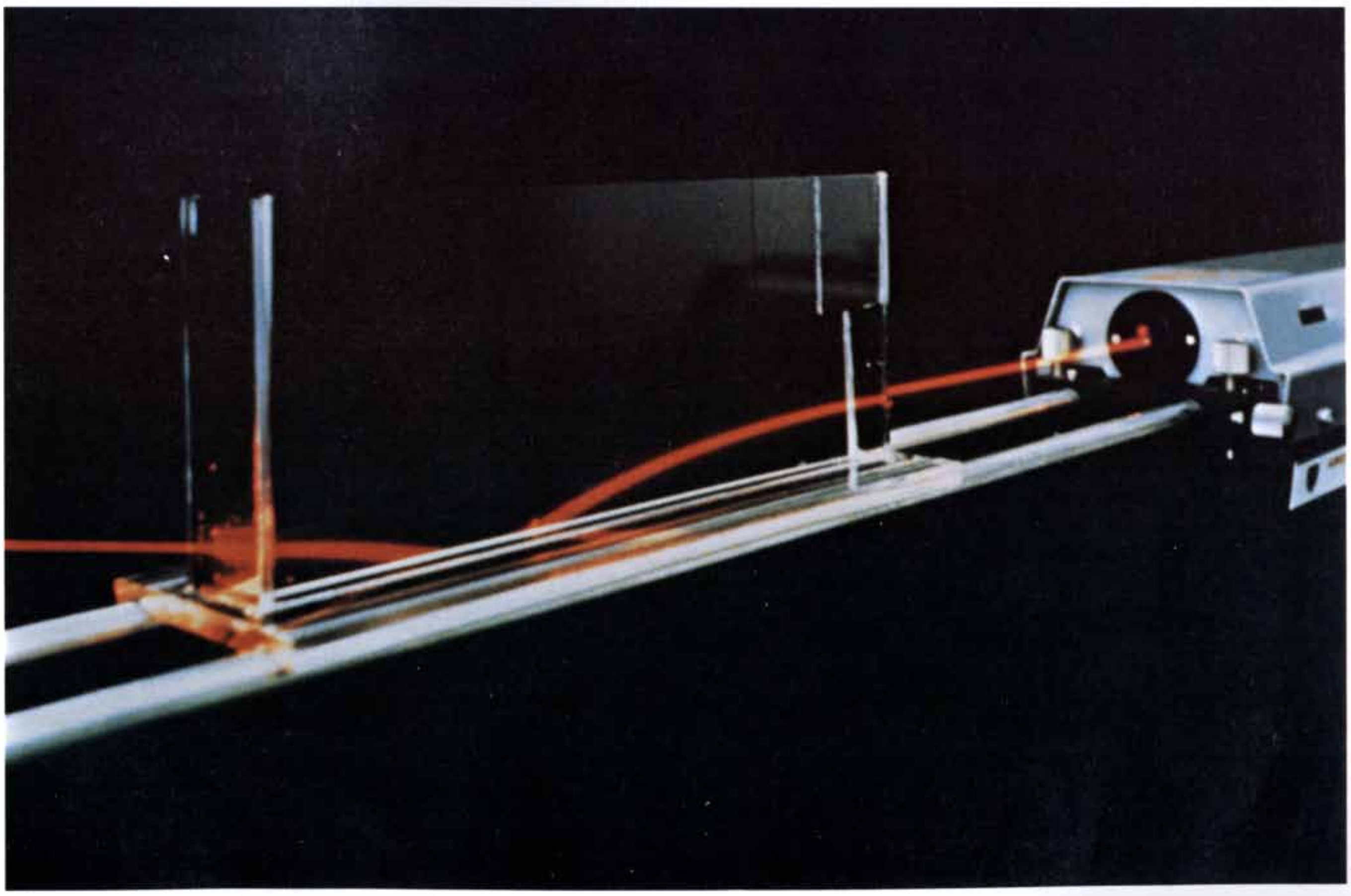
The amount of light emitted by sugar seemed to diminish as I continued grinding. Sugar made slightly damp by the humidity in the room gave off a fainter glow. Sugar dissolved in water did not glow unless I hit an undissolved crystal with the grinding tool.

Many of the liquids and powders for washing clothes have a component that absorbs ultraviolet and emits in the visible spectrum. (The aim is to make the clothes look whiter.) I added one of these products to sugar but saw no difference in the glow.

I also studied the glow from several flavors of Lifesavers. To spare my teeth I followed Zink's suggestion of crushing the candy with a pair of pliers. Wintergreen yielded the most easily seen light, apparently because of the emission from the methyl salicylate.

To study the spectrum I taped a replica diffraction grating to my eyeglasses. If I looked directly at the glow, I received the central peak of the pattern developed by the grating. This peak overlaps all the colors from a light source, however, and prevents a determination of the spectrum. The first side peak from a diffraction grating separates the colors from a source, but I could not make out the side peak, apparently because the light from the candy was so faint. Even a brighter side peak might not be discerned because the light is so faint that the retina would have difficulty distinguishing the colors.

The glow from a wintergreen Lifesaver as it is crushed with pliers



A sugar solution acts as a gradient-index lens and refracts a laser beam in William M. Strouse's experiment

Physician, did you miss any of these significant developments in medical science?

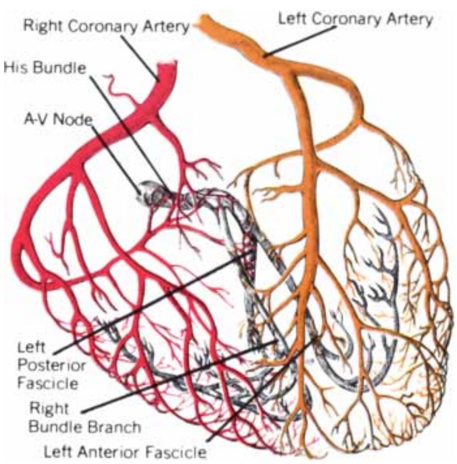
• *Campylobacter fetus* **subsp.** ***jejuni*** **is associated with a colitis that can clinically and sigmoidoscopically resemble acute idiopathic ulcerative colitis. Stool cultures are in order for** ***C. fetus*** **before beginning nonspecific anti-inflammatory therapy.**

• **Coumarin derivatives cross the placenta. A recent study shows that the consequences for the fetus can be severe. These include embryopathy, stillbirth, and premature delivery.**

• **Nonsteroidal anti-inflammatory drugs may produce a marked reduction in glomerular filtration rate; with termination of the drug, GFR returns to normal.**

• **The first documented incident of indigenous transmission of dengue in the continental United States since 1945 has been reported in Brownsville, Texas.**

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.



Branches of the right and left coronary arteries supply blood to the A-V 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.

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 2,500-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.

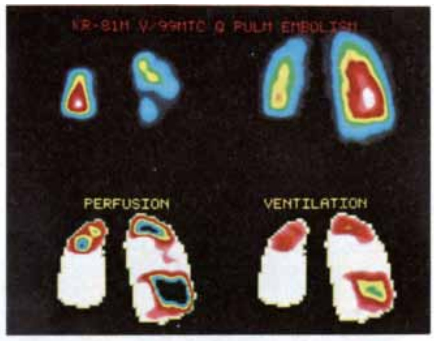
New material replaces old material in the living text, so that the information is there — up-to-date, at your fingertips.

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 has been reviewed and is acceptable for 32 prescribed hours by the American Academy of Family Physicians. This program has been approved by the American College of Emergency Physicians for 32 hours of ACEP Category 1 credit.

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 $220 for the full 12 months (renewal is currently $170); otherwise return the two volumes.

Please mail the coupon today and let us take the hassle out of keeping up.



Computerized scintigraphy reveals pulmonary thromboembolism.

Abdominal computed tomogram reveals large renal carcinoma replacing part of right kidney.

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

Douglas Wilmore, M.D., Harvard Medical School and Brigham and Women's Hospital

5. Hematology

Stanley L. Schrier, M.D., Stanford University School of Medicine

6. Immunology

John David, M.D., Harvard Medical School and Harvard School of Public Health

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; Beth Israel Hospital, Boston; and Brigham and Women's Hospital

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 Brigham and Women's Hospital

10. Nephrology

Roy H. Maffly, M.D., Stanford University School of Medicine and Palo Alto Veterans Administration Medical Center

William Bennett, M.D., Health Sciences University, Portland, Oregon

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

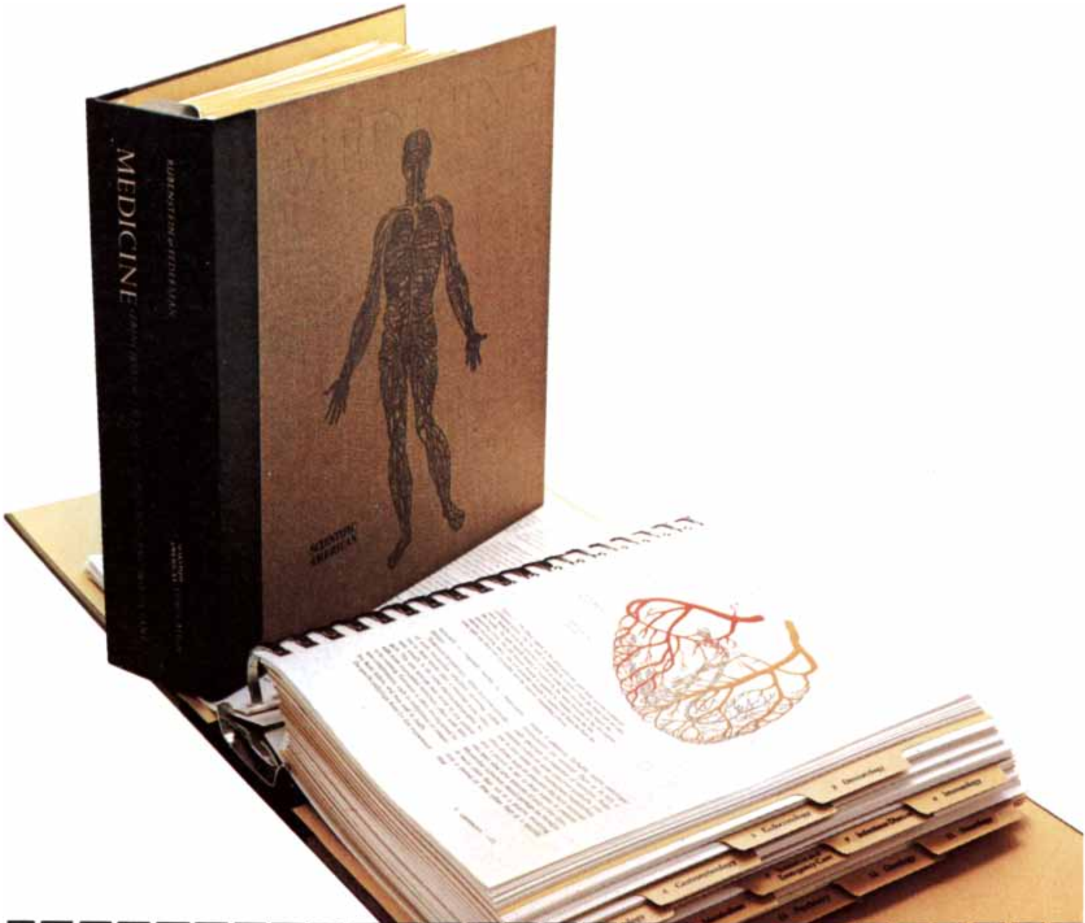
Stephen M. Krane, M.D., Harvard Medical School and Massachusetts General Hospital

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: 1-800-345-8112 (in Pennsylvania, call 1-800-662-2444);** 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 $220 for first year of service is tax deductible, as is current renewal at $170. 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 $220*
☐ Bill me ☐ VISA ☐ MasterCard

Expiration Date________ Account Number________

*Please add sales tax for California, Illinois, Massachusetts, Michigan and New York

Name________

MD Specialty________

Address________

City________ State________ Zip________

Signature________

Please allow six to eight weeks for delivery. All payments must be in U.S. dollars. CME is available outside the U.S. and its possessions for an additional $65. Surface routes used unless airmail delivery is prepaid.

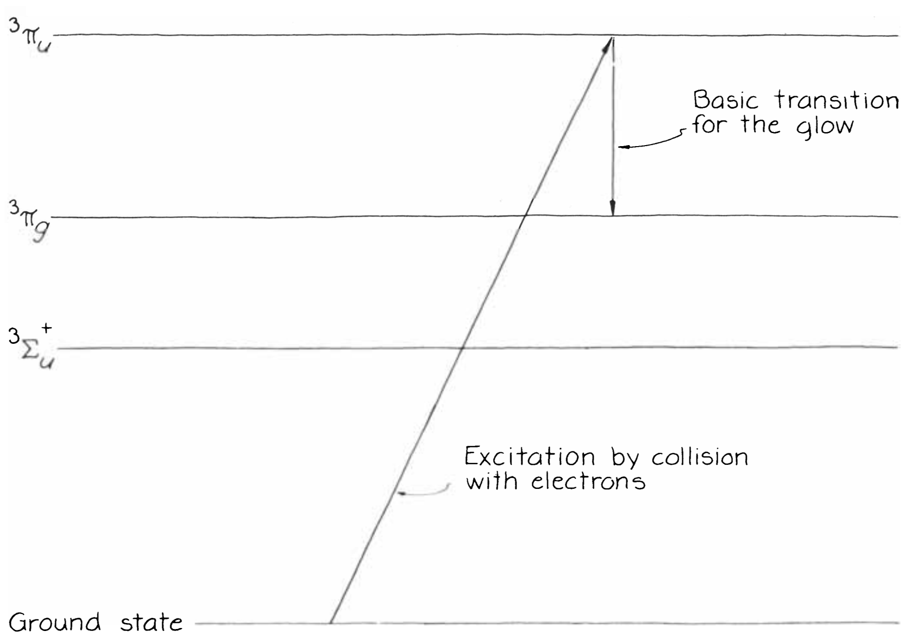
Q1

You can buy an inexpensive replica diffraction grating from the Edmund Scientific Company (101 East Gloucester Pike, Barrington, N.J. 08007). If you have made any of the spectrum analyzers I have described in other articles for this department, you will be able to separate the spectrum of a piece of hard candy. A procedure for doing such an experiment is given in the article by Rebecca Angelos, Zink and Gordon E. Hardy cited in this month's bibliography [*page 154*].

I found triboluminescence in tablets of saccharin and crystals of tartaric acid and rock salt. Table salt and all soft candies failed to glow. Although soft candy is likely to contain sugar, the crystals may be too small or too difficult to stress (because of the plasticity of the candy) for a glow to result.

Sometimes hard candy will glow as a result of a sudden shock from a source other than crushing. If a Lifesaver is dropped into liquid nitrogen, the sudden drop in temperature creates enough stress to generate a brief glow. (If you want to repeat the experiment, be extremely careful with the liquid nitrogen. It can be dangerous if it is handled improperly.)

Equipment for investigating triboluminescence



Levels of excitation and deexcitation of nitrogen in the glow

Many other substances display a triboluminescence that originates in the material itself rather than in the ambient gas. You can recognize such an origin if the spectrum of the triboluminescence matches the spectrum emitted when the substance is heated or illuminated. Here again the connection between the applied pressure and the excitation of the emitting molecules is not well understood. It could be the result of chemical reactions. It could also be thermal, arising because the pressure distorts the molecules, alters the interactions of adjacent molecules or moves dislocations in crystals of the substance.

The glow from hard candy seems to be due to electric discharge rather than to chemical or thermal mechanisms. The discharge results from a separation of charges brought about by the stress on the material. Even though the material is usually not strongly charged, stress liberates ions and semifree electrons. The glow may then result from the recombination of positive and negative ions. Or it could result when the semifree electrons combine with positive sites in the material, giving rise to light. A third possibility is that the electrons are accelerated by electric fields between the charged sites and collide with molecules, exciting them so that they emit light when they drop to lower energy levels.

The glow from sugar appears to be attributable to an electrical action that results in molecular excitation by collision. When the sugar is pressed, electrons are freed at various points within it. The electrons jump across a crystal of the material or between crystals to reach positively charged sites. In the process they collide with molecular nitrogen that has been adsorbed on or absorbed by the sugar. The electrons also collide with molecules of nitrogen gas. Whatever the type of collision, the electrons transfer some of their kinetic energy to the molecular nitrogen, raising it to an excited state.

In sugar the creation of charged surfaces appears to be caused by the fracture of the sugar crystals. A fracture leaves charged surfaces that are almost immediately coated with molecular nitrogen from the air. The stage is set for moving electrons to make the nitrogen glow.

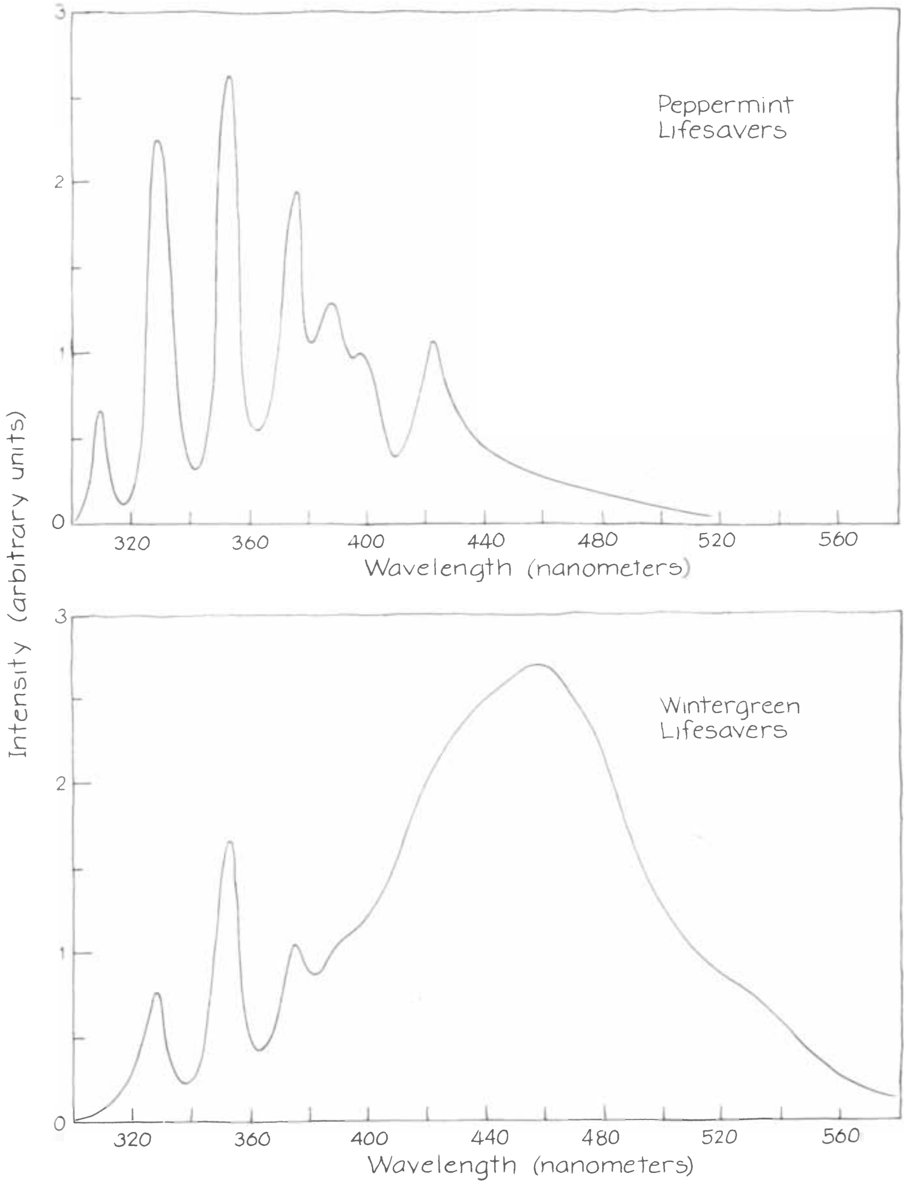
Zink and his associates have investigated the rate at which cracks propagate through crystals of sugar. They find that a single crack and its associated glow should persist for about a microsecond. Chewing on a Lifesaver or crushing it with pliers generates a longer glow because of a progression of cracks. The brightness of the glow depends on the size of the crystals. I believe the glow diminishes as I continue crushing a piece of hard candy because the crystals become smaller.

Another material that displays triboluminescence is tape coated on one side with an adhesive. E. Newton Harvey of Princeton University demonstrated in 1939 that several kinds of tape glow in the dark when a strip is pulled rapidly from the roll or off certain surfaces such as glass. The glow can also be seen if two layers of tape are stuck together and then pulled apart. Occasionally it appears if you run your finger along the uncoated side of a length of tape.

As with sugar the glow results from a discharge through the ambient nitrogen gas. The spectrum matches that of molecular nitrogen. If a tape is unrolled in an atmosphere of neon at low pressure, the glow is red instead of blue.

Apparently the manipulation of the tape builds up small areas of charge. When these areas are separated, the electrons in the negative areas leap through the gas to reach the positive areas, exciting the ambient nitrogen on the way. Similar glows can be seen in mica as it is split and on rare occasions in a stretched rubber band as the stretch is released suddenly.

Harvey described how to demonstrate that pulling a piece of tape from a substrate results in a separation of charge. He stuck a length of tape on a sheet of glass and pulled it off. If he then held the tape near the glass, it was attracted to the glass by the charged sites that had developed on the two surfaces. Harvey ascertained the sign of the charge on each surface by means of charged pith balls. A positively charged pith ball was attracted to the tape, indicating that the tape had been made negative by its removal from the glass. Similarly, a negatively charged pith ball was attracted to the glass. This is not to say that the tape is necessarily all negative and the glass all positive, merely that localized pockets of charge on each surface are dominated by charge of one sign.



Triboluminescence spectrograms obtained by Jeffrey I. Zink and his associates

Conventional optical devices such as lenses and prisms work by refracting light, that is, by redirecting it at their surfaces. Light is refracted when it crosses a boundary between two materials in which its effective speeds differ. One way of keeping track of the degree of refraction is to assign to each material an index of refraction; the slower light moves in the material, the higher the index of refraction is. For example, glass has an index of about 1.5, air an index of slightly more than 1.

Light is refracted on entering and leaving a glass lens but is not refracted inside the lens, since the index of refraction does not change within the glass. If the sides of the glass were parallel and flat, the refraction of light at the surface where the light enters the glass would be canceled by the refraction at the surface where it emerges. A lens has curved surfaces in order to create a net change in the direction in which the light propagates. A convex lens redirects parallel rays of light so that they cross at a focal point after passing through the lens.

A gradient-index lens (called, for obvious reasons, a GRIN lens) is unconventional in that its index of refraction varies internally. In one type of GRIN lens the index varies along an axis perpendicular to the optical axis (the axis along which the light is directed). The center of such a lens might have an index of refraction higher than that of the periphery. When light passes through the lens, the rays near the optical axis are deflected weakly but those away from the optical axis are directed strongly toward it. Then even if the lens has flat, parallel exterior surfaces, it can focus light like a conventional convex lens. If the surfaces are curved, the light is refracted even more.

A second type of GRIN lens has its gradient along the optical axis rather than perpendicular to it. The index at the front surface is different from the one at the rear surface, and it varies smoothly between them. A third type of lens would have a spherical gradient in its index of refraction. For example, the index might decrease outward from the center of the lens in a spherically symmetrical way. Although I do not know if any lens of this third type has ever been made, James Clerk Maxwell studied its optics more than a century ago. A prototype is the crystalline lens of the eye.

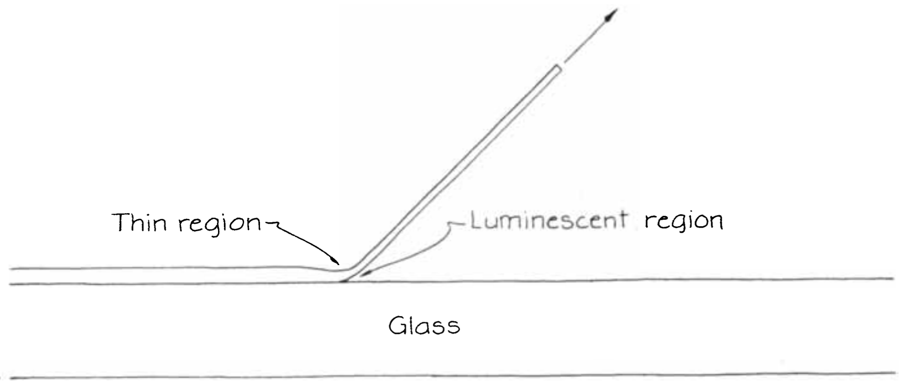
Several simple applications have been found for GRIN lenses, but the lenses are not common because it is hard to make them. Jurgen R. Meyer-Arendt of Pacific University and his student Mark R. Zilm recently described to me an easy way of making a simple GRIN lens. They have made one that behaves like a concave cylindrical lens.

In essence their lens consists of clear

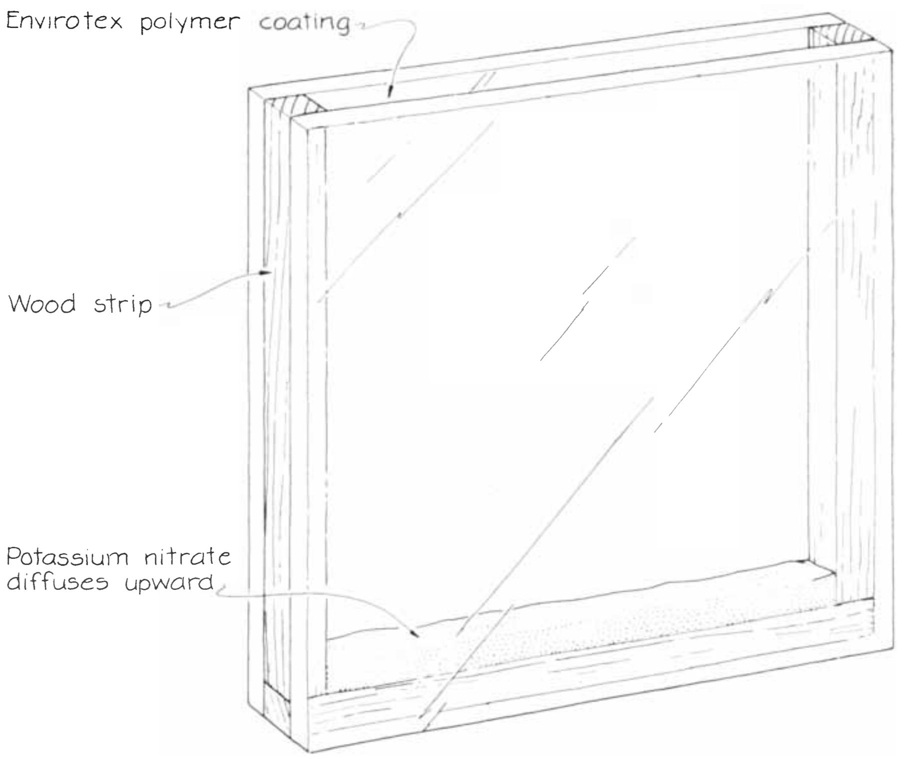
epoxy into which potassium nitrate has been allowed to diffuse, thereby creating the gradient in the index of refraction. A container for the material is made with two square pieces of glass, each piece 50 millimeters on an edge and 2.05 millimeters thick. They are held parallel and separate by thin pieces of wood (matchsticks) that are 50 millimeters long and three millimeters square in cross section. The wood was originally balsa, but Meyer-Arendt and Zilm later found that the hard wood of matchsticks is better. The wood is glued along three edges of a piece of the glass with a silicone seal glue. (Other types of glue might not work as well because the glue might diffuse into the epoxy.)

The fourth side of the container is left open so that the epoxy and potassium nitrate can be poured in. Potassium nitrate (KNO_3), otherwise known as saltpeter, can be bought in a drugstore. The epoxy chosen by Meyer-Arendt and Zilm is Envirotex polymer coating, which is sold in stores that deal in materials for paper-cutout decoration (decoupage). The potassium nitrate is packed to a uniform depth of about two millimeters at the bottom of the container, which is then filled with the epoxy. Meyer-Arendt and Zilm modified the usual procedure for mixing epoxy because they wanted the material to take longer to harden so that some of the potassium nitrate would have time to diffuse into it. The standard ratio of hardener to epoxy is 1 : 1; they made it 1 : 4. They tried even smaller amounts of hardener but found that air bubbles were trapped in the mixture when it hardened.

After three weeks in which the container was left undisturbed the potassium nitrate had diffused upward into the epoxy and the entire mixture had solidified. The result was a GRIN lens with an index of refraction that varied upward. The index was highest at the bottom, where the potassium nitrate was concentrated, and lowest near the top, where no potassium nitrate had penetrated.



Where a glow appears in tape being pulled from glass



The gradient-index lens made by Jurgen R. Meyer-Arendt and Mark R. Zilm

Meyer-Arendt and Zilm devised a simple experiment to demonstrate the refractive characteristics of their device. First they etched a grid on one of the pieces of glass at the side of the assembly. Each rectangle in the grid was three millimeters high and four millimeters wide. The GRIN lens was then set on one edge so that what had been the base was now vertical.

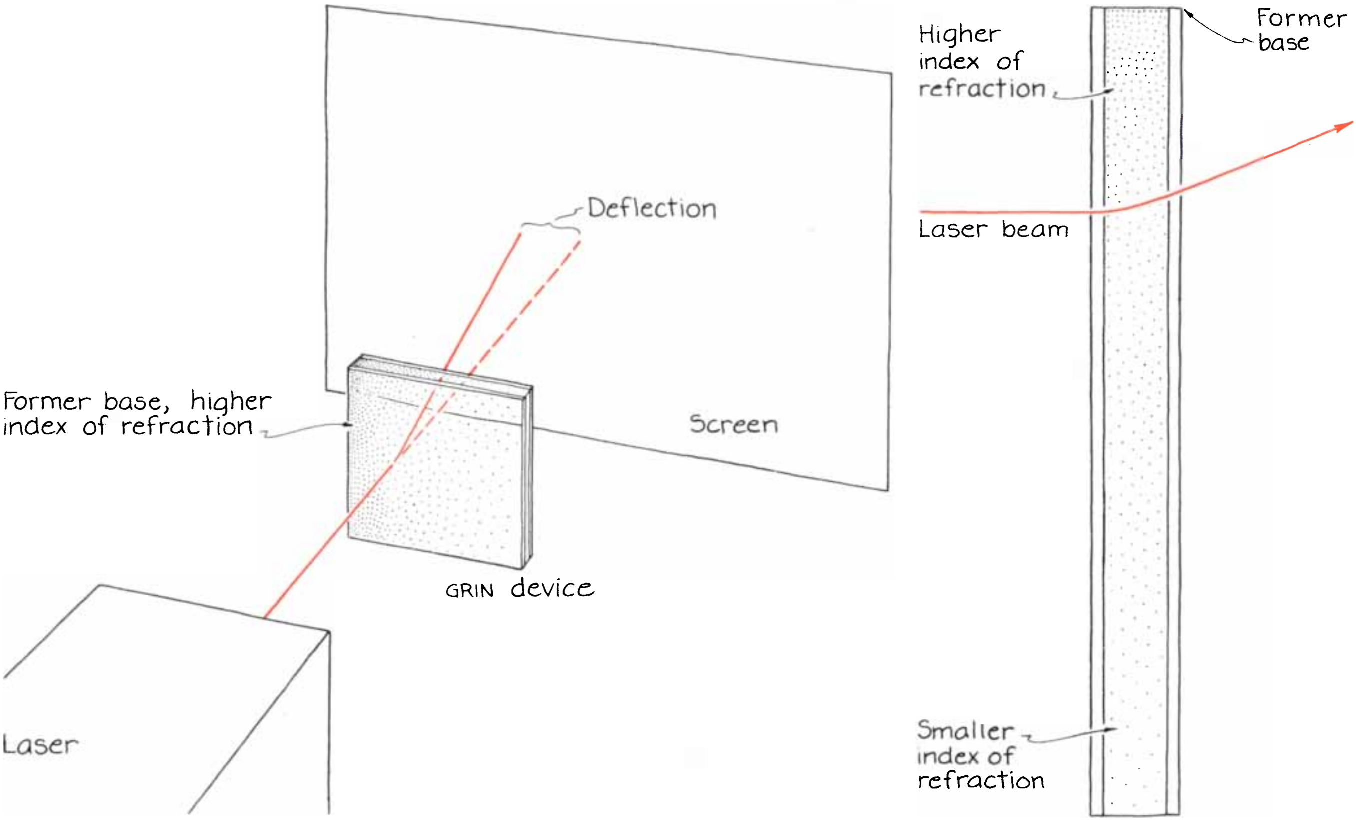
The next step was to direct a laser beam through one section of the grid. The light passed through the lens and fell on a screen 10 meters away. Ascertaining first where the laser beam hit the screen when the lens was not in place and then where the beam hit after passing through the lens, Meyer-Arendt and Zilm could measure how much the lens deflected it. They repeated the measurement along a line of grid sections ranging from the high-index part of the lens to the low-index part.

The refraction of a laser beam passing through the GRIN lens near what was originally the bottom is shown in the top illustration on the opposite page. The beam travels through a gradient in the index of refraction. The part of the beam passing through the area with a higher index is delayed with respect to the part that goes through the area with a smaller index. As a result the beam is deflected toward the side that was originally the base.

The optical properties of this GRIN lens are something like those of half of a concave cylindrical lens, as is shown in the bottom illustration on the opposite page. Near the areas with a high gradient in the concentration of potassium nitrate light is strongly deflected, as it would be in the outer part of the concave lens. The deflection in the areas with a smaller gradient of potassium nitrate is much as it would be in the center of the concave lens. If the diffusion of the potassium nitrate into the epoxy could be precisely controlled, one could presumably make a GRIN lens to match any conventional lens.

If you make a GRIN lens the way Meyer-Arendt and Zilm did, you might want to investigate its chromatic properties. Unless you have access to laser light of various colors, you will have to work with partially coherent light from a pinhole. The source could be any bright light filtered to narrow the range of color in the beam. You will need a mask or a stop to confine the light to only a small area of the GRIN lens.

A good demonstration of gradient-index refraction published by William M. Strouse is cited in the bibliography [*page 154*]. He directed a beam of light through a pair of lenses (to narrow it)

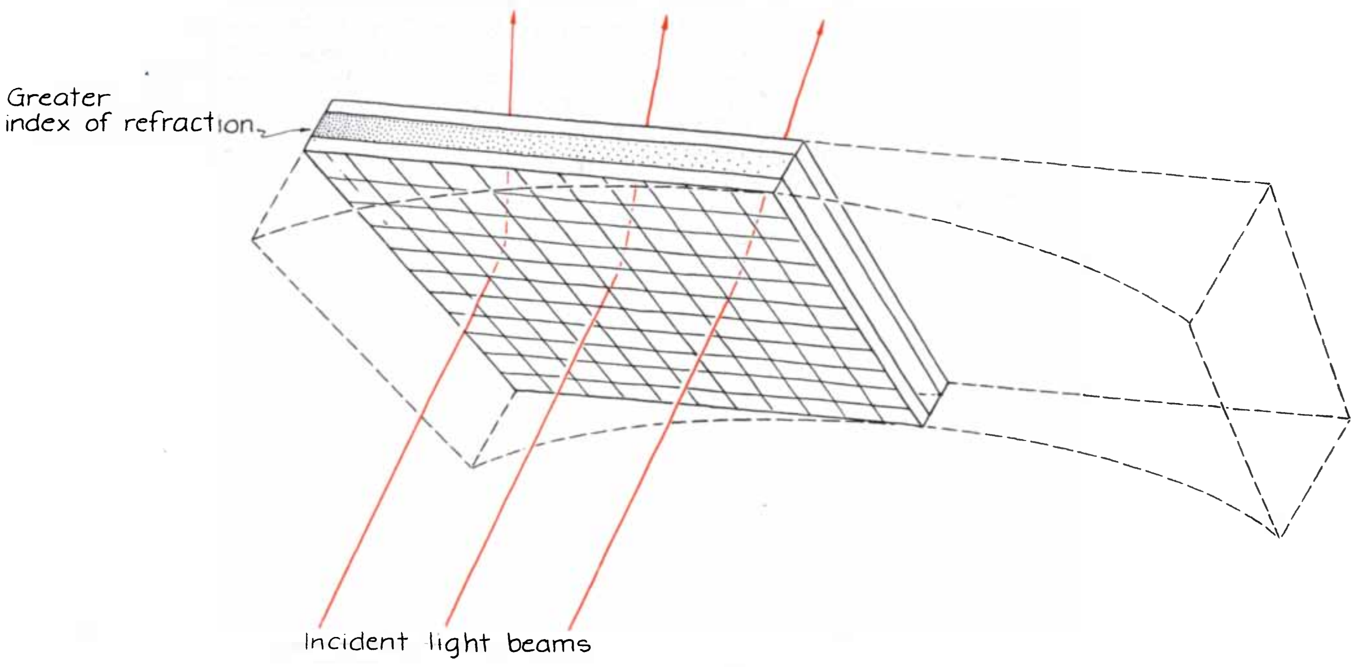


How the lens refracts a laser beam

and then into a tank filled with a solution of sugar in water. The tank was two centimeters wide, eight high and 20 long. Strouse put in three centimeters of water and added three lumps of sugar, which he allowed to dissolve undisturbed. As the sugar diffused through the water it created a gradient that was highest near the bottom and lowest near the top. The index of refraction therefore changed from high to low in the same direction.

When the beam was directed into the solution near the bottom of the tank, it curved downward and was reflected off the bottom. The reflected beam continued to curve. With a long tank the beam would bounce along the bottom.

Seen from above, the beam maintains the same width as it crosses the tank. Seen from the side, it narrows periodically. This partial focusing is also due to the gradient of the index of refraction. The gradient redirects the rays in differing degrees, so that the beam narrows in some parts of the solution and widens in others. Strouse said the phenomenon is enhanced if the lenses installed to narrow the initial beam are removed. The beam gets wider and is therefore easier to follow.



The similarity of the GRIN device and a concave lens

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

METAMAGICAL THEMAS

HANDBOOK OF CUBIK MATH. Alexander H. Frey, Jr., and David Singmaster. Enslow Publishers, 1982.

INSIDE RUBIK'S CUBE AND BEYOND. Christoph Bandelow. Birkhauser Boston, Inc., 1982.

PUZZLE IT OUT: CUBES, GROUPS AND PUZZLES. John Ewing and Czes Kośniowski. Cambridge University Press, 1982.

THE SPACE TELESCOPE

SCIENTIFIC RESEARCH WITH THE SPACE TELESCOPE: IAU COLLOQUIUM NO. 54. Edited by M. S. Longair and J. W. Warner. National Aeronautics and Space Administration CP-2111, 1979.

THE SPACE TELESCOPE SCIENTIFIC INSTRUMENTS. David S. Leckrone in *Publications of the Astronomical Society of the Pacific,* Vol. 92, No. 545, pages 5–21; February–March, 1980.

ORGANIC SUPERCONDUCTORS

SUPERCONDUCTIVITY AT ROOM TEMPERATURE. W. A. Little in *Scientific American,* Vol. 212, No. 2, pages 21–27; February, 1965.

ELECTRIC AND MAGNETIC PROPERTIES OF LINEAR CONDUCTING CHAINS. I. F. Shchegolev in *Physica Status Solidi,* Vol. (a)12, No. 1, pages 9–45; July 16, 1972.

THE DESIGN AND SYNTHESIS OF ORGANIC METALS. Anthony F. Garito and Alan J. Heeger in *Accounts of Chemical Research,* Vol. 7, No. 7, pages 232–240; July, 1974.

THE DIFFERENCE BETWEEN METALLIC AND INSULATING SALTS OF TETRACYANOQUINODIMETHANE (TCNQ). Jerry B. Torrance in *Accounts of Chemical Research,* Vol. 12, No. 3, pages 79–86; March, 1979.

THE PHYSICS AND CHEMISTRY OF LOW DIMENSIONAL SOLIDS: PROCEEDINGS OF THE NATO ADVANCED STUDY INSTITUTES HELD AT TOMAR, PORTUGAL, AUGUST 26–SEPTEMBER 7, 1979. Edited by Luis Alcacer. D. Reidel Publishing Co., 1980.

EXTENDED LINEAR CHAIN COMPOUNDS. Joel S. Miller. Plenum Press, 1982.

ULTRALIGHT AIRPLANES

PROGRESS IN FLYING MACHINES. Octave Chanute. *The American Engineer and Railroad Journal,* 1894. Reprint available from Ultralight Publications, P.O. Box 234, Hummelstown, Pa. 17036.

ULTRALIGHT AIRCRAFT: THE BASIC HANDBOOK OF ULTRALIGHT AVIATION. Michael A. Markowski. Ultralight Publications, P.O. Box 234, Hummelstown, Pa. 17036, 1981.

ENERGETIC OUTFLOWS FROM YOUNG STARS

THE MILKY WAY GALAXY. Bart J. Bok in *Scientific American,* Vol. 244, No. 3, pages 92–120; March, 1981.

GIANT MOLECULAR-CLOUD COMPLEXES IN THE GALAXY. Leo Blitz in *Scientific American,* Vol. 246, No. 4, pages 84–94; April, 1982.

DNA TOPOISOMERASES

STRUCTURE AND FUNCTIONS OF DNA and BINDING AND UNWINDING PROTEINS AND TOPOISOMERASES. Arthur Kornberg in *DNA Replication.* W. H. Freeman and Company, 1980.

SUPERCOILED DNA. William R. Bauer, F. H. C. Crick and James H. White in *Scientific American,* Vol. 243, No. 1, pages 118–133; July, 1980.

DNA TOPOISOMERASES. Martin Gellert in *Annual Review of Biochemistry,* Vol. 50, pages 879–910; 1981.

ANEURYSMS

SURGICAL CONSIDERATIONS IN THE TREATMENT OF ANEURYSMS OF THE THORACO-ABDOMINAL AORTA. Michael E. DeBakey, E. Stanley Crawford, H. Edward Garrett, Arthur C. Beale, Jr., and J. F. Howell in *Annals of Surgery,* Vol. 162, No. 4, pages 650–662; October, 1965.

CONTRIBUTION OF ABDOMINAL AORTIC ANEURYSMECTOMY TO PROLONGATION OF LIFE. D. Emerick Szilagyi, Roger F. Smith, Franklin J. DeRusso, Joseph P. Elliot and Frederick W. Sherrin in *Annals of Surgery,* Vol. 164, No. 4, pages 678–699; October, 1966.

ANEURYSMS OF THE AORTA. Crile Crisler and Henry T. Bahnson in *Current Problems in Surgery,* December, 1972.

SURGICAL MANAGEMENT OF ABDOMINAL AORTIC ANEURYSMS: FACTORS INFLUENCING MORTALITY AND MORBIDITY—A 20-YEAR EXPERIENCE. Jesse E. Thompson, Larry H. Hollier, R. Don Patman and Alfred V. Persson in *Annals of Surgery,* Vol. 181, No. 5, pages 654–661; May, 1975.

INFRARENAL AORTIC ANEURYSMS. Robert B. Rutherford in *Vascular Surgery,* edited by Robert B. Rutherford. W. B. Saunders Company, 1977.

ORIGIN, GROWTH, AND RUPTURE OF SACCULAR ANEURYSMS: A REVIEW. Laligan N. Sekhar and Roberto C. Heros in *Neurosurgery,* Vol. 8, No. 2, pages 248–260; February, 1981.

THE COOPERATIVE BREEDING BEHAVIOR OF THE GREEN WOODHOOPOE

THE EVOLUTION OF COOPERATIVE BREEDING IN BIRDS. J. T. Emlen in *Behavioral Ecology: An Evolutionary Approach.* J. R. Krebs and N. B. Davies. Sinauer Associates, Inc., 1978.

AVIAN COMMUNAL BREEDING SYSTEMS. J. L. Brown in *Annual Review of Ecology and Systematics,* Vol. 9, pages 123–155; 1978.

THE COMMUNAL SOCIAL SYSTEM OF THE GREEN WOODHOOPOE IN KENYA. J. David Ligon and Sandra H. Ligon in *Living Bird,* Vol. 16, pages 159–197; 1978.

DEMOGRAPHIC PATTERNS AND COMMUNAL BREEDING IN THE GREEN WOODHOOPOE, *PHOENICULUS PURPUREUS.* J. D. Ligon in *Natural Selection and Social Behavior: Recent Research and New Theory,* edited by R. D. Alexander and D. W. Tinkle. Chiron Press, Inc., 1981.

ECOLOGICAL FACTOR AND KIN SELECTION IN THE EVOLUTION OF COOPERATIVE BREEDING IN BIRDS. W. D. Koenig and F. A. Pitelka in *Natural Selection and Social Behavior: Recent Research and New Theory,* edited by R. D. Alexander and D. W. Tinkle. Chiron Press, Inc., 1981.

RENAISSANCE INTARSIA: THE ART OF GEOMETRY

QUATTROCENTO SCIENCE IN THE GUBBIO STUDY. Emanuel Winternitz in *The Metropolitan Museum of Art Bulletin,* Vol. 1 N.S., No. 2, pages 104–116; October, 1942.

CONFLICT OF THE MAJOR ARTS AND NEW TECHNIQUES. André Chastel in *The Flowering of the Italian Renaissance.* The Odyssey Press, 1965.

TARSIE: DESIGNS AND DESIGNERS. M. J. Thornton in *Journal of the Warburg and Courtauld Institutes,* Vol. 36, pages 377–382; 1973.

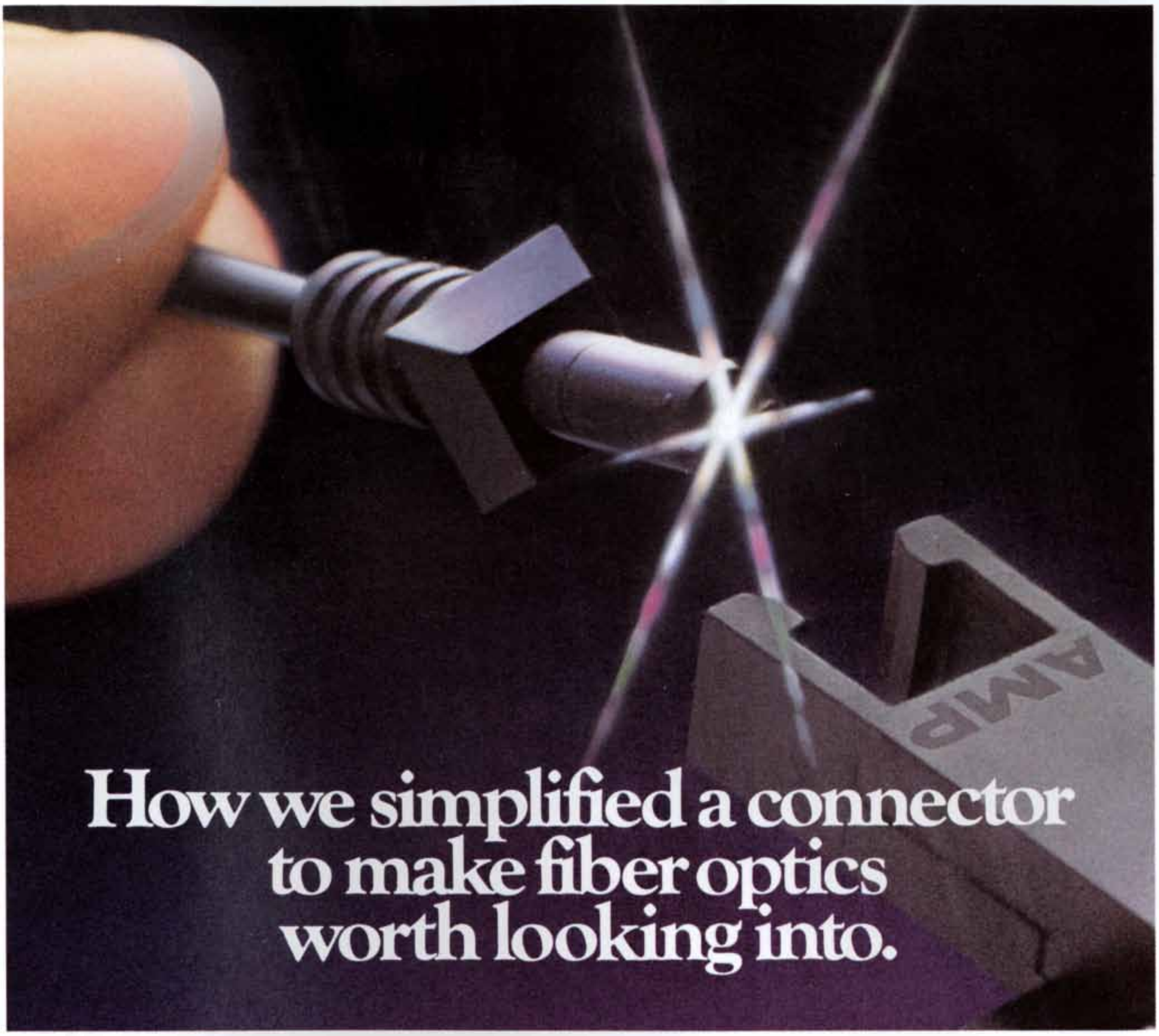
THE RENAISSANCE REDISCOVERY OF LINEAR PERSPECTIVE. Samuel Y. Edgerton, Jr. Basic Books, 1975.

THEORIES OF VISION FROM AL-KINDI TO KEPLER. David C. Lindberg. University of Chicago Press, 1976.

THE AMATEUR SCIENTIST

BOUNCING LIGHT BEAM. William M. Strouse in *American Journal of Physics,* Vol. 40, No. 6, pages 913–914; June, 1972.

TRIBOLUMINESCENCE SPECTROSCOPY OF COMMON CANDIES. Rebecca Angelos, Jeffrey I. Zink and Gordon E. Hardy in *Journal of Chemical Education,* Vol. 56, No. 6, pages 413–414; June, 1979.



If you're a company that uses electronic connectors, you need tomorrow's technology—today. And at a realistic price.

That's why you should look to AMP.

A good example is this fiber optic connector.

By simplifying its design, we made it compatible with the latest cable designs, made its installation easier and reduced the total cost of making a fiber optic link by 88 percent.

The result? Now our customers can take advantage of the future with a connector that dramatically increases productivity.

How can we help you?

Call us for solutions. Or call us for connectors. Either way, you'll get what the world has come to expect from the industry's leader in research and development. Products that are designed to 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, "Why AMP Means Productivity." AMP Incorporated, Harrisburg, PA 17105.

Auckland • Barcelona • Brussels • Buenos Aires • Dublin • Frankfurt • Harrisburg • Helsinki • s-Hertogenbosch • Kowloon • London • Luzern • Mexico City
Montreal • Oslo • Paris • San Juan • Sao Paulo • Singapore • Stockholm • Sydney • Taipei • Turin • Toronto • Tokyo • Valley Forge • Viby • Vienna

AMP means productivity.

The Most Astounding Car In The

The Probe III, an experimental Ford, drew some of the biggest crowds at the Frankfurt Auto Show. And it got great press. Germany's leading automotive publication, *Auto Motor und Sport*, called it "a dream car–slippery as a futuristic study in aerodynamics."

Aerodynamics is what the Probe III is all about. The airstream over the car does not symbolize the wave of the future, rather it shows how the Probe III looks in a wind tunnel.

It looks spectacular. A completely enclosed underbody, unique twin deck biplane rear spoiler (to maximize aero-dynamic efficiency), sculptured door mirrors and flat wheel covers all contribute to Probe III's aerodynamic excellence.

Seat belts save lives-buckle up.

German Auto Show Was A Ford.

Compare it to most '82 European models and you get a glimpse of the near future. Drag coefficient is reduced by 50%, which conceivably could result in a 27% rise in fuel efficiency at higher speeds.

But aerodynamics is not merely a future concern of Ford Motor Company. It's a science in which we are assuming a leading role right now.

Our Ford Escort and Mercury Lynx, here in America, for example, have less wind resistance than the VW Rabbit, Mercedes Benz 300D and Audi 5000.

So the Ford Probe III is not a beginning, it's another move forward.

There's A Ford In America's Future.


THE TASTE BEYOND 12-YEAR-OLD SCOTCH
Discover more.
Let your mature taste lead you from the finest premium Scotch on up to the most expensive 12-year-old Scotch in the world. The Glenlivet. The ultimate in Scotch.
Most premium Scotch is blended and depends on several whiskies for taste and smoothness.
Yet the unblended character of The Glenlivet is noble enough to stand alone. In this distinctive Scotch whisky you'll experience an exceptional smoothness and full-bodied richness, unmatched in all other premium Scotch.
Try the taste beyond premium Scotch.
The Glenlivet
Unchanged since 1824.
The GLENLIVET
AGED 12 YEARS
AGED 12 YEARS
BOTTLED IN SCOTLAND
The GLENLIVET
12 YEARS OLD
Unblended all malt
Scotch Whisky
Distilled and Bottled in Scotland
GEORGE & J. G. SMITH LIMITED
THE GLENLIVET DISTILLERY · GLENLIVET SCOTLAND
U.S. IMPORTERS
THE GLENLIVET DISTILLING COMPANY
NEW YORK, N.Y.
750 ML. (25.4 FL. OZ.) 86 PROOF