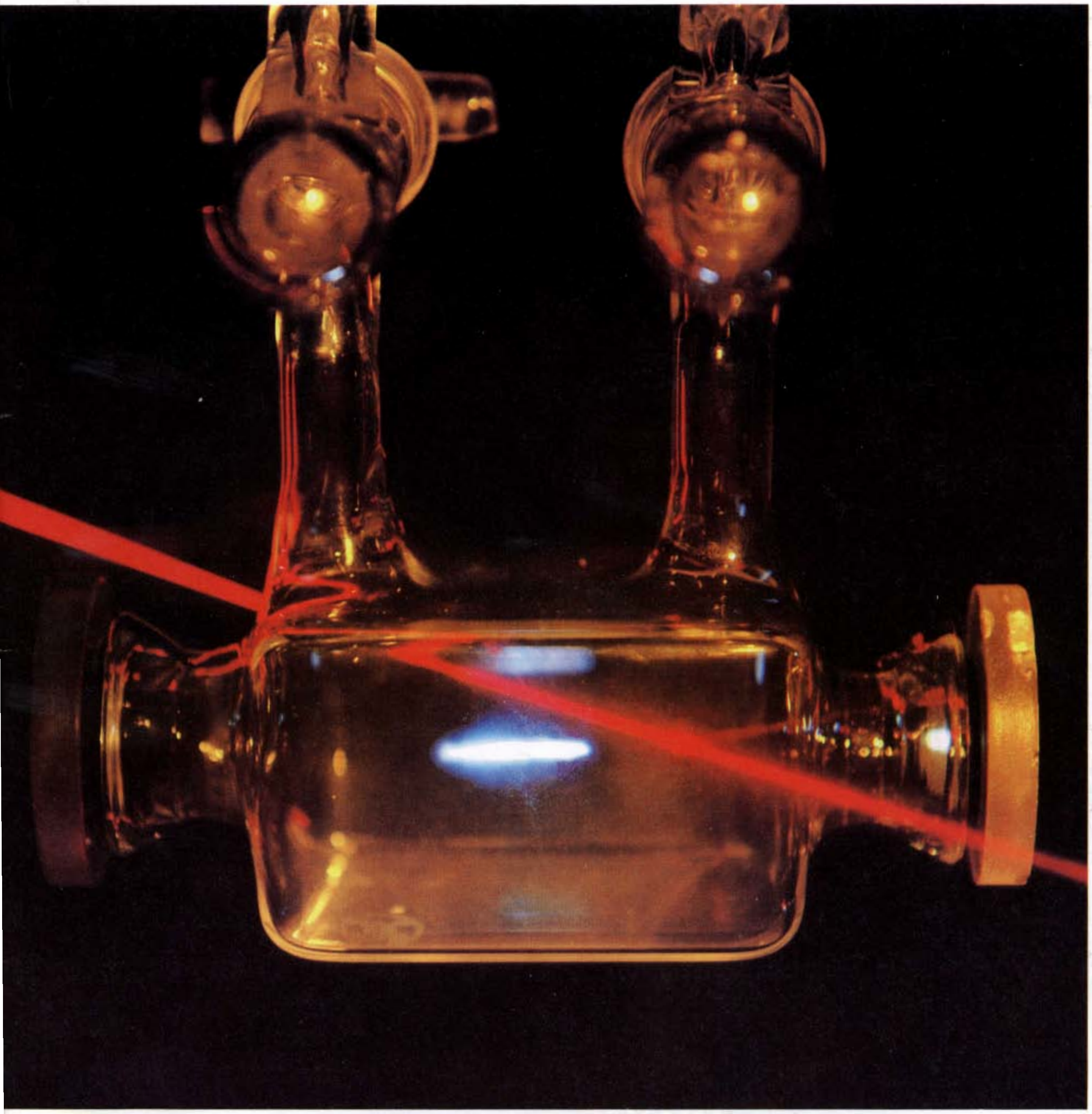
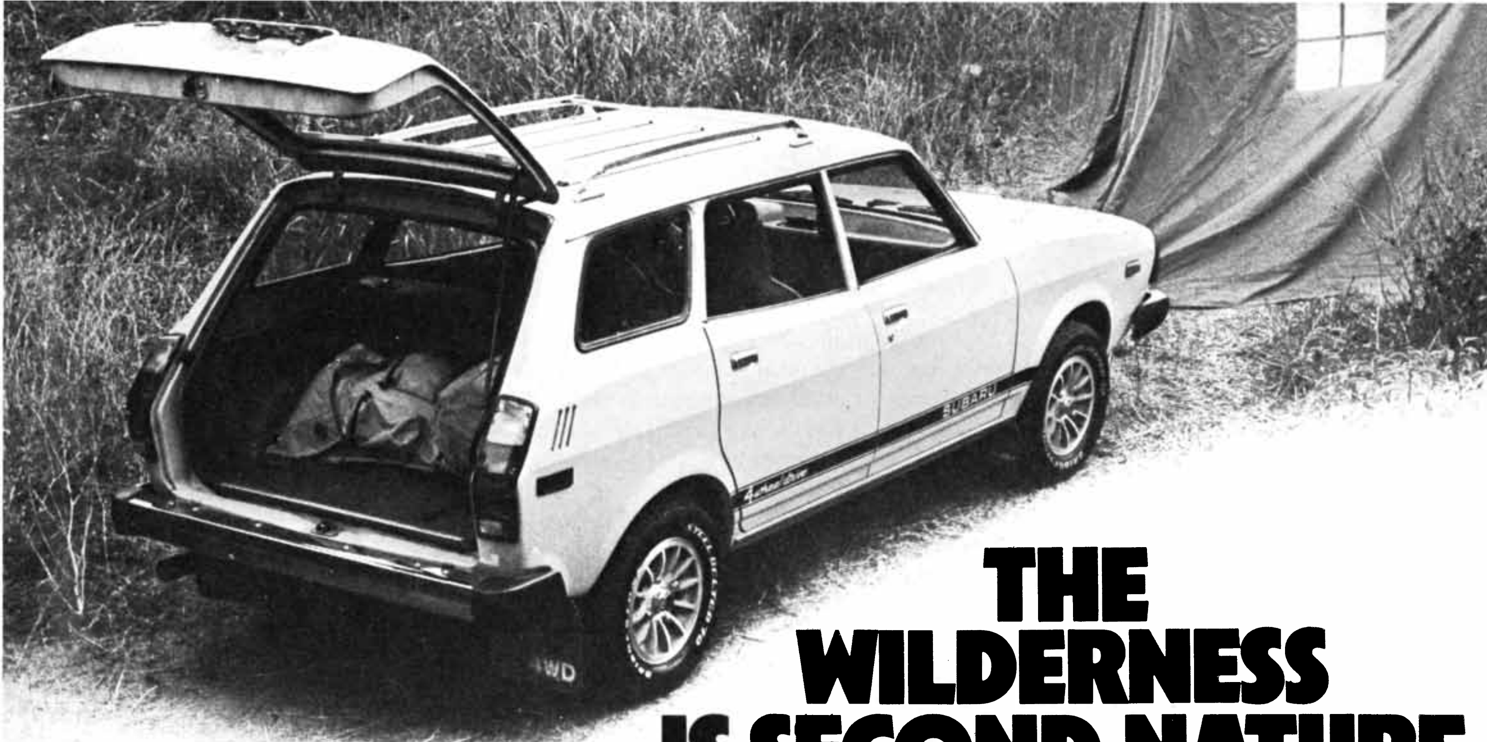
SCIENTIFIC AMERICAN



LASER CHEMISTRY

$1.50

May 1979



THE WILDERNESS IS SECOND NATURE TO OUR WAGON.

Until now, the station wagon's place was near the home. Sure it was great for picking up the kids and shopping. But when the going got tough, the station wagon didn't have what it took to get going: four wheel drive.

Our remarkably inexpensive Subaru wagon, on the other hand, is like having two cars in one.

Besides full-time front wheel drive, at a flick of a lever from inside the car, it swings right into four wheel drive. Something that makes our wagon at home even when it's far, far away from home.

What's more, the Subaru 4 wheel drive wagon has plenty of room and delivers great gas mileage. An estimated 34 highway mpg and 25 estimated city mpg using lower cost regular gas. (In Calif., it's 32 estimated hwy. and 21 estimated city mpg using unleaded).*

We look at it this way. There are times you want to get away from it all. Shouldn't your station wagon go with you.

For your nearest Subaru dealer call 800-243-6000† toll free.

*1979 EPA estimates for our 4 wheel drive vehicle. Use estimated city mpg for comparisons. Your mileage may differ depending on driving speed, weather conditions and trip length. Actual hwy. mileage will probably be less than hwy. estimate. Actual city mileage will be less in heavy traffic. †In Conn., call 1-800-882-6500. Continental U.S. only. ©Subaru of America, Inc. 1979

SUBARU
WE'RE CHANGING THE FACE OF 4 WHEEL DRIVE

Established 1845 **SCIENTIFIC AMERICAN** May 1979 Volume 240 Number 5

ARTICLES

46 **STANDARD AND DAYLIGHT-SAVING TIME, by Ian R. Bartky and Elizabeth Harrison** The long effort to achieve a unified time system in the U.S. is still in progress.

54 **CARBON MONOXIDE IN THE GALAXY, by M. A. Gordon and W. B. Burton** Its radio signal provides evidence for a vast ring of cold star-forming clouds in the Milky Way.

68 **THE SPIN OF THE PROTON, by Alan D. Krisch** Collisions between spinning protons indicate that there are objects within them that also spin.

100 **THE MOLECULAR BASIS OF CELL MOVEMENT, by Elias Lazarides and Jean Paul Revel** Not only muscle cells but also other cells incorporate fibers that enable them to move.

114 **LASER CHEMISTRY, by Avigdor M. Ronn** A laser can provide energy at the frequency needed to drive reactants down a preferred pathway.

130 **A BRAIN-COOLING SYSTEM IN MAMMALS, by Mary Ann Baker** A heat-exchange net enables some mammals to withstand lethal extremes of heat and exertion.

140 **INTRINSICALLY DIFFICULT PROBLEMS, by Larry J. Stockmeyer and Ashok K. Chandra** Such mathematical problems are solvable only by a computer as big as the universe.

160 **GIOTTO'S PORTRAIT OF HALLEY'S COMET, by Roberta J. M. Olson** One of the frescoes in the Arena Chapel in Padua shows the comet during its apparition of 1301.

DEPARTMENTS

11 LETTERS
13 50 AND 100 YEARS AGO
14 THE AUTHORS
18 MATHEMATICAL GAMES
26 BOOKS
82 SCIENCE AND THE CITIZEN
172 THE AMATEUR SCIENTIST
184 BIBLIOGRAPHY

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

There should be a word for people who take up hospital space they don't really need.

THE COVER

The photograph on the cover shows the gas carbonyl sulfide (OCS) being dissociated by an intense microsecond burst from a carbon dioxide infrared laser. The molecule of the compound is decomposed by dielectric breakdown, which occurs when the electric field associated with a laser pulse reaches a strength of a million volts per centimeter. Under these conditions molecules in the field are simply torn apart (see "Laser Chemistry," by Avigdor M. Ronn, page 114). The flash of blue-white light in the center of the reaction cell indicates the violent decomposition of carbonyl sulfide in the focused beam of the infrared laser. (The infrared radiation, of course, is invisible and leaves no record on the film.) The red beam is the path of light from a helium-neon laser that serves to monitor the decomposition process. Outside the reaction cell the path of the beam is made visible by dust particles in the air (and is enhanced by an extended exposure of the film). Inside the cell the beam is invisible until the firing of the carbon dioxide laser decomposes the carbonyl sulfide, filling the cell with particles of sulfur, which scatter the helium-neon beam and make it visible.

THE ILLUSTRATIONS

Cover photograph by Fritz Goro

Page	Source
20	Ralph Morse
47	Jerome Kuhl
48	Jerome Kuhl (*top*), Association of American Railroads (*bottom*)
50–53	Jerome Kuhl
54–56	Alan D. Iselin
57	Kitt Peak National Observatory (*top*), Alan D. Iselin (*bottom*)
58–59	Alan D. Iselin
60	Kitt Peak National Observatory
61	National Radio Astronomy Observatory
62–66	Alan D. Iselin
67	Kitt Peak National Observatory
69	Fritz Goro
70–80	Dan Todd
101	Elias Lazarides
102–103	Jean Paul Revel
104	Ilil Arbel
105	Elias Lazarides (*top*), Keigi Fujiwara and Thomas D. Pollard (*bottom*)
106–109	Elias Lazarides
110–112	Ilil Arbel
115	Fritz Goro
116–128	Gabor Kiss
131–132	Carol Donner
133	Mary Ann Baker
134	Carol Donner
135	Ilil Arbel
136	Carol Donner
137	Ilil Arbel
141	Patricia J. Wynne
142	Patricia J. Wynne (*top*), Albert Miller (*bottom*)
145	Patricia J. Wynne (*top*), Albert Miller (*bottom*)
146	Patricia J. Wynne
148–154	Albert Miller
156	Albert Miller (*top*), Patricia J. Wynne (*bottom*)
161	Scala/Editorial Photocolor Archives
162	Scala/Editorial Photocolor Archives (*top*), Hale Observatories (*bottom*)
165	New York Public Library
166	Bettmann Archive (*top*); Master and Fellows of Trinity College, Cambridge (*bottom*)
167	National Gallery of Art (*top*), Alinari/Editorial Photocolor Archives (*bottom*)
168	Kupferstichkabinett, Berlin (*top*); Metropolitan Museum of Art (*bottom*)
170	Dan Todd
172–182	Michael Goodman

You owe it to yourself to drive this car.

Introducing the all-new 1979 Mercury Marquis.

Take the Ten Minute Test Drive and discover a new Marquis standard of driving comfort.

You owe it to yourself to spend ten minutes comparing the 1979 Marquis to last year's Marquis and comparing its EPA Interior Volume Index ratings to any full-size car.

Minute 1: Check the luggage capacity.

No American car has more. With the new mini spare, there's even more room than last year.

Minute 2: Settle into a more spacious interior.

More shoulder room and hip room than the biggest '79 Buick or Olds: there's nearly half a foot more hip room in front. The '79 Marquis is even roomier than last year.

Minute 3: Take corners flatter.

Marquis has a new rear suspension. A four-bar link design with axle centered coil springs, front mounted shock absorbers and larger bushings at all suspension points.

Minute 5: Take bumps and dips with more stability.

Lincoln-Mercury took years creating a tradition of Marquis riding comfort. Ten minutes on the road and you'll discover how far we've come in 1979.

Minute 7: Handle with new maneuverability.

A reduced steering ratio makes steering more responsive, maneuvering in tight places easier.

Minute 9: View the road through expanded glass area.

More glass, a lower hood and new seating give you a commanding view of the road.

Minute 10: Read more visible instruments. Operate more accessible controls.

Dials are positioned to be read easily. Horn, wipers, high beam switches are on the steering column virtually at your fingertips.

Listen to Buick and Olds owners who drove Mercury Marquis and bought it. Dial toll-free.

(800) 228-2400.

In Nebraska, (800) 642-8888. Alaska and Hawaii, (800) 228-2424.

Then take the Ten Minute Test Drive. You owe it to yourself.

A new Marquis standard of driving comfort.

MERCURY MARQUIS

LINCOLN-MERCURY DIVISION Ford

The NASA Space Telescope: Getting ready for the clearest look yet into space

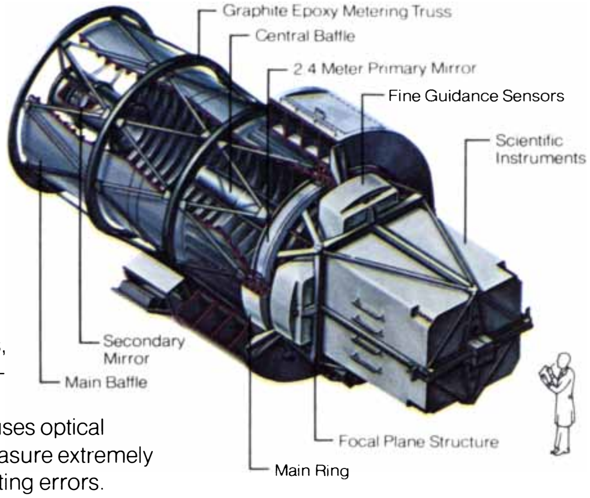
The NASA Space Telescope, scheduled for launch by the Space Shuttle in the 1980's, will orbit the earth at an altitude of 310 miles. Unlike ground-based telescopes which are restricted to a narrow spectral window and subject to distortions by the earth's atmosphere, the Space Telescope will provide astronomers with the clearest view yet of the universe.

The optical telescope assembly for the Space Telescope is being designed and built by Perkin-Elmer. It will permit astronomical observations over a wide range of the electromagnetic spectrum, including ultraviolet, visible, and infrared wavelengths. The telescope's large 2.4 meter aperture will permit observation of stars and galaxies 1/50th as bright as can now be observed by the largest ground-based telescopes. Faint objects are of particular interest to astronomers because of their importance in observational tests of theories of cosmology and the evolution of galaxies, including our own Milky Way.

The telescope will be able to point anywhere in the celestial sphere and hold onto a target for extended periods within .007 arc second.

To accomplish this, Perkin-Elmer has designed a fine guidance system which uses optical interferometry to measure extremely small telescope pointing errors.

Once in orbit, the telescope system and its instruments will operate entirely under remote control from the ground. The system is designed for a reliable operational lifetime of 15 years. Its science instrument packages are replaceable in orbit by astronauts, or the entire Space Telescope can be returned in the Shuttle for ground-based refurbishment.



The telescope's primary mirror will be figured on a precision optical polisher developed by Perkin-Elmer and controlled by a Perkin-Elmer computer.

New, computerized infrared data system: Fast, precise, low-cost tool for chemical analysis

The dispersive infrared spectrophotometer is used to identify unknown substances as diverse as impurities in a cosmetic or rust flecks in an oil spill. Very simply, the spectrophotometer organizes (or "disperses") invisible infrared radiation by wavelength, the same way nature organizes visible light in a rainbow, and beams it at an unknown substance. Since no two materials absorb infrared (IR) in quite the same way, graphing the amount of radiation that *is* absorbed at every frequency creates an absorption spectrum, a wavy line that identifies the unknown material as precisely as a fingerprint identifies a human being.

The only way a spectrum can be matched with the substance it identifies is to compare it visually with spectra of known substances. Up to now this has been a time-consuming process. Now, Perkin-Elmer has introduced a low-cost computing

Perkin-Elmer IR Data Station can overlay two spectra and highlight differences between them (bright band on screen).

and display facility—we call it the IR Data Station—which, linked to one of our IR spectrophotometers, greatly improves the ease, speed and accuracy of the process. As one early user put it, the Data Station "does for IR spectroscopy what the slide rule did for mental arithmetic."

Only moments after the operator puts a sample into the spectrophotometer's sample compartment, a spectrum of the substance traces across the Data Station's CRT screen. At the touch of a button, this spectrum can be accumulated or smoothed to get rid of electronic baseline noise (which may be present if the sample was microscopic or ill-prepared). It can then be compared with standard reference spectra for identification. Since there are spectra for literally hundreds of thousands of substances on record, this is easier to say than to do.

So the Data Station enumerates the peaks of a sample's spectrum and compares them with lists of peaks stored in its memory. It then gives the operator a "hit list" of probable identifications which can be checked quickly against a spectrum library. The saving in time and effort is immense.

For some commercial applications, the Data Station can summon known spectra to the screen, one or two at a time, scaled and oriented for comparison with the unknown sample. It can also lay one spectrum over the other and highlight the difference.

In brief, a remarkable system. If you are involved in chemical analysis of any kind, we'd be delighted to send you more detailed information about it. See below for where to write.

First of a new generation of minicomputer systems: The Perkin-Elmer 3220 and FORTRAN VII

Only a few years ago, big data manipulation jobs such as handling the reservations for an airline were strictly the province of large, powerful, and expensive mainframe computers. Minicomputers did "mini" jobs—handling the data generated by a single scientific instrument, for example.

Then, in 1974, Perkin-Elmer introduced the world's first 32-bit minicomputer. Priced far lower than a mainframe, it nonetheless could perform many of the same tasks.

Now Perkin-Elmer has introduced the first of the next generation of 32-bit "superminis"—the 3220. The 3220 incorporates 500 man-years of software development and all the advanced 32-bit technology we've developed during the past five years in over 2000 installations. It can handle massive real-time simulation, data acquisition, and scientific data reduction more economically than mainframes. Yet the 3220 costs only half as much as other systems in its class.

For cognoscenti, some key facts:

The fast Metal Oxide Semiconductor (MOS) memory is expandable from 256 kilobytes to four megabytes.

Processor architecture is based on 128 registers, divided into eight sets, enabling the processor to switch and execute tasks more rapidly than conventional architecture.

An optional high-speed cache memory reduces main memory access time to accelerate program execution.

A floating point processor performs single- and double-precision arithmetic with unprecedented accuracy.

But it is more than hardware that makes the 3220: it is the interaction of hardware with software. For example, the Perkin-Elmer developed FORTRAN VII compiler "globally" optimizes programs. That is, every flow path is analyzed to minimize the demand on system resources. The result: exceptionally fast run times.

System capacity is exceptional, too: up to 32 users can control as many as 255 different programs simultaneously in a 3220 system.

In short, the 3220 is ideal for the system builder and user whose applications require the computational power of a 32-bit processor coupled with reliable system software at a very reasonable price.

The 3220 is the first of a new series of low-cost, high-performance 32-bit superminis from Perkin-Elmer.

For more information on these products, please write. Corporate Communications, Perkin-Elmer Corporation, Main Ave., Norwalk, CT 06856.

PERKIN-ELMER

Responsive Technology

LETTERS

Sirs:

In Donald B. Redford's article "The Razed Temple of Akhenaten" [SCIENTIFIC AMERICAN, December, 1978] he dismisses the possibility that Akhenaten's monotheism was a forerunner of Judaism, Christianity and Islam with the observation that "the spiritual and ethical content of Hebrew religion and nascent Christianity are entirely lacking in Akhenaten's program, just as his religion's solar aspects are equally foreign to the teachings of both Moses and Christ." J. H. Breasted pointed out years ago that "recent excavations at Samaria have revealed the fact that... Egyptian conceptions of the righteous Sun-god were common in Palestinian life.... The winged Sun-god of the Nile was not only known to the Hebrews as a God of righteousness but also as the beneficent protector of his worshippers" (*The Dawn of Conscience,* by James Henry Breasted, Charles Scribner's Sons, 1933). Ezekiel speaks of solar worship (Ezek. 8:16 ff.), and the Solomonic temple is known to have a solar orientation and an inscription implying sun worship.

Far from being foreign to the spiritual content of Judaism, Christianity and Islam the religion of Akhenaten is the only possible origin of the benevolent, protective aspects of these monotheisms. The religion of primitive pastoral monotheists is always punitive and chastising, and such characteristics in Judaism, Christianity and Islam are easily accounted for as being the legacy of the ancient Hebrew nomads. For an explanation of the providential aspects of the deity in these religions we can do little else than look to ancient Egypt.

CHRISTOPHER BADCOCK

London School of Economics
and Political Science
University of London
London

Sirs:

Mr. Badcock has taken a generalization made by a gifted scholar more than 65 years ago, long before modern archaeology had clarified the picture in the Middle East, and has pressed it beyond all measure. True, the Levant in antiquity knew of the Egyptian solar cult, as we know of Hinduism, but it had as much impact on Canaanite or Hebrew religion as Hinduism has on Judaism or Christianity today. Solar worship among the Canaanites centered on Shaphash, the sun-goddess, her cult and iconography, or to a lesser extent on the Mesopotamian Shamash, but these owed next to nothing to Egyptian inspiration. As for the solar cult Ezekiel mentions, it is clearly anathema not only to this prophet but also to normative Hebrew religion at the time. And if Mr. Badcock cares to inquire of more authoritative sources than he has given evidence of perusing, he will discover numerous temples, including some of Osiris and sundry other mortuary figures radically antithetical to sun worship, oriented toward the rising sun.

The fact is, and anyone who consults the primary sources rather than the vast and unreliable secondary literature on the subject will soon find it out, that the cult of the sun disk was a cold, elitist monotheism, which delighted in the wonders of nature but displayed no compassion. It was far more political in inspiration than most people realize, and it stressed the monarchical principle above all. Certainly there were divinities in the Egyptian pantheon who were said to take compassion on the widow, to hear the cry of the poor man wronged in the law court and to care for the orphan, but the chief of these was the god Amun, and it was he who was the prime target of Akhenaten's wrath!

DONALD B. REDFORD

Director
Akhenaten Temple Project
Department of Near Eastern Studies
University of Toronto
Toronto

Scientific American, May, 1979; Vol. 240, No. 5. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, vice-president and treasurer; C. John Kirby, vice-president; Arlene Wright, assistant treasurer.

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

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

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 40 cents.

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

Name ____________

New Address ____________

Old Address ____________

50 AND 100 YEARS AGO

MAY, 1929: "Matter and energy have merged into each other. Both are treated geometrically as if they were properties of space, or rather of the greater generalization called space-time. There is beginning to be a great unification that seems likely to lead to an ultimate simplification. The whole of activity consists in the transmutation of forms of energy, although what energy is in itself remains an open question. Space is discovered to have physical properties, and just as Faraday called our attention from the conductors supporting electric charge to the space surrounding them, and showed that all the observed phenomena really went on in that apparently empty space, so it would appear to be with our wider outlook in the near future. Matter is turning out to be an insignificant portion of the entire physical universe, a rare and occasional perturbation of its vast extent, and probably the more important, although certainly the more elusive, activities occur in the inter-atomic, inter-planetary and inter-stellar space."

"A very red and cool star such as Alpha Herculis has a surface temperature of about 2,300 degrees Centigrade. A moderately red one such as Betelgeuse is at 2,800 degrees. As the temperature rises the proportion of the energy that is radiated within the limits of the visible spectrum rapidly increases, until at about 6,000 degrees the region of strongest radiation from the star's surface coincides with that in which the radiations affect our eyes most powerfully. For still hotter stars an increasing fraction of the radiant energy moves over into the ultra-violet. If we could observe from a station on the moon, there is no doubt that we would find this to be the case, but our terrestrial measures have to be made through miles of air, which is quite opaque to ultra-violet radiation only a little beyond the limit to which the eye can see. If a star's temperature is 10,000 degrees, only 40 per cent of its heat gets through the earth's atmosphere. At 15,000 degrees the transmission is 21 per cent, and at 20,000 degrees it is only 11 per cent."

"The reason the anti-aircraft gun did not make a more favorable record in the World War was that its shells exploded too high or too low, too far to the right or the left, too far to the rear or too much in advance of the plane. After the war the Ordnance Department of the Army inaugurated the development and construction of entirely new weapons of all types. In the new method of anti-aircraft fire control the altitude of the airplane is measured from the ground by means of the stereoscopic height finder, an optical instrument developed specially for the purpose. The necessary firing data for the guns are computed automatically by the anti-aircraft director, a calculating machine that, if kept directed at the target by means of two telescopes, will not only continuously compute the range, speed and course of the plane but also continuously indicate the elevation and direction at which the gun should be set so that it will point ahead of the target by the correct amount at all times. The prevalent idea that the airplane will in future wars ride roughshod over all existing defenses unscathed will, as years go by, be forgotten, and the airplane, like the submarine, the tank and the gun, will take its place as an important auxiliary tool in the compli cated machinery of war."

MAY, 1879: "Now that the publication of Mr. Edison's patents for electric illumination has made the public acquainted with the details of his process, it is well to recall what was done on this subject many years ago. Dr. John W. Draper, in a memoir published in the *American Journal of Arts and Sciences*, 1847, and also in the *London, Edinburgh, and Dublin Philosophical Magazine* of the same year, made an exhaustive examination of the subject. He used a strip of platinum, brought to incandescence by the passage of a voltaic current through it, and showed that the light emitted increases in brilliancy far more rapidly than the temperature increases. This memoir treats of the whole subject of the incandescence of platinum very exhaustively, measuring the heat emitted, the light emitted and the spectrum analysis of the light. Gas companies and others interested in the rivalry between electric and gas illumination will do well to examine it closely. At that time the method could not be recommended for public use, because it required a nitric acid battery. The dynamo-electric machine has of late years removed that difficulty."

"At the U.S. Naval Academy, Ensign A. A. Michelson has begun (under orders from the Naval Department and with funds supplied by Mr. A. H. Heminway of New York) the erection of apparatus for the more accurate determination of the velocity of light. The method to be employed by Ensign Michelson is described as essentially that of Foucault, with the exception that a lens of great focal length and a plane mirror are used instead of a concave mirror. This arrangement permits the use of a considerable distance, giving a longer interval of time and insuring greater accuracy. The displacement of the image of a slit is the quantity to be measured, and this in Foucault's experiment was a fraction of a millimeter. In the experiments made at Annapolis by Ensign Michelson the displacement has been increased to more than 100 millimeters. Hence the error introduced by the measurement would be less than a thousandth of the whole."

"Geologists, astronomers and physicists alike have hitherto been baffled in their attempts to set up any satisfactory kind of chronometers that will approximately measure geological time and thus afford us some clew to the antiquity of the globe. Mr. Millard Reade has recently contributed to the Royal Society a very suggestive paper in which he endeavors to grapple with the question by employing the limestone rocks of the earth's crust as an index of geological time. Mr. Reade concludes that the elimination of the calcareous matter now found in all the sedimentary strata must have occupied at least 600 millions of years. This, therefore, represents the minimum age of the world. Mr. Reade is led to believe that geological time has been enormously in excess of the limits urged by certain physicists, and that it has been ample to allow for all the changes that on the hypothesis of evolution have occurred in the organic world."

"In his letter published in the April 5th issue of *Scientific American*, Mr. Gary intimates that the world is not indebted to 'learned professors' and to 'laboratories' for knowledge of the laws of magnetism and of electricity. Let us see who has done the work and given us the laws of magnetism and electricity: Gilbert, Fellow of the College of Physicians, London; Galvani, Professor of Anatomy, University of Bologna; Volta, Professor of Natural Philosophy, University of Pavia; Oersted, Professor of Natural Philosophy, University of Copenhagen; Ampère, Inspector General of the University of Paris; Ohm, Professor of Mathematics, College of Cologne; Weber, Professor of Natural Philosophy, Goettingen; Faraday, Professor of Chemistry, Royal Institution, London; Thomson, Professor of Natural Philosophy, University of Glasgow; Maxwell, Professor of Natural Philosophy, University of Cambridge, and Henry, Professor of Natural Philosophy, Princeton College. These are the men who have discovered virtually all we know about these matters, and so it would appear that 'learned professors' have done the work, and the work was done in 'laboratories.'"

THE AUTHORS

IAN R. BARTKY and ELIZABETH HARRISON ("Standard and Daylight-saving Time") have served in an advisory capacity to the Interstate and Foreign Commerce Committee of the House of Representatives on a number of topics, including daylight-saving time. Bartky is scientific assistant to the director of the National Measurement Laboratory of the National Bureau of Standards. He did his undergraduate work at the Illinois Institute of Technology and received his Ph.D. in chemistry from the University of California at Berkeley in 1962. He then went to the National Bureau of Standards, where he has spent most of his career reviewing programs in areas such as aluminum house wiring, asbestos regulation, energy-related inventions, artificial football turf and consumer products. In 1973 Bartky was a science and technology fellow with the Department of Commerce. Harrison, an attorney, is counsel to the House Interstate and Foreign Commerce Committee. She did her undergraduate work at Columbia University and obtained her J.D. in 1970 from the Georgetown University Law Center. Her main legislative areas include the departments of Energy, Transportation and Commerce, the Federal Trade Commission and the Consumer Product Safety Commission.

M. A. GORDON and W. B. BURTON ("Carbon Monoxide in the Galaxy") have collaborated on studies of galactic structure based on observations with radio telescopes. Gordon is an assistant director of the National Radio Astronomy Observatory and manages the observatory's facilities at Tucson and Kitt Peak, Ariz. He studied physics at Yale University and received his Ph.D. in astrogeophysics from the University of Colorado in 1966. His research interests have included the earth's aurora, radio emissions from Jupiter, the physics of the interstellar medium and the structure of the galaxy. Gordon is a bicycling enthusiast of long standing; in 1976 he bicycled 1,076 miles from Missoula, Mont., to Pueblo, Colo. Burton is professor of astronomy and chairman of the department of astronomy at the University of Minnesota. He studied mathematics and astronomy at Swarthmore College and received a Fulbright Scholarship to do graduate work in astronomy and physics at the University of Leiden, which gave him his doctorate in 1970. He then worked at the National Radio Astronomy Observatory's offices in Charlottesville, Va. After eight years there he moved to Minnesota last fall. His principal interests are galactic structure and computer analysis of radio-telescope observations. Burton's major avocation is woodworking, and he is also an accomplished horseman.

ALAN D. KRISCH ("The Spin of the Proton") is professor of physics at the University of Michigan. He did his undergraduate work at the University of Pennsylvania and received his Ph.D. from Cornell University in 1964. In that year he joined the Michigan faculty. His research has focused on the study of proton-proton interactions in a search for smaller objects within the proton. Since 1972 Krisch has concentrated on developing and utilizing the polarized proton beam in the zero-gradient synchrotron at the Argonne National Laboratory.

ELIAS LAZARIDES and JEAN PAUL REVEL ("The Molecular Basis of Cell Movement") are cell biologists at the California Institute of Technology. Lazarides is assistant professor of biology. Born in Greece, he came to the U.S. in 1968 to study chemistry and biology at Wesleyan University. He then did graduate work in biochemistry at Harvard University, obtaining his Ph.D. in 1975. After a postdoctoral fellowship in Keith R. Porter's laboratory at the University of Colorado at Boulder, Lazarides went to Cal Tech in 1977. Revel is professor of biology. He was born in France and studied at the University of Strasbourg. He then came to the U.S. to do graduate work in biochemistry at the Harvard Medical School. After receiving his Ph.D. in 1957, he joined the Harvard anatomy department and rose from research associate to professor. He moved to Cal Tech in 1971. Revel's research has focused on the use of electron microscopy in the study of the ultrastructure of cell membranes.

AVIGDOR M. RONN ("Laser Chemistry") is professor of chemistry at Brooklyn College of the City University of New York. He was born in Tel Aviv, where he completed his high school education and served in the artillery corps of the Israel Defense Forces from 1956 to 1958. He then came to the U.S. to study chemistry at the University of California at Berkeley and went on to get his Ph.D. from Harvard University in 1966. After a postdoctoral fellowship at the National Bureau of Standards, he joined the faculty of the Polytechnic Institute of Brooklyn. He spent the academic year 1971–72 as a Sloan Foundation fellow at Tel Aviv University, and in 1973 he was a fellow of the Organization of American States at the University of São Paulo. Ronn joined the Brooklyn College faculty in September, 1973; his research has since focused on infrared lasers and the laser separation of isotopes.

MARY ANN BAKER ("A Brain-cooling System in Mammals") is associate professor of biomedical sciences and biology at the University of California at Riverside. She obtained her M.S. in exercise physiology at the University of California at Santa Barbara and her Ph.D. in anatomy from the University of California at Los Angeles in 1968. Following a two-year postdoctoral fellowship at the University of Washington she joined the physiology faculty at the University of Southern California School of Medicine. She moved to Riverside in 1976 as one of the original faculty members in the Biomedical Sciences Program, an accelerated medical-school training curriculum. Baker writes: "When not running pigs on treadmills or taking blood samples from burros, I spend all the time I can spare riding horses—pleasure riding in the foothills and jumping over things at horse shows."

LARRY J. STOCKMEYER and ASHOK K. CHANDRA ("Intrinsically Difficult Problems") are computer scientists at the Thomas J. Watson Research Center of the International Business Machines Corporation. Stockmeyer is on the staff of the mathematical sciences department. He did his undergraduate work in mathematics at the Massachusetts Institute of Technology and pursued his interest in the theory of computation as a graduate student there, receiving his Ph.D. in electrical engineering in 1974. His main research interests are computational complexity and theoretical studies of data storage. Stockmeyer also plays trumpet in a jazz ensemble made up of IBM scientists and engineers. Chandra is manager of the Theory of Programming Project at the Watson Research Center. A native of India, he took his bachelor's degree in electrical engineering at the Indian Institute of Technology at Kanpur. He then came to the U.S. and obtained his master's degree at the University of California at Berkeley and his Ph.D. in computer science from Stanford University in 1973.

ROBERTA J. M. OLSON ("Giotto's Portrait of Halley's Comet") is assistant professor of art history at Wheaton College, where she specializes in Italian Renaissance art. She attended Carleton College, Hebrew University of Jerusalem and the University of Oxford and was graduated with a Phi Beta Kappa in 1969 from St. Olaf College. She received her M.A. in art history at the University of Iowa and her M.F.A. and Ph.D. degrees in art history from Princeton University, the last in 1976. Olson's doctoral thesis concerned the late paintings of Sandro Botticelli.

extend your possibilities.

New HP instrument measures transmitter frequency, power, and modulation to help keep signals in their place.

Microprocessor controlled HP 8901 modulation analyzer brings new convenience to transmitter testing by combining the functions of several instruments, and achieves a new level of accuracy in associated modulation measurements.

Today our airwaves are crowded as never before. Jostling for space in the 150 kHz to 1300 MHz frequency range alone are signals from all AM/FM broadcast stations, satellite command stations, CB and ham radio transmitters, mobile radios used by fire, police, and other public services, medical telemetry, cellular telephones, maritime and military communications, as well as air navigation and instrument landing systems.

To maintain order, these various users are restricted by law to certain frequencies, power levels, and modulation limits which are built into transmitter design specifications. Transmitters and components must be thoroughly tested during design, manufacturing, and installation phases to make sure the intended performance specifications are met, and further periodic testing is usually performed to ensure that these specifications are maintained.

The new HP 8901 modulation analyzer reduces this test time and increases measurement accuracy. It is a complete measurement system, combining the capabilities of frequency counter, power meter, and modulation meter in a compact, portable package. The 8901 can automatically characterize signals in the 150 kHz to 1300 MHz range with a single keystroke. Fully programmable, it can be remotely controlled via the Hewlett-Packard Interface Bus (HP-IB) to automate test procedures. High accuracy and portability simplify testing during field maintenance operations.

New accuracy standards for modulation measurements.

The accuracy of modulation depth and deviation measurements is within an unprecedented 1 percent. Combined with low internal noise, such accuracy gives the 8901 the ability to characterize high-precision signal sources, and makes it an excellent choice for metrology labs in calibrating signal generators.

An optional internal calibrator provides an easy means to verify the instrument's modulation measurement accuracy.

Prices start at $7500*.

HEWLETT hp PACKARD

1503 Page Mill Road, Palo Alto, California 94304

For assistance call: Washington (301) 948-6370, Chicago (312) 255-9800, Atlanta (404) 955-1500, Los Angeles (213) 970-7500, Toronto (416) 678-9430

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

( ) HP System 45B computer
( ) HP 8901 modulation analyzer

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

*Domestic U.S. prices only.

00944

MATHEMATICAL GAMES

How to be a psychic, even if you are a horse or some other animal

by Martin Gardner

Conjuring is the art of entertaining people by doing things that seem to violate natural laws. A woman levitates. An elephant vanishes. Doves materialize. Spoons bend when stroked with a finger. And so on.

A small branch of magic, much in the news these days because of public enthusiasm for self-styled psychics, is concerned not with doing the impossible but only with doing something extremely improbable. Feats of this kind usually call on things that serve as randomizers in games of chance, such as a deck of cards or a pair of dice, so that there is an understandable overlap between the methods followed by psychic charlatans and those followed by gambling hustlers and honest charlatans (magicians).

A full account of magic and probability would involve a multivolume discussion of ways of controlling dice, cards, flipped coins, roulette wheels, bingo and other games of chance, and of "gaffing" carnival games. In all these areas there are cheating stratagems of incredible ingenuity. As an example, a vertically spinning carnival wheel can be gaffed with the aid of a loose floorboard. One end of the board presses upward against the bottom of a vertical board that runs behind the wheel and that has a hole in it through which the rotating axle of the wheel extends. The carnival operator stands at the side of his booth opposite the wheel. By shifting his weight imperceptibly from one leg to the other, he can manipulate the loose board as a long lever and can push the vertical board up to put friction on the axle. The beauty of this particular gaff is that no one can prove that it is intentional even if the entire booth is torn apart.

This month I shall skim briefly over a few of the hundreds of ways in which fake psychics can boost their odds when they work with playing cards or ESP cards. Today any sincere parapsychologist would make certain that a subject does not see or handle the cards, but in the early days of parapsychology, when the most sensational results were obtained, such controls were often not imposed. Even today in informal testing and in public demonstrations by dishonest psychics the ESP cards are sometimes in full view.

What can a clever psychic do to raise his scores above chance levels? One method is to secretly mark some of the cards in the course of handling them during preliminary tests. There are many ways of doing it. Card hustlers often resort to what in the trade is known as daub, a waxy substance that can be put on the back of a coat button for easy access to a fingertip. It leaves on the border of the card an extremely faint smudge indistinguishable from the smudges cards acquire in ordinary handling. Another method is to put a tiny nick on the edge of the card with a fingernail. A third method is to scrape a thumbnail about an eighth of an inch along the edge to produce a whitish line. It is only necessary to mark a few cards in order to raise a score to significant levels.

Sometimes it is not even necessary to mark cards to recognize them from the back. After playing cards or ESP cards have been used they tend to acquire all kinds of minute imperfections: a dirt mark, a slightly bent corner and so on. That is why in professional card games new decks are constantly being produced. Suppose a sharp-eyed poker player notices that the king of hearts has a tiny speck of dirt at one corner. If in a game of stud poker he sees the speck on an opponent's hole card, is he cheating or just being alert?

It is not generally known, even by magicians, that the official ESP cards now in use (authorized by J. B. Rhine) have what card magicians call "one-way backs." This means that if you examine the backs carefully, you will find they are not the same when the card is rotated 180 degrees. For example, the upper right-hand corner of the back of an ESP card either has a star there or it does not. In the course of a trial run there are many ways a psychic can set a 25-card deck so that all its cards are "one way." For example, he will try to guess only the cards that are turned one way and will not try on all the others. The unguessed cards are dealt to a separate pile. After the test (on which he is likely to score at the level of random chance) one pile is turned around before the pack is reassembled. The cards are now all one way, and the psychic is ready to perform miracles.

The literature of card magic is filled with clever tricks based on the one-way principle. A psychic may spread the cards on a table, turn his back and ask someone near the left end of the table to draw a card. That person is then told to hand the card to someone near the right end of the table to verify the symbol. The second person then returns the card to the spread and gives the deck a shuffle. This maneuver nearly always reverses the card. (There are many other subtle procedures for causing one or more chosen cards to become reversed in a one-way deck.) The cards are then dealt in a row. The psychic turns around, and moving his hand slowly down the row to "feel the vibes," he easily locates the chosen card. The card is turned face up by rotating it end for end and is then replaced in the row by turning it face down from side to side, so that the cards are one way again.

It is not necessary to have the ESP deck all one way to perform similar feats. Suppose the deck is random with respect to its back patterns and that you are the psychic. Have five cards dealt in a row. Simply memorize the pattern as a binary number, say 11010. Follow the above strategy for having a card selected and returned to the row while your back is turned. You can find the card easily, and you can memorize the new binary pattern and repeat the performance as often as you like.

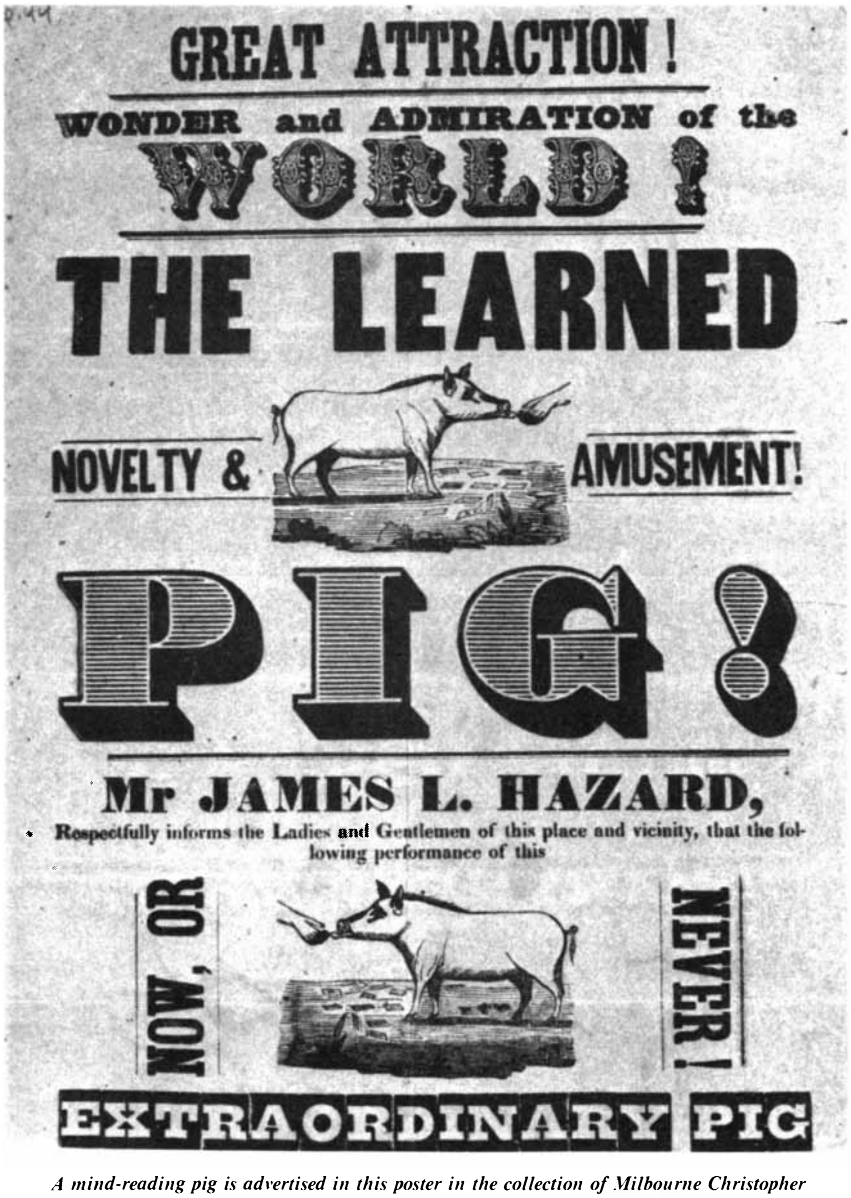
A standard deck of 25 ESP cards has five cards for each symbol: a star, a cross, a circle, wavy lines and a square. It is only necessary to have the 10 cards for two symbols one way and the 15 cards for the other three symbols the other way to make an impressive score in guessing each symbol as someone deals the shuffled pack face down. For an unprepared deck the expected score is five hits, assuming that there is no feedback information about the cards until the dealing is completed. With two symbols set one way you have a chance of 1/2 of guessing correctly between the two. This is an expectation of 10/2, or five, hits. You have a 1/3 chance of guessing correctly on the other three symbols, which is an expectation of 15/3, or again five, hits. The total expectation is five plus five, or 10, hits for the run of 25 cards. If the run is repeated four times, for the standard four-run test you can expect 40 hits out of 100. Parapsychologists regard a score of 30 as indicating excellent ESP, and so a score of 40 is sensational. It is much

more impressive than a perfect score, which would strongly suggest cheating. Note the curious fact that if only one symbol is marked, your expectation is still 40 hits in a run of 100. That, however, requires always guessing one symbol correctly, which would be noticed on score sheets and give the game away.

If the cards are not visible to the psychic in a test, a common method of cheating is to rely on what magicians call a "stooge": someone who is watching behind a screen and sending secret signals to the psychic by any one of scores of little-known techniques. (One of the hardest to detect involves the electronic sending of pulses that the psychic receives by means of a tiny device inserted in the anus. Few examinations of psychics take this possibility into account.)

The use of secret stooges is common in the history of psychic phenomena. Almost all "mind reading" animals, which in the past included famous vaudeville horses, dogs and even pigs, rely on secret sound signals that can be picked up easily by species with large ears and a sense of hearing better than man's. For example, the trainer has a playing card selected and learns the card by standard magic methods. The animal then "checks" each of 52 cards spread in a large circle on the floor or displayed on a long easel. When it comes to the right card it paws it, noses it or picks it up in its mouth. The card is of course cued by a signal only the animal can hear. It may be almost anything, from a faint sniffing sound to the clicking of one fingernail against another by a hand held in a pocket or behind the back. The signaler need not be the person on stage; it can be someone sitting in the first row.

The trained animal can also pick alphabet cards to spell a word or numbered cards that display the solution to a mathematical problem called out by a spectator. Many old books go into great detail on how to train animals for such acts. One of the best is *Haney's Art of Training Animals* (Jesse Haney & Co., Publisher, 1869), which recommends fingernail signaling. Another is *The Expositor: Or Many Mysteries Unravelled,* by William Frederick Pinchbeck (privately published, 1805). It has an excellent section on training mind-reading pigs by the sniffing-sound method.



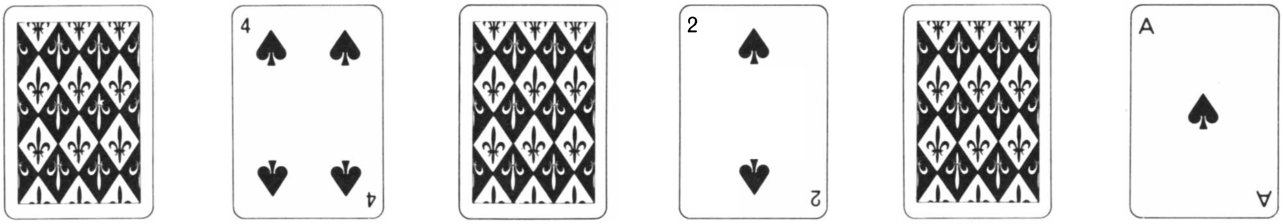
A mind-reading pig is advertised in this poster in the collection of Milbourne Christopher

When J. B. Rhine was a young man he was completely taken in by Lady Wonder, a mind-reading horse in Virginia whose psi abilities he and his wife enthusiastically described in *The Journal of Abnormal and Social Psychology* (Vol. 23, No. 4, pages 449–466; January–March, 1929). A more sophisticated Rhine later repeated his tests with Lady Wonder and found that her trainer was signaling. To this day Rhine contends that Lady Wonder did have genuine extrasensory perception at one time, and that after she had lost it her owner began signaling to her in secret. (For details see my *Fads and Fallacies in the Name of Science,* a Dover paperback, pages 351–352.)

It is usually impossible to know whether or not stooges were present by reading a parapsychologist's official report of an experiment. If you go over the papers by Harold E. Puthoff and Russell Targ that recount their clairvoyance tests of the Israeli magician Uri Geller at the time he was visiting the Stanford Research Institute, you will find no hint that Geller's best friend, Shipi Strang, was always present during the tests. When this fact came out, Puthoff and Targ agreed that he was there, but they wrote that Strang was carefully "excluded from the target area." By this they meant that he was not in the room where the randomly selected target pictures were being "sent" to Uri. But Strang, it turns out, was unattended and moving around freely, and there are many ways he could have learned of the targets and signaled them. The contrast between the elliptical, sanitized reports of the experiments and the chaotic conditions that actually prevailed, as described in John Wilhelm's *The Search for Superman* (Pocket Books, 1976), and James Randi's *The Magic of Uri Geller* (Ballantine Books, 1976), is startling. That Strang often stooged for Geller is well established, having been described by Strang's sister in an interview she gave an Israeli reporter and also by Geller's former manager, Yasha Katz, in an Italian television interview last year.

If a psychic is guessing ESP cards that are concealed from him, the cards do not have to be marked or turned one way for a stooge to send valuable information. It is only necessary for the stooge to signal the symbol after the card is turned to enable the psychic to score well above chance. With complete information about each card after the guess has been made a psychic who


JOHN
JAMESON
ESTD
1780
SINE MET
IRISH WHI
A BLEND
DISTILLED IN IRELA
LIMITED
BOW STREET DISTI
DUBLIN
IMPORTE
SKEY



FIVE OF HEARTS THREE OF SPADES RED BACK SIX OF SPADES

The red-card prediction trick

adopts his best strategy can raise his expected score on 25 cards from five to 8.65. (See Ronald C. Read, "Card-guessing with Information—a Problem in Probability," *The American Mathematical Monthly,* Vol. 69, No. 6, pages 506–511; June–July, 1962.)

If the feedback is no more than whether the psychic made a hit or a miss, it allows a strategy capable of producing an expected score of 6.63 hits. (See Persi Diaconis, "Statistical Problems in ESP Research," *Science,* Vol. 201, No. 4351, pages 131–136; July 14, 1978.) In many classic tests of ESP it is impossible to tell from published reports whether a subject was given any kind of feedback or whether friends of the subject were present who could have sent feedback by secret signaling.

In most card games feedback from dealt cards gives valuable information to a skilled player. That this is the case in stud poker and bridge is obvious. In the casino game of blackjack, or twenty-one, a player capable of memorizing the values of dealt cards can actually win consistently by betting high when the odds favor him and low when they do not. Such players are called counters, and in recent years they have done so well that most casinos now refuse to let known counters play the game. It hardly seems fair, because someone who is counting is not in any reasonable sense of the word cheating.

Intuition can go wildly astray in evaluating feedback from dealt cards. I know of no more startling example than a card-betting game discovered by Robert Connelly, a mathematician at Cornell University who made news last year by disproving a famous conjecture about polyhedrons. If a polyhedron has rigid faces but is hinged along all its edges, can it be "flexed," that is, made to alter its shape? It had been known since 1813 that the answer is no if the polyhedron is convex, and it had been conjectured that the same was true of all nonconvex polyhedrons with nonintersecting faces. Connelly found a counterexample with 18 triangular faces. (For a nontechnical description see his article "A Flexible Sphere" in *The Mathematical Intelligencer,* Vol. 1, No. 3, pages 130–131; 1978.)

Connelly calls his game Say Red. The banker shuffles a standard deck of 52 cards and slowly deals them face up. The dealt cards are left in full view where they can be inspected at any time by the player. Whenever the player wants, he may say "Red." If the next card is red, he wins the game, otherwise he loses. He must call red before the deal ends, even if he waits to call on the last card. What odds should the banker give to make it a fair game, assuming that the player adopts his best strategy on the basis of feedback from the dealt cards? The player must announce the size of his bet before each game begins. Next month I shall give Connelly's surprising answer.

Another well-kept secret of psychic charlatans for increasing their probability of success is the use of what Diaconis, in the article mentioned above, calls multiple end points. Magicians know them more informally as "outs." The basic idea is for the psychic not to specify in advance exactly what he plans to do and then let the outcome depend on what happens. As Diaconis puts it, the chance for some kind of coincidence is very much better than the chance for a coincidence specified in advance. Diaconis describes the work of a psychic known in the literature as "B. D.," whose playing-card feats were favorably written about in three papers in Rhine's *Journal of Parapsychology.* Diaconis, a former magician of great skill with cards, was present as an observer during one of B. D.'s demonstrations at Harvard University. Diaconis recognized at once that B. D.'s main secret was his reliance on outs.

Although I cannot go into detail here about how magicians who pretend to be psychics make extensive use of outs, I can illustrate the technique with an anecdote followed by two remarkable "precognition" tricks with which readers can amaze their friends. Paul Curry, a New York amateur magician, likes to remember the time a magician friend had been asked to do some card tricks at a party. He noticed that the deck handed to him included a card, say the eight of clubs, that was badly torn. Since such a card would have interfered with his manipulations, he surreptitiously removed it and put it in his pocket. During his performance he asked someone to name a card. A woman called out: "The eight of clubs." Without thinking he responded: "I'm sorry, madam, but you'll have to call another card. You see, before I started performing I noticed that the eight of clubs was torn and so I—" He broke off and almost fainted when he realized he had missed a chance to work a miracle: to make the named card vanish from the deck and reappear in his pocket, albeit damaged in transit.

It is hard to believe, but magicians have actually devised tricks with a stacked deck (a deck in prearranged order) with 52 different outs depending on the card named! The card can be shown on the top, on the bottom or as the only reversed card in the deck. One can spell or count to the card in various ways, discover a duplicate of the card tacked to the ceiling, take from the breast pocket a handkerchief and open it to show on it a large picture of the card, have a spectator look under the cushion of his seat and find a duplicate of the card, show that it is the only card missing from the deck, toss the deck against a window blind and have the blind shoot up to reveal the card pasted outside the window, and so on.

The following clever card trick, calling for six cards and six outs, has several forms. The version given here was devised by Tom Ransom of Toronto. The illustration above shows how the six cards are arranged in a row. You can tell your audience honestly that the cards bear values from 1 through 6. All the cards except one have a blue back. The two of spades has a red back and is face up. All the cards are spades except the five of hearts, which is of course face down. Write on a piece of paper, "You will choose the red card," and put the prediction face down.

Ask for any number from 1 through 6. If you like, you can hand someone an "invisible die" and ask him to pretend to roll it and tell you what number comes up. Here are the six outs for each of the six numbers:

1. From your left count to the first card. Turn it over to show its red face. Turn over the other face-down cards to show that all their faces are black.

2. Turn over the deuce to show its red back. Reverse the other two face-up cards to show that all the other cards have blue backs.

3. Ask someone to count to the third card from his left. Finish as above.

4. Count to the fourth card from your left. Finish as above.

5. Turn over all the face-down cards

"OF COURSE I OWN A RABBIT. WHO'D WANT A MERE COPY?"

The Volkswagen Rabbit is the most copied car in America.

So many cars look like Rabbits, you can hardly count them.

But you can sure tell they aren't Rabbits if you know what to look for.

Look for front-wheel drive. Very few cars have it. Which is too bad. Because front-wheel drive will help you up, around and through hills, turns, curves and nasty weather.

Look for fuel injection. Which is mighty hard to find, except for the Rabbit. You will not find fuel injection on any Honda. Or Renault. Or Mazda. Or Subaru. Or any Ford or Chevrolet or Plymouth or Dodge or Buick in Rabbit's class. So there.

Look for performance. For capacity. For handling. Very few cars in Rabbit's class combine the ability to carry as much and accelerate in the 8.3 seconds it takes the Rabbit to go from 0 to 50 m.p.h.

Even General Motors named Rabbit the best of five economy cars tested. And one of them was their own.*

So it's up to you.

You can drive the original or you can end up in one of a million copies.

But you can't say we didn't warn you.

VOLKSWAGEN DOES IT AGAIN

*Based on 1977 test.

to disclose that the five-spot card is the only one with a red face.

6. Ask the other person to count from his left to the sixth card. Finish as above.

No matter what number is chosen your prediction will be accurate. You cannot, of course, repeat the trick for the same audience.

Another ingenious trick involving outs is currently on sale in Japanese magic shops. It is the invention of Shigeo Futigawa, a Tokyo mathematics teacher and amateur magician.

You will need four identical cards that are blank on both sides. On one card print the number 17 and on the back of it print the number 30. On the other three cards print the number pairs 26/39, 28/41 and 45/58. In the first row of the illustration below are four cards that have been put down with their low number uppermost. You must memorize these four low numbers so that you can identify them quickly, or if you prefer, you can pencil a dot somewhere near each of them to distinguish them from the high numbers on the back of the cards.

Hand the four cards to someone with the request that he mix them thoroughly by turning the cards over any way he likes and that he then place the cards on the table so that the four visible numbers have been randomly selected. Before he does so you write a predicted total on a piece of paper and put it face down to one side.

When the randomized cards are on the table, there are just three possibilities. For each you follow a different procedure:

1. The cards may show two low numbers and two high ones. In the long run this will happen three out of eight times. Have the four numbers added. The sum will be 142. That is what you wrote on the paper. Let someone check your prediction.

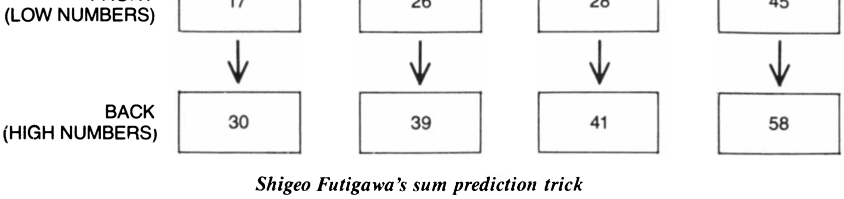
2. All four cards may be high or all four low. This happens one out of eight times. Stand up, turn your back and ask someone to randomize the numbers a bit more by turning over any pair of cards. That, of course, alters them to two high and two low, and so you finish as before.

3. The cards may show either three high and one low or three low and one high. This, perhaps surprisingly, is the commonest pattern, occurring with a probability of exactly 1/2. When it happens, pick up the single card (either the sole low card or the sole high card) and as you turn it over say: "Please notice that the numbers on opposite sides of each card are not the same. If you had placed this card down the other way, the sum of the four visible numbers would have been entirely different." You now stand up, turn your back and ask someone to reverse any two cards to randomize them further. Because that yields an even split of high and low numbers, you can finish as you did in the first case.

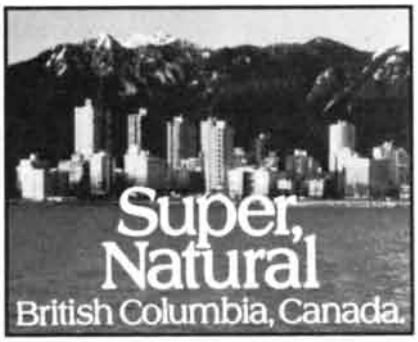
In brief, the procedure you adopt always yields a sum of 142, so you can't miss. Note that the difference between each pair of numbers on a card is 13. I leave it to the reader to prove algebraically that the trick must work. The difference can be any number. The predicted total will be the sum of the four low numbers plus twice the difference. This makes it easy to prepare cards with numbers other than the eight I have given here.

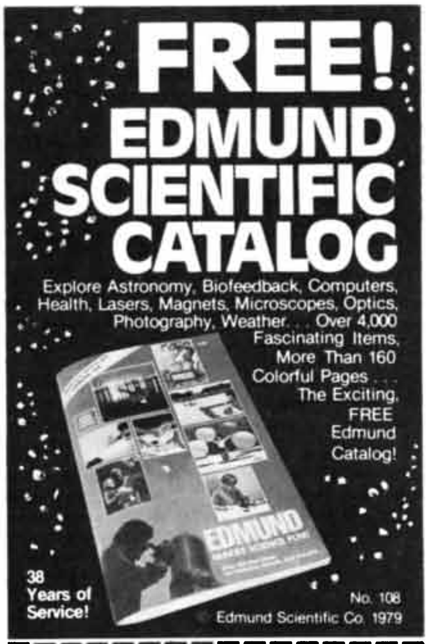
Is it possible to prepare four cards with eight different numbers so that the trick can be done exactly the same way except that the four selected numbers are multiplied instead of added? You must be able to predict the product the same way you predicted the sum, and by following the same three outs. The answer is yes, and Futigawa has designed such a set of cards as a variant of his earlier trick. Next month I shall describe this second set and explain the secret of its construction.

Two comments on my January column about tangent circles are in order. Benjamin L. Schwartz tells me that a proof for the optimal packing of six unit circles into a square first appeared in his paper "Separating Points in a Square" in *Journal of Recreational Mathematics* (Vol. 3, No. 4, pages 195–204; October, 1970). The circle-packing problem as I gave it is equivalent to the problem of placing n points inside a square so that the minimum distance between any pair of points is maximized. And my description of H. S. M. Coxeter's spiral of tangent circles as being "unique" was misleading, since it is unique only when the ratio between consecutive radii is the one specified. For details see Coxeter's paper cited in the bibliography of this issue [*page 184*].



Shigeo Futigawa's sum prediction trick





BOOKS

The health of the poor, image formation and the character of scientific progress

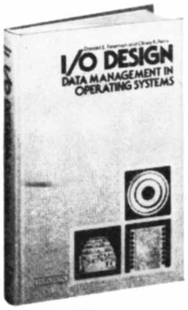
by Philip Morrison

TECHNOLOGIES FOR RURAL HEALTH: A ROYAL SOCIETY DISCUSSION, organized by D. A. J. Tyrrell, D. P. Burkitt and Sir William Henderson. The Royal Society, London (6.45 pounds). The stage is cluttered and diverse: two or three million villages, mainly tropical, and a noisy sprawl of 10,000 periurban slums. The cast is their people, with the lead roles surely played by a couple of hundred million mothers of young families. The time is the present. What we actually hear is a chorus offstage: two dozen clinical researchers, field workers and officials of international agencies gathered in the London house of the Royal Society to recount what they have seen, measured, lamented and hoped. Only a few of the papers are stilted by policy generalities, which can convey little even when they control much; the major papers treat of improving crops and feeding children, of fetching, using and safeguarding water, of direct therapies and preventive medicine, of family-planning ways and means and of organizational issues.

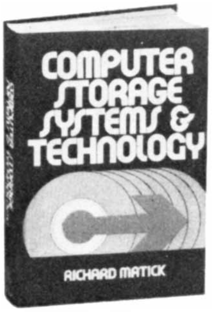
So far the typical villager has air enough and generally does not suffer much from heat loss. Water is the next requirement of human life, and it is widely wanting. "The most important innovation...for the great mass of the rural populations... would be bringing clean water to homes and taking soiled water away." The women who carry water home on heads held high know of their labor and its grace and take pride in it; they may also enjoy the deft social exchanges at the well. Statistics, however, speak coldly. Bringing water is hard work; the carriers skimp on the supply. Whereas city people flush water away heedlessly (a minimum of 100 liters per person per day and much more in the U.S.), rural and urban carriers provide only 10 to 15 liters of water per person per day, and considerably less when they must carry it a mile or more. The delicious amenity of easy washing is obvious; more than that, enough water for personal hygiene is a requisite for health. Fortunate villages may have access to streams; in drier places the extreme prevalence of infective skin and eye diseases ("often affecting over half of all school children on any day"), not to mention the dysenteries, is highly dependent on inadequate access to water. All of these can be washed away in plentiful clean water.

The need is for low-cost reticulation systems; the major gain to health is achieved by a "single tap inside the compound or home, perhaps with a shower extension.... In some places this is too expensive." Plastic pipe is the greatest innovation of recent years; it is cheap and easy to use and repair, albeit vulnerable to damage by stock and small boys. If peak flows are kept low by some cheap limiting device that is safe against bypass or other clever defeat, the cost is kept down. Taps, pumps, cheap catchments for rainwater and artful handling of well entrances to reduce parasite contamination are called for, along with other devices suited to the setting and capable of long service with simple maintenance. Water-supply development at the localized, self-help level is a natural basis for rural development. The danger is that enthusiasm may arouse expectations that cannot be met; any self-help project needs long-term support and a central commitment for maintenance and supplies.

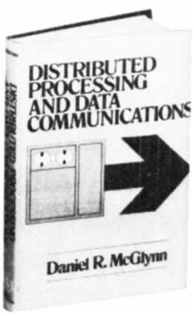
A major cause of infant mortality and a steady drain on all health resources is gastroenteritis of childhood, the commonest of the diarrheal diseases. (Cholera is the deadliest.) A remarkable study reported from Gambia shows that faltering weight gain in children month by month reflects the presence of diarrheal disease more than it does the lack of food. In the dry months food is in short supply but weight remains at an optimum level; for most of the year for most Gambian children the best way to ensure normal growth is to reduce infections, not to supplement the diet. Breast feeding for infants is safe and satisfying, but beyond three months, even when lactation is prolonged for another year or more, most mothers would have to use a supplementary food, usually the adult staple made more liquid by dilution with the water of the house. Safe preweaning feeding procedures that maintain some long-term breast feeding "should be a main priority." This system shows strong interactions: food, water, family size, disease.



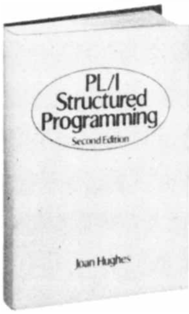
54465. I/O DESIGN: Data Management in Operating Systems. *Freeman and Perry.* Covers I/O devices, channel programming, device allocation, control statements, data organization and staging, and much more. **$18.95**



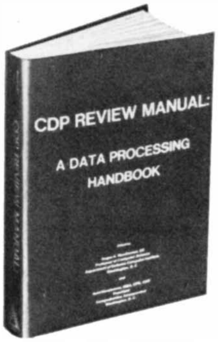
40065-2. COMPUTER STORAGE SYSTEMS AND TECHNOLOGY. *Richard Matick.* A thorough examination of available storage devices. *Counts as 2 of your 3 books.* **$29.95**



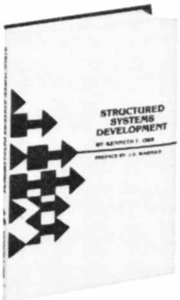
42410-2. DISTRIBUTED PROCESSING AND DATA COMMUNICATIONS. *Daniel R. McGlynn.* The technical and organizational components of data communications. *Counts as 2 of your 3 books.* **$20.75**



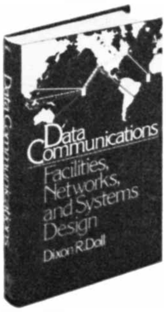
69297. PL/1 STRUCTURED PROGRAMMING. 2nd Ed. *Joan K. Hughes.* Covers the *full* language with detailed case studies, practice problems, coding explanations, a glossary of over 300 terms, and an extensive chart listing of all PL/1 keywords. **$18.95**



37400. CDP REVIEW MANUAL: A Data Processing Handbook. 2nd Edition. *Lord and Steiner.* In-depth coverage of data processing equipment, computer programming and software, data processing management, and mathematics and statistics. *Softcover.* **$19.95**



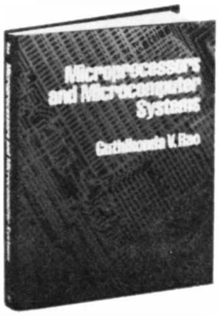
80845. STRUCTURED SYSTEMS DEVELOPMENT. *Kenneth I. Orr.* The basic steps for analyzing, developing and installing systems with the emphasis on a structured "problem-solving" approach to systems building. **$17.00**



41645-2. DATA COMMUNICATIONS: Facilities, Networks, and Systems Design. *Dixon R. Doll.* A straightforward guide that covers everything from basic terminology to design trade-offs. *Counts as 2 of your 3 books.* **$25.95**

80150. STANDARDIZED DEVELOPMENT OF COMPUTER SOFTWARE. *Robert C. Tausworthe.* How to incorporate principles and techniques of structured programming in the development of large-scale systems. **$19.95**

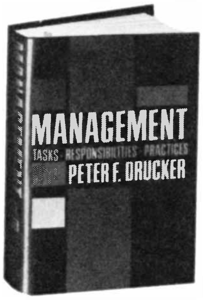
54455. IMS PROGRAMMING TECHNIQUES: A Guide to Using DL/1. *Kapp and Leben.* Spells out specific techniques to write application programs in an IMS DL/1 data base environment. **$17.95**



62651-2. MICROPROCESSORS AND MICROCOMPUTER SYSTEMS. *G. V. Rao.* Highlights the latest microelectronic and microcomputer technology. *Counts as 2 of your 3 books.* **$24.50**

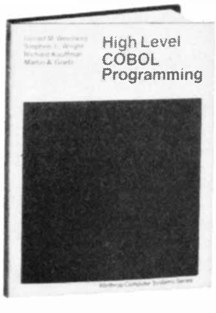
81845. SYSTEM/370 JOB CONTROL LANGUAGE. *Gary DeWard Brown.* Includes descriptions of hardware devices and access methods and reference to many useful JCL features. *Softcover.* **$12.50**

56265. INTRODUCTION TO DECSYSTEM-10 ASSEMBLER LANGUAGE PROGRAMMING. *Singer.* Softbound. **$8.95**



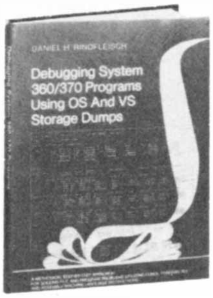
59920. MANAGEMENT: Tasks, Responsibilities, Practices. *Peter F. Drucker.* The skills, techniques, and tools a good executive needs to stay effective in modern management. **$17.50**

54450. THE IBM 5100 PORTABLE COMPUTER: A Comprehensive Guide for Users and Programmers. *Harry Katzan, Jr.* Includes BASIC and APL programs for payroll/inventory control applications. **$19.95**

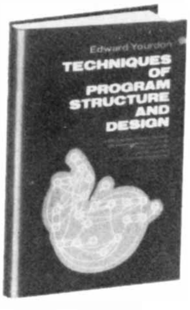


52900-2. HIGH LEVEL COBOL PROGRAMMING. *Gerald M. Weinberg et al.* Introduces "sheltered programming" concepts, stresses modular design and maintainability of code. *Counts as 2 of your 3 books.* **$21.95**

(Publishers' Prices shown)

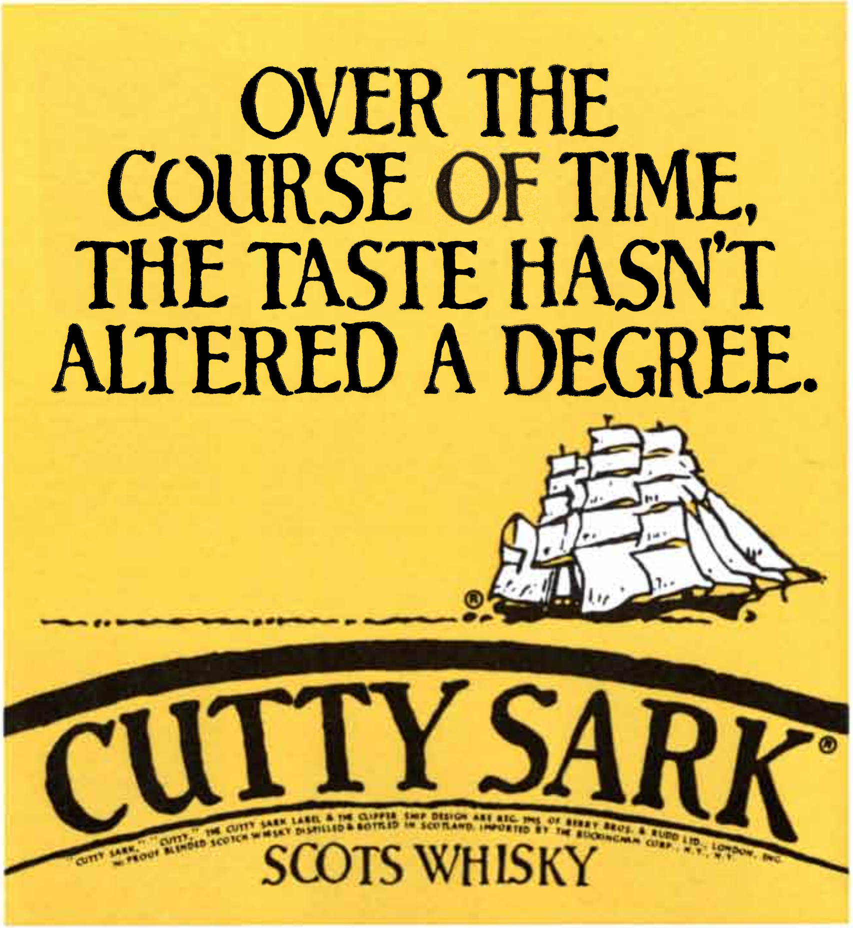


41785. DEBUGGING SYSTEM 360/370 PROGRAMS USING OS AND VS STORAGE DUMPS. *D. H. Rindfleisch.* A superior guide to storage dump debugging. Illustrations, examples, sample dumps. **$19.95**



82560. TECHNIQUES OF PROGRAM STRUCTURE AND DESIGN. *Edward Yourdon.* Over 300 illustrative problems, 90 valuable references, and four demonstration projects. **$19.50**

If the reply card has been removed, please write to
The Library of Computer and Information Sciences
Dept. 7-A7F, Riverside, N.J. 08370
to obtain membership information and an application.



The tale of the near-eradication of smallpox over 12 years (at a bargain total cost to all countries and organizations involved of only $2.5 million a year) is well told. The U.S. alone was spending $150 million a year for vaccination and quarantine. The World Health Organization vaccinator is newly freed by direct test from the "traditional ritual of cleansing the skin . . . with a cotton sponge dipped in alcohol." He carries only a simple shirt-pocket kit: two identical little plastic tubes, one for sterile needles and the other for used ones, and a vial each of vaccine and diluent. The treatment of vaccine supply as a global problem allowed a flexible approach to manufacture, some countries making their own vaccine with international technical aid, others receiving bilateral donations and still others drawing on the WHO pool, to which eventually 25 countries were contributors. The wise guiding principle was that all vaccine from whatever source be subject to testing by a WHO laboratory.

For a decade we have understood the rational treatment of nearly all acute diarrhea, both the classic killer, cholera, and the commonplace infections of children. Drugs, including antibiotics, are rarely of any use; rehydration is the only treatment necessary, and for that one standard intravenous fluid (for severe cholera) and one standard oral fluid suffice. In one typical case plotted in the book a man received 30 liters intravenously and almost 40 by mouth during the three and a half days to his recovery. Even these heroic methods cost little (say $10 worth of locally prepared fluids) and are simple—well suited to rural health centers. Such rehydration posts widely spread in small towns would provide both a solution to endemic cholera and a basis for emergency reinforcement to cope even with the terrible epidemic form.

Only one network of aid can feather out finely enough, however, to reach every sick child quickly, by night as by day, in every village. It is already in place: the mothers. A few ounces of the physiological salts (potassium, sodium and buffer) plus the indispensable 20 grams of glucose per fluid liter that so well promotes oral absorption, in a labeled packet, along with universal preaching of the motto "Replace that which is lost," would be a big step toward reducing childhood mortality at small cost. Rougher approximations to ready-mixed salts are workable as well. Here is a challenge to designers, educators, marketers and health officials.

Family planning is critically surveyed, and the strong recommendation is that many options should be widely available. "Success can be and indeed has been achieved." The primary need is not a perfect contraceptive but flexible, perceptive and consumer-centered delivery systems. Birth spacing is no new

idea; extended lactation, for example, has always been a major factor in the mean spacing of births. What contraception does is offer a choice to the one-tenth of the women who become pregnant before their first postpartum menstrual period. Motive and intent are the essentials; it appears to be good sense to integrate family planning with maternal- and child-health services. A child who survives is a clear piece of evidence to every family.

These accounts at Thameside lack neither experience nor concern, and yet they have an ex parte quality, being so far from the Niger or the Mekong. Even authoritative people from Harrow and Geneva do seem a little out of it. One discussant observed dryly: "If isolated missionaries care for the poor, something may be done locally, but unless governments care nothing can be achieved in countries at large." Governmental lack of concern is not mere neglect: villages are not simple, nor are their people uniformly in distress. In most of the world the adequately landed villagers manage rather well in sound houses, even when they live away from the paved roads, because they share in the crops of others. Here is a knot of questions that transcend even this wide discussion. The book will open up the world's real health problems in an up-to-date way for most readers; it outlines solutions less tidily. Certainly we in the capital-rich north have heavy responsibilities, beyond charity and expertise, that we have not yet fully accepted.

Life-and-death matters often lead the speakers beyond the expected clear exposition. One Liverpool clinician teaches his medical students biochemical doggerel that is right to the point. "Babies who have D. and V. / Shrivel up and fail to wee. / When this happens, then we oughta / Fill them up with salt and water." And one matter-of-fact remark sounds an authentic note of poetry. "It had been found useful to teach mothers to test the fluid by taste—they should not use it if it tasted more salt than tears."

IMAGES: A UNIFIED VIEW OF DIFFRACTION AND IMAGE FORMATION WITH ALL KINDS OF RADIATION, by Charles A. Taylor. Crane, Russak & Company, Inc. ($19.50). That historians in each generation can recount the same past with always a new vision is a truism. Changing concerns and new knowledge lead to new selections from the overflowing store of past events. This small volume (surprisingly high in price, even allowing for its many fascinating illustrations in black and white) is just such a fresh and contemporary retelling of an old theme. It is no piece of history, however. It presents the beginning physics of the familiar thin-lens equation and double slit, of pinhole and mirror—but a revisionist physics sung in our heroic time of lasers and tomography. The contemporary unified view of wave physics and information and the full range of practical image making from ultrasonics to neutron diffraction offer a deep, compelling and comprehensive way to understand physical images and how they are made. The author has cleared a broad, smooth path along that way; his exposition and its illuminating illustrations, most of them specially made for the purpose, should mold the conventional textbook explanations of the future. Professor Taylor's readers will enter that bright future at once; they need carry along the lightest of mathematical baggage, mainly the firm concepts of adding and multiplying.

Begin with a slide projector. Turn on the lamp (obviously some physical flow must carry the information) but remove the lens. Even a coarse slide, say a thick black cross, will show then only as an uninformative patch of light on the screen. The lens can clarify it all, but the lens surely does not know what is on the slide! All the information is already there on the screen, impressed firmly on the beam by the scattering of the waves as they pass the complex obstacle. That broad patch of light is a hologram; each point on the screen is receiving coded information—by way of light from every point on the object. The lens need not know anything; it simply recombines the information present in the scattered light to form a clear image. Such a decoding process, from a one-to-many map down to a one-to-one image, is the universal second step in image formation; first there is scattering, then recombination.

Three or four more or less direct recombination procedures can be outlined. A simple pinhole in an opaque card, placed anywhere in the light patch, will immediately give us a complete image. A converging lens placed anywhere in the beam will work even better. Or let the pinhole touch the slide itself. The screen will be illuminated with a dimmer but still nearly uniform patch. Choose any point on the screen and record the intensity there. Now scan the slide by moving the pinhole; the time sequence of the recorded intensity will plainly decode the image. All these direct methods are familiar and have their domains of choice.

How is the image detail decoded? It cannot be in the near-uniform intensity, or amplitude, over the screen; how could that crude variation ever yield a crisp image? It is coded rather in the phase of the waves as they strike the screen, each little train bringing the time relations of the particular path it takes from every point on the source, each producing a unique pattern of wave interferences, the summed resultant of its crests and troughs, across the screen. The patterns shift randomly and swiftly as the wave pulses from each emitting

We are proud of the birthdates of our children, the grapes of Almadén.

On our classic varietal wines, you will find a vintage date. A date that means the wine in the bottle comes from a particular year's harvest.

For us, it is a source of much pride and satisfaction. Because each vintage is like a new child. A child we lovingly care for as it matures and develops.

For you, the vintage date adds yet another dimension to the enjoyment of our classic varietal wines. Because each vintage expresses

subtle differences that are uniquely its own. Differences you may taste and enjoy.

Also, while all our wines are ready to drink when purchased, you may enjoy setting a few bottles aside. Here you will find the vintage date invaluable in giving the wine the additional aging you desire.

On our classic varietal wines, you will also find the birthplace of our children. For example, Monterey County on our Cabernet Sauvignon. And San Benito County on our Pinot Chardonnay. Each variety of grape has its own special needs. And so we take great care in finding the best possible home for each of our children, the grapes of Almadén. A home that provides ideal conditions for them to grow and ripen to perfection.

We invite you to enjoy the incomparable quality of these fine vintage-dated varietal wines. And you will find this same dedication to quality in our entire family of Almadén wines.

ALMADÉN

Celebrate Arthur Fiedler's 50th Anniversary concert with the Boston Pops on PBS Tuesday evening May 1; made possible by a grant from Almadén Vineyards, Inc. Please check your local listing for times

PENTAX SYSTEM 10.

THE FIRST 110 SLR AS PRECISE AS A FINE WATCH AND ALMOST AS SMALL.

It might make you feel good to know that if you sometimes think you are hard to satisfy, we believe *we* are almost impossible to satisfy. Ever.

This attitude was the reason we spent years and millions of dollars seeking a way to drastically improve the 110 camera.

And we have succeeded. Almost beyond our own expectations. And certainly far beyond the expectations of the photographic press.

The new Pentax System 10 is the first 110 SLR (single lens reflex) with interchangeable lenses, a power winder, a dedicated auto flash and a host of accessory close-up lenses, filters, lens shades and carrying case.

We've overcome the limitation of the fuzzy 110 picture.

In fact, this unretouched enlargement was actually taken with the Pentax System 10 camera, using its telephoto lens.

The image is razor sharp and free of perceptible grain.

What also makes System 10 unique and so astonishing is its incredible lack of size and weight, coupled with the most sophisticated SLR technology.

It measures $2\frac{1}{5} \times 3\frac{9}{10} \times 1\frac{4}{5}''$ and weighs less than 6 oz.

System 10 is designed to be the ideal complement to your 35mm SLR. The whole system can fit in a small area of your current camera case, in place of a zoom lens (but your pocket will do in a pinch).

And frankly, there are times when you just don't want to carry a 35mm SLR around. But up to now, everything else has been a compromise.

Now, with the Pentax System 10, you'll never again have to say "I wish I had my camera with me."

Pentax System 10. Available in limited quantities only at select fine photographic stores.

TECHNICAL DATA

Type: TTL metering 110 SLR camera, bayonet mount (80-degree setting angle).
Film: Cartridge 110 film format; ASA set automatically for film in use.
Lenses: Pentax-110 24mm f/2.8; focusing range: 0.35m to ∞. Pentax-110 50mm f/2.8 telephoto; focusing range: .9m to ∞. Pentax-110 18mm f/2.8 wide angle; focusing range: .25m to ∞.
Viewfinder: Eye-level pentaprism finder with instant-return mirror; green or yellow LEDs indicate correct handheld exposure, or need for flash/tripod.
Exposure Metering: Center-weighted, TTL metering at full aperture via SPD cell; EV Range: 3-17 (24mm f/2.8 lens; ASA 100).
Shutter/Automatic Diaphragm: Programmed electronic shutter with programmed body diaphragm; Range: f/2.8 to f/13.5.
Other Features: Two-stroke film advance lever, synch terminal for automatic flash synch with AF 130P auto flash; tripod socket, lens bayonet release button, battery holder tray; exposure count via back cover window, winder capability.
Power Source: Two 1.5V silver-oxide batteries.
Weight: 172 grams (6.1 ozs) w/lens. **Size:** 2.2"x 3.9"x 1.8" w/lens.

PENTAX®
SYSTEM 10

atom come randomly in time from random points within the glowing source. How we can replace the usual beam by a superposition of coherent fronts, each well defined both in space and in time, is a key issue, beautifully explained and demonstrated in pictures. Even a fully coherent beam, however, encodes the image into that wide hologram patch, which would then show a clear, if enigmatic, structure. It is that hidden reciprocal structure that joins all the forms of image making: the simple ones outlined above, the familiar but remarkably subtle lens that works by differentially delaying the transmitted wave trains in such a way as to restore a one-to-one mapping of slide on screen, and also the indirect two-stage atomic images of the X-ray crystallographer, magnified 100 million times.

The heart of this exposition is the careful, nonmathematical and visually engrossing study of coherently coded interference patterns, optical transforms of the scattering object. Professor Taylor's laboratory at University College, Cardiff, has an elegant device for making and recording such optical transforms (Fraunhofer diffraction patterns to some), which is deftly employed in the service of a rich pedagogy. (Some years ago the author and his colleagues published an admirable atlas of such transforms with a text aimed at advanced students, and he has also written a volume on sound; both books were reviewed in these columns.)

The most important results of this transform analysis are two. One is the understanding in depth of why "it is virtually impossible to impress on a wave any information about detail less in size than its wavelength." The other is an understanding of the limitations on the detail yielded by any image-forming device that arise from its finite aperture in relation to the wavelength employed; this effect—illusions of the microscope—is particularly well shown. The story ranges more widely over its broad field, however. We come to understand depth of focus, apodization, zone plates, Bragg's law (a conceptual breakthrough that paradoxically tended to overemphasize the importance of periodicity in the diffraction process, and hence to separate unduly the techniques of X-ray diffraction from image forming in more ordinary instruments), electron microscopes of both great classes, radar, and laser holography.

Some pictures that are hard to forget: a tomograph picking out one layer of a multilayered X-rayed object, an ultrasonic echo scan showing the skulls of twins in utero, an optical analogue to the field-emission electron microscope, a set of diffraction studies made looking through an umbrella, a study of pinhole portraits, and optical analogues of X-ray-diffraction patterns. The key role of partial knowledge—of knowing what the object ought to look like—is emphasized for X-ray crystallography in a simple account of the heavy-atom and other X-ray reconstruction techniques.

This book marks a new high point in the literature of basic optics. It is not perfect; there are a few editorial deficiencies that can annoy a reader, in the labeling of and references for photographs in particular. American readers could use a little help—in understanding "liquorice all-sorts," for example. Not everyone will be convinced that focusing by trial and error is in principle possible only given some notion of the object, or that the null regions in an interference pattern are receiving energy—but such doubtful points are few. Even the optical expert will profit from this account, if in no other way than by learning how he can really explain to his friends what it is he is doing! An expert physics teacher, G. E. Foxcroft of the Rugby School, has collaborated on some details as well as on the broader issues.

SCIENTIFIC PROGRESS: A PHILOSOPHICAL ESSAY ON THE ECONOMICS OF RESEARCH IN NATURAL SCIENCE, by Nicholas Rescher. University of Pittsburgh Press ($18.95). Operations research had a finest hour: after the fall of France, when the hard-pressed British fighting forces were served by unlikely backroom wizards. Ingenious and daring physicists, crystallographers and mathematicians, all but innocent of military expertise but keen to win the war, made out dimly some quantitative nubbins of truth amidst the fogs of war and were able to manipulate them to tactical advantage. Nowadays that trade has grown greatly in sophistication and has passed beyond the gifted amateurs poring over the logs of submarine attacks. So it is with econometrics and ecology: the plausible simple curves of yesteryear are today's long-running programs. It may be true still, however, that most influence lies with the simplest functions. Malthus himself treated only of addition and multiplication, and in the maze of human affairs even a thin thread of quantity may serve reason well.

Way back in Theodore Roosevelt's day it became part of the education of Henry Adams that science grows exponentially. By now the measurers have been in the field, and their data turn the early perception (shared in some measure by bearded worthies such as Lord Kelvin and Friedrich Engels) into a truism. We know it for scientific journals, papers, authors, doctorates, research and development budgets and much more, all small parts of what it is fashionable to call the *problématique*. The future is plain too: research growth must in this finite world submit to limits. The budget curve slows first, then the job market, then the doctorates. . . .

By now growth and decline are widely philosophized about and rendered rational, not to say inevitable, in the nature of things.

In this crisp yet many-sided little book we are treated to a virtuoso analytical performance, as plausible and bold as one of P. M. S. Blackett's wartime papers, although more responsive to what others have written. The philosopher-analyst author has sought to quantify the almost indefinable and to project its future from a few empirical results and a boxful of argument, based on a sweep of thought from Gottfried Wilhelm Leibniz, C. S. Peirce and E. A. Du Bois-Reymond to R. P. Feynman and Derek de Solla Price. Only straight-line plots are invoked—plotted, to be sure, on the appropriate choice among two or three kinds of graph paper!

The historical issue is clear: Are the days of scientific discovery within the natural sciences numbered? The author's answer is a ringing no, a challenge flung to the defeatism of our times in the spirit of old Leibniz' vision of an "open and hopeful future for inquiry." Rescher casts a wide net over this issue. First of all he addresses the idea that the world itself may set limits to inquiry, either because it is limited in extent and complexity or because it is fundamentally incomprehensible at some level. He is a realist; he accepts that the future of our kind is hardly infinite, but the point is whether, before human thought itself ends, science will have been played out. None of the asserted causes suffices to end our search for meanings; there is no "metaphysical hidden hand" that dooms our entry into ever deeper levels to become trivial or to lose interest. Yet exponential growth must plainly cease.

Let us assume the exponential past and the eventual constant future, both measured in the resources made available to inquiry, resources both human and economic. What can be foreseen within this reasonable framework, if no inner logic can force an end to human inquiry?

The key to Rescher's argument is the diminishing return on scientific effort. Science occupies more minds, it costs much more, it demands much aid, yet its findings of the first rank, he argues, do not increase proportionately. Topmost quality is distributed more sparsely the wider the effort and the bigger the harvest of the merely routine. He concludes—the argument is well put if less than compelling—that the decisive intellectual output has grown only uniformly with time even in the face of exponentially rising inputs. From this it follows that in a future marked by a plateau of merely uniform input the increment of first-rate new findings in science will not cease but rather will fall steadily toward an asymptotic limit, like the curve $1/t$.

This result is no mere happenstance of history. It rests on the inner nature of

real scientific discovery. Discovery is no passive intellectual act in a library chair; it demands the pursuit of new experience logarithmically across a vast parameter space, to distances great and small, energies and temperatures high and low, times long and minute, and all the rest. The author allows for the behavior of microfields that individually saturate, but that together can aggregate to a steadily richer overall discipline. His active way of understanding science, highly congenial to the physicist with growing machines and bigger computers, makes it plain that science is in an "arms race against nature," a race that demands increasing effort as human beings push beyond their own parametric neighborhood. They have kept going smoothly while they could still afford exponential input growth; therefore this has been the half millennium (say from A.D. 1600 to 2100) of a great transformation in human history, the science explosion, "as unique in its own historical structure as the Bronze Age or Industrial Revolution or the Population Explosion," and indeed not unconnected.

Of course, quantitative forecast is a bit shaky. That everything should really go as the simplest fits prescribe is not very likely; even the German submarines might have escaped their airborne fate with a lucky radar detector or two, and science is far more complex than war in the Bay of Biscay was. Russian authors, following up a tip from Engels, have studied an even more optimistic model that is also plausible enough. Their view is that science is fully cumulative, that its results add on year after year. Rescher is rather more of a revolutionary. He holds that great new findings simply overturn the old ones. For him it is the increment that is to be measured.

The difference is not unimportant. The very same simple regularities lead, in the Marxist view, to the forecast that in the coming age of a fixed rate of inputs the treasure trove of science will still grow uniformly in time. (This implies that in the past the storehouse has grown quadratically with time, as Engels said, not linearly, as Rescher argues.) A reflective reader, delighted and bemused by all this fascinating conjecture, is pretty sure to opt for the middle way. Science in fact lies between constant revolution and steady cumulation, and its plateau of input growth will therefore see intermediate—probably oscillatory—output behavior, the annual yields still dwindling over the long run, but even more slowly than $1/t$.

Classroom, laboratory bench and lunch table should ring to debates over this provocative essay. It ends naturally enough with a set of injunctions, like most operations research: Do not await the end of history; neither be a pessimist, nor depreciate achievement, nor overgeneralize the analysis (to the social sciences or the humanities, say), nor expect miracles, nor excuse inadequate effort. Rescher cites Sir Peter Medawar: "To deride the hope of progress is the ultimate fatuity, the last word in poverty of spirit and meanness of mind." Here is nourishment for that hope.

SEISMIC SEA WAVES: TSUNAMIS, by T. S. Murty. Bulletin 198 of the Fisheries Research Board of Canada. Available in the U.S. from Unipub, Box 433, Murray Hill Station, New York, N.Y. 10016 ($18.50). The broad Pacific often reverberates with great surface waves. They rattle from shore to shore, slowly decaying, with some 1.7 reflections a day; on the high seas the waves measure a foot or two at the crest, and they spread out from the source at jet-aircraft speed, some 750 kilometers per hour. The isolated mid-ocean islands mark only a gentle swell on their tide records for half an hour or so, but on the continental shores the waves can become terrible: a wall of water 20 meters high, capable of sweeping away a heavy locomotive and the rails on which it rests. The Minoans suffered heavy losses to fleet and state when the violent eruption of the volcano of Thera flooded their ports; in Voltaire's time seismic sea waves wrought such destruction in Lisbon as to confound reason; the explosion of Krakatoa sent waves high over the surrounding coasts, carrying away tens of thousands of people.

Most large, shallow earthquakes under the ocean floor are on the Pacific ring of fire, so that the tsunami rightly bears a Japanese name, which means harbor wave. About 10 percent of the energy of the right kind of earthquake is radiated away in the tsunamis it initiates, presumably by a sudden large-scale motion of the sea floor. Nuclear explosions, volcanic eruptions and landslides can also generate the waves. The postwar years have seen a small oceanographic industry of tsunami studies, motivated pretty surely as much by the temptation to apply sophisticated wave physics to such grand-scale phenomena as by the obvious need to set up a warning system that works.

T. S. Murty is an Ottawa physical oceanographer. His strong mathematical interests are clear in the theoretical chapters of this comprehensive work, but he includes many regional data and many details of the various national and international warning networks, the largest of which is the Tsunami Warning System based in Honolulu. Eleven Pacific nations share in the growing scheme. It includes about 80 tidal or seismic stations spanning the entire vast ocean, from Cape Horn to the Aleutians and from New Zealand to Tucson. Through a comprehensive communications network the stations alert the authorities to any event that is a possible source and then seek by the hour

BERLINETTA

A NEW WAY TO TAKE YOUR PULSE.

The New Camaro Berlinetta offers a superb balance between the qualities of a sporty road car and the comfort of a fine touring car.

Beneath its classic Camaro styling, Berlinetta also offers a wealth of engineering advances. For example, special insulation surrounds you to help reduce engine, wind, and road noise. A new suspension system, unique to Berlinetta, gives you a smooth, responsive ride with a road-hugging pleasure you would expect from a car that wears the Camaro name.

And if its roadability and ride are special, Berlinetta's interior is equally so. There you'll find custom-tailored cloths in fine cuts and colors. Freshly designed front bucket seats carry you in comfort. Berlinetta's instrument panel—including tachometer, voltmeter, temperature gage and electric clock—is newly designed for quick, easy reading.

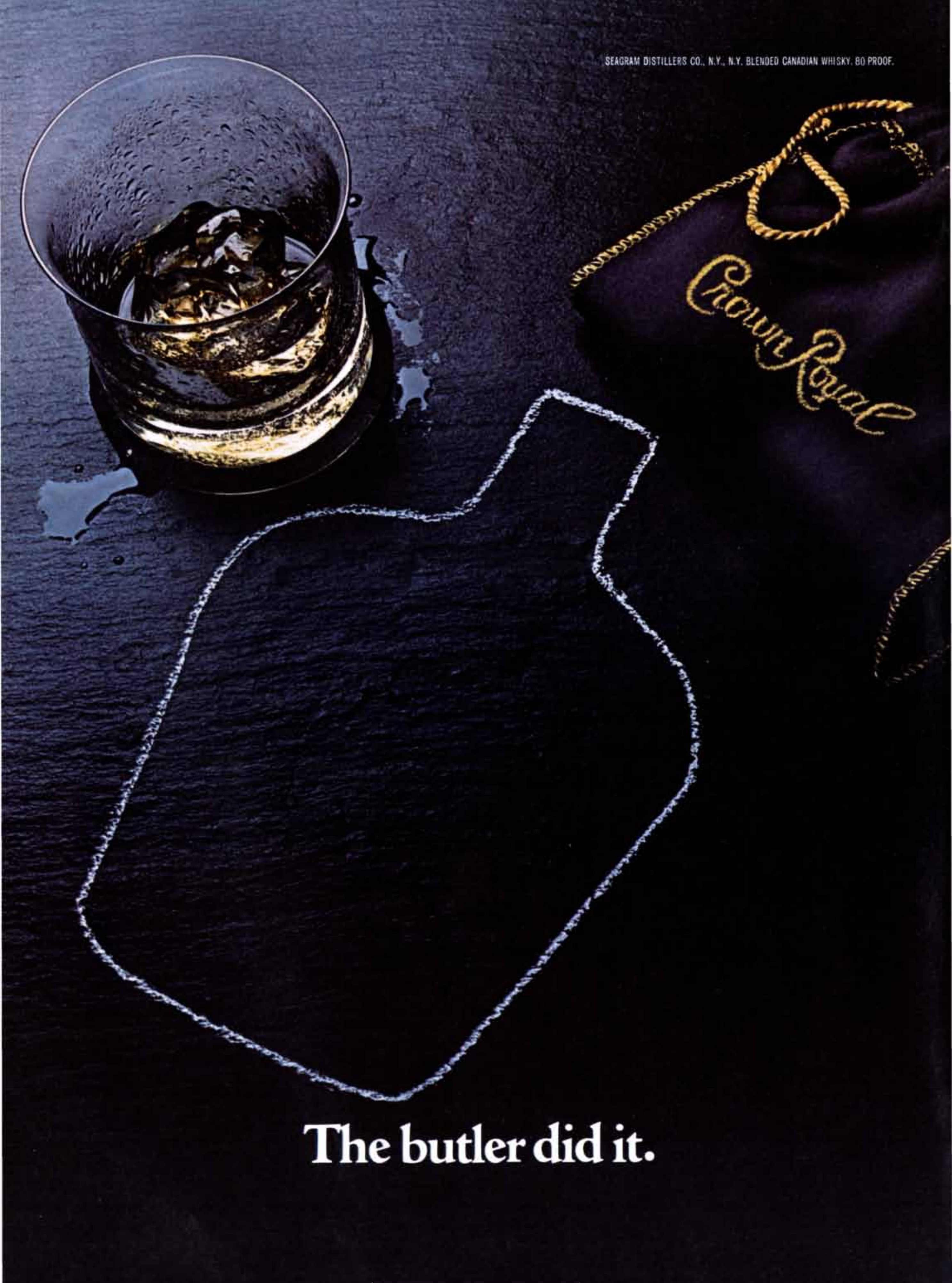
Outside, there's a special Berlinetta grille treatment, dual pin striping, blackened rocker panels, color-keyed custom-styled wheels, sport mirrors, and distinctive Berlinetta nameplates, front, side, and rear.

So, for a new way to take your pulse, buy or lease a new Camaro Berlinetta. Then go out and hug a road—in comfort and style.

Aluminum wheels shown are available at extra cost.
Available white-lettered tires are supplied by various manufacturers.

CAMARO. THE HUGGER.

Chevrolet


SEAGRAM DISTILLERS CO., N.Y., N.Y. BLENDED CANADIAN WHISKY. 80 PROOF.
Crown Royal
The butler did it.

to sharpen predictions for the strength, position and time of wave arrivals. There are national systems in Japan, the U.S.S.R., Tahiti and other places at risk. One official reminder is worth citing: "Never go down to the beach to watch for a tsunami; when you see the wave you are too close to escape it."

This large-amplitude gravity wave has taxed the best physics and has by no means been reduced to routine understanding. The energy in the waves can be estimated by a number of methods based on amplitude, correlations of period with seismic-wave energy and vertical displacement of the source. There is usually either a large tsunami (the quake was well coupled to the sea) or none at all. Propagation times, the effect of refraction and reflection from coasts and islands, trapping phenomena, resonances and many other details have been studied by a battery of methods from analytical to digital, from careful simulation to idealized model, through a variety of ambitious laboratory experiments.

Three disparate measures are plainly relevant: the height of the wave, the depth of the water and the length of the wave. Dispersion in both amplitude and phase can become important in the various domains of approximation, and the tricky solitary-wave solutions, for which the distortions tend to cancel, are of high interest. The dangerous piling up in shallow waters is the most difficult part of the intellectual problem as well as the part that has the greatest practical importance. Clear transoceanic reflections, showing up as delayed second arrivals of the wave, are not at all marked, although they may occur. A generalized absorption, probably rather localized in origin, rules the decay of the energy, some of which can even leak out through the narrow passages from ocean to ocean. The tendency for preferred directions of energy radiation is often remarkable: of two island stations at the same distance over clear ocean paths but 45 degrees apart in azimuth, one has registered a 16-meter wave and the other no wave at all! The effect is held to be a result of the bottom topography right at the source. Other events have been quite isotropic.

All of this is set out by Murty through critical citations of scientific papers from a large literature. The work is generally authoritative, although it falters here and there in small matters such as explosion units or in confusing notations in related equations. Noticeable secondary effects of these big water waves on the atmosphere are alleged, but the point seems vexed. There is a reasonable hope that small acoustical disturbances of the ionosphere might become useful elements of a novel system for the remote detection of tsunamis, with some relevant experience already at hand. The book badly misses an index; even one for authors would have made the long bibliography much more useful. Perhaps in exchange the author and publisher have bestowed an unusual bonus: stored on the inside cover is a microfiche offering a 40-page summary of the seismological background to the main topic. The book is agreeably designed by Christine Rusk; the watercolor on the dust jacket is a small prize.

MURMURS OF EARTH: THE VOYAGER INTERSTELLAR RECORD, by Carl Sagan, F. D. Drake, Ann Druyan, Timothy Ferris, Jon Lomberg and Linda Salzman Sagan. Random House ($15). A couple of months ago the well-aimed *Voyager 1* spacecraft skimmed past Io and Ganymede, sped up as it felt the strong pull of Jupiter himself and was then flung out for a second bounce off the Saturn system before the end of 1980. After that the spacecraft, its work done, losing its voice as its radioactive power sources decay year by year, will permanently escape from solar gravitational control to loop the galaxy forever—or at least until the insensible rasp of the interstellar gas erodes it into a swarm of smooth fragments.

Securely mounted on the side of the spacecraft (and also on its fellow Voyager, due at Jupiter this summer) are two 12-inch phonograph records within a golden anodized-aluminum cover. They are not the usual vinyl pressings; a hydrogen-bonded plastic is too unstable for the aeons. They are electroplated copper positives like the ones that yield the molds from which the usual discs are pressed. The total weight is about five pounds, including the playing stylus and cartridge generously supplied for the beings who someday might pick up this piece of galactic flotsam and wonder at its nature.

Murmurs of Earth reproduces the recorded contents. There are about 120 pictures (recorded in the audio spectrum) representing the earth and its living cargo; voiced greetings in 55 main languages of the human species; a selection of "sounds of Earth," including a pulsar, the noises of flintworking and a kiss (recorded under strict NASA orders to "keep it heterosexual... it felt and sounded fine"); a few pages of ceremonial and functional text, and about 90 minutes of the music of humankind from every continent and most of human history.

The lively and intimate text by the adroit authors listed above—astronomers, writers and artists—chronicles their breathless, serious, successful effort (too little time, too little money, plenty of bureaucratic inertia) to select and record this hopeful, quixotic message in a galactic bottle. The whole story amounts to a self-examination that is droll, challenging, evocative, informative and in the end absorbing.

How might we give some account of

ourselves on a couple of discs that would augment the material story provided by the spacecraft itself for interstellar archaeologists, if they should exist and should happen on the tiny waif? A date? That is easy: a sample of pure uranium 238 will mark a pretty good epoch by the content of its decay products. A place? The pulsar directions help, and a photograph of the galaxy in Andromeda is orienting and also fixes the handedness of the pictures. The time unit given by the radio-frequency line of hydrogen is implied, and binary multiples of that time define the playback rate well enough. A simple circular form provides a means of deciphering the visual raster system and fixes the ratio of the horizontal scale to the vertical.

The music puts our best foot forward. Among others, Bach, Mozart, Beethoven and Blind Willie Johnson sample the post-Renaissance period; the silken strings of the *ch'in* are heard in "Flowing Streams," a Chinese classical piece perhaps older than the Han; a solo morning raga, peerlessly sung by Surshri Kesar Bai Kerkar with a 14/4 drumbeat and a drone accompaniment, recorded when she was over 70, admonishes a young one not to go alone during Holi, when people throw saffron at one another.

How such things are in fact realized, in this world of prudery, copyrights and regulations, of engineering standards, deadlines and disputes, of the separation of constitutional powers, the sovereignty of each UN state and the dignity of its official language, is the burden of the narrative. Voyager is well described. You can read a White House message and a list of the members of the relevant committees of Congress. (They get equal time, even if they are not as interesting as the whale voices or the Balinese dancer.) Somehow the most endearing of all the warm simplicities this task evoked from the friendly volunteer participants seems the greeting spoken in Amoy, a language of the Chinese coast: "Friends of space, how are you all? Have you eaten yet? Come visit us if you have time."

The recording expertise was provided by CBS Records, and some key advice came from RCA Records. Structural conflict between the not-for-profit principles of NASA and the very-much-for-profit organizations of the American record market has so far blocked the public release of a vinyl pressing containing just what is out there now in the orbit of Jupiter. Perhaps this problem too can someday be resolved, before the Voyager recordings degrade by micropitting (a mean time of a billion years) or possibly even before they approach the *M*-class dwarf star named AC + 79 3888, the nearby star in the Little Bear to which, as far as can be foretold, the craft will come closest: within two light-years after 50,000 years of drift.

Treasure Detector

A new computerized metal detector that actually selects treasure from trash may uncover America's long lost relics and precious metals.

The new breakthrough in metal detectors makes finding treasure much easier.

There's a lot of treasure right under our feet. There's also a lot of garbage.

And the problem with most metal detectors is that they're dumb. They can't tell treasure from trash.

The new Techna metal detector is different. It has both a sensing system and a brain that can tell the difference between a foil gum wrapper and a coin–between a bottle cap and a diamond ring.

The new breakthrough was made possible by the use of a "discriminator IC"–a computer-type integrated circuit that can compare the ferrous and non-ferrous relationships that distinguish treasure from trash.

There are discriminator-type metal detectors now on the market, but they cost between $170 and $400. And no matter what price you pay, the detector is usually difficult to operate.

The new Techna Discriminator represents several breakthroughs. First, it is inexpensive–only $69.95. Secondly, it uses a new (patent pending) phase compensation system of metal detection, whereas other discriminators use either the off-resonance or inverse discrimination principle.

This system utilizes a microprocessor circuit which replaces the conventional electronics, mode switch, and multiple tuners that added to the cost and weight of a discriminator unit.

Finally, the Discriminator is very easy to operate and understand. You simply set it to sense treasure, trash, or both and it automatically tunes itself and starts operating. Whenever you scan treasure, a loud speaker will emit a sound and you start digging.

DETECTORS ARE BIG BUSINESS

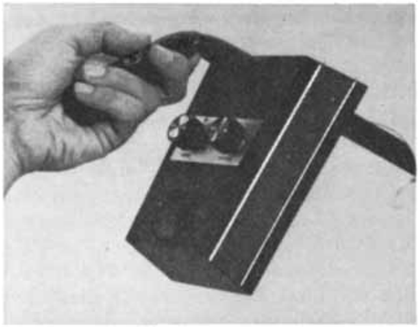
Metal detectors are big business. When we investigated the field, we discovered an entire new sport–treasure hunting. Treasure clubs exist and conduct contests. There's a national magazine and an association, and hundreds of thousands of units are in use every day.

Treasure hunting doesn't just mean looking for buried pirate chests. There's great interest now in discovering articles of historical significance such as old coins, military buttons, and old pistols.

Long ago when people distrusted banks, they buried their valuables somewhere on their property. If they died suddenly or became senile, their treasures were lost forever. Many treasure hunters are now visiting ghost towns or going through older sections of cities looking for both historic and valuable articles.

WORLD WAR II STARTED IT

Metal detectors first saw extensive use during World War II. Back then, they were called mine detectors and were used to uncover enemy land mines. They were heavy, often weighing hundreds of pounds, and had to be carried on the backs of soldiers along with separate and heavy power supplies.



The Techna Discriminator is light and easy to operate with only two controls to adjust.

The new Techna Discriminator is light. It weighs only 2½ pounds and is powered by two readily-available 9-volt batteries. As you glide the sensing head over the ground, the unit remains silent until it uncovers a precious metal or whatever type metal you are searching for. An electronic sound is emitted. Then just dig in the area of the sound.

If you already own an expensive metal detector, you know that most of your "discoveries" turn out to be bottle caps or gum wrappers. With the Techna, you discover just what's worth digging up. While others are digging up bottle caps, you're covering more ground faster and are more likely to discover something worthwhile.

BREAKTHROUGH PRICE

The fully computerized Techna Discriminator is available from JS&A for only $69.95 complete with batteries and all components. We suggest you order one just to try it out. Try it in your back yard. Take it to a sandy beach where many coins and jewelry are lost. See how the system can tell the difference between treasure and trash, and then after you have discovered the fun of treasure hunting and how advanced this new product really is, decide whether or not you wish to keep it.

If you feel the Techna Discriminator does not meet all your expectations for any reason, we will gladly accept the return of your unit within our 30 day trial period and even refund your $3.50 postage and handling. If you decide to keep your unit, you will own the world's most advanced metal detector. No competitive model even comes close.

Techna is America's largest manufacturer of metal detectors in the United States, and JS&A is America's largest single source of space-age products–further assurance that your modest investment is well protected.

Each Techna detector is backed by a solid one-year parts and labor limited warranty. We doubt if you'll ever have a problem with the unit because of its solid-state construction, but if service is ever indeed required, Techna's service-by-mail center will fix your unit and have it back to you quickly.

To order your Techna Discriminator detector, send your check for **$69.95** plus $3.50 for postage and handling (Illinois residents please add 5% sales tax) to the address shown below. Credit card buyers may call our toll-free number below. We will promptly send you your Techna detector with batteries, 90-day limited warranty, and instructions.

Why not join the legion of treasure hunters worldwide with the world's most advanced space-age metal detector. Order the Techna Discriminator metal detector at no obligation, today.

JS&A® NATIONAL SALES GROUP

Dept.SA One JS&A Plaza
Northbrook, Ill. 60062 (312) 564-7000
Call TOLL-FREE 800 323-6400
In Illinois Call .... (312) 564-7000

Established 1845 SCIENTIFIC AMERICAN May 1979 Volume 240 Number 5

Standard and Daylight-saving Time

The four standard-time zones of the U.S. are the result of a long effort to achieve a compromise among competing demands. Congress is still confronted with the task of doing the same for daylight-saving time

by Ian R. Bartky and Elizabeth Harrison

Everyone is aware of the changes in the time of sunrise and sunset as the seasons advance. Only if one were near the eastern edge of a time zone on one day and near the western edge on the next would one notice another change: the sunrise and sunset would be distinctly later on the second day than they were on the first. This change results from the nation's system of civil time. Most of the population also see an abrupt change in the time of the sunrise and sunset when daylight-saving time begins and ends. These phenomena were even more apparent in the U.S. in 1974 and 1975, when in response to the Arab oil embargo Congress put the country on extended daylight-saving time. Parents found that their children had to leave for school in the dark, and many complained that the schedule was unsafe. The episode focused attention on the nation's time system and in particular on the question of what is the best period of the year for daylight-saving time to be in force, a subject that is currently under review in Congress.

Until as recently as 1883 the system of time in the U.S. was quite different from the one now in effect. Local time was precisely that: noon by the clock was the time when the sun reached the observer's meridian. (We refer here to the sun's mean or average position, which is sometimes ahead of and sometimes behind its real one.) As a result the time differed from city to city; when it was noon in Washington, it was 12:24 P.M. in Boston and 11:43 A.M. in Savannah. The local clocks along their particular meridians were related to the Naval Observatory in Washington. The meridian there was the reference for the time in the U.S. Similarly, the observatory in Paris fixed the time in France, the Greenwich Observatory the time in Britain, and so on for other countries.

By the middle of the 19th century the growth of the railroads (and to a lesser extent of the telegraph system) led to a reconsideration of the time system. The expansion of the rail lines had begun to bring the local time of one city into regions that had the local time of another city, giving rise to much confusion and annoyance. The solution was to adopt a system in which the clocks in large regions would read the same. Such a system had been suggested in 1828 by the astronomer Sir John Herschel, and by the middle of the century Greenwich mean time had been widely adopted in England, Scotland and Wales.

In the U.S. it was clear to most people that a single time standard would be unworkable because of the country's great expanse from east to west. It was therefore proposed (apparently first in 1869 by C. F. Dowd, the principal of a seminary for girls in New York) that the country be divided into zones, each with its own standard of time. The questions raised by this proposal included how many zones should be established, how large a zone should be and what meridian should control the time in each zone. All these questions gave rise to debate, since the answers were necessarily arbitrary.

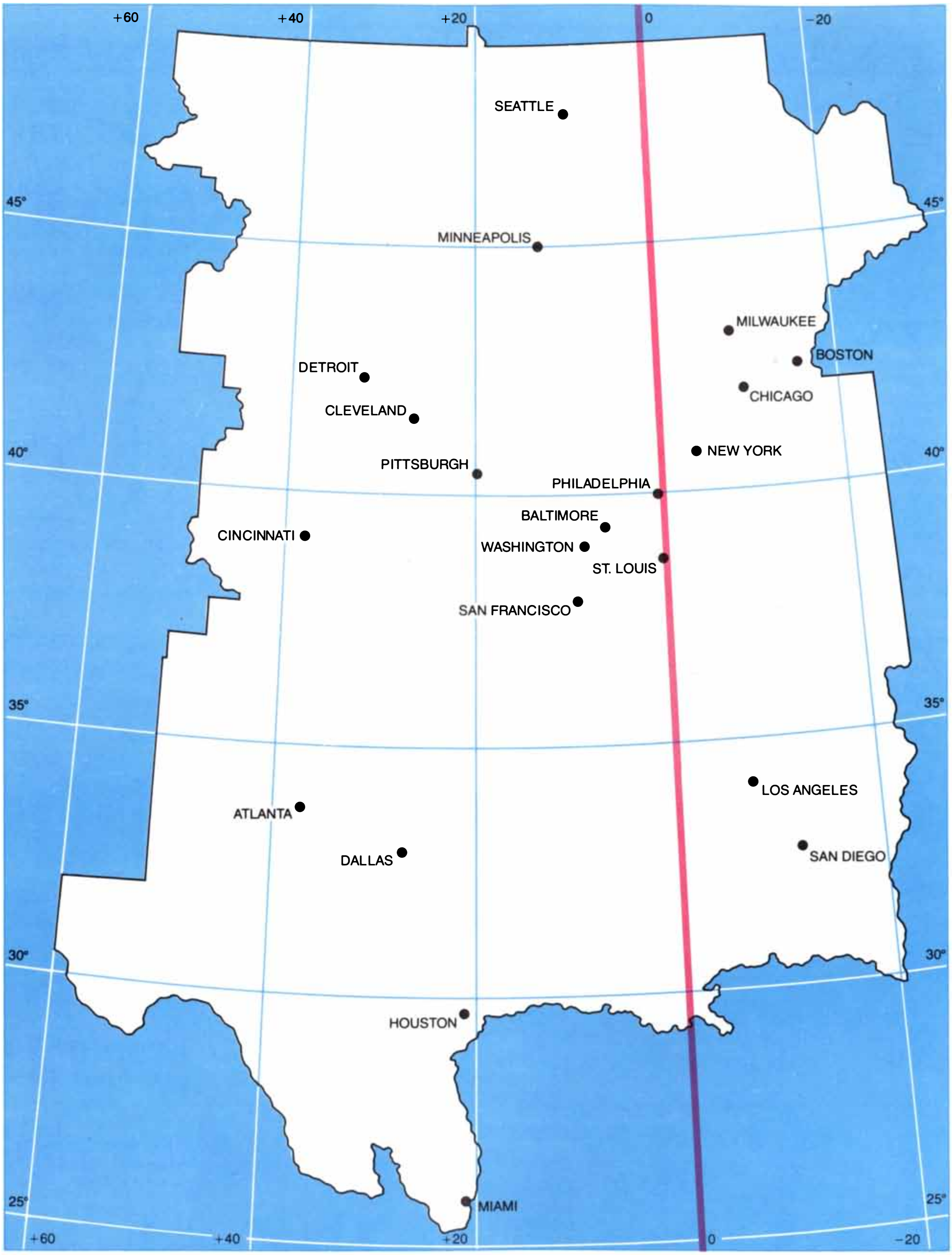
In 1878 Sandford Fleming, who was the chief engineer for the Canadian government, suggested that such a system of zones be made worldwide. There would be 24 meridians, each 15 degrees (one hour) apart in longitude, starting from Greenwich. Each zone would keep the local time of the meridian that bisected it.

Credit must also be given to W. F. Allen, editor of the *Travelers' Official Guide* and secretary of the General Time Convention, which was a North American railroad organization concerned with scheduling and safety. Allen showed how the zone idea could be adapted to correspond to the divisional arrangement of the railroads. His argument was persuasive. On November 18, 1883, most of the railroads in the U.S. and Canada began to operate on Standard Railway Time, reducing the number of railroad times from at least 56 to four.

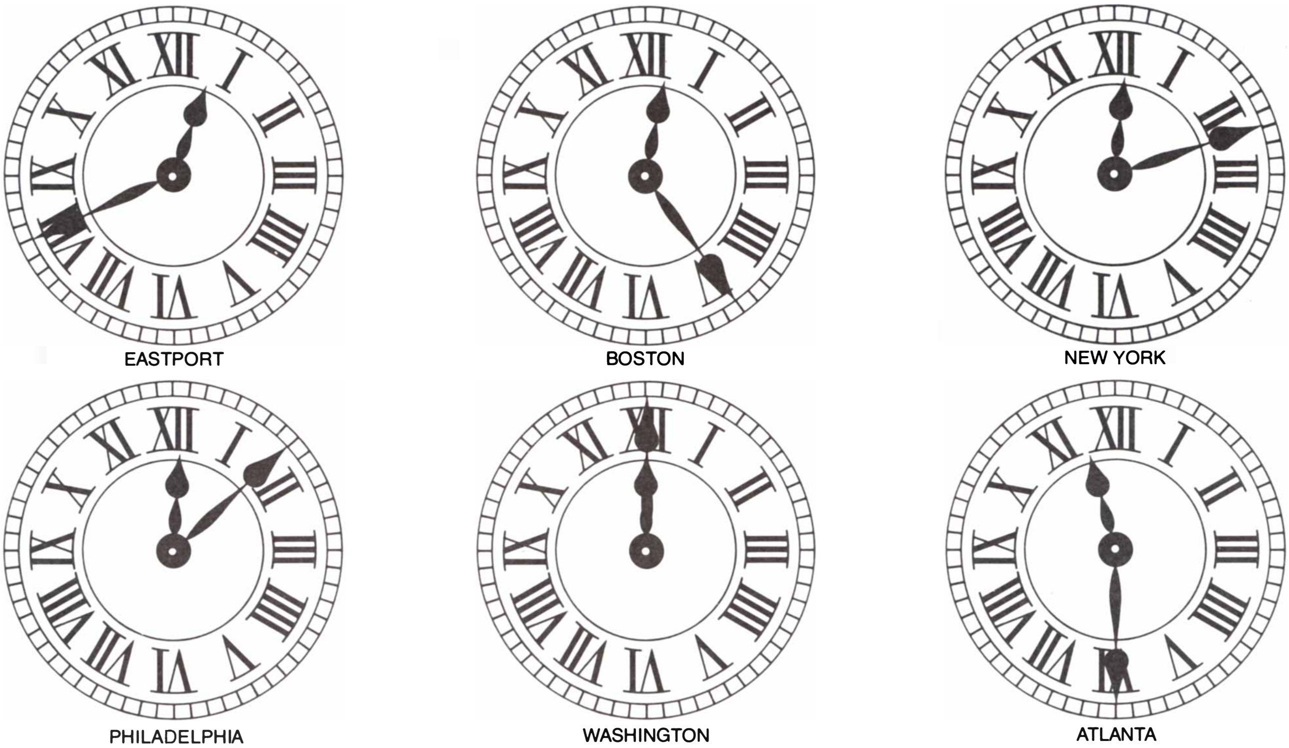
The four time zones were designated Eastern, Central, Mountain and Pacific. The time at the controlling meridians was respectively exactly five, six, seven and eight hours later than the mean solar time at the Greenwich meridian. Within five months 80 of the 100 principal cities in the U.S. had adopted the system; 70 of them did so immediately.

As one might expect, this voluntary system met opposition. Some people argued that local time was God's time. Others complained about the zone they were in and continued keeping the old local time. When the system actually went into effect on that Sunday in the fall of 1883, however, the only discernible effects were (1) two "noons" in New York and other places as clocks were adjusted to the new time standard from the old local time and (2) significant shifts for some people in the clock times at which the sun rose and set. Within a few years the zonal system became virtually worldwide, and the meridian at Greenwich became the reference or prime meridian for the entire standard-time system.

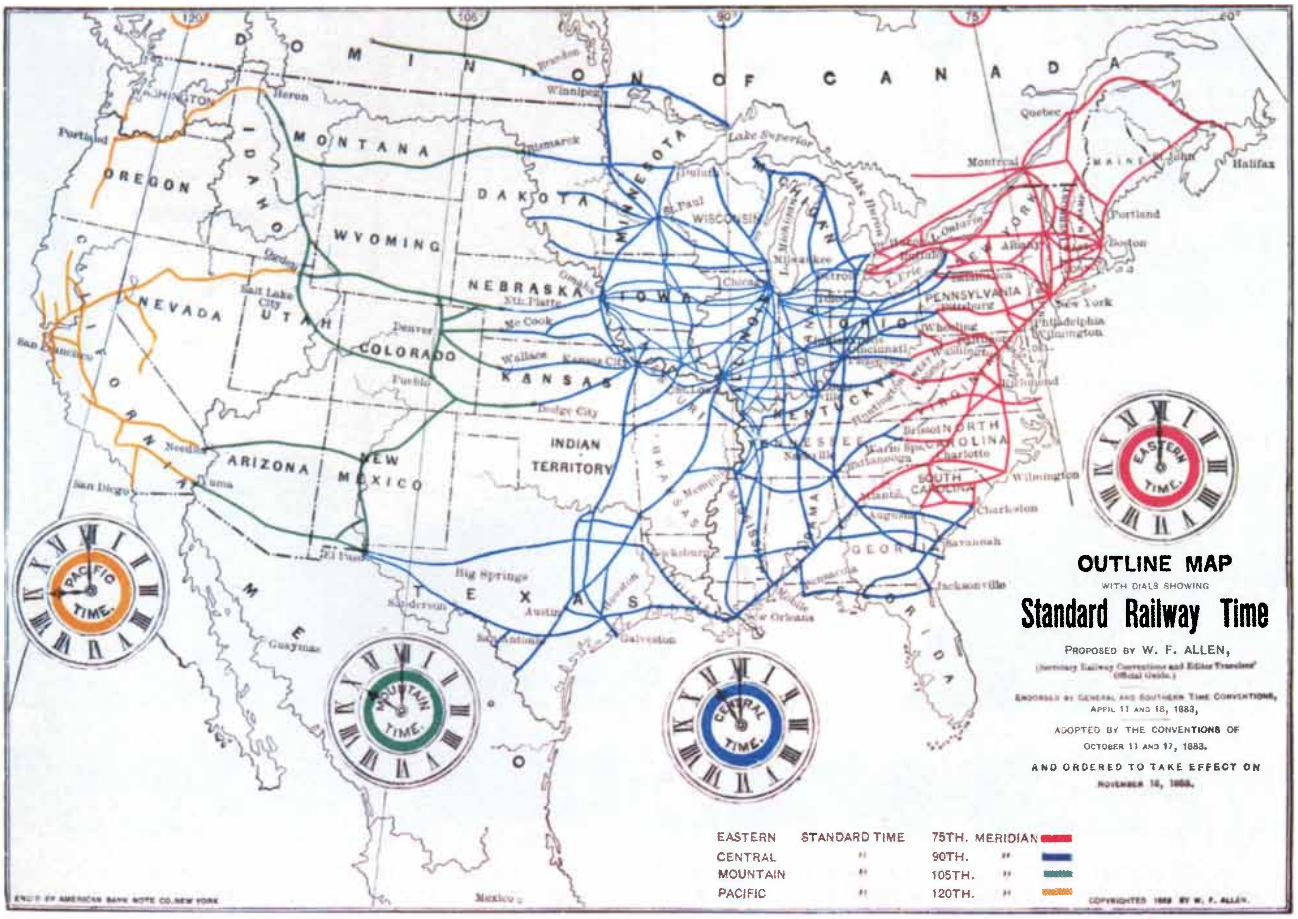
The next major change involving clocks was daylight-saving time. It would not be an issue if the rotational axis of the earth were perpendicular to the plane of the earth's orbit around the sun. Then the daily amount of daylight and darkness would be equal everywhere all year. With the axis inclined



SUPERPOSED TIME ZONES indicate how people in the 48 coterminous states of the U.S. are affected by daylight-saving time. On this map the four time zones have been superposed and each city representing a metropolitan agglomeration of 1.5 million people or more has been plotted according to its location in its own time zone. The colored line represents a standard meridian of time and the vertical lines to the left and right of it show how much (in minutes) the local mean time there differs from the time at the meridian. The boundaries at the left and right of the map represent the eastern and western extremes of the time zones. Many people in the U.S. live in the western portions of the various time zones and are therefore most affected by late sunrises that often result from daylight-saving time.



CLOCK VARIATIONS that existed in the U.S. before 1883, when the first system of standard time was introduced by the railroads, are shown for six cities in what later became the Eastern time zone. Noon in local time was when the mean sun reached the local meridian.



STANDARD TIME instituted in the U.S. by the railroads in 1883 is shown on this map published at the time. The standard meridians were (and are) 75, 90, 105 and 120 degrees of longitude west from Greenwich; 15 degrees is equal to one hour in mean solar time. Railroads introduced the system because of confusion and safety hazards caused by the many (at least 56) railway times in effect previously.

from the plane of the ecliptic, however, the summer hemisphere receives considerably more daylight than the winter one, to the point where at the poles summer has no night and winter has no day. Even with the tilt the question of daylight-saving time would not arise if everyone lived near the Equator (where day and night are equal throughout the year) or near the poles (where the range of daylight is from zero to 24 hours and a one-hour shift in clock time would be superfluous). The 48 coterminous states of the U.S. lie between 25 and 49 degrees of latitude, however, so that the seasonal variations in daylight and darkness are sufficient to cause dislocations in the relation between the hours of daylight and the hours when people are normally active.

The principle of daylight-saving time is simple. During part of the year the sun rises before most people do. If the clock is advanced, sunrise by the clock will be later (as will sunset). This clock shift does not increase the amount of daylight, but it "saves" daylight in the sense that it reduces the number of daylight hours when people are normally asleep and increases the number when they are normally awake.

The idea of daylight-saving time originated with William Willett, who was a British builder and also a fellow of the Royal Astronomical Society. In 1907 he began a campaign for the adoption of daylight-saving time. From the outset the anticipated benefits were stated as additional time for recreation, less crime and a reduction in the need for artificial light. Opposition arose at the same time from such groups as farmers and from parents who were concerned about the sleeping habits of their children. Willett did not live to see the introduction of summer daylight-saving time, first by Germany in 1916 and soon afterward by a number of other European countries. World War I provided the impetus; the expected saving of electricity and coal through the adoption of daylight-saving time was seen as an aid to the war effort.

The U.S. had a more difficult problem in World War I: standard time and daylight-saving time were introduced simultaneously. Summer daylight-saving time began on March 31, 1918, only 11 days after the law establishing both of them went into effect. National daylight-saving time lasted for only two summers. Congress abolished it in 1919, partly because the war was over and partly because of public complaints. From 1920 on daylight-saving time was a matter of local option, and it was not widely adopted.

Standard time creates clock-time discontinuities across three roughly **north**-south lines in the U.S. When it is **noon on** the eastern side of a zone **boundary**, it is 11:00 A.M. on the western side. The location of the boundary lines becomes an important matter for the people living near them.

In the 1918 law Congress assigned the resolution of the boundary problem to the Interstate Commerce Commission. After six months of study the ICC decided that new zones had to be established. The railroad time zones were so different from the timekeeping of the cities that it was impossible to define them "even approximately." Some cities had a choice of two times, because their railroad time had changed; other cities were keeping the time of another "zone." Some railroads were keeping time standards different from those of other railroads using the same tracks.

In the 1918 law Congress gave the ICC guidance: follow the principle of the standard-time system and define the zones "having regard for the convenience of commerce and the existing junction points and division points" of the railroads. The ICC understood that the public was strongly affected by clock changes and took "the convenience of commerce" in its broadest sense. The zone boundaries were set close to the meridians that lie halfway between the standard (time-controlling) meridians but somewhat to the west to "secure the greatest amount of daylight for the active hours." The zones were made compact and their boundaries were drawn to avoid heavily populated areas. Natural commercial regions were kept on the same clock time and the clocks of the cities and states were given more weight than those of the railroads.

The zone boundaries could be modified by the ICC and were modified somewhat even before January 1, 1919, the effective date of the ICC's initial decision. During the next 48 years the ICC approved 35 boundary changes and rejected seven. All 42 of the changes extended boundaries westward, reflecting the orientation of commerce eastward in many communities at the time.

The period from January 1, 1919, to March 28, 1920, is one of only two periods when everyone in the U.S. kept the same clock time. Standard time began on January 1, 1919; everyone kept it until daylight-saving time began that spring and was observed by everyone for seven months, and then everyone kept standard time for five months until Massachusetts and New York City instituted local daylight-saving time in the spring of 1920.

Daylight-saving time was gradually adopted in other states, and the period of its observance began to vary. By 1941 it was observed in large regions of the Eastern time zone and in a few areas of the Central zone. No region west of the Mississippi River observed it, however, and we can find no region 30 minutes or more west of the standard meridian in the Eastern zone that observed it for any length of time.

Year-round daylight-saving time, or "war time," as it was named by President Roosevelt, began on February 9, 1942. Its purpose was to help dissipate evening peaks in the consumption of electric power, which at that time were higher in the winter than they were in the summer. This obligatory advanced time put an end to local-option daylight-saving time, but it also ended the observance of standard time in some places. The ICC ruled that it did not have the authority to shift zone boundaries merely to reduce the effects of advanced time for regions at the western sides of zones, and so a number of places simply shifted to another time standard. The failure to observe the standard time of the zone in which a community was located persisted in some regions until 1966.

At the end of the war the system of war time was repealed, effective on the last Sunday in September of 1945. The following spring various states and municipalities reinstituted summer daylight-saving time. The resulting confusion became worse in 1954 when the New England states extended daylight-saving time through October, whereas most other places remained on the more general period from May through September.

In 1962 several groups in the transportation industry began a campaign to resolve the lack of uniformity in the observance of time through new Federal legislation. Their aims were to have standard time apply to all civil activities in the nation and to provide uniform dates for the beginning and end of daylight-saving time. As part of their argument the campaigners cited a 35-mile bus trip from Steubenville, Ohio, to Moundsville, W.Va., that required seven time changes; a Government office building in Minneapolis–St. Paul where the time was different on different floors for part of the daylight-saving period, and a statement by the Department of the Navy that "it is in general not possible to tell with certainty what time is kept in a particular locality on a given date."

In 1966 Congress resolved these problems by occupying the area of time completely, expressly superseding state and local laws. Standard time was made mandatory, and the one-hour advancement of standard time had to be observed from the last Sunday in April through the last Saturday in October. A state could exempt itself from daylight-saving time and (starting in 1972) a state split by a zone boundary could exempt part of itself, provided the exempted area was entirely in one time zone. Zone boundaries could still be changed, and the procedure for changing them was simplified. Penalties were established for failure to observe the time system. Responsibility for administering the law was assigned to the newly created Department of Transportation, and the Secretary of Transportation was directed "to foster and promote widespread and uniform adoption and observance of the same standard of time within and throughout each zone."

The law became effective on April 1, 1967, opening the second period of time uniformity in the U.S. Uniformity lasted barely a year, until April 20, 1968, when Arizona became the first state to exempt itself from advanced time. In the first seven years 12 boundary changes were made, reflecting the struggle in certain states to adjust to the dual problem of a required system and daylight-saving time. The record shows more exemptions by states, the first eastward boundary movements (because of dissatisfaction with late sunrises in the western parts of time zones) and the evolution of the present system: the standard-time zones as they stand now, with Arizona and the part of Indiana that is in the Eastern time zone exempting themselves from daylight-saving time.

The next period of adjustments to the system began when Congress adopted year-round daylight-saving time for a two-year trial period in response to the Arab oil embargo. The period began on January 6, 1974. It lasted for one season; because of public complaints Congress modified the law to provide for eight months of daylight-saving time in 1975. In 1976 Congress reviewed the results of the preceding two years. Both houses of Congress proposed new seven-month periods of daylight-saving time, but opposition in the House of Representatives prevented the passage of a bill. The U.S. therefore reverted to the six-month period of daylight-saving time that had been established in 1966. Since then the system has remained unchanged.

Part of the reason for extending daylight-saving time in 1974 and 1975 was to explore the question of what advanced-time period would be best for the country. The experiment yielded new perceptions of how people's lives are affected by clock time, particularly by daylight-saving time.

One known effect is that under daylight-saving time a number of amplitude-modulation radio stations that are allowed to operate only after sunrise (the U.S. has more than 2,000 stations in this category) are off the air during part of the morning when people are commuting and children are going to school. Other effects can be postulated. For example, recreational facilities might receive more use, since people have more light in the evening. Effects of this kind are thought to be exceedingly small and therefore hard to assess.

The proponents of the experiment with year-round daylight-saving time had argued that the change might affect the consumption of energy, the incidence of traffic accidents and the safety of schoolchildren. The data that were collected during the experiment on energy and traffic safety show no pronounced effects. On the safety of schoolchildren, however, there is enough evidence of a trend to be disquieting.

The issue is the safety of children who are waiting for school buses or walking to school in the morning. Evening is not a concern, since most school days end by midafternoon and an additional hour of daylight in the evening is not seen as reducing the hazard associated with the return from school. On the other hand, the loss of an hour of light in the morning is seen as increasing the risk to children as they go to school.

In 1974 advanced time began on January 6, during the period when sunrise comes at the latest times. The hour added by daylight-saving time made the earliest sunrise on that day at 7:36 A.M., and at the northwestern edges of the time zones the latest sunrise was at 9:49 A.M. State and local officials in many places resorted to a variety of strategies to mitigate the effects.

Data compiled since the experiment show a statistically significant increase in fatal accidents between 5:30 and 9:30 A.M. in January and February (1974 compared with 1973). Although the increase cannot be interpreted as an effect caused solely by daylight-saving time, one cannot dismiss daylight-saving time as a cause either. In any event it seems probable that these findings, tentative though they are, will prevent the acceptance of daylight-saving time during the late-sunrise months (November through February).

The fact that the experiment with extended daylight-saving time leaves one almost as uncertain about the effects of advanced time as before might lead to the conclusion that the experiment was a failure. It at least had the result, however, of bringing about a general recognition that further experiments would serve no useful purpose, since most of the effects of advanced time are small and difficult to quantify. Moreover, the recent period was the first occasion that any analysis had been made in the U.S. of the effects of different advanced-time periods. It was also the first time that Congress had fairly complete

TIME ZONE							
EASTERN	1.369	25.906	25.629	22.061	26.148	5.451	.03
CENTRAL	1.059	5.497	15.375	15.457	22.167	1.494	
MOUNTAIN	.123	.611	3.228	2.233	2.145	.044	
PACIFIC			.331	11.971	13.363	.419	
TOTAL	2.55	32.01	44.56	51.72	63.82	7.41	.03
PERCENT	1.3	15.8	22	25.6	31.6	3.7	

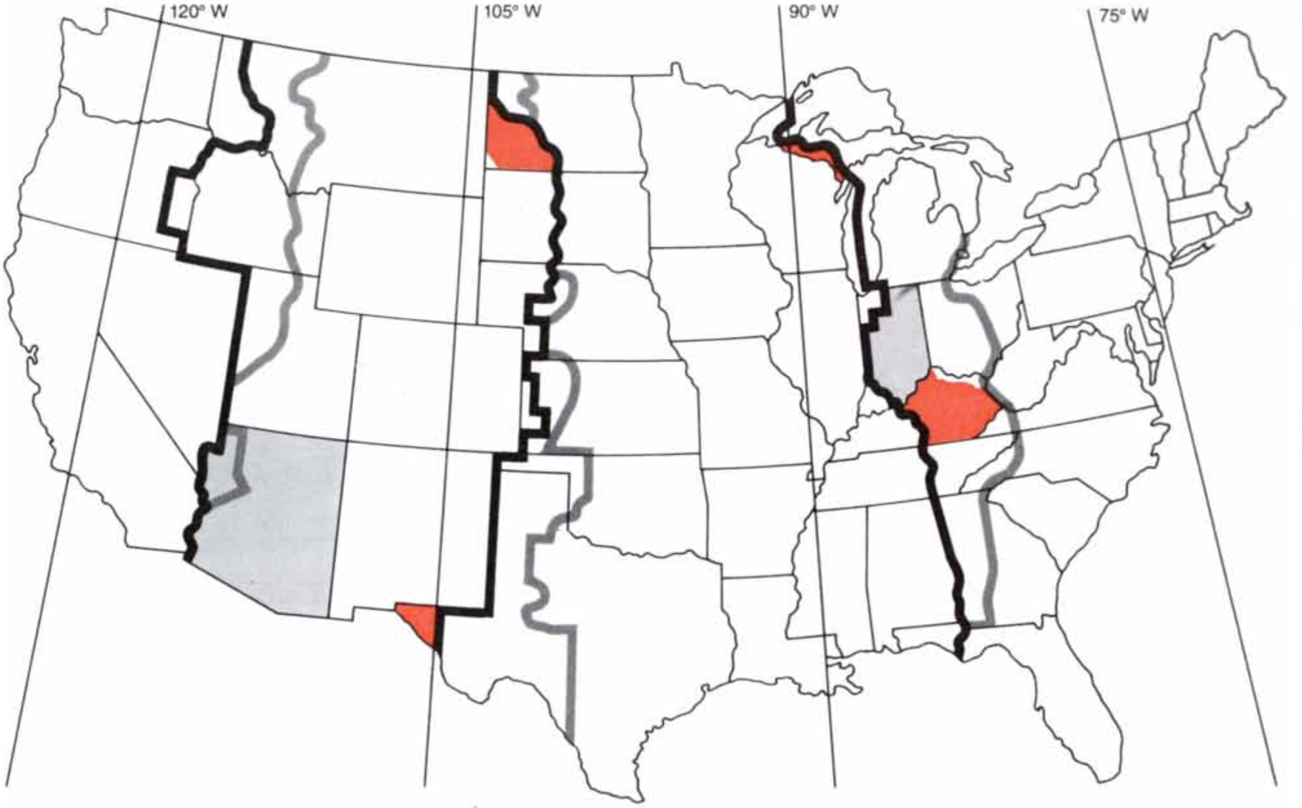
+60 +45 +30 +15 0 −15 −30 −45

MINUTES FROM THE STANDARD MERIDIAN

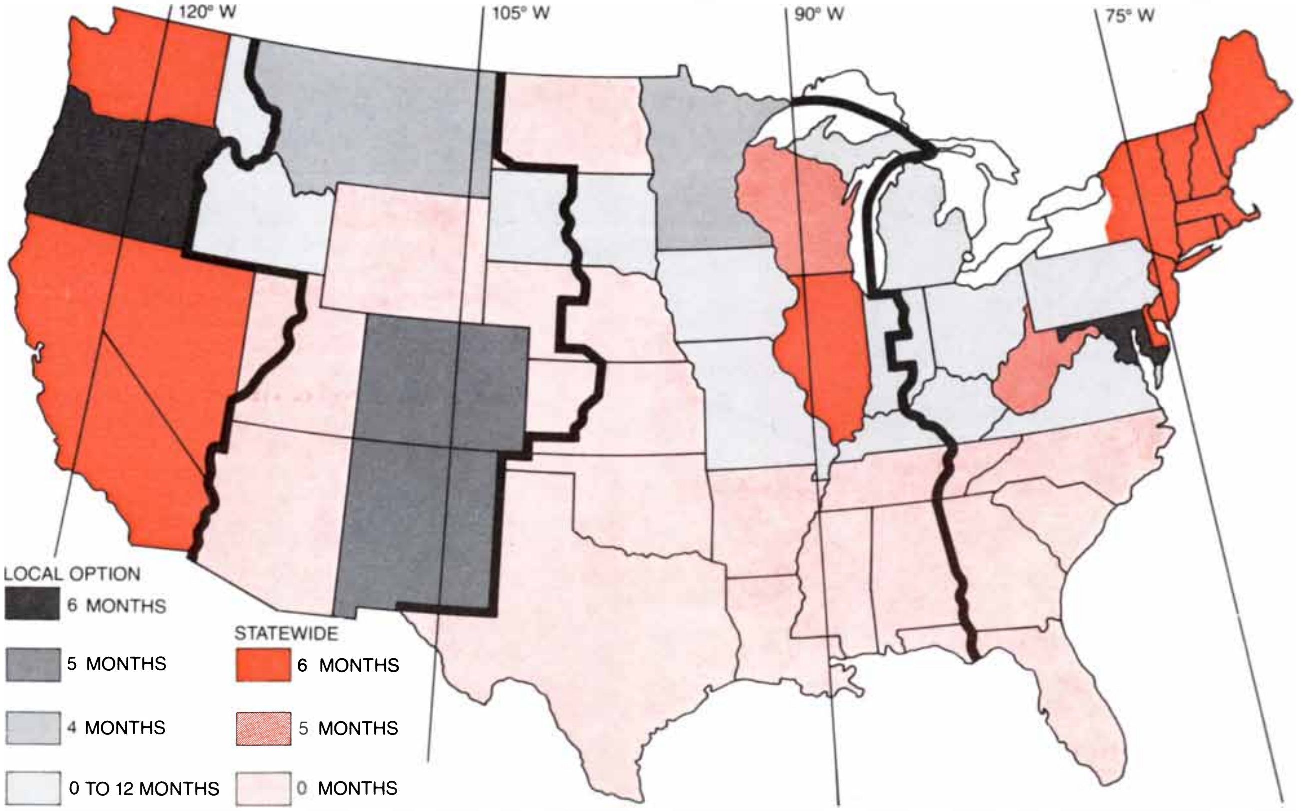
DISTRIBUTION OF POPULATION (in millions) in the four time zones of the U.S. shows that most people live in western parts of the zones and that 17 percent live more than 30 minutes west of a standard meridian, where late sunrises under advanced time evoke complaints.

NO SUNRISE AFTER	PERIOD OF DAYLIGHT SAVING TIME	LENGTH IN MONTHS
9:00	FEBRUARY 18 — NOVEMBER 13	8.8
8:45	FEBRUARY 26 — NOVEMBER 3	8.2
8:30	MARCH 6 — OCTOBER 25	7.7
8:15	MARCH 14 — OCTOBER 15	7.1
8:00	MARCH 23 — OCTOBER 4	6.4
7:45	APRIL 5 — SEPTEMBER 18	5.5
7:30	APRIL 17 — AUGUST 22	4.2
7:15	MAY 3 — JULY 28	2.9
7:00	MAY 27 — JUNE 28	1.1

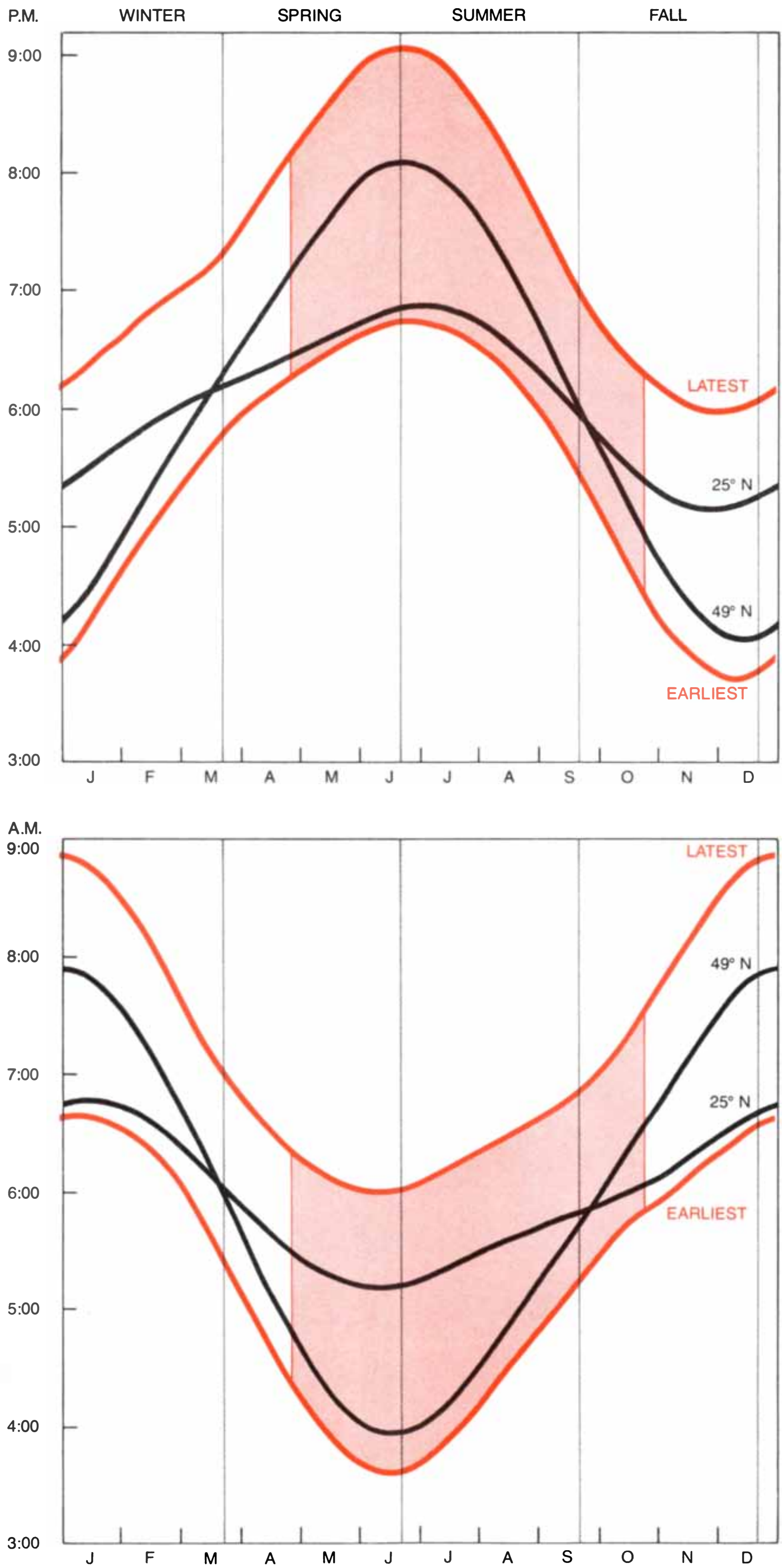
TIME OF SUNRISE for nine periods of daylight-saving time is indicated. The table covers the four time zones of the 48 coterminous states. Lateness of sunrises after October and before March makes public acceptance of advanced time from November through February unlikely.



BOUNDARY CHANGES in the standard-time zones from 1918, when the standard-time system was made official, to the present are indicated. The original boundaries are in gray, the present ones in black. All the 49 changes have been westward except for the four shown in color. The regions depicted in gray, representing about 3 percent of the population, have exempted themselves from daylight-saving time. Any state can exempt itself or, when it is divided by a zone boundary, can exempt a part of itself that is entirely in one zone.



VARIED OBSERVANCE of daylight-saving time in 1964 is shown (with the zone boundaries then in effect). Transportation groups, which were strongly affected by the resulting confusion, urged Congress to adopt a uniform system, particularly for observing daylight-saving time. Congress did so in the Uniform Time Act of 1966, which superseded state and local laws and set the present six-month period for observing daylight-saving time (from the last Sunday in April through last Saturday in October) but allowed exemptions by states.



CLOCK TIMES of sunsets (*top*) and sunrises (*bottom*) in the standard-time zones of the U.S. are charted by the colored curves. The black curves show the clock times at the extreme latitudes along the standard meridians that govern the clock time in the zones. One hour is added to the clock during the six-month observance of daylight-saving time, a period indicated by shading. Variations in the amount of daylight through the year can also be seen from the chart.

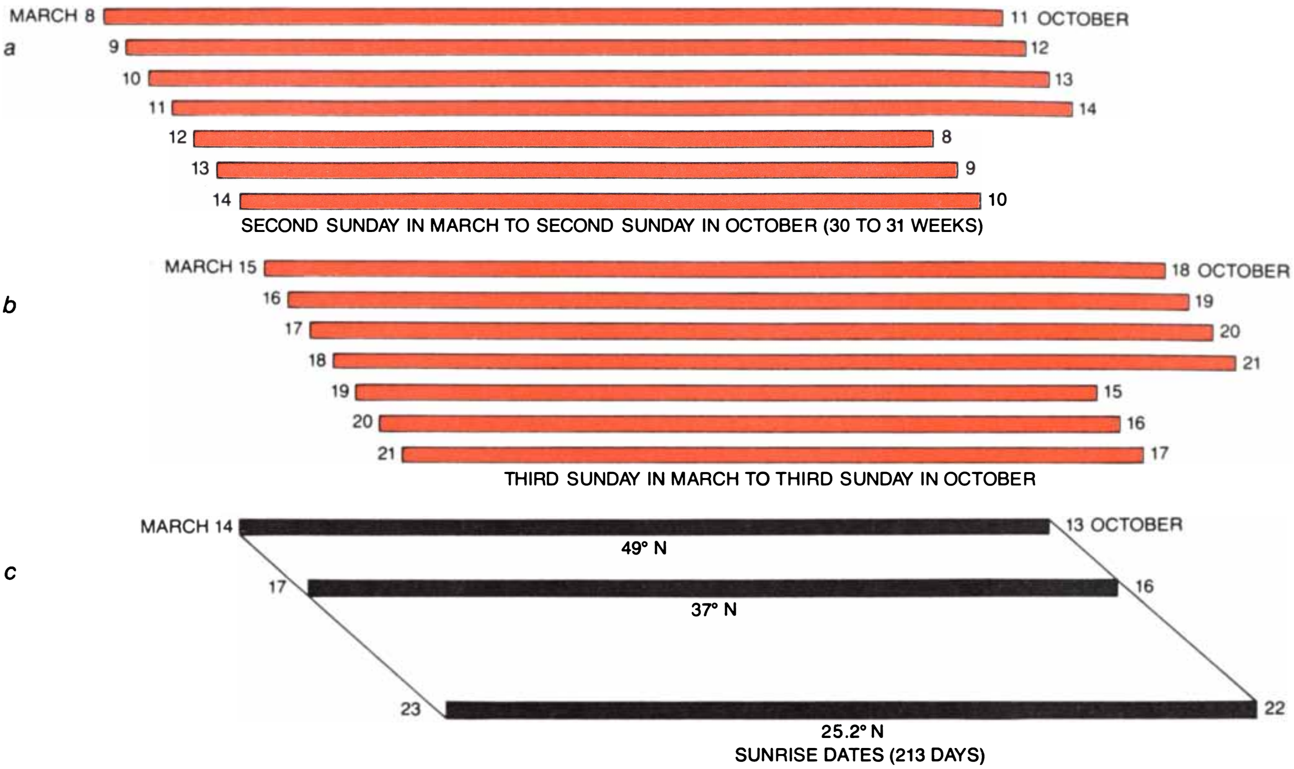
astronomical, geographical and demographic data. On the basis of this information one can set some guidelines for the observance of daylight-saving time.

Clearly the time system should be acceptable to the public, because civil time is what people live by. Surveys of public opinion show that most people want daylight-saving time for at least part of the year. The question is when. In answering the question one cannot assume that people will be governed in their attitude by the region or the kind of place (city, small town or rural community) they live in; the governing factor is more likely to be their location in their time zone. For example, the attitude of a resident of a rural area in Maine will probably differ from that of a resident of a rural area in Michigan. They both receive the same amount of light every day, but the resident of Maine receives it earlier by his clock.

Hence one guideline should be to consider where people live with respect to the standard-time system they must observe. Data from the 1970 census show that more people live in the western parts of the time zones than in the eastern parts and that 17 percent of the population (34.6 million people) live more than 30 minutes from a standard meridian. Somewhat surprisingly, since debates on daylight-saving time have suggested that most of the opposition to it is from Middle Western farmers, the vast majority of these 34.6 million people live in the Eastern time zone: most of Michigan and Indiana, western Ohio, eastern Kentucky and Tennessee, most of Georgia, the central portion of Florida along the Gulf Coast and small western regions of West Virginia, North Carolina and South Carolina. These people, with corresponding groups in the Central and Mountain zones, have been quite vocal in expressing their concerns about daylight-saving time.

Another guideline is to consider the astronomical data pertaining to the U.S. Late sunrises caused by daylight-saving time are the main problem, as the Department of Transportation has pointed out in its reports to Congress on the experiment of 1974 and 1975. The department has provided Congress with a criterion (it can be described as a sunrise-symmetrical arrangement) that can serve to minimize the lateness of sunrises under advanced time during any period. The criterion depends on latitude, so that no one set of dates is valid for the entire U.S., but the National Bureau of Standards has shown how it could be applied to select one period from any group of fixed periods of daylight-saving time. The bureau also presented two ways of applying astronomical data.

The first way leads to the concept of a tolerable sunrise as the determining factor in choosing an advanced-time period. The strong opposition to year-round daylight-saving time almost everywhere



ALTERNATIVE PERIODS of daylight-saving time are compared. The first period (*a*) was adopted by the Senate in 1976, the second (*b*) by a committee of the House of Representatives. The full House did not approve the bill, and so the change to seven months of advanced time rather than the present six months (from the last Sunday in April through the last Saturday in October) was not enacted. Both proposals embody astronomical data matching the advanced-time period to sunrises in the U.S. (*c*), a 213-day period that varies only with latitude. The 49-degree line of latitude and 25.2-degree line are at approximately northern and southern limits of 48 coterminous states.

shows that certain clock sunrise times are unacceptable to the public. The public-opinion surveys indicate that half of the adult population have risen by 7:00 A.M. and three-fourths have risen by 7:30. If the public position was that only the hours of daylight during normal sleeping time could be "saved," the period of daylight-saving time would have to be quite short. The overwhelming acceptance of daylight-saving time shows, however, that what actually happens is that reduced morning light is accepted in exchange for more evening light for most of the summer. One can generate a table of advanced-time sunrises for the U.S. and can use it to choose a period for daylight-saving time after people have answered a question about how late a sunrise they would tolerate during such a period.

The second way was to consider sunrises and sunsets simultaneously. The assumptions inherent in the adoption of daylight-saving time are (1) sunrises are early enough with respect to people's morning habits so that few problems arise when the clock is advanced and (2) most people want additional evening light. The assumptions are equivalent to asking for sunrises that are not too late and sunsets that are not too early. One can have both when the period of daylight is long and the time of sunrise is early. In addition certain months are comparable from the standpoint of the amount of light: April and September, March and October, and so on. When one examines the possible periods for daylight-saving time on the basis of the available daylight, the major problem is the trade-off of late sunrises and early sunsets in March and October.

This trade-off has certain peculiarities, benefiting one region at the expense of another. For example, in the middle of October the sun rises in Portland, Me., at 5:55 A.M. and sets at 4:58 P.M., whereas in Grand Rapids, Mich., the times are 6:56 A.M. and 6:01 P.M. Advancing the clock will help the people in Portland by giving them a sunset at 5:58, but it would put sunrise at 7:56 in Grand Rapids. The seasonal changes and the size of the time zones make it impossible to satisfy both preferences everywhere in October or March. After October and before March the effects are larger, so that the acceptable period of daylight-saving time is limited to eight months by astronomy and people's habits.

The final guideline is the need for uniformity in the country's time system. The present level of 97 percent took more than 50 years to achieve. The experiment in 1974 and 1975 caused a temporary reduction in this level, and it is not difficult to predict that a permanent increase in the length of the advanced-time period would have the same effect. The country cannot go back to the 1880's or even to the 1950's and early 1960's because the transportation and communications networks require a uniform time system. Uniformity was the major goal of the legislation establishing the time zones in 1918, and it has been preserved by Congress in all the bills bearing on daylight-saving time.

The present advanced-time period of six months (from the last Sunday in April through the last Saturday in October) is flawed by being asymmetrical with respect to the sunrise curve. Better advanced-time periods that are symmetrical with the amount of available daylight have been considered: the seven-month period proposed independently by both houses of Congress and a period of eight months. Neither would lead to sunrises later than the present ones.

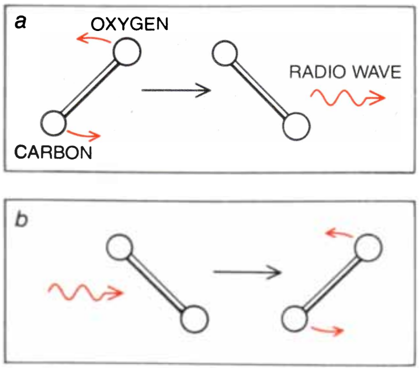
For people who live in Washington, D.C., or in any other place near a standard meridian the choice of a period for daylight-saving time is not critical. The choice must be made, however, with consideration given to the entire population. In our opinion, which takes into account people's concerns and preferences and the need for time uniformity, the best choice is the light-symmetric period of six months starting on the first Sunday in April and ending on the last Sunday in September. The latest sunrise during this period would be at 7:52 A.M., 48 minutes earlier than the latest one in the present period.

Carbon Monoxide in the Galaxy

Radio telescopes tuned to the characteristic emission wavelength of carbon monoxide molecules reveal the presence of a vast ring of cold, star-forming clouds in the invisible inner reaches of the Milky Way

by M. A. Gordon and W. B. Burton

The exploration of our galaxy, the Milky Way, is a comparatively recent venture in astronomy. The problem had been one of not being able to see the forest for the trees, that is, of not being able to visually perceive the large-scale structure of the galaxy through two of its principal components: clouds of gas and dust. The advent of radio astronomy some four decades ago provided a way around the problem. Radio waves, which are emitted by atoms and molecules of gas everywhere in the galaxy, penetrate the interstellar clouds much more readily than light waves do, enabling radio astronomers to map out the invisible structural features of the galaxy. So far the principal radio wavelength used for this purpose has been the 21-centimeter signal emitted by atomic hydrogen (H). We shall report here the results of the newest phase of this exploration, which is based largely on the much shorter radio waves emitted by interstellar molecules of carbon monoxide (CO). It turns out that the picture of the galaxy derived from carbon monoxide molecules is quite different from the one derived from hydrogen atoms.



RADIO WAVE IS EMITTED by a molecule of carbon monoxide at a wavelength of 2.6 millimeters when the molecule stops rotating at the lowest speed allowed by quantum mechanics (*a*). The molecule can absorb radiation at the same wavelength by starting to rotate again at the identical minimum speed (*b*).

It is scarcely more than half a century since astronomers generally agreed that the Milky Way is only one of almost countless stellar systems, or galaxies, some of which contain as many as 300 billion stars. The recognition of this fact was hard won, taking nearly two centuries of observation and discussion. Although the largest telescopes of the 19th century could resolve scattered patches of nebulosity into individual stars, the sizes and distances of these stellar associations were the subject of debate. Some astronomers maintained that our galaxy is the only one, constituting the entire universe, and that the stellar associations seen around it are nothing more than small and relatively nearby zones of active star formation. Others agreed with the view originally suggested in 1755 by Immanuel Kant: some of the nebulas are enormously distant "island universes" comparable in size and structure to the Milky Way.

The experimental resolution of the dispute followed soon after the discovery that certain regularly pulsating variable stars, the Cepheids, could serve as reliable indicators of astronomical distance. In 1925 Edwin P. Hubble announced the observation of Cepheid stars in Messier 31, the Great Nebula in Andromeda. The characteristic relation between the absolute luminosity and the pulsation period of these variable stars showed irrefutably that such nebulas are very far away indeed and hence that they must be regarded as galaxies in their own right.

Once the existence of Kant's island universes was acknowledged, the study of individual galaxies, including our own, got under way. Galaxies incorporating substantial amounts of gas and dust in addition to stars, it was noted, often have a spiral structure. Hubble devised a classification scheme in which the spiral galaxies are arranged from those with a less developed spiral structure (designated S0) through a sequence of intermediate stages (Sa, Sb, Sc) to those with a more developed structure (Sd). Although S0 galaxies are sometimes called "early" and Sd galaxies "late," actually nothing is known about the relation, if any, between the prominence of a galaxy's spiral structure and its age. (To complicate the classification task, about a third of the spiral galaxies have a barlike structure across the middle. These galaxies are called barred spirals and are labeled SB0, SBa and so on.)

One might think the Milky Way galaxy would be an ideal place to begin the study of spiral galaxies, since the solar system is embedded in it. It is not, for the very reason that it has been so difficult to study optically: the disk of any spiral galaxy contains a great deal of gas and dust (as much as 10 percent by mass). In a disk seen edge on the gas and dust efficiently extinguish the starlight of the galaxy. The fact that the solar system lies within such a disk is revealed by plots showing the distribution of the visible stars within the galaxy.

Although the presence of all this gas and dust in a flat disk indicates that the Milky Way galaxy is probably a spiral galaxy (by reference to the situation in external galaxies), the resulting obscuration limits conventional optical investigation of the galaxy to a range of about 6,000 light-years. The Milky Way galaxy appears, however, to have a diameter of more than 100,000 light-years. One cannot even determine what type of spiral the Milky Way galaxy is by means of optical evidence, much less determine the details of its overall structure and composition.

The new tool that enabled astronomers to deal with the galaxy arrived on the scene in 1951. In a single issue of *Nature* that year three groups, one American, one Dutch and one Australian, announced the detection of radio waves from interstellar hydrogen atoms at a wavelength of 21 centimeters (corresponding to a frequency of 1,420 megahertz). Radio waves at this wavelength have two distinct assets: their propagation through space is comparatively unhindered by interstellar gas and dust, and the particles that emit most of the radiation at this frequency, namely hydrogen atoms, are known to be among the commonest constituents of interstellar gas. Therefore by studying the 21-centimeter emission radio astronomers can determine the distribution of the atomic-hydrogen component of the interstellar gas at great distances, far beyond the region accessible to optical telescopes. Until recently it was widely thought the distribution of atomic hydrogen in the galaxy would reveal the distribution of other constituents of the interstellar gas, but that assumption has turned out to be wrong.

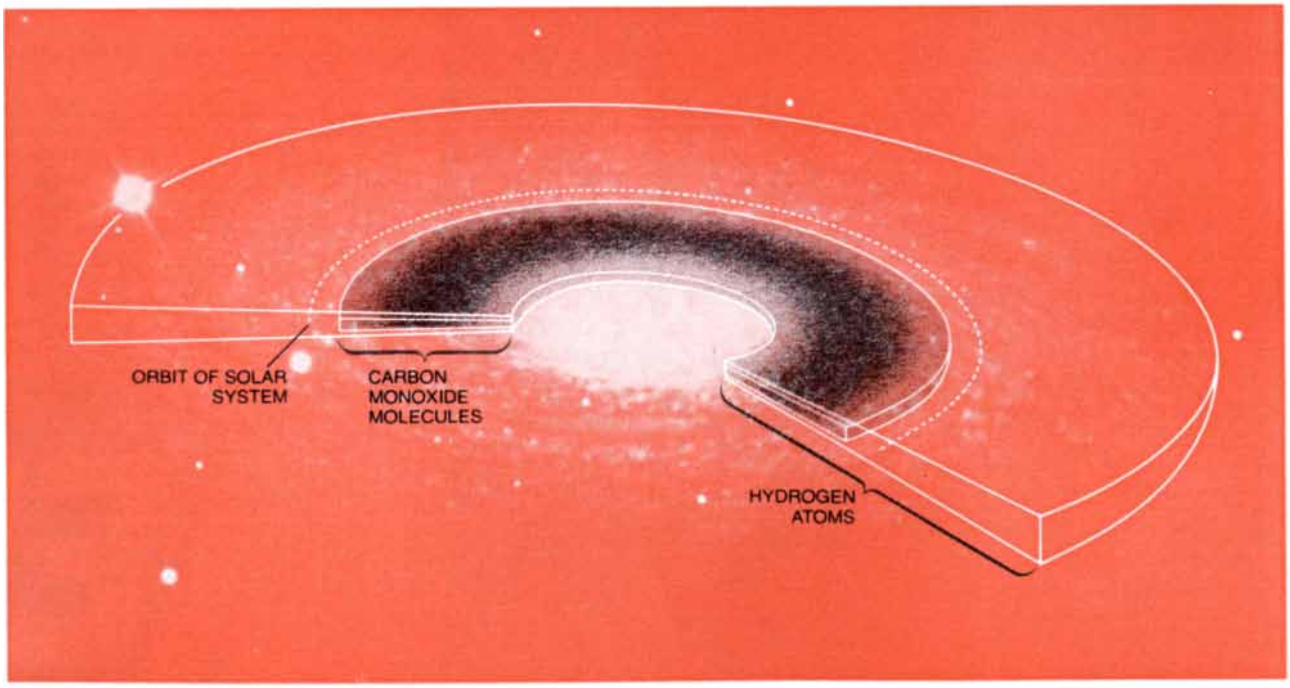
In general the radio waves from a particular kind of atom are emitted at a precise frequency, which is determined by the characteristics of the atom. Because every atom in a gas is moving, however, an observer detects the wave at a slightly different frequency, which is shifted one way or the other from the characteristic frequency in proportion to the relative motion of the atom with respect to the observer. This property, the Doppler effect, means that atoms moving away from the observer will appear to radiate at lower frequencies whereas those moving toward the observer will appear to radiate at higher frequencies. By comparing the observed frequency with the standard one measured in the laboratory the astronomer can calculate the velocity of the gas atoms with respect to the telescope. A velocity away from the telescope is termed positive and a velocity toward the telescope negative.

To explore the galaxy the radio astronomer proceeds by first measuring the intensity of an atom's radiation as a function of frequency. The resulting plot of intensity against frequency is a radio spectrogram, and the process of making such an observation is radio spectroscopy. In practice the astronomer tunes the radio receiver to the standard frequency. The incoming radio signals are then directed to a device that simultaneously measures the intensity at many different frequencies. To simplify the interpretation a computer attached to the receiver automatically corrects the spectrograms for frequency shifts due to known velocities, such as the rotation of the telescope about the center of the earth, the orbital motion of the earth around the sun and even the motion of the sun with respect to its neighboring stars. The computer labels the frequency scale of the spectrogram in terms of the average velocity of the stars in the sun's vicinity, an inertial framework astronomers call the local standard of rest.

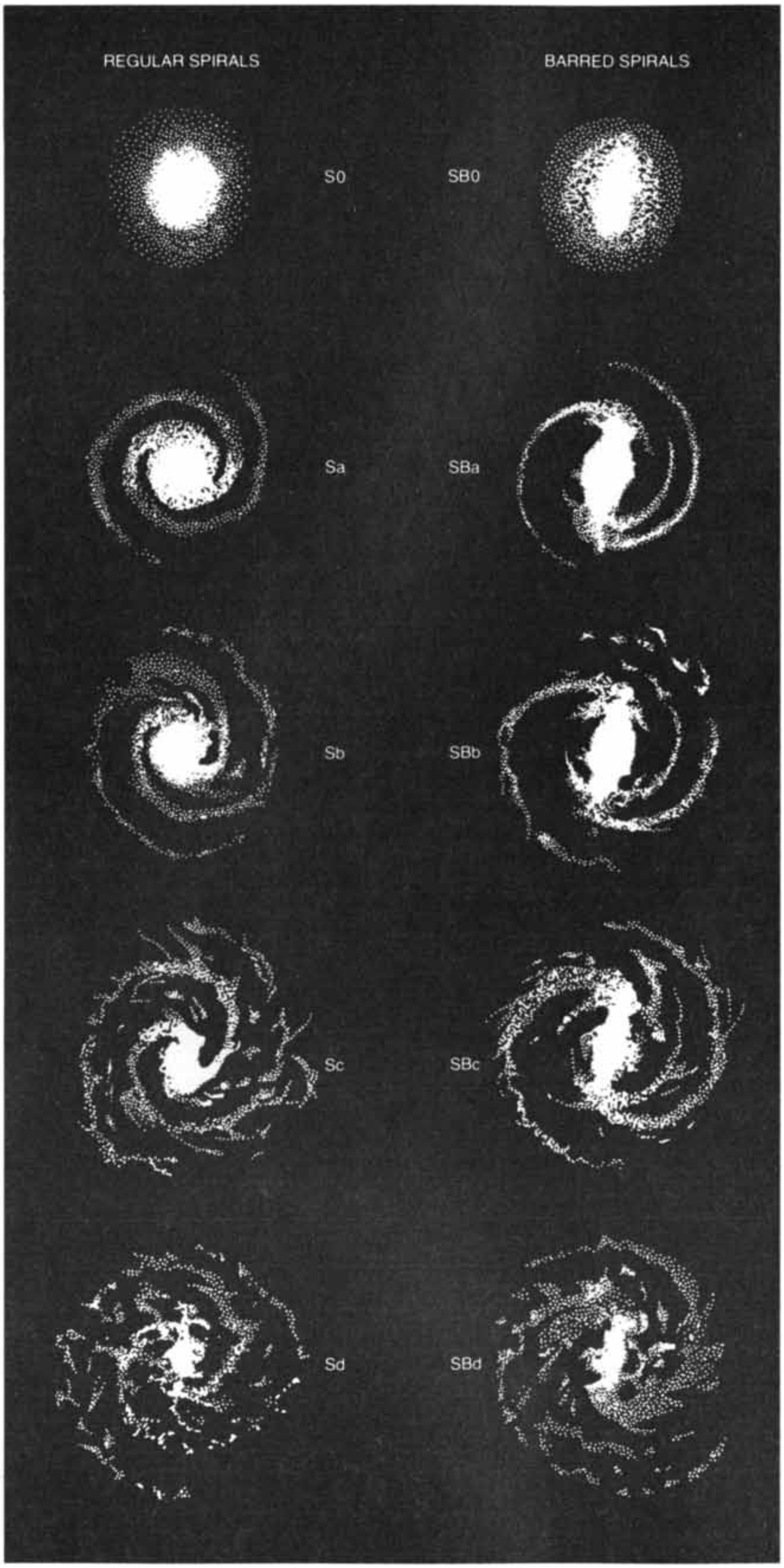
In astronomy distances are generally difficult to measure. One possible indicator of distance is the velocity of the gas under observation. For any line of sight from the telescope through the plane of the Milky Way galaxy it would be helpful to be able to relate velocity to distance from the solar system. In this way one could convert the intenisity of a spectrogram at a given velocity into the amount of gas lying at the corresponding distance along the line of sight. Unlike the problem of converting radio frequency into velocity, however, this problem is not entirely straightforward. It requires that one first find a suitable relation between velocity and distance.

As it happens, the galaxy does provide a way of relating the velocity of the interstellar gas to distance from the solar system, making it possible to explore the inner reaches of the galaxy by means of radio spectrograms. The situation is as follows. Suppose the galaxy were to rotate as a rigid disk. Each position along a line of sight through the disk would then move at exactly the same velocity with respect to the telescope. There would be no way of relating velocity to distance.

In actuality, however, galaxies rotate differentially. In other words, the angular rate of rotation of the interstellar gas and dust changes with distance from the galactic center, decreasing nonuniformly outward. Accordingly the velocity of the gas along the line of sight also changes with distance from the solar system. In particular the velocity of the gas along the line of sight increases with distance up to the point where the radius of the galaxy is perpendicular to the line of sight, and then it decreases. At the intersection of the line of sight with the orbit of the sun on the far side of the galaxy the velocity becomes zero with respect to the sun's velocity. Proceeding even farther away from the sun the velocity becomes increasingly negative with respect to that of the sun. The differential



RING OF MOLECULES around the core of the Milky Way galaxy extends from a galactic radius of about 12,000 light-years to one of 24,000 light-years, attaining its highest density at approximately 17,000 light-years (*gray area*). In contrast, the radius of the sun's orbit around the galactic center is 30,000 light-years (*broken line*). The layer of cold, star-forming clouds defined by the observations of carbon monoxide molecules appears to be about 300 light-years thick. The molecular ring lies within a much larger and thicker disk of interstellar gas, composed predominantly of hydrogen atoms, which radiate at a characteristic radio wavelength of 21 centimeters. The overall radius of the galaxy is estimated to be some 50,000 light-years. The colored background is a photograph of a comparable spiral galaxy.



CLASSIFICATION OF SPIRAL GALAXIES follows a scheme originally proposed by Edwin P. Hubble in 1926. The different types of galaxies shown in this drawing are arranged according to the prominence of their spiral structure. From evidence gathered by radio astronomers over the past 25 years or so the Milky Way galaxy is either an Sb or an Sc galaxy. Whether or not the nuclear region of the galaxy is barred like some others is currently the subject of debate.

rotation of the galaxy is the key that enables one to relate the velocity of the gas represented by a given spectral feature to distance from the sun.

It is not quite that simple. For every line of sight through the inner part of the galaxy there are generally two positions moving at the same velocity with respect to the sun: one position on the near side of the perpendicular point and one on the far side. Both positions, however, are at the same distance from the center of the galaxy. By simply converting velocity into distance from the galactic center rather than distance from the sun one can avoid the distance ambiguity.

Relying on this technique, groups working in the Netherlands and Australia combined their data in 1965 to produce a map of the distribution of atomic hydrogen in the galaxy [*see illustration on page 58*]. To make the map they had to determine distances along the line of sight. The distance ambiguity was resolved by considering many spectrograms arrayed in slices perpendicular to the flat galactic disk and trying to follow a given feature from one spectrogram to another. A feature extending over many lines of sight was presumed to be on the near side of the perpendicular point; one extending over only a few adjacent lines of sight was presumed to be on the far side. The resulting map did not reveal the galaxy's structure very clearly. Either our galaxy is substantially more irregular than had been thought or the method of resolving distance ambiguities was not effective. Nevertheless, the map was one of the first pictures of the overall structure of the Milky Way galaxy.

In the vicinity of the solar system the radio investigations of galactic structure have by no means obviated optical studies. The keystone of the radio method is the change of velocity with distance, a change that is extremely small within 5,000 light-years of the sun. In this neighborhood the optical techniques are vastly superior, because they do not depend on relative velocity. Recently W. E. Herbst of Wesleyan University examined all the optical "tracers," objects that include superluminous young stars, associations of stars in dusty regions and gaseous nebulas. His map shows lanes of stars in the neighborhood of the sun. Unfortunately the efficient extinction of light waves by the interstellar dust limits the size of his map and therefore any independent knowledge of the overall structure of the galaxy.

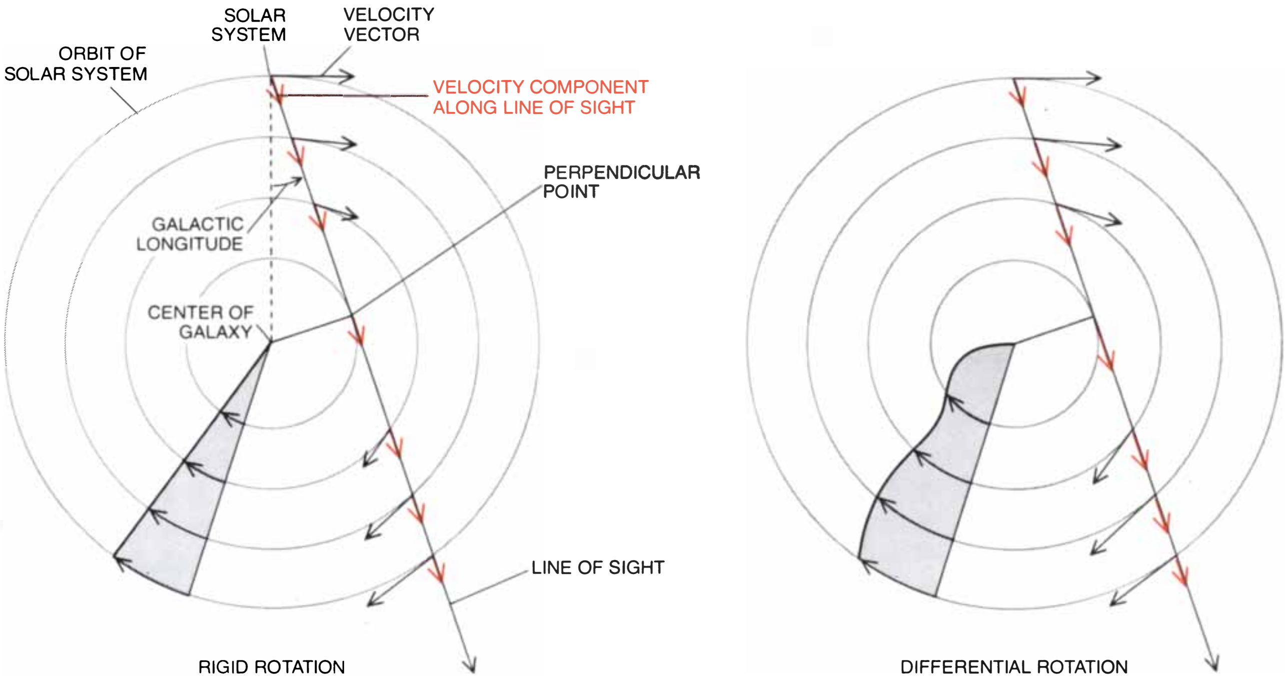
As a probe of the galaxy's structure the 21-centimeter spectral line of hydrogen has certain flaws. For one thing there is simply too much atomic hydrogen in the interstellar gas to provide a sharp picture. For another the tiny mass of the hydrogen atoms makes them particularly vulnerable to buffeting by thermal effects. This buffeting spreads out

SEEN EDGE ON, spiral galaxies such as this type-Sb disk, NGC 4565 in the constellation Coma Berenices, are distinguished by the efficient extinction of starlight by clouds of interstellar gas and dust. **Photograph was made with the 2.1-meter telescope at the Kitt Peak National Observatory. Star plots of the Milky Way galaxy show that the solar system lies within a similar obscuring disk of gas and dust.**

the radio emission in frequency, thereby masking the large-scale motion of the galaxy that is the basis for assigning distances. Furthermore, in the regions of star formation, which are among the most interesting regions of a galaxy, most of the hydrogen is usually in the diatomic molecular form (H_2), and molecular hydrogen cannot be detected by radio telescopes. Although the radio waves from hydrogen atoms are effective in penetrating the dusty horizon of the galaxy, the galactic structure deduced from observations of atomic hydrogen alone is unsatisfactory.

The recent discovery of radio waves from interstellar carbon monoxide has provided the key to a new understanding of the structure of the Milky Way galaxy. Face-on photographs of spiral galaxies show that the most conspicuous features of these galaxies are the extremely young stars, the emission nebulas and the dust lanes, all of which trace out the galaxies' spiral structure. These features coincide with regions of star formation, where the gas clouds get cold and dense, and turbulent eddies within them fall together to form stars. The regions of star formation are also typically the sites of molecules and not of atomic hydrogen. Hence the detection of intense and ubiquitous radio signals from one of these molecules, carbon monoxide, is an exciting finding. It means that regions of star formation can now be investigated directly on a galactic scale.

The carbon monoxide molecule emits its fundamental radio signal at the very short wavelength of 2.6 millimeters. The emission process is as follows. The carbon monoxide molecule can be regarded as a dumbbell consisting of

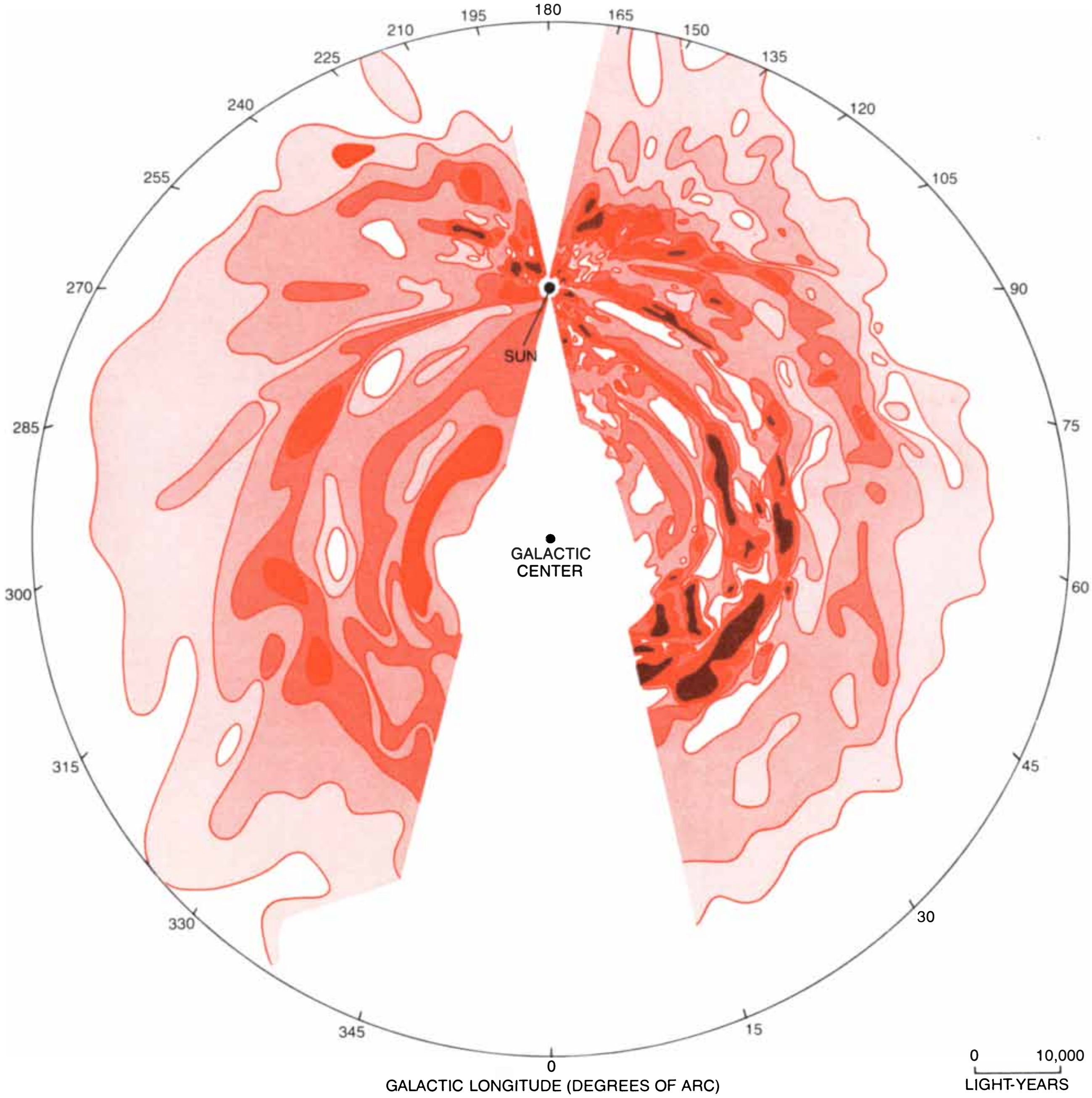


DIFFERENTIAL ROTATION of the galaxy is a key factor in the exploration of galactic structure by means of radio waves. In both situations depicted here the outermost circle marks the orbit of the solar system around the center of the galaxy and the shaded area represents the movement of interstellar material within this orbit during some fraction of a galactic year. If the galaxy were to rotate as a rigid disk, as is assumed in the diagram at the left, then every position along a line of sight through the disk would move at exactly the same velocity with respect to the radio telescope, and there would be no way of relating velocity to distance from the solar system. The fact that the galaxy rotates faster in its inner regions, as is shown in the diagram at the right, means that the velocity of the gas along the line of sight changes as a function of distance from the telescope. For the particular line of sight chosen in this case the rotational velocity of the gas increases with distance up to the point where the galactic radius is perpendicular to the line of sight, and then it decreases. At the intersection of the line of sight with the orbit of the solar system on the far side of the galaxy this velocity becomes zero with respect to the velocity of the solar system; farther away the velocity becomes increasingly negative. The vector component of the velocity along the line of sight (*colored arrows*) is determined by measuring how much the characteristic wavelength of a radio signal is shifted from its laboratory value (as a result of the Doppler effect). Galactic longitude is measured in degrees of arc from the direction of the center of the galaxy.

a carbon atom and an oxygen atom connected electrically. (Actually only the nuclei of the atoms form the dumbbell; the electrons associated with the atoms surround the dumbbell like a swarm of bees.) A molecule of this type can store energy in three ways: the vibration of the atoms with respect to each other, end-over-end rotation and changes in the size of the electron cloud. Because the temperature of dark clouds in the interstellar medium is typically only a few degrees above absolute zero, however, there is little energy available to be stored. Therefore the principal energy-storing activity of the carbon monoxide molecule in dark clouds involves the lowest-energy mechanism, namely the end-over-end rotation of the molecule.

Quantum mechanics gives the details of the process. The smallest amount of energy that can be stored is equal to the energy the molecule acquires when it shifts from not rotating at all to rotating at the lowest rate allowed by quantum mechanics, a rate that is determined by the moment of inertia of the carbon monoxide molecule about its rotation axis. To radiate at a wavelength of 2.6 millimeters the molecule simply stops rotating. To absorb radiation at the same wavelength it starts rotating again. The emission or absorption of energy can be stimulated by collisions with other particles or by the radio waves themselves.

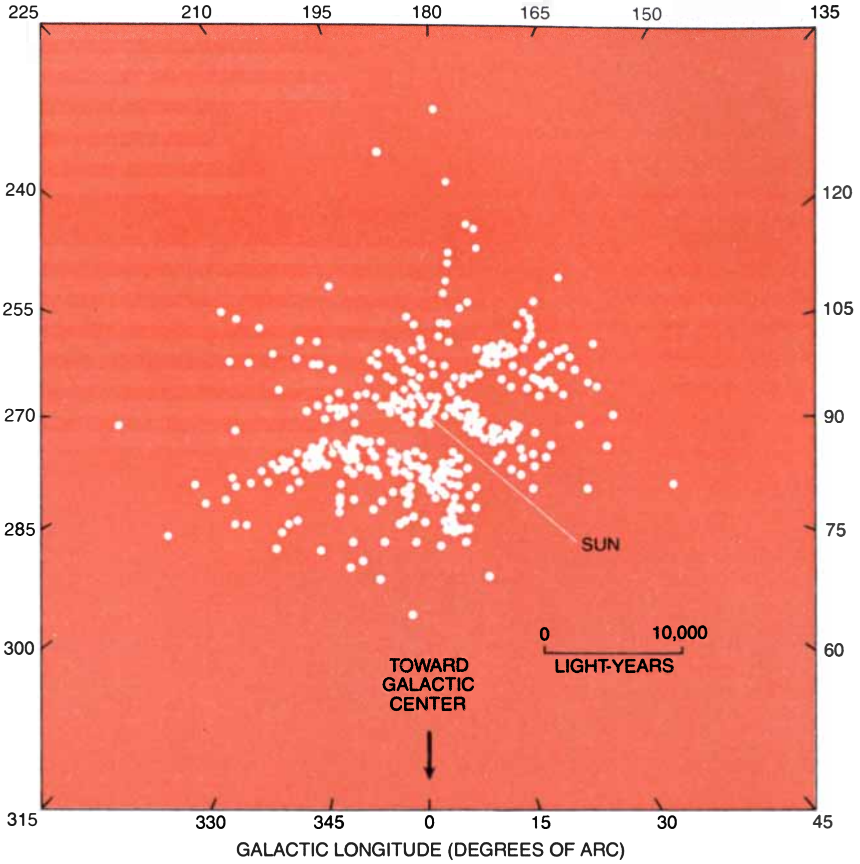
Radio waves from interstellar carbon monoxide were first observed in 1970 by Robert W. Wilson, Keith B. Jefferts and Arno Penzias of Bell Laboratories, working with the 11-meter short-wave radio telescope of the National Radio Astronomy Observatory (NRAO) on Kitt Peak in Arizona. Subsequent observations revealed that the carbon monox-



EARLY PICTURE of the structure of Milky Way galaxy is based on a large number of radio-telescope observations made in the late 1950's of 21-centimeter radio waves emitted by hydrogen atoms distributed along a wide range of galactic longitudes. The picture combines observations made in the Netherlands (mostly by Gart Westerhout and Maarten Schmidt) with those made in Australia (mostly by Frank J. Kerr). The empty wedges in the picture show where the analytic technique cannot be applied because the line-of-sight component of the velocity attributable to the rotation of the galaxy is close to zero. The map fails to reveal a clear spiral structure in the galaxy.

ide signals could be detected over a wide span of the galaxy. Moreover, the signals could be detected from several different isotopes, or nuclear species, of carbon monoxide. The ratio of the intensity of the signals from two of the isotopes was found to be substantially different from the ratio normally found on the earth. Either the isotopic abundances are very different outside the solar system or carbon monoxide is so abundant within the galactic clouds that only a small fraction of its radio energy can escape.

The more likely hypothesis is the latter one. If the radiation from carbon monoxide molecules interacts closely with the motions of the gas cloud, its study should give the temperature of the cloud directly. On the other hand, because all the radiation emitted by one of the less abundant carbon monoxide isotopes escapes from the cloud, the total radiation is proportional to the total number of emitters within the cloud. Hence unlike radiation from hydrogen the radiation from the isotopes of carbon monoxide can be analyzed to yield both the temperature and the amount of gas in distant, invisible regions of the galaxy.



LANES OF STARS in the vicinity of the sun are evident in this map, on which the positions of various optical "tracers" have been plotted with respect to the plane of the galaxy. The map, compiled by W. E. Herbst, includes gaseous nebulas, bright young stars and associations of stars in dusty regions. The size of such maps is limited by the efficient extinction of starlight by the interstellar dust and gas. The area of the galactic disk represented by this map, for example, is less than 5 percent of the area that is covered by the radio-wave map on opposite page.

More than three years elapsed between the discovery of radio waves from interstellar carbon monoxide and their introduction as a probe of large-scale galactic structure. The delay was due entirely to technical difficulties, not to any tardiness in acceptance on the part of astronomers. The frequency corresponding to the 2.6-millimeter wavelength of the carbon monoxide signal is 115,271 megahertz, about 1,000 times higher than the frequencies broadcast by commercial frequency-modulation radio stations. Observations at this frequency call for sophisticated receivers and antennas. The total energy to be measured is small. (All the nonterrestrial signals detected during the entire history of radio astronomy amount to not much more than the energy released by a cigarette ash falling from a height of 10 inches!)

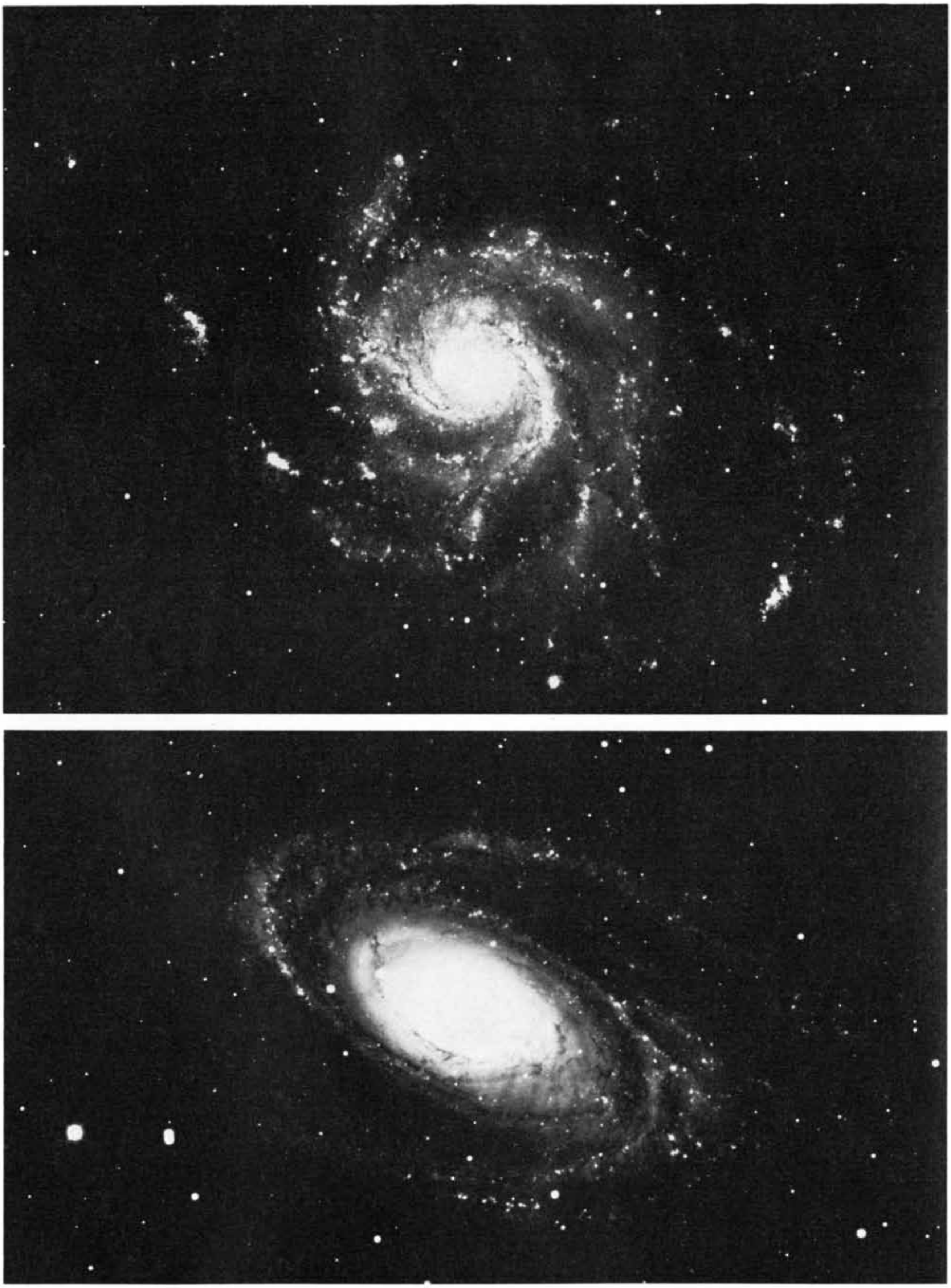
Lines of sight through the central plane of the Milky Way galaxy can traverse a range of velocities higher than 100 kilometers per second, which is equivalent to more than 38 megahertz at the frequency of carbon monoxide. To separate this broadband emission from instrumental background effects a spectrometer would have to cover a velocity range of 300 kilometers per second, which corresponds to a frequency range of 115 megahertz. Furthermore, the radio receiver itself would need to have great sensitivity and stability. At the time of the initial detection of the carbon monoxide signal, however, most millimeter-wave radiometers were barely adequate for detecting intense emissions from individual clouds, far short of the minimum requirements for a galactic survey.

Fortunately engineers at the NRAO were able to satisfy these requirements. The present receiver operating for this research in conjunction with the 11-meter NRAO telescope on Kitt Peak is cooled to 15 degrees Kelvin (−258 degrees Celsius) in order to maximize its sensitivity to astronomical radio signals. The equivalent temperature of the thermal noise generated by this receiver is about 200 degrees K., roughly 100 degrees below room temperature. The receiver and the spectrometer cover a bandwidth of 256 megahertz, corresponding to a velocity range of 660 kilometers per second at the frequency of carbon monoxide.

Our direct observations of carbon monoxide in the galactic plane include spectrograms recorded every .2 degree of arc at galactic longitudes between 10 and 80 degrees, augmented by similar observations for the galactic center conducted by Thomas Bania, who was then working at the NRAO [*see illustration on page 61*]. Spectrograms of this type are often aligned vertically and presented in such a way that the intensity of the recorded signal modulates the colors in the illustration. The resulting diagram is simply a convenient way to look at a large number of spectrograms. Although it contains the essential spectroscopic information from which one deduces the physical distribution of gas in the galaxy, it is in no way analogous to a photographic image of an external galaxy.

The accompanying diagram in the illustration represents comparable observations of the 21-centimeter line of atomic hydrogen. The map is based on observations made by Gart Westerhout with the 100-meter radio telescope of the Max Planck Institute for Radio Astronomy in West Germany and by Bania with the 43-meter NRAO telescope in Green Bank, W.Va.

Keeping in mind the geometry of the galaxy's rotation, one can learn a lot about the galactic clouds just by looking at the raw spectroscopic data. The most striking differences between the carbon monoxide diagram and the hydrogen diagram are the extremely "spiky" appearance of the carbon monoxide data, the absence of carbon mon-

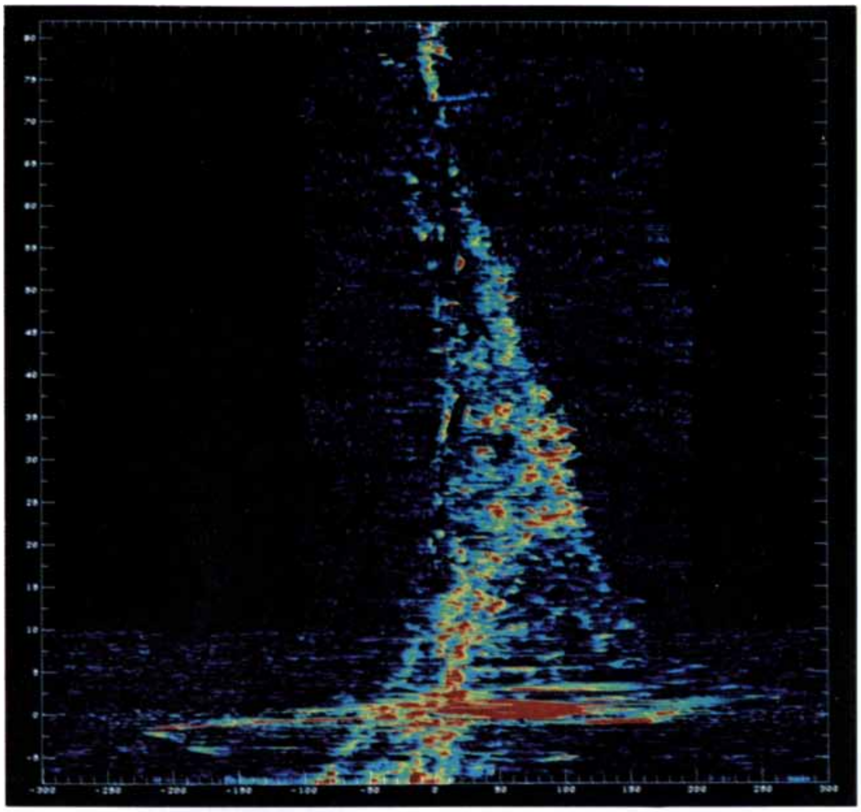


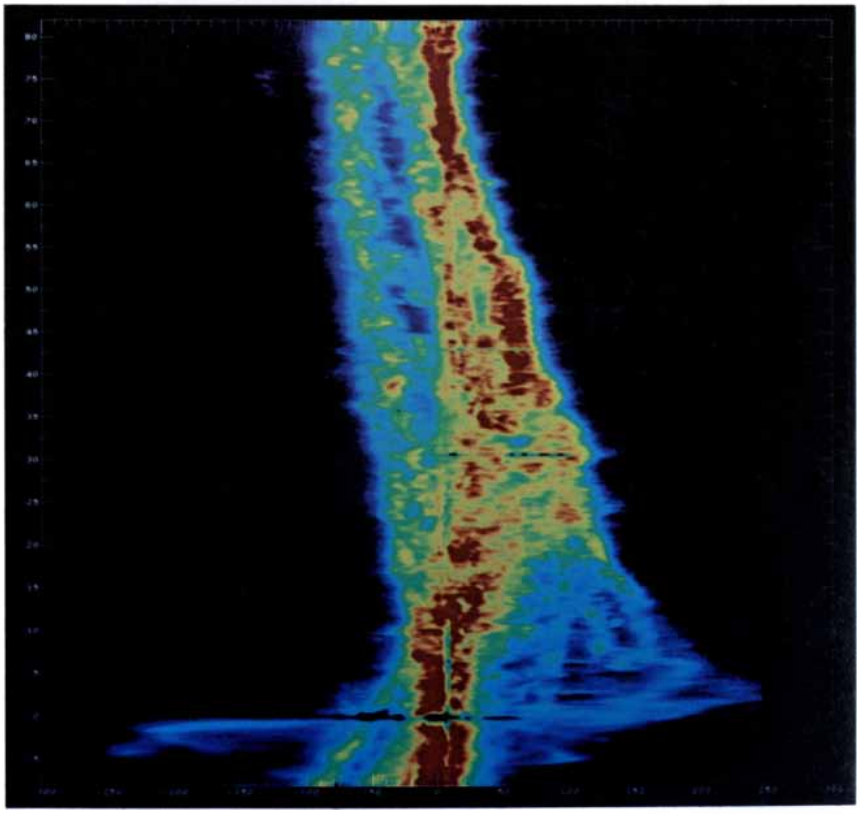
SPIRAL STRUCTURE of two galaxies is clearly delineated by lanes of bright young stars and gaseous nebulas. Seen more or less face on, our galaxy might resemble one of these. The photograph at the top shows a type-Sb galaxy, NGC 3031; the one at the bottom shows a type-Sc galaxy, NGC 5457. Both galaxies are in Ursa Major. The photographs were made with the four-meter telescope on Kitt Peak.

oxide moving at negative velocities at galactic longitudes greater than 10 degrees and the gradual decrease in the total amount of the carbon monoxide emission with increasing longitude. The spiky appearance shows that the radio waves from carbon monoxide are emitted nonuniformly over the galactic plane; presumably the emission is from distinct dense clouds in which stars may be forming.

The geometry of the motions observed along lines of sight shows that negative velocities encountered along lines of sight through the inner galaxy are related to material that lies outside the sun's orbit around the center of the galaxy. The absence of emission from carbon monoxide molecules moving at negative velocities and the decrease in emission toward higher galactic longitudes indicate that the carbon monoxide is confined to the inner region of the Milky Way. (This situation contrasts sharply with that of atomic hydrogen, in which the strong emission at negative velocities indicates an abundance of hydrogen gas in the outer regions of the galaxy.) In addition the marked deficiency in the carbon monoxide emission at positive velocities at longitudes less than 20 degrees means that carbon monoxide molecules are not abundant within a galactic radius of about 12,000 light-years. The same deficiency is observed for hydrogen atoms. Evidently both the molecular and the atomic forms of interstellar gas are scarce in the central region of the galaxy.

To analyze the carbon monoxide observations quantitatively it is necessary to convert the velocity information into distance from the galactic center. In other words, one must know how the galaxy rotates as a function of galactic radius, a physical relation known as the galactic rotation curve. With a couple of assumptions, the spectrograms themselves can give the rotation curve. As we mentioned above, the maximum velocity along any line of sight through the inner galaxy is that of the material at the perpendicular point. The distance of that point from the galactic center changes with the longitude of the line of sight. The right-hand edge of the spectrograms represents the gas at the perpendicular point, and the velocity at that point (after correcting for the motion of the sun) gives the rotation speed of the galaxy at that distance from the center of the galaxy. By measuring these maximum positive velocities as a function of longitude one can determine the rotation curve.





RADIO SPECTROGRAMS were recorded every .2 degree of arc at various longitudes along the galactic equator and then aligned vertically to produce these two diagrams, in which the intensity of the recorded signal modulates the colors, ranging from black (no emission) to red (maximum emission). The spectrograms at the top were made by Thomas Bania and the authors, based on the 2.6-millimeter emission from carbon monoxide molecules; those at the bottom were made by Bania and Westerhout, based on the 21-centimeter emission from hydrogen atoms. The carbon monoxide data are clumpier, indicating discrete clouds. Observed velocity of the gas (*horizontal scales*) can be interpreted in terms of distance from galactic center.

The preceding analysis is based on two major assumptions. The first is that all the galactic motion is circular: the material of the galaxy is moving only in circular orbits around the galactic center. Actually not all the material moves in this way. There are large-scale variations in velocity about the smooth rotation curve that are caused by "winds" in the galactic plane driven by gravity and by shock waves. For the purpose of this discussion, however, neither of these mechanisms is important.

The second assumption is that material is so smoothly distributed over the galactic plane that there is always gas at the perpendicular points. The distribution of atomic hydrogen is ubiquitous enough for this assumption to hold. The distribution of carbon monoxide, however, is clumpy. The empty regions between the clumps are responsible for much of the scatter in the measured velocities from which the rotation curve is derived. An examination of the velocity-longitude diagram for carbon monoxide shows that the scatter does not result from instrument noise. The nature of the scatter provides direct information on the extent of the empty regions and hence on such matters as the probability of collisions among the clumps. The scatter also contains information on the random motions of the individual clouds.

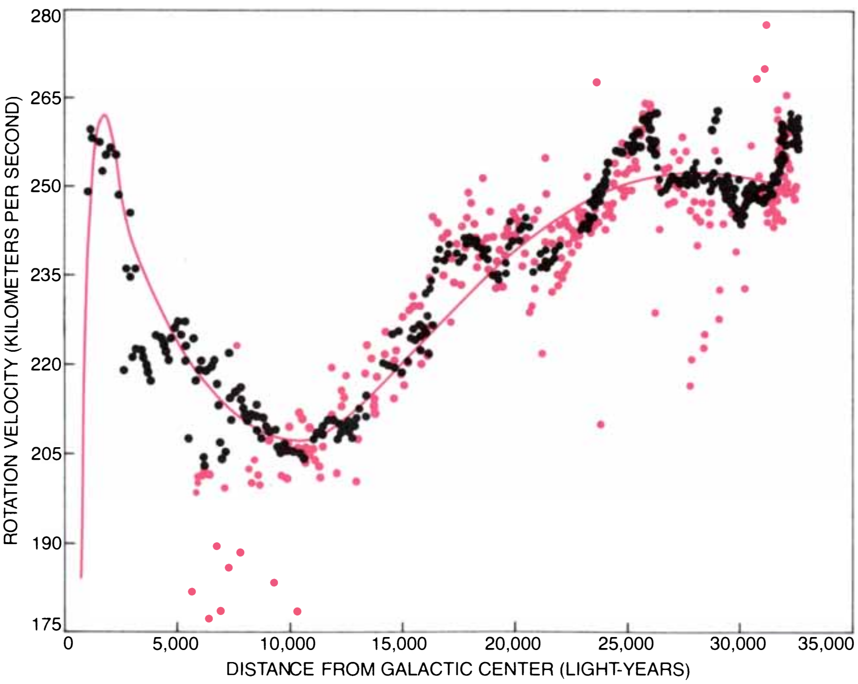
The carbon monoxide molecules trace out regions of star formation because stars form in cold, dense clouds of gas. The turbulent eddies of the clouds can cause local increases in the density of the gas, to the point where the gas molecules are gravitationally attracted to one another. They then begin to condense into a protostar. As the gas contracts it also heats up. Eventually the temperature gets high enough to ignite a nuclear reaction, which releases the energy from the atoms. This energy is what causes a star to glow and to expand until the inward gravitational force is just balanced by the outward radiation pressure. The carbon monoxide molecules are within the dense gas clouds, which are sufficiently opaque to be self-protected from dissociation by ultraviolet radiation from the newly formed hot stars. Therefore mapping the distribution of carbon monoxide in the galaxy is in effect mapping the regions of active and potential star formation in the galaxy.

The quantitative analysis is straightforward. From the rotation curve for the galaxy it is possible to determine the total intensity of the carbon monoxide emission (and hence the number of cold, star-forming clouds) as a function of galactic radius. The procedure is to consider successive intervals along a given longitudinal line of sight, to calculate the velocity within each interval and to ask whether the appropriate spectrogram shows the carbon monoxide emission feature at that velocity. If it does, the intensity is subtracted from the spectrogram and recorded as a function of the galactic radius of the interval.

The resulting bar chart shows a striking variation in the intensity of the carbon monoxide emission with galactic radius [*see illustration on opposite page*]. Most of the carbon monoxide is within the orbit of the sun, which is 30,000 light-years from the galactic center. The amount of cold material increases with distance toward the center up to a galactic radius of 12,000 light-years, where it decreases abruptly. The maximum intensity of the radiation from carbon monoxide is at 17,000 light-years from the galactic center. Assuming that the galaxy is symmetrical, the carbon monoxide is evidently distributed in a ring around the galactic center with the sun near the outer edge.

The carbon monoxide molecules are excited by collisions with hydrogen molecules, so that the data show indirectly how the abundance of molecular hydrogen varies with galactic radius. The relation between molecular hydrogen and carbon monoxide is intimate, because it is only molecular hydrogen that is present in the galaxy at sufficient densities to excite the observed transition between the rotational energy states of the carbon monoxide molecule. In the interstellar gas molecular hydrogen is approximately 10,000 times more abundant than carbon monoxide, which is the next most abundant molecule.

One of the most important aspects of the carbon monoxide findings is what they imply about the total amount of molecular hydrogen in the galaxy. From the known absorption of diffuse X rays near the nucleus of the galaxy, together with our data, it is possible to estimate the density of molecular hydrogen and to compare it with the density derived from a similar analysis of atomic hydrogen. Because of the predominance of hydrogen in the interstellar gas the sum of these two densities yields a good estimate of the total density of the interstellar gas. The difference in radial distribution between atomic and molecular hydrogen is striking. At distances from the galactic center greater than the sun's orbit hydrogen exists primarily in atomic form; it gradually changes into the molecular form with distance toward the galactic center. Even though the concentration of hydrogen in any form suddenly decreases within a radius of 12,000 light-years, the gas there is predominantly molecular hydrogen. Since star formation is most likely to proceed in the cold, dense regions rich in hydrogen molecules, the solar system must lie on the outer edge of the regions of major star formation in the galaxy; therefore comparatively few stars must form in the galactic disk farther from the galactic center than the sun. (This generalization is strictly statistical, because our analysis has smoothed the gas distribution in galactic azimuth. It does not rule out the exis-



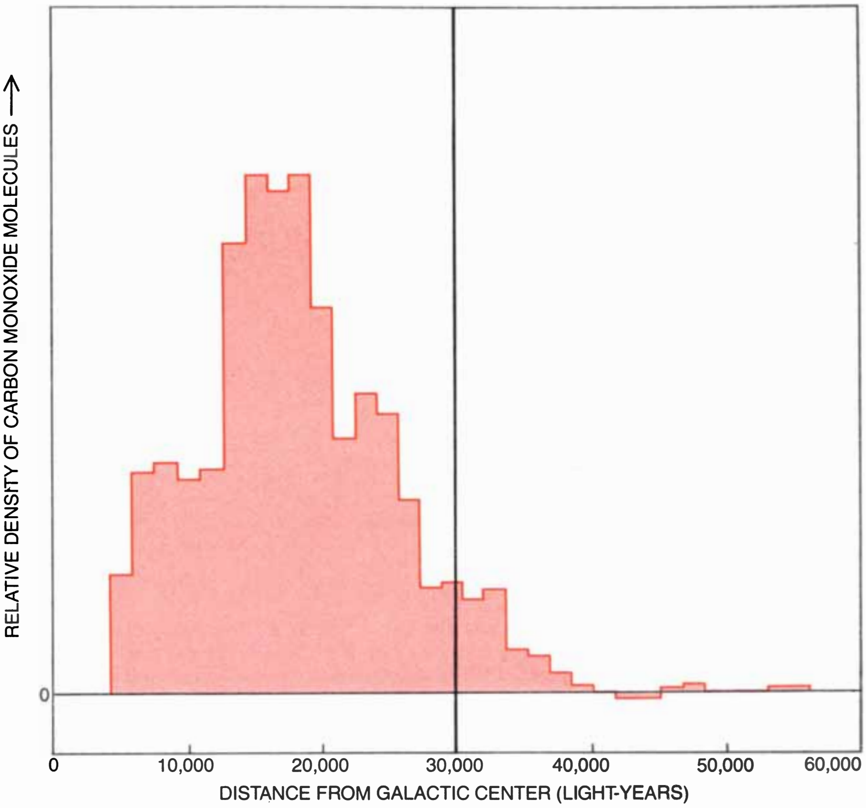
DIFFERENTIAL-ROTATION CURVE **of the galaxy was constructed on the basis of the radio observations summarized in the diagrams on the preceding page. The velocities measured along the right-hand edge of the spectrograms in those diagrams were first added to the line-of-sight component of the orbital velocity of the sun and then plotted with respect to galactic longitude. Carbon monoxide data are in color, atomic-hydrogen data in black. Curvilinear nature of both sets of data demonstrates that the galaxy does not rotate as a rigid body.**

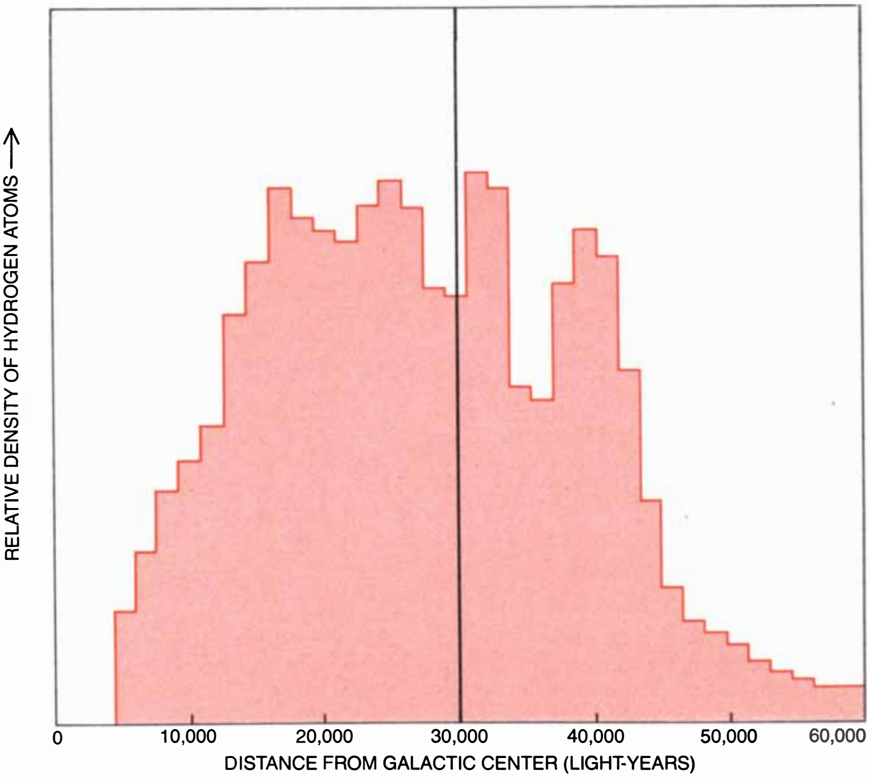
tence of isolated patches of vigorous star formation in the outer parts of the galaxy.) If one were outside the galaxy and looking at it face on, the sun would be at the edge of the visible disk.

Findings from several branches of astronomy have recently converged in agreement with this conclusion. The distribution of atomic hydrogen revealed in the 21-centimeter data had been thought by many to be the prototype for the distribution of all the other constituents of the interstellar medium. It is now apparent that the atomic-hydrogen distribution is unique, defining a much larger galaxy than is implied by the molecular data and by the associated region of very young stars. A similar situation is being revealed in external spiral galaxies as the increasing sensitivity of 21-centimeter observations makes it possible to map them. The maps show that the atomic-hydrogen region extends out to at least twice the radius of the photographic image, which records the light emitted by the brightest newly formed stars and by the clouds of ionized gas surrounding them. The dynamics of this distribution is poorly understood. It is not known, for example, what confines the atomic hydrogen in the outer parts of the galaxy to a thin layer, keeping it bound to the galaxy and rotating in circular orbit about the galactic center.

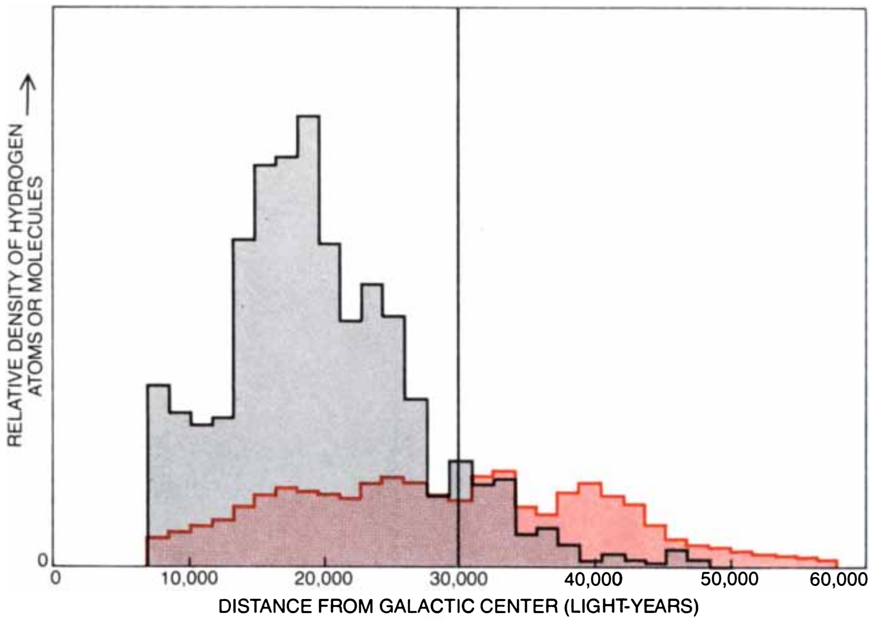
What about the thickness of the carbon monoxide layer? In order to determine the total amount of interstellar gas one must measure the variation of both hydrogen atoms and carbon monoxide molecules with respect to galactic latitude. We have made such a measurement along a galactic longitude passing very nearly through the region of maximum carbon monoxide intensity, 17,000 light-years from the galactic center. The results show that although the atomic-hydrogen disk is perhaps 650 light-years thick, the carbon monoxide ring is only 300 light-years thick, coinciding with the estimated thickness of the layer of very young stars seen optically in the vicinity of the sun. Here is additional evidence for the close relation between the dense gas clouds and the young stars.

Richard Cohen and Patrick Thaddeus of the National Aeronautics and Space Administration and Columbia University have investigated the thickness of the carbon monoxide layer at a number of longitudes. Their findings show that the thickness increases from the galactic center outward toward the sun, a result consistent with the measured behavior of atomic hydrogen, even though the hydrogen layer is twice as thick. Because the thickness of the carbon monoxide layer supplies information on the confinement of the total gravitational po-





DISTRIBUTION of carbon monoxide molecules (*top*) and hydrogen atoms (*bottom*) in the galaxy is plotted in these bar charts as a function of distance in light-years from the galactic center. The vertical black line in each chart represents the radius of the sun's orbit around the center of the galaxy at a distance of 30,000 light-years. Most of the carbon monoxide is concentrated in a ring around the galactic center with the sun situated near the outer edge. The intensity of the radio emission from the carbon monoxide reaches a maximum at about 17,000 light-years from the galactic center. The distribution of atomic hydrogen is not confined to this ring.



DIFFERENCE between the galactic distribution of atomic hydrogen (*colored bars*) and the distribution of molecular hydrogen (*gray bars*) is striking. Close to the galactic center most of the hydrogen is apparently present in the molecular form (H_2), whereas at distances greater than the radius of the sun's orbit (*vertical black line*) most seems to be in the atomic form (H). Processing of the interstellar material into complex molecules, cold compressed dust and eventually into stars, reflected by this distribution, is occurring in the inner part of galactic ring.

tential to the galactic equator, this result is of interest in the context of galactic dynamics.

The ratio in a galaxy of the mass of gas to the mass of stars can reveal much about the efficiency of star formation. If one takes a nominal figure for the thickness of the carbon monoxide layer and of the hydrogen layer, one can calculate the distribution of gaseous mass (excluding stars) as a function of galactic radius, since atomic hydrogen and molecular hydrogen are by far the dominant gaseous constituents of the galaxy. Calculating the total mass (including the stars) is more difficult. To this end numerous efforts have been made to combine observations of stars in the neighborhood of the sun with the deduced galactic rotation to generate models of the distribution of the unseen stars in the Milky Way galaxy.

A recent theoretical model of the galaxy put forward by Kimmo A. Innanen of York University agrees surprisingly well with the observed distribution of the gaseous mass. Innanen's data and our carbon monoxide results suggest that at a distance of between 30,000 and 12,000 light-years from the galactic nucleus the ratio of gas to stars is approximately 4 percent by mass, which implies a uniform rate of star formation in the middle region of the galaxy. This result calls for changes in earlier concepts of how star formation proceeds in a spiral galaxy. Formerly much of what was known about the distribution of gas in the galaxy came from radio observations of atomic hydrogen, the abundance of which with respect to stars decreased with distance toward the center of the galaxy. The underlying assumption was that star formation proceeded more vigorously in the central galactic regions, thereby depleting the interstellar hydrogen. The new data show that the interstellar gas has not been depleted; it exists in the form of the molecular gas. In fact, the total mass of the molecular hydrogen in the galaxy is equal to two trillion times the mass of the sun and is the same as the total mass of the atomic hydrogen. If a new theory of star formation is to be successful, it must account for this unexpectedly constant ratio of gas to stars.

Any astronomical model is a blend of observable facts and reasonable hypotheses. In our spectrograms we have observed thousands of characteristic carbon monoxide signals, representing an equal number of dark clouds. Because we were only able to sample the gas within the galactic plane at intervals that were large compared with the cloud sizes there must be many more dark clouds than we observed, far too many to deal with on an individual basis. It would take the 11-meter NRAO telescope years to completely map the distribution of molecules in the Milky Way galaxy. That would be out of the question, not least because the instrument is in great demand for other investigations. Instead we analyze our data in a statistical manner, following what are called Monte Carlo techniques. We supply a computer with such basic parameters of individual clouds as their size and with such general parameters of all the clouds as the rotation curve of the galaxy and the thickness of the carbon monoxide layer. We also supply first guesses at what we want to determine: the relative velocities of the clouds with respect to one another, the number of clouds and their spatial separation.

The computer places the clouds randomly in the model galaxy, which it then "observes" in a way that mimics our actual observations. By comparing the actual observations with the computer observations we can test some of our hypotheses. By varying the hypotheses and parameters slightly we can see how sensitive the model is to our initial assumptions and correspondingly how good our guesses were. High sensitivity favors a good result and vice versa. The manipulation of various parameters leads not only to new insights into the large-scale characteristics of dark clouds in the galaxy but also to new information about the structure of the galaxy.

The success of our final model can be judged by examining the illustration on page 66. The top set of curves shows the simplest comparison between the observations and the model: it compares the intensities of the observed and the simulated carbon monoxide emission integrated over all velocities plotted against longitude. The two curves agree well, showing that we have correctly modeled the carbon monoxide clouds along each line of sight, not only in actual number but also in the degree of randomness (as is shown by the range of scatter).

Another way to test the model is to examine the maximum positive velocities as a function of galactic longitude. This comparison is much more sensitive than the one described above, because it tests both the velocity characteristics and the spatial distribution of the model clouds. The fact that the average velocities match those of the observations means the assumed rotation curve is the correct one.

At galactic longitudes between 25 and 10 degrees the scatter in maximum positive velocity increases sharply. In these directions the line-of-sight velocity changes rapidly in the vicinity of the perpendicular point. Because the number of dark clouds decreases on the inner side of the carbon monoxide ring and because the line-of-sight velocity changes rapidly there the probability of a cloud's being at the perpendicular point becomes small. Clouds removed from the perpendicular points generally will have a lower velocity, and their maximum velocities will fall below the idealized value shown. A similar situation holds for longitudes greater than 55 degrees. Here the clouds become so sparse that the probability of one being at the perpendicular point is low. In both regions the model agrees well with the observations.

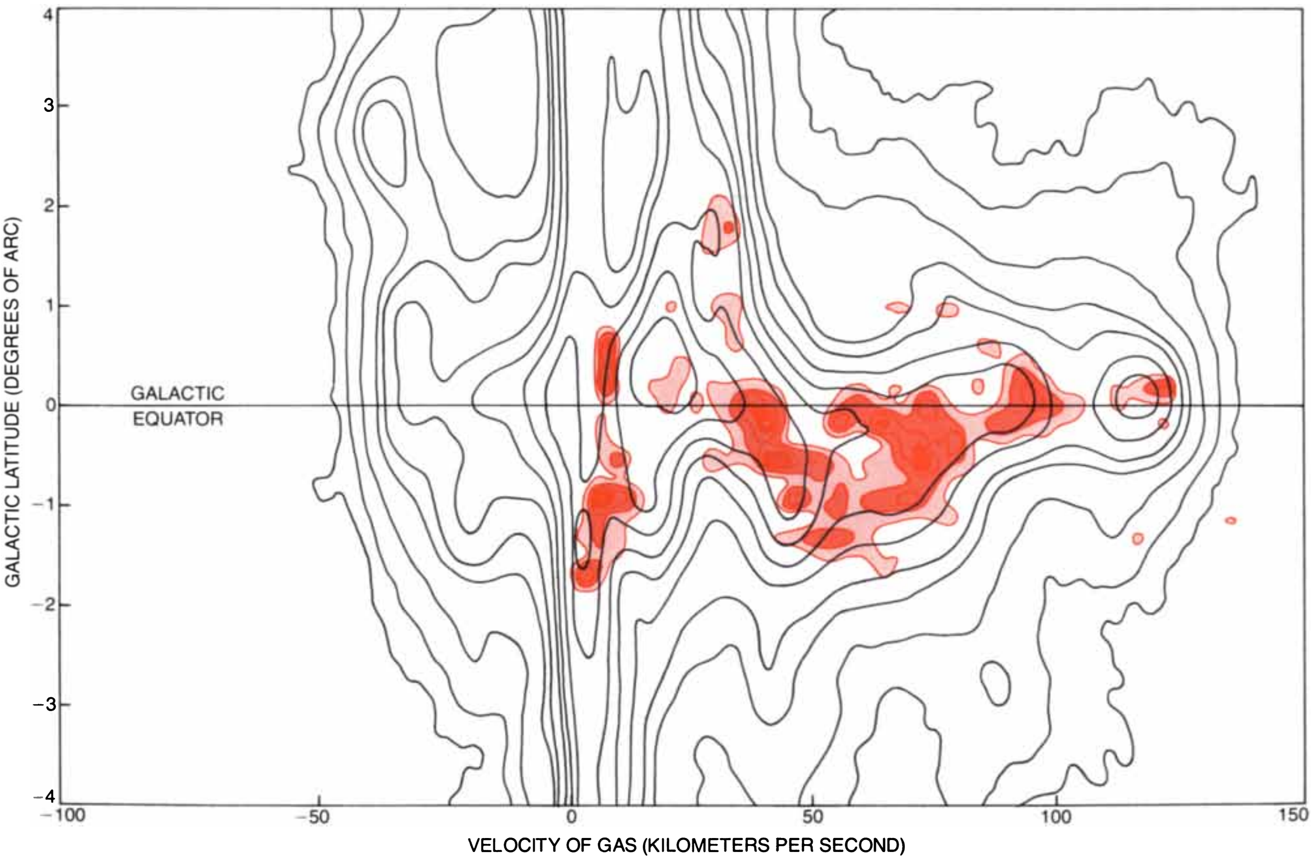
Although the model reproduces many of the observations rather well, it required a number of modifications to achieve its present success. Some of the initial assumptions proved to be incorrect, but reasonable adjustments could be found to bring the model in line with the observations. For example, in order to match the scatter in the observations we had to assume that the random component of the velocity of one cloud with respect to another has an amplitude of about four kilometers per second. This high relative velocity, combined with the average cloud separation of only about 10 times an average cloud's diameter, means that collisions and subsequent accretion of clouds into large clumps is not uncommon. The finding leads to an interesting conclusion: the giant clouds that are probably assembled in this way have total gaseous masses on the order of 100,000 times the mass of the sun, which is a fairly typical star. In short, these recently discovered objects are the most massive objects in the galaxy.

Our final model contains some hundreds of thousands of clouds, each cloud with a representative diameter of 50 light-years. The mean separation of the clouds in the densest region at the middle of the molecular ring is about 400 light-years, a distance that increases toward the edges of the ring. Actually most of the clouds are known to range in size from 15 to 100 light-years, but incorporating a range of cloud sizes into the model would probably not affect the results substantially. The total mass of the molecular-cloud material in the galaxy is estimated to be equal to several billion solar masses.

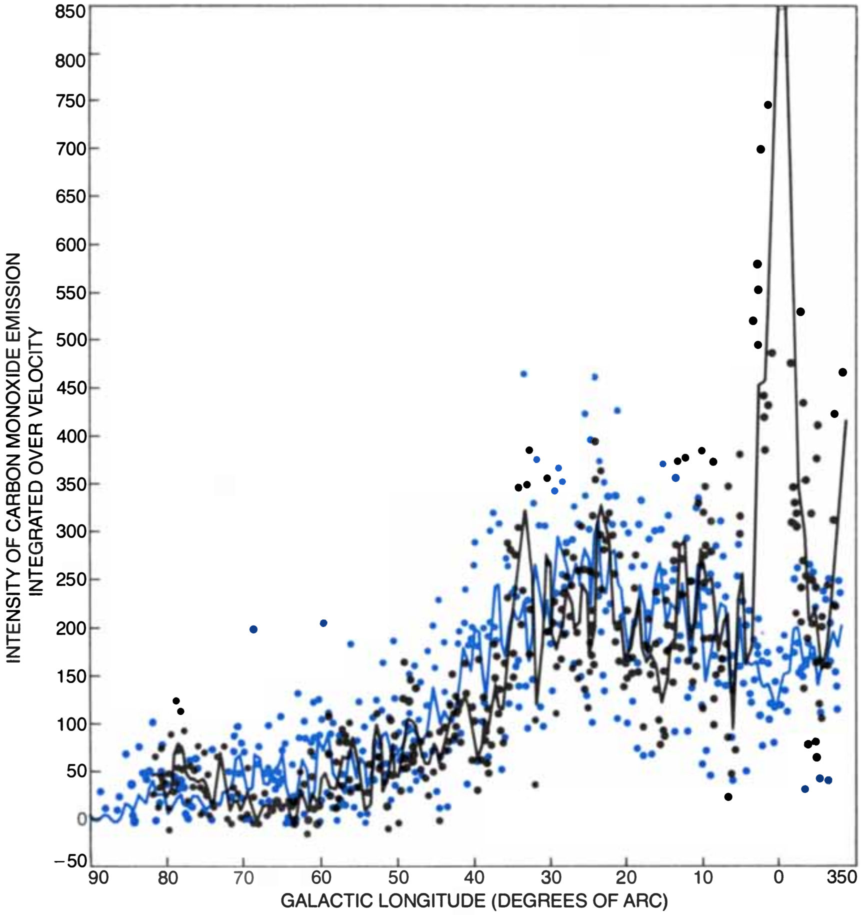
There is still a serious flaw in the model. We have assumed that the clouds are distributed throughout the carbon monoxide ring more or less uniformly in azimuth. When one compares the actual carbon monoxide spectrograms with those produced from the model by the computer, however, the actual data show a higher-order clumpiness than the computer-produced clouds, which are more uniformly distributed over the velocity-longitude region.

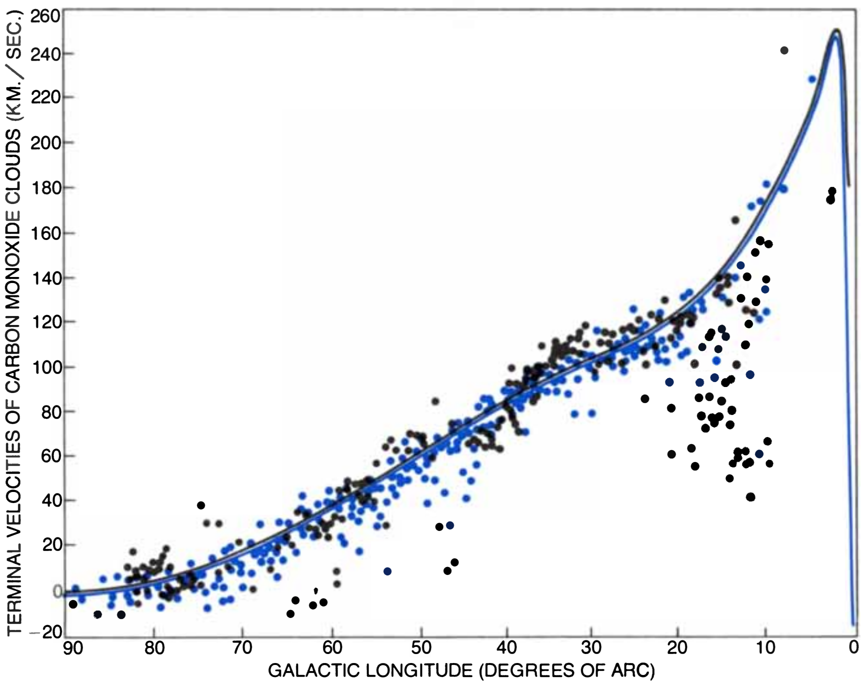
The most likely explanation is that the dark clouds lie along fragments of the galaxy's spiral arms and as such are not distributed in an azimuthally symmetrical way about the galactic center. W. W. Roberts of the University of Virginia has devised theoretical descriptions of the spiral arms in terms of gravitationally induced shock waves in the interstellar gas of galaxies. The central concept of this density-wave theory is that the interstellar gas and stars move through the spiral waves in their rotation about the galactic center. Being massive, the stars pass through the wave unaffected, but the interstellar gas is compressed, forming dark clouds in the process. The molecular clouds would themselves be so dense that they would respond to subsequent perturbations of the density wave, behaving more like stars than like interstellar gas. Because the lifetime of dark clouds is thought to be short compared with the time required for them to migrate from one spiral wave to another, however, an observer would see the dark clouds slightly downstream from the spiral wave, thereby delineating the wave. The clouds would presumably be visible for another reason: stars would form within them, so that the spiral wave would also be marked by highly luminous young stars.

Incorporating the notion of spiral waves into our computer model im-



THICKNESS of the galaxy's molecular layer can be compared with that of its atomic-hydrogen layer on the basis of this contour diagram, which plots the velocity of carbon monoxide and atomic hydrogen as a function of galactic latitude along adjacent lines of sight arrayed at right angles to the gas layer and passing very nearly through the region of maximum carbon monoxide intensity at the radius of 17,000 light-years from the galactic center. The results show that the carbon monoxide layer (*colored contours*) is about half as thick as the hydrogen layer (*black*). The clumpier appearance of the carbon monoxide data again suggests that these molecules form discrete clouds.





TWO COMPARISONS are made here between the actual radio observations of carbon monoxide in the galaxy (*black*) and the properties predicted by the authors' computer model of the galactic distribution of carbon monoxide clouds (*color*). The graph at the top, which gives the total intensity of the carbon monoxide emission integrated over all velocities plotted against galactic longitude, shows that in this respect the model agrees well with the observations. The graph at the bottom, which gives the most positive velocities of carbon monoxide (that is, those at the perpendicular points) as a function of galactic longitude, shows an even closer agreement. The latter test is much more sensitive than the former in establishing the correct number of clouds, their characteristic separation and their velocities with respect to one another.

proves the agreement with the observations. The illustration on the opposite page shows the observations side by side with a uniform-distribution model and with a density-wave model of the cloud distribution calculated in collaboration with Roberts. The long feature at a velocity of 50 kilometers per second observed between the 50-degree and the 25-degree longitude is now present in the density-wave model, but the higher-order clumpiness of the data is still not reproduced by the model. Moreover, the velocity width of the wave-produced features in the model is rather narrow compared with that of the observations.

A possible explanation of the failure of the model to reproduce the higher-order clumpiness of the data is that there are regions in the galaxy, possibly along the spiral arms, where the conditions for star formation are unusually favorable. In these regions there is a surfeit of dark clouds. Photographs of spiral galaxies lend support to this picture. Typically they show distinct regions of high luminosity in places where the dark clouds would be located. Because the clouds are long-lived, large and move with respect to one another, it is likely that they occasionally collide and thereby collect to some degree.

Although we have not yet perceived the details of the grand design of the Milky Way galaxy, the observations have established important constraints for star formation in the optically inaccessible inner regions of the galaxy. In particular they have isolated the regions of active star formation.

Comparing our estimates of the ratio of the mass in stars to that in the interstellar gas, we find that between 12,000 and 24,000 light-years from the galactic center the value of the ratio stays remarkably constant at 25 to one. What accounts for this constancy? Has star formation reached the saturation point in this part of the galaxy, and is it now limited by the small amount of gas available? According to the density-wave theory of galactic structure, the strength of the compression should increase inward toward the galactic nucleus. One might assume that the rate of star formation would increase accordingly. Yet the only available evidence, even though it is tentative, suggests that the ratio of stars to gas over this region is constant.

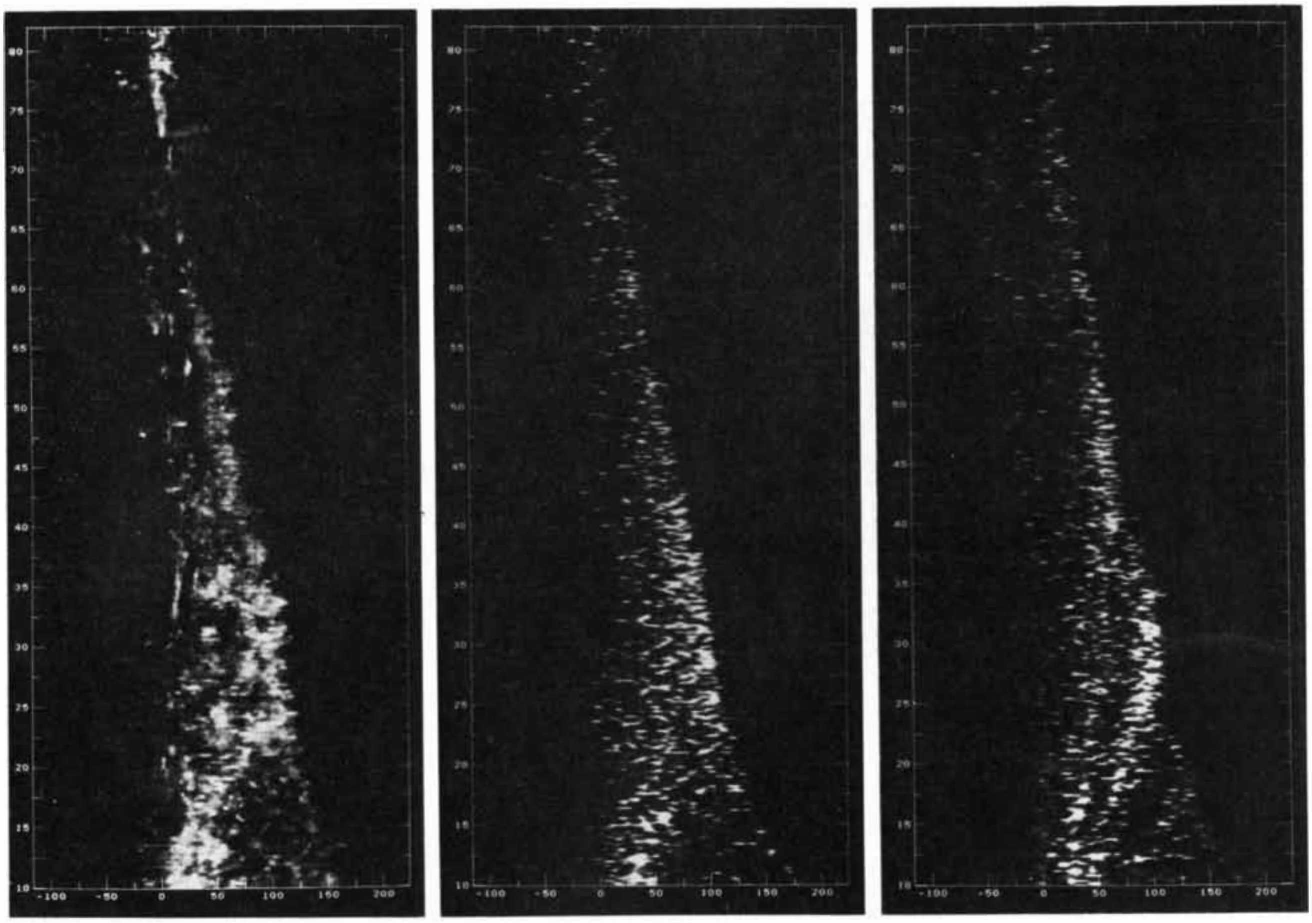
Another curious finding is that the thickness of the galaxy, at least the thickness measured by the dark-cloud layer, is not very great: about 300 light-years. Moreover, the thickness remains nearly constant over the entire width of the carbon monoxide ring. The thickness is determined by the equilibrium between gravitational compression and the dynamic and magnetic forces of expansion. Yet over a large range of cloud conditions and motions the thickness of the carbon monoxide layer apparently remains the same. What accounts for this uniformity?

As we have mentioned, our measurements place the solar system near the edge of the visible region of the galaxy, at the outer edge of the most active star-forming region. The dark clouds seen optically from this perspective represent only a small number of those in the inner galaxy. If the solar system were closer to the galactic nucleus, life on the earth might be imperiled, because the chances of the solar system's colliding with a dark cloud would be increased. If such a collision were to occur, the increase in the extinction of solar radiation between the sun and the earth would substantially decrease the amount of sunlight falling on the earth. In fact, some astronomers believe chance encounters with interstellar dust clouds caused ice ages on the earth. If this were the case, it would be ironic that the dust clouds required for the formation of stars and planets, which is presumably a precondition for the evolution of life, are also capable of causing the extinction of life.

Does life exist on the earth because the solar system is outside the main dusty regions of the galaxy or in spite of it? The rate of star formation appears to be much higher in the inner region of the galaxy, and so one would expect that the chance of life arising would be greater there. The chances of a collision with a dust cloud, however, are also greater. Which effect dominates and whether the effect is crucial are still subjects of speculation.

The new tool of carbon monoxide radio spectroscopy that we and others are exploiting to explore the galaxy is in some respects still primitive. Millimeter-wave astronomy is only about 10 years old. Most carbon monoxide data have been gathered by radio telescopes in the U.S. The next step is to make observations of the other half of the galaxy, a study that requires a millimeter-wave telescope located at a more southerly latitude. More sensitive receivers are also needed. The results of the efforts to date have sharply changed ideas about star formation and galactic evolution. As a result of discoveries in millimeter-wave astronomy the new field of astrochemistry has arisen. One can now investigate chemical processes involving a variety of molecules at the extraordinarily low temperatures and densities of the interstellar gas, conditions that cannot be duplicated in the laboratory.

Several countries are now building large millimeter-wave radio telescopes. The U.S. plans to build a national facility of this type in Hawaii in order to follow up on the discoveries of the present 11-meter telescope on Kitt Peak. Future observations of carbon monoxide and other interstellar molecules will surely lead to a fuller understanding of how stars form and how galaxies evolve.



THREE SETS OF SPECTROGRAMS are compared. The velocity-longitude diagram at the left is composed of a large number of actual radio spectrograms of galactic carbon monoxide. The other two diagrams were generated with the aid of a computer and are based on the authors' Monte Carlo models of the distribution of carbon monoxide clouds in the galaxy. The clouds represented by the diagram in the middle are distributed in a random but basically uniform way. The agreement between many aspects of the observations and the predictions of this computer model indicates that the values of such parameters as the number of clouds and their separation have been correctly chosen. The most conspicuous failure of this model is its inability to account for the observed higher-order clumpiness. In the model represented at the right, calculated by the authors in collaboration with W. W. Roberts, the distribution and motions of the clouds are assumed to be influenced by the gravitational effects of spiral arms. The fact that this model satisfies the observations better than the middle one suggests that in our galaxy star-forming regions describe a spiral pattern, a situation observed in external spiral galaxies.

The Spin of the Proton

When two protons collide, the outcome depends on which way the particles are spinning. The effect seen in violent collisions suggests there are objects in the proton that spin very rapidly

by Alan D. Krisch

All the fundamental particles of matter—the proton, the neutron and the electron—seem to be spinning perpetually. The spinning is much like that of a top, with one significant difference: there is no need for the particles to be wound up. Rotation is one of their intrinsic properties. Each particle has a fixed spin angular momentum in the same way that it has a definite mass and electric charge.

When two spinning particles collide, the outcome often depends on how they are spinning. For example, the path taken by a proton after a collision can be affected by the proton's spin just as the "english" applied to a billiard ball can alter the ball's trajectory. It has long been thought, however, that the influence of spin should decline as the energy of the collision increases. The reasoning behind this assumption is simple: the energy associated with a proton's spin is constant, and so it becomes an ever smaller fraction of the total energy as the collision becomes more violent. At a sufficiently high collision energy it should make no difference whether two colliding protons are spinning the same way or in opposite directions.

Only in the past few years have experimental techniques been devised for testing this assumption. It has turned out to be quite wrong. The influence of spin does not diminish as the energy of a collision increases; on the contrary, spin seems to become more important as the collision becomes more violent. A recent series of experiments has shown that protons spinning in the same direction are much more likely to rebound violently than protons spinning in opposite directions. Protons with opposite spins often seem to pass through each other without interacting at all.

The interpretation of these experiments is still uncertain. They seem to imply that inside the proton there are some smaller objects that carry most of the particle's spin angular momentum. The objects must be spinning very rapidly. Some years before these results were known it had already been proposed that the proton has an internal structure. One model suggested that each proton has a small, dense core. A theory popular today supposes every proton is made up of three of the small entities called quarks. Independent evidence has accumulated in support of the quark hypothesis. On the other hand, recent experiments suggest that the internal constituents of the proton have properties associated with their spin that are somewhat different from the properties predicted by the quark hypothesis. The differences are not easily reconciled.

The angular momentum of an object is proportional to its rotational velocity and is also influenced by the distribution of its mass. An elementary particle can have two kinds of angular momentum: orbital angular momentum and spin. Orbital angular momentum is most easily visualized in the planetary model of the atom proposed by Niels Bohr in 1913. The orbital angular momentum of an electron in such an atom is proportional to the velocity with which it revolves around the nucleus and to the radius of its orbit and to its mass. Orbital angular momentum also appears in any glancing collision between particles; even if the particles never complete a full circle, they still briefly revolve around their common center of mass.

Spin angular momentum measures the rotation of a single particle about its own internal axis. In quantum mechanics spin differs in a fundamental way from orbital angular momentum. A particle can gain or lose orbital angular momentum depending on its circumstances, as when an electron in an atom jumps from one energy level to another and hence from one orbit to another. Spin angular momentum, on the other hand, is a fixed property of each particle. The magnitude of the spin can be changed only by altering the identity of the particle itself.

Spin angular momentum is usually described as a vector, a quantity that has both a magnitude and a direction. The spin vector can be represented as an arrow parallel to the axis of rotation and with a length proportional to the magnitude of the spin. The direction of the arrow is defined by the arbitrary convention called the right-hand rule. If the fingers of the right hand are wrapped around the particle in the direction in which it is rotating, then the thumb indicates the direction of the spin vector. According to this convention, the spin angular momentum of the earth could be represented by a vector at the North Pole pointing up.

The concept of spin was introduced into the physics of elementary particles in 1925 by Samuel A. Goudsmit and George E. Uhlenbeck. In order to account for the splitting of certain lines in the spectra of atoms, they proposed that every electron has an intrinsic spin. Their proposal was made in the context of the quantum-mechanical theory of the atom, which was then just taking form. They assumed that both the magnitude and the direction of the spin angular momentum could take on only discrete values.

The magnitude of the spin vector is measured in units of $\hbar$, a symbol that is read "h bar" and that is equal to the Planck constant, h, divided by 2π; $\hbar$ is the most fundamental unit of angular momentum. In Bohr's 1913 model of the atom an electron could have only certain values of orbital angular momentum such as 0, $1\hbar$, $2\hbar$, $3\hbar$ and so on. Goudsmit amd Uhlenbeck suggested that the spin angular momentum of the electron is exactly $1/2\ \hbar$, a result that has since been confirmed by numerous experiments. Moreover, the proton and the neutron have been found to have the same spin as the electron.

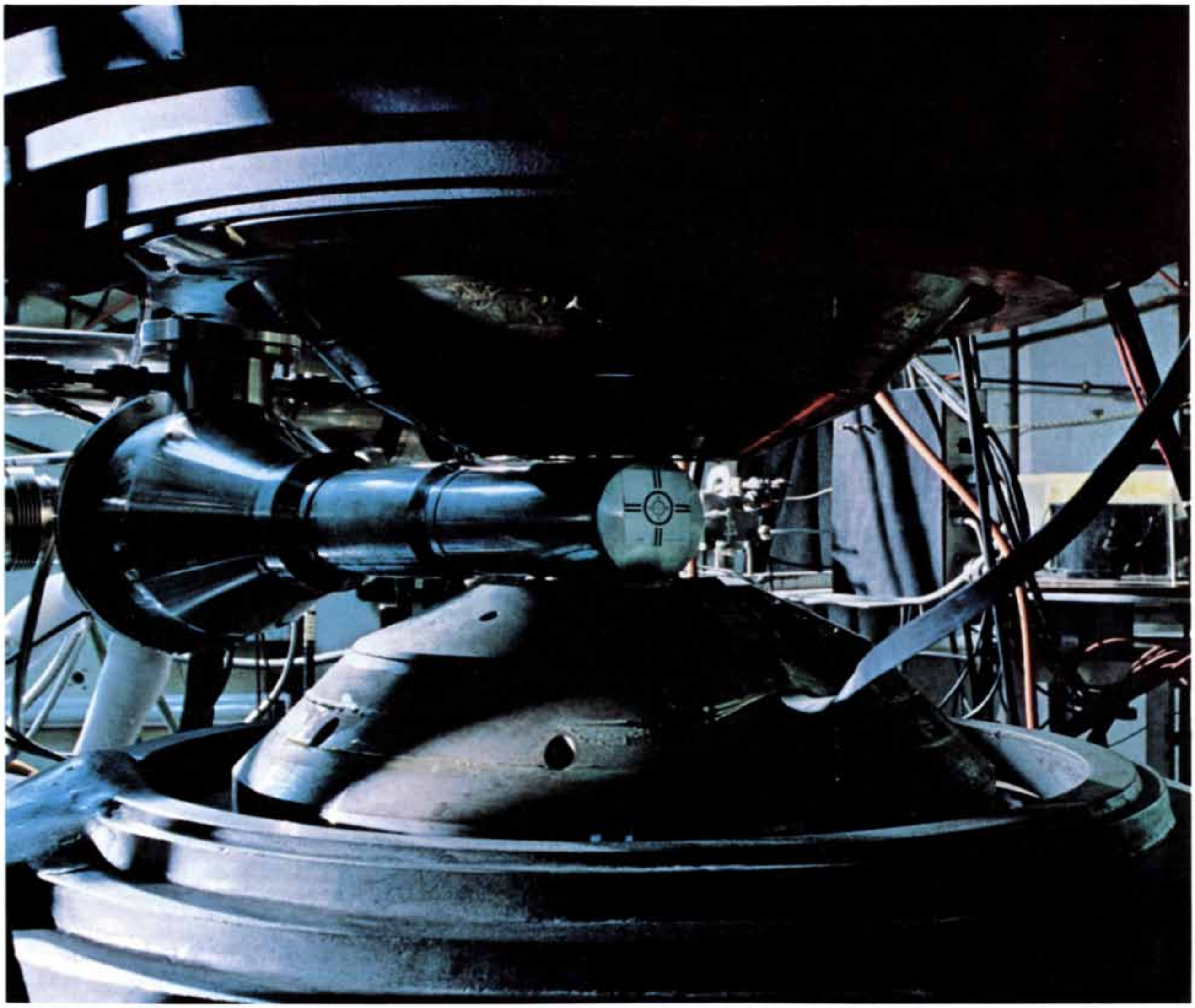
Just as the spin vector can have only a few possible magnitudes, so it can assume only a few possible directions. The number of allowed directions is related to the magnitude; for the proton, the neutron and the electron (and for all other particles with a spin of $1/2\ \hbar$) the vector can point in just two directions, which are always diametrically opposed. The two directions can be defined with respect to any suitable frame of reference, but it is customary to refer to them simply as "up" and "down."

Although spin angular momentum is well established as a quantum number and as a descriptive attribute of elementary particles, it is not known whether the particles are actually rotating in the same sense that a planet or a top rotates. The particles are too small for their rotation to be observed directly. The proton and the neutron have a radius of roughly one fermi, a unit of length equal to 10^{-13} centimeter. It is possible to imagine, at least, that such a particle might be spinning like an ordinary top. The electron and the neutrino, however, have the same spin of $1/2\ \hbar$ and are still smaller; indeed, they may even be point particles, with no size at all. The meaning of spin angular momentum in such objects is far from certain.

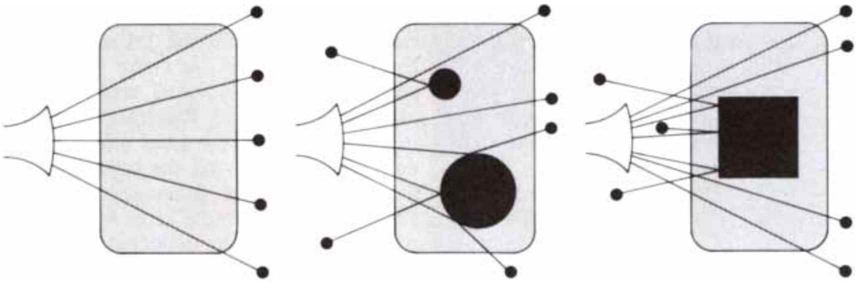
What has been demonstrated since 1925 is that in all observable properties the proton, the neutron and the electron behave exactly as if they were spinning with an angular momentum of $1/2\ \hbar$. Any more fundamental understanding must await experiments in which the effects of spin angular momentum are examined directly. The experiments I shall describe here may represent the first step in such a program.

The prototypical experiment in the physics of elementary particles was carried out in 1911 in the laboratory of Ernest Rutherford. A beam of alpha particles, or helium nuclei, was directed at a thin foil of gold. Particles scattered in collisions with the gold atoms were detected by observing the scintillations they caused on a fluorescent screen. From the unexpectedly large number of particles that were deflected through large angles, Rutherford deduced that the gold atom has a small, dense core: the atomic nucleus.

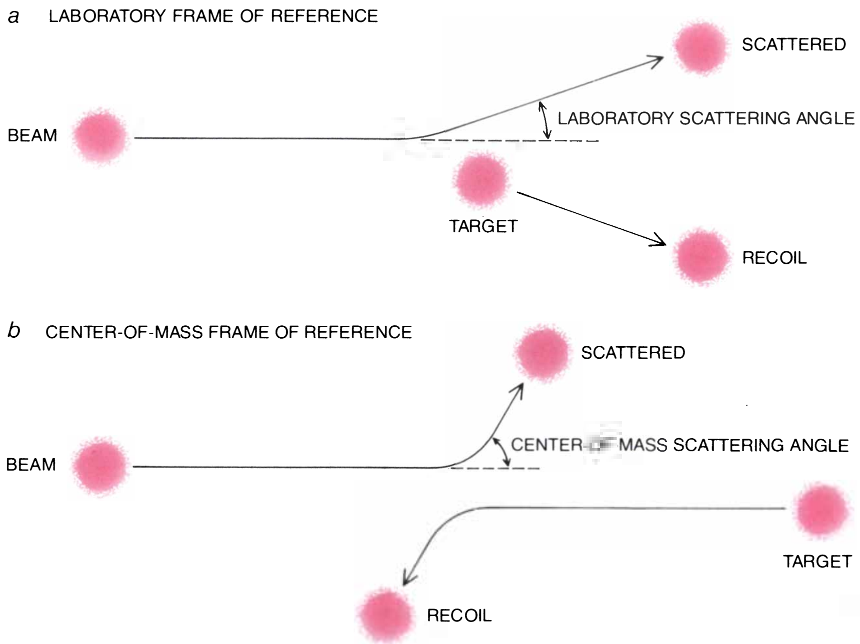
Most modern scattering experiments are identical with Rutherford's in conception; they differ only in scale. The apparatus has grown larger and the particles investigated have grown



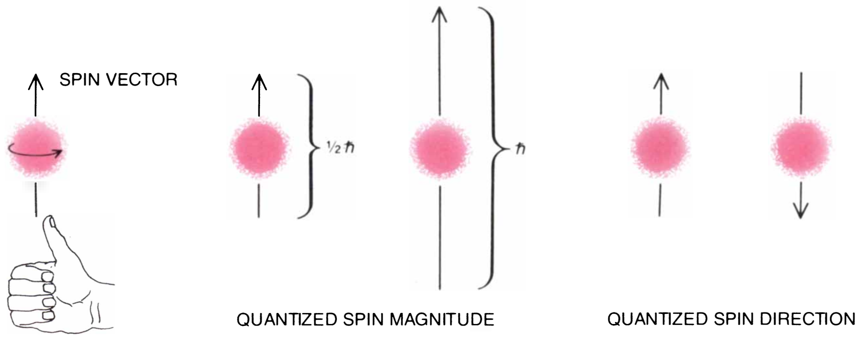
INFLUENCE OF SPIN on collisions between protons was investigated at the Argonne National Laboratory with this apparatus. The metal cylinder at the center encloses a target material in which protons are oriented so that most of them spin in the same direction; the protons are then said to be polarized. The target is struck by a beam of protons that are also polarized. The beam is accelerated by the Zero Gradient Synchrotron at Argonne and passes through the target from left to right. The target is placed between the poles of a 25-ton magnet, which is part of the system that aligns the spins of the target protons. The conical enlargement at the end of the cylinder holds a liquid-helium coolant. The target symbol near the tip of the cylinder is a surveyor's mark that aids in aiming the proton beam.



SCATTERING EXPERIMENTS probe the structure of the proton by a method that is illustrated here by means of an analogy. The proton can be regarded as being somewhat like a bale of cotton whose interior is inaccessible to direct examination. An experimenter might study the interior by shooting many bullets through the bale. If all the bullets pass through undeflected, there is probably nothing in the bale. A finding that some bullets are deflected would indicate there is a hard object inside. The number deflected would measure the object's size, and the angles at which the bullets bounce would provide information about its shape. In the Argonne experiments the "bullets" employed as probes of the proton were themselves protons.



ELASTIC SCATTERING takes place when two protons collide and rebound, carrying off all the collision energy, without creating new particles. In the laboratory frame of reference (*a*) a scattering event consists in an accelerated proton striking a stationary one. The trajectories are more symmetrical in another coordinate system (*b*), which moves with the center of mass of the two-particle system. In this moving frame the protons collide head on with equal speed. Scattering angle in the center-of-mass coordinate system measures the violence of a collision.



SPIN ANGULAR MOMENTUM is represented as a vector, a quantity with a magnitude and a direction. The spin vector is parallel to the axis of rotation and its direction is given by a right-hand rule: if the fingers of the right hand curl in the direction of rotation, the thumb points in the direction of the vector. For the proton the magnitude of the spin vector is $1/2\ \hbar$, where $\hbar$ is a unit of angular momentum equal to Planck's constant divided by 2π; for other particles the magnitude is different, but it is always a multiple of $1/2\ \hbar$. The direction of the vector is also quantized. The spin of the proton has two possible directions, "up" and "down."

smaller. The projectiles employed today are most often protons or electrons accelerated to energies conveniently measured in billions of electron volts, or GeV; some of the machines needed to reach those energies are a few miles in circumference. Through these accelerators it is possible to study not only the nucleus as a whole and the protons and neutrons that make up the nucleus but also the internal structure of the proton and the neutron.

It seems ironic that the largest of scientific instruments should be trained on the smallest of all objects, but there is a straightforward explanation. The uncertainty principle of quantum mechanics, formulated by Werner Heisenberg, establishes a relation between the precision with which an object can be measured and the energy available for the measurement. The principle states that the product of these two quantities is constant; if the precision is given in fermis and the energy in GeV, then the product is equal to .197. Suppose one were to attempt a study of the proton with a precision equal to 1 percent of the proton's radius. This precision is about .01 fermi, or 10^{-15} centimeter, and so the required energy is 19.7 GeV.

Actually the energy required of an accelerator in such an experiment is still greater. In Newtonian mechanics the kinetic energy of a moving particle is proportional to the square of its velocity. If the particle were stopped entirely by a collision, all its kinetic energy would be given up, but a proton colliding with a stationary proton does not stop. Instead the scattered and recoiling protons together carry off half of the available energy in their motion. The same effect can be observed when a moving automobile strikes a parked one: the collision is considerably less violent than a head-on one.

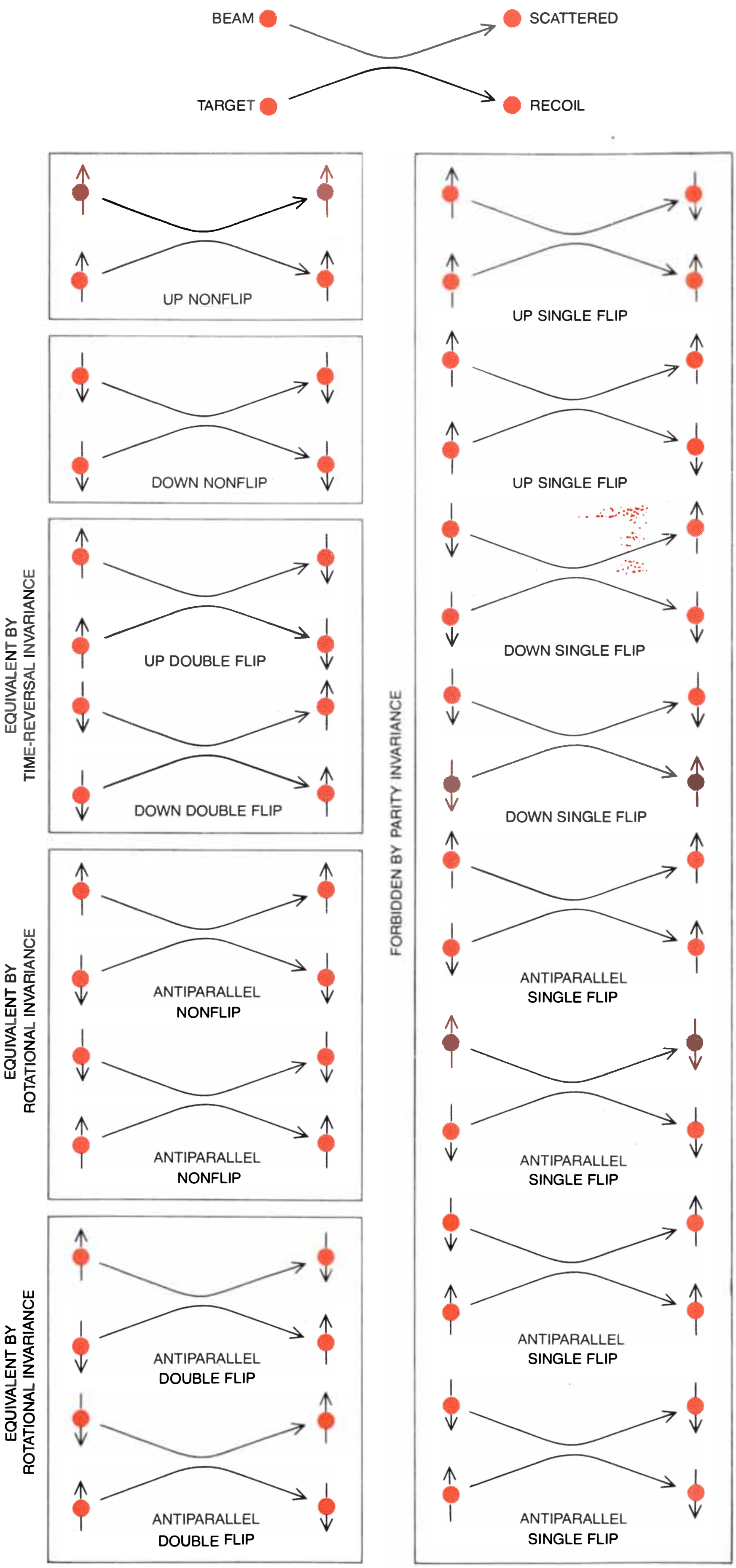
The analysis of these collisions can be simplified by shifting to another frame of reference. The coordinate system in which the collisions are actually observed is the laboratory frame of reference, where the observer and the accelerator are stationary. A more natural coordinate system is the one that moves along with the center of mass of the two particles. In this frame of reference each proton has half of the beam velocity; since a particle's energy is proportional to the square of its velocity, each proton has a fourth of the beam energy. It is then easy to see that only half of the beam energy (as measured in the laboratory frame of reference) is available for the collision.

The special theory of relativity adds a further complication. At beam energies higher than about 6 GeV the accelerated proton is already moving at more than 99 percent of the speed of light, and further increases in energy add little to the speed; instead they increase the effective mass of the particle. As a result the proportion of the beam energy available for the collision is sharply reduced. An appropriate analogy is the collision of a moving freight train with a stationary automobile: only a small portion of the energy invested in accelerating the train is made available in the collision. In fact, to attain a center-of-mass collision energy of 19.7 GeV would call for a beam energy of some 800 GeV, roughly double the energy available from the largest of the particle accelerators operating today.

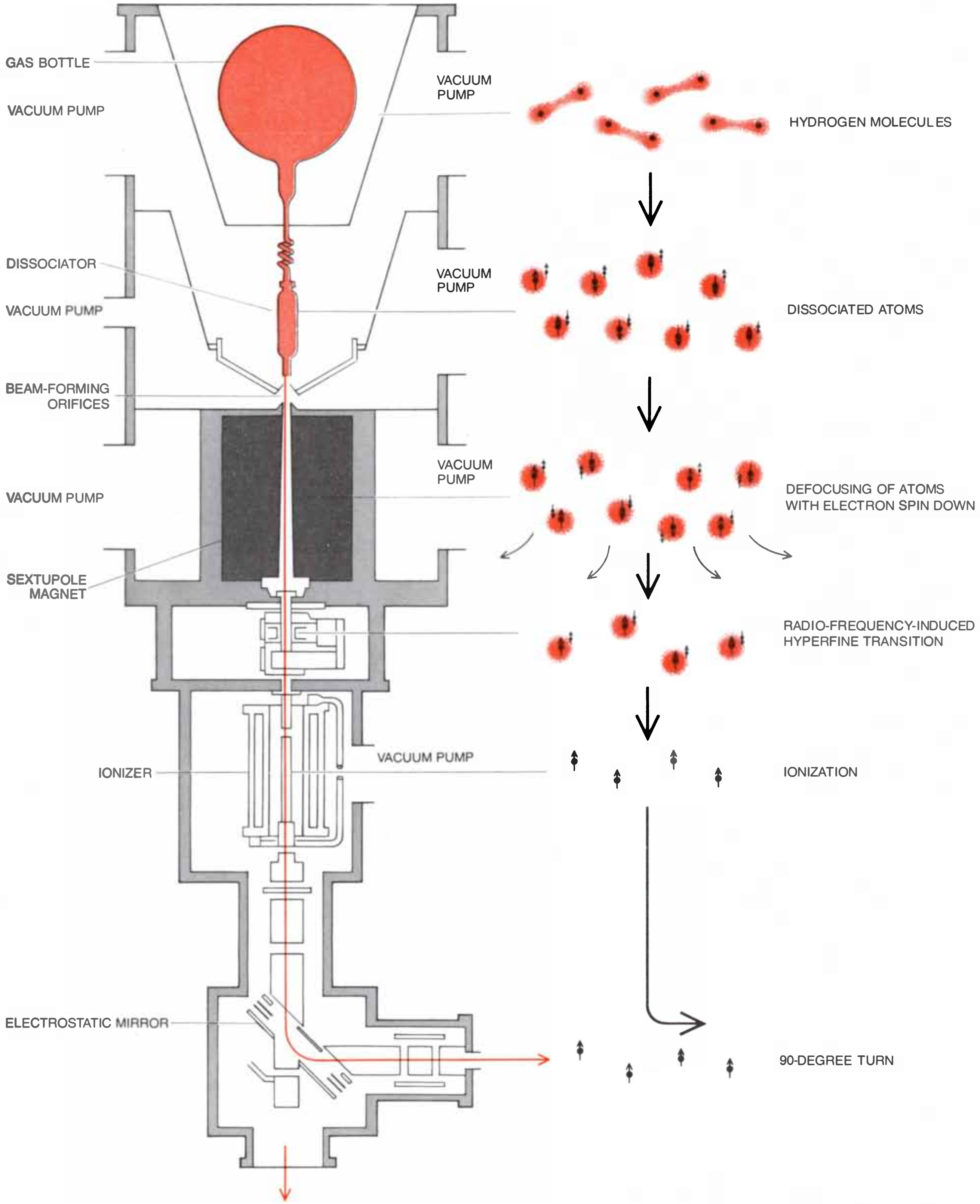
One factor besides the center-of-mass energy affects the resolving power of an accelerator experiment: it is the angle at which the particles scatter. Glancing collisions, in which there is little change in the particles' energy or direction, reveal only the long-range interaction between protons. The finest features of the proton's structure are observed only in the most violent collisions, where the two protons bounce apart with large angular deflections. By counting the number of protons scattered at each angle, the size and shape of the colliding protons can be estimated. Indeed, it was from just such an experiment that Rutherford deduced the size of the nucleus.

What is actually measured in a scattering experiment is the quantity called cross section. As the name suggests, a cross section can be expressed in units of area, but it is actually a measure of the probability that a collision will take place. For example, the experiments recently done by my colleagues and me measured the cross section for elastic scattering when two protons collide. A scattering event is said to be elastic if all the energy of the incident proton is carried off by the two rebounding protons; in inelastic events some of the energy goes into creating new particles. The cross section was measured at various scattering angles at an energy of 12 GeV. The experiments were conducted by a group of about a dozen physicists, mainly from the University of Michigan and the Argonne National Laboratory near Chicago. The accelerator we employed was the Zero Gradient Synchrotron (ZGS) at Argonne.

The cross section for proton-proton elastic scattering had been measured



PURE-SPIN CROSS SECTIONS measure the probability that two protons will scatter when their spin vectors point in specified directions before and after the event. There are 16 possible spin configurations, but eight correspond to events forbidden by the requirement that parity be conserved. Among the remaining eight configurations there are three pairs made up of cross sections that cannot be distinguished from each other because of symmetries, or invariance principles. Each pair can be considered a single cross section, and so there are five distinct pure-spin cross sections. Experiments at Argonne measured differences among these five cross sections.



SOURCE OF POLARIZED PROTONS for the Zero Gradient Synchrotron lines up proton spins by first polarizing electrons. Hydrogen molecules are dissociated into atoms, which emerge as a beam and enter the nonuniform field of a sextupole magnet. All the atoms whose electron spins happen to be oriented down are drawn toward the periphery of the field and are thereby eliminated from the beam. In the remaining atoms all the electron spins are up and the proton spins are oriented randomly, half up and half down. The atoms are next exposed to radio-frequency radiation tuned so that it can be absorbed only by atoms in which the electron and proton spins are opposite. The absorption of the radiation flips both spins, an event called a hyperfine transition. After the double spin flip all the proton spins are up and the electron spins are random (half up and half down). Finally the electrons are stripped away and the resulting beam of polarized protons is reflected through a 90-degree angle, so that the spin vectors are perpendicular to the direction of beam motion. Contamination by stray atoms would spoil the polarization; therefore the beam source requires an elaborate system of vacuum pumps.

many times before, and at energies higher than those available at the ZGS. What distinguished our experiments was that we knew the directions of the spin vectors of both the incident and the target protons before they scattered. We could therefore measure a separate cross section for each combination of spins. In earlier experiments the spins were orientated randomly, and information about their influence could not be recovered from the data recorded.

Just two protons participate in an elastic-scattering event, but four spins must be considered. They are the spins of the beam, target, scattered and recoil protons. The beam and scattered protons are in fact the same particle, and so are the target and recoil protons; their spins must be determined separately, however, because during the collision they can change direction, or in other words make a 180-degree flip.

Since there are four spins and each spin can have either one of two possible directions, there are 2^4, or 16, possible combinations of spins. The 16 corresponding pure-spin cross sections are the fundamental cross sections for proton-proton elastic scattering. If a theory that accurately describes the interactions between protons is ever developed, it is these pure-spin cross sections it will predict. Experiments that fail to distinguish the spin states of the proton measure only a complicated average of all the cross sections.

Among the 16 pure-spin cross sections some can be eliminated and others consolidated. Eight of the cross sections are for events in which a single spin flips. If the two scattering protons make up an isolated system (as they must in elastic-scattering experiments), then all single-flip events are forbidden by a conservation law. The cross sections for these events must therefore be exactly zero; the events are not observed.

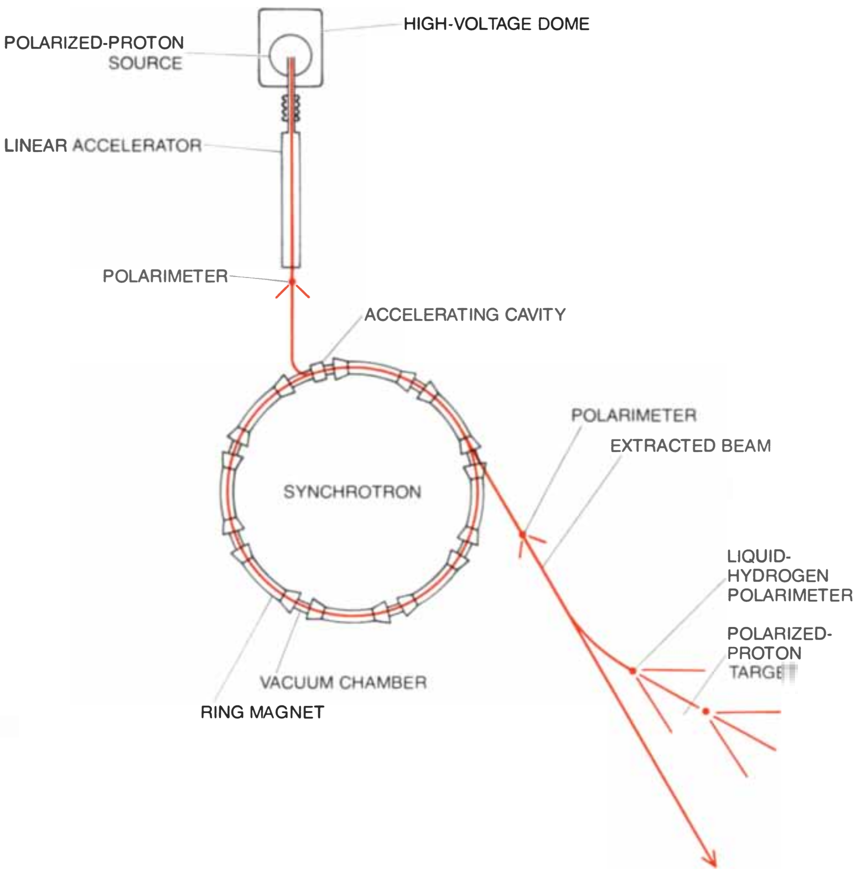
Among the remaining eight cross sections not all are experimentally distinguishable from one another. Two of them are equivalent by the postulate of time-reversal invariance, which states that any microscopic physical process must remain unchanged when the direction of time is reversed. Two more pairs of cross sections are equivalent by rotational invariance: one member of each pair is made identical with the other member by turning the entire laboratory upside down.

The grouping together of indistinguishable events leaves five distinct cross sections. In a convenient shorthand they are designated up nonflip, down nonflip, parallel double flip, antiparallel nonflip and antiparallel double flip [*see illustration on page 71*]. It might seem that the first two cross sections, which respectively represent events with all four spins up and all four spins down, could also be consolidated by rotational invariance. Indeed, these cross sections are very similar, but they are distinct. Because spin angular momentum has an effect somewhat like that of english on a billiard ball, the up-nonflip and down-nonflip scattering events are both asymmetrical. Turning the laboratory upside down would not transform one into the other.

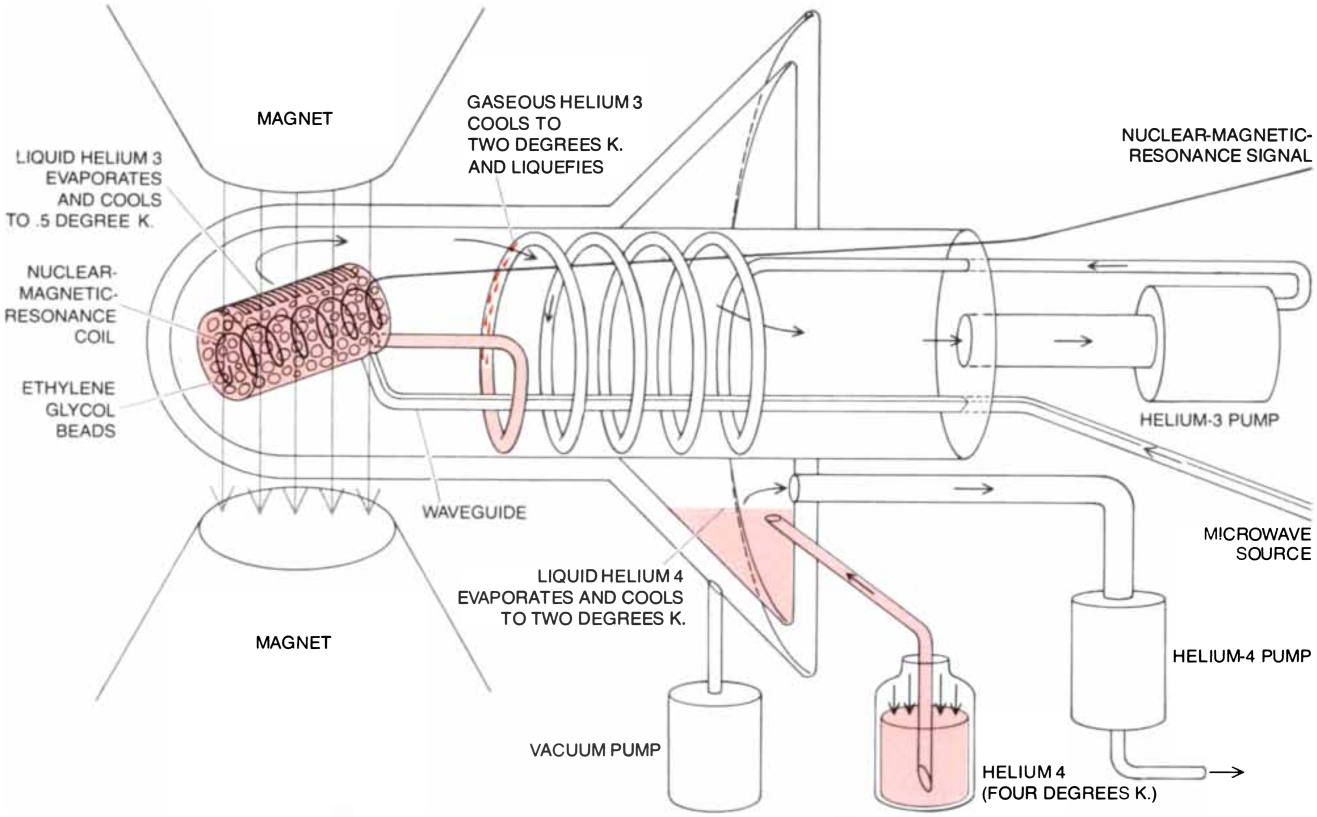
In practice it is not always necessary to distinguish even these five cross sections. The most important distinction is apparently between the events in which the beam proton and the target proton are spinning the same way and those in which their spins are opposite. These two classes of events are conveniently noted by the terms parallel and antiparallel. Many of the results I shall present here observe only this distinction.

An experimental investigation of the influence of spin on proton-proton scattering requires both a beam and a target that consist of protons whose spins are all or almost all lined up the same way. The beam and the target are then said to be polarized. (The borrowing of this term from optics is entirely legitimate, since the polarizing of light consists in lining up the spins of photons.) For now the Argonne Zero Gradient Synchrotron is the only accelerator in the world where both a polarized proton beam of high energy and a polarized proton target are available.

The only practical way to get a "grip" on a particle and control its spin direction is through the particle's magnetic moment, which is closely connected with its spin and electric charge. In an external magnetic field the magnetic moment of the proton gives rise to a small force that tends to align the spin vector along the field direction. Thus a direct means of polarizing protons is by the brute-force application of a strong magnetic field, but it is not the best means. The magnetic moment of the proton is exceedingly small and the magnetic field needed to polarize a large



ZERO GRADIENT SYNCHROTRON accelerates polarized protons to an energy of 12 billion electron volts (GeV). The polarized-proton source is enclosed in a high-voltage dome; protons leaving the dome acquire an energy of 750,000 electron volts. A linear accelerator then raises their energy to 50 million electron volts before they are injected into the synchrotron itself. On each pass through the accelerating cavity of the synchrotron the particles gain 12,000 electron volts; in order to reach 12 GeV they must therefore circle the ring about a million times, which requires a little less than a second. The accelerator is called the zero gradient synchrotron because the fields of the ring magnets are uniform, a feature that aids in maintaining beam polarization. When the particles reach 12 GeV, they are extracted from the ring and directed by steering magnets to the target. Polarization is monitored just before injection and just after extraction by polarimeters that measure the left-right asymmetry in scattering events. The asymmetry is related to the spin orientation and hence to the beam polarization.



TARGET OF POLARIZED PROTONS is cooled to a temperature of .5 degree Kelvin, placed in a magnetic field of 25,000 gauss and exposed to microwave radiation that has a frequency of 70 gigahertz. The refrigeration is supplied by two isotopes of helium. The common isotope helium 4 cools the entire apparatus to two degrees K., and further cooling of the target itself is provided by the light isotope helium 3. Both cooling systems employ vacuum pumps to reduce the temperature of the liquids by speeding their evaporation. The microwave energy is brought to the target by a waveguide. A coil inside the target canister measures the spin polarization of the protons by the method called nuclear magnetic resonance. The target material is ethylene glycol doped with chromium, but only the protons of certain hydrogen atoms are polarized. At the temperature of the target the ethylene glycol is a solid, and it is formed into small beads.

fraction of the protons in a specimen of matter would be impractically large.

At the ZGS a more complicated but also more efficient procedure is employed in an ingenious polarized-proton source constructed by the ANAC Corporation of New Zealand. The basic principle of operation is to act on the magnetic moment of the electron, which is some 660 times larger than that of the proton. The electrons are polarized and then their spin alignment is transferred to the protons.

The protons begin as nuclei in molecules of hydrogen gas (H_2), which are first dissociated into atoms. In each atom both the proton and the electron have a spin with two possible orientations, and so there are four spin combinations. The atoms emerge from a nozzle as a low-energy beam, then pass into the strongly nonuniform field of a sextupole, or six-pole, magnet. The field has little direct effect on the proton spins, but it discriminates strongly between spin-up and spin-down electrons. The atoms with spin-up electrons are repelled by the field and are therefore focused along the central axis of the magnet, where the field is weakest. The atoms with spin-down electrons are drawn to the region of stronger field at the periphery and are thereby removed from the beam.

Emerging from the sextupole magnet is a beam of atoms with the electron spins fully polarized and the proton spins completely unpolarized, that is, half up and half down. The next stage polarizes the protons and depolarizes the electrons. That is done by flipping both the electron and the proton spins in one population of atoms, those in which the spins are antiparallel. The selective spin flipping is accomplished by applying a uniform magnetic field and irradiating the atoms at a frequency of 1.5 gigahertz (billions of cycles per second). This frequency is absorbed only by those atoms in which the electron is spin up and the proton is spin down, and it induces a double spin flip called a hyperfine transition. After the transition the electron spin is down and the proton spin is up. The only other atoms present have both electron and proton spin up, and they are unaffected by radiation with this frequency.

The proton spins are now polarized and the electrons are depolarized. In the final stages of the procedure the electrons are stripped away and the protons are accelerated to an energy of roughly 20,000 electron volts. The ANAC source emits some 10^{11} protons per second with a polarization of about .75. The polarization is defined as the difference between the number of spin-up particles and the number of spin-down ones, divided by the total number of protons. Hence a polarization of .75 implies that the ratio of spin-up to spin-down protons is 7 to 1.

The subsequent acceleration of the protons takes place in three stages. The source itself is mounted inside a dome maintained at a potential of 750,000 volts; as the protons leave the dome they therefore acquire an energy of 750,000 electron volts. The next stage is a linear accelerator 200 feet long, where the protons are brought to an energy of 50 MeV (million electron volts). Finally they are injected into the ring of the main proton synchrotron.

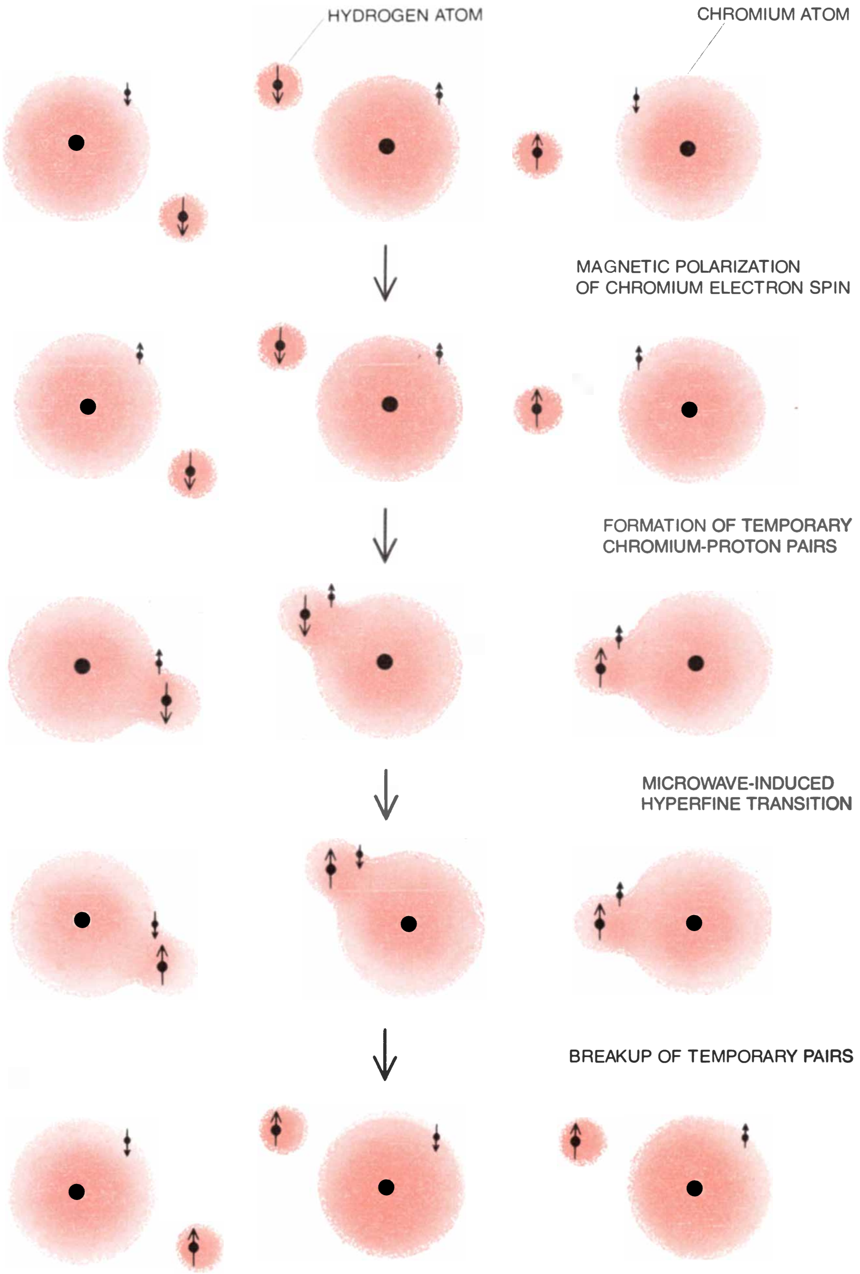
The accelerator at Argonne is called the zero gradient synchrotron because the magnetic fields that confine the particles are all uniform fields; the ZGS lacks the quadrupole and sextupole magnets that are employed for focusing in many other accelerators. As in all other synchrotrons, the actual work of acceleration is done only in a comparatively small cavity, where a radio-frequency field sets up a potential difference of some 12,000 volts. The rest of the machine, which has a circumference of almost 500 feet, consists of magnets that bend the trajectories of the protons to a circular path. In this way each proton can pass through the accelerating cavity many times, receiving a 12,000-electron-volt boost in energy each time. To reach 12 GeV a proton must make about a million turns around the ZGS.

The ZGS was not designed for the acceleration of polarized protons, but the unusual uniformity of its magnetic fields simplifies the task. The proton spins are oriented vertically, and any horizontal fields can depolarize them. Even in the ZGS there are weak horizontal fields, but after a million revolutions they cause a depolarization of only about 1 percent. Of more serious concern are depolarizing "resonances" encountered at certain energies where the small depolarizing force acts in the same direction on each turn and so is reinforced coherently. When these coherent forces add in phase, they can quickly obliterate the polarization of the beam. The resonances have been overcome by installing two quadrupole magnets and a horizontal magnet in the accelerator. The magnets are turned on abruptly just before each resonance energy is reached and remain on for only .003 second. They slightly change the path of the beam, enabling the protons to jump quickly through the resonance energies. Such precisely timed corrections must be applied for 29 resonances in each acceleration cycle. The technique of "resonance jumping," which was first applied at the ZGS, has now been so carefully refined that protons can be accelerated to 12 GeV while maintaining a polarization of .70.

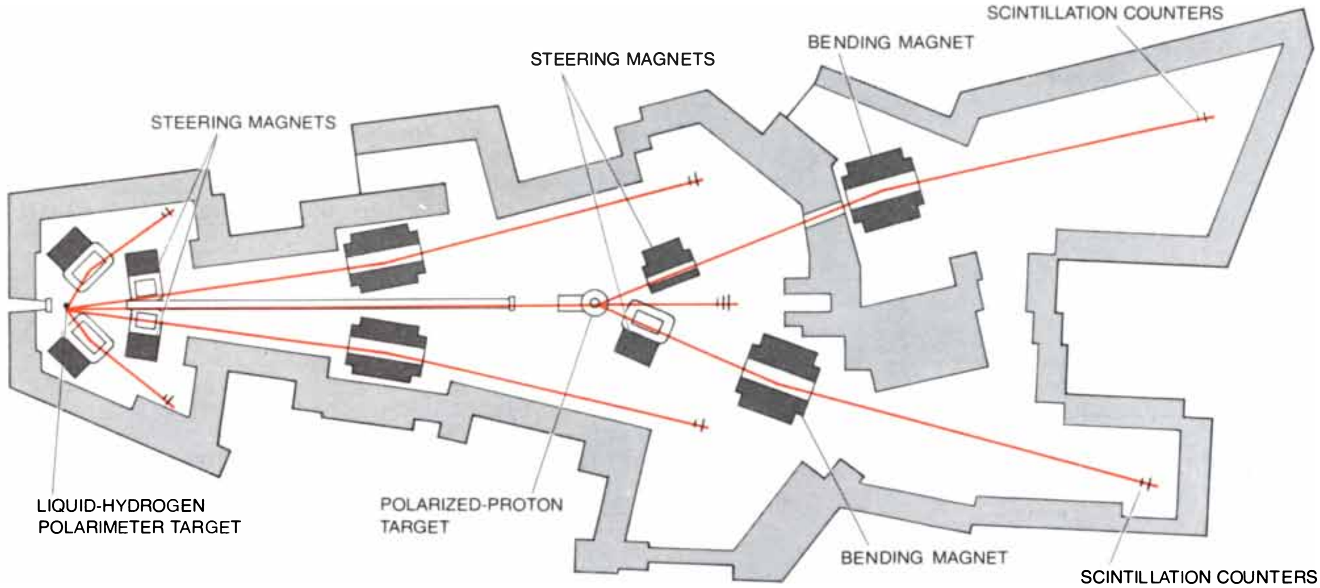
The polarized target works in much the same way that the source of the polarized beam does. The polarization is initiated by a strong magnetic field that acts on electrons, which are easily polarized because of their large magnetic moment. This method of spin polarization was developed in the late 1950's by two low-temperature physicists, Anatole Abragam of the Collège de France and Carson D. Jeffries of the University of California at Berkeley. It was applied to experiments in particle physics at Berkeley, at the European Organization for Nuclear Research (CERN) and at Argonne. In adapting the technique to particle physics Owen Chamberlain of Berkeley and Michel Borghini of CERN made important contributions.

The polarization of the target protons is directly proportional to the applied magnetic field and to the magnetic moment of the proton; it is inversely proportional to the absolute temperature. In our target the applied field is 25,000 gauss and the temperature is .5 degree Kelvin (.5 degree Celsius above absolute zero). Under these conditions protons alone would have a polarization of only .003; electrons, on the other hand, have a polarization of .995.

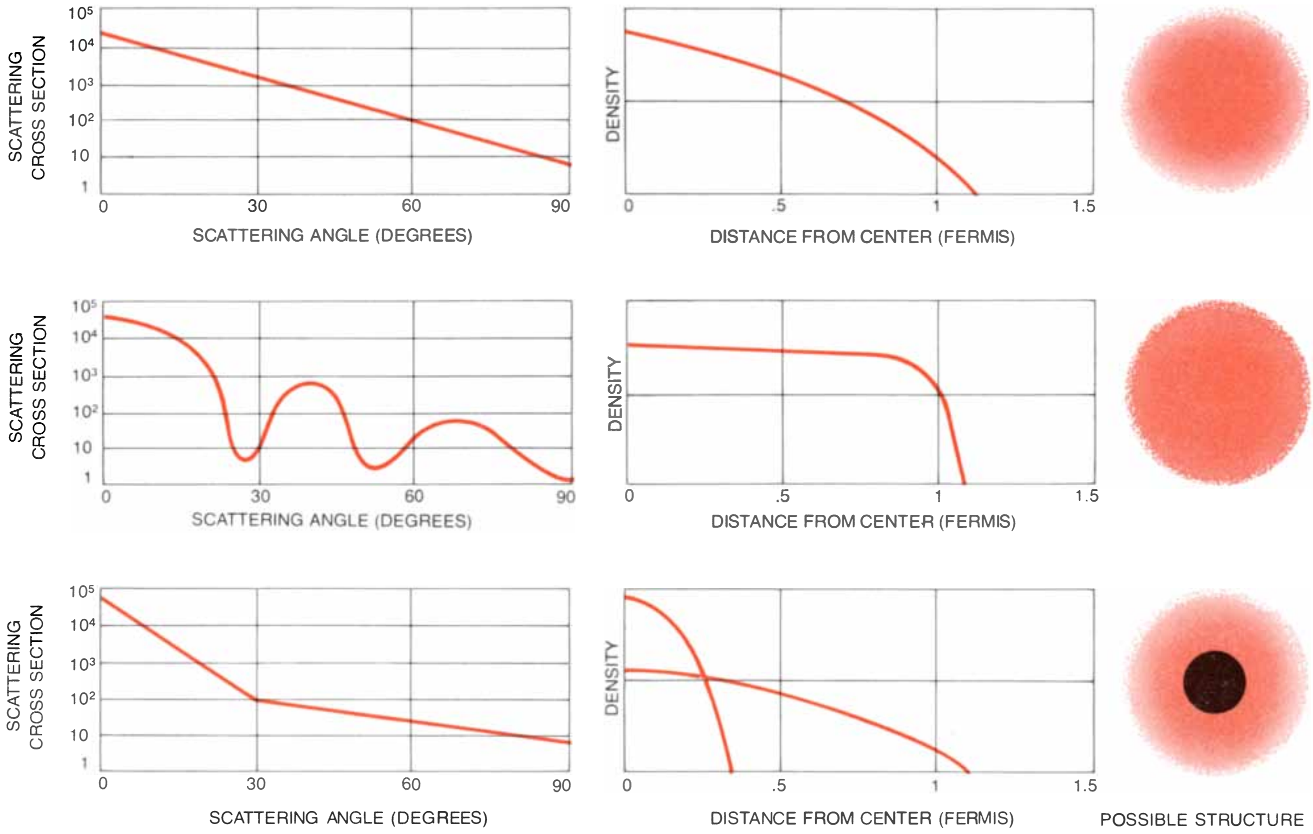
The protons to be polarized are the nuclei of hydrogen atoms in an organic substance such as ethylene glycol. (Ironically ethylene glycol is a common antifreeze, but in our apparatus it is frozen into small beads.) What the applied magnetic field actually polarizes are the unpaired electrons of chromium atoms, which can form temporary pairs with the hydrogen protons. In each chromium-hydrogen pair there are only two possible spin states: electron up, proton up, and electron up, proton down. A microwave field with a frequency of 70 gigahertz then flips both spins in the lat-



MECHANISM OF TARGET POLARIZATION is indirect, as in the beam source: first the spins of electrons are lined up, then the polarization is transferred to protons. This method is adopted because the electron has a magnetic moment some 660 times larger than that of the proton and so is more easily polarized. The electrons in question are unpaired electrons of chromium atoms, which are aligned in the spin-up configuration by the strong magnetic field that permeates the target. The chromium electrons and the protons of hydrogen atoms form temporary pairs in which the electron spins are all pointing up but the proton spins are random. A hyperfine transition induced by microwave radiation then flips the spins of both the proton and the electron, but only in those pairs where the two spins are opposite. When the pairs break up, all the protons are left in the spin-up state and the electrons are depolarized. The process is continuous: the electron spins are realigned and again transfer polarization to nearby protons.



DOUBLE-ARM SPECTROMETER measures the angle and the energy of both the scattered and the recoil protons in the aftermath of a collision. The beam first strikes a target of unpolarized liquid hydrogen, and scattering events there are monitored for a left-right asymmetry that measures the beam polarization. At the polarized target, protons scattering at selected angles pass through a series of magnetic fields and are ultimately detected by scintillation counters. The experiment measures the probability that a proton will scatter at each angle between zero degrees and 90 degrees for each possible spin configuration. Instead of moving the detectors to intercept particles scattered at various angles, steering magnets deflect the protons to stationary detectors. The bending magnets measure the protons' energy.



POSSIBLE CROSS SECTIONS for proton-proton elastic scattering are plotted as a function of scattering angle. The cross section at any given angle is a measure of the probability that a proton will be deflected through that angle by a collision. All cross sections decline sharply as the angle increases, which merely signifies that a proton is much more likely to undergo a small deflection than a large one. The form of the curve carries information about the internal structure of the proton. A smooth decline in the cross section (*top*) would suggest that the density of the proton drops off continuously from the center to the surface; it would describe a "fuzzy" particle. A cross section with several pronounced peaks and troughs (*middle*) would be recorded if the proton were a hard-edged sphere, in which the density is constant from the center to the surface and then drops quickly to zero. A cross section with two components separated by a discontinuity where the slope of the curve changes (*bottom*) would correspond to a structure with two distinct regions: a dense core and a more diffuse peripheral cloud. The actual elastic-scattering cross section of the proton seems to resemble the two-component curve.

ter pairs, depolarizing the electrons and polarizing all the protons in the spin-up condition. The bandwidth of the microwave signal must be narrow enough to avoid the frequency 70.2 gigahertz, which stimulates the equivalent double-flip transition in the other spin pairs.

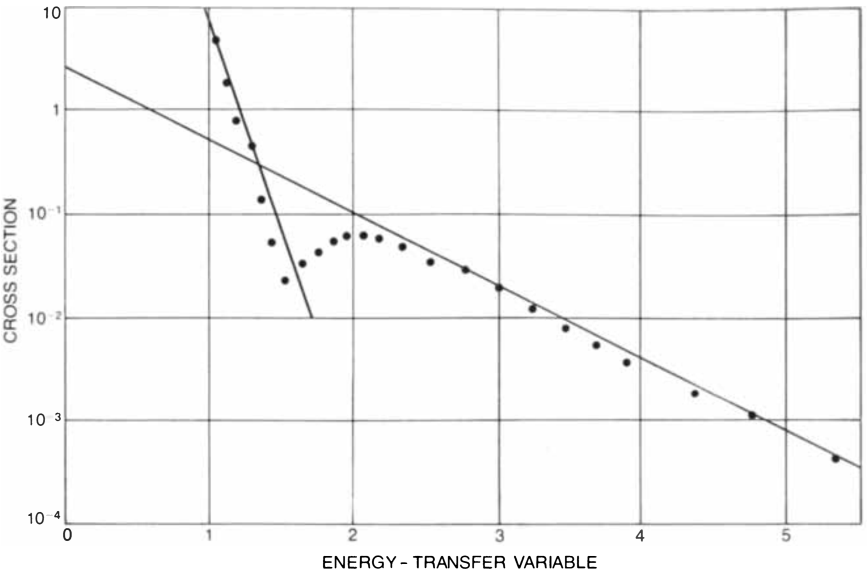
Cooling is provided by the rare helium isotope helium 3, which circulates among the ethylene glycol beads. The temperature of the helium 3 is maintained at .5 degree K. by a large vacuum pump that reduces the vapor pressure of the helium and thereby accelerates its evaporation. The entire target, with its helium-3 refrigerator, is immersed in a vessel of helium 4, the ordinary isotope, which is cooled to two degrees K. by similar means.

The polarization of the Argonne target has been as high as .85, but that value cannot be sustained when the beam is operating because of radiation damage to the ethylene glycol beads. An average polarization of about .7 can be achieved by annealing the beads at a temperature of 160 degrees K. every eight hours and by replacing them every few days.

An essential component of any scattering experiment, whether or not it employs polarized particles, is a spectrometer: a system of detectors and bending magnets arranged to measure the position and the energy of the scattered particles. We employed a double-arm spectrometer: one with detectors for both the scattered protons and the recoil protons. It consists of four magnets, which bend the protons' paths and thereby help to determine their energy, and six scintillation counters, which detect the angular position of a particle in the same way that Rutherford's fluorescent screen did.

The greatest resolution is provided by the arm that intercepts the scattered particle. The scattering angle is measured with a precision of about .5 degree and the energy of the scattered proton is determined with a precision of about 7 percent. The resolution of the other counters is not as great, but they serve the important function of identifying extraneous events so that they can be rejected. These background events include accidental or coincidental events, in which particles are detected from two independent but simultaneous collisions. Other background signals come from inelastic collisions and from scattering in the carbon and oxygen nuclei of the ethylene glycol.

The usual procedure in such an experiment is to measure a cross section at one angle and then move the arms of the spectrometer along an arc centered on the target to the next angular position. The movement of the apparatus is particularly inconvenient and time-consuming in our experiment. (The roof must be removed from the building.)



ELASTIC-SCATTERING CROSS SECTION was measured with unpolarized protons at the European Organization for Nuclear Research (CERN) near Geneva at an energy equivalent to 1,500 GeV. Because the spins of the colliding particles were not controlled the observed cross section is an average of all the pure-spin cross sections. The data are plotted as a function of an energy-transfer variable, which measures the violence of the scattering and is related to the scattering angle. The curve can clearly be resolved into two components with different slopes. The steep upper portion reflects glancing collisions that involve only a diffuse outer layer of the proton, whereas the more gradual decline of the lower segment suggests violent collisions between small, hard objects inside the proton. The objects seem to have a radius of about one-third fermi (one fermi being equal to 10^{-13} centimeter). The trough between the two segments of the curve could represent interference between the two modes of scattering.

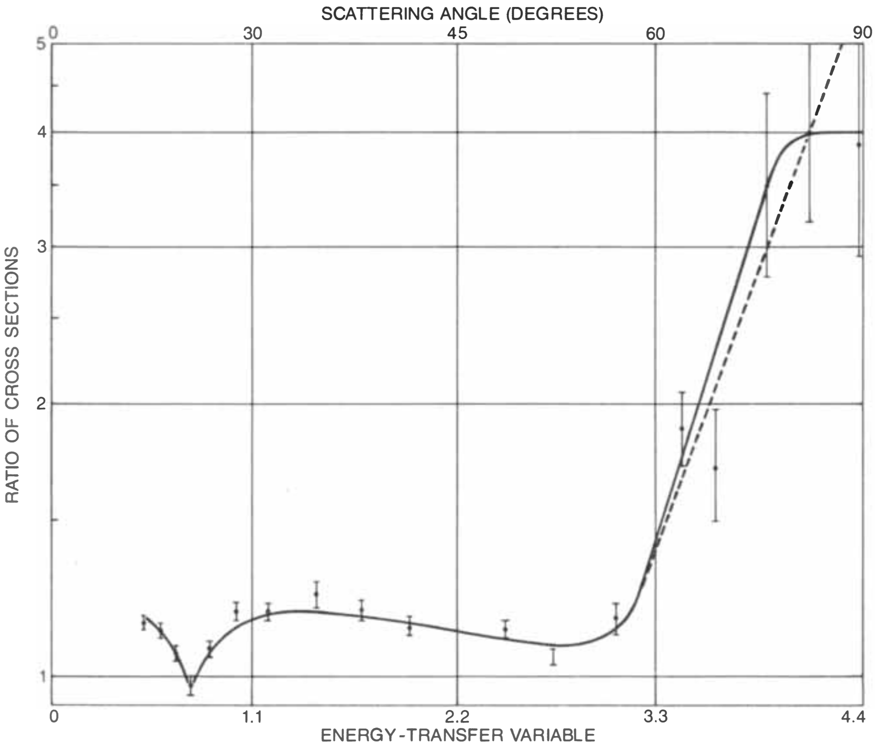
We simplified the angle changes by placing a bending magnet with a large aperture near the target in each arm of the spectrometer. The bending magnet captures the scattered particles and steers them to the fixed scintillation counters. A large range of angles can therefore be observed without moving any part of the apparatus; only the current in the magnet has to be changed.

We counted events at scattering angles ranging from zero degrees to 90 degrees, measured with respect to the beam direction in the center-of-mass coordinate system. (The angles observed in the laboratory frame of reference are smaller.) The protons can scatter at angles larger than 90 degrees; indeed, the most violent collisions are those at 180 degrees, when the scattered proton bounces directly back along the beam axis. Because all protons are identical, however, it is not possible to determine which is the scattered proton and which is the recoil proton. In order to avoid this ambiguity we classify scattering events according to the angle of the forward proton, which must always be 90 degrees or less.

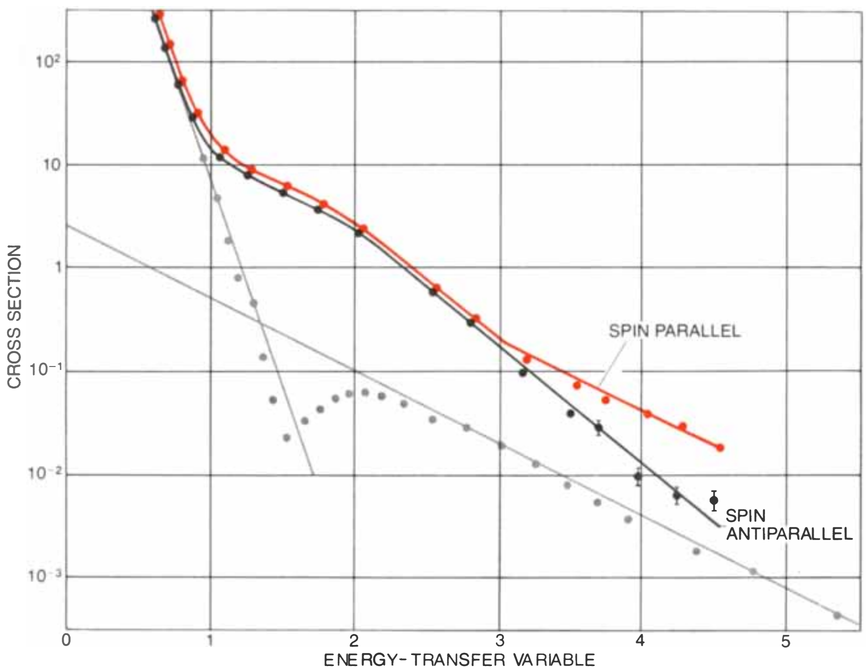
The experiment ran 24 hours a day for about six months, recording one or two events per hour at large scattering angles. The spin of the beam protons was reversed with each cycle of the synchrotron (about once every three seconds), and the polarization of the target was reversed each time the beads were annealed or replaced (about once every eight hours). Tests were also conducted with unpolarized beams, with unpolarized targets and with targets that included no hydrogen. The consistency of the results obtained under these varied conditions gave assurance there was no systematic bias in the apparatus. The polarizations of the beam and of the target were also monitored continuously. Because the polarizations were always less than 1 we corrected the recorded data to derive pure-spin cross sections.

As I pointed out above, a cross section measured at some definite angle represents the probability that two colliding protons will scatter elastically at that angle. The variation in the cross section as the angle is changed yields information about the size, shape and internal structure of the proton. Only the violent collisions that give rise to large-angle scattering can probe the structure of the proton at the smallest scale. It is the large-angle scattering events that will reveal the presence of any hard objects inside the proton.

All the data collected in an elastic-scattering experiment with unpolarized protons can be reduced to a single graph, in which the cross section labels one axis and the scattering angle labels the other. When pure-spin cross sections are measured, a separate graph must be plotted for each spin state. Information about the structure of the proton is ex-



FINDINGS OF THE MICHIGAN-ARGONNE STUDY of polarized-proton scattering are presented in the form of a ratio: the cross section measured when the spins of the beam and the target protons were parallel divided by the cross section when they were antiparallel. It is notable that the ratio is greater than 1 at all scattering angles with one possible exception; hence protons are always more likely to interact when they are spinning the same way. The abrupt increase in the ratio at an angle of about 60 degrees indicates that spin is more important in violent collisions, which presumably involve the hard objects in the proton. From the data it is not clear whether the ratio continues to climb at larger scattering angles or levels off at a value of 4.



EFFECTS OF SPIN on proton-proton scattering are also apparent when the cross sections for spin-parallel events (*color*) and spin-antiparallel events (*black*) are plotted separately. The two cross sections differ only slightly for small-angle, glancing collisions but diverge at large scattering angles. The data are consistent with the hypothesis that there is no discontinuity at all in the cross section for spin-antiparallel scattering, whereas the spin-parallel cross section has a conspicuous change in slope. At large values of the energy-transfer variable the slope of the spin-parallel curve is the same as that of the 1,500-GeV measurements made at CERN (*gray*).

tracted from these graphs by the mathematical procedure called Fourier analysis. The details of the technique will not be given here, but some of the conclusions that follow from it are intuitively clear.

One form the cross-section graph might conceivably take is that of a straight line descending from a large cross section at zero degrees to some minimum value at 90 degrees. Such a featureless graph would suggest that the proton is a soft and smooth sphere with no abrupt boundaries and no hard objects inside it. The density would decline smoothly from a maximum value at the center to zero at a radius of about one fermi.

Another possibility is a curve with the same overall trend but with pronounced peaks and troughs superposed on it. Such a graph would be analogous to the diffraction pattern formed when a beam of light is intercepted by a hard edge; the Fourier analysis of the graph describes the proton as a hard-edged sphere. The density is essentially constant from the center to a boundary near the one-fermi radius; then it falls quickly to zero.

A third possible graph is best described as a composite of two distinct curves. At small scattering angles (say less than 60 degrees) the cross section falls off smoothly but quickly. At 60 degrees there is a discontinuity in the curve, and at all larger angles the cross section declines more slowly. Such a two-component curve suggests that the proton has two regions, the inner region being dense and the outer one diffuse; the discontinuity between the two curves indicates that the two regions are clearly defined. The size of each region can be calculated from the rate at which the cross section declines. The actual cross-section graph for proton-proton elastic scattering has a form much like this third example, with two curves separated by an abrupt change in slope.

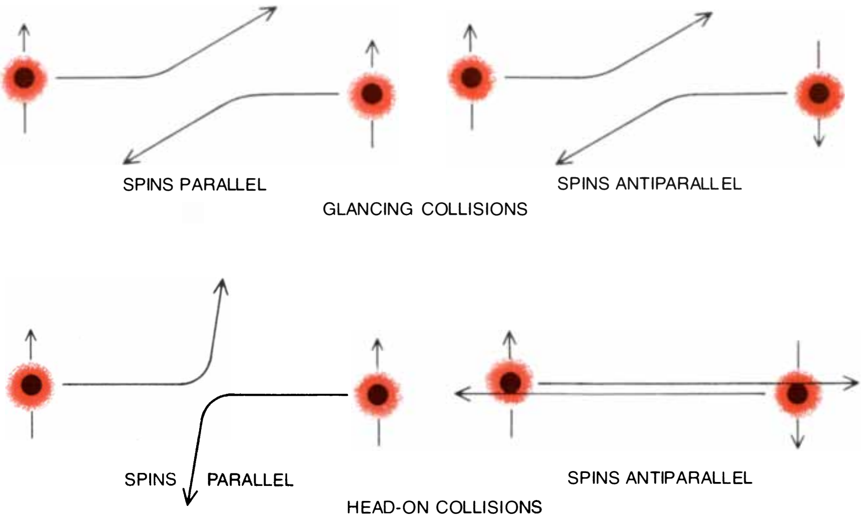
In 1963 I suggested that the proton has a "core" surrounded by a "cloud" of lesser density. Indeed, there might be several layers of decreasing density, and so my proposal has come to be known as the onion model. The first experimental evidence for an internal structure of the proton was obtained in a 1966 experiment with unpolarized protons carried out at the ZGS by me and my colleagues from Michigan and Argonne. The experiment had a slightly different form from the ones described above. Instead of studying events at a single beam energy and a variety of scattering angles, we kept the angle constant at 90 degrees and measured the cross section as the energy of the incident particle was varied from 5 to 13 GeV. At energies of up to 8 GeV the cross section decreased quickly and smoothly; at that energy, however, there was an abrupt change to a more gradual slope. The sharp break in the cross-section curve was interpreted as evidence for objects inside the proton with a size of one-third fermi.

Recent experiments with the Intersecting Storage Rings at CERN support the hypothesis of one-third-fermi objects. In the Intersecting Storage Rings protons circulate in opposite directions, so that they collide head on. By this stratagem extremely high center-of-mass energies are achieved; to yield equivalent results a fixed-target accelerator would require a beam energy of 1,500 GeV. The CERN investigators observed a distinct break in the cross-section curve as they passed from small-angle to large-angle scattering. Indeed, at these high energies the distinction between the core and the cloud seems to be quite sharp. It may be particularly significant that the apparent size of the core is the same in the 1,500-GeV findings as it is in the experiment carried out at from 5 to 13 GeV.

In about 1970 another line of evidence for internal structure emerged from experiments at the Stanford Linear Accelerator Center (SLAC), where 20-GeV electrons were made to strike a target of stationary protons. Similar experiments have since been completed elsewhere with high-energy neutrinos scattered by protons. In these experiments all interactions were counted rather than only the elastic ones. Such experiments are called "inclusive," and although the resulting cross sections are somewhat harder to interpret, they are often much easier to measure than an elastic cross section is. A surprising number of large-angle inclusive events were detected at SLAC, which suggests that the proton has an internal structure of some kind, but there is no direct way to calculate the size of the objects from the scattering data.

In spite of the difficulties of interpretation, the electron-scattering and neutrino-scattering experiments have been widely accepted as evidence for certain models of proton structure. One of these models, put forward in 1968 by Richard P. Feynman of the California Institute of Technology, proposes that the proton is made up of many of the tiny objects called partons. Another model is the quark theory, formulated in 1963 by Murray Gell-Mann and by George Zweig, who are also of Cal Tech. In this model the proton, the neutron and many related particles are thought to represent distinct combinations of three quarks each. The model has been quite successful in explaining the membership of several large families of particles and has served as a guide in the search for new particles.

From the results of 15 years of scattering experiments I think it is reasonably well established that the proton does have some smaller constituent objects. The elastic-scattering experiments strongly suggest that the size of these objects is about one-third fermi. In my opinion all other properties of these objects are highly uncertain, including their shape, mass, spin and electric charge. Even the number of constituent objects is conjectural. I see little direct evidence linking the quarks that have proved useful in accounting for the families of particles with the objects seen in scattering experiments. I should point out, however, that I am in the minority in holding this opinion.



INTERPRETATION OF SPIN EFFECTS observed in elastic-scattering experiments suggests that most of the proton's spin angular momentum is carried by the one-third-fermi interior objects. When only the outer regions of the proton interact, as in glancing collisions (*top diagrams*), the influence of spin is slight and the two cross sections are almost equal. In head-on collisions, where the hard objects participate (*bottom diagrams*), spin is of decisive importance. When the two proton spins are parallel, the one-third-fermi objects usually bounce apart violently; when the spins are antiparallel, the objects seem to pass straight through each other. The proton is shown as having a single dense core, a structure proposed by the author in 1963. The data do not exclude the possibility that there is more than one such object in each proton.

Experiments in which pure-spin cross sections are measured provide a new test for models of the proton's structure. In the past two years we have measured pure-spin cross sections at an incident-beam energy of 12 GeV and at scattering angles (in the center-of-mass coordinate system) ranging from near zero degrees to exactly 90 degrees. Knowing the polarization of the beam and of the target protons, but without determining the spins of the scattered and the recoil particles, we were able to distinguish four spin states: beam spin up and target spin down, beam spin down and target spin up, both spins up and both spins down.

The cross sections for the first two of these states, in which the spins are antiparallel, are required by symmetry principles to be exactly equal. The experiment confirmed that prediction, thereby providing another check for systematic bias. The two parallel spin cross sections, up up and down down, need not be the same. We found some difference between them in scattering at small and medium angles, but in the most violent collisions they are almost exactly equal.

Our most important findings are adequately represented by consolidating all the measured data in two cross sections, one for all states in which the incoming particles have parallel spins and one for the corresponding states with antiparallel spins. Both of these quantities are averages of four pure-spin cross sections, but they seem to display the largest spin effects observed so far in proton-proton scattering.

The results of our two-year experiment can be presented succinctly in a graph that plots the ratio of the spin-parallel cross section to the spin-antiparallel cross section as a function of the scattering angle. Even at small and medium angles (up to 60 degrees) the curve has interesting features. The first is simply that the ratio is never smaller than 1; no matter what the scattering angle is, parallel-spin scattering is more probable than antiparallel-spin scattering. On the other hand, at angles of less than 60 degrees the parallel cross section is never more than 25 percent larger than the antiparallel cross section. The most curious feature in this region is a narrow dip at about 25 degrees; only at that angle do the two scattering probabilities become equal. The dip is so narrow and so clearly resolved that it seems it must have some simple explanation, but so far there are no good candidates.

At small angles the detailed form of the curve is intriguing, but the approximate magnitude of the spin effects is unexceptional. At large scattering angles, however, the behavior of the ratio is astonishing. Large-angle scattering events are violent collisions, in which the particles exchange a large quantity of energy. We had expected that the small spin energy would be swamped in these events by the huge scattering energy and that spin would be unimportant. Instead the influence of spin grows dramatically. At angles greater than 60 degrees the ratio of the parallel cross section to the antiparallel cross section darts upward, and by 90 degrees it reaches a value of 4. It may climb still higher; the uncertainty in the measurements is still too large for us to determine whether or not 4 is the terminal value.

The same data can be presented in the form of separate curves for each cross section. Some of the properties observed in the earlier 1,500-GeV experiments with unpolarized protons then become apparent. Most notably the spin-parallel curve exhibits the familiar two-component form, with a sharp break separating the two segments. Furthermore, the slope of the large-angle segment, which corresponds to hard-scattering events, gives the same result for the size of the interior objects: one-third fermi.

There is another remarkable feature of these two curves. The abrupt transition in the spin-parallel cross section comes at a scattering angle of about 60 degrees: the same angle at which the ratio of parallel-to-antiparallel cross sections rises suddenly. Although the data are not yet precise enough for certainty, it may be that there is no corresponding break at all in the spin-antiparallel cross section. If that is so, the sudden increase in the ratio could readily be explained. At a scattering angle of about 60 degrees the one-third-fermi objects become the major influence on elastic proton-proton scattering, and these objects apparently scatter violently only in spin-parallel protons.

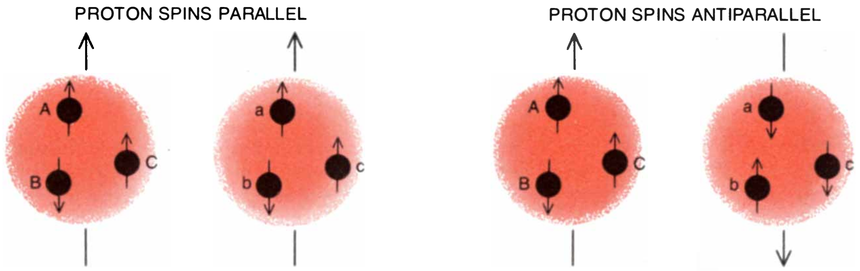
If these preliminary findings prove to be correct, the interpretation of the spin-resolved cross sections is simple but startling. When two protons spinning in the same direction collide head on, they almost always scatter violently. When the same particles are spinning in opposite directions, they often pass through each other as if they were transparent. The interactions at these large angles are probably mediated almost entirely by the one-third-fermi objects. Presumably these objects are spinning rapidly, and for some reason they rarely interact unless their spins are parallel.

The size of the spin effect may reveal something of the nature of the interior objects. At the maximum scattering angle (90 degrees) the contribution from the hard objects seems most important. At that angle the ratio of parallel-to-antiparallel cross sections was found to be 4, and this value may reflect in some way the number of objects in the proton and the nature of their spins.

For the onion model, with a single hard core, these observations merely suggest that the core carries most of the proton's spin angular momentum. Interactions that involve only the outer cloud are only slightly influenced by spin. The numerical value of 4 observed in the experiment is explained as the ratio of the spin-parallel to spin-antiparallel cross sections for direct interactions between the cores. At smaller scattering angles the direct core-core scattering is obscured by the moderating influence of the slowly spinning cloud.

The analysis is more complicated, and apparently less satisfactory, for a proton made up of three quarks. Each quark must have a spin of $1/2\,\hbar$; since the proton as a whole has the same value of spin angular momentum, the quarks must always combine with two of their spins parallel and one spin in the opposite direction. Large-angle scattering presumably results from the direct collision of one quark in the beam proton with one quark in the target proton. For such quark-quark collisions the pure-spin cross sections for proton-proton scattering must equal the sum of the pure-spin cross sections for all nine possible combinations of colliding quarks. When the ratio of the two sums is evaluated, it should yield the value of 4 observed for proton-proton scattering. This evaluation leads to an impossibility: a negative cross section, or in other words a probability of less than zero. A more complex model of quark-quark scattering might eliminate this difficulty, but it would also compromise the elegance of the model itself.

The large and unexpected influence of spin on large-angle scattering strongly constrains any theory that would explain violent scattering experiments in terms of the fundamental constituents of the proton. My colleagues and I hope soon to extend the investigation to collisions of greater violence, where the interior structure of the proton may be resolved even more clearly.



CONSTITUENT SPINS PARALLEL ($\uparrow\uparrow$)

$$\left.\begin{array}{l}\sigma(A\times a)\\ \sigma(A\times c)\\ \sigma(B\times b)\\ \sigma(C\times a)\\ \sigma(C\times c)\end{array}\right\}5\sigma(\uparrow\uparrow)$$

CONSTITUENT SPINS ANTIPARALLEL ($\uparrow\downarrow$)

$$\left.\begin{array}{l}\sigma(A\times b)\\ \sigma(B\times a)\\ \sigma(B\times c)\\ \sigma(C\times b)\end{array}\right\}4\sigma(\uparrow\downarrow)$$

CONSTITUENT SPINS PARALLEL ($\uparrow\uparrow$)

$$\left.\begin{array}{l}\sigma(A\times b)\\ \sigma(B\times a)\\ \sigma(B\times c)\\ \sigma(C\times b)\end{array}\right\}4\sigma(\uparrow\uparrow)$$

CONSTITUENT SPINS ANTIPARALLEL ($\uparrow\downarrow$)

$$\left.\begin{array}{l}\sigma(A\times a)\\ \sigma(A\times c)\\ \sigma(B\times b)\\ \sigma(C\times a)\\ \sigma(C\times c)\end{array}\right\}5\sigma(\uparrow\downarrow)$$

$$\frac{\sigma\,(\text{PROTON SPINS PARALLEL})}{\sigma\,(\text{PROTON SPINS ANTIPARALLEL})}=\frac{5\sigma(\uparrow\uparrow)+4\sigma(\uparrow\downarrow)}{4\sigma(\uparrow\uparrow)+5\sigma(\uparrow\downarrow)}=4$$

$$5\sigma(\uparrow\uparrow)+4\sigma(\uparrow\downarrow)=4\,[4\sigma(\uparrow\uparrow)+5\sigma(\uparrow\downarrow)]$$

$$0=11\sigma(\uparrow\uparrow)+16\sigma(\uparrow\downarrow)$$

$$\sigma(\uparrow\downarrow)=-\frac{11}{16}\sigma(\uparrow\uparrow)$$

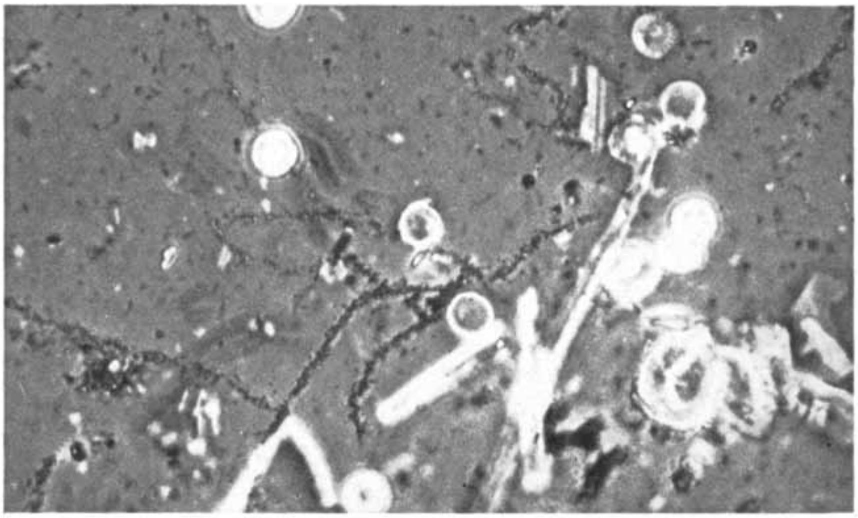
QUARK HYPOTHESIS suggests another model for the structure of the proton, but it is one that cannot easily be reconciled with the results of the spin-resolved scattering experiments. If quarks are the observed one-third-fermi objects, then at large angles the cross section for proton scattering should equal the sum of the cross sections for all events in which one quark is scattered by another. It is assumed that each proton consists of three quarks, and so there are nine cross sections for quark-quark scattering. Here a cross section is denoted by the Greek letter sigma (σ), and the symbols $\uparrow\uparrow$ and $\uparrow\downarrow$ respectively represent spin-parallel events and spin-antiparallel events. Since the proton cross sections for the two spin configurations are known to be in a ratio of 4 to 1, the sums of all the possible quark cross sections should have the same ratio. Setting the ratio for quark-quark scattering equal to 4, however, leads to an impossibility: one of the quark cross sections must be negative, and no probability can be less than zero. The difficulty can be avoided only in models with more complex quark-quark interactions.

Is dissolved organic carbon involved in eutrophication?

Possibly so. Possibly phosphates or nitrates are not alone to blame. But Hans W. Paerl's studies with Kodak nuclear track emulsions suggest that algae seldom use glucose or acetate from fresh or salt water. CO_2 seems to provide all the carbon they need, at least during daylight.

Dr. Paerl takes water from various depths in lakes of various kinds as well as the open ocean. He adds tritiated glucose and acetate to his samples and incubates. Then he kills all the microorganisms. Following preservation he filters out everything particulate, gets rid of salts that could crystallize, and dries. A quick pass through acetone vapor makes his filters transparent.

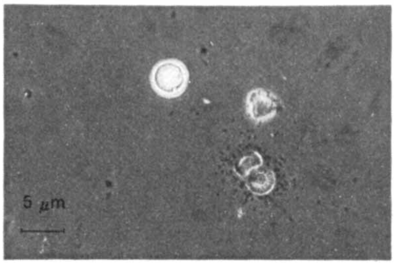
In the dark he melts Kodak nuclear track emulsion, either Type NTB-2 or NTB-3, at 40°C, dilutes it 1:1 with distilled water, dips in slides bearing his cleared filters, pulls them out after about 5 seconds, lets them dry on edge for about 15 minutes, then packs them away for two weeks in a lighttight box containing desiccant. After Kodak developer D-19 has done its work, he fixes and washes in deionized water. Then he looks in the microscope at ×1000 for black silver grains that show where the tritium was incorporated.



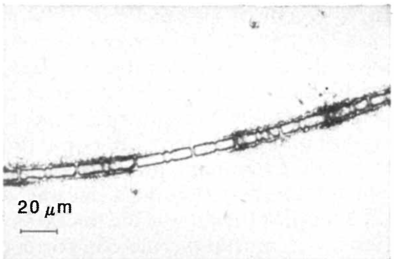
In scenes like this the black dots form over filamentous and coccal bacteria attached to detritus. Only bacteria and fungal hyphae—both free-living and attached—give such evidence of having extracted the dissolved organic carbon. Algae do not.

In working with algae such as diatoms and dinoflagellates it has been hard to tell those that were alive when the sample was taken from dead siliceous skeletons that abound in natural waters. Dr. Paerl solves that with microautoradiography too, as seen at right.

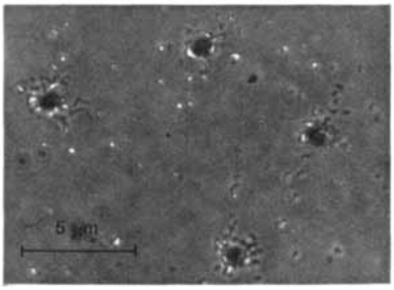
His address is Institute of Marine Sciences, University of North Carolina, Morehead City, N.C. 28557, phone 919-726-6841. The person to talk to about acquiring some Kodak nuclear track emulsion can be reached at 716-724-4633. Or Publication P-64 will give product and ordering information—drop a postcard to Dept. 55W, Kodak, Rochester, N.Y. 14650.



Living *vs* non-living cells of the diatom genus *Cyclotella*. Cells which have actively incorporated $^{14}CO_2$ through photosynthesis are capable of exposing Kodak nuclear track emulsions. The emulsion was placed over a filtered plankton sample after the sample was incubated with $^{14}CO_2$ under sunlight. Non-living *Cyclotella* (bright unlabeled cell) fail to expose this emulsion.



Living *vs* non-living cells present in a filamentous diatom, a species of *Melosira*. A majority of individual cells in the filament are capable of photosynthesis as determined by autoradiography. Photosynthetic ^{14}C-labeled cells are surrounded by clusters of exposed silver grains in Kodak NTB-2 nuclear track emulsion.



The detection of ultraplankton (0.7 µm diameter) by microautoradiography. Photosynthetically-active cells are surrounded by groups of exposed silver grains in Kodak NTB-3 nuclear track emulsion. Such cells escape detection during routine microscopic enumeration.

Kodak

SCIENCE AND THE CITIZEN

Cardiovascular Countertrend

Fewer Americans are dying of diseases of the heart and blood vessels. Whether because of reduced incidence or better care, the mortality rate from cardiovascular disease, adjusted for the effects of an aging population, has declined a surprising 32 percent in the past three decades. The decline has been accelerating: in just the nine years from 1968 through 1977 the age-adjusted cardiovascular death rate fell by 24.3 percent, and the decline appears to have continued in 1978. (The overall death rate has fallen too, but more slowly: by 17.7 percent in the nine years. The difference is accounted for by a rise in deaths from some other causes, including cancer, which accounted for 16.5 percent of all deaths in 1968 and for an estimated 20.5 percent last year.) Even the absolute number of annual deaths from major cardiovascular diseases has declined in spite of the increasing size and age of the U.S. population. In 1975 the number was below one million for the first time in a decade; the 1977 total, 960,000, was the lowest since 1963. The mortality rate for coronary heart disease, the major component of the cardiovascular group and the leading cause of death in the U.S., declined by 23.4 percent from 1968 through 1977. Most striking of all was the drop in the death rate for the second-largest component, cerebrovascular disease (stroke). It fell by 32.1 percent in the nine years and has been declining by an average of more than 5 percent per year since 1972.

Commenting on such statistics in an editorial in *The New England Journal of Medicine,* Robert I. Levy of the National Heart, Lung, and Blood Institute points out that the important question is how much of the decline can be attributed to lower rates of incidence and how much to better patient care. The answer is important to any assessment, on the one hand, of preventive programs (including control of high blood pressure and education as to risk factors such as smoking) and, on the other, of aggressive treatment (including coronary-care units and cardiac surgery). Levy cites the report of a conference on coronary heart disease last fall. The report concluded that probably both primary prevention and improved treatment had played a part, but that their relative roles could be determined only by further studies, particularly "those designed to document whether the frequency of nonfatal coronary events is changing" along with the fatality rate. Levy believes the situation is about the same for cerebrovascular disease, namely that only quantitative data on the incidence of strokes, whether fatal or nonfatal, can help to define the causes of the sharp decline in mortality.

Such data are hard to come by because of problems of comparability among various studies and of long-term continuity within individual studies, according to a team of investigators at the Mayo Clinic, writing in the same issue of *The New England Journal.* W. Michael Garraway, Jack P. Whisnant and their colleagues exploited their institution's medical-records system, which covers all hospitalizations and almost all office visits in Rochester, Minn., to collate stroke-incidence data for residents of the city over the period from 1945 through 1974. The overall incidence of stroke was found to have decreased in each five-year period, from an average annual rate of 190 cases per 100,000 of population (adjusted for age and sex to the 1950 U.S. white population) in 1945–49 to 104 cases in 1970–74. The extent of the reduction was similar for men and women, but the rate was usually higher for men than it was for women. Age-specific rates showed the greatest reductions for the very elderly. Analysis by birth cohorts showed no difference in the rates of decline for people born in successive five-year periods from 1865 to 1915. The 30-year decrease was about 34 percent for strokes brought on by obstruction of blood vessels supplying the brain (the major category) and about 44 percent for strokes caused by hemorrhage within the brain; the rate for subarachnoid hemorrhage (between a membrane covering the brain and the brain itself) showed almost no change.

The Rochester survey, the authors point out, is the first to document "the declining incidence of cerebrovascular disease in a well-defined population over such a long-term period." The reduction appears to have taken place in two phases. The average annual decline in the overall stroke rate, adjusted for age and sex, was 3.1 per 100,000 between 1945 and 1959. There was not much change for the next five years. Then the rate of decline accelerated, averaging 4.8 per 100,000 in 1965–69 and 5.3 in 1970–74.

Advancing Periastron

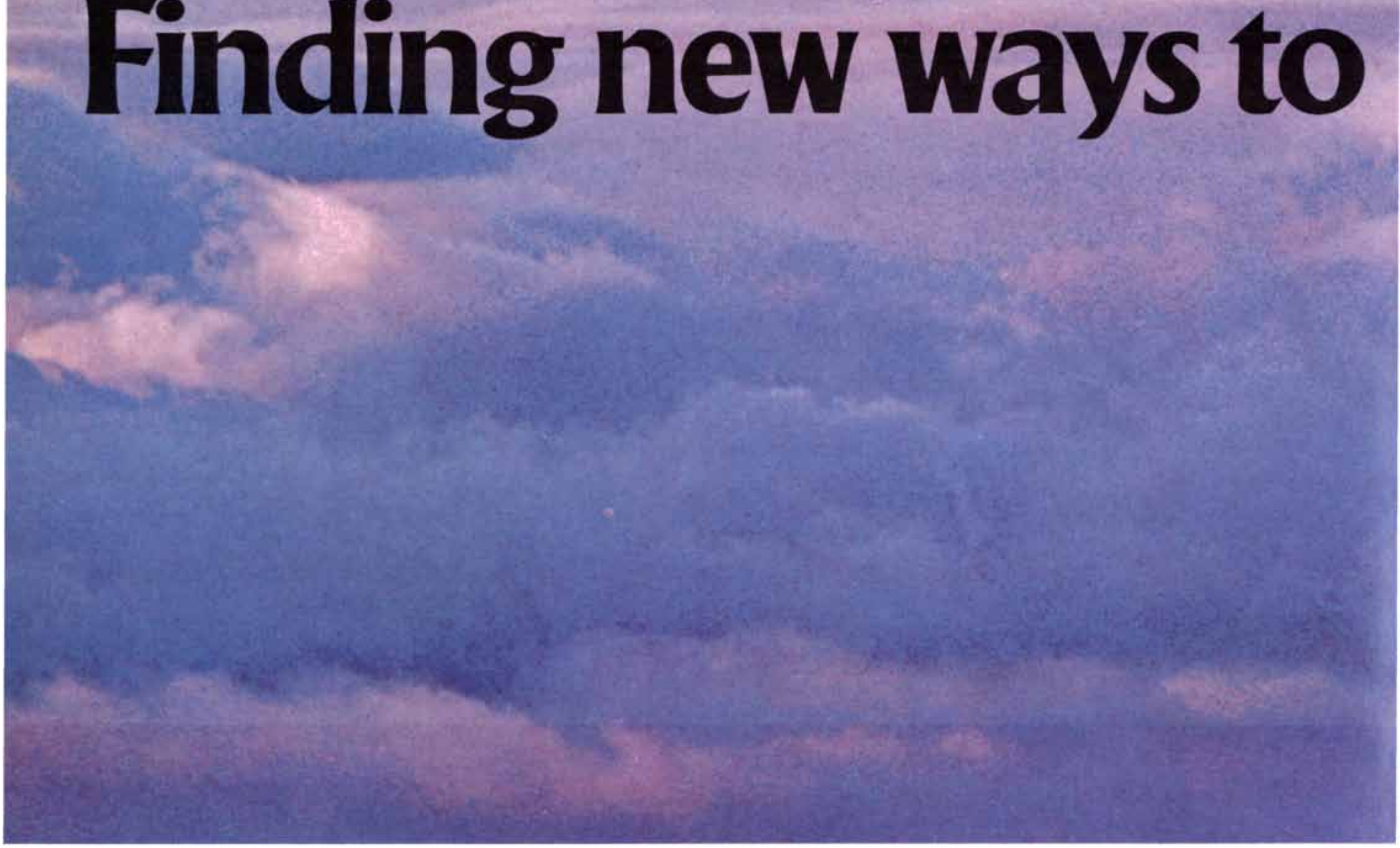
The first test of Einstein's general theory of gravitation to be made on objects outside the solar system was reported shortly before the 100th anniversary of Einstein's birth. The opportunity for such a test presented itself with the discovery in 1974 of a radio pulsar that is a member of a binary pair. Of more than 300 pulsars discovered to date only one radio pulsar, PSR1913 + 16, is in orbit around another object. (Although there are a dozen or more X-ray binary pulsars, they do not lend themselves to tests of relativity.) The unmistakable presence of an unseen partner is revealed by a systematic Doppler shift in the pulsar's sharply defined radio signal as it travels in its 7.75-hour orbit alternately toward and away from the terrestrial observer. A pulsar's radio emission takes the form of a beam, possibly a hollow cone, that sweeps the sky locked in synchrony with the rapidly spinning pulsar source. Pulsar rotation periods range from about a thirtieth of a second to several seconds. The binary pulsar PSR1913 + 16 has a rotation period of .059029995269 second. A pulsar is thought to be the stellar remnant of a supernova explosion. Slightly more massive than the sun, the remnant contracts to a radius of a few kilometers, at which point its constituent particles are squashed into neutrons. They survive as rapidly spinning neutron stars.

Because the arrival of pulsar emissions can be timed with high precision it was immediately clear that several predictions of Einstein's special and general theories of relativity could be tested rather quickly in a binary system in which two large masses are interacting at high speed over a short distance. A major prediction, never before accessible to test, is that the interacting bodies should radiate part of their kinetic energy in the form of gravitational waves. The general theory predicts that the amount of energy radiated gravitationally by the binary pulsar pair should be comparable to the electromagnetic energy radiated by a typical star, roughly 10^{33} ergs per second. Although there are no existing instruments sensitive enough to detect the gravitational radiation directly, the energy loss should reveal itself by a gradual decrease in the distance separating the two orbiting bodies and a corresponding decrease in the pulsar's orbital period.

Since 1974 the signal emitted by PSR1913 + 16 has been closely monitored by its codiscoverer, Joseph H. Taylor of the University of Massachusetts at Amherst, with the 305-meter radio telescope at Arecibo in Puerto Rico. In a recent issue of *Nature* Taylor, L. A. Fowler and P. M. McCulloch report the results of some 1,000 observations over four years. With gradual improvements in technique pulse-arrival times can now be established with an accuracy of about 50 microseconds.

As the binary pair have radiated energy over the four-year period the time required for the objects to circle each other has been decreasing at the rate of 101 ± 19 microseconds per year. On the basis of all other measurements of the system, including estimates of the mas-


Evenings that memories are made of...
DRAMBUIE
LIQUEUR
Prince Charles Edward's Liqueur
TRADE MARK
DRAMBUIE
A LINK WITH THE "45"
80 PROOF LIQUEUR. IMPORTED BY
W. A. TAYLOR & CO., MIAMI, FLORIDA © 1978



Lockheed knows how.

"Biggest advance since the autopilot."

That's one widely held opinion of an exclusive fuel-saving system now being installed in many L-1011 TriStars.

Called the Lockheed Flight Management System, it senses altitude, speed, and other flight factors. Then it can act automatically to control the engine throttles for best fuel-use efficiency at all times during flight—takeoff to landing.

Flight management control panel.

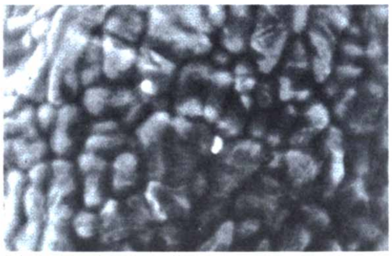
Though a new feature, the system isn't just for new planes; it can easily be added to existing L-1011s. And the savings, over the life of each plane, could total millions of dollars.

Making aircraft lighter.

Since the 1930s, most aircraft have been structured of strong, lightweight aluminum. Future generations of aircraft may also be made mostly of aluminum, but of a very different nature.

A new technology being explored by Lockheed uses powdered aluminum, alloyed with such elements as iron, lithium or cobalt, to create lighter, stronger and stiffer metals than can be achieved with conventional aluminum technology.

Results indicate this 'new aluminum' could reduce present-day jetliner weight by 10%. That in turn means savings of millions of gallons of fuel over each aircraft's life, a big benefit in energy-short years.



Aluminum lithium alloy magnified 1,000 times.

For this new-material process, Lockheed scientists even had to employ astonishing techniques for turning the powdered aluminum and alloys into superior metal. After heating and melting, for example, the mixture is rapidly cooled *a million times faster* than the cooling rate used with standard aluminum ingots.

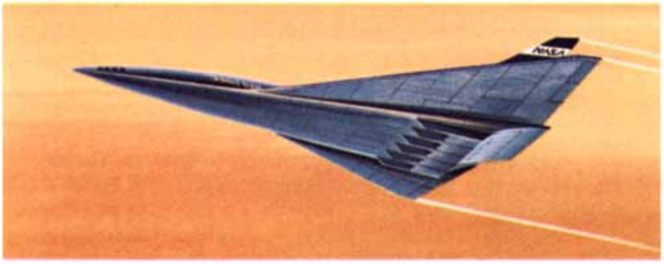
Down the road: liquid hydrogen.

Lockheed scientists are deep into a NASA-sponsored study of a future aircraft, powered by a clean-burning liquid hydrogen, that could cruise at a flashing hypersonic speed of about 4,000 mph, or six times the speed of sound.

Because liquid hydrogen can be made from coal or even water, its future use in aircraft could eliminate the need for scarce, expensive petroleum.

The aircraft would cruise at altitudes as high as 120,000 feet and have ocean-spanning range. Its 200 passengers

save fuel in flight.



Design for liquid hydrogen fueled aircraft.

could be whisked from New York to London in under two hours, and could make the Los Angeles/Tokyo flight in a mere two hours, 18 minutes—takeoff to landing.

Fine-tuning the flight plan.

When a pilot calls Lockheed's JetPlan, he's in touch with a computerized encyclopedia of information affecting his proposed flight. And the benefit he gets is eye-opening economy—savings of both time and fuel.

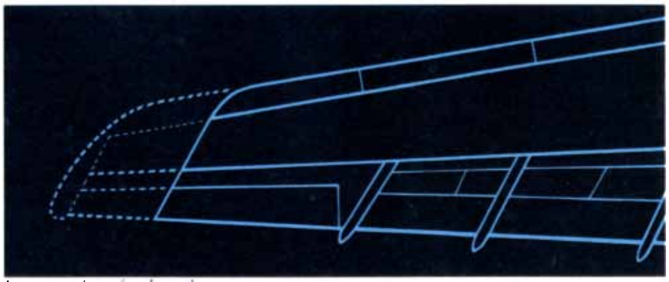
To tap a constantly updated reservoir of data, a pilot makes preflight contact via telephone and a small printer terminal. In minutes, he gets back his most efficient flight plan. Given origin, destination, and type of aircraft, JetPlan reckons its answer from such factors as route analysis, current weather, recommended fuel load, cruise altitude, speed and shortest route.

How well does it work? Over a recent 12-month test period, U.S. military and corporate aviation saved more than 30 million gallons of fuel, or better than $1.5 million. But that's just the beginning. The number of JetPlan users and the savings are both growing steadily.

Stretching wings.

Longer wings mean reduced drag and consequently a substantial cut in fuel use. But among big wide body jetliners, only the Lockheed L-1011 TriStar can add nine feet of wingspan without costly structural redesign.

That's because of the L-1011's advanced technology, in this instance an exclusive system of Active Control ailerons. Controlled by computers, these ailerons will act automatically to reduce the structural loads on the lengthened L-1011 wings.



Longer wings for less drag.

The longer wings and Active Control ailerons will mean a healthy savings of fuel when they enter airline service in 1980. And they'll also help the L-1011 have the smoothest ride of any jetliner through gusts and turbulence.

When it comes to devising farseeing programs to meet the growing fuel crisis, Lockheed knows how.

Lockheed

ses of the two objects and the inclination of their orbital plane with respect to the line of sight, general relativity predicts that the orbital period should be decreasing at the rate of 76 microseconds per year. It is considered impressive that the observed value falls hardly more than two standard deviations from the predicted value. In any event the observations seem to rule out some rival theories of general relativity, notably one advanced by Nathan Rosen of the Technion-Israel Institute of Technology in Israel, who collaborated with Einstein in the 1930's. The Rosen theory, like several others, predicts dipole gravitational radiation from an orbiting system such as PSR1913 + 16, whereas Einstein's theory predicts quadrupole radiation. Dipole theories in general predict much larger values of gravitational radiation than Einstein's theory does; hence they also predict a higher rate of decrease in orbital period than the observed 101 microseconds per year. The Brans-Dicke theory of relativity, developed by Carl Brans and Robert Dicke of Princeton University, makes about the same prediction as Einstein's theory for a binary system consisting of two similar masses, as is the case with PSR1913 + 16. The Brans-Dicke theory, however, has had difficulty meeting solar-system tests.

A second major prediction of general relativity involves the rate at which the axes of an object's elliptical orbit will precess, or rotate in space, over time. The greater the eccentricity of the orbit and the shorter the orbital period, the higher the rate of precession. It had been known long before Einstein's work that there was a discrepancy of 43 seconds of arc per century between observations and the prediction of the classical Newtonian theory of gravity in the rate of precession of the perihelion of Mercury. The perihelion, or point of closest approach to the sun, defines the orientation of a planet's orbital ellipse. The general theory of relativity accounted for the 43-arc-second discrepancy with an accuracy of about 1 percent.

Because the binary pulsar circles its companion more than 1,100 times per earth year and because its orbit is about three times as eccentric as Mercury's (.617 as opposed to .207), the precession rate of the binary's orbit provides a far more dramatic and stringent test of general relativity than the precession rate of Mercury does. The binary's periastron, or closest approach to its unseen companion, has been found to precess at the rate of $4.226 \pm .002$ degrees per year. The value is now determined with an accuracy of better than .1 percent. Newtonian theory predicts no precession at all for two bodies that act as point masses. Assuming that the measured value of precession is accounted for entirely by general relativity, one can calculate that the pulsar has a mass $1.39 \pm .15$ times the mass of the sun and that the

Step up to excellence. Step up to Series E.

Why settle for anything less than excellence. Especially now that it's available at a very affordable price in Series E from Hewlett-Packard—five precision calculators for science, engineering and business starting at just $60.*

EXCELLENCE BY DESIGN. Experience the HP difference in Series E: a new larger display with commas for easy reading; built-in diagnostic systems that help catch errors; RPN logic for ease-of-use and efficiency; a complete documentation system—Owners' Manual and Applications Books; plus a host of HP extras—low battery warning light, rechargeable batteries, positive feedback keys, sturdy impact-resistant case. In sum—excellence by design.

A SERIES E FOR EVERY NEED. Series E gives you more power and features at a lower price than any comparable calculator HP has ever offered. You'll find them all outstanding values:

HP-31E. Scientific. $60.*
HP-32E. Advanced Scientific with Statistics. $80.*
HP-33E. Programmable Scientific. $100.*
HP-37E. Business Management. $75.*
HP-38E. Advanced Financial with Programmability. $120.*

SEE YOUR HP DEALER. For details, send in the coupon or visit your nearest HP dealer. For the name and address, CALL TOLL-FREE 800-648-4711 except from Alaska or Hawaii. In Nevada, call 800-992-5710.

HEWLETT hp PACKARD

Dept. **236H** 1000 N. E. Circle Blvd., Corvallis, OR 97330

HEWLETT-PACKARD
Dept. **236H**
1000 N. E. Circle Blvd.
Corvallis, OR 97330
Please send details on Series E:
______ Business Calculators
______ Scientific Calculators
Name ____________________
Address ____________________
City ____________________
State __________ Zip ________

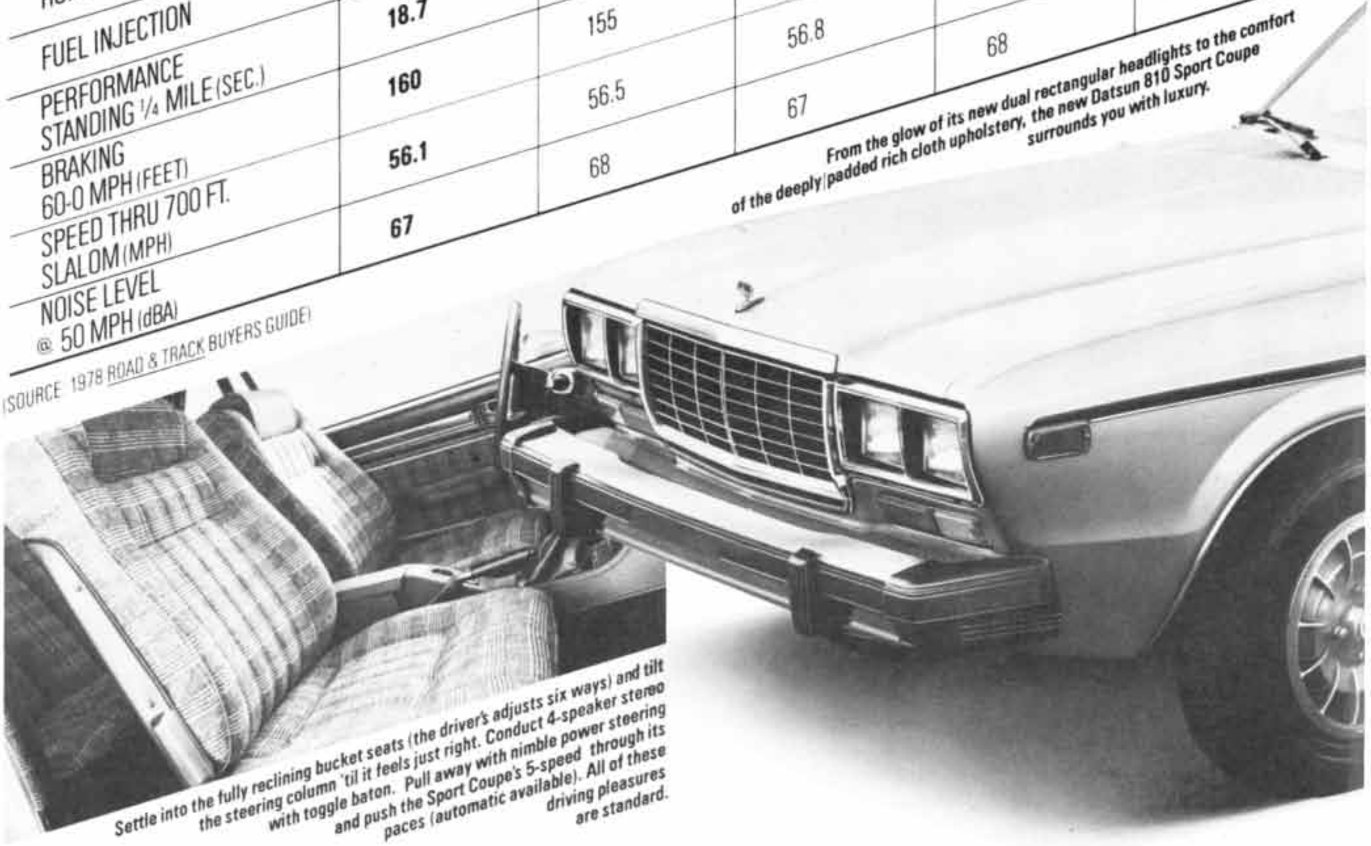
*Suggested retail price excluding applicable state and local taxes—Continental U.S.A., Alaska and Hawaii

Displays photographed separately to simulate typical appearance. #619/03

DATSUN ASKS: WHY PAY THE DIFFERENCE IF YOU CAN'T TELL THE DIFFERENCE?

	DATSUN 810	MERCEDES-BENZ 280-E	AUDI 5000	BMW 320i	VOLVO 244 DL
CONFIGURATION	**F/ENGINE R/DRIVE**	F/ENGINE R/DRIVE	F/ENGINE F/DRIVE	F/ENGINE R/DRIVE	F/ENGINE R/DRIVE
WHEELBASE (INCHES)	**104.3**	110.0	105.7	100.9	104.0
BRAKES	**DISC/DRUM**	DISC/DISC	DISC/DRUM	DISC/DRUM	DISC/DISC
SUSPENSION	**IND/IND**	IND/IND	IND/BEAM	IND/IND	IND/LIVE
ENGINE	**SOHC 6**	DOHC 6	SOHC 5	SOHC 4	SOHC 4
HORSEPOWER (BHP)	**127**	137	103	112	114
FUEL INJECTION	**BOSCH**	BOSCH	BOSCH	BOSCH	BOSCH
PERFORMANCE STANDING 1/4 MILE (SEC.)	**18.7**	18.6	19.5	18.7	19.8
BRAKING 60-0 MPH (FEET)	**160**	155	148	167	180
SPEED THRU 700 FT. SLALOM (MPH)	**56.1**	56.5	56.8	55.1	53.0
NOISE LEVEL @ 50 MPH (dBA)	**67**	68	67	68	65

(SOURCE: 1978 ROAD & TRACK BUYERS GUIDE)



From the glow of its new dual rectangular headlights to the comfort of the deeply padded rich cloth upholstery, the new Datsun 810 Sport Coupe surrounds you with luxury.

Settle into the fully reclining bucket seats (the driver's adjusts six ways) and tilt the steering column 'til it feels just right. Conduct 4-speaker stereo with toggle baton. Pull away with nimble power steering and push the Sport Coupe's 5-speed through its paces (automatic available). All of these driving pleasures are standard.

mass of the companion (presumably also a neutron star but conceivably a white dwarf or a blackhole) is $1.44 \pm .15$ solar masses. These values are consistent with other observed properties of the binary system and yield the first direct determination of the mass of a pulsar.

How Interferon Interferes

The remarkable ability of the protein interferon to inhibit the multiplication of viruses in animal cells has tantalized biochemists and virologists ever since its discovery in 1957. Experiments have demonstrated that interferon does not prevent virus particles from infecting cells but rather inhibits replication and expression of the viral genes. Acting much like a hormone, interferon binds to a specific receptor site on the cell surface, triggering the manufacture within the cell of enzymes that mediate the antiviral effect. The mechanism of action of these enzymes is now beginning to emerge. Writing in *Proceedings of the National Academy of Sciences,* Charles E. Samuel of the University of California at Santa Barbara describes how two interferon-induced enzymes selectively prevent the translation of viral messenger RNA (mRNA) into viral proteins within the host cells.

Protein synthesis, whether of cellular or of viral proteins, always begins with the formation of an "initiation complex" consisting of (1) the strand of mRNA to be translated, (2) the smaller of the two subunits of the ribosome, the subcellular particle that coordinates translation and (3) a molecule of initiator transfer RNA, which provides the first link in the growing chain of amino acids. Several additional protein initiation factors hold the complex together. Samuel has found that interferon induces the manufacture in the host cells of an enzyme, protein kinase, that transfers a phosphate group to a specific subunit of the initiation factor designated eIF-2. The phosphorylated form of eIF-2 is no longer able to function efficiently in the formation of the initiation complex, so that the translation of the viral mRNA is blocked before it can begin.

Although interferon induces the manufacture of the protein kinase, an additional ingredient is necessary to activate it: double-strand RNA. Samuel believes this requirement is supplied by the viral mRNA, which, although it is a single strand, loops back on itself to form double-strand regions. Once activated, the kinase first phosphorylates itself (an unusual reaction called autophosphorylation) and then phosphorylates eIF-2, thereby preventing the initiation of protein synthesis.

Interferon also induces the manufacture of a second enzyme, a synthetase that catalyzes the polymerization of adenine nucleotides into a long chain of adenine units called 2,5-oligoadenylic acid. This simple nucleic acid in turn activates a ribonuclease (an enzyme that degrades RNA) that is present in the cell in an inactive form. The activated ribonuclease degrades the viral mRNA so that it cannot be translated into protein.

How are the protein kinase and the ribonuclease able to selectively prevent the expression of the viral mRNA without interfering with the mRNA of the host cell? Without such a distinction the enzymes would block the protein synthesis of the host cell and ultimately kill it. Samuel has proposed that the two enzymes work in concert to destroy the viral mRNA selectively. According to this scheme, the protein kinase is activated by the viral mRNA but not by the host-cell mRNA, since only the viral mRNA has the appropriate double-strand regions needed to activate the enzyme. (Whether there are in fact significant differences in the secondary structures of viral and cellular mRNA's will be resolved only when a large number of nucleotide sequences of both types of mRNA have been determined.)

Assuming that the protein kinase is in fact activated only by viral mRNA, the enzyme would prevent the viral mRNA from forming an initiation complex but the host-cell mRNA would continue to form stable initiation complexes. As a result the host-cell mRNA would be rapidly enclosed within a polyribosome (a series of closely spaced ribosomes translating the same mRNA strand), whereas the viral mRNA would remain uncoated by ribosomes. In this "naked" state it would be vulnerable to degradation by the activated ribonuclease. Although this hypothesis is far from proved, it provides an elegant explanation for how interferon could accomplish the selective destruction of the viral mRNA.

Another interesting feature of Samuel's hypothesis is that the phosphorylation of eIF-2 is also known to inhibit protein synthesis in reticulocytes: immature red blood cells present in the bone marrow. In these cells the phosphorylation of eIF-2 is triggered by a deficiency of hemin (a form of heme); it effectively blocks the manufacture of all cellular proteins, including the globin chains that combine with heme to form hemoglobin. (This regulatory pathway makes eminent sense because there is no point in the cells' manufacturing globin if there is not enough heme.) The specific phosphorylation of eIF-2 is therefore responsible for controlling translation in two quite different physiological conditions: interferon treatment and hemin deficiency.

QCD

The idea of the quark is one that has explained much in the physics of elementary particles but has itself proved difficult to explain. The idea states that protons, neutrons and the many related particles called hadrons are not elementary but are made up of smaller objects: the quarks. In this way all the known hadrons (and there are more than 200) can be accounted for as simple combinations of a few quarks. It has taken more than 15 years, however, to devise a satisfactory theory of how the quarks themselves move inside a hadron and how they interact with one another. The theory is still not complete, but a consensus now seems to be forming that it is correct.

The theory is called quantum chromodynamics, or QCD. It is formulated on the model of quantum electrodynamics, or QED, the modern theory of electromagnetic interactions. In QED the basic quantity to be explained is the force that arises between two particles bearing an electric charge, such as two electrons. The theory accounts for the force by stating that the electrons can exchange an intermediary particle. The intermediary is a photon, a massless particle that moves at the speed of light. The photon itself carries no electric charge, but it transmits the electromagnetic force from one particle to another.

The forces that bind quarks together are not electromagnetic; they result from another of the four basic forces of nature: the strong force. (The remaining forces are the weak force and gravitation.) The mechanism of interactions among quarks is nonetheless much the same as it is in QED. A force arises between particles that have a property analogous to electric charge, and the force is transmitted by other particles, which in this case are called gluons because they hold the quarks together. Like the photon, gluons are massless and move at the speed of light.

One difference between QED and QCD is that quarks can have not merely a single kind of charge but any one of three kinds. The charges are called colors, whence the name chromodynamics. Whereas an electrically charged particle can be only positive or negative, a particle bearing a color charge might be red or antired, blue or antiblue or green or antigreen. Moreover, several intermediary particles are needed to transmit the forces between the variously colored quarks. A single photon suffices for electromagnetism, but eight kinds of gluon are required for the strong interaction.

The multiplicity of color charges and of intermediary particles makes QCD more complicated than QED, but it does not change the form of the theory. What does introduce a fundamental change is that gluons are themselves colored, unlike the electrically neutral photon. The presence of a color charge on the gluons alters the way the strong force varies with distance.

The electromagnetic force between two particles declines in proportion to the square of the distance between them.


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
U.S. IMPORTERS
THE GLENLIVET DISTILLING COMPANY
NEW YORK, N.Y.
750 ML. (25.4 FL. OZ.) 86 PROOF
ESTABLISHED 1824
THE TASTE BEYOND
PREMIUM SCOTCH

1873

The photograph of Eroded Sandstones shown here is one of the first views Americans saw of the West.

It was made back in 1873 by William Henry Jackson at Monument Park, Colorado.

Jackson was an official photographer for the first U.S. geological survey.

Traveling with 300 pounds of bulky photographic gear strapped to mules, Jackson journeyed thousands of miles through hazardous territory in Colorado, Wyoming, Utah and the Southwest.

He made hundreds of photographs like this one by the wet-plate process. On location, a glass plate was coated with a light-sensitive solution, exposed, and developed in a portable dark-tent.

The finished glass-plate negatives were carried back from the wilderness and printed on albumen paper.

Jackson's pictures were instrumental in the creation of the National Park System by Congress during the 1870's.

This is an original picture of Monument Park, Colorado made by William H. Jackson.

The force between quarks is quite different. When two quarks are very close together, they seem to move about freely and independently, and a working hypothesis states that the strength of the coupling between them falls to zero when they are close enough to touch. At longer range (and in this context the diameter of a proton is a long range) the coupling becomes stronger. In other words, as the quarks are pulled apart the bonds between them get tighter. As a result of this bizarre force law it is difficult, and perhaps impossible, to separate a single quark from a hadron.

Actually the force law is not as bizarre as it seems at first. Even in QED the force between particles does not vary strictly as the inverse square of the distance. That relation is observed only at long range, and it results in part from the screening of the electron's "bare" charge by surrounding opposite charges that appear even in a vacuum. If an electron could be approached closely enough, the effective charge and the associated force would increase. In QCD the effect is exactly the opposite, again because the gluons bear a color charge. At long range the cloud of gluons surrounding a quark tends to increase the quark's effective color charge; when two quarks are close together, the rapid exchange of gluons tends to neutralize the quark charges.

QED is the most precise theory in science, having predicted one physical constant to an accuracy of one part in 100 million. QCD has not approached that level of precision, but in the past few years some of its predictions have been confirmed. For example, the theory suggests that if two hadrons collide violently enough, a quark and an antiquark should be emitted back to back. The quark and the antiquark each give rise to other hadrons, which form "jets" of particles all moving in approximately the same direction. Such jets have been observed in several experiments. Another experiment at the European Organization for Nuclear Research (CERN) near Geneva has provided quantitative evidence in support of the theory. The experiment employed high-energy neutrinos to examine the quarks inside protons and effectively measured the variation of the color charge with distance. The result was in close agreement with the predictions of QCD.

In the past few months new techniques for calculating the probability of certain interactions in QCD have been developed. They may allow additional quantities to be predicted and hence may lead to further experimental tests of the theory. New experimental apparatus might also help to verify certain predictions. At higher energies than those available today a violent collision might give rise to three jets of particles rather than two. Such three-jet events should be observed in experiments with several new storage-ring accelerators about to go into operation.

All the experimental evidence for QCD involves short-range interactions of quarks. The long-range interactions are much harder to understand. QCD does account in a qualitative way for the strong binding of distant quarks. Ironically, for that very reason the theory has so far proved to be useless for predicting the detailed motion of quarks at long range. The method for making such calculations in both QED and QCD consists in summing a series of terms, and the series converges rapidly toward the correct value only if the coupling between the particles is weak. In QED this constraint is unimportant because the electromagnetic force is never very strong. The large magnitude of the strong force confines the utility of QCD to short-range interactions.

Several solutions to this problem are being explored. One solution introduces the confinement of quarks as a postulate by placing them inside a flexible but impermeable bag or bubble. Another suggestion is that the fields of force binding quarks together may have topological twists, called instantons, that alter the long-range interaction. Still another approach is through the mathematical technique called the renormalization group, which might allow a numerical solution of the problem.

Other investigators point out that the inability of QCD to predict the long-range interactions of quarks is not a fundamental flaw in the theory but merely a complication. By way of analogy, the development of quantum mechanics in the 1920's did not lead to the immediate understanding of all atomic structures but only to that of the simplest atom: hydrogen. It was nonetheless possible to employ quantum mechanics in formulating further theories of the atomic nucleus and eventually of the particles in the nucleus. In the same way the effort is now under way to incorporate QCD, in spite of its difficulties, into a larger theory that would unify at least three of the four forces in nature: the strong, the weak and the electromagnetic.

Sea-Floor Ore

One of the principal geochemical mechanisms for the formation of metallic ores is believed to be hydrothermal activity at oceanic rifts, where hot magma is welling up from the earth's mantle and spreading out to form new ocean floor. In such activity metals are extracted by the interaction of hot water and mantle minerals. The mechanism has now been observed in action on a massive scale along the rift at the crest of the East Pacific Rise off Baja California. The ore deposits may extend along the rift for thousands of kilometers.

Writing in *Nature*, Jean Francheteau of the Oceanographic Center of Brittany and 14 colleagues from nine other institutions in France, the U.S. and Mexico have reported their reconnaissance of the East Pacific Rise aboard a deep-diving research submersible in February and March of last year. At a depth of more than 2,600 meters they found a thermally active zone where mineral-rich hot water from vents in the sea floor was precipitating sulfides of iron, zinc and copper as it mixed with the cold seawater. The precipitates took the form of cone-shaped structures, some as much as a meter high, rising from a foundation consisting mainly of pillow lava. The crew of the submersible took 14 samples of these future ore bodies in the course of two dives. The samples represented five distinct types of precipitate, which were analyzed for their elemental composition by means of X-ray fluorescence and atomic absorption spectrometry.

Four of the five types of precipitate, consisting mainly of differing proportions of the minerals sphalerite and pyrite, proved to contain 28 and 23 percent zinc and 29 and 15 percent iron respectively. The fifth type, consisting of globular aggregates of iron oxide, contained 42 percent iron. All but the aggregates also contained a proportion of copper ranging from .2 to 6 percent. The investigators suggest that a similar metal-rich zone may exist along the Mid-Indian Ocean Ridge.

Astounding Asteroids

The asteroids, or minor planets, have always had a peculiar charm. Although none has ever been seen close up, a good idea of what asteroids look like is provided by spacecraft pictures of Deimos and Phobos, the moons of Mars. Asteroids are lumpy, irregular bodies, not at all smooth orbs like the major planets. Now it seems they have other unexpected characteristics. Recent investigations indicate that some asteroids have satellites and others may be paired, orbiting around each other like a double star. At least one asteroid, 624 Hektor, may have the shape of a dumbbell, presumably formed billions of years ago when two asteroids gently collided and stuck together.

On rare occasions an asteroid may pass between a star and the earth, briefly eclipsing the star. The diameter of the asteroid can then be estimated from the length of time the star is occulted. On June 7, 1978, a star was eclipsed by the asteroid 532 Herculina, write R. P. Binzel of Macalester College and T. C. Van Flandern of the U.S. Naval Academy in *Science*. The occultation lasted for 20.6 seconds, but six additional reductions in starlight were also observed within two minutes of the main eclipse; their duration ranged from .5 second to four seconds. The secondary occultations could not have been caused by some atmo-

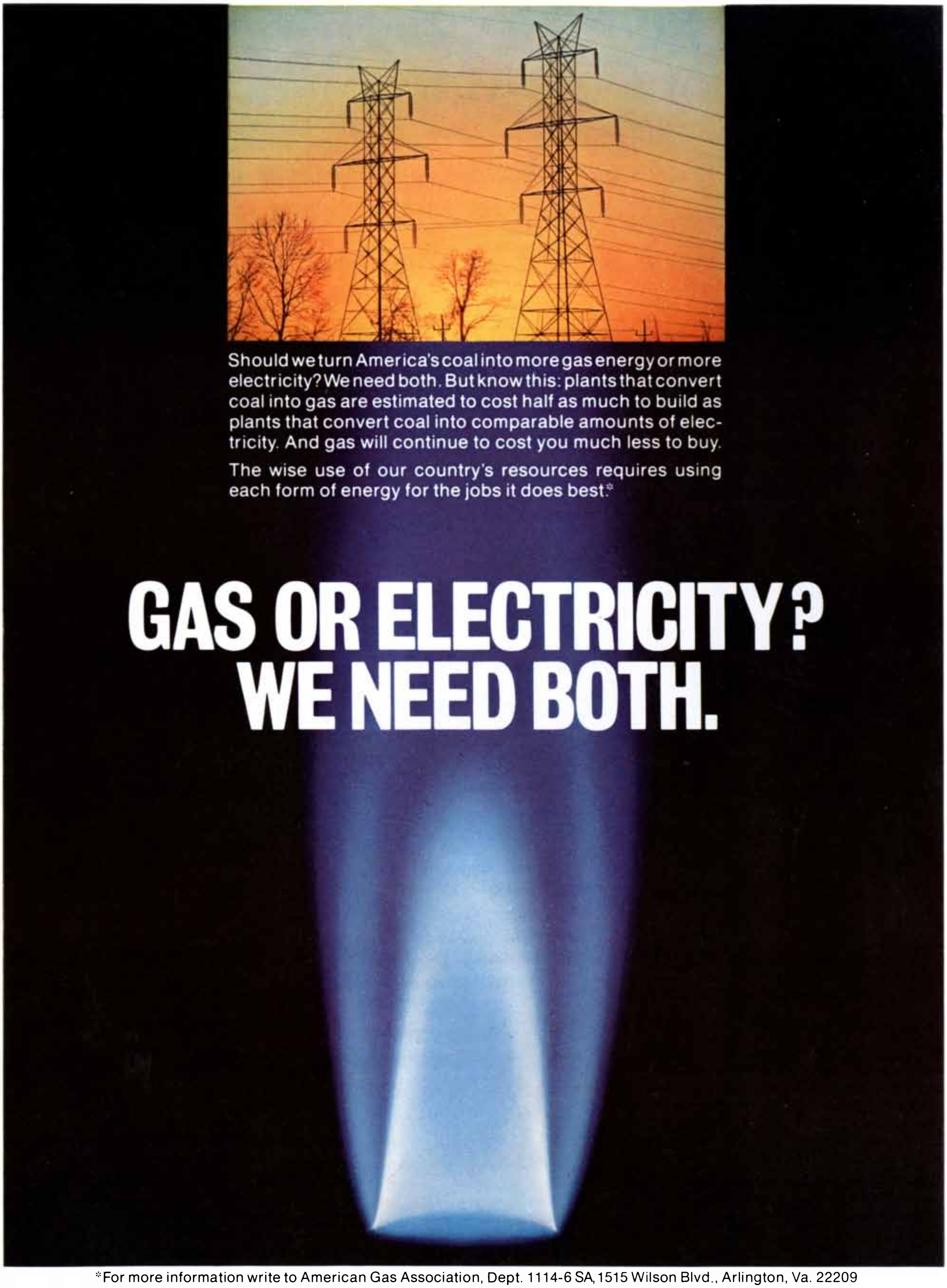

ROLL = 30.00
PITCH = 30.00
YAW = 30.00
DESIGN YOUR FUTURE.
Engineering
Electronics
Laser Technology
TECHNOLOGICAL CHALLENGES. Lasers. Space systems design. Nuclear energy. Air Force engineers and scientists are at the forefront of today's fast-moving technological frontiers.
WANT TO JOIN THE TEAM? If you're graduating with or have a degree in engineering, computer science, mathematics, meteorology, or physics, Air Force Officer Training School could be your first step.
Air Force officers receive good salaries, medical care, 30 days paid vacation a year, and often postgraduate degree opportunities.
Air Force ROTC scholarships may be available if you have at least two years of college remaining. They offer full tuition, fees and $100 a month tax free.
TALK TO YOUR AIR FORCE RECRUITER OR AFROTC CAMPUS REPRESENTATIVE. Mail in the coupon. Or call toll free 800-523-5000 (PA: 800-362-5696).
It's a great way of life!
AIR FORCE
Air Force Opportunities
ASA059
P.O. Box 1776
Valley Forge, PA 19481
Please send me information on engineering opportunities and becoming an Air Force Officer. I am especially interested in:
Air Force Officer Training School ☐ (maximum age 29½—college degree required).
Air Force ROTC ☐ (age 16-28).
Name ____ Sex( )M( )F
(Please Print)
Address ____ Apt. # ____
City ____ State ____
Zip ____ Phone ____
Date of Birth ____
Engineering Discipline ____
College Attending ____
Date of Graduation ____ Degree ____

spheric phenomenon because the reduction in starlight of almost four magnitudes was too large.

It is now thought that the six secondary eclipses were caused by six satellites of Herculina. The longest secondary occultation has been confirmed by independent observation. This occultation was caused by a secondary body about 50 kilometers in diameter and about 1,000 kilometers from Herculina, whose diameter is 220 kilometers. Theoretical considerations indicate that Herculina could hold in its gravitational field an object in orbit as much as 30,000 kilometers away. Since the confirmed minor satellite is only 1,000 kilometers from Herculina, the gravitational interaction is highly stable. Occultation data yield 23 other candidates for minor satellites orbiting eight asteroids. Independent observations of these occultations are needed before they can be definitely attributed to satellites.

Binary asteroids were detected because of the peculiar characteristics of the light reflected from them. Since the shape of an asteroid is irregular, the body varies in brightness as it rotates. Irregularity of shape, however, cannot account for the light curves (plots of variation in brightness as a function of time) of the asteroids 49 Pales and 171 Ophelia, according to another communication in *Science* from Edward F. Tedesco of the University of Arizona. Both light curves show a constant brightness for slightly more than half of the time and a *V*-shaped variation in brightness for the rest of the time. The simplest interpretation of the light curves is that they were generated by a self-eclipsing binary system. Two asteroids in synchronous orbit around each other would appear to be of constant brightness when they were not eclipsing each other (just over half of the time) and of varying brightness when they were eclipsing each other (just under half of the time).

The light curve of 624 Hektor, one of the Trojan asteroids whose orbits are locked in resonance with Jupiter, suggests that the asteroid has the shape of either a dumbbell or a fat cigar 3.1 times as long as it is wide. The irregular form of an asteroid is thought to be the result of collisions with other asteroids, and it is unlikely that a collision would give rise to an oblong object. Accordingly William K. Hartmann of the Planetary Sciences Institute in Tucson, Ariz., and Dale Cruikshank of the University of Hawaii propose that Hektor is two asteroids in contact. Their hypothesis will be put to the test when the body's reflection spectra are examined. If Hektor consists of two asteroids, the probability is high that they differ slightly in composition. In that case the reflection spectrum of one end of Hektor should differ from that of the other. If, on the other hand, Hektor is a single oblong object, the two spectra should be the same.


Should we turn America's coal into more gas energy or more electricity? We need both. But know this: plants that convert coal into gas are estimated to cost half as much to build as plants that convert coal into comparable amounts of electricity. And gas will continue to cost you much less to buy.
The wise use of our country's resources requires using each form of energy for the jobs it does best.*
GAS OR ELECTRICITY?
WE NEED BOTH.

The Molecular Basis of Cell Movement

Not only muscle cells but also other cells are capable of movement, from creeping over a surface to the internal streaming of cytoplasm. All these movements are engineered by a small number of proteins

by Elias Lazarides and Jean Paul Revel

Movement is one of the most striking properties of living matter. It underlies nearly every manifestation of life and is often the most obvious sign that something is alive. Muscle cells are notably motile, but many nonmuscle cells are able to creep about and actively change their shape. For example, lymphocytes and macrophages, the white blood cells that combat infection, migrate from the bloodstream into an open wound to engulf invading bacteria. The cells of the developing embryo perform a precisely choreographed series of movements that gives rise to the different tissue layers. Cells also display a wide variety of internal movements, such as the streaming of the cytoplasm, the secretion of cell products from vesicles, the engulfment of fluid or solid matter and the separation of the paired chromosomes in the course of cell division.

In the past few years the study of the mechanisms of cell motility has been facilitated by the development of new methods, in particular the use of fluorescent antibodies, to localize specific structural proteins within living cells. Such approaches have led to the conclusion that the diverse movements of cells, from the contraction of muscle fibers to the migration of lymphocytes, have common molecular mechanisms. Cell motility is mediated by a small set of contractile proteins, present in both muscle and nonmuscle cells, whose function is controlled by a larger set of regulatory proteins. The subtle and complex interactions of these proteins give rise to the broad diversity of movements of which cells are capable.

The fact that certain cells move has been known since Anton van Leeuwenhoek observed swimming sperm cells in his primitive microscope in the 17th century, but the study of cell motility began in earnest with the development in this century of methods for culturing animal tissues in laboratory glassware. Of particular importance was the work of Warren H. Lewis of the Department of Embryology of the Carnegie Institution of Washington, who made time-lapse motion pictures of cells grown in culture. In order to grow fibroblasts (embryonic connective-tissue cells) he treated the tissue with protein-cleaving enzymes to break up the adhesive contacts and networks of extracellular material holding the cells together. He then separated the resulting clumps of cells mechanically and grew the individual cells in a nutrient medium.

How Cells Move

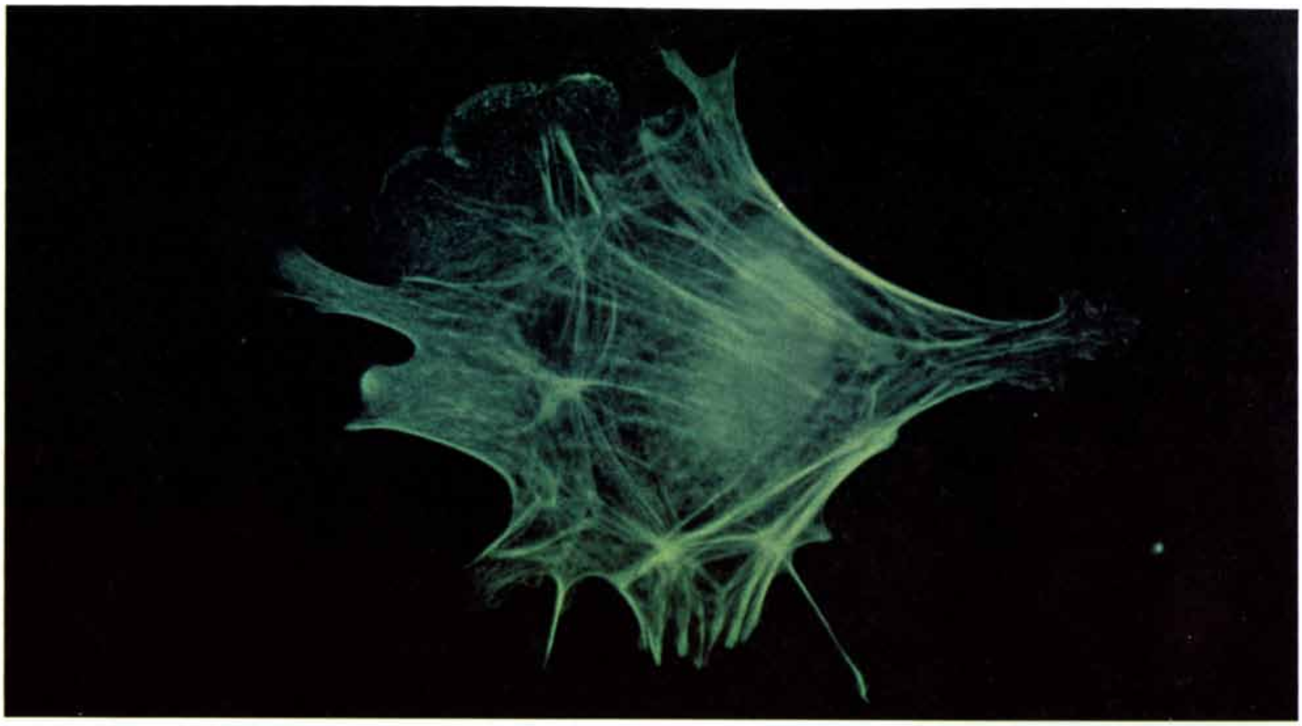
When Lewis examined the fibroblasts under the microscope after a few hours in culture, they had acquired a polygonal shape. Along the shortest side of the polygon were lamellipodia: delicate sheetlike extensions of the cytoplasm. The movements of the lamellipodia were so slow that they were difficult to discern, but in the time-lapse motion pictures it became apparent that, in Lewis' words, "the ruffled pseudopodia [slowly bend] back and forth like the ruffles of a dress in a slight breeze." The word "ruffle" is so evocative that it is still often used to describe the lamellipodia and their movements.

Vernon M. Ingram, Michael Abercrombie and their colleagues at University College London subsequently made time-lapse motion pictures of migrating fibroblasts from the side rather than from above, which showed that the ruffled membrane is the main locomotory organ of the cells. The lamellipodium forms transient contacts with the substrate on which the cell rests, stretching the rest of the cell among several temporary adhesions. When the cell is spread out to its fullest extent, it maintains one or two primary ruffles, each of which tends to lead the cell in a particular direction. The cell therefore migrates with frequent changes of direction, which typically come at intervals of several hours.

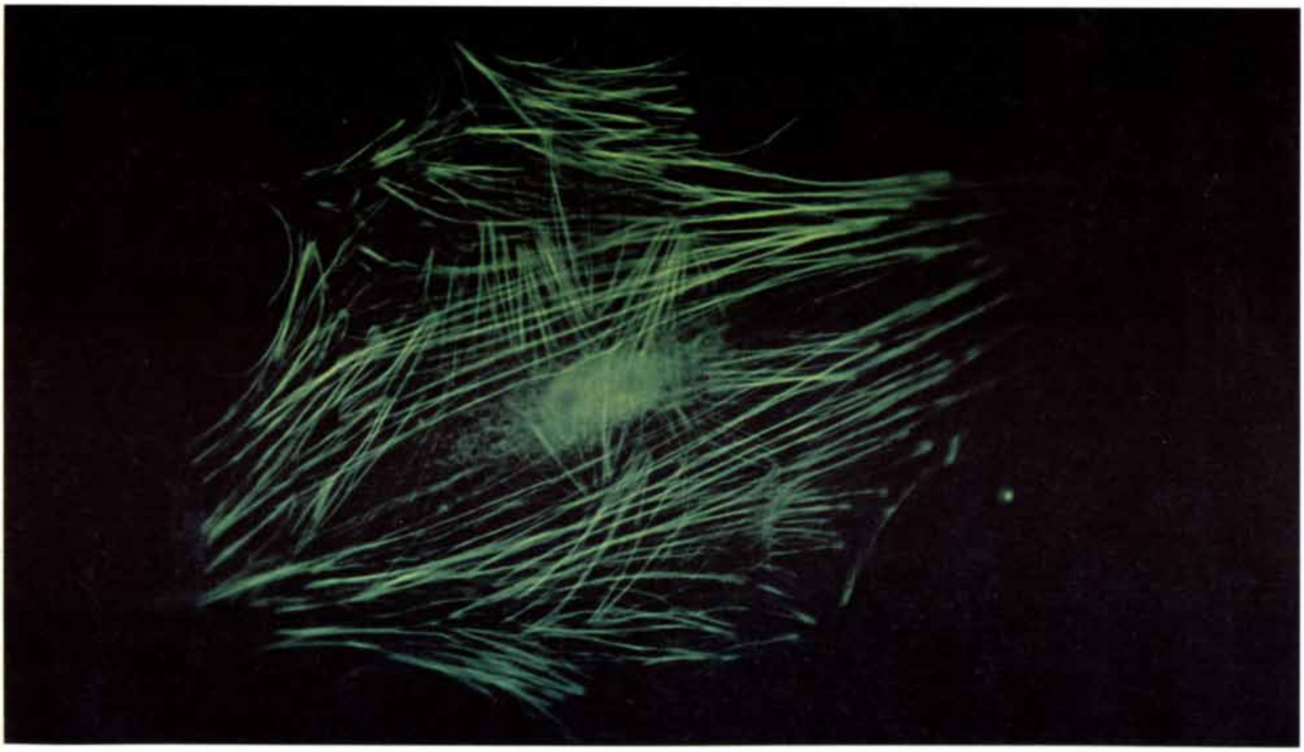
The sequence of events in the movement of a fibroblast across a glass culture dish is as follows. The lamellipodium first extends forward more or less parallel to the bottom of the dish and makes contact with the glass. The length of the ruffle is usually between two and 10 micrometers, but it may be as long as 20. If the tip of the ruffle adheres to the substrate, the cell tends to pull in its rear margins. Hence as the leading edge moves outward the rest of the cell moves with it. If the ruffle does not stick to the substrate, it folds back over the upper surface of the cell, collapses onto it and then disappears into the cell.

The structural details of the ruffling membrane are seen most dramatically in the scanning electron microscope, which provides both a higher resolution and a greater depth of field than the light microscope. At high magnification it becomes apparent that the lamellipodium is a somewhat irregular sheet of cytoplasm about .1 micrometer thick, decorated along the edges by small, regularly spaced protrusions resembling stubby fingers. These protrusions attach themselves to the surface of the dish. They can also quickly extend several times their initial length into rodlike filopodia. Filopodia are highly dynamic structures that can change from a fluid state to a rigid one in less than a second. In growing out from the cell surface they provide spatial guides between which the cell membrane flows to form the ruffle. As the filopodia form they wave freely in the medium and either attach themselves to the substrate and become rigid or melt back into the cell.

The cell does not adhere to the substrate with the entire surface of its bottom membrane. Instead it makes contact at a small number of sites, leaving



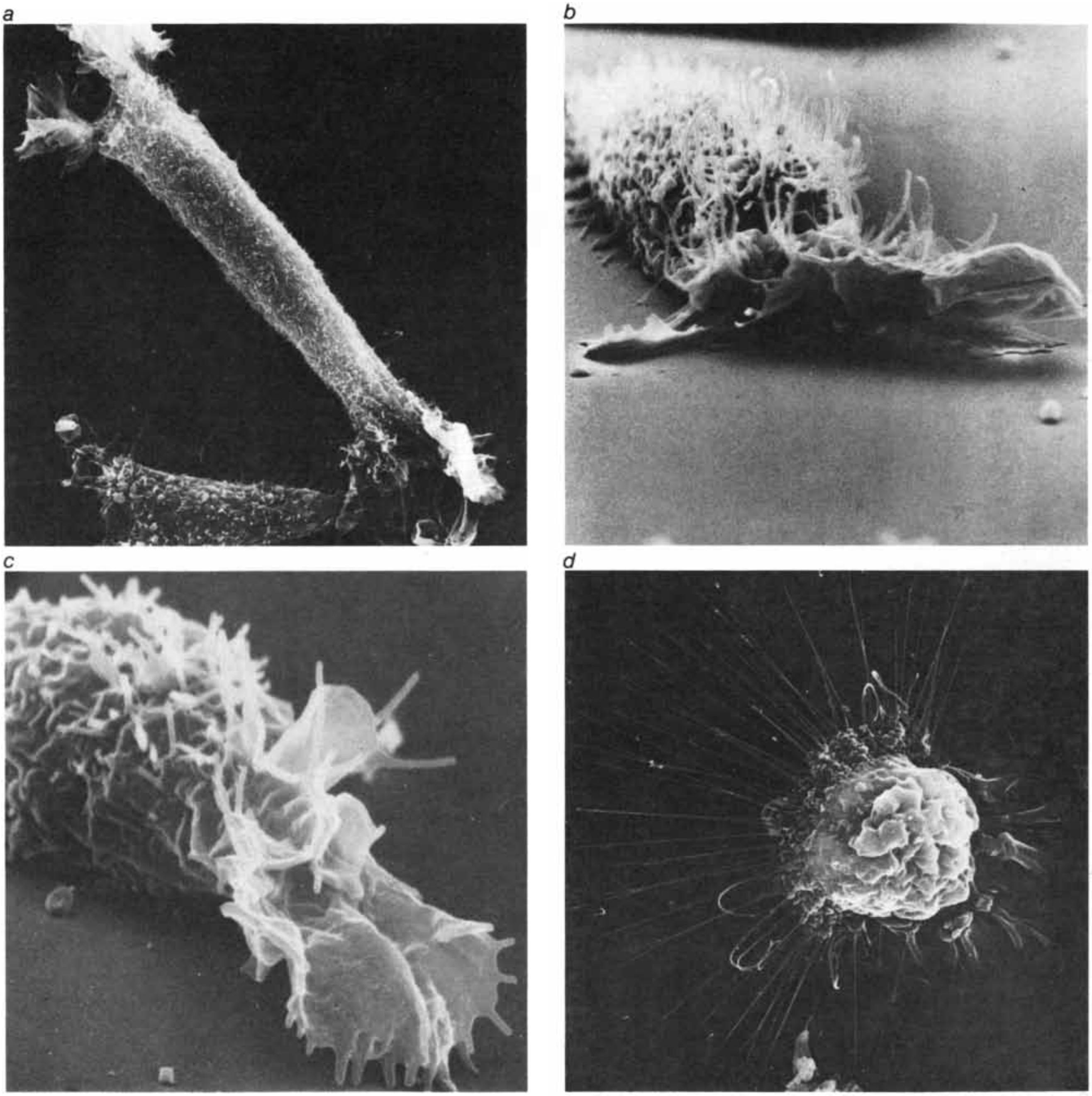
MOVING CELL in tissue culture displays a characteristic pattern of fluorescence when it is treated with two sets of antibodies: the first set specific to the contractile protein actin and the second set specific to the first antibody and labeled with a fluorescent chemical group. This technique, known as indirect immunofluorescence, reveals that most of the actin filaments in the moving cell are in a diffuse meshwork state. Also apparent at the margins of the cell are thin, ruffling sheets of cytoplasm called lamellipodia, which form transient adhesions with the substrate and slowly pull the cell across the surface of the culture dish. This particular cell has several lamellipodia and a few thin extensions called filopodia, which appear to serve a "sensory" function by detecting other migrating cells. The cell nucleus, located in the center of the cell, is brightly fluorescent. The magnification of this immunofluorescence micrograph is 1,500 diameters.



RESTING CELL spread out over the surface of the culture dish has its actin filaments organized into linear bundles, some of which span the entire length of the cell, as is clear in this immunofluorescence micrograph. When the cell begins to move, the filament bundles will disassemble to form a diffuse mesh like that of the motile cell in the illustration at the top of the page. The form of the actin filaments in the cell is directly correlated with its state of motility, so that actin appears to play a central role in the molecular mechanism of cell movement.

the rest of the membrane unattached, rather like a hand placed flat on a table with only the fingertips in contact with the table. Depending on the degree of motility of the cell, the attachment sites, known as adhesion plaques, may be numerous or very few; more motile cells have fewer plaques. Any permanent contact between the ruffling membrane or a filopodium and the substrate results in the formation of an adhesion plaque, and as the cell moves across the substrate the plaques are continuously formed and broken. The plaques tend to stick, so that as the cell moves, long fibers attached to the plaques are dragged behind it. These traction fibers stretch until they are broken by tension and are left behind.

In addition to their role in cell adhesion and motility the ruffles, and in particular the filopodia, serve a sensory function. In the developing embryo when the filopodium of one migrating nerve cell makes contact with the surface of another, the cells stop moving toward each other. Similarly, when the filopodium of a cell spreading in tissue culture touches the surface of a cell that is already flattened, the membrane of



INDIVIDUAL CELLS in tissue culture are depicted in this gallery of scanning electron micrographs made by one of the authors (Revel). In *a* two *LAq* cells (a cell line derived from mouse fibroblasts) are migrating across the substrate. The delicate feathery structure at the top left is a "ruffle," or lamellipodium, which marks the leading edge of the cell. In *b* a cell is shown from an oblique angle as it moves across the substrate, extending its ruffle to form new adhesions. In *c* a single ruffle is shown close up. The fingerlike projections spaced regularly along the edge of the cytoplasmic sheet form adhesion sites with the substrate. In *d* a flattened cell has been induced to round up and retract into itself by incubating it for 30 minutes at four degrees Celsius. The cell has left behind long retraction fibers that are still attached to the substrate through large flattened pads called adhesion plaques. The rounding up can also be induced by the enzyme trypsin.

the first cell will flow in a direction opposite to the point of contact. Cell movements therefore appear to be determined by some kind of chemical computer, the nature of which is beyond our present understanding [see "The Tracks of Moving Cells," by Guenter Albrecht-Buehler; SCIENTIFIC AMERICAN, April, 1978].

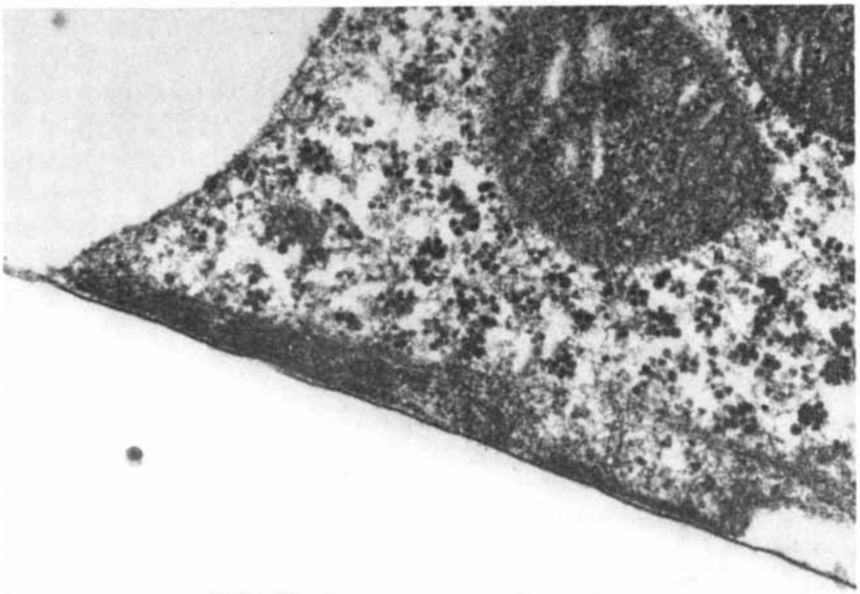
Mechanisms of Motility

What are the molecular mechanisms that underlie cell motility? The existence of one such mechanism was suggested in 1952 by the experiments of Ariel G. Loewy of Haverford College with *Physarum polycephalum*, a slime mold consisting of a large streaming mass of cytoplasm. Loewy wanted to understand how an unspecialized tissue such as this one could convert chemical energy into mechanical work. At that time how muscle accomplishes the same feat was only beginning to be investigated, but the primary cell components responsible for muscle contraction had been identified; they were the proteins actin and myosin.

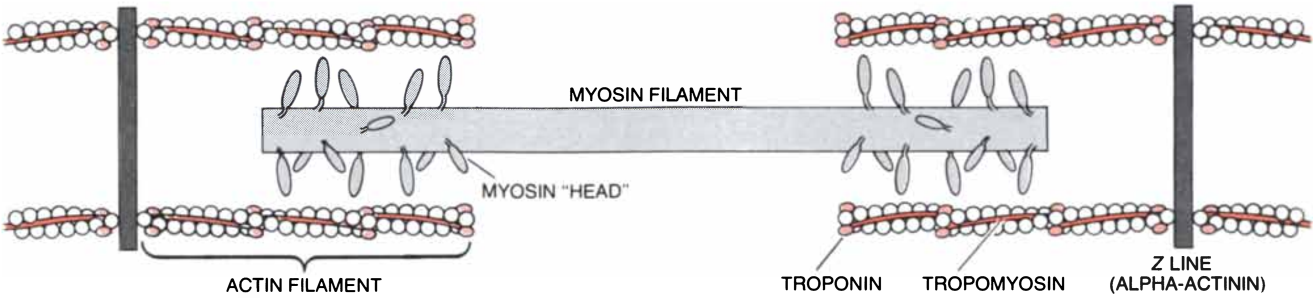
Loewy found that extracts of the slime mold responded to the addition of the energy-rich substance adenosine triphosphate (ATP) with a decrease in viscosity and the removal of the ATP molecule's energy-carrying terminal phosphate group. He concluded he was characterizing a system that was capable of structural change under the influence of ATP and hence was similar to muscle. The reversibility of the changes in viscosity led him to hypothesize that an adequate supply of ATP could cause the system to undergo cyclic alterations in structure and thereby mediate the conversion of chemical energy into the mechanical work required for the locomotion of the slime mold.

By the late 1950's it was known from the work of H. E. Huxley and his coworkers at University College London that the contraction of muscle is achieved by the sliding of filaments of actin and myosin past one another. The process is driven by ATP and is controlled by the presence or absence of calcium ions. In the presence of calcium ATP drives the cyclic formation and dissociation of bridges between the actin and myosin filaments, so that the filaments are moved past one another in a ratcheting action.

The idea that contractile proteins might also be involved in the motility of cells other than muscle cells was in the air before 1960, but it was far from widely accepted. A major stumbling block was the fact that actin and myosin from nonmuscle cells had not yet been purified and shown to be similar to the actin and myosin isolated from muscle. The turning point came in 1966, when



BUNDLES OF ACTIN FILAMENTS are associated with the membrane of animal cells, as is demonstrated in these two electron micrographs. The image at the top was made by fixing the cell and then removing all but the underlying membrane with a jet of water. A replica of the remaining material was prepared by coating it with carbon, and the replica was then examined in the electron microscope. The large fragment of membrane that appears in the micrograph is anchored to the substrate by only a few adhesion plaques, but because the membrane is reinforced by the filament bundles these discrete contacts were sufficient to keep it attached. The micrograph at the bottom is of a thin section of a cell attached to the substrate. The actin filament bundles that are associated with the underlying membrane of the cell are readily apparent.



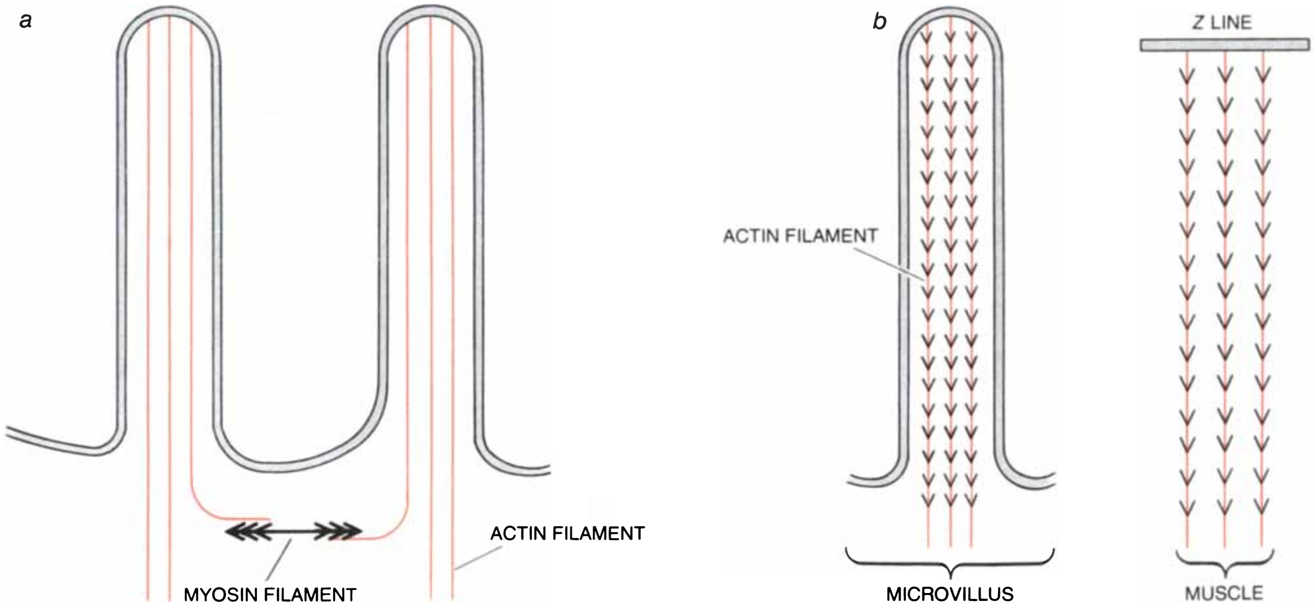
INTERDIGITATING FILAMENTS of actin and myosin form an orderly array within muscle cells. Here a single actin-myosin complex is diagrammed. The actin filaments have a specific polarity, with filaments of opposite polarity jutting out from each side of the *Z* lines, or attachment sites. The myosin filaments are bipolar, with globular "heads" projecting from each end; these heads form cross bridges with the actin filaments. Two regulatory proteins associated with actin, tropomyosin and troponin, allow the actin filaments to interact with the myosin heads only in the presence of calcium ions. Muscle contraction is mediated by the sliding of the actin and myosin filaments past one another powered by a ratcheting action of the myosin heads. A similar sliding interaction seems to take place in nonmuscle cells.

Sadashi Hatano and Fumio Osawa of Nagoya University purified actin from the slime mold *Physarum polycephalum* and demonstrated that the molecule had physical, chemical and biological properties very similar to those of muscle actin.

Nonmuscle Actin and Myosin

Actin was subsequently identified in nonmuscle cells covering the entire evolutionary range of organisms from the most primitive to the most advanced. It constitutes as much as 15 percent of the total protein in actively motile cells such as amoebas and human blood platelets, from 5 to 10 percent of the protein in migrating fibroblasts grown in tissue culture and about 2 percent of the protein in organ tissues such as those of the mammalian liver. In 1969 Harunori Ishikawa and his colleagues at the University of Pennsylvania School of Medicine demonstrated that actin is the major component of microfilaments, an important class of intracellular fibers.

To demonstrate that microfilaments contain actin Ishikawa and his co-workers turned to a technique that had been developed by Huxley for studying the mechanism of muscle contraction. Huxley had found that projecting "heads" on the myosin molecule, which act to cross-link the actin filaments in muscle cells, could be cleaved from the rest of the molecule with the enzyme trypsin. The head fragments, termed heavy meromyosin, retained the ability to bind to actin and formed characteristic complexes that, when they were viewed in the electron microscope, resembled strings of arrowheads pointing along the axis of the actin filaments. Ishikawa and his colleagues treated various kinds of nonmuscle cells with heavy meromyosin and observed arrowhead complexes associated with one class of subcellular fibers: the microfilaments. The complexes were not associated with other classes of fibers, such as those known



MICROVILLI are tiny cytoplasmic protrusions that extend out from the epithelial cells lining the small intestine, greatly increasing the surface for the absorption of nutrients. Microvilli contain both actin and myosin filaments and are known to contract much like muscle cells, and so they provide a convincing example of nonmuscle movement mediated by sliding filaments of actin and myosin. As is shown in *a,* bundles of actin filaments project upward inside each microvillus; the myosin filaments are localized at the base of the microvilli. In *b* the orientation of the actin filaments was determined by treating the microvilli with isolated head fragments from muscle myosin, termed heavy meromyosin; these fragments retain the ability to bind to actin filaments. When the head fragments are applied to muscle cells, they form "arrowhead" complexes with the actin filaments that point in the direction of the filaments. When heavy meromyosin was added to microvilli, the head fragments formed arrowhead complexes with the actin filaments that pointed downward from the attachment sites in the tips of the microvilli. The actin filaments within the microvilli are therefore analogous to the actin filament arrays of muscle cells.

as microtubules and 100-angstrom filaments. These findings indicated that the various types of fibers differ in their chemical composition. The "decorated actin" technique has since been widely applied as a specific probe for actin-containing filaments in cells.

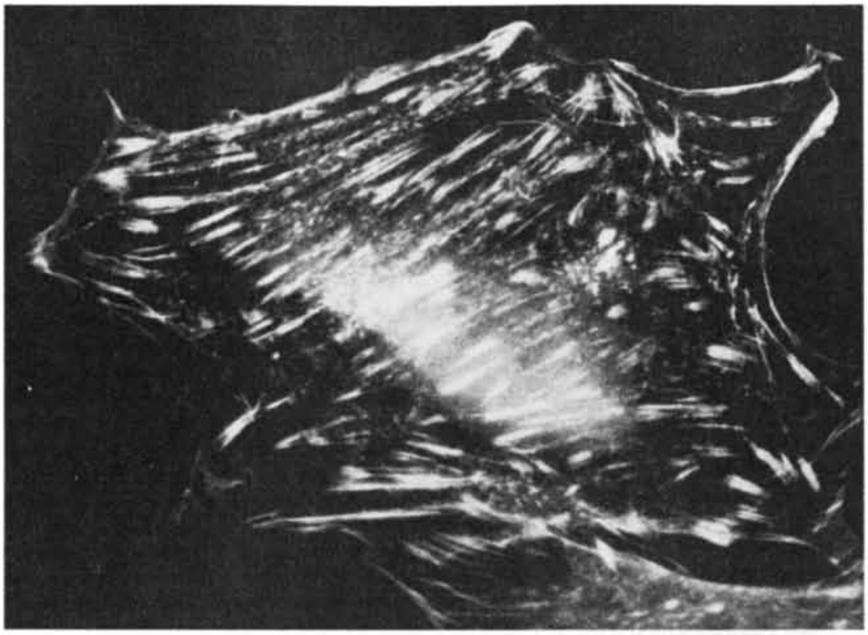
Microfilaments have an average diameter of 60 angstrom units, and they assume a variety of configurations depending on the type of cell and the state it is in. The examination of cultured cells in the electron microscope has shown that microfilaments run a considerable distance through the cytoplasm in the form of bundles. Ian Buckley and Keith R. Porter of Harvard University, who were the first to identify the microfilament bundles, realized that they must be important for determining the elongated shape of the cell and for enabling the cell to adhere to the substrate and spread out on it. In a single migratory or resting cell most of the microfilament bundles (which Warren Lewis had proposed to call stress fibers) are located just below the cell membrane and are frequently in close association with it.

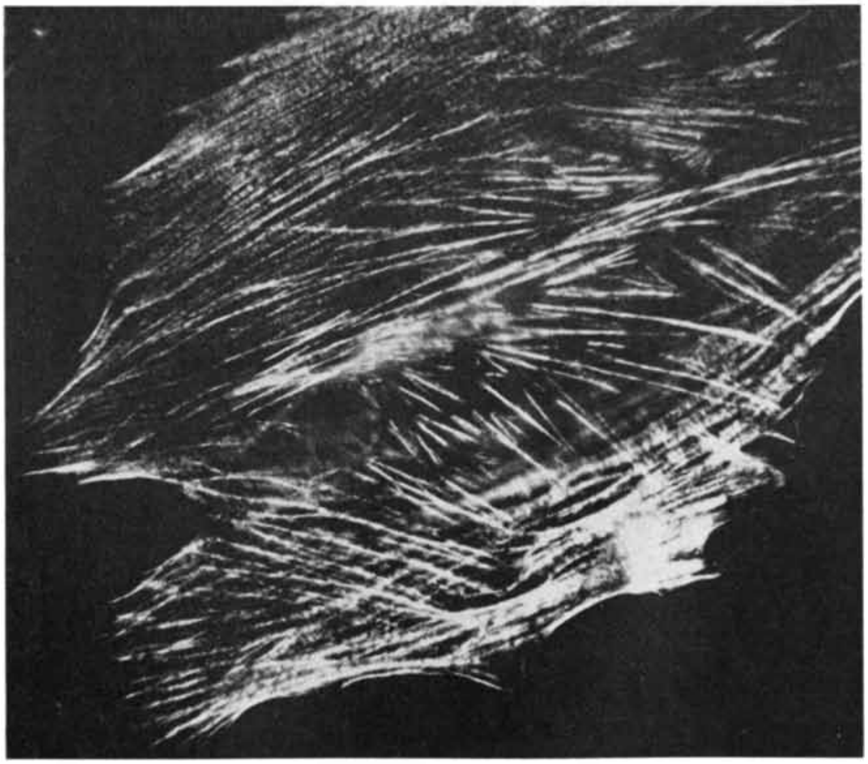
The adhesion plaques formed by moving cells are the sites where the microfilament bundles make contact with the cell membrane. Most of the bundles are oriented parallel to the long axis of the cell and shift their orientation as the cell changes direction. It has been estimated from time-lapse motion pictures that the bundles reorient themselves over a period of from 15 to 60 minutes. The oriented microfilament bundles rarely extend into the outer parts of the ruffling membrane.

Microfilaments can exist in forms other than straight bundles. In rounded cells that do not adhere strongly to the substrate (such as dividing cells and tumor cells) the microfilaments form an amorphous meshwork that is quite distinct from the highly organized bundles. The meshwork is commonly located in the cell cortex just below the cell membrane (a part of the cytoplasm that is usually free of cellular organelles such as ribosomes and mitochondria) and is also seen in areas of membrane ruffling in motile cells.

Although microfilament bundles appear to be highly organized structures, they can disaggregate rapidly into the meshwork form. The two filamentous states appear to be the extremes of a gradient, and the ratio between them depends on a variety of factors, including the extent of cell motility, the degree of adhesion to the substrate, the particular region of the cytoplasm in which the filaments are found and the stage of the cell cycle between divisions. These observations indicate that at least in a tissue-culture cell there are two interconvertible systems of actin filaments: bundles, which give the cell tensile strength, an adhesive capability and structural support, and meshwork, which adds elastic support. The bundles are found almost exclusively in immotile cells, the meshworks in actively moving ones. Slowly moving cells contain a mixture of the two forms.

Myosin, the other structural protein involved in muscle contraction, has also been isolated from nonmuscle cells. Hatano and Masashi Tazawa of Osaka University were the first to isolate myosin from slime mold in the late 1960's, and early in 1971 Robert S. Adelstein





IMMUNOFLUORESCENCE MICROGRAPHS reveal the distribution of contractile proteins within single cultured fibroblasts. The micrograph at the top was made with antibodies to alpha-actinin, a protein specific to the *Z* lines of muscle cells that is localized in nonmuscle cells at the sites where the actin filament bundles adhere to the plasma membrane. Bright patches of fluorescence appear at the attachment sites, which are widely distributed over the inside of the membrane. The micrograph at the bottom reveals the distribution within a cell of myosin. Micrograph was made by Keigi Fujiwara and Thomas D. Pollard at Harvard Medical School.

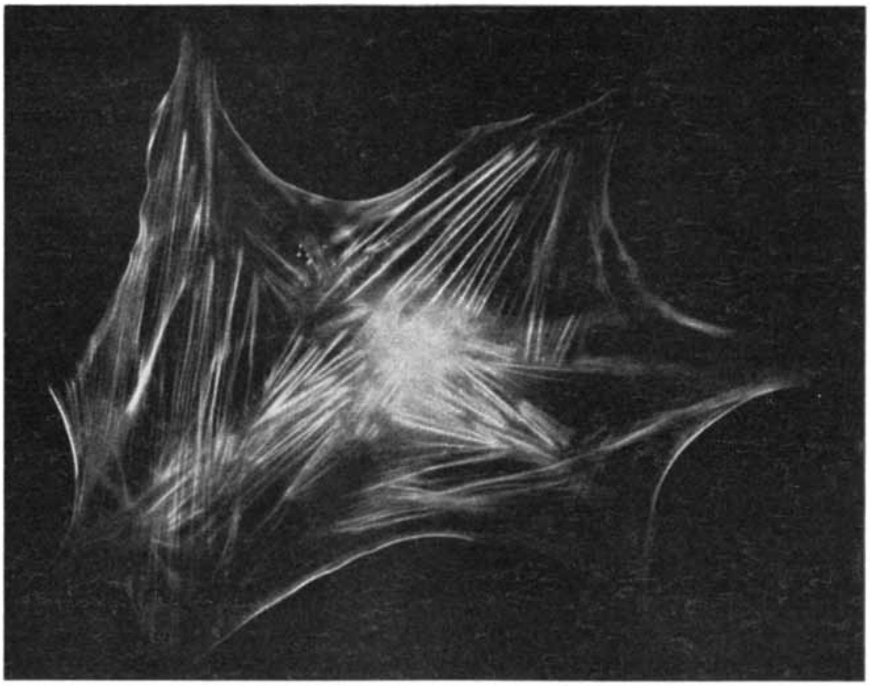
and Thomas D. Pollard, working at the National Institute of Heart and Lung Diseases, purified and characterized myosin from human blood platelets. All the myosins extracted from nonmuscle cells share the ability to cleave ATP in the presence of actin, a characteristic feature of muscle myosin. Nonmuscle myosins are not, however, uniform in their ability to cleave ATP under different ionic conditions or in the presence or absence of actin.

Although these differences among nonmuscle myosins have not been closely examined, they may reflect differing functional roles for such proteins. For example, the migration of a cell across a substrate takes place over a period of minutes, whereas the contraction of a muscle fiber is achieved in milliseconds. Nature may therefore have evolved an entire battery of myosins, each of which has a different rate of ATP consumption and is regulated in a different way, depending on the type of motile activity in which it is involved.

Like actin, myosin has the inherent ability to form long filaments. Muscle myosin forms thick filaments with an average diameter of 150 angstroms; the filaments formed by nonmuscle myosin are shorter and thinner, with an average diameter of less than 60 angstroms, about the same as the diameter of actin filaments. Thus whereas muscle myosin can be easily identified in electron micrographs, nonmuscle myosin filaments cannot be distinguished from actin filaments.

The unique feature of the myosin filaments is that the "heads" at their ends are oriented in opposite directions. The heads are therefore able to cross-link actin filaments running in opposite directions. As we have mentioned, muscle contraction is achieved by the sliding of actin and myosin filaments past one another. This sliding-filament model was proposed by Huxley and has essentially been adopted to explain the involvement of actin and myosin in the motility of nonmuscle cells.

One difference between muscle cells and nonmuscle cells is that in muscle fibers the actin filaments are anchored to flat protein structures called *Z* lines, which are emplaced between every two contractile units. The actin filaments extend perpendicularly from both surfaces of the *Z* line and run in opposite directions on opposite sides. In nonmuscle cells, however, the actin filaments are anchored directly to the cell membrane. It therefore seems conceivable that in nonmuscle cells myosin molecules could cross-link two actin filament bundles anchored to the inside of the cell membrane; the sliding of the myosin and actin filaments past one another could then move the two attachment sites closer together and thereby effect changes in the shape of the cell.



TROPOMYOSIN, a regulatory protein of muscle, is also present in nonmuscle cells, as is apparent in this immunofluorescence micrograph of a cultured fibroblast treated with antibodies to muscle tropomyosin. The fluorescence is evenly localized along the actin filament bundles. In muscle cells tropomyosin regulates the ability of actin filaments to form cross bridges with adjacent myosin filaments. In nonmuscle cells tropomyosin appears to play a different role: it stabilizes the actin filaments into bundles and is absent from the diffuse actin of motile cells.

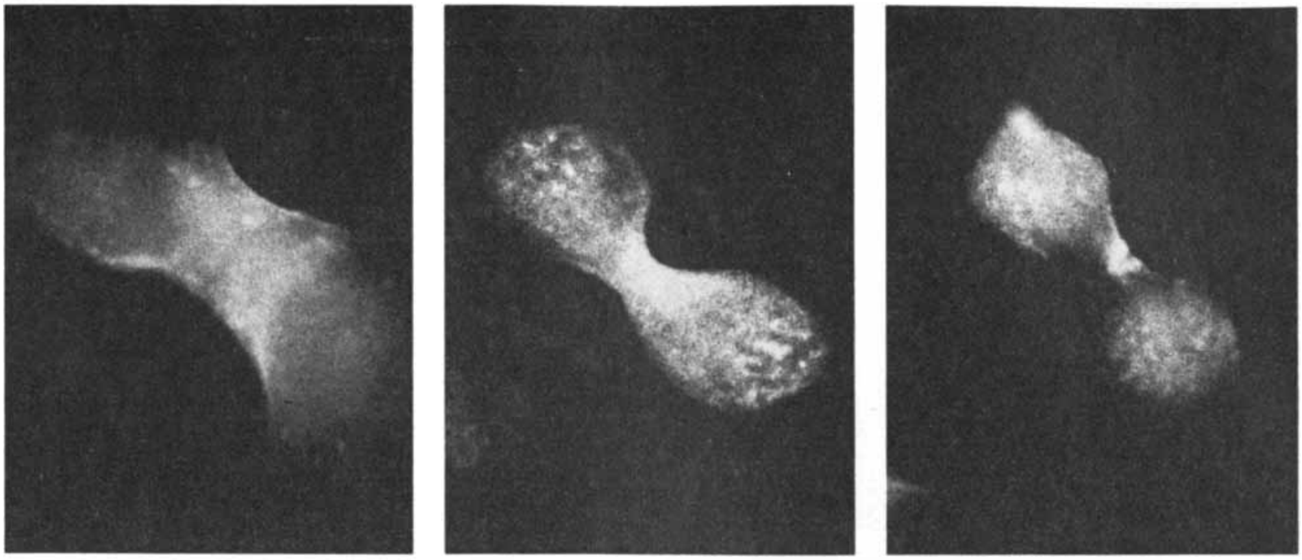
The many questions that remain about the molecular mechanisms of cell motility could be better defined if there were a means of following the distribution of molecules such as actin and myosin within cells during their motile activities. Such a method became a reality when one of us (Lazarides), working on his Ph.D. thesis at the Cold Spring Harbor Laboratory, adapted a technique known as indirect immunofluorescence, which had been developed by Albert H. Coons and his colleagues at the Harvard Medical School in the early 1950's.

Indirect Immunofluorescence

Coons's approach was first to prepare an antibody that would specifically combine with a particular protein he sought to localize in cultured human cells. This was done by injecting the purified human protein into rabbits. Once the rabbit antibody against the human protein had itself been purified he fixed the cultured human cells, made them permeable to antibody by treating them with acetone and then incubated the cells with the rabbit antibody. After an hour of incubation the cells were washed and treated with a second antibody, this one made by injecting the rabbit antibody into goats. The goat antibody, which had been coupled to a fluorescent dye, bound itself to the rabbit antibody, which was already bound to the human protein. With the fluorescence microscope, in which cells can not only be viewed but also be exposed to ultraviolet radiation, it was possible to see the glow of the bound goat antibodies and to determine from it the precise location of the human protein within the cells. The technique was named indirect immunofluorescence because two kinds of antibody were employed rather than one. The advantage of this approach was that it greatly amplified the fluorescence, because several fluorescent goat-antibody molecules often bound to a single rabbit-antibody molecule.

Although indirect immunofluorescence was soon widely applied in biomedical research, for a long time it seemed unlikely that it could reveal the location of actin and other structural proteins within cells. The problem was that the actin of one species closely resembles that of another, and hence actin is a poor inducer of antibody. For example, when native human actin is injected into rabbits, only antibodies to the contaminating proteins are manufactured.

One of us (Lazarides) took the following approach to the problem. Actin from fibroblasts was separated from its contaminating proteins on the basis of molecular weight by means of electrophoresis across a gel in the presence of a strong detergent. The detergent caused the actin to denature and assume the



CONTRACTILE RING made up of bundles of actin and myosin filaments divides a parent cell into two daughter cells in this sequence of immunofluorescence micrographs. The ring is a short-lived structure (lasting for about 10 minutes) that maintains a constant width and thickness even as its circumference and volume steadily decrease. Some of the actin and myosin filaments in the ring must therefore disassemble after they have interacted, a feature that cannot be explained by simple sliding-filament models. Indeed, since most kinds of nonmuscle-cell movement are highly dynamic, they may involve a variety of different interactions among the contractile filaments.

molecular configuration of a random coil. The denatured actin was eluted from the gel and injected into rabbits for the manufacture of antibody.

Surprisingly, the antibodies against the denatured actin reacted not only with it but also with the native (undenatured) actin, and not only with native actin from fibroblasts but also with actin from a wide variety of other kinds of cells. Apparently the denaturation of actin with detergent was sufficient to increase its antibody-inducing power severalfold, yet the antibodies against the denatured antigen were still able to react with the native antigen. The anti-actin antibodies could now be utilized to examine the intracellular distribution of actin with indirect immunofluorescence.

The Distribution of Actin

The fluorescent antibodies revealed the presence inside the cells of an elaborate array of actin filament bundles with a distribution closely resembling that seen in electron micrographs. The bundles frequently spanned the entire cytoplasm in different directions, converged to focal points and were observed in close association with the cell membrane. Their orientation depended on the motile activity of the cell: in actively moving cells they were mostly parallel to the long axis of the cell in the direction of movement, whereas in resting cells they pointed toward the nuclear area in the center of the cell. The most impressive observation was that each cell had its own individual arrangement of actin filament bundles; in this respect no two cells were alike.

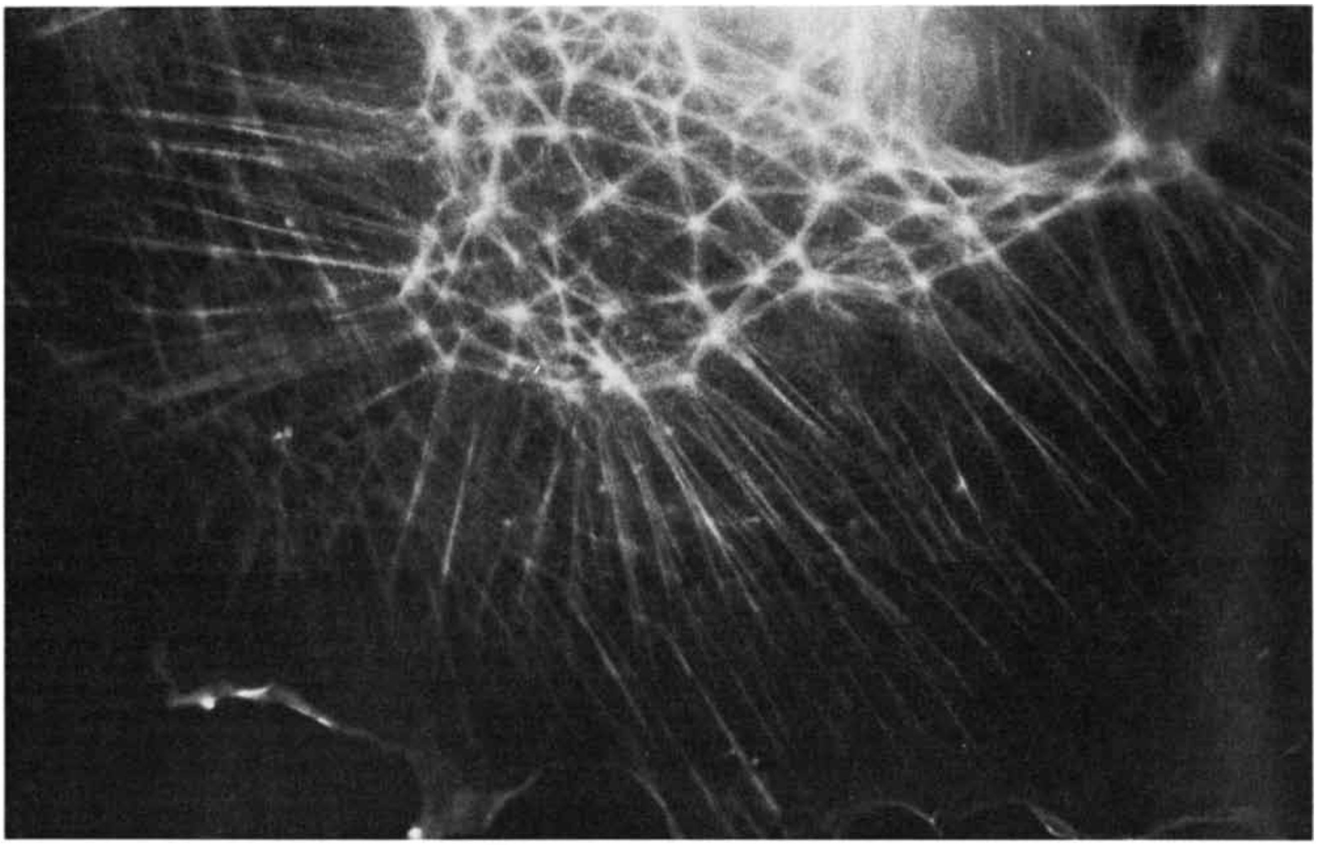
These dramatic results established immunofluorescence as a powerful tool in the study of contractile and other structural proteins within cells. The technique provides a bridge between the biochemistry and morphology of cell structure and the observations of cell motility. Although the electron microscope still provides the ultimate in resolution, the convenience of the immunofluorescence technique makes it even more valuable. Potentially any protein for which an antibody is available can be localized in cells and its function defined in detail. Moreover, immunofluorescence is the only way to study the localization of certain proteins, such as nonmuscle myosin, that cannot be identified in the electron microscope.

Immunofluorescence makes it possible to screen a large number of cells under a variety of experimental conditions, and so now it became feasible to learn about the function of actin, its accessory proteins and other filamentous systems in the cytoplasm by studying their distribution within the cell during different types of cell movement. From even our first experiments many broad generalizations could be made. The fact that in nonmotile cells actin was highly organized into bundles lent credence to the idea that the function of the bundles in this state of the cell is primarily structural: to keep the cell flat and stretched out on the substrate. In highly motile and poorly adherent cells, on the other hand, the opposite extreme prevailed: very few if any organized bundles could be seen, and for the most part the actin filaments were in a meshwork state showing only a diffuse fluorescence. This was particularly true of the actin filaments within membrane ruffles or in close association with the cell membrane. The two populations of actin filaments are not mutually exclusive, however, and any population of cultured cells reveals a gradient between the two extremes.

A second conclusion that emerged was that the high sensitivity of the indirect-immunofluorescence technique would make it possible to detect proteins in the cell in amounts so small that they could not be detected by any other means. Furthermore, it was discovered that actins from a wide variety of cells and species react with antibodies against the actin of muscle cells. This finding implied the possibility of making antibodies against other structural proteins of muscle and using them to localize related proteins in nonmuscle cells. These hopes were confirmed in the experiments that followed.

Alpha-Actinin

The first question we chose to investigate was whether the adhesion plaques of nonmuscle cells, where the actin filaments attach themselves to the cell membrane, are biochemically similar to the *Z*-line structures to which actin filaments are attached in muscle cells. In addition we wanted to know whether the actin filament bundles themselves incorporated structures that could function as organization centers for the fil-



REGULAR NETWORK of actin filament bundles is an intermediate step in the conversion of the diffuse actin filament mesh of an actively moving cell into the highly organized actin filament bundles of a spread-out, immotile cell. This three-dimensional network surrounds the cell nucleus, and its vertexes provide organization sites for actin filament bundles that project toward the periphery of the cell. Once these peripheral bundles have formed, the network itself is converted into linear bundles. Because of the striking similarity between this network and the geodesic domes designed by R. Buckminster Fuller, the authors have termed it the cellular geodome.

aments. To test these hypotheses we needed a fluorescent marker that was present in the *Z* lines of muscle cells. The ideal marker was alpha-actinin, a protein known to be a component of the *Z* lines (although its function is a mystery).

We found that antibodies against alpha-actinin from muscle reacted with a protein present in fibroblasts and epithelial cells grown in tissue culture. Alpha-actinin was intimately associated with the actin filaments in a highly ordered fashion, not only with the filament bundles but also in areas where the filaments were associated with the cell membrane. The protein was particularly prominent in areas of contact between two cells and at adhesion plaques.

The localization of alpha-actinin clearly demonstrated that the adhesion plaques of nonmuscle cells are similar to the *Z* lines of muscle cells. Furthermore, the highly ordered and periodic localization of alpha-actinin within the actin filament bundles demonstrated that the bundles are organized along their length by structures resembling *Z* lines. The impressive point was that in spite of the high degree of organization of the filament bundles they could disaggregate in a matter of seconds. Alpha-actinin was localized not only in the adhesion plaques and the filament bundles but also in the membrane ruffles and the filopodia. We concluded that any part of the cell membrane can function as an attachment and organization site for actin filaments.

Orientation of the Actin Filaments

Before the attachment sites for actin filaments on the plasma membrane of nonmuscle cells could be considered strictly analogous to the *Z*-line structures of muscle cells we had to know the orientation of the actin filaments. Do they point away from the filament attachment sites or toward them? Richard D. Rodewald and his co-workers in Morris J. Karnovsky's laboratory at the Harvard Medical School, and also Mark Mooseker, working in Lewis G. Tilney's laboratory at the University of Pennsylvania, set about answering this question by reacting the actin filaments with heavy meromyosin and determining the orientation of the arrowhead complexes.

As a model system these workers chose the microvilli, or cylindrical protrusions of the cell membrane, in the "brush border" of epithelial cells that lines the small intestine. The brush border contracts in the presence of calcium ions and ATP, much as a muscle cell does, and it continues to do so when it is removed from the intestine and bathed in a nutrient solution. Each microvillus in the brush border has from 20 to 30 actin filaments; they are attached to the cell membrane at the tip of the microvillus and along its length. Filaments resembling those of myosin are also present at the base of each microvillus, where the actin filaments terminate. When the microvillus contracts, the length of the actin filaments does not change, which strongly suggests that the movement is mediated by a sliding-filament system of actin and myosin resembling that of skeletal muscle.

When Mooseker treated microvilli with small "head" fragments of myosin similar to heavy meromyosin, arrowhead complexes appeared along the actin filaments. The arrowheads were uniformly oriented downward away from the filament attachment site at the tip of the microvillus, indicating that the microfilaments had a common orientation. The filopodia of blood platelets and sea-urchin eggs have also been shown to contain membrane-bound actin filament bundles that point inward from the attachment sites. It therefore appears that the function of at least some of the attachment sites in nonmuscle cells is analogous to that of *Z* lines in muscle.

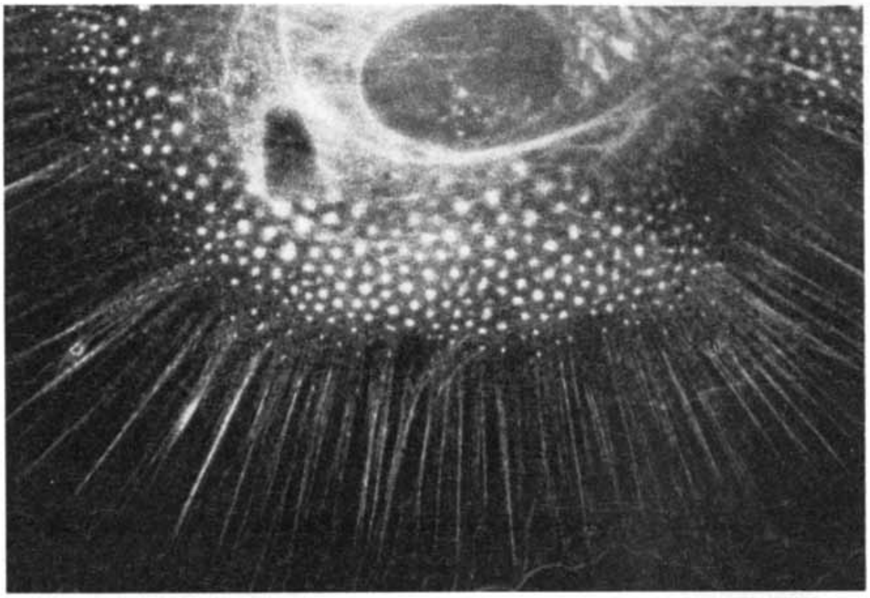
The Contractile Ring

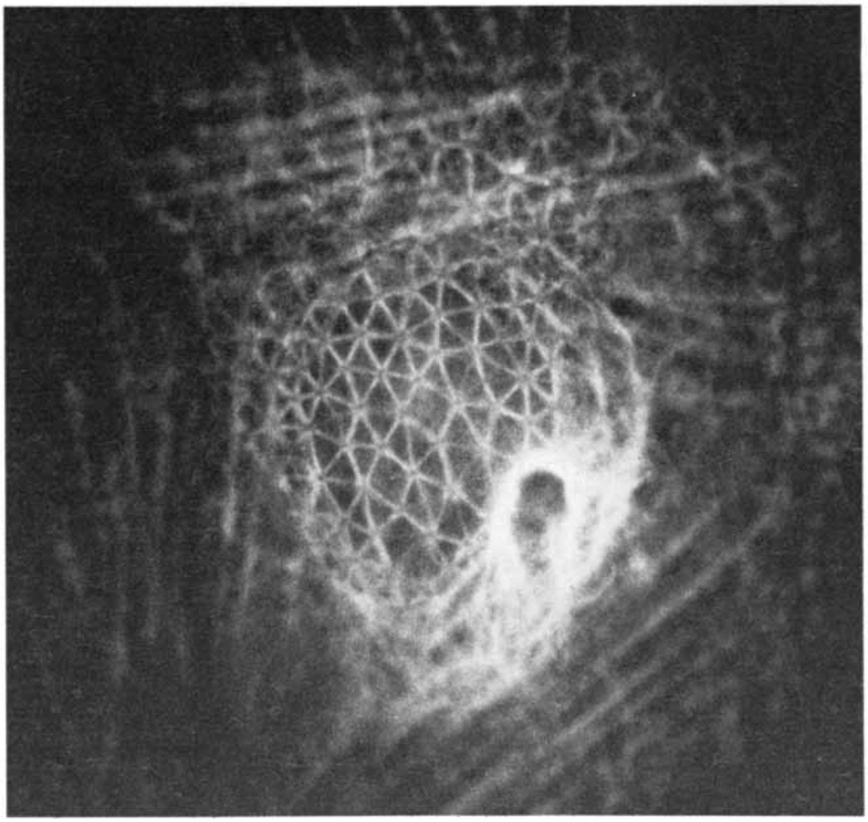
So far we have focused on the role of actin filaments in maintaining cell shape and mediating cell motility. Actin also appears to be involved in the movement of specialized parts of the cytoplasm. For example, during the final stages of cell division, after the duplicated chromosomes have been pulled apart by the mitotic spindle, a constriction develops along the equator of the mother cell perpendicularly to the long axis of the spindle. This constriction, known as the cleavage furrow, progresses until it divides the cytoplasm and its organelles in half, giving rise to two daughter cells.

A measurable tension is exerted at the cleavage furrow. In sand-dollar and sea-urchin eggs this tension has been estimated at about 3×10^5 dynes per square centimeter, a value comparable to the tension developed in muscle. Thomas E. Schroeder of the University of Washington demonstrated that the tension along the cleavage furrow is generated by a contractile ring of circumferentially aligned actin filaments, which encircle the zone of cell constriction. The actin filaments are closely associated with the cell membrane and are confined to the area of the cleavage furrow. Keigi Fujiwara and Pollard, working at the Harvard Medical School, subsequently demonstrated the presence of myosin filaments in the contractile ring of dividing human cells.

Schroeder proposed that the constriction of the ring is mediated by mutual sliding interactions of the actin and myosin filaments. The existence of such interactions was supported by elegant experiments done by Issei Mabuchi and his colleagues at the University of Tokyo. Instead of doing immunofluorescence studies, which require that the entire cell be fixed, they injected specific antibodies into unfixed living cells and observed the effects of this treatment on cytoplasmic motility. When they injected antibodies against myosin into starfish embryos at the two-cell stage in doses sufficient to block all actin-myosin interactions in the cells, the formation of a new cleavage furrow was prevented and the egg stopped dividing. These results clearly demonstrated that an actin-myosin interaction is necessary for cleavage to take place.

The contractile ring is a temporary structure that exists for only about 10 minutes; its rapid assembly and disassembly attest to its formidably dynamic nature. Schroeder found that although the width and thickness of the contractile ring remain constant during its constriction, the volume of the ring decreases. The ring must therefore disassemble even as it contracts, meaning that any sliding interaction of the actin and myosin filaments must be followed by the disaggregation of some of the filaments. Since the width and thickness of the ring remain constant during cleavage, however, the disassembly must take place uniformly throughout the ring during its entire brief lifetime. Therefore in spite





ACCESSORY PROTEINS involved in the construction of the cellular geodome are revealed in these two immunofluorescence micrographs, one micrograph made with antibodies to alpha-actinin (*top*) and the other made with antibodies to tropomyosin (*bottom*). The images show that the geodome network consists of foci containing actin and alpha-actinin, which are connected by actin filaments associated with tropomyosin. The linear actin filaments extending from the vertexes of the geodome network incorporate both tropomyosin and alpha-actinin.

of the presence of actin and myosin in the contractile ring, a simple sliding-filament model is insufficient to explain cell cleavage.

Although the molecular unknowns are still too great for a detailed model of the contractile ring to be formulated, these findings underscore the fact that most motile events within nonmuscle cells are dynamic and transient and hence not strictly analogous to muscle contraction. Moreover, merely associating the presence of actin and myosin with a particular motile event does not necessarily mean that these molecules play a functional role. For example, both actin and myosin have been shown to be constituents of the mitotic spindle, which is made up primarily of microtubules. Antibody-microinjection studies have demonstrated, however, that an actin-myosin interaction is probably not necessary for chromosome movement.

Actin Regulatory Proteins

The next question to be asked was whether any of the proteins involved in the regulation of muscle contraction would be found in association with actin in nonmuscle cells. The interaction of the actin and myosin filaments in muscle is regulated by two proteins: tropomyosin and troponin. When the level of calcium in the muscle cell is low, a complex of the two proteins inhibits the interaction of actin and myosin. When the level of calcium is high, the calcium binds to troponin, which then acts on tropomyosin, initiating a sequence of conformational changes that allows actin and myosin to interact.

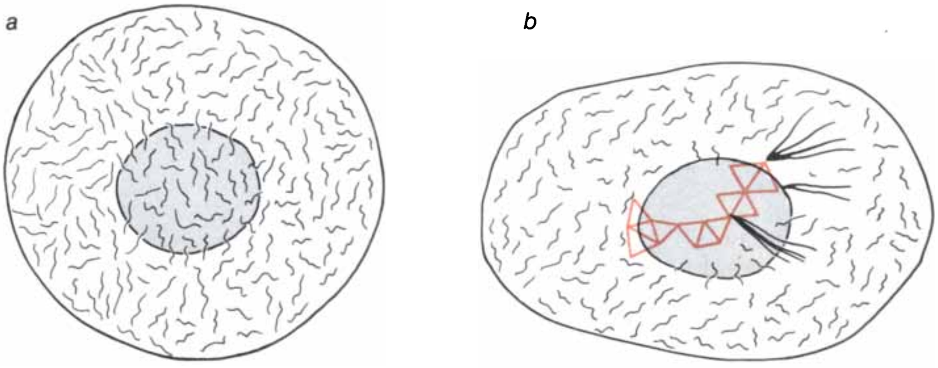
Tropomyosin is one of the most universal regulatory molecules: it is associated with the actin filaments of the muscles in organisms ranging from mollusks to human beings. Immunofluorescence now provided us with a unique opportunity to find out whether the actin filaments observed in nonmuscle cells also had tropomyosin associated with them. We reasoned that the presence or absence of tropomyosin might reflect important regulatory differences between sets of actin filaments involved in different cellular activities.

To our surprise we found that active membrane ruffles and filopodia did not show much, if any, tropomyosin fluorescence, nor did the actin filament bundles of highly motile cells. Although the absence of fluorescence does not necessarily mean that tropomyosin is absent from these areas (they might contain a tropomyosin not immunologically related to muscle tropomyosin), other observations support the conclusion. We found that when the cell is fully adherent or immotile, fluorescent tropomyosin is indeed present along the actin filament bundles and in the membrane ruffles.

The role of tropomyosin in nonmuscle cells therefore appears to be different from that in muscle cells. In nonmuscle cells the protein binds to actin filaments that are involved not in motile activities but rather in maintaining cell shape. The presence of tropomyosin therefore reflects the presence of actin filaments with structural functions, and its absence reflects the presence of actin filaments with motile functions. Thus certain actin filaments in nonmuscle cells do not bind the muscle form of tropomyosin as the actin filaments in muscle cells do.

Since nonmuscle cells contain tropomyosin, by analogy with muscle they should also contain troponin. So far the evidence for the existence of troponin in nonmuscle cells is inconclusive, but proteins analogous to the calcium-binding subunit of muscle troponin known as troponin *C* have been identified. Such observations indicate that calcium ions are important regulators of motility in nonmuscle cells as well as in muscle cells. Indeed, calcium appears to govern the "on-off" state of many molecules involved in one way or another with cell motility.

With the aim of studying in greater detail the association of molecules such as alpha-actinin and tropomyosin with



CONVERSION OF GEL INTO BUNDLES in a flattening tissue cell involves the temporary organization of the actin filaments into a regular geodome structure. The vertexes of this network then serve as organization centers for the actin filament bundles. Most of the actin fila-

the actin filaments, we looked for an experimental system that would enable us to follow the conversion of a cell from a highly motile state into an immobile, adherent state. Such a system was provided by treating cells with a low concentration of a protein-cleaving enzyme such as trypsin. The enzyme-treated cells retract their cytoplasm and cell membrane toward the nuclear area, adhering to the substrate only by means of their retraction fibers. As the cells change their shape from flat to round the actin filament bundles are rapidly converted into a filamentous meshwork.

When the detached cells are allowed to resettle onto the substrate, they slowly flatten out over a period of from six to eight hours, with intensive ruffling activity at their periphery. During this period actin is almost exclusively in the meshwork form and is localized primarily in the areas of membrane ruffling. Gradually, however, the actin filaments begin to be organized into bundles, with a concomitant reduction in the ruffling activity of the margins. This process enables one to study how the bundles are assembled.

The Cellular Geodome

We observed that a number of cells assembled their actin filament bundles into a strikingly regular network, which looked remarkably like a geodesic dome. The network encompassed the entire area above and around the nucleus. Similar regular networks were observed when the cells were treated with fluorescent antibodies against alpha-actinin and tropomyosin: alpha-actinin was localized primarily at the vertexes of the network and tropomyosin was localized along the short fibers connecting the vertexes. We also noted longer fibers that were attached to the vertexes of the network at one end and extended to the edge of the ruffling membrane. These latter fibers contained actin, alpha-actinin and tropomyosin in an arrangement indistinguishable from that of the actin filament bundles in fully spread-out cells.

Quite frequently actin filament bundles originating from the vertexes of the geodesic-dome network were seen to extend into filopodia or lamellipodia. The vertexes of the network therefore act as organization centers for actin filament bundles involved in maintaining the structure of these cellular extensions. When we followed the formation of the network structure with indirect immunofluorescence, we found that both actin and alpha-actinin came together at one vertex in the early stages of cell spreading. There was then a considerable time lag before tropomyosin could be detected in association with the network, suggesting that alpha-actinin determines the binding sites for tropomyosin on the actin filament bundles. These observations also provided further evidence that actin filaments can exist in the cell without any tropomyosin being bound to them.

Soon after the application of indirect immunofluorescence for the localization of actin the technique was adopted by other workers for the localization of other contractile and structural proteins in the cytoplasm, such as the microtubules and the 100-angstrom filaments. A number of laboratories successfully demonstrated the localization of myosin along the actin filament bundles and showed that its distribution varies with the motile activity of the cell. The presence of both myosin and tropomyosin in close association with the actin filament bundles reinforces the idea that these filaments mediate certain motile phenomena by the sliding-filament mechanism. Whether the bundles are in fact capable of contraction by such a mechanism, however, is still undetermined.

Actin Filament Gels

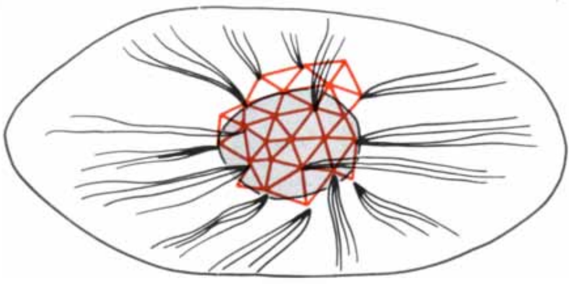
A cell in tissue culture can be found at any point between two extremes: fully resting or fully motile. Parts of the cell such as the filopodia and the membrane ruffles also vary in the degree of their motility. In slowly moving cells parts of the cytoplasm are immobile, and the actin filaments shift, rapidly or slowly, between bundles and meshworks. What are the cytoplasmic factors that regulate the transition of the actin filaments between these two states? How is the structure of the cell maintained once the actin filament bundles are transformed into a meshwork? And how can the meshwork induce movement?

Answers to some of these questions came unexpectedly from test-tube experiments on the polymerization of actin by Robert E. Kane of the University of Hawaii. Kane observed that cytoplasmic extracts of sea-urchin eggs would spontaneously form a gel when they were warmed from four degrees Celsius to 37 degrees, and that this gel was dense with polymerized actin filaments. He then proceeded to show that the gelation was induced by the conversion of monomeric (unpolymerized) actin into actin filaments. Gelation also required the presence of two additional proteins: a protein with a molecular weight of 58,000 daltons that interacted with the individual actin filaments to form bundles, and a 220,000-dalton protein that cross-linked the bundles into a three-dimensional gel.

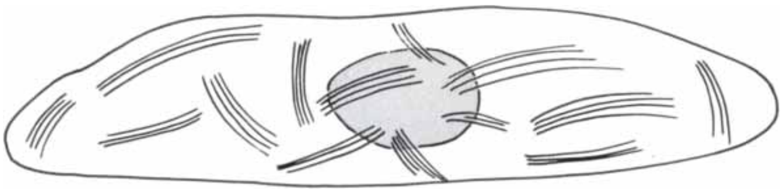
Soon afterward Thomas P. Stossel of the Children's Hospital Medical Center discovered that cytoplasmic extracts of rabbit-lung macrophages would also form a gel. Here, however, only the actin filaments and a 220,000-dalton actin-binding protein were needed for the formation of a three-dimensional cross-linked gel. In the electron microscope the actin gel looked very similar to the meshwork of actin filaments that other workers had described in actively motile areas of the cell, such as membrane ruffles and filopodia. Moreover, the actin gel contracted reversibly in the presence of myosin. Pollard and D. Lansing Taylor obtained similar results with actin extracted from amoebas.

These results gave new insight into the

c



d



ments in a rounded, motile cell are in the diffuse meshwork state (*a*). When the cell stops moving and begins to flatten, some of the filaments organize into a geodome around the nucleus (*b*). Actin filament bundles then assemble at the vertexes of this three-dimensional network (*c*). Finally the network disassembles to leave only linear bundles in the fully flattened cell (*d*). The process takes seven hours.

molecular basis of cell motility. Actin filament gels may be responsible for providing mechanical support to dynamic cytoplasmic structures such as ruffles, filopodia and the cell cortex. Motility could then be achieved by reversible contractions of the actin gel induced by myosin. The cytoplasm of highly motile cells such as macrophages would be permeated by such an actin gel, rendering the cytoplasm highly viscous and providing elastic support for the cell's motile activities.

Other investigators soon identified proteins from a variety of muscle cells that resembled the actin-binding protein of macrophages. Kuan Wang, working in Jon Singer's laboratory at the University of California at San Diego, isolated a protein named filamin from smooth-muscle cells that had properties much like those of the actin-binding protein. In indirect-immunofluorescence studies filamin was localized along the actin filament bundles of nonmuscle cells in a periodic manner and was found in diffuse form in the membrane ruffles and the cell cortex. In the test tube filamin has been shown to induce actin to aggregate into a filamentous gel.

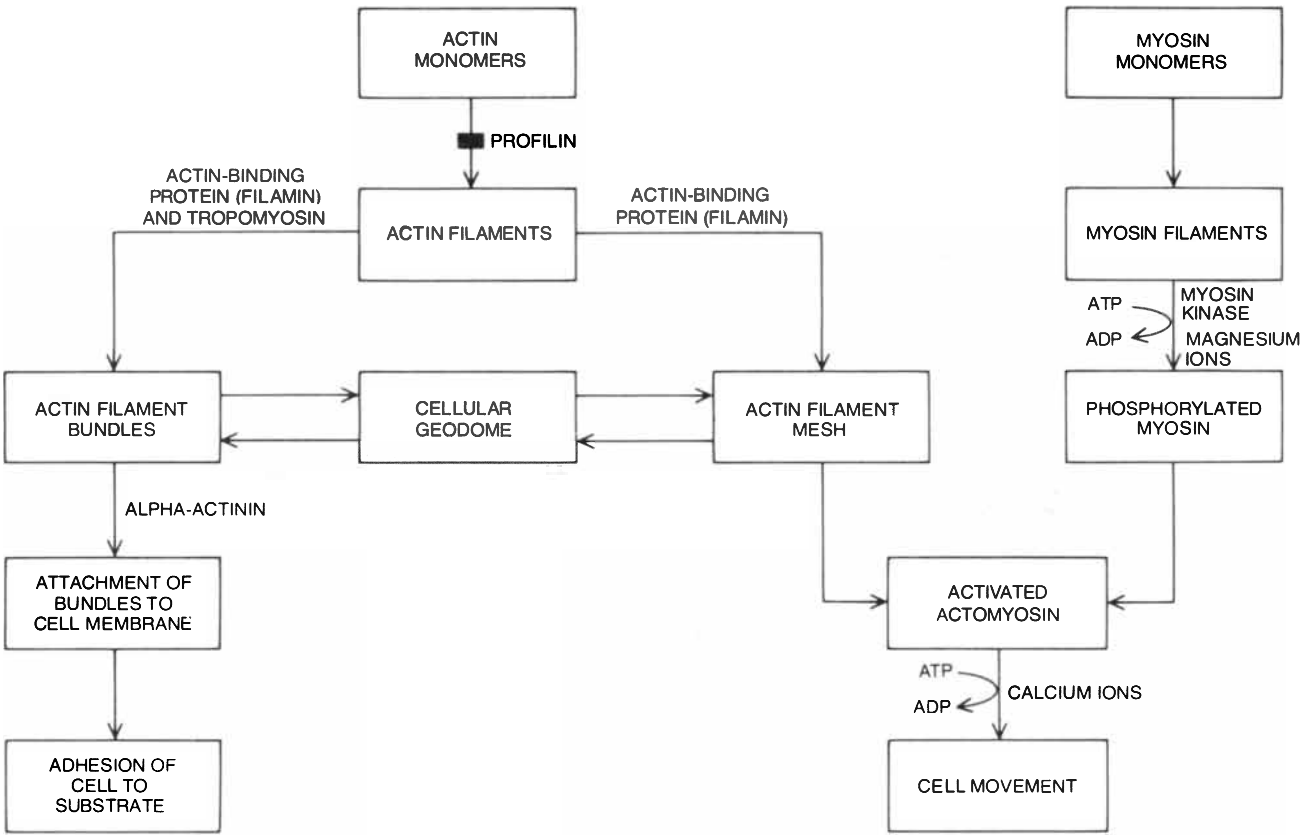
Control of Actin Polymerization

Actin does not always exist in one or another form of filament; it is also present in considerable amounts in the cytoplasm in the monomeric form. An understanding of how the level of unpolymerized actin is regulated came from the work of Uno Lindberg and his colleagues at the University of Uppsala. In 1974 they purified monomeric actin from calf spleen and succeeded in crystallizing it. Structural analysis of the actin crystals revealed that the monomeric actin was complexed with equal amounts of a protein of low molecular weight. Subsequent experiments showed that this protein is responsible for keeping actin in a monomeric state and that its removal allows actin to polymerize into filaments. Lindberg named it profilin. The discovery of this regulatory protein has opened up yet another vista in the understanding of cell motility at the molecular level.

The findings we have discussed so far reveal how diverse the regulatory pathways of actin are. Proteins have been identified that regulate (1) the transformation of monomeric actin into polymeric filaments, (2) the aggregation of these filaments bundles or into a meshwork gel, (3) the transformation of the bundles into the gel and (4) the association of the actin filaments with the cell membrane. In addition to these protein controls there are chemically different actins that polymerize under different ionic conditions, such as the presence or absence of calcium ions. The cell may also impose further controls such as changes in *p*H or the activation and deactivation of regulatory proteins by enzymatic cleavage or chemical modification.

The Regulation of Myosin

Since myosin is involved in at least some types of cytoplasmic motility, it may be regulated by its own set of controls, either independently of or in conjunction with the actin filaments. For ex-



REGULATION OF CELL MOVEMENT at the molecular level in nonmuscle cells is outlined in this flow chart. Actin is regulated at two levels: the polymerization of actin monomers into filaments is inhibited by the regulatory protein profilin, and two additional regulatory proteins, actin-binding protein (filamin) and tropomyosin, respectively favor the assembly of the actin filaments into an amorphous gel and highly organized bundles. A fourth accessory protein, alpha-actinin, is involved in the attachment of the filament bundles to the cell membrane. Myosin, on the other hand, is regulated primarily by chemical modification. An enzyme, myosin kinase, transfers the terminal phosphate group of ATP to myosin, thereby activating it. The activated myosin then interacts with the actin filaments in the cytoplasmic gel to give rise to cell movement. Some types of cell movement do not appear to require myosin, however, and may be mediated instead by the rapid assembly and disassembly of actin filaments or by other kinds of filaments in the cell, such as microtubules.

ample, one mechanism for the activation of myosin ATPase, the enzyme that cleaves the terminal phosphate group of ATP to provide the energy for the contraction of the actin-myosin system, seems to be modification of the enzyme itself. David J. Hartshorne and his colleagues at Carnegie-Mellon University and Adelstein and his co-workers discovered that the myosin ATPase from muscle cells and human platelets could be activated by filamentous actin only after myosin had been phosphorylated by a specific enzyme called a kinase. This regulatory system is common to muscle cells, fibroblasts and myoblasts. Adelstein and his colleagues have pursued the point further and have discovered a second enzyme in nonmuscle cells: a phosphatase that removes the phosphate that was attached to the myosin by the kinase. The dephosphorylated myosin no longer retains its ability to interact with actin, and its actin-activated ATPase is inhibited. This kinase-phosphatase system therefore provides an "on-off" regulation of myosin-dependent motile phenomena.

In addition recent biochemical examination of both muscle and nonmuscle myosins has revealed that a particular type of cell may contain two or more distinct myosins. Therefore if one takes into consideration the fact that the cell contains at least two and perhaps as many as five different types of actin, these molecules may combine in from four to 10 different ways. Moreover, certain actins polymerize in the presence of calcium or magnesium ions, whereas other actins polymerize only in their absence. It therefore becomes clear that the cell has a large repertory of actin-myosin interactions to choose from for its different motile activities.

To sum up, nonmuscle cells possess control mechanisms that regulate monomeric actin, actin in a filamentous bundle form, actin in a meshwork-gel form and the enzymatic activity of myosin. Some motile activities are certainly mediated by an actin-myosin interaction, whereas others may be mediated by an explosive polymerization of actin, a rapid disaggregation of actin filaments or an interconversion of filament bundles into a meshwork. In a dynamic system such as the cytoplasm all these regulatory modes of cell motility and maintenance of cytoplasmic structure coexist, and many of them operate simultaneously.

All these findings demonstrate how multiphasic and multilevel the regulation of cell motility is, and the surface has only been scratched. The future is bound to yield many new perceptions, and a broad interdisciplinary approach should lead to a thorough understanding of this fundamental cellular phenomenon.

Laser Chemistry

Most chemical reactions must be initiated by an input of energy. A laser can often supply energy at the precise frequency needed to drive the reactants down the most favored chemical pathway

by Avigdor M. Ronn

Chemistry is the science that deals with molecular transformations. An essential feature of such transformations is the gain or loss of specific amounts of energy. Chemical changes cannot be understood on a fundamental level or turned to practical purposes unless the specific amounts of energy that accompany chemical reactions are known and controlled. When atoms or molecules in a reaction vessel are excited conventionally by heat or elevated pressures, they follow a variety of reaction pathways, yielding a miscellany of by-products in addition to the desired substance. In order to force the reaction along a particular pathway the chemist often employs a catalyst, which may take any one of a wide variety of forms. It can be metallic, nonmetallic or organometallic; it can be a solid, a liquid or a gas.

The day may not be far distant when the empirical search for optimum combinations of temperature, pressure and catalytic materials to bring about a particular chemical result will seem almost as primitive as alchemy. Since a basic objective of chemistry is to weaken or break specific chemical bonds, as the prelude to a reaction or a rearrangement, energy should ideally be introduced only at the particular energy level, or frequency, necessary to accomplish the objective. Such explicit control of energy changes is becoming feasible with the advent of the laser. Lasers provide extremely intense sources of energy at virtually any desired wavelength in the ultraviolet, visible, infrared or microwave regions of the spectrum. The laser's inherent properties of monochromaticity and high intensity are tailor-made for the control of the specific energy changes that induce or catalyze chemical reactions.

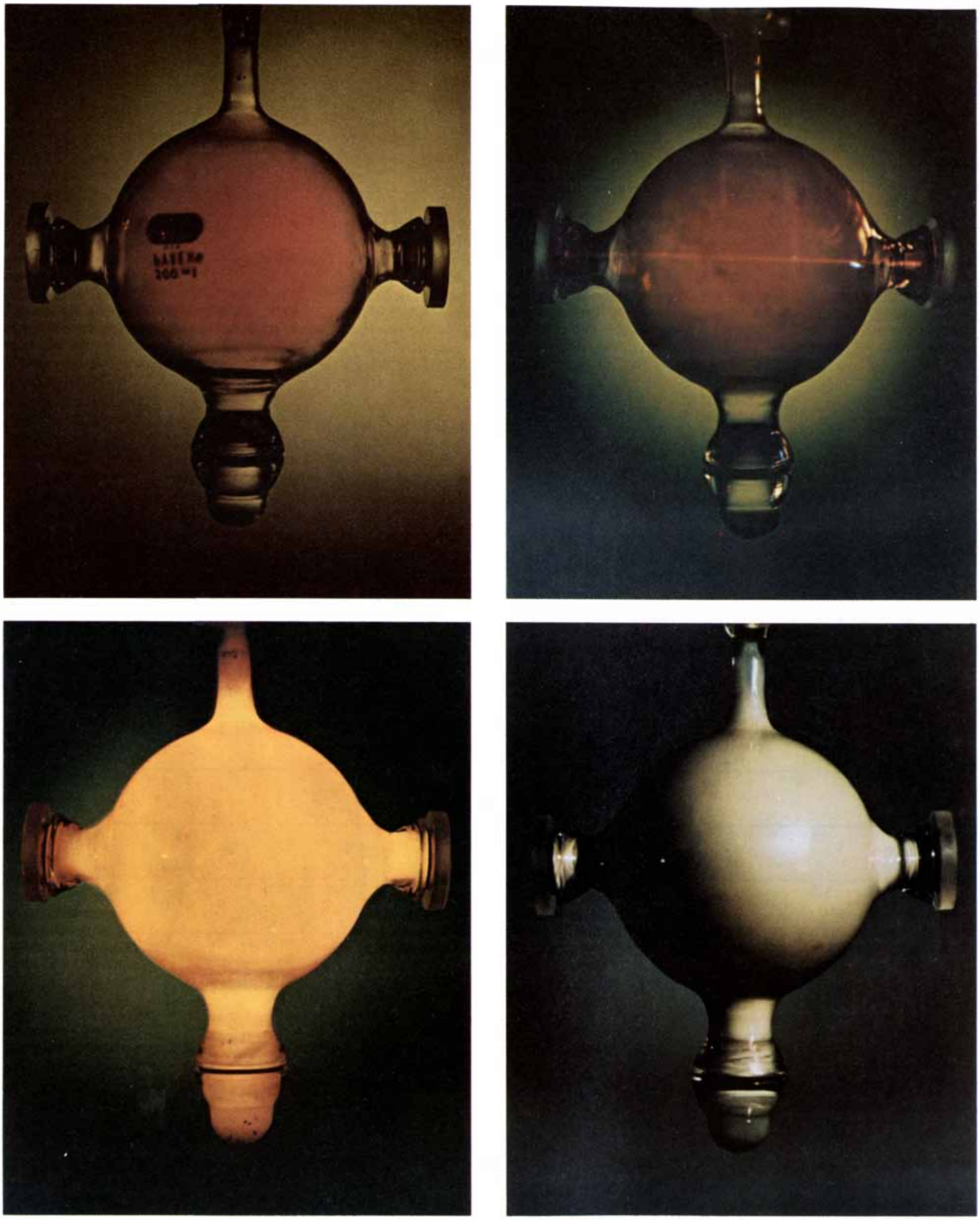
In order to appreciate the impact lasers have already had on chemistry and to speculate on what is sure to come, it will be helpful to look closely at how the energetics of molecules, atoms and electrons involved in chemical reactions are affected by the input of laser energy. A brief examination will reveal how most of the shortcomings of conventional chemical technology can in principle be overcome or bypassed by the appropriate laser system.

Chemical reactions can and do occur in a wide variety of mediums under a great range of conditions. To be sure, certain reactions take place spontaneously as soon as the gases, solutions or solids are brought in contact, requiring no additional energy input. Such reactions are obviously not candidates for laser intervention. Other reactions, and probably the majority of reactions of current and future industrial interest, must be driven by heat, pressure or catalysts or combinations of these agents. The efficiency with which the driving force is coupled to the reacting atoms or molecules determines the yield of wanted products and hence the efficiency of the reaction. The goal is to deliver energy directly into the reaction pathway desired and not to deposit it randomly and wastefully into many possible pathways. Since the laser can supply copious amounts of energy at a given and selectable frequency, or energy level, corresponding to a particular pathway, the laser is emerging as a new precision tool of chemical transformations.

Each of the major regions of the electromagnetic spectrum to which lasers can be tuned, from the ultraviolet to visible light to the infrared and beyond, has unique attributes for depositing energy in a reacting system. It is a fundamental property of atoms and molecules that they absorb or emit energy only in quanta, or discrete packets. When an atom or a molecule absorbs a quantum of energy, it jumps from a lower energy state to a higher one. The jump may be small or large and will affect different properties of the atom or molecule depending on the amount of energy in the quantum.

The absorption of a quantum of visible or ultraviolet radiation alters the state of excitation of the electronic structure of an atom or a molecule. The electronic structure describes the configuration and size of the orbits of the electrons circulating around the central nucleus of an atom or, in the case of a molecule, around the nuclei of two or more atoms. On the absorption of a quantum of visible or ultraviolet energy an electron is raised to a higher orbit. The absorption of a quantum of infrared radiation, on the other hand, will alter energy levels in the system represented by the atoms of the molecule. Such changes involve vibrations: the relative motion of one or more atoms with respect to the common center of mass of the molecule. When a single atom absorbs a quantum of infrared radiation, it simply moves faster, which is described as an increase in the translational energy of the atom. A quantum in the microwave region of the spectrum carries an amount of energy some 10^4 times less than that carried by a quantum of visible radiation, but still the absorption of a microwave quantum can raise the translational energy of either an atom or a molecule. If the absorber is a molecule, the microwave quantum can also cause the atoms of the molecule to rotate faster about their common center of mass.

It is therefore clear that a laser source, depending on the spectral region to which it is tuned, can deposit a precisely measured packet of energy in a molecule, giving rise to a distinct transition from one specific energy state (normally the unexcited ground state) to a higher excited state. In the standard quantum-mechanical picture the various energy states of an atom or a molecule are drawn as a vertical array of horizontal lines [*see illustration on page 117*]. The largest separation between levels, corresponding to the absorption or emission of the largest amount of energy per quantum, is between electronic levels, corresponding to the excitation states of electrons. Many vibrational energy states are contained within one electronic level and many rotational energy states are contained within one vibrational level. Over the past several decades spectroscopy—a discipline of physics and chemistry devoted to the study of atomic and molecular structure—has



DECOMPOSITION OF CHROMYL CHLORIDE, induced by a burst of infrared photons from a carbon dioxide laser, is depicted in these photographs. Chromyl chloride (CrO_2Cl_2), a reddish brown gas at room temperature, is shown in the flask at the top left before the introduction of the laser beam. The gas is at a pressure of .005 atmosphere. The picture at the top right was made the instant the laser beam, consisting of photons with a wavelength of 10 micrometers, entered the flask. The red beam is created by molecules of chromyl chloride that have absorbed 16 to 18 photons in rapid succession and have been raised to an electronically excited state. Some of the excited molecules release a single photon of visible light, incorporating all the energy of the 16 to 18 absorbed photons, and fall back to the initial ground state. When hydrogen is present along with chromyl chloride in the flask, a different reaction results. On absorption of about 28 photons the CrO_2Cl_2 decomposes into CrO_2Cl and an atom of chlorine. Subsequent chain reactions trigger an explosive decomposition (*bottom left*). In picture at bottom right flask is filled with particles of chromium oxide (Cr_2O_3), hydrogen chloride and water vapor.

gathered a wealth of detailed information on the locations of energy states of the known atoms (and their isotopes, or nuclear species) and of thousands of molecules.

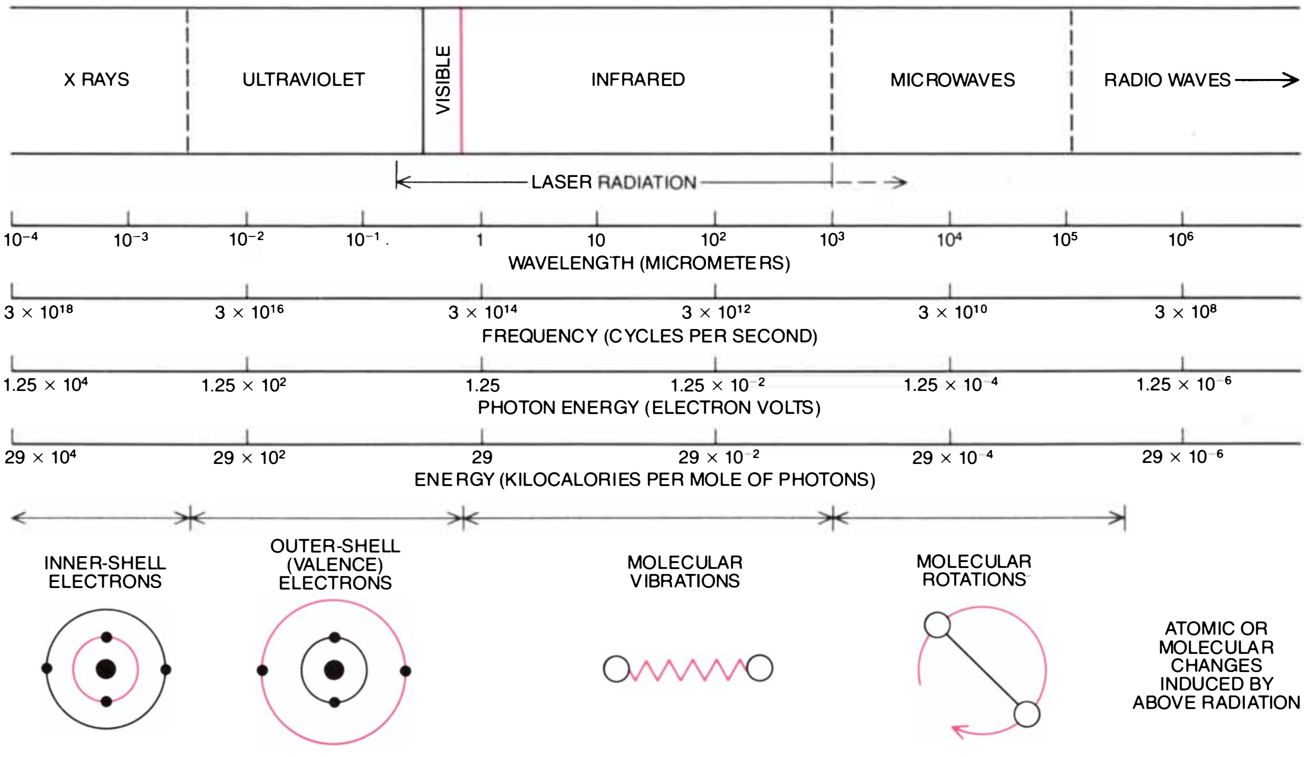
A detailed knowledge of energy levels is mandatory if the laser is to serve effectively as a tool for exciting specific states with the goal of bringing about a particular chemical reaction. Before the advent of the laser the chemist who wanted to introduce energy into a reaction had little choice but to use a conventional heat source, which indiscriminately excites essentially all the vibrational states of a molecule simultaneously. It is true that conventional sources of light and heat—flames, electric arcs, incandescent lamps and the like—can be controlled to a certain extent by passing their radiant output through devices that select narrow bands of wavelengths, but such control can be achieved only at the expense of losing most of the energy generated by the source. The flux of energy that can be obtained from a laser in any narrow frequency range desired is millions or trillions of times more intense than the flux that can be obtained in that range from a conventional source by filtering. As an example, even a modest-sized laboratory laser emits many more photons per unit area of the cross section of its beam than the sun emits per unit area of its surface at the same wavelength.

At first, then, the advent of the laser would seem to be a chemist's dream come true, enabling him in principle to selectively excite reactants—electronically, vibrationally or rotationally, as he wishes—and to drive a reaction toward the desired pathway without the formation of by-products. I should stress "in principle" because in practice laser chemistry often involves additional constraints that must be dealt with before the dream can become a reality.

An atom or a molecule can absorb radiation only in discrete quanta and can do so only at the energy states allowed by quantum mechanics. Sequential absorptions can elevate an atom or a molecule to increasingly high energy levels. Ultimately the absorbed energy will be sufficient to strip the atom of an electron (or electrons), at which point the atom is said to be ionized. Similarly, a molecule will ultimately be dissociated into fragments or even into its constituent atoms. Ionization and dissociation can be regarded quantum-mechanically as the energy state at which there is a continuum of allowed energy levels rather than discrete levels.

The critical question for the laser chemist is what happens to the excess energy that has been selectively deposited by a laser in an atom or a molecule during the time interval required for the reaction to occur. If the energy is dissipated too quickly into nonproductive channels, the reaction either will never proceed at all or will proceed only weakly.

An atom or a molecule can lose its excitation through a variety of mechanisms. It can simply reemit a photon with the same energy as that of the photon it had absorbed, thereby returning to its initial state. Alternatively it can dispose of its excess energy through one or more collisions with neighboring atoms or molecules. The entire excess can be transferred in one quantum or disposed of sequentially in smaller quanta. Alternatively the system can ionize or dissociate if the energy input is high enough. Interesting opportunities arise if the ion-



ATOMIC AND MOLECULAR RESPONSES to electromagnetic radiation depend on the wavelength of the radiation and hence on the amount of energy carried by individual photons. Extremely energetic photons in the X-ray region of the spectrum can alter the state of excitation of the innermost electrons of an atom. Less energetic photons from the ultraviolet and visible regions of the spectrum interact with outer-shell electrons: the electrons that participate in chemical reactions. When such an atom absorbs a photon, or quantum, of ultraviolet or visible radiation, the electron is raised to a higher orbit. Infrared photons characteristically alter the vibrational state of atoms within a molecule. Still less energetic photons from the microwave region alter the rotational state of a molecule. Because lasers can be tuned to provide photons of accurately determined frequency, or energy, they can induce precisely known changes in the state of excitation of atoms and molecules. The laser thus becomes a tool for creating the state of excitation most favorable to a particular chemical reaction. Lasers can now be tuned to yield almost any wavelength from .2 micrometer in the ultraviolet to 100 micrometers in the infrared. Lasers can also generate microwaves, but they are then usually called masers. Chemists find it useful to specify the energy content of laser radiation in terms of kilocalories per mole. In chemistry a mole is the weight in grams equal to the molecular (or atomic) weight of a particular substance and consists of 6.02×10^{23} molecules (or atoms) of that substance. Similarly, a mole of photons is 6.02×10^{23} photons.

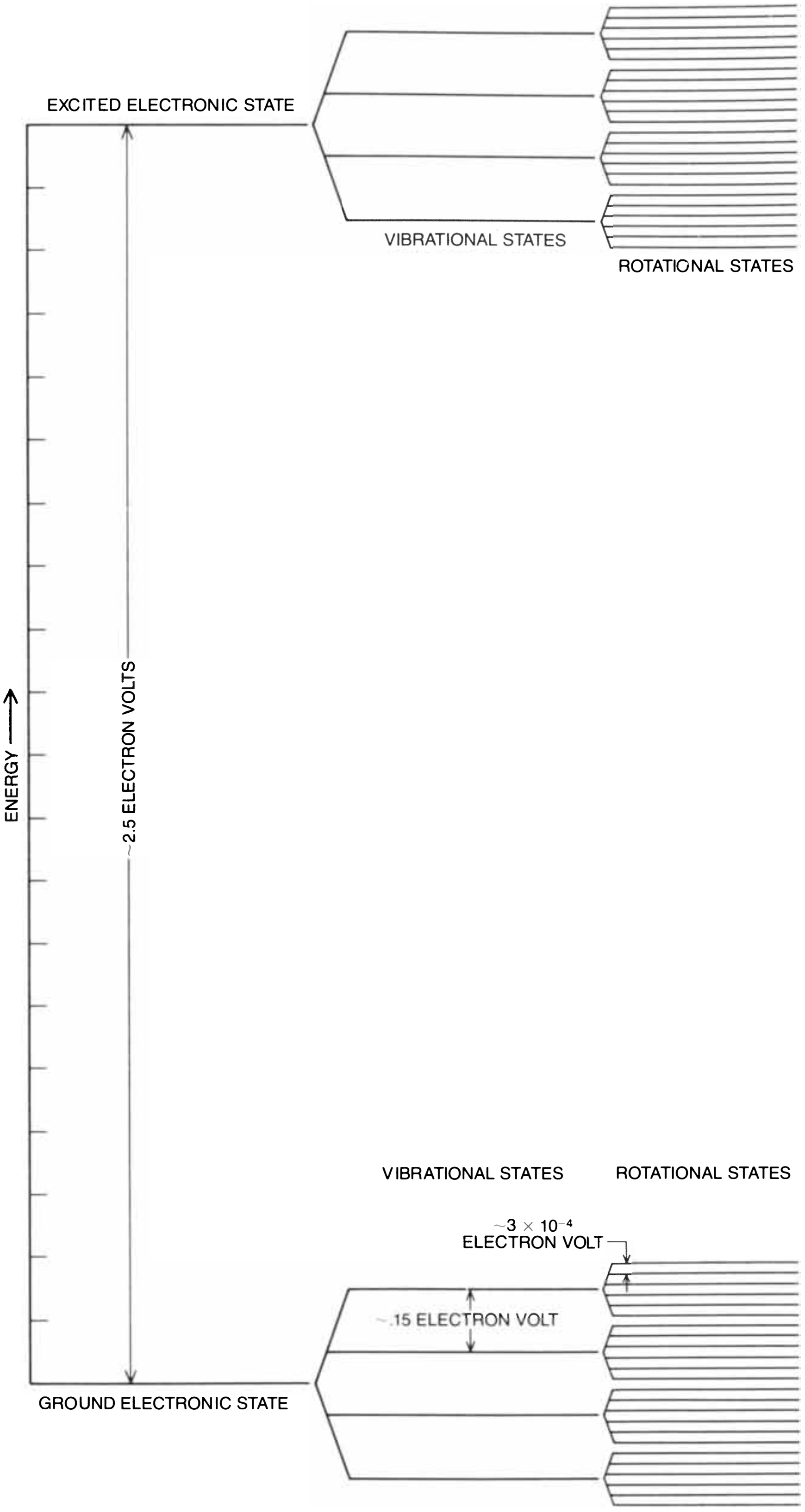
ization or dissociation takes place before any atoms or molecules are able to collide.

Consider what may happen to a molecule that absorbs an infrared quantum capable of raising it from a lower vibrational state to a higher one. If the chemist exposes a molecular species at low pressure in the gas phase to the beam of an infrared laser, he can measure quite accurately the time course of the energy flow into and out of the energy states that are susceptible to excitation. Such experiments are known as vibrational-relaxation studies. In most of the systems analyzed so far (many hundreds) a certain fraction of the irradiated molecules are quickly elevated to a particular energy level, after which the energy is either used for decomposition or dissipated within the same molecule collisionally. Alternatively collisions can dissipate the initial energy among neighboring molecules. In either case a great many vibrational energy states become excited.

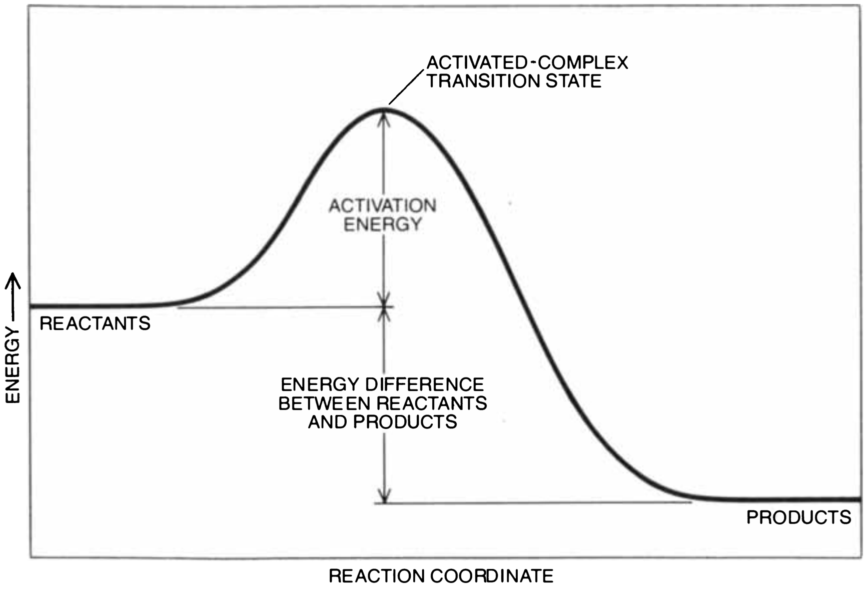
Two kinds of pathway are favored for redistributing the energy. In one of them energy is transferred from the excited vibrational state to one or more other vibrational states of the molecule by collisions. The process continues until all vibrational energy has been transformed into rotational and translational levels of the ground vibrational state. In the second type of redistribution the excess energy of the initially excited level is transferred directly into rotational and translational states. Generally speaking the first kind of energy redistribution, the transfer of vibrational energy to other vibrational states within the molecule, proceeds from 10 to several thousand times more rapidly than the one-step degradation from a vibrational state to a rotational or a translational state.

Thus one can conceive of three distinctly different time regimes in which a laser experiment can be done. In the first type of experiment a reaction must be found that proceeds in a time frame either shorter than or equal to the time required for the transfer of vibrational energy from the initially excited vibrational state to any other vibrational state within the same molecule. In the second type of experiment a reaction must be found that will proceed within the longer time frame compatible with the transfer of energy from the initial vibrational state to the much lower rotational and translational states. The third time frame is of considerable importance because it defines a molecular dissociation (or atomic ionization) that takes place before there is any degradation of energy by molecular (or atomic) collisions. Under such conditions the laser energy is used fully and selectively to accomplish particular reactions.

It is clear that reactions of the first



QUANTUM-MECHANICAL PICTURE specifies that a molecule can occupy only discrete energy states. These can involve the orbits of electrons, vibrations between individual atoms in the molecule and rotations of the entire molecule. For each electronic state there are many possible vibrational states, and for each vibrational state there are many possible rotational states. For example, in the ground state the absorption of a low-energy microwave photon is sufficient to raise a molecule to a higher rotational state. A more energetic infrared photon is needed to raise the molecule to a higher vibrational state. A photon of light or ultraviolet radiation is needed to raise the molecule from the ground state to an excited electronic state. When a molecule drops from a higher state to a lower one, a photon of the appropriate energy is emitted.



INPUT OF ENERGY is normally required to bring about a chemical reaction. This diagram represents the conversion of a single reactant into one or more reaction products. Activation energy must be introduced in order to raise the reactant to an elevated transition state. Once the reactant is in that state there is a high probability that it will descend the far side of the energy barrier and proceed to decompose. There is a chance, however, that it will release its newly acquired energy and return to its original state. The diagram represents an exoergic reaction, so that energy is released when the reaction goes to completion. A similar energy path is followed in bimolecular reactions, where two substances are made to form new reaction products.

type, those that proceed swiftly after the excitation of only one vibrational state, are well suited for laser activation because only one pathway is driven preferentially. In the second type of reaction, in which all vibrational states are excited, one cannot hope to achieve such a high degree of reaction specificity. It is therefore important that the experimenter choose his reaction sequence only after he has established the time frames for the transfer of vibrational energy from the initial state to other vibrational states and ultimately down to rotational and translational states for a given system. One can see that the second type of experiment is basically a "heat" experiment, because the reaction depends less on a specific vibrational state and more on the high excitation of all vibrational states. Such a condition may be achieved equally well by conventional heating. The main difference lies in the time required for achieving total excitation. With a laser source the time can be less than a microsecond, as against a second or much more with a conventional heat source. The short time can be important for many reactions of interest because it may well result in different and unique intermediate states in the reaction pathway.

Consider now the two types of reactions that can be profitably studied in the gas phase: reactions involving only molecules of a single kind and reactions involving molecules of two kinds. In the unimolecular reaction the reactant, after being excited either conventionally or by a laser, is transformed into a new product or is decomposed into two or more products, with or without molecular collisions. In a bimolecular reaction one or both of the reactants may be excited, after which collisions between the two give rise to a new product.

The two types of reaction lend themselves differently to laser induction and laser catalysis. A reaction can be said to be induced if the laser not only triggers the reaction but also supplies all the energy thermodynamically required to drive the reaction to completion. A reaction is said to be catalyzed if the laser provides only the initial "spark," after which the reaction proceeds on its own without further need of energy. Obviously catalyzed reactions, since they make a much smaller demand on the laser as an energy source, tend to be easier to study, understand and evaluate for large production schemes. On the other hand, laser-induced reactions are by and large unique; they cannot be duplicated by conventional means.

In general unimolecular reactions call for a high initial deposit of energy before the reactant is converted into a particular product. Thus unimolecular reactions commonly fall into the category of reactions that can potentially be laser-induced rather than laser-catalyzed. Bimolecular reactions can fall into both categories. The case where only a small deposit of laser energy is needed to accelerate a reaction that occurs naturally at a low rate is comparable to conventional catalysis in chemistry; the laser photons serve as the catalyst.

The energetic pathway followed in a typical chemical reaction, whether the reaction is unimolecular or bimolecular, can be simply visualized. Imagine a road over a hill drawn in profile. The start of the road on the left (representing the reactants) is at a higher elevation, or energy level, than the end of the road at the bottom of the hill on the right (representing the products of the reaction). As always in chemistry, lower energy states are the stablest; therefore stable products will have a lower energy content than the reactants. The hill represents the energy of activation: the amount of energy that must be supplied to the system in order to drive the reaction along a particular reaction coordinate, or pathway.

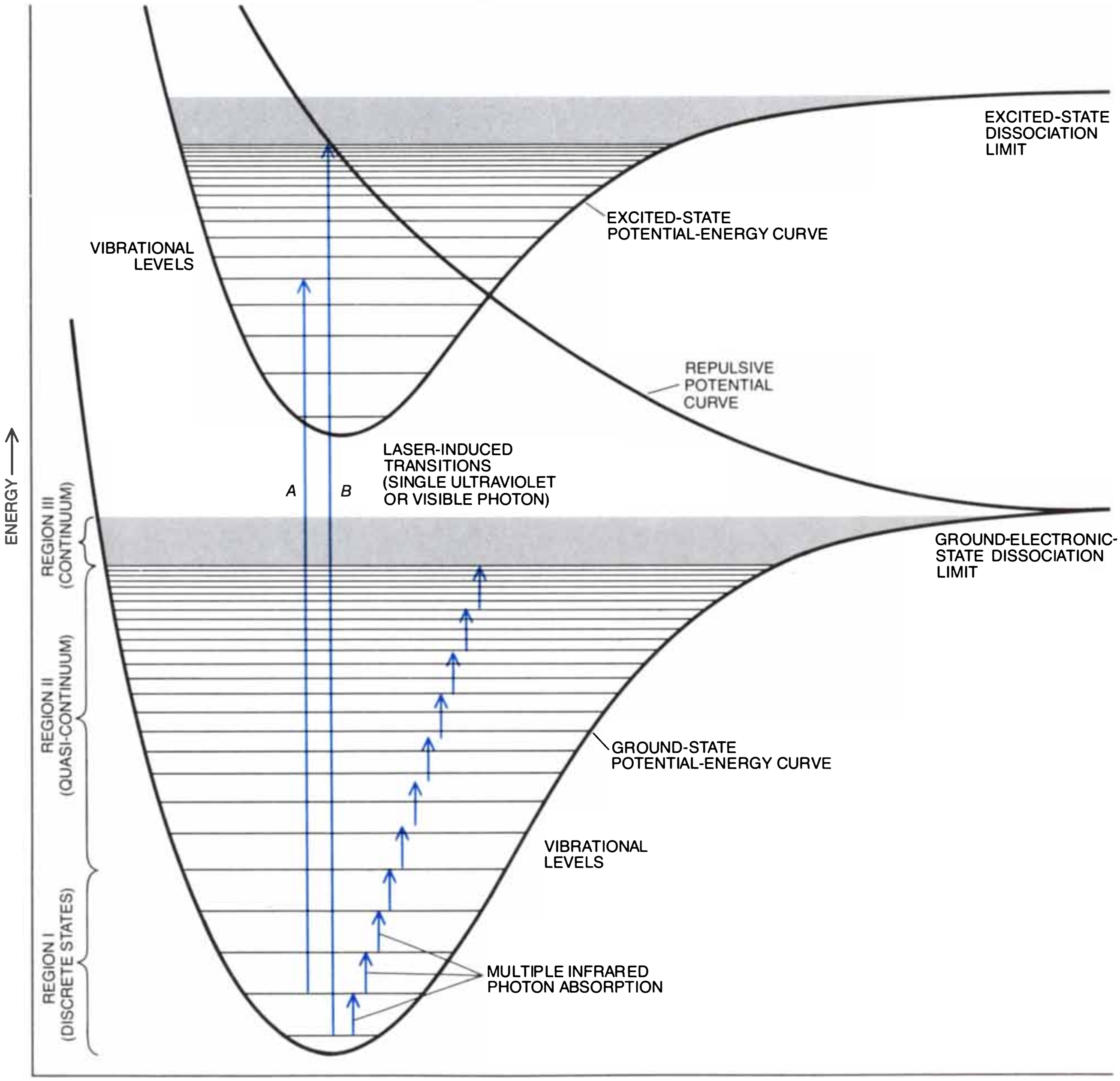
If the reaction is unimolecular, the single reactant must be supplied with enough energy to drive it to the top of the hill, whence it can slide down the other side to the valley of stable products. It should be noted, however, that once the reactant has reached the top of the hill it has a distinct probability of returning to its original state. If it loses its recently acquired energy by some deexcitation process, it will slide back to its unexcited state. If it continues down the far slope, it will be transformed into its products.

If the reaction is bimolecular, both reactants must be raised to the top of the hill. Each reactant can be raised to the top by absorbing energy independently, or one of them can acquire the energy sufficient for both and transfer part of the energy to its partner. At the top of the energy barrier the two reactants form a transition state from which again there is a finite probability of their reverting to their original states and a higher probability of their reacting to form new and stabler products.

The transition state represents a critical configuration. One would like to know for many reactions of interest whether the transition state arrived at by laser excitation is identical with the state achieved by conventional means. The answer to the question is not yet clear; the body of data for laser-driven chemical reactions is not large enough to justify a general statement.

Over the past few years much attention has been given to unimolecular reactions driven by infrared lasers. The point of interest is that the energy carried by a single infrared photon is insufficient to dissociate a molecule from the ground state, or state of lowest energy. The ground-state potential curve can be visualized as being filled with many vibrational levels [*see illustration on opposite page*]. The curve simply represents the fact that the molecule is bound and exists in the ground electronic state.

The absorption of a single infrared photon is sufficient to raise the molecule only one step in the energy ladder. Thus if the molecule is to be dissociated, many infrared photons must be absorbed in sequence. On the other hand, if the molecule absorbs a far more energetic photon from a visible-light or ultraviolet laser, it can be raised in one step from the ground state into an excited electronic state well above the ground state, from which dissociation can proceed more readily. An even smaller amount of energy may be required for dissociation from that higher state if, after the molecule has been excited, it happens to cross over to what is called the repulsive-potential curve. That curve represents the molecule in a repulsive configuration rather than in a bound one, and so the molecule finds its way to dissociation without having to acquire all the energy normally needed for that purpose. Much of the most interesting work in recent years has dealt with laser-induced or laser-catalyzed reactions involving multiphoton infrared-laser excitation. Among the many reasons for studying multiphoton infrared-laser excitation two are compelling. The first is the availability of dependable high-power, tunable lasers in the infrared range. The carbon dioxide laser, the workhorse of the industry, can conservatively deliver a mole of photons (6.02×10^{23} photons) for a total cost of three cents. The second reason is the basic need to understand the role highly



MOLECULAR FRAGMENTATION can be induced by laser excitation in more than one way. Potential-energy curves show the variation in molecular binding energy plotted against interatomic distance for a typical molecule. Each curve has many vibrational levels. On absorption of a light (*A*) or ultraviolet (*B*) photon a molecule is raised from the ground electronic state to a higher excited state. The probability of dissociation increases as a molecule is elevated to vibrational states approaching the dissociation limit. The repulsive potential curve represents an unstable molecular configuration. If after excitation the molecule ends up in such a configuration, dissociation is immediate. Dissociation can also be induced directly from the ground electronic state by the multiple absorption of infrared photons. In the discrete region (I) of the curve infrared photons can be absorbed only if they precisely match the quantum interval between vibrational states. In the quasi-continuum region (II) the requirement for energy matching is less critical. There is no mechanism by which multiple absorption of infrared photons can give rise to dissociation from the excited electronic state, unless the upper state dips into bottom state.

excited vibrations play in chemical reactions, a need that had gone unsatisfied for several decades.

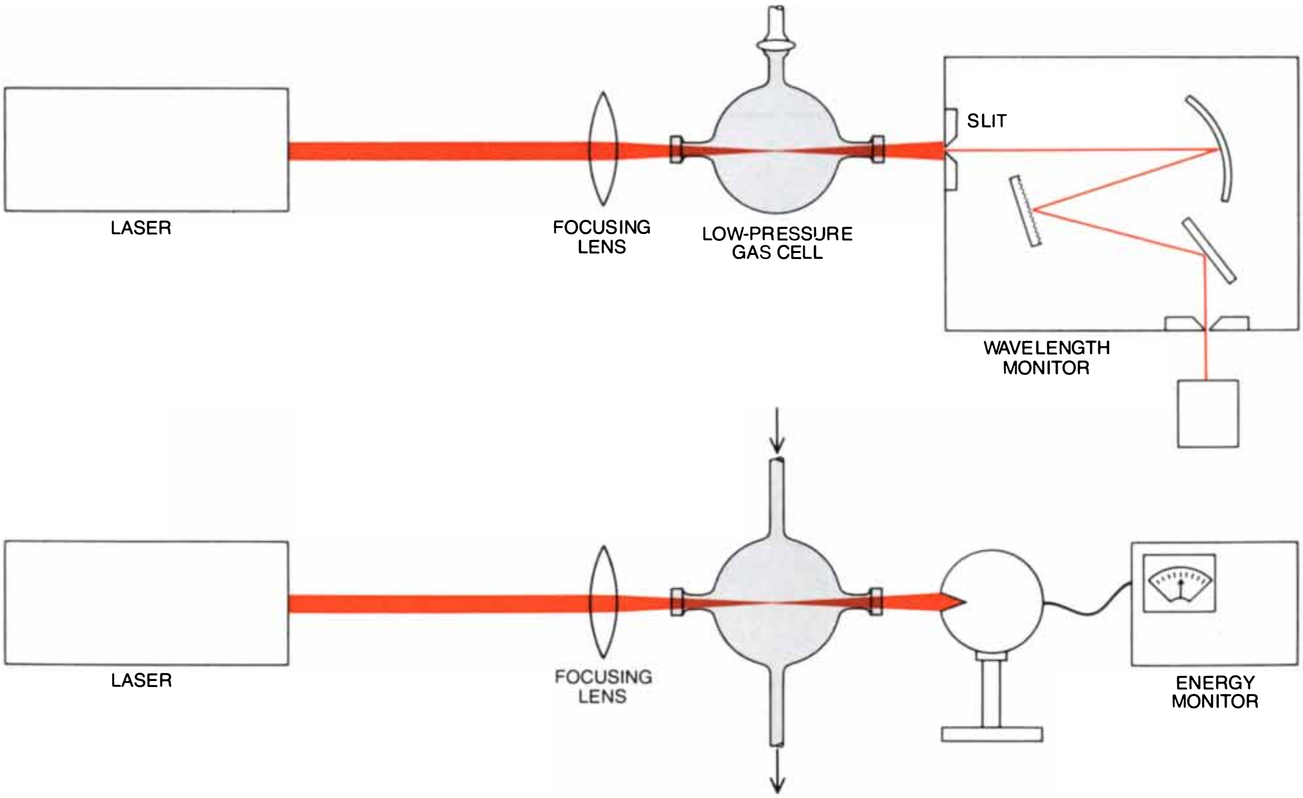
Let me describe two well-studied laser-excited reactions. The first is the dissociation of sulfur hexafluoride (SF_6), which is perhaps the best-studied example of laser-induced unimolecular dissociation involving the multiple absorption of infrared photons. The second is a laser-catalyzed bimolecular reaction in which nitric oxide (NO) reacts with ozone (O_3) to form nitrogen dioxide (NO_2) and oxygen (O_2). The second reaction is catalyzed by a single infrared photon appropriately tuned for absorption either by a molecule of NO or by a molecule of O_3.

The dissociation of SF_6 proceeds as follows. The first three or four infrared photons are absorbed almost simultaneously and at the same frequency, raising the SF_6 molecule from the ground state to three or four successively higher vibrational states. Above the last of these states the number of available energy states becomes extremely high (more than 10^9 states per electron volt), which means that constraints on the wavelength of the entering laser radiation are much reduced. In this region, known as the quasi-continuum, the additional rotational and translational states provide all the "fine tuning" necessary to match the photon frequency with the quantum gap between vibrational levels. In other words, exact resonant conditions between radiation frequency and energy states must be met for only the first three or four photons; after that photons with a much broader distribution of wavelengths can be absorbed, which will eventually promote the molecule to the dissociation limit. The total number of infrared photons absorbed is 35. This number of photons corresponds to the energy required to break a single sulfur-fluorine bond (80 kilocalories per mole).

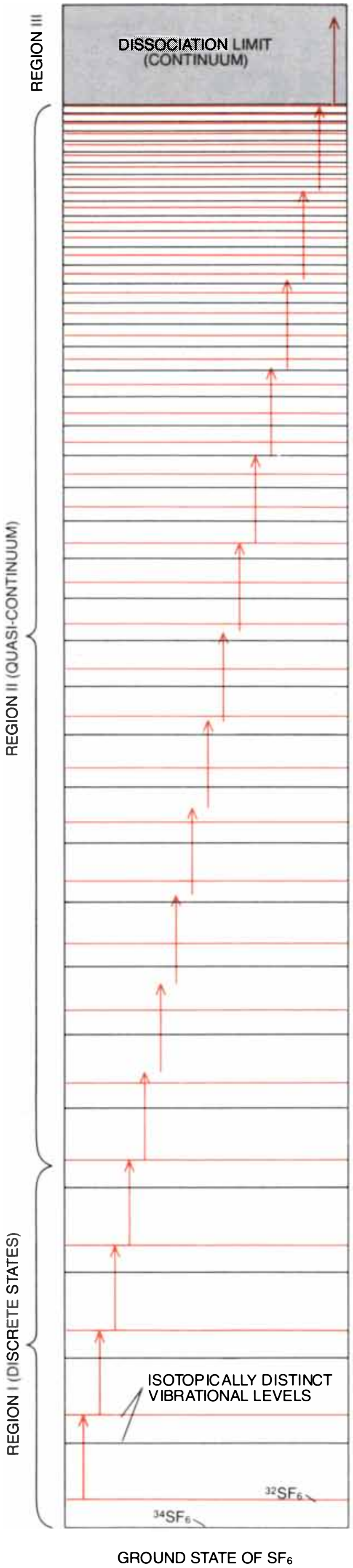
It has been shown that dissociation begins with the breaking of a single sulfur-fluorine bond, leading to the formation of the radical SF_5 and a fluorine atom. The dissociation is so fast that there is no time for energy to be transferred by collisions from one molecule to another. The molecule that absorbs the laser photons is the molecule that breaks apart. No complicating by-products arise, and it would seem that the dissociation of SF_6 represents an ideal case of bond-specific excitation and bond-specific dissociation.

The clean dissociation of SF_6 was one of the first laser-induced reactions exploited to demonstrate that laser radiation can be used to separate isotopes, in this case the two nuclear species of sulfur: sulfur 32 (written ^{32}S) and sulfur 34 (^{34}S). The latter isotope, which has two more neutrons in its nucleus than the former, constitutes only 4 percent of sulfur atoms in nature: 96 percent are the isotope ^{32}S. The aim of the experiment is to selectively excite and decompose molecules of $^{32}SF_6$, leaving the residual gas enriched in molecules of $^{34}SF_6$ (or vice versa). A successful separation was first reported by R. V. Ambartzumian and V. S. Letokhov of the Institute of Spectroscopy in Moscow. They irradiated sulfur hexafluoride with a pulsed carbon dioxide laser tuned to coincide with the vibrational absorption of $^{32}SF_6$. Virtually all the $^{32}SF_6$ was decomposed, leaving the residual gas enriched almost 3,000-fold in $^{34}SF_6$. The effective laser frequency was 10.61 micrometers. When the frequency was shifted slightly to 10.82 micrometers, the $^{34}SF_6$ molecules were selectively decomposed. This method of separating isotopes by laser is being intensively studied for the separation of fissionable uranium 235 from nonfissionable uranium 238 [see "Laser Separation of Isotopes," by Richard N. Zare; SCIENTIFIC AMERICAN, February, 1977].

The decomposition of SF_6 raises an important question for laser chemistry: Does the decomposition truly represent bond-specific excitation followed by bond-specific dissociation? Considering the very short time between the ab-



LABORATORY APPARATUS for studies of laser chemistry takes two general forms. In one arrangement (*top*) the starting material is sealed in a flask or cell and exposed to a laser beam. In the second arrangement (*bottom*) the reactants are irradiated as they flow through the reaction vessel continuously. With the second system the efficiency of the reaction can be followed by sampling the exit gases. Two devices for monitoring the laser radiation are illustrated. In the device at the top the wavelength is determined with the aid of a diffraction grating. In the device at the bottom the energy in the energy beam is simply collected and measured without determining the wavelength.



ISOTOPE SEPARATION can be achieved in some cases by preferentially dissociating molecules incorporating a particular isotope of an element without disturbing similar molecules incorporating a different isotope. In this way sulfur 34 (^{34}S), the rarer isotope in nature, can be separated from sulfur 32 (^{32}S). A normal mixture of the two isotopes exists in sulfur hexafluoride ($^{34}SF_6$ and $^{32}SF_6$). An infrared laser tuned precisely to 10.61 micrometers will excite the vibrational energy states of $^{32}SF_6$ but not those of $^{34}SF_6$. Vertical arrows represent successive absorption of photons. When $^{32}SF_6$ reaches the continuum region, it dissociates into $^{32}SF_5$ and fluorine.

sorption of laser photons and dissociation of the molecule, the answer would seem to be yes.

Uncertainty arises, however, because the molecule contains six chemically equivalent fluorine atoms. How can one establish that the bond that breaks is precisely the one that has absorbed all, or even most, of the laser photons? A decisive experiment, but one that cannot be conveniently attempted, would be to replace one of the fluorine atoms with an atom of another isotope of fluorine. It happens, however, that there is only one stable isotope of fluorine (^{19}F) and that the longest-lived radioactive isotope (^{18}F) has a half-life of less than two hours.

An alternative experiment is able to resolve the question. One of the atoms of fluorine can be replaced with chlorine to obtain the compound SF_5Cl. The experiment is then repeated with the laser tuned to excite the same sulfur-fluorine vibration that is used for SF_6. The unexpected result is that the molecule now decomposes into SF_5 and Cl instead of releasing a fluorine atom. The apparent explanation is that one of the sulfur-fluorine bonds absorbs the laser energy but that the vibrations are quickly and collisionlessly shared by all the atoms linked to the central sulfur. When chlorine is one of the atoms, its bond with sulfur, which is known to be weaker than the others, is the bond that breaks. When all six atoms are fluorine, the bond that breaks is evidently a matter of chance; all the bonds are equally strong. The conclusion is that the dissociation of SF_6 is not truly bond-specific. A similar conclusion has been reached in many other multiple-photon infrared-absorption experiments with other molecular species. The implication of such studies for laser chemistry is that if a transient intermediate state is prepared by multiple-photon absorption, the bond most likely to rupture will usually be the weakest one.

This statement is true only for present-day studies with relatively long laser pulses (.01 microsecond to three microseconds). It is theoretically predicted that pulses about 10,000 times shorter (one picosecond to 100 picoseconds) will indeed provide bond-selective excitation and rupture. The argument is that a certain minimum time, estimated to be about 100 picoseconds, is needed for energy to be shared within an excited molecule. Thus laser pulses shorter than 100 picoseconds should dissociate a single bond of the molecule before there is time for the absorbed energy to be shared.

Let us now look more closely at the bimolecular reaction catalyzed by absorption of a single laser photon: the reaction of nitric oxide and ozone. This reaction between a diatomic molecule and a triatomic one has been studied by the selective infrared excitation of a selected vibration in either molecule. The ozone was selectively excited by a carbon dioxide laser at a wavelength of 9.48 micrometers. The nitric oxide was excited in a separate study by a carbon monoxide laser at a wavelength of 5.2 micrometers. The reaction requires only a small energy of activation: about five kilocalories per mole of the reactants.

Because the reaction proceeds at an appreciable rate even at room temperature it was expected that the deposition of about 2.3 kilocalories of radiant energy in a specific excited vibration of the ozone by a carbon dioxide laser would greatly enhance the reaction rate. Still greater enhancement was expected from the deposition of about four kilocalories in the vibrational mode of nitric oxide by a carbon monoxide laser. In each experiment the input energy corresponded to about one photon per molecule. In both experiments the reaction rate was enhanced by a factor of 20 over the normal rate at room temperature.

Such effective and selective enhancement of a reaction would again appear to be a bond-specific excitation deposited in either of the two reactants by the laser and transferred by way of a reaction intermediate to the products. Even in this instance, however, the actual mechanism turned out to be more complex than had been expected. Energy-transfer studies have indicated that collisions play a significant part in this infrared-catalyzed reaction. Within the ozone molecule more than one vibrational state can be collisionally activated following the specific laser excitation of one vibration. This activation occurs within the time course of the reaction. Therefore at least two and possibly more reaction coordinates corresponding to the number of excited vibrations are responsible for the observed catalytic enhancement. Moreover, it is difficult to be sure whether the excitation is strictly vibrational or whether translational excitation, the result of collisions, also plays a significant part. The latest studies indicate that vibrations and translations play an equal part in the rate enhancement.

Although such findings tend to take

NO RUM REFLECTS PUERTO RICO LIKE RONRICO.

Puerto Rico is the Rum Island, the world's foremost rum-producing region. And Ronrico is *the* rum—authentic Puerto Rican rum since 1860. Ronrico's smooth, light taste has been the pride of six generations of Puerto Rican rum masters. One sip will tell you why.

RONRICO: AUTHENTIC RUM OF PUERTO RICO.

General Wine & Spirits Co., N.Y.C.

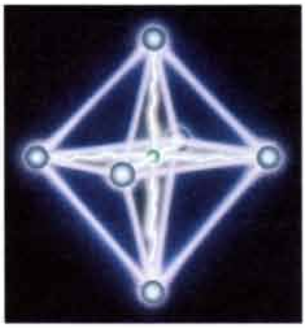
Man has been plating chromium for over a century . . . even though puzzled by the electrochemistry involved.

For example, the normal process uses a bath containing Cr^{+6} ions. During plating, the Cr^{+6} ions reduce to Cr^{+3} and then to metallic Cr. Now you might wonder: If the process goes through the Cr^{+3} stage anyway, why not start with those ions? Well, strangely enough, if you try, the process won't work unless considerably modified.

Our scientists have long been intrigued by this enigma here at the General Motors Research Laboratories. And they now believe they've not only explained the Cr^{+3} mystery, but developed a correct theory of the entire chromium plating process as well.

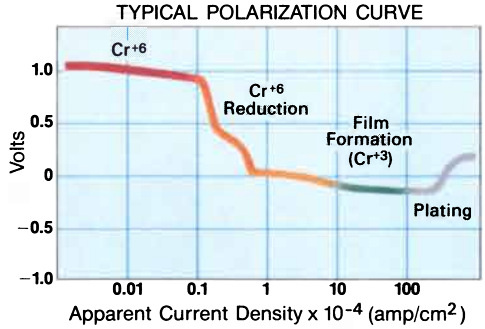
By analyzing polarization curves obtained from carefully designed experiments, they concluded that:

Cr^{+3} (GREEN BALL) BOUND BY WATER MOLECULES AS A STABLE COMPLEX ION



- Starting with Cr^{+3} fails because it immediately forms a stable complex with water molecules (see drawing) from which Cr cannot be deposited.
- Starting with Cr^{+6} succeeds because during reduction a chemical film forms around the cathode (the part being plated); and since Cr^{+3} is bound in that film, it does not react with water . . . but, instead, plates out as chromium metal.

TYPICAL POLARIZATION CURVE

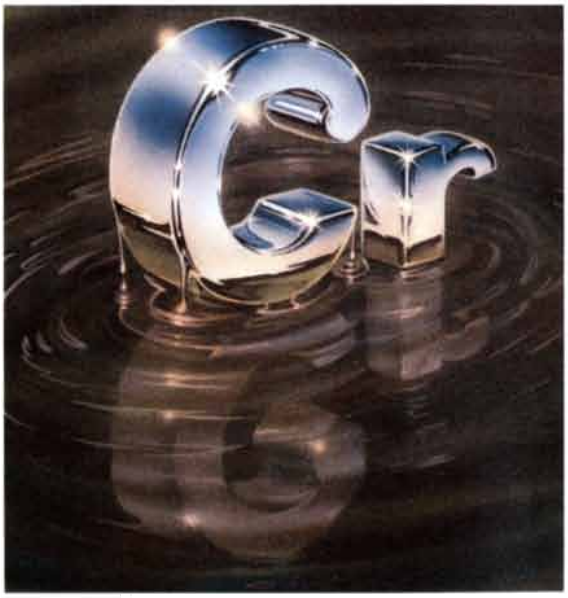


Our researchers have, in fact, determined all 10 steps that take place as Cr^{+6} reduces to bright Cr, pinpointed the step at which catalysis begins, and identified the active catalyst. (It's not the sulfate ion, as commonly held, but the bisulfate ion.)

So have we merely solved a stubborn puzzle? No. More than that, we've gained important new insight to guide us toward a more efficient chromium electroplating process.

We currently have openings for Ph.D.s in engineering or the physical, mathematical, or biomedical sciences. If interested, please send your resume to: GMR Personnel, Dept. 510. An Equal Opportunity Employer.

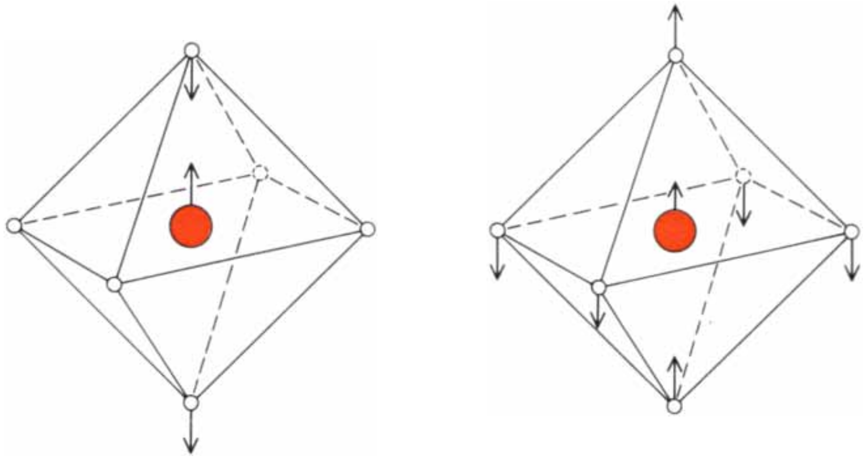
From a murky chromium plating bath, some bright new answers.



GM

General Motors Research Laboratories
Warren, Michigan 48090

some of the luster off the bright future described for the laser in chemical technology, they can also be considered a valuable brake on premature enthusiasm. Early published results are subject to understandable but hasty misinterpretations. Laser chemistry is an explosively growing field; it is an old lesson that the deepest understanding often comes from a reinterpretation of preliminary results.



VIBRATIONS OF SULFUR HEXAFLUORIDE can be excited by infrared radiation. The molecule consists of a sulfur atom (*color*) surrounded by six fluorine atoms (*open circles*), which form the corners of an octahedron. For clarity sulfur-fluorine bonds are omitted. In one fundamental vibrational mode (*left*) the sulfur atom moves toward and away from two fluorine atoms that are 180 degrees apart. In a second mode (*right*) the sulfur atom and two fluorine atoms on a common axis move together as other fluorine atoms move in the opposite direction.

The foregoing examples have stressed some of the problems inherent in analyzing reactions that can proceed on a microsecond time scale and that involve complex bond vibrations, molecular collisions and energy sharing within a molecule that dissociates with or without molecular encounters. Nevertheless, the laser synthesis of special chemicals, albeit in small volume, is already a reality. Products have been prepared that to the level of the detection methods appear to be completely free of by-products. Other compounds, which are hard to isolate by conventional means, have been successfully separated by laser.

The principal obstacle to the large-scale introduction of the laser into the chemical industry is economic feasibility. Laser photons are substantially costlier today than photons that enter a reaction vessel from a jacket of steam or from the direct application of a flame of a burning fossil fuel. The cost of a mole of thermal photons, where they will do the job, is infinitesimal. The cost of a mole of laser photons is on the order of several cents for photons of infrared wavelengths and several dollars for photons of visible or ultraviolet wavelengths. Commercial carbon dioxide lasers operating at infrared wavelengths with high efficiency yield about 5×10^{19} photons per microsecond pulse (equivalent to one joule of energy); therefore about 10,000 pulses are needed to yield a mole of photons. If the reaction proceeds by laser induction, on a mole-for-mole basis, the energy cost in round numbers will come to several cents for several ounces (or several tens of grams) of product. Clearly the first applications will have to be directed toward specialty products—fine chemicals, drugs, catalysts and rare isotopes—whose cost far exceeds the dollar per pound that is typical of large-volume chemicals.

On the other hand, if the laser can be used as a catalyst, yielding many moles of product for each mole of photons consumed, the economics can be much more attractive. The need is to find reactions that can be driven rather easily above the energy threshold necessary for the formation of an excited state, which then sets off a chain reaction. Such reactions, exploited in the production of many industrial chemicals and plastics, require only a tiny amount of catalyst per mole of product. Perhaps the best-known use of catalysts is in the petrochemical industry, where they help to crack large molecules into smaller ones, to polymerize small molecules into larger ones and to rearrange molecular structures into more desirable configurations.

As an example of a reaction triggered by laser photons on a less than mole-for-mole basis let us look closely at a reaction that is most unusual: the decomposition of chromyl chloride (CrO_2Cl_2), a reddish brown gas, either by itself or in the presence of hydrogen (H_2). The reaction is of interest not only because of its economical use of photons but also because the chromium in the product ends up combined with oxygen in different proportions, depending on the presence or absence of a chemical scavenger, in this case hydrogen.

In the first version of the experiment, with hydrogen absent, chromyl chloride gas at reduced pressure is irradiated by an infrared carbon dioxide laser at a wavelength of 10 micrometers. On absorption of the infrared photons the chromyl chloride strongly emits a broad spectrum of visible light, predominantly orange in hue. The reaction is the first example of the conversion of infrared photons into visible photons by a mechanism of absorption and reemission. Independent studies in two laboratories (my own at Brooklyn College of the City University of New York and that of Richard N. Zare at Columbia University) indicate that the visible emission originates with chromyl chloride molecules that have been driven to a low-lying excited electronic state. Following a more intense excitation the chromyl chloride dissociates into chromium oxide (CrO_2) and chlorine gas (Cl_2). The reaction proceeds in two steps, with the two chlorine atoms being released sequentially.

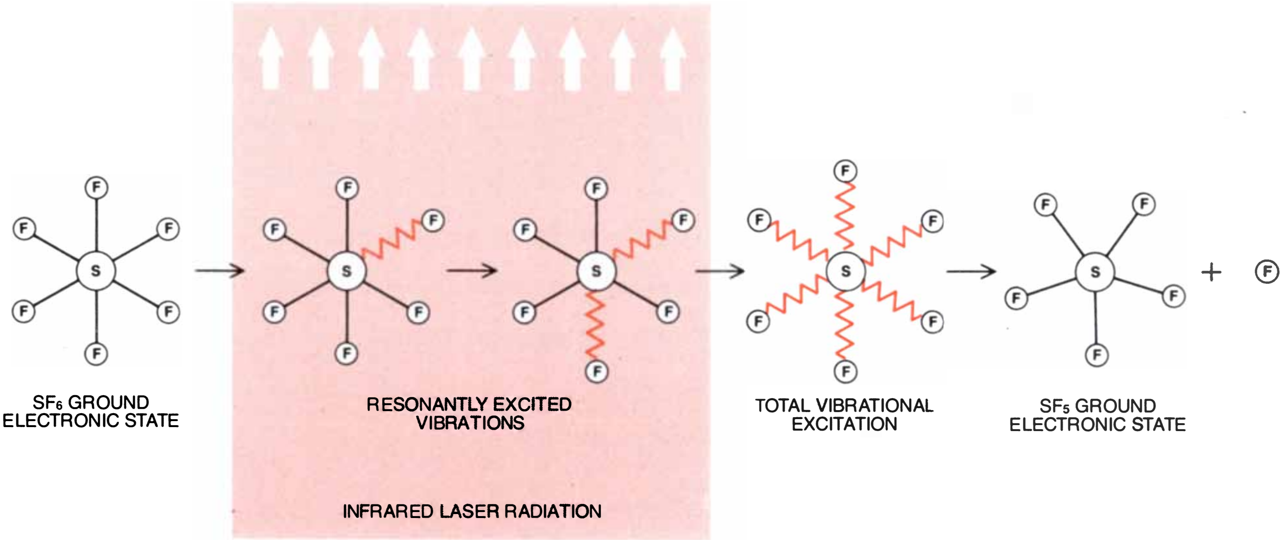
When no hydrogen is present, the emission of visible light coincides precisely with the duration of the laser pulse. The implication is that each absorbing molecule is raised directly, without molecular collisions, from a vibrational energy level in the ground state to a vibrational energy level in the excited electronic state. To reach the higher excited state a molecule of chromyl chloride must absorb 16 to 18 infrared photons in rapid succession. At that point the molecule has been driven to an energy state from which it can either release a single high-energy visible photon and return to its original ground state or absorb an additional 10 photons and use the total energy it has absorbed to dissociate into CrO_2Cl and an atom of chlorine.

The CrO_2Cl fragment, for the duration of its existence, is also subject to the laser excitation. It absorbs some 10 to 16 photons and in turn emits visible radiation or uses the absorbed energy to dissociate into CrO_2 and another chlorine atom. Thus the final products are CrO_2 (which is a brown powder) and chlorine gas (Cl_2).

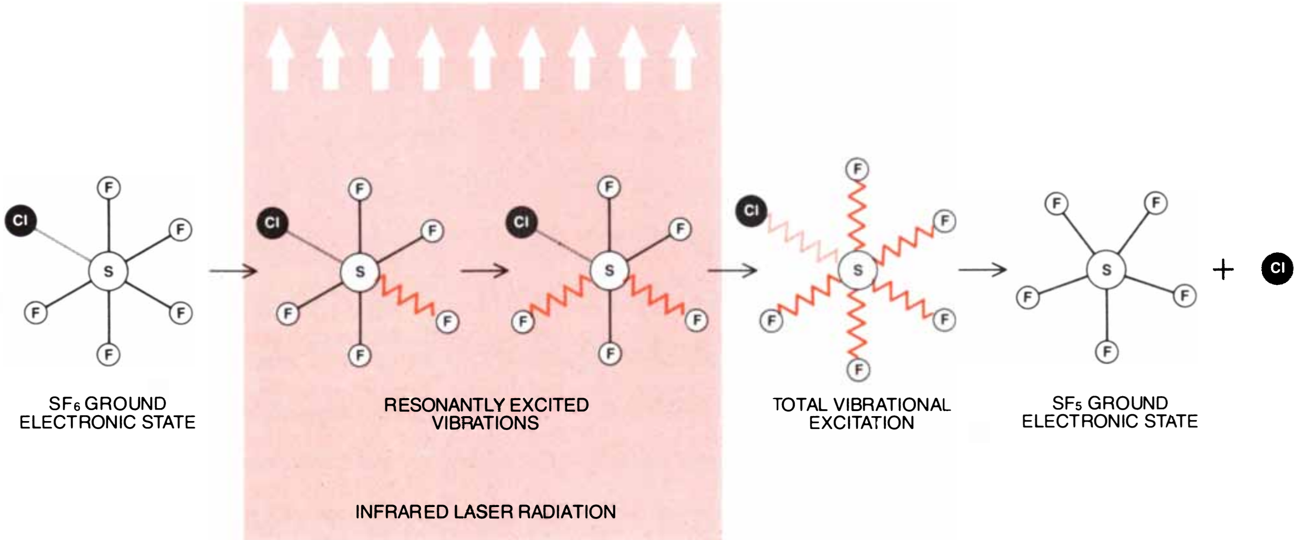
When 1.5 moles of hydrogen is added to the reaction chamber for every mole of chromyl chloride, the reaction follows the same initial first step: a chlorine atom is released from a small fraction of the chromyl chloride molecules. Thereafter the reaction proceeds very differently. The chlorine atom that is dissociated reacts with molecular hydrogen in a well-known chain reaction that splits the H_2 molecule into hydrogen atoms and forms hydrogen chloride (HCl). The extra hydrogen atom reattacks the partly dissociated CrO_2Cl, forming more HCl and water in addition to more complex hydroxylated species. The products are HCl, water and a different form of chromium oxide (Cr_2O_3).

In both reactions chromium was originally in the sixth oxidation state, meaning that to restore it to a neutral state six electrons would have to be added to each atom. In the reaction without hydrogen present chromium is converted into the fourth oxidation state. And in the reaction with hydrogen present chromium ends up in the third oxidation state. In the second reaction only a fraction of a mole of photons is needed to convert a mole of chromyl chloride into a mole of chromium oxide. The dissociation of chromyl chloride in the presence of hydrogen is one of the few explosive branched chain reactions that have been studied with laser techniques.

There is still another mechanism by which chemical reactions can be activated by lasers. The phenomenon was discussed in the early days of lasers, was then neglected and has been returned to only recently. The phenomenon is known as laser-induced optical breakdown or as dielectric breakdown. An extremely intense laser pulse has a very strong alternating-current electric field. When the strength of the field is above a million volts per centimeter, the gas breaks down into a plasma, in which electrons are stripped from the atoms or molecules present. Under these conditions photons from either an infrared or a visible-light laser can interact with a molecule without matching the molecule's specific absorption frequencies and still bring about a fairly specific reaction through the breakdown phenomenon. The reactions that result from such laser-induced electric fields, which interact with the molecule as a whole, are similar to the kinds of chemical reactions that can be induced by very high temperatures (for example in electric



DISSOCIATION OF SULFUR HEXAFLUORIDE, shown schematically, is induced by the sequential absorption of 35 infrared photons from a carbon dioxide laser. The energy carried by the photons is transferred to the sulfur-fluorine bonds (*color*) of the molecule, raising the molecule to successively higher vibrational states. Finally one of the six sulfur-fluorine bonds breaks, releasing a fluorine atom. The remaining molecular fragment, sulfur pentafluoride (SF_5), carries with it any excess energy remaining after breaking of the bond.



SUBSTITUTION OF CHLORINE ATOM for one of the fluorine atoms in sulfur hexafluoride yields SF_5Cl, a compound that can be used to clarify the mechanism involved in the laser dissociation of the molecule. If the laser is tuned to excite sulfur-fluorine bonds specifically, one would expect one of those bonds to break. Instead the sulfur-chlorine bond breaks, releasing a chlorine atom. Although the photon energy must be absorbed initially by the sulfur-fluorine bonds, the vibrational excitation is rapidly shared by all the bonds in the molecule. Because the sulfur-chlorine bond is weaker than any of the sulfur-fluorine bonds it is the bond in the molecule that breaks.

arcs) and that take place on a time scale of microseconds.

In a variety of studies it has been shown that optical breakdown in an intense carbon-dioxide-laser field involves a chain-reaction mechanism known as avalanche ionization. The mechanism requires the presence of a small initial number of electrons. When they are accelerated by the laser field, they start an avalanche of electrons that culminates in a "spark" of ionized plasma. The original few electrons may be supplied by dust particles, cosmic rays or even the windows of the reaction cell.

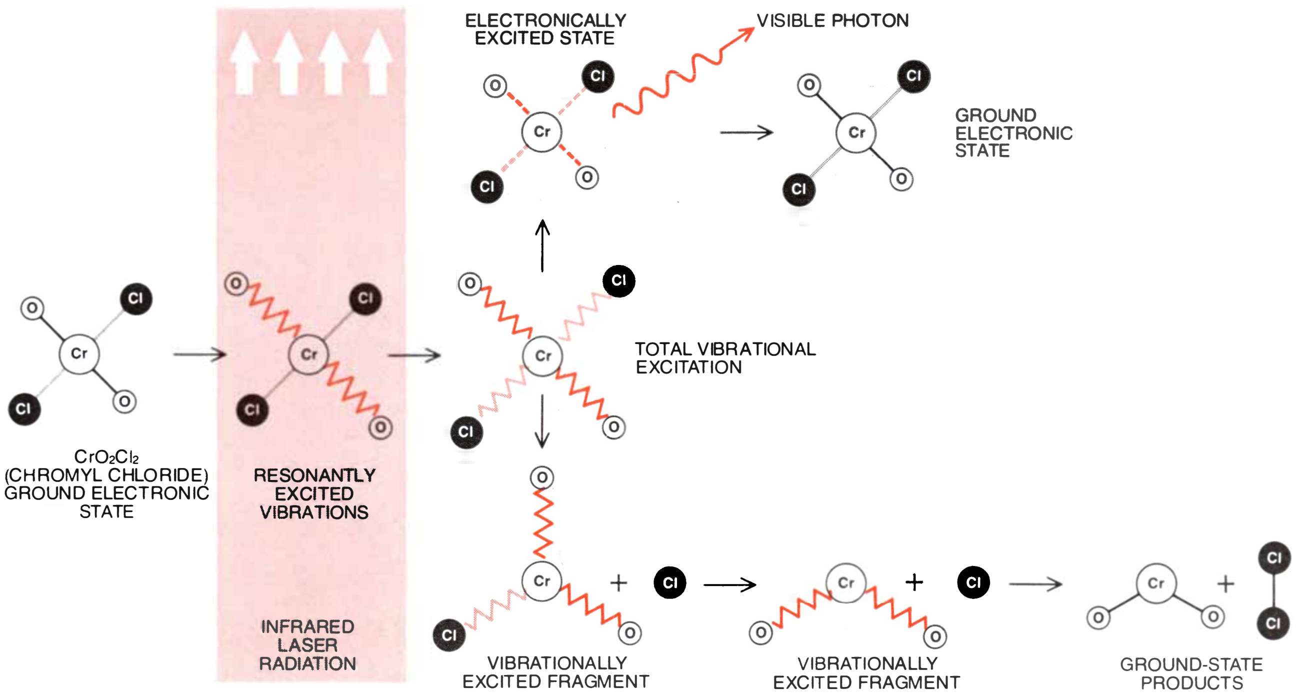
Regardless of the precise mechanism, one can say that laser-induced breakdown can be effected with virtually any molecular species capable of being put into the gaseous state. This means that laser-induced chemical reactions can be studied in many substances lacking energy states capable of interacting with and absorbing photons of available and specific laser wavelengths. A case-by-case study is needed to determine whether or not laser-induced breakdown offers any advantage over the direct application of heat or simple continuous electric discharges. Nevertheless, a number of studies have shown that laser-induced breakdown yields specific products less readily obtainable, or obtainable only in admixtures, by other excitation methods.

Reactions induced by laser dielectric breakdown may offer certain economic advantages over excitation-specific reactions. Reactions of the latter type are most effective when the gaseous reactants are at a fairly low pressure. This places a sharp limitation on the volume of product obtainable from a reaction chamber of a given size. Laser-induced dielectric breakdown, on the other hand, allows the reactant gases to be at elevated pressures. The potential yield of product per pass or per batch is therefore inherently higher than it is for energy-specific reactions. If one adds the possibility that dielectric breakdown may also lead to a chain reaction, the economic picture brightens further.

A number of studies in my laboratory on systems in which reactions have been driven both by specific multiphoton absorption and by laser-induced dielectric breakdown have resulted in the formation of identical products. This does not mean that the reactions proceed by the same mechanism in both cases but merely that the thermodynamic constraints are such that the reaction proceeds along the lowest energetic channels available to the same end result. There is some speculation that the mechanism is actually the same since there exists subtle evidence that multiphoton-absorption processes are controlled by the same constraints that control processes induced by dielectric breakdown. For example, radicals, or molecular fragments, such as C_2, CH, NH_2, CrO_2Cl and CrO_2 are observed in particular reactions whether they are induced by multiphoton absorption or by dielectric breakdown. Moreover, the same statistical thermodynamic rate equations can be applied to both types of experiment as theoretical explanations for the observed products.

Laser-induced optical breakdown can also be applied to the preparation of finely divided solids. For example, when carbonyl sulfide (OCS) is decomposed by this mechanism, finely divided sulfur is precipitated in the reaction chamber [*see illustration on cover of this issue*]. The precipitation of the sulfur can be closely controlled by the manipulation of the gas pressure, the laser conditions and the quenching rate. The principles seem applicable to the preparation of other substances desired in a finely divided state.

Various branches of the chemical industry are closely studying potential applications of lasers to commercial processes. The evident laboratory successes of laser isotope separation, chemical separation, chemical synthesis and removal of trace substances from gas streams of high required purity all warrant serious development work. It is solely for economic reasons that lasers have yet to make a major impact on the chemical industry. Laser technology continues to advance swiftly, however, bringing the cost of devices down steadily as their output efficiency rises. Even now competitive costs can be forecast for drugs and other fine chemicals produced in whole or in part by laser technology. One should not be surprised to learn of major commercial applications of lasers within the next few years.



DISSOCIATION OF CHROMYL CHLORIDE (CrO_2Cl_2) is unusual among laser-induced reactions in that once the molecule has been raised to a transient state of total vibrational excitation by the absorption of 16 to 18 infrared photons it can release all the absorbed energy in a single energetic photon of visible light and then return to the ground electronic state. Alternatively, with the absorption of another 10 photons the excited molecule can decompose in two steps into chromium oxide (CrO_2) and a molecule of chlorine (Cl_2). The laser dissociation of chromyl chloride and of a mixture of chromyl chloride and hydrogen is depicted in color photographs on page 115.

How to buy a typewriter.

-Clip and save-

Choosing a portable typewriter isn't hard if you know what to look for. This brief guide will help you make the best choice.

Test the feel. Check the slope and height of the keyboard. Check the size and shape of the keys. Make sure the controls are uncrowded and easy to reach.

Test the feel of a Smith-Corona® electric typewriter against several other brands. We welcome the comparison.

Try the touch. A responsive touch makes for better, easier typing. Look for a touch that is prompt, easy and dependable.

When you test a Smith-Corona, for instance, note how smartly the carriage returns. Press a button—zip—the carriage is back where it started.

Listen to the sound—the typewriter is trying to tell you something. If it sounds tinny, beware. This may indicate that the construction is too light.

Note the look of the type. Lines and individual letters should be straight. The impression should be crisp, clean and even. The print quality should not vary over the page.

Check the overall design. Good design is part of good value, so choose an attractive modern instrument. The Smith-Corona shown is an example of classic, good design.

Look at the carrying case. Does it have double walls for air-cushioned protection? Does it have sturdy latches and hinges? The Smith-Corona case does.

Check the price. A typewriter that sells for substantially less than others might be substantially less typewriter. If the price difference is minimal, you're probably better off paying a few dollars extra for the typewriter that tests best.

Ask who makes it. Smith-Corona makes every single typewriter that bears its name, which is not true of most other brands. So consider the maker's reputation. A company that has a solid reputation will still be around tomorrow and in the future to give your typewriter necessary service and maintenance.

A note about ribbon systems. Smith-Corona offers a unique cartridge ribbon and cartridge correction system. It lets you change ribbons in seconds without touching the ribbon. It also lets you correct typing errors neatly, quickly and easily. Not all correction systems produce equally good results. Test and compare.

Be sure to try the Smith-Corona carbon film ribbon. We offer a re-usable nylon fabric ribbon, excellent for ordinary typing jobs. This is the only kind of ribbon most portable typewriters offer. But Smith-Corona also offers carbon film ribbon in five colors. It's the kind of ribbon the most expensive office typewriters use, and it's perfect for jobs requiring a crisp, professional look such as term papers or a resume.

More people prefer Smith-Corona electric portables than all other brands combined. After these tests, we think you'll know why.

Patented Correction Cartridge

SCM® SMITH-CORONA
SCM CORPORATION

A Brain-cooling System in Mammals

Carnivorous mammals and some of their mammalian prey endure lethal extremes of heat and exertion because they have a rete, or heat-exchange network, that keeps the brain from getting too hot

by Mary Ann Baker

On a hot day a dog can keep a rabbit on the run until the rabbit dies. How is this possible? The answer is that although running raises the temperature of both animals, the dog's brain has a cooling system and the rabbit's brain does not. If the rabbit cannot find a hiding place where it can cool off, the temperature of its brain soon reaches a lethal level.

Specifically what the dog has that the rabbit does not is a countercurrent heat-exchange plexus at the base of its brain, an anatomical feature found in some mammals but not in all. This plexus, consisting of a network of small blood vessels branching from the carotid arteries, was first recognized more than two millenniums ago by the Greek physician Herophilus. As he was dissecting the head of an animal (probably a sheep) in about 300 B.C. he discovered a prominent network of blood vessels at the base of the brain.

Herophilus described the structure, but it was the anatomist Galen who made it famous nearly half a millennium later. Working in Rome in the middle of the second century A.D., he ascribed functions to the arterial network that appear to be responsible for the term that was soon applied to it: the *rete mirabile*, or wonderful net. Galen was not able to obtain human cadavers for dissection, and so his descriptive anatomy came mainly from the study of domesticated animals. In his view the carotid *rete* was a key anatomical structure: it acted to transform the "vital spirit," which was carried through the arteries, into the "psychic spirit." That spirit was then transported through the body by the nervous system, which Galen believed consisted of hollow tubes.

Galen had a profound influence on later anatomy and medicine. From his day until the time of the Renaissance, anatomists who had human heads to dissect were expected to find the wonderful net because Galen had described it, even though it is not present in man and other primates. The great Renaissance anatomist Vesalius, writing in about 1538, describes the dilemma: "I who so much labored in my love for Galen... never undertook to dissect a human head in public without that of a lamb or ox at hand, so as to supply what I could in no way find in that of man, and to impress it on the spectators, lest I be charged with failure to find that plexus so universally familiar by name."

Countercurrent heat-exchange networks are not limited to the brain of certain mammals; they are found in other parts of the body throughout the animal kingdom. They are responsible for the ability of wading birds to stand for long periods in cold water, for the ability of some sea mammals and fishes to live in polar seas and for the ability of such tropical mammals as sloths and anteaters to conserve body heat at night. Here, however, we shall be concerned only with the first wonderful net to be discovered, the carotid *rete*, and the role it plays in keeping the brain cool.

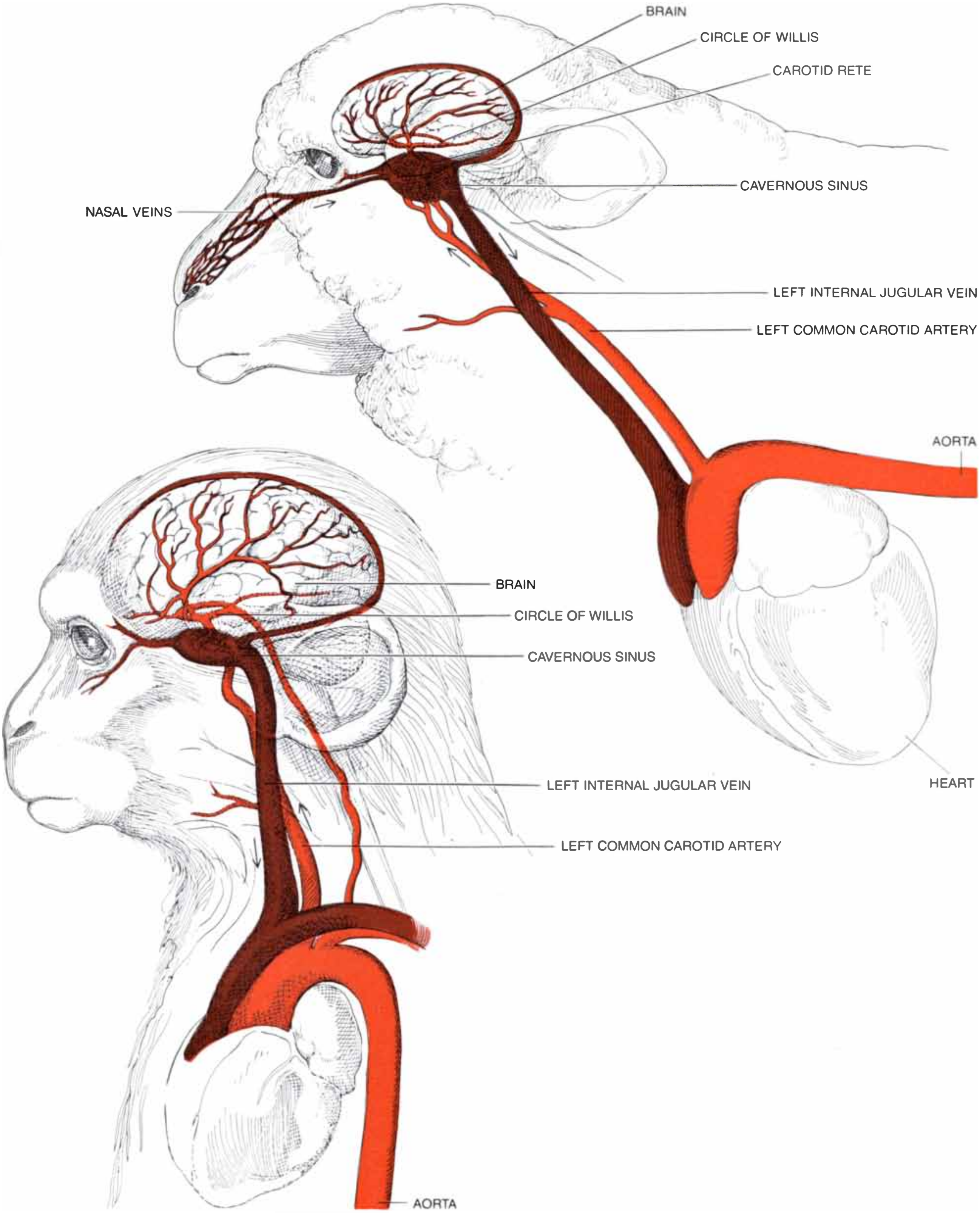
No carotid *rete* is found among the monotremes (the order that includes the platypus), the marsupials, the perissodactyls (the order that includes the horse, the rhinoceros and the tapir), the rodents, the lagomorphs (the order that includes rabbits and hares) and, as we have seen, the primates. For example, in man blood is supplied to the brain through the right and left internal carotid arteries and the right and left vertebral arteries. At the base of the brain the four arteries are connected in the circle of Willis, named for the English anatomist who described it in the 17th century. It is from this circle that all the major arteries supplying blood to the brain arise.

Among the artiodactyls (the order that includes cattle and the other hoofed mammals with even-numbered toes) and many carnivores the pattern of the blood supply to the brain is quite different. Instead of receiving blood from the two pairs of cranial arteries the circle of Willis in these animals is supplied mostly or entirely by the carotid *rete*. The *rete* in turn receives blood either from branches of the left and right internal carotid arteries or from branches of the external carotid arteries or from both. The right and left vertebral arteries are present, but they are often very small. Thus the *rete* is interposed between the two common carotid arteries and the circle of Willis, and almost all the arterial blood that flows into the brain passes through it.

The size and shape of the carotid *rete* varies among the species of mammals that have it, although in some of these species the *rete* has not been examined in detail. For example, among the carnivores all members of the cat family have a well-developed *rete:* a network of vessels 200 to 300 microns in diameter. These are medium-size vessels, much smaller than the arteries that enter and leave the *rete* but much larger than the capillaries. Among dogs, wolves and hyenas and among some seals and sea lions the *rete* is small and consists of no more than a few twisting medium-size arteries.

All mammals, those with a carotid *rete* and those without it, show the same anatomical arrangement at the base of the brain. There, where the arteries that supply the brain with blood enter the cranial cavity, large reservoirs of venous blood are found. Known as the venous sinuses, the reservoirs receive blood from veins both outside and inside the skull. One of the reservoirs is known as the cavernous sinus; in mammals that have no *rete* the internal carotid arteries run through the cavernous sinus on their way to the circle of Willis. In animals that have a *rete* the arterial plexus either lies inside the cavernous sinus or is associated with a similar plexus of venous blood vessels that is connected to the cavernous sinus.

In both arrangements the arteries are bathed by venous blood but the arterial blood and the venous blood do not mix. That is not surprising with respect to the internal carotid arteries, which have normally thick walls. The walls of the *rete* arteries, however, are unusually thin, as I have determined in studies



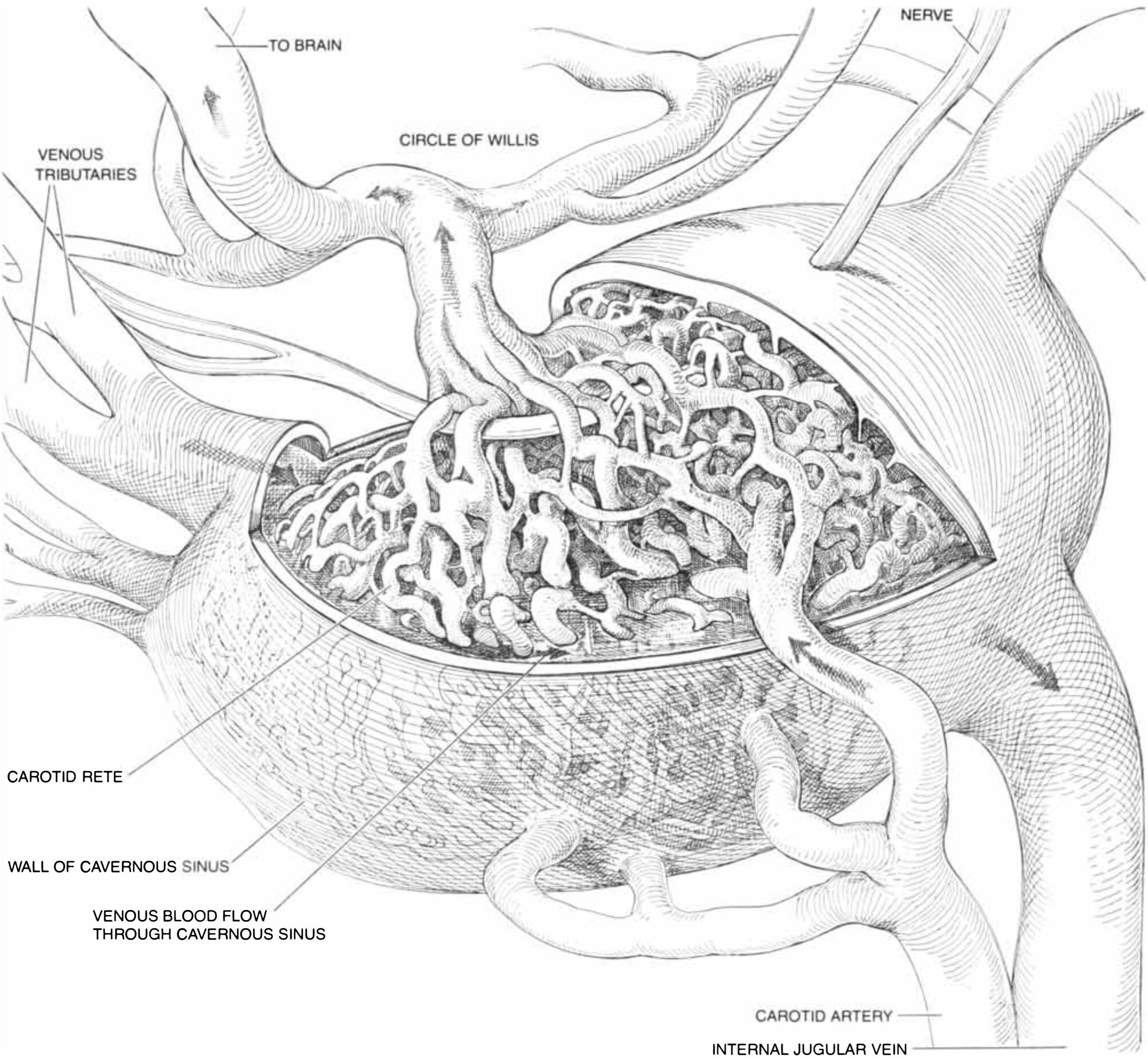
SUPPLY OF BLOOD to the brain of a mammal moves from the heart in one of two basic ways, depending on whether or not the mammal has a carotid *rete*. As illustrated schematically here, the sheep (*top*) has a *rete* but the monkey (*bottom*) does not. In the sheep arterial blood is distributed to the brain from the network of vessels known as the circle of Willis. Some of the blood arrives at the circle via the vertebral arteries but most arrives via the common carotid arteries and their branches, detouring through the network of vessels that forms the carotid *rete*. This network, located within a pool of venous blood in a cavity known as the cavernous sinus, acts as a heat exchanger: the warmer arterial blood loses heat to the cooler venous blood. As a result most of the arterial blood that reaches the sheep's brain is at a lower temperature than it was when it was pumped from the heart. In the monkey the arterial blood travels from the heart via the vertebral and carotid arteries and their branches directly to the circle of Willis. The internal carotid arteries pass through the cavernous sinus with a minimum of heat lost to the pool of venous blood there, and so the blood is supplied to the brain at heart temperature.

conducted in collaboration with Wendelin J. Paule and Sol Bernick at the University of Southern California School of Medicine. Why is the barrier between the arterial blood and the venous blood so insubstantial? The answer to this question helps to explain the major function of the carotid *rete.*

Since the supply of blood to the brain is of major physiological and clinical importance, the anatomy and physiology of the system have been studied in much detail. The metabolic processes of the brain operate at a high rate, and the organ must be supplied with blood at a rate of flow high enough to supply oxygen and nutrients and carry away waste products. If the supply is interrupted, the brain cells can be irreversibly damaged; an interruption may be damaging within a matter of minutes or in no more than a few seconds, depending on the circumstances. It is not surprising that the carotid *rete,* intimately associated with the supply of blood to the brain, has intrigued investigators for hundreds of years.

The question of whether the *rete* does or does not play some role in regulating the rate of blood flow to the brain still remains unanswered. What is now understood, however, is its function in the quite different role of a heat exchanger for cooling the arterial blood destined for the brain. The work leading to an understanding of this role began for me some 10 years ago in the laboratory of James N. Hayward at the University of California at Los Angeles. It was known that in vertebrate animals generally the temperature of the brain was important for the regulation of body temperature. As a result we were interested in finding out how brain temperature was controlled.

Our first studies arose from the observation that rather large changes in brain temperature, as much as one or two degrees Celsius, take place when mammals go from one behavioral state to another. For example, when a relaxed rabbit is suddenly alerted, its brain temperature rises, and when the rabbit relaxes again or goes to sleep, its brain temperature falls. Some investigators had suggested that such temperature



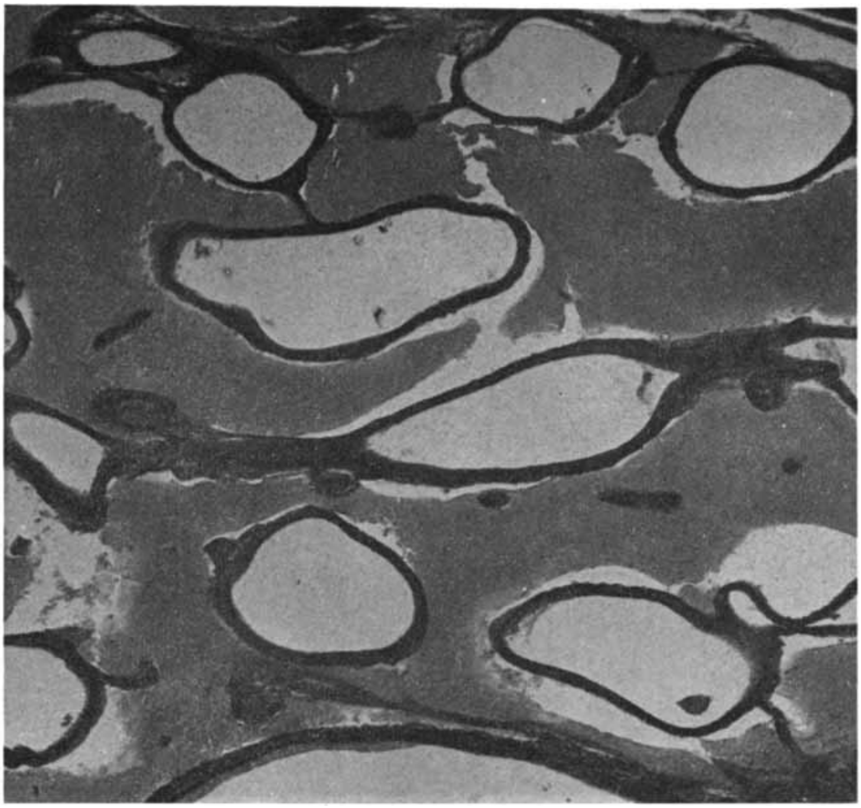
CAROTID RETE, located within a hollow known as the cavernous sinus at the base of a sheep's brain, is exposed in this cutaway drawing. Venous blood, draining from the nose and mouth, enters the sinus from the left, bathes the small arteries of the *rete* and leaves via veins that enter the internal jugular vein at the right. At the same time arterial blood from the heart enters the *rete* via branches of the external carotid artery (*right*) and travels on to the circle of Willis (*top*) after losing some of its heat to the pool of cooler venous blood.

changes were caused by increases or decreases in the metabolic rate of the brain; our studies showed that this was not so.

Selecting monkeys and rabbits as our experimental animals, we implanted very small thermocouples in each subject: one thermocouple near the arteries of the circle of Willis, a second in the animal's brain and a third in a carotid artery near the aorta. The leads to the thermocouples were attached to plugs fixed to the animal's head; long wires led from the plugs to automatic temperature recorders, making it possible for us to monitor changes in temperature as we observed the animal's behavior. We found that all the changes in brain temperature were preceded by changes in the temperature of the arterial blood being supplied to the brain.

What caused the changes in the arterial blood temperature? We found that they were the result of changes in the rate at which heat was being lost from the animals' skin. When the flow of blood through the skin increases, so does the heat loss; when the flow decreases, the heat loss is reduced.

The dilation and constriction of the blood vessels of the skin are controlled by the autonomic nervous system, a system that is affected by emotion and activity. For example, when an animal is excited, the autonomic nerves cause constriction of the blood vessels in the skin; heat loss is reduced and the body-core temperature, including the temperature of the arterial blood, rises. Soon afterward so does the temperature of the brain. Conversely, when an animal relaxes, the autonomic nervous system allows the blood vessels in the skin to dilate, heat loss through the skin increases and the temperature of the body core falls. In rabbits one of the most important heat-exchange areas is the ear; the fur there is thin and the tissues are richly supplied with blood. In monkeys the hand and the foot are important heat-exchange areas; in human beings the hand is. Indeed, this same autonomic reaction occurs in human beings. When we are excited or upset, our hands tend to get cold. When we relax, they warm up.



CAROTID RETE of a cat is seen in section in this micrograph. Varied in size, the thin-walled arteries lie in a lake of cool venous blood, secured in place by strands of connective tissue.

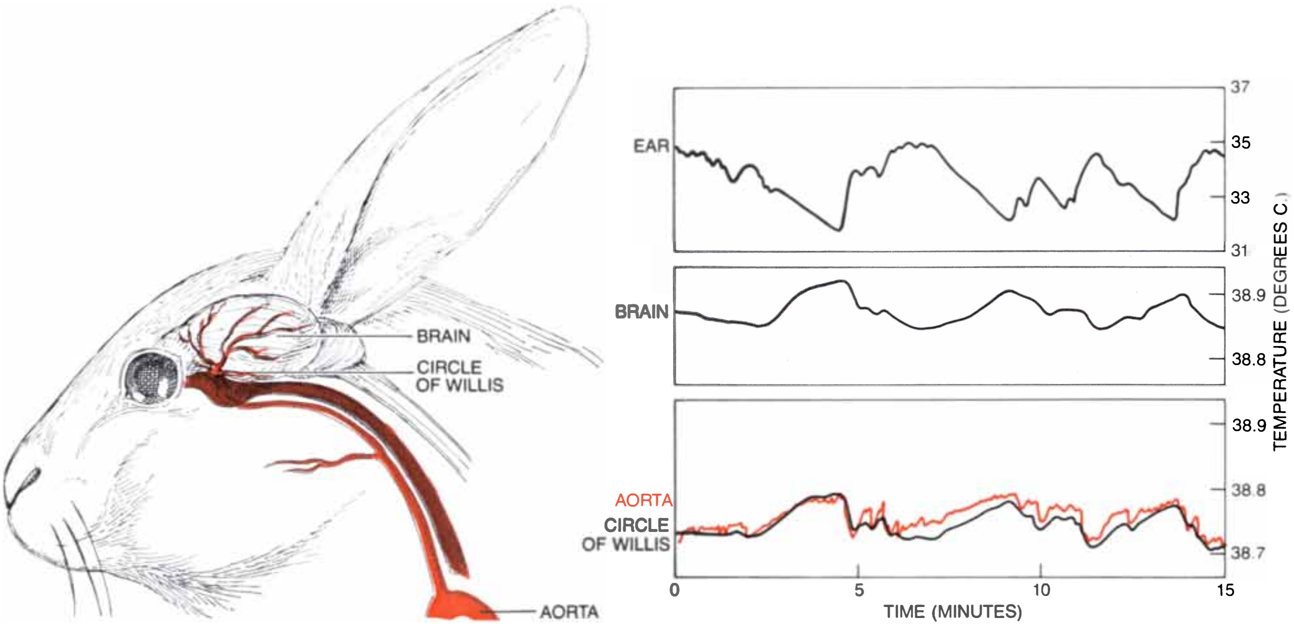
While we were studying the brain and blood temperatures of rabbits and monkeys we found that the temperature of the arterial blood in the carotid artery and of the arterial blood in the circle of Willis were the same. We wondered whether the same pattern would be found in other species. Our next experimental animals were cats and sheep, species that unlike rabbits and monkeys have a carotid *rete*. We implanted three thermocouples in the new subjects. When either animal was in a relaxed state, we found that, as in rabbits and monkeys, the temperature of the brain was higher than the temperature of the arterial blood at the circle of Willis. When the animals were active, the temperatures rose, as they also had in the rabbits and monkeys. We found, however, that during activity the arterial blood at the circle of Willis remained cooler than the blood at the third thermocouple, which was emplaced near the aorta. Somehow in cats and sheep the arterial blood was being cooled as it flowed from the heart to the brain. This was quite different from our findings in rabbits and monkeys.

The only major difference between the two sets of experimental animals as far as the blood supply to the brain was concerned was the absence of a carotid *rete* in monkeys and rabbits and the presence of the *rete* in cats and sheep. This fact suggested that in our new subjects the *rete* was functioning as a countercurrent heat exchanger. To test this hypothesis we implanted an additional thermocouple in the animals' cavernous sinus, where the venous blood surrounds the arteries of the *rete*. We found that the venous blood also underwent temperature changes: it was cooler when the animals were relaxed or sleeping and warmer when they were alert or excited. Regardless of the animals' state, however, the venous blood was cooler than the arterial blood entering the *rete*, and so a countercurrent heat exchange was taking place.

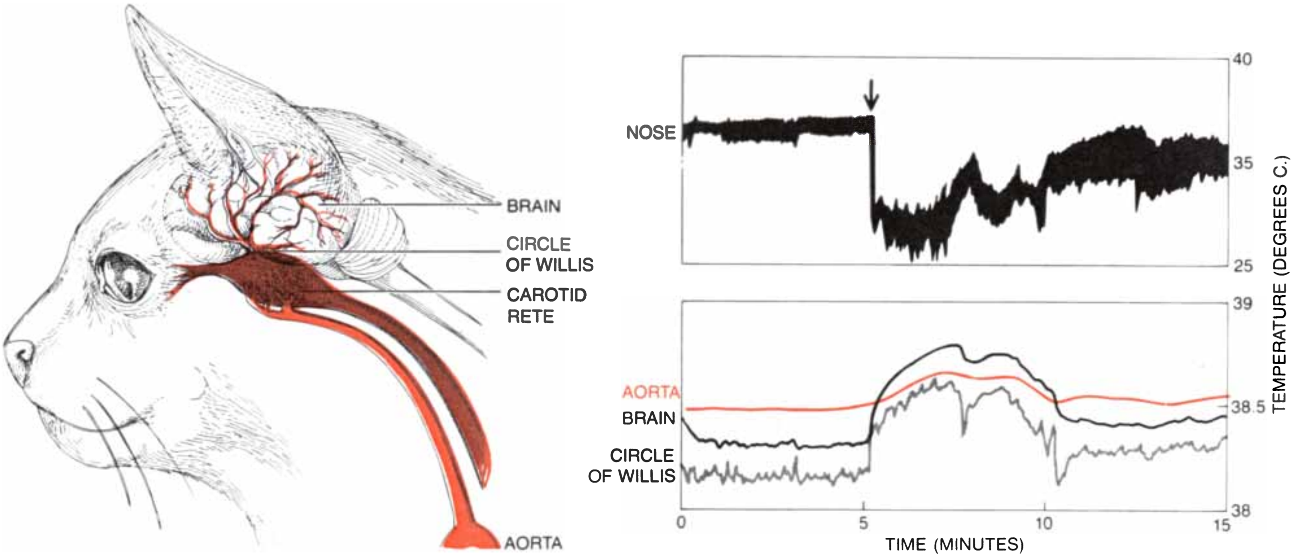
This led us to two further questions. What is the source of the venous blood in the cavernous sinus? Why does the temperature of the venous blood change when the animal's activity changes? As to the first question, venous blood arrives at the cavernous sinus from several sources. Some of it comes from the base of the brain; we would expect this blood to be warmer than arterial blood because the brain has added heat to it. Much of the venous blood, however, comes from areas outside the cranial cavity. For example, anatomists have known for a long time that the venous blood draining from the nose and parts of the mouth can flow into the intracranial sinuses.

James H. Magilton and Curran S. Swift of Iowa State University have studied the venous drainage of the nose, with dogs as their subjects. They found that changes in temperature within a dog's nose will change its brain temperature by changing the temperature of the nasal venous blood that flows into the cranium. In an anatomical study of our own we traced a similar flow in sheep. Working with preserved sheep heads, we injected colored latex into the veins of the nose; we found that the latex filled the cavernous sinus at the base of the brain where the sheep's carotid *rete* is located.

Such studies enabled us to answer the first of our questions: the venous blood that was affecting the temperature of the arterial *rete* came from the animals' nasal passages. As for the question of changes in temperature, these changes reflect the condition of the nasal mucosa: the moist lining of the nasal passages. The blood vessels of the nasal mucosa, like the blood vessels of the skin, are under the control of the autonomic nervous system. When an animal, including the human animal, is relaxed, the blood vessels of the nasal passages are dilated and the venous blood is cooled by the evaporation of moisture from the mucosa. When the animal is excited, the blood vessels of the nasal mucosa are constricted by the autonomic nerves. The action decreases the amount of blood flowing through the nasal passages and also the amount of venous blood flowing from the nose to the cavernous sinus in the cranium. The decrease in the supply of cool venous blood diminishes the heat exchange be-



BLOOD-TEMPERATURE CHANGES in a rabbit proved to be correlated with changes in the heat loss from the ear. The dilation and constriction of the blood vessels in the ear of the rabbit, a major heat-exchange surface, are controlled by the autonomic nervous system. When the blood vessels of the ear are unconstricted and the air temperature is 25 degrees Celsius, the temperature of the ear is generally above 34 degrees. When the autonomic nerves constrict the blood vessels of the ear, the temperature of the ear drops. The drop is followed by a rise in the temperature of the arterial blood flowing from the heart through the aorta to the circle of Willis. The temperature of the brain, which is normally somewhat higher than the temperature of the blood at the circle of Willis, rises and falls similarly.



SIMILAR TEMPERATURE CHANGES in a sleeping cat that was wakened once during a 15-minute period are shown. The major heat-exchange surface, the mucosa of the cat's nose and mouth, is similarly under the control of the autonomic system. While the cat sleeps the heat loss through the nose is high and evaporation cools the venous blood draining into the cavernous sinus. There the warmer arterial blood moving from the aorta to the circle of Willis is cooled by the venous blood as it flows through the carotid *rete*. When the cat wakes (*arrow*), the autonomic nerves constrict the mucosal circulation, the nasal heat loss is reduced, as is the heat exchanged through the *rete*. The temperature of the arterial blood at the circle of Willis rises to almost match the temperature of the arterial blood from the aorta, whereas the temperature of the brain, normally lower than that of the aortic blood, briefly exceeds it. When the cat returns to sleep, the dilation of the mucosal blood vessels enhances heat exexchange in the *rete*, and the other temperatures decline accordingly.

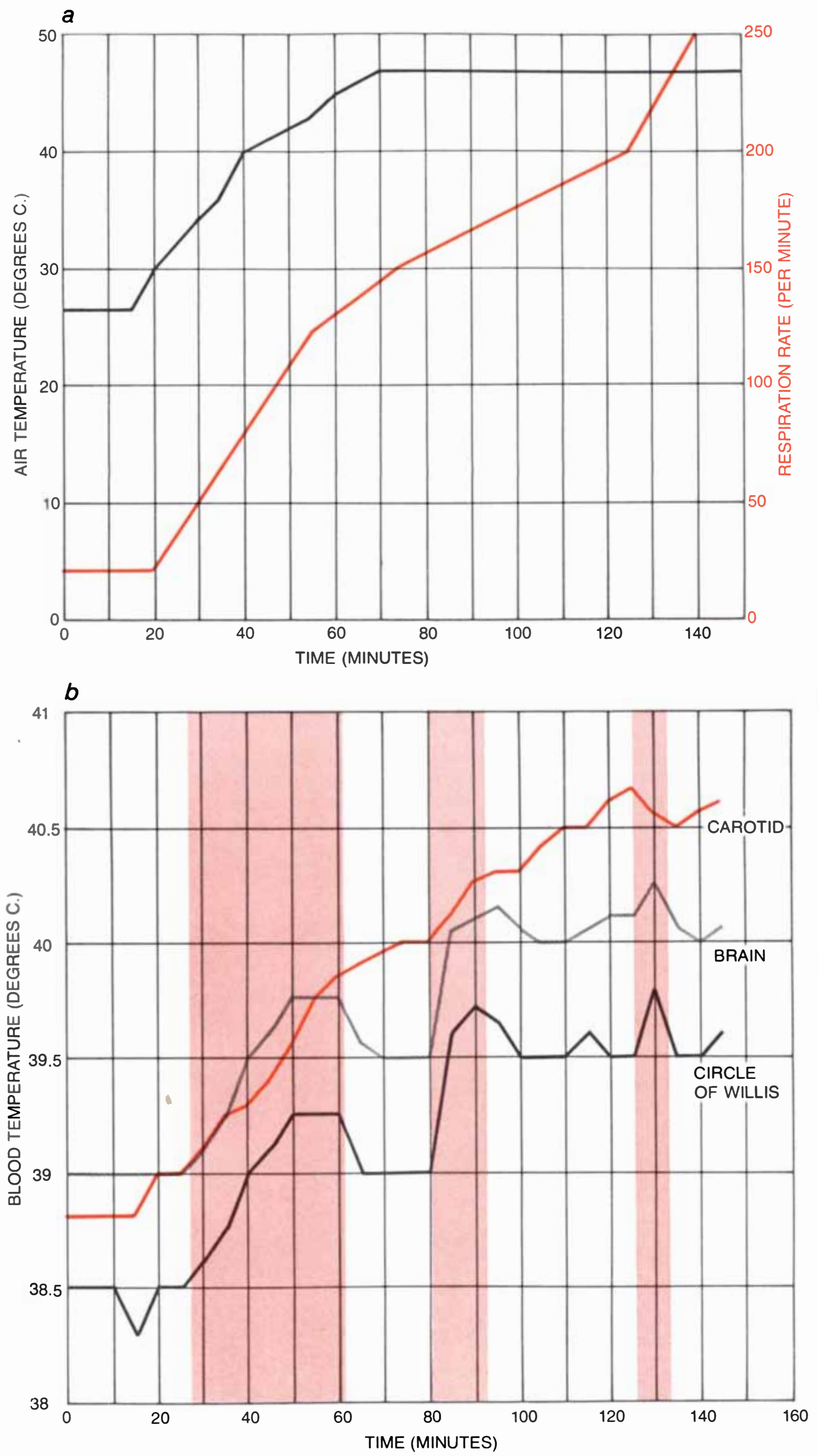
tween the venous blood and the arterial blood passing through the carotid *rete*. This results in a rise in the temperature of the arterial blood reaching the brain until this blood approaches a temperature equal to that of blood in the body core.

We confirmed the validity of this sequence of events by a series of experiments measuring the temperature of the nasal mucosa with thermocouples implanted in the nostrils of both cats and sheep. When the blood flow through the nasal mucosa increased, the temperature of the mucosa rose; when the blood flow decreased, so did the temperature of the mucosa. This exactly paralleled the changes in skin temperature we had observed in association with the constriction and dilation of the skin blood vessels. (It should be mentioned that the ambient air temperature and the animals' respiration rate remained constant during these measurements.)

Exactly how does evaporation cool the nasal mucosa? The mucosa is always moist. With each breath the animal draws relatively dry air across the moist surface, the moisture evaporates and the cooled mucosa cools the blood flowing through it. Therefore one could expect that an increase in the rate of evaporation would further decrease the temperature of the venous blood flowing from the nose and produce a consequent decrease in the temperature of the arterial blood in the carotid *rete*.

That is exactly what happened when we conducted a further experiment. We anesthetized cats and sheep and pumped air into their noses. The arterial blood supplied to the head by the heart remained constant in temperature, but the increased evaporation made the temperature of both the *rete* and the brain drop. When we repeated the experiment with a rabbit as the subject, there was no such rapid cooling. Instead the first drop in temperature took place in the carotid artery; it was followed by a lowering of the brain temperature. The sequence indicates that in the absence of a carotid *rete* the cooled venous blood from the nose must return to the heart and there cool the arterial blood moving from the heart to the brain before the temperature of the brain is affected.

Would an increase in respiratory evaporation have the same effect on animals in normal circumstances that it had on anesthetized animals? To find out we placed cats and sheep in a hot room; the flow of air through their nasal passages would naturally increase when they began to pant in response to overheating. By the time the animals were panting at a rate of 250 to 300 respirations per minute the rate of cooling of the arterial blood passing through the *rete* had also increased. Both the carotid arterial blood flowing to the input side of the *rete*



RESPONSE TO HOT ENVIRONMENT by a *rete*-protected mammal is shown in this pair of graphs. An unrestrained sheep was placed in a test chamber at an air temperature of 26 degrees C. Over a period of an hour the air was gradually heated to 47 degrees C. (117 degrees F.) and kept at that temperature for more than an hour (*a*); the respiration rate of the sheep, only 20 breaths per minute at the start, increased rapidly and continued to rise after the air temperature leveled off. The temperature of the blood in the sheep's carotid artery also rose steadily (*b*), but the animal's panting cooled the venous blood surrounding the *rete* sufficiently to keep the blood temperature in the circle of Willis below that of the blood in the carotid artery. As a result the sheep's brain, warmer than carotid blood to start, remained at or below the carotid temperature for most of the experiment. When the sheep stood up (*colored bands*) instead of lying down, both the brain and the blood at the circle of Willis became temporarily warmer.

and the cerebral arterial blood flowing from the output side showed a rise in temperature, but the rise was less for the blood that had passed through the *rete*.

As a result the difference in temperature between the arterial blood on the input side and that on the output side of the *rete* became greater. When the animals were kept at a normal temperature, the temperature of the input blood had been about .25 degree C. higher than that of the output blood. Now, when the animals were panting in a hot room, the input blood was a full degree warmer. The hot-room observations demonstrated the significance of the carotid *rete* as a heat exchanger. In conditions of heat stress the brain of a panting animal that has a carotid *rete* will remain cooler than the rest of its body.

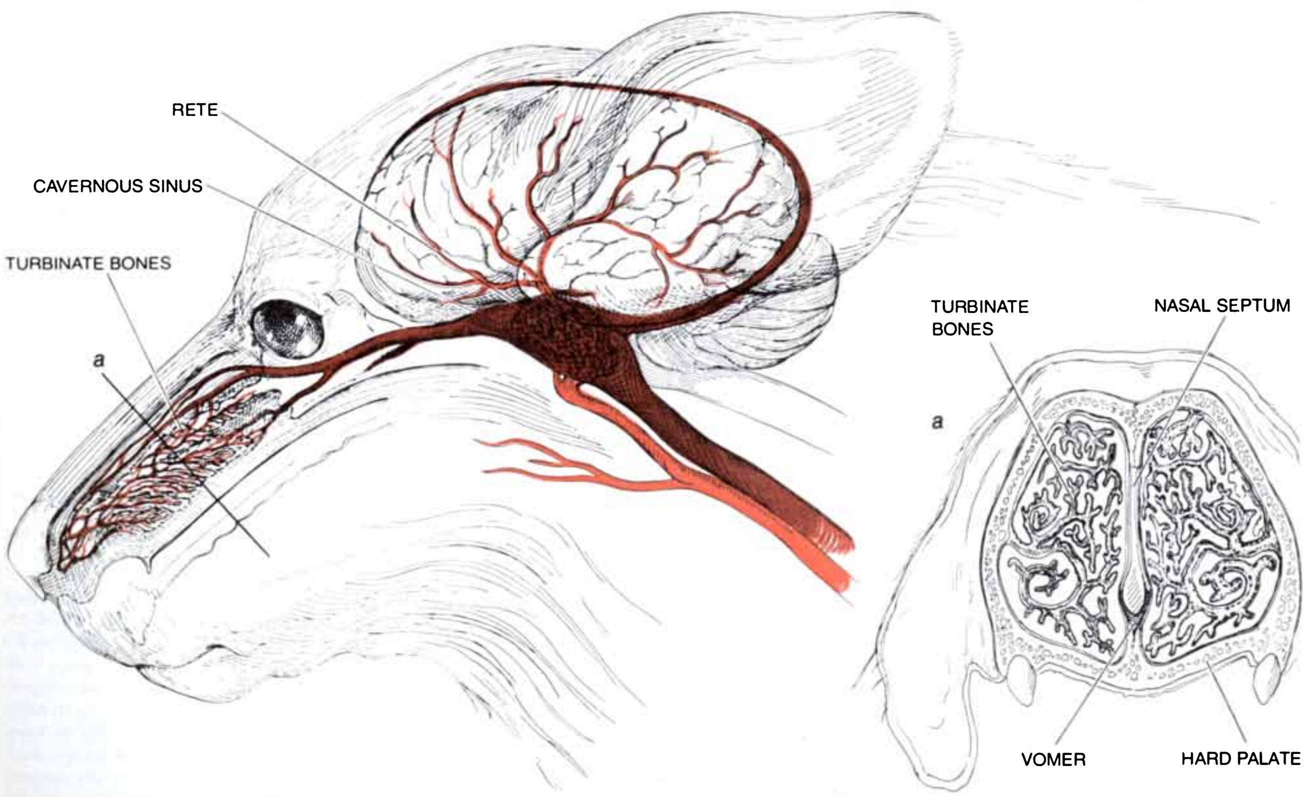
The brain is particularly sensitive to abnormal temperatures. A rise of only four or five degrees C. above normal begins to disturb brain functions. For example, high fevers in children are sometimes accompanied by convulsions; these are manifestations of the abnormal functioning of the nerve cells of the overheated brain. Indeed, it may be that the temperature of the brain is the single most important factor limiting the survival of man and other animals in hot environments.

When the air temperature is higher than the body temperature, mammals would overheat rapidly if it were not for evaporative cooling. Some mammals cool themselves by both panting and sweating. Others, man and the rest of the primates included, depend exclusively on sweating. Still others rely entirely on the evaporation from the nasal and oral cavities that panting enhances. Of the mammals in this last group those with a carotid *rete* have the greatest advantage. This, we now realized, is because the venous blood that drains the evaporative surfaces of the nose and mouth almost immediately comes in close contact with the arteries of the carotid *rete*. Hence the blood supplying the brain receives the full benefit of the evaporative cooling. At the same time, when this venous blood returns to the heart to mix with venous blood from the rest of the animal's body, it provides further cooling for the circulatory system in general.

In most panting animals both the nose and the mouth are anatomically well suited to efficient evaporative cooling. The surface areas are large and moist, and the numerous connections between arteries and veins allow a high rate of blood flow. Considering the nose first, the complex infolding of the turbinate bones presents a very large surface area; in some mammals, such as dogs, the total surface of the nasal turbinates has been calculated as being larger than the rest of the body surface. The nasal interior not only is covered with mucosa richly supplied with interconnected veins and arteries but also houses an organ unique to panting mammals: the lateral nasal gland. The gland secretes fluid onto the mucosa, providing a constant supply of water for evaporation. In dogs the rate of secretion of the lateral nasal gland, as studies by Charles Blatt, C. R. Taylor and M. B. Habal of Harvard University have shown, is directly proportional to the air temperature.

As for the other evaporative region, the mouth, water for cooling the mucosa that lines the oral cavity is provided by the salivary glands. These glands too increase their secretion when the ambient temperature increases. For the evaporative process to reach maximum efficiency the animal's rate of ventilation must increase at the same time the glandular secretions and the rate of blood flow through the mucosa increase; the act of panting produces the required increase in ventilation.

Does exposure to high air temperatures represent the most severe thermal stress a mammal can encounter? For most mammals the answer is no. For example, in mammals the size of cats, dogs and even sheep, Taylor has found that it is exercise rather than heated air that presents the most serious threat of overheating. In these medium-size mammals the very high rate of metabolism during heavy exercise causes an explosive rise in body temperature. For the purpose of comparison, consider our finding that when cats and sheep are at rest under conditions of abnormal



DOG'S NOSE, shown in side and front view, contains an array of complexly infolded turbinate bones that support the nasal mucosa, a membrane that is cooled by evaporation. Cool venous blood, draining from the mucosa into the cavernous sinus, cools the dog's *rete*.

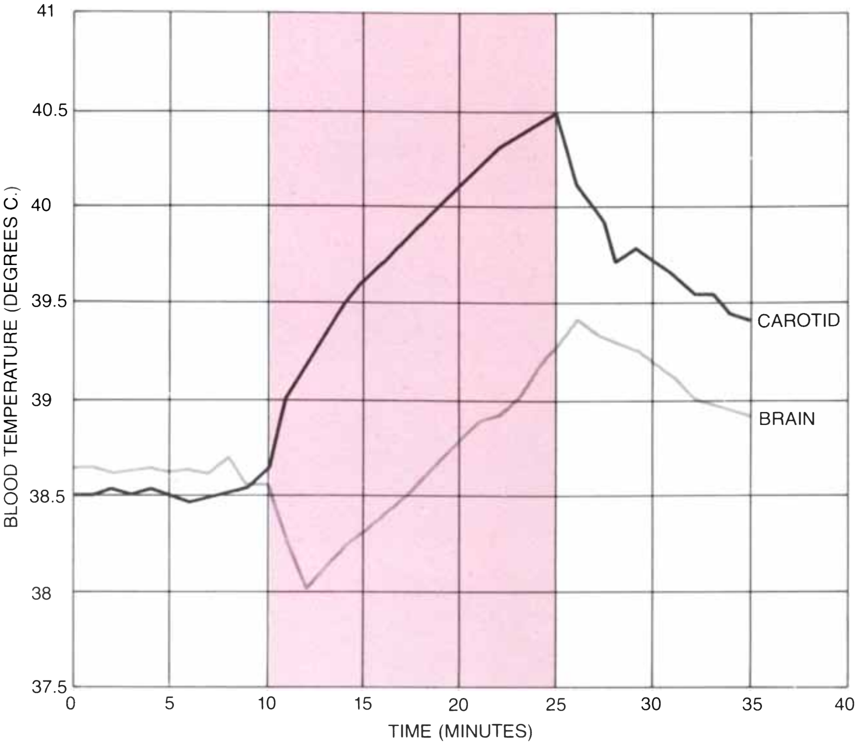
heat, they display a brain temperature one degree C. lower than their body temperature. Taylor and his colleague Charles P. Lyman made similar measurements of brain and body temperature in Thompson's gazelle (an African antelope with a carotid *rete*) when the animals were running at full speed. They found that the brain temperature dropped to almost three degrees C. below the body temperature.

This unexpectedly large cooling effect led me and my colleagues at the University of California at Riverside to suspect that the carotid *rete* might not reach its maximum efficiency as a heat exchanger except during exercise. Recent studies we have conducted with exercising dogs as subjects have confirmed the conjecture, even though dogs do not have a carotid *rete* as well developed as that of antelopes, sheep or cats.

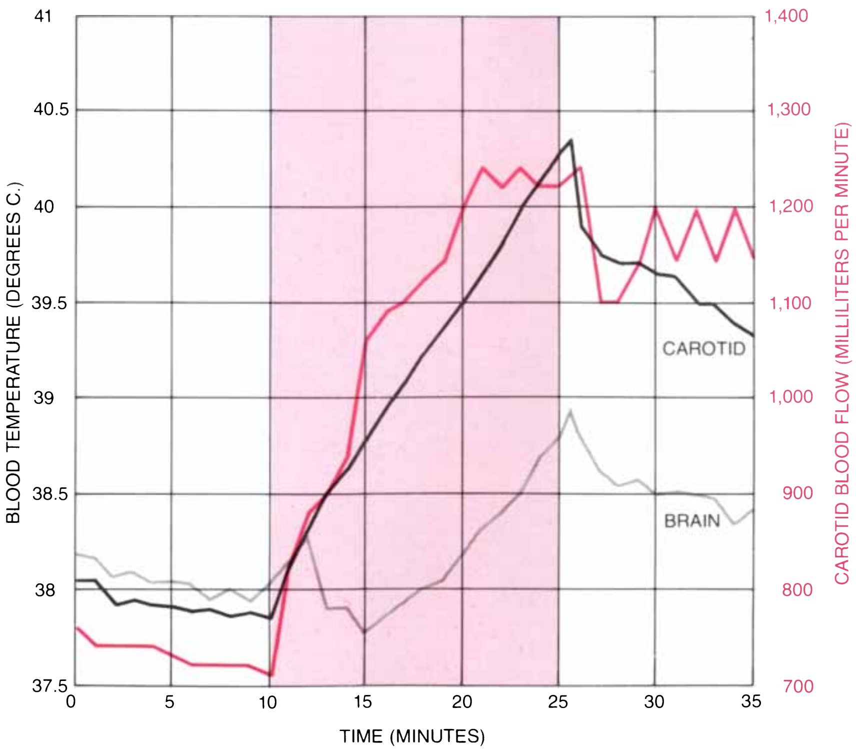
When the dogs were put in hot rooms, they reacted scarcely at all to the abnormal temperature. Both the body and the brain rose in temperature, and the brain remained only slightly cooler than the body. We found this puzzling, because dogs are known to be among the most heat-tolerant of mammals and to have a marked capacity for enduring long periods of exertion in extreme desert conditions.

We set up a treadmill in a warm, rather than hot, room and measured the dogs' brain and body temperatures as they ran on the treadmill. From the moment the dogs began to run their body temperature rose rapidly. Within the first few minutes of exercise, however, the brain temperature fell; it remained about 1.3 degrees C. below the body temperature throughout the run. That is almost three times the cooling we had observed when the dogs were at rest in a hot room.

Two of the physiological factors responsible for this high rate of brain cooling have recently been measured. The first is a large increase in the rate of ventilation when dogs exercise. R. Flandrois and his colleagues at the University of Lyons have found that a dog at rest breathes five to six liters of air per minute. In the first few seconds of hard exercise (about the same degree of exertion that is demanded of the dogs in our laboratory) Flandrois's dogs increased their rate of ventilation to 30 liters per minute. The rate thereafter rose less dramatically and reached 40 liters per minute after 15 minutes of exercise. When exercise was terminated, the rate of ventilation dropped abruptly. The primary purpose of the increase in ventilation, of course, is to deliver more oxygen and flush out more carbon dioxide in order to meet the physiological demands associated with exercise. At the same time the increase has the secondary effect of accelerating evaporation in the dogs' nasal and oral pas-



EXERCISING DOGS maintained a brain temperature significantly lower than body temperature during a 15-minute trot at 7.2 kilometers per hour (*colored band*) on a treadmill with a slope of 13 degrees. Warmer than the body while the dogs were at rest, the brain rapidly grew cooler at the onset of exercise and remained more than one degree C. cooler throughout the exercise period. The two traces show the mean values recorded for two dogs in four trials each.



INCREASED CIRCULATION during exercise is traced in another exercising dog, this one trotting on a treadmill with a slope of 18 degrees. At rest the blood flowed through the dog's left common carotid artery (*color*) at a rate of less than 800 milliliters per minute. During the 15-minute exercise period (*colored band*) the rate of blood flow increased to between 1.2 and 1.25 liters and remained at 1.1 liters or above for several minutes after exercise ended. The enhanced flow of blood to the nose and mouth caused an increase in the heat exchange in the carotid *rete*, as is evidenced by the consistently lower temperature of the animal's brain.

sages, thereby lowering the temperature of the venous blood.

The second brain-cooling factor to be measured is the increase in the rate of blood flow to the nasal and oral mucosa during exercise. Working in collaboration with Roland D. Rader and William Kirtland of the University of Southern California School of Medicine, we have measured the flow of blood through the common carotid artery in dogs during exercise, using an ultrasonic flow probe designed by Rader. In a large dog at rest the flow of blood through the common carotid artery is between 600 and 700 milliliters per minute. In our experiments we found that the rate began to rise at the onset of exercise and eventually reached a level in excess of 1,200 milliliters per minute.

There is no reason to believe the rate of blood flow to the brain increases during exercise. Accordingly we can assume that the additional blood flowing through the common carotid artery during exercise circulates to areas other than the cranium. Subtracting the flow destined for the brain (roughly 250 milliliters per minute), we can calculate that the flow destined for extracranial areas is about 400 milliliters per minute when the dog is resting and a full liter per minute when the dog is exercising. Since the arterial circulation passes through both common carotid arteries, the combined flow of two liters per minute represents about a sixth of the total output of the heart of an exercising dog.

Most of this extracranial flow of blood is destined for the mucosa of the dog's nasal and oral cavities, where the increased rate of ventilation during exercise has accelerated the process of evaporative cooling and the dog's lateral nasal glands and salivary glands are secreting increased amounts of water for evaporation. Hence the combination of physiological factors, by significantly cooling the brain of a panting mammal that has a carotid *rete*, provides an increased tolerance to exertion, particularly in a hot environment.

What about panting mammals that do not have a carotid *rete?* We have studied the brain and body temperatures of domestic rabbits while the animals were running. We found that their brain temperature increased in parallel with their body temperature. The animals' panting nonetheless can do something to offset the rise in brain temperature, as M. Caputa and his colleagues at Nicholas Copernicus University of Toruń in Poland have demonstrated. Measuring the temperature of venous blood returning from the nose of panting rabbits, they found that it was cooled by evaporation. As a consequence cooling by conduction lowered the temperature of regions of the rabbits' brain adjacent to the cool venous blood; these parts of the brain proved to be .5 degree C. lower in temperature than the rabbits' body.

In nature this offsetting factor appears to be trivial. Vaughan H. Shoemaker and his colleagues at Riverside have used radio telemetry to measure the body temperature of jackrabbits living unrestrained in the Mojave Desert. On a hot summer day a jackrabbit at rest has a body temperature of about 41 degrees C. When the animal is flushed from its resting place and chased, it shows a rapid rise in body temperature. Between five and 10 minutes of exertion can raise its body temperature above 43 degrees. Knut Schmidt-Nielsen and his colleagues at Duke University have demonstrated that if a jackrabbit's body temperature rises much above 44 degrees, the animal will die. It is therefore apparent that a jackrabbit unable to rest and cool off between brief periods of running cannot survive for long.

The study of the heat-exchange function of the carotid *rete* in panting mammals has raised intriguing questions about mammalian temperature regulation in general. It has long been known that there is a heat-sensitive region in the brain stem of mammals and that by sensing the temperature of the arterial blood flowing from the core of the body to the brain this region probably acts as a thermostat controlling the body-core temperature. This seems to be a valid assumption for mammals that lack the carotid *rete;* in those mammals the arterial blood does not change temperature as it passes from the body to the brain. What, however, does the brain-stem thermostat sense in a mammal that has a carotid *rete?*

What it senses is at the very least equivocal. For example, during exertion the temperature sensed by the thermostat is quite different from the actual temperature of the body core; the brain temperature and the body temperature change in opposite directions at the onset of exercise and remain at different levels during most of the exercise period. How, then, does the brain thermostat know what the body temperature is? In mammals that have a carotid *rete* are the temperatures of the brain and the body regulated independently?

We can speculate that in such animals temperature-sensitive nerve cells in parts of the body other than the brain are important in the regulation of temperature during exertion. For example, such sensors exist in the spinal cord and in the abdominal cavity and have been shown to play a thermoregulatory role in some mammals. In mammals with a carotid *rete* the input from these extracranial sensors to the region of the brain that controls body temperature may be more significant during a period of exertion than during a period of rest. What the right answers are to such questions will be learned from further study of thermoregulation in a wide variety of mammals.

Intrinsically Difficult Problems

Some kinds of computational problems require for their solution a computer as large as the universe running for at least as long as the age of the universe. They are nonetheless solvable in principle

by Larry J. Stockmeyer and Ashok K. Chandra

Some problems in mathematics (particularly in number theory and logic) can be solved in principle but only by methods that exceed the capacity of the most powerful computer imaginable. Certain intrinsic properties of the universe will always limit the size and speed of computers. The most powerful computer that could conceivably be built could not be larger than the known universe (less than 100 billion light-years in diameter), could not consist of hardware smaller than the proton (10^{-13} centimeter in diameter) and could not transmit information faster than the speed of light (3×10^8 meters per second). Given these limitations, such a computer could consist of at most 10^{126} pieces of hardware. Albert R. Meyer of the Massachusetts Institute of Technology and one of us (Stockmeyer) have proved that this ideal computer, regardless of the ingenuity of its design and the sophistication of its program, would take at least 20 billion years to solve certain mathematical problems that are known to be solvable in principle. Since the universe is probably not 20 billion years old, it seems safe to say that such problems defy computer analysis.

Problems that would stump the ideal computer for at least 20 billion years constitute the most extreme cases of intractable problems. Short of these extreme cases are problems that would take modern high-speed computers years to solve regardless of the methods of solution. Then there are problems for which fast methods of solution have not been found, although such methods may exist. In recent years mathematicians and computer specialists have formalized the idea of intractability, have precisely identified several kinds of intractable problem and have proved many theorems about each kind. The results of this work raise fundamental questions about the capabilities and limitations of computers.

Intractable problems seem to come up often in games such as chess, checkers and go. Suppose two people are playing a game of checkers in which there are only four pieces left on the board: three red kings and one black king. If the player with the red kings has the next move, he has a strategy that will lead ultimately to victory no matter what counterstrategy black adopts. If instead perhaps a dozen pieces were left on the board, then in most cases it would not be obvious whether or not red had an unbeatable strategy. It is possible in principle to determine the answer by "brute force": by examining all possible continuations for red and all possible countermeasures for black.

The strategies and counterstrategies can be investigated systematically by constructing a "game diagram" consisting of a collection of points connected by arrows. Each point corresponds to a checkers position, that is, to the entire arrangement of the pieces together with the information about who has the next move. Each possible move in a given position is represented by an arrow that connects corresponding points. When a position includes no black checkers, it is a winning position for red and is marked as such. When a position includes no red checkers, it is marked as a winning position for black.

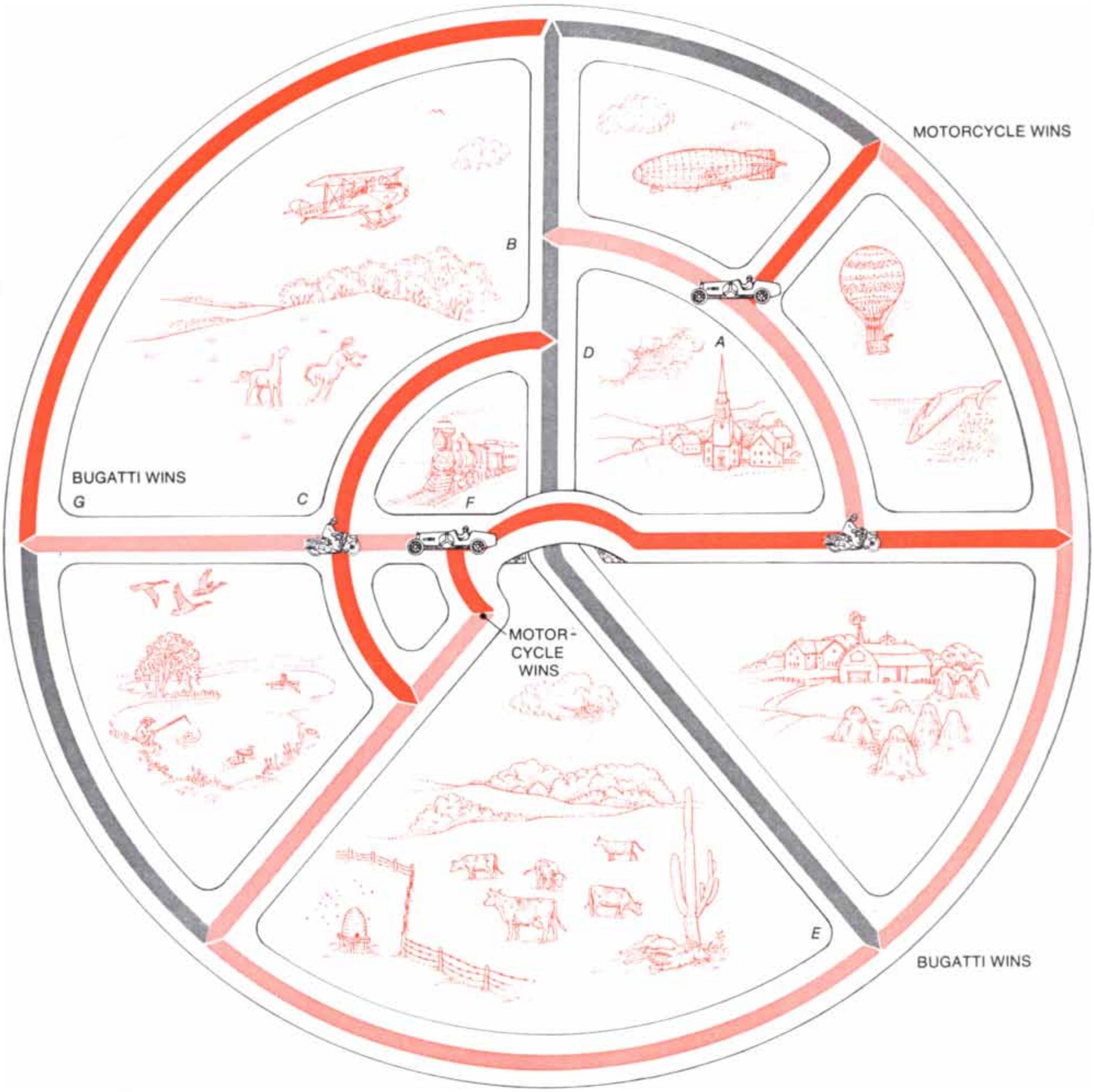
The other points (positions) in the diagram can now be identified as winning positions for either red or black on the basis of a few simple rules. If red can move from position *A* to position *B* and *B* has already been established as a winning position for red, then *A* too can be marked as a winning position for red. By the same token, if from position *A* red can move only to positions that have already been identified as winning for black, then *A* too can be marked as winning for black. The marking procedure continues until no more points can be marked. Positions that are not marked are destined to be a draw. If care is taken so that different points in the diagram do not correspond to the same position, the diagram can be constructed and marked in a finite amount of time. In this way any initial position can be analyzed either to a draw, a win for red or a win for black.

Such a method of analysis is called an algorithm: a precise set of instructions that can be mechanically applied to yield the solution to any given instance of a problem. If a problem is to be solved by a computer, an algorithm is indispensable, because the computer can follow only instructions that are stated in the unambiguous form of an algorithm. Examples of algorithms are the procedures taught in elementary schools for adding and multiplying whole numbers. When these procedures are mechanically applied, they always yield the correct result for any pair of whole numbers. The construction of a game diagram is an algorithm that analyzes any checkers position. Essentially the same algorithm can be used to analyze games, such as chess and go, that do not involve dice, cards, spinners or other elements of chance. In general the algorithm for a game determines which player (if any) has an unbeatable strategy in any given position.

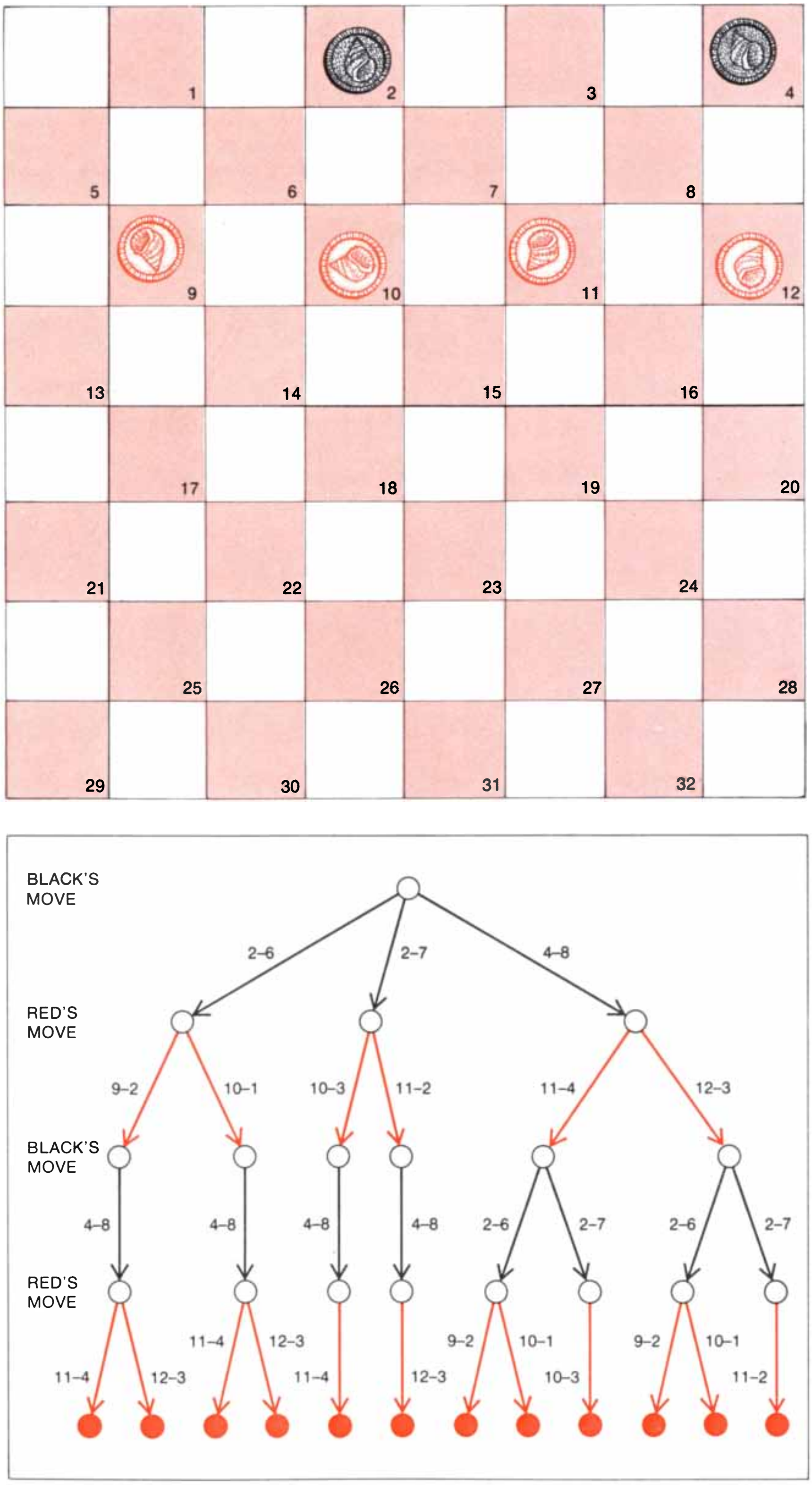
Why, then, can computers not play perfect checkers and perfect chess? The answer is simply that for typical positions in these games the diagram consists of a tremendous number of points, so many in fact that the fastest existing computers would take years to examine them all. Computer programs that play chess and checkers do not generate the entire diagram but only small parts of it, in much the same way that a human player does when he "looks ahead" in these games. As a result the programs do not play perfectly, although some of them play fairly well compared with human players.

Mathematicians have not been able to rule out the possibility, unlikely though it may be, that computer programs will someday incorporate algorithms for playing perfect chess and checkers that are more efficient than the brute-force method of constructing entire game diagrams. In fact, clever and efficient algorithms have already been devised for several games. One such game is nim, in which a number of stones or other counters are initially arranged in a number of piles. Two players go in turns, and on his turn a player can remove any number of stones from any one of the piles (although he must remove at least one stone). The player who manages to remove the last stone is the winner.

Although the positions of nim could be analyzed satisfactorily by constructing a game diagram, there is a much more efficient algorithm that can analyze any position by performing a few calculations that are as simple as the adding of numbers [*see illustration on page 146*]. The algorithm is so efficient that a human player can evaluate positions consisting of thousands of stones in just a few minutes with only pencil



"ROADBLOCK" is a board game played on a network of roads in which each road is drawn in one of three colors. Two players, who alternate moves, each have a fleet of cars that initially occupy certain designated intersections. On his turn a player moves one of his cars to a new intersection by traversing roads of the same color that are free of other cars. The winner is the first player to move a car to a designated goal intersection. In the illustration the player with the Bugattis could move one of them from intersection *A* to intersection *B*, threatening to win on his next turn by moving from *B* to *E*. If his opponent tries to block the road from *B* to *E* by moving one of his motorcycles from *C* to *D*, the first player wins by moving a Bugatti from *F* to *G*. This means that if the player with the Bugattis has the first move, he has an unbeatable strategy. In general it is not this easy to analyze an arbitrary position in roadblock. There exist algorithms, or mechanical routines, for determining which player has an unbeatable strategy in any given position, but all of them are "inefficient": the time it takes to execute any one of them grows exponentially as c^n does, where *c* is a constant greater than 1 and *n* is the number of intersections in the position. Although every algorithm will not take an exponential amount of time to find an unbeatable strategy for one of the players in every position, for each algorithm there are infinitely many positions that take an exponential amount of time to evaluate.



CHECKERS POSITION (*top*) can be analyzed by a "game diagram" (*bottom*) to determine whether the outcome of perfect play is a win for red, a win for black or a draw. Each point in the diagram corresponds to a checkers position, with the first point standing for the position in the illustration. Each possible move is represented by an arrow that is labeled by two numbers, the first indicating a square a checker occupies and the second indicating a square to which that checker could be moved. Because all the points (*red*) in the bottom row correspond to positions that do not include any black checkers, the initial position is obviously a win for red. In general the method of making a game diagram cannot be used to analyze any given position because the size of the diagram grows exponentially as a function of the number of checkers.

and paper. To evaluate such positions by constructing a game diagram would take years, even with the aid of the fastest computer. The game of nim illustrates an important point, namely that there can be both fast and slow algorithms for solving the same problem.

The efficiency of an algorithm depends on how the time it takes to execute it increases as the size of the input increases. The input is the particular instance of the problem that the algorithm is supposed to solve, such as a position in checkers or nim. The inefficiency of analyzing checkers by constructing a game diagram is revealed by the fact that the number of points in the diagram grows exponentially as a function of the number of checkers in the position. In other words, for positions consisting of n checkers the diagram has at least c^n points, where c is a constant greater than 1. A small increase in the number of checkers results in an explosive increase in the size of the diagram and hence an explosive increase in the time required for constructing and marking the entire diagram. Although n is at most 24 for regular checkers, the algorithm applies to arbitrarily large versions of the game, for example a 20-by-20 board with 60 checkers. Of course, the exponential growth in execution time for this brute-force algorithm also applies to these larger values of n. In contrast, the execution time of the fast algorithm for nim grows as a linear function, such as $6n$ or $10n$, where n is the number of stones in the position. The rapid growth of exponential functions and the comparatively modest growth of linear functions accounts for the disparity in efficiency between the two algorithms.

It is generally agreed that an algorithm whose execution time grows as c^n does is not practical. If c is only slightly greater than 1, say 1.1, such an algorithm might be useful for very small inputs. Even so, the exponential growth explodes for larger inputs; the exact value of c affects only the point at which the exponential growth takes effect and makes the algorithm useless. If one wants to solve a certain problem for which only exponential-time algorithms are known, there is good reason to look for a more efficient algorithm. If a much faster algorithm is discovered, the resources invested in finding it will pay interest every time the algorithm is applied.

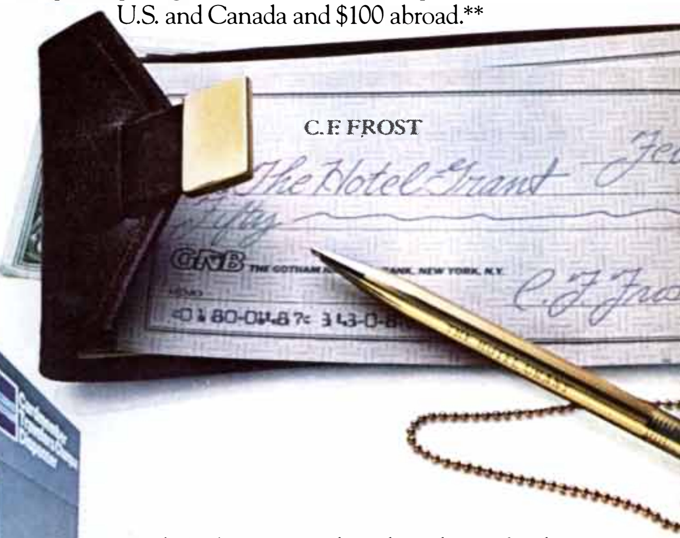
There is always the danger, however, that the investment will not pay off because an efficient algorithm for solving the problem may not exist. It is possible that the complexity of the problem makes it "inherently exponential": all its algorithmic solutions require an exponential amount of time. This possibility

If you're out of town and out of cash, you're not out of luck if you're an American Express Cardmember.

"I was U.S. Treasurer, so I know money isn't everything. But when I travel, I feel more secure knowing that with the American Express Card and my personal check I can get some money at an American Express Travel Service Office.* Up to $800 in American Express® Travelers Cheques and $200 in cash."

Francine Neff

If you're a stranger in town and need to cash a check, the American Express Card backs your personal check at participating hotels and motels for up to $250 in the U.S. and Canada and $100 abroad.**



Look for this at selected major airports around the country. It's the American Express Cardmember Travelers Cheque Dispenser. Day or night, weekends and holidays, you can get from $100 to $500 in American Express Travelers Cheques. Cardmembers must first enroll for this special service. For details just call 800-528-8000.

Did you bring enough cash with you for the taxi ride into town from the airport? Don't get caught stranded. Your personal check for $50 is guaranteed at the ticket counters of participating domestic airlines when you have the Card and a valid ticket.

Being able to cash a check where people don't know you is only one example of the travel security of carrying an American Express® Card. Along with the worldwide recognition and charge convenience that lets you buy almost anything, almost anywhere. For more information and an application, just call 800-528-8000, toll-free.

The American Express Card. Don't leave home without it.℠

AMERICAN EXPRESS

"If we lose our American Express Cards, we're not as lost as you'd think. Because an American Express Travel Service Office can replace our Cards in a day or two, and help arrange for emergency cash, travel and lodging. We just have to convince them that we are who we say we are. I'm Bob and he's Ray. No, I'm Ray..."

Bob Elliot & Ray Goulding

*At most of the more than 750 Travel Service Offices of American Express Co., its subsidiaries and Representatives. (Overseas, subject to local currency laws/regulations and cash availability.)
**Subject to cash availability.

Introducing the last word in First Class.

Today, with the development of our long-range 747SP, Pan Am can take you to places nearly halfway around the world non-stop.

But non-stop flying doesn't mean much without non-stop service.

So, for First Class passengers on our 747SPs, we are proud to introduce new ideas you won't find on any other airline.

Like our new Sleeperette® seat. When it reclines, the Sleeperette® stretches out to the length of four-and-a-half of our windows. Perfect for reading, enjoying stereo or getting a good night's rest.

Another First Class exclusive: a choice of banquette table-for-two seating or individual dining at your Sleeperette® seat. And, of course, superb international cuisine.

Whether you fly First Class on our 747SP with the Sleeperette® seat, or on our traditional 747, you'll experience the same high standard of service that our First Class passengers have always enjoyed.

You see, we want to be the world's first choice in First Class service.

See your Travel Agent or transportation department.

PAN AM®

We fly the world the way the world wants to fly.

has motivated mathematicians to try to prove that certain problems are inherently exponential. For many problems, such as the analysis of positions in chess and checkers, the fastest-known algorithms use an exponential amount of time, although it has not been proved that efficient algorithms do not exist.

The first proof that a problem was inherently exponential was put forward by Meyer and one of us (Stockmeyer) in 1972 for a problem in the area of computer science known as formal language theory. Similar proofs for problems in other mathematical disciplines soon followed. In 1976 we proved that the analysis problems for several games were inherently exponential. One of these games is "roadblock," which is played on a board consisting of a network of intersecting roads. Each road is drawn in one of three colors. The two players, who make alternate moves, each have a collection of markers that are placed initially on certain designated intersections. On his turn a player moves one of his markers along the roads to a new intersection, with the proviso that the traversed roads all be the same color and free of other markers. The winner is the first player to move a marker to a designated "goal" intersection.

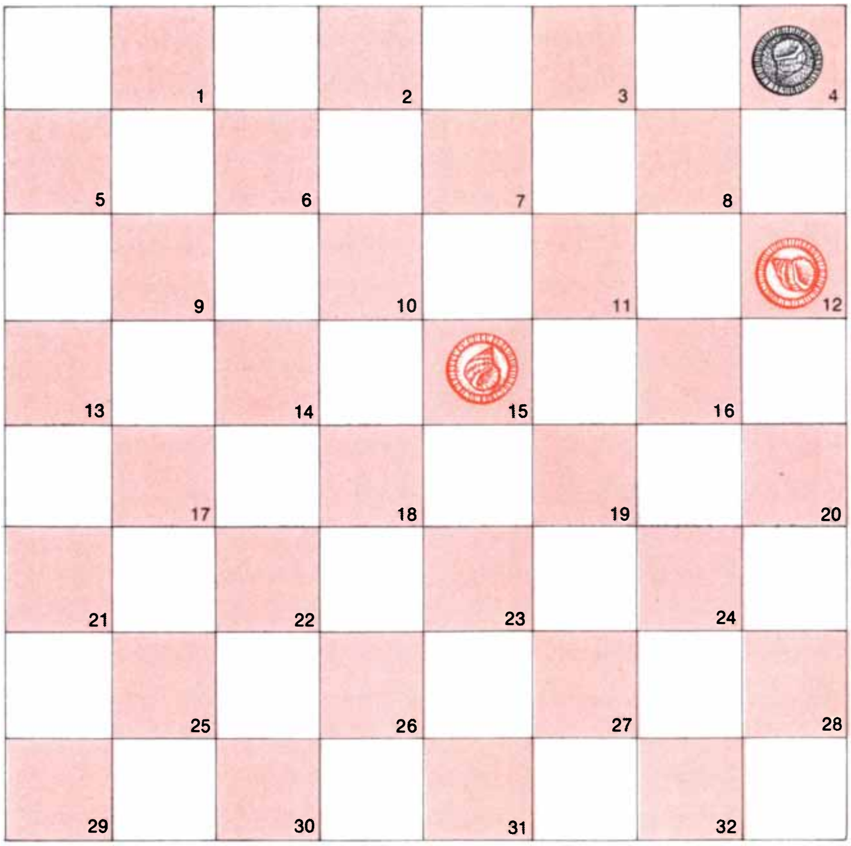
Given any position in roadblock (an arbitrary network of roads and an arbitrary initial placement of the markers), the problem of analysis is to determine which player has an unbeatable strategy. We proved that the execution time of every algorithm solving the analysis problem grows as c^n does, where c is a constant greater than 1 and n is the number of intersections in the position. It should be emphasized that this conclusion is not just a good guess that has been substantiated empirically by trying several algorithms and failing to find an efficient one; the conclusion has been verified by a mathematical proof.

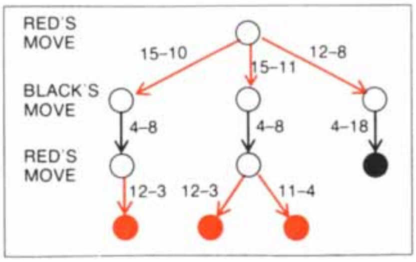
It is important to realize that the proof does not demonstrate that every position with n intersections will cause every algorithm to consume an exponential amount of time. Like the position in checkers of three kings against one king, there are positions in roadblock that are easy to analyze. Most of the problems of the game are not complex enough to fool all of the algorithms all of the time. What has been established is that the analysis problem for roadblock is complex enough to fool all of the algorithms some of the time. In other words, regardless of the algorithm chosen there are difficult positions that will cause it to consume an exponential amount of time. We know almost nothing about the frequency of these difficult positions. The most we can say is that there is at least one such position for each value of n.

The concept of an algorithm was first captured in precise mathematical terms in the 1930's by logicians who investigated the conditions under which the proofs of mathematical theorems could be generated automatically by a mechanical process. A. M. Turing, the noted British mathematician, was one of the pioneers in developing a precise formulation of such a process. He postulated that any algorithm could be executed by a machine consisting of an infinitely long paper tape divided into squares, a printing mechanism that writes and erases marks on the tape and a scanner that senses whether or not a given square is marked. Known as a Turing machine, this imaginary device can be programmed to find the solution to a problem by executing a finite number of scanning and printing actions.

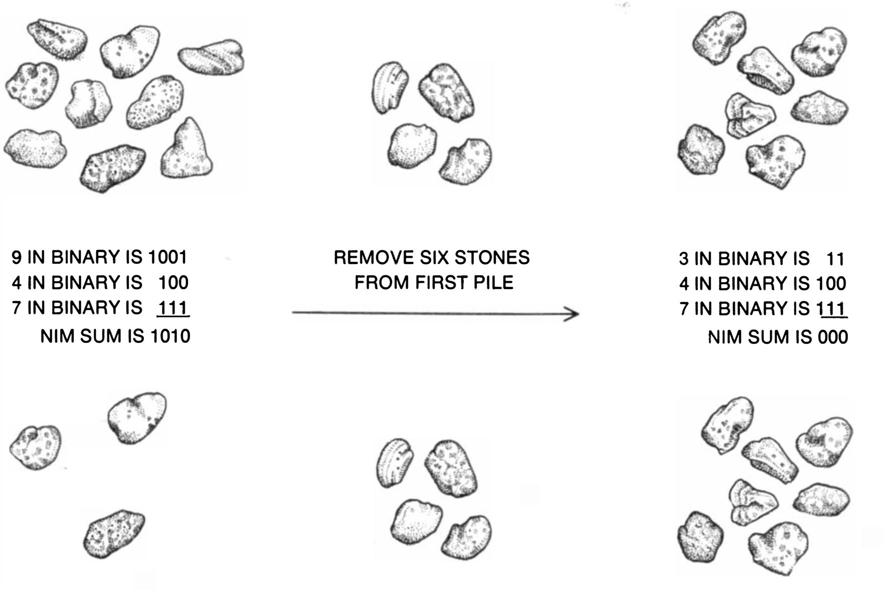
What is remarkable about the Turing machine is that in spite of its simplicity it is not exceeded in problem-solving ability by any other known computing device. If the Turing machine is given enough time, it can in principle solve any problem the most sophisticated computer can solve. As a result the fact that a problem can be solved by a Turing machine has been generally accepted as a necessary and sufficient condition for the solution of the problem by an algorithm.

Like a Turing machine, a modern high-speed computer has a memory and a set of instructions for retrieving stored information, processing the information and entering new information in the memory. The memory is usually divided into a sequence of individual cells that can be addressed separately. Because these computers have the ability to directly address any memory cell, they are called random-access machines. It has been proved that a ran-





GAME DIAGRAM (*bottom*) of a checkers position (*top*) may have both final positions that are wins for black (*black points*) because they include no red checkers and final positions that are wins for red (*red points*) because they include no black checkers. Red has the move in the initial position, and so red can force a win by moving from square 15 to either square 10 or square 11. Hence in the initial position red has an unbeatable strategy.



NIM is a game for two players in which stones or other counters are arranged in piles. On his turn a player removes any number of stones from one of the piles. (He must remove at least one stone.) The player who removes the last stone is the winner. Although nim can be analyzed completely but inefficiently by the construction of a game diagram, it can also be analyzed by a much more efficient algorithm. To determine whether or not the player with the move in the top position has a winning strategy, the number of stones in each pile should be expressed as binary numerals that are written one above the other (*middle*). Next the nim sum is computed by writing a 1 below each column that includes an odd number of 1's and a 0 below each column that includes an even number of 1's. If the nim sum is zero, the first player can be beaten. If the nim sum is not zero, the first player can always win by moving on every turn to a position that yields a zero nim sum. For example, in the top position the first player should take away six stones from the first pile in order to reach the bottom position. This nim-sum algorithm works efficiently for any number of piles and for any number of stones.

dom-access machine can solve only those problems that can be solved by a Turing machine. Although the Turing machine may require many steps to simulate one action of the random-access machine, this disparity in computational power disappears in the realm of exponential amounts of time. If a random-access machine can solve a problem in an exponential amount of time, so can a Turing machine, and vice versa. When a problem is described as requiring c^n time to solve regardless of the algorithm, the description applies to algorithms for Turing machines, random-access machines and all other computers in existence or likely to be built in the future. In general, however, the exact value of c depends on the specific computing device.

How is it proved that a problem requires an exponential amount of time to solve regardless of the algorithm? The direct approach is to imagine all possible algorithms for a problem and prove that all of them require an exponential amount of time. For most problems the direct approach has not been successful. The difficulty is that one must discern not only the obvious methods of solution but also the subtle ones. Algorithms can operate in strange and mysterious ways, and it is difficult to know when all these ways have been discovered and examined. For example, anyone who is not familiar with game theory would probably overlook the fast solution of nim and erroneously suppose the solution of the game requires an exponential amount of time.

The direct approach has been useful, however, in proving that problems of inherently exponential complexity do in fact exist. The existence of such problems was established in the early 1960's by Michael O. Rabin of the Hebrew University of Jerusalem and Juris Hartmanis and Richard E. Stearns of the General Electric Research Laboratory. The problems are constructed by the technique of diagonalization, which was invented by the 19th-century German mathematician Georg Cantor to prove that the real numbers cannot be put in one-to-one correspondence with the positive integers [*see top illustration on page 154*]. The inherently complex problems constructed by diagonalization are problems only in an abstract sense, because they do not relate to anything tangible. Nevertheless, they play a fundamental role in proving that certain more down-to-earth problems are inherently complex.

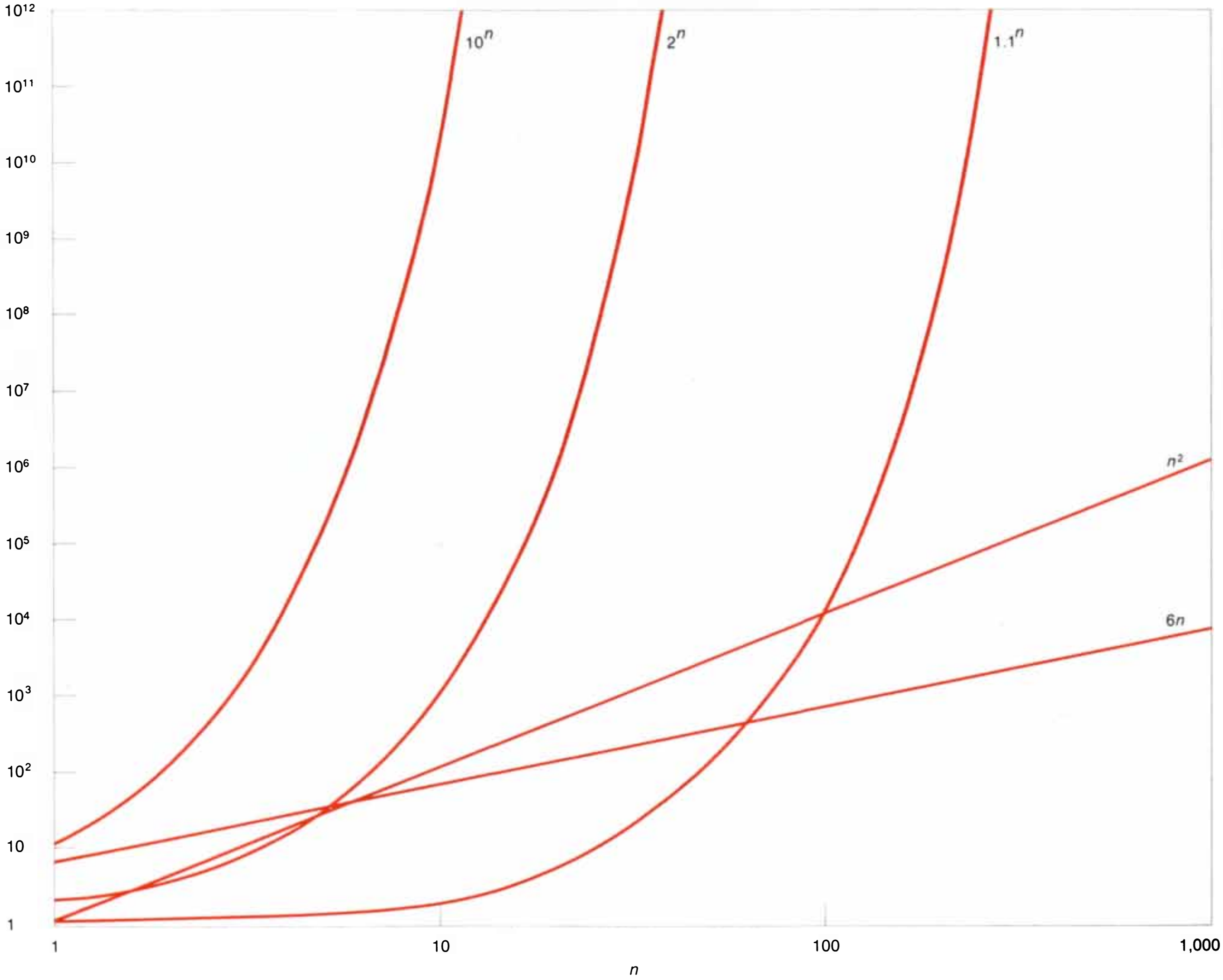
The complexity of the solutions of two seemingly unconnected problems can sometimes be related by demonstrating that one of the problems is "efficiently reducible" to the other. A problem P is efficiently reducible to a problem Q if an efficient solution for Q would provide an efficient solution for P. The usual way of demonstrating that this relation holds between P and Q is to find an efficient algorithm for P that, in the course of its computations, seeks answers from an algorithm for Q. It is necessary that the algorithm for Q provide its answers "for free"; in other words, determining answers to Q must not increase the execution time of the algorithm for P. For example, the problem of squaring two numbers can be efficiently reduced to the problem of multiplying two numbers, because given a number to be squared the multiplication algorithm can simply multiply the number by itself. It is less obvious that the multiplication problem can be efficiently reduced to the squaring problem, but in fact the product of two numbers consisting of n digits can be calculated in an amount of time proportional to n by applying the squaring algorithm twice, because $xy = (x + y)^2/4 - (x - y)^2/4$. Apart from the double application of the squaring algorithm the computations in the formula take an amount of time proportional to n, which in this context is considered to be efficient because there is no known way to multiply or square numbers in that amount of time.

If a problem P is known to be efficiently reducible to a problem Q, then the complexity of P cannot be much more than the complexity of Q. This fact enabled us to prove that the analysis problem for roadblock is so complex that it is inherently exponential. Let EXP stand for the class of all problems that can be solved in an exponential amount of time. We have proved that any problem in EXP is efficiently reducible to the analysis problem for roadblock. Here the method of reduction is somewhat more complicated than that of the multiplication and squaring problems, but the time required to perform the reduction is negligible compared with an exponential amount of time. This result indicates that the complexity of any problem in EXP is not much more than the complexity of the analysis problem for roadblock. Since EXP consists of all problems that can be solved in an exponential amount of time, it includes the problems of inherently exponential complexity constructed by the technique of diagonalization. Such problems can be efficiently reduced to roadblock, and so the complexity of the analysis problem for roadblock must also be inherently exponential.

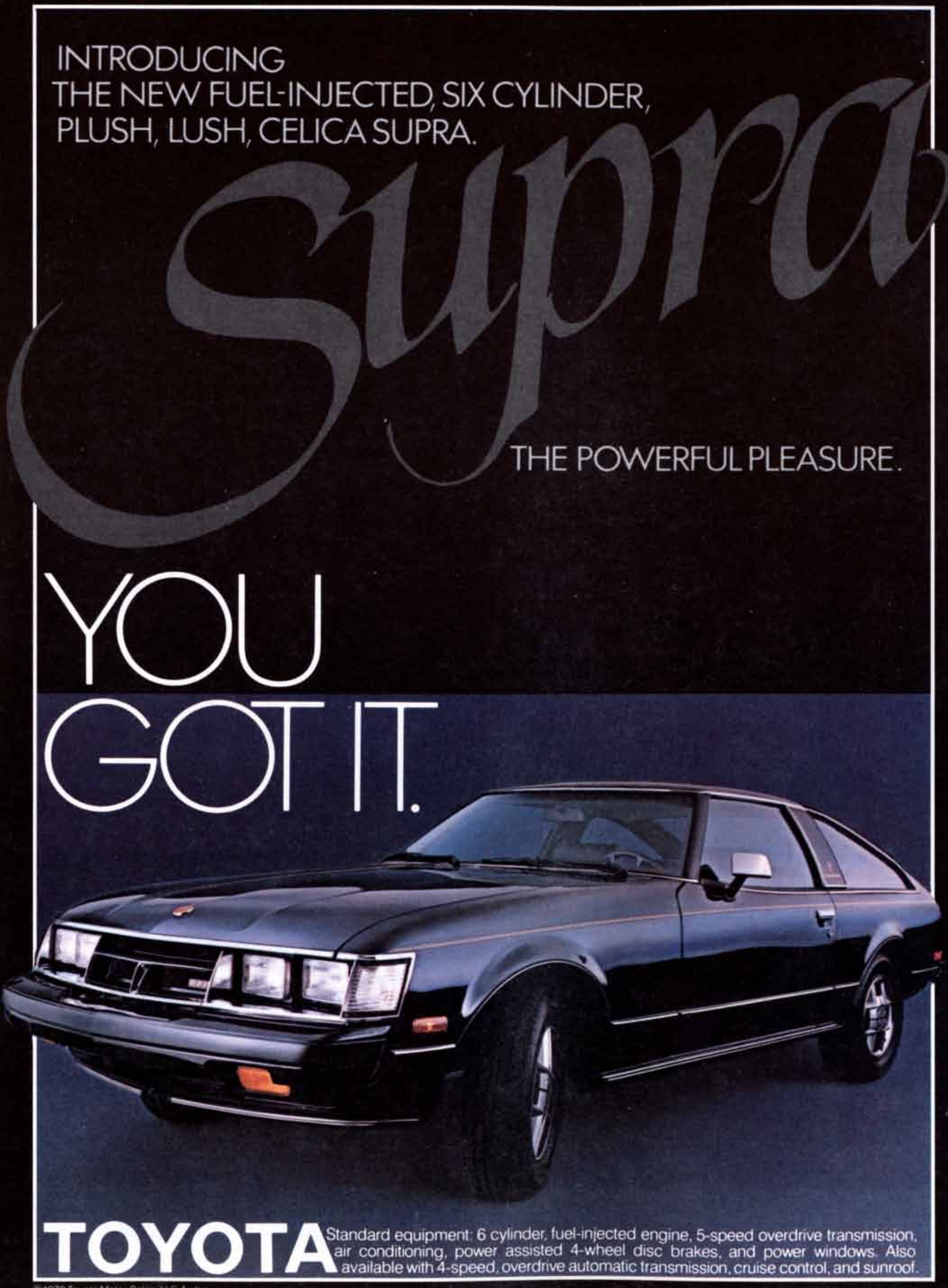
EXP includes the problem of recognizing winning positions in checkers, and so this problem is reducible to that of roadblock. Each position in checkers corresponds to a position in roadblock such that the checkers position is a win for the first player if the corresponding roadblock position is a win for the first player. The correspondence holds not only for checkers but also for chess, go, hex, nim and many other games not governed by chance. That makes roadblock a universal game: its solution would provide the solution to a great many other games.

The technique of efficient reducibility has been exploited to prove theorems about the inherent complexity of recognizing winning positions in go and checkers when those games have been extended to arbitrarily large boards. Go has been studied by David Lichtenstein and Michael Sipser of the University of California at Berkeley and checkers by A. S. Fraenkel and Y. Yesha of the Weizmann Institute of Science in Israel, M. R. Garey and D. S. Johnson of Bell

FUNCTION	TYPE	$n = 1$	$n = 3$	$n = 10$	$n = 30$	$n = 100$	$n = 300$	$n = 1{,}000$
$6n$	LINEAR	6	18	60	1.8×10^2	6×10^2	1.8×10^3	6×10^3
n^2	QUADRATIC	1	9	10^2	9×10^2	10^4	9×10^4	10^6
1.1^n	EXPONENTIAL	1.1	1.33	2.59	17.5	13,781	2.62×10^{12}	2.47×10^{41}
2^n	EXPONENTIAL	2	8	1,024	1.07×10^9	1.27×10^{30}	2.04×10^{90}	1.07×10^{301}
10^n	EXPONENTIAL	10	10^3	10^{10}	10^{30}	10^{100}	10^{300}	$10^{1{,}000}$



EFFICIENCY OF AN ALGORITHM depends on how its execution time grows as the size of the problem increases. The rate of growth is described by a mathematical function. Algorithms that have exponential execution times are of limited use in computer programs because such execution times grow explosively as the size of the problem increases. The game-diagram algorithm for checkers is inefficient because its execution time grows exponentially, whereas the nim-sum algorithm is efficient because its execution time grows linearly. Although a linear function may exceed an exponential function for small *n*, beyond a certain *n* the latter function is always greater.


INTRODUCING
THE NEW FUEL-INJECTED, SIX CYLINDER,
PLUSH, LUSH, CELICA SUPRA.
Supra
THE POWERFUL PLEASURE.
YOU
GOT IT.
TOYOTA
Standard equipment: 6 cylinder, fuel-injected engine, 5-speed overdrive transmission, air conditioning, power assisted 4-wheel disc brakes, and power windows. Also available with 4-speed, overdrive automatic transmission, cruise control, and sunroof.
©1979 Toyota Motor Sales, U.S.A., Inc.

New Books in Developmental Psychology from W. H. Freeman and Company

Social Development

The Origins and Plasticity of Interchanges

Robert B. Cairns, The University of North Carolina

"My goal was—and is—to bring together in a readable and coherent fashion the ideas and evidence central to a contemporary perspective on social behavior and personality. . . . Fifteen of the chapters are concerned mostly or exclusively with child social patterns. . . . The remaining seven chapters provide an introduction to parallel issues in the social development of nonhumans."—from the Preface

1979, 438 pages; 86 illustrations, cloth: 0195-2 $15.00

Human Development

T. G. R. Bower, University of Edinburgh

T. G. R. Bower emphasizes the individuality of development and the importance of finding general principles that still preserve the uniqueness of the individual. Unlike many books in this area, *Human Development* is not a catalog of various age-specific behaviors; instead, it offers a thorough outline of development, incorporating advances made in other developmental sciences and giving students a theoretically and practically sound understanding of the subject.

1979, 464 pages (approx.), 234 illustrations, cloth: 0058-1 $15.00

Understanding Infancy

Eleanor W. Willemsen, University of Santa Clara

"We have written this book for people who are interested in research about the psychological development of a baby—an audience that includes students, pediatricians, parents, and all others who care for and about babies."
—from the Preface

1979, 331 pages, 36 illustrations, cloth: 1002-1 $15.00;
paper: 1001-3 $7.50

Studying Children

An Introduction to Research Methods

Ross Vasta, State University of New York at Brockport

Studying Children gives the reader a thorough introduction to the fundamentals of research in developmental psychology, drawing on all research traditions and both cognitive and behavioral theories. Examples range from infancy through adolescence, and many are taken from well-known published articles. The book first establishes the basic foundations of psychological research and then focuses on specific techniques and issues.

1979, 217 pages (approx.), 188 illustrations (approx.),
cloth: 1067-6 $13.00; paper: 1068-4 $6.50

The Development of Memory in Children

Robert Kail, University of Pittsburgh

"My aim in writing this book was to provide an overview of our knowledge of memory development in children which is comprehensible and interesting to individuals whose background in psychology is limited to the introductory course."—Robert Kail

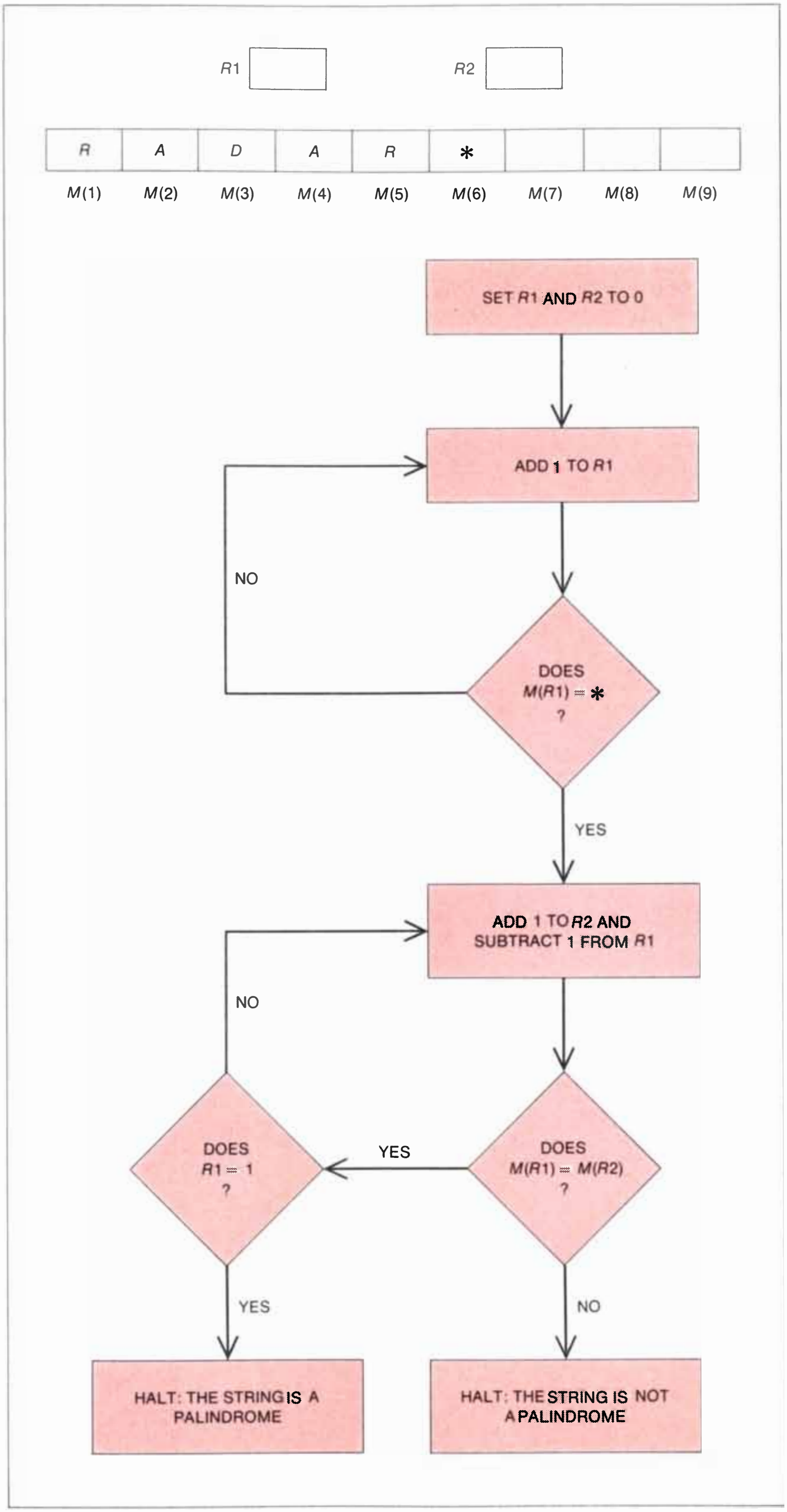
May 1979, 168 pages (approx.), 36 illustrations (approx.),
cloth: 1097-8 $12.00; paper: 1098-6 $5.50

W. H. Freeman and Company
660 Market Street, San Francisco, California 94104

Laboratories and T. Schaefer of the CALMA Company. They have proved that an efficient solution to go or checkers would provide an efficient solution to a great many other problems for which fast algorithms have been sought but none have been found. The remote possibility that these problems have efficient algorithms cannot yet be ruled out, and so the proof does not demonstrate conclusively that checkers and go are inherently exponential, although it is an important step in that direction. It is possible the method we applied to establish the inherent complexity of roadblock will not work for go and checkers and more ingenious methods will have to be invented. Nevertheless, the method of establishing the inherent complexity of problems by diagonalization and efficient reducibility works not only for roadblock but also for certain problems in logic and number theory.

There are many mathematical conjectures that can be stated simply and unambiguously, although their truth remains a mystery. Perhaps the most famous example of such a conjecture is Fermat's last theorem. In the 17th century the prodigious French mathematician Pierre de Fermat, who is the father of modern number theory, wrote extensive notes in his copy of a Greek mathematical treatise. One note presents the proposition that has come to be called Fermat's last theorem: "It is impossible for a cube to be the sum of two cubes, a fourth power to be the sum of two fourth powers or in general for any number that is a power greater than the second to be the sum of two like powers. I have discovered a truly marvelous demonstration of this proposition that this margin is too narrow to contain." In other words, there do not exist positive integers x, y, z and n, where n is greater than 2, such that $x^n + y^n = z^n$. Although the ablest mathematicians of the past three centuries have tried to prove (or disprove) this proposition, no one has succeeded [see "Fermat's Last Theorem," by Harold M. Edwards; SCIENTIFIC AMERICAN, October, 1978].

In other instances the proof of a conjecture was finally found after it had eluded the best mathematicians for years. One such conjecture is the four-color-map theorem, which was put forward in 1852 but was not proved until 1976 by Kenneth Appel and Wolfgang Haken of the University of Illinois. The theorem states that it is possible to color every country on any flat map one of four colors without bordering countries having the same color. Appel and Haken developed their proof with the massive aid of high-speed computers, which spent 1,200 hours in effect coloring some 2,000 maps in as many as 200,000 ways [see "The Solution of the Four-



RANDOM-ACCESS COMPUTER MEMORY has a sequence of cells M(1), M(2), M(3)... that can be addressed directly. In this example R1 and R2 are two additional storage registers. The algorithm shown here determines whether or not a string of alphabetic characters is a palindrome: a string of characters that reads the same backward and forward, for example "radar." A string is put in the memory followed by a symbol (*) marking the end of the string. When the algorithm is given a string of n characters, it tests the string by executing at most $6n + 5$ steps.

	ALGORITHM 1	ALGORITHM 2	ALGORITHM 3	ALGORITHM 4 • • •
INPUT 1	YES	NO	NO	YES
INPUT 2	YES	NO	NO	YES
INPUT 3	NO	YES	NO	YES
INPUT 4 ⋮	YES ⋮	NO ⋮	YES ⋮	YES ⋱

	DIAGONAL PROBLEM
INPUT 1	NO
INPUT 2	YES
INPUT 3	YES
INPUT 4 ⋮	NO ⋮

DIAGONALIZATION is a conceptual process for constructing inherently complex problems, that is, problems that can be solved only by algorithms with exponential execution times. Imagine a table (*left*) whose first row lists all the algorithms that answer either yes or no to input problems by executing at most 2^n steps, where n is the size of the input. All the possible inputs to these algorithms are listed in the first column. Both lists are infinite in length. The entries on the inside of the table show the answers the algorithms would give to each input. An inherently complex "diagonal problem" is constructed by taking the answers along the diagonal (*color*) and changing each no to a yes and each yes to a no (*right*). For example, the answer to the diagonal problem for input 3 is yes because the third diagonal entry is no. This ensures that the answers to the diagonal problem do not match the answers to any problem whose algorithm has an execution time of at most 2^n. Hence the diagonal problem can be solved only by an algorithm that consumes at least an exponential amount of time.

Color-Map Problem," by Kenneth Appel and Wolfgang Haken; SCIENTIFIC AMERICAN, October, 1977].

In view of the central role computers played in proving the four-color-map theorem, it is worth considering whether computers could ever completely take over the job of deciding the truth of mathematical propositions. Could there exist an algorithm such that when it was given a statement written in precise mathematical language, it would report eventually whether the statement was true or false? The answer is no. For a powerful formalized language known as the predicate calculus it has been shown that no such algorithm exists. A problem not amenable to algorithmic solution is said to be unsolvable.

Nevertheless, the goal of computers' automatically proving theorems is so alluring that its impossibility in general has not dissuaded mathematicians from trying to attain it in part. One approach for circumventing unsolvability is to limit the kinds of statements given to the computer. The restriction should not be so stringent that only trivial statements remain, but it must be strong enough to eliminate the blight of unsolvability. Logicians have investigated several such restrictions. If the computer is limited to considering statements of only a certain kind, then with the aid of an algorithm it can decide correctly the truth value of any of these statements. In other words, the decision problem for these restricted statements is solvable.

It is time to examine a particular class of statements, the second-order language of one successor ($S1S$), that is accessible to algorithmic solution. This language deals with the natural numbers (0, 1, 2, 3 . . .). The simplest kind of statement in $S1S$ makes assertions about the relative magnitudes of numbers or about the sum of numbers. An example of a true statement is "3 is less than 4" and that of a false statement is "$2 + 2 = 5$." Such sentences can be combined by the logical connectives "and," "or" and "implies." The preceding examples can be connected to yield the proposition "'3 is less than 4' or '$2 + 2 = 5$.'" Statements that deal with sets of numbers, such as "The numbers in the set (0, 5, 7) are less than the numbers in the set (9, 12, 15, 19, 24, 100)," are also permissible.

The real power of $S1S$ lies in its ability to express mathematical generalities through the use of alphabetic variables that stand for numbers, such as x, y and z, and of quantifiers, such as "all," "every" and "there exists." For example, $S1S$ includes the sentences "Every nonempty set of numbers has a smallest number" and "For every number x there is a number y such that $y = x + 1$." In order for the truth value of these statements to be determined by an algorithm, $S1S$ must be limited to statements that do not include the sum of two variables.

In 1960 J. R. Büchi of the University

VARIABLES		CONNECTIVES	QUANTIFIERS	MEMBERSHIP
$a,b,c,\ldots,z$	stand for numbers	$\wedge$ = and	$\forall$ = for all	$x \epsilon S$ = x is in the set S
$A,B,C,\ldots,Z$	stand for sets of numbers	$\vee$ = or	$\exists$ = there exists	$x \notin S$ = x is not in the set S
		$\rightarrow$ = implies		
		$\leftrightarrow$ = logically equivalent		

STATEMENT 1 = $(\forall B)[(\exists x)(x \epsilon B) \rightarrow (\exists s)(s \epsilon B \wedge (\forall y)(y \epsilon B \rightarrow s \leq y))]$

STATEMENT 2 = $(\exists D)(\exists E)[0 \notin D \wedge 1 \notin E \wedge (\forall b)(b \notin D \rightarrow b+2 \notin D) \wedge (\forall b)(b \notin E \rightarrow b+2 \notin E) \wedge (\exists t)(t \epsilon D \wedge t \epsilon E)]$

SECOND-ORDER LANGUAGE OF ONE SUCCESSOR ($S1S$) is a mathematical language for expressing precisely statements about numbers and sets of numbers. The symbols in $S1S$ for variables, connectives, quantifiers and membership in sets are listed here. Statement 1 is the true assertion that "every nonempty set B contains a smallest member," and statement 2 is the false assertion that "there exist sets D and E and a number t such that D consists only of odd numbers, E consists only of even numbers and t belongs to both D and E." The decision problem for $S1S$ is to determine whether an arbitrary statement is true or false. It has been shown that the decision problem can be solved by algorithm, but any algorithm will consume a superexponential amount of time to decide the truth of some statements.

Yesterday you could admire all-band digital tuning in a short wave receiver.* Today you can afford it.

RF—4900

Tune in the Panasonic Command Series™ top-of-the-line RF-4900. Everything you want in short wave at a surprisingly affordable price. Like fluorescent all-band readout with a five-digit frequency display. It's so accurate (within 1 kHz, to be exact), you can tune in a station even before it's broadcasting. And with the RF-4900's eight short wave bands, you can choose any broadcast between 1.6 and 31 MHz. That's all short wave bands. That's Panasonic.

RF-2900

And what you see on the outside is just a small part of what Panasonic gives you inside. There's a double superheterodyne system for sharp reception stability and selectivity as well as image rejection. An input-tuned RF amplifier with a 3-ganged variable tuning capacitor for excellent sensitivity and frequency linearity. Ladder-type ceramic filters to reduce frequency interference. And even an antenna trimmer that changes the front-end capacitance for reception of weak broadcast signals.

To help you control all that sophisticated circuitry, Panasonic's RF-4900 gives you all these sophisticated controls. Like an all-gear-drive tuning control to prevent "backlash." Separate wide/narrow bandwidth selectors for crisp reception even in crowded conditions. Adjustable calibration for easy tuning to exact frequencies. A BFO pitch control. RF-gain control for improved reception in strong signal areas. An ANL switch. Even separate bass and treble controls.

And if all that short wave isn't enough. There's more. Like SSB (single sideband) amateur radio. All 40 CB channels. Ship to shore. Even Morse communications. AC/DC operation. And with Panasonic's 4" full-range speaker, the big sound of AM and FM will really sound big. There's also the Panasonic RF-2900. It has most of the features of the RF-4900, but it costs a lot less.

The Command Series from Panasonic. If you had short wave receivers as good. You wouldn't still be reading. You'd be listening.

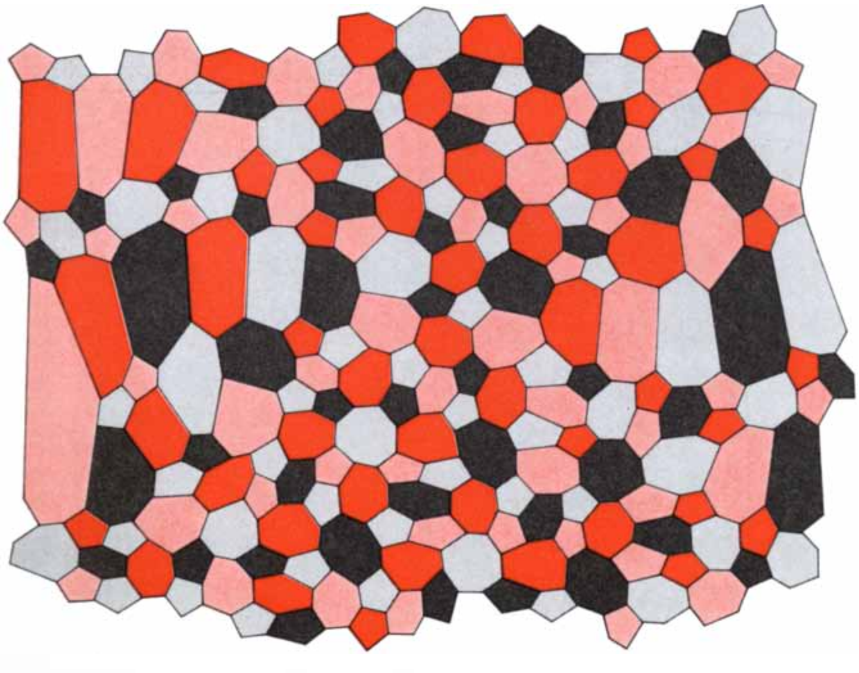
*Short wave reception will vary with antenna, weather conditions, operator's geographic location and other factors. An outside antenna may be required for maximum short wave reception.

Panasonic®
just slightly ahead of our time.

of Michigan proved there is an algorithm that can determine correctly the truth value of any given statement in *S1S*. That *S1S* is solvable is somewhat surprising because of the open-ended nature of the quantifiers. For example, to determine the truth value of the statement "For all x, 2 is less than $x + 3$," an algorithm could not try substituting each natural number for x because there are infinitely many of them. Instead the algorithm must resort to some other kind of reasoning, perhaps to algebraic manipulation or to mathematical induction. Büchi has in effect shown that an algorithm can perform such reasoning in order to ascertain the truth value of any statement in *S1S* regardless of the statement's length.

CLASS OF STATEMENTS		EXECUTION TIME	
STATEMENTS INVOLVING NUMERIC VARIABLES	ARE SETS ALLOWED?	AT LEAST	AT MOST
VARIABLE + NUMERAL	YES	$\left.2^{2^{\cdot^{\cdot^{\cdot^{2^{n}}}}}}\right\\} cn$	$\left.2^{2^{\cdot^{\cdot^{\cdot^{2^{n}}}}}}\right\\} dn$
VARIABLE + VARIABLE	NO	$2^{2^{cn}}$	$2^{2^{2^{dn}}}$
VARIABLE × VARIABLE	NO	$2^{2^{2^{cn}}}$	$2^{2^{2^{2^{dn}}}}$

INHERENT COMPLEXITIES of decision problems in logic have been classified for many cases. The first two columns in this table describe classes of statements, and the last two columns give the known lower and upper bounds on the execution time of algorithms that solve the decision problem for each class. The time is expressed as a function of the statement length *n*; *c* and *d* are constants greater than zero. The statements in the first row correspond to those in *S1S*. The statements in the second row constitute a language called Presburger arithmetic.



MAP OF 210 COUNTRIES illustrates the four-color-map theorem, which states that it is possible to color every country on any flat map one of four colors without bordering countries having the same color. The theorem was put forward in 1852 but was not proved until 1976 (by Kenneth Appel and Wolfgang Haken of the University of Illinois). The proof was unusually long and relied on the quite extensive aid of high-speed computers, which spent 1,200 hours of computer time in effect coloring each one of some 2,000 maps in as many as 200,000 ways.

Our proof that the decision problem for roadblock is amenable to algorithmic solution says nothing about the efficiency of the algorithm, which has turned out to be inherently exponential. Similarly, Büchi's proof of the existence of an algorithm that solves the decision problem for *S1S* promises nothing about the algorithm's efficiency. In fact, in 1972 Meyer proved that the inherent complexity of the decision problem for *S1S* grows even faster than exponentially. The double-exponential function 2^{2^n} is an example of a function that exhibits superexponential growth. To evaluate the double-exponential function at a number n, the number is exponentiated to yield $m = 2^n$, which in turn is exponentiated to yield the final result 2^m. By the same token a triple-exponential function involves three successive exponentiations, a quadruple-exponential function involves four successive exponentiations and in general for any number k the k-fold-exponential function involves k successive exponentiations. It is difficult to imagine just how rapidly k-fold-exponential functions grow when k is greater than 2, because no physical process is known to exhibit such explosive growth.

Meyer's remarkable theorem states that no matter how large the number k of successive exponentiations is, the inherent complexity of deciding the truth value of statements in *S1S* grows faster than a k-fold-exponential function of the statement length n. Although Büchi's theorem asserts that the decision problem for *S1S* is solvable in principle, Meyer's theorem has much the same effect as an assertion of unsolvability. Calling a problem unsolvable means that any algorithm will fail to solve the problem for infinitely many inputs. Large inherent complexity means that any algorithm needs an intolerably long time to solve the problem for infinitely many inputs. In either case the practical effect is the same: the mathematician who asks the computer to apply the algorithm to one of these inputs will have to wait too long (either forever or practically forever) for the solution.

S1S is not the only language whose decision problem is solvable in theory but not in practice. Consider the set of statements, called Presburger arithmetic, that consists of all the statements in *S1S* with the exception of those mentioning sets of numbers and with the addition of those dealing with sums of numeric variables, such as "$x + y = z$" and "$x + z = 30$." In 1973 Rabin and Michael J. Fischer of M.I.T. demonstrated that the inherent complexity of the decision problem for Presburger arithmetic

Astronomy Book Club invites you to take

ANY 3 BOOKS FOR ONLY $1.00 EACH

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

(values to $69.35)

73939. THE RESTLESS UNIVERSE/WHITNEY'S STAR FINDER. *The 2 count as one book.* **$21.90**

49240. FROM VINLAND TO MARS. *Richard S. Lewis.* A sweeping history of a thousand years of exploration. **$15.00**

36395. BLACK HOLES, QUASARS AND THE UNIVERSE. *Henry L. Shipman.* An absorbing exploration of the frontiers of astronomy. **$15.50**

55000. THE ILLUSTRATED ENCYCLOPEDIA OF ASTRONOMY AND SPACE. $16.95

64157-2. THE RAND-McNALLY NEW CONCISE ATLAS OF THE UNIVERSE. *Patrick Moore.* Nearly 1,000 illustrations with 400 in spectacular color. 11" x 14". *Counts as 2 of your 3 books.* **$29.95**

52850. THE HIGH FRONTIER: Human Colonies in Space. *Gerard K. O'Neill.* Convincing look at the plausibility of space colonies in the near future. **$8.95**

40685. THE COSMIC FRONTIERS OF GENERAL RELATIVITY. *William J. Kaufmann III.* Nontechnical introduction to post-Einsteinian physics. **$12.95**

80250. STARS: Their Birth, Life, and Death. *Josif S. Shklovskii.* **$17.50**

63217. MURMURS OF THE EARTH. *Carl Sagan et al.* The story of the recorded messages carried by the *Voyager* space probe, man's first attempt at interstellar communication. **$15.00**

79735. SPACESHIPS OF THE MIND. *Nigel Calder.* Scientifically plausible proposals for man's self-directed evolution, his colonization of space, and the restoration of the earth to a more natural state. **$14.95**

35740. THE BEGINNER'S GUIDE TO THE SKIES. *Clarence H. Cleminshaw.* Illustrated reference covering planet positions, constellations, paths of the sun, moon and stars. **$7.95**

Basic Astronomy Library

Here's a carefully selected set of five books that will provide you with a basic, well-rounded astronomy library. First, Henry L. Shipman's outstanding **THE RESTLESS UNIVERSE**—a more than 450-page tour of the ideas, principles, and processes of modern astronomy, containing more than 250 illustrations and tables. Then, a marvelously readable overview of the latest advances in the "new astronomy," Shipman's **BLACK HOLES, QUASARS, AND THE UNIVERSE.** Next, Richard S. Lewis's sweeping survey of man's place in space, **FROM VINLAND TO MARS.** Then, an unequaled field guide to the heavens, **WHITNEY'S STAR FINDER.** Finally, a just-published reference that's going to become a "must" for any astronomy reader, the **ILLUSTRATED ENCYCLOPEDIA OF ASTRONOMY AND SPACE.** Publishers' price for 5-book set: $69.35. *Counts as 3 of your 3 books.* To order check **BASIC ASTRONOMY LIBRARY** on reply card.

34678. ASTRONOMY DATA BOOK. *Robinson and Muirden.* Superb single-source guide to technical information: list of constellations, laws of planetary motion, magnitude, telescope types, much more. Well illustrated with over 50 pages of tables and charts. **$12.95**

(Publishers' Prices shown)

80585. STELLAR EVOLUTION. *A. J. Meadows.* The modern theories of the origin and evolution of stars explained in understandable language. **$16.50**

45050. THE ENDS OF THE EARTH. *Isaac Asimov.* Everything about the frozen polar wastes, from their exciting discovery to their effect on world climate. **$15.00**

69303. PLANETARY ENCOUNTERS. *Robert M. Powers.* The case for all-out exploration of our solar system in this century. **$13.95**

82735. TELESCOPES FOR SKYGAZING. *Henry E. Paul.* Classic reference to buying, building, or using telescopes. Packed with photos, diagrams, and charts. **$10.95**

59690. MAN AND THE STARS. *Hanbury Brown.* Complete and thoroughly entertaining history of man's constant fascination with the secrets of the heavens. **$14.95**

79725. SPACE TREK: The Endless Migration. *Glenn and Robinson.* Presents the economic, political, and philosophical effects of space colonization. **$9.95**

53015. THE HISTORICAL SUPERNOVAE. *Clark and Stephenson.* First detailed study of supernova remnants based on historical observations, particularly Far Eastern. **$16.50**

85461. UNTIL THE SUN DIES. *Dr. Robert Jastrow.* Account of the creation of the universe and the evolution of life as uncovered by twentieth-century science. **$8.95**

57833. THE LAROUSSE GUIDE TO ASTRONOMY. *David Baker.* A thorough summary of the universe. Packed with color photos and illustrations. **$15.95**

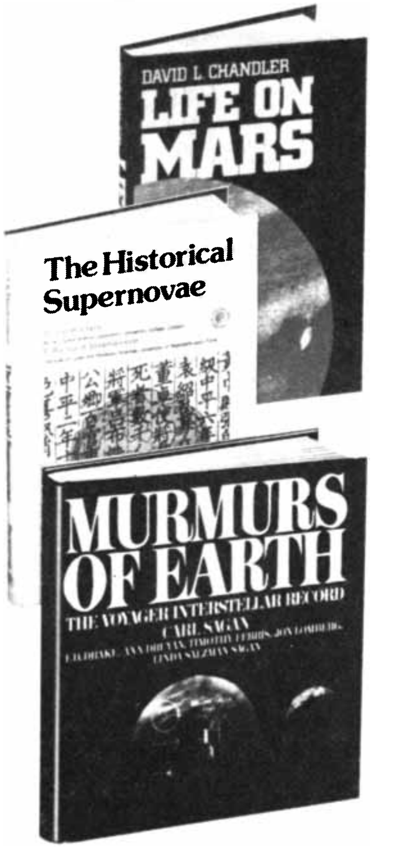
80255. STARS AND NEBULAS. *William J. Kaufmann, III.* Covers everything from radio signals in space to time as the fourth dimension. **$13.00**

73015-2. PULSARS. *Manchester and Taylor.* A comprehensive guide to pulsars. Hundreds of charts, tables, illustrations, and equations. *Counts as 2 of your 3 books.* **$22.50**

53585. A HOUSE IN SPACE. *Henry S. F. Cooper, Jr.* Skylab's unique experiments and observations as well as a behind-the-scenes look at the astronaut's experiences. **$8.95**

59800. MAN DISCOVERS THE GALAXIES. *Richard E. Berendzen, et al.* A study of what is known about the galaxies, and how we came to know it. Over 200 illustrations. **$15.95**

58585. LIFE ON MARS. *David L. Chandler.* Speculates that life may indeed be present. **$9.95**



If the reply card has been removed, please write to **Astronomy Book Club Dept. L-A4D, Riverside, N.J. 08370** to obtain membership information and application.

Now JAL offers full-fare passengers special privileges before, during and after the flight.

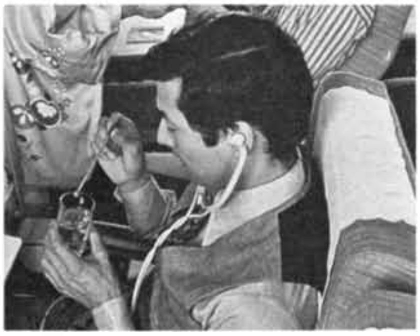
JAL feels that the full-fare passenger deserves better than just a couple of free in-flight gestures. That's why JAL's Executive Service starts working for you the minute you book. Ask your Travel Agent, Corporate Travel Dept. or Japan Air Lines for further details. And find out what it is like to be pampered Japanese style, all the way.

We never forget how important you are.
JAPAN AIR LINES

Bilingual business cards, English on one side, Chinese or Japanese on the other, at a nominal fee.

Special Executive check-in at selected airports.

True Japanese hospitality in the Executive Service section.



Free earphones—for in-flight entertainment.

Free alcoholic beverages—cheers!

1st Class offers the Sky Sleeper Service, a private bed at an additional charge, on all 747 Transpacific flights.

JAL's Executive Hotel Service offers a room at a special rate and allows check-out as late as 6 p.m.

At hotels all over the Orient, JAL's Hospitality Desks give any assistance you may need.

The Executive Service Lounge in the Imperial Hotel, Tokyo, "your office away from the office."

Executive Service

grows at least as fast as a double-exponential function of the length of the statement. In recent years the superexponential complexity of many other decision problems has been discovered. Such discoveries present serious obstacles to the possibility of avoiding unsolvability for decision problems in logic merely by restricting the kinds of statements that can be considered.

Another possible way to evade the blight of unsolvability is based on the assumption that in actual applications people are interested not in solving all instances of a problem but in solving a finite number of them. It is unlikely, for example, that anyone would want to know the truth of a statement more than a million symbols long. The results about inherent complexity we have described so far have dealt with the eventual explosive growth rate of the execution time as a function of the size of the input. We have been fairly specific in stating growth rates (exponential, double-exponential, triple-exponential and so on) but less specific in discussing the size of the inputs for which the growth rates start to explode. It might be suspected that if the inherent complexity of a problem grew at least exponentially, the problem would remain too complex even if it were limited to inputs of only several hundred symbols in length. In the case of the decision problem for *S*1*S* this suspicion has been confirmed by mathematical proof.

Consider the statements in *S*1*S* that consist of no more than 675 symbols. What is the complexity of determining the truth value of any statement in this finite class of statements? Before we can answer the question it is necessary to clarify the concept of complexity for problems involving only a finite number of instances. Consider the problem of finding the sine of an angle to five decimal places when the angle ranges from 0 to 90 degrees in increments of one minute of arc. Before the proliferation of inexpensive hand-held calculators the most popular method for solving this problem was to look up the answer in a table. Looking in a table is much faster than calculating the sine from a formula with pencil and paper. The increase in speed was achieved at the price of an increase in the size of the algorithm: the table of sines takes up much more space than the formula.

Once a table is constructed for a finite problem by finding and recording the answer for each possible input, the method of solving the problem by looking in the table qualifies as an algorithm, since it can be mechanically applied for every instance of the problem. The catch is that the number of instances of a finite problem can be so large as to be unwieldy. In *S*1*S* there are considerably more than 10^{100} statements of length 675. It is obvious that constructing a table with an entry for each statement is out of the question. The point we want to make is that in addressing the complexity of finite problems it is necessary to take into account the size of an algorithm as well as the length of its execution time.

One conceptually simple way to measure the size of an algorithm is to count the number of components in a computer that is capable of executing the algorithm. It is assumed the computer is wired to solve only that problem, so that there is no excess circuitry. In real computers the basic unit of electronic logic is a "gate," or switch, that takes two electric signals, each of which is either on or off, and sends out a third signal, which is also either on or off. Such gates are the simplest of electronic components, although thousands of them can be connected to form quite complex systems. When one of the signals that enter a gate changes from off to on or from on to off, there is a slight delay before the output signal reaches a stable value. The time it takes a computer to execute an algorithm is just this switching delay multiplied by the number of times the gates must switch in order to execute the algorithm.

The largest high-speed computers have roughly half a million gates with switching delays of about 10^{-9} second. Of course, larger and faster computers will inevitably be built, but none will ever exceed the capacity of the largest imaginable computer we discussed above. This ideal computer would consist of at most 10^{126} gates no smaller than a proton. It is reasonable to assume that the switching delay of a gate is not less than the 3×10^{-24} second it takes light to traverse the diameter of the proton. Meyer and one of us (Stockmeyer) proved that this ideal computer would take at least 20 billion years to determine the truth value of certain *S*1*S* statements consisting of 675 symbols. This means there are statements of length 675 whose truth value can never be determined by an algorithm.

It is possible that these inscrutable statements are few and far between. Perhaps for practical purposes it is unnecessary to hope for an algorithm that can handle either any given statement or any given statement up to a certain length. In the case of *S*1*S* one could imagine an algorithm that would give one of three answers: "true," "false" or "don't know." The possibility of a "don't know" response would not seriously impair the usefulness of such an algorithm if it were issued only rarely. Although this three-valued algorithm seems to neatly finesse the intractability of *S*1*S*; there is no guarantee that such an algorithm would be efficient. It is an open question whether or not the possibility of an occasional indeterminate response would make the decision problem for *S*1*S* any easier.

Giotto's Portrait of Halley's Comet

A blazing comet represents the star of Bethlehem in one of Giotto's famous Arena Chapel frescoes in Padua. It is a naturalistic portrait of Halley's Comet as seen during its spectacular apparition of 1301

by Roberta J. M. Olson

In the late summer of 1301 a spectacular comet splashed across the dome of the night sky. Its luminous spherical coma surrounded a starlike central incandescence that seemed to spew a diffused flare of radiant particles, like a mane of hair, directly away from the sun. We know now that it was Halley's Comet, making one of its periodic visits to the center of the solar system. And we know how it looked to amazed observers in Italy because its "portrait" was rendered in faithful detail by Giotto di Bondone (1267–1337), the Florentine pioneer of naturalistic painting.

Giotto was in Italy in 1301 (although exactly where is not certain), and he surely saw the comet. Within perhaps a year and no more than four years he executed the remarkable series of frescoes on the walls of the Scrovegni (Arena) Chapel in Padua, in northern Italy. In a scene depicting the Adoration of the Magi he represented the star of Bethlehem, in a remarkable departure from iconographic tradition, not as a stylized, many-pointed little star but as a blazing comet. The coincidence of dates, the naturalistic representation of the comet and its similarity to photographs of the comet made on its most recent appearance, in 1910, constitute strong evidence that Giotto's comet is indeed Halley's Comet.

There are billions of comets (the word is from the Greek *kometes,* "long-haired") orbiting the sun in a thinly distributed cloud that extends up to thousands of times as far as the outermost planets. Each comet nucleus is a small, icy aggregate of frozen gases and interstellar dust. Occasionally one of these "dirty snowballs," its orbit perhaps having been perturbed by a star, enters the inner regions of the solar system, where its gases and dust, activated by solar radiation and the solar wind, form a visible coma, or head and tail [see "The Nature of Comets," by Fred L. Whipple; SCIENTIFIC AMERICAN, February, 1974]. Some of these comets attain a new elliptical orbit that periodically brings them within sight of the earth.

Halley's Comet was the first comet to be recognized as being periodic. In 1705 the English astronomer Edmund Halley noted striking similarities in the orbits he had calculated, based on earlier observations, for major comet appearances in 1682, 1607 and 1531. He postulated that the three apparitions represented return visits, approximately 76 years apart, of a single comet describing an elongated elliptical orbit around the sun. He predicted that it would return in 1758. Later astronomers, allowing for perturbations of the orbit by Jupiter and other planets, refined Halley's calculations to predict that the next perihelion passage (the closest approach to the sun) of Halley's Comet would be in April of 1759. The comet was first sighted by a German peasant on Christmas Day in 1758, and it reached perihelion in March of 1759.

Subsequent calculations have established the date of every perihelion passage back to 239 B.C.; examination of old records (largely those of Chinese observers for passages before the 15th century) reveals observations of a comet at the right time and in the right part of the sky to match most of the calculated returns to perihelion. The mean period of the comet is just under 77 years, but the precise period varies as much as two and a half years on each side of the mean because of perturbations by the planets. Halley's Comet will return to view during 1985, with perihelion predicted for February 9, 1986.

It is estimated that in every century perhaps 20 or 30 comets are visible from the earth. In the 13th century, two centuries before the invention of the telescope, more than 20 unaided-eye sightings of "comets" were recorded. Most of the "comets" mentioned by Western chroniclers are referred to only once, however, and without precision; some of them were certainly meteors or other celestial phenomena. The only reliable comet apparitions are those that have been verified by careful correlation of several observations, including sightings recorded in Chinese annals brought to the West by returning Jesuit missionaries in the 18th century and first translated in 1846. Presumably these were the most spectacular apparitions of their time, since they drew enough attention to give rise to multiple references in Europe and in Asia. A meteor might be recorded as a comet by one person, but a genuine comet would not be likely to confuse several witnesses. A spectacular apparition would astonish all observers.

Of the comets with calculated orbits securely recorded for the 13th century only two have been established as spectacular apparitions. The first was seen over both China and Europe in 1264. The second was definitely seen over China in 1299 (and a treatise attributed to one Peter of Limoges records such a comet over Europe). Giotto was born in about 1267, and so he could not have observed the comet of 1264. The apparition of 1299 probably was not seen over Italy or was unremarkable there. The first securely established spectacular apparition within Giotto's lifetime was therefore that of Halley's Comet in 1301. The next spectacular comet apparition was not until 1337, the year of Giotto's death. Moreover, the apparition of 1301 was an impressive one, commented on by many contemporary historians. The eminent 14th-century Florentine chronicler Giovanni Villani wrote in his *Chroniche Storiche* that a comet appeared in the heavens in September of that year "with great trails of fumes behind" and remained visible until January of 1302. There are various discrepancies in the dating, with some writers reporting that the apparition remained for only six weeks, but all agreed that the comet's tail was impressively long; it is estimated to have subtended an angle of as much as 70 degrees in the sky. Such an apparition accords well with Giotto's representation.

Throughout recorded history the appearance of a comet has been an occasion for fear. For millenniums it was believed that all aspects of life on the earth were ruled by the positions of the stars and planets. Once the locations of the fixed stars and the changing positions of the planets were known the appearance of a "new star" of any kind seemed to violate the order of the heavens and hence to be prescient of disaster (literally a "bad star"). Such appearances were often associated with significant human events, yet before the 16th century comets were only rarely represented in Western art. (One example is a sword-shaped comet shown hovering in the sky on a Roman cameo.) In medieval times comets are sometimes seen in subsidiary positions in large astrological works, such as the 12th-century zodiacal cycle on the west facade of the Cathedral of Piacenza in central Italy. Such representations of comets are generic, however; "portraits" of identifiable comets—explicit representations of historical apparitions—are even rarer.

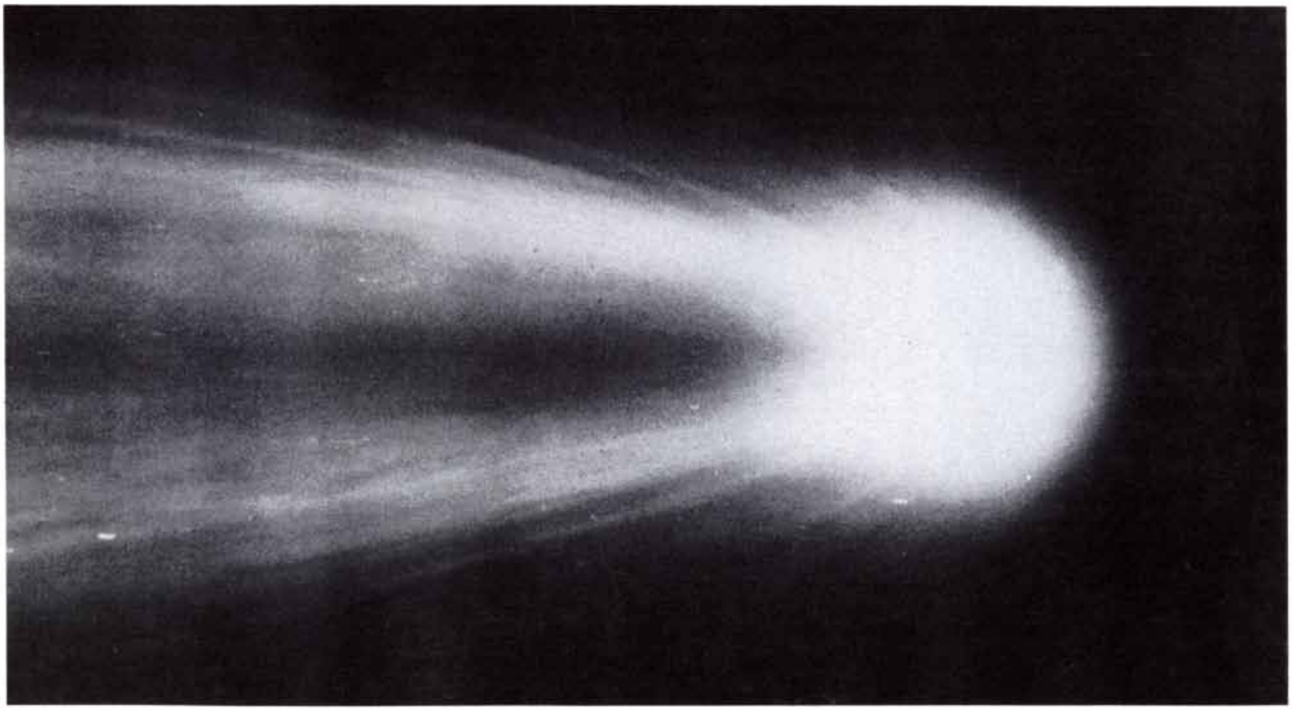
Three such "portraits" of apparitions of Halley's Comet before Giotto's time have been identified. The earliest one represents the apparition of A.D. 684, although it was not executed until eight centuries after the fact, in the *Liber Chronicorum,* or *Weltchronik,* of Hartmann Schedel, known in English as the *Nuremberg Chronicles* because it was printed in Nuremberg. The book was first issued in 1493 and was illustrated with woodcuts by the German artist Michel Wohlgemuth and his stepson Wilhelm Pleydenwurff. The printer was Anton Koberger, the godfather of Albrecht Dürer (who was later to portray a spec-

THE ADORATION OF THE MAGI is one scene in a fresco cycle executed by the Florentine master Giotto di Bondone. The cycle decorates the interior of the Arena Chapel, which was commissioned by the Paduan businessman Enrico Scrovegni (perhaps to expiate the sin of his father, who was identified by Dante in the *Inferno* as the archusurer). Scrovegni obtained permission to erect the building in 1302, the site was dedicated in 1303 and the frescoes appear to have been begun that year; the *Adoration,* on the second tier of the many-tiered cycle, was probably completed in about 1304. The scene exemplifies Giotto's major innovations, his naturalism and the humanity of his figures, and is notable for its representation of the star of Bethlehem not as a stylized star but as a dynamic comet. Halley's Comet, which returns to center of solar system about every 77 years, made an appearance in 1301. It served as the model for Giotto's comet.

CLOSE-UP OF COMET in the *Adoration* shows how the artist applied tempera and gold pigments to the plastered wall in textured strokes approximating the luminescent appearance of the coma and tail of the actual comet, which he must have observed carefully a few years earlier. He depicted the intensely bright center of the coma, or head, with what appears to be an eight-pointed star, building up layers of pigment over the star to diffuse the image. Some pigment is lost, revealing the red adhesive by which it was applied to the plaster.

tacular comet apparition in his engraving *Melencolia I*). A crude representation of Halley's Comet appears with accompanying text on the page dealing with the year 684. (It is repeated haphazardly throughout the book because the limited repertory of small woodblock designs functioned not only as specific illustrations but also as "finger posts" for the reader searching for his place in the large unpaged volume.)

The earliest contemporaneous, albeit stylized, portrait of Halley's Comet depicts the apparition in the spring of 1066. It is found in a scene of the Bayeux Tapestry, which was commissioned by Queen Matilda, the wife of William the Conqueror, to illustrate her husband's victory at the Battle of Hastings. The tapestry (actually a crewel embroidery on eight narrow strips of coarse linen,



HEAD OF HALLEY'S COMET was photographed with the 60-inch reflecting telescope on Mount Wilson on May 8, 1910, during the comet's most recent return to the vicinity of the earth. Major comets have distinctive, although somewhat variable, configurations, and this photograph is recognizably similar to Giotto's depiction. The comet proper is a small (and hence invisible) nucleus of frozen gases and dust. When the comet comes near enough to the sun, some of the gas is evaporated, carrying dust particles with it; the glow of the comet's coma is from sunlight that is scattered by the dust or reradiated by fluorescing gases. The tail, which always points away from the sun, is a stream of dust impelled by solar radiation pressure and of gas molecules ionized by solar radiation and by electrons in the solar wind and impelled by magnetic fields within the solar wind. Halley's Comet will return to perihelion in 1986 (*see illustration on page 170*).

measuring almost 231 feet long by 19½ inches wide) was executed between 1073 and 1083 and now hangs in the town hall of Bayeux in Normandy. Numerous accounts by contemporary chroniclers date the 1066 apparition and remark on its spectacular nature, which is attested to in the tapestry by the amazed expressions on the faces of its beholders and by the embroidered legend: "They are in awe of the star." The comet itself is represented, in the prevailing Romanesque style of the period, as a highly stylized geometric structure, a flat and purely decorative configuration of lines and planes.

The next portrait of Halley's Comet is found in the *Eadwine* (*Canterbury*) *Psalter,* a 12th-century English manuscript that is a copy of the ninth-century *Utrecht Psalter,* a collection of the Psalms. At the bottom of the page giving the text of Psalm 5 the comet is seen in bare outline: a stylized rosette inscribed in a circle, with a tail of four wavy rays. The position of the sketch at the bottom of the page and its large scale suggest that it was added, along with its Old English textual description, as a passing notation; it appears to have no connection with the three Latin versions of the Psalm on the page. The style of the manuscript and the dates of its scribe, the monk Eadwine, correspond with the 1145 apparition of Halley's Comet. In spite of its sketchiness this drawing is revolutionary for the 12th century in that it is a contemporary representation of a natural phenomenon.

The most remarkable aspect of Giotto's portrait of Halley's Comet, in sharp contrast to the schematic nature of the earlier representations, is its naturalism. This is fitting, because Giotto's contemporary reputation and his immense significance in the history of painting (he was the first artist to exert an almost universal influence on Western painting) proceed from his startlingly naturalistic innovations. He was born in the small village of Colle di Vespignano, near Florence. Little is known of his early life, but he is thought to have been a pupil of the Florentine painter Cimabue, a progressive master who began to break with the prevailing Italo-Byzantine style. Giotto rebelled more dramatically against the established style and developed a new three-dimensional plasticity that must have seemed even more overwhelming in his day than it does to connoisseurs today. His majestic sculptural figures, calmly positioned to form simple masses of glowing color, communicate an emotional intensity unprecedented in European painting. And it was in the chapel frescoes in Padua that Giotto's revolutionary style achieved not only maturity but also its most lucid statement.

Giotto had worked in Rome at about the turn of the century and is thought to

THE SMALL CAR JUST GREW UP.

Do you want to own a small car with Oldsmobile looks, comfort and room? You're gonna like our new Omega a lot. A whole lot.

TRANSVERSE ENGINE.

Omega's inside size will startle you. We got the room inside by turning the engine sideways and moving it forward. Your legs will appreciate the difference.

Omega's engine is a transverse-mounted 4-cylinder, or you have the option of the world's first transverse V6. Both are GM-built by various divisions, and your Olds dealer has details on sources and availability.

There are other nice surprises.

FRONT-WHEEL DRIVE.

Whichever engine you choose sits above the drive wheels, which is what gives Omega the impressive traction of front-wheel drive.

RACK-AND-PINION STEERING.

Omega is quick and agile, with rack-and-pinion steering.

If the name MacPherson doesn't mean anything to you in strut suspension, Omega's ride will. It is smooth.

GOOD GAS MILEAGE.

With the transverse 4-cylinder, Omega's EPA estimate is 24 mpg. The highway estimate is 38 mpg. Very good indeed, for a car this size.

Remember, the circled EPA estimate is for comparison to other cars. Your mileage depends on speed, weather and trip length; your actual highway mileage will probably be lower than the highway estimate. California estimates are lower.

Are you starting to believe this is an amazing small car? You should.

The 1980 Omega is the Oldsmobile of small cars.

You can buy it or lease it. But you gotta drive it.

You're gonna love it.

The small car just grew up.

Oldsmobile

Have one built for you.

THE 1980 OMEGA
THE OLDSMOBILE OF SMALL CARS.

GM

have been in Padua by 1302, working on a fresco that has since been lost. A wealthy Paduan merchant, Enrico Scrovegni, commissioned him to decorate the interior of a small family chapel being built adjacent to the family palace and near the ruins of a Roman amphitheatre (whence the familiar name Arena Chapel). It has been suggested that Giotto was the architect of the chapel, in which case the conceptual date of the frescoes would be even closer to the apparition of the comet; in any case they appear to have been started in 1303 at the latest. Giotto covered the walls with tiered rows of scenes illustrating the life of Christ, from the events before his birth (the lives of St. Joachim and St. Anne, parents of the Virgin Mary, and the life of the Virgin) to the infancy of Christ, his adult life, his passion, crucifixion and resurrection and the Last Judgment, all within a larger symbolic program. Most of the 38 narrative scenes, including the *Adoration*, measure about 200 by 185 centimeters, with figures about half life size. Because of its position in the second tier the *Adoration* can be dated to 1303 or 1304.

When Giotto came to paint the star of Bethlehem, he rejected the strictures of both astrological symbolism and medieval convention and rendered the comet as he had actually seen it a few years before, theatrically illuminating the Italian night sky. The large, fiery comet dominates the sky of his fresco. The coma pulses with energy; in its center is a hard-edged star shape representing the bright "center of condensation" that is often seen within the more diffuse coma. (It is not the nucleus, which is too small to be visible within the coma.) The striated tail imparts a dynamic impression of the arc traced by the comet's passage across the sky. Spectacular comets do look like this to the unaided eye, and Halley's Comet would have looked like this to Giotto. One conjectures that he might have recorded his observations of 1301 in a drawing to which he referred later (although no drawings made by him are extant). Whether that was in fact the case or whether he simply recalled the apparition vividly, his naturalistic comet harmonizes perfectly with the naturalistic aesthetic that characterizes the *Adoration* and the entire cycle of frescoes.

What could have motivated Giotto to represent the star of Bethlehem as a comet? No one knows, of course, whether the biblical star was a comet, a nova or a conjunction of planets, or whether it was apocryphal. Matthew is the only Evangelist who mentions a star leading the Magi to the Nativity, and he gives no detail: "For we have seen his star in the east, and are come to worship him.... And, lo, the star, which they saw in the east, went before them, till it came and stood over where the young child was. When they saw the star, they rejoiced with exceeding great joy" (Matt. 2:2; 2:9–10). The star is also mentioned in the apocryphal *Protoevangelium of James,* but again without characterization. For centuries, however, the star of Bethlehem was depicted in Nativity and Adoration scenes. Almost always it was shown as a small, stylized and quite imaginary star, often with rays of light shining down on the Infant Jesus, signifying God's blessing on the birth.

Renaissance painters depended on well-established rules of iconography. Style and technique might vary, but the choice of what to put in a painting, even where to place the figures and how to pose them, was largely established by tradition (or innovative interpretations of tradition), which in turn shaped current expectations. Giotto was therefore not likely to have substituted a comet for the conventional star on mere impulse, even in response to a vivid personal observation.

Certainly there was a lengthy literary tradition concerning comets. Long before Giotto such writers as Aristotle, Virgil, Seneca and the Roman poet Lucan, among others, had speculated about comet apparitions. In general comets were regarded with apprehension and interpreted as portents of catastrophe, plagues or the death of kings. Less often they were considered positive signs—of victory, bounty or the birth of kings. The interpretation varied with circumstance. The apparition of Halley's Comet in 1066 could have been taken by the English as a portent either of their King Harold's victory over Harold III of Norway in late September of that year or of his defeat the following month by William of Normandy at Hastings; the Normans, the eventual victors in the three-way conflict, obviously came to regard the comet as being prescient of William's triumph. Comets that were regarded as positive omens were generally believed to have been of divine rather than natural origin—created by God for a particular benevolent purpose. This view was echoed in Giotto's own time by Aegidius of Lessines, in his treatise of 1264 *On the Essence, Motion and Signification of Comets.*

At the time of the writing of the Gospels at the end of the first century it was common to associate the appearance of a "new star" with the birth of a king. Much later the writings of the church fathers, which would have been familiar to Giotto's contemporaries, expanded on Matthew's account of the star of Bethlehem. Two innovative Christian thinkers, Origen in the third century and John of Damascus in the seventh and eighth centuries, remarked on the cometlike nature of the star described by the Evangelist.

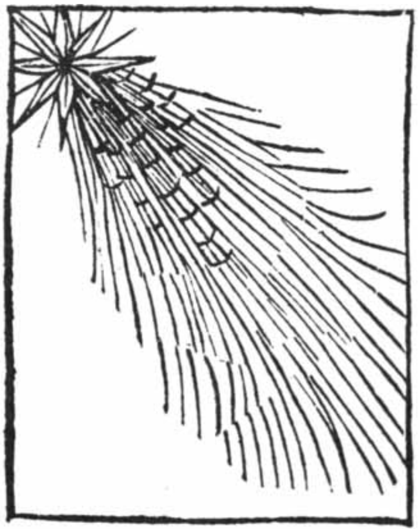
In his treatise *Against Celsus* Origen wrote: "The star that was seen in the

RETURNS TO PERIHELION OF HALLEY'S COMET

239 B.C.	MARCH 30
163	OCTOBER 5
86	AUGUST 2
11	OCTOBER 5
A.D. 66	JANUARY 26
141	MARCH 20
218	MAY 17
295	APRIL 20
374	FEBRUARY 16
451	JUNE 24
530	SEPTEMBER 25
607	MARCH 13
684	SEPTEMBER 28
760	MAY 22
837	FEBRUARY 27
912	JULY 9
989	SEPTEMBER 9
1066	MARCH 23
1145	APRIL 22
1222	OCTOBER 1
1301	OCTOBER 23
1378	NOVEMBER 9
1456	JUNE 9
1531	AUGUST 25
1607	OCTOBER 27
1682	SEPTEMBER 15
1759	MARCH 13
1835	NOVEMBER 16
1910	APRIL 20

RETURNS TO PERIHELION, or closest approach to the sun, have been calculated back to 239 B.C. and most of the dates have been correlated with records of observations. The average period of the comet's long elliptical orbit is almost 77 years, but successive periods vary as the orbit is perturbed by the gravitational effect of planets, Jupiter in particular.

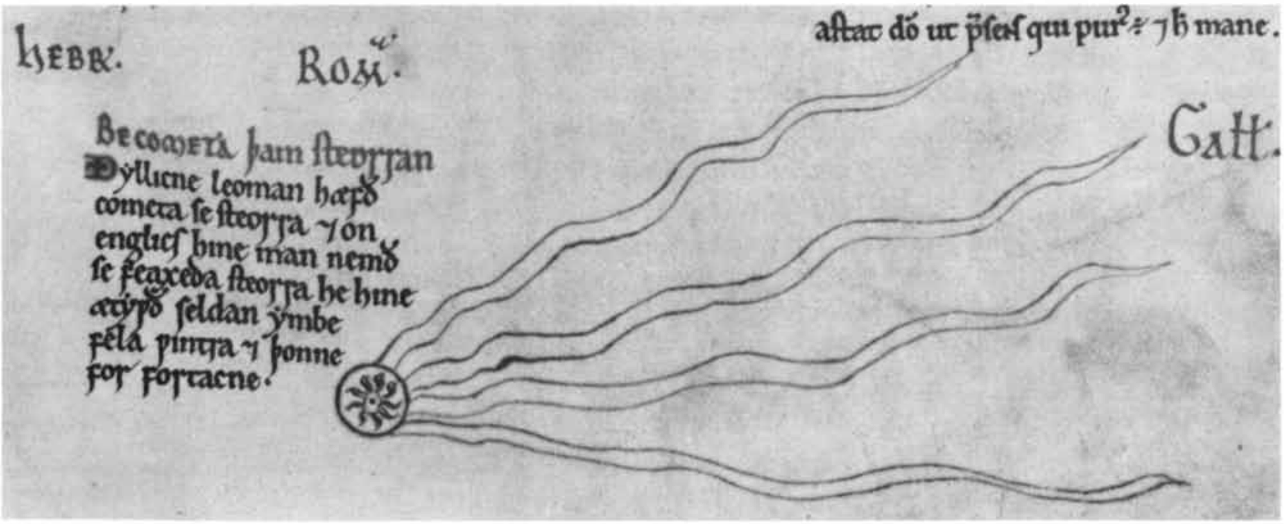
Cometa



NUREMBERG CHRONICLES, published in 1493, contains this woodblock illustration of a comet on the page recounting the events of A.D. 684, a year in which Halley's Comet appeared. The accompanying text tells of calamitous events brought about by the comet: three months of rain, thunder and lightning, during which people and flocks died, grain withered in the fields and an eclipse of the sun and moon was followed by a plague.

east we consider to have been a new star . . . such as comets, or those meteors [comets and meteors were not then differentiated] which resemble beams of wood, or beards or wine jars." And he pointed out that "we have read in the *Treatise on Comets* by Chaeremon the Stoic, that on some occasions also, when *good* was to happen comets made their appearance. . . . If then, at the commencement of new dynasties . . . there arises a comet . . . , why should it be a matter of wonder that at the birth of Him who was to introduce a new doctrine to the human race . . . a star should have arisen?" Origen acknowledged that there was no specific early prophecy that a particular comet would "arise in connection with a particular kingdom or a particular time, but with respect to the appearance of a star at the birth of Jesus

BAYEUX TAPESTRY, which commemorates the events of 1066, records the apparition of Halley's Comet in the spring of that year. At the left a crowd of Englishmen point to the stylized comet in the upper border; the legend above them reads: "They are in awe of the star." At right King Harold II of England, on being told of bad omen, imagines ghostly invasion ships (*bottom*) that foreshadow his defeat.



EADWINE PSALTER, an illuminated-manuscript collection of the Psalms, was copied by the monk Eadwine from the earlier *Utrecht Psalter*. Judging from Eadwine's dates, stylistic evidence and this representation of a comet at the bottom of a page, the copy was made in or soon after 1145, when Halley's Comet appeared. The legend refers to the radiance of the comet, or "hairy star," and remarks that comets appear rarely, "and then as a portent." Above comet are three Latin versions of Psalm 5: *Hebraicum, Romanum* and *Gallicanum*.

there is a prophecy of Balaam recorded by Moses to this effect, 'there shall arise a star out of Jacob and a man shall rise up out of Israel' (Numbers 24 : 17)." (It is interesting to note that *The New English Bible* translates the passage: "A star shall come forth out of Jacob, a comet arise from Israel.")

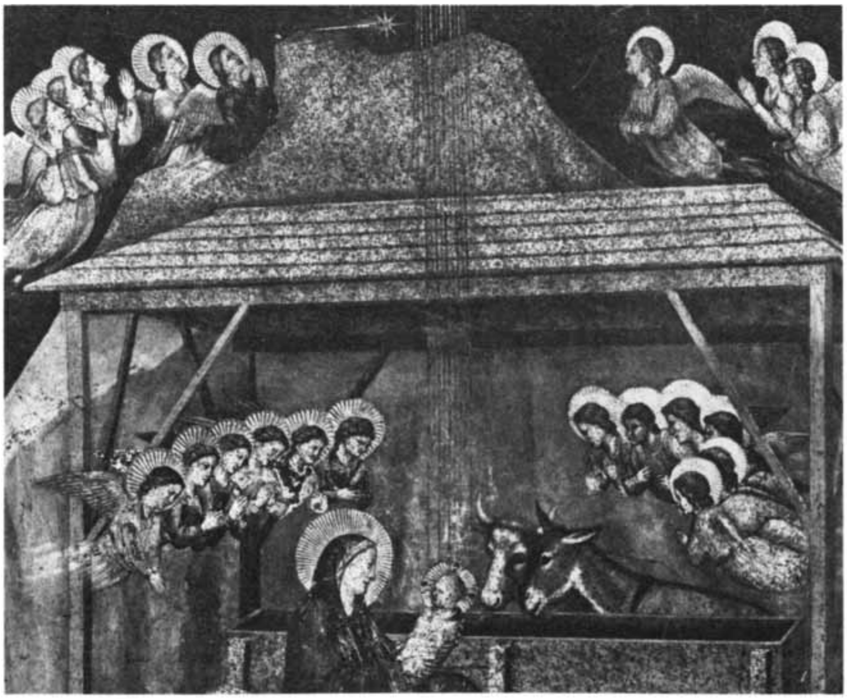
Origen thereby contributed three important concepts to the legend of the star of Bethlehem. By repeating the Old Testament prophecy he implied a connection with the old tradition that the birth of prophets such as Abraham and Moses was heralded by a star, stressing the continuity between the two testaments. He clearly associated the star with a comet. And he stated unambiguously that comets could portend good.

John of Damascus, a Byzantine dogmatician whose writings were translated into Latin in the 12th century and who was one of the few Greek church fathers known to 12th- and 13th-century Europeans, related the star to a comet in his influential *Exact Exposition of the Orthodox Faith,* albeit less directly. "It often happens that comets arise. These . . . are not any of the stars that were made in the beginning, but are formed at the same time by divine command and again dissolved. And so not even the star which the Magi saw . . . is of the number of those that [were] made in the beginning. And this is evidently the case because sometimes its course was from east to west, and sometimes from north to south. At one moment it was hidden, and at the next it was revealed: which is quite out of harmony with the order and nature of the stars." That description surely connoted a comet to a reader.

The writings of Flavius Josephus, a Jewish historian of the first century A.D., support the association of the star with a comet. In *The Jewish War* Josephus mentioned in conjunction the portent of a broadsword-shaped star over Jerusalem and "a comet that remained a whole year." This was probably a confused reference to the apparition of Halley's Comet in A.D. 66, just before the outbreak of the Jewish revolt against Rome. Josephus listed other portents, among them a cow that gave birth to a lamb. Writers on Christian doctrine later read these details as an allegory of the star of Bethlehem announcing the birth of Christ (the lamb), whereas Jewish writers considered the passage to be prescient of the burning of the Temple in A.D. 70. (The apparition of A.D. 66 might well have influenced Matthew, whose Gospel was set down after the fall of Jerusalem.)

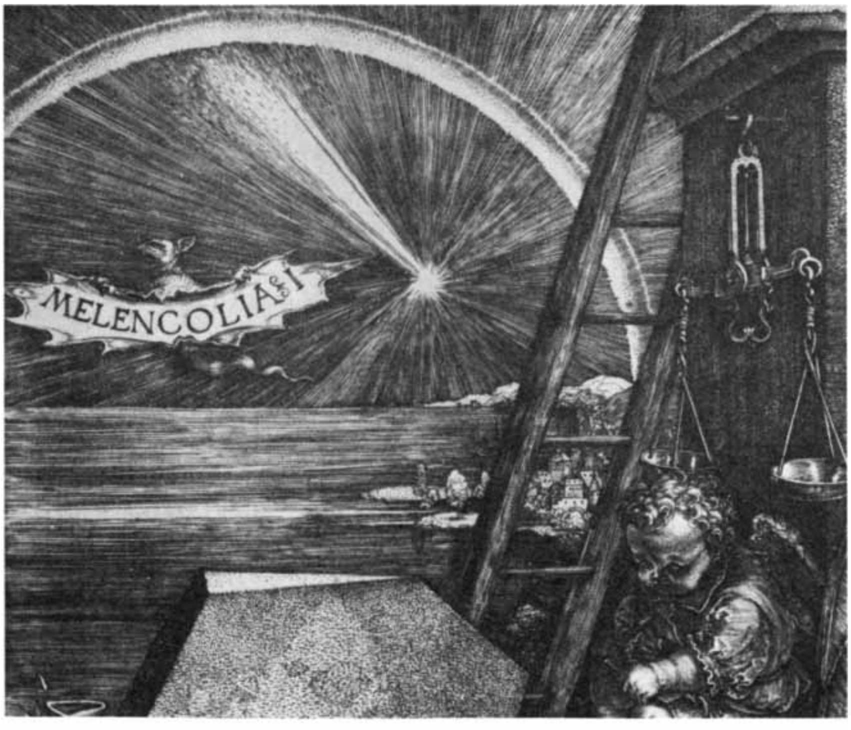
There was a very popular tradition in Giotto's own time linking the biblical star with a comet. In the widely read *Golden Legend* of 1275 (which has been suggested as a source for other details of the Arena Chapel cycle but not up to

NATIVITY BY DUCCIO DI BUONINSEGNA is a small panel from the predella, or base, of a very large altarpiece, the *Maestà,* executed between 1308 and 1311 for the Siena Cathedral; the panel (with two side panels, not shown here) is now in the National Gallery of Art in Washington. Duccio's star of Bethlehem, in contrast to the comet in Giotto's *Adoration,* is a small star with faint rays directed toward the Child. Duccio was Giotto's senior, from a rival school.



NATIVITY BY A FOLLOWER OF GIOTTO was painted in about 1316 in the Lower Church of San Francesco in Assisi; the fresco, of which this is a detail, is a pastiche of motifs from scenes in the Arena Chapel. There is a comet, but it is small and stylized, a pale echo of Giotto's, and it has been subordinated to the traditional heavenly rays descending to the Child below.

SCENES FROM INFANCY OF CHRIST are depicted on a page of an illuminated concordance of the four Evangelists executed in 1399 by the Paduan Jacopo Gradenigo. The painting contains two representations of the star of Bethlehem. Over the manger (*upper left*) it is a stationary star; above the Magi (*top center*) it is shown as a schematic comet, part angel and part star, with swooping rays that project forward as well as backward to form the tail.



ALBRECHT DÜRER included a luminous and dynamic comet, much less naturalistic than Giotto's, in his famous engraving *Melencolia I* of 1514. This is a detail from the upper left-hand corner of the complex print, which is interpreted as a "spiritual self-portrait" depicting the artist's melancholy at his inability to unite theory and practice. The comet and rainbow may symbolize the disease, insanity and miraculous events once associated with melancholics.

now for this one) the Genoese theologian and chronicler Jacobus de Voragine described the star in a way that would have suggested a comet to his readers: "It was a star new created and made of God. . . . Fulgentius [a sixth-century defender of orthodoxy] saith: It differed from the other stars in three things. First, in situation, for it was not fixed in the firmament but it hung in the air nigh to the earth. Secondly, in clearness. . . . Thirdly, in moving, for it went alway before the kings." Jacobus also commented on the etymology of "epiphany," stating that it was derived from *epi* ("above") and *phanes* ("apparition") and thereby identifying the Epiphany (the adoration of Christ by the Magi) with the apparition of a divine star. "Magi" itself is derived from the Greek *magoi*, which in turn comes from a Persian word for sorcerers and astrologers, the wise men who studied the stars.

Language as well as literature, then, supported the close linking of the biblical star and the Adoration of the Magi with an irregular and spectacular apparition of a "new" and miraculous star, which was most likely to have been a comet. Giotto or his patron could have cited doctrinal authority for the representation of the star as a comet. It is also noteworthy that in Giotto's time there arose a more dispassionate—one might almost say scientific—attitude toward comets. One contemporary chronicler of the 1301 apparition concluded his description with an affirmation of the freedom of man's will as opposed to the power of the stars. Padua was a center of mathematics, the discipline that would eventually elevate astronomy from the superstitions of astrology; already in the late 13th century scholars had begun to observe the heavens assiduously. (Galileo was to hold a chair of mathematics at the University of Padua in the 16th century.) Giotto, then, worked in a philosophically progressive community that might well have encouraged the spirit of individualism and objectivity with which he regarded the comet and represented it in his fresco.

The revolutionary nature of Giotto's rendering is apparent when his comet is compared with representations of the star of Bethlehem by his contemporaries and successors. Giotto's chief rival of the period, the innovative but more Byzantine painter Duccio di Buoninsegna of Siena, executed an enormous altarpiece, the *Maestà*, for the Siena Cathedral between 1308 and 1311. In Duccio's *Nativity*, a panel from the predella, or base, of the altarpiece, he painted a static, symbolic eight-pointed star in the conventional iconographic tradition, with thin gilt rays pointing down to the Christ Child.

In about 1316, a decade after the com-

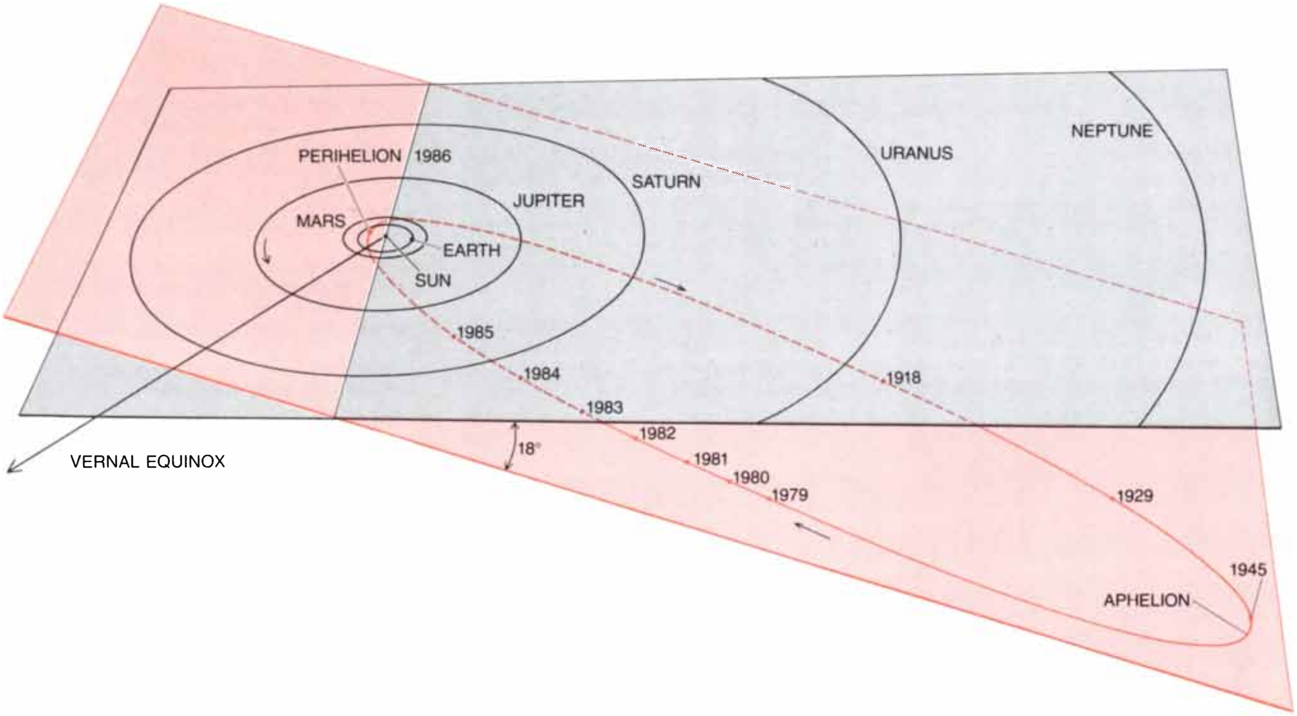
pletion of the Arena Chapel cycle, a follower of Giotto painted a Nativity fresco in the Lower Church of San Francesco at Assisi, combining in the one scene an assortment of motifs from Giotto's Arena cycle. The anonymous artist confused his stellar images, painting a small conventionalized comet and a separate and more dominant stream of gold rays descending from heaven. (He painted an identical star-comet in the Adoration scene of the cycle, but it is damaged and barely visible.)

In contrast to Giotto's analytical depiction of Halley's Comet, the tiny Assisi comet is a stylized eight-pointed star like Duccio's, but with a wispy, linear tail; it seems to have been included as a timid afterthought mimicking the Arena comet. The artist either had never studied a comet apparition or had no desire to represent it convincingly; he wanted only to show the star as a symbol, relying on his knowledge that Giotto had painted it as a comet. Perhaps the separate rays from heaven implied theological criticism of Giotto's unorthodox depiction; by stressing the glorifying rays the Assisi artist may have meant to reinforce in a dogmatic way the divine nature of the event and the blessing conferred on it. Perhaps the less daring Assisi interpretation was required by the very public nature of this major pilgrimage church. The Arena Chapel, on the other hand, was private, providing Giotto with a more relaxed environment for both theological and artistic experimentation.

The advanced status of Giotto's portrait of the comet is accentuated by comparison with a much later work by another Paduan artist. In 1399 Jacopo Gradenigo executed an illuminated manuscript, a concordance of the Gospels. The star of Bethlehem over the Magi in his Infancy cycle is not a real comet but a symbolic representation, part angel and part star, culminating in a long, sweeping tail. It is highly schematic and has little to do with Giotto's naturalistic conception, even though it postdates the Arena Chapel fresco by almost a century. Illuminated manuscripts tend to reflect earlier innovations in more monumental painting. Gradenigo was not an innovator: the conservative nature of his art is evident in his adherence to the old device of "continuous narrative," with six temporally different scenes shown in the same illustration. Gradenigo's representation of the star as a comet, stylized though it is, demonstrates that by the end of the 14th century the comet interpretation of the star of Bethlehem had gained acceptance in Padua.

The comet did not supplant the more usual interpretation of the star in European art in general, however. We do not even know if Giotto repeated it, because the Arena Chapel painting is the only extant depiction of the star of Bethlehem securely attributed to him. It seems futile to speculate about whether the idea of representing the star of Bethlehem as a comet was conceived by Giotto, by his patron Scrovegni or by a theological adviser. Apart from the question of iconographical innovation the significance of the *Adoration* comet lies in its striking naturalism, which surpasses even the comet depicted by Dürer in 1514 in his *Melencolia I.*

It is clear that by Giotto's time a substantial body of tradition associated the star of Bethlehem with a comet. Giotto drew on that tradition and on his own observation of Halley's Comet in 1301 when he painted his *Adoration,* and in doing so he depicted a celestial body that reflected naturalistic truth even as it heightened emotional intensity. Moreover, by painting a historical comet Giotto enhanced the contemporary impact of his depiction of the Adoration. He encouraged his viewers to identify with the biblical witnesses of the miraculous event of Christ's birth: they too had experienced a comet's spectacular apparition, in 1301.



ORBIT OF HALLEY'S COMET (*color*) is an elongated ellipse with a major axis of about 35.6 astronomical units (1 A.U. is the mean distance from the earth to the sun) and a minor axis of about 9 A.U. The comet's motion is retrograde, that is, in a direction opposite to the motion of the planets around the sun; the plane of the orbit (*light color*) is inclined 18 degrees from the plane of the earth's orbit (*gray*). The comet will reach its perihelion, about .6 A.U. from the sun, on February 9, 1986. The colored dots show the position of the comet at perihelion and on the same date in some other years; the black dot shows where the earth will be when the comet reaches its perihelion.

THE AMATEUR SCIENTIST

How to measure the size of the earth with only a foot rule or a stopwatch

by Jearl Walker

Two ingenious procedures for measuring the radius of the earth have been sent to me independently by Joseph L. Gerver of Honolulu and Dennis Rawlins of San Diego. The procedures are elegant but simple enough for an amateur to be able to repeat easily. You might enjoy seeing how close you can come to the true size of the earth and how subtle variations in the methods alter the results.

Gerver's procedure calls for three stones, a line of string and a ruler. To follow his directions you will need to find a small hill that gives you a clear view of the ocean horizon in opposite directions. Ideally the hill should be between 100 and 1,000 feet above sea level and should afford a clear view of a section of the beach. Gerver did his work on Makapuu Point on the eastern tip of Oahu. You might try to find a similar peninsula. Since part of the experiment involves viewing two stones placed on the beach, you also need to be able to see the shoreline. The final requirement is a pointed rock on the hill.

Mount the ruler upright about 15 feet from the pointed rock, measuring the distance with the string. Gerver labels the distance b. For the first measurement stand on the other side of the ruler from the rock and adjust your eye level until you can line up the ocean horizon, the top of the rock and a point on the ruler. Note where your line of sight intersects the ruler. Move to the rock and adjust your line of sight until the top of the rock is lined up with the ocean horizon in the opposite direction. Again note where your line of sight intersects the ruler. The distance between the two points of intersection on the ruler is labeled a.

The last measurement you need is your height above sea level. Place two stones on the beach so that they are equidistant from your observation point on the hill and a measured distance apart. Return to the observation point, hold the ruler at arm's length, sight on the stones and note their apparent separation on the ruler. Then, shifting the ruler to a vertical position by rotating your fist, measure the distance from the stones to the ocean horizon in units of the apparent separation. Your height above sea level (call it h) is the number you get multiplied by the true separation between the stones. Finally, calculate the radius of the earth as being approximately equal to $8hb^2/a^2$.

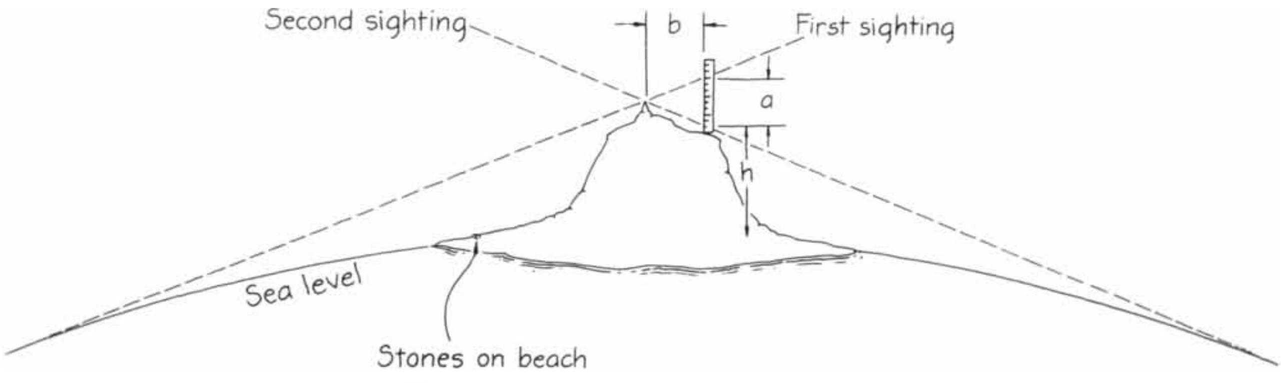
For example, in one of Gerver's observations the distance between the two points of intersection on the ruler was about an inch, the distance between the pointed rock and the ruler was 144 inches and the observation point was 208 feet above sea level. The calculated radius for the earth was 6,540 miles, or 10,500 kilometers, which is appreciably higher than the official radius of 6,378 kilometers at the Equator. (The official polar radius is 6,357 kilometers.) With repeated observations the random errors incorporated in the results could be reduced, but several systematic errors will still remain.

Why does Gerver's formula work? The geometry of the experiment, which is shown in the illustrations on pages 176 and 178, includes four important triangles. They are similar triangles, that is, they have equal angles. Since the triangles are similar, Gerver could employ them to devise several ratios involving the earth's radius r and the three measurements in the experiment: a, h and b. He was then able to solve for r.

I shall sketch out his proof here but leave the details to you. From triangles *1* and *2* you can write the ratio $CO/r = r/(r + h)$. From triangles *2* and *3* you have the ratio $AC/BC = CO/AC$. From triangles *2* and *4* the ratio to be used is $b/(a/2) = CO/AC$. Finally, BC is $r + h - CO$. By eliminating CO, AC and BC through appropriate substitutions in these ratios you can solve for r. At one point in the calculations Gerver makes the approximation that the altitude of the hill is insignificant compared with the diameter of the earth, an approximation that introduces into the results an error of less than .01 percent.

The height h of the experimental position above sea level was measured by sighting on the stones on the beach. Gerver makes the approximation that the ratio of h and the true separation between the stones is the same as the ratio of the apparent separation between the stones and the apparent distance from the stones to the nearest ocean horizon. In Gerver's scheme for measuring the distance you measure the distance to the ocean horizon in terms of the apparent separation of the stones, and so you automatically have the second of the two ratios. When you multiply by the true separation of the stones, you get the height above sea level.

In addition to random errors in the observations several systematic errors



Joseph L. Gerver's method of calculating the radius of the earth


IMPORTED BY W.A. TAYLOR & CO., MIAMI, FLORIDA, 80 PROOF, SOLE DISTRIBUTORS U.S.A. © 1978
PAINTING: THE CALVIN BULLOCK COLLECTION.
Courvoisier
VSOP
Fine Champagne
COURVOISIER
The Brandy of Napoleon
COGNAC
Produced and Bottled by
Courvoisier, Jarnac, France
COURVOISIER VSOP. THE BRANDY OF NAPOLEON

Physician, did you miss any of these significant developments in medical science?

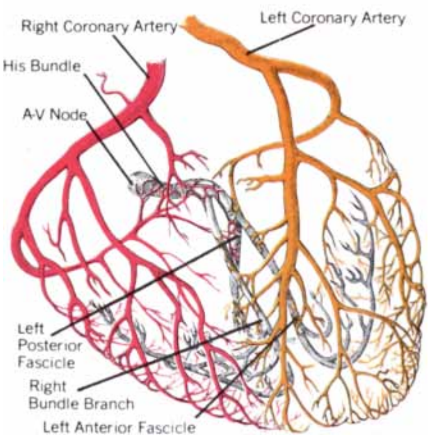
• The lidocaine level in the blood may fall below the therapeutic range when the drug is given as a single bolus followed by a continuous infusion. To avoid a "therapeutic hiatus," a second bolus should be given 20 to 30 minutes after the first bolus...

• Evidence is mounting that there is widespread over-diagnosis of pulmonary embolism at a time when heparin, in therapeutic doses, has become an important cause of drug-related death in the relatively well, hospitalized patient...

• The widely used street drug, phencyclidine, appears to cause persisting and possibly permanent brain dysfunction after repetitive use...

• There is alarming incidence of acute leukemia which may reach five percent in patients who appear to have been successfully treated with chemotherapy or radiation therapy for Hodgkin's disease...

Every day medical science evolves in ways that affect patient management and therefore patient well-being. But with more than 2,000 journals published, innumerable meetings held and hundreds of courses given, it is all too easy for developments that compel change in diagnosis and treatment to get by even the most conscientious reader and meeting-goer. And since such discoveries are today's news, they will not appear in the standard texts for years.



Branches of the right and left coronary arteries supply blood to the AV node and intraventricular conduction system.

SCIENTIFIC AMERICAN *Medicine* is lucidly illustrated with drawings and photographs. Some examples are seen here and on the facing page.

We have a solution to this problem: It is called SCIENTIFIC AMERICAN *Medicine*. It makes such developments accessible to you when and where you need them, month after month.

Here is how SCIENTIFIC AMERICAN *Medicine* works. When you subscribe, you receive a two-volume, 2,000 page text, organized according to the familiar subspecialties so that you know where to find the answers to your questions.

The authors of SCIENTIFIC AMERICAN *Medicine* are leading authorities from the faculties of the Harvard Medical School and the Stanford University School of Medicine. These seasoned scholar-practitioners, close to the creation of new medical knowledge, write for you the practitioner, not the student. Therefore, at the time you receive your copy of SCIENTIFIC AMERICAN *Medicine*, you will hold in your hands the world's most advanced text of medicine. As long as you continue to subscribe it will stay that way.

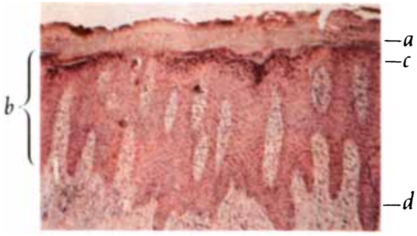
Each month you will receive several updated chapters. The new chapters which replace the old chapters contain reports of the significant advances in medical science that affect patient management. The index will also be updated each month. Furthermore, updated material will be summarized in a monthly bulletin accompanying the updated chapters so that you can quickly scan the information.

As the updated material is incorporated into the text, older more familiar material is deleted. Thus the amount of useful knowledge increases but the number of pages does not.

A Continuing Medical Education Program

For subscribers who wish it, SCIENTIFIC AMERICAN *Medicine* offers at no extra charge a continuing medical education program sponsored by the Stanford University School of Medicine's Office of Postgraduate Medical Education and administered by the Educational Testing Service of Princeton, New Jersey. The program consists of eight patient management problems per year, each written by a member of a Panel of Examiners drawn from leading clinical institutions. By completing all eight PMPs in a 12-month period the subscriber can earn 32 Category 1 credits per year, from the Stanford University School of Medicine, toward the AMA's Physician's Recognition Award.

This complete, concise medical knowledge service is yours for the drop of the accompanying card. Please fill it out and mail it to us. We think you will find that SCIENTIFIC AMERICAN *Medicine* is your answer to the challenge of bringing all that modern medicine has to offer to your patients.



Histopathologic changes in psoriasis.

Extensive psoriasis (top) cleared completely (bottom) after four weeks of treatment with anthralin.

THE DISTINGUISHED AUTHORS, AND THE FIFTEEN SECTIONS OF SCIENTIFIC AMERICAN *MEDICINE*.

1. Cardiovascular Medicine

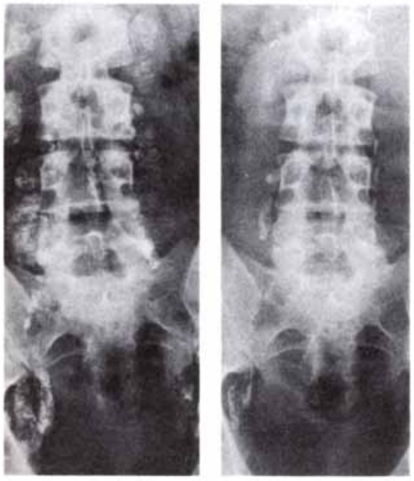
Edgar Haber, M.D., Professor of Medicine, Harvard Medical School, Physician and Chief of Cardiology, Massachusetts General Hospital

E. William Hancock, M.D., F.A.C.P., Professor of Medicine, Stanford University School of Medicine

Roman W. DeSanctis, M.D., Professor of Medicine, Harvard Medical School, and Physician and Director, Coronary Care Unit, Massachusetts General Hospital

Adolph M. Hutter, Jr., M.D., F.A.C.P., Associate Professor of Medicine, Harvard Medical School, and Assistant Physician and Associate Director, Coronary Care Unit, Massachusetts General Hospital

Eve Elizabeth Slater, M.D., Instructor in



Extensive retroperitoneal Hodgkin's disease before and after MOPP chemotherapy.

Medicine, Harvard Medical School, and Chief, Hypertension Unit, Massachusetts General Hospital

2. Dermatology

Eugene M. Farber, M.D., Professor and Chairman, Department of Dermatology, Stanford University School of Medicine

3. Endocrinology

Daniel D. Federman, M.D., F.A.C.P., Professor of Medicine and Dean for Students and Alumni, Harvard Medical School

4. Gastroenterology

Gary M. Gray, M.D., Professor of Medicine and Head, Division of Gastroenterology, Stanford University School of Medicine

5. Hematology

Stanley L. Schrier, M.D., Professor of Medicine and Head, Division of Hematology, Stanford University School of Medicine

6. Immunology

John David, M.D., Professor of Medicine, Harvard Medical School, and Physician, Robert B. Brigham Hospital

7. Infectious Disease

Thomas C. Merigan, M.D., Professor of Medicine and Head, Division of Infectious Diseases, Stanford University School of Medicine

Morton N. Swartz, M.D., F.A.C.P., Professor of Medicine, Harvard Medical School, and Chief, Infectious Disease Unit, Massachusetts General Hospital

Cyrus C. Hopkins, M.D., Assistant Professor of Medicine, Harvard Medical School, and Hospital Epidemiologist and Associate Physician, Massachusetts General Hospital.

Adolf W. Karchmer, M.D., F.A.C.P., Assistant Professor of Medicine, Harvard Medical School, and Associate Physician, Massachusetts General Hospital

Robert H. Rubin, M.D., Assistant Professor of Medicine, Harvard Medical School, and Assistant Physician, Massachusetts General Hospital

Harvey B. Simon, M.D., F.A.C.P., Assistant Professor of Medicine, Harvard Medical School, and Assistant Physician, Massachusetts General Hospital

8. Intensive And Emergency Care

Edward Rubenstein, M.D., F.A.C.P., Associate Dean of Postgraduate Medical Education, and Professor of Clinical Medicine, Stanford University School of Medicine

9. Metabolism

George F. Cahill, Jr., M.D., Professor of Medicine, Harvard Medical School, Director of Research, Howard Hughes Medical Institute, and Physician, Peter Bent Brigham Hospital

10. Nephrology

Roy H. Maffly, M.D., Professor of Medicine, Stanford University School of Medicine, and Chief, Renal Service, Palo Alto Veterans Administration Hospital

11. Neurology

Robert W.P. Cutler, M.D., Professor of Neurology, Stanford University School of Medicine

12. Oncology

Saul A. Rosenberg, M.D., F.A.C.P., Professor of Medicine and Radiology, and Chief, Division of Oncology, Stanford University School of Medicine

13. Psychiatry

Ned H. Cassem, M.D., Associate Professor of Psychiatry, Harvard Medical School, and Chief, Psychiatric Consultation—Liaison Service, Massachusetts General Hospital

14. Respiratory Medicine

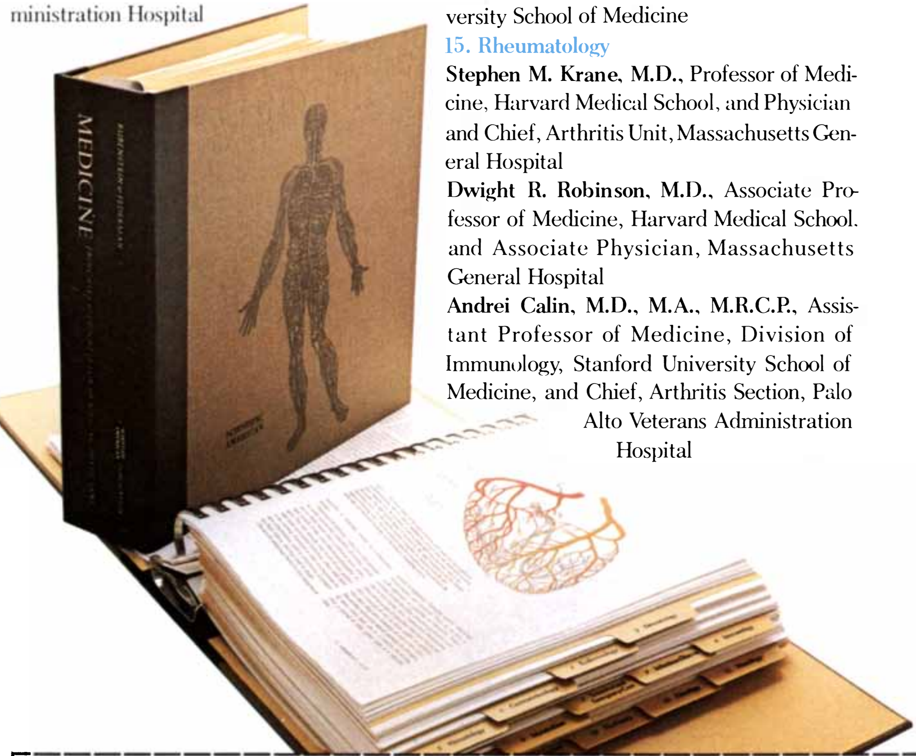
Eugene D. Robin, M.D., F.A.C.P., Professor of Medicine and Physiology, Stanford University School of Medicine

15. Rheumatology

Stephen M. Krane, M.D., Professor of Medicine, Harvard Medical School, and Physician and Chief, Arthritis Unit, Massachusetts General Hospital

Dwight R. Robinson, M.D., Associate Professor of Medicine, Harvard Medical School, and Associate Physician, Massachusetts General Hospital

Andrei Calin, M.D., M.A., M.R.C.P., Assistant Professor of Medicine, Division of Immunology, Stanford University School of Medicine, and Chief, Arthritis Section, Palo Alto Veterans Administration Hospital



SCIENTIFIC AMERICAN MEDICINE

Post Office Box 5911, New York, N.Y. 10017

Gentlemen: Please enroll me as a subscriber to SCIENTIFIC AMERICAN *Medicine.* On receipt of this coupon you will send me the advanced two-volume text described in your announcement, and update it regularly by sending me new monthly subsections. As part of this service, I understand that the annual subscription of $150 for SCIENTIFIC AMERICAN *Medicine* is tax-deductible. If I am not entirely satisfied, I may cancel at any time during the first 60 days, returning all materials for a complete refund. SA5-79

☐ Please enter my subscription for SCIENTIFIC AMERICAN *Medicine.*
☐ I shall also enroll in the CME Program
☐ I enclose a check made out to SCIENTIFIC AMERICAN *Medicine* for $150
☐ Please bill me

Name______________________________

MD Specialty______________________________

Address______________________________

City________________ State__________ Zip________

Subscribers outside of the U.S., its possessions, and Canada will be charged extra for shipping and handling; updates will be sent by surface routes unless air mail delivery is requested.

figure in the results. One error involves the bending of light rays that skim the horizon, slightly raising it in your field of view. This refractive effect of the atmosphere makes Gerver's computed radius too large by about 20 percent. I shall discuss the effect further when I describe Rawlins' method of measuring the size of the earth.

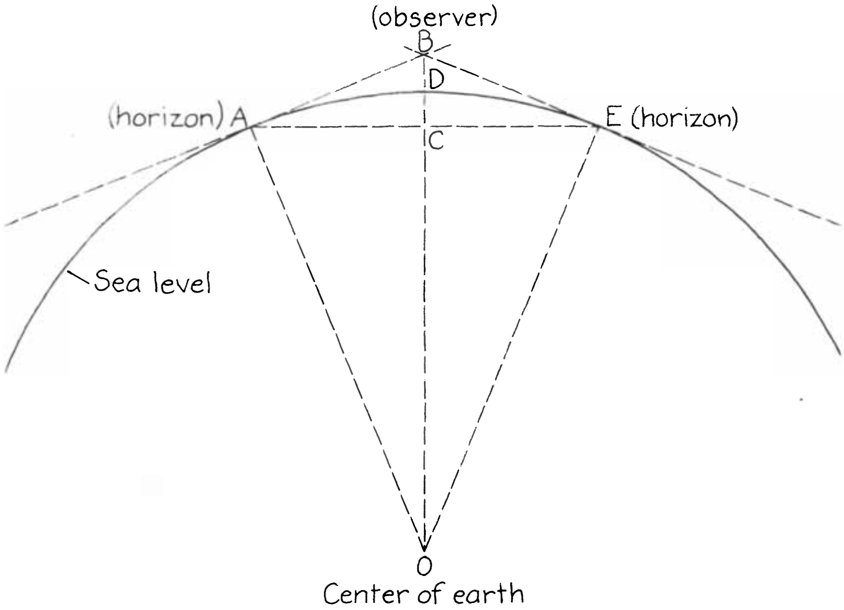
Another factor to consider is the waves on the water; they too act to raise the horizon in your field of view. This effect can be taken into account by subtracting half of the mean wave height (half of the crest-to-trough distance reported by the Coast Guard and other services) from the observation height h determined by Gerver's method.

One source of random error in the experiment arises in judging the alignment of the tip of the rock, the ocean horizon and a mark on the ruler. It is not only the usual kind of error inherent in measuring the length of anything but also the error arising from the fact that these particular things do not cast sharp images on the retina.

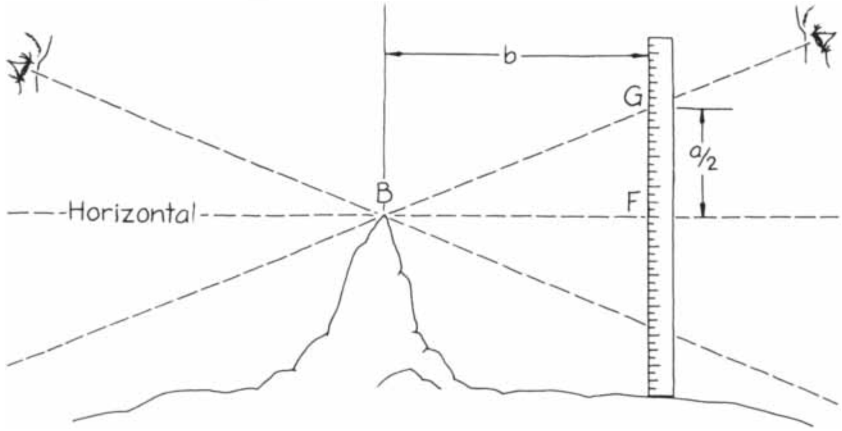
Still another source of error comes in estimating the height of the observation site above sea level. The method of estimating the height on the basis of the distance between the stones on the beach assumes that a line to the ocean horizon is horizontal. It is actually angled downward because of the curvature of the earth. This angle, which is called the dip angle, is proportional to the square root of the observer's height above sea level.

With proper care in your observations and with multiple observations Gerver's method will enable you to come close to the true radius of the earth. (What you are actually measuring is not, of course, the equatorial radius of the earth but the local radius of curvature of the ocean surface.)

Rawlins published his method for determining the size of the earth in the February issue of *American Journal of Physics*. Basically he has you view two sunsets in quick succession, one as you are lying down on a beach and the other after you leap to your feet. By measuring the time that elapses between the two sunsets you should be able to calculate the radius of the earth to an accuracy of about 10 percent.



The geometry of Gerver's method



Triangle **BFG** *between the rock and the ruler*

To follow Rawlins' procedure you need a clear view of the sunset from a beach overlooking a relatively calm ocean or large lake. The eastern side of one of the Great Lakes may be good, and places around Great Salt Lake, which typically has low waves, may be ideal. A stopwatch is preferable for the timing, but a sweep second hand on a wristwatch will also serve. You should be very careful about how you view the sun. Even at sunset looking directly at the sun for any length of time can harm your eyes. The greatest danger is from the invisible radiation, which can damage the eye without causing any immediate pain. It is best if you avoid gazing at the sun's disk until it is mostly below the horizon. You can keep track of the disk's motion by looking off to the side of it or by giving it an occasional quick glance.

The basic geometry involved in Rawlins' procedure is shown (for an observer on the Equator) in the top illustration on page 179. You observe the first sunset while you are lying at point M. As soon as the last visible arc of the sun's disk disappears below the horizon you jump up and wait for the same thing to happen from an elevation of h, the height of your eyes. The last ray in the first observation is tangent to the earth's surface at your location and therefore forms an angle of 90 degrees with the vertical, but the last ray in the second observation is not tangent. With your eyes above sea level the ocean horizon forms an angle (labeled θ in the illustration) with the vertical; it is 90 degrees minus the "dip angle."

The time between the disappearance of the last ray in the first view and the disappearance of the last ray in the second view is proportional to the dip angle θ through which the sun appears to descend during the double sunset. (Actually, of course, it is the angle through which the earth turns in that time.) If the time is measured in seconds, the proportionality is found from the fact that a full rotation of the earth corresponds to 86,400 seconds, or 24 hours.

Rawlins recognized that the triangle QNO was a right triangle and thus that the sides were related by the Pythagorean theorem. One of the legs, QN, is approximately equal to the earth's radius multiplied by the angle through which the earth turned between the two sunsets. (The leg is therefore approximately equal to the arc between points Q and

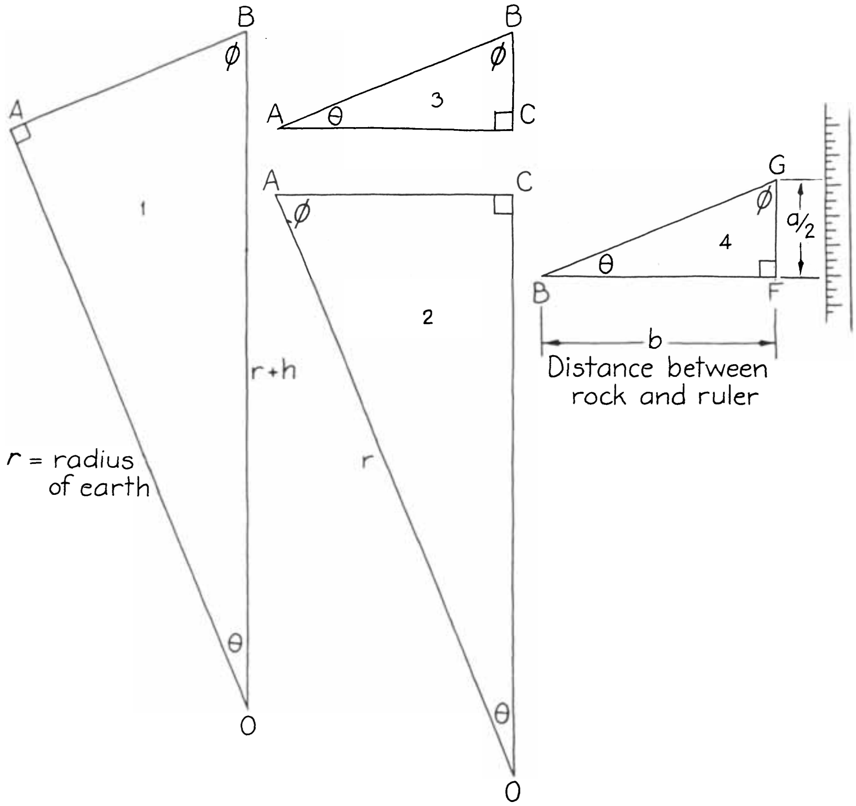
M.) Putting this information together with the correspondence between the angle and the elapsed time, Rawlins devised an equation relating the time between the two sunsets, the earth's radius r and the eye-level height h of the observer. (He made one additional approximation: the eye-level height is negligible compared with the earth's radius.) The earth's radius in kilometers is equal to a proportionality constant of 3.78×10^5 multiplied by the eye-level height (expressed in meters) and divided by the square of the time between sunsets (measured in seconds).

If the first of the observations is not made at sea level, the height employed in the calculation is less easily determined. Suppose you make the first observation at normal eye level h while you are standing on the beach and then make the second observation after running up a flight of steps. The effective height difference for the calculation is the square of the difference of the square roots of the altitudes of your two observations. I shall give an example of this calculation below.

Although in the ideal situation you would make your observations on a calm ocean or lake, it is more likely that you will have to deal with waves. They complicate the observations in two ways. First, waves near the shore obscure the ocean horizon when you lie on the beach for your first observation. Hence the first observation must be made from a point higher than the height of the waves. Second, waves at the ocean horizon, which should be a few miles away, raise the horizon. The best you can do to account for them is to estimate their average height and subtract it from each of the heights from which you make your observations. Rawlins takes the average height for the distant waves to be .6 meter, which he bases on oceanographic observations.

A subtler correction has to do with the curvature of light rays by refraction in the earth's atmosphere. When a light ray passes from one medium to another medium with a different index of refraction, the path of the light is bent at the interface separating the two materials. The only exception comes when the ray is perpendicular to the interface; then the direction does not change. The extent of the bending in all other instances depends on the angle at which the ray meets the interface and on the difference between the refractive index of one medium and that of the other. The direction of the bending is governed by whether the refractive index increases or decreases across the interface. If it increases, the ray is closer to being perpendicular to the interface. If it decreases, the ray bends the other way.

As a ray of sunlight passes through the earth's atmosphere it is likely to be refracted continuously as it meets layers of air that have somewhat different densities and therefore different refractive indexes. The extent of the bending is insignificant if the light has an almost vertical path through the atmosphere, but if it enters the atmosphere at a large angle to the vertical, as a ray of light from the setting sun does, the refraction can be significant.



The similar triangles employed in Gerver's calculations

If you are an amateur astronomer, you may have noticed the effect with stars. When a star is overhead, it is seen in its correct position because light from it is not refracted much. When a star is just above the horizon, however, it is seen slightly out of position (too high by about half a degree at the horizon) because the light rays the eye eventually receives enter the atmosphere at such a large angle with respect to the vertical that they are significantly refracted. The refraction makes the star seem to take longer to set because when it is actually a fraction of a degree below the horizon, the bending of its light rays extends the image over the horizon.

So it is with the sun. You can see it even after it should be out of sight below the horizon. Since the same thing happens at sunrise, the daylight lasts a little longer than it would without refraction. For example, since the sun sets perpendicularly to the horizon at the Equator, it can be seen for at least an additional 2.3 minutes because of refraction. Near the poles the additional time is much longer because the sun sets at a considerable angle to the vertical and therefore drops behind the horizon much more slowly.

Refraction makes trouble both for Gerver's method and for Rawlins'. A ray leaving the horizon travels not in a straight line to the observer but roughly in a circle that has a radius six times the radius of the earth. (A good discussion of this effect appears in a classic work by Simon Newcomb, *A Compendium of Spherical Astronomy,* which was first published in 1906. It can be found on pages 198–203 in the current Dover reprint.)

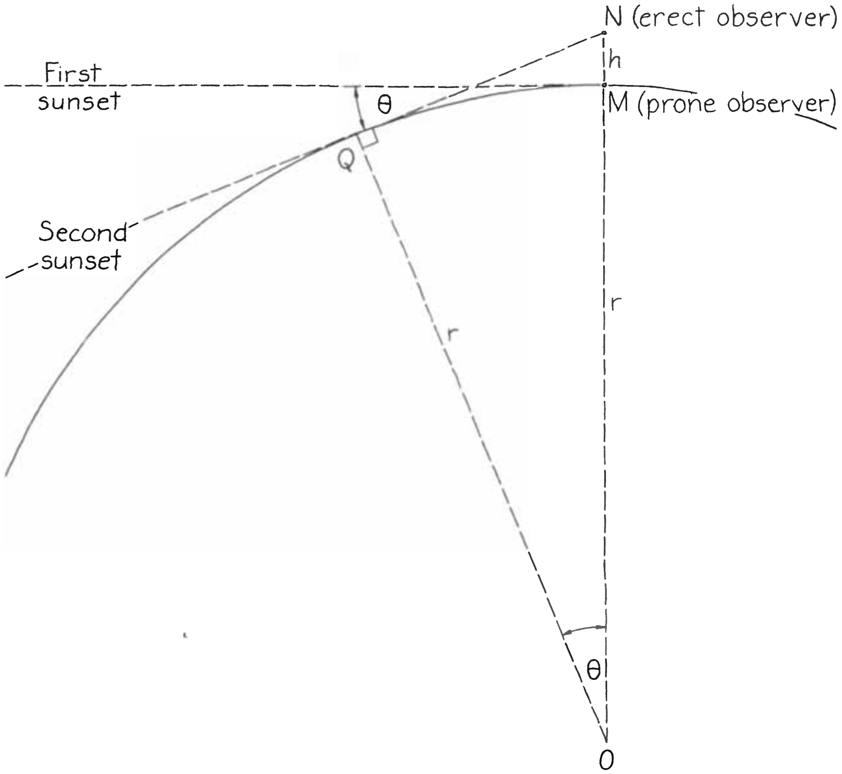
The effects of refraction on the Gerver computations are the opposite of those on the Rawlins ones. In Gerver's method the apparent dip of the horizon is contracted, thereby decreasing length a. Since that length is needed for several calculations of proportionality made to find the earth's radius, the error is maintained, finally yielding a radius that is about 20 percent too large because of the refraction. In Rawlins' technique the distance to the ocean horizon is increased, enlarging the angle through which the earth must rotate to give the observer the second sunset. Hence the refraction increases the time between sunsets and decreases the computed radius of the earth by about the same 20 percent.

One cannot account for the effect of refraction in any consistent way because it depends on the temperature and pressure in the air layers through which the light rays travel. Those quantities change frequently, not only from place to place in the atmosphere but also at a given place. As an approximate measure of the refraction Rawlins multiplies his basic equation for the radius by 1.2. This correction rests on the assumption that the effects of refraction lower the computed radius by about 1.2. The correction effectively reduces the measured time between sunsets (more precisely the square of the time) to the value one would obtain in the absence of atmospheric refraction.

Without elaborate refinements of Gerver's and Rawlins' procedures the only way to get around the daily variation in refraction is to repeat the experiments over many days and average the calculated radii. The corrective factor of 1.2 might normally be off by a few percent. If the atmosphere is strongly stratified into layers with sharp differences in refractive index, the setting sun will even appear to be chopped into layers. The corrective factor of 1.2 will then probably be quite incorrect and your computation of the earth's radius will be far off the mark.

Rawlins makes two more alterations in his basic equation. They both have to do with the fact that in most places the sun does not set perpendicularly to the horizon. The ideal vertical setting can be seen only at the Equator at the time of the spring and fall equinoxes. For any other latitude or date sunsets are slower because the sun sets at an angle to the vertical. Rawlins handles this effect with two factors he calls A and B. The first factor is designed to take care of the date, the second to handle the latitude. A is the square of the cosine of the sun's declination on the day of the experiment. B is the square of the sine of the latitude. (Whether it is north or south latitude does not matter.)

The sun's apparent path through the sky for the year lies in the plane of the ecliptic, which is tilted with respect to the plane of the Equator by 23.44 degrees. The sun crosses the equatorial plane twice a year: on the vernal equinox, which is about March 20, and on the autumnal equinox, which is about September 23. (The precise times of the equinoxes shift backward every four years except when leap year is skipped.) The sun's declination is the angular distance north or south of the earth's equatorial plane. For example, at the equinoxes the declination is zero, at the winter solstice (December 21) it is about 23.4 degrees south, and at the summer solstice (June 21) it is about 23.4 degrees north. Precise tables of the sun's declination can be found in *The American Ephemeris and Nautical Almanac,* which is published yearly by the Government Printing Office. You can probably find your approximate latitude in an atlas.



The geometry of Dennis Rawlins' sunset-watching method

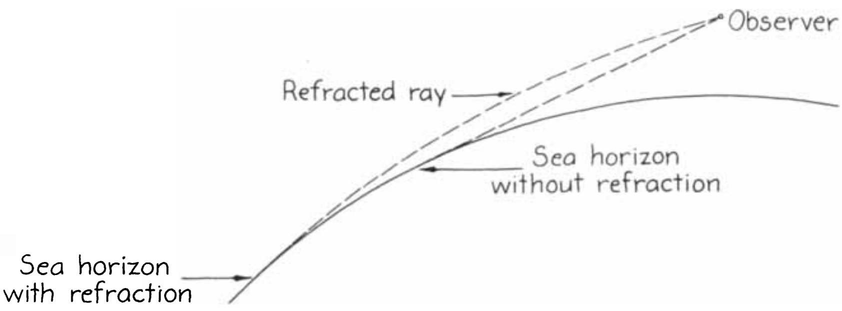
With this latitude and the date of your observations you can then take the associated factors A and B from the illustration on the next page. To use these corrective factors subtract B from A and then divide the result into Rawlins' basic equation.

As an example of the full formula Rawlins sends along the data from one of his observations. On April 5, 1978, he timed a double sunset at Cove Park in La Jolla, Calif. The latitude of the park is 32°51′ N. Rawlins made the first sunset observation at 6:11:55 P.M. with his eyes at a height of 1.72 meters above sea level. After running up a stairway he made his second observation at 6:12:15 P.M. at a height of 8.95 meters above sea level. The measured time on his stopwatch between the last rays of the first sunset and those of the second was 19.6 seconds.

Rawlins next approximated the height added to the ocean horizon by the distant waves as being .6 meter. This number was subtracted from both of his observation heights, which then became 1.12 and 8.35 meters. To get the effective height difference needed for the calculation he computed the square root of both of these numbers, found the difference between them and squared the result. The answer was 3.35 meters.

Squaring the time, dividing the result into the proportionality constant of 3.78×10^5 and then multiplying by the effective height difference of 3.35 meters gave Rawlins 3,300 kilometers (2,100 miles) as the radius of the earth. That was about 48 percent too low, but the factors involving the date, the latitude and the refraction were still to be taken into account.

Interpolating from the tables, Raw-



The role of refraction

lins found that April 5 has an *A* factor of .988. The *B* factor for his latitude was approximately .295. *A* minus *B* was .693, which he divided into the previous result to get a radius of 4,800 kilometers (3,000 miles), a result about 25 percent too low. To take refraction into account he multiplied this result by 1.2 to get his final value for the radius: 5,700 kilometers (3,500 miles). This is about 10 percent less than the official value.

If you live near the ocean or a large lake, you might like to repeat either Gerver's procedure or Rawlins', gathering enough data to enable you to find an average value for the radius of the earth. Since atmospheric refraction affects their procedures in opposite directions, by following both procedures and averaging the results you could almost eliminate the effects of refraction. I find it impressive that one can come anywhere near the earth's true radius with rough measurements and without any equipment other than a ruler or a watch and even more impressive that values within 10 to 20 percent of the official ones can be achieved.

Long ago some people realized that the earth was round because of several clues. Stars were in noticeably different positions at places widely separated from north to south. An outbound ship disappeared below the horizon. The earth's shadow on the moon during a lunar eclipse was round. Nevertheless, the idea of a round earth was difficult for many people to accept.

Legend records that the Greek scholar Eratosthenes, working in Alexandria, made the first measurement of the size of the earth about 2,200 years ago. He employed a simple (even crude) method but obtained surprisingly good results. At Syene (now Aswan) he either noted or was told that at noon on the summer solstice sunlight entered a well vertically. He reasoned that he could calculate the circumference of the earth if he knew the angle (from the vertical) of the sun's rays at noon on the same date in Alexandria, which was north of Syene and was situated on about the same meridian.

According to the story, Eratosthenes' measurement in Alexandria showed that the rays were slanted from the vertical by one-fiftieth of a circle (7.2 degrees). Since the rays at Alexandria were parallel to those at Syene (he must have realized that the distance to the sun greatly exceeded the radius of the earth), the difference in angle had to be the result of the curvature of the earth. The arc between the two cities along the earth's surface therefore occupied an angle of 7.2 degrees with respect to the earth's center.

To find the circumference of the earth all Eratosthenes needed to do was to find the ratio of this angle and 360 degrees and equate it with the ratio of the arc distance between the cities and the circumference of the earth. It was not easy, however, to find the distance between the cities. Eratosthenes resourcefully measured the distance in terms of the travel time of camels, since he knew that a caravan typically took 50 days to travel between the cities. Knowing how

Day of the year (month / day)				A	Latitude	B
3/20	9/22	9/22	3/20	1	0°	0
3/27	9/15	9/30	3/13	.9975	5°	.0076
4/4	9/7	10/7	3/6	.99	10°	.030
4/10	9/1	10/14	2/27	.98	15°	.067
4/15	8/27	10/19	2/22	.97	20°	.117
4/20	8/22	10/23	2/18	.96	25°	.179
4/24	8/18	10/27	2/14	.95	30°	.250
4/28	8/14	10/31	2/10	.94	35°	.329
5/1	8/10	11/3	2/6	.93	40°	.413
5/5	8/7	11/7	2/3	.92	45°	.500
5/9	8/3	11/11	1/30	.91	50°	.587
5/13	7/30	11/14	1/27	.90	55°	.671
5/17	7/26	11/18	1/23	.89	60°	.750
5/21	7/22	11/22	1/19	.88	65°	.821
5/26	7/17	11/27	1/14	.87	70°	.883
5/31	7/12	12/2	1/9	.86	75°	.933
6/7	7/5	12/8	1/3	.85	80°	.970
6/12	6/29	12/13	12/29	.845	85°	.9924
6/21	6/21	12/21	12/21	.8417		

The **A** *and* **B** *factors employed by Rawlins* far a camel walked in a day, he was able to compute the distance between Alexandria and Syene and thereby calculate the circumference of the earth. His result (in modern units) was 46,250 kilometers, which is roughly 16 percent too high.

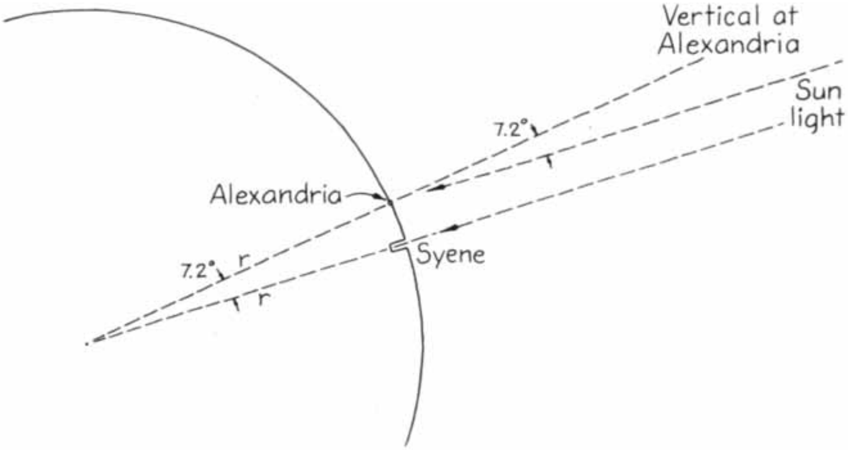
According to history, the next measurements of the circumference of the earth were made in about 85 B.C. by Posidonius of Apamea, who used the distance between Rhodes and Alexandria as Eratosthenes did the distance between Alexandria and Syene. Posidonius' means of measuring the distance was a boat that traveled between the two places. The angle was measured by sighting on the star Canopus; when it was on the horizon at Rhodes, it was 7.5 degrees above the horizon at Alexandria. By doing the same type of calculation that Eratosthenes is said to have done Posidonius computed the circumference of the earth. By modern reckoning his figure is low by about one part in six, or about 17 percent.

Still another measurement was reportedly made later by Abdullah al-Mamun, who obtained a value that was within 3.6 percent of the true figure. He had fairly accurate astronomical data for finding the needed angle. He also went to the trouble of measuring the distance of his arc by laying sticks down end to end.

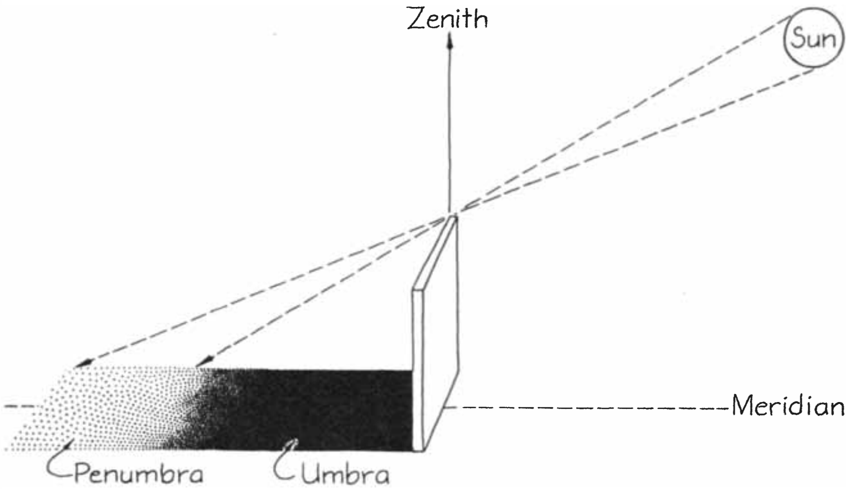
Modern techniques are of course far more accurate, exploiting artificial satellites, lasers and reflectors on the moon. As a result the size of the earth is well known, as is the fact that the earth is not exactly a sphere. We have come a long way from the time of the early Greeks, who believed the earth was flat and was held up by four huge elephants standing on the back of a huge turtle. We also no longer have the philosophical problem of deciding what the turtle stands on.

You might like to measure the earth's radius by the technique attributed to Eratosthenes. Your two cities should be close to the same meridian, although this condition is not critical. Try to choose cities as far apart on the meridian as is practical. Ascertain the distance by means of a map or a table of distances.

To measure the orientation of the sun's rays in the two cities you will need a friend to do the experiment in one city on the same day you make your measurements in the other city. Each of you should erect a gnomon, a vertical rectangular post with one face perpendicular to the meridian. When the sun reaches its zenith, measure the length of the shadow cast by the gnomon. (It will be shortest then.) The tangent of the sun's angular distance from the zenith is equal to the length of the shadow divided by the height of the gnomon. Although you could then compute the angle from a table of trigonometric functions, I suggest you work with a pocket calculator



Geometry of the method attributed to Eratosthenes



Gnomon method for determining the orientation of the sun's rays

to avoid the problem of interpolation between the values in a table. The angle between the vertical and the sun's rays is of course the same angle.

After making the measurements, subtract from them the angles you and your friend have obtained. Follow Eratosthenes' procedure by dividing the angular difference into 360 degrees and then multiplying by the arc distance between the two cities. The result is the circumference of the earth.

A problem with this method of measurement is that the far edge of the shadow is indistinct. The shadow fades from total brightness through an area of partial illumination (the penumbra) and into an area of maximum darkness (the umbra). The penumbra is partly illuminated because the plate blocks the rays originating from the lower section of the sun but does not block those from the upper section. Inside the umbra all the direct rays from the sun are blocked by the plate.

To find the angular distance between the zenith and the center of the sun you should measure the shadow to the center point of the penumbra. This measurement is difficult to make with precision. It inevitably introduces error, perhaps as much as a quarter of a degree if you use either edge of the penumbra in your determination of the angle of the sun's rays. The penumbra is .5 degree wide because the sun subtends an angle of .5 degree in one's field of view. The error inherent in locating the center of the penumbra makes the gnomon an imprecise instrument for measuring the size of the earth.

One way to get around the problem is to replace the gnomon with another device. Mount a horizontal pin at the center of a graduated circle placed so that its plane is vertical and aligned parallel to the meridian. The shadow cast by the pin has penumbras running the length of the shadow. Since they are symmetrical on the two sides of the shadow, you can easily find the center of the shadow and read its position on the graduated circle. When the circle is properly oriented, your reading will be the angular distance of the sun from the zenith and so will be the angle of the incident rays, without any systematic error introduced because of the penumbras.

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

MATHEMATICAL GAMES

LOXODROMIC SEQUENCES OF TANGENT SPHERES. H. S. M. Coxeter in *Aequationes Mathematicae,* Vol. 1, pages 104–121; 1968.

MAGIC AND PARAPHYSICS. Martin Gardner in *Technology Review,* Vol. 78, No. 7, pages 42–51; June, 1976.

GELLER, GULLS, AND NITINOL. Martin Gardner in *The Humanist,* Vol. 37, No. 3, pages 25–32; May/June, 1977.

STANDARD AND DAYLIGHT-SAVING TIME

TABLES OF SUNRISE, SUNSET AND TWILIGHT. Supplement to *The American Ephemeris.* U.S. Naval Observatory, 1946.

REVIEW AND TECHNICAL EVALUATION OF THE DOT DAYLIGHT SAVING TIME STUDY. National Bureau of Standards Report Prepared for the Chairman of the Subcommittee on Transportation and Commerce of the Committee on Interstate and Foreign Commerce, House of Representatives; April, 1976.

FROM SUNDIALS TO ATOMIC CLOCKS. James Jesperson and Jane Fitz-Randolph. National Bureau of Standards Monograph No. 155. National Bureau of Standards, 1977.

CARBON MONOXIDE IN THE GALAXY

THE HISTORICAL QUEST FOR THE NATURE OF THE SPIRAL NEBULAE. J. D. Fernie in *Publications of the Astronomical Society of the Pacific,* Vol. 82, pages 1189–1230; 1970.

THE MILKY WAY. Bart J. Bok and Priscilla F. Bok. Harvard University Press, 1974.

MAN DISCOVERS THE GALAXIES. Richard Berendzen, Richard Hart and Daniel Seeley. Science History Publications, 1976.

THE LARGE-SCALE CHARACTERISTICS OF THE GALAXY. Proceedings of the International Astronomical Union, Symposium No. 84, edited by W. B. Burton. D. Reidel Publishing Company, 1979.

THE SPIN OF THE PROTON

PROTON-PROTON SCATTERING AND STRONG INTERACTIONS. Alan D. Krisch in *Physical Review Letters,* Vol. 11, No. 5, pages 217–220; September 1, 1963.

ELASTIC PROTON-PROTON SCATTERING AT 90° AND STRUCTURE WITHIN THE PROTON. C. W. Akerlof, R. H. Hieber, A. D. Krisch, K. W. Edwards, L. G. Ratner and K. Ruddick in *Physical Review,* Vol. 159, No. 5, pages 1138–1149; July, 1967.

ACCELERATION OF POLARIZED PROTONS TO 8.5 GEV/C. T. Khoe, R. L. Kustom, R. L. Martin, E. F. Parker, C. W. Potts, L. G. Ratner, R. E. Timm, A. D. Krisch, J. B. Roberts and J. R. O'Fallon in *Particle Accelerators,* Vol. 6, No. 4, pages 213–236; 1975.

SPIN-SPIN INTERACTIONS IN HIGH $P_\perp^2$ ELASTIC *P-P* SCATTERING. J. R. O'Fallon, L. G. Ratner, P. F. Schultz, K. Abe, R. C. Fernow, A. D. Krisch, T. A. Mulera, A. J. Salthouse, B. Sandler, K. M. Terwilliger, D. G. Crabb and P. H. Hansen in *Physical Review Letters,* Vol. 39, No. 12, pages 733–736; September 19, 1977.

THE MOLECULAR BASIS OF CELL MOVEMENT

ADHESION OF CULTURE CELLS TO THEIR SUBSTRATUM. J. P. Revel, Pamela Hoch and D. Ho in *Experimental Cell Research,* Vol. 84, pages 207–218; 1974.

CELL MOTILITY. Edited by R. Goldman, T. Pollard and J. Rosenbaum. Cold Spring Harbor Conferences on Cell Proliferation, Cold Spring Harbor Laboratory, 1976.

ACTIN, α-ACTININ, AND TROPOMYOSIN INTERACTION IN THE STRUCTURAL ORGANIZATION OF ACTIN FILAMENTS IN NONMUSCLE CELLS. Elias Lazarides in *The Journal of Cell Biology,* Vol. 68, No. 2, pages 202–219; February, 1976.

FLUORESCENT ANTIBODY LOCALIZATION OF MYOSIN IN THE CYTOPLASM, CLEAVAGE FURROW, AND MITOTIC SPINDLE OF HUMAN CELLS. Keigi Fujiwara and Thomas D. Pollard in *The Journal of Cell Biology,* Vol. 71, No. 3, pages 848–875; December, 1976.

LASER CHEMISTRY

INFRARED LASER CATALYZED CHEMICAL REACTIONS. A. M. Ronn in *Spectroscopy Letters,* Vol. 8, No. 5, pages 303–328; 1975.

CHEMICAL AND BIOCHEMICAL APPLICATIONS OF LASERS: VOL. 3. Edited by C. B. Moore. Academic Press, 1977.

SELECTIVE DISSOCIATION OF POLYATOMIC MOLECULES BY INTENSE INFRARED LASER FIELDS. Rafael V. Ambartzumian and Vladilen S. Letokhov in *Accounts of Chemical Research,* Vol. 10, No. 2, pages 61–67; February, 1977.

LASERS AND CHEMISTRY. Sol Kimel and Shammai Speiser in *Chemical Reviews,* Vol. 77, No. 4, pages 437–472; August, 1977.

INFRARED INDUCED UNIMOLECULAR REACTIONS. N. Bloembergen and E. Yablonovitch in *Physics Today,* Vol. 31, No. 5, pages 23–30; May, 1978.

A BRAIN-COOLING SYSTEM IN MAMMALS

STUDIES OF THE CAROTID RETE AND ITS ASSOCIATED ARTERIES. P. M. Daniel, J. D. K. Dawes and Marjorie M. L. Prichard in *Philosophical Transactions of the Royal Society, Series B,* Vol. 237, pages 173–208; 1954.

INFLUENCE OF THE CAROTID RETE ON BRAIN TEMPERATURE IN CATS EXPOSED TO HOT ENVIRONMENTS. M. A. Baker in *Journal of Physiology,* Vol. 220, No. 3, pages 711–728; February, 1972.

RAPID BRAIN COOLING IN EXERCISING DOGS. M. A. Baker and L. W. Chapman in *Science,* Vol. 195, No. 4280, pages 781–783; February 25, 1977.

INTRINSICALLY DIFFICULT PROBLEMS

THE DESIGN AND ANALYSIS OF COMPUTER ALGORITHMS. Alfred V. Aho, John E. Hopcroft and Jeffrey D. Ullman. Addison-Wesley Publishing Co., 1974.

WEAK MONADIC SECOND ORDER THEORY OF SUCCESSOR IS NOT ELEMENTARY-RECURSIVE. Albert R. Meyer in *Lecture Notes in Mathematics: No. 453,* edited by A. Dold and B. Eckmann. Springer-Verlag, 1975.

AN INTRODUCTION TO THE GENERAL THEORY OF ALGORITHMS. Michael Machtey and Paul Young. Elsevier North-Holland Publishing Co., 1978.

GIOTTO'S PORTRAIT OF HALLEY'S COMET

THE NATURE OF COMETS. Nikolaus B. Richter. Methuen & Co Ltd., 1963.

THE COMPLETE PAINTINGS OF GIOTTO. Edi Baccheschi. Weidenfeld & Nicolson, 1969.

GIOTTO: THE ARENA CHAPEL FRESCOES. Edited by James H. Stubblebine. W. W. Norton & Co., Inc., 1969.

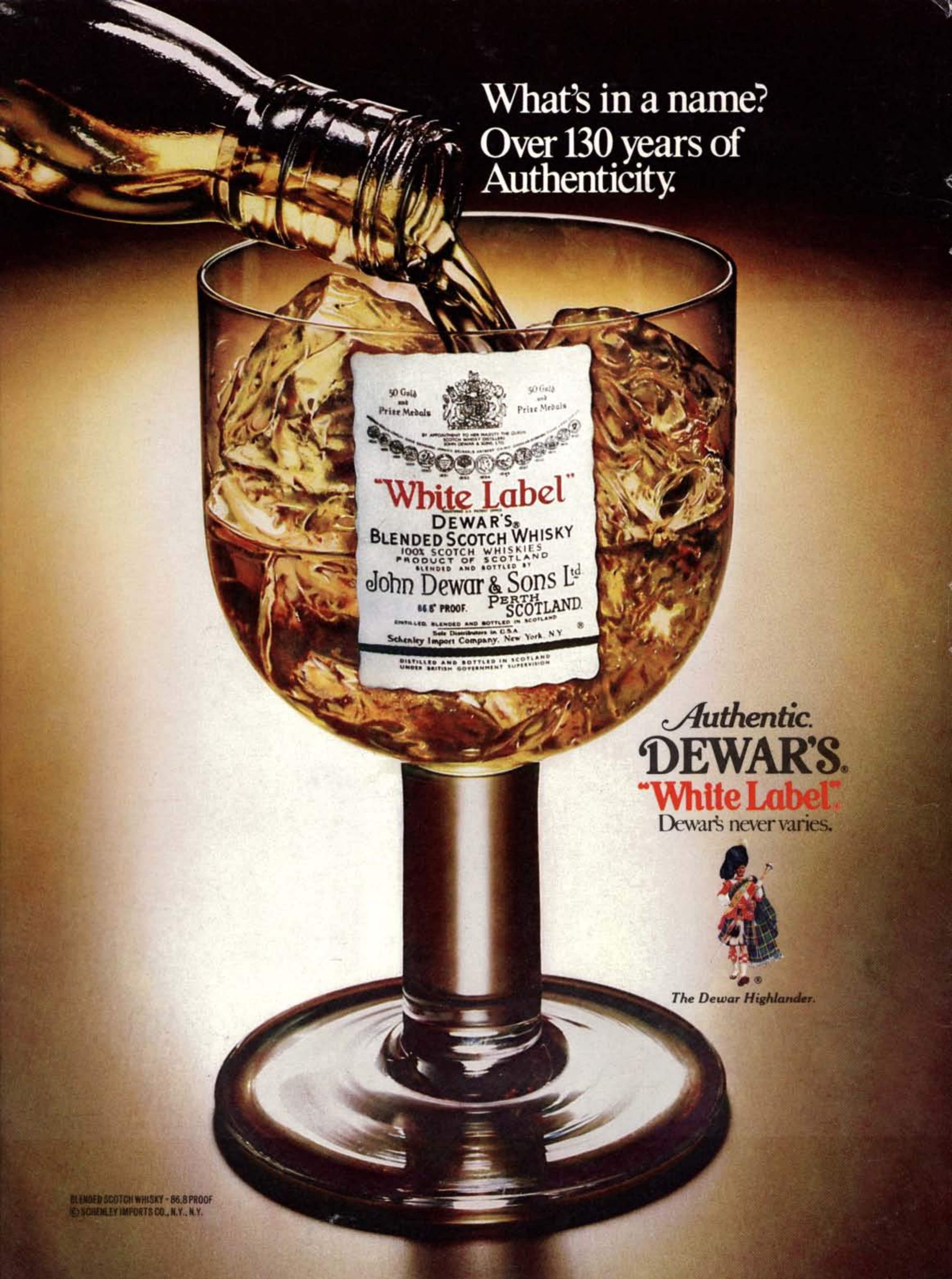
THE AMATEUR SCIENTIST

REFRACTION IN A CLOUD-FREE ATMOSPHERE. Hans Neuberger in *Introduction to Physical Meteorology.* The Pennsylvania State University, 1957.

A COMPENDIUM OF SPHERICAL ASTRONOMY. Simon Newcomb. Dover Publications, Inc., 1960.

DOUBLING YOUR SUNSETS, OR HOW ANYONE CAN MEASURE THE EARTH'S SIZE WITH A WRISTWATCH AND METERSTICK. Dennis Rawlins in *American Journal of Physics,* Vol. 47, No. 2, pages 126–128; February, 1979.


"pretty spectacular"


What's in a name?
Over 130 years of Authenticity.
"White Label"
DEWAR'S
BLENDED SCOTCH WHISKY
100% SCOTCH WHISKIES
PRODUCT OF SCOTLAND
John Dewar & Sons Ltd
PERTH SCOTLAND
Schenley Import Company, New York, N.Y.
Authentic.
DEWAR'S.
"White Label"
Dewar's never varies.
The Dewar Highlander.
BLENDED SCOTCH WHISKY - 86.8 PROOF
© SCHENLEY IMPORTS CO., N.Y., N.Y.