SCIENTIFIC AMERICAN

BUMBLEBEE ENERGETICS

ONE DOLLAR

April 1973

This year we'll plant two trees for every family in America.

Our plan is for more than 100 million seedlings in 1973.

This means an average planting of about 275,000 seedlings a day on Weyerhaeuser lands.

The program is ambitious. And it's expensive. But it's necessary. Particularly if we are to help keep up with your individual demands for wood products.

As a matter of fact, every man, woman and child in the U.S. will use the equivalent of a 100-foot tree this year, and every year.

But forests do more than provide wood. They add beauty. Enrich the ecology for all living things. And provide land for recreation.

Right now we have 5.7 million acres in all stages of growth. Some are mature trees ready for harvest this year. Others are for the future.

This is another way we keep our forests green and growing. Forever.

For a free booklet on Weyerhaeuser High Yield Forestry, write us at Box P-10-1, Tacoma, Wash. 98401.

Weyerhaeuser

Established 1845 SCIENTIFIC AMERICAN April 1973 Volume 228 Number 4

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. COPYRIGHT © 1973 BY SCIENTIFIC AMERICAN, INC. ALL RIGHTS RESERVED. PRINTED IN THE U.S.A. NO PART OF THIS ISSUE MAY BE REPRODUCED BY ANY MECHANICAL, PHOTOGRAPHIC OR ELECTRONIC PROCESS, OR IN THE FORM OF A PHONOGRAPHIC RECORDING, NOR MAY IT BE STORED IN A RETRIEVAL SYSTEM, TRANSMITTED OR OTHERWISE COPIED FOR PUBLIC OR PRIVATE USE WITHOUT WRITTEN PERMISSION OF THE PUBLISHER. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $10 PER YEAR, U.S., POSSESSIONS AND CANADA; $12.50 PER YEAR, ALL OTHER COUNTRIES. EUROPEAN SUBSCRIPTION COPIES MAILED c/o INTERNATIONAL DISTRIBUTORS, KOEPOORTBRUG 1, B 2000 ANTWERP, BELGIUM.

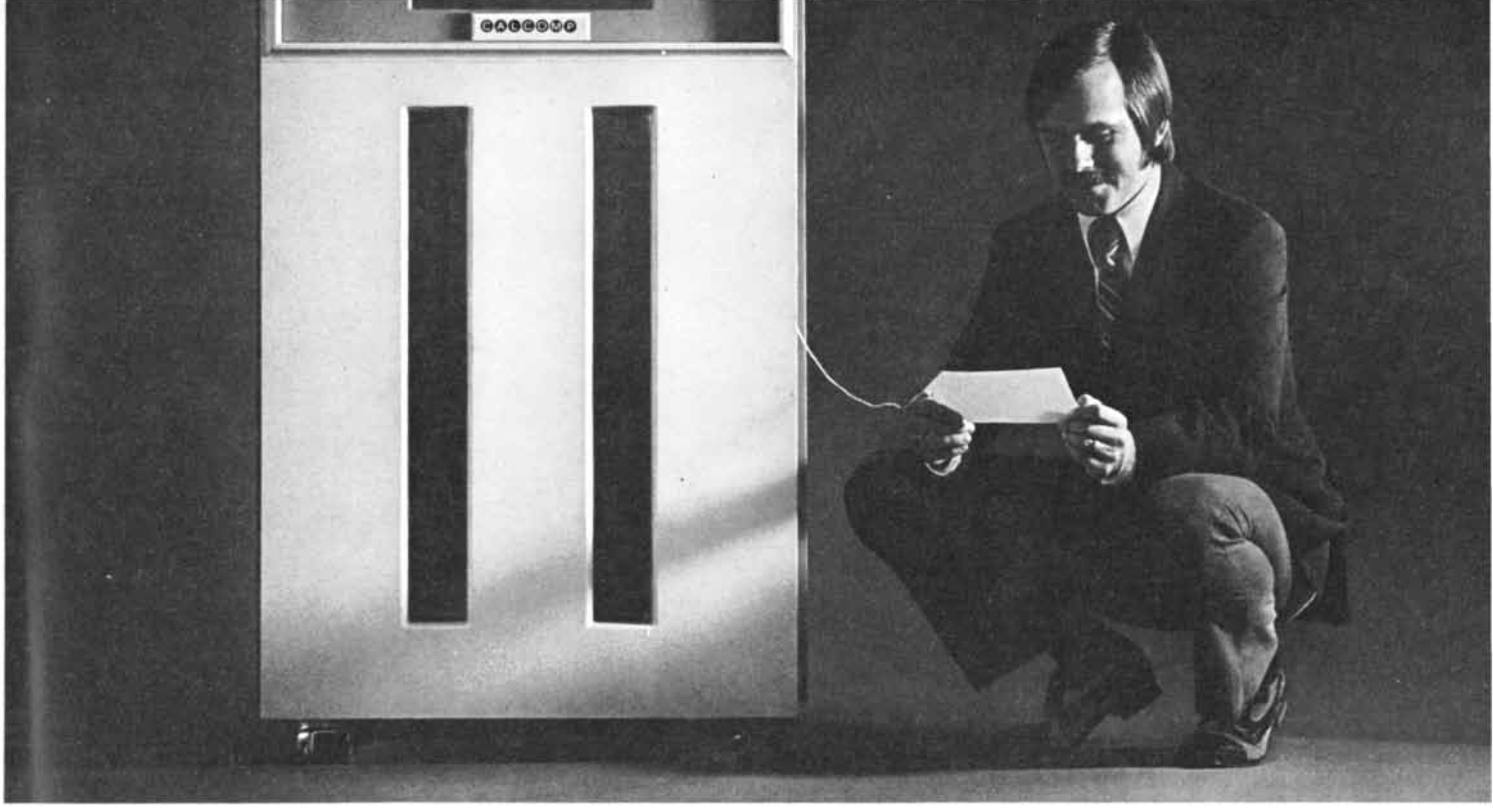
THE COVER

The painting on the cover shows a native American bumblebee, *Bombus ternarius,* perched on a bloom of the familiar yellow weed the dandelion, a plant native to Europe that has widely colonized North America. For millions of years insects and flowering plants have evolved in intimate union; many of the plants depend on bees for cross-pollination and have evolved scents and colors to attract them. Of all insect pollinators the bumblebees are the hardest workers. By generating enough heat to remain active they can forage at temperatures so low that other insects are immobilized (see "The Energetics of the Bumblebee," by Bernd Heinrich, page 96). Dandelions are not pollinated by insects but this American bee has been attracted to the alien plant because its bright yellow resembles the color of a related native plant that the bee does pollinate (*see illustration on page 96*).

THE ILLUSTRATIONS

Cover painting by Peter Parnall

Page	Source	Page	Source
13–17	Ilil Arbel	50–53	Gabor Kiss
18	Michiko Mitani, University of Texas at Dallas	54–59	Eric Mose
20–25	Bunji Tagawa	61	EROS Data Center, Geological Survey, U.S. Department of the Interior
26	Christine Milliken, University of Texas at Dallas	64	F. W. Goro
27	Bunji Tagawa	66–68	Jim Egleson
28–29	Cerro Tololo InterAmerican Observatory	69	F. W. Goro
30–34	Allen Beechel	70	Jim Egleson
35	Lick Observatory (*top*), Allen Beechel (*bottom*)	71	Frederick F. Morehead, Jr., and Billy L. Crowder, International Business Machines Corporation
36	Hale Observatories	74	Nai-San Wang
37	Allen Beechel	76–79	Carol Woike
38	Hale Observatories	80–82	Nai-San Wang
39–40	Allen Beechel	87–92	Harvard College Library
46–47	EROS Data Center, Geological Survey, U.S. Department of the Interior	96	Bernd Heinrich
48	Eric Mose	98–99	Tom Prentiss
49	Ilil Arbel	100–111	Ilil Arbel
		112–116	Roger Hayward



Buy your own tape drive.

Can you afford to? Can you afford <u>not</u> to?

We have a set of figures worked out by one of our accountants that shows you a savings of 20% over a three year period if you *buy* instead of lease.

But your figures will be different from his. They always are.

Here's what you positively *can* afford to do. You can call (714) 821-2011 collect, and talk to Rod Murray. Get together with him or one of his people and your accountant.

Find out if you really can save something by buying instead of leasing.

But if you prefer to lease, check to see if you're leasing from us. If not, you're still spending too much.

Call your local CalComp office, or contact California Computer Products, Inc., SA-M4-73, 2411 West La Palma Avenue, Anaheim, California 92801. (714) 821-2011.

CALCOMP

NT ZERO "S."

performance, the responsiveness, the unmatched roadholding, that make BMW the most exciting sports sedan you can drive.

If you really appreciate fine machinery, you owe it to yourself to visit your BMW dealer and test drive the new 3.0 S. You'll say the S stands for *spectacular.*

BMW's are available in four or six-cylinder models. U.S. Sole Importer: **Hoffman Motors Corporation.** 375 Park Avenue, New York, N.Y. 10022; 12541 Beatrice Street, Los Angeles, California 90066. Dealers Coast to Coast. Advantageous European Delivery Program. Write for free brochure.

Canada: East: BMW Distributors Eastern Canada, Ltd. 801 Progress Avenue, Scarborough, Ontario. West: BMW Distributors (Western) Co. 2040 Burrard Street, Vancouver 9, B.C.

Bavarian Motor Works

BMW

LETTERS

Sirs:

The article by Derek P. Gregory titled "The Hydrogen Economy" [SCIENTIFIC AMERICAN, January] literally begs for comment to put it in perspective. The central message of the article seems to be that if we replaced our fast-disappearing natural-gas system with a system based on hydrogen, we would create a source of energy cheaper than electricity. This is just not so. Hydrogen cannot possibly compare with electricity costwise, since at a minimum one extra step, namely the electrolyzing step, is required with inherent efficiency losses to produce it. Second, our studies show that distribution costs of electricity are less, not more, than those that would be inherent in piping hydrogen, except for distances in excess of 500 miles, although this is a relatively unimportant point.

The true test comes in the cost of each useful unit of work output. Here electricity has a clear advantage over hydrogen in virtually every conceivable use. In heating, gas furnaces have operating efficiencies of between 40 and 60 percent (although laboratory tests will yield 75 percent after three hours of warm-up, which is hardly the case in a home). Electric-resistance heating is virtually 100 percent efficient, and heat pumps boast apparent efficiencies of more than 200 percent. A similar comparison is valid for every other heating use of gas. In cars, electric vehicles are some 60 percent efficient from input to the battery charger to useful force at the rear wheels, whereas a typical internal-combustion engine is some 10 percent efficient and, in the case of hydrogen, requires liquefaction, which is some 50 to 60 percent efficient. This means that an electric car requires some 5.3 units of nuclear energy to create one unit of driving force at the car wheels, whereas a liquid-hydrogen-powered car requires some 75 units of input to deliver the same unit of useful work! Today it is the cheapness of gasoline that overrides the higher efficiency of electricity. The hydrogen economy would obviate this advantage.

Thus relying too heavily on a hydrogen economy could bring catastrophic wastes of our precious nuclear fuel. In an era of energy crisis this seems ludicrous. There appears to be no logic for using electricity to create hydrogen except on a very limited basis for end uses where form is highly critical, such as jet aircraft and heavy trucks. Even here substantial engineering problems that make these applications impossible today must be overcome. In virtually all other applications there are cheaper and more efficient ways to deliver the energy. These should be stressed rather than novel ways to use archaic energy-conversion devices.

J. HUNTER CHILES III

Power Systems Planning
Westinghouse Electric Corporation
East Pittsburgh, Pa.

Sirs:

I think Mr. Chiles is quite correct in putting the hydrogen economy into perspective. In my opinion our basic problem is to find the best and fastest way of using nonfossil energy to satisfy our increasing energy demands, which cannot now be met by coal, oil and gas. There seem to be only three alternatives: the all-electric economy, synthesis of conventional hydrocarbon fuels from carbon dioxide and hydrogen, and the direct use of hydrogen. We shall most likely develop some combination of all three.

I believe it may be possible to develop a process for the conversion of nuclear or solar heat into chemical (hydrogen) energy with about the same efficiency as we now obtain in power stations. This may come about either from an integrated nuclear-electrolysis plant or from a thermochemical process. Our studies, based on natural-gas pipeline and distribution experience, indicate a very considerable cost advantage in moving hydrogen, particularly if the electricity alternative is forced underground. On this point Mr. Chiles and I will have to disagree, but as I pointed out in my article, it appears to me to be likely that we could manufacture and deliver hydrogen energy to the average consumer at a cost below that of delivering electricity, because the savings in transmission and distribution would be greater than the extra cost of turning electricity into hydrogen.

I like to think that a hydrogen space-heating system could operate without a flue, and could thus be virtually 100 percent efficient, comparing well with electric systems. Many industrial uses of energy could apparently be supplied by hydrogen at very high efficiencies. On the other hand, Mr. Chiles's figures for automobile efficiencies are indeed dramatic, and even the significant increases expected in hydrogen-engine efficiencies fail to justify the hydrogen car. However, the technical problems facing the development of a high-performance, cheap, fast-charging, long-range electric vehicle are formidable compared with the use of the "archaic" gasoline engine, suitably modified.

My understanding of the nuclear-fuel supply is that, in the absence of the breeder reactor, our nuclear fuels are indeed "precious," and this gives cause for alarm regardless of the way we use them. I hope that the breeder reactor and the more distant prospects of the fusion reactor and solar power will bring us into an era of more abundant and less expensive fuel, where capital costs rather than fuel-supply limitations will dominate. I am sure that in the future we shall all be in such a scramble for clean energy in any form that there will be ample room for both of our approaches.

DEREK P. GREGORY

Institute of Gas Technology
Chicago

Sirs:

I do not wish to deride Robert L. Birch's detecting that Mount Chimborazo is the highest mountain because its peak is the farthest from the center of the earth ["Science and the Citizen," February]. However, my favorite author, Isaac Asimov, noted the same thing in "Up and Down the Earth," an article in the February 1966 issue of *The Magazine of Fantasy and Science Fiction.*

ROBERT LINDEN

Tacoma, Wash.

Sirs:

. . . the calculation was done by O. M. Miller of the American Geographical Society in 1968. . . .

ROSTISLAV DIDKOWSKY

New York

Sirs:

Referring to your "Science and the Citizen" note, your reader Robert L. Birch must also have been a reader of the June 1972 issue of *Explorers Journal,* wherein Terris Moore related the highest-mountain proposition. The idea is in fact a very old one, having been presented by Dr. Moore in various writings since the early 1930's.

ELLIOTT ROBERTS

Explorers Journal
Washington

Sirs:

The mountainous "Fact" presented in February's "Science and the Citizen" is so up in the air that it calls forth the thinking man's pique. The hugest hill in the world is obviously neither Everest nor Chimborazo but Mauna Loa. For the only physically impressive measure of a mount is its immediate prominence. The Hawaiian high rises more than 33,000 feet from an adjacent ocean deep.

Of course, your search for superlatives did not call for the highest mountain in the world but just the highest mountain. That is Nix Olympica on Mars, which rises perhaps 100,000 feet from its surroundings.

JAMES ENTERLINE

New York

Scientific American, April, 1973; Vol. 228, No. 4. 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, treasurer; 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 Harry T. Morris, 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 25 cents.

Change of address (or other subscription correspondence) should be addressed to Circulation Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. Please notify us four weeks in advance of change. If possible, kindly furnish an address imprint from a recent issue. Send both old and new addresses or complete and mail Post Office Form 3578.

☐ Please enter change of address.

☐ Please enter subscription.
U.S. and Canada:
☐ 1 year, $10 ☐ 2 years, $18 ☐ 3 years, $25
All other countries:
☐ 1 year, $12.50 ☐ 2 years, $23 ☐ 3 years, $32

NAME

NEW ADDRESS

OLD ADDRESS

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

APRIL, 1923: "We shall remember the day in 1895 when the news was cabled over that a German professor had made a discovery by which it was possible to see the bones in human beings and animals and even to photograph the same. The man who made this epoch-making discovery that has saved thousands of lives and mitigated untold suffering recently died in Munich at the age of 77. Wilhelm Conrad Roentgen was born in Lennep, Prussia, in 1845 and studied in Holland and at Zurich, where he took his doctor's degree in 1869. He became a professor in physics at several German universities and finally went to Würzburg in 1885, where 10 years later he made the discovery by which he will always be remembered—the Roentgen rays, commonly known as X-rays. The new science progressed by leaps and bounds, so that in 1901 a Roentgen Ray Congress was held in Paris and he received the Nobel prize for physics. Although the X-rays are used mostly in medicine and dentistry, industrial uses were quickly found for the testing and analysis of all kinds of material and for the detection of smuggled articles."

"Edwin Hubble at Mount Wilson has lately published papers dealing with nebulae of the galactic type. The diffuse or irregular nebulae, which are confined to the Milky Way, do not all show a gaseous spectrum. About half of them do, but the spectra of the rest are continuous and show dark lines, just as they should do if the nebula was not self-luminous but was an opaque cloud reflecting the light of neighboring stars. In most cases the stars from which the light comes are easy to find, and the similarity of the spectra of such stars and of the surrounding nebula clinches the argument. What is more remarkable, the nebulae that have bright-line spectra are also associated with stars. This suggests very strongly that it is the stars that set the nebulae shining, whether simply by the reflection of their own light or by some other form of influence, which, emanating from the hottest stars alone, is capable of energizing the gases of the nebulae and making them glow. In some cases the influence cannot be ordinary light; it may be ultraviolet light of very short wavelength or electrons emitted at high speed, either or both of which we might expect to find given off from the very hottest stars."

"We have just heard of the latest improved method of causing an airplane to take off and alight at low speed, while at the same time making it possible for the airplane to maintain the maximum speed in flight. The manner in which this extreme difference in speed is accomplished is by varying the surface of the wings by withdrawing the rear part of the wings from the front part to increase the surface, and sliding the rear part back into the front part, into which it telescopes, when the airplane is in flight and it is desired to attain the maximum speed. The altering of the surface of the wings is accomplished by the aviator turning a crank within the body, which he can do very readily. This new improvement in airplane construction is the invention of a Frenchman, M. Bille, and it seems to be a far neater and better solution than that of the variable angle of incidence devised by Hollander Lanzius, or even that of the slotted wing, invented independently by Mr. Handley Page and by Herr Lachmann."

SCIENTIFIC AMERICAN

APRIL, 1873: "We are at present in such a stage of the development of the industry of all civilized nations that the increase in producing capacity far outstrips the increase of population, so that the amount produced and consumed on the average by every person far exceeds in quantity and value that which was ever before known. If we look at the statistics of the increase of productive capacity in various branches among different nations, we are especially struck at the development that has taken place during the past decade. For instance, let us take the production of steel manufacture in Germany; in 1860 only 250 tuns of manufactured steel, worth $3,000,000, was produced by 4,000 workmen, while in 1870 2,000 tuns, worth $20,000,000, was the result of 14,000 workmen. This view of productive capacity and its results is the best argument against that conservative class of people who sometimes raise their voices against useful inventions and new patents, under the pretext that such improvements often take the bread out of the mouth of the workmen, who are unable to compete with hand labor against machine labor. Experience has proved that all such fears are totally groundless, and in every case have the machines which increased production been a blessing in the end, giving more labor and higher wages to those using them than they could obtain by their unimproved methods and much smaller productive capacities."

"M. Fizeau communicates to *Le Monde* the results of a series of very elaborate experiments made with a view toward the most accurate determination of the velocity of light. The source of the ray was a jet of burning oxyhydric gas, and the distance between the two stations, as found by careful triangulation, was 33827.1 feet, with a probable error of a tenth of 1 per cent. Six hundred and fifty satisfactory observations were made, the mean of which, multiplied by the index of refraction, 1.003, gives 185,368 miles per second as the velocity of light, with a probable error of three tenths of 1 per cent. This result agrees with that determined previously by Foucault. M. Fizeau considers that, with stations separated a distance of 12 miles, the velocity of light could be determined to within a tenth of 1 per cent."

"There is an extensive demand for small portable steamers for farm use, with attachments to render them applicable to various purposes. For example, such machines should be capable of use in working mowers and reapers in lieu of horses. The application of steam to farm labor is at present in its infancy. One of the first requisites for its general introduction will be the instruction of farmers in the principles and economies of the steam engine, studies that are at present almost wholly neglected in our common schools."

"The bill for an underground railway beneath the great thoroughfare of New York City, known as Broadway, has finally passed both branches of the State Legislature, has received the Governor's signature and has become a law. The wonder is, in a community like New York City, so noted for the number of its intelligent, active and vigorous men, that such an important enterprise should have been so long postponed. No city in the world has more pressingly needed facilities for rapid transit."

THE AUTHORS

WILLIAM H. GLAZIER ("The Task of Medicine") is associate professor of community health and acting chairman of the department of community health at the Albert Einstein College of Medicine of Yeshiva University. Following his graduation from Harvard College in 1939 and military service during World War II he spent the period from 1946 to 1962 with the International Longshoremen's and Warehousemen's Union, first as legislative representative in Washington and then as administrative officer. From 1962 to 1968 he was at the Salk Institute for Biological Studies, serving as assistant to the director from 1962 to 1965 and assistant to the president for academic and scientific affairs from 1965 to 1968. Toward the end of that period he also acted as executive director of the New York City Commission on the Delivery of Personal Health Services. Glazier notes that he is "a dry-fly trout fisherman and a chamber-music lover (Mozart and Schubert)."

ROYSTON C. CLOWES ("The Molecule of Infectious Drug Resistance") is professor of biology and chairman of the department of biology at the University of Texas at Dallas. Born in Wales, he took his undergraduate degree in chemistry at the University of Birmingham in 1942 and then spent five years in the British army. Returning to Birmingham, he received his Ph.D. from the medical school in 1951; in 1965 the university awarded him a D.Sc. in genetics. For several years he did research in microbiology at St. Mary's Hospital in London and on microbial genetics at Hammersmith Hospital. After a period of time as a visiting professor at the University of California at Berkeley he took up his present position. Clowes writes that his major hobbies are skin diving, reading, gardening "and going to the theater—live, not movies."

G. NEUGEBAUER and ERIC E. BECKLIN ("The Brightest Infrared Sources") are at the California Institute of Technology; Neugebauer is professor of physics and Becklin is senior research fellow in physics. Both are also associated with the Hale Observatories: Neugebauer is a staff member and Becklin a staff associate. Neugebauer was graduated from Cornell University and obtained his Ph.D. from Cal Tech in 1960. He has been involved with the *Mariner 2* probe of Venus and three of the Mariners making radiometric measurements of the surface of Mars and is now working on future probes to Mars, Jupiter and Mercury. Becklin was graduated from the University of Minnesota in 1963 and obtained his Ph.D. from Cal Tech in 1968.

JOSÉ P. PEIXOTO and M. ALI KETTANI ("The Control of the Water Cycle") are respectively professor of physics at the University of Lisbon and associate professor at the University of Petroleum and Minerals in Saudi Arabia. Peixoto, who is also director of the Geophysical Institute at the University of Lisbon, received part of his education in Portugal (he holds a D.Sc. in physics from the University of Lisbon) and part at the Massachusetts Institute of Technology. He maintains an association with the department of meteorology at M.I.T. Kettani, who was born in Morocco, received his higher education in Switzerland and the U.S. He obtained his Ph.D. in electrical engineering from Carnegie-Mellon University. He writes that he originated "the idea of a heliohydroelectric power system that would use solar energy to produce electricity on a large scale after its conversion to hydraulic energy."

FREDERICK F. MOREHEAD, JR., and BILLY L. CROWDER ("Ion Implantation") are on the staff of the Thomas J. Watson Research Center of the International Business Machines Corporation. Morehead was graduated from Swarthmore College in 1950 and received his master's degree and his Ph.D. in physical chemistry from the University of Wisconsin in 1951 and 1953 respectively. He taught chemistry at Union College for a year and did research on powder electroluminescence for the General Electric Company for five years before joining IBM. "My outside interests," he writes, "include activities in local government and education and an attempt to understand (and change) the interaction of housing, property taxes and public education." Crowder, who was graduated from Duke University in 1957 and received his Ph.D. (in physical chemistry) from Cornell University in 1962, joined IBM in 1964 after two years of research work with the Union Carbide Corporation.

MARY ELLEN AVERY, NAI-SAN WANG and H. WILLIAM TAEUSCH, JR. ("The Lung of the Newborn Infant"), are at McGill University: Dr. Avery is professor of pediatrics and chairman of the department of pediatrics in the faculty of medicine (and physician-in-chief of the Montreal Children's Hospital); Wang is assistant professor of pathology and Taeusch is assistant professor of pediatrics. Dr. Avery was graduated from Wheaton College in 1948 and received her M.D. from the Johns Hopkins University School of Medicine in 1952. She is the author of a book, *The Lung and Its Disorders in the Newborn Infant,* and coauthor (with A. Schaffer) of *Diseases of the Newborn.* She writes that her principal avocation is "to fish and swim in a lake in Maine," where she spends part of each summer. Wang received his M.D. from the National Taiwan University in 1960 and his Ph.D. (in pathology) from McGill in 1971. He writes that he plays tennis and enjoys watching basketball and football. Taeusch majored in English literature at Harvard College, from which he was graduated in 1961, and obtained his M.D. at Case Western Reserve University in 1965.

LAWRENCE S. LERNER and EDWARD A. GOSSELIN ("Giordano Bruno") are at California State University at Long Beach: Lerner is associate professor in the department of physics and astronomy and Gosselin is assistant professor of history. Lerner's degrees are from the University of Chicago; he obtained his bachelor's degree in 1953 and his Ph.D. (in solid-state physics) in 1962. He writes of his efforts to encourage women to take up work in the physical sciences "in the hope that the social stereotypes and injustices that have so effectively excluded women from a deeply satisfying, reasonably lucrative and fairly prestigious profession can be substantially overcome well within my lifetime." Gosselin was graduated from Yale College in 1965, received his master's degree at Columbia University in 1966 and is scheduled to receive his Ph.D. from Columbia this year. He writes that in pursuit of his main outside interest, which is music, he is active as a clarinetist.

BERND HEINRICH ("The Energetics of the Bumblebee") is assistant professor of entomology at the University of California at Berkeley. He came to the U.S. from Germany as a boy and grew up on a farm in Maine, where he spent much time in outdoor activity. At the University of Maine he began studying forestry but was graduated in zoology. In 1970 he obtained his Ph.D. from the University of California at Los Angeles.

SCIENCE/SCOPE

A domestic satellite system for Brazil which would provide substantially greater telephone, teletype, TV, educational TV, and commercial broadcast communications services has been proposed by Hughes. The satellite antenna's beam would be shaped to serve vast regions whose populace is separated by dense jungle and mountainous terrain. International communications traffic with Western Europe, Canada, Africa, and the U.S. has more than doubled in the four years since Hughes designed and built Brazil's first satellite ground station at Itaborai, near Rio de Janeiro.

A corona pulse counter for high-voltage components, developed by Hughes, provides a far more sensitive method of measuring corona intensity than previously achieved. It makes quantitative measurements of the corona spectrum under DC or AC voltage possible for the first time. Corona, commonly observed as the faint glow around the surface of a high-voltage conductor in air, is the most important cause of internal degradation of transformers, capacitors, and cables. By accurately measuring the corona discharge energy, manufacturers will be able to improve the energy density, reliability, and lifetime of high-voltage components.

A new data terminal display console developed by Hughes to monitor and control the Naval Tactical Data System (NTDS) computers was delivered to the U.S. Navy recently. Designed to make functional changes in NTDS equipment assignments as required, the Monitor Control Console (MCC) combines a data terminal with a CRT display and typewriter keyboard. It is one of more than 1700 varying types of NTDS consoles built by Hughes. The MCC is the first to incorporate microprocessing techniques made possible by MSI/LSI technology. NTDS uses radars, computers, and communications equipment to gather, process, and exhibit action within tactical combat zones instantaneously on the display consoles.

A prototype series of portable ground stations which Hughes is developing for the U.S. Dept. of Health, Education and Welfare will be used in experiments designed to relay TV for medical assistance and educational programs via satellite to isolated areas of Alaska and the United States. The 300 lb. stations, using 10-foot-diameter dish antennas, can be carried by small van or helicopter and easily erected by two men. They are forerunners of low-cost stations that could be used for television distribution and two-way telephony by developing and emerging nations.

Hughes has immediate openings for Field Engineers. Qualifications include U.S. citizenship, BSEE degree, willingness to travel, and experience in any one of the following systems: electro-optical, infrared detection, laser ranging and target designation, or low-level-light TV detection. Please write: Professional Staffing, Hughes Aircraft Company, Field Service & Support Division, P.O. Box 90515, Los Angeles, CA 90009. Hughes is an equal opportunity M/F employer.

Quicker turn-arounds for jetliners and substantial improvements in aircraft and manpower utilization are being achieved by Eastern Airlines on aircraft circuit tests made with a Hughes-built Flexible Automatic Circuit Tester (FACT). Based at Miami International Airport, Eastern's FACT system is housed in an air-conditioned van which rolls up to the jetliner, where technicians attach cables from the aircraft's connectors to the six portable remotely-operated FACT switch modules. Tests are performed automatically by computer-generated punched-tape programs.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY

Established 1845 SCIENTIFIC AMERICAN April 1973 Volume 228 Number 4

The Task of Medicine

American medicine is traditionally geared to respond to acute illness. Now that chronic illness plays a much larger role, it will be necessary to intervene more actively in its causes

by William H. Glazier

The medical system of the U.S. is able to meet with high efficiency the kind of medical problem that was dominant until about 40 years ago, namely infectious disease. It also deals effectively with episodes of acute illness and with accidents that call for advanced, hospital-based biomedical knowledge and technology. The system is much less effective in delivering the kind of care that is more often needed today: primary (first-contact) care and the kind of care needed at a time when chronic illnesses predominate. They are the degenerative diseases associated with aging and the diseases that can be characterized as man-made because they are associated with such things as smoking and environmental contaminants. For these diseases medicine has few measures and not even much comfort.

My colleagues and I in the department of community health at the Albert Einstein College of Medicine of Yeshiva University have been much concerned with this shortcoming of the medical system. Since the college is in a community (the borough of the Bronx in New York City) with large numbers of people who are poor, elderly, Puerto Rican or black, we see the deficiencies regularly. When these people seek medical help (and part of the weakness of the present medical system is that it deals only with people who present themselves as sick), they are likely to find themselves very sick indeed and in a long-term regime of treatment involving costs that they cannot possibly meet out of their own resources. The problem, which is by no means limited to the Bronx, is to fashion a medical system that deals as well with the chronic diseases as the present system deals with the infectious diseases and the complicated hospital cases and that handles the costs in a way that is acceptable not only to the patient but also to the society at large.

Until as recently as the beginning of the present century the expectation of life was short in the U.S., largely because of the toll exacted by the infectious diseases, particularly among children. The first advances against the infectious diseases resulted from a series of developments in which medical technology and the work of physicians with individual patients played a relatively minor role. The more important factors were the improvement in the standard of living and the effectiveness of public health measures. Because of the rise

DEATH RATE in the U.S. declined steadily until about 1950 and then leveled off. Infectious diseases, which took a heavy toll among the young, are much reduced as a cause of death, whereas the chronic diseases that are associated with aging have become a major cause.

in the standard of living, people had a more balanced diet, received better education (so that they knew more about taking care of themselves and how to avail themselves of medical help) and had more money to spend on medical care if they needed it. The decline of tuberculosis, for example, is believed to be largely attributable to improved diet.

Meanwhile governments at all levels were moving to purify public water supplies, handle sewage in such a way that it would not be a contaminant, improve housing conditions, drain mosquito-breeding swamps and provide education in health care, particularly for mothers and children. Efforts of this kind led to the decline of such infectious diseases as typhus, typhoid, cholera and malaria.

The role of medicine during this period cannot be dismissed as insignificant. The conquest of smallpox, for example, surely resulted from the discovery of an effective vaccine and the provision of vaccination as a medical procedure. Nonetheless, it was as recently as 1915 that Lawrence J. Henderson of the Harvard Medical School could justifiably make his celebrated remark that the average patient had no better than a 50-50 chance of benefiting from an encounter with the average physician.

The beginning of what might be described as the "golden age" of clinical medicine in the U.S. can be put at about 1922, when the discovery of insulin made possible the control of diabetes. Clinical medicine came into its own as a force against infectious disease with the discovery of the sulfa drugs in 1932. It was in that period too that the effects of earlier advances in physiology and bacteriology and the improvement of medical education resulting from the Flexner report of 1910 began to tell.

The great change in the nature of the health problem during this century is illustrated by a comparison of the 10 leading causes of death in 1900 and 1970 [*see illustration on opposite page*]. In 1900 three of the 10 (tuberculosis, influenza-pneumonia and diphtheria) were directly infectious and three more (gastroenteritis, chronic nephritis and diseases of early infancy) were closely related to infectious processes. By 1970 none of the first 10 causes of death was an infectious disease except for influenza-pneumonia and certain diseases of early infancy, and in both of these groups the mortality rate was far below the level of 1900.

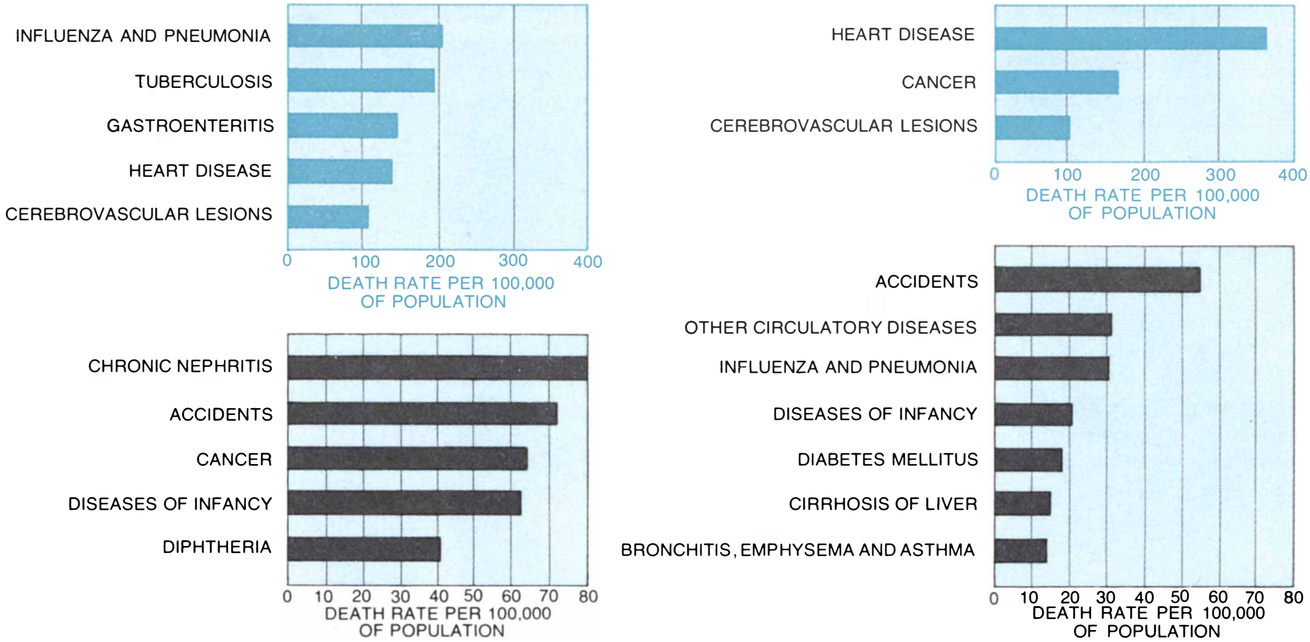
Today the list is headed by heart disease, cancer and cerebrovascular lesions—all chronic diseases. The first two have increased by 268 percent and 240 percent respectively since 1900 in terms of deaths per 100,000 people. Other chronic diseases, such as general arteriosclerosis and diabetes, have emerged as leading causes of death. Of every 100 males born in the U.S. this year, 83 are likely to die eventually of a chronic disease; in 1901 the rate was 52 in 100. The likelihood of dying of an infectious disease is now about six in 100, which is about one-sixth of the rate in 1901.

The death rates tell only part of the story, since the chronic diseases also afflict large numbers of the living for long periods of time. According to estimates made in 1968, heart and circulatory disorders afflicted 26.2 million people in the U.S., mental and emotional disorders 20 million and arthritis and rheumatic diseases 16 million. The inescapable legacy of improved health in early and middle life is the increased prevalence of these less tractable forms of disease and disability in middle and later life.

The infectious diseases have a fairly abrupt onset and a finite duration. In the medical structure that has evolved to deal with them the person who feels sick visits or calls in a physician, who prescribes and administers a treatment that nowadays usually ends fairly soon in recovery. The chronic diseases, on the other hand, typically have a gradual onset and an indefinite duration. A patient with a disease such as cancer, arteriosclerosis or emphysema may go about for months or years before he realizes that he is afflicted. Indeed, the case of the person who is unaware that anything is wrong until he is struck down by a heart attack has become common.

The medical system encounters several different types of problem in dealing with the chronically ill. In the first

CHANGING ROLE of infectious diseases (*black*) and chronic diseases (*color*) as causes of death in the U.S. during the period from 1900 to 1970 is portrayed. Chronic diseases include both degenerative ailments associated with aging and diseases that can be called man-made because they are associated with such behavior as smoking and with such phenomena as environmental contamination.



MAJOR CAUSES of death in the U.S. in 1900 and 1970 are charted. Each list contains the 10 principal causes of death, and the bars show the death rate associated with the disease. In 1900 (*left*) six of the diseases were either infectious or related to infectious processes whereas in 1970 (*right*) none of diseases listed were infectious except influenza-pneumonia and certain diseases of early infancy.

place, it is essentially a passive system, that is, it does not go into operation until a patient takes the initiative by visiting a physician or a clinic. Often by the time a patient with a chronic illness takes this step it is late in the progress of the disease. For many of the chronic diseases much of the treatment is directed to symptoms rather than being curative. The regime of treatment is also likely to be protracted and costly. Another type of problem is that the system is geared to the one-to-one, episodic relationship in which the patient sees a physician, receives treatment and pays a fee. The system is unwieldy and inefficient when, as is often the case with chronic disease, the patient requires care by several physicians with different specialties, by other professional people such as nurses, therapists and social workers and by different institutions. Finally, the system is in a better position to take care of the patient who is so incapacitated that he has to be in a hospital bed than the patient who is ill but more or less able to go about his normal business—and such patients constitute about 85 percent of the total.

In sum, the technology and medical procedures available for coping with the diseases that afflict the population have been expanded and improved, but the structure of the system that deploys the technology and resources has tended to remain fixed in a mold determined by medical and social circumstances that are quite different from those that exist today. The result is a mismatch between the technology of medicine and the apparatus that delivers care. Many observers see this mismatch as the cause of the crisis in medical practice, medical education and health care in the U.S. today.

It should be emphasized that the crisis does not arise from any shortage of medical personnel or the other resources employed in providing medical care. On the contrary, the people and the facilities (hospital beds, sophisticated machines, nursing homes and so on) have increased in number over the past decade, both absolutely and per capita, although the distribution of the components of the system is conspicuously uneven. The nation suffers from a misuse (sometimes an overuse) and a maldistribution of medical technology and medical services. Indeed, the root of the crisis is the growing social demand for a more effective and fair distribution of the medical potential created by the scientific and technological advances of the past 50 years. A related problem is the growing lack of primary care, which is what used to be provided by the general practitioners whose number is dwindling steadily as physicians turn increasingly to specialization. General practitioners also used to help patients adjust to the increasing disability of chronic, degenerative disease.

The crisis is articulated mainly in dollar terms, which is understandable in view of the enormous flow of public funds into the health care industry during the past few years. For example, both the demand for medical care and the cost of it have been heavily affected by the 1965 amendments to the Social Security Act, which created Medicare and Medicaid. (Medicare is for the aged, Medicaid for the poor; Federal funds are employed to help pay for the medical care of both groups.) Five years after the inauguration of these programs the Federal outlay for personal health care had risen above $14 billion per year from a starting level of about $3 billion.

In fiscal 1971 the amount of money spent in the U.S. for all types of medical care was $75 billion. At $358 per person, this sum represented 7.4 percent of the gross national product. Five years earlier medical expenditures were $212 per person, or a bit less than 6 percent of the gross national product. One can therefore see why concern over what is being delivered for the money is rising, particularly in the light of evidence that the delivery system is not as efficient or equitable as it might be.

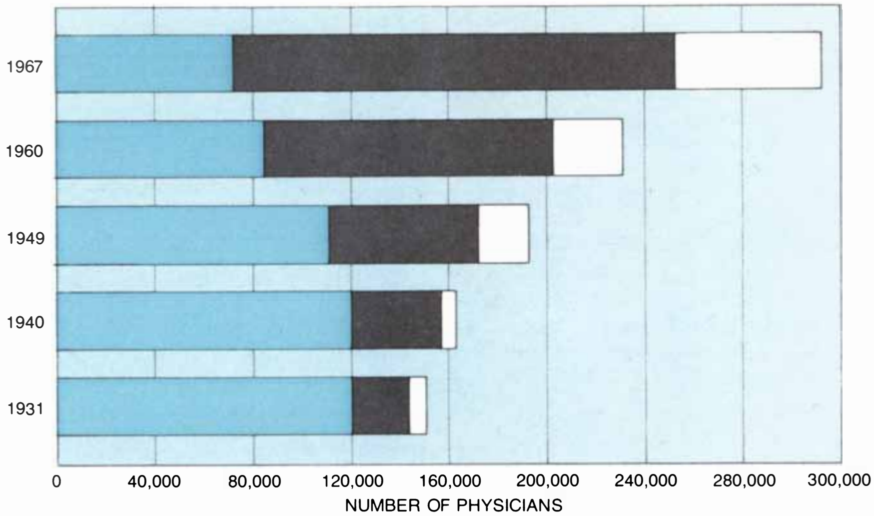
Most of the legislative proposals addressed to the health care crisis seek to moderate or control expenditures by trying to make each dollar provide more medical care. The reasoning behind many of the health bills in Congress seems to be to find ways to make the individual physician more available to the individual patient and to do so, if possible, without spending more. The goal is apparently to provide everyone with the kind of services now available to the people who can afford to buy the best in medical care, either out of their own resources or because they have ample medical insurance.

Even if the best of existing physician-patient relationships could be provided for everyone, which is financially and practically out of the question, it can be argued that this approach is wrong. It would seem more appropriate to think in terms of setting different priorities for the delivery of health care, considering the problems raised by the growing prominence of the chronic diseases and the fact that some 85 percent of the sick can be cared for at the primary and ambulatory level. The priorities I have in mind would rest on the assumption that the present organization of the medical system—a fee-for-service arrangement for treating discrete episodes of illness—is incompatible with the need for more primary care and with the continuity of care that is necessary for the emerging patterns of disease.

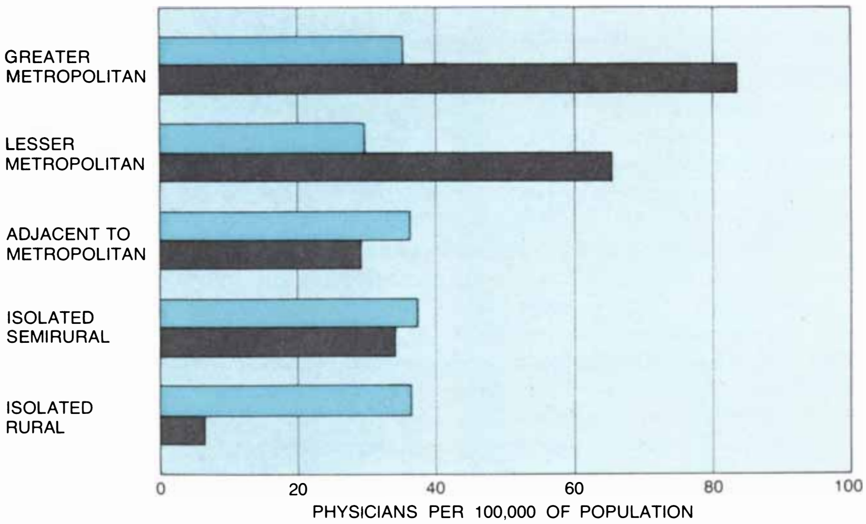
The modern "epidemics" of cancer, heart disease, arthritis, mental illness and so on present medicine with a more complex task of organization and treatment than it has had before. One reason is that the diseases appear to have multiple causes, including behavioral and social factors. Smoking, overeating, overdrinking and lack of exercise are patterns of behavior that seem to contribute to the onset and severity of certain chronic diseases. Social factors, including the tensions resulting from the pace of modern life and the alienation and despair of ghetto life, are reflected in such problems as alcoholism and drug addiction. The complexity of the medical task is also increased by the long-term, sophisticated and expensive nature of the treatment often required.

The answer to the mismatch of technology and delivery is for medicine to orient itself toward a more interventionist approach, by which I mean that the physician and the medical system should be prepared to take the initiative in delivering medical care, rather than leaving the initiative to the patient. The system should reach out to people, seeking out those who for genetic reasons or because of their work or way of life may have a predisposition to a disease, carrying on health education among those at risk, pressing among the poor the case for better nutrition for children so as to forestall crippling disabilities in later life and reminding patients who are identifiably ill of the need for specific measures of treatment. In taking this approach medicine would be recognizing the fact that in the chronic diseases its present capability is more for helping than for curing. Since cures are lacking for most chronic disease, the need is for intervention at the point in time for each patient or group of patients when intervention would do the most good in maintaining functional capability under the existing circumstances.

A related need is to recognize that the task of preventive medicine has changed. Preventive medicine was once a wholly public concern: public health officials took the lead in the administration of inoculations, the drainage of swamps, the control of disease-bearing insects and other measures to protect large segments of the population from disease. Now it has also become a private concern. The onset of certain diseases, including some of the chronic ones, can be prevented or postponed by action directed at individuals or groups of individuals who have characteristics of age, sex, occupation or behavior that put them at high risk. Screening of groups and selective testing for specific diseases according to an individual's risk factor will inevitably become a significant part of medical care. The risk of cancer associated with smoking comes to mind in this context; one could also cite other diseases that arise from exposure of susceptible individuals to a particular environmental hazard.



ACTIVE PHYSICIANS in the U.S. have tended increasingly to be in specialty practice (*gray*) rather than in general practice (*color*). The trend makes it more difficult for people to obtain primary care, which is what was traditionally provided by general practitioners. The white portions of the bars represent hospital physicians who are interns and residents.



DISTRIBUTION OF PHYSICIANS is shown according to type of community and type of practice, with colored bars representing general practitioners and gray bars representing specialists. The uneven distribution is an obstacle to efficient delivery of medical care.

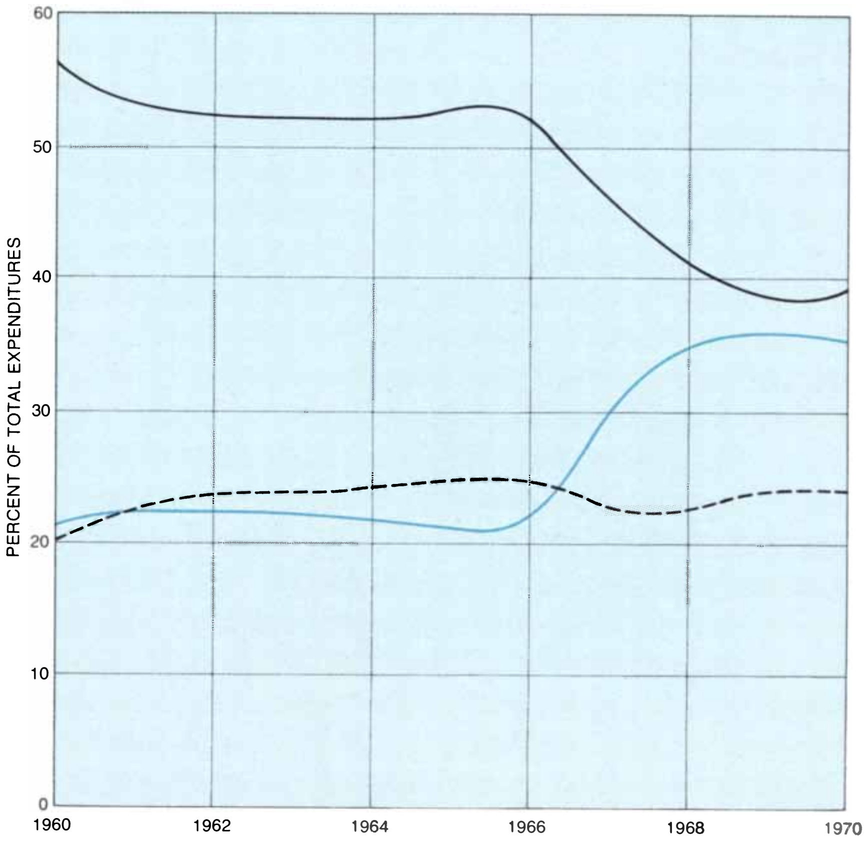
A further need is to redesign medical education so that physicians and other health personnel are trained and challenged to provide primary care and to deal with the chronic diseases. Most medical education still emphasizes the treatment of acute illness and injury (which is not to say that such treatment should be slighted). Few medical people are trained in long-term management of chronic illness and in home care. Medical services must be restructured along similar lines.

A particular problem to be taken into account is the demonstrated reluctance of many people to take the positive step of seeking medical advice, even when, as in the case of national health services such as the one in Britain, the cost will be borne by the society rather than the individual. A study of the British system has estimated that only one person in four with symptoms seeks medical advice. A related problem is that health statistics, being measures of people who go to doctors, fail to show the full extent of social pathology and of undiagnosed diseases.

In my view the medical delivery system would work best if groups of medical personnel were organized to deal with defined population groups, whether in communities or in places of employment. The medical task would involve not only the care of the sick but also the monitoring and health education of the not yet sick. The medical group would have a broad base of knowledge of the population group that it was serving—knowledge including the demographic characteristics of the population, the environmental factors affecting different subgroups, the economic situation of the people and the behavioral patterns in the group. People who presented themselves as sick would of course receive treatment, as they do now. In addition, however, the system would also be in touch with people who were not sick, or not known to be sick, providing them with information on health, suggesting periodic checkups and screening individuals thought to be susceptible to particular illnesses.

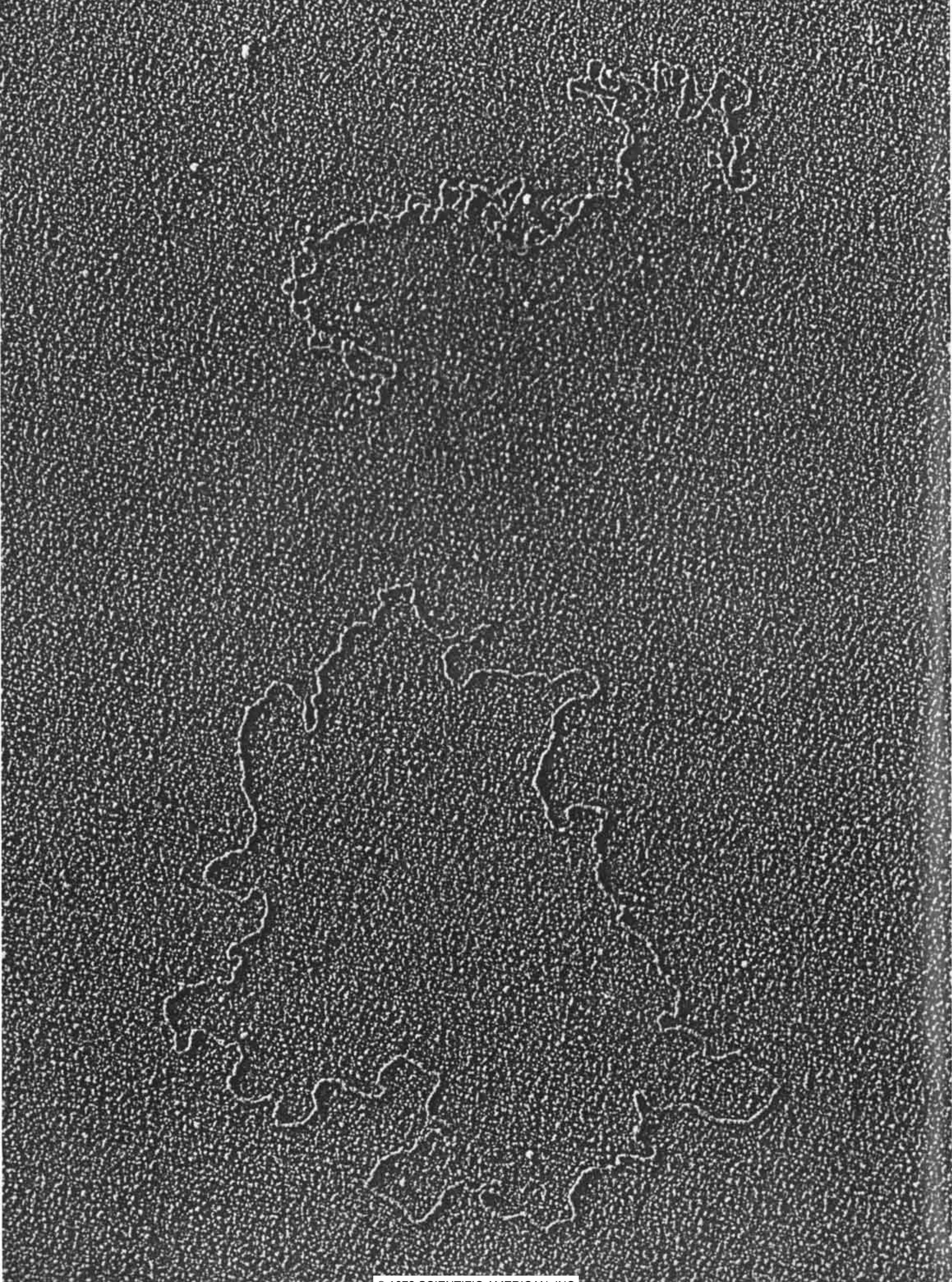
One can envision a useful role for modern computer regimes in such a system. The memory of the computer would contain the name of each member of the group, the record of the system's contacts with him and an indication of his particular susceptibilities or needs if he had any. If a specific crisis arose, such as a period of heavy air pollution, the computer could produce the names of the people who ought to be urged to come in for a checkup because they were vulnerable to respiratory afflictions. If certain members of the group had been fitted with a cardiac pacemaker, the computer would contain information indicating when each pacemaker might be expected to run down, taking into account the condition of the patient and the nature of his activity, so that notice could be sent to him whenever it was time for the pacemaker to be checked.

In such a system it would be feasible for each member to pay (or for a public or private agency to pay in his behalf) what might be called a health-maintenance fee, probably on an annual basis, in exchange for which he would receive the continuing comprehensive attention I have described. The significance of such an arrangement is that it shifts the focus of medical care away from treatment of illness, although that is still provided, and toward the maintenance of health. A few arrangements of this kind are being tested, notably the Kaiser-Permanente program, which has operated successfully on the West Coast for many years [see "The Delivery of Medical Care," by Sidney R. Garfield; SCIENTIFIC AMERICAN, April, 1970].



PATTERN OF PAYMENT for medical care has changed in recent years, largely as a result of such Federal programs as Medicare and Medicaid. Chart shows distribution of expenditures for personal health care according to payments made by patients (*solid black curve*), Government payments (*color*) and payments by insurance companies (*broken curve*).

It is implicit in what I have been discussing that the distinction between health and illness, which was clear when the infectious diseases were the main problem, is much less clear now that chronic diseases predominate. In this situation both the medical profession and the society at large confront questions about what levels of health maintenance to seek and about deploying medical resources to achieve them. How far should society and the medical system go to maintain life in the face of incurable disease? If a costly procedure such as dialysis is available to prevent death from kidney failure, to whom and in what way should dialysis be made available? How much of the gross national product should go into health care? What are the acceptable bench marks for analyzing and evaluating the program? Since poverty and poor health are so closely linked, to what extent is it better to seek improved health by attacking poverty rather than by deploying medical resources? Such questions cannot be answered by medicine alone; they require the formulation of social policies and their expression through the political system.



The Molecule of Infectious Drug Resistance

R factors, which transmit resistance to antibiotics from one bacterial strain to another, are carried on extrachromosomal genetic elements called plasmids, some of which have now been isolated and measured

by Royston C. Clowes

In 1971, only a few months after the antibiotic gentamycin had been introduced for the treatment of bacterial infections, strains of bacteria had developed resistance to the antibiotic. This development was not too surprising. Ever since sulfonamides were first administered to treat bacterial infection it had been known that repeated exposure of bacteria to an antibiotic results in the development of a resistant strain through the processes of mutation and selection. A bacterium that has spontaneously undergone a gene mutation that makes it no longer susceptible to the antibiotic's effect will live and multiply in the presence of the drug, and its similarly resistant progeny constitute the resistant strain. Mutations are rare, however; usually only one bacterial cell in a population of several hundred million will have mutated, so that the physician has a chance to treat the infection successfully by administering antibiotics before resistant cells develop in large numbers.

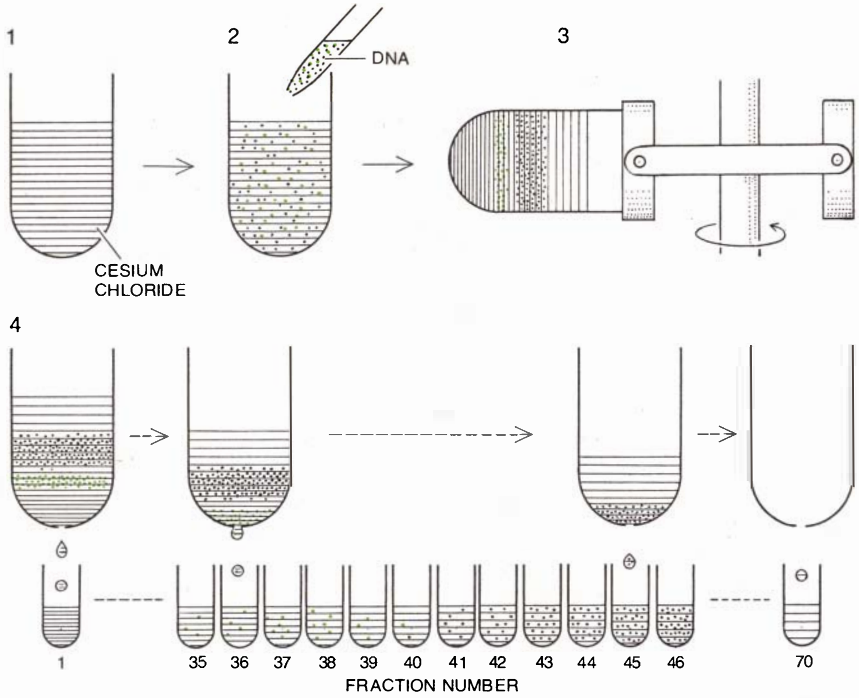
The resistance to gentamycin was different. The bacteria from the first patient in whom a gentamycin-resistant strain was detected were resistant not only to gentamycin but also to three other common kinds of antibacterial agent: chloramphenicol, ampicillin and sulfonamide. Then, from a second patient, investigators isolated four different bacterial strains, each of which carried resistance to the same four antibiotics. Moreover, when they grew any of these bacteria in laboratory glassware with a bacterial strain that was susceptible to the antibiotic's effects, the susceptible strain also became resistant to all four antibiotics in a matter of hours. From the investigators' knowledge of many similar cases reported during the past decade they realized that they were dealing with a case of infectious antibiotic resistance. Each strain isolated from the patients carried an "infectious antibiotic resistance factor," or *R* factor: a complex consisting of a gene producing gentamycin resistance, genes determining resistance to the other three antibiotics and genes that facilitate the transfer of the *R* factor.

Intestinal bacteria that give rise to such diseases as enteritis or typhoid fever and carry *R*-factor-controlled resistance to all known antibiotics have now been isolated in most parts of the world. Because the genes of these *R* factors are transferred independently of the genes on the bacterial chromosome, and because they are "cured," or lost, if bacteria carrying them are grown in the presence of certain dyes or other agents, they are known to exist independently of the chromosome; they are carried on the extrachromosomal genetic elements called plasmids. A strain carrying an *R* factor is able to conjugate, or pair, with other bacterial cells. When that happens, a copy of the *R* plasmid is transmitted to the second cell, which thereby acquires all the drug resistances carried on the plasmid and also the potential of transmitting those resistances to still more cells.

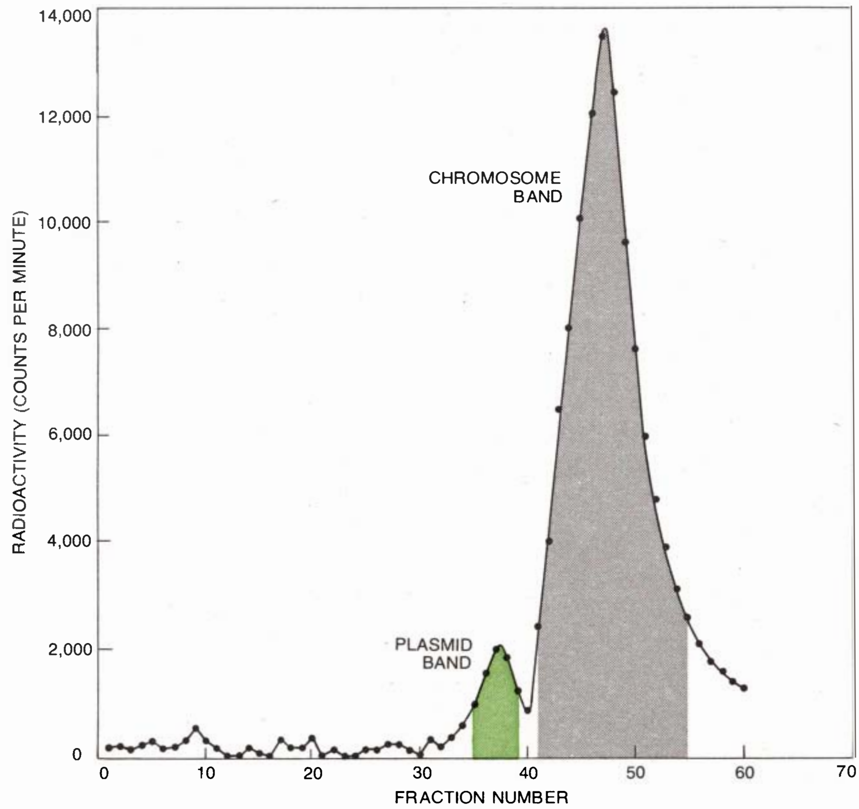
During the 1960's genetic experiments in a number of laboratories established the mode of transference of *R* factors and revealed something of their composition and structure [see "Infectious Drug Resistance," by Tsutomu Watanabe; SCIENTIFIC AMERICAN, December, 1967]. More recently the effort in my laboratory at the University of Texas at Dallas and in other laboratories has been to learn more about the physical nature of the plasmids that constitute the *R* factor: their size and structure, the ways in which they multiply and how they may evolve.

It seemed likely that *R* factors are composed of the genetic material DNA, and that had been confirmed by transferring *R* factors to *Proteus mirabilis,* a strain of bacteria that has DNA of unusually low density. When DNA extracted from a *Proteus* culture is suspended in a solution of the salt cesium chloride and spun at high speed in a centrifuge, the DNA molecules move to positions in the centrifuge tube where their density is equal to the density of the salt surrounding them. The contents of the tube are separated into a number of fractions, each from a different level in the tube. Fractions containing DNA are then identified either by measuring the extent to which they absorb ultraviolet radiation of a certain wavelength or by first incorporating radioactive subunits in the DNA of the live culture and then measuring the radioactivity of each fraction [*see illustrations on next page*]. If the DNA is extracted from a *Proteus* strain to which an *R* factor has been transferred, one finds in addition to the main band of DNA characteristic of the *Proteus* chromosome a smaller "satellite" band at a greater density. If the strain of *Proteus* is then "cured" of its *R* factor, this denser satellite band is no longer seen, and so it is clear that the satellite band must be the DNA of the *R* plasmid.

R-FACTOR PLASMIDS, shadowed with platinum vapor, are enlarged 62,000 diameters in an electron micrograph made by Michiko Mitani. The plasmids are molecules of looped DNA. The extended one (*bottom*) is a "nicked duplex loop"; the distance around it can be measured to establish its molecular weight. The other one (*top*) is a "supercoiled" form.



***R*-FACTOR DNA is separated from bacterial DNA by density-gradient centrifugation. A solution of cesium chloride (*1*) to which DNA from a culture of *Proteus mirabilis* carrying an *R* factor has been added (*2*) is spun in an ultracentrifuge for 48 hours (*3*). The DNA's move to points in the centrifuge tubes where their density is equivalent to that of the cesium chloride. The contents of a centrifuge tube are collected drop by drop in smaller tubes (*4*); each fraction contains material from a different level and hence of a different density.**



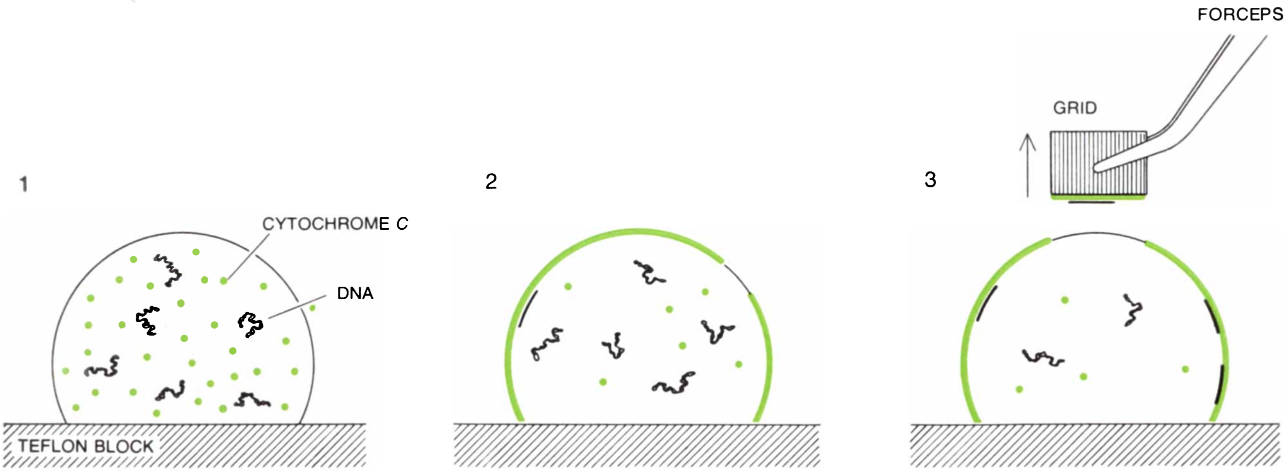
DNA-CONTAINING FRACTIONS can be identified if radioactive subunits were previously incorporated in the bacteria; the amount of DNA in each fraction is proportional to the radioactivity. In the case of *Proteus mirabilis* carrying an *R* factor there are two DNA peaks. The major, less dense band (*gray*) is the DNA of the *Proteus* chromosome. The satellite band (*color*), which comes from earlier fractions, represents the denser DNA of the *R* factor.

The bacterial chromosome is a single giant DNA molecule in the form of a closed loop about a millimeter long (about 1,000 times as long as the cell). After DNA isolated from a culture of *Proteus* harboring an *R* factor had been centrifuged in cesium chloride, my colleagues Taizo Nisioka and Michiko Mitani and I took samples from the fractions containing the *R*-plasmid DNA and prepared them for examination in the electron microscope by what is called the microdrop technique. In this method the DNA samples are mixed with a solution of a basic protein such as cytochrome *c*. When small drops of the mixture are put on a material such as Teflon, the protein molecules spread over the surface of the drops as oil does on water, in a film one molecule thick. The acidic groups of the DNA become bound to the basic groups of the protein, so that the long DNA molecules, teased out gently without breaking, become extended at the surface of the film [*see top illustration on opposite page*]. A sample is taken from this surface to an electron-microscope specimen grid and dried, and the DNA molecules can be viewed in the microscope.

We found that the DNA of the *R* factor consists of closed looped molecules, some of them extended and some of them tightly coiled [*see illustration on page 18*]. The distance around all the extended loops was about the same. After storage over many months the proportion of tightly coiled molecules decreased, but the number of extended loops, which were the same size as the original ones, increased. It was clear, therefore, that the DNA of the *R* factor is made up of looped (or what are often called "circular") molecules of a specific size.

Knowing the length of other DNA molecules of known molecular weight, and thus the molecular weight per unit length of DNA, we could calculate the molecular weight of the *R*-plasmid molecules. By this method and others all *R* factors so far investigated have been shown to be looped DNA molecules, each with a characteristic size. Their molecular weights range from about one megadalton to more than 100 megadaltons. (A dalton, the unit of molecular weight, is the weight of one hydrogen atom. One megadalton is a molecular weight of one million.) This spectrum of sizes is about the same as that found for the DNA of bacterial viruses and is at most only a small fraction of the size of the bacterial chromosome.

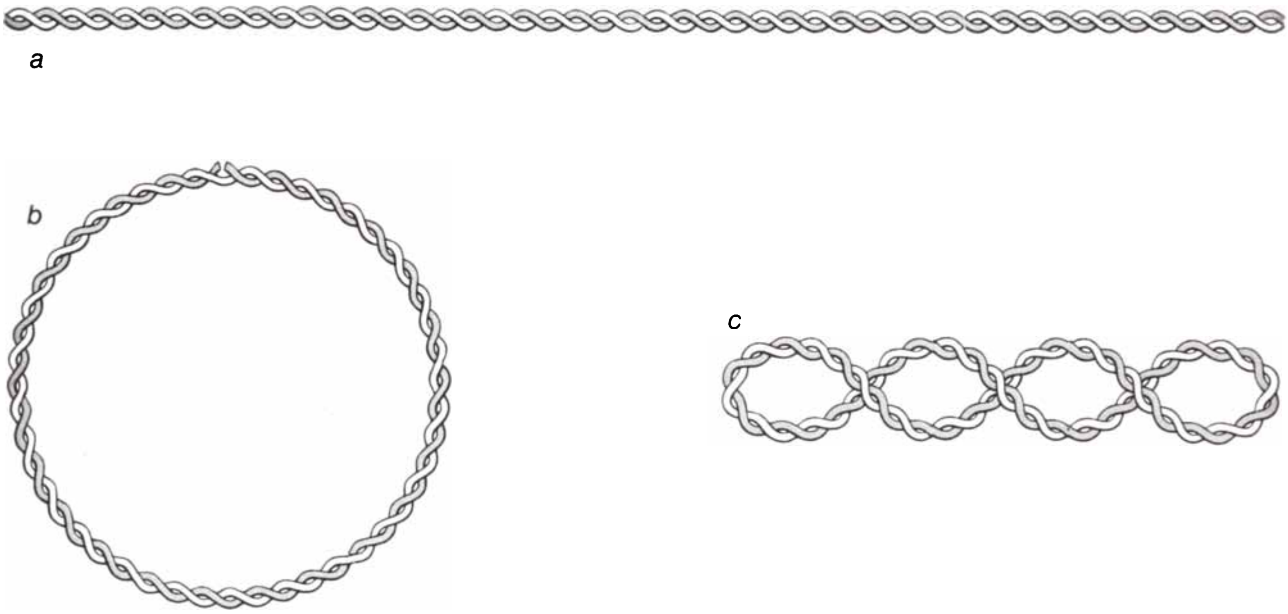
The fact that *R*-plasmid molecules are circular and tightly coiled was of great



MICRODROP TECHNIQUE developed by Dimitrij Lang and Michiko Mitani at the University of Texas at Dallas mounts DNA molecules for electron microscopy without distorting them. A drop of water containing DNA (*black*) from a plasmid-rich centrifuge fraction, salt, formaldehyde and the protein cytochrome *c* (*color*) is placed on a water-repellent surface (*1*). The cytochrome *c* molecules form a monomolecular film at the surface of the drop and the positively charged protein attracts the negatively charged DNA molecules, which unfold against the film (*2*). A part of the film, with the DNA, is picked up on a microscope specimen grid (*3*).

value in our subsequent work. When DNA molecules are extracted from cells, they are subjected to shearing forces that may break them. The small size and tight coiling of plasmid molecules make them resistant to these forces. More important, a circular molecule must by its nature have remained unbroken during extraction, and so its size and weight are true measures of what the native molecule was inside the cell. The special properties of various species of circular DNA have been investigated intensively by two groups headed by Robert L. Sinsheimer and Jerome Vinograd at the California Institute of Technology. The Cal Tech groups found circular DNA molecules of two kinds. In one the two strands of the DNA double helix are joined to form a duplex loop; these structures are "supercoiled," presumably so that the long DNA molecules can be fitted inside a cell. In the other kind only one of the DNA strands is joined; these molecules are not supercoiled. If a single break is made in one strand of a supercoiled molecule, the molecule loses its coils and is converted to the uncoiled structure, which is therefore called a "nicked duplex loop" [*see illustration below*]. The two kinds of circular DNA we had found in *R* factors were just such nicked duplex loops and supercoiled molecules.

There is an effective method of distinguishing supercoiled DNA from other DNA, which depends on the insertion of molecules of ethidium bromide, one of the dyes that can "cure" plasmids, between adjacent DNA subunits. This intercalation, as it is called, results in two



DNA MOLECULE is found in three forms. In each the DNA consists of two twisted strands that form a double helix, or duplex; the strands are connected, as by the rungs of a ladder, by paired subunits (nucleotides) whose particular sequence constitutes the genetic code. The basic form is the linear duplex (*a*). If the ends of the duplex are joined through one strand, the structure is a nicked duplex loop (*b*). If both strands are joined, the structure is an intact duplex loop, usually found in the supercoiled form (*c*).

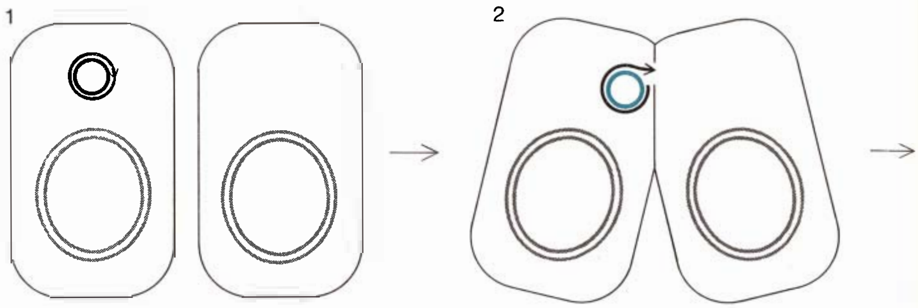
changes: the length of the DNA molecule is increased, so that it becomes less dense, and the helix becomes slightly less twisted. Large numbers of ethidium bromide molecules can intercalate in a nicked duplex loop or a linear duplex, without the untwisting effect leading to coiling. In contrast, the number of dye molecules that can be inserted in a supercoiled loop is strictly limited because there is a limit to the further coiling that is produced as a result of untwisting. The limitation on intercalation limits the associated extension in length and thus leads to a smaller decrease in density for supercoiled loops than for nicked duplex loops or linear duplex DNA. Therefore if ethidium bromide is added to a mixture of these three types of DNA molecule, the supercoiled loops can be separated from the other forms as a denser band after centrifugation.

Since, as we have seen, a major portion of *R*-factor DNA exists in the cell as supercoiled loops, it can be isolated and separated by this procedure from the chromosomal DNA. That DNA, being about 100 times longer, is generally broken into linear fragments when it is extracted from the cell. The procedure provides a means of separating plasmid DNA even in host bacteria such as *Escherichia coli,* in which the chromosomal DNA does not differ in density from the *R* plasmid as it does so conveniently in *Proteus.* Since *E. coli* and related bacteria are the natural hosts of many *R* factors, the procedure enables one to study the plasmids as they are found in nature.

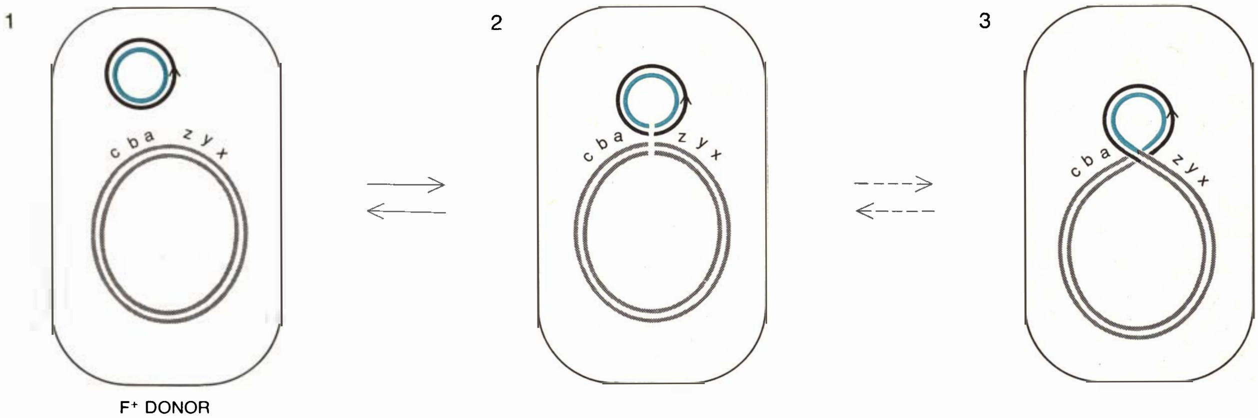
The same experiments can be used to measure the number of *R*-plasmid molecules in each cell. When bacteria are grown in the presence of a radioactive subunit of DNA such as thymine, the DNA in the cell takes up this radioactive label. If we assume that the uptake of radioactive thymine into plasmid DNA occurs to the same extent as it does into chromosomal DNA, and if most of the plasmid DNA exists as supercoiled loops, then the relative amounts of DNA in the plasmid and in the chromosome can be measured by the relative amounts of label in the denser satellite peak and in the main peak. Since we know the molecular weights of a number of *R* plasmids, and other experiments have shown that the molecular weight of the *E. coli* chromosome is 2,500 megadaltons, we can estimate the relative number of *R*-plasmid molecules per chromosome from their relative DNA content in the cell. By this method and others many *R*-plasmid molecules have been found to exist in an approximate one-to-one relation with the chromosome. Although the DNA's of the *R* plasmid and of the chromosome are physically separated, then, there appears to be some regulatory mechanism in the bacterium such that the *R* plasmid replicates only once for every chromosomal replication. How this regulation is effected is not known and is the subject of intensive work in a number of laboratories.

Before 1955 infectious antibiotic resistance factors were rare. By 1965 it was found that up to 60 to 70 percent of all common intestinal bacteria from hospitals and other clinical sources carried *R* factors, usually determining resistance to three or more antibiotics. *R* factors can also be transferred to bacterial species responsible for cholera and plague, and similar elements are now common in staphylococci. The epidemic spread of these factors among pathogenic bacterial strains has made an understanding of how they have evolved—and how they evolve so quickly—an important problem in clinical bacteriology today.

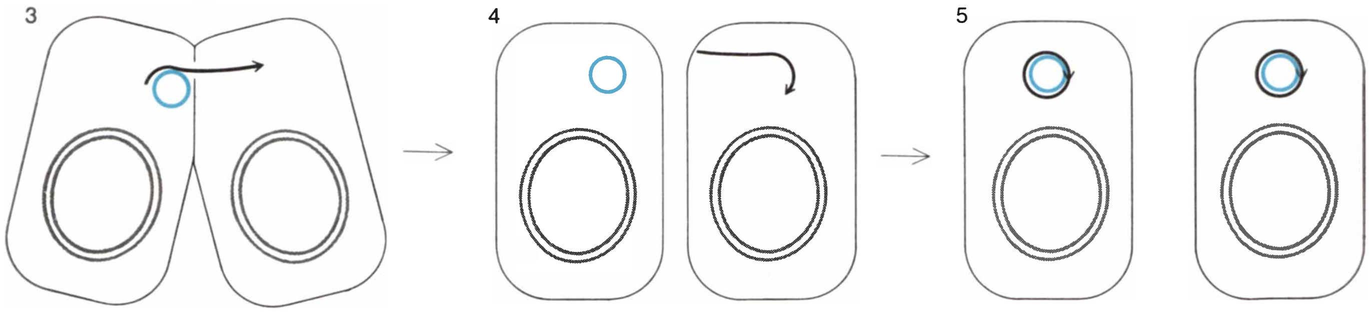
One of the original hypotheses for *R*-factor evolution suggests that it depends on a process known as gene pickup. To explain this concept we must consider another bacterial plasmid, the *F* factor, which determines sexuality and conjugation in *E. coli.* Like the *R* factors, the *F* factor is infectious, can transmit itself se-



INFECTIOUS TRANSFER of a bacterial plasmid such as an *F* factor or one of the *R* plasmids is diagrammed. Two bacteria are shown about to conjugate, or pair (*1*); each of them has a circular chromosome (*gray*), and one of them, the donor cell, also has a



DONOR CELLS can be of three types, the relations among which are illustrated here with the *F* factor as a model. An F^+ donor (*1*) usually transfers *F* as in the illustration at the top of the page. Sometimes, however, a break occurs in both strands of the *F* plasmid and of the chromosome (*2*) in a region where the nucleotide sequences of plasmid and chromosome are similar enough so that the two sets of ends join (*3*) and the plasmid is inserted into the chromosome (in this case between genes *x, y, z* and *a, b, c*) to form

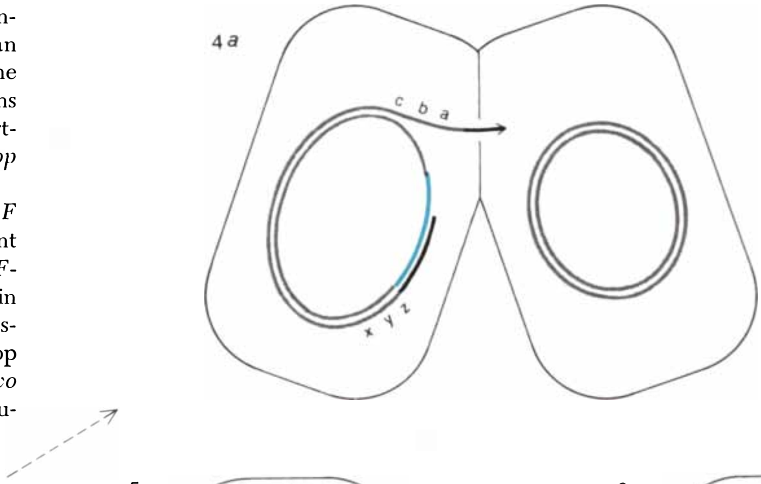


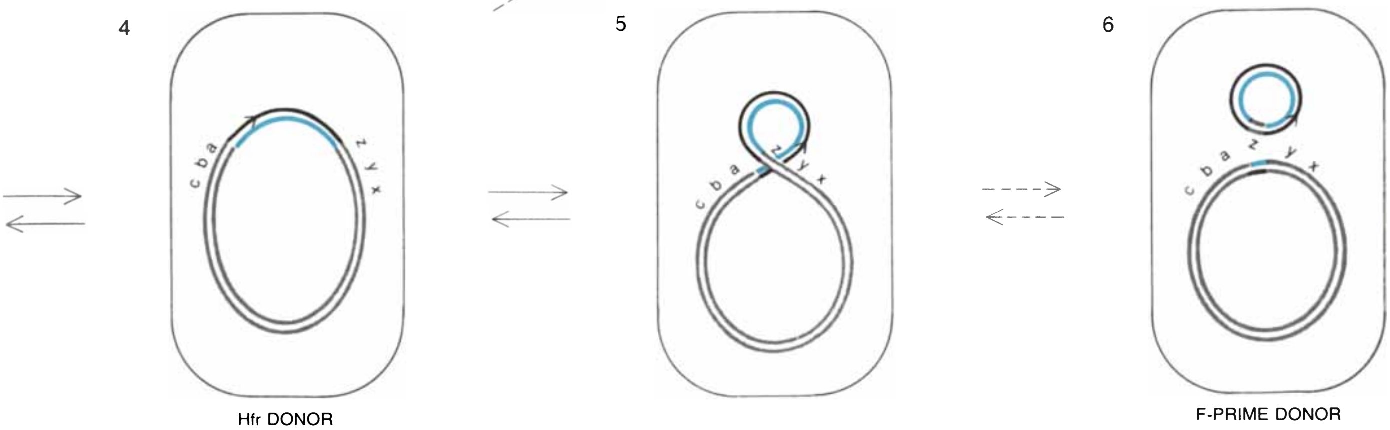
plasmid, an intact duplex loop (*color and black*). On pair formation (*2*) one strand of the plasmid DNA breaks at a particular point (*arrow*) and passes into the recipient cell (*3, 4*). After transfer each strand of DNA synthesizes its complementary strand, so that two intact duplex loops are formed. Each cell now has a copy of the original plasmid and is therefore a potential donor cell (*5*).

quentially by conjugation and is an intact duplex DNA loop (of about 62 megadaltons), and there is about one of these *F* plasmids for every chromosome. The infectious transfer of the *F* factor and of a number of *R* factors has recently been elegantly defined in molecular terms by Daniel Vapnek and W. Dean Rupp of the Yale University School of Medicine. They have shown that conjugation leads to a break in one of the two strands of the intact *F*-factor duplex loop and the transfer of that strand to the recipient cell. There it synthesizes its complementary strand and subsequently forms an intact duplex loop; meanwhile the strand retained in the donor also forms its complementary strand and is converted into an intact duplex loop [*see top illustration on these two pages*].

From donor cells that harbor an *F* factor (called F^+ donors) rare mutant cells can be selected in which the *F*-plasmid DNA has been inserted within the bacterial chromosome without causing a break in the chromosomal loop [*see bottom illustration on these two pages*]. Conjugation by one of these mutant strains results in a similar break within the DNA of the inserted *F* plasmid. Thereafter a strand of DNA consisting of a part of the *F* plasmid is transferred to the recipient cell, taking with it the attached chromosomal strand. This chromosomal transfer occurs very frequently, so that the mutant strains are called *Hfr* (for high-frequency recombination). Different *Hfr* cells are formed after the insertion of the *F* plasmid at different sites along the chromosome.

Still a third type of donor is derived from *Hfr* donors by a reversal of the insertion process, leading to the release of the integrated *F* factor from the chromosome. This release usually occurs at a point slightly different from the point of insertion, in which case a segment of the bacterial chromosome is incorporated within the DNA of the released *F* plas-





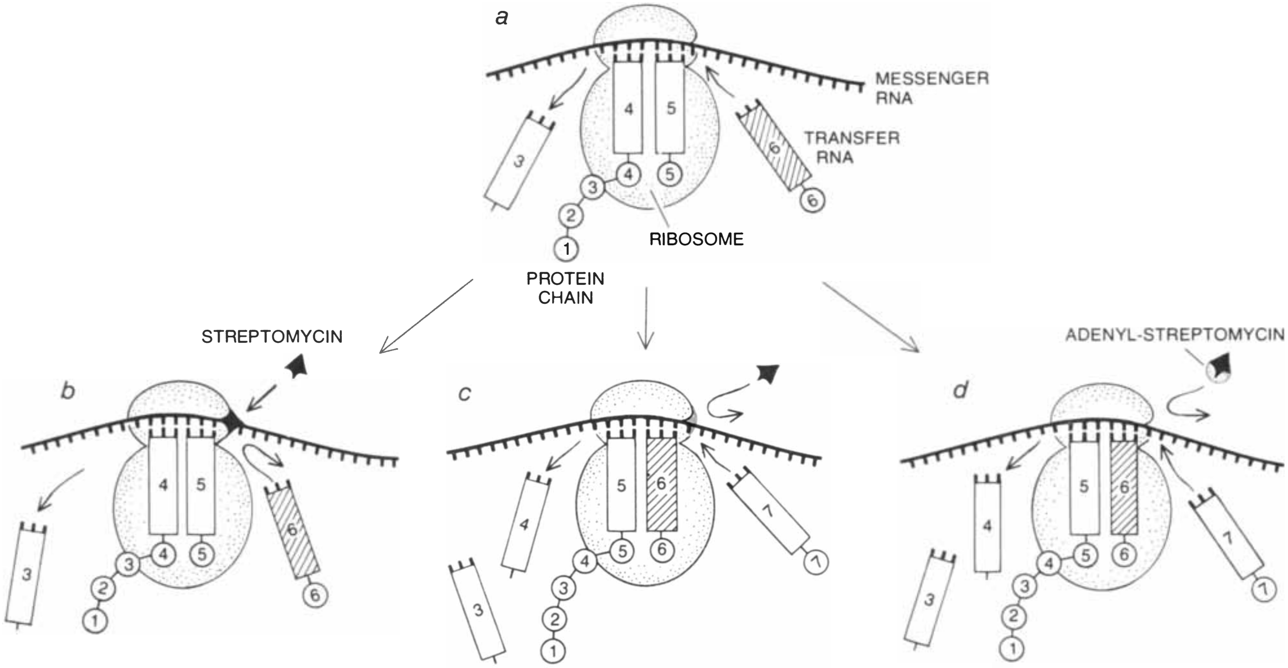
an *Hfr* donor cell (*4*). When an *Hfr* cell conjugates, one strand of the *F* plasmid breaks and the first part of the plasmid strand is transferred to the recipient along with the chromosome (*4a*). Occasionally the *F* plasmid is released from the chromosome of an *Hfr* donor. If this occurs by breakage and reunion at a point different from where the plasmid was incorporated, segments of chromosomal and plasmid DNA are interchanged. The plasmid, having picked up a chromosomal region, becomes an *F*-prime factor (*5, 6*).

mid and some of the plasmid DNA remains in the chromosome. These released *F* factors that have "picked up" chromosomal DNA are known as *F*-prime factors, and bacterial strains in which they are present are called *F*-prime donors. A similar mechanism of gene pickup was suggested as the basis for the evolution of *R* plasmids. The idea was that a "transfer factor" resembling *F* was integrated in the chromosome and then released, together with a segment of adjacent chromosomal DNA on which a gene controlling drug resistance was located.

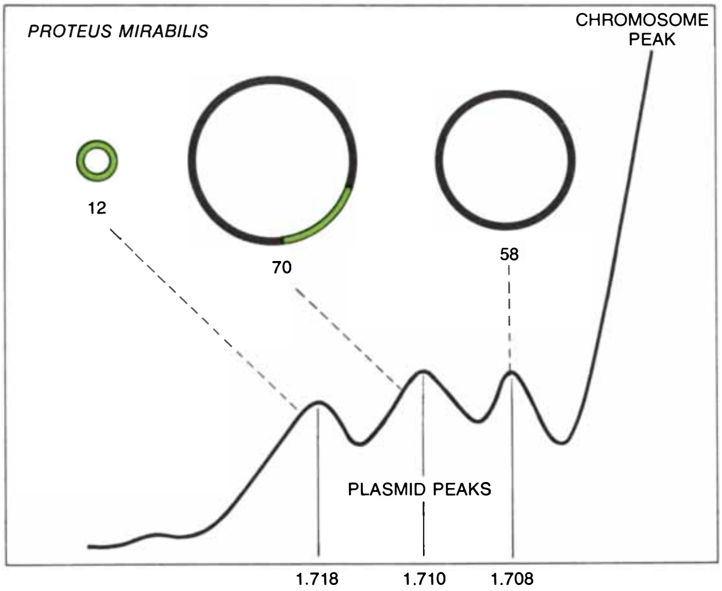
Some recent findings make the idea of evolution of *R* plasmids through gene pickup seem less likely. The most telling argument comes from discoveries about the mode of action of antibiotic resistance mediated by *R* factors. Many antibiotic molecules exert their lethal effect by becoming attached to bacterial ribosomes: the cellular particles where proteins are synthesized. In this way they interfere with protein synthesis. In all bacteria where they have been analyzed resistance to these antibiotics that arises from mutations of chromosomal genes is effected by changes in the structure of the ribosomes such that the antibiotic can no longer interfere with protein synthesis [*see illustration below*]. In contrast, *R*-factor resistance to these same antibiotics involves the direct modification of the antibiotic molecule in such a way that it can no longer interact with the ribosome, which remains unchanged. Since the mode of action of *R*-factor resistance appears to be so different from that of chromosomal resistance, *R*-factor genes (unless they arise from chromosomal genes as yet unidentified) must be extrachromosomal in origin.

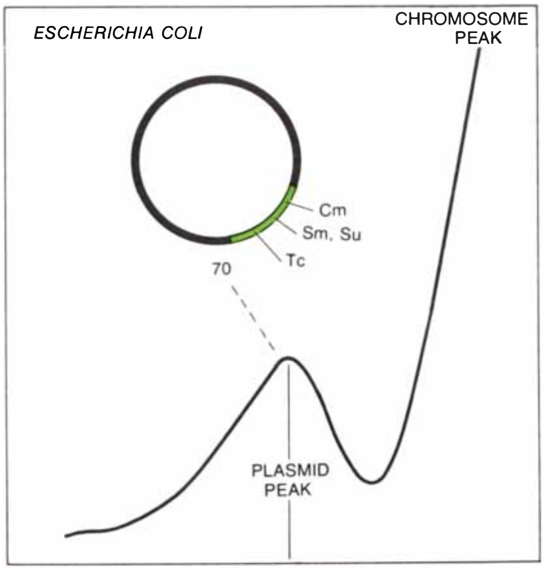
Although the DNA of any one of a number of *R* factors isolated from either *E. coli* or *Proteus mirabilis* consists of molecules of the same size, certain exceptions are found. One of the first reported was in the case of the 222/*R* plasmid, which determines resistance to tetracycline, chloramphenicol, streptomycin and sulfonamide. DNA from a *Proteus* host harboring this plasmid produces a satellite band with three distinct peaks [*see illustration on opposite page*]. When we examined a sample from each peak in the electron microscope, we found molecules of 12 megadaltons in the peak of greatest density, molecules of 70 megadaltons in the middle peak and molecules of 58 megadaltons in the lowest-density peak. The sizes and densities of these molecules suggested that some of the 222/*R*-factor plasmids (which in *E. coli* were all 70 megadaltons in size) were divided into two smaller molecules of 12 and 58 megadaltons in *P. mirabilis*. Since the peak of greatest density is about the same size as the other two peaks (that is, it has the same amount of DNA) in spite of the fact that the 12-megadalton molecules in this peak are much smaller than the molecules in the other peaks, there must be many more of the 12-megadalton molecules than of the larger molecules. It is therefore likely that after the 70-megadalton molecule subdivides into two smaller ones the 12-megadalton molecule can replicate independently. Other experiments show that a modified plasmid similar to the 58-megadalton molecule can be derived from the 70-megadalton molecule, and this plasmid can also replicate normally.

The natural replication of the bacterial chromosome (or any other DNA molecule) appears to be possible only beginning at a unique point of origin on the chromosome, and this point appears to be essential for replication. If an *R* factor such as 222/*R* had in fact evolved by sequential chromosomal pickup, it would be unlikely that enough of the chromosome would have been picked up to have much chance of including the unique point of origin. The *R* factor would therefore be expected to have only the one point of origin for DNA replication that was present in the original transfer-factor molecule. Since in the *Proteus* host the 222 factor segregates into two molecules, both of which



MODE OF ACTION is different in chromosomal and in plasmid-mediated antibiotic resistance. Streptomycin inhibits protein synthesis in bacteria. Ordinarily transfer RNA assembles amino acids on ribosomes according to instructions recorded in messenger RNA (*a*). Streptomycin attaches itself to a site on the ribosome and prevents protein synthesis, perhaps by blocking transfer-RNA attachment (*b*). Mutation of a chromosomal gene changes the ribosome's structure so that the streptomycin cannot bind (*c*), but *R*-factor genes make an enzyme that attaches an adenine molecule to the streptomycin, which no longer fits the unchanged ribosome (*d*).





PLASMIDS representing the 222/*R* factor in *Proteus mirabilis* and *Escherichia coli* are compared after centrifugation and fractionation, with ethidium bromide present in the case of *E. coli*. In *Proteus* there are three peaks in the satellite band. In each peak a different plasmid, measuring respectively 12, 70 and 58 megadaltons, is found. In *Escherichia*, on the other hand, there is only one plasmid peak, containing 70-megadalton molecules. The genes for resistance to four antibiotics are on the 12-megadalton plasmid, those for the transfer factor on the 58-megadalton plasmid. The two plasmids are integrated in the 70-megadalton plasmid cointegrate.

may replicate independently, this again does not fit the idea of evolution by sequential chromosomal pickup. On the other hand, these facts can be reconciled with an alternative idea: that the genes on *R* factors arise from molecules that were originally extrachromosomal and capable of independent replication.

The *R* factors first studied by genetic and physical methods gave results similar to those of 222/*R* in *E. coli,* that is, all the antibiotic-resistance genes were transmitted together with the genes determining transfer properties as though they were all part of the same genetic structure. The physical analysis of the DNA of these plasmids confirmed this unitary structure, disclosing only a single molecular species for each *R* factor, as shown in the case of 222/*R* in *E. coli.* Since all the antibiotic-resistance genes and the genes determining transfer are integrated in a single structure, these *R* factors are called plasmid cointegrates. A number of other factors have recently been shown to have different properties, in particular an *R* factor studied in *E. coli* by E. S. Anderson and Malcolm Lewis at the Central Enteric Reference Laboratory in London. That factor controls resistance to streptomycin, ampicillin and tetracycline but transfers only some of its properties to some recipient cells [*see illustration on page 27*]. After conjugation some of the recipient cells carry infectious resistance to only one of the three antibiotics; others carry either streptomycin resistance or ampicillin resistance that is no longer infectious. Still other cells have lost all antibiotic resistance but, when they are incubated with the noninfectious streptomycin- or ampicillin-resistant cells, can potentiate the transfer of this noninfectious resistance. Anderson and Lewis therefore suggested that these antibiotic-sensitive cells contain a plasmid (called delta) that controls only transfer-factor properties.

In my laboratory I undertook a biophysical study of the DNA molecules isolated from these cultures in collaboration with Christine Milliken. This study has fully confirmed the conclusion, reached by Anderson and Lewis from their genetic studies, that the original strain harbors several independent plasmids. The delta plasmid controls transfer-factor properties, a second plasmid (called *S*) controls noninfectious streptomycin resistance and a third (called *A*) controls noninfectious ampicillin resistance. The delta plasmid was found to be joined with a tetracycline factor (*T*) to form a composite delta-*T* plasmid controlling infectious tetracycline resistance. When *A* or *S* is present in the same cell as a transfer factor (delta or delta-*T*), it too can be transferred to recipient cells.

Satellite DNA separated from *E. coli* host cells by centrifugation showed the following features: DNA from delta$^+$ cells (cells with transfer-factor properties but no drug resistance) contained a single species of duplex looped DNA molecules of 60 megadaltons; DNA from *S* or *A* cells (respectively carrying noninfectious streptomycin or ampicillin resistance) was also a single species but was a tenth as large (six megadaltons); DNA from delta (*S*) or delta (*A*) cells (respectively carrying infectious streptomycin or ampicillin resistance) was found to contain two distinct species of molecules, 60 megadaltons and six megadaltons in size.

It appears, then, that an *R* factor can be composed of an aggregate of two or more plasmid molecules. That is, delta (*S*) has the properties of an *R* factor, since it can transfer infectious streptomycin resistance, but those properties are due to two independent molecules, one (delta) controlling transfer-factor properties and the other (*S*) controlling noninfectious streptomycin resistance. As a result of conjugation brought about by the presence of the delta factor, delta itself is transferred and the probability of simultaneous transfer of *S* is very high. The majority of recipient cells therefore inherit both delta and *S* molecules and show the characteristics of infectious streptomycin resistance. Occasionally some of the recipient cells acquire only one or the other of the two molecules, and then they acquire either transfer-factor properties or noninfectious streptomycin resistance. Similar properties are proposed for delta (*A*) cells.

In distinction to delta and delta-*T*, which like all other *R* plasmids were found in a one-to-one molecular ratio with the chromosome, there were more than 10 copies of either the *A*- or the *S*-plasmid for each chromosome. This new type of *R* factor, delta (*S*) or delta (*A*), is termed a "plasmid aggregate." Its evolution is difficult to explain by gene pickup and is clearly more consistent with extrachromosomal evolution. The plasmid aggregate may be a more primitive evolutionary type that can evolve, through a joining of the separate molecules, into a plasmid cointegrate such as delta-*T*.

E. coli is a normal inhabitant of the human alimentary tract; it seems unlikely that during its evolutionary history in the preantibiotic era such an organism would ever have come into contact with an antibiotic such as streptomycin, which is produced by a very different genus of microorganism usually found only in soil. How then can the bacterium produce an enzyme that recognizes streptomycin and inactivates it? (The evolution of chromosomal, as opposed to *R*-factor, resistance is easier to explain, since chromosomal resistance stems from a change in the structure of a normal bacterial product, the ribosome, through mutation.)

DNA MOLECULES of the *R* factor delta (*S*), a plasmid aggregate, are enlarged 40,000 diameters in an electron micrograph made by Christine Milliken. The large nicked duplex loop (*top*) is the delta plasmid, the transfer factor. The small molecules (*bottom*) are *S* plasmids, carrying streptomycin resistance. One is supercoiled, one is a nicked duplex loop.

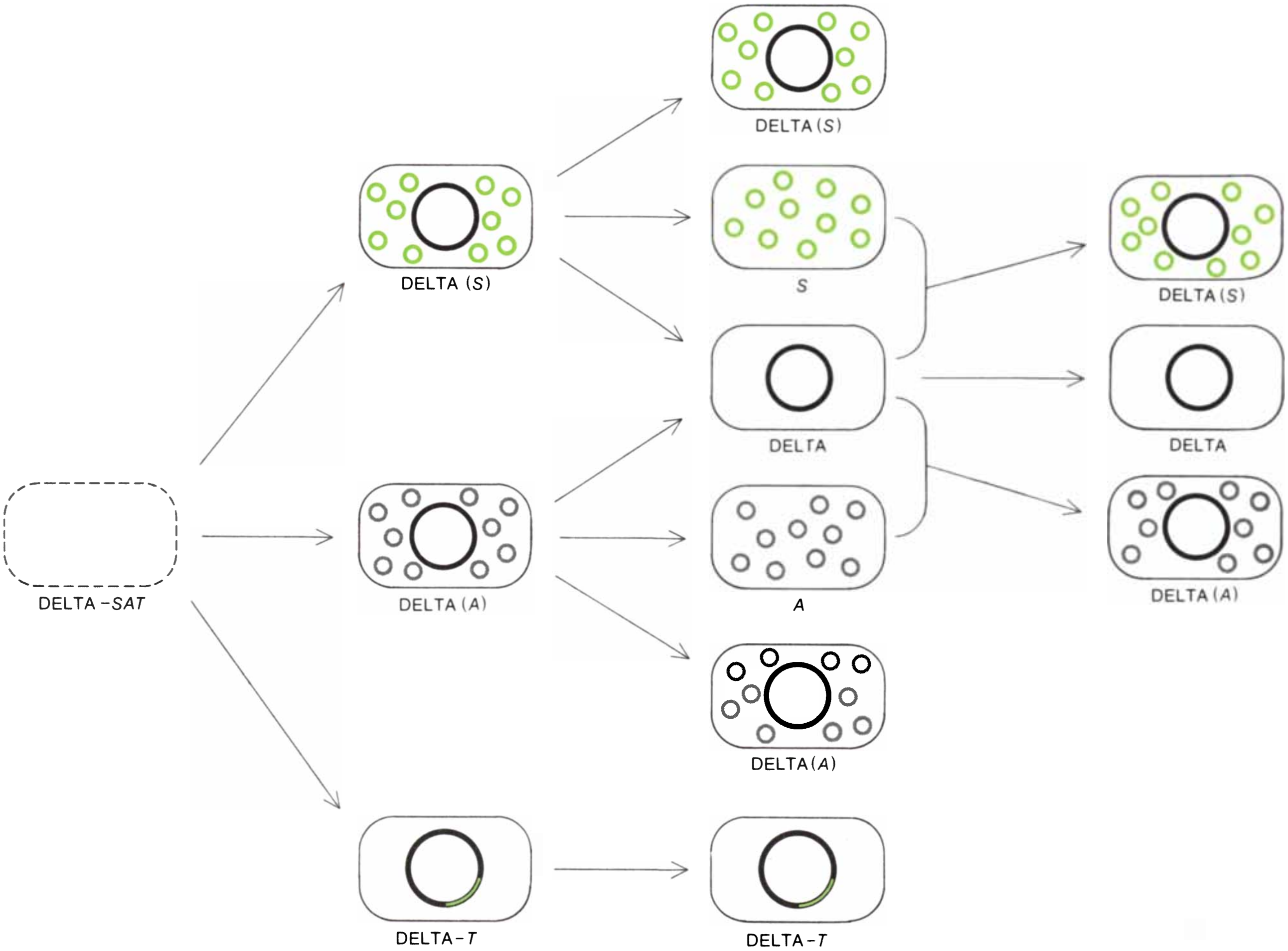
William Shaw of the University of Miami has pointed out that many *R*-factor resistances depend on the inactivation of antibiotics through such enzymic processes as acetylation, phosphorylation or adenylation, processes that normal metabolic compounds undergo in the normal bacterial cell. Shaw suggests that enzymes coded for by *R*-factor genes and operating to inactivate antibiotics may have evolved from enzymes that originally acted on normal bacterial molecules. He has found that resistance to chloramphenicol by a number of plasmids of different origin is due in all cases to its acetylation by enzymes that are similar in size and constitution and appear to differ only in the efficiency with which they can bind to chloramphenicol. He suggests that these enzymes may have evolved from one that originally acetylated a normal cell component. Mutations of the gene determining this enzyme might have modified the enzyme so that it bound to chloramphenicol with increasing efficiency, until now it binds to chloramphenicol rather than to the original molecule and can inactivate it effectively enough to prevent its antibiotic activity.

The rapid proliferation of *R* factors and the identification of *R*-plasmid aggregates suggest that similar entities could have importance in the normal regulatory control of bacteria. Genes on the bacterial chromosome determine the activities that enable the bacterial cell to grow and multiply in many very different environments. Under most circumstances most of these genes are not active; they are "switched off" except in the particular environment in which their function is needed [see "The Control of Biochemical Reactions," by Jean-Pierre Changeux; SCIENTIFIC AMERICAN, April, 1965]. Now, if a certain function determined by several sequentially acting enzymes is required very infrequently, the continual replication of the genes that determine it may impose more of an evolutionary burden than would be compensated for by the rare need for the function. If, however, these determinant genes were carried on a plasmid present in only a few cells of a population, they would need to replicate only in those rare cells. Any time there arose a need for the function in question, these special genes could be acquired rapidly by all the cells in a progeny population if the plasmid also had transfer-factor properties or (even more economically) if other rare cells contained transfer factors that could interact to establish a plasmid aggregate.

The control of a number of activities in this way by their distribution to a number of different plasmids carried by different cells in the population would enormously increase the total gene pool and extend the overall metabolic potential of the species. This would appear to represent a logical extension of genetic regulation: from the present concept of individual cellular control by the switching on and switching off of chromosomal genes to a species-wide control of gene-pool synthesis.

As a matter of fact, entities that are similar to *R* factors but that control enzymes involved in metabolism have recently been recognized. For example, genes for enzymes catalyzing the breakdown of complex organic molecules to enable them to be used as a source of carbon have recently been shown to be situated on plasmids in *Pseudomonas*, a genus of bacteria well known for its ability to grow on almost any kind of organic substance. Certain strains of *Pseudomonas* are enabled to grow on camphor by a number of enzymes determined by one plasmid, and a plasmid controlling a sequence of genes that allow another strain to grow on salicylic acid has been identified. Since *R* factors have recently been found in *Pseudomonas* strains (which are also being seen more often in postoperative infections), it would not be too surprising if much of the nutritional versatility of *Pseudomonas* turns out to depend on a series of plasmids, each controlling enzymes that make it possible for the organism to grow on unusual organic compounds.

Plasmids, which are being found in an increasing number of bacterial species, share many properties with bacterial viruses, particularly the "temperate" type of bacterial virus, which in the latent state establishes a stable symbiotic relation with its host bacterium. Bacterial plasmids can be thought of as symbionts without even the limited lethal effects of temperate viruses, whose spread depends on the eventual destruction of their host cells. Since it has been recognized that certain components of plant and animal cells, such as chloroplasts and mitochondria, also rely for their maintenance partly on extrachromosomal DNA, the possibility exists that plasmids not only may be common in many bacterial species and be concerned with activities other than antibiotic resistance but also may be closely related to other extrachromosomal elements in all cells.



PLASMID AGGREGATE is an *R* factor composed of several independent plasmids. Here the original donor is resistant to streptomycin (*S*), ampicillin (*A*) and tetracycline (*T*) and has a transfer factor (delta). On conjugating it gives rise to three kinds of new donor, each carrying infectious resistance to one antibiotic. When these donor cells conjugate, the strains that result either have the same characteristics as the donor or have noninfectious resistance (either *S* or *A*) or have only the transfer factor (delta); tetracycline resistance, however, is always infectious. These properties have now been explained by the presence of four distinct plasmids: delta, measuring 60 megadaltons; *S* and *A*, each of six megadaltons, and delta-*T*, a plasmid cointegrate of 67 megadaltons.

THE BRIGHTEST INFRARED SOURCES

Certain celestial objects radiate thousands of times more energy at infrared wavelengths than the sun does at all wavelengths. They evidently include stars being born and the debris of dying stars

by G. Neugebauer and Eric E. Becklin

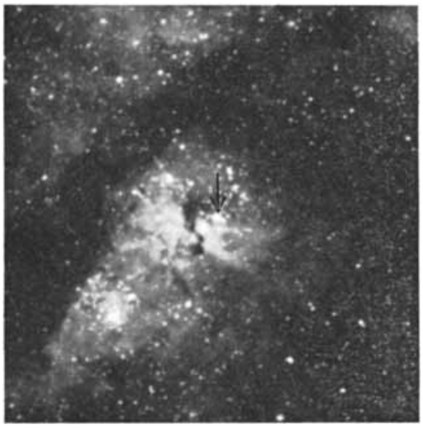
During the past century large telescopes equipped with photographic plates have provided most of what we know about normal stars, stars whose temperatures range from 2,000 degrees Kelvin, or a third the temperature of the sun, up to 50,000 degrees K. Until the past 10 years little was known about the possible existence of stars and other celestial objects whose temperatures might range from 2,000 degrees K. down to room temperature or even lower. The reason is that objects in this temperature range emit the bulk of their energy at wavelengths between one micron and 1,000 microns (one millimeter), which lie in the infrared part of the electromagnetic spectrum and are therefore largely blocked by the earth's atmosphere.

To observe objects that radiate predominantly in the infrared one must either make measurements with special detectors through the few narrow infrared "windows" that exist in the atmosphere or send instruments above the atmosphere in balloons or rockets. If one does so, one finds that even at a wavelength of 20 microns the brightest objects are no longer ordinary stars. Instead the principal sources of 20-micron radiation are large clouds of dust and gas. In some cases the clouds appear to have been ejected from old, evolved stars; in other cases the clouds seem to represent protostars: cool, tenuous masses of dust and gas in the earliest stages of stellar evolution. The nucleus of our galaxy, which is invisible even with the largest optical telescopes, is actually a conglomerate of millions of stars. It stands out so conspicuously at wavelengths of 20 and 100 microns, however, that it seems likely that sources more exotic than ordinary stars are contributing to the infrared emission.

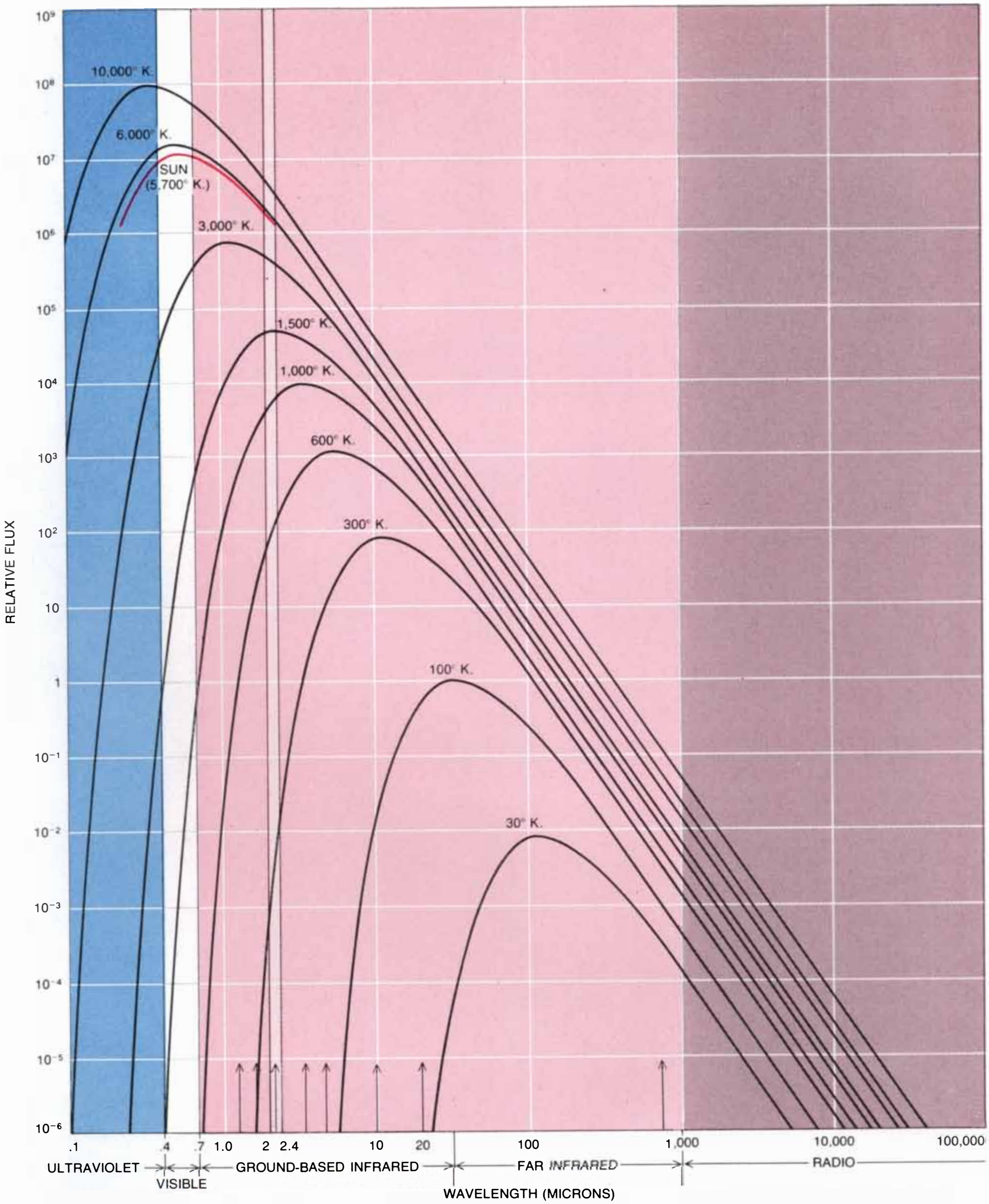
Since the sky presents a very different appearance when it is observed at different wavelengths, it is necessary to conduct surveys with instruments of different spectral sensitivity in order to sample as many different kinds of object as possible. The choice of observational wavelengths, however, is determined less by what one would prefer than by what is technically feasible. Between one micron and 20 microns there are only seven windows in which the atmospheric transmission exceeds 50 percent. At longer wavelengths there are poor windows near 35, 350 and 450 microns that are just beginning to be explored from very dry sites, such as the Mauna Kea Observatory in Hawaii. Finally at 800 microns a broad transmission window opens and extends to wavelengths that are accessible to radio telescopes. The important region between 50 and 300 microns, totally obscured at ground level, has been explored in the past two years from high-flying aircraft, balloons and rockets.

It was only five years ago that we learned from the California Institute of Technology infrared survey how the northern three-fourths of the sky appears at a wavelength of two microns [see "The Infrared Sky," by G. Neugebauer and Robert B. Leighton; Scientific American, August, 1968]. Only a small fraction of the 6,000-odd stars visible to the unaided eye are prominent at two microns. The brightest objects at that wavelength have temperatures that fall primarily between 1,000 and 4,000 degrees K. and thus radiate very little of their total energy in the visible part of the spectrum. In the several years since the Cal Tech survey was conducted the sensitivity of instruments for making infrared observations has increased nearly tenfold. In the same period the number of astronomers observing in the infrared part of the spectrum has grown significantly.

In this article we shall undertake to summarize the current thinking about various types of infrared objects by describing the brightest and most conspicuous sources in the sky as they are observed at two, 20 and 100 microns. These sources encompass a wide range



BRIGHTEST INFRARED SOURCE at a wavelength of 20 microns is associated with Eta Carinae, a star embedded in a well-known emission nebula in the southern half of the sky. The nebula appears in the color photograph on the opposite page. The location of Eta Carinae is indicated by an arrow in the black-and-white photograph above, which represents a small section of the color photograph. Eta Carinae flared up brilliantly in 1843, but is now almost invisible to the unaided eye. Infrared measurements indicate that Eta Carinae is surrounded by a cloud of dust about a tenth of a light-year across that is heated to 250 degrees Kelvin (some 20 degrees Celsius below the freezing point of water). An object at that temperature has a peak emission at about 15 microns. The central star is probably hotter than 10,000 degrees K. The color photograph was taken by G. Araya at the Cerro Tololo InterAmerican Observatory in Chile.



WAVELENGTH OF RADIATION emitted by heated bodies varies with the absolute temperature. The wavelength of the most intense emission is given by Wien's displacement law, which states that for a thermal radiator the product of the peak wavelength and the absolute temperature is a constant. In the units used here the constant is about 3,000 micron-degrees. Thus the sun, whose temperature is about 5,700 degrees K., has a peak emission near .5 micron. The eye responds only to the narrow band of wavelength between .4 micron (violet) and .7 micron (red). The arrows along the bottom of the diagram indicate the center of narrow infrared "windows" where the atmosphere allows the passage of 50 percent or more of the incident radiation. There are poor windows near 35, 350 and 450 microns that can be exploited from observatories in dry locations. The atmosphere becomes essentially transparent to incoming radiation at about 800 microns. Astronomers use photographic emulsions and photomultipliers out to one micron in the near-infrared. Infrared techniques are used between one micron and a few thousand microns, where they overlap with radio techniques.

of objects and include a sample of practically every interesting infrared source currently being studied in detail.

In most cases the radiant energy emitted by a heated body has a continuous spectrum with a broad peak whose maximum is related to the body's absolute temperature. The wavelength of the most intense emission can be computed by Wien's law, which states that for a thermal source the product of the absolute temperature and the wavelength of the maximum spectral intensity is a constant. This constant has a value of about 3,000 micron-degrees. Thus the sun, which has a surface temperature of a little less than 6,000 degrees K. (actually about 5,700 degrees), radiates most intensely at about .5 micron. The visible spectrum extends from .4 micron (violet) to .7 micron (red). Objects at room temperature (300 degrees K.) radiate most intensely at about 10 microns [*see illustration on opposite page*].

As with visible celestial objects, infrared sources can appear bright either by virtue of their closeness or, if they are far away, by virtue of their high intrinsic luminosity. Luminosity refers to the total power radiated by an object at all wavelengths.

Stars in our galaxy have luminosities that range from one ten-thousandth to one million times the luminosity of the sun. The effect of closeness is well illustrated by the sun; although it is hot and thus relatively weak in the infrared, it is so close that it is by far the brightest source in the sky at all infrared wavelengths. Moreover, the moon and several of the planets (at the right phases) are conspicuous objects in the infrared because of both their closeness and their temperature. We shall not, however, deal with them further here.

Infrared observations at wavelengths between one micron and five microns have now become fairly routine. The most commonly used detectors are lead sulfide photoconductors in which charge carriers (electrons and "holes") are knocked into conduction bands of the crystal by incident photons (quanta of radiation). The cells are cooled to 77 degrees K. by liquid nitrogen to decrease the number of random electrons that jump into the conduction bands without being hit by a photon.

We know from the Cal Tech survey that the brightest star at two microns observable from southern California is Betelgeuse, the well-known red star that is the second-brightest object in the constellation Orion and the 12th-brightest among all stars at visual wavelengths.

STAR	RELATIVE BRIGHTNESS	APPARENT TEMPERATURE (DEGREES KELVIN)
1. SIRIUS	6.6	10,000
2. CANOPUS	3.4	7,000
3. ALPHA CENTAURI	2.2	6,000
4. ARCTURUS	1.8	4,000
5. VEGA	1.7	10,000
6. CAPELLA	1.6	5,000
7. RIGEL	1.6	12,000
8. PROCYON	1.3	6,000
9. ACHERNAR	1.2	14,000
10. BETA CENTAURI	1.0	20,000

BRIGHTEST VISIBLE STARS have surface temperatures that range from 4,000 to 20,000 degrees K. All but four are significantly hotter, hence bluer, than the sun. Only the coolest, Arcturus, emits strongly enough in the infrared to rank among the 10 brightest objects observable at a wavelength of two microns (*see list below*). The star with the highest absolute luminosity on the list, Rigel, emits more than 10,000 times as much energy as Alpha Centauri, our nearest stellar neighbor, which is only slightly more luminous than the sun.

OBJECT	TYPE	RELATIVE BRIGHTNESS	APPARENT TEMPERATURE (DEGREES KELVIN)
1. BETELGEUSE	S.G.	3.0	3,000
2. R DORADUS	L.P.V.	3.0	2,000
3. ANTARES	S.G.	2.5	3,000
4. ALPHA HERCULES	S.G.	1.9	3,000
5. W HYDRAE	L.P.V.	1.9	2,000
6. MIRA	L.P.V.	1.2	2,000
7. GAMMA CRUSIS	G.	1.2	3,000
8. ARCTURUS	G.	1.2	4,000
9. CHI CYGNI	L.P.V.	1.0	2,000
10. ALDEBARAN	G.	1.0	3,500

BRIGHTEST OBJECTS AT TWO MICRONS include three supergiant stars (*designated S.G.*), three giants (*G.*) and four long-period variables (*L.P.V.*). Betelgeuse, the second-brightest star in the constellation Orion, ranks 12th in brightness among all visible stars. Because of its low temperature it is distinctly reddish. The list is complete for the northern sky but is incomplete for the southern hemisphere. Nevertheless, R Doradus and Gamma Crucis in the southern sky are known to be very bright at two microns, hence are included.

OBJECT	RELATIVE FLUX	SIZE (ARC SECONDS)	APPARENT TEMPERATURE (DEGREES KELVIN)
1. ETA CARINAE	40	6	250
2. OMEGA NEBULA	35	500	250
3. GREAT NEBULA IN ORION	25	100	100-500
4. IRC +10216	10	2	500
5. VY CANIS MAJORIS	5	<2	500
6. G 333.6-0.2	3	30	200
7. CENTER OF GALAXY	2	40	200
8. NML CYGNI	2	<2	500
9. IRC +10420	2	<2	250
10. W3	2	100	150-350

BRIGHTEST OBJECTS AT 20 MICRONS are no longer ordinary stars. The list may omit some bright sources since a complete sky survey has not yet been made at this wavelength. Six of the 10 sources have a substantial angular size. Several radiate more than 10,000 times as much energy as the sun does, yet are below room temperature (300 degrees K.). The relative fluxes are given on a scale in which 1 equals the brightness of Betelgeuse at 20 microns.

In the southern sky R Doradus, which is just visible under favorable conditions, is about as bright as Betelgeuse at two microns [*see top and middle illustrations on preceding page*].

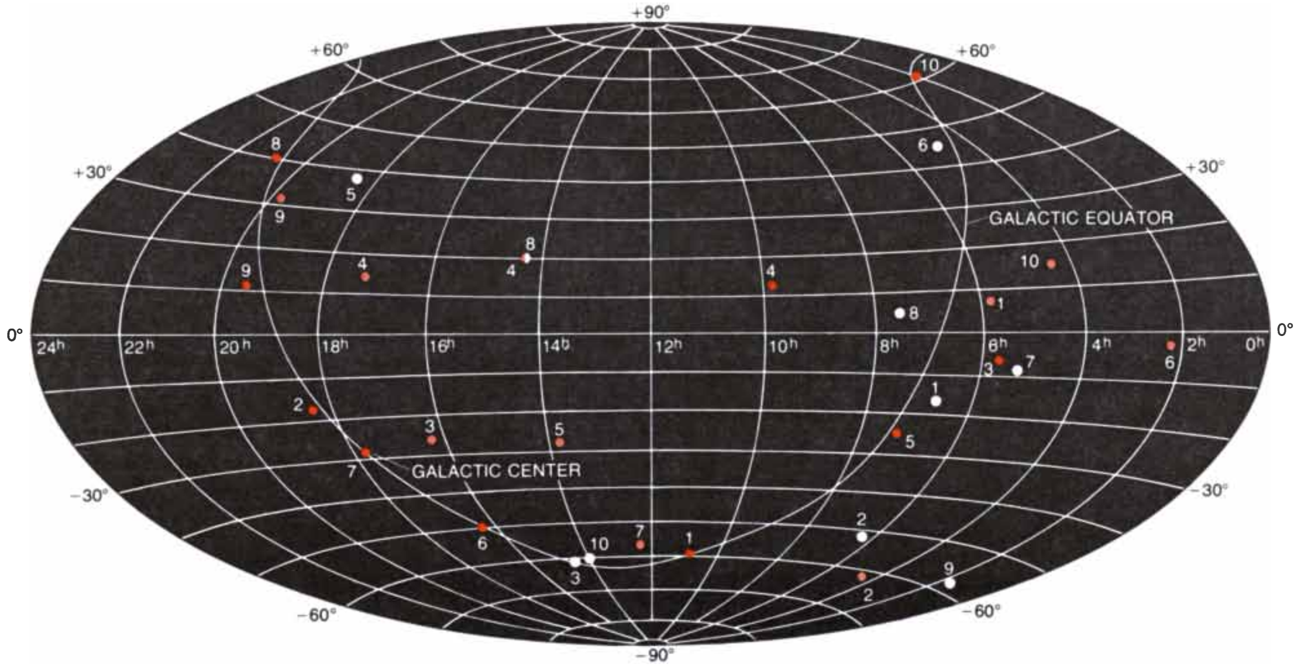
The brightest sources at two microns are well-understood objects. Betelgeuse is visibly red because its temperature is only about 3,000 degrees K., which places it among the coolest stars visible to the unaided eye. It is a very luminous supergiant, a star that has burned much of its hydrogen and has dropped in temperature while simultaneously expanding and becoming more luminous. R Doradus, on the other hand, is a long-period variable star whose energy output changes by a large factor in the course of about a year. It is as bright as Betelgeuse at two microns not only because of its intrinsic luminosity but also because at 2,000 degrees nearly 15 percent of its total energy is emitted in the band between two and 2.4 microns. The third- and fourth-brightest stars at two microns, Antares and Alpha Herculis, are visually bright stars in Scorpio and Hercules and are also luminous red supergiants. In contrast, the fifth-brightest star on the two-micron list, W Hydrae, is too faint to be easily seen with the unaided eye. Like R Doradus, W Hydrae is a long-period variable star. The infrared radiation emitted by all the bright two-micron sources originates in the luminous gas that forms the outer envelope of the stars.

At 20 microns the brightest sources are much more varied and complex than those that stand out most brilliantly at two microns. Furthermore, it is significantly more difficult to make observations at 20 microns than it is at two microns. At 20 microns the most commonly used detectors are heat detectors—really thermometers. They are often made of semiconducting germanium "doped" with a suitable impurity, such as gallium, that yields a crystal whose electrical resistance changes with temperature. Such detectors must be cooled to about two degrees K. with liquid helium.

The principal difficulty in making astronomical observations at 20 microns is that objects at room temperature emit about half of their thermal energy in the region between nine and 20 microns. Thus the telescope and the atmosphere itself radiate strongly at just the wavelengths one wants to observe. A side effect of the large background flux at 10 to 20 microns is that if more than a small area of the sky is observed, temperature fluctuations in the telescope and in the atmosphere introduce prohibitive amounts of noise. As a result ground-based sky surveys in this wavelength region have not been notably successful. Within the past year, however, Russell G. Walker and Stephan D. Price of the Air Force Cambridge Research Laboratories have surveyed much of the sky at four, 10 and 20 microns with a helium-cooled telescope in a rocket.

A list of the 10 brightest sources at 20 microns, compiled from our own data and from the work of others, includes none of the sources on the list of objects that are brightest at two microns [*see bottom illustration on preceding page*]. Although the list will almost certainly be altered by the discovery of other strong 20-micron sources, it is doubtful that a comprehensive sky survey would include any new class of objects. In fact, most of the bright 20-micron sources in the northern three-fourths of the sky were picked up by the two-micron survey; at two microns, however, the 20-micron sources are generally inconspicuous. Nevertheless, all but one of the bright 20-micron objects were selected for special study on the basis of some unusual characteristic, such as color or association with either a visible feature or a radio feature. Therefore they were known to be bright at 20 microns before the rocket survey. This repeats the experience of the two-micron survey: all the brightest two-micron sources had previously been observed in the visible region and were known to be strong infrared emitters.

Perhaps the most striking feature of the strong 20-micron sources is that they



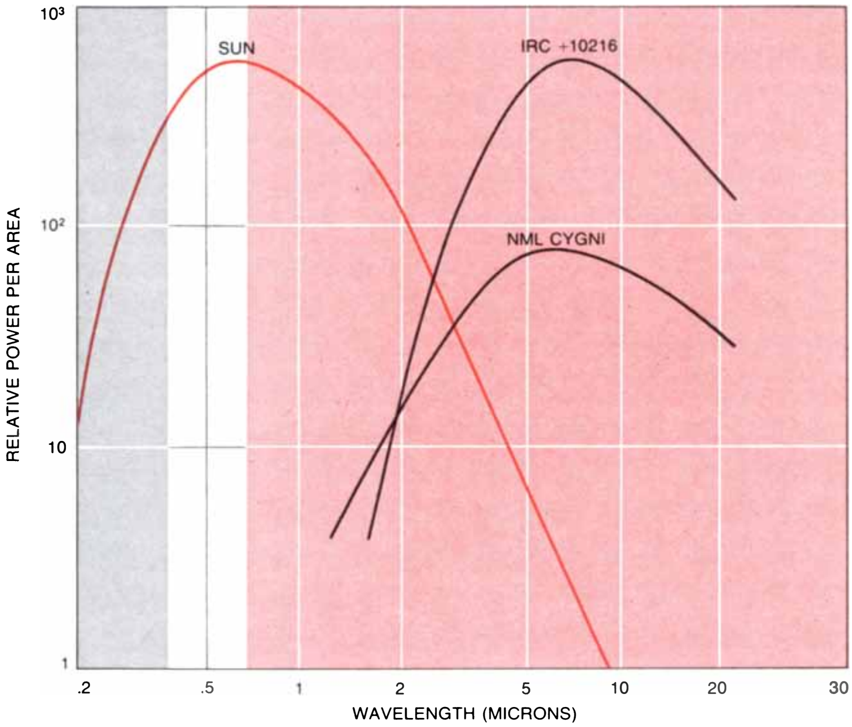
BRIGHTEST VISIBLE AND INFRARED OBJECTS are plotted on an equal-area projection of the celestial sphere. The 10 brightest visible stars are shown as white dots; the 10 brightest objects at two microns are shown as pink dots. Arcturus, which ranks No. 4 among visible objects and No. 8 among two-micron sources, is half white and half pink. The 10 brightest sources at 20 microns are shown as dark red dots. Unlike the two-micron sources, which are all less than one second of arc in diameter, several of the 20-micron sources are complex objects with diameters ranging up to seven minutes of arc; all but one of them are near the galactic equator.

display a large range of angular sizes. Whereas all the bright two-micron sources are single objects less than one second of arc in diameter, the bright 20-micron sources are in some cases complex objects with a diameter of several minutes of arc. As a consequence the ordering of any list of bright 20-micron objects must depend on the angular size specified.

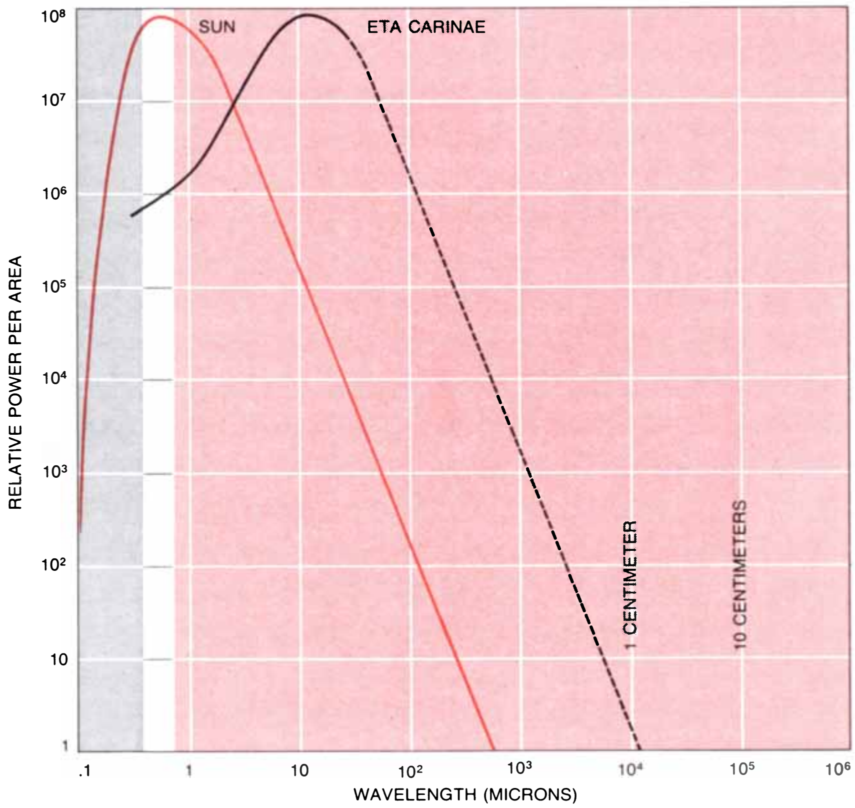
Like the bright two-micron sources, all the major 20-micron sources are within our galaxy, but most of them are much more distant than the two-micron sources. The brightest visible and two-micron stars lie within about 1,000 light-years of the sun. In contrast the brightest 20-micron sources lie between 1,000 and 30,000 light-years. It follows that the 20-micron sources must have high intrinsic luminosities, generally more than 10,000 times the luminosity of the sun. The bright 20-micron sources span a wide range of ages and include, we think, both dying stars and stars just being born.

The object designated IRC +10216, fourth in our list of brightest 20-micron sources, appears to be typical of a large class of infrared objects. It is barely visible even on red-sensitive photographs taken with large telescopes and thus has no name other than its listing in the Cal Tech infrared survey. (IRC +10216 stands for the 216th source listed in the Cal Tech Infrared Catalogue in the declination band centered on +10 degrees.) The spectral distribution of the energy emitted by IRC +10216 closely follows the curve that would be produced by a thermal source with a temperature of 600 degrees K.; hence the peak of its emission comes at about five microns [*see top illustration on this page*]. A temperature of 600 degrees K. (+327 degrees Celsius) is almost exactly the melting point of lead. Although the flux of radiation from IRC +10216 between two and 20 microns is devoid of spectral features, spectra taken at about .8 micron show bands produced by cyanogen, a simple molecule consisting of two atoms of carbon and two of nitrogen. This indicates the presence of a star rich in carbon whose temperature is about 2,000 degrees K.

One model that explains most of the properties of IRC +10216 consists of a carbon-rich star surrounded by a shell of dust particles. Interstellar and circumstellar dust particles, which typically are smaller than a micron, absorb visible and ultraviolet photons efficiently but are quite transparent to infrared photons. A dense shell made of such



TWO OBJECTS BRIGHT AT 20 MICRONS, IRC +10216 and NML Cygni, are thought to be stars surrounded by thick shells of dust that absorb most of the short-wavelength radiation from the central star. The heated dust produces a spectral distribution with a peak near five microns. In this illustration and three similar ones including the one below, the curve for the sun is included only to show the spectral distribution of the sun's energy and is not to be compared with the curves for the infrared objects on the relative flux scale.



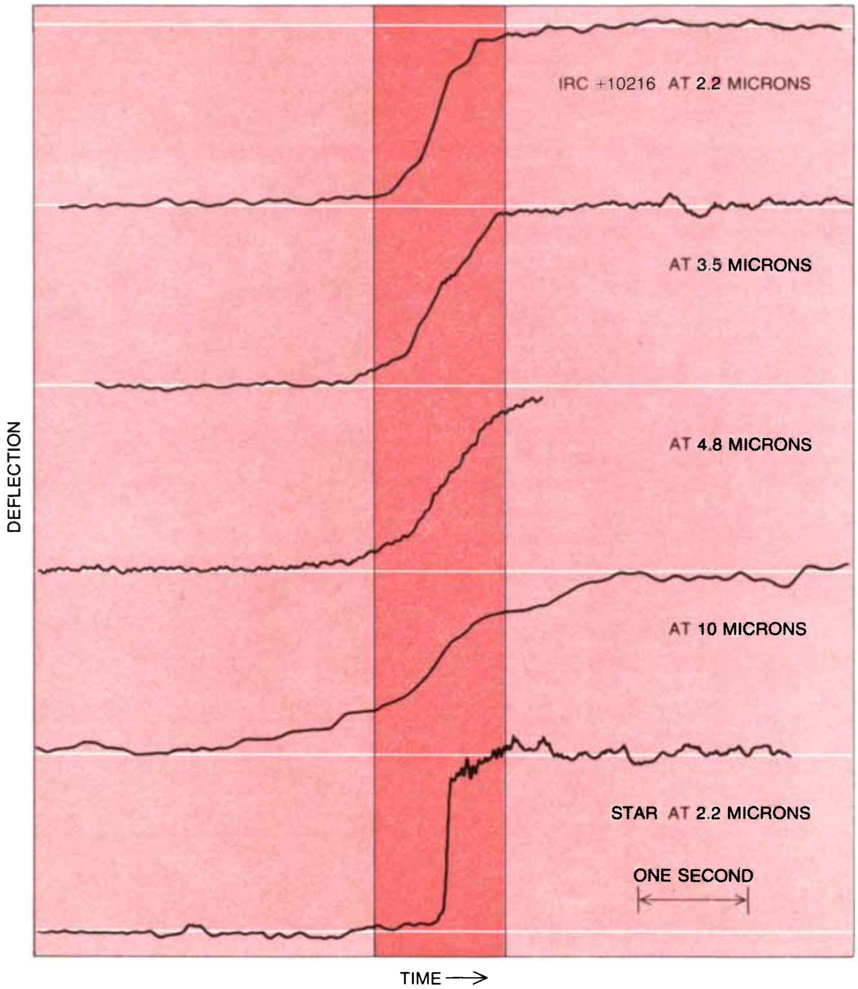
BRIGHTEST 20-MICRON SOURCE, Eta Carinae (*see photograph on page 29*), has a total emission integrated over all wavelengths exceeding that of the sun by a factor of more than a million. The portion of the curve shown in broken line is a guess; no measurements exist.

particles would absorb enough of the visible radiation from a central star to hide it. The radiation absorbed by the dust particles would heat the shell to a temperature of about 600 degrees K., making it glow in the infrared.

It was possible to test this model directly when IRC +10216 was occulted by the moon. Measurements were made at 2.2, 3.5, 4.8 and 10 microns of the time required for radiation to reappear when the source emerged from behind the moon. The results showed that most of the radiation comes from a disk .4 second of arc in diameter, or essentially the size predicted for an opaque dust shell radiating at about 600 degrees K. [*see illustration below*]. A reasonable guess for the distance of IRC +10216 is 1,000 light-years, in which case the radius of the shell is about 60 astronomical units (60 times the mean distance from the earth to the sun), or about the diameter of Neptune's orbit. Surprisingly, the occultation measurements indicate that at 10 microns about half of the flux comes from a more transparent outer shell of dust that is some five times larger than the inner shell. The outer shell has a temperature of only about 300 degrees K.

The general picture of an absorbing dust shell that is heated by a central star and reemits thermal radiation in the infrared apparently explains many of the sources that are bright at 10 and 20 microns. IRC +10216 is thought to represent an extreme example, with a very opaque shell, but evidence for similar shells has been found associated with stars of diverse characteristics. Although the precise chemical composition of the dust remains to be established, the energy distribution observed in several cases seems to point to particles made up of silicates.

The source of the dust is also open to speculation; we cannot say whether it represents matter ejected from the star or whether it is accreted from material in interstellar space. Both possibilities must be considered. In several instances the shells seem to be associated with stars that are believed to be losing mass. Moreover, often (as in the case of IRC +10216) there is no evidence for nearby interstellar dust. In other instances, however, the stars with shells are in regions that are obviously dusty.



SIZE OF INFRARED SOURCE IRC +10216 was determined when it was occulted by the moon. The first four curves show the time required for the radiation at four different wavelengths to return to full strength as the source emerged from behind the moon. The bottom curve shows how the 2.2-micron curve looks when the object occulted is an ordinary star. The time of emergence of IRC +10216 at a few microns corresponds to the time (*shaded band*) that would be required for a disk .4 second of arc in diameter. Assuming IRC +10216 to be about 1,000 light-years distant, such a disk corresponds to a source, presumably a vast shell of dust, whose radius is 60 astronomical units, or 60 times the distance from the earth to the sun. At 10 microns about half of the flux comes from a cooler and more transparent shell about five times larger. The shells appear to surround a star of 2,000 degrees K.

A visually beautiful example of a star with a dust shell is Eta Carinae [*see illustration on page* 29]. Located deep in the southern skies in a nebulous region, Eta Carinae itself is a tiny nebulosity a few seconds of arc across. Recognized in the early 19th century as a faint variable star, it flared up dramatically in 1843, rivaling Sirius as one of the brightest stars in the sky. Subsequently it faded until in this century it has been almost invisible to the unaided eye. Observed in the infrared at 20 microns, however, it is the brightest object we know outside the solar system.

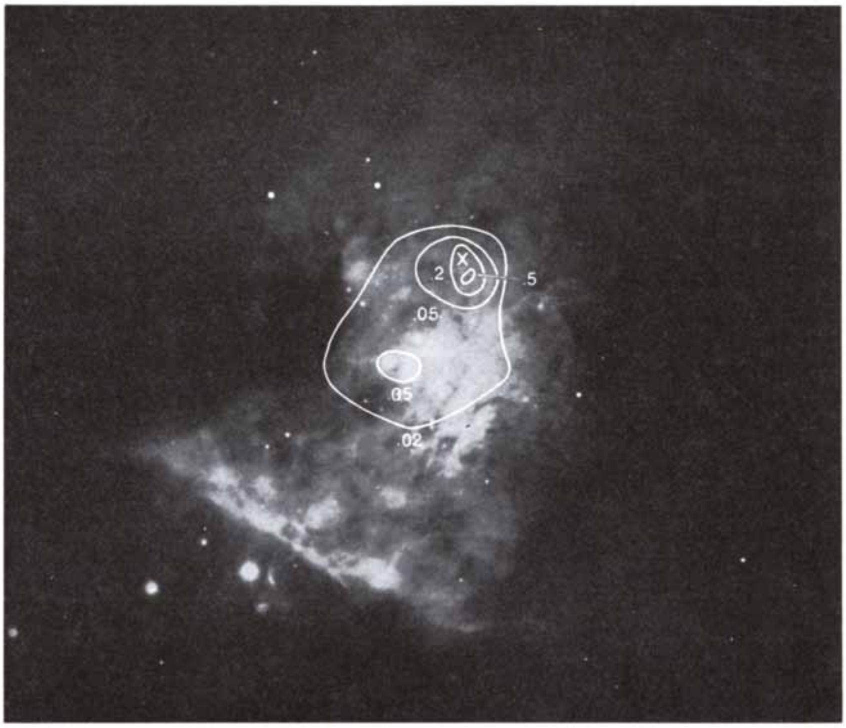
The total energy output of Eta Carinae can be estimated from observations at several wavelengths out to 20 microns and by making reasonable guesses for the shape of the curve beyond 20 microns [*see bottom illustration on preceding page*]. It appears that the total power emitted now is essentially equal to the output of the star at the time of its flare-up in 1843. It is attractive to speculate that material ejected from the central star is now condensing into a dust shell. Indeed, photographs taken a few decades apart do indicate that visible condensations are still moving from the central core. The estimated mass of the material in the shell is about a tenth the mass of the sun.

Eta Carinae is large enough so that one can scan across it and determine its size at various wavelengths. At two microns the source is about one second of arc in radius; at 10 and 20 microns the radius is some three times larger. Apparently at 20 microns one is sampling a cool outer shell with a temperature of about 250 degrees K. and a radius of 6,000 astronomical units, or about a tenth of a light-year. At two microns one is evidently seeing dust particles that are hotter (about 450 degrees K.) because they are only 2,000 astronomical units from the stellar heat source. The central source is probably hotter than 10,000 degrees K. and is possibly two million times more luminous than the sun.

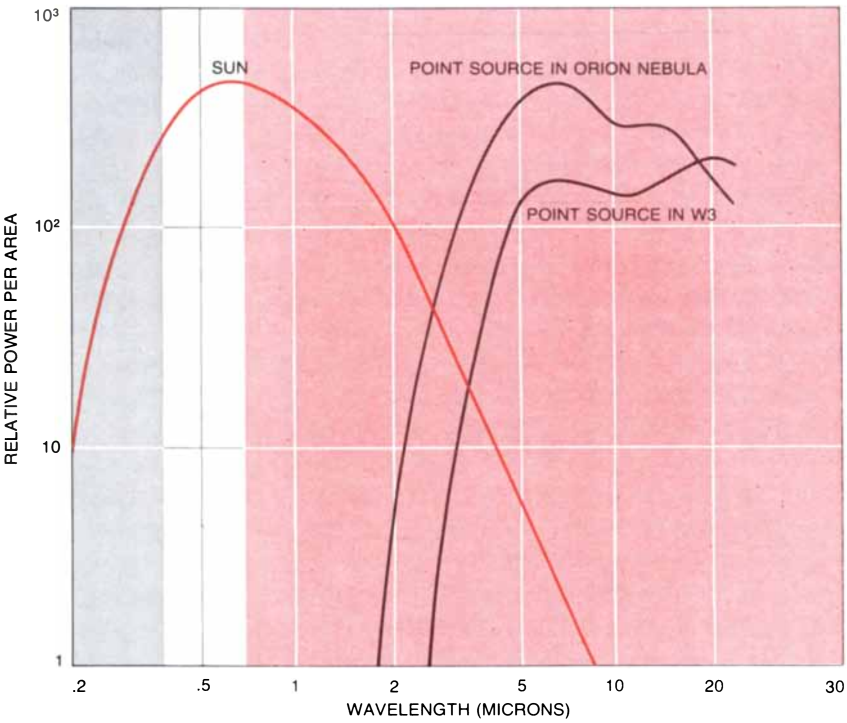
It has been speculated on the basis of photographic observations that Eta Carinae is a massive supergiant star that is rapidly evolving by the loss of mass and by nuclear processes. Another suggestion is that Eta Carinae may be a "slow" supernova, a more massive version of the supernova in the Crab Nebula that was observed by Chinese astronomers in A.D. 1054. Since the size of Eta Carinae varies with wavelength, one must conclude that whatever the ultimate source of energy, the infrared radiation almost surely comes from heated particles of dust.

Among the 10 brightest 20-micron sources three others resemble IRC +10216 and Eta Carinae in having a diameter of less than 10 seconds of arc. Thus it seems likely that all five are stars surrounded by glowing shells of dust. The three remaining small-diameter sources are of interest in their own right for other reasons. For example, NML Cygni and VY Canis Majoris are probably supergiants surrounded by thick dust shells. Not only are they bright at 20 microns but also they emit intense radio waves near 18 centimeters. Such emission is evidently produced by a maser amplification process involving the hydroxyl (OH) radical [see "Interstellar Molecules," by Barry E. Turner; SCIENTIFIC AMERICAN, March]. In addition VY Canis Majoris can be observed at visible wavelengths and has long been known for its complex structure and its association with nebulosity. The first hint of the brightness of IRC +10420, which showed up as a weak source in the Cal Tech two-micron survey, came from the 20-micron rocket survey carried out by Walker and Price. Except for its lesser strength the infrared emission of IRC +10420 closely duplicates the energy distribution of Eta Carinae. It is probably a similar kind of object, although it is not in a nebulous region.

The remaining half of the 10 brightest 20-micron sources range in diameter from about 30 seconds of arc to 500. All the sources except the center of the galaxy are identified with H II regions: regions containing ionized hydrogen atoms (bare protons) and electrons. The hydrogen atoms are ionized by ultraviolet photons emitted by very young hot stars. The H II regions also contain dust that is heated by the ultraviolet radiation with-



TWENTY-MICRON FLUX IN ORION NEBULA is superposed on a photograph taken at the Lick Observatory. The innermost contour is equal to half the 20-micron flux of Betelgeuse as measured with an aperture of five arc seconds. The outermost contour equals a fiftieth the flux of Betelgeuse. Trapezium stars are within lower .05 contour. *X* denotes a point source that seems to be a protostar with a surface temperature of about 600 degrees K.



WAVELENGTH DISTRIBUTION OF TWO POINT SOURCES, one in Orion and one in a region known as W3 (*see photograph on next page*), shows similar characteristics. The peak emission at about seven microns in the curve for the Orion source indicates that it has a temperature of about 600 degrees K. The peak emission at about 18 microns in the W3 source indicates that its temperature is no more than about 350 degrees K. Nevertheless, the total energy output of the W3 source is 30,000 times that of the sun. As noted under the first illustration of this type on page 33, the vertical positioning of the solar curve is arbitrary.

in the nebula; the heated dust presumably emits the infrared radiation. Some H II regions, such as the Great Nebula in Orion, are conspicuous visible objects as well as strong radio sources. Other H II regions are not visible optically because of obscuration by interstellar dust and are identified only through their radio emission; the presence of the hot ionizing star is merely inferred.

The 20-micron radiation from the Orion nebula comes from an enormous region roughly one light-year in diameter. Within this region the radiation appears to be concentrated in two distinctly different areas [*see top illustration on preceding page*]. One concentration is centered on a bright group of four stars known as the Trapezium, but it clearly extends beyond the group. Evidently the infrared radiation in this region arises from the nebular dust, which has been heated to between 100 and 200 degrees K. by the ultraviolet and visible radiation originating in the Trapezium stars. Two observations support this notion. First, the emission near 10 microns includes a broad spectral feature that resembles the spectral emission produced by silicate dust particles. Second, the observed temperature decreases with distance from the Trapezium stars in a manner expected for dust particles heated by stellar radiation.

The second strong concentration of infrared radiation in the Orion nebula is about one minute of arc northwest of the Trapezium. Curiously, this source is not associated with visible bright stars. Furthermore, there is no indication from radio observations of any concentrations of the ionized gas one would expect near luminous hot stars. At about 20 microns, however, the flux from the second region is greater than the flux from the Trapezium region!

What is the origin of this intense flux?

REGION AROUND W3 is shown in this photograph from the sky survey made with the 48-inch Schmidt telescope on Palomar Mountain. The box encloses the region of maximum brightness at radio wavelengths. At visual wavelengths the region is almost completely obscured by dust. A comparison of the radio emission and the infrared flux at 2.2 and 20 microns is given on the opposite page.

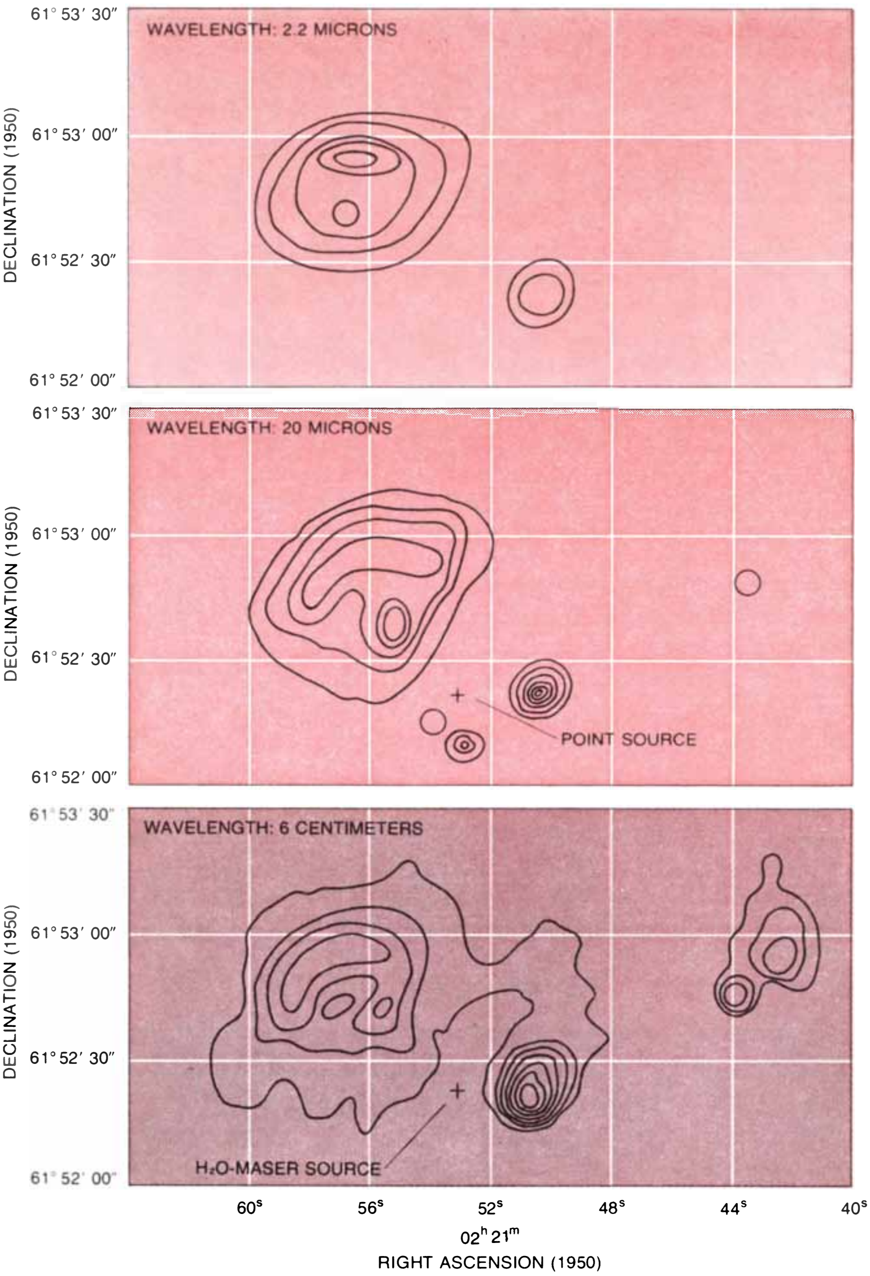
At five and 10 microns most of the radiation comes from an unresolved point source that is less than two seconds of arc in diameter and is, as far as we can tell, starlike. In contrast, D. E. Kleinmann and Frank J. Low of the University of Arizona found that at 20 microns the strongest signal comes from an extended infrared nebulosity.

In initiating infrared observations one of our earliest hopes was to discover evidence for a protostar, that is, an object being formed from the condensation of dust and gas. A protostar would be expected to glow in the infrared by virtue of the energy released by gravitational collapse; nuclear burning would at most have barely begun. One of our first thoughts was that the pointlike source in the Orion nebula was a protostar with a surface temperature of about 600 degrees K. Since the Orion nebula cannot be much more than 10,000 years old, we discounted the possibility that the energy could come from an evolved dust-embedded star like IRC +10216, which must be many millions of years old.

For a time a rival hypothesis seemed more attractive than the protostar idea. Perhaps we were observing a bright, hot supergiant with a temperature of about 7,000 degrees K. lying deep within the Orion nebula and so shrouded by dense clouds of dust that its visible light would be diminished by the enormous factor of perhaps 10^{30}. (If the light of the sun were reduced by such a factor, it could not be seen from the earth even with the 200-inch Hale telescope.) Radio observations at wavelengths of about two millimeters established that gas and dust are indeed present. They are revealed by emission lines characteristic of the formaldehyde molecule (H_2CO). The characteristic signature of absorption by silicate dust particles has also been found at about 10 microns.

Quite recently we have made scans at 20 microns with a spatial resolution high enough to isolate the "point" source from the surrounding nebula. We find that the starlike source does emit strongly at 20 microns in its own right. In fact, the flux at this wavelength is so intense that if it were being produced by a giant hot star whose radiation at ultraviolet and visible wavelengths was being almost totally blocked by dust, the hidden source would have to be from 10 to 100 times more luminous than any known supergiant. Since this seems unlikely we have returned to the view that the point source is probably a protostar in the process of collapsing from the cold interstellar material.

Similar high-resolution studies of the infrared nebula to the south of the point source indicate that much of its radiation is concentrated in isolated spots. These spots are probably similar to the point source except that they are embedded in a much thicker envelope of dust. Conceivably we are observing stars in various stages of formation.



INFRARED AND RADIO MAPS OF W3 SOURCE show considerable similarity. The radio flux (*bottom map*) is associated with H II regions, areas surrounding stars so hot they ionize hydrogen atoms into free protons and electrons. The radio map also includes an intense source of 18-centimeter line radiation (*cross*) produced by a maser amplification process involving water molecules. The maser source shows up as a point source at the same location in the 20-micron map. The evidence suggests that the point source is a star in the process of being born. Radio map is by G. Wynn-Williams of the University of Cambridge.

The infrared source that is 10th on the list of the brightest objects at 20 microns lies in an H II region known as W3. It was first catalogued as a strong radio source by Gart Westerhout of the University of Maryland. At visible wavelengths the center of this H II region is completely obscured by dust [*see illustration on opposite page*]. More than half of the radio emission at a wavelength of six centimeters originates in three or four small compact regions. When one compares the emission from W3 at two microns, 20 microns and at the radio wavelength of six centimeters, one can see a considerable similarity in the general distribution of radiation [*see illustration on preceding page*]. All the sources that are common to the infrared and radio maps are consistent with their being compact H II regions embedded in dust. They are evidently very young, and their energy is being supplied by newly formed massive hot stars. Eventually the pressure of the radiation from the stars may blow away much of the dust, leaving an object that is similar in visual appearance to the Orion nebula.

The W3 region also contains a point source that closely resembles the one we think may be a protostar in Orion. The two objects have an energy distribution that is very similar [*see bottom illustration on page 35*]. The point source in W3 is one of the most remarkable objects yet discovered through infrared observations. Although its temperature appears to be no more than 350 degrees K. (well below the boiling point of water), it emits 30,000 times as much energy as the sun. To emit so much energy at such a low temperature the source must be larger than the solar system. It seems clear that nuclear burning has not yet begun in the source; if it had started as recently as 1,000 years ago, an H II region would already have been created and we should therefore be able to observe its radio emission.

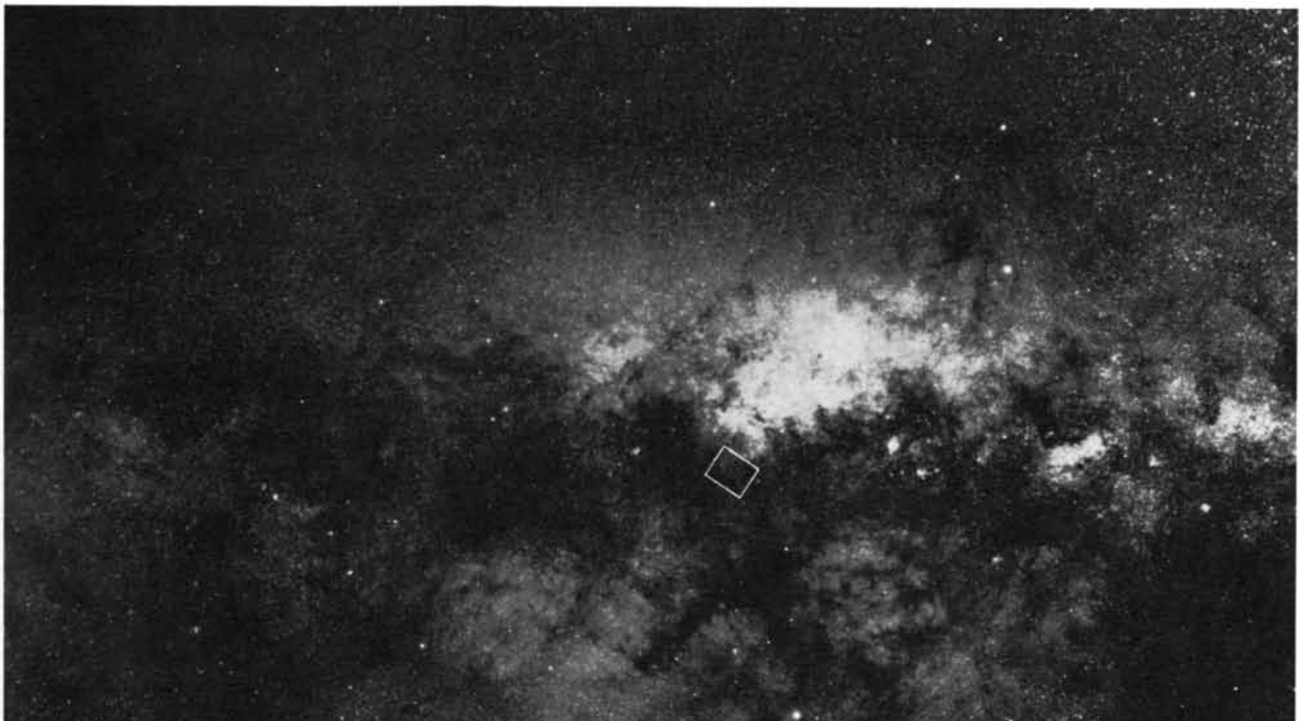
A fascinating property of the W3 point source and the Orion point source is that both are near sources that produce emission lines attributable to maser processes involving either water or hydroxyl molecules. The physical connection between the masers and the infrared sources is still uncertain, but it has been suggested that infrared photons provide the "pump" for the maser phenomenon.

The seventh-brightest source at 20 microns, and one that is possibly unique, is the source or complex of sources located in the nucleus of the galaxy. The center of the galaxy lies in the constellation Sagittarius in a region of the Milky Way that is totally hidden by intervening dust at visible wavelengths [*see illustration below*]. In the infrared the galactic center stands out strongly. At two microns it is likely that we are observing the infrared flux from millions of stars concentrated in the galactic nucleus. The strong radiation observed at 20 microns is too powerful, however, to represent the added output of even millions of stars.

The 20-micron radiation from the galactic nucleus is concentrated in a core six or seven light-years in diameter, corresponding to roughly .02 percent of the distance from us to the nucleus of the galaxy. The energy distribution of this core rises steeply between five and 20 microns and probably peaks at a wavelength beyond 50 microns [*see top illustration on opposite page*].

High-resolution infrared maps made by us with the 200-inch Hale telescope and by Low and G. Rieke at the University of Arizona have shown that the apparently unitary core of the galactic nucleus is in fact made up of at least four separate sources, each with different relative intensities at different wavelengths. At shorter wavelengths the structure of the sources becomes confused with the background radiation of stars at two microns. At wavelengths longer than 20 microns the available instruments do not have the resolution to define the sources clearly [*see bottom illustration on opposite page and illustrations on page 40*].

At present we lack the information needed to understand the sources unambiguously. They could be concentrations of gas, dust and stars in which the dust has absorbed starlight and is reemitting the absorbed energy. In that case the sources would be similar to the other bright 20-micron sources. On the other hand, the radiation could result from some nonthermal process. For example, it could be a form of synchrotron emission: radiation produced by high-



CENTER OF THE GALAXY, marked by a rectangle, is hidden from view by vast clouds of intervening dust. At infrared and radio wavelengths, however, the area inside the rectangle can readily be mapped, as shown in the illustrations on the next two pages. The center of the galaxy is the seventh-brightest source at 20 microns. It is located 27,000 light-years away in the constellation Sagittarius.

energy electrons moving in a magnetic field.

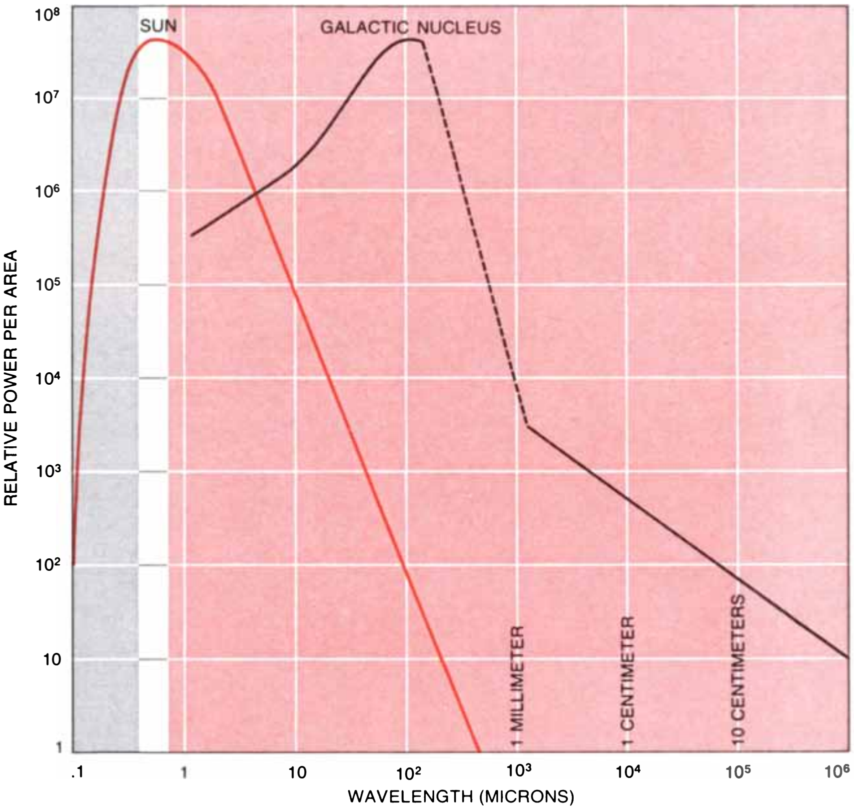
Whatever the source of the infrared radiation, it may serve as a nearby example of processes taking place in the nuclei of other galaxies. Such nuclei exhibit a variety of phenomena, some of which are shared by the nucleus of our own galaxy. Specifically, many galactic nuclei show evidence for strong nonthermal radiation at both visible and radio wavelengths, and in a number of them one can observe outflows of gas and other signs of violent activity. At radio wavelengths the center of our galaxy also shows strong nonthermal emission and outflow.

A number of galaxies exhibit radiation at 20 microns well in excess of that expected from ordinary stars. The emission is particularly prominent in Seyfert galaxies (galaxies that have highly luminous pointlike nuclei) and in certain other types of galaxy. The amount of power such galaxies radiate in the infrared corresponds to as much as 10^{11} times the power output of the sun. This is approximately the amount of power radiated by all the stars in our galaxy at all wavelengths. It therefore seems reasonable to suspect that much of the infrared radiation produced by galaxies that are exceptionally luminous in the infrared must be nonstellar in origin.

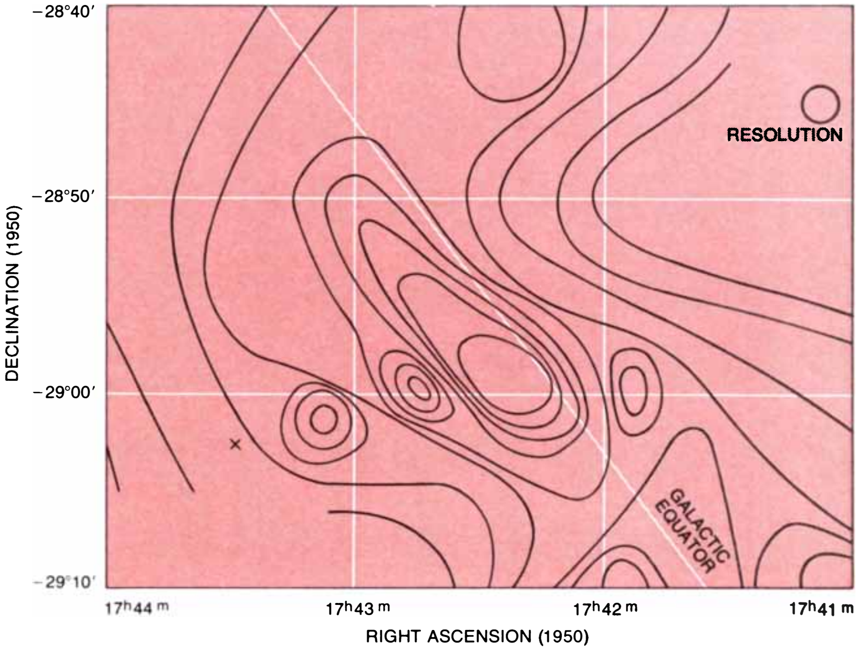
Since the nearest spiral galaxy resembling ours is 1.5 million light-years away and most are much more distant, the infrared signals we receive from them are quite weak. Thus our best hope of understanding the violent events taking place in the center of distant galaxies is to probe the center of our own galaxy with ever higher resolution and at many different wavelengths on the assumption that our galaxy is more or less typical. Enough has already been observed to indicate that the center of our galaxy is far from quiescent.

It is still too early to say much about how the sky looks to instruments sensitive to 100-micron radiation. We cannot even list the strongest sources at this wavelength because no comprehensive survey has yet been made. From limited searches made in the central plane of the galaxy, however, we expect that the H II regions providing some of the brightest objects in the 20-micron sky also dominate the sky at 100 microns.

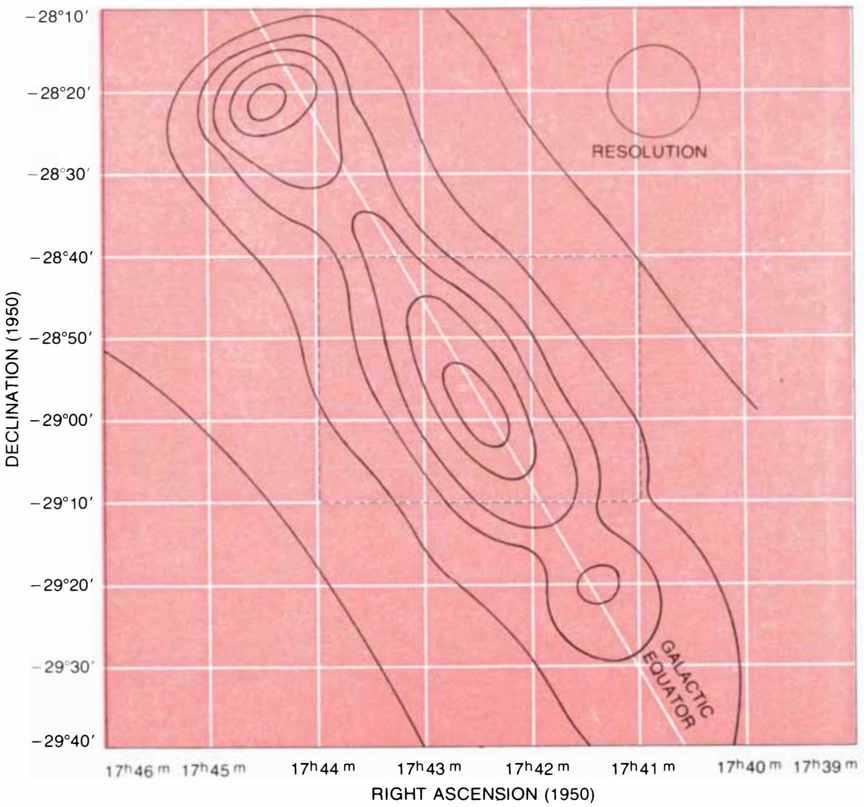
To observe at 100 microns one uses the same kinds of detectors and techniques that are employed at shorter infrared wavelengths. In order to avoid the obscuration caused by water vapor in the atmosphere, however, it is neces-



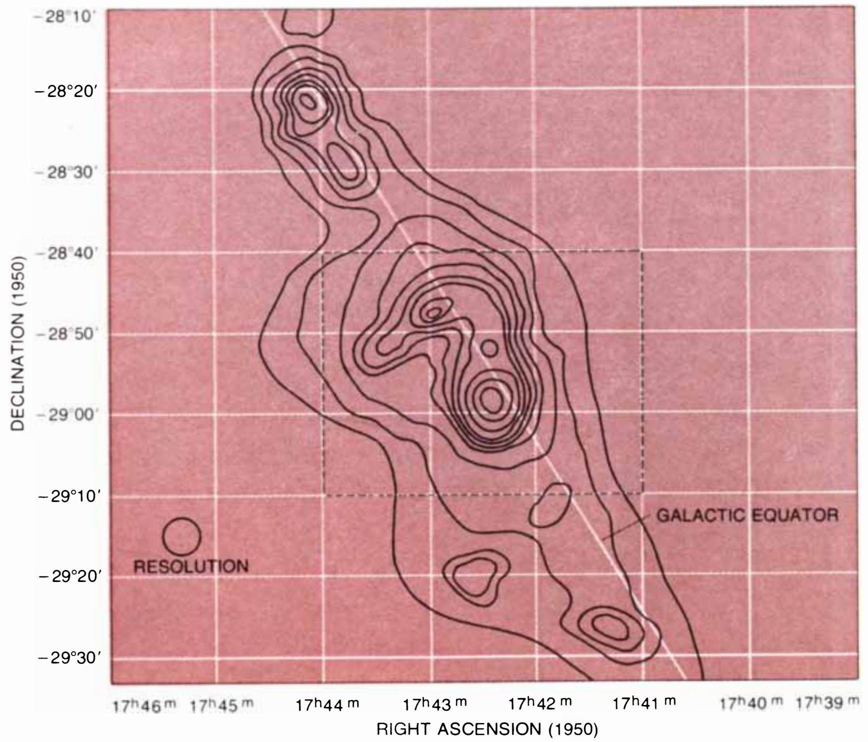
WAVELENGTH DISTRIBUTION OF GALACTIC CENTER is now reasonably well established except for the region between 200 microns and 1,000 microns (one millimeter). The energy emission beyond 1,000 microns has been extensively studied by radio astronomers. At two microns the emission from the galactic center probably represents the summed output of millions of stars. It seems, however, that the intense flux at longer wavelengths must arise from other sources since it is much too strong to come from ordinary stars.



GALACTIC CENTER AT TWO MICRONS was mapped in 1968 by the authors using the 24-inch telescope on Mount Wilson. The map covers the area within the rectangle in the photograph on the opposite page. The center of most intense two-micron emission is just to the left of the galactic equator. About a minute farther to the left is an intense point source of infrared radiation, marked by a cross. The map has a much higher resolution than the surveys of the same area at 100 microns and 3.8 centimeters, depicted on the next page.



GALACTIC CENTER AT 100 MICRONS was mapped by William F. Hoffmann and his collaborators at the NASA Goddard Institute for Space Studies with a small balloon-borne telescope. The resolution is necessarily much coarser than the authors were able to obtain at two microns with their 24-inch ground-based telescope. The rectangle outlines the area depicted in the two-micron map in the illustration at the bottom of the preceding page.



GALACTIC CENTER AT 3.8 CENTIMETERS, in the radio part of the spectrum, was mapped by D. Downes and A. Maxwell of Harvard University and M. L. Meeks of the Massachusetts Institute of Technology using a 120-foot dish antenna. Resolution is about half as good as in two-micron map of galactic center and much higher than in the 100-micron map.

sary that the telescopes be carried by high-flying aircraft, balloons or rockets. Observations between 50 and 300 microns have been made in this way.

Although the largest amount of information has come from surveys with balloons, the poor pointing capabilities of balloon-borne telescopes place a limit on the angular resolution attainable. More detailed observations have been carried out from airplanes, even though a flight may last only a few hours, because an observer can go along to guide the telescope. Rockets, of course, provide the shortest observation times of all: only a minute or two. In compensation rockets reach such high altitudes that the thermal background noise can be reduced to a very low level, making it possible to scan large areas of the sky much more rapidly than can be done from either a balloon or an airplane. So far the largest telescope carried aloft for observation near 100 microns has had an aperture of only 12 inches; as a result the angular resolution achieved is measured in minutes of arc rather than in the seconds of arc attainable at shorter wavelengths with large ground-based telescopes.

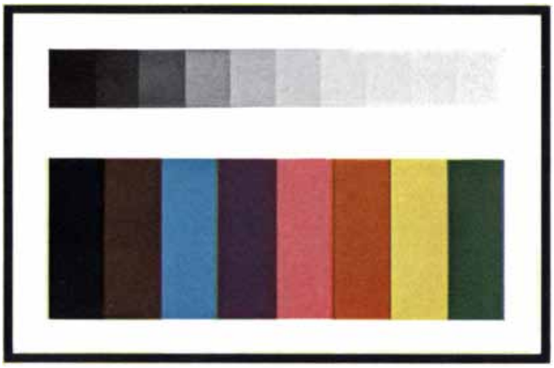
The most general survey at 100 microns was carried out from balloons under the direction of William F. Hoffmann of the NASA Goddard Institute for Space Studies. He and his colleagues surveyed 750 square degrees, roughly 2 percent of the entire sky, mostly along the plane of the galaxy. They found a total of 72 sources, of which at least 12 are not identifiable with objects observable at either visible or radio wavelengths. Two-thirds of the objects, including the brightest, are tentatively identified with known H II regions. By far the brightest object at 100 microns is the complex of sources associated with the nucleus of the galaxy.

At 100 microns the map of the galactic nucleus qualitatively resembles the maps made at two microns. This suggests that the 100-micron flux is thermal radiation from dust heated by starlight. That is also suggested by the close resemblance between the 100-micron map and maps made at radio wavelengths, since the radio flux is thought to come mainly from large H II regions produced by luminous stars.

In summary, we think we understand the brightest sources at two microns. We have much more to learn about the sources at 20 microns. We are just beginning to explore the sky at 100 microns. Nonetheless, for an infant discipline infrared astronomy has already proved to be highly informative.

We want to be useful...and even interesting

Film to please and film to differentiate



A swatch of inks as rendered by Kodak color slide films that maximize consumer satisfaction with the greenness of grass, the blueness of sky, and the healthy glow of complexions.

The same swatch photographed on the new KODAK Photomicrography Color Film 2483,* our first designed not for mere pleasure but for accentuating the ability of biological stains to differentiate structures visually indistinguishable without artifice. Note on the neutral step tablet the fewer greys between black and white. In photography for pleasure this high gradient would make most pictures harsh; in photography to inform and teach, this high contrast, along with the 100-line/mm resolving power for 1.6:1 test-object contrast—remarkable resolution for a color film—brings out detail more conspicuously than do our other films.

No unusual processing is required. KODAK *Process E-4 is easily run in any darkroom and routine with commercial labs. The film comes in 36-exposure 135 magazines, 35mm x 125-foot rolls, and 4 x 5-inch sheets in packages of 25. Many dealers in professional photographic supplies stock it. An inquiry to Kodak, Dept. 916, Rochester, N.Y. 14650 will bring help if needed in locating one.*

*If the package happens to be labeled "SO-456," please forgive us. It's the same product.

Properly exposed, but the flower color is wrong.

Badly underexposed, but the flower color is approximately right. (What Dr. Rickett considers approximate would strike others as rather close.)

◀ Right and wrong

H. W. Rickett of the New York Botanical Garden, author of the monumental and magnificent multi-volume *Wild Flowers of the United States,* made these two exposures of *Eupatorium coelestinum* on Kodachrome-X film.

What is right and what is wrong in color reproduction is intimately bound up not only with what goes on in the individual human head but with the physical conditions of viewing the original and the reproduction. Between a projected image in a darkened room and a piece of paper or plastic in the hand or on the wall the objective differences seem almost too much for subjective comparison. But not really. Kodak scientists have spent whole careers in putting subjective color phenomena on an objective basis. Their work, though published, is little known to many who might find it useful. To follow it calls for a bit of mental concentration by a newcomer to this field, but the human intellect is known to be capable of leaping higher hurdles.

With this warning, we would be pleased to send a reprint of a paper entitled "Objectives in Color Reproduction," by R. W. G. Hunt of the Research Laboratories of Kodak Limited in Harrow, England. Address request to Dept. 55W, Kodak, Rochester, N.Y. 14650.

Kodak

Colorful

Three basic reasons why the Pinto Wagon is the best-selling economy wagon in America.

Shown here is the Pinto Wagon with Squire Option, optional whitewall tires, luggage rack, deluxe bumper group, automatic transmission.

In less than one year, the Pinto Wagon became far and away America's favorite small wagon. Ahead of Vega Kammback and VW Squareback; ahead of all Toyota, Datsun and Opel wagons combined.

The reasons are simple, and basic.

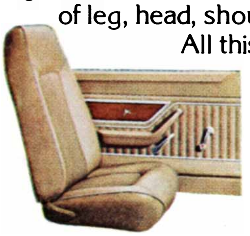
YOU GET A LOT FOR YOUR MONEY.

First and foremost, the Pinto Wagon is an economy car: low in price, easy on gas, low on upkeep.

But it's also a roomy little wagon. Put the rear seat down and you get over 60 cubic feet of cargo space. (VW and Vega give you about 50.)

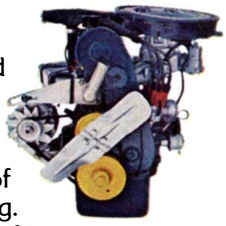
The lift gate swings up, the spare tire is stored under the cargo floor to maximize carrying space.

You get comfort and luxury, too: highback front bucket seats, plush carpeting, and plenty of leg, head, shoulder and hip room.

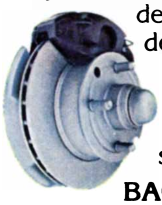


All this at a base sticker price of $2,319 for the basic Pinto Wagon; $2,556 for the handsome Squire Option version. (Prices exclude dealer preparation, destination charges, if any, title and taxes.)

IT'S ENGINEERED FOR DEPENDABILITY.



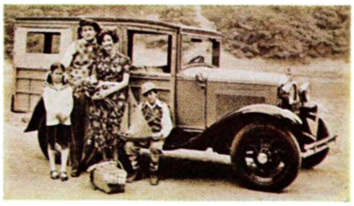
A rugged 2,000 cc cam-in-head engine (right) is standard on Pinto Wagon. It delivers all the power and pep you need to pull the wagon, fully loaded, almost anywhere. Its dependability has been developed in millions of miles of actual driving.



Front disc brakes (left) are standard, too—for dependable stopping and minimum fading.

BACK TO BASICS: DURABILITY.

Virtually every part of the Pinto Wagon was designed with durability in mind—from the welded unitized steel body to the extra strength ball and universal joints and the full synchromesh transmission.



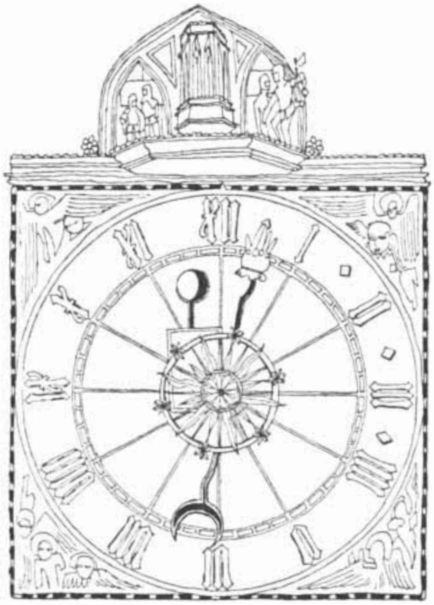
Economy, dependability and durability: those are the basics Ford made famous in the days of the Model T and Model A. See your Ford Dealer.

When you get back to basics, you get back to Ford.

FORD PINTO

FORD DIVISION Ford

SCIENCE AND THE CITIZEN



Homo Sapiens

In the middle of 1971, the latest date for which figures are available, the earth was inhabited by 3.7 billion people. They were having children at the rate of 34 per 1,000 per year and dying at the rate of 14 per 1,000. Thus they were increasing their total number by 2 percent per year, which would double the world population by the year 2006 if the rates were maintained (which would be an unlikely occurrence). Population is increasing most rapidly in Latin America (2.9 percent per year), Africa (2.6 percent) and Asia (2.3 percent), and least rapidly in North America (1.2 percent), the U.S.S.R. (1 percent) and Europe (.8 percent). The figures are from the United Nations *Demographic Yearbook* for 1971, which was published in February.

Birth rates are decreasing in much of the world, according to the UN, particularly in Europe. The lowest rates in 1971 were in West Germany (12.8 per 1,000), Luxembourg (13) and Finland (13.1). In 1971 the U.S. rate was 17.3, but last year it dropped to 15.6. Twelve countries in Africa and Asia still had birth rates above 50 per 1,000; the highest rate, 52.3, is shown for Swaziland.

Official figures from China (the first for the *Yearbook*) show that Shanghai is the world's most populous city, with 10,820,000 people. It is thus considerably bigger than Tokyo (8,841,000) and New York (7,895,000). Peking, London and Moscow are the only other cities with more than seven million inhabitants. When it comes to urban agglomerations rather than central cities, the largest are New York (11,572,000), Buenos Aires and Paris.

Urbanization continues to advance throughout the world. Definitions of urban places vary and so it is somewhat misleading to compare different countries, but on the basis of national definitions Belgium is the country with the highest degree of urbanization: 86.8 percent. Australia (which has a broad definition encompassing built-up areas) reports that 85.5 percent of its people live in urban places. Other countries that are more than 80 percent urbanized are Sweden, Israel and Uruguay. The U.S. figure is 73.5. The highest degree of urbanization in Africa is 47.9. Urbanization does not, of course, equal crowding. Australia seems almost empty (two people per square kilometer), whereas (apart from a number of small islands and city-states such as Macau and Monaco) Belgium is one of the most crowded countries (319 inhabitants per square kilometer), along with England and Wales, the Netherlands and South Korea.

The highest life expectancy at birth is found in Sweden and the Netherlands: female infants born in either country can expect to live for 76.5 years. Among males the longest-lived are Swedish (71.9 years), and male life expectancy is more than 70 in only four other countries: Norway, Iceland, the Netherlands and Denmark. Female life expectancy, on the other hand, is more than 70 in 43 of the 108 countries for which data are available. There are seven countries in which men have a longer life expectancy than women. They are all African or Asian nations with high birth rates and high maternal mortality.

Infant mortality, considered a good indicator of a population's health, is decreasing generally. The Netherlands and Sweden had the lowest rate in 1971: 11.1 per 1,000 live births. Finland (11.8), Japan, Iceland, Norway, Denmark and France (14.4) followed in that order; the U.S. rate was 19.2.

The marriage rate appears to be declining in West Germany, Austria and Sweden; it is going up in Cuba, France, Italy, Ireland, Egypt, Israel and the U.S.S.R. As for divorce, the U.S., with 3.7 divorces per 1,000 of population in 1971, is the undisputed leader, yielding only to its "divorce mill," the Virgin Islands (4.6). Puerto Rico, Cuba and the Spanish Sahara come next, followed by the U.S.S.R. (2.6) and Hungary.

Decisions, Decisions

A computer can "crash," or fail, for any number of reasons. Usually the cause can be traced either to the failure of an electronic component or to a flaw in the program. When the machine crashes for no discernible reason, computer engineers often describe the failure as a "glitch."

A number of computer workers, most of whom are outside the computer industry, have begun to suspect that a significant fraction of glitches are caused by flip-flop circuits that have the task of synchronizing the receipt of information from peripheral equipment (such as tape drives) or from satellite computers that are not running on the same clock as the central processor. To ensure synchronization during data transfer, computer designers try to provide an ample time cushion. Thus a flip-flop might be allotted several microseconds to receive an incoming message and "make up its mind" whether or not it has received one; meanwhile the computer's internal clock may have ticked several hundred times. At some tick of the clock, however, the flip-flop is interrogated and asked if it has received the message. It is at that moment, according to the synchronizer-glitch hypothesis, that the flip-flop may still be in a metastable state, undecided about whether or not it received an input pulse at the precise moment that the time allotted for accepting it ran out.

It is difficult to evaluate the seriousness of the problem, because if a flip-flop gets hung up in a metastable state and causes the computer to crash, it leaves no record of the fact. "Ninety-nine out of 100 engineers would probably deny the existence of the problem," says Alan Kotok of the Digital Equipment Corporation. "They believe that conservative design is all you need; you simply allow enough time for a flip-flop to reach a decision. But as computers are designed to run faster and faster, the safety factor gets squeezed out. I know of a recent experimental machine in which synchronizer flip-flops were allotted only 30 nanoseconds to do their thing. The machine often died in an hour. In other words, maybe once in every few hundred million decisions a flip-flop went into a metastable, or undecided, state that lasted more than 30 nanoseconds. To me the problem looks unresolvable except on a probabilistic basis."

"I think it's an incredibly deep problem," says Anatol Holt, an information theorist at Massachusetts Computer Associates. "It's Zeno's paradox in 20th-century dress or a continuation of the philosophical battle over the discrete and the continuous. A computer ought to be able to decide whether it has received a message or not, but in the mushy world we live in it sometimes can't. Metastability is not a freak problem but a basic conceptual one. We are confronted with defining what we really mean by 'choice.' "

"In general," says Charles E. Molnar, director of the Computer Systems Laboratory at Washington University, "it does not appear to be physically possible for a synchronizer decision to be made with complete assurance that a definite conclusion will be reached within a fixed time interval. The result is an undecided condition that is fatal to a finite-state machine."

Molnar and others believe the synchronizer problem is bound to get worse as more and more individual computers, each with its own clock, are tied together in interactive networks. One engineer working on an advanced system of this kind, Severo Ornstein of Bolt Beranek and Newman Inc., doubts that present synchronization techniques are adequate for the task. "Some people say this is a tempest in a teapot," he says, "but I believe the problem is serious and is still being neglected."

A No to NO

Among the most troublesome air pollutants produced by automobiles are the chemically active nitrogen oxides (NO_x). These substances are manufactured when the high temperatures of the automobile engine cause atmospheric nitrogen and oxygen to combine. Now workers at Bell Laboratories have found catalysts that react nitrogen oxides with a reducing gas (hydrogen or carbon monoxide), converting them to nitrogen and such harmless by-products as water and carbon dioxide.

The catalysts employ oxides containing manganese, rare earths and lead. Prepared in the form of powder, they can be coated on a ceramic support to make a filter-like device that could be installed in an automobile. So far such devices have been tested only in the laboratory; further tests are necessary to see if they will stand up under the severe conditions in an exhaust system of an automobile running for extended periods. According to the Bell Laboratories workers, the catalysts produce much less ammonia (a compound of nitrogen and hydrogen) than more familiar catalysts do, particularly at low temperatures. Any significant output of ammonia would make a catalytic device unsuitable for automobiles.

The automobile industry faces increasingly strict Federal standards for reducing carbon monoxide, unburned hydrocarbons and oxides of nitrogen in exhaust emissions by 1976. Experience so far indicates that effective control of carbon monoxide and hydrocarbons tends to increase the yield of nitrogen oxides, whereas effective control of nitrogen oxides tends to increase the yield of carbon monoxide and hydrocarbons. A catalyst that dealt effectively with the nitrogen oxides without impairing the efficiency of devices that controlled carbon monoxide and hydrocarbons could ease the difficulty of meeting the standards.

Newton Fudged

One of the classic temptations of the scientist is to "fudge" data to fit a theory. To be sure, it sometimes turns out that the theory is right and the original data were wrong. Be that as it may, it seems that even the great Isaac Newton fudged some of his data.

According to Richard S. Westfall, writing in *Science,* Newton was a master at manipulating observations so that they exactly fit his calculations in the *Principia.* Westfall considers three examples in which Newton's theoretical predictions matched independent measurements to a degree of accuracy that was then unheard of. The first example is Newton's calculation that the acceleration of gravity at the earth's surface is equal to the centripetal acceleration of the moon in its orbit; the second is his computation of the velocity of sound; the third is his calculation of the precession of the equinoxes. Newton in effect claimed an accuracy of about one part in 1,000 in the second example, and of one part in 3,000 for the first and third examples. In fact, in the first example he carried his result to seven decimal places.

At the turn of the 18th century such precision was unprecedented. Only some 20 years before the *Principia* was published Robert Boyle had proposed his gas law (Boyle's law) on the basis of measurements that frequently deviated from the law by one part in 100. Boyle believed this lack of precision was unavoidable. Newton insisted that it could be and must be avoided. The accuracy of his determinations supported his concept of gravitational force, or "action at a distance," a concept that seemed mystical to some of his contemporaries and was vigorously attacked by Gottfried Wilhelm von Leibniz.

All three of the examples appear in the first edition of the *Principia.* In the second edition Newton introduced major changes that radically increased the level of apparent precision by mere numerical manipulation of the same basic body of data. Roger Cotes, the editor of the *Principia,* was not only fully aware of what Newton was doing; his correspondence shows that he aided Newton in doing it. At one point in his work Cotes came to Proposition 37, which concerns the tides. When he corrected one calculation on the basis of more recent data, he noted that the new result would change the figures in the corollaries to the proposition, particularly the comparative masses of the moon and the earth. He wrote to Newton: "This alteration will very much disturb Your Scholium of y^e^ 4th Proposition [the correlation of the gravitational acceleration with the moon's centripetal acceleration] as it now stands; neither will it well agree with Proposition 39^th^ [the precession]." Newton replied that he had lost his copy of the amended Proposition 39 and directed: "If you can mend the numbers so as to make y^e^ precession of the Equinox about 50″ or 51″, it is sufficient."

The second edition of the *Principia* was a justification of Newtonian science as well as an amended version of the first edition. Cotes composed a preface for the beginning of it and Newton added a "General Scholium" at the end; both were devoted to the defense of Newtonian philosophy, of exact quantitative science rather than speculative hypotheses. Near the end of their collaboration Cotes wrote to Newton: "I am satisfied that these exactnesses ... are inconsiderable to those who can judge rightly of Your book, but y^e^ generality of Your Readers must be gratified w^th^ such trifles, upon which they commonly lay y^e^ greatest stress." Newton knew perfectly well that his case rested largely on whether or not he could provide better answers than his opponents. After having insisted that a precise correlation between his calculations and the observations was the criterion of truth, he made sure that the precise correlation was presented, whether or not it was properly achieved.

Bologna Hero

How might one find out how quickly organic matter decays in the high-pressure environment of the deep ocean? The question is not only a scientific one; it is raised by the dumping of garbage. Part of the answer has been supplied by an inadvertent experiment. In 1968 workers from the Woods Hole Oceanographic Institution were launching the small research submarine *Alvin* in the Atlantic when the vessel sank because a hatch had been left open. Inside the submarine was a box lunch consisting of a thermos of bouillon, a bologna sandwich with mayonnaise and two apples. The *Alvin* remained on the bottom at a depth of 5,000 feet until it was raised some 10 months later.

Writing in the Woods Hole magazine *Oceanus,* H. W. Jannasch and C. O. Wirsen report on the condition of the box lunch after its high-pressure submersion. Although the foods had contained the usual numbers of bacteria, they were practically untouched by decay. The most perishable item, the bouillon, was still potable. The bread, bologna and mayonnaise of the sandwich, although soaked with seawater, seemed unspoiled. The bologna remained pink on the inside, indicating that the nitrate added to preserve its color had also escaped degradation. The apples had the appearance of having been pickled, but their juice was equal in acidity to fresh apple juice.

Intrigued by these examples of retarded decay, Jannasch and Wirsen arranged to anchor samples of a number of pure organic compounds, inoculated with known kinds and numbers of bacteria, at even greater depth and pressure. Duplicate samples, kept in the laboratory at just above freezing temperature, served as controls. After submersion for two to five months the samples were retrieved. Their rate of decay proved to be from 100 to 600 times slower than that of the unpressurized controls. Evidently even bacteria that thrive at low temperatures find a high-pressure environment inhospitable.

The practical lesson, the two investigators note, is obvious. Since natural recycling processes are nearly at a standstill in deep water, these areas are not appropriate dumping grounds for organic wastes.

Electricide

Thomas Edison and George Westinghouse sternly disagreed about how electricity should be distributed. Edison's company pressed for direct-current distribution, Westinghouse's for alternating current. In 1890 the two systems had a curious confrontation: the first use of the electric chair. The episode is recounted in the February issue of *IEEE Spectrum* by Theodore Bernstein of the University of Wisconsin.

Westinghouse espoused the alternating-current system because it would allow the use of transformers, and therefore higher voltages and lighter conductors. Edison argued that, in the event of electric shock, alternating current would be more dangerous than direct current of the same voltage. Thus when the electric chair was proposed as a humane substitute for hanging, Edison suggested that alternating current would be best for the purpose. Westinghouse strongly objected. Indeed, he tried to prevent the State of New York from getting one of his alternating-current generators for the first electric chair.

The chair was nonetheless built in 1889 with a generator that had been obtained from a second-hand dealer. The first man to be put to death by electrocution was William Kemmler, who had murdered his girl friend. The electrocution was botched. The current was first applied for 17 seconds, after which Kemmler was observed to be twitching slightly. It was then turned on for an additional 70 seconds, at the end of which time one electrode began to smoke. Kemmler was now pronounced dead.

Westinghouse said: "It has been a brutal affair. They could have done better with an axe." Edison, it seems, had devoted considerable thought to what the new form of execution should be called. According to Bernstein, he suggested "ampermort," "dynamort" and "electromort." One of his associates sought the opinion of Eugene H. Lewis, a New York attorney. Lewis favored "electricide," but he also pointed out a parallel in the naming of the guillotine after the French physician Joseph Ignace Guillotin. A good term for being executed by electricity, he proposed, would be "westinghoused."

The Control of the Water Cycle

The recent emphasis in hydrologic studies on the crucial role played by the general circulation of the atmosphere has led to schemes for altering this vast natural process

by José P. Peixoto and M. Ali Kettani

"We made from water every living thing."

—The Koran (Prophets 30)

"All the rivers run into the sea; yet the sea is not full."

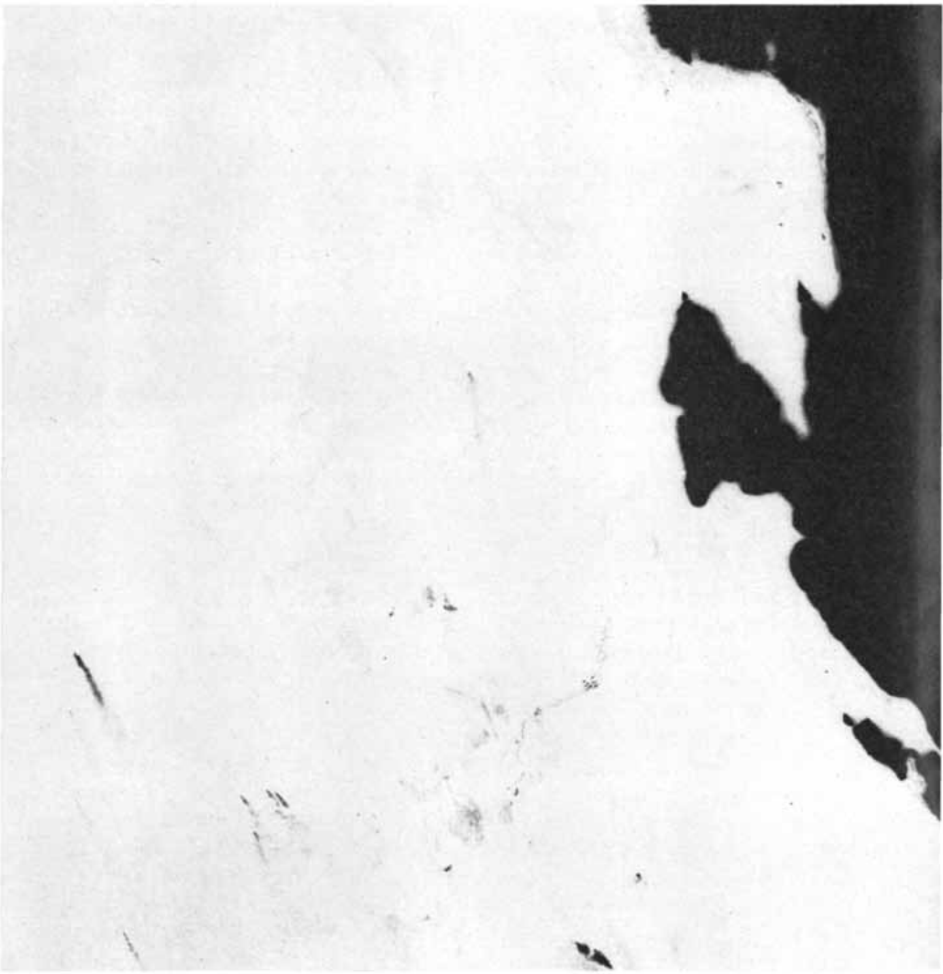
—Ecclesiastes 1:7

The importance of water to the creation and sustenance of every living thing, including prosperous human communities, was profoundly appreciated by the civilization-builders of the ancient Middle East. So was the vital role of the hydrologic cycle, the wondrously fickle process by which the terrestrial supply of water is constantly replenished, thus ensuring the continued existence in a given place of man and his creations.

Indeed, the loss of an adequate water supply can be implicated in the demise of many human settlements in the ancient world, most notably in the disappearance of the original high civilizations in the formerly "Fertile Crescent" of the Tigris and the Euphrates river valleys. In more recent times the fall of the Roman Empire and the decline of the Arabian caliphates have been blamed on a more or less sudden decline in the amount of water available to their populations. In the modern world vast, potentially prosperous regions of Africa, Asia, Australia and South America—regions stretching over millions of square kilometers in the subtropical latitudes—are sparsely populated today because of a lack of water. Even in prosperous countries such as the U.S. the increasing danger of an imbalance between water production and water consumption is acutely felt.

Yet in spite of the obvious importance of water to the life and prosperity of humankind, societies have gone on for centuries using this precious commodity in a basically inefficient and wasteful way. It is only recently that some tentative efforts have been made to consider the possibilities of controlling the water cycle on a large scale to better meet human needs. In order to understand why this long overdue development has finally begun, it is necessary to review some recent trends in the study of hydrology.

During the past few decades the water balance at the earth's surface has been the subject of intensive research. Until the mid-1930's most of these in-

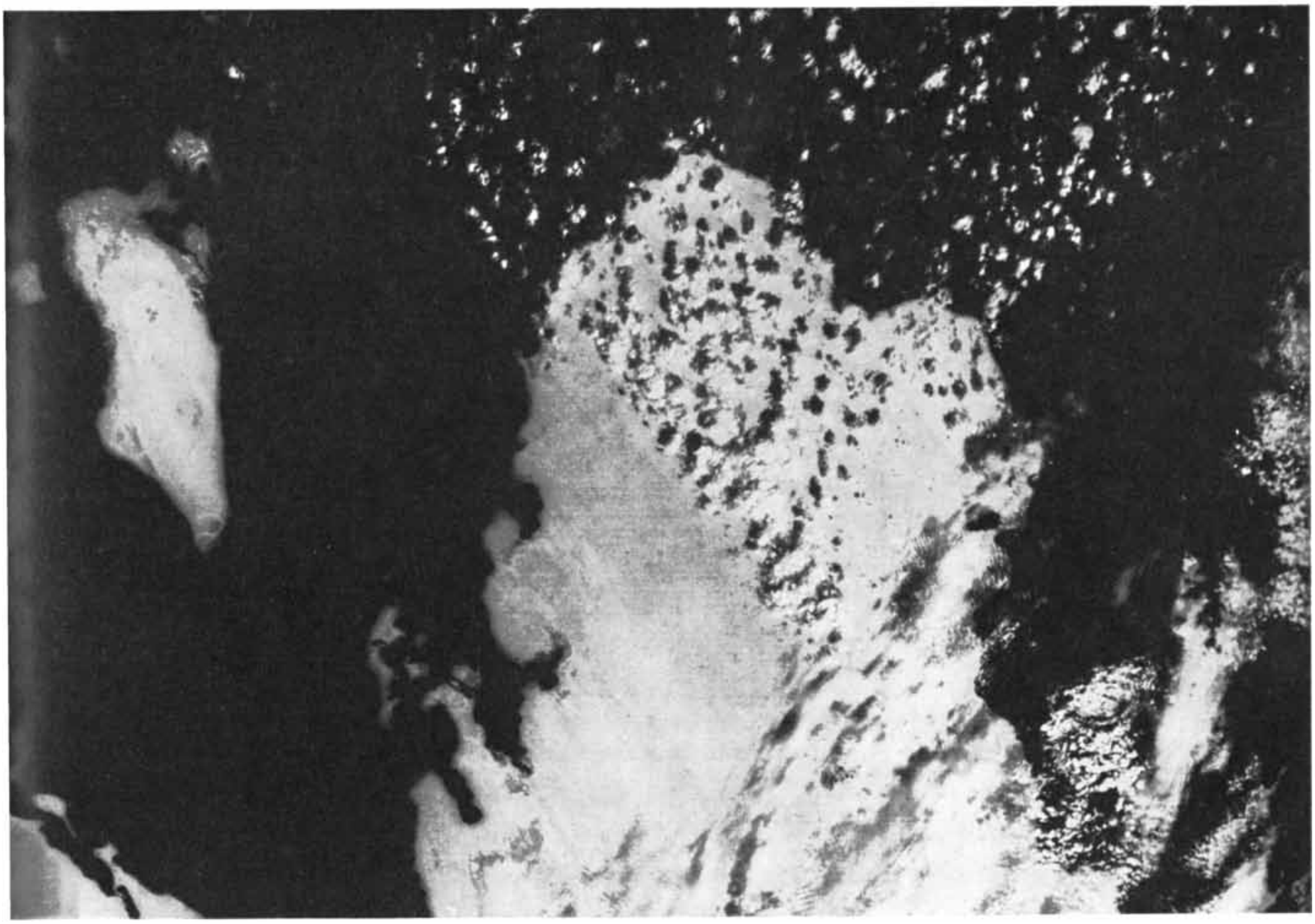


SITE OF PROPOSED PROJECT for controlling the water cycle over a section of the Persian Gulf is shown in this photograph, made from an altitude of almost 600 miles by a camera on board the National Aeronautics and Space Administration's Earth Resources Technology Satellite (ERTS). The proposed "heliohydroelectric" scheme, put forward in vestigations dealt only with the terrestrial branch of the water cycle, examining the balance of precipitation, evaporation, transpiration, runoff, underground flow and changes in ground-water storage. This conventional approach to the water budget often yielded results that were not widely accepted by other hydrologists; at best the results were representative of the water budget only in limited regions or watershed systems. Actually the problem was that evaporation from the surface of the earth is not uniform; the water released is continuously removed by turbulent diffusion near the surface, by convection currents due to temperature differences in the air and by the wind. All these factors condition the phenomenon of evaporation. Moreover, there is no single instrument that can measure natural evaporation directly. Measuring precipitation by direct means in many parts of the world is possible, but precipitation tends to be spotty, and simultaneous widespread sampling is at times inadequate.

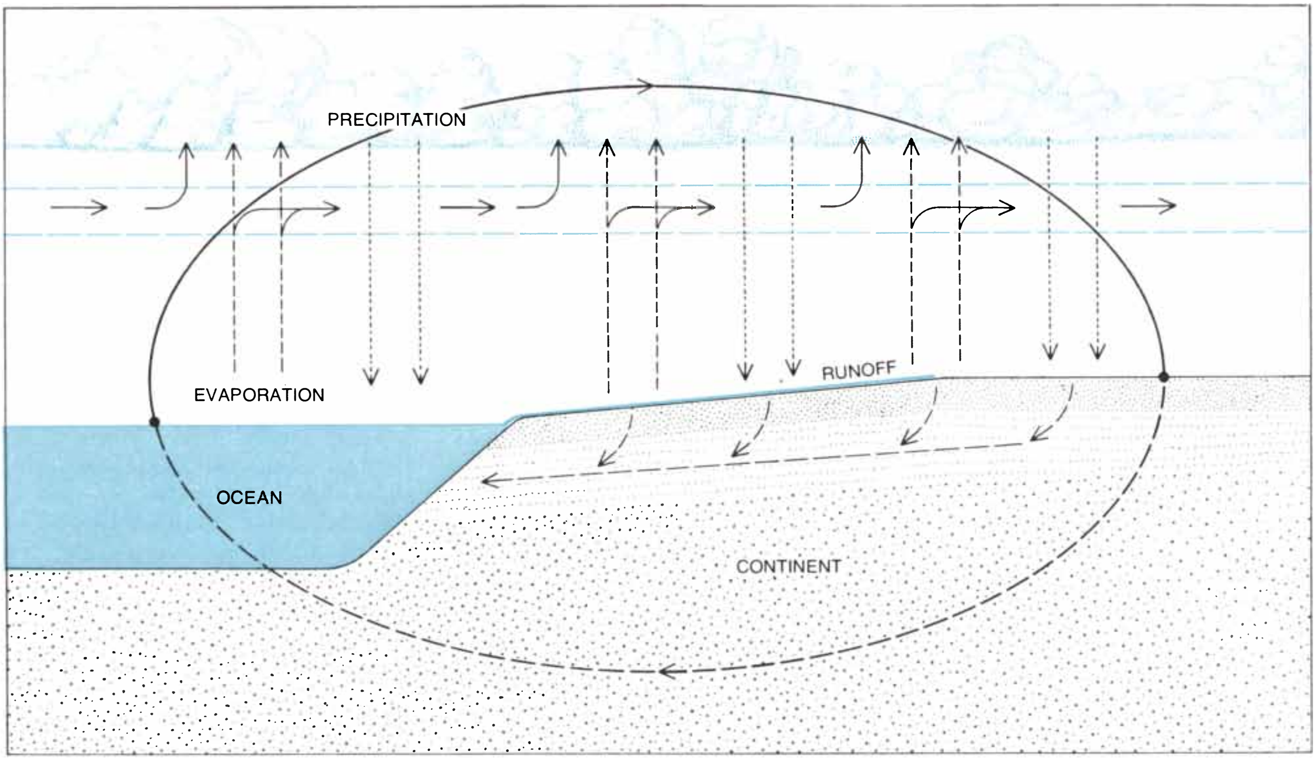
In view of these difficulties some recent studies have approached the problem of the water balance by considering the atmospheric branch of the hydrologic cycle instead of the terrestrial branch. This approach is based on the principles of continuity and conservation of mass in the atmosphere rather than at the earth's surface. It has become feasible in recent years as a result of the rapid expansion of the upper-air sensor network and the resulting improvement in the quality of aerological observations, which are now adequate to deal quantitatively with the large-scale processes of the atmospheric branch of the water cycle. This information in turn has become available for evaluating the water balance at the earth's surface. For example, the Planetary Circulations Project at the Massachusetts Institute of Technology, with which we have both been associated, is engaged in studying the general circulation of the atmosphere. This work has given impetus to a large number of investigations of the relations between the flux of water vapor in the atmosphere and the hydrologic cycle. It is out of investigations such as these that specific schemes for the control of the water cycle are beginning to emerge.

Why Water Is Where It Is

In primeval days our planet was devoid of any atmosphere, since elements such as helium and hydrogen were too light to be trapped by the earth's gravitational field. The earth had no oceans, and its entire surface was covered with active volcanoes from which poured lava, gases (mainly hydrogen and hydrogen-rich compounds) and vapors (mostly water vapor). Water molecules in the volcanic emissions were promptly photodissociated, or split by sunlight, into their constituent atoms: hydrogen and oxygen. The hydrogen escaped and the oxygen reacted with ammonia (NH_3) and methane (CH_4) to form nitrogen (N) and carbon dioxide (CO_2). Thus the earth's original atmosphere is believed



1970 by one of the authors (Kettani) in collaboration with L. M. Gonsalves, would utilize solar energy to generate through evaporation an inflow of water from the open sea into an artificial reservoir called Dawhat Salwah (*see illustrations on pages 58 and 59*). The reservoir would be formed by building causeways between Saudi Arabia (*landmass at left*) and Bahrain (*island at center*) and between Bahrain and Qatar (*peninsula at right*). The inflow of water through a conduit could be used to generate electricity.



WATER CYCLE OF THE EARTH, depicted in this highly schematic diagram, can be regarded as consisting of two parts: an atmospheric branch (*solid arc*) and a terrestrial branch (*broken arc*). The earth yields water to the atmosphere mainly through evaporation from the oceans. This water is transported by the general circulation of the atmosphere in both the gaseous phase (water vapor) and the condensed phases (clouds). Eventually the water precipitates back to the earth, either as a solid or as a liquid. If it falls on land, it can infiltrate underground, run off in rivers and streams into the ocean or evaporate directly back to the atmosphere.

to have been a mixture of the last two gases. Its composition later changed to its present state when plants appeared and started absorbing the carbon dioxide and liberating oxygen. After the first process of atmospheric formation had been stabilized, the excess water began accumulating in depressions, forming the oceans.

The amount of water on the earth can be considered constant during the evolution of man. True, some new water is being produced by the remaining few volcanoes and hot springs. Most steam ejected by volcanoes, however, is actually either rainwater that had saturated the upper layers of rock or seawater that was trapped at the time marine sediments were deposited. A certain amount of water may be destroyed in the upper atmosphere by the photodissociation of water vapor by solar radiation. Both effects are negligible on the geological time scale.

At present the water existing on the earth is distributed in three separate reservoirs. In order of importance they are the oceans, the continents and the atmosphere. The interior of the earth contains an appreciable amount of water, either dissolved or chemically combined with solid or molten rock, but there is no satisfactory estimate of the amount of water thus locked up.

About 97.3 percent of all the water in the hydrosphere is in the oceans; the remaining 2.7 percent is on the continents, mostly in the glaciers of the Arctic and the Antarctic. The atmosphere contains a mere hundred-thousandth of the total water present on the earth. In terms of volume the total water in the oceans amounts to $1,350 \times 10^{15}$ cubic meters; the entire atmosphere normally contains a constant quantity of water on the order of $.013 \times 10^{15}$ cubic meters.

The water on the continents is in turn distributed in several reservoirs, namely glaciers (29×10^{15} cubic meters), ground water (8.4×10^{15} cubic meters), lakes and rivers ($.2 \times 10^{15}$ cubic meters) and the living matter of the biosphere ($.0006 \times 10^{15}$ cubic meters). The amount of water locked in the polar ice is impressively large, totaling some 1.8 percent of all the water in the hydrosphere. Of the total amount of underground water, vadose water (water present in soils) accounts for only $.066 \times 10^{15}$ cubic meters. The remainder is about evenly divided between reservoirs deeper than 800 meters and reservoirs shallower than that level.

This distribution of terrestrial water changes with time. For instance, continental ice caps have periodically developed and melted again during the past two million years. If they were to melt altogether, the sea level would rise by as much as 60 meters and submerge a large area of the continents. During periods of maximum glaciation the sea level may have been lowered by as much as 140 meters, uncovering large areas of the continental shelves. These fluctuations are accompanied by a disturbance in the balance of the water cycle.

The Water Cycle

The earth is constantly releasing water vapor into the atmosphere, both through evaporation from the surface (seas, lakes, rivers, glaciers, snowfields, man-made reservoirs, soil and so on) and through the transpiration of living matter. This natural transfer of water from the earth's surface to the atmosphere is extremely important, since it conditions the characteristics of the air masses, the energetics of the atmosphere, and the establishment and evolution of the hydrologic cycle.

The water cycle is the fundamental concept of hydrology. It is a consequence of the principle of conservation of water in its three phases on the earth. It describes a closed sequence of natural phenomena by which water enters the atmosphere from the surface in the gaseous phase and returns to the surface in the liquid or solid phase. There it is partly retained or runs off, to evaporate again into the atmosphere. Energy is continuously exchanged between the atmosphere and the surface. It is this exchange that induces the circulation of water in a continuous cycle.

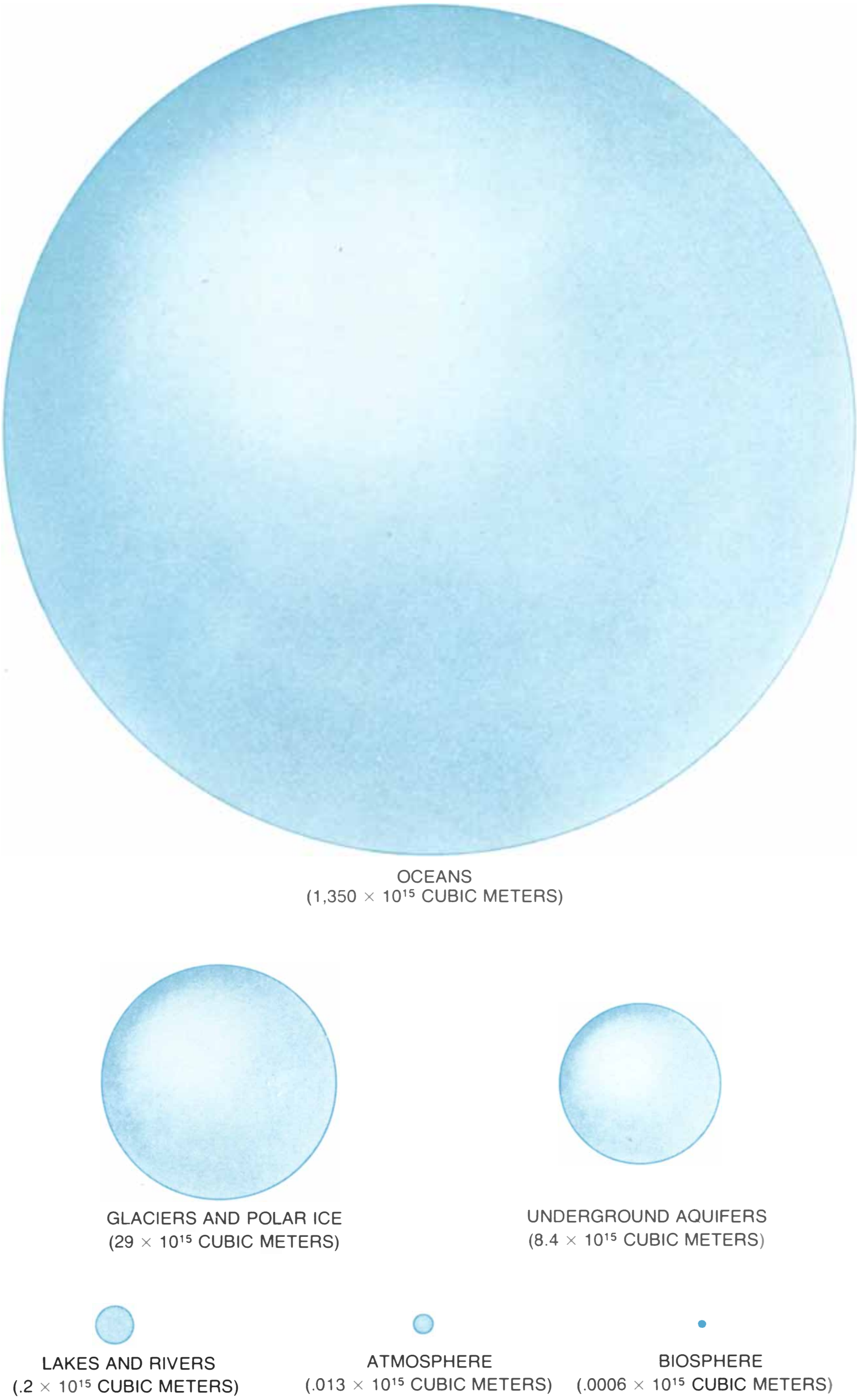
Except for a negligible outflow from the interior of the earth, all the energy that drives the water cycle comes from the sun. The solar radiation reaching the earth's orbit amounts to two calories per square centimeter per minute, a quantity termed the solar constant. The amount of solar energy actually received by the earth in one minute is equal to the product of the solar constant and the cross-sectional area of the earth: 2.55×10^{18} calories. This amount corresponds to an average energy of .5 calorie per minute per square centimeter, or a fourth of the solar constant. In terms of power that energy flux amounts to a total continuous input of about $.178 \times 10^{18}$ watts. As the solar radiation proceeds toward the surface of the earth, its ultraviolet component is strongly absorbed by the ozone in the upper atmosphere, heating that part of the atmosphere. More radiation is absorbed by air molecules, is reflected back into space by clouds or is simply scattered at lower altitudes. Nevertheless, on the average about 81 percent of the incoming radiation manages to reach the earth's surface. Roughly a third of this energy is immediately reflected back into space. The remaining two-thirds is absorbed by the continents and the oceans. Of the absorbed radiation 77 percent is ultimately reradiated as long-wave radiation through the atmosphere back into space. The remaining 23 percent goes into the evaporation of water.

The most important meteorological factor in the hydrologic cycle therefore is sunlight, the supply of energy that causes and maintains the circulation of water in the cycle. Air temperature, humidity and wind mainly condition the evapotranspiration process. Clouds, the visible deposits of water in the atmosphere, in turn condition the solar radiation that reaches the earth.

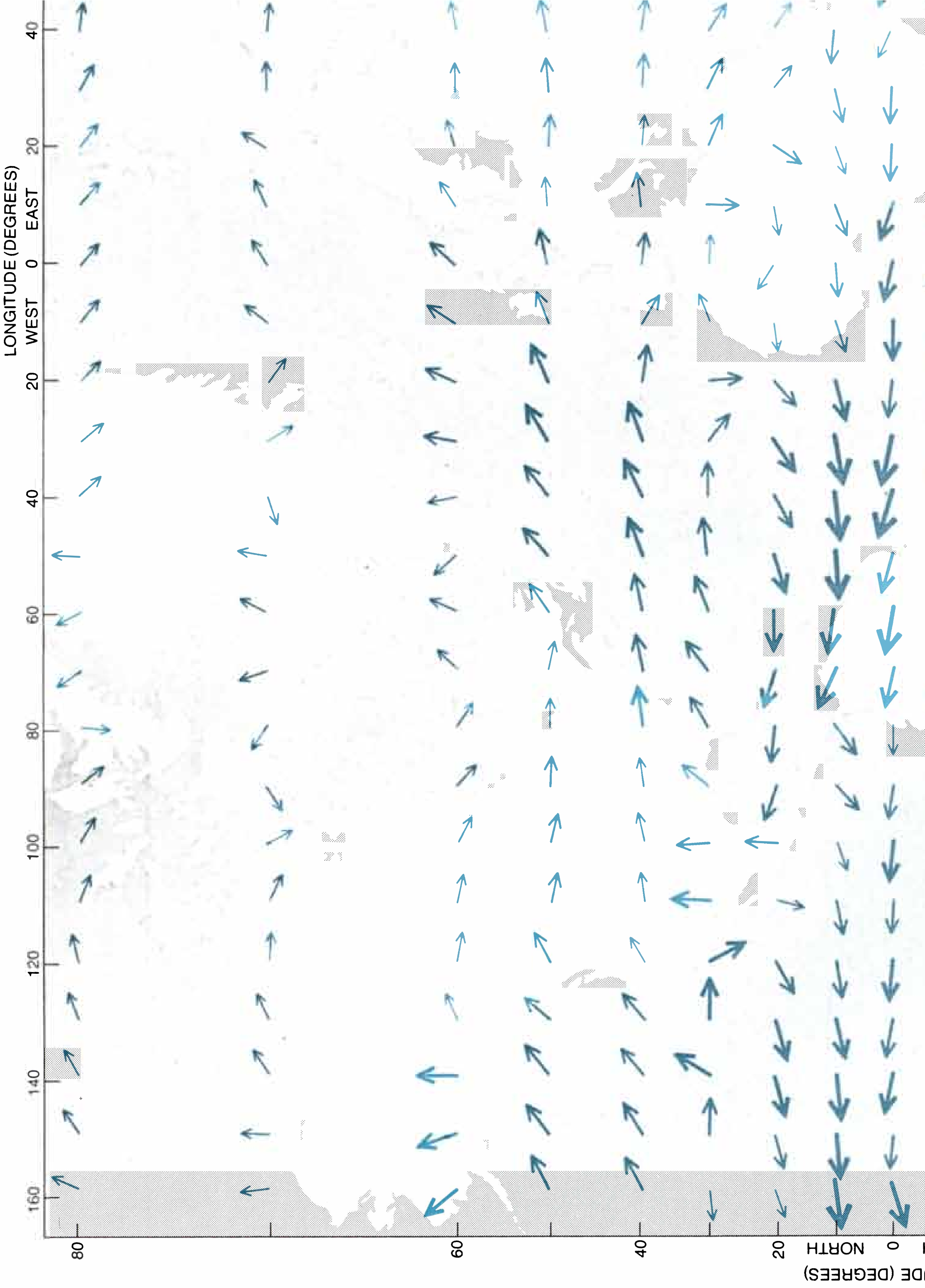
The hydrologic cycle has two distinct branches: the atmospheric branch, in which the horizontal flux of water is mainly in the gaseous phase, and the terrestrial branch, in which the flux of water in the liquid phase and the solid phase predominates. At the surface—the interface that separates the two branches of the cycle—complex processes arise owing to boundary conditions. Even in each branch of the cycle the physical processes are not simple, because neither the atmosphere nor the soil is homogeneous. The influence of regional and local factors in hydrologic phenomena must therefore be taken into account.

The Role of the Atmosphere

The amount of water vapor in the atmosphere varies widely, both in space

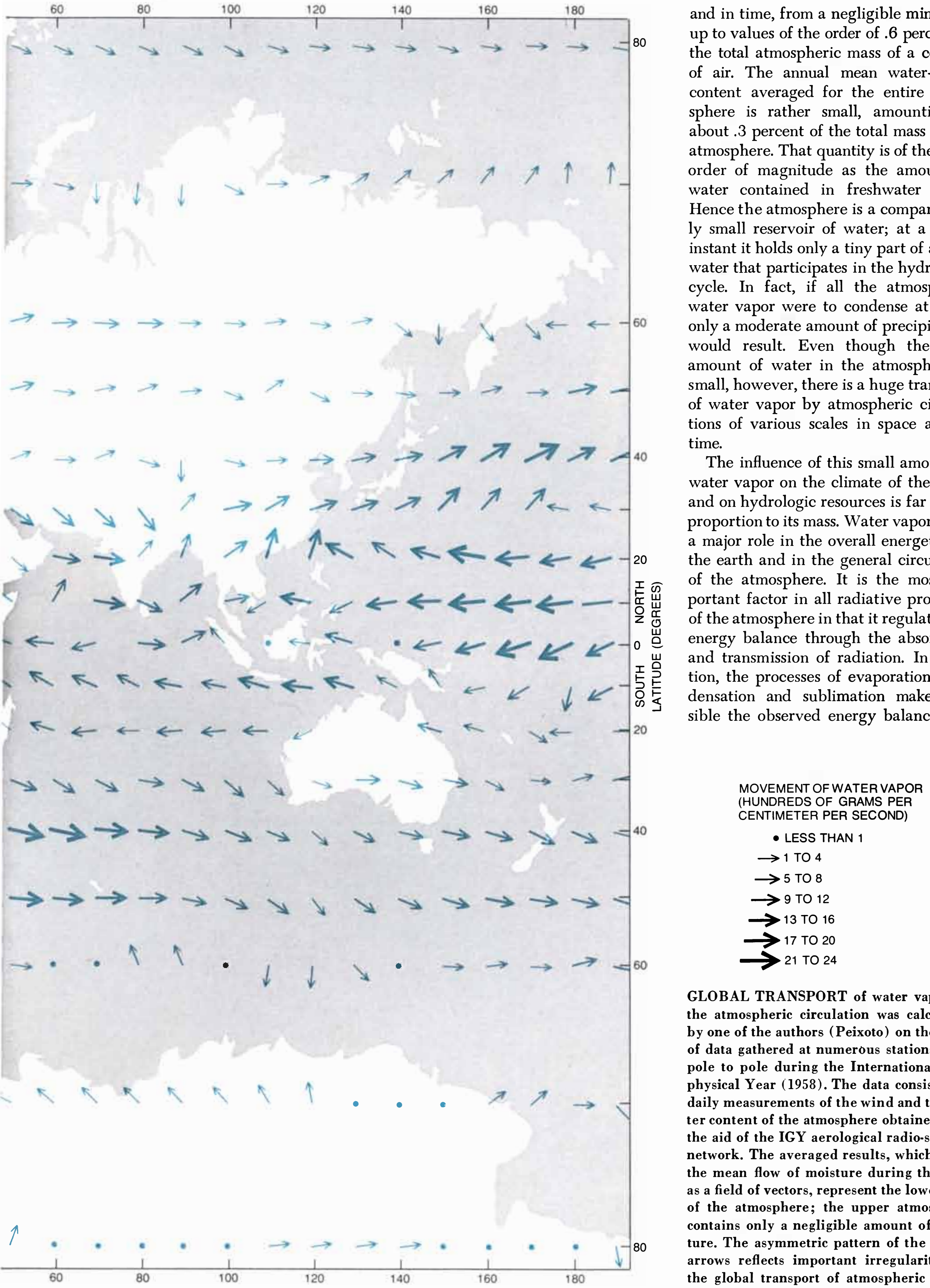


DISTRIBUTION OF WATER on the earth is indicated in this illustration, in which the amount of water present in various natural reservoirs is represented in terms of comparative spherical volumes. The number under the name of each reservoir denotes the contents of that reservoir in cubic meters. Although the atmosphere contains a mere hundred-thousandth of all the water in the hydrosphere, the influence of this small amount on the climate of the earth and on the location of hydrologic resources is far out of proportion to its mass.

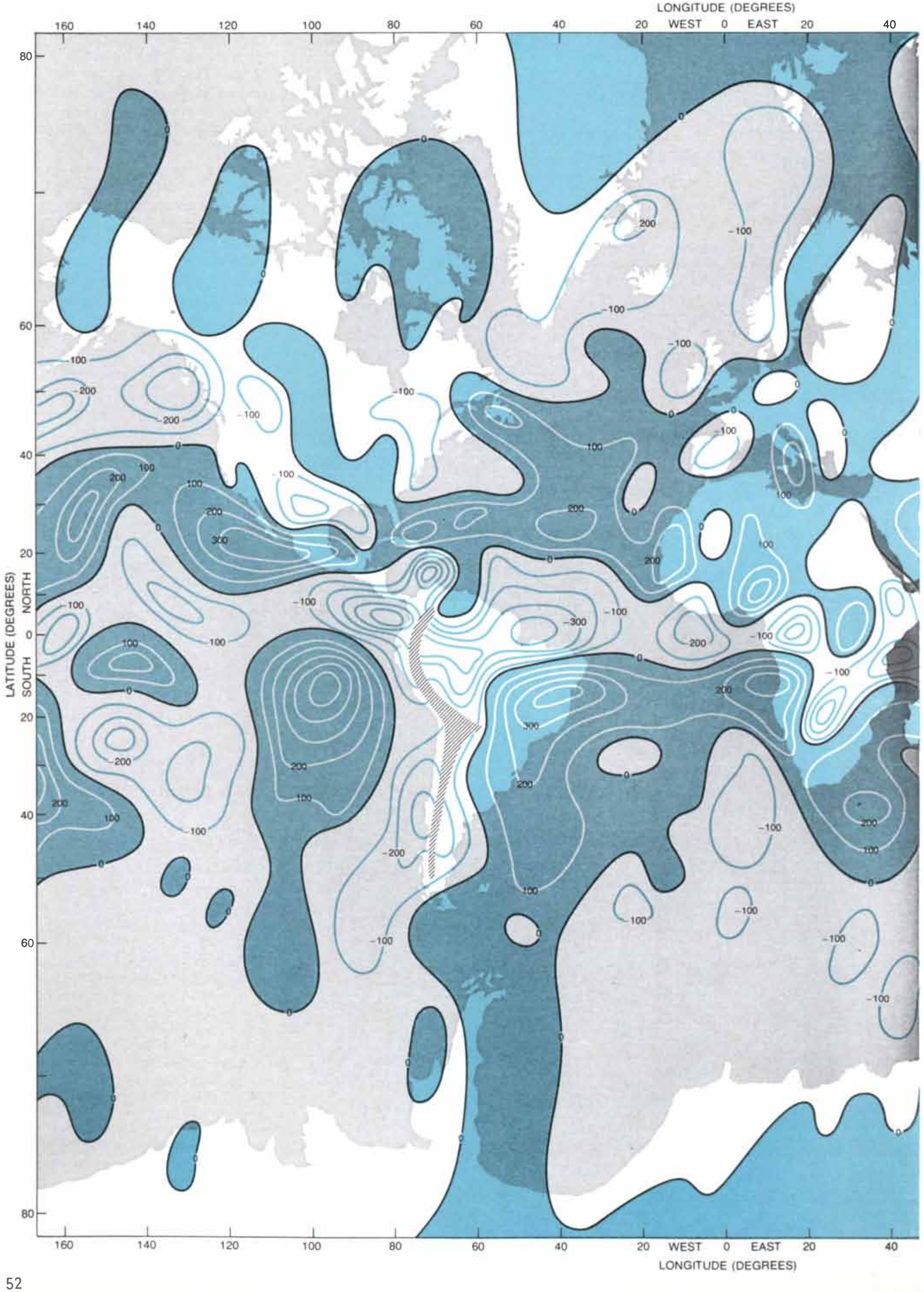

LONGITUDE (DEGREES)
WEST
EAST
NORTH

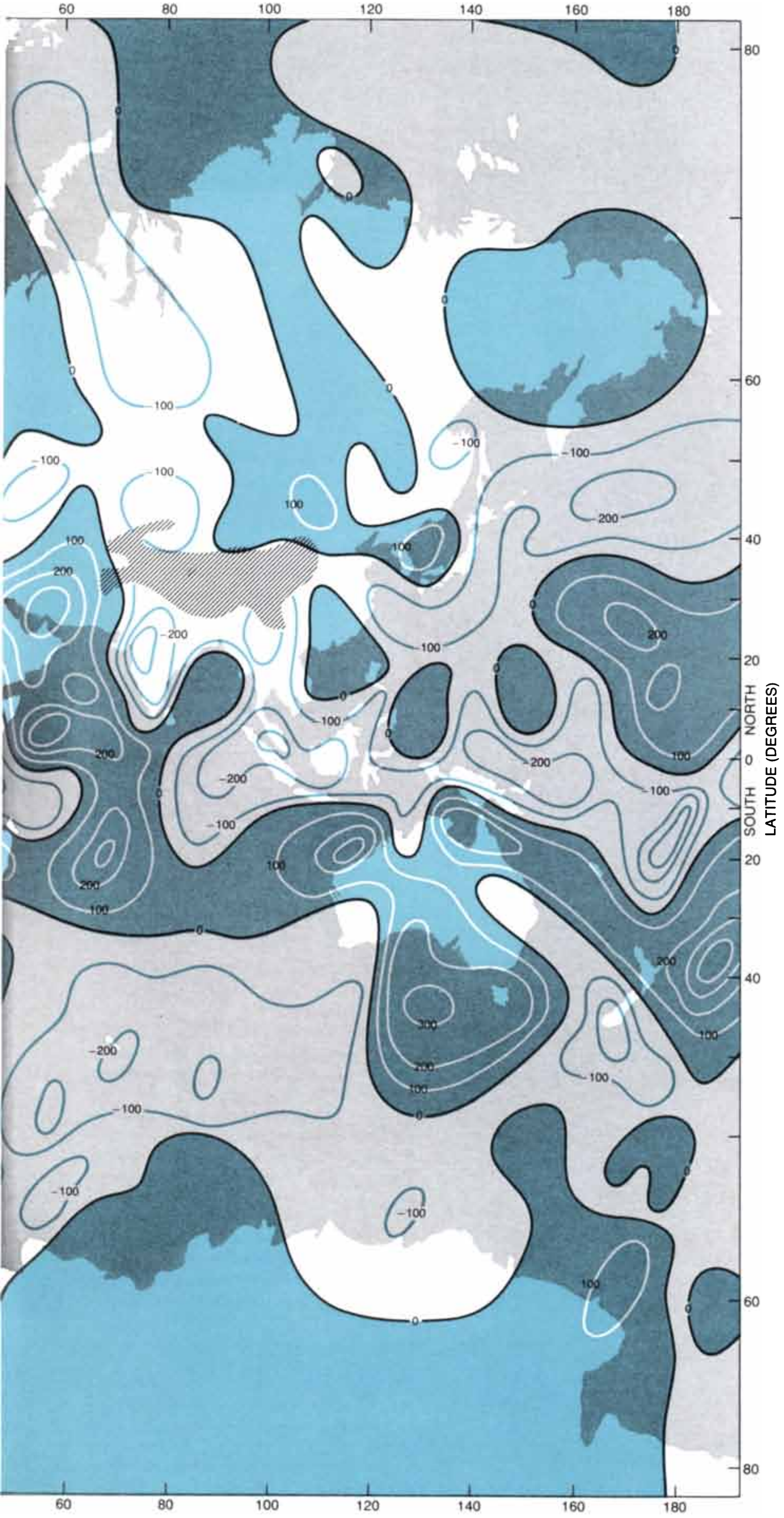
and in time, from a negligible minimum up to values of the order of .6 percent of the total atmospheric mass of a column of air. The annual mean water-vapor content averaged for the entire atmosphere is rather small, amounting to about .3 percent of the total mass of the atmosphere. That quantity is of the same order of magnitude as the amount of water contained in freshwater lakes. Hence the atmosphere is a comparatively small reservoir of water; at a given instant it holds only a tiny part of all the water that participates in the hydrologic cycle. In fact, if all the atmospheric water vapor were to condense at once, only a moderate amount of precipitation would result. Even though the total amount of water in the atmosphere is small, however, there is a huge transport of water vapor by atmospheric circulations of various scales in space and in time.

The influence of this small amount of water vapor on the climate of the earth and on hydrologic resources is far out of proportion to its mass. Water vapor plays a major role in the overall energetics of the earth and in the general circulation of the atmosphere. It is the most important factor in all radiative processes of the atmosphere in that it regulates the energy balance through the absorption and transmission of radiation. In addition, the processes of evaporation, condensation and sublimation make possible the observed energy balance and



GLOBAL TRANSPORT of water vapor by the atmospheric circulation was calculated by one of the authors (Peixoto) on the basis of data gathered at numerous stations from pole to pole during the International Geophysical Year (1958). The data consisted of daily measurements of the wind and the water content of the atmosphere obtained with the aid of the IGY aerological radio-sensing network. The averaged results, which show the mean flow of moisture during the year as a field of vectors, represent the lower half of the atmosphere; the upper atmosphere contains only a negligible amount of moisture. The asymmetric pattern of the vector arrows reflects important irregularities in the global transport of atmospheric water.


LONGITUDE (DEGREES)
160 140 120 100 80 60 40 20 WEST 0 EAST 20 40
LATITUDE (DEGREES)
NORTH
SOUTH
80 60 40 20 0 20 40 60 80

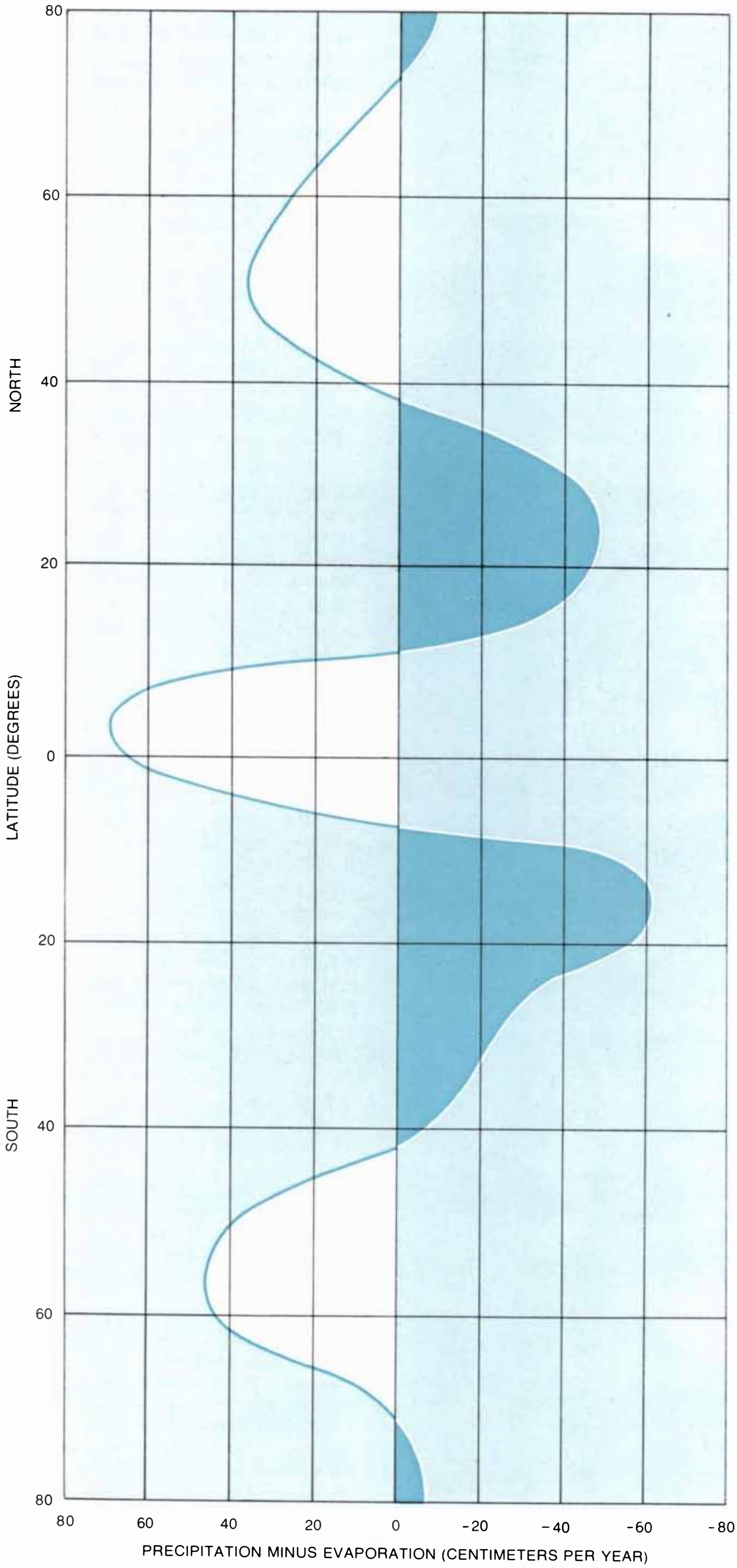


the water balance within the physical system of the earth and the atmosphere.

On the average, whenever there is an imbalance between precipitation and evaporation, there is a need for a net transport of water vapor to or from the locality by atmospheric circulations. Can such water transports in the atmosphere be assessed from the knowledge of the wind and the water content of the atmosphere? If the answer is affirmative, then an independent check on the precipitation-evaporation difference is available. That would be a most welcome development in view of the shortcomings of the existing methods of measuring these quantities separately. It must be understood, however, that for this purpose time averages of the wind and the humidity will not do, because it turns out that the effect of the deviations of these factors from their arithmetic mean cannot be neglected. The process must be evaluated instantaneously, say once each day; then an average of the results must be taken over a long enough interval. That must be done for several levels in the atmosphere up to an elevation of about five or six kilometers. Therefore, with the use of many stations at which soundings are taken, the volume of the calculations soon becomes enormous. Fortunately modern data-processing techniques are powerful enough to cope with the task.

Some years ago one of us (Peixoto) made an attempt to perform this new type of moisture-transport calculation over the entire globe on a daily basis from pole to pole for the International Geophysical Year (1958). The project was part of the comprehensive M.I.T. study of the general circulation of the earth's atmosphere and of its energetics. The results in one form showed the mean flow of moisture during the year as a field of vectors [*see illustration on*

MATHEMATICAL OPERATION was performed on the field of vectors in the illustration on pages 50 and 51 in order to compare the aerological data with conventional climatological estimates. The outcome is a contour map in which regions where there is a divergence in the field of vectors (*white contours on color background*) correspond to net sources of moisture and hence to places where there is on the average an excess of evaporation over precipitation, whereas regions where convergence prevails (*color contours on white background*) correspond to net sinks of water vapor and hence to places where precipitation exceeds evaporation. The numbers on the contours indicate the equivalent depth of liquid water in centimeters per year.



pages 50 and 51]. These indicators are interesting to examine in themselves, but a second operation must be performed on them in order to compare the outcome with the conventional climatological estimates, which are arrived at by evaluating the excess of water that crosses the boundaries into any unit area of the earth's surface. In mathematical terms this operation is known as the convergence of a field of vectors.

When there is a divergence of water vapor in a given region of the atmosphere, there will be a source of moisture. This means that for the same interval there is on the average an excess of evaporation over precipitation. On the other hand, when convergence prevails, there is a sink of water vapor, and precipitation exceeds evaporation. Therefore the numerical value of the divergence equals the rate of evaporation minus the rate of precipitation. The divergence-convergence distribution corresponding to the vector picture is shown on the map on pages 52 and 53. This is the first such global chart ever to be drawn from actual aerological data.

The distribution of divergence and convergence can be averaged around the world for various latitudes [*see illustration at left*]. Inspection of these figures shows that on the average there is (1) a net convergence in the equatorial zone, where heavy precipitation occurs; (2) a net convergence in the middle and high latitudes, where the large excess precipitation over evaporation is associated with the migratory cyclones, the polar front and the alternation of air masses; (3) a net divergence in the subtropical

AVERAGE DISTRIBUTION of divergence zones (*dark color*) and convergence zones (*white*) around the world were computed for various latitudes from the contour map on the preceding two pages. The resulting curve is almost symmetric with respect to the Equator, where there is a comparatively narrow zone of net convergence (excess of precipitation over evaporation). Flanking this intertropical convergence zone are two broader zones of net divergence over the subtropical latitudes, where the large excess of evaporation over precipitation, particularly over the oceans, accounts for most of the water vapor supplied to the atmosphere. The net convergence in the middle and high latitudes is associated with the migratory cyclones, the polar front and the alternation of air masses. There is a small net divergence in the polar regions. The main effect of this alternation of zones on the hydrologic cycle is a net transport of water vapor northward and southward from primary sources of moisture in subtropical belts.

regions in both hemispheres, where there is strong evaporation, associated with the large semipermanent anticyclones. The results agree quite well with those obtained from independent earlier estimates of evaporation minus precipitation by climatological methods.

Thus the primary and most important sources of moisture for the entire atmosphere are found over the subtropical regions, mainly in the oceans, where evaporation proceeds continuously. The moisture supplied to the atmosphere is transported from the sources and carried by the atmospheric circulation into regions of prevailing convergence, where it condenses and falls as precipitation. Therefore the theory of the formation of precipitation from evaporation in a given place cannot be accepted. Judging from the present data, it is probable that the path length of the atmospheric branch of the water cycle varies from one locality to another. Taking as a criterion the average distance between divergence and convergence centers, it appears that the mean path length in temperate latitudes is of the order of 1,000 kilometers.

The Role of the Terrestrial Branch

Regions of convergence must have a means for disposing of the excess of precipitating water. Over land this means generally drainage by rivers, and hence convergence regions should coincide with the catchment basins of large streams. This conclusion agrees with findings on the divergence-convergence map, an outcome that lends support to the validity of the aerological method used.

The equatorial regions of the Atlantic and Pacific oceans show a general convergence. Marked centers of strong convergence are found just south of Panama, off the east coast of South America near the Equator and over Brazil (with high values in the headwaters of the Amazon River, as would be expected). This belt of convergence extends through the Atlantic Ocean with a center over the Gulf of Guinea. In equatorial Africa the regions of convergence are found in the general vicinity of the headwaters and drainage basins of many large rivers. The areas of largest convergence are found in Ethiopia and the Somali highlands, from which the Blue Nile derives most of its water; in northern Zaire over the headwaters of the Ubangi and Congo rivers; in southern Angola, Rhodesia and the eastern part of southwest Africa, where the Zambezi, Orange and Limpopo rivers originate.

In Ghana and the Ivory Coast there is convergence, and much of the excess of water associated with it is probably fed into the Niger, the Volta and perhaps the Senegal River. An extensive area of convergence extends from India through Southeast Asia, Indonesia and the entire Central Pacific to the west coast of Central America, with several intense centers. Major river systems found in this belt include the Indus, the Ganges, the Brahmaputra, the Salween, the Mekong and the Yangtze. All these areas are known to have extremely high values of precipitation.

The mid-latitude regions around the Southern Hemisphere also show many areas of convergence. The convergence areas are related to the polar front that is associated with the tracks of extratropical storms across the South Atlantic, the South Indian Ocean and the South Pacific. The extensive area of convergence that covers most of South America corresponds to the basins and headwaters of a large system of rivers (including the Magdalena, the Orinoco, the Amazon, the Paraguay, the Paraná and the Uruguay). There is a convergence center that extends through the eastern South Pacific and along the Chilean coast, where strong thunderstorm activity predominates. In the neighboring regions the salt content of the ocean is low.

The subtropical regions of the South Atlantic, the Indian Ocean and the South Pacific show strong and extensive areas of divergence (excess of evaporation over precipitation). In the Atlantic region the divergence belt is elongated in an east-west direction without any interruption. The strong divergence off the coast of central Brazil is reflected by a correspondingly higher salt content in the waters of that region of the Atlantic. By and large all the divergence regions over the oceans are associated with high salinity. These areas and the corresponding ones in the Northern Hemisphere are areas where evaporation attains its maximum value. They are associated with the anticyclonic belt of the subtropical latitudes and are generally situated on the margins of the semipermanent high-pressure cells where clear skies predominate. On land, areas of large divergence are found over northeastern Brazil, a region noted for droughts; in Angola, Uganda, Kenya and South Africa and over most of Australia. In the Northern Hemisphere there are strong divergences over southern Mexico and Lower California in North America, over the Sahara in Africa and over Arabia and Iran in Asia. In the polar regions it seems that there is a small net divergence. With the presence of ice there is comparatively little evaporation. The divergence is higher over the North Pole than over the South Pole.

The regions of strongest divergence tend to be found over the subtropical regions, although there are exceptions such as the center over the eastern coast of Asia. When these regions are over ocean surfaces, there is of course no shortage of water to supply the net evaporation; it is always replenished by ocean currents. The situation is otherwise where the regions of strong divergence are over land, as they are, for example, over the Sahara.

Where does the water come from in such places? It is now known that under this seemingly arid region huge amounts of water are buried in large formations of porous rock. Evaporation from these aquifers has been found to amount in certain areas to as much as 200,000 cubic meters per year per square kilometer. Minerals formed by the deposition of salt in the course of evaporation are common in the Sahara, where such evaporites are found in the salt flats called chotts. Deposits of minerals in the form of "desert varnish" are the result of strong evaporation followed by the active oxidation of a thin layer that consists mainly of iron compounds that have accumulated on exposed rock surfaces.

The excess of evaporation over precipitation has to be made up by underground drainage from the less arid areas that supply the water. The study of the flow of water below the surface in desert areas is demanding, but it is an enterprise that has increasing practical significance. As early as 1940 Bo Hellström of the Royal Institute of Technology of Stockholm made such a study of the eastern Sahara near the Nile. It seems from historical and archaeological evidence that the Khârga Oasis in Egypt was more abundantly supplied with water from springs in ancient times than it is today. Apparently the decline was rather sudden and can be attributed to the puncturing through erosion and subsequent leakage of a subterranean water-bearing rock stratum in the bed of the Nile. This had the effect of decreasing the hydraulic pressure of the underground water over the entire region and thus reducing the discharge of the Khârga springs. Hellström suggested that the erection of a high dam at Aswan in Egypt would in effect repair the damage. The higher head of the impounded waters would then prevent the leakage from the break, which is located just upstream from Aswan.

In a more recent study Robert P. Ambroggi of the United Nations Food and Agriculture Organization also analyzed the problem of water under the Sahara. He reached the conclusion that there are several internal drainage areas below the surface, indicating the existence of large sources of underground water [see "Water under the Sahara," by Robert P. Ambroggi; SCIENTIFIC AMERICAN, May, 1966]. The existence of similar underground flows passing the oases said to be connected with the Nile River were reported by F. M. Ali of the Egyptian Meteorological Department at the Symposium on Tropical Meteorology held in Nairobi in 1960.

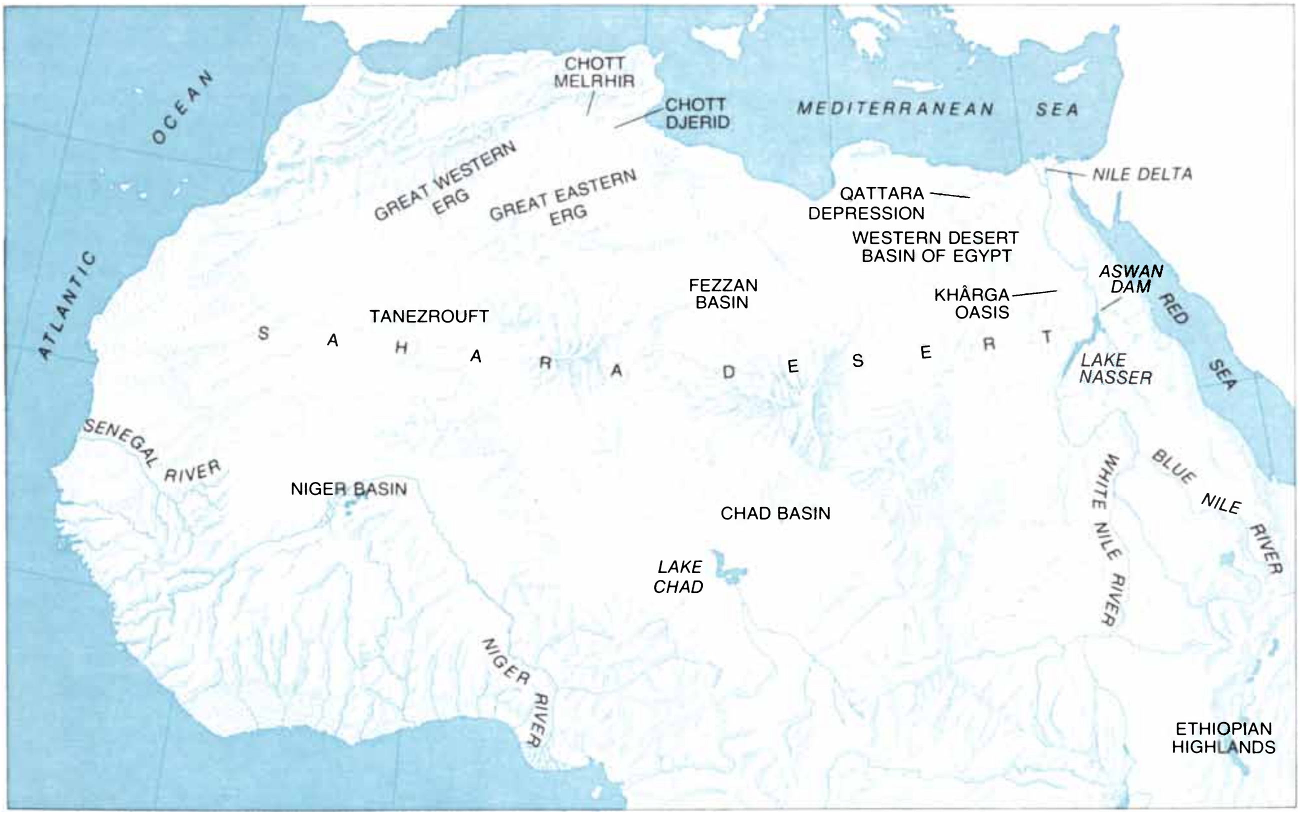
It has been suggested that there may be another example of underground drainage in the Chad basin. According to the French hydrologist G. Drouhin, although Lake Chad receives water from the Sahara and Lagone rivers, its area and salinity content remain constant. Without a sizable subterranean drainage, the salinity due to the combined effects of evaporation and the transport of salts by the rivers should increase. Hydrological observations indicate that there is a flow toward the north-northeast in an aquifer that extends under the lake. In a more recent work based on hydrological and geological evidence Eugene M. Rasmusson of the National Oceanic and Atmospheric Administration found some indications of large sources of underground water in the desert areas of the U.S.

Control of the Water Cycle

In order to decide in what direction the water cycle should be controlled, it is first necessary to determine just what human needs for water are. Water is a necessary component of all industrial or agricultural development. In the Tropics, for example, to raise one metric ton of wheat requires some 8,000 metric tons of water. In industry 200 tons of water are needed to make one metric ton of steel, and 20 tons of water are consumed in refining one ton of gasoline.

The water needs of an "average" person range from 900 cubic meters per year in an agricultural society such as Iran to about 2,700 cubic meters per year in a highly industrial society such as the U.S. Of the 2,700 cubic meters of water consumed by the average American, a mere .5 cubic meter is needed for drinking, 200 cubic meters is used in the home and the rest is spent in industry and agriculture. In the Tropics three cubic meters per year per square meter would be needed for agriculture. Assuming, then, that the water needs of an average man are about 1,500 cubic meters per year, the total world consumption of water would be 4.5 trillion cubic meters per year. Just to keep up with the increase in world population the water demand will increase by about 100 billion cubic meters per year. Since the amount of water that traverses the water cycle in a year is limited, a solution to the water problem should clearly be found soon.

What do we really mean by water need? Water molecules in the water cycle, just like electrons in an electric circuit, do work in traversing a system. For example, the water drunk by a man is not destroyed; it passes through his body, doing work in the process, and then rejoins the water cycle. In this system the body of the man acts as the load in the cycle, whereas the water acts as the flow of electrons. Similarly, the water applied in the irrigation of a crop



SAHARA DESERT has been found in recent years to conceal vast underground water resources concentrated in seven major storage reservoirs: the Great Western Erg, the Great Eastern Erg, the Tanezrouft, the Niger basin, the Fezzan basin, the Chad basin and the Western Desert basin of Egypt (the last being the largest). The total water-storage capacity of these seven basins is estimated at 15.3 trillion cubic meters. Other important hydrologic features of the Sahara region discussed in the text are designated on map.

does work by passing through the plants before rejoining the cycle through evapotranspiration. Basically the sun is the equivalent of the generator in the electric circuit.

Continuing with the electric-circuit analogy, if the work done by water is to be increased, there are three possibilities: (1) to increase the amount of water in each cycle by improving the efficiency of the cycle at the load and at the generator, (2) to increase the number of cycles possible and (3) to accelerate the cycle by increasing the rate of both evaporation and precipitation.

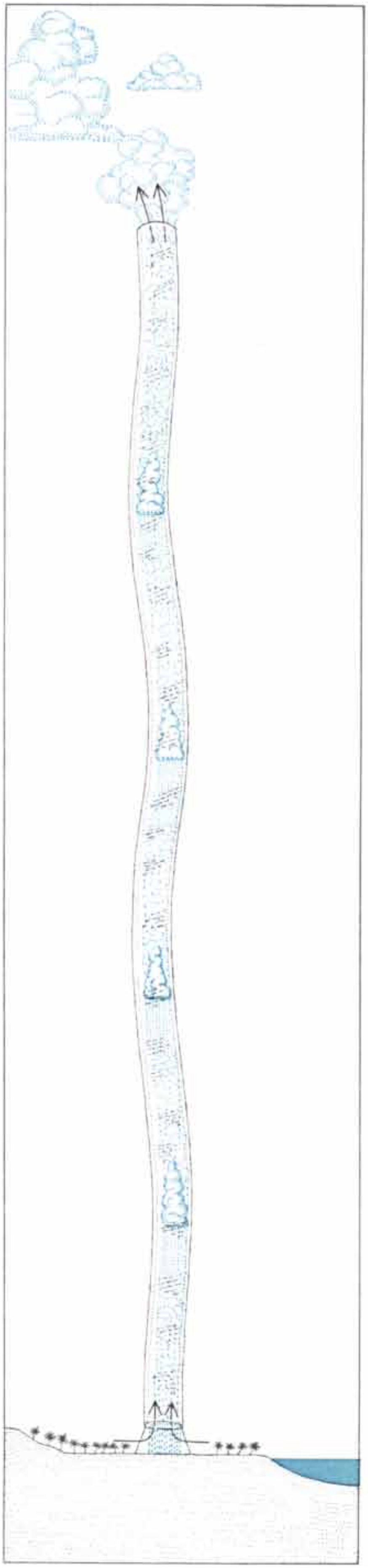
At this point some further considerations should be taken into account in discussing the control of the water cycle. First, water as a carrier is able to do work only when it is in its pure or quasi-pure state. Second, since man is a land dweller, he uses the water on the continents rather than the water in the seas; any water falling on the ocean is lost for human purposes. Third, water is mostly used in the liquid phase. Under these conditions one can establish three principles for the control of the water cycle: (1) Evaporation from the continents, from freshwater surfaces and from underground aquifers should be minimized as much as possible. (2) Since seawater cannot be used as such, evaporation from the ocean surfaces should be accelerated as much as possible. (3) No water should be allowed to proceed in the cycle, flowing into the oceans or evaporating into the atmosphere, until it has done the largest possible amount of useful work. The desalting of seawater is a way to obtain fresh water outside the natural cycle, through the application of nuclear energy, fossil-fuel energy or concentrated solar energy.

Control of the Atmospheric Branch

One way of attacking the problem of lack of water in the arid and semiarid regions of the globe would be to speed up the hydrologic cycle in the atmospheric branch by the artificial inducement of precipitation over the region to be benefited, at times and in amounts previously fixed and with the assurance that the water thus formed would not fall on adjacent regions or on the sea. In fact, the atmosphere can be regarded as a resource for the extraction of liquid water by artificial means. There have been some attempts in this direction through the seeding of clouds with silver iodide, frozen carbon dioxide ("dry ice") and other substances. In spite of the good results obtained under certain conditions, the technique cannot be used at present because artificially induced precipitation is not yet economically justified or even physically defined. Nonetheless, this field of research remains very active. One can be confident that the ultimate success of methods for the artificial release and control of precipitation will depend on the acquisition of a much deeper knowledge of cloud physics, namely the mechanisms of nucleation and the ways in which nature produces rain and snow. Present attempts at rainmaking must thus be undertaken cautiously.

The formation of dew is another way of extracting moisture from the atmosphere. Dew has been a source of water since the Middle Ages, but only recently has it been collected on a larger scale. For example, it is a source of water for human uses in Gibraltar. The capture of water through the interception of fog has also been attempted with some success.

A different line of attack on the problem of control of the atmospheric branch of the water cycle has been advanced by Victor P. Starr and David Anati of M.I.T. They have proposed that liquid water could be recovered from the atmosphere through a partial duplication of natural moist-convection processes, in which air warmed and loaded with water vapor by the heat of the sun rises until it reaches a point at which the water vapor condenses. They suggest an experimental device consisting essentially of a large chimney, perhaps 100 meters in diameter, erected to some height above the free condensation level of a parcel of ground [*see illustration at right*]. This device, called an aerological accelerator, would be self-energizing; no source of energy would be necessary other than the automatic release of the latent heat of the condensing water vapor. Starr and Anati consider a hypothetical experiment within a tube extending to a



HYPOTHETICAL DEVICE for recovering liquid water from the atmosphere by partially duplicating natural moist-convection processes has been proposed by Victor P. Starr and David Anati of the Massachusetts Institute of Technology. Their device, called the aerological accelerator, would consist essentially of a long, double-walled plastic tube filled with helium and allowed to float freely in the atmosphere at a fixed height above a parcel of land in a normally arid coastal region. They suggest a tube with a height of about 3,000 meters and a cross-sectional area of 10,000 square meters. Moist air ascending within the tube should reach the condensation level, leading to precipitation. The scheme could act in part as a mechanism for the desalting of seawater.

height of 3,000 meters and having a cross-sectional area of 10,000 square meters. Assume that the vertical velocity of the moist air is of the order of 20 meters per second, and that one gram of water is released from each kilogram of air in the ascending column. A simple calculation shows that the water obtained would be sufficient to irrigate several square kilometers to a depth of 30 centimeters per year, if meteorological conditions are favorable during a fair fraction of this period.

Starr and Anati have considered Arabia and the Gulf of Mexico as possible locations for experiments with the aerological accelerator. A computation based on aerological data for Aden on June 5, 1958, shows that about 1,000 metric tons of water per hour can conceivably be extracted from the gaseous hydrosphere with a device having about the same dimensions as the one described above. It should be mentioned that no precipitation has been recorded in Aden for the month of June during the 10-year period from 1951 through 1960, and that the mean annual precipitation barely exceeds four centimeters per year.

One can take advantage of the dependence of the convection efficiency on the moisture content of the intake air to appreciably reduce the necessary minimum height of the tube. That would require the evaporation of additional moisture from the sea surface in a proper manner at the intake. The scheme would then act, at least in part, as a mechanism for the desalting of ocean water. From a water-production point of view the "captive rainstorm" would for its size have some advantages over a natural one, since the lateral entrainment of drier air, which has a damping effect on the moist-convection dynamics, would be absent. Thus the presence of the tube should favor the desired process.

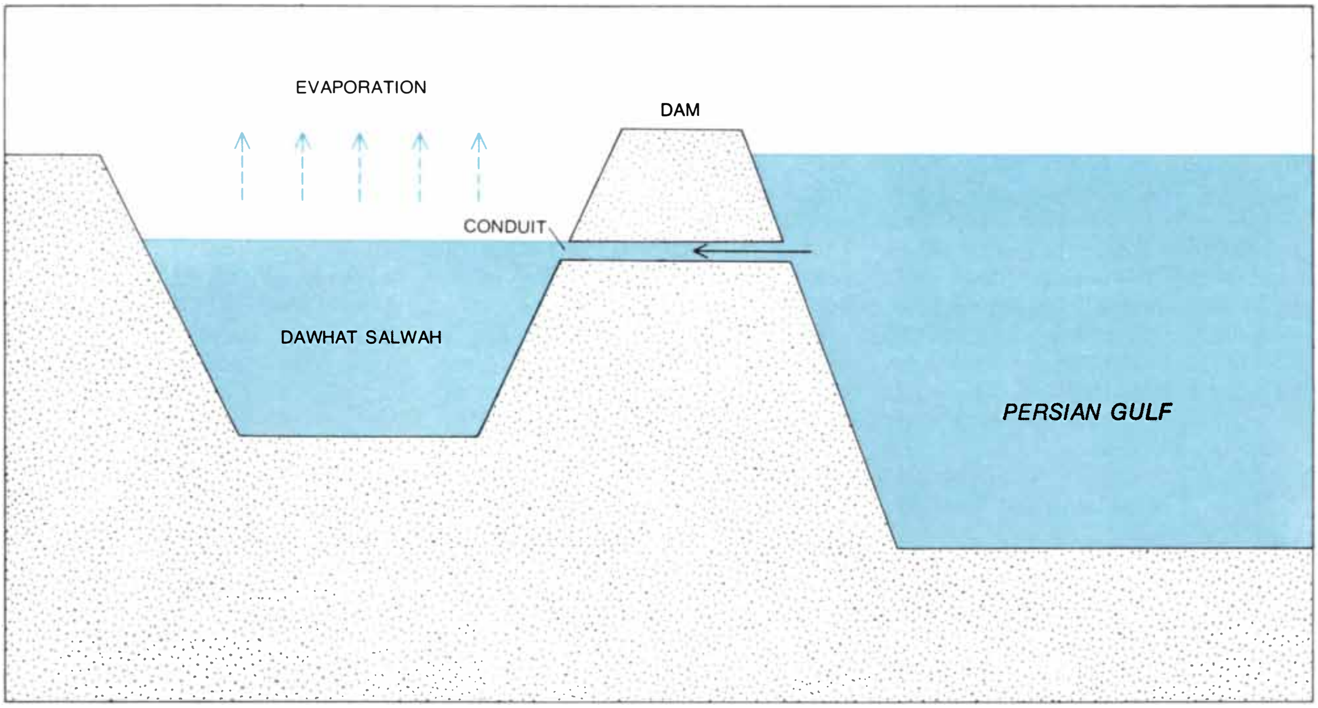
The influence of such a device might extend beyond its immediate location: natural convection at higher levels could develop as a result of its action, and individual cells of free convection could possibly detach themselves from the top of the tube and drift off as autonomous thunderstorms. Obstacles to the construction of an experimental unit include the need for much more study to determine the best theoretical design, the best location and so on, the engineering difficulties connected with the construction, and the high cost of the enterprise. The scientific and practical interest of such a project could nonetheless be great enough to warrant a considerable effort under the proper circumstances.

Real prototypes of the aerological accelerator are found in ventilation shafts of deep gold mines in South Africa. A single such shaft can produce as much as 33,000 cubic meters of water per year. Water obtained in this way has actually been put to agricultural use. The greatest obstacle to such schemes is the need for a mechanism to remove the raindrops from the rising air.

SCALE of the heliohydroelectric project advanced by Kettani and Gonsalves is suggested by the map of the entire Persian Gulf area at top and by the more detailed map at bottom. The total surface area of the proposed reservoir (*lighter shades of color*) would be 6,460 kilometers. The marginal lands that could be reclaimed from the sea in Dawhat Salwah, amounting to some 4,100 square kilometers, are indicated in the lightest shade of color.

Control of Surface Runoff

The arid regions around the world include most of the African continent outside its equatorial zone, the entire Mid-



CONDUIT would link the Dawhat Salwah reservoir with the open sea. The rapid loss of water due to evaporation in the confined reservoir would create a hydraulic head across the conduit, which could be exploited to produce electric power by controlling the flow of water in such a way that water is admitted to the conduit at the same rate at which it evaporates in the reservoir. For example, with a net evaporation rate of 3.5 meters per year and an optimum head of 12.4 meters, 300 million kilowatt-hours of electricity could be produced. In addition it has been suggested that magnesium could be obtained from the reservoir by precipitation methods.

dle East, all of central Asia, most of Australia and most of the western part of the Western Hemisphere. Many of these regions lie near the shore, where water is plentiful. One might therefore be tempted to seek a solution to the water problem in desalination projects. The largest desalting plant in the U.S., however, produces no more than 3.6 million cubic meters of fresh water per year, merely enough to satisfy the needs of a town of 3,600 people. The largest desalting plant in the U.S.S.R. is expected to yield 55 million cubic meters per year, enough for a city of 55,000 people.

These figures compare poorly with the discharges of many rivers that are allowed to dump huge quantities of fresh water into the sea, often in areas where these waters are most needed. For example, the water discharged by the Amazon (6.8 trillion cubic meters per year) is enough to satisfy the needs of almost twice the present population of the entire earth. The Congo River would satisfy the needs of four times the population of Africa (1.35 trillion cubic meters per year). The Mississippi alone discharges into the ocean every year as much water as the entire consumption of the U.S. (560 billion cubic meters per year). This waste goes on even in areas where water is most badly needed, such as the Middle East. There the Tigris and the Euphrates dump into the Persian Gulf as much water as the entire consumption of Iran. If this water could be brought to Arabia by pipeline, the entire economy of the Arabian peninsula would be changed.

To cite just one example where the control of running water might boost the economic capacity of a country, consider the case of Egypt. That country is literally a gift of the Nile; it has a population of 34 million people, two-thirds of whom live in the 25,000 square kilometers of the Nile delta. The remaining third live in about 14,000 square kilometers along the banks of the river. Thus the entire population is concentrated in a mere 39,000 square kilometers in a country with a total area of one million square kilometers. The Nile itself has an average yearly discharge at Aswan of 92 billion cubic meters, enough for a population of 100 million people if utilized fully and efficiently. The waters of the Nile, which cross Egypt on a 1,550-kilometer path, come from the Ethiopian highlands (84 percent) and from the lakes plateau of central Africa (16 percent).

The construction of the high dam at Aswan certainly reduced the annual amount of water lost to the Mediterranean. It also brought 28,000 square kilometers of desert land into cultivation (of which 8,000 square kilometers are in Egypt and the rest in the Sudan) and generates 10 billion kilowatt-hours of electric energy. The lake behind the dam, however, introduces a new loss, a loss due to evaporation. (Seepage is not considered a direct loss because it helps to refill the underground aquifers.) Assuming an evaporation rate of three meters per year, one can see that as much as 24 billion cubic meters of water would be evaporated yearly from the 7,800-square-kilometer surface of the lake. This amount is huge, representing about a fourth of the entire discharge of the Nile at Aswan. Either water reservoirs presenting such a large free surface to the sun should be avoided in arid zones or a solution to the evaporation problem should be found. It might be well to examine the possibility of covering the entire evaporating surface by a layer of some material that would eliminate or greatly reduce evaporation, such as a layer of liquid hexadecanol, plastic covers, wood plates or dyes to retard heat absorption.

The Qattara Depression, some 200 kilometers west of Cairo and 60 kilometers south of the Mediterranean, has long fascinated engineers and scientists. The depression has a surface area at sea level of about 20,000 square kilometers, and its lowest point is 134 meters below sea level. It has been proposed that the depression be used for power generation by admitting seawater from the Mediterranean. It would seem more beneficial, however, to devote this large area to cultivation and settlement. If only a ninth of the Nile were diverted into it, the depression could be made the home of as many as 10 million people, and the silt brought by the Nile could transform large areas into farmland. The waste water would reenter the water cycle through evaporation.

Control of Underground Water

In sloping soils about 20 percent of precipitation infiltrates the ground. In flat, permeable soils the figure can be as high as 80 percent. These waters feed underground aquifers and make up for their loss through natural springs and evapotranspiration.

Underground seas have been discovered in arid regions such as Iran, Arabia and North Africa. Ancient peoples were already acquainted with these aquifers; for example, some 3,000 years ago a system of qanats, or underground channels, was built in Iran for tapping underground water [see "The Qanats of Iran," by H. E. Wulff; SCIENTIFIC AMERICAN, April, 1968]. The system was expanded under the Islamic caliphates to the rest of the Middle East and to North Africa. The qanat system consists of a simple network of underground channels bringing water by gravity from the highland aquifers to the lowlands. Even today this system supplies 75 percent of the water needs of Iran.

The underground water in Arabia was discovered as a result of drilling for oil. At present a great effort is being made to develop its potentialities. The eastern region of Arabia alone has seven major aquifers in Cenozoic and Mesozoic strata, with hydraulic gradients decreasing from west to east. Near the Persian Gulf the Mesozoic strata reach a depth of 1,000 meters. The strata are not too deep to be mined efficiently.

The discovery of the vast underground water potential of the Sahara by Ambroggi and others confirmed a prediction made by Starr and one of us (Peixoto) in 1958 on the basis of aerological findings. There are seven major storage basins under the Sahara, having a total area of 4.5 million square kilometers and a total water-storage capacity of 15.3 trillion cubic meters. They are the Great Western Erg, the Great Eastern Erg and the Tanezrouft in Algeria; the Niger basin in Mali; the Fezzan basin in Libya; the Chad basin in Chad, Niger and Nigeria, and the Western Desert basin of Egypt [*see illustration on page 56*]. The last is the largest, with a total area of 1.8 million square kilometers and a total water-storage capacity of six trillion cubic meters. It extends into neighboring African countries.

Consider the Great Eastern Erg of Algeria, with an area of 375,000 square kilometers and a water-storage capacity of 1.7 trillion cubic meters. Most of this water is fossil water gathered after the last ice age, when the Sahara must have been a tropical region with heavy rainfall. At present the aquifer is being recharged by the flood rains from the northern Tell mountains. The total natural recharge is estimated to be about 900 million cubic meters per year. Since the total capacity of the aquifer does not change appreciably, there has to be some discharge. That discharge is evaporation from a large depression called Chott Melrhir-Djerid. The chott is dry except during rainy seasons, and its lowest point is 31 meters below sea level. It extends to Tunisia, where only about 20 kilometers of sandy soil separates it from the Mediterranean. Chott Melrhir proper is entirely below sea level and extends over an area of about 10,000 square kilometers.

If Chott Melrhir were flooded entirely by water from the Mediterranean, the discharge of the Great Eastern Erg through evaporation would be greatly reduced, thus limiting the loss of usable underground water. Moreover, the oil of eastern Algeria could be shipped directly from where it is produced, and great quantities of salt minerals could be extracted from the newly formed gulf. The annual supply of fresh water made available in this way would be enough to support the settlement of about a million people in a region that now has scarcely 200,000 inhabitants.

Control of the Oceans

Some attempts have been made to control the water cycle over the oceans. The desalting of seawater is only an example of such attempts. Schemes on a broader scale have been proposed, such as the "heliohydroelectric" project put forward in 1970 by one of us (Kettani) in collaboration with L. M. Gonsalves. The proposed scheme utilizes solar energy to generate through evaporation an inflow of water from the Persian Gulf to an artificially made reservoir called Dawhat Salwah. The reservoir, which would be formed by building causeways between Saudi Arabia and Bahrain and between Bahrain and Qatar, would have a total surface area of 6,460 square kilometers [*see illustrations on page 58*].

The system formed would consist of two reservoirs linked by a conduit [*see illustration on preceding page*]. The rapid loss of water due to the evaporation in the confined reservoir would generate a hydraulic head across the conduit, since the water level in the open sea would not be affected. Thus the flow of water across the conduit could be controlled in such a way that if water were admitted in the conduit at the same rate at which it evaporated in the enclosed reservoir, electric power could be generated. With a net evaporation rate of 3.5 meters per year and an optimum head of 12.4 meters, 300 million kilowatt-hours of electricity could be produced. Four other avenues of economic exploitation have been considered: mineral production, road communication, land reclamation and sea farming.

The marginal lands that would be gained from the sea in Dawhat Salwah would amount to about 4,100 square kilometers. The effect on Bahrain would be considerable: its area would increase almost four times. Because the amount of water in the atmosphere cannot be increased, the effect of creating such an artificial depression would be to raise the level of all the oceans of the world by .25 millimeter. Presumably nobody would object to this rise.

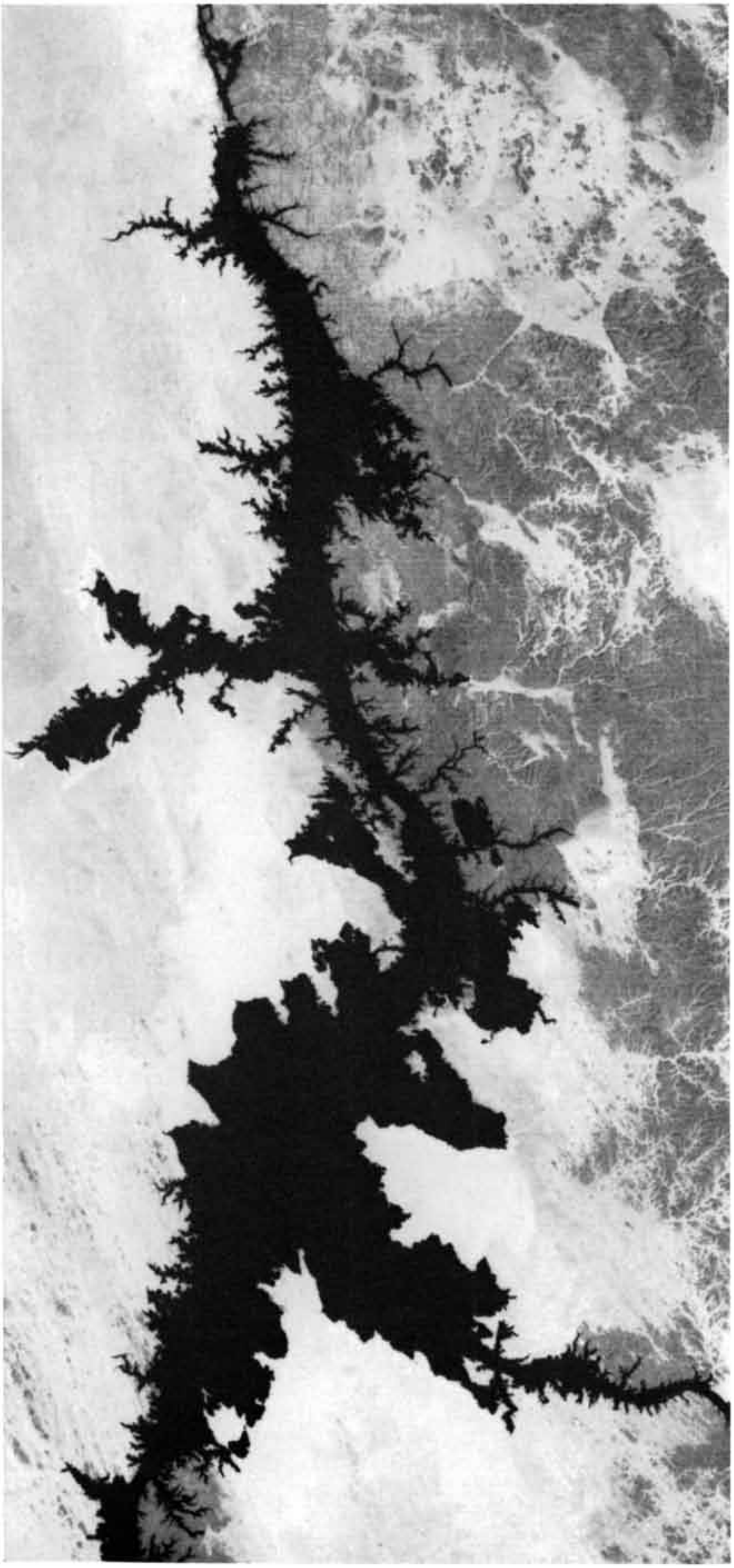
The project is actually a replica of a natural phenomenon in central Asia, where the Caspian Sea sends 8.4 billion cubic meters through a narrow strait into a 10,400-square-kilometer desert basin called the Kara Bogaz Gol [see "The Sea that Spills into a Desert," by Maurice A. Garbell; SCIENTIFIC AMERICAN, August, 1963]. The Caspian, however, is a closed sea, and whatever water it loses goes directly into the overall water cycle. As a result the surface level of the Caspian has been going down for the past 90 years. The flow into the Kara Bogaz Gol has not been large enough to keep up with the evaporation rate, and the area of that desert lake is still decreasing.

Antarctic ice is the most important source of fresh water waiting to be tapped. The feasibility of towing Antarctic icebergs to arid areas in America, Africa, Asia and Australia has been considered. It has been suggested that it would be possible to tow a train of 10 icebergs containing a total of about 10 billion cubic meters of fresh water to southern California in about 10 months, with a loss of water content of only 10 percent. The icebergs could be moored at their destination and then quarried for human use.

Epilogue

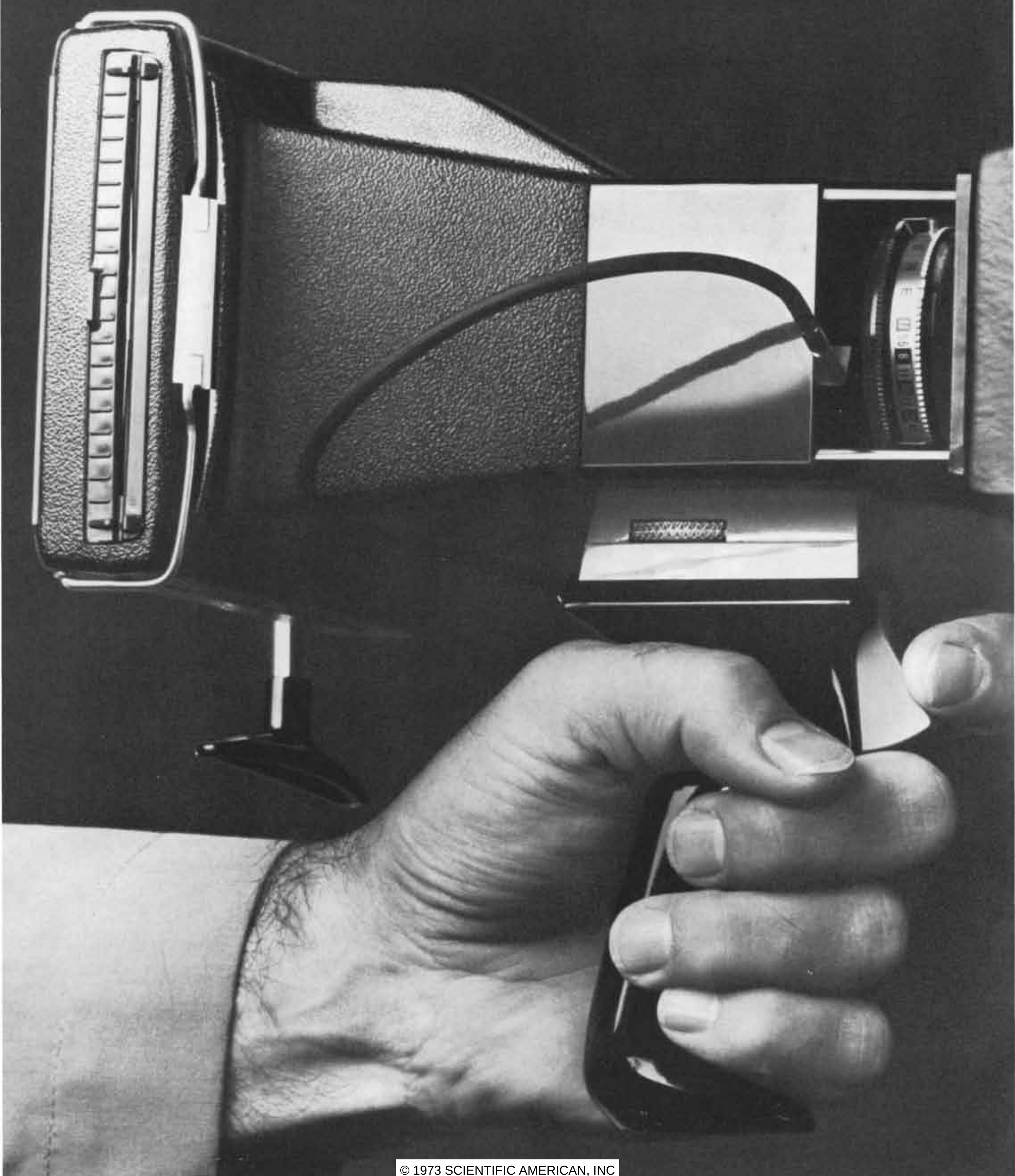
The hydrologic cycle on a planetary scale can be regarded as the result of a gigantic distillation scheme extending over the entire earth. The heating of the subtropical regions by solar radiation leads to a continuous evaporation of water, which is released into the atmosphere and is removed by the wind circulation and transported, mostly in the lower half of the atmosphere, northward and southward into other latitudes. During the transfer process in the gaseous phase, because of cooling due to expansion, some of the water vapor eventually condenses to form clouds, which lead to precipitation over the equatorial regions and over the middle and high latitudes. Finally the combined action of oceans, rivers and underground flows provides the compensating return of water in the liquid or solid phases to the source regions, closing the sequence of natural phenomena involved in the transition phase of the water. Thereafter the water is once again available to be released into the atmosphere and to start the cycle anew.

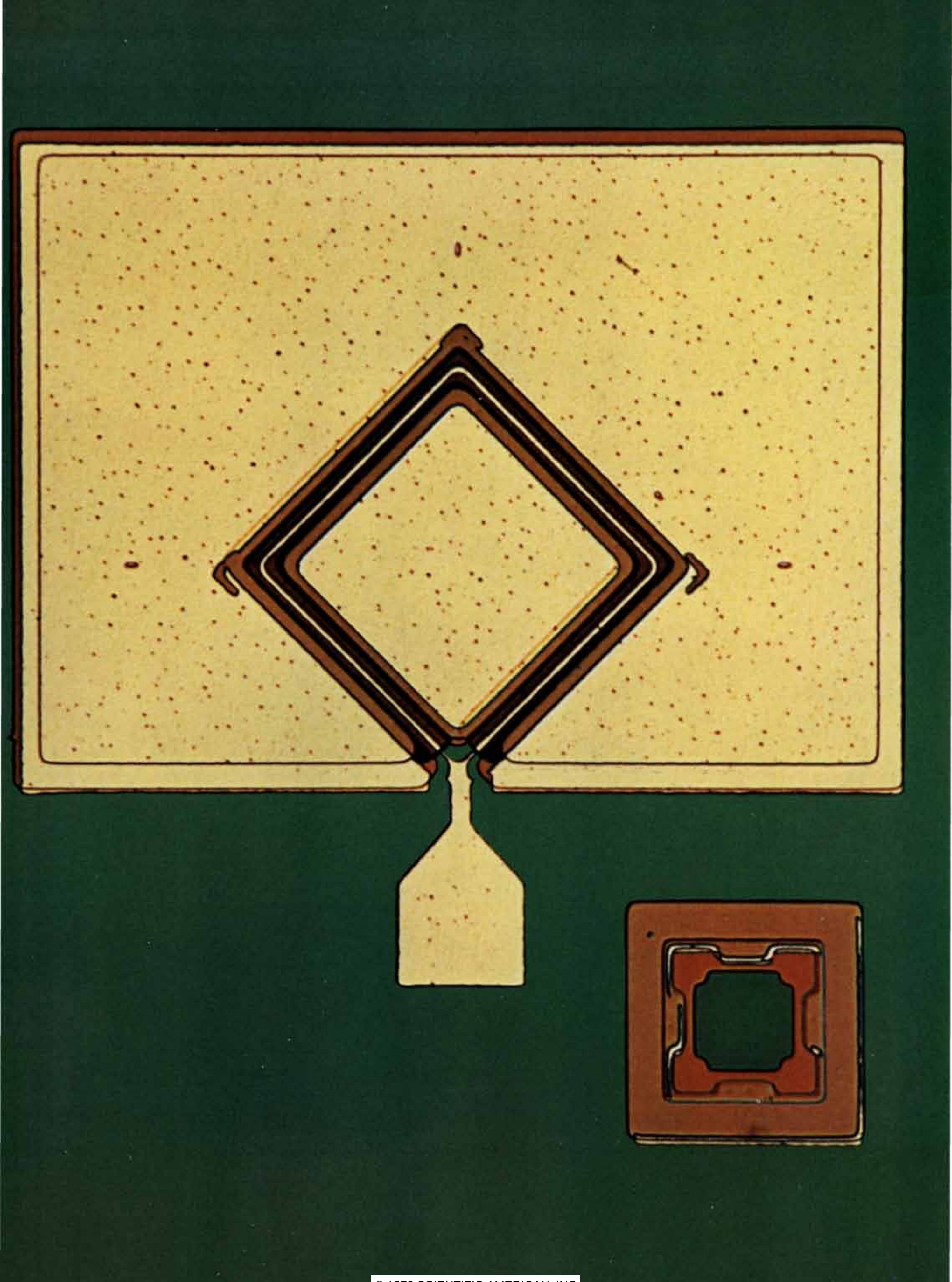
Thus the hydrologic cycle is maintained by the general circulation of the atmosphere. We cannot understand the mechanisms of the hydrologic cycle unless we understand the mechanisms of the general circulation, and vice versa. This complex feedback process leads to a highly nonlinear coupling of the hydrologic cycle and the general circulation, which if well understood will make their control a foreseeable possibility.



ASWAN DAM has contributed significantly to the economic capacity of the Nile valley region by reducing the annual amount of water lost to the Mediterranean, by bringing thousands of square kilometers of desert land into cultivation and by generating prodigious amounts of electric energy. The loss of water due to evaporation from the lake behind the dam, however, is huge, representing about a fourth of the entire discharge of the Nile at Aswan. One solution to the evaporation problem that has been suggested is to cover the entire evaporating surface of the lake by a layer of some material that would retard heat absorption. The dam is at upper left in this ERTS photograph.

5 good reasons why it had to look this





ION IMPLANTATION

Microelectronic devices are now being "doped" with impurity atoms by firing ionized atoms into the device with an accelerator. The technique offers greater precision than the older diffusion method

by Frederick F. Morehead, Jr., and Billy L. Crowder

The size of microelectronic circuits has decreased to the point where their surface dimensions are measured in microns (thousandths of a millimeter) and their thickness in angstroms (ten-millionths of a millimeter). In general shrinking the surface dimensions of an electronic device such as a transistor also requires shrinking the vertical dimensions. For example, to make transistors with surface dimensions on the order of a micron requires the introduction of impurity atoms to within less than a micron of the surface. Such precise "doping" cannot readily be accomplished with the thermal-diffusion techniques that currently dominate the semiconductor market. Now, however, what was once only a research tool of nuclear physicists, the ion accelerator, offers a means of precisely implanting atoms of any desired species. The ion-implantation technique consists in accelerating ions in an electric field to energies high enough so that the ions will penetrate for some distance into a solid semiconductor.

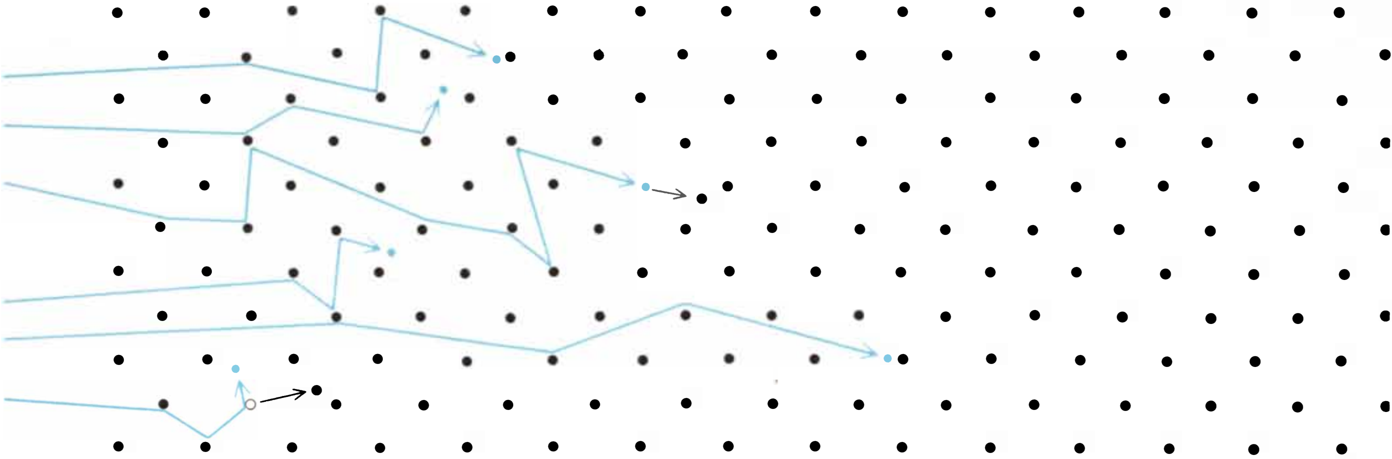
Many of the advantages of ion implantation for doping semiconductors were anticipated in a patent submitted in 1954 by William B. Shockley, one of the inventors of the transistor. Shockley recognized that the depth of the introduced ions could be controlled by controlling the acceleration energy, and that the position of the ions could be controlled by masking the semiconductor surface. He was right in his belief that the precision of ion implantation would be much greater than that of thermal diffusion, but he was wrong in his expectation that the depth of the implanted ions could be controlled within a few angstroms. It turns out that the depth at which projectile ions come to rest within the solid target can be predicted only within a few hundred angstroms.

In the past few years the technique of implanting accelerated ions in silicon has been intensively examined in nearly every major laboratory involved in semiconductor research. Current methods of controlling the area of ion implantation rely on masking those portions of the semiconductor in which doping is not desired. Masks can be made from various types of material, for example insulators such as silicon dioxide, silicon nitride and aluminum oxide or metals such as aluminum, molybdenum and gold. To expose the regions in which ion implantation is desired the mask is removed by chemical etching. This removal is usually accomplished by coating the mask with a photosensitive or electron-sensitive material. Areas of the sensitive material exposed to light or electrons become soluble in a special solvent and are etched away to uncover the masking material. The uncovered areas of the mask are in turn etched down to the surface of the semiconductor. The semiconductor is then bombarded by accelerated ions, which penetrate the semiconductor only in the areas that are not protected by the masking material.

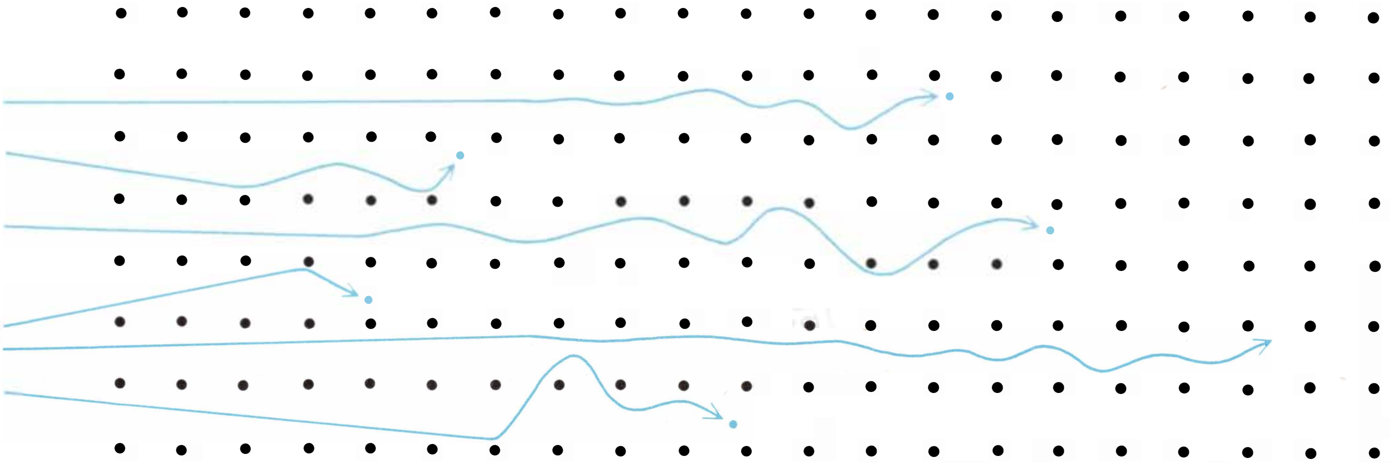
Since the accelerated-ion technique offers fairly precise control of both the number of ions implanted and the depth of implantation, it is gradually replacing the older diffusion method of fabricating electronic devices with dimensions on the order of several microns. To produce devices with dimensions of one micron or less it will be necessary to combine the ion-implantation technology with the new electron-beam method of delineating patterns on the masking material [see "Microcircuits by Electron Beam," by A. N. Broers and M. Hatzakis; SCIENTIFIC AMERICAN, November, 1972].

The ultimate ion-implantation system would be one in which no masks are used. A tiny ion beam less than one micron in diameter would write patterns directly on a semiconductor surface. Such a system, however, calls for tremendous advances both in ion-source design and in ion-beam control. The best ion sources available are not nearly intense enough to produce a sufficient number of ions in a micron-sized beam. Moreover, the control of an ion beam is inherently more difficult than the control of an electron beam because electrostatic lenses must be used rather than electromagnetic lenses. Fortunately for the foreseeable future the combination of ion implantation and electron-beam-generated patterns provides sufficient preci-

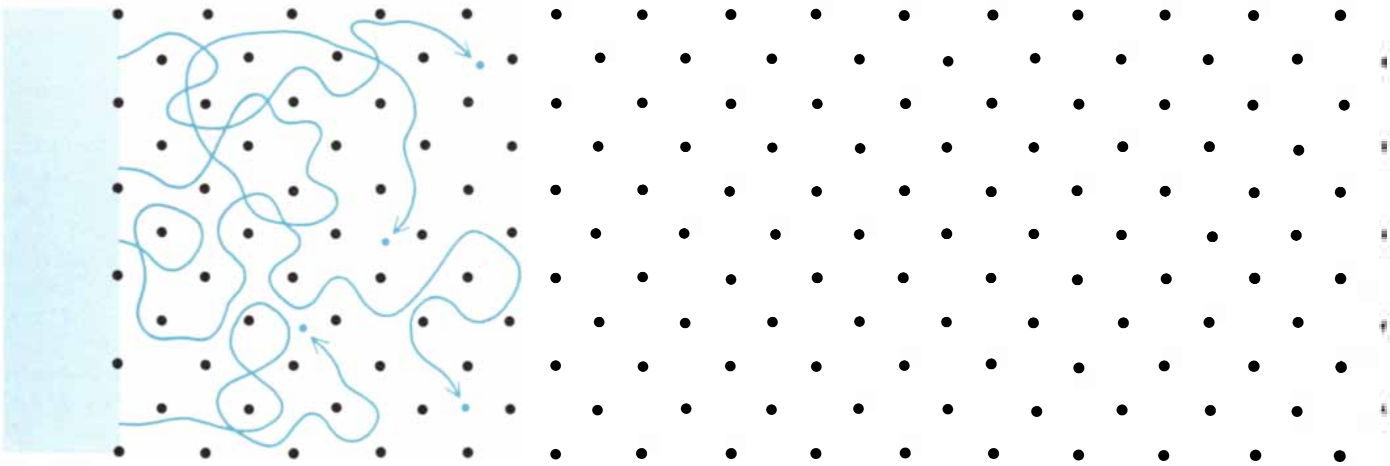
MOSFET DEVICE, a metal-oxide semiconductor field-effect transistor, shown in the photograph on the opposite page is used in high-speed switching applications. The device, which has a maximum oscillation frequency of 14×10^9 hertz, owes its high speed to doping by the implantation of accelerated ions. The ion implantation was done in the area that appears as a brown diamond. The thin, light-colored line in the center of the brown diamond is the gate electrode, which measures one micron in width and about 100 microns on a side. The gold tab below the diamond is the gate contact. The gate electrode and all the gold regions are aluminum. The gold diamond in the center is the drain electrode and the surrounding gold area is the source electrode. The square at bottom right is a registration pattern that is used in the photoengraving techniques for the fabrication of the aluminum electrodes. The green areas are silicon oxide. More than a million MOSFET devices with gates smaller than the one that is shown here can be fabricated simultaneously on a silicon wafer that is one inch in diameter. The photograph was made by F. W. Goro with a Leitz Orthoplan microscope using differential interference contrast illumination.



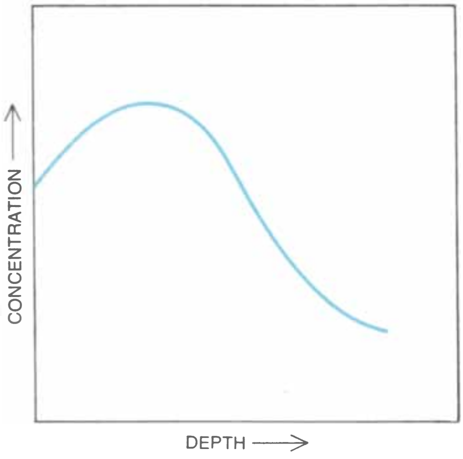
PATH OF AN ACCELERATED ION in a solid is affected by its collisions with the nuclei and electrons of the solid. Here boron nuclei (*color*) are shown penetrating a crystalline lattice of silicon atoms (*black*). The ions are deflected and slowed by collisions with silicon nuclei and are also slowed by interactions with electrons in the solid. Some of the silicon nuclei are displaced by the



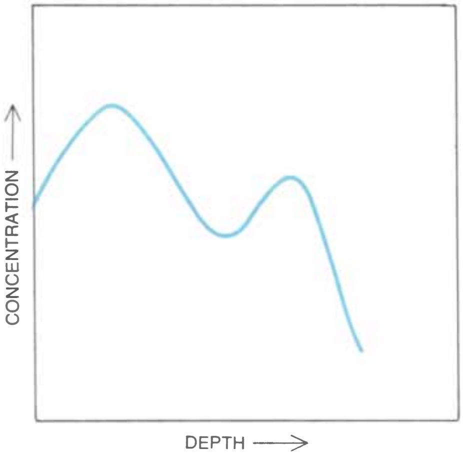
CHANNELING increases the penetration of implanted ions. In a crystalline solid an ion may channel between rows of the lattice, as shown here, and such an ion is slowed by its interactions with electrons. The normal-curve distribution for depth of penetration, shown at the top of the page, is therefore actually typical only in noncrystalline solids. If ions are aimed to penetrate along an



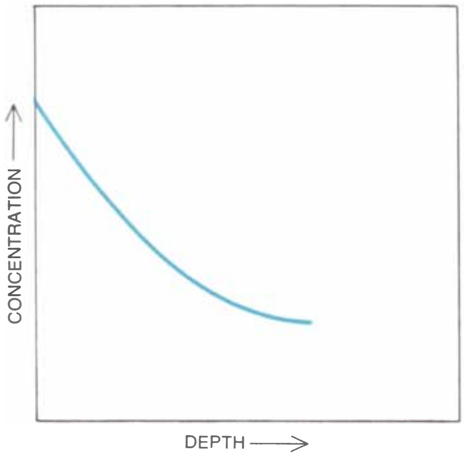
DIFFUSION of impurity atoms (*color*) into a solid, a more conventional process than implantation by acceleration, is usually done by applying a vapor of the impurity metal to the surface of the semiconductor solid. The atoms penetrate the solid in a "ran-



boron ions (*short arrows*). The final distribution of the implanted boron ions in the solid approximates normal curve (*right*).



"open" direction in the lattice, there are two distribution peaks: one for deflected ions and one for channeled ions (*right*).



dom walk." Most of the impurity atoms remain near surface, as curve (*right*) indicates.

sion to meet the miniaturization requirements of integrated electronic circuits.

When a highly accelerated ion strikes a solid target, it loses energy within the solid by collision with the nuclei of the target atoms and by interaction with the electrons in the solid. The collisions with nuclei deflect the ion from its original course; the interactions with electrons merely subtract energy from the ion without changing its direction. For a given kinetic energy the range of penetration of an ion into a solid depends principally on the atomic weight of the ion and on the number of atoms per unit volume in the solid. Penetration by heavier ions is always much less than that by lighter ions. Because of the random nature of the collisions of the ions and the nuclei of the atoms in the target, the final distribution of the implanted ions in the solid approximates a bell-shaped curve. The peak of the distribution is called the projected range. The range distribution of ions implanted into noncrystalline solids can be predicted fairly accurately on the basis of a theory worked out by J. Lindhard, M. Scharff and H. E. Schiøtt at the University of Aarhus in 1963.

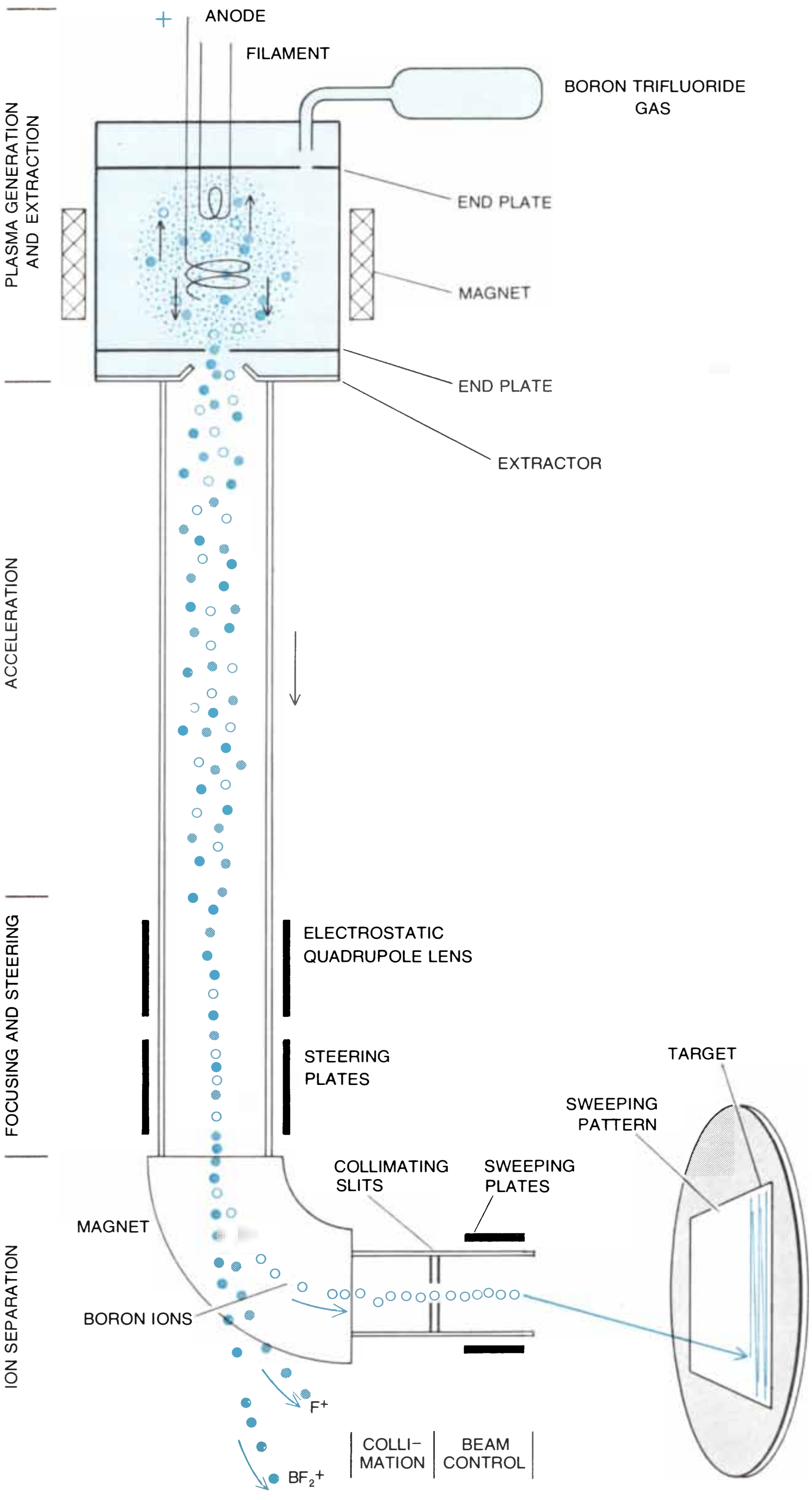
Crystalline solids present many channels that ions can follow. Along these channels the probability of an ion colliding with a nucleus is greatly diminished, and hence the range of the ion is greatly increased. A deeply penetrating ion is slowed down only by interaction with the electrons in the channel. When the dose of ions is large, however, the channels become clogged and the median range of the ions is decreased.

Some of the atoms in the target struck by the ions will be displaced from their sites in the crystal lattice. This radiation damage must be repaired by annealing if the target is to have the desired semiconducting properties. The temperatures required for the annealing of crystals implanted with ions are much lower than those required for introducing the same ions into the crystal by thermal diffusion. With silicon, for example, temperatures between 1,000 and 1,100 degrees Celsius are needed in order to introduce dopants by diffusion. If the same dopant is put into the crystal by ion implantation, temperatures as low as 550 degrees C. are sufficient for annealing the radiation damage. It is an interesting fact that a large amount of radiation damage is more effectively annealed than a lesser amount. If the radiation damage is great enough to completely destroy the silicon's lattice structure near the surface, the amorphous layer will recrystallize when it is annealed at 550 degrees. A pattern is imposed on the amorphous layer by the underlying crystalline silicon. The process of regrowth removes most of the defects in the crystal, leaving only the implanted ions as the electrically active sites. If, on the other hand, the radiation damage is not great enough to produce a continuous amorphous layer, defects will remain until the material is annealed at temperatures above 800 degrees C.

Much of the work on ion implantation in semiconductors has been done with equipment that was originally designed for research in nuclear physics. High-voltage ion accelerators were developed by physicists primarily for the study of the nuclear reactions produced by bombarding nuclei with high-energy particles. With modifications these accelerators can be used for implanting ions in solids. For example, in our laboratory at the Thomas J. Watson Research Center of the International Business Machines Corporation we have 50-kilovolt, 300-kilovolt and 2.5-million-volt accelerators for ion implantation.

The 50-kilovolt accelerator serves for studies that do not require high-energy ions. Most of the implantation work is carried out with the 300-kilovolt machine, which has a power source of the Cockcroft-Walton type, a combination of high-voltage transformers and rectifiers submerged in a large vat of oil. The ions to be accelerated are produced at the high-voltage end of the system. The desired ion species is introduced as an un-ionized gaseous compound into a chamber that contains an electron source. Electrons boiled out of a heated filament are attracted toward a positively charged electrode in the middle of the chamber. The electrons pass through the anode and are repelled by a negatively charged end plate, whereupon they go back through the anode and are repelled by another end plate. Each such oscillation results in some collisions between electrons and gas atoms. These collisions ionize the gas. A magnetic field confines the ionized gas and increases the efficiency of the ionizing process.

The positive ions are now extracted and accelerated in a high-voltage drift tube. The resulting ion beam is focused and passed through a strong magnetic field that bends the ions through a 90-degree turn. Since the desired ion species is introduced in a compound, more than one kind of positive ion emerges from the ionized gas, and it is necessary to separate the unwanted species. This separation is accomplished by adjusting the strength of the magnetic field so



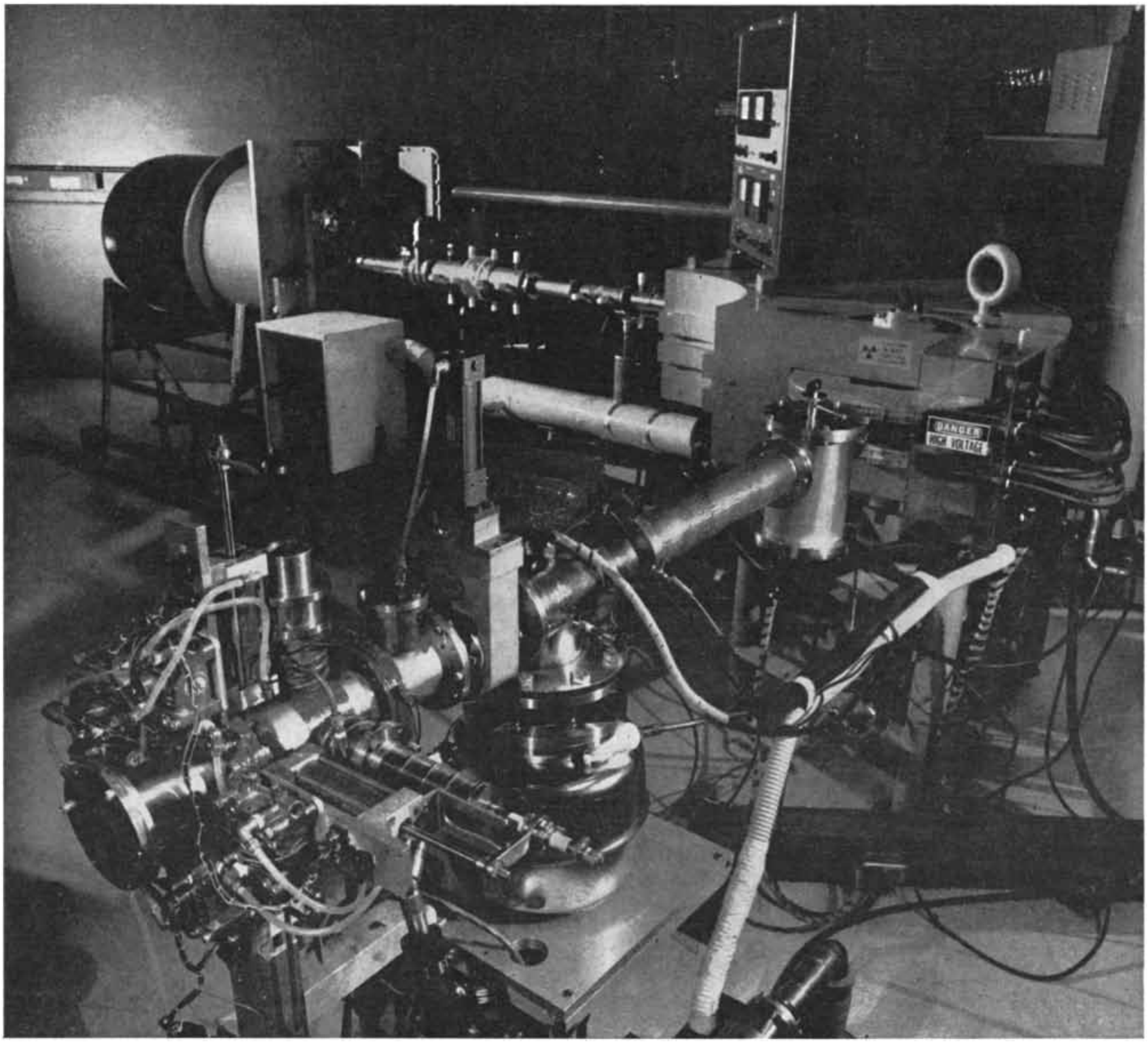
ION IMPLANTATION is accomplished with an ion accelerator modified for the purpose. The ions are generated in a plasma at the high-voltage end of the system (*top*). A gas containing the desired atoms is introduced into the chamber. Electrons from a heated filament are accelerated back and forth between two end plates by an anode; the electrons collide with gas atoms, ionizing them. The ions are extracted by applying a negative potential between the end plate with a central aperture and the extractor, and are accelerated by the difference in potential between the two ends of the system. The desired ion species (boron in this case) is selected by bending the beam in a magnetic field (*bottom left*). The boron-ion beam is collimated and focused and is scanned back and forth across the target.

that only the desired ions are bent toward the target. After collimation and focusing, the ion beam is about a millimeter in diameter. In order to get uniform coverage of the two-centimeter-square target area the beam is deflected both horizontally and vertically so that it sweeps across the target area in a raster pattern similar to that of the electron beam in a television set.

Since the target is at ground potential, the current representing the number of ions hitting it can be measured directly. The 300-kilovolt accelerator can handle four different targets at one time; each target is inserted into the beam and then withdrawn by remote control. The targets can also be heated electrically to 600 degrees C. or cooled by liquid nitrogen to minus 170 degrees. Keeping the target at a high temperature during ion bombardment prevents the formation of amorphous zones. Moreover, if the hot crystal is carefully oriented, the number of ions that are channeled to greater depths is increased. Keeping the target at low temperature greatly decreases the dose of ions needed to produce a continuous amorphous layer, particularly with light ions such as those of boron.

Commercial machines that combine powerful ion sources and large target areas are available. The Lintott ion-implantation system, for example, can produce an isotopically pure ion beam at 40 kilovolts and at high current. The target can be electrically charged to as much as 150 kilovolts, so that the total accelerating potential can be as high as 190 kilovolts. Whereas in our 300-kilovolt system the ion beam is swept over the target by deflecting it electrostatically, in the Lintott system the beam is positioned by moving the target. As many as 400 one-inch silicon wafers can be loaded into the machine at a time and processed to a uniformity of within 1 percent in two hours. If each silicon wafer contained 200,000 devices (the present limit of diffusion methods for producing them), this machine could produce up to 400 billion devices a year. As methods for increasing the number of devices per wafer are developed, it may be possible for a single machine to produce more than a trillion devices a year, which is roughly the current total world production of semiconductor devices.

The 2.5-million-volt Van de Graaff accelerator makes it possible to implant ions much deeper. It can be employed to bombard target atoms with light but energetic particles, such as helium nuclei. The specific nuclear reactions and

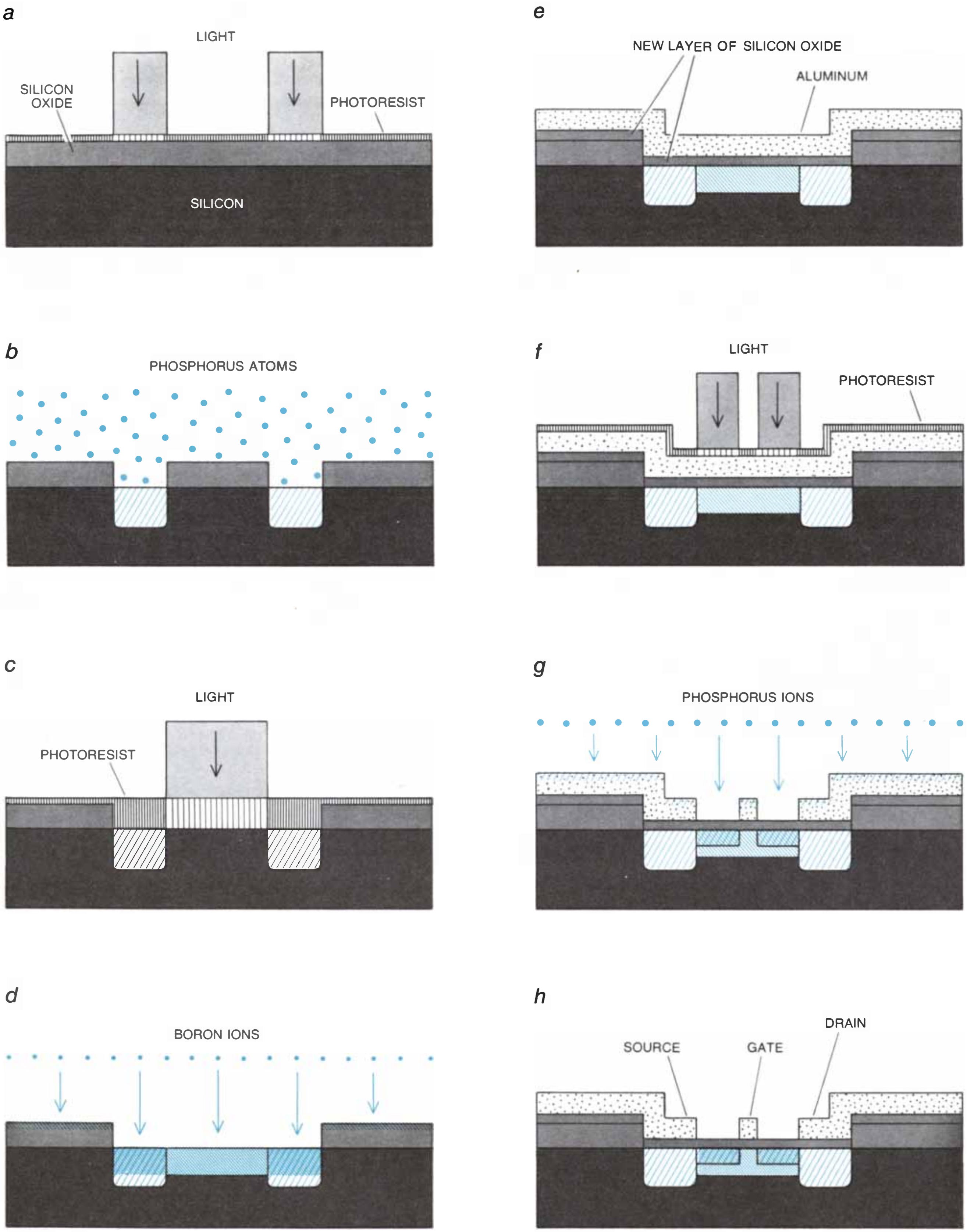


THREE-HUNDRED-KILOVOLT ACCELERATOR at the Thomas J. Watson Research Center of the International Business Machines Corporation was modified so that up to four targets can be inserted into the ion beam and withdrawn by remote control. In the target area (*foreground*) the targets can also be heated electrically or cooled by liquid nitrogen. The plasma is generated in the dome (*top left*), which also encloses the high-voltage drift tube in which the ions are accelerated. The beam goes through an evacuated tube in which it is partially focused to a large magnet (*upper right*), which bends desired ions 90 degrees to aim them toward the target.

the backward scattering of particles that result from this kind of bombardment can provide valuable information about such things as the distribution of ions in a crystal lattice. These techniques are beginning to give us a better understanding of the components of thin-film electronic devices, such as transistor circuits involving several metals, switches between metals and semiconductors, and junctions between semiconductors and insulators.

One intriguing aspect of ion implantation lies in adding extremely small amounts of dopant (on the order of parts per million) to compound semiconductors, many of which are used as light-emitters. With such small amounts of dopant material it is often difficult to determine if the dopant is responsible for altering the characteristics of a semiconductor. For example, incorporation of oxygen in zinc telluride makes it an efficient emitter of red light. In order to identify the precise role played by the oxygen James L. Merz and L. C. Feldman of Bell Laboratories attempted to incorporate the rare oxygen isotope O-18 into the zinc telluride crystal while it was being grown. The experimental results were complicated by the fact that the crystals were contaminated with large amounts of the omnipresent common oxygen isotope O-16.

It was possible to circumvent this problem with an ion accelerator. The deflection magnet of the machine can produce a beam of isotopically pure ions, so that the quantity of atoms of that isotope in the beam is known. Hence Merz and Feldman could now shoot into a crystal of zinc telluride a known quantity of O-18 atoms. Incorporation of the heavier oxygen into the semiconductor causes a slight but measurable shift in the spectrum of the emitted light. The shift was easily detectable following ion implantation. Furthermore, by masking half of the semiconductor from the ion beam they were able to show that it was the implanted oxygen that gave rise



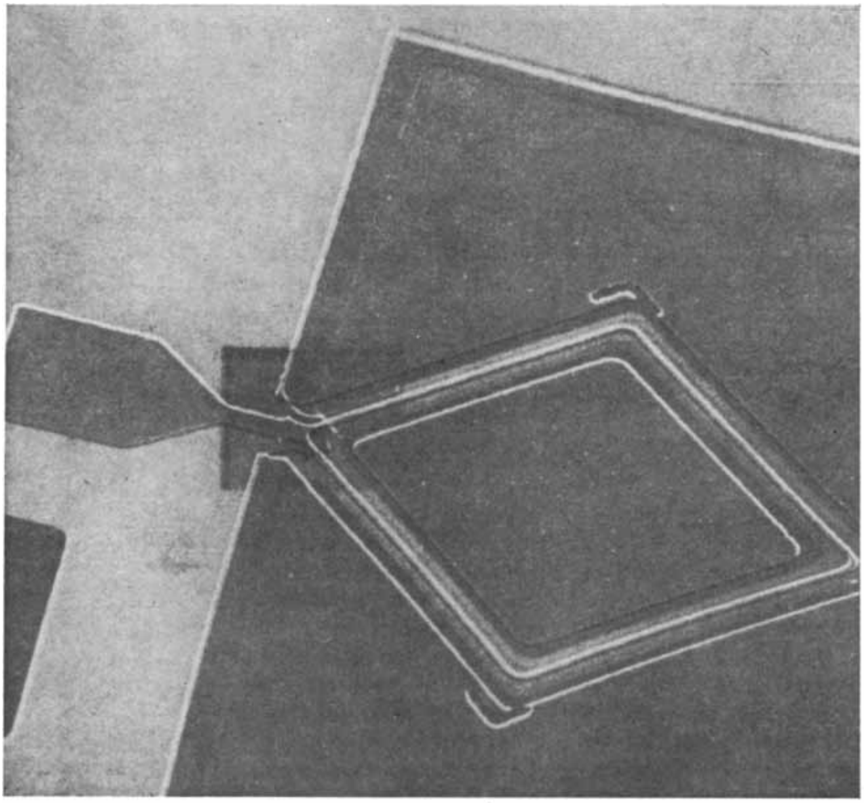
HIGH-PERFORMANCE MOSFET is made in these steps. Light admitted through a mask sensitizes a "photoresist" protecting a silicon oxide layer grown on a silicon wafer (*a*). Unprotected silicon oxide is etched away and phosphorus atoms are diffused into them to produce "source" and "drain" areas (*hatching, b*). A new area is exposed to light (*c*) and etched and the wafer is subjected to implantation by accelerated boron ions (*d*); the boron ions implanted in the central area (*fine hatching*) provide the required level of *p*-type doping. A new protective layer of silicon oxide is grown and a layer of aluminum is evaporated over the wafer (*e*) and exposed to light (*f*). After etching, phosphorus ions are implanted to extend source and drain regions to edge of the gate electrode (*g*). After being annealed at 550 degrees Celsius to remove radiation damage and activate phosphorus ions, device is complete.

to the observed luminescence and not something else.

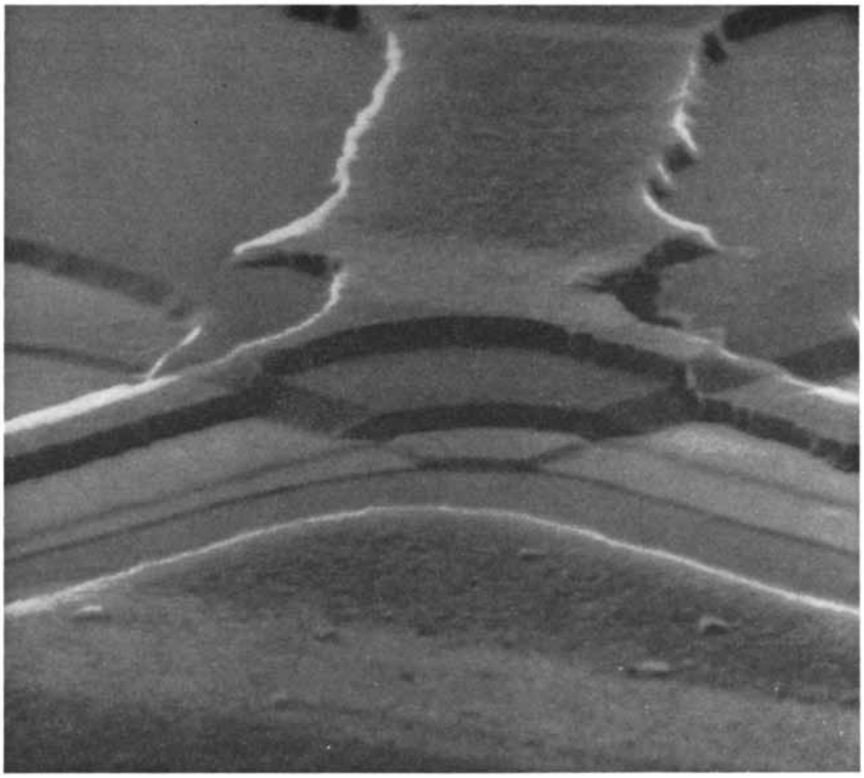
Another research application of ion implantation is in the study of diffusion of boron in silicon. The classic technique of measuring diffusion in solids is to add a small amount of a radioactive isotope of the dopant to a uniform concentration of nonradioactive dopant. The diffusion constant under known conditions can be obtained by measuring the motion of the radioactive tracer. That is done by removing thin layers of the solid and measuring the radioactivity. It happens, however, that there is no suitable radioactive isotope of boron. Hence other methods are needed to measure the element's diffusion constant in silicon. When the rare isotope boron 10 is irradiated with slow neutrons, its nucleus fissions into an energetic alpha particle (a helium-4 nucleus) and an energetic nucleus of lithium 7. The common isotope boron 11 and the isotopes of silicon do not undergo this reaction. To get the diffusion constant of boron, ion implantation is first used to uniformly dope a region of a silicon crystal with boron 11. Then tracer quantities of boron 10 are implanted. The diffusion constant can now be determined by irradiating the silicon with slow neutrons and measuring the energy of the alpha particles that emerge. Alpha particles that originate deep in the crystal lose more energy than those that originate near the surface; hence the depth distribution of the boron 10 can be determined from the energy distribution of the emitted alpha particles. Boron in silicon exhibits a diffusion behavior that is concentration-dependent, that is, the rate of diffusion increases with the increasing concentration of boron.

An interesting application of ion implantation has been in the production of amorphous (noncrystalline) semiconductors by ion bombardment. Amorphous materials produced by evaporation, sputtering and chemical deposition often have a surface that is too rough for certain studies. With the ion-bombardment method we start with a mirror-smooth crystalline surface and make the surface layer amorphous. The surface remains smooth, and the technique produces amorphous specimens of excellent quality.

Ion implantation has taken nearly two decades to reach its maturity. It is clearly a process at the frontier of semiconductor technology, but at the very least it is a valuable research tool that is helping us to increase our understanding of the solid state.



SCANNING ELECTRON MICROGRAPH shows a MOSFET of the kind whose manufacture is diagrammed on the opposite page. The diagrams are of cross sections, whereas the micrograph is an oblique view of the surface. The five-sided tab (*left center*) is the gate-electrode contact; the gate electrode is the bright line that runs along the center of the diamond-shaped pattern. The MOSFET is enlarged 520 diameters in this micrograph.



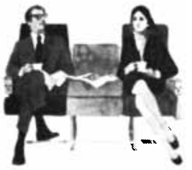
GATE ELECTRODE AND CONTACT are seen in sharp relief in this scanning electron micrograph, in which their junction is enlarged 6,400 diameters. The highest surface is a layer of aluminum; the lower steps are silicon oxide layers of various thicknesses.

TWA is what

TWA's Ambassador Service.*
What a long flight should be:
Available on over 250 flights a day from 23 cities throughout the U.S.A. and on all flights between the U.S.A. and international destinations.

Curbside Check In.
Check in your bags at the terminal curb.

Ground Ambassador.
Get any assistance you need at the departure lounge.



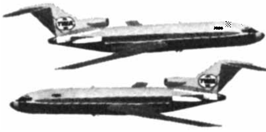
The Twin Seat in 707 Coach.
If nobody's sitting next to you, fold down the middle seat and sit two across.

Choice of Movies.†
Only TWA gives you a choice of movies on every movie flight across the U.S.

Choice of Meals.
We offer a choice of three main courses in coach and five in first class.

8 Channel Audio.
Only TWA offers it on all 727 flights of 90 minutes or more. There is a special channel for executives and headsets are free.

Choice of Meals.
A full-size meal or just a snack on most meal flights.



Out and Back Scheduling.
To most destinations you can go and come back the same day on TWA.

TWA's L-1011.
What an airplane should be:

The most modern plane in the world, featuring every comfort and convenience TWA ever put on an airplane.

Coach.
You're never more than one seat from the aisle. If the seat next to you is empty, you can turn that seat-back into a table, or you can raise the armrests and turn your seat into a couch. Only TWA coach has this flexibility.

First Class.
It feels like a big lounge, because there are club seats in the center area that swivel. Now you can chat and dine the way you would in a restaurant.

TWA's Ambassador Express.*
What a short flight should be:
Available on over 250 flights a day from 18 cities throughout the U.S.A.

Gate-Area Ticketing.
If you're late, buy your ticket right at the gate.

Carry-on Luggage Compartments on all 727s.
Take a bulky package, even a full-size bag, right on the plane with you.

The Twin Seat in Coach.
If nobody's sitting next to you, fold down the middle seat and sit two across.

†Entertainment at nominal cost in coach. Movies by Inflight Motion Pictures, Inc. *Service marks owned exclusively by TWA.

travel should be.

What a vacation in the U.S.A. should be:

Fares.
This summer no other non-stop scheduled airline will have lower fares for you and your family over comparable routes and for comparable service between TWA cities.

Inflight Service.
In addition to Ambassador Service, there are special games and movies to keep your children occupied.

On the Ground.
After you stay with friends or relatives, TWA can arrange for you to stay in a hotel, sightsee and rent a car for just a little more than the cost of your flight.

Getaway* Adventures.
Low airfare. First-rate tours. Hotels. Sightseeing. And more combinations of vacation places than most other airlines.

What a vacation in Europe should be: (On a Tour.)

Low airfare. Complete sightseeing if you wish. Each ground arrangement backed by a warranty.

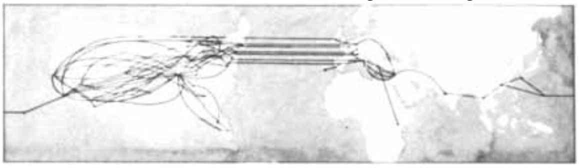
TWA's TOUR WARRANTY:
If any tour element or service is not delivered as described in TWA's Getaway Adventure brochure. TWA certifies that an appropriate refund will be made of that portion of the tour price covering the undelivered tour element as outlined in the brochure.

Travel by air, by driving your own rented car, or by luxurious "Meteor" motor coach.

(On Your Own.)
We can help you with hotel reservations and car rentals. And you can take along a Getaway Guidebook with valuable discount coupons.

TWA's Getaway Card.
You can charge your trip on your Getaway Card and take up to two years to pay.

What airline should you fly?



TWA. And here's one more reason why: Only TWA can take you on a short flight within the U.S., a long flight across the U.S., a vacation to Europe or a trip around the world.

Which means you can fly from a TWA city to Europe or around the world without the nuisance of having to switch airlines and terminals along the way.

What the future should be:

We'll be doing a lot more to make travel better than ever.

To make vacations here and abroad more available, better organized and quality controlled.

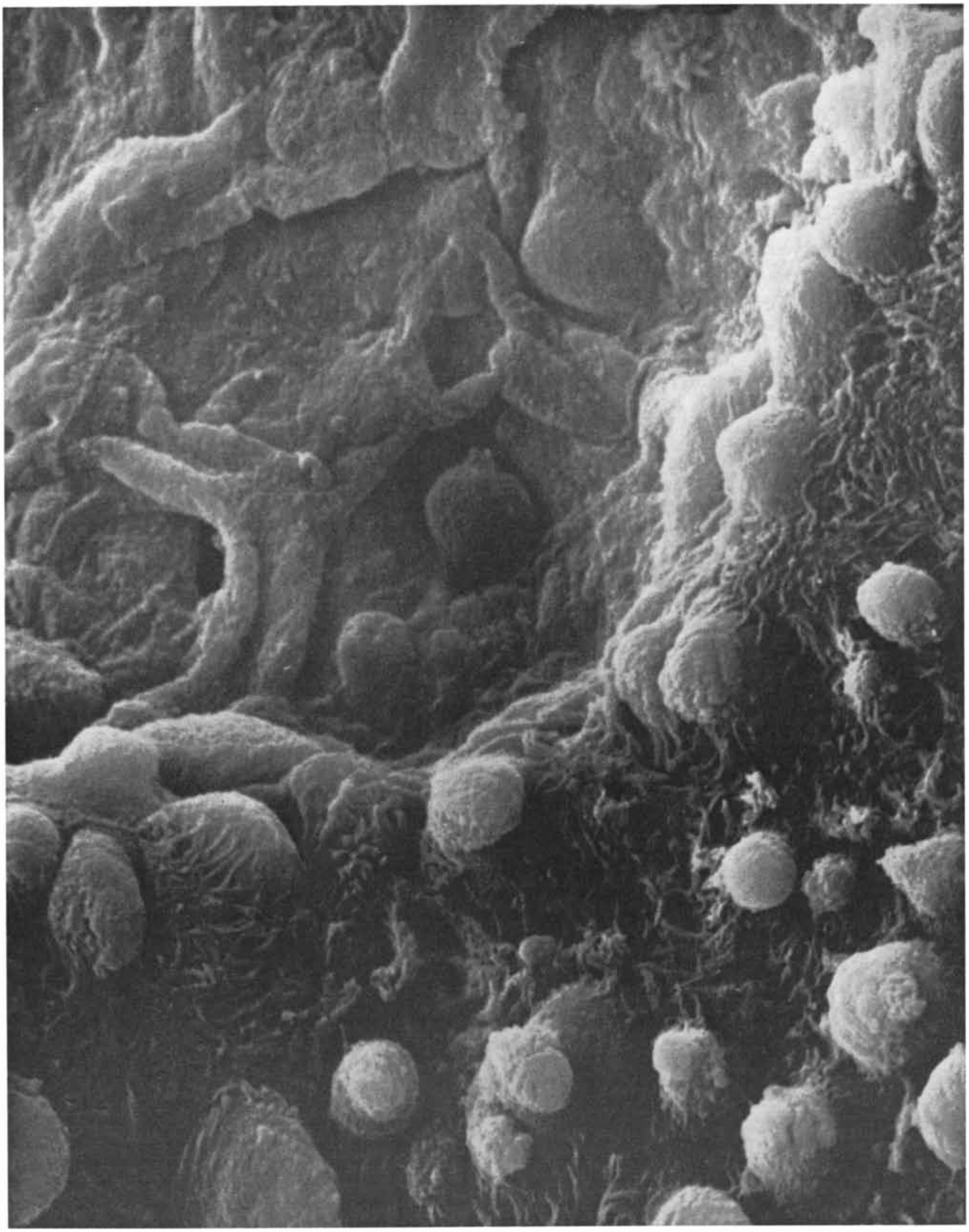
To develop innovative new fares so that more people can travel for less.

TWA

So call your Travel Agent or TWA.

TWA is what travel should be.

That isn't just a slogan. It's a commitment.



CLIFFLIKE BOUNDARY marks the margin between the ciliated cells of a bronchiole, with their hairlike appendages (*right and bottom*), and the unciliated cells that line a cup-shaped alveolus, one of the terminal air spaces of the lung. The tissue, magnified 3,250 times in this scanning electron micrograph, is not from the lung of an infant but from a laboratory animal with analogous lung structures; the animal is an adult rabbit. The micrograph was made by one of the authors (Nai-San Wang). The hole to the left of center is one of many "pores of Kohn" that open onto adjacent alveoli; the boulder-like bodies among the cilia are Clara cells.

The Lung of the Newborn Infant

Certain cells essential to lung function mature only shortly before birth. A way has been found to accelerate the development of these cells in order to prevent respiratory distress in premature infants

by Mary Ellen Avery, Nai-San Wang and H. William Taeusch, Jr.

The major challenge faced by an infant during birth is switching from one oxygen source to another. In the U.S. until recently some 25,000 infants died each year because they could not convert from oxygen obtained through the placenta to oxygen obtained through the lungs. Most of these infants were born before the normal 40 weeks of fetal growth and maturation. When such respiratory distress occurs, it has a predictable time of onset: at the moment of birth or within minutes thereafter.

Every infant's first breath requires great effort. For infants with immature lungs the effort continues. Immature lungs cannot retain air, and as the first breath is exhaled the lung's air spaces empty completely and collapse. As a result the infant finds each inhalation as difficult and demanding as the first. After a time plasma may begin to leak out of the lung tissue and coat the air spaces. This glassy pink coating gives respiratory distress in the newborn the name hyaline membrane disease.

In the past some infants with hyaline membrane disease recovered spontaneously. Many became exhausted and died a few hours after birth. Today intensive care (including assistance in respiration) can often carry such infants through the critical first five or six days of life, and once this period has passed most of them recover completely. In an effort to reduce the losses still further interested investigators have in recent years directed their attention to the maturation of the fetal lung, and in particular to possible ways of accelerating the maturation process.

Consider the mature lung. The total surface area of its 300 million terminal air spaces is between 70 and 80 square meters, or nearly the size of a badminton court. Throughout this enormous field of tissue the blood courses constantly, separated from the atmosphere by a membrane only a thousandth of a millimeter thick. Waste carbon dioxide diffuses through the membrane into the atmosphere as atmospheric oxygen diffuses in the opposite direction.

In tracing the lung's air pathways outward from their many inner terminals we can begin with a single alveolus (Latin for "little hollow"). This cup-shaped depression is from two- to three-tenths of a millimeter in diameter; its dimensions at any particular moment depend partly on the degree of inflation. Each of the 300 million alveoli is surrounded by a network of capillaries and by supporting connective tissue. It is estimated that in the network surrounding each alveolus there are 2,000 segments of capillary.

Clusters of alveoli are connected by numerous ducts to the next-larger airways of the lung: the richly branched minor passageways known as bronchioles. Bronchioles by the hundreds open onto each of the small bronchi, and small bronchi by the score join the two main bronchi, one in each lung, that lead to the trachea [*see illustration on next page*]. All these minor and major conduits are surrounded by supporting connective tissue and by a dense network of lymphatic channels; these channels are more abundant in the lungs than in any other organ of the body.

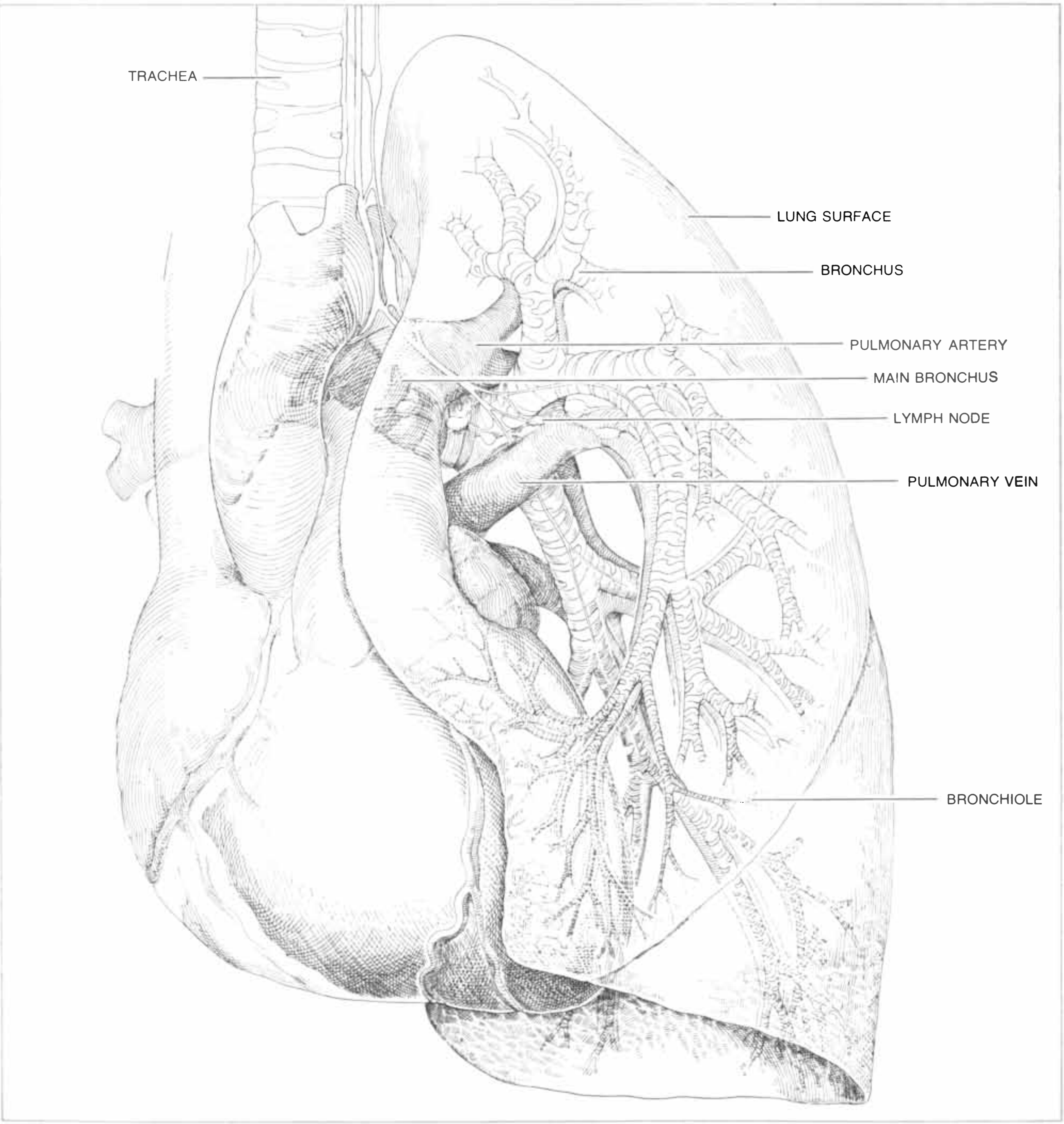
From the most distant alveolus to the trachea the airways are lined with a protective layer of epithelial cells. The cells differ in type according to their location. For example, the epithelium of the trachea and the main bronchi is composed of a layered array of four kinds of cells: tall columnar cells with hairlike cilia and a prismatic cross section, mucus-secreting goblet cells, unciliated brush cells and short basal cells [*see illustration on page 78*]. Farther along, in the small bronchi, the epithelium is reduced to a single unstratified layer of cells, and where the bronchioles branch from the small bronchi the cells that form the layer are more cubic than columnar. Where the ducts that lead to clusters of alveoli branch from the bronchioles the boundary is marked by the presence of unciliated cells that were once thought to have a secretory function; these cells are called Clara cells after M. Clara, who first described them in 1937.

Three more kinds of cells, each distinctly different, are found within the alveolus. For convenience two of the three are now simply referred to as Type I and Type II cells. The epithelium consisting of Type I cells is so fine-spun that the cells' existence was not even acknowledged by some workers until electron microscopy provided incontrovertible proof of their reality. Type II cells are readily visible under the light microscope; over the years they have been variously named niche cells, dust cells, corner cells, great alveolar cells and granular pneumocytes. The third kind of cell in the lining of the alveolus is the macrophage, or scavenger cell. Macrophages are dislodged by the millions every day and are swept away by the blanket of mucus that is continuously moved along all the airways of the lung by the cilia of the epithelial cells.

The lungs of a newborn infant are necessarily much smaller than an adult's. The reduction in lung size, however, is by no means directly proportional to the reduction in body size. A normal newborn infant is only a twenty-fifth the size of an adult. Its trachea, however, is between one-third and one-fourth the size of an adult's, being some six millimeters in diameter as compared to 20 millimeters in the adult. The smaller airways, from two to four millimeters in diameter in the adult, are about half that size in the newborn. The alveoli too are half-size; moreover, the newborn infant has only a tenth as many alveoli as the adult, or a total of about 30 million. (The growing child requires about eight years to develop the other 270 million alveoli.)

If the newborn infant's lungs were a twenty-fifth the size of an adult's, the infant would soon be dead. Airways that small would offer a fatal resistance to the passage of air. A comparison of the airway resistance in the lungs of a newborn infant with that in the lungs of an adult shows how rapidly resistance increases as airway size is reduced. Although the infant's airways are only some 50 percent smaller than the adult's, the resistance is five to six times greater. Resistance is customarily expressed in terms of the pressure exerted by a column of water of variable height. The force required to overcome airway resistance in an infant at rest is equivalent to the pressure of a water column 30 centimeters high.

This value scarcely compares with the resistance offered by the liquid-filled airways of the lung at the time of a newborn infant's first breath. A liquid, of course, is much more viscous than air. Moreover, the surface tension at the interface between air and a liquid increases as the diameter of the airway



ADULT HUMAN LUNG is shown schematically, starting with the trachea (*top left*) and continuing with the many ramifications, upward and downward, of the left main bronchus. The total area of the lung's 300 million cup-shaped alveoli is some 70 square meters.

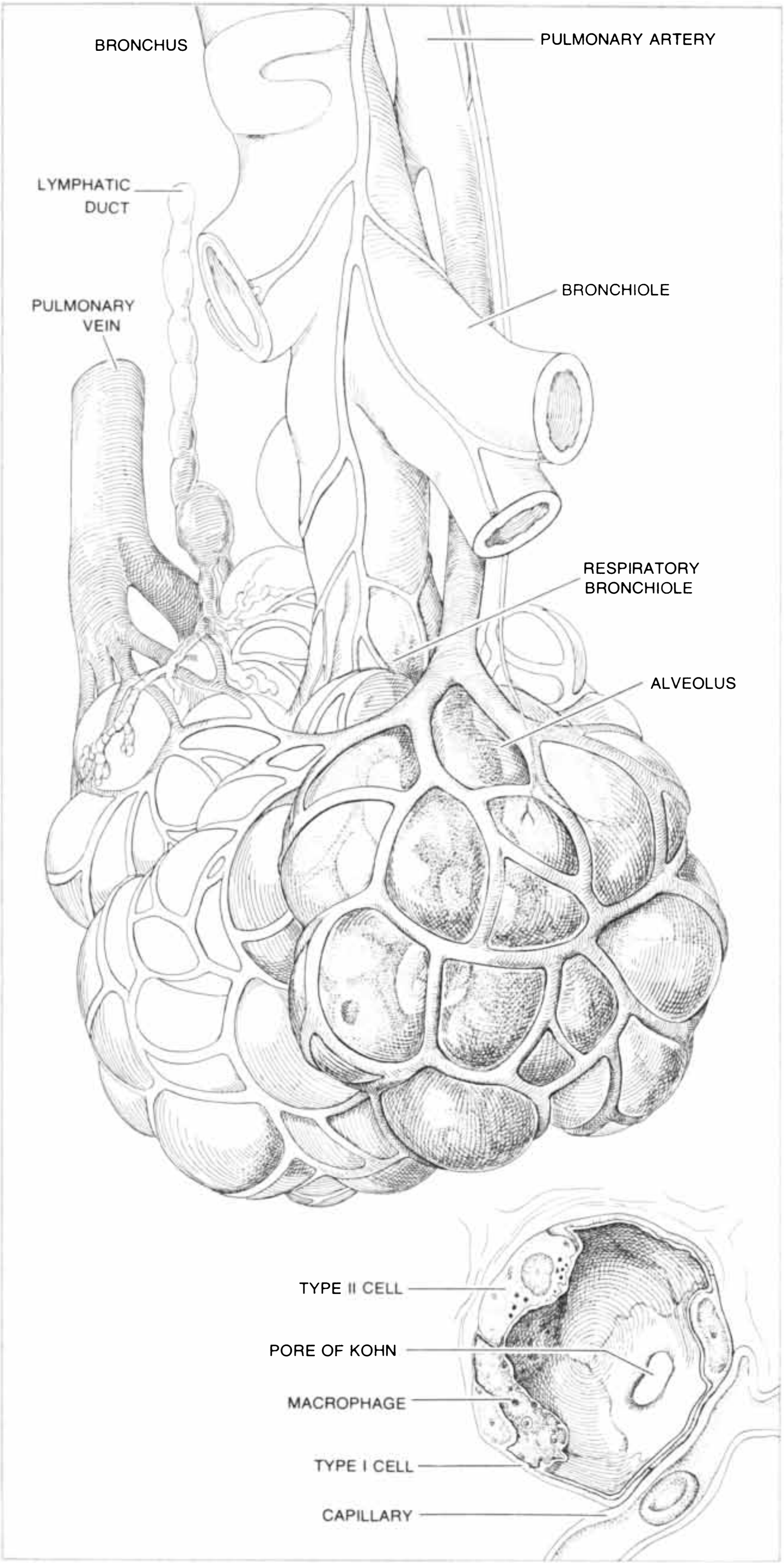
diminishes. As a result the pressure required to achieve the first breath may be the equivalent of 80 centimeters of water. Even under the best of circumstances the infant's first few breaths call for pressures 10 to 15 times greater than the pressure needed for breathing once the lungs are aerated. With these values in mind it is easy to understand why the infant with hyaline membrane disease, whose successive breaths are often as difficult as the first, is quickly in danger of exhaustion.

How are the lungs of the fetus just before birth readied to receive the infant's first breath? At this time the potential air spaces of the fetal lung are distended to about a third of their total capacity by a liquid termed the alveolar fluid. Secreted by the fetal lung, the fluid resembles blood plasma except that it is more acid and contains only a fraction as much protein.

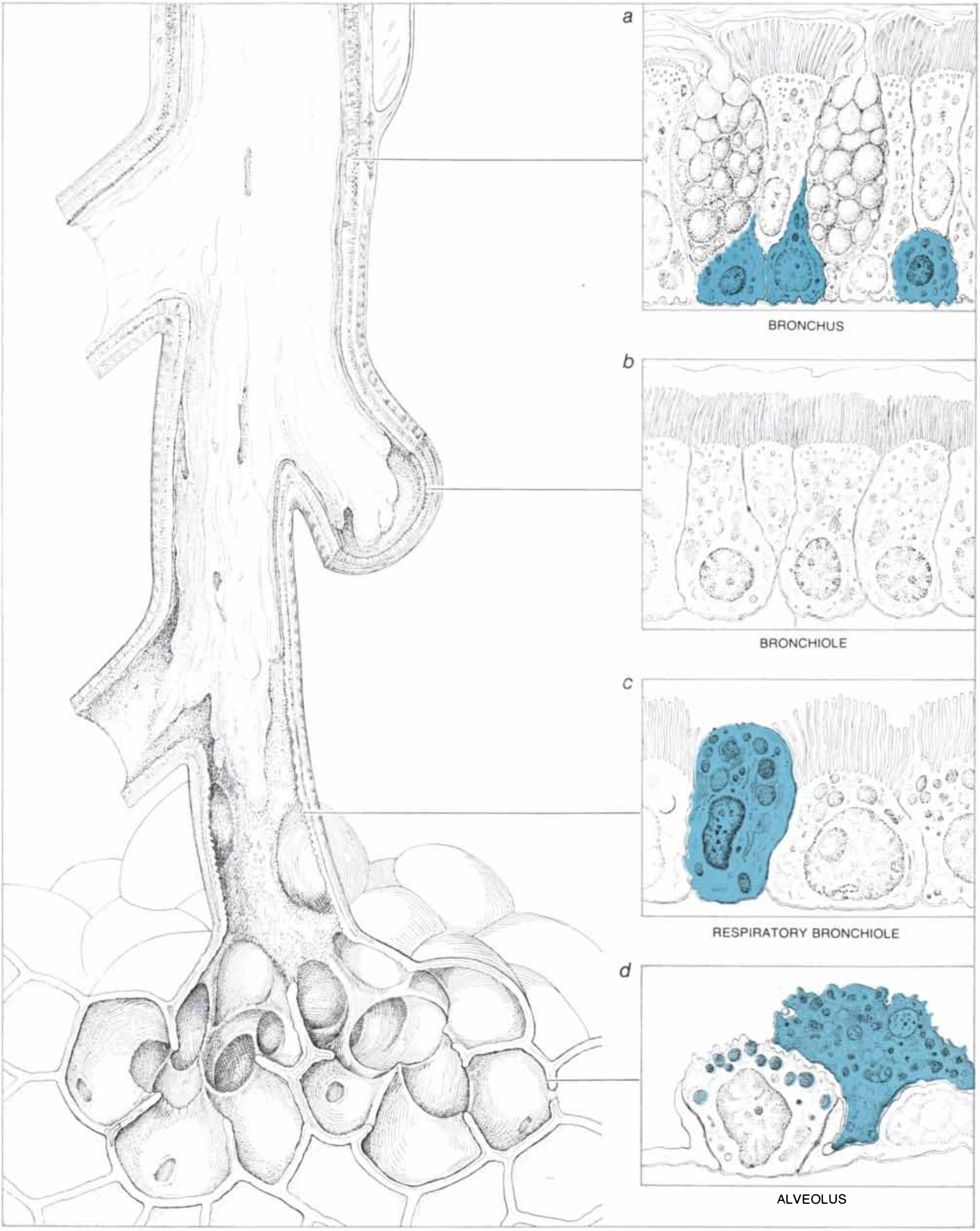
The alveolar fluid must leave the lung at birth. A number of studies have suggested that it departs by as many as three separate routes. Of an estimated total of 50 milliliters of fluid, perhaps 15 are passively emptied from the upper air spaces and pharynx of the newborn infant during its expulsion from the birth canal. At the same time as much as 25 additional milliliters of fluid (and most of the fluid's protein content) probably enters the infant's lymphatic system. Finally, with the introduction of air into the lungs and the accompanying increase in blood flow through the newly activated organ, a part of the alveolar fluid probably enters the bloodstream; at least the difference in osmotic pressure between the fluid and the blood favors movement in that direction.

There are two reasons why the flow of blood through the lungs increases immediately after birth. In the fetus a shunt known as the *ductus arteriosus* connects the right ventricle of the heart directly with the aorta. Studies of fetal lambs have shown that 80 percent of the blood from the right ventricle passes through this shunt and never enters the lung at all. The effectiveness of the shunt is augmented by the resistance to the flow of blood offered by the still undilated arterioles of the lung.

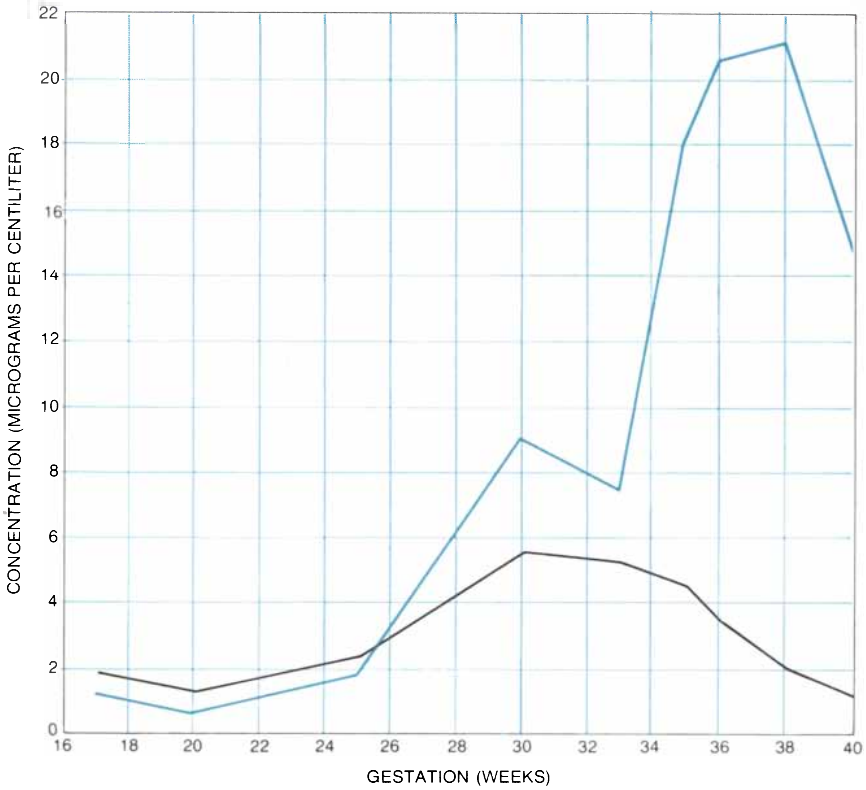
It would, of course, be of little use to the newborn infant to breathe in air if there were no flow of blood through the lungs. Indeed, an occasional infant whose lungs are normal in all respects but who for unknown reasons shows persistent low pulmonary blood flow simply dies of asphyxia. Such deaths are rare, because normally the *ductus arteriosus*



CLUSTERS OF ALVEOLI, forming alveolar sacs, are connected by ducts to the smallest air pathways of the lung: the respiratory bronchioles. Several sacs and the network of capillaries that surrounds them are illustrated here. Seen in cross section is a single alveolus with its lining of three distinctive kinds of cell. Cells labeled Type II secrete lipids that reduce surface tension during exhalation and keep the lung from deflating completely.



CELLS OF DIFFERENT KINDS comprise the lining of various parts of the lung. Both in the trachea and in the bronchi (*a*) the lining consists of a deep array. Short basal cells (*color*) are most prominent at the bottom. Rising between them are (*left to right*) a brush cell, which lacks surface cilia, a goblet cell, which secretes mucus, and a ciliated cell, columnar in form. More of these same cells continue the sequence. The lining of the smaller bronchi and the bronchioles (*b*) consists of only a single layer of cells; they are ciliated and remain columnar in form. At the boundary between bronchiole and alveolar duct (*c*) ciliated cells, columnar in form elsewhere, are shortened and more cube-shaped. Among them appear Clara cells (*color*) with an unknown function. Lining the individual alveolus (*d*) are scavenger cells (*color*), the attenuated surface cells known as Type I and the more prominent cells of Type II. Organelles (*light color*) found within Type II cells may be sites where surface-active lipids the cells secrete are stored.



TWO LIPIDS that comprise the pulmonary surfactant, lecithin (*color*) and sphingomyelin (*black*), are found in roughly equal quantities in human amniotic fluid until the 28th week of pregnancy. Normally the quantity of lecithin increases sharply about the 35th week. If amniotic fluid shows no increase, the fetal lung is not maturing normally and the child when born is in danger of developing hyaline membrane disease. Data for this graph were obtained by Louis Gluck of the University of California at San Diego School of Medicine.

and the arterioles of the lung react in an opposite and complementary manner soon after the infant's first breath. The increase in the amount of oxygen in the infant's internal environment causes the arterioles to dilate, so that resistance to the flow of blood into the lungs diminishes. At the same time the *ductus* reacts to the increased oxygen by constricting, thereby allowing all the blood formerly shunted away from the lung to enter the pulmonary circulation. In a newborn lamb the blood flow to the lungs increases from about 100 milliliters per minute before birth to some 500 milliliters soon thereafter; the increase is probably about the same in the newborn infant.

Pound for pound, a newborn infant consumes oxygen at nearly double the adult rate. The reason is that two organs with a high rate of metabolism—the liver and the brain—represent a disproportionately large percentage of the infant's total weight. The infant's need for more oxygen is met in part by its proportionately large lung surface, a tenth the size of an adult's. Whatever deficit may remain is overcome by the infant's high rate of respiration, which is two to three times more rapid than the adult's.

If at birth the immature lung collapses with each exhalation, how does the mature lung avoid similar collapse? Two investigators have supplied the answer to this question, which is related to surface tension. In 1955 Richard Pattle, who was then working with the Chemical Defense Experimental Establishment in Britain, pointed out that the alveoli of the lung are coated with a complex substance that made bubbles from the lung long-lived. Other workers later identified lipids and proteins that act as "surfactants" in the substance. This is to say that these components are "surface active," and that their activity affects surface tension. For example, when the surface of the lung area is enlarged during inhalation, the increase in surface tension contributes significantly to the elasticity of the lung tissue, enhancing the organ's elastic recoil. Conversely, when the surface area of the lung decreases during exhalation, the surfactant brings about a sharp decrease in surface tension; this stabilizes the airways of the lung, prevents their collapse and allows the organ to remain partially aerated. The fact that the surface tension can change when the surface area does, thereby preventing collapse of the lung during exhalation, was first appreciated by John A. Clements in 1957, when he was associated with the U.S. Army Chemical Center at Edgewood, Md. [see "Surface Tension in the Lungs," by John A. Clements; SCIENTIFIC AMERICAN, December, 1962].

Investigation has shown that the source of this crucial secretion is the Type II cells of the alveolus. The Type II cells contain in abundance not only the requisite organelles (mitochondria, endoplasmic reticulum and Golgi bodies) but also the enzymes necessary for the synthesis of fatty substances (lipids). Their capacity for such synthesis can be demonstrated in the laboratory by incubating lung tissue with glucose labeled with tritium, the radioactive isotope of hydrogen. The radioactive glucose is promptly incorporated within the lipids that accumulate in the cytoplasm of Type II cells. Studies with rats show that the half-life of these radioactive lipids is between 14 and 18 hours; this suggests that the supply of pulmonary surfactant is being renewed continuously.

Now that we know the presence of a natural surfactant accounts for the success of the newborn's first breath, we may ask when it is that an adequate supply of the surfactant becomes available to the maturing fetus. Most of our information about the stages of development of the fetal lung comes from the study of laboratory animals: lambs, rabbits, rats, mice and others. Autopsies of human fetuses, spontaneously aborted or prematurely born, suggest that the human sequence of development, if not the timetable, is approximately equivalent. In what follows the time intervals given are those that apply to human fetuses.

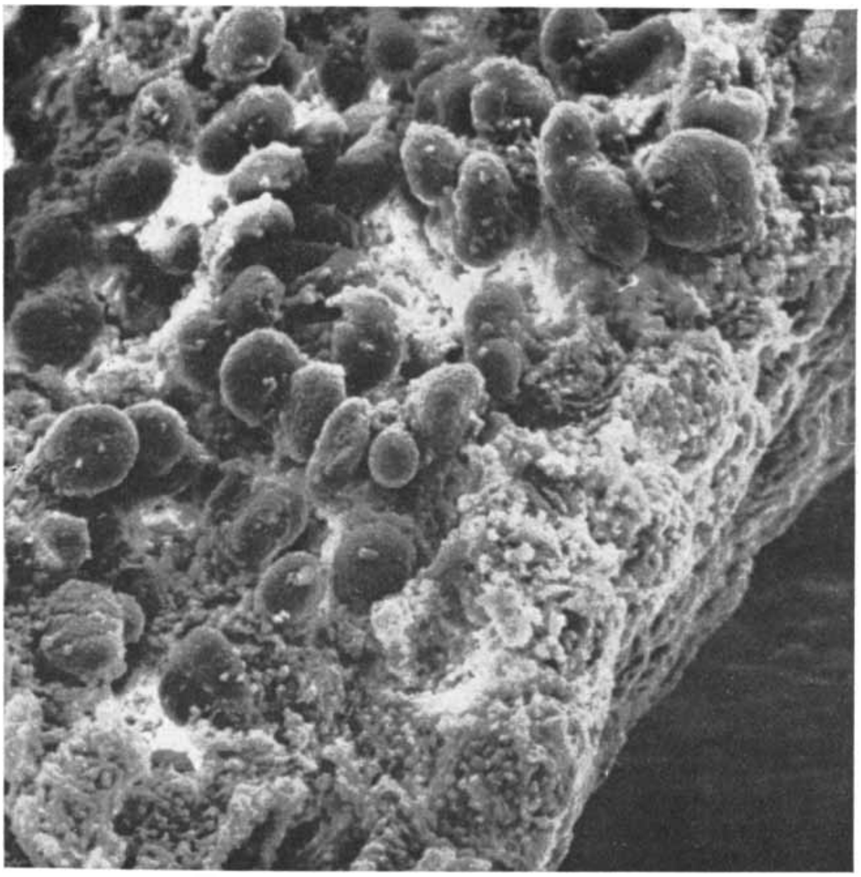
The lung first arises in the 24-day embryo as a pouch in the primitive gut. Thereafter the sequence of development follows a "wavelike" pattern. This pattern was first suggested by Sergei Sorokin of the Harvard Medical School and describes a process of maturation that proceeds generally outward from the tracheal region until it finally involves the most distant alveolar clusters.

There is evidence that some three weeks after the formation of the pouch in the gut the treachea has developed some cartilaginous support. Three weeks or so later mucus glands appear, and the surfaces of the growing airways have become lined with columnar epithelial cells. Within 15 weeks the columnar cells develop cilia; within 16 weeks goblet cells have appeared among them. At about the same time the process of ramification has brought into being a full array of bronchioles. The terminal air spaces that open off the bronchioles, however, do not yet show any true alveoli. The cell lining in the spaces consists only of somewhat shortened columnar cells, and the supporting connective tissue is far more prominent than the alveoli-to-be are.

By 24 weeks the cartilaginous supports of the bronchi are almost fully formed. In the terminal air spaces lacelike airways are developing as the formerly columnar epithelial cells assume a squat, cubic form. During the remaining 16 weeks of gestation the lacelike airways continue to develop, and the number of capillaries associated with the developing airways increases. Most significant, the epithelial cells continue to differentiate, becoming mature alveolar cells of Type I and Type II.

In lambs and rabbits Sorokin's wavelike pattern of development is readily apparent. The upper lobes of the lungs, that is, the lobes nearest the animal's head, tend to differentiate earlier than the lobes that are lower and farther to the rear. This may be explained by the fact that in lambs the lower lobes have as many as 25 divisions of bronchi as compared with a maximum of 13 divisions in the upper lobes. Hence the distance from trachea to alveolus is generally greater in the lower lobes.

Maturation of the lung does not halt at birth. One process that continues after birth is "budding," whereby more and more clusters of alveoli form along the terminal bronchioles. This process, which begins during the late fetal period, continues in infancy and childhood. Edward A. Boyden and D. H. Tompsett of the University of Washington Medical School, studying the casts of infant lungs, have found that the newly budded alveoli are larger in diameter than more mature ones. The reason is that the new alveolus is more saucer-shaped than cup-shaped; as the months pass and the number of elastic fibers that support the alveolus gradually increases, it assumes a mature cup-shaped configuration that decreases its diameter. Another event during the first year of infancy is the appearance of small holes between adjacent alveoli. They are known as the pores of Kohn, after Hans Kohn, who first described them in 1893. The pores allow air to pass between air spaces if an air passage becomes obstructed.



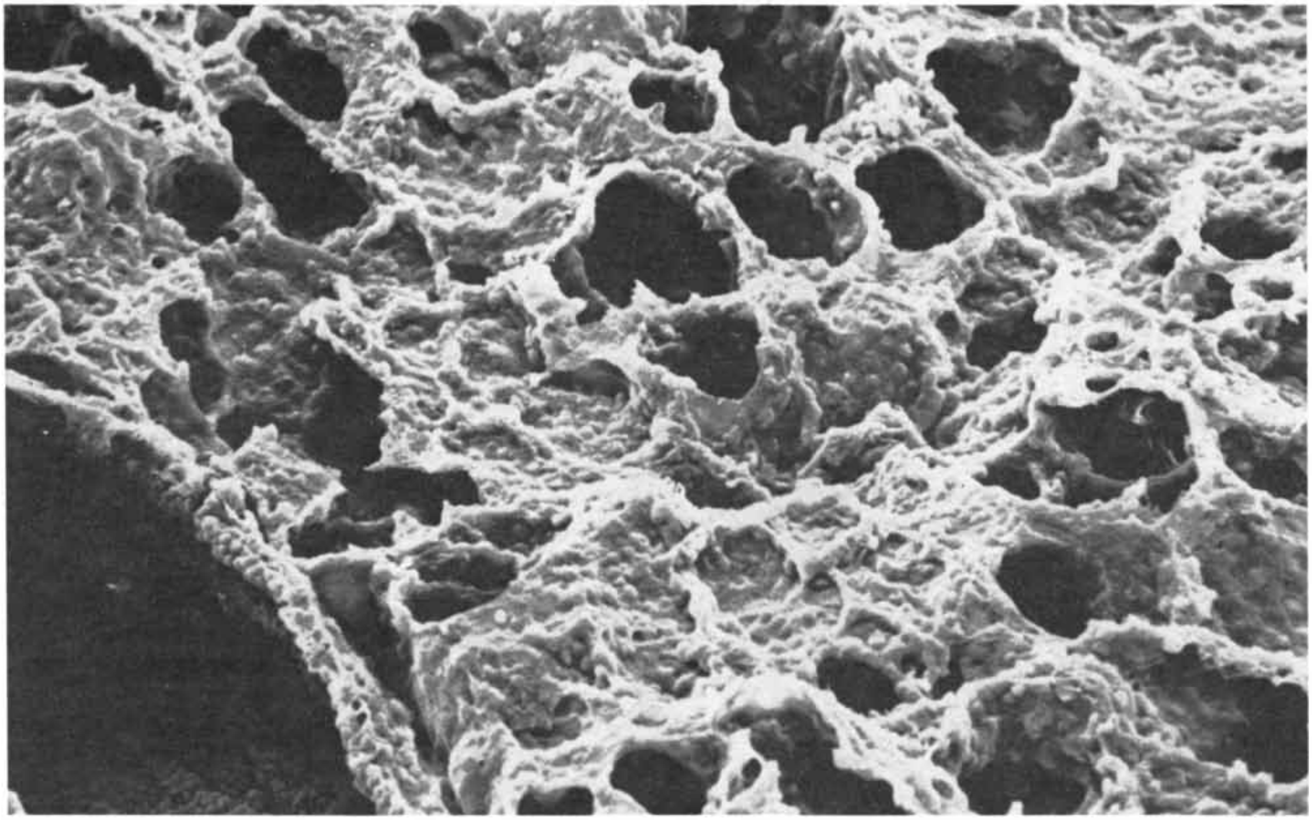
SMALL LUNG BUDS look like clusters of spheres in this scanning electron micrograph of the cut lung surface of a 20-day-old rabbit fetus, seen magnified 250 times. As the lung continues to mature over the next 11 days the buds will elongate into airways terminating in alveolar sacs. This micrograph and the ones that follow were also made by Wang.

We have seen that the performance of the newborn infant's lung depends on certain natural lipid surfactants that influence the surface tension of the air spaces and also that the surfactants are secreted by the Type II cells of the alveoli. It follows that the fetal lung will be deficient in lipids until after the time that Type II cells differentiate in the normal course of fetal maturation. At the Harvard School of Public Health in 1959 one of us (Avery) and a colleague, Jeremiah Mead, showed that the premature infant's fatally unstable lung was deficient in surfactants. At the Johns Hopkins Hospital in 1961 Avery and a second colleague, Sue Buckingham, examined the lungs of fetal mice at various stages of development and found that the first appearance of surfactant in the mouse lung coincided with the differentiation of Type II alveolar cells and with the presence within Type-II-cell cytoplasm of certain organelles that stained readily when exposed to osmic acid. Such organelles are called osmiophilic.

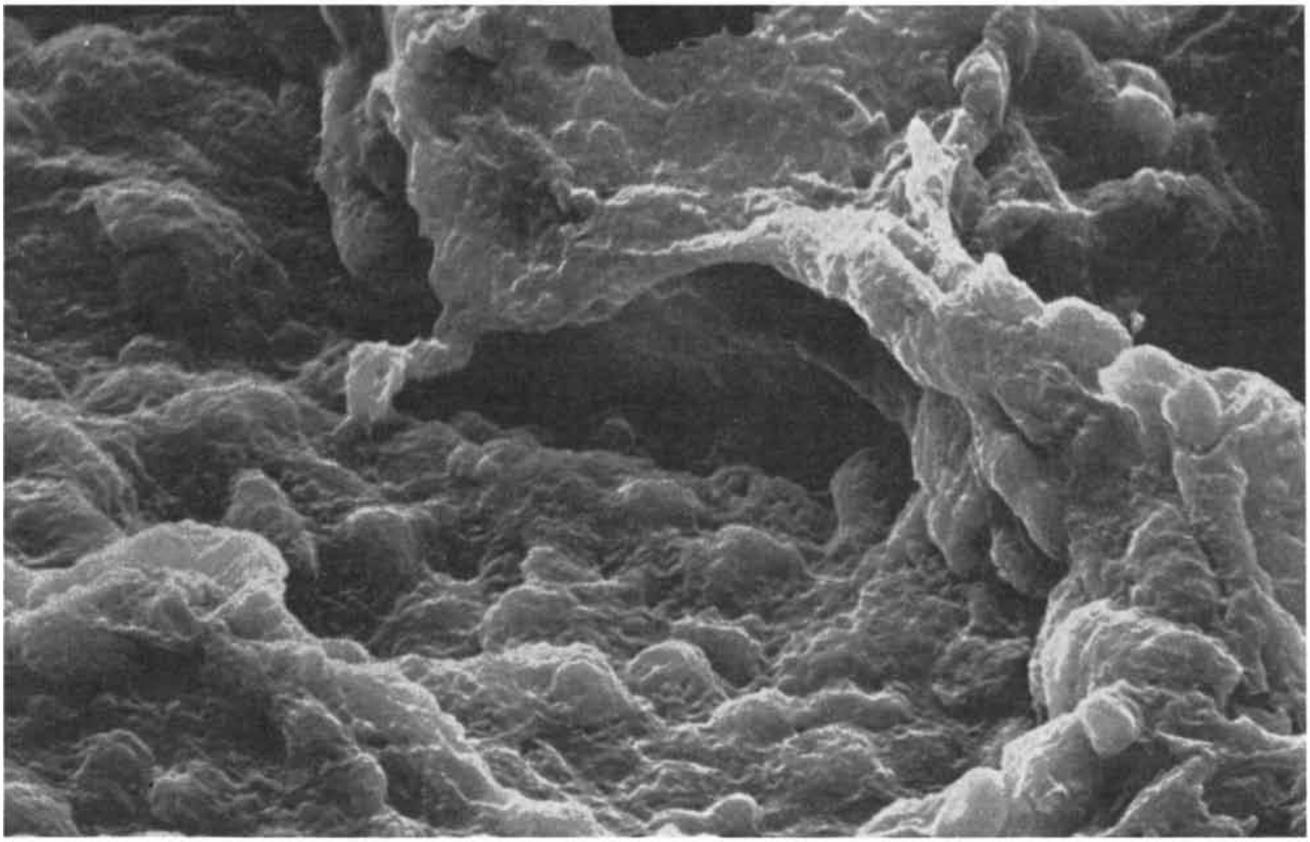
Later work has demonstrated that the osmiophilic bodies are associated with the surfactant. In the absence of surfactant in the alveolar tissue the bodies are either altered in appearance or few in number, suggesting that they are organelles that store the surfactant after the Type II cell synthesizes it.

In human fetuses osmiophilic bodies may appear in small numbers as early as the 20th week of gestation. The bodies are usually present after 26 weeks. Surfactant too is found in some fetal lungs as early as the 26th week and is evident in most lungs between the 28th and 32nd week. In contrast to this picture, examination of the lung tissue of premature infants who have died soon after birth of hyaline membrane disease shows that the cells of the terminal air spaces are still cubic in form and that the cell cytoplasm contains few, if any, osmiophilic bodies. When premature infants do not die until a few days after birth, autopsy shows that the cell cytoplasm of the terminal air spaces often includes osmiophilic bodies. The obvious inference is that development of these surfactant reservoirs is a prerequisite to spontaneous recovery from hyaline membrane disease.

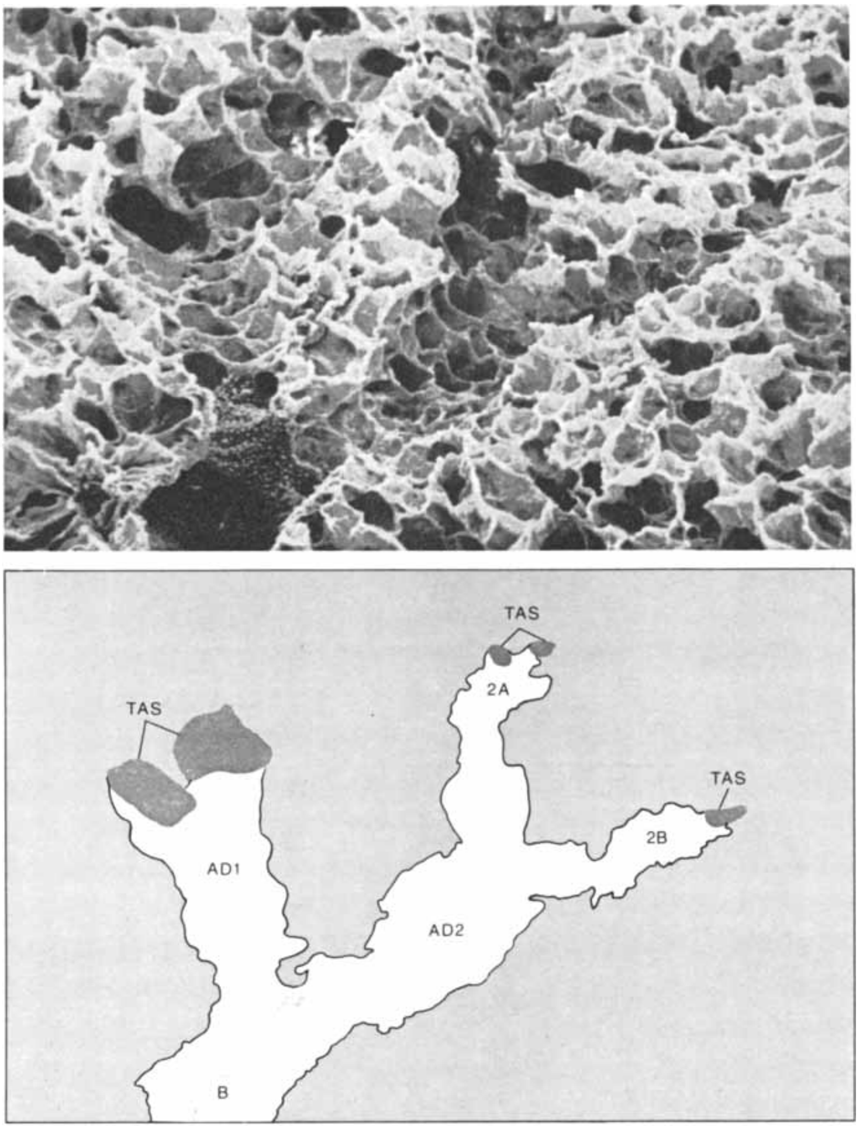
Since the lung arises from a pouch in



SAUCER-SHAPED DEPRESSIONS in the stretches of cratered tissue are potential alveoli; shown in this scanning electron micrograph is the cut surface of a rabbit lung the day before birth, magnified 450 times. One of the depressions is illustrated below.



SINGLE ALVEOLAR DEPRESSION in the tissue illustrated above appears magnified 2,000 times in this scanning electron micrograph. The cells of the alveolar surface are more cuboidal and more closely packed than in an adult (*see illustration on page 74*).



TISSUE SECTION from the lung of an adult rabbit shows a few of the final branchings of lung air pathways; the section is magnified 200 times in this scanning electron micrograph (*top*). At lower left a respiratory bronchiole (*B in diagram, bottom*) forks to left and to right into two alveolar ducts (*AD 1 and AD 2 in diagram*). The left fork ends in two terminal air spaces (*TAS*); the right fork forks again (*2A, 2B*) and ends in other air spaces.

the gut and is made effective by a secretion from a differentiated epithelial cell, Buckingham was led to wonder whether lung epithelium and gut epithelium have any responses in common. Florence Moog of Washington University had found that she could accelerate the appearance of one kind of gut-enzyme action in newborn mice by injecting the animals with adrenal steroids. When she gave nine-day-old suckling mice doses of steroid, the activity of the enzyme phosphatase in the gut increased, reaching a level normally seen in 20-day-old mice.

Before we could do more than speculate what might happen if lung epithelium were similarly treated, we received information from New Zealand, where G. C. Liggins, an obstetrician in Auckland, was using sheep to study the role of the fetal adrenal gland in the onset of labor. To induce premature delivery Liggins injected a number of fetal lambs with steroids. His method worked, and as an entirely unexpected by-product of the experiment he found that some of the premature lambs survived even though they were born as early as the 118th day of a normal 147-day gestation period.

Liggins' results surprised workers at Johns Hopkins, where some time earlier George Brumley and his colleagues had concluded that the amount of pulmonary surfactant present in fetal lambs before the 125th day of gestation is insufficient to stabilize the air spaces of the lung. Robert deLemos and his associates at Johns Hopkins immediately decided to test the possibility that Liggins' steroid injections had accelerated the maturation of the fetal lambs' lungs. Selecting ewes that were carrying twin lambs, they injected one twin in several pairs with a corticosteroid but left the other twin untreated. They then examined both twins from 16 hours to three days after the time of injection. The lungs of all the lambs that had received the injection were more mature than the lungs of their twins.

At McGill University one of us (Avery) and Robert V. Kotas, both of whom had formerly worked at Johns Hopkins, decided to repeat and expand on the earlier lamb studies, using rabbits as the experimental animals. It was just possible, we thought, that the phenomenon was confined to sheep. In our first experiment we gave one or more of the fetuses in the litters of several doe rabbits a single injection of steroid on the 24th day of gestation. (The normal gestation period is 31 days.) The litters were then examined on either the 26th or the 27th day. The lungs of the injected fetuses were approximately twice as mature as the lungs of their littermates; they respectively resembled the lungs of 28- and 29-day rabbits.

In a second experiment we found that injected fetal rabbits, when delivered prematurely, were able to keep their lungs inflated and hence survived on the average four times longer than their uninjected littermates.

In all the injected rabbits it was clear that the steroid had accelerated the differentiation of the lung epithelial cells. Microscopic examination revealed that the cells in the alveoli had changed from a cubic form to a more flattened one and that osmiophilic bodies were abundant in the cell cytoplasm. As we had anticipated, pulmonary surfactant also had appeared ahead of schedule. The injections had no effect, however, on the Clara cells in the bronchioles; precisely what role the Clara cells play in the lung remains uncertain.

As a part of our experiment we inflated a number of fetal lungs to test their capacity to be distended. We found that at any given distending pressure the lungs from cortisone-treated fetuses would hold more air per gram of lung tissue than those from untreated littermates. The finding provided further

UOP TEAMWORK IN TECHNOLOGY

"THE END OF THE BEGINNING."

Highlights of a dialog between John O. Logan, President of Universal Oil Products Company, and W. Robert Price, Jr., Corporate Vice President.

Mr. Price: By 1976 air pollution from new cars will be nearly eliminated by factory-installed controls using, in many cases, our UOP emission control catalysts. This won't help the *older* cars, though. A single 1968 car, on the average, will pollute the atmosphere as much as thirty-two 1976-model cars.

Mr. Logan: That's why there's a big potential for effective retrofit devices such as our Mini-Verter™. Millions of people live in severely-polluted areas that can't wait ten years, Bob, until new cars with emission controls replace the older cars.

Mr. Price: The Mini-Verter certainly has a lot of advantages for use in regional air quality programs. It's simple, easy to install and effective. Road tests show that on a 1972 automobile the Mini-Verter reduces harmful reactive emissions by about 75%. On older, pre-controls cars, the Mini-Verter Retrofit System will remove about *one ton* of pollution per year!

Mr. Logan: Tests also show the Mini-Verter catalyst remains active over 50,000 miles, with only minor maintenance at 25,000. But the best news is that use of the Mini-Verter makes it possible to tuneup a car for peak performance while reducing pollution at the same time.

Mr. Price: Fleet operators will be happy to learn the Mini-Verter System costs one-third to one-fifth as much as liquefied gas conversions and with none of the inconveniences. Emissions are even lower, too!

Mr. Logan: Perhaps the end of the battle for cleaner air is in sight.

Mr. Price: At least, John, with the help of our new Automotive Products Division, it may be the end of the *beginning* of the fight.

Write for free booklets on UOP's stake in the environment crisis, Ten UOP Plaza, Des Plaines, Ill. 60016.

evidence of the surfactant's effect on the elasticity of the lung.

Just how adrenal steroids trigger the differentiation of Type II alveolar cells is unclear. Moog's findings with respect to gut epithelium lead us to suspect that a steroid or some metabolic product of a steroid activates the enzyme systems that mediate the synthesis of the surfactant. In any event it is clear that accelerated maturation of the Type II cells enables an otherwise immature organism to live at a gestational age when survival is normally impossible.

How might this fact be put to use in terms of medical application? One answer may lie in the results of work by Leonard B. Strang and his colleagues at University College Hospital in London. Some time ago Strang's group demonstrated that the lung itself secretes the alveolar fluid that partially fills the fetal lung at the time of birth. Some of the fluid continuously overflows into the amniotic fluid bathing the fetus. The alveolar fluid contains trace amounts of various large molecules; among them are the two principal lipids—lecithin and sphingomyelin—that represent the bulk of the pulmonary surfactant.

Louis Gluck of the University of California at San Diego School of Medicine has shown that as gestation advances the relative abundance of the two lipids changes. The closer the lung is to maturity, the more abundant lecithin is in the alveolar fluid and the scarcer sphingomyelin is. For this reason Gluck has suggested sampling the amniotic fluid, where both lipids should be present because of the overflow from the lung late in pregnancy. By determining the relative abundance of the two lipids it should be possible to assess the degree of lung maturation in the fetus.

At the least such an assessment would help the physician to identify the fetus with a greater than average prospect of hyaline membrane disease. At the most an adverse finding might allow treatment of the fetus to hasten the differentiation of Type II cells in the immature lung. Already preliminary studies from Liggins' group in Auckland are promising: they find that treatment of the mother with steroid before the 32nd week of pregnancy can prevent hyaline membrane disease. The ability to assess lung maturation in combination with the ability to accelerate the maturation process before (and even after) birth should substantially advance the physician's capacity to improve the outlook for the prematurely born.

GIORDANO BRUNO

It is generally assumed that he was burned at the stake because he espoused the Copernican system. It appears, however, that he believed in the system more for mystical than scientific reasons

by Lawrence S. Lerner and Edward A. Gosselin

Early in the morning of February 17, 1600, after eight years of interrogation and solitary confinement, Giordano Bruno was burned at the stake for heresy and monastic apostasy on the Campo dei Fiori in Rome at the behest of the Roman Inquisition. Among other offenses, Bruno had espoused the Copernican view that the earth revolved around the sun and had maintained that most, if not all, heavenly bodies were populated with living beings just as the earth was. Thus, as is well known, ecclesiastical obscurantism reduced to ashes a leading forerunner of modern science. Within a generation the same forces would attempt to extinguish the last luminary of the Italian Renaissance: Galileo Galilei. When the Risorgimento returned light and learning to Rome almost three centuries later, the intellectuals of the world were not slow to acclaim Bruno as the first martyr of the new science and as a precursor of the persecuted Galileo.

It is not difficult to see how this view of Bruno arose from the constant warfare in the 18th and 19th centuries between the large majority of intellectuals, who were liberal and anticlerical in outlook, and the strongly reactionary Papacy. It is nonetheless superficial and quite misleading. Its generalizations obscure both Bruno's fascinating personality and his rich contribution to the philosophical foundations of the scientific revolution. The real reasons for his condemnation have likewise been obscured, as has the intricate connection between his persecution and the persecution of Galileo.

Giordano Bruno was born in 1548 in the small city of Nola, near Naples. At the age of 15 he joined the Dominican order and received a thoroughly Scholastic education. He began his career as a scholar in Naples, where he appears to have been in demand as a teacher of the then fashionable art of mnemonics. It was at this time that his unorthodox streak first came to the surface. In 1576 he was forced to flee Naples as a renegade monk under a cloud of accusations of heterodoxy, one of the charges being that he had read the prohibited works of Erasmus and had hidden them in his privy.

By this time Bruno's religious and philosophical views had already solidified. He had embraced the system of belief called Hermetic-Neoplatonism, and on it his entire career was based. We shall discuss the Neoplatonic view and its bearing on Bruno's physical science in due course. For the moment we need only remark that Bruno saw Neoplatonism as a basis for reconciling Catholics and Protestants in an era of violent religious warfare.

Bruno's flight took him to Rome, but he did not remain there long. He wandered through Italy, Switzerland and France, making his living by lecturing on and tutoring in the art of memory. He seems to have abandoned his monk's habit on several occasions and then resumed it. We have few details, but because of his tactlessness and belligerence in dispute he left less than cordial feelings behind him in most of the places where he stayed. In 1581 Bruno's wanderings brought him to Paris. There he attached himself to the Palace Academy, a group of literary men and philosophers in the court of Henri III.

Henri is often represented as a weak and indecisive king, wavering continually and dishonorably between the ultraorthodox Catholic faction led by the ambitious and powerful Guise family and the equally powerful Protestant faction led by the Bourbon-Navarre family. Henri's personal life, particularly the fact that he was a homosexual and a transvestite, has not enhanced his reputation. Yet Henri labored—not entirely without success—to avert the utter disintegration of a France polarized and torn by civil war. His purpose in founding the Palace Academy was to nurture a nondoctrinal, philosophical basis for the political reconciliation of the Catholic and Protestant factions. It was in this atmosphere that Bruno's message of religious reconciliation fell on receptive ears.

In 1583 Bruno received royal letters of recommendation to Michel de Castelnaud, Marquis de Mauvissière, who was Henri's ambassador at Queen Elizabeth's court. Bruno moved to London and remained attached to Mauvissière's retinue until 1585, when the ambassador was recalled to Paris as a result of political realignments at home. During this English sojourn Bruno wrote his most important works: six Italian dialogues. One of them, *The Supper of Ashes* (*La Cena de le ceneri*), is Bruno's major exposition on the Copernican theory. It is a strange and difficult dialogue, full of rich and denunciatory polemic. The view that Bruno was a scientific precursor of Galileo is based mainly on the apparent subject matter of this book.

After 1585 Bruno resumed his wanderings through Europe, lecturing and writing and almost always managing to remain a center of violent controversy wherever he went. He continued to develop his "true philosophy," which he hoped would be the basis for a reconciliation not only of Catholic and Protestant but also of man and God. Believing that the triumphal accession of the tolerant, newly converted Henri IV to the Catholic throne of France presaged the achievement of his mission, Bruno rashly returned to Italy in 1592 with the idea that he could eventually convert the Pope to his philosophy. His Venetian host betrayed him to the local Inquisition, which then turned him over to its Roman counterpart. Eight years later Bruno was led from his dungeon and, after scoffing at his executioners and averting his eyes from the proffered crucifix, was burned.

The Inquisition had been in no hurry to dispatch Bruno, since it did not regard him as an imminent threat. Indeed, his long imprisonment was in large measure a macabre colloquium, partly oral and partly written, in which each side tried to convince the other of the error of the other's views and the correctness of its own. It is interesting to note that the Inquisition gave scant attention to Bruno's advocacy of Copernicanism. When Bruno once raised the issue, perhaps hoping to sidetrack the Inquisitors, they dropped the subject immediately, passing instead to the question of Bruno's relations with Henri III, Henri IV and Queen Elizabeth. Moreover, Bruno's eventual execution seems to have been the Pope's part of a minor political *quid pro quo* with Spain, whose antipathy toward Bruno was far stronger and more immediate than that of the Papacy.

Bruno stated on several occasions that his mission to London in the 1580's was an official one. Although there is no incontrovertible evidence to support the claim, there is no doubt that Bruno's philosophical aims were entirely in line with the political goals of the French monarchy. In the interminable three-sided struggle for political ascendancy between Spain, France and England the Spanish had gained the upper hand. It was thus only natural for France to seek closer ties with England as a counterbalance. The French crown, however, was committed at least nominally to the Catholic side in the convulsions that shook Europe in the aftermath of the Protestant Reformation. As a result such a political commitment was a very touchy matter indeed. Ultra-Catholic Spain stood ready to back a take-over by the Guises in France, or by their kinswoman Mary Queen of Scots in England, at the first opportune moment.

The liberal Catholic intellectuals in France and their influential liberal Protestant counterparts among the English courtiers urgently sought to find out what bonds could be forged between them. Such an intellectual link could become a powerful counterbalance to the doctrinal objection to a French-English alliance. The English courtiers included such many-sided Renaissance men as Sir Philip Sidney, the poet-philosopher; John Dee, the astronomer-technologist-magician and royal astrologer; Thomas Digges, the Copernican savant; William Gilbert, the royal physician, mystic philosopher, pioneer in magnetism and one of the few men to be singled out by Galileo for scientific praise, and Sir Fulke Greville, the courtier-intellectual-poet. It was to this circle that Bruno gravitated in England, and it is with respect to this circle that we must seek Bruno's aims.

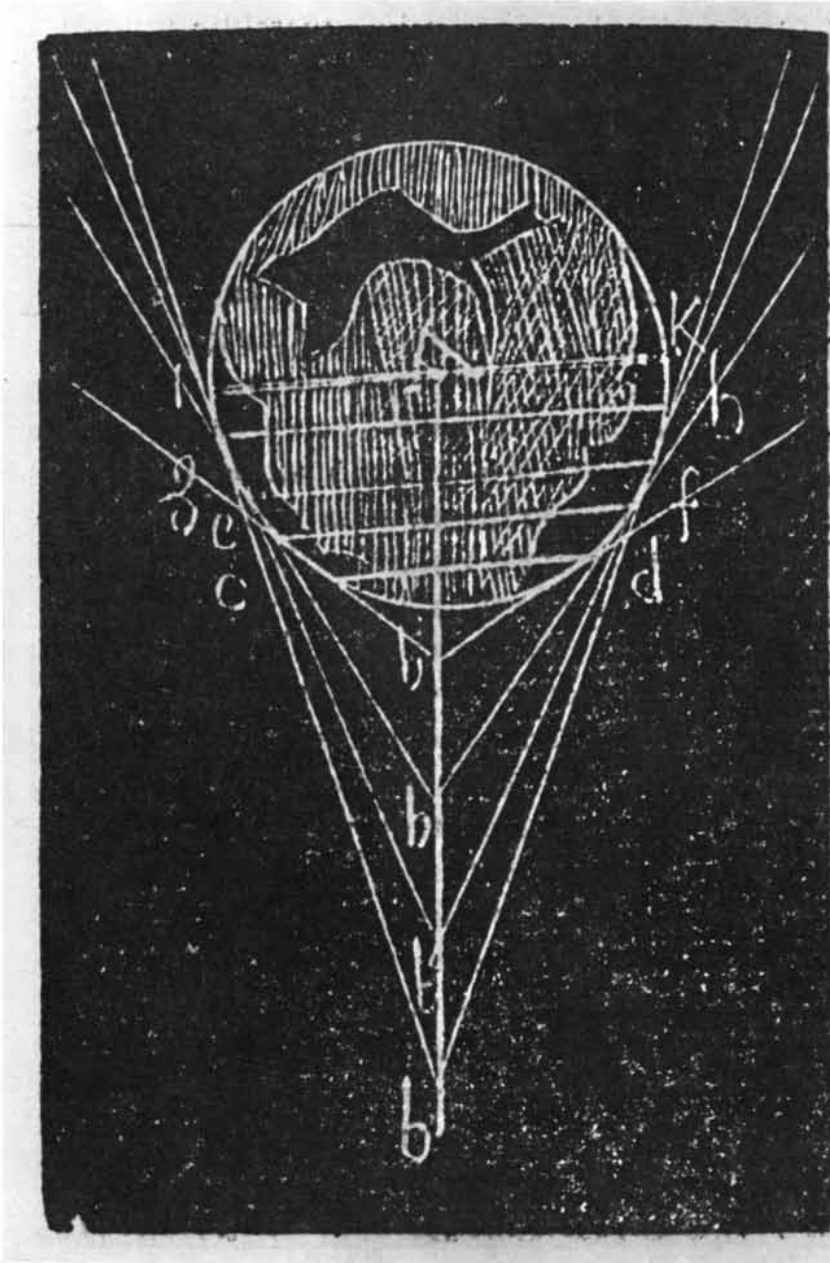
Like Galileo's two largest works, *The Supper of Ashes* is in the form of a dialogue in Italian. Most of the intellectuals at Elizabeth's court were fluent in that language; the Queen herself enjoyed showing off her proficiency. Knowledge of Italian was not otherwise widespread in England (or for that matter in France), and this suited Bruno's purpose. That purpose was nothing less than the establishment of a Christian commonwealth, guided by his philosophical principles

PTOLEMAIC SYSTEM OF THE UNIVERSE (*top half*) was compared with the Copernican system (*bottom half*) by Giordano Bruno in his dialogue *The Supper of Ashes*. Bruno used the diagram to demonstrate the ignorance of his anti-Copernican adversaries Nundinio and Torquato; in a real debate at Oxford in 1583, Bruno was defeated by the Ptolemaicists largely on the basis of his own evident ignorance of the Copernican system. For example, in the bottom half of the illustration he put the earth and the moon on a common epicycle about an imaginary point in orbit around the sun, rather than placing the earth in an orbit around the sun and the moon in an orbit around the earth. Symbols from the outside in represent the fixed stars, Saturn, Jupiter, Mars, the sun, Venus and Mercury. Original wood-block engravings are white on black; they are reproduced with permission of Harvard College Library.

under the benign hand of Queen Elizabeth. Her tolerant and ambiguous stand on theological matters, together with her intellectual leanings, made her in Bruno's view a likely candidate. At other times in his career Bruno seems to have favored Henri III and particularly Henri IV in the same role.

The dialogue revolves around the recounting of a stormy debate supposed to have taken place at Greville's house on Ash Wednesday in 1584. The dialogue within a dialogue involves "The Nolan" (that is, Giordano Bruno of Nola) in a defense of the Copernican theory against Torquato and Nundinio, two thinly disguised caricatures of pedantic, logic-chopping, hairsplitting, orthodox Protestant Oxford scholars.

In making their way to Greville's house Bruno and his friends encounter hilarious tribulations. To begin with no one arrives to pick them up at the appointed hour. They fall into the hands of feeble, insane boatmen whose boat leaks like a sieve. After disembarking they find themselves in an alley where they sink into a bottomless mudhole. Having struggled through the mire, they find themselves nearly where they started. They make their way through crowds of bestial, ill-natured townsfolk who go out of their way to push, trip and molest them. When they arrive at Greville's house, they are rudely treated by hangers-on until finally they enter the sanctum of the genteel Elizabethan courtiers. Bruno intends this passage to symbolize the decay into which society has fallen as a result of the total obscuration of the true religion and philosophy.



ILLUMINATION OF A SPHERE (presumably the earth) by a small luminous object is one of Bruno's thought experiments from *The Supper of Ashes*. The small luminous object, if placed a short distance above the nonluminous sphere, will illuminate a small part of the sphere. At successively more distant locations (labeled *b*) the source illuminates larger and larger areas of the sphere (labeled *cd*, *ef*, *gh* and *ik*). According to Bruno, when the source is removed to infinity, the entire sphere will be illuminated. Curiously, the points on the diagram are denoted by symbols different from those used in the accompanying text.

The debate finally begins. The nature of the debate and of Bruno's opponents is best demonstrated by Bruno's mocking prose in translation. Teofilo, Bruno's mouthpiece, describes Torquato as he rises to speak:

"With the same emphatic attitude in which Jove appears in Ovid's *Metamorphoses,* sitting in the midst of the council of the gods on the point of pronouncing that tremendous sentence against the profane Lycaon, . . . Torquato, having admired his own gold chain. . . , rose, withdrew his arms from the table, shook his back a little, puffed and sprinkled a little with his mouth, arranged the velvet biretta on his head, twirled his moustache, put his perfumed face in order, arched his brows, expanded his nostrils, settled himself with an oblique look. . . , pointed the first three fingers of his right hand and began to wag it back and forth, saying: *'Tunc ille philosophorum protoplastes?* [Are you then the foremost of philosophers?]' "

Such are Bruno's preposterous anti-Copernican opponents. Yet it is clear from Bruno's remarks that he regards Copernicus as a great mathematician. Nevertheless, according to Bruno, Copernicus failed to understand the true metaphysical meaning underlying his phenomenological calculations. Conveying the "true" meaning is of course the aim of *The Supper of Ashes.*

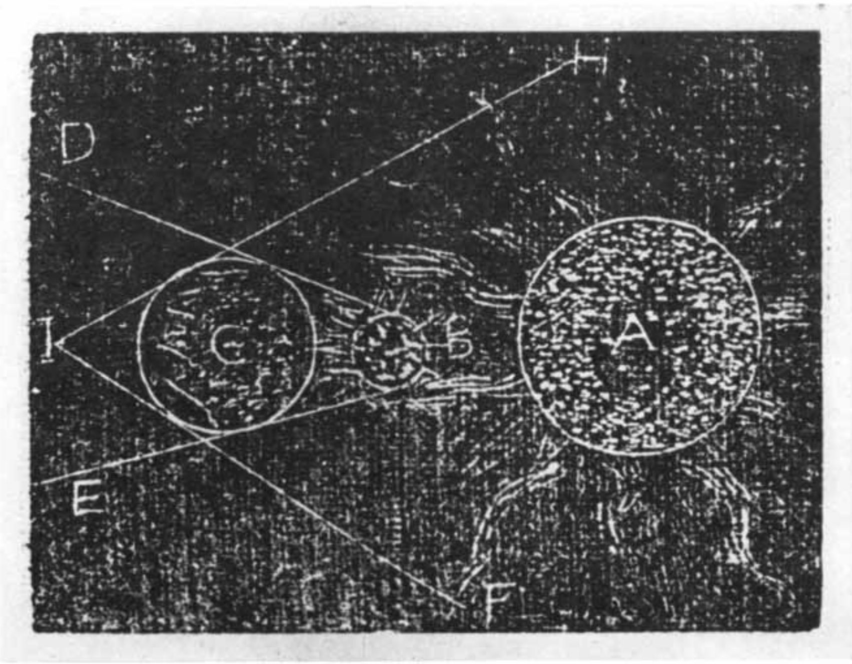
As we have mentioned, the dialogue purports to be a defense and enlargement of the Copernican view that the earth moves around the sun, rather than the opposite view, on which the Ptolemaic system was based. When one begins to inspect the argument carefully, however, one is surprised to find that

Copernicanism is simply not what the dialogue is about. Taken literally, Bruno's scientific arguments, two of which we shall discuss further, are a congeries of almost pure nonsense and *non sequitur*. The fact that we find them so from the vantage point of modern scientific knowledge is not to the point. What is important is that Bruno's contemporaries would also have perceived them as nonsensical, had they chosen to interpret the arguments literally. If they had, his enemies might well have ridiculed him, but they hardly would have bothered to burn him at the stake.

Since it is evident that the scientific arguments could not have been intended to be taken literally, what was the sense in which they were intended? The answer to that question makes it possible to understand the connection between Bruno and the scientific revolution. It is also vital for an understanding of the precise way in which Bruno's ghost hung as an incubus over Galileo in his vain attempt three decades later to convince the Inquisition that his views were neither heretical nor politically dangerous.

Bruno appears to have deliberately chosen the Copernican theory as a vehicle for introducing his own philosophical views, partly because there was considerable current interest in it and partly because it could be adapted to his own theories. In addition Bruno and his contemporaries had reason to perceive Copernicus as a reviver of magical Pythagoreanism, which they regarded as one of the sources of the Hermetic-Neoplatonic tradition. First, the Pythagoreans had held a heliocentric view; second, Copernicus had alluded in his *De revolutionibus orbium coelestium* to Hermes Trismegistus, the reputed founder of the Hermetic-Neoplatonic tradition.

Bruno's "scientific" arguments range widely. He discusses astronomical, physical, geological, archaeological and geometrical questions. They are entirely nonscientific in the modern sense and must be evaluated in terms of the place they occupy in late-Renaissance thought. One aspect of his approach is revealed in his treatment of optical thought experiments. Bruno discusses the objection to the Copernican theory that as the distance between the earth and Venus varies, the apparent size of Venus should change. He replies that, although the size of a nonluminous or opaque body diminishes with distance, this is emphatically not true of a luminous body, because the visibility of the luminous body depends on its brightness. As a result its distance cannot be deduced. Bruno scoffs at his opponents, demonstrating the absurdity of their arguments by means of a fallacious argument of his own. He maintains that if the distance from the observer to a luminous object (such as a candle) were repeatedly halved, the size of the object would increase to "whatever size you can think of." Obviously this would be absurd. Bruno ignores the fact that this argument applies to opaque objects as well as to luminous ones.



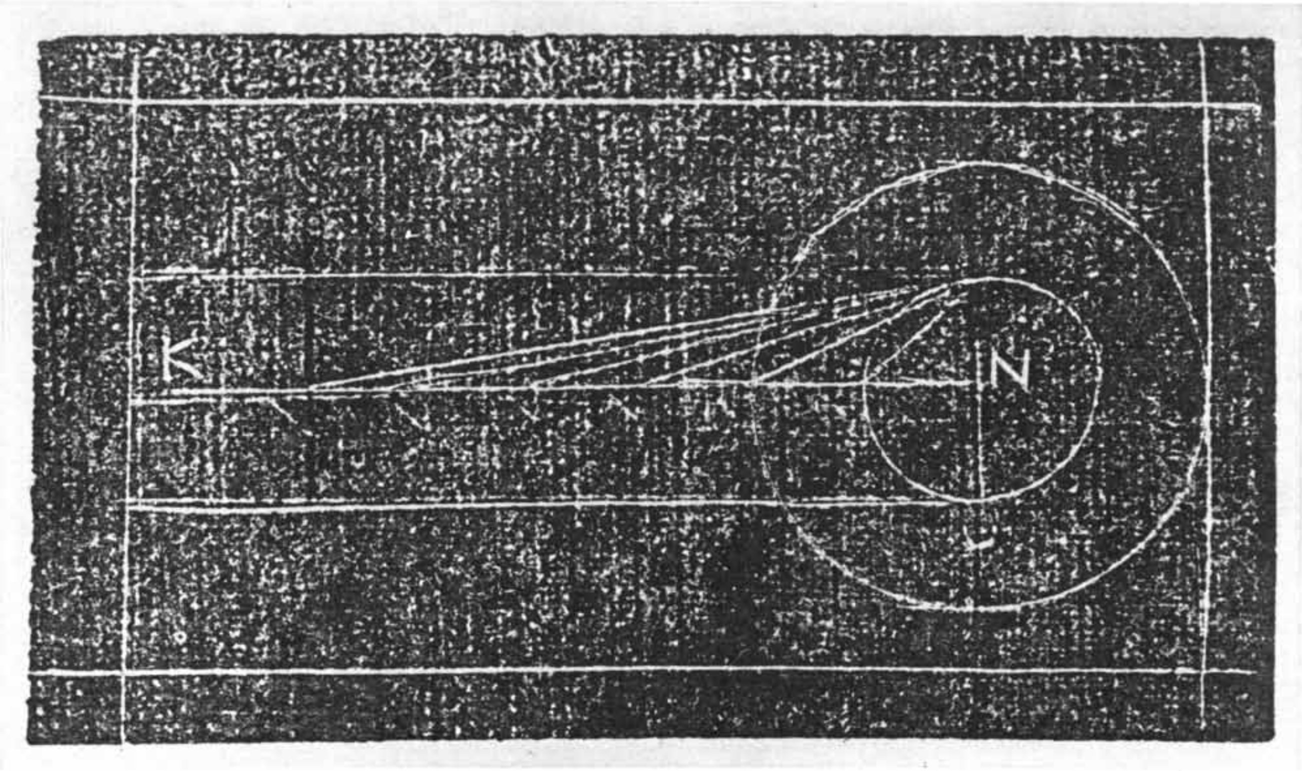
SUN IS LARGER THAN THE EARTH according to one of Bruno's proofs. If the sun (*B*) were smaller than the earth (*C*), then the shadow cone (*DE*) cast by the earth would diverge. If, on the contrary, the sun (*A*) is larger than the earth (*C*), the shadow cone (*HIF*) will converge. Bruno reasoned: We know that the earth's shadow never eclipses Mercury, therefore the sun must be bigger than the earth. He overlooked the fact that Mercury lies *between* the sun and the earth in both the Copernican and the Ptolemaic systems. If he had used Mars instead of Mercury, his argument would have been correct, at least in principle.

Strangely enough, Bruno pursues the fallacious argument to a correct conclusion: To a distant observer on another "star" the earth would look like a star. The size of the dark continents would apparently shrink as the distance increased, whereas the shiny oceans would not. Ultimately the earth would appear entirely starlike. Conversely, Bruno concludes that many of the stars are essentially the same as the earth, only their great distance from us obscures their admixture of land and water.

Bruno generalizes the argument in a curious way. A small luminous body a short distance above a larger nonluminous sphere will illuminate a small part of the sphere [*see illustration on opposite page*]. As the distance between the two bodies increases, a greater part of the larger body will be illuminated until an entire hemisphere is illuminated. Now, says Bruno, increase the distance still further. More than a hemisphere will be illuminated, and when the distance becomes infinite, the entire nonluminous body will be illuminated!

Bruno has a further generalization. Consider two small luminous bodies between which lies a larger opaque body, so that the light from either luminous body cannot reach the other. If the distance between the bodies is increased without limit, the opaque body will shrink to nothing, whereas the two luminous bodies will remain the same size. At a sufficient distance the opaque body will no longer impede the passage of light between the luminous bodies.

In modern terms these pseudogeometric demonstrations have nothing at all to do with the Copernican theory. In Bruno's terms they turn the Copernican theory into one grand metaphor (or, as Bruno would put it, an emblem) of his own philosophy. Once we realize that Bruno was using the Copernican theory purely as a philosophical metaphor, his mathematical and physical blunders fade into insignificance. If more proof is needed of his total unconcern for his own lack



AN OPAQUE BODY between two smaller luminous bodies will not impede the passage of light from one to the other if the two luminous bodies are far enough apart. As one of the sources is moved to successively more distant points from the opaque body (*smaller circle*), the angle at the source becomes more acute. At a sufficient distance *K* the angle is in Bruno's words "so acute that it will . . . appear . . . as a line. . . . So it follows necessarily that two more luminous bodies . . . will not be impeded from seeing each other."

of understanding of the essential details of the heliocentric view, we can refer to his ignominious defeat in a debate on the Copernican hypothesis held at Oxford in 1583. In appealing to *De revolutionibus* at one point in the debate, Bruno appears to have misinterpreted a diagram. More remarkable, his defense of Copernicus was based not so much on *De revolutionibus* as on a work on magic by the Florentine Neoplatonist Marsilio Ficino. The incident is transformed in *The Supper of Ashes;* there Bruno is made to appear as the victor, although his misinterpretation is preserved. He puts the earth and the moon on the same epicycle, instead of having the moon's epicycle revolve around the earth. Still more convincing evidence of Bruno's lack of interest in astronomical details is a flat statement by Bruno himself in the dialogue. When the argument gets down to details, the Nolan says impatiently: "I do not care much about Copernicus, and I care little whether you or others understand him." Elsewhere he states that he cares not at all for such things as optics.

Bruno's real interest in heliocentrism was that it implies that the earth is not the only center of the universe. The implication allowed him to put forth his own views that the universe is infinite in extent, and that it contains an infinite number of worlds, each of which can be considered as much the center of the universe as any other. Bruno's belief in a centerless or multicentered universe was derived from the 15th-century Platonist Nicholas of Cusa, who had spoken philosophically of a universe of indeterminate dimensions whose "center is everywhere and whose circumference is nowhere." Indeed, this slogan embraced the Hermetic (and hence the Brunian) definition of man. According to Bruno, the universe exhibits the same infinitude physically as the human mind does intellectually. The universe is therefore a fitting creation of an infinite God who is All in All. It is also a fitting object of contemplation for the infinite receptacle that is man's mind. The mystical element in all of this is perfectly obvious. According to Aristotelian-Neoplatonic psychology, the mind becomes what it contemplates. Thus man-microcosm becomes universe-macrocosm and as a result is brought closer to the Creator.

In the light of these arguments a picture of Bruno emerges that is quite different from the one with which we began. He was truly a Renaissance man and in no sense a modern one. He reflected the holistic approach of the Renaissance in that he believed all knowledge is interesting because all knowledge is interrelated. The aim of, and the key to, this interrelation is understanding the place of man in the universe, that is, the relationship between man and God. Knowledge is useful insofar as it elucidates that essential relationship. Like all knowledge, physical knowledge is seen as part of a vast metaphor. The details are of little importance as long as they do not disturb the metaphor.

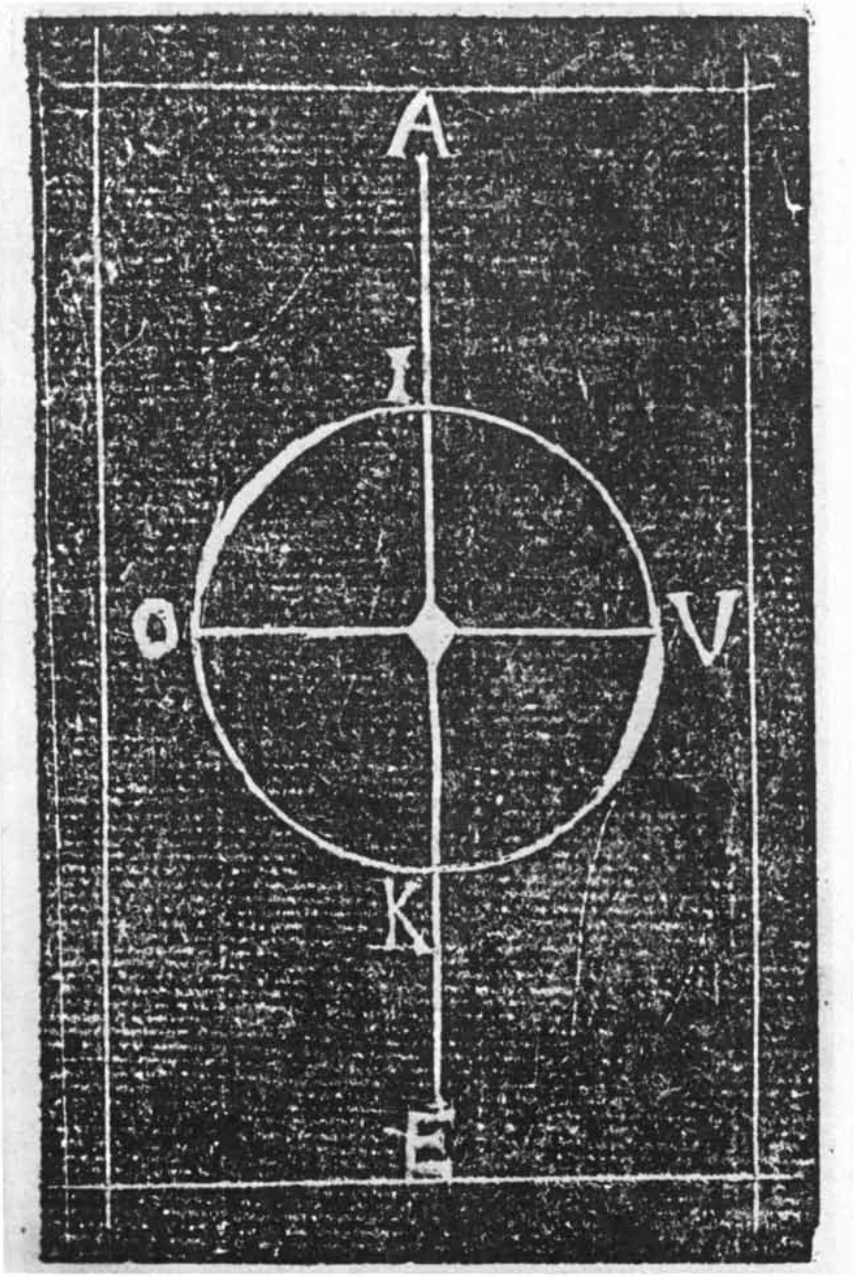
Although Bruno eventually does describe some central aspects of Copernicus' theory, his description of the earth's motions differs significantly from that of Copernicus. This discrepancy is attributable in part to Bruno's limited understanding of astronomical details and in part to his desire to improve the fit between Copernicus' astronomy and his own philosophy. Bruno's real interest lies in constructing an image that is large in scope, convincing and aesthetically pleasing that leads to metaphysical enlightenment. It is in these nonmodern terms that Bruno invoked the Copernican theory.

We are now in a position to consider the way in which Bruno's science and philosophy were expressed in his program for practical action. Bruno's aim, as we have noted, was the reconciliation of man and God. The 16th century was a time of religious splintering. What was worse, Christians everywhere were killing one another in the name of God and seemed destined to continue doing so indefinitely. The theological view that emphasized the evil side of man's nature appeared to be amply justified by experience. For all that the new Protestant religion contended that it harked back to original Christian principles, it seemed just as impotent as Catholicism to produce a universal public morality. Neither religion appeared capable of effectively inhibiting corrupt and cynical behavior on the part of either the individual or the public. Under both religions the parallel state of affairs in the world of learning was the proliferation of arid, nit-picking scholarship.

Bruno's Hermetic-Neoplatonism proposed a way out. He believed, as did many other Renaissance thinkers, that man had once lived in a golden age that was both physically and morally superior to his own times. Bruno and the other Hermetic-Neoplatonists used a series of works called the *Corpus Hermeticum* as the basis of their philosophical and historical perspective. The *Corpus Hermeticum* was attributed to Hermes Trismegistus (Hermes the Thrice Great), an Egyptian magician-philosopher-theologian believed to have been an approximate contemporary of Moses. (The works actually originated with the Gnostics of the third century, a point demonstrated conclusively in 1614. It is of no moment, however, insofar as the influence of the *Corpus Hermeticum* on the Hermetic-Neoplatonists is concerned.)

The "true religion" revealed to Hermes Trismegistus was believed to have been passed down by the Greek Pythagorean and Platonic philosophers on the one hand and by the Judeo-Christian theologians on the other. (This opinion ultimately led Bruno to conclude that Christ was not divine but was a magician.) In the process the truth had been corrupted and diluted, accounting for the degraded state in which 16th-century man found himself. It was in a revival of the true religion that reconciliation of Catholics and Protestants could be sought and society reformed.

In contrast to the orthodox Christian view that man had fallen from a state of grace through the original sin, the Hermetists believed man had descended voluntarily from the nonmaterial world of the Divine Mind to earth and continued to partake of the divine nature that had been his before the descent. There was indeed a channel for continual communion between man in his divine aspect and God. This channel could be deepened and widened by acquiring a kind of recondite knowledge we would call occult, which relied on certain affinities between man and the cosmos that Nicholas of Cusa had discussed. As a result of the curious juxtaposition of belief



A BALL THROWN UPWARD **illustrates the four Copernican motions of the earth, according to Bruno. The upward motion of the ball from *A* to *B* (*not shown*) corresponds to the annual revolution of the earth around the sun. The earth's daily rotation on its axis corresponds to the spin of the ball around its center, resulting in the continual interchange of the points represented by *I* and *K*. In addition points *1, 2, 3* and *4* (*not shown*) on the circumference turn into the points *5, 6, 7* and *8* (*also not shown*). Finally, the ball proceeds obliquely, so that *I* and *K* do not always pass through the same line and points *O* and *V* exchange positions. The last two ambiguously described movements are meant to correspond to Copernicus' third motion (an *ad hoc* one due to Copernicus' lack of understanding of inertia) and his fourth motion (precession). Bruno's eccentricity in failing to label points referred to in his text is common and is probably meant to have mystical significance.**

in magic and in reason as paths to knowledge of God, many serious scholars, including Dee, Gilbert, Ficino and Paracelsus, took part simultaneously (and with no sense of incongruity) in magical and scientific investigations.

It was by realizing that there is an intimate connection between man and God that man was to attain both material and spiritual redemption. Moreover, the mystical connection provided a means for manipulating the universe. If God could use the connection in a "downward" fashion to perform supernatural deeds as well as natural ones, man could do the same by working "upward" through the performance of magic. From this standpoint the boundary between magic and scientific investigation, which in modern times is clear, becomes indistinct. The Neoplatonists saw no inconsistency between the two activities; indeed, they were regarded as being complementary and mutually reinforcing.

Let us now reexamine Bruno's "optical" arguments in the light of his philosophy. It is nonsensical to maintain that a light source sufficiently far removed can illuminate an entire sphere. Bruno, however, is talking about divine light rather than physical light. Surely the laws of physical optics do not restrict the flow of this spiritual light from God to man, wherever the latter may be, figuratively as well as physically. The statement conceals a rich multiple image that is worth considering in detail, both for its poetic value and for the light (in the Brunian sense!) it sheds on how the Renaissance man integrated what we would call science into the totality of his thinking.

There is good reason to believe that Bruno is implicitly referring first of all to arguments put forward by Lactantius and Augustine, the early church fathers. Either the earth is flat or, if it is spherical, there can be no people living at the antipodes (that is, on the other side). If there were, their up would be our down and their summer our winter. Moreover, the descent of Christ at the Second Coming could not be seen by all. Since the ancients could not travel to the antipodes, people there could not be children of Adam, tainted by the original sin and capable of redemption. By extension such people would not be subject to a 16th-century pope. Knowledge gained through the great explorations had of course exploded this theory. Bruno is saying figuratively that through the new knowledge he is imparting, the metaphysical counterpart of those explorations can be accomplished. Divine illumination can reconcile the seemingly opposite poles of Catholicism and Protestantism.

The argument of the opaque body separating the two luminous bodies is similar in intent. The outstanding doctrinal difference between Catholics and Protestants was (and still is) a dispute over the precise nature of the communion Host. The Catholic view is that the physical existence of the bread and wine is obliterated and miraculously transformed into the physical substance of the body and blood of Christ. Protestant views vary, but whatever their precise significance, the bread and wine retain their own physical existence. What Bruno is saying is that a mere opaque physical substance such as the communion Host, whether or not it is transubstantiated, cannot (or ought not to) stand in the way of the reconciliation and unity of souls, Catholic and Protestant, through the Divine Mind. In this image the souls are represented by the luminous bodies and the light is divine illumination, transmitted by each to the other. Arguments similar to Bruno's had already been advanced in the mid-14th century by Nicholas of Autrecourt. They too had been applied to the doctrine of transubstantiation in an unorthodox manner.

The strange title of Bruno's dialogue, which appears entirely irrelevant if the subject matter is indeed the Copernican theory, becomes entirely appropriate in terms of Bruno's arcane message. *The*

"SHIP EXPERIMENT" was another of Bruno's thought experiments. A ship drifts smoothly with the current of a river. A man at the masthead and a man on the bank drop stones from the same point at the same time. Where will the stones strike the deck of the ship? Bruno concludes, correctly, that the stone dropped from the masthead will strike at the foot of the mast whereas the other stone will strike farther astern. There is a remarkable set of discrepancies between Bruno's text and his woodcut. The text seems to refer to a geometrical diagram, citing points *A, B, C* and so forth, none of which appear in the picture. This is not mere carelessness but can be interpreted in the iconographic traditions of the 16th century. Neither of the two experimenters described in the text is visible, nor are the stones that they drop. The ship is shown in a stern view, and this would be the worst place from which to observe the difference in the paths of the two stones. The ship, rather than being on a smooth river, is shown in a turbulent sea, representing the political and religious conflicts of Bruno's age. Wind god represents their underlying spiritual causes. Flames issuing from the yardarms represent the star deities Castor and Pollux, the Twins, who stand for spiritual calm. Bruno introduced such symbols not merely to illustrate the text but to propel the reader beyond it in a quasimystical way that his contemporaries would have understood.

Supper of Ashes is about the eucharistic, or communion, feast. Indeed, the Eucharist is called *cena* in Italian, in contrast to the more explicit "Lord's Supper" in English. The symbolic meaning of "ashes" comes from an ancient myth quoted by Bruno: "Alexander of Aphrodisias says that Mount Olympus displays, through the ashes of the sacrifices, the condition of a very high mountain and of the air lying above the extremes and limbs of the earth." According to the myth, a man who had offered a sacrifice on the summit of Olympus wrote some words in the cold ashes before departing. The next year he returned, only to find the ashes completely undisturbed and the words just as legible as when he had written them. This picture is of course appropriate to the perpetual calm present in that high and quasispiritual region. Man can achieve this calm by undertaking Bruno's program of spiritual regeneration. Ash Wednesday is the beginning of Lent, a period intended for such regeneration, and the appropriateness of the allusion to ashes is again clear.

The modern mind can easily appreciate the rich imagery of Bruno's writings. It is not so easy to give credence to the notion that Bruno would attempt to carry out a serious practical mission through such an arcane message. It is essential to realize, however, that the mentality of the 16th-century Neoplatonists had a strongly arcane bent, and that Bruno's message was aimed at the Hermetic-Neoplatonic court circle of Queen Elizabeth. Bruno's emblems and metaphysical optical symbols would have been clear to them. Regarded in this poetic and metaphysical light, Bruno's use of the Copernican theory as an emblematic hieroglyph would have been clear to his readers. The use of such devices was not uncommon, and Bruno himself employed them frequently in his writings.

Bruno describes the Copernican theory as a "hellebore," an herb supposed to be specific for the treatment of mental illness and thus a healing drug for the religious and political madness of his age. We can now understand why Bruno's inquisitors did not press him on his Copernicanism in the literal sense. They clearly realized that the scientific theory of Copernicus was entirely unrelated to Bruno's metaphorical use of that theory philosophically, religiously and politically.

We also wish to suggest here, without elaboration, that the generally accepted theory concerning the reasons for the trial of Galileo needs to be modified. There is considerable evidence suggesting that in similar political circumstances during the Thirty Years' War, Galileo was construed by those in authority as a resurrected Bruno. Galileo was well aware of the danger that his works might be interpreted as being Brunian. He protested repeatedly and vigorously that his intent was limited to scientific investigation, and that his writings were to be taken quite literally. His protestations were in vain. The condemnation of Galileo in 1633 could well have had as its principal (although misguided) aim the condemnation of Protestant-Catholic alliances of the kind Bruno had advocated. Such alliances had already been consummated, to the great discomfiture of the Papacy.

Let us cite here one example of the Inquisition's deep suspicions that Galileo's ultimate aims were philosophical, theological and political—that is, Brunian—rather than scientific. When Galileo's *Dialogue Concerning the Two Chief World Systems* first appeared, Niccolò Riccardi, the Master of the Holy Palace, expressed much concern over a symbol printed on the title page. The symbol consisted of three dolphins in a circle, each holding the tail of the next in its mouth. The dolphin was a common astrological symbol in Renaissance iconography. On hearing from a friend of Galileo's that this was the printer's colophon and not the work of Galileo, Riccardi was much relieved and said that Signor Galilei would have nothing to fear. Unfortunately Galileo's friend was unable at the moment to find another book published by the same printer that contained the dolphin symbol, and Riccardi's suspicions were revived.

Such preoccupation with symbols rather than with the intent of the book is illustrative of the kind of suspicions that were likely to be held by the contemporaries of Galileo and Bruno. In spite of his own great efforts and those of his ecclesiastical friends, Galileo was never able to allay such suspicions. When the Inquisition forced him to swear that the earth did not move, they were far less interested in the physical motion than in the metaphorical motion they were sure Galileo had in mind, and whose significance was certain to be both subversive and heretical.

In view of the great gulf between Bruno's mode of thought and Galileo's on what superficially appears to be the same subject, we must ask just what, if anything, was Bruno's influence on the scientific revolution. Gilbert, like Bruno, was a Hermetist, and it is possible that Bruno's influence on him was considerable. Bruno's influence was not, however, merely personal. Even though he was neither a scientific nor a methodological precursor of Galileo, we believe that for two reasons his influence on the birth of the scientific revolution was profound.

First, Bruno was supremely confident that man was at least in part a divine being and not merely the detestable product of the original sin, destined to fall lower and lower in the absence of some more or less capricious divine intervention in human affairs. Such a belief was heretical from the standpoint of both orthodox Catholicism and orthodox Protestantism, which were pessimistic concerning natural man. Bruno's advocacy of the Neoplatonic view made him a leading figure in the rebirth of man's confidence in himself, the like of which had not existed since classical antiquity. That confidence came to its greatest flowering in the 18th and 19th centuries, when serious men could believe that an earthly paradise was in prospect. Although this view has been somewhat clouded in the 20th century, it still forms the basis of action for most Western people.

Second, Bruno believed the path to perfection is the path of knowledge. It is true that Bruno's conception of what kind of knowledge is important was quite different from the modern view and would not in itself have led to much progress in the modern sense. Nevertheless, the conviction that man must look to knowledge rather than to faith in order to solve his earthly problems is one without which the scientific and technological revolutions could not have come into being.

Furthermore, Bruno believed in the ability of the human mind to comprehend the universe. He asserted that with such knowledge the savant can operate fruitfully on the world of nature in order to improve the human condition. That belief was synthesized into the highly influential views of Francis Bacon and has come to lie at the foundation of the goals of modern science. We have not been able to trace in detail all the threads of Bruno's philosophy. We hope nonetheless that we have been able to show the spirit of his approach, and to demonstrate that Bruno as Bruno was a richer contributor to the modern world than Bruno as proto-Galileo ever could have been.

The Electronic Scratch-Pad for scientists and engineers. $495.

The technical problems you handle are tough enough without unnecessary distractions and delays. Like the inaccuracy of a slide rule (!) Or the time wasted with a mechanical calculator. And the bothersome chore of looking up log and trig functions in various tables.

So Victor designed a calculator that helps you keep your mind on the overall problem by taking care of the complex mathematical operations for you. Eleven digits more accurately than a slide rule. Considerably faster than a mechanical calculator. With instant, push-button access to log and trig functions accurate to 14 digits. Without interpolation.

What's more, it's even capable of handling long sequential operations exactly as they're written. So you don't have to break a problem down and store intermediate answers.

All this without sacrificing the practical features of a 4-function calculator or the rugged dependability of a compact, desk-top model. And all for $495.

Victor's Electronic Scratch-Pad. It takes care of the figures while you take care of the problems.

For more complete information on Victor's Series 1800 Scientific/Engineering Electronic Calculator, simply call the Victor office listed in your Yellow Pages or send us the coupon.

VICTOR ELECTRONIC CALCULATORS

Victor Comptometer Corporation
Business Machines Division
3900 North Rockwell Street,
Chicago, Illinois 60618

Please send me full details about your New Series 1800 Scientific/Engineering Electronic Calculator.

Name ______________________

Title ______________________

Firm ______________________

Address ______________________

City __________ State ______ Zip ______

Victor Comptometer Corporation, Chicago, Ill. 60618
Also Galt, Ontario, Canada and Zurich, Switzerland.
SA-43

The Energetics of the Bumblebee

The bumblebee, like many other insects, evolved in interaction with the flowers it feeds on and pollinates. This evolutionary interaction is apparent in the energy budget of its activities

by Bernd Heinrich

Charles Darwin discovered that when red clover was sequestered from visits by bumblebees, the plant failed to produce seed. This led two 19th-century contemporaries, the German biologists Karl Vogt and Ernst Haeckel, to observe that the British Empire owed its power and wealth largely to bumblebees, since its power resided mainly in the Navy, whose sailors subsisted on beef, which came from cattle that subsisted on clover, which, as the mother of the chain, depended for its propagation on pollination by bumblebees. The bumblebee's ecological puissance was demonstrated in a practical way by ranchers in New Zealand, who in 1885 imported the insect and found that it indeed brought about large increases in the production of clover seed.

Few members of the living world are linked together as intimately as bees and flowers. For bees the nectar of flowers is the sole source of food energy; for flowers bees serve essentially as copulatory organs, delivering pollen from one plant to another and thereby effecting fertilization. By mutual evolution bees and flowers have become exquisitely attuned to their mutual convenience. Many flowers have evolved features making them attractive to those bees that are efficient in pollinating them and making them unattractive or essentially inaccessible to other foragers. The flowers attract by signals such as scents and colors that identify them for their partners, and they minimize or avoid visits from other pollinators by blooming at times when the latter are not active or by holding their nectar in a long, tubular corolla where it can be reached only by the favored animals.

Some ecological effects of the evolutionary collaboration of bees and flowers have been investigated by Theodore Mosquin of the Canadian Department of Agriculture and by Herbert Baker of the University of California at Berkeley, working with Gordon W. Frankie at Texas A & M University and Paul Opler in Costa Rica. They have found that in each region the various plants requiring cross-pollination by bees bloom at different times spaced out over the growing season, so that they avoid excessive competition for the attentions of the available pollinators. This separation of blooming periods no doubt has been brought about by natural selection as influenced by the bees' season-long feeding needs.

The nature of the pollinating agent has determined the properties of flowers in another way. If the nectar-seeking insect could obtain a full meal at one sitting, there would be no need to flit from flower to flower and consequently no cross-pollination. Evolutionary selection has therefore dictated that the individual flower offer only a small part of the pollinator's requirement, thus compelling the forager to visit many blossoms to fill its needs. Each blossom must provide enough nectar to make a visit worthwhile but not enough to allow the pollinator to get its fill from just a few blossoms.

I have investigated the matter quantitatively by observing the feeding of a species of bumblebee (*Bombus vagans*) on the compound blossoms of hawkweed, which have only minute amounts of nectar at any one time and which are usually not very attractive to bumblebees. To test the bumblebee's capacity I placed a large deposit of nectar in a single blossom and found that 100 microliters or more was needed to sate a bee. Since a large fraction of the flowers on which bumblebees forage contain less than half a microliter of nectar, this indicated that a bumblebee in the field probably visits 200 or more such flowers during a foraging trip. I followed one bee during part of such a trip and observed that it visited 337 blossoms before I lost sight of it.

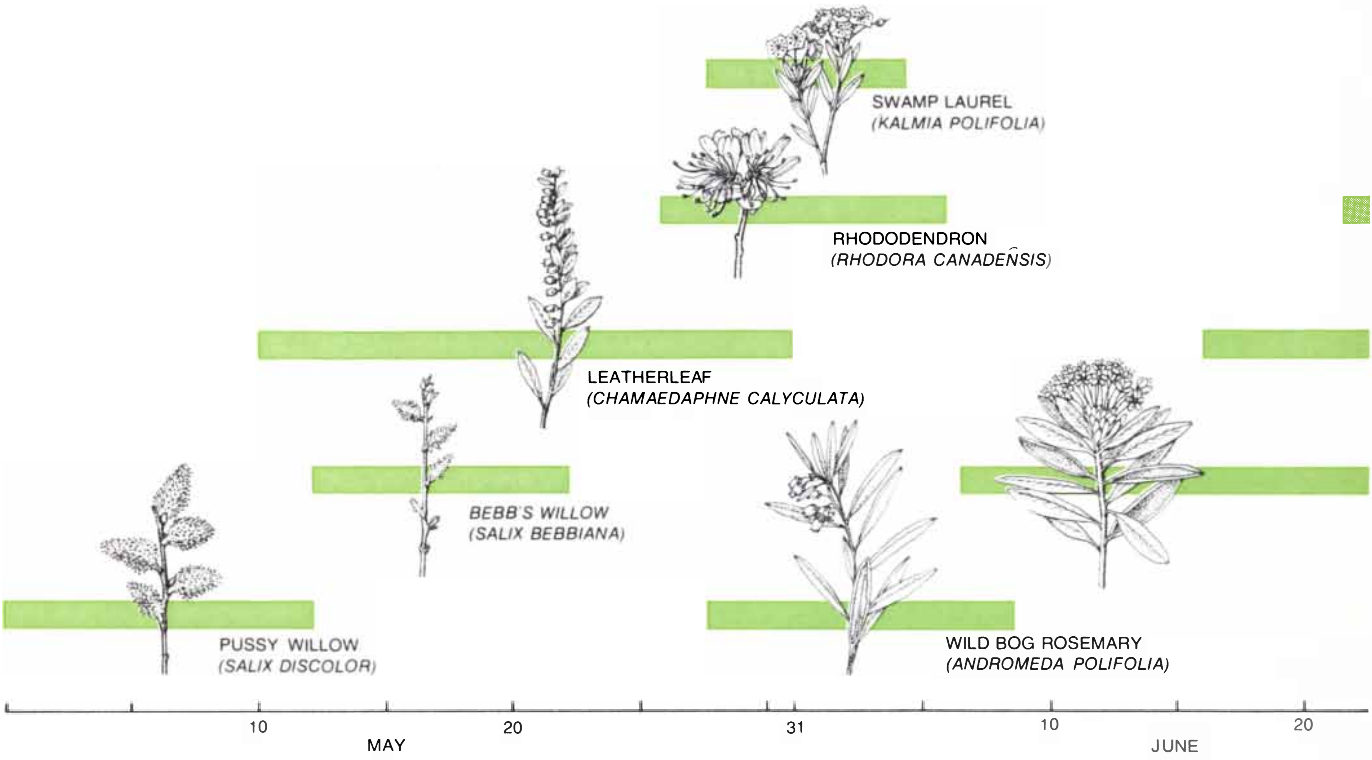
Bees of course are well known to be indefatigable workers. The nature of their interdependence with flowers therefore suggests interesting questions concerning the bee's energy budget and the balance sheet it shares with flowers. I had already been impressed by the remarkably high energy expenditure of the foraging sphinx moth and by its virtuosity in regulating its body temperature [see "Temperature Control in Flying Moths," by Bernd Heinrich and George A. Bartholomew; SCIENTIFIC AMERICAN, June, 1972]. I was therefore drawn to a detailed study of the energetics of bumblebees.

The bumblebee is a particularly interesting subject because it can forage at temperatures as low as zero degrees Celsius, where all other insects are immobilized. Bumblebees are found not only in warm regions but also on high mountains and in arctic climates. As a cold-blooded animal ordinarily restricted in body temperature to the temperature of the environment, the bee must step up its generation of energy to heat its muscles to the temperature necessary for flight; it does this by exercising the muscles of its thorax (essentially shivering) before it takes off. Thanks to the bumblebee's ca-

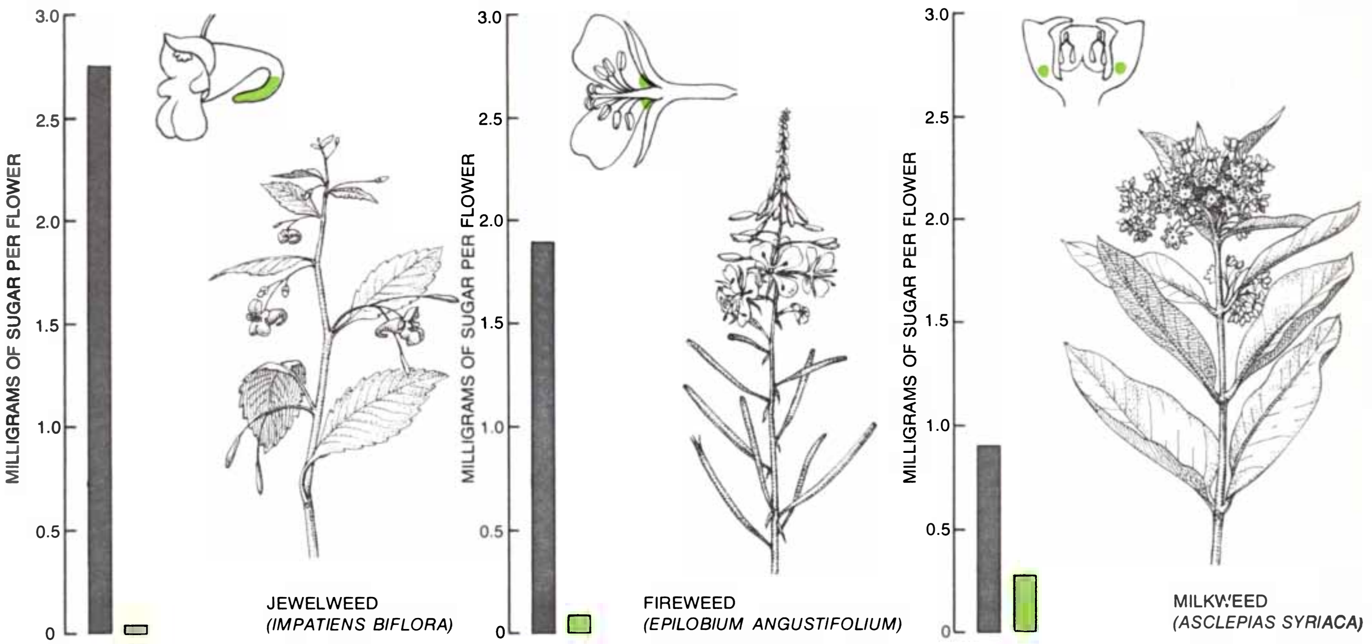
CLINGING UPSIDE DOWN to a dangling groundsel blossom in a shady Maine bogland (*opposite page*), a bumblebee of the species *Bombus ternarius* thrusts its proboscis (*left*) into the flower to gather nectar.

pability for generating enough heat to keep it active over a wide range of environmental temperatures, it is possible to obtain comparative information about its energy requirements at various ambient temperatures. Furthermore, since essentially its sole metabolic source of energy is the sugar in the nectar it collects from flowers, and since the extent of its foraging and the flowers it visits can be observed, one can estimate how much energy the bumblebee spends in its foraging trips.

I conducted my field studies of bumblebee energetics at my family's farm in Maine and followed up with laboratory investigations at the University of California at Berkeley. The area in Maine (near Farmington) is a haven for a variety of flowers and many species of bees, among which bumblebees seem to be the



SEQUENCE OF BLOOMING among 10 common plants in a bog in Maine is staggered, beginning in late April (*far left*) with pussy willow and continuing without interruption into August (*far right*), when meadowsweet continues to bloom. At least one plant species



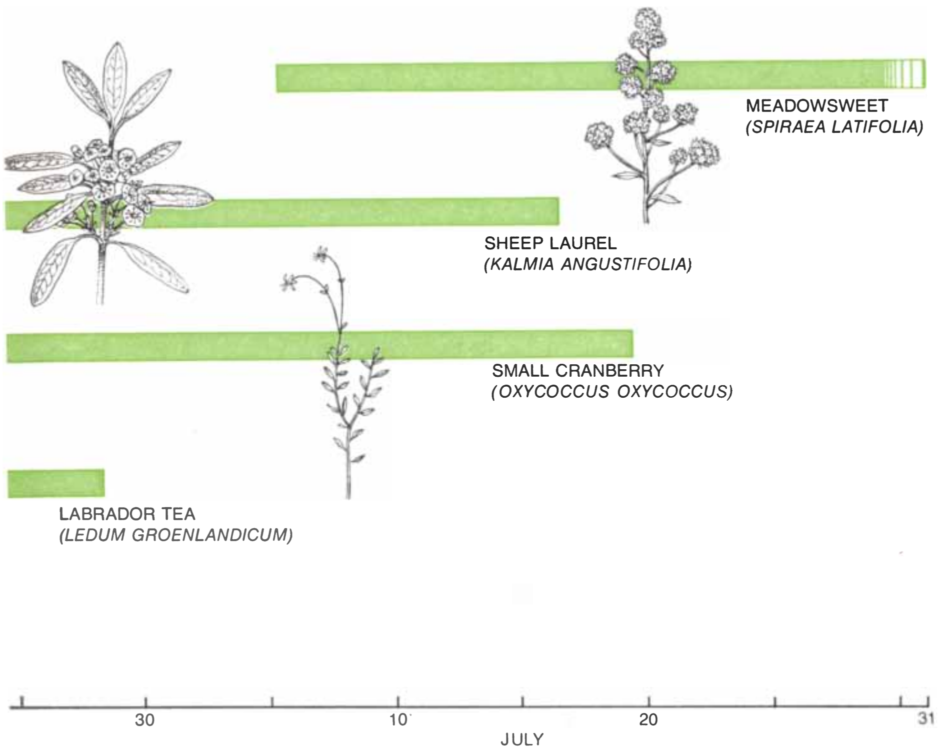
QUANTITY OF NECTAR produced by different species of plants is generally proportional to the size of the blossom. The 24-hour production of nectar per floret of five plants that were shielded from foragers is expressed here in milligrams of sugar (*gray*). A most common flower-visitors. A bumblebee visits about twice as many blossoms per unit of time as other bees, and it works from dawn to dark. Active at all ambient temperatures, bumblebees serve as pollinators on nearly all days over the entire season, visiting the flowers that come into bloom one after the other.

Although bumblebees are potentially promiscuous, being attracted to many species of flowers, they do show definite preferences, presumably based on the comparative nectar reward of the favored flower. I came across an interesting case of individual taste: In a field where two almost identical species of hawkweed were blooming, the only apparent difference being that one was yellow and the other orange, one bumblebee I followed for 146 consecutive visits chose only the orange flowers, whereas another bee of the same species showed almost as definite a preference for the yellow, selecting during a foraging trip the yellow species 170 times and the orange only 14 times. Evidently each bee had developed its own preference on the basis of its foraging experiences. Whatever the reasons, most of the bumblebees I studied tended to be faithful to particular species of flowers. This was convenient for my research, because it enabled me to estimate how much sugar the bees collected in a foraging trip.

In order to investigate the bees' energy budget I had to find a means of measuring their rate of energy expenditure in the field. The usual method of obtaining this rate in an animal is to measure its oxygen consumption, a direct indication of its rate of metabolism. That method obviously would not be applicable to a bee engaged in complex behavior in the field. A good index to a bee's energy expenditure could, however, be obtained by measuring its temperature as it perched on a flower and timing the duration of its flights between flowers. The energy expended by a bumblebee in flight had been measured in the laboratory by observing the bee's rate of oxygen consumption. About 80 percent of the energy a bee expends goes into heat as a by-product of the heightened rate of metabolism. To get an indication of the bee's rate of energy expenditure when it was perched on a flower I therefore adopted a simpler technique: I would grasp it with a gloved hand between my thumb and forefinger and quickly thrust a tiny thermistor into its thorax, thus obtaining the temperature before the body could cool appreciably.

One striking finding that soon emerged was that whereas the thorax, where the muscles for the wings are located, shows a considerably elevated temperature, the abdomen (the rear part of the body) remains at just about ambient temperature. The bee does not waste energy on the body part that is not involved in powering flight, and thereby it saves itself half the metabolic cost of the energy it would have to produce if it heated its entire trunk.

Another interesting observation was that, regardless of the length of time the bumblebees perched on a flower and regardless of the proportion of time they

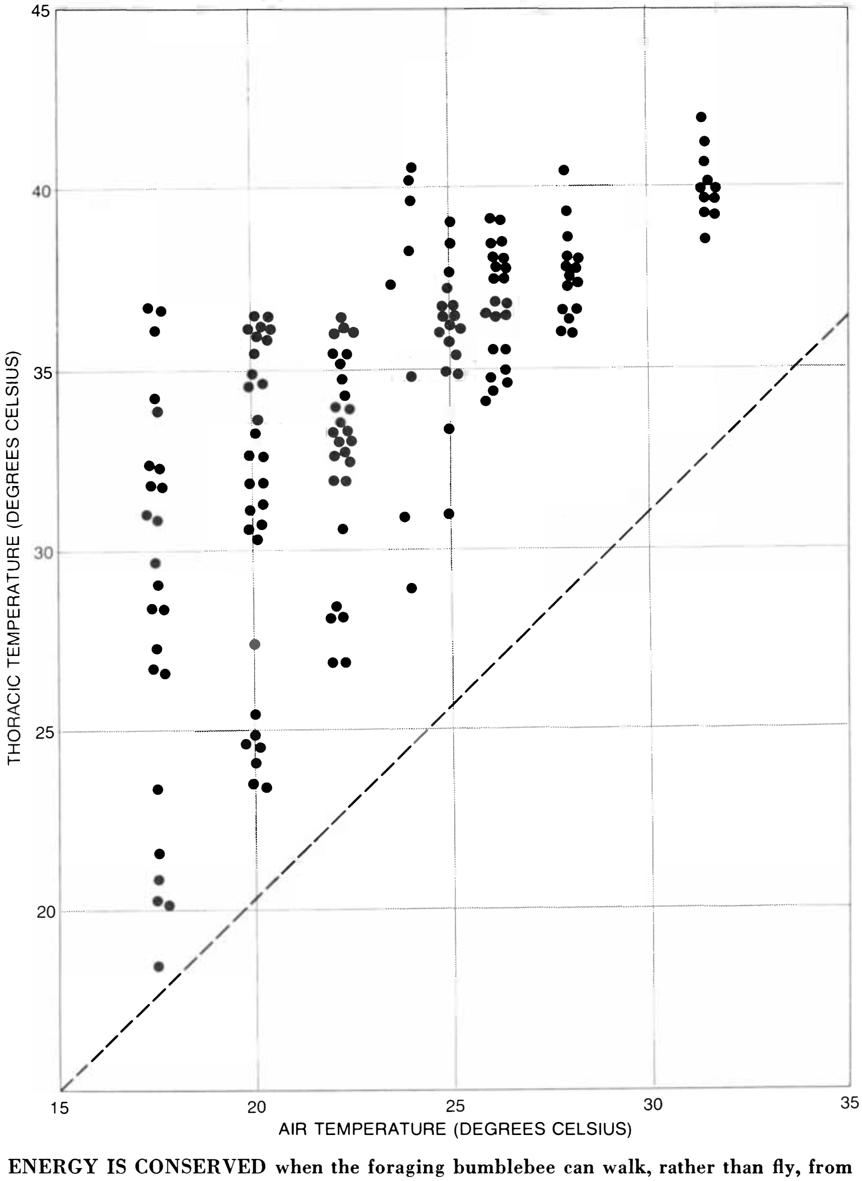


is in bloom at any one time for more than 90 days and so the bumblebees in the area have a steady supply of nectar and pollen. Foraging simultaneously cross-pollinates the plants.

second bar (*color*) shows the amount in unshielded florets. Meadowsweet and goldenrod may seem low producers, but the one may have 45 florets per bloom and the other 1,300.

spent in flight during a foraging trip, they usually maintained their thoracic temperature at a high level: between 30 and 37 degrees C. (86 to 98.6 degrees Fahrenheit). This requires an exorbitant rate of heat production. A bumblebee may elevate its thoracic temperature by as much as 35 degrees C. when the ambient temperature is only about two degrees. For a bumblebee, only one or two tenths of a gram in weight, maintaining such an elevation of body temperature implies a prodigious rate of energy expenditure, in view of its high rate of heat loss to the atmosphere. This rate of loss from the bee's small body is unavoidably high even though it is covered with fur (which serves to pick up pollen as well as to help to conserve heat in some measure).

One might suppose the bee cools off while it is perched on a flower. As a rule it does nothing of the sort. I found that in most cases bumblebees kept their body temperature at the level required for flight even during long periods of stationary feeding, which I would stretch out by filling blossoms with a large amount of sugar syrup. Why do they spend all that energy to regulate their temperature while they are not in flight? The answer seems obvious: the engine remains revved up so that the bee does not need to spend time warming up in order to take off. In the field the bee usually keeps itself ready to take off at an instant's notice, whether for the purpose of flying on to another blossom or to escape some threat.

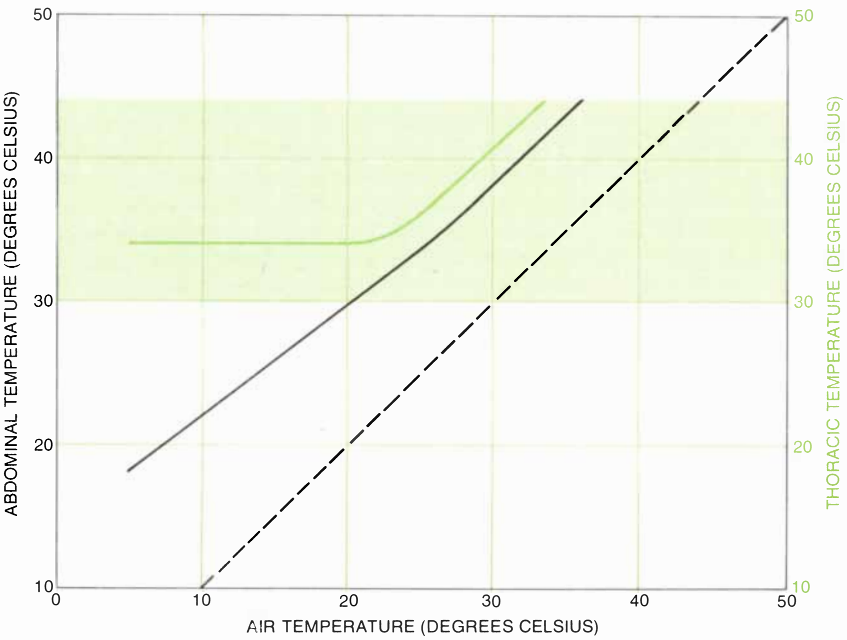


ENERGY IS CONSERVED when the foraging bumblebee can walk, rather than fly, from one nectar source to the next. The graph shows the thoracic temperature of bees that were foraging on meadowsweet. When the air temperature was below 24 degrees C., the thoracic temperature of the bees frequently fell below the 30-degree level required for flight as they crawled from floret to floret. Bees that had been most recently airborne were the warmest.

How much energy must a bumblebee spend to maintain an elevated temperature while it is not flying? On the basis of its passive rate of cooling when it is not foraging, I calculated that at an ambient temperature of five degrees C. a medium-size bumblebee must produce a little more than half a calorie of heat per minute to keep its thoracic temperature at 30 degrees, or 25 degrees higher than the air temperature. This rate of heat production is close to that required for flight itself. At high air temperatures, of course, the need for generating heat in the thoracic muscles is reduced. At ambient temperatures higher than 25 degrees C. the heat-regulation factor in perching bees is eliminated, and the total energy cost comes down to what is needed to power flight, with heat as an incidental product. My measurements of oxygen consumption in the laboratory show that in bumblebees, as in sphinx moths, the energy cost of flight is the same at all air temperatures.

Most other insects heat up their thorax only during and just before flight; the bees are almost unique in their ability to maintain an elevated thoracic temperature while remaining still. As the Danish physiologist August Krogh and other investigators showed many years ago, bees produce a considerable quantity of heat by activating their flight muscles in a shivering mode. Ann E. Kammer of Kansas State University and I have investigated the matter by electrophysiological methods and have confirmed that the activation of these muscles is in fact solely responsible for the heating of the thorax. Curiously, when the bee is stationary, this warm-up is not accompanied by any visible vibration of the wings, such as is seen in moths. We have not yet found out how the bumblebee uncouples its wings from the flight muscles during its warm-up. We think, however, that it does so by means of a mechanical clutch mechanism in the complex articulations at the base of its wings.

Armed with the foregoing background of theory and information and with a thermistor and a stopwatch, I was ready to investigate the bees' energy cost of foraging in the field. First I measured the energy income they would obtain from the flowers. I collected samples of the nectar from a given flower species, measured the sugar concentration in the nectar with a pocket refractometer and then computed the energy this represented in terms of calories. Such a measurement showed, for example, that on the average a fireweed blossom filled with nectar (having been shielded from feeders for 24 hours) contained enough sugar to support a bee's maximum metabolic rate (the rate required for continuous flight or for the maximum elevation of its body temperature when it is stationary) for about 13 minutes. Ordinarily, of course, a bee does not often come on such a bonanza in the field. I found that in field conditions the same blossom in the afternoon, having already fed a series of visitors, would on the average provide only enough nectar to supply a bee with energy for about a minute's flight. As a result to obtain a profit beyond the energy consumed in the foraging activity the bee would have to visit flowers at the rate of at least one or two per minute. Actually I observed that bumblebees commonly fed on fireweed blossoms at the rate of 20 to 30 per minute, which was more than sufficient to supply the bee's energy needs for flying and for keeping the engine revved up while it fed on the blossoms. Thus the flower amply repaid the bee's efforts, particularly since it might also provide a dividend of pollen that the bee would take back to its nest to feed its larvae.



ENERGY INVESTMENT required to keep up the body temperature that enables a bumblebee to fly varies with the air temperature. The graph shows the temperature of the bumblebee's abdomen, averaged in both sunlight and shade, at various air temperatures (*black curve*) and the average temperature of the thorax, the site of the bee's flight muscles, which must not fall below 30 degrees Celsius or rise above 44 degrees (*colored curve*). The temperature of the abdomen, only a few degrees higher than the air temperature, is maintained by passive heat flow only. Until the air temperature approaches 25 degrees C., however, maintenance of the thoracic temperature at an average of 33 degrees requires an investment of energy far greater than the one required for intermittent flight from flower to flower.

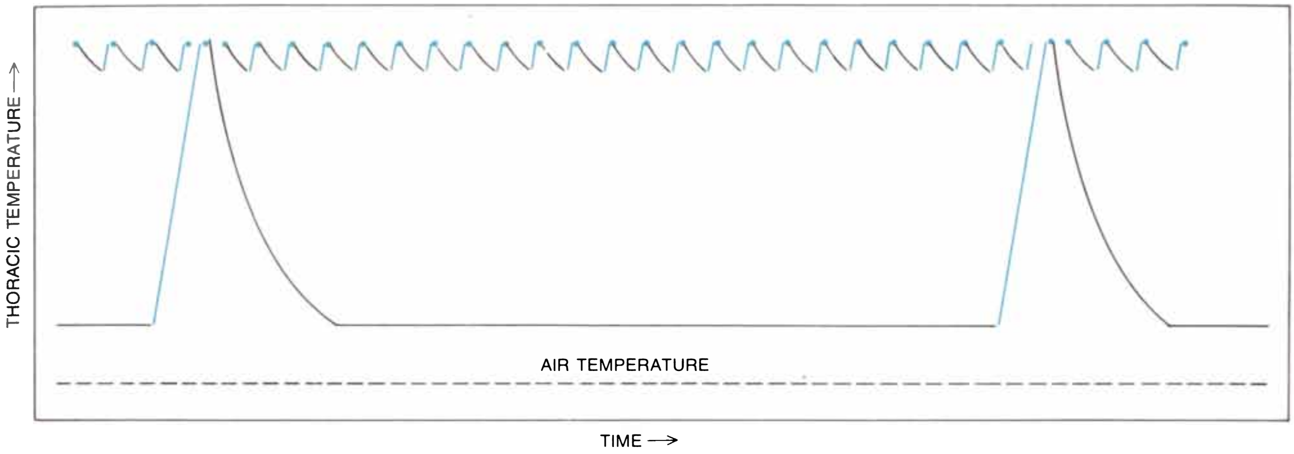
This example only indicates a general picture. In practice there are several factors complicating the question of whether or not a given foraging expedition will be worthwhile for the bee. That depends not only on the amount of caloric reward offered by each flower but also on the distance between blossoms, on the air temperature at the time and on other factors.

In general, as we have seen, if a bumblebee must maintain the necessary body temperature by shivering continuously while it is resting on each blossom, it needs to obtain enough sugar to provide more than .54 calorie per minute throughout the duration of its foraging trip, regardless of how much of the time it spends in flight. If it does not spend energy for warming up while it is feeding but needs to spend more than 50 percent of the time flying from blossom to blossom, it requires an intake of at least .27 calorie per minute. If it can dispense with the stationary warm-ups and spend no more than 10 percent of the foraging time in flight, it can make an energy profit on an intake of only about .05 calorie per minute. Let us see how these requirements work out for the behavior of the bees and the pollination chances of various flowers.

In the spring, when the bumblebees are forming their colonies and the queens and workers presumably have their highest need for energy, the early-blooming flowers tend to be relatively rich in nectar. The foraging bees have ample energy for flying rapidly from flower to flower and keeping themselves well warmed up while feeding. Speed of foraging is at a premium at this time of year, and the bees are alert and difficult to capture even in the cool of dawn with the air temperature almost at the freezing point of water.

That is in sharp contrast to the situation in late summer, when the bee colonies are breaking up, the energy demands and the need to bring in a profit at a high rate are not so large and the sources of nectar tend to become less abundant. The foragers in our Maine area then have to depend on the nectar-poor blossoms of plants such as meadowsweet and goldenrod. Each floret has only a minute amount of nectar, but providentially the blossoms occur in dense panicles, clusters consisting of hundreds or thousands of florets. The bee can crawl from blossom to blossom and may spend a long time on each cluster, using a significant amount of energy only when it flies from one panicle or plant to another. While feeding on such plants the bee does not trouble to keep itself warmed up. I found that in relatively cool air (20 degrees C. or less) the bees' thoracic temperature was generally below the threshold for flight. At such air temperatures many of the bees were so unready for flight while feeding that they could be shaken off the blossoms onto the ground. Bumblebees seldom ventured, however, to feed on meadowsweet or goldenrod unless the air temperature was above 20 degrees C. or they were warmed above this temperature by the sun. If passing clouds cut off the sunshine on a cool day, the bees lost flight readiness within minutes.

In short, these observations and similar ones on other plants showed that the bumblebee accommodates itself to a meager supply of nectar by practicing drastic economies in the use of energy. By means of these economies it is able to render the service of pollination to

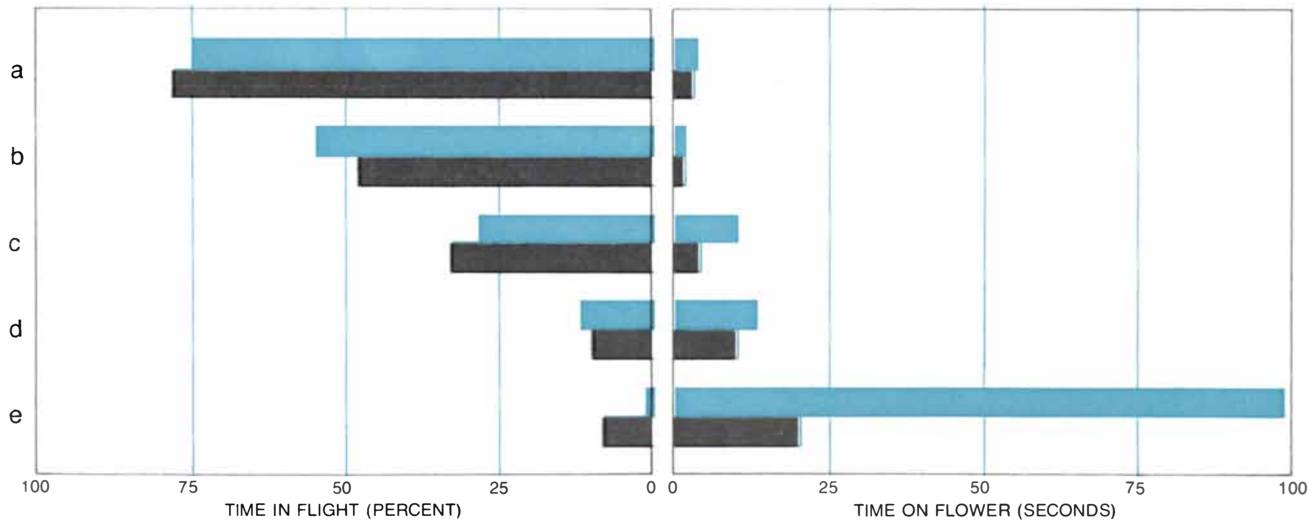


COMPARATIVE COST of foraging on plants with single or few blossoms and on plants with many florets per bloom is shown schematically. Given a low air temperature, a bumblebee that crawls around on the blooms of meadowsweet or goldenrod will only warm up its flight muscles (*color*) and take flight twice during a span of time when another bee, feeding on jewelweed or fireweed, will have warmed up and taken off many times. The second bee, however, will probably gather a larger amount of nectar.

plants offering only a spartan diet when richer sources are not available. In the case of the fireweed, a relatively rich provider (about .09 milligram of sugar in each floret), a bee can afford to spend 50 percent of its foraging time in flight from flower to flower and to keep itself warmed up in readiness for flying; in contrast, meadowsweet (offering only .005 milligram of sugar per floret) limits the bee to only a fifth as much flying and requires that the engine be cut to a low idling level during the stationary interval. The sphinx moth, lacking the adaptability and options of bees, cannot forage profitably on meadowsweet or goldenrod, because it hovers over a flower while feeding instead of settling on it and therefore must continuously produce heat even at high air temperatures.

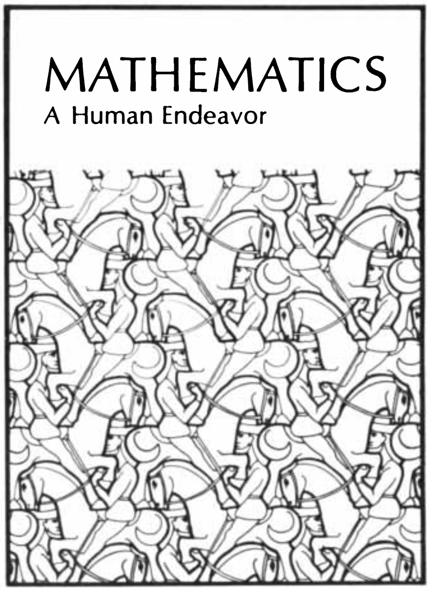
The bumblebee may sometimes forage on some plants without taking nectar at all and instead collecting pollen to take to the nest. On meadowsweet, for example, the bee simply sweeps the entire panicle and flies rapidly from one to another; on bittersweet it shakes the pollen loose by flight-muscle vibrations that vibrate its body and the flower and make a buzzing sound. For such operations the bee must, of course, use a large amount of energy. If the plant from which the pollen is collected does not furnish enough nectar, the bee setting out on a pollen-foraging trip must either use honey from the nest for its energy supply or collect nectar from other plants before it visits the pollen suppliers.

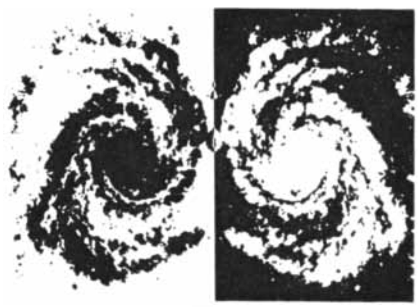
We have seen enough to discern the outlines of an intricate web of interrelations among the bees and the flowers, governed by the energy needs and payoffs as its basic plan. The competition among plants for the pollinators' service, and among the pollinators for the plants' food rewards, decisively shapes the behavior and the structure and physiology of both the plants and the pollinators. The investigation of the energetics of the bumblebee therefore offers a beautiful case study for insight into the operations of evolution and the way a cooperative balance can be established in the ecology of a community.



PROPORTION OF TIME spent by bees in flight and in foraging is more closely related to the kinds of bloom visited than it is to the air temperature. Black bars show both activities at an air temperature of 20 degrees C. in shade; colored bars, both at 30 degrees. The plants are jewelweed (*a*), fireweed (*b*), milkweed (*c*), meadowsweet (*d*) and goldenrod (*e*). The only significant increase in foraging time at low temperature occurred when bumblebees fed at length on the florets of goldenrod, which contain little nectar.

ANNOUNCING THE BEST-SELLING TITLES IN OUR 1972 BOOK OFFER





Last fall we mailed a special book offer to all *Scientific American* subscribers in the United States and Canada. In the brochure we announced that a 15-volume *Scientific American* Resource Library would be given to anyone who could select the four books that would be ordered in the greatest numbers by recipients. The results have now been tabulated, and the winners have been notified by mail. For the interest of others who mailed in a selection, we list below the ten best-selling titles in the offer.

1. **Mathematics**
   A Human Endeavor
   A Textbook for Those Who Think They Don't Like the Subject
   Jacobs, $8.50

2. **Worlds-Antiworlds**
   Antimatter in Cosmology
   Alfven, $3.50

3. **The Psychology of Consciousness**
   Ornstein, paperbound, $3.50

4. **Continents Adrift**
   Readings from SCIENTIFIC AMERICAN
   With Introductions by Wilson,
   clothbound $7.00, paperbound $3.50

5. **Spacetime Physics**
   Taylor and Wheeler,
   clothbound $5.50, paperbound $2.75

6. **Altered States of Awareness**
   Readings from SCIENTIFIC AMERICAN
   With Introductions by Teyler,
   clothbound $6.00, paperbound $2.95

7. **Martin Gardner's Sixth Book of Mathematical Games from Scientific American**
   Gardner, $9.95

8. **Computers and Computation**
   Readings from SCIENTIFIC AMERICAN
   With Introductions by Fenichel and Weizenbaum,
   clothbound $10.00, paperbound $4.95

9. **Oceanography**
   Readings from SCIENTIFIC AMERICAN
   With Introductions by Moore
   clothbound $11.00, paperbound $5.75

10. **The Social Animal**
    Aronson, paperbound $3.50

Thanks to everyone who participated.

W. H. FREEMAN AND COMPANY
660 Market Street, San Francisco, California 94104
58 Kings Road, Reading, England RG1 3AA

SOME THINGS ARE CHANGING FOR THE BETTER:

A lot has happened to interferometry since Professor Michelson.

In 1892 Professor A. A. Michelson used his recently invented interferometer to determine the number of wavelengths of cadmium light in the standard meter. His error was 5 x 10^{-8} meter. It took another 70 years until the invention of the laser permitted the application of interferometry with this accuracy to everyday industrial measurements.

Today, HP's laser interferometer has found a place in metalworking and integrated circuit manufacturing as the tool to guarantee that a given dimension is precisely achieved. This modern version of Michelson's invention is enhancing accuracies of all types of new and existing machines and, as a consequence, lowering inspection costs and reducing rejection rates.

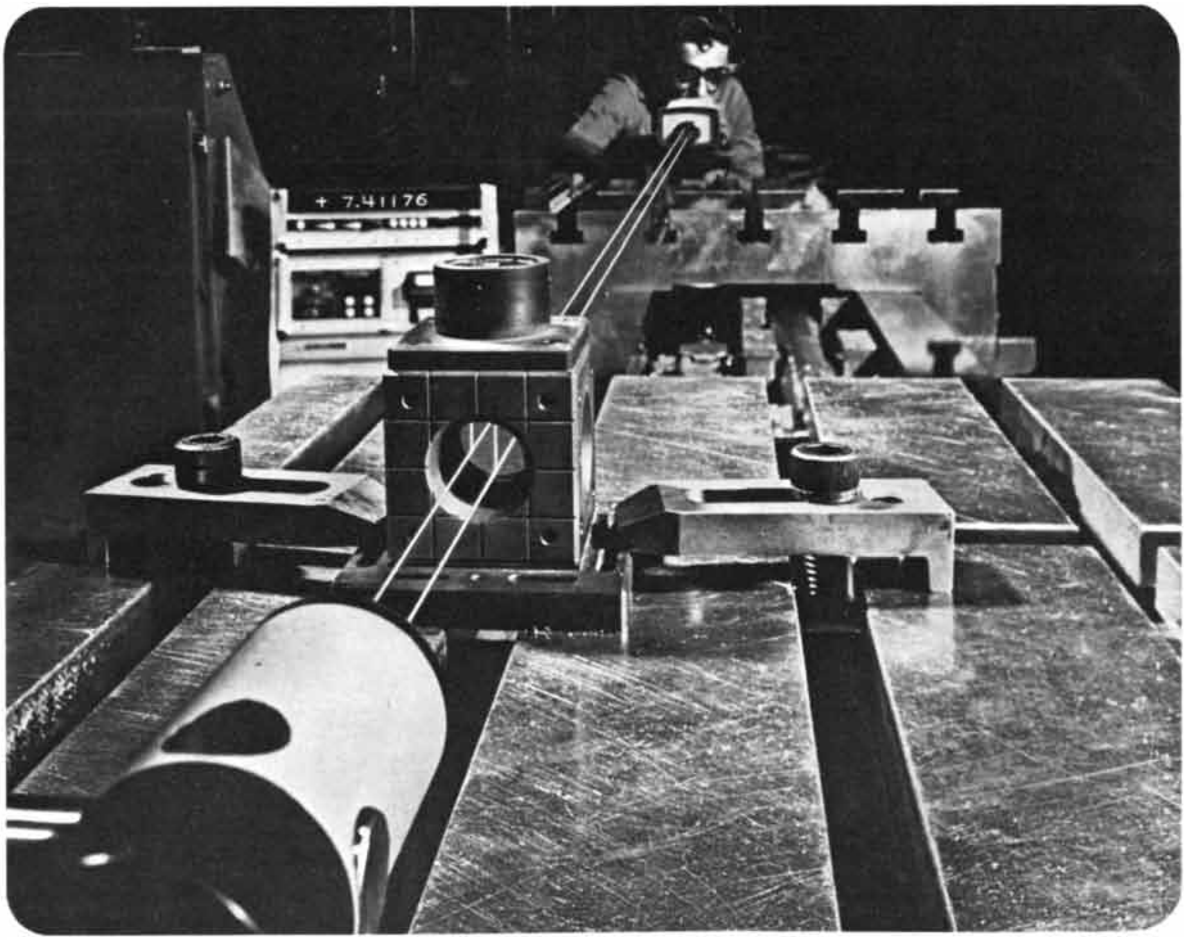
A little over a year ago, a group of HP engineers seeking to improve the quality and flexibility of this measurement technique developed the "HP Magic Cube." In essence, it is an interferometer that can have a laser beam aimed at it from some distance away.

With it many kinds of measurements can now be taken without moving the laser head — pitch, yaw, flatness, straightness, squareness, and non-contact measurements — all that's needed to define a machine tool or measuring machine.

It can even make measurements around a corner, in tight places, or simultaneously measure x and y axes.

Moreover, it's portable. You can set it up, plug it in and start making error-free measurements in minutes wherever it's required in your metrology and quality control laboratories.

If you would like to know more about laser interferometry, we'll be glad to mail you our comprehensive brochure on the HP 5526A Laser Measurement System.



High-performance liquid chromatography: a new HP analytical tool for life scientists.

Gas chromatography (GC) has revolutionized analytical chemistry in practically every discipline. However, it has been of somewhat limited help to the life scientist because he works mostly with large molecules that cannot be vaporized for GC without breaking them down.

As a result, liquid chromatography (LC) has always been more attractive to the life scientist because it separates a sample in the liquid state at relatively low temperatures. But only recently has LC instrument technology advanced enough to give him the analytical speed and sensitivity that he also needs.

At the forefront of this new technology is HP's new Model 1010 Liquid Chromatograph. Already well established with European chromatographers, the 1010 is capable of performing all forms of LC, including ion-exchange and gel permeation. It delivers constant flow through the columns at pressures up to 3700 psig. A septum-less injector permits reproducible sample injection against full-column pressure; column change and solvent flush are exceptionally fast and convenient. Completely modular in design, the 1010 can be equipped for solvent programming, gradient elution and fully automatic preparative operation. Ask for full information.

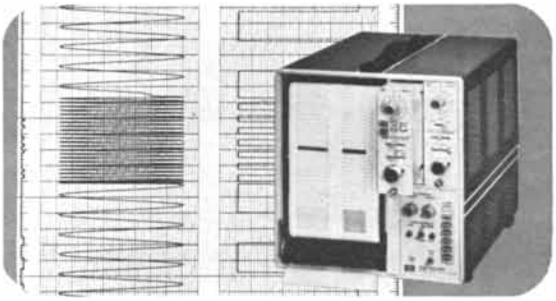
HEWLETT *hp* *PACKARD*

Sales, service and support in 172 centers in 65 countries.
Palo Alto, California 94304. Offices in principal cities throughout the U. S.

Just a better trace for your money. That's all.

There's a great temptation to talk about the features on our new HP 7402A two-channel oscillographic recorder as though they were the last word in recorder design. But as we both know, the field is in a great state of flux, with new technology appearing in every model.

Common sense prevails: We simply want you to understand that the HP 7402A delivers one of the clearest, most readable, traces you will find on any currently available machine irrespective of price. Yet the price of the HP 7402A can be as little as $1740 for a complete recorder. Domestic USA price only.



For your money, you get a big, wide, easy-to-read 50 mm trace on each channel. A high-pressure inking system puts down an instant drying, fine image with no overshoot, no skipping and no smear at any speed (in tests, the carbide tips showed no wear after 2,000 miles of travel over 80,000 feet of chart paper; you can expect them to last the life of the machine).

You also get the flexibility and economics of plug-in amplifier modules and a wide range of options to accommodate your exacting recording needs.

We'll be happy to send you more information on this sensible recorder.

00304

Mail to: Hewlett-Packard, 1509 Page Mill Road, Palo Alto, CA 94304
Please send me information on the following:
( ) HP 5526A Laser Measurement System
( ) HP Model 1010 Liquid Chromatograph
( ) HP 7402A Oscillographic Recorder

Name ______

Title ______

Company ______

Address ______

City ______ State ______ Zip ______

MATHEMATICAL GAMES

How to turn a chessboard into a computer and to calculate with negabinary numbers

by Martin Gardner

"Napier's bones" (last month's topic) are the calculating rods that were invented by John Napier, the 16th-century Scottish mathematician who discovered logarithms. In *Rabdologia,* the book in which Napier first explained his "bones," he also described a curious method of calculating by moving counters across a chessboard. This method, which seems to have been completely forgotten, deserves to be remembered for several reasons: it is not only a pleasant recreation but also a valuable teaching device and of considerable historic interest. It is the world's first binary computer, and it came almost 100 years before Leibniz explained how to calculate with binary numbers! Although Napier did not express numbers explicitly in binary notation, we shall see how his counting board is equivalent to doing so.

The use of checkered boards and cloths for calculating was widespread in Europe during the Middle Ages and the Renaissance. English words such as "exchequer," "check" and "counter" derive from these boards; even "bank" comes from the German word for counting board, *Rechenbank.* The algorithms for calculating on these boards, however, were clumsy. By adopting a binary system and basing his algorithms on old methods of multiplying by "doubling," Napier created a remarkably efficient counting board unlike any that had been in use before.

Napier's counting board is a chessboard of arbitrary size, with columns and rows labeled by the doubling series 1, 2, 4, 8, 16, 32, ... These numbers are, of course, successive powers of 2. Before explaining Napier's methods for multiplying, dividing and extracting square roots, let us see how his board can be used for addition and subtraction. Suppose we want to add $89 + 41 + 52 + 14$. Each number is expressed by placing counters on a row of the board [*see "a" in top illustration on opposite page*]. A counter has the value of its column. (Ignore the row numbers on the right margin.) Thus the fourth row shows 89 as the sum of $64 + 16 + 8 + 1$. If you think of each counter as 1 and the empty cells as 0, then 89 is represented in binary notation as 1011001, and similar notations can be made for the other three numbers. The counters can be positioned rapidly because any positive integer is uniquely represented as a sum of the powers of 2. Start at the left and put a counter on the largest power less than the number to be represented, then move right and place a counter on the next larger power that, when added to the previous power, will not exceed the desired number. Continue in this way until the unique binary representation is obtained.

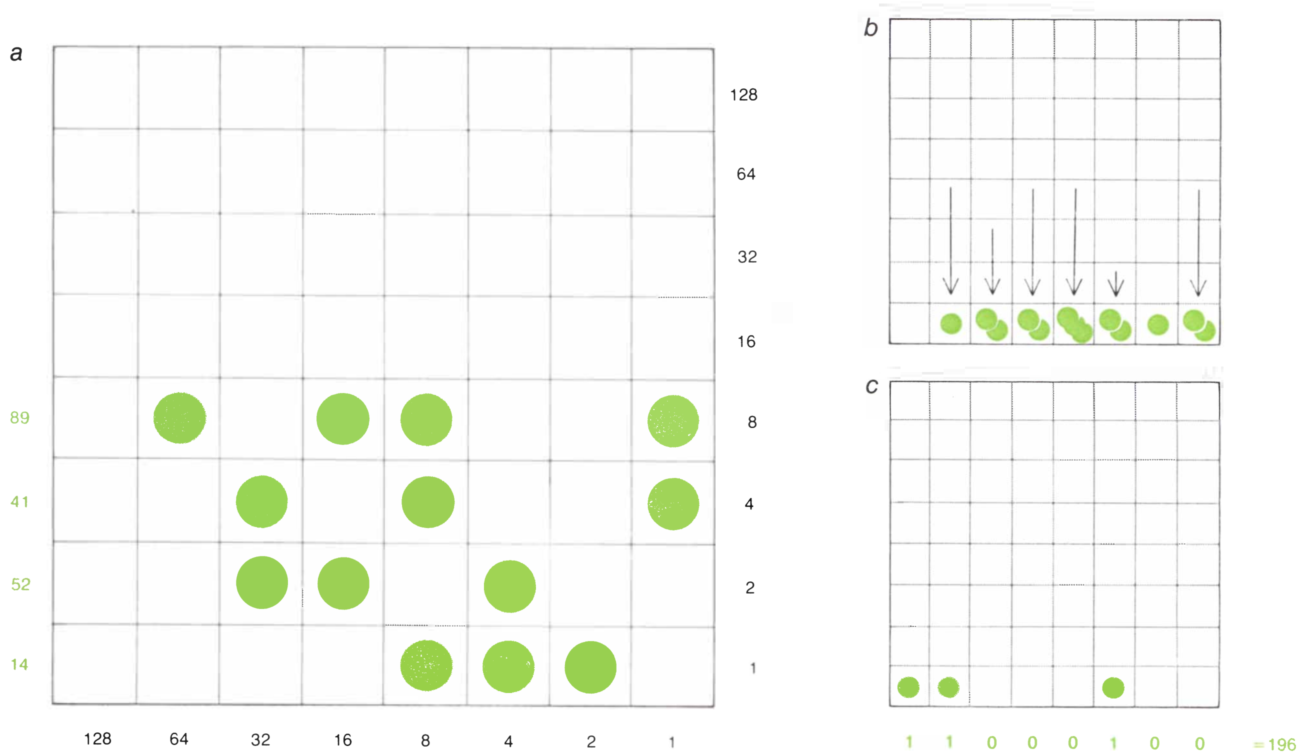
To add the four numbers, first move all the counters down like rooks in chess to the bottom row [*see "b" in top illustration on opposite page*]. Adding the values of all these counters will give the correct sum, but we want to express the sum in binary notation. To do this, "clear" the row of multiple counters on a cell by the following procedure. Start at the right, taking each cell in turn. Remove every pair of counters on a cell and replace them with a single counter on the adjacent cell to the left. We shall call this "halving up." Clearly it will not affect the sum of the counters' values because every pair of counters of value n is replaced by one counter of value $2n$. The final result after clearing is the binary number 11000100, or 196 in decimal notation [*see "c" in top illustration on opposite page*].

Subtraction is almost as simple. Suppose you want to take 83 from 108. Represent the larger number on the second row and the smaller on the bottom row, as shown in *a* in the bottom illustration on the opposite page. You can now do subtraction in the usual manner, starting at the right and borrowing as you go, but I prefer to alter the entire second row (preserving the total value of its counters) until each counter on the bottom row has one or two counters above it, and no empty cell on the bottom row has more than one counter above it. This is done by "doubling down" on the second row: removing a counter and replacing it with two counters on the next cell to the right. How the top row looks after it is transformed to meet the two specified conditions is shown in *b*. The next step is to "king" (as in checkers) each counter on the bottom row by moving a counter on top of it from the cell directly above. After this is done the second row shows in binary notation the difference between the two numbers. In this case it is 11001, or 25, as shown in *c* in the illustration.

A different subtraction method is to "complement" the smaller number and add. A number is complemented by placing a counter on each of its empty cells and removing all counters originally there. In other words, each digit in its binary notation is changed: 0 to 1, and 1 to 0. (If the subtrahend has fewer digits than the minuend, before complementing you must add zeros to the left of the subtrahend until it is the same length as the minuend.) Add the two numbers, clear the row by halving up, then transfer the counter at the extreme left to the extreme right. Clear again if necessary. To use the preceding example, 1010011 is changed to its complement 101100. Adding and clearing produces 10011000. Shifting the counter from left to right gives 11001, or 25, the correct difference.

Multiplication is delightfully easy. Napier explains it with the example $19 \times 13 = 247$. One number, say 19, is indicated below the board by marking the proper columns, and the other number, 13, by marking the proper rows. A counter goes on each cell at the intersection of a marked row and column [*see "a" in illustration on page 108*]. Every counter not on the column at the extreme right is moved diagonally up and right (like a chess bishop) until it is on the rightmost column. The result is shown in *b*. The sum of the values of these counters (as indicated on the right margin) is 247, the desired product, but we wish to express it in binary notation. That is quickly done by halving up until the column is cleared. The final result is 11110111, or 247.

It is easy to see why it works. Counters on the first row keep their values when moved to the right, counters on the second row double in value, counters on the third row quadruple in value and so on. The procedure is equivalent to



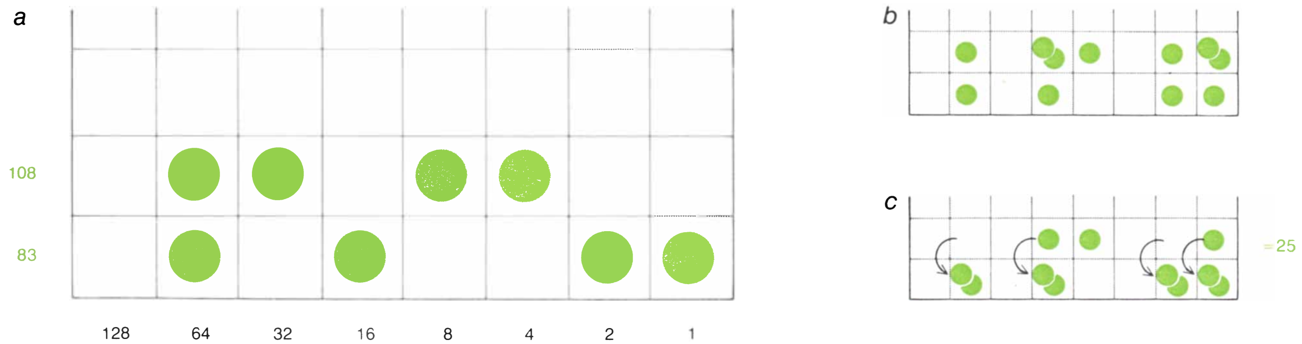
Binary addition: 89 + 41 + 52 + 14

multiplying with logarithms to base 2. In our example, 19 is expressed as $2^4 + 2^1 + 2^0$ and 13 as $2^3 + 2^2 + 2^0$. Cross multiplying in the familiar manner (remembering the basic law of exponents: $x^n \times x^m = x^{n+m}$) yields $2^7 + 2^6 + 2 \cdot 2^4 + 2 \cdot 2^3 + 2^2 + 2^1 + 2^0$. This corresponds exactly to Napier's procedure. Indeed, moving the counters is equivalent to cross multiplying. We are, in effect, multiplying by adding exponents.

Napier was not the first to recognize that powers of 2 can be multiplied by adding their exponents. As early as 1500 it had been clearly explained with exponential notation by Nicolas Chuquet, a physician of Lyons, in the algebraic part of his *Triparty en la sciences des nombres*. It is Napier, however, who gets the credit for the first mechanical device operating with logs with the base 2.

Napier next explains how to do long division on his abacus, using the example $250 \div 13$. The procedure, as one would expect, is the reverse of multiplication. Complications arise that make it difficult to explain, although in practice one soon learns to do it quickly. The divisor, 13, is marked at the bottom of the board, and the dividend is indicated by counters on the column at the extreme right [*see "a" in top illustration on page 109*]. You must now move the dividend counters like chess bishops, down and left, to produce a pattern that has counters (one to a cell) only on marked columns, and each marked column must have its counters on the same rows. Only one such pattern can be formed, but to do so it is necessary at times to double down on the right column, that is, remove single counters, replacing each with a pair of counters on the next lower cell.

Start with the top counter and move it diagonally to the leftmost marked column. If you see that you cannot proceed to form the desired pattern, return the counter to its original cell, double down and try again. If the first attempt fails, the second will succeed in beginning the required pattern, although more doubling down may be necessary. Continue in this manner, doubling down whenever you see that you must, gradually filling in the pattern by extending it down and right until finally the unique



Binary subtraction: 108 — 83

pattern is constructed [*see "b" in top illustration on opposite page*]. After the final counter at the bottom right corner of the pattern is in place you will have three counters left over. They represent the remainder. The rows containing counters are marked on the right margin, symbolizing 10011, or 19, the correct quotient. The three extra counters give the fraction, 3/13.

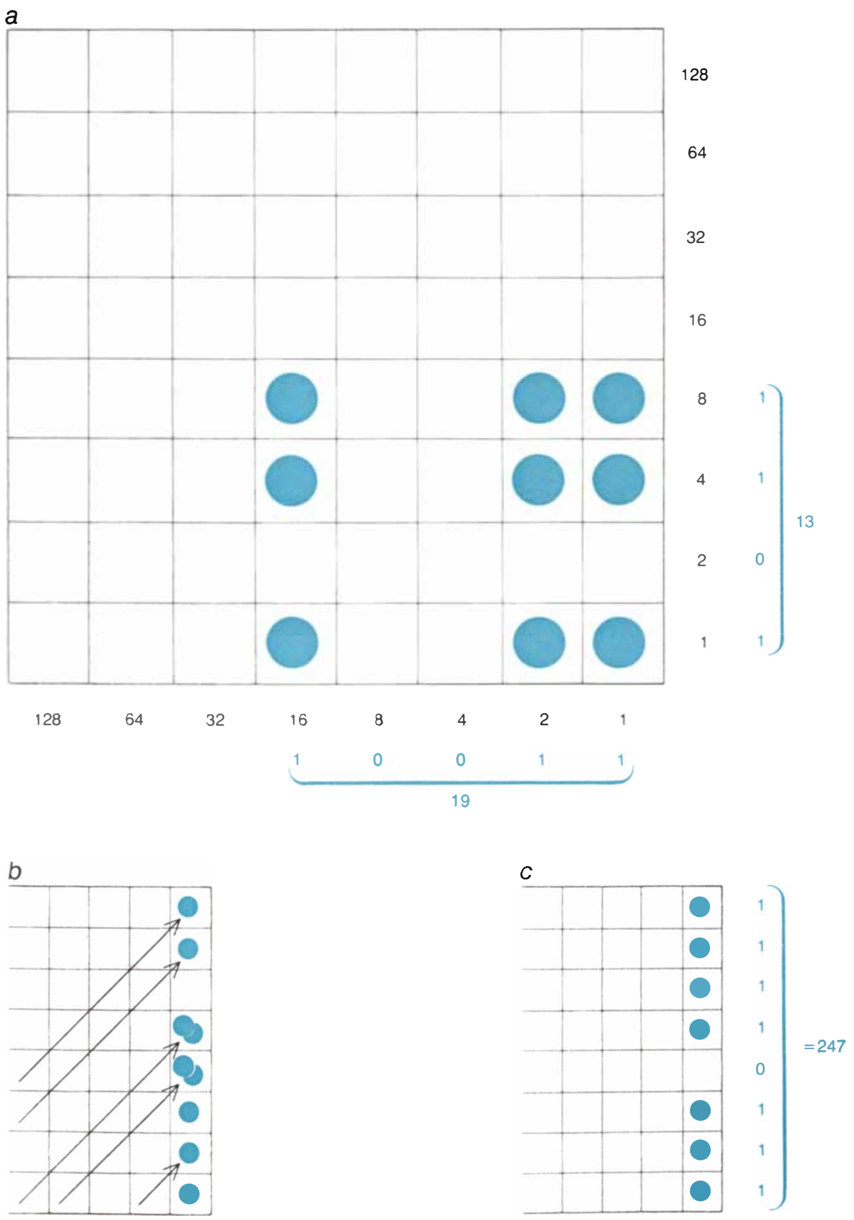
A similar procedure is used to find integral square roots. If the root is not integral, the procedure gives the root of the largest square less than the original number. Counters left over then represent the difference between that number and the original. Napier demonstrates by finding the square root of the largest square less than 1,238. This requires a board extended higher than the standard chessboard. As in division, the number is represented by counters on the rightmost column [*see "a" in bottom illustration on opposite page*]. Since no divisor is marked on the bottom, how do we form a pattern? We must move counters diagonally down to produce a pattern with two properties: (1) every column with counters must have its counters on the same rows, and (2) the pattern must have bilateral symmetry along the diagonal passing through the board's lower right corner. This ensures, of course, that multiplier and multiplicand are identical. As before, start with the top counter and see if you can move it to the diagonal of symmetry. If you can, that is the correct first move. If you cannot, double down and move one of the counters to the diagonal of symmetry. Continue in this fashion, doubling down when necessary, until the required symmetrical pattern is achieved. The result is 35 × 35 = 1,225, with 13 leftover counters that represent the difference between the square and 1,238.

The 15 patterns that generate all squares from 1 through 225 are shown in the upper illustration on page 110. Studying them will familiarize you with the kind of pattern that must be formed for square roots. Note that in every pattern each row and column has a counter on the diagonal of symmetry.

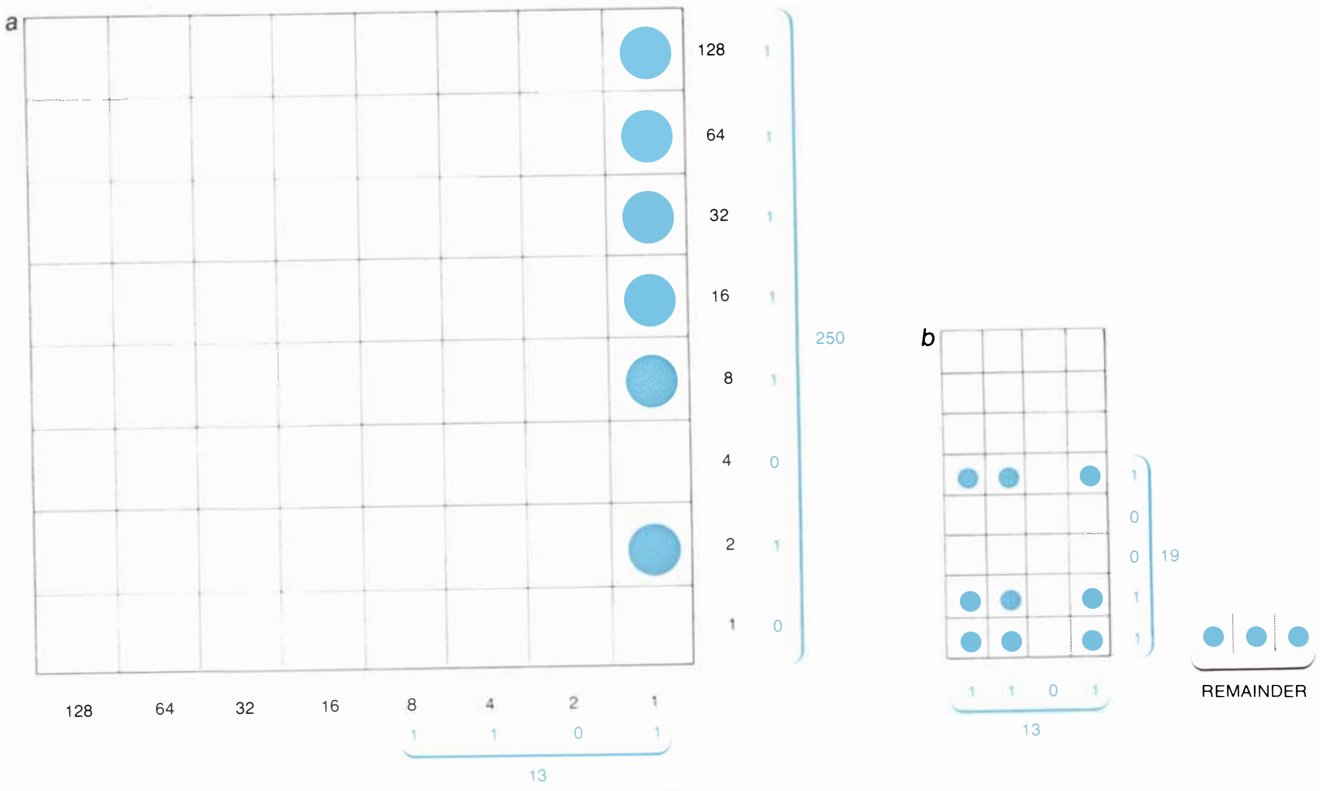
Napier's device will operate with any base notation, but above base 2 it is necessary to work with multiple counters on single cells. As the base increases, the system becomes progressively more cumbersome and uninteresting, and more multiplying must be done in the head. For example, to multiply 77 by 77 in decimal notation each of the four cells at the lower right corner must hold 7 × 7 = 49 counters. After moving them to the right column you have 49 counters on the bottom cell, 98 on the next and 49 on the next. Then every set of 10 counters on a cell is replaced by a single counter immediately above it, resulting finally in counters on four cells that signify the product, 5,929.

The most interesting extension of Napier's board was suggested by Donald E. Knuth, whose latest book on computer programming was reviewed in this column in February. A checkered board can be used very efficiently for calculating in the "negabinary system." Because this remarkable notation is based on powers of −2, the rows and columns of the board are labeled with the series +1, −2, +4, −8, +16, −32, . . . , in which alternate powers are negative. The main virtue of negabinary is that every positive and negative integer can now be uniquely represented in binary notation without the use of signs. Examples are 13 = 11101 (16 − 8 + 4 + 1) and −13 = 110111 (−32 + 16 + 4 − 2 + 1).

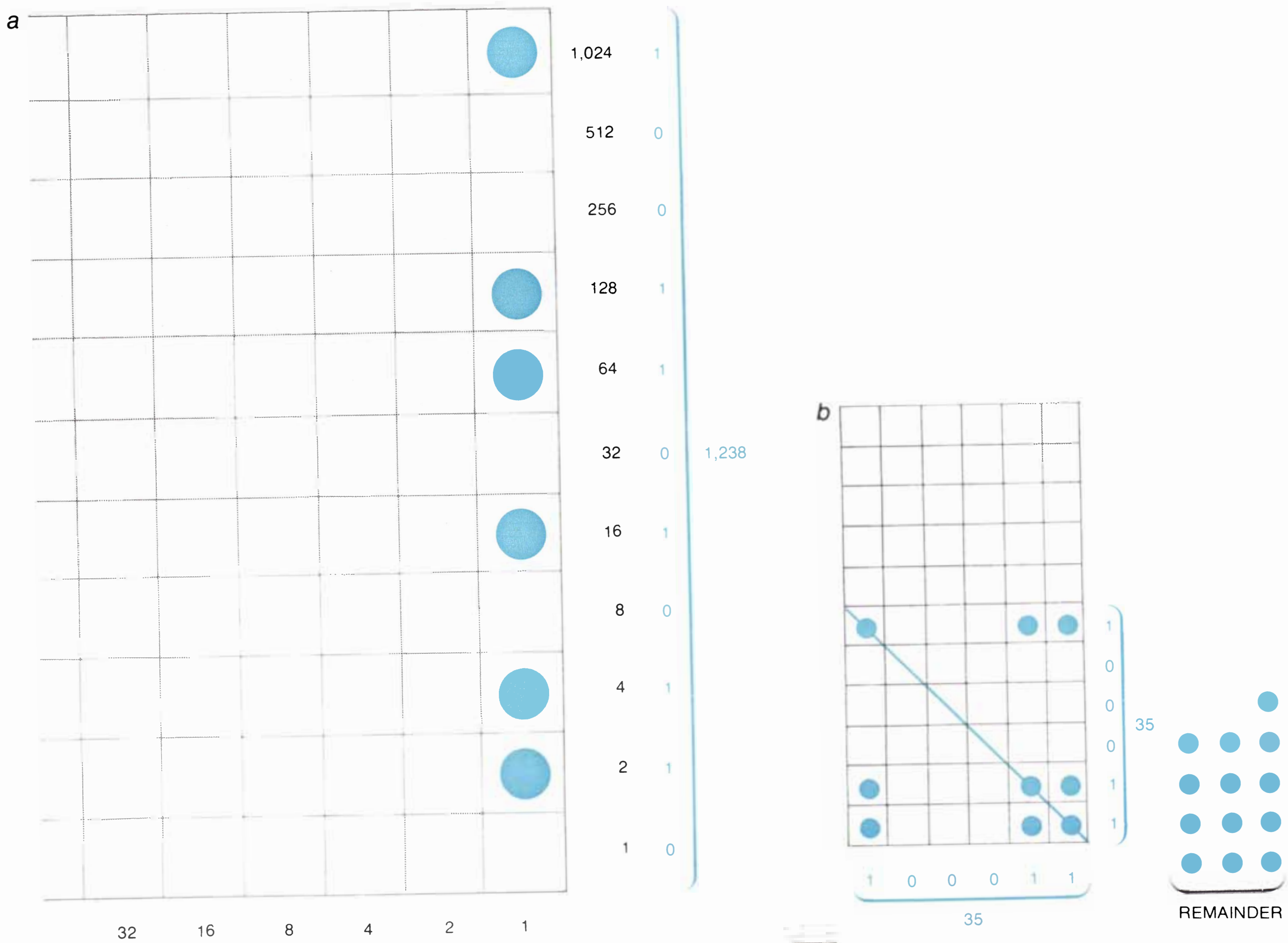
The negabinary forms of positive and negative integers from 1 through 20 are shown in the top illustration on page 111. Note that every positive number has an odd number of negabinary digits and every negative number has an even number of negabinary digits. Every odd number, regardless of sign, ends in 1; every even number, regardless of sign, ends in 0. Many other basic theorems are easily discovered. For example, a negabinary number is divisible by 3 if and only if its number of 1's is a multiple of 3. Observe that every palindromic nega-



Binary multiplication: 19 × 13



Binary division: 250 ÷ 13



Binary extraction of square root: $\sqrt{1{,}238}$

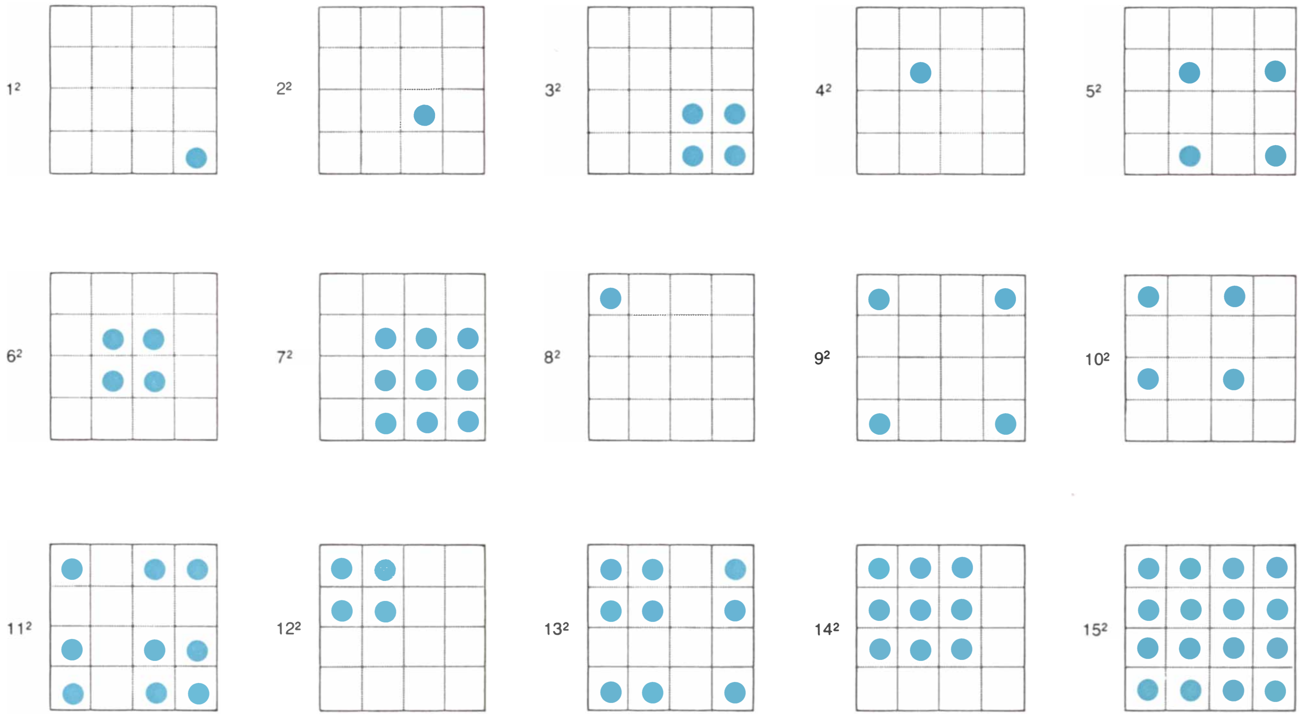
binary number on the list (a number that is the same in both directions) is a positive or a negative prime. Is this true in general? If not, what is the first exception?

I know of no better way to become acquainted with this extraordinary notation (so rich in recreational possibilities) than to calculate with it on Napier's board. Addition is handled exactly as before except that in clearing the sum the following two rules are observed: (1) A pair of counters on one cell and a single counter on the next higher cell cancel one another. Remove all three. (2) If any cells still have double counters, remove each pair and put single counters on each of the *two* next higher cells. The clearing procedure, thanks to the cancellation rule, is unusually rapid [*see lower illustration below*].

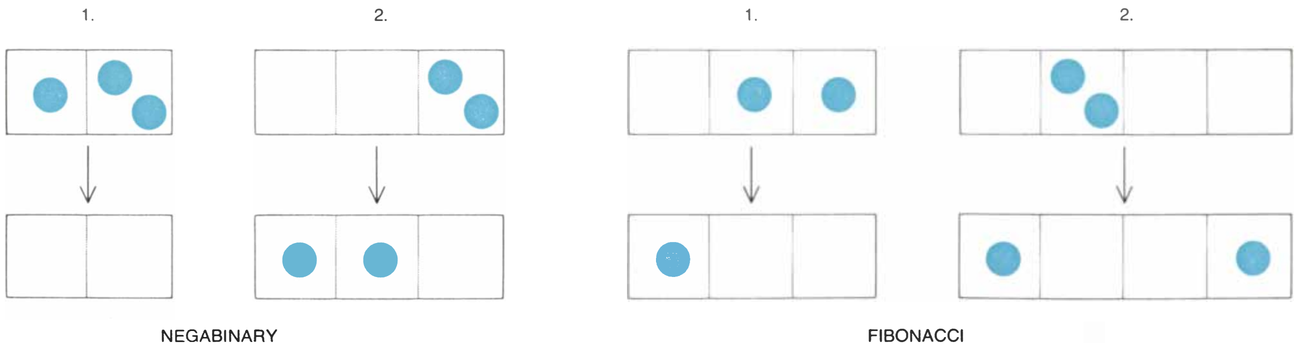
The fastest way to do subtraction is to change the sign of the subtrahend and add! Changing the sign is the same as multiplying by −1, or 11 in negabinary. Since multiplying by 11 is the same as adding a number to itself, with one replica shifted one cell to the left, we can reverse the sign of any negabinary number by the following simple algorithm: add a new counter to every cell that is immediately to the left of a counter originally there, then clear the row as explained. For example, 11 (−1 in decimal notation) becomes 121, but the first two digits cancel (by rule 1), leaving 1, which is positive. Applying the algorithm again restores 11, or −1. When this algorithm is used on standard binary numbers, by the way, it is the same as multiplying by 3. (Do you see why?)

Any two negabinary numbers can be multiplied by using Napier's procedure, then clearing the result according to negabinary rules. The product will have the correct sign when translated into decimal notation. Try multiplying −4 and −6. They are 1100 and 1110 in negabinary [*see bottom illustration on opposite page*]. After multiplying and clearing you get 1101000, or +24. If you had multiplied −4 and +6, or +4 and −6, the result would have been 111000, or −24.

Division and square-root procedures are much trickier, although interested readers should be able to devise them. In finding square roots both positive and negative roots appear as solutions. Are there ways to use Napier's board efficiently for converting a signed binary number to negabinary, and vice versa? Yes; we can exploit two simple algo-



Patterns for squaring 1 through 15



Clearing rules for negabinary and Fibonacci notations rithms given by Knuth as the answer to Exercise 12 on page 177 of his *Seminumerical Algorithms* (Addison-Wesley, 1969). Readers are invited to work them out before they are disclosed next month.

It is hard to believe, but the idea of negative-base notation (it applies to any radix) did not occur to anyone until the 1950's, when many people independently thought of it. In 1955, when Knuth was a high school senior, he wrote a short paper on it for a science talent search, but the first published account (at least in English) seems to be a short letter by Louis B. Wadel in *IRE Transactions on Electronic Computers* (Volume EC-6, June, 1957, page 123). The term "negabinary" was coined by Maurits P. de Regt, whose series of pioneering articles on negative radix arithmetic is listed in "Bibliography" [*page 124*].

Knuth also suggests the Fibonacci labeling, 1, 2, 3, 5, 8, 13, . . . , for Napier's board. You cannot multiply or divide with it, but addition and subtraction can be handled by representing each integer as the sum of the fewest possible Fibonacci numbers. Start by putting a counter on the column with the highest value less than the number to be represented, then work downward until the desired sum is obtained. For example, 19 is uniquely indicated in Fibonacci notation by 101001, or 13 + 5 + 1. The adding procedure is the same as Napier's except that a row is cleared by the following two rules: (1) If single counters are on adjacent cells of the board, remove them and put one counter on the next higher cell. (2) For every pair of counters on the same cell, remove them and put one counter on the next higher cell and one on the *second* lower cell. For example, two counters on cell 13 are replaced by one on cell 21 and one on cell 5 [*see bottom illustration on opposite page*].

If you imagine the row extended two more cells to the right, with values of 1 and 0 (or, alternatively, that the columns are labeled 0, 1, 1, 2, 3, 5, . . .), then the two rules above suffice. Otherwise there are two exceptions. A pair of counters on 2 is replaced by one on 3 and one on 1, and a pair on 1 is replaced by one on 2.

To subtract I know of no better way than the "kinging" procedure explained for binary subtraction. You must, of course, first change the minuend to the required pattern by applying the two clearing rules in reverse. There may be a better method. Indeed, there may be all kinds of clever algorithms for calculating on Napier's board, in various notations, that no one has yet discovered.

The answer to last month's question about Napier's bones is that one set of his original 10 rods will form every multiplicand of 11,110 or less, and two sets will form every multiplicand of 111,111,110 or less.

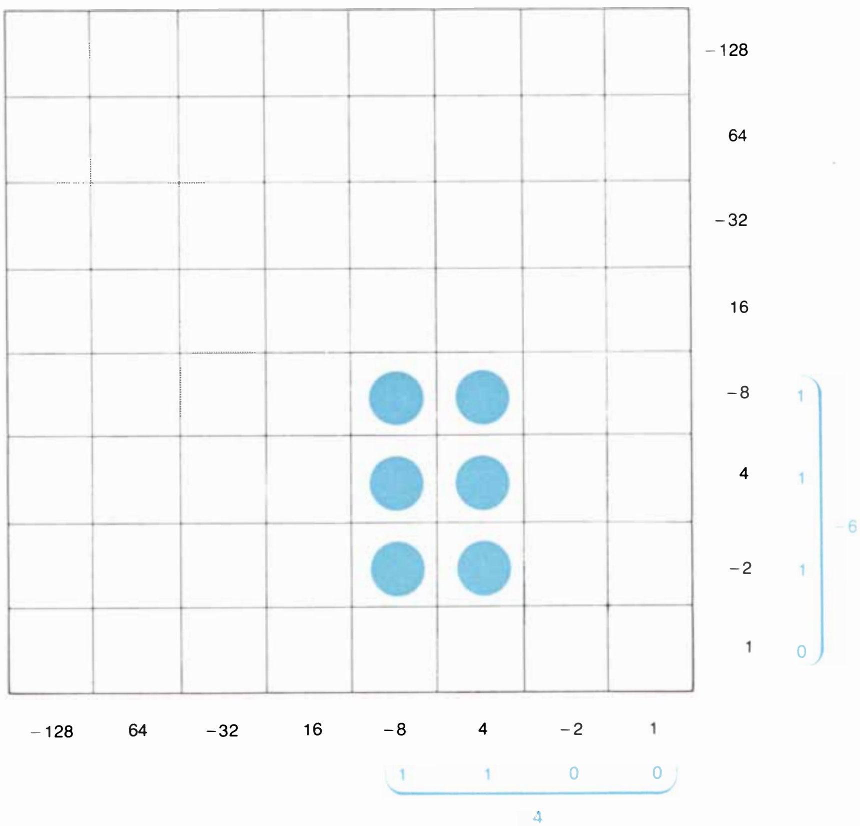
In last September's column on Gray codes I reported that the number of five-digit binary Gray codes (which correspond to the number of Hamiltonian paths on the edges of a five-dimensional hypercube) had not yet been determined. This problem has since been solved by David Vanderschel of Houston, Tex., using a PDP-11/20 computer owned by his employer, Schlumberger Well Services. His results for the subset of cyclic Hamiltonian paths was first confirmed by Steve Winker of Naperville, Ill., and later by Alexander G. Bell and Peter Hallowell at the Rutherford High Energy Laboratory in Chilton, England. The results for "cubes" of n dimensions ($n = 1$ through 5) can be summarized as follows:

The number of Hamiltonian paths starting at the 0 corner and ending on a vertex adjacent to 0: 1, 2, 12, 2,688 and 1,813,091,520. That is the same as the number of ways (including reversals) a fly can start at a corner, make a round trip visiting each corner once and return to where it started. If you wish to obtain the number of such paths starting at any corner, multiply each number by 2^n.

The number of Hamiltonian paths starting at 0, visiting each vertex once only and ending on any vertex except 0: 1, 2, 18, 5,712 and 5,859,364,320. To obtain the number of such paths starting at any corner, multiply each number by 2^n. Thus, under the broadcast definition of "different," the number of five-digit binary Gray codes, including reversals, is 187,499,658,240.

1	1	−1	1 1
2	1 1 0	−2	1 0
3	1 1 1	−3	1 1 0 1
4	1 0 0	−4	1 1 0 0
5	1 0 1	−5	1 1 1 1
6	1 1 0 1 0	−6	1 1 1 0
7	1 1 0 1 1	−7	1 0 0 1
8	1 1 0 0 0	−8	1 0 0 0
9	1 1 0 0 1	−9	1 0 1 1
10	1 1 1 1 0	−10	1 0 1 0
11	1 1 1 1 1	−11	1 1 0 1 0 1
12	1 1 1 0 0	−12	1 1 0 1 0 0
13	1 1 1 0 1	−13	1 1 0 1 1 1
14	1 0 0 1 0	−14	1 1 0 1 1 0
15	1 0 0 1 1	−15	1 1 0 0 0 1
16	1 0 0 0 0	−16	1 1 0 0 0 0
17	1 0 0 0 1	−17	1 1 0 0 1 1
18	1 0 1 1 0	−18	1 1 0 0 1 0
19	1 0 1 1 1	−19	1 1 1 1 0 1
20	1 0 1 0 0	−20	1 1 1 1 0 0

Negabinary notation of integers



Negabinary multiplication: −4 × −6

THE AMATEUR SCIENTIST

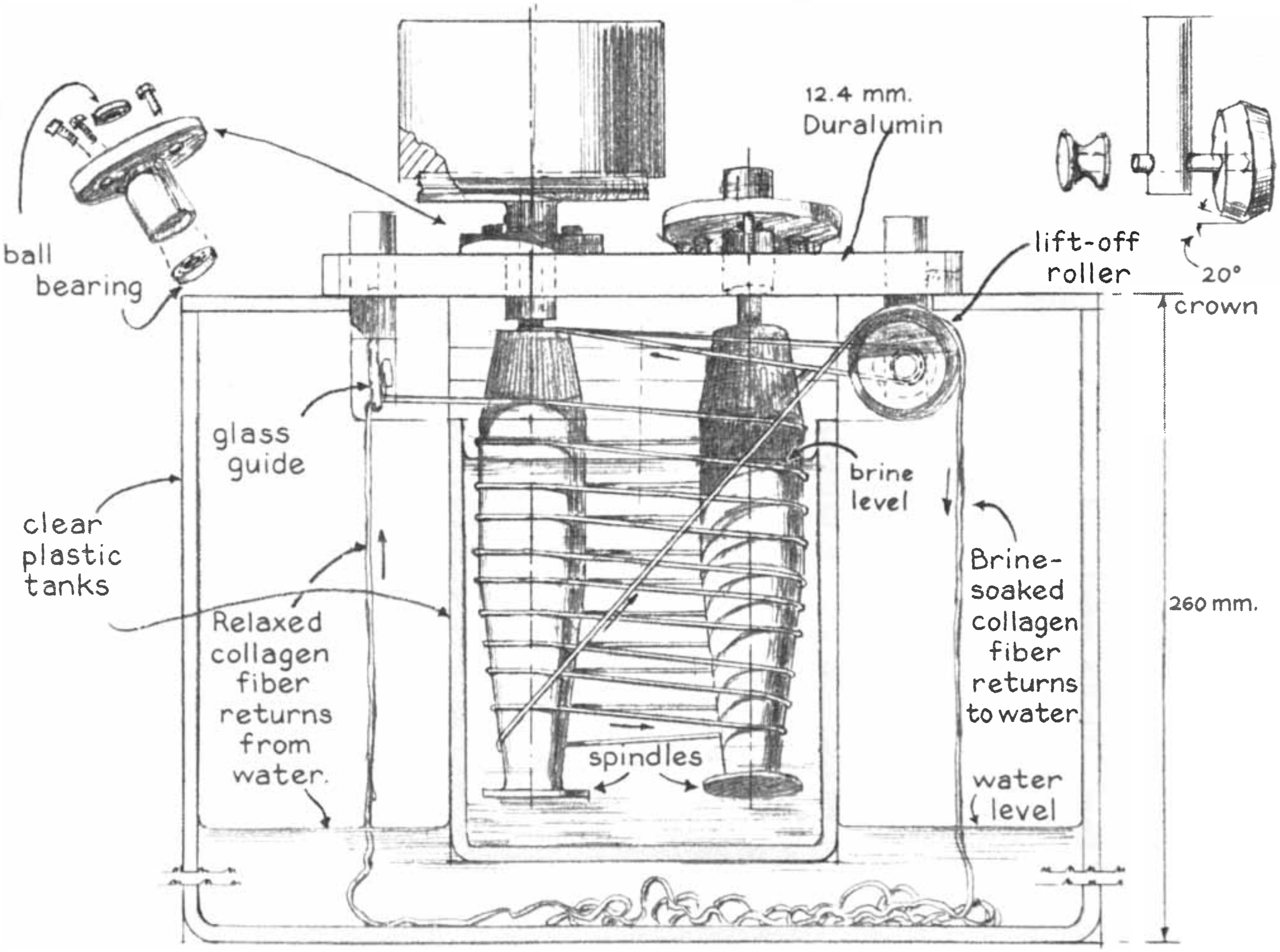
Machines that work like muscles, and how Maxwell's demon was captured in a bottle

Conducted by C. L. Stong

The concept of an engine that would transform chemical energy directly into mechanical work has been pursued by experimenters for many years. Such a machine could be expected to be much more efficient in terms of fuel consumption than are the conventional systems that first convert chemical energy into heat or electricity and then into mechanical work. A one-step conversion could also be expected to produce less waste product.

Engines of this type exist in the form of muscle. An animal's muscles convert almost half of the energy from food into useful work, whereas the best modern steam-turbine systems have an efficiency of about 40 percent. A promising step toward the development of an engine based on synthetic muscle was taken about 20 years ago by a group of investigators at the Weizmann Institute of Science in Israel. The group found that a certain gel resembles muscle in that it contracts and lifts a weight when it is immersed in a solution of appropriate chemicals. Immersion in a different solution causes the gel to relax [see "Muscle as a Machine," by A. Katchalsky and S. Lifson; SCIENTIFIC AMERICAN, March, 1954]. A comparable material is now manufactured in the form of collagen tape. With this material Martin V. Sussman of the Tufts University department of chemical engineering made, while he was a guest of the Weizmann Institute, a mechanochemical turbine that rotates as smoothly as an electric clock. The output power can be taken from a pulley on one of the rotating shafts. Sussman writes:

"The new engine is actuated by a collagen fiber in the form of a loop, a por-



Martin V. Sussman's mechanochemical turbine tion of which is wrapped helically around a pair of contoured spindles that are canted with respect to each other [*see illustration on opposite page*]. The spindles are partly submerged in a tank of salt solution and are free to rotate. The brine serves as the fuel. The remaining portion of the fiber passes through a tank of fresh water below the fuel.

"The salt solution causes the collagen to shrink and thus to develop a tensile force that is relieved as the fiber passes around the rotating tapered spindles. After spiraling through the brine the fiber is pulled from the solution by a belt-driven roller and lowered into the water bath. The diffusion of salt from the collagen into the water restores the material to its original length. Power developed by the turbine pulls the relaxed fiber from the water through a glass ring that guides the material to the top of the spindles for the next cycle.

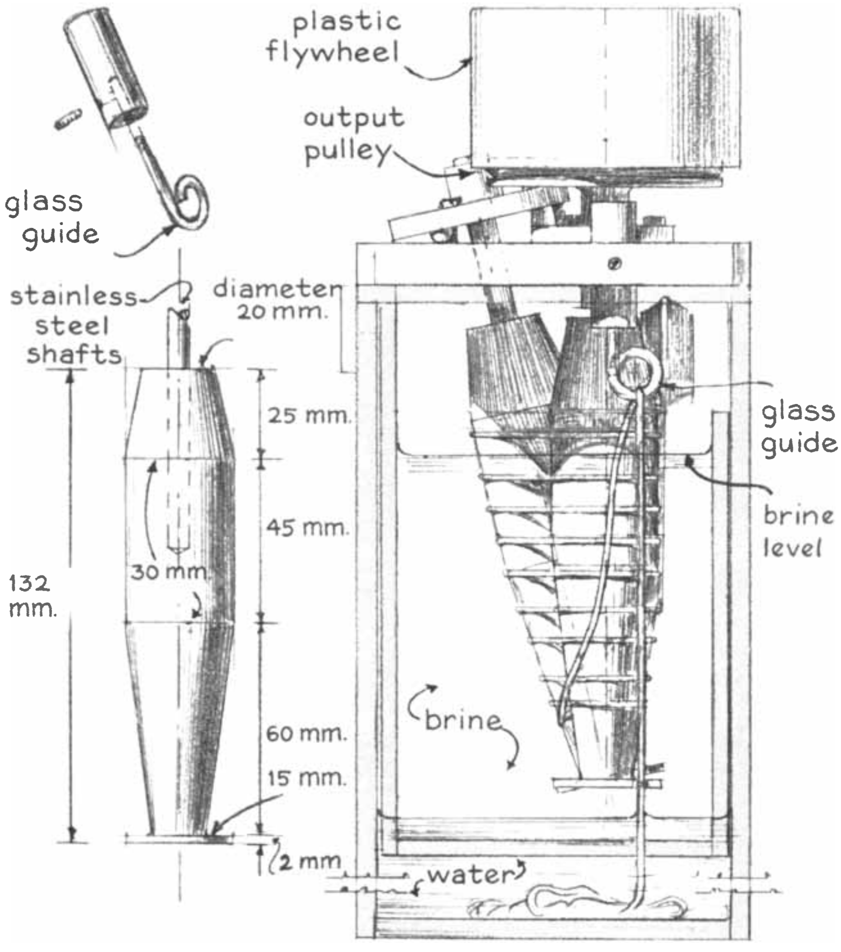
"In its present state of development the turbine generates mechanical power measured in milliwatts at the output shaft. Even so, it demonstrates a basic thermodynamic cycle of some interest. The cycle is outlined by the accompanying graph [*bottom of this page*], which shows how the length and the tension of one centimeter of collagen fiber vary in pure water and in brine.

"In this experiment the brine consisted of an eight-molar aqueous solution of lithium bromide. (A liter of solution at this concentration contains 695 grams of the salt.) The fiber shrinks to about 60 percent of its normal length when it is transferred from a bath of fresh water to a lithium bromide solution of this concentration. Collagen retains its flexibility when it is saturated with salt; it can be stretched to its salt-free length by exerting on its ends a tension of 52 kilograms per square centimeter of cross-sectional area. It is this property that the machine exploits to generate power.

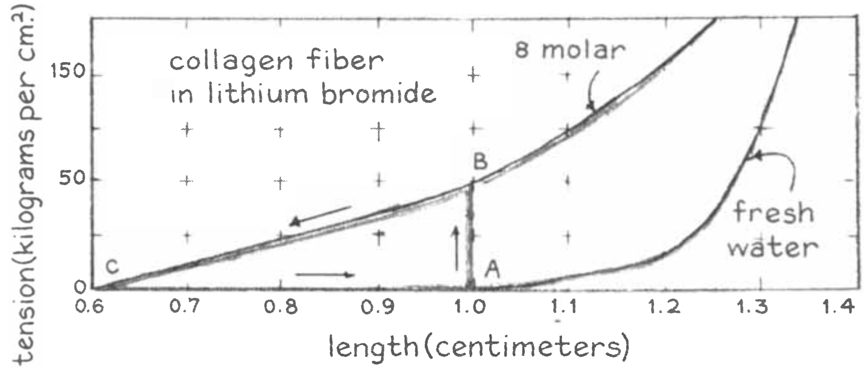
"As shown by the graph, the cycle begins at point *A*, proceeds through points *B* and *C* and returns to *A*. At point *A* salt has been washed from the collagen. The fiber is at its normal length and in zero tension. In this state it is carried to the top of the spindles, which are cylindrical. The cylindrical shapes constrain the fiber so that it cannot shrink. The fiber makes two turns around the cylinders to establish solid frictional contact between the slippery collagen and the surface of the spindles.

"The material is then carried into the salt solution [*point B*]. Tension increases as the fiber absorbs salt. Tension is relieved as the shrinking material spirals through the brine and exerts torque on the tapered part of the rotating spindles [*from point B to point C*]. The taper of the spindles enables the force against which the fiber acts to decrease gradually as tension in the fiber decreases. The arrangement allows efficient release of the tensile energy in the fiber. The cycle is closed by washing salt from the collagen, which restores the fiber to its normal length [*from point C to point A*]. The cycle can then be repeated, although some degeneration of the collagen becomes evident after about 100 cycles.

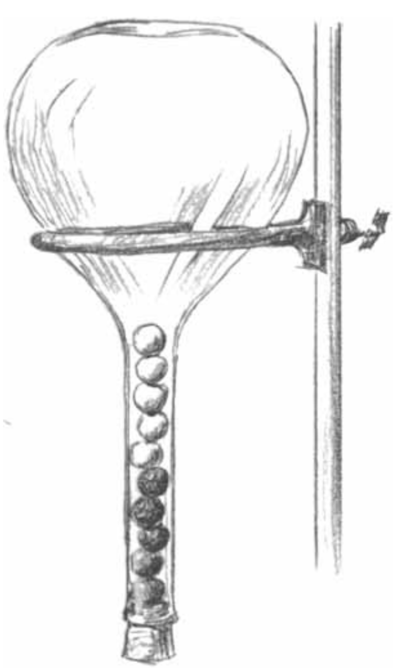
"The maximum power that an ideal turbine of this type can generate is proportional to the area enclosed by points *A*, *B*, *C* and *A* of the graph. For a collagen fiber with a normal dry length of 270 centimeters per gram and a density of 1.3 grams per centimeter the area enclosed by points *A*, *B*, *C* and *A* of the graph indicates an energy conversion of 7.85×10^6 ergs of work per gram of



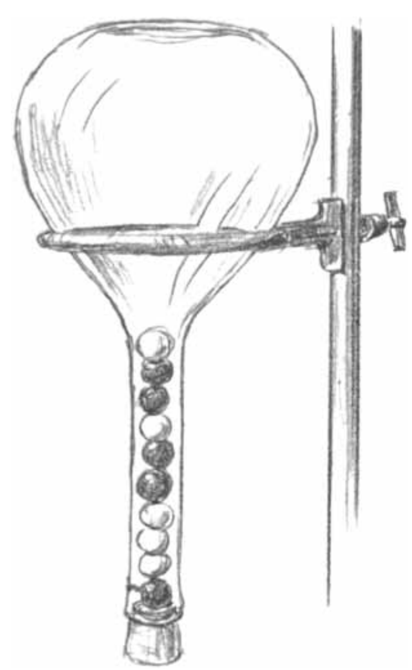
Dimensions of the turbine



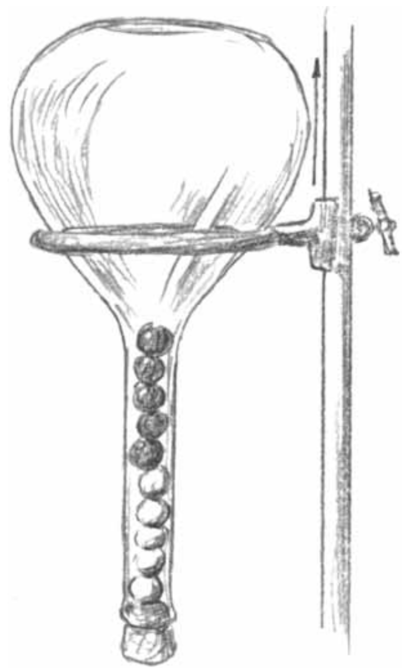
Thermodynamic cycle of the turbine



Demon demonstrator



Intermediate energy level



Highest energy level

fiber. The turbine that was used in the experiments has a collagen throughput of .09 gram per second, which corresponds to an ideal power output of 70 milliwatts.

"The machine delivers about 30 milliwatts at its output shaft, which corresponds to a mechanical efficiency of about 40 percent. A substantial portion of the difference between the ideal energy conversion and the actual output of the machine can be traced to power spent in overcoming friction in the bearings of the machine and in pulling the fiber through the liquids. Another factor that impairs efficiency arises from the transfer of liquid between the tanks of brine and wash water, both of which tend to cling to the collagen. Liquids so mixed do not contribute to the output. Doubtless efficiency could be increased by installing wipers at the points where the fiber is withdrawn from the tanks.

"The output of work per cycle can be increased somewhat by pretensioning the fiber before it is immersed in the brine. Pretensioning can be accomplished by feeding the washed collagen onto short inverted cones at the top of the cylindrical sections of the spindles. Salts other than lithium chloride can be used as fuel, including calcium chloride, magnesium chloride and potassium thiocyanate. I have even operated the turbine on brine from the Dead Sea!

"Any substance that shrinks reversibly or that develops tensile force when it is transferred from one environment to another might be substituted for collagen, although I have not found any other material that works as well. Fibers of crystalline polyethylene or of rubber might be used as the working fiber. In that case the environmental change would have to be thermal rather than chemical.

"The fact that stretched rubber tends to shrink when it is warmed was first observed by the 19th-century physicist James Prescott Joule. A number of fascinating engines have been based on this effect [see "The Amateur Scientist"; Scientific American, April, 1971]. My turbine can be operated on a heat cycle by substituting a fiber of rubber for the collagen and replacing the salt solution with hot water. Output power will vary with the difference in temperature between the beginning and the end of the cycle. In theory it is possible to operate the contraction cycle in reverse; cranking the output shaft should cause the turbine to operate as a chemical pump, which would remove salt from water, or as a refrigeration machine.

"The principal dimensions of the turbine appear in the accompanying illustration [*top of preceding page*]. The working cones of the spindles taper at an angle of 10 degrees. The inverted cones for pretensioning the fiber taper at an angle of 15 degrees. A suitable angle for the axes of the spindles (the cant) is approximately seven degrees. The angle can be altered to adjust the pitch of the fiber helix, which will in turn control the extent of useful contraction of the fiber. The fiber will spontaneously follow a stable helical path when the spindles are rotated. Adjust the angle of the canted spindle so that the fiber contracts completely just before it leaves the helix.

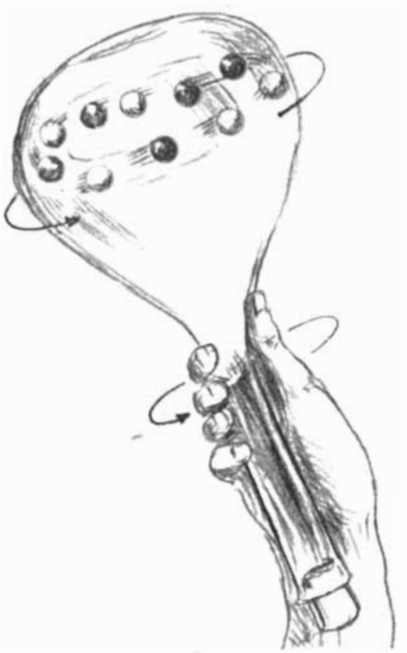
"The collagen I have used is supplied in the form of tape normally employed for making absorbable sutures. The material is untanned. I tan, or cross-link, the molecules by soaking the fiber for 24 hours in a 1 percent solution of formaldehyde at room temperature. The treatment shrinks the collagen slightly.

"In the machine with which I have experimented most I use 32 feet of size 3/0 collagen tape in the form of a loop. The loop is closed by tying the ends with a square knot just before the material is soaked in formaldehyde. The slippery ends can be tied after soaking, but the tying is then best done by someone with surgical skills.

"The spindles were machined from plastic bar stock. They are supported by shafts that turn in high-precision ball bearings of stainless steel. The bearings seat in flanges of stainless steel. The supporting plate, the flywheel and most of the other parts are made from corrosion-resistant metal. It is best to make the tanks of clear plastic so that all parts of the mechanism are visible.

"At present the turbine does not hold great promise as a prime mover, but its deficiency in this respect is partly offset by its attractions as a demonstration apparatus. I am grateful to Ethicon, Inc., of Somerville, N.J., for its contribution of the collagen tape."

Sussman has also devised a scheme for illustrating what Maxwell's demon was supposed to do. The demon was in-



Invoking Maxwell's demon

Demon's work on molecules

voked by James Clerk Maxwell in one of his more puckish moments for a hypothetical experiment relating to apparent differences in the statistical behavior between a gas in bulk and a gas as an assemblage of molecules. The experiment required a pair of adjacent rooms at equal temperature. The partition between the rooms was to contain a small hole. Maxwell proposed to station his demon at the hole with instructions to let all randomly moving fast molecules (hot molecules) pass through freely and collect in one room and to allow all randomly moving slow molecules (cold molecules) to collect in the other room. Without expending physical work the demon would thus cause one room to become hot and the other one to become cold—in clear violation of the second law of thermodynamics, which stipulates that heat always flows from hot regions to cold regions and never in the opposite direction.

Maxwell never succeeded in laying hands on the elusive demon. Sussman, however, offers a trick for imprisoning the demon in a bottle that illustrates various aspects of thermodynamics. Sussman describes the arrangement as follows:

"My demonstrator of Maxwell's demon involves a relatively simple variation of an apparatus long used by teachers to illustrate the nature of irreversible processes. Essentially the apparatus consists of a flat-bottomed flask with a long neck that contains five white spheres and five black spheres [*see illustration at left on opposite page*]. The neck can be sealed shut with a flame or with a stopper. A good stopper is a cork secured with sealing wax.

"Suppose the flask is inverted and the spheres have fallen into the neck so that the five black spheres are on the bottom and the five white spheres are above them. An 'irreversible' process can be illustrated by tilting the flask upright so that the spheres run into the bulb and mix, whereupon the flask is again inverted so that the mixed spheres fall into the neck. The mixing will doubtless alter the initial sequence of the spheres, because there are 252 ways in which five black spheres and five white spheres can be arranged in a row. (The number of possible permutations can be found by expressing the total number of spheres as a factorial number and dividing it by the product of the number of black spheres and white spheres, which are also expressed as factorial numbers: $10!/5! \times 5!$. The factorial symbol ! indicates the product of all integers from 1 through the number of interest. For example, $5! = 1 \times 2 \times 3 \times 4 \times 5 = 120$. Similarly $10! = 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10 = 3{,}628{,}800$. The number of ways in which two groups of five objects each can be arranged in a row is therefore equal to $3{,}628{,}800/120 \times 120 = 3{,}628{,}800/14{,}400 = 252$.)

"When the number of objects increases to several thousand, the possible permutations become so many that the chance of a recurrence of an observed sequence is vanishingly small. That is why it is easy to mix cream into a cup of coffee but difficult to unmix the two. In such a case at least 10^{23} particles of fluid have been intermingled. The number of possible permutations is astronomically large.

"The apparatus can also illustrate the fine structure of energy in a system, showing that the energy carried by myriad bits of matter can change only in discrete steps. Assume that each of the black spheres has unit mass and that the white spheres symbolize massless spacers. Support the flask neck down with the black spheres at the bottom.

"An energy level can now be assigned to each sphere according to its elevation above the closed end of the neck. The bottom sphere cannot fall below its present position. Hence it can do no work. It occupies the zero energy level, which is termed the ground state.

"An energy level can be similarly assigned to the four remaining black spheres. For example, the second sphere in the column could fall a distance equal to its diameter if the bottom sphere were removed. It could do work. One unit of energy is assigned to it. Two units of energy are assigned to the sphere in the next higher position, and so on. This scheme results in a total energy carried by the five black spheres of $0 + 1 + 2 + 3 + 4 = 10$ units.

"Assume now that the spheres run into the body of the flask, mix and return to the neck in a different array [*see center illustration on opposite page*]. In this array black spheres occupy the ground level and the fourth, fifth, seventh and eighth energy levels, so that the new array has a total energy of 24 units. It is of course possible that the white spheres will find their way to the bottom of the neck and be surmounted by the black spheres. This pattern would represent the maximum energy configuration that the system could reach: $5 + 6 + 7 + 8 + 9 = 35$ units.

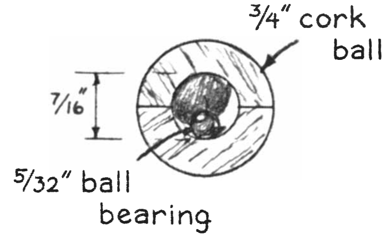
"Engaging now in a flight of fancy nomenclature, let each array be called a 'state.' As I have noted, a system of five black spheres and five white spheres can have 252 states. The energies of the states can range from 10 to 35 units. In the absence of factors prejudicial to one state or another one would expect the average energy of the flask to be 22.5 units.

"The energy of the system can be altered by performing 'work' on the system as a whole. For example, loosen the ring supporting the bottle and slide it and the bottle upward on the stand. All black spheres, since they have mass, gain energy when the bottle is lifted. The effect suggests a tenet of thermodynamics: The energy of a system can be increased by performing work on the system.

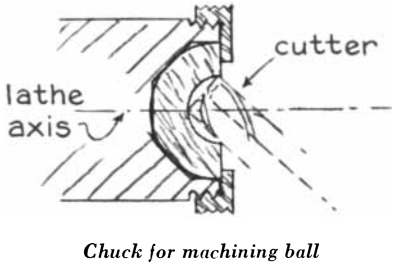
"Just as the energy of the black spheres is increased with respect to the rest of the universe by lifting the bottle, so the work spent in compressing a gas adds to the energy of its constituent molecules. The similarity goes even further, because the work of compressing the gas requires a change in the geometric coordinates (volume) of the gas, just as the work done on the bottle requires a change in the geometric coordinates of the bottle. Continuing the analogy, the effect of increasing the energy of molecules by adding heat to the system can be simulated by altering the probability that a given state of the black spheres will appear. Simply add or remove white spheres.

"In reality the addition of white spheres has two effects. Their presence allows individual black spheres to climb higher into the neck of the flask and so reach higher levels of energy. The white spheres also introduce indeterminacy into the system. Each white sphere that is added increases the number of ways that black spheres and corresponding energy levels can be permuted. Hence the addition of white spheres to the bottle increases both the freedom of the black spheres to move through the accessible energy levels and the difficulty or uncertainty of describing the exact state of the system at any instant.

"In thermodynamics the measure of that uncertainty is known as entropy. With no white spheres in the bottle there is only one possible state for the black spheres. There is no uncertainty and hence zero entropy—a condition analogous to achieving a temperature of



Cross section of black ball



Chuck for machining ball

absolute zero in a perfect crystal. The second law of thermodynamics implies that any physical system left to itself can move only in the direction of increased randomness. In other words, the 'white spheres' of nature permeate and spread, destroying organization, structure and uniqueness. In short, entropy increases with time.

"Having thus illustrated by analogy the irreversible processes of nature, the granular character of energy and the classical concept of entropy, the bottled particles can be returned to their initial state of minimum energy for another demonstration by invoking Maxwell's demon. Grasp the flask at the base of the neck in the neck-up position and swirl the contents with a circular motion of the wrist. The rotating spheres will climb to the middle of the bulb, where they will be held by centrifugal force.

"While continuing the swirling movement slowly turn the flask upside down. Gradually decrease the amplitude of the swirling movement. As the balls lose velocity the black spheres will spiral into the neck, followed by the white spheres. Why? The explanation lies in an inoffensive deception. The black spheres are built so as to lose momentum rapidly. They are hollow and contain a small ball bearing. Friction between the inner surfaces of the sphere and the steel ball dissipates some momentum as heat. The white spheres are solid.

"For the bottle I use a two-liter flask with a long neck; it is available from distributors of chemical glassware. The cork spheres are 3/4 inch in diameter. They are stocked by novelty stores.

"The black spheres are altered by cutting the cork in half with a razor blade and removing material from the center. Improvise a semicircular template of 7/32-inch radius from a piece of sheet metal. With a draftsman's compass inscribe a carefully centered circle on the flat face of each hemisphere. Material within the circle can be removed smoothly and quickly with a high-speed burr 3/8 inch or less in diameter. The inscribed circle defines the limit of the cut. The depth of the cut can be gauged with the template. The high-speed cutter and the motor-driven chuck are available from dealers in craftsmen's supplies. The hollow center can also be cut with a high-speed lathe. Chucks of various kinds for supporting the hemispheres can be improvised from either wood or metal. The complete Maxwell's demon bottle is also available commercially from the Sargent-Welch Scientific Company (7300 North Linder Avenue, Skokie, Ill. 60076)."

WHY MISS OUT?

The important books you want to read and own are yours...in

The Library of Science

How many times have you passed up the books you wanted—only to discover they're impossible to find when you decided to get them? Let The Library of Science bring you the important books you want for reading or reference NOW—books that cover *every* scientific discipline, and all at worthwhile savings. Why miss out?

INTRODUCTORY OFFER:

Take any 3 books **(values to $49.45)**
all only $3.95

if you will join now, and agree to take only 3 more books (at member discount prices) in the next twelve months.

(retail prices shown)

64771. NO LONGER ON THE MAP. *Raymond H. Ramsay.* A delightful voyage to "places that never were." Profusely illustrated with hundreds of old maps, this is an adventure in historical geography. **$10.95**

42520. THE DOOMSDAY SYNDROME. *John Maddox.* The editor of *Nature*—discussing controversial topics such as the fate of our natural resources, ecology, pollution—argues persuasively, with evidence, that the facts may well prove the doomsayers wrong. **$6.95**

48400. THE FASCINATION OF REPTILES. *Maurice Richardson.* Handsomely illustrated appreciation of these scientifically neglected and often maligned animals. Includes information on care and feeding. **$10.00**

57280. KINGDOM OF THE SEASHELL. *R. Tucker Abbott.* Expert introduction to the fabulous world of marine mollusks—how they grow, reproduce and defend themselves. Magnificent illustrations, many in color. **$14.95**

88120. YELLOWSTONE: A Century of the Wilderness Idea. *Ann and Myron Sutton.* A lavish celebration—in evocative words and breathtaking photographs (72 of them in color)—of deer, bison and bears, awesome canyons, the spectacular lakes and forests of this great century-old national park. *Counts as 2 of your 3 books.* **$25.00**

61340. MATHEMATICAL STATISTICAL MECHANICS. *Colin J. Thompson.* A mathematical approach to classical statistical mechanics with important applications to biology. Author is Professor of Physics and Mathematics at Northwestern University. **$12.75**

39891. THE COMPUTER FROM PASCAL TO von NEUMANN. *Herman H. Goldstine.* The world-famous mathematician and computer scientist traces graphically and absorbingly the history of the major scientific and technological development of our times. **$12.50**

63950. THE NATURE OF THE CHEMICAL BOND. *Linus Pauling.* The winner of both the Nobel Prize in Chemistry and the Nobel Peace Prize updates and revises his classic work on the resonance theory. **$17.50**

87410. WORDS FOR BIRDS. *Edward S. Gruson.* This ornithological, etymological and historical guide to 800 birds is a readable treasuretrove of fun as well as accurate information. Illustrations. **$8.95**

63191. THE MOUNTAIN PEOPLE. *Colin M. Turnbull.* Compellingly readable first-hand account of Uganda's strange Ik people—an isolated tribe forced into inhuman cruelty by an inhospitable environment, and by an indifferent government. Photographs. **$7.95**

85830. THE VIEW FROM SPACE: Photographic Exploration of the Planets. *Merton E. Davies and Bruce C. Murray.* A wealth of information supplemented with useful detail on space photography and camera equipment. **$14.95**

38940. THE CLOSING CIRCLE. *Barry Commoner.* The man who has been called "the Paul Revere of ecology" analyzes our steadily worsening environmental dilemma and offers some cogent prescriptions for change. **$6.95**

78340. SHAPES, SPACE, AND SYMMETRY. *Alan Holden.* By word and picture, the author explains, analyzes, and provides amazing insights into hundreds of mathematical solids ranging from the tetrahedron to the trapezoidal hexacontahedron. **$11.00**

50570. GRADUATE SCHOOL IN THE SCIENCES. *Jack G. Calvert, James N. Pitts, Jr., and George H. Dorion.* Three experts from academic life and from industry offer some useful and practical advice for those who want to get into graduate school as well as for those who want to stay in. **$9.95**

48090. THE FACE OF THE DEEP. *Bruce C. Heezen and Charles D. Hollister.* Superb, scientific account of the strange and fascinating portion of the world's surface that is deeply covered by the sea. Beautifully illustrated with more than 600 photographs. *Counts as 2 of your 3 books.* **$25.00**

32571. ALBERT EINSTEIN: Creator and Rebel. *Banesh Hoffmann.* More than a fascinating biography, this book is probably the most rounded and best-balanced picture ever to appear of the great scientist. Many documents and intimate family photographs. **$8.95**

48531. FINITE MATHEMATICS WITH APPLICATIONS. *Abshalom Mizrahi and Michael Sullivan.* Brilliant work that puts such tools as probability, linear programming, matrices, directed graphs, game theory within the reach of the non-specialist. **$10.95**

39960. COMPUTER HARDWARE THEORY. *W. J. Poppelbaum.* Introduction to the theory used in designing computers and their input/output equipment. Covers advanced devices, electro-optics, noise theory, and non-linear circuits. **$16.95**

78190. SEXUAL SELECTION AND THE DESCENT OF MAN. *Edited by Bernard Campbell.* In one seminal volume, ten of the world's outstanding biologists, anthropologists and sociologists assess some of the problems originally posed by Darwin in 1871. Unanimously hailed as a modern masterpiece. **$14.75**

57930. LAWS OF FORM. *G. Spencer Brown.* "Spencer Brown has succeeded in doing what, in mathematics, is very rare indeed. He has revealed a new calculus of great power and simplicity. I congratulate him."—*Bertrand Russell.* **$9.00**

61590. THE MATURE SOCIETY. *Dennis Gabor.* The Nobel Prize winner and discoverer of holography offers prescriptions for the future of man. Raises many controversies and provokes intense and purposeful thought. **$7.50**

47260. EXPLORING NEW ETHICS FOR SURVIVAL. *Garret Hardin.* Frank look at many modern attitudes which affect man's chances for survival. Elaborates on his essay "The Tragedy of the Commons," which burst like a bombshell on the scientific community a few years ago. **$7.95**

37411. CHALLENGING BIOLOGICAL PROBLEMS: Directions Toward Their Solution. *Edited by John A. Behnke.* A number of distinguished contributors examine 17 vital biological problems, including genetic intervention, abnormal growth, aging, the biology of aggression and much more. **$10.95**

62460. THE METAPHORICAL BRAIN: An Introduction to Cybernetics as Artificial Intelligence and Brain Theory. *Michael A. Arbib.* Introduces a new concept of the brain as an action-oriented, distributed computer, based on the latest experimental research and from recent advances in the design of robots. **$14.95**

34630. ASIMOV'S GUIDE TO SCIENCE. *Isaac Asimov.* The incomparable Isaac Asimov has done it again with this exciting general account of modern science—an encyclopedic work of 945 pages that is meant to be read and enjoyed. **$15.00**

34610. ASIMOV'S BIOGRAPHICAL ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY (New Revised Edition). *Isaac Asimov.* A fascinating reference work—the lives of 1,195 of the world's great savants from ancient Greeks to molecular biologists. **$12.95**

MEMBERSHIP APPLICATION

The Library of Science, Riverside, N.J. 08075

2-39Y

Upon acceptance of this order, please enroll me as a member and send the 3 books I have indicated. Bill me only $3.95, plus postage and handling. If not delighted, I will return all books within ten days and this membership will be cancelled.

As a member, I need accept only three more selections during the next 12 months at reduced member prices, plus postage and handling. Savings range up to 30% and occasionally even more. I understand that I will receive free advance Reviews which fully describe each month's Main Selection and Alternates. If I wish the Main Selection, I need do nothing and it will be sent automatically. If I prefer an Alternate — or no book at all — I need only return the convenient reply card you send me by the date specified. Send no money. Members are billed when books arrive.

(Offer good in Continental U.S. and Canada only. Prices slightly higher in Canada.)

3 books for 3.95. (write in numbers). Some expensive books count as 2 choices.

Name ____________________

Address ____________________

City __________ State __________ Zip __________

BOOKS

The man who named the plants and animals, and two studies of glaciers and their geology

by Philip Morrison

The Compleat Naturalist: A Life of Linnaeus, by Wilfrid Blunt, with the assistance of William T. Stearn. The Viking Press ($14.95). The hydra specimen, with seven heads, each gaping mouth set with jaws of horrid sharp teeth, was a young one, perhaps cat size. It is here in a colorful reproduction of an engraving out of the four-volume natural history of Albert Seba of Amsterdam. Seba was a famous traveler and collector, but he had no hydra in his fine collection of curios; the one he published was copied from a drawing sent him by the burgomaster of Hamburg, who owned the rare stuffed monster, once loot from Prague. Although Seba had never seen the specimen, he had no doubt of its genuine importance; his book lists and describes similar wonders. Then Carl Linné, passing through Hamburg, inspected the relic and pronounced it a fake. The claws and gaping jaws were those of weasels; the body was covered with joined, glued snakeskins. Linné (better known today in the Latinized form Linnaeus) was not the first or the only doubter; the price had been dropping steadily ever since the burgomaster had begun seeking a wealthy buyer. No one, however, had been so sharp and convincing a skeptic as the 30-year-old traveler from Uppsala. His tactless clarity brought the price to zero, and he "thought it best to leave Hamburg forthwith" for Amsterdam.

So did the Swedish naturalist begin his foreign travel "in search of fame," as the biographer puts it. Linnaeus was not new to travel. He had made some mark by a summer's journey to Lapland, the northernmost part of Sweden. There Christian missions and a few official posts were scattered through the arctic mountains of a reindeer-herding people; it was a raw land not yet explored by naturalists. Linnaeus spent five months of his 25th year on a collecting trip, for which the Royal Academy of Science had somewhat stingily paid, covering 3,000 miles on foot and by inland waters. His collecting, his eventual systematic volume on the flora of Lapland, his wide and knowing interest in the entire natural world, his clear and lively writing (although the drawings were awkward)—all set the pattern for his long lifework.

No one who goes to the zoo or looks up an identification in herbal or bird book can now miss the name of Linnaeus. His systematic listing and description of plants and animals remain the first resort of the official canon of biological nomenclature. He first systematically employed the terse binomial scheme—each species' name affixed to the name given a genus, or set of related species, like a person's surname followed by a given name, instead of the longish descriptive phrases that had been in general use among botanists. His first version of *Systema Naturae* appeared in Leyden in 1735; it arranged the entire system of nature into its three kingdoms according to classes, orders, genera and species in 14 large, crowded pages.

Today the mineral kingdom has left the rule of Linnaeus, but because resemblance, perspicaciously used, within the evolving world of life means real physical descent, biologists are still Linnaean. His work grew with him; 12 editions came out in his lifetime, the last one running to 2,300 pages in three volumes. It is the 10th edition of this book that remains "the starting-point for the modern scientific naming of animals," as his *Species Plantarum* is for plants. Both books appeared in the 1750's. He included then almost 8,000 species of plants and more than 4,000 of animals. The million and a half or more species we now count are still named in the same way; for plants the first descriptions are still written in the same technical Latin, derived from Renaissance sources, that Linnaeus developed.

Linnaeus was a physician by training and trade, like many a scientist of the day. His chief practice was among the venereally ill young rakes of Stockholm. ("I hardly have a moment to snatch even a hasty meal. This fills my purse but... eats up my time.... Aesculapius is the giver of all good things; Flora bestows nothing on me.") By good luck he fell in with the nobility and at last became professor of medicine and botany at Uppsala, "released from the wretched drudgery." As such he spent his life, writing prodigiously and acquiring a following of students, some of whom became his "apostles": serious botanists. One of these was Daniel Solander, who sailed around the world with Captain Cook and became an official at the British Museum. Another was Carl Thunberg, whose work in South Africa and Japan is classic. Kept isolated, like most foreigners, on a tiny island in the harbor of Nagasaki for most of the 15 months he spent in Japan, poor Thunberg searched daily for plant specimens in the fodder the Japanese brought for the cattle and swine. (The black-eyed Susan is *Thunbergia alata.*)

Linnaeus is for Sweden a kind of Newton. He was heaped with the honors of the learned of Europe. In his prime he would lead a retinue of 150 much-impressed students from many lands out every Saturday to botanize. They would assemble at a set hour, and the master would benignly identify their plants. "A table was spread for twenty, provided with fruit and syllabubs, and those who had found the rarest plants sat with the Master at this table." At the end of the day they all marched back into town with horns and drums, cheered Linnaeus and dispersed. That is the Swedish image of the Prince of Botanists. The English, who came to half-adopt him and who still hold his manuscripts and collections in the Linnaean Society of London, seem to have smothered the man with Victorian sentiment. Oscar Wilde retold the legend that "Linnaeus fell on his knees and wept for joy when he saw for the first time the long heath of some English upland made yellow with the tawny aromatic blossoms of the common furze." Perhaps more revealing is Wilfrid Blunt's glimpse: the clever young man on the make, Enlightenment come from the rude north, full of new ideas, brash, conceited, something of a womanizer, whose costume on great occasions was his Lapland dress, complete with a shaman's magic drum, to set off his stocky frame and "piercing hazel eye" with the glamor of his travels.

Blunt is an artist and a writer; he is at home with botany and its tradition, but he does not spend much space on science, and curiously he does not seem to like Linnaeus, although he handles him fairly enough. The writing is evocative and rich with learning; it supplies a clear look into the times no less than the man and it treats as well, perhaps a little too thoroughly, the opinions of Linnaeus held since his time. William T. Stearn, a distinguished botanist, supplies an excellent, if brief, survey of Linnaeus' work and its influence on modern systematics.

Linnaeus was, to be sure, not beyond his time. He believed, for example, that swallows wintered under water. He knew, however, what evidence is; he marked the species he listed to show whether he had seen the living plant, a dried specimen or only a written account of it. There exist an engraving and a letter demonstrating that he, and not his countryman Anders Celsius, invented the centigrade scale. He had an eye for all nature. He first saw the use of tree rings for dating, silicosis caused by the quarrying of grindstones, holes made by fly grubs in reindeer hides—all new to science. "He adored Nature in all her manifestations; and above all he never lost his sense of wonder."

GLACIER ICE, by Austin Post and Edward R. LaChapelle. University of Washington Press ($20). GLACIAL AND QUATERNARY GEOLOGY, by Richard Foster Flint. John Wiley and Sons, Inc. ($26.75). Mankind and glaciation go together. Today we know that the hominids began some 10 million years ago, and that species ancestral to our own were making and using tools several million years back. The old Quaternary period, the latest geological epoch, a million years old at most and once considered the time of humans and of ice ages, now is too confining. The time of the ice too has been moved back. In the rocks of Antarctica the fossil record shows a warm climate early on, but for the past 10 or a dozen million years we find clear mark of the unstratified deposit of glacier ice: the glacial till. Professor Flint, whose wonderfully readable big text came out the year of his retirement, confides that his title is a little misleading now, but he keeps it out of old associations; it would be a bit pedantic to say late-Cenozoic geology. So the hominids line up in the drawing, from *Ramapithecus* to *Homo sapiens*, getting taller and brainier as the temperature falls.

The main evidence for this cooler climate is the traces of a rare mineral: ice. No solid component of the geological kingdom is so rapidly changing as ice, which is near its melting point even at the coldest spots and is transported in vapor and liquid phases. It flows in big sluggish streams; its most turbulent rapids, the ice falls, advance perhaps a kilometer a year, and occasional explosive runaways, the glacial surges, reach speeds six times greater. Thus it erodes the rock floor, using rock tools like the eoliths of the clever apes; thus it grinds out rock flour, carries strange boulders snatched from scraped cliffs, sends meltwater flowing in lakes and streams to build piles and walls of crushed rock. Ice loads the landmass, deflecting it downward; the melt augments the sea, flooding the coasts. The plant world reflects the cool winds of the icy highlands, and the land and sea animals change in response. From all these lines of argument the geologists have reconstructed the past when Boston and London lay under the ice cap, when the English Channel was dry, when the woolly rhinoceros roamed the steppes of Germany. The changes are swift, as epochs go; during this time the shifting continental plates could not drift far enough to close or open the oceans much.

We can see the ice landscapes today if we travel to the right places. Indeed, the Alpine peasants have watched glacier ice for centuries; they could not have missed the similarity of the moraine across the green valley floor to that along the icy margin of the glacier. Nor did they. In 1815, well before Louis Agassiz visited the Rhône glaciers, a man named J. P. Perraudin orally communicated the inference from the moraines—that once the glaciers had been far more widespread—to a Swiss civil engineer.

Only a few people live near the ice, and almost none live on it. *Glacier Ice* is a joint product of two men who for a dozen years have made aerial photographs, mainly of the glaciers that abound between Puget Sound and Bering Strait, in a systematic study of the details of glacier ice on the kilometer scale that is natural to it and well suited to the airborne camera. (They pay warm tribute to their pilot, William R. Fairchild, killed in a tragic crash in 1969, for the skill that "made these projects possible.") The book is a rather thin volume whose large pages display more than 130 crisp, revealing and beautiful photographs of the ice, mainly from the air but with some from the mountaineer's viewpoint. Some photographs are taken from other sources and show Himalayan or Alpine ice. The text is a masterpiece of scientific pedagogy, fitting and explaining the photographs very well; the book will speak eloquently to the tyro but should not let down the expert, although it stops short of the quantitative. (Flint goes further, but still in an introductory way.)

A few sights insist on description here. The glacier sole is the debris-laden layer at the very bottom of the glacier, its working face against the rock. It is an unlikely subject for aerial photography, but here is an area of the sole, the rock pieces showing strong parallel grooves. The glacier at the head of a long Alaskan fjord had just discharged an iceberg and several pieces from the bottom had floated to the surface and overturned to expose the sole as the photographer above made his exposure; only on the print did he learn of his remarkable subject. Life is scarce on the endless ice but not absent. The mountaineer is there (one dizzying shot shows a pair crossing a knife-edge ridge between two slotlike crevasses falling off into the dark) and the ski-equipped aircraft. Of course, the photographs are themselves the sign of the flying camera crew overhead. There is also the ice worm—no joke at all but a "common inhabitant" of the glacier snow surfaces of early summer along the North Pacific Coast. Like little earthworms an inch long, these animals live in the surface meltwater, feeding on windfalls of pollen or algae. They flee both the sun's heat by day and the freezing cold of night by retreating into the snow; they feed in the evening or in cloudy weather. They overwinter deeper in the snow, creatures adapted to permanent snow and ice.

The most striking of all the manifold forms of the glacier is the ogive band, a regular arcuate transverse waving or banding of the flowing ice, on the scale of tens of yards, convex downstream, forming each year as regularly as summer and winter, below its icefall source. One remarkable view of an Alaskan glacier shows the year-by-year advance plainly marked and dated. The "orderly highways" of the usual medial moraine stripes or the looping "marble cake" of the moraines that form on a surging glacier is visually more striking. No form in cloud or sea, however, is more regularly waving than the close-spaced ogives of some of these glaciers. Their origin is unclear, although a plausible story or two has been spun by the ice-watchers.

The Flint volume, "a compromise between an encyclopedic treatment and a generalized summary," is an introduction to a handful of sciences. You will see a dozen curves marking temperature trend with recent geologic time, all put side by side on a page, to make clear how far these interpretations of oxygen-isotope abundance and diatoms, of pollen and silt layers, agree with one another. The verdict is: Tolerably well. The climate has fluctuated worldwide over the past 70,000 years, north and south together. The glacial ages are temperature-controlled and not times of extra precipitation. Nor are they essentially polar phenomena. During glacial maximum the northern ice sheet grows to double the rather fixed volume of the Antarctic ice; the inventory is here in quantitative estimate. The maximum ice sheet has three times the present ice volume, the bulk of it Canadian and Antarctic; Europe and the Andes have not much room for the ice. Even if the southernmost plateau were at lower latitudes, it would be an ice cap; glaciers are phenomena of continental climates, high altitudes with enough precipitation to feed the flow.

The causes? Will the ice sheet grow again? We do not know. Professor Flint has for a generation made the constant wager that a piled-up highlands geography—such as the colliding continents have built over the past 100 million years—is needed, but is not sufficient, for the ice. Probably, he holds, the fault is not in ourselves but in our star: the sun. Does that furnace dwindle? Perhaps the neutrinos underground will yet tell us the answer. Meanwhile we can admire the ice and its creatures; on a few pages Flint reproduces a marvelous menagerie of the Quaternary, little scaled drawings the size of a thumbnail, of the beasts that lived on the earth while the clever ape struggled in the savanna. Did Siberian hunters, far afield, ever see the giant American browsing ground sloth, feeding in the treetops like a giraffe but standing like a bear 16 feet tall?

LEARN SCIENCE THROUGH BALL GAMES, by C. B. Daish. Sterling Publishing Co., Inc. ($3.95). The British edition of this penetrating and original little book has a more straightforward title: *The Physics of Ball Games.* The U.S. publishers have seen the matter more subtly, and no doubt there is truth on their side, but the direct aim of the book is rather to learn ball games through science. Never mind that all such interactions are reciprocal; the book hits the mark with the skill of a Nicklaus on the approach shot. With succinct tables, almost no formulas and the expertise both of thoughtful theory and a good many experiments, Mr. Daish illuminates the way of bat, paddle and toe with balls from table tennis to soccer. Internal evidence implies that he is associated with the golf-equipment industry; certainly the experiments he cites center there. (The absence of cricket suggests the pruning that seems to have produced the abridged but bargain-priced American edition.)

Ball games have three physical stages: the propulsive (or ending) impacts, the long flight in the air and the contact with surfaces such as walls or playing field. The impact stage is swift: up to a few milliseconds span the contact time, so brief that in a golf stroke "the ball must have left the club face before the hands of the player feel the blow," let alone the signal to his brain. By the time the player is conscious of the impact the ball is two feet ahead of the club. These short times mean large forces. But they are dynamic: they depend on club-head speed and mass, not at all on shaft strength. A neat, if approximate, theory based on momentum balance and the bounciness of the ball shows that an increase in club-head mass does no real good, since the club head already dominates the mass of the golf ball. Indeed, a heavier club is harder to swing. Direct experiment bears this out: an increase by almost a factor of three in head mass does not noticeably affect the length of drives.

The flight through the air is a major portion of any ball game. The crude drag-free approximations of the beginning physics text are not worth much in practice; even a soccer kick reaches only about 210 feet on the field, although in a vacuum it would carry 410 feet. Table tennis is air-dominated. Swerves and curves are air-produced; they are of course real, not illusory. In golf they become the hooks and slices that ruin shots when the deflections are horizontal. When the air gives lift (all golf shots, even the drive, are undercut), it is the golfer's aid, as it is the pitcher's. The drag of the turbulent wake and the sidewise forces of the Magnus effect (the name given the transverse forces produced by the flow pattern around a spinning ball) are carefully considered. They are at the heart of baseball, tennis, golf and high-quality soccer. The drawings of real trajectories and the explanation of their nature give conviction to this excellent section, which is made simple but can repay sophisticated attention. Squash and table tennis are played below the critical Reynolds number, soccer is well above it and golf (with dimpled balls but not with smooth ones) takes place above it as well. This regime is that of the turbulent boundary layer, when the wake is narrowed and drag drops sharply.

Even more subtle are the effects of surface impact. Spin and its frictions, rolling and sliding, rule there. The billiard cue is a tool for producing spin; the slides and swerves and hesitations the real cueman can produce have their analysis here not in mathematical but in plain physical language. It is fair to remind the reader that understanding is not the same as performance; a hustler is made by eye and hand, not by the explicit statement of spin axis and slip. Cue balls given "top" or "bottom" can act surprisingly on impact, but they must collide while still slipping. That control of distance and speed is not for beginners.

Energy and power in the end determine motions. Since games are muscle-powered, a chapter is devoted to human forces and energies. A golfer during the fraction of a second of a full stroke generates about 2.5 kilowatts, comparable to that of a high jumper, and with skill can store about half of the energy in the club head. He needs to bring into play some 10 kilograms of active muscle. Jack Nicklaus put it well: "You hit the golf ball with your legs." Most of the body's muscles are in the legs and thighs.

There is much more in detail and in broad sketch in this fine book. More work remains to be done. One hopes two kinds of reader will come to the text: those who can learn from it for the first time, like any general reader, how to keep mind as well as eye on the ball, and those with the training and wherewithal to elaborate and extend these fascinating insights.

REMARKABLE FLYING MACHINES: A PICTORIAL ACCOUNT OF CERTAIN EXTRAORDINARY AIRCRAFT THAT HAVE BEEN BUILT AND FLOWN SINCE 1783, by Henry R. Palmer, Jr. Superior Publishing Company, Seattle ($12.95). The subtitle makes the contents clear enough. Put another way, these photographs, more than 200 of them (plus one engraving from the prephotographic days of the brothers Montgolfier), exhibit the curious species evolved chiefly to fill unusual niches in the ecology of human flight. A few show rather odd ways to fly—here is a metal-clad "tin blimp" in New Jersey before World War II, and the 1955 Army "flying platform," a rifleman standing on top of an engine with a ducted fan, nothing else. Mostly, however, we see the dodos and the hummingbirds, the fleas and the toucans of the engineering zoology of flight.

Dodo? The underpowered Dornier DO-X flying boat ("more engines than any aircraft ever built before or since") came to New York from her builders in Friedrichshafen in 1930. She took 10 months, what with weather, fires and just not enough zip to take off with all the fuel she needed for those dozen air-cooled Siemens Jupiters: 6,000 horsepower for 52 tons of weight, including two shower baths. (The takeoff power of a recent propeller-driven airliner of similar weight is two or three times greater, with four engines.) The hummingbird? Consider the one and only Wee Bee, which for the photographer's benefit could be carried in the pilot's arms (with some straining). In flight he lay in belly-flop position as though on a sled. (The Wrights too cut drag by prone flight.) Flea? Well, out of the bomb bay of the huge, slow, propeller-driven postwar B-36 bomber there was to have unfolded a trapeze, and off that structure there was to buzz a tiny folding-wing jet fighter, swift and short-ranged, that could protect the mother craft—perhaps. That symbiosis was not a success, although three hook-ons were accomplished. Piggyback flight is neither new nor obsolete: the American dirigibles *Macon* and *Akron* carried up to five fighters each, and a British piggyback scheme flew the Atlantic in 1938, one big flying boat carrying a smaller pontoon plane into the air. The multistaged rocket is the same solution, and the space shuttle plans to work it out on a grand scale with pilots aboard. As for the toucan, finally, various craft remind one of it with their special grotesqueries of form to house radar or bulky cargo; once "even a 49-foot yacht" flew in a Super Guppy, an oversize, hinged modification of a Boeing Stratocruiser.

Most of the artifacts shown are the serious work of professional design teams and not of eccentric inventors, although a few of those appear as well. Flight demands cooperative skills, if not always much common sense, and is too social an enterprise for the odd dreamer. It is impressive to see, though, what the taxpayers have paid for in one or another military project. The inflatable reconnaissance planes of the late 1950's are rather endearing somehow. You take a small engine and some wrappings out of a not very big box, unroll the wing, fuselage and control surfaces made of layers of nylon, inflate and fly. The idea appears to have proved impractical ("due, perhaps, to . . . a proneness to deflation in mid-air if holed"). Uglier, if no less heroic, is the Northrop Flying Ram, a small twin-jet intended to "dive on enemy bombers, slicing off tail assemblies. . . . The wings were built of heavy magnesium plate. . . . This plane went out of control on its first test flight and crashed."

ULTRAPURITY: METHODS AND TECHNIQUES, edited by Morris Zief and Robert Speights. Marcel Dekker, Inc. ($37.50). Once the concept of chemical purity had to be wholly constructive: a sample was only as pure as you knew how to make it. Nowadays we have an absolute standard, almost Platonic; a volume of space that contains only repeated modules of one single molecular structure, or of one repeated lattice unit, with all nuclear species of a single isotopic nature, is a pure sample. Such a sample scarcely exists; this technical symposium surveys the 20-year history of ever closer approaches to the ultimate. The contributors are English, German and American; they represent university and government research centers and the specialized preparative industries from which the best reagents come. The treatment is at a fully technical level, but the topic is so rich in ingenuity and understanding that it deserves some attention from many readers, certainly from teachers and students of chemistry at almost any level.

The alchemy of World War II, dealing with intensely toxic and scrupulously purified new radioactive elements, began the modern trends. The key idea then was to handle materials dust-free, for the protection both of the sample and of the worker. No sample is pure if it contains dust specks drawn from the air, even though they are insoluble and often slip past the ordinary analyst. The filters developed by the Manhattan project and its successors are still the standard: they are elaborate structures of submicron glass fibers. The individual filters are tested by measuring the passage of a carefully controlled organic smoke of known particle size. An approved filter catches 99.97 percent of the .3-micron particles of the smoke. Build the ceiling of the laboratory entirely of such filters, use a perforated grating for a floor and ventilate the room with a vertical draft of air, moving straight from ceiling to floor in smooth parallel flow at about 100 feet per minute. Now you should have a tolerably dust-free room, one with fewer than 100 particles as large as half a micron in diameter per cubic foot. It is routinely achieved, although not without painstaking attention to detail.

Ultrapure substances approach being the "universal solvent," not because of their high corrosiveness but because the almost inevitable small reaction with the container walls is ruinous to purity. The semiconductor industry must produce the purest of products, because the tiny mass of electrons and holes that course through their samples is disturbed by very slight impurities indeed. The chemical apparatus the industry uses is therefore mostly made of vitreous silica, purified silica sand or quartz crystal, electrically fused, with various surface treatments. For the highest grades of such material the silica is prepared artificially by oxidizing some pure silicon compound. This hard, clear, inert glass is perhaps the most broadly resistant engineering material commercially available. If you store acid or water in a small silica flask, you can expect one part per million of silica in solution in 10 years. In a good bottle of resistant normal glass the effect is 10 times faster. For a trace contaminant such as iron there is not much difference between silica and glass, but for a leachable component of glass—sodium—the difference is a factor of 100,000. Ultrapure water is surprisingly reactive; it attacks pure gum rubber "quite severely." Yet the ideal bottle for such water has been found. It is ice! The water is stored cold enough to promote "the formation of ice from pure water around the walls of a vessel." (The expert writing on the preparation of ultrapure water has the reassuring name of Verity C. Smith.)

The initial preparation of pure substances is of course highly specific to the substance. Nearly a dozen classes are treated here: water, alkali metals and their halides, organic solvents, proteins and more. Parts per billion of contaminants give the broad range of practical analysis today, whether for special impurities or for insoluble particulates. Preparation under an inert atmosphere or in a vacuum can remove even gaseous contaminants, which are generally hard to exclude.

This book is expensive in spite of its composition by typewriter; those in the trade are likely to find themselves directly repaid, but these technical chapters are also full of surprisingly broad insights for any browser with a concern for the nature of matter.

INDEX OF ADVERTISERS

APRIL 1973

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

THE TASK OF MEDICINE

MEDICAL INSTITUTIONS AND MODIFICATION OF DISEASE PATTERNS. Walsh McDermott in *The American Journal of Psychiatry,* Vol. 122, No. 12, pages 1398–1406; June, 1966.

AN INTRODUCTION TO SOCIAL MEDICINE. T. McKeown and C. R. Lowe. Blackwell Scientific Publications, 1968.

THE THEORY AND PRACTICE OF PUBLIC HEALTH. Edited by W. Hobson. Oxford University Press, 1969.

DYNAMICS OF HEALTH AND DISEASE. Edited by Carter L. Marshall and David Pearson. Appleton-Century-Crofts Educational Division/Meredith Corporation, 1972.

THE MOLECULE OF INFECTIOUS DRUG RESISTANCE

EXTRACHROMOSOMAL INHERITANCE IN BACTERIA. Richard P. Novick in *Bacteriological Reviews,* Vol. 33, No. 2, pages 210–235; June, 1969.

THE PROBLEMS OF DRUG-RESISTANT PATHOGENIC BACTERIA. Conference of the New York Academy of Sciences, edited by Eugene L. Dulaney and Allen I. Laskin in *Annals of the New York Academy of Sciences,* Vol. 182; June 11, 1971.

MOLECULAR STRUCTURE OF BACTERIAL PLASMIDS. Royston C. Clowes in *Bacteriological Reviews,* Vol. 36, No. 3, pages 361–405; September, 1972.

DRUG-RESISTANCE FACTORS AND OTHER BACTERIAL PLASMIDS. G. G. Meynell. The MIT Press, 1973.

THE BRIGHTEST INFRARED SOURCES

INFRARED SOURCES OF RADIATION. Gerry Neugebauer, Eric Becklin and A. R. Hyland in *Annual Review of Astronomy and Astrophysics,* Vol. 9, pages 67–102; 1971.

THE CONTROL OF THE WATER CYCLE

ON THE GLOBAL BALANCE OF WATER VAPOR AND THE HYDROLOGY OF DESERTS. V. P. Starr and J. P. Peixoto in *Tellus,* Vol. 10, No. 2, pages 189–194; May, 1958.

POLE-TO-POLE MOISTURE CONDITIONS FOR THE IGY. Victor P. Starr, José P. Peixoto and Robert G. McKean in *Pure and Applied Geophysics,* Vol. 75, pages 300–331; 1969/IV.

FEASIBILITY OF A HELIOHYDROELECTRIC POWER PLANT ON THE EASTERN SHORE OF SAUDI ARABIA. M. A. Kettani and L. M. Gonsalves in *Proceedings of the 6th Annual Conference of the Marine Technology Society,* Vol. 2; 1970.

THE WATER CYCLE. H. L. Penman in *Scientific American,* Vol. 223, No. 3, pages 98–108; September, 1970.

ION IMPLANTATION

ION IMPLANTATION STUDIES IN SILICON. L. Eriksson, J. A. Davies and J. W. Mayer in *Science,* Vol. 163, No. 3868, pages 627–633; February 14, 1969.

ION IMPLANTATION IN SEMICONDUCTORS: SILICON AND GERMANIUM. James W. Mayer, Lennart Eriksson and John A. Davies. Academic Press Inc., 1970.

THE APPLICATION OF ELECTRON/ION BEAM TECHNOLOGY TO MICROELECTRONICS. George R. Brewer in *IEEE Spectrum,* Vol. 8, No. 1, pages 23–37; January, 1971.

APPLICATIONS OF ION IMPLANTATION FOR NEW DEVICE CONCEPTS. Billy L. Crowder in *The Journal of Vacuum Science and Technology,* Vol. 8, No. 5, pages S71–S75; September/October, 1971.

THE IMPACT OF ION IMPLANTATION ON SILICON DEVICE AND CIRCUIT TECHNOLOGY. H. G. Dill, R. M. Finnila, A. M. Leupp and T. N. Toombs in *Solid State Technology,* Vol. 15, No. 12, pages 27–35; December, 1972.

THE LUNG OF THE NEWBORN INFANT

POSTNATAL GROWTH OF THE LUNG. M. S. Dunnill in *Thorax,* Vol. 17, No. 4, pages 329–333; December, 1962.

MORPHOMETRY OF THE HUMAN LUNG. Ewald R. Weibel. Academic Press Inc., 1963.

DEVELOPMENT OF THE LUNG. Ciba Foundation Symposium, edited by A. V. S. de Reuck and Ruth Porter. Little, Brown and Company, 1967.

FOETAL AND NEONATAL PHYSIOLOGY. Geoffrey S. Dawes. Year Book Medical Publishers, Inc., 1968.

THE LUNG AND ITS DISORDERS IN THE NEWBORN INFANT. Mary E. Avery. W. B. Saunders Company, 1968.

THE SURFACTANT SYSTEM OF THE LUNG. Emile M. Scarpelli. Lea & Febiger, 1968.

PULMONARY FUNCTION TESTING IN CHILDREN: TECHNIQUES AND STANDARDS. George Polgar and Varuni Promadhat. W. B. Saunders Company, 1971.

GIORDANO BRUNO

LA CENA DE LE CENERI. Giordano Bruno, edited by Giovanni Aquilecchia. Giulio Einaudi Editore, 1955.

THE CRIME OF GALILEO. Giorgio de Santillana. University of Chicago Press, 1955.

BRUNO. Paul Oskar Kristeller in *Eight Philosophers of the Italian Renaissance.* Stanford University Press, 1964.

GIORDANO BRUNO AND THE HERMETIC TRADITION. Frances A. Yates. University of Chicago Press, 1964.

WAS GIORDANO BRUNO A SCIENTIST?: A SCIENTIST'S VIEW. Lawrence S. Lerner and Edward A. Gosselin in *American Journal of Physics,* Vol. 41, No. 1, pages 24–38; January, 1973.

THE ENERGETICS OF THE BUMBLEBEE

THE FERTILIZATION OF FLOWERS. Verne Grant in *Scientific American,* Vol. 184, No. 6, pages 52–56; June, 1951.

BUMBLEBEES. John B. Free and Colin G. Butler. The Macmillan Company, 1959.

TEMPERATURE REGULATION IN THE BUMBLEBEE *BOMBUS VAGANS:* A FIELD STUDY. Bernd Heinrich in *Science,* Vol. 175, No. 4018, pages 185–187; January 14, 1972.

ENERGETICS OF TEMPERATURE REGULATION AND FORAGING IN A BUMBLEBEE, *BOMBUS TERRICOLA* KIRBY. Bernd Heinrich in *Journal of Comparative Physiology,* Vol. 77, No. 1, pages 49–64; 1972.

ENERGETICS AND POLLINATION ECOLOGY. Bernd Heinrich and Peter H. Raven in *Science,* Vol. 176, No. 4035, pages 597–602; May 12, 1972.

MATHEMATICAL GAMES

NEGATIVE RADIX ARITHMETIC. Maurits Pieter de Regt in *Computer Design,* Vol. 6, Nos. 5–11, Vol. 7, No. 1; May–December, 1967, January, 1968.

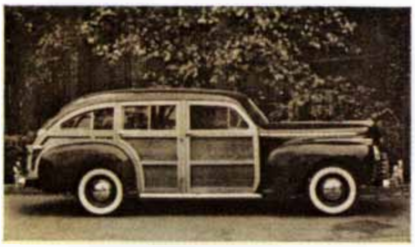
THE AMATEUR SCIENTIST

MECHANOCHEMICAL TURBINE: A NEW POWER CYCLE. M. V. Sussman and A. Katchalsky in *Science,* Vol. 167, No. 3914, pages 45–47; January 2, 1970.

ELEMENTARY GENERAL THERMODYNAMICS. M. V. Sussman. Addison-Wesley Publishing Co., 1972.

The Chrysler Town & Country Wagon was created to be the most luxurious wagon in the world.

That idea has never changed.



The original Chrysler Town & Country wagon was introduced in 1941. It was one of the finest wagons of its day, offering such luxuries as deep-cut pile carpeting in the rear seat, an automatic transmission, even an electric clock and directional signals, all as standard equipment.

Beyond the shadow of a doubt, it was the most luxurious wagon money could buy.

Today's 1973 Town & Country follows this same tradition of luxury. It's big. And comfortable. With such standard features as a powerful 440 cubic inch, four-barrel V-8, TorqueFlite automatic transmission, power steering, power front disc brakes, and a three-way tailgate.

In front, there is a split-bench seat with a center armrest and passenger recliner. All of the seats are upholstered with rich vinyl, and cushioned with thick foam padding.

The floor is covered wall-to-wall with deep-loop pile nylon carpeting. Including the entire cargo area. And the Chrysler Town & Country has more room for cargo than most other station wagons in the world.

If this sounds like a wagon that comes up to your standards, see your Chrysler-Plymouth dealer for a test drive.

You'll be surprised at how easy it is to own the worlds most luxurious wagon for a little more than the price of an ordinary wagon.

CHRYSLER
Plymouth

Chrysler Town & Country
Extra care in engineering...it makes a difference.

Low insertion. Low extraction. Low cost.

Integral cantilever springs inside each socket keep insertion and extraction forces to less than one pound per circuit, on .165-inch centerlines. And contact termination by automated machinery at rates of 2,000 to 12,000 an hour ensures low applied cost.

Add to these lows the high versatility of connector-to-connector mating and connector-to-pc header mating, and you have quite a connector. It's available in 1 through 36 positions, in nylon housings with optional hinged strain relief and tin or gold plated contacts. Contacts accept #26 through #18 AWG wire and are rated to 10 amp.

Internal cantilever springs provide each socket with excellent wiping action and pin retention, yet require less than one pound of force for pin insertion and extraction.

For more information on these reliable miniature rectangular connectors, just write AMP Industrial Division, Harrisburg, Pennsylvania 17105.

AMP
INCORPORATED