SCIENTIFIC AMERICAN

JUDAISM AT THE TIME OF CHRIST

ONE DOLLAR

January 1973

IF WE GOT A CALL TODAY AT 3PM TO BUILD A MID-OCEAN TERMINAL, WE COULD GO TO WORK ON IT AT 3:05.

It's night. And an English container ship slips out of its berth for America. That same night an American container ship sails for England.

But part of the English ship's cargo will never reach America. And part of the American ship's cargo will never reach England.

Instead, both ships will dock in the mid-Atlantic.

At a mid-ocean transfer terminal.

Here's why.

Both ships were loaded to capacity, but some of their cargo was bound for other destinations. So they off-load some containers and take on others.

Then each ship continues on, completely filled with cargo for one specific destination.

Think of it. Mid-ocean terminals for all the oceans on earth. And we could start to build them today.

Look at it this way.

We've been innovators in marine systems and shipbuilding for nearly a century. And today, we're the only shipyard that builds all three types of Naval nuclear ships. Submarines. Frigates. And carriers.

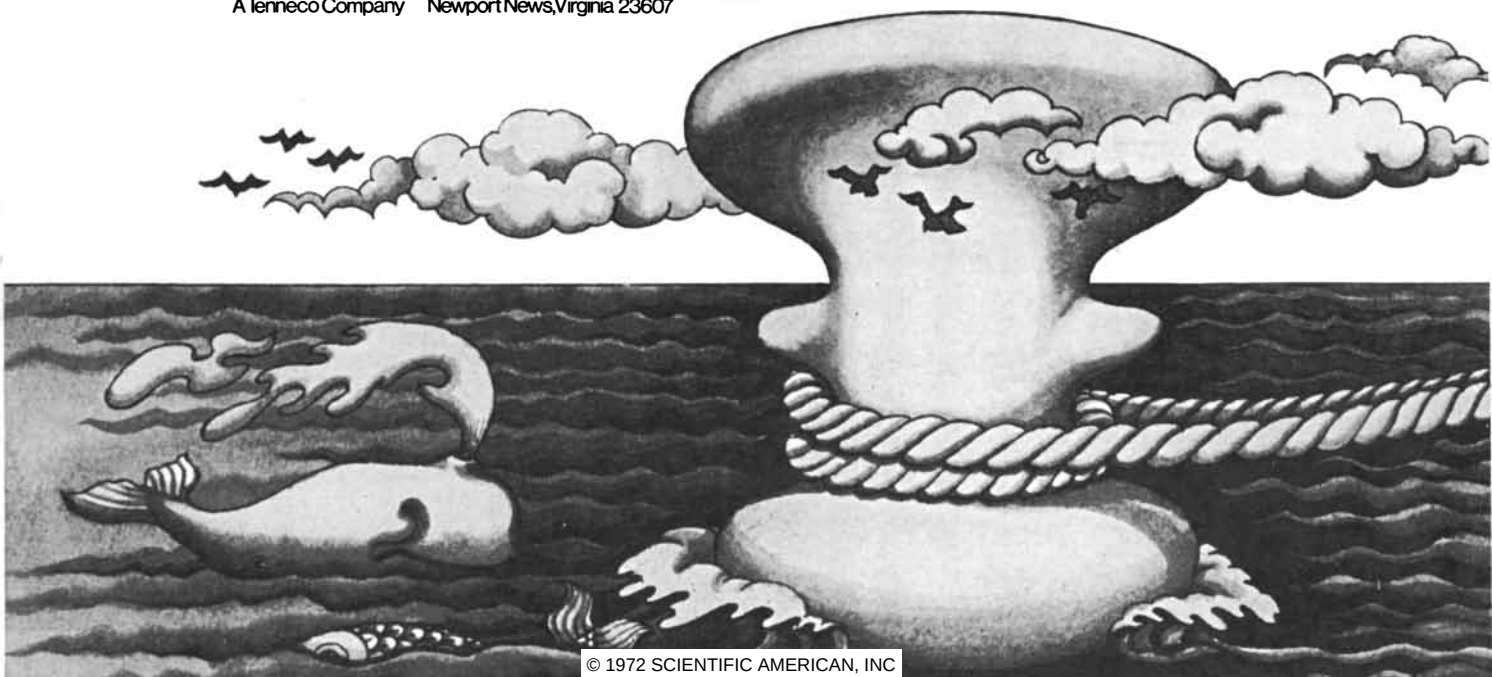
And our interest and expertise in commercial markets continues to expand. Especially in the marine and energy industries. From super oil tankers and high-technology liquefied natural gas carriers to components and services for land-based nuclear power plants.

So you see, we have the experience and confidence. Experience and confidence to go to work on tomorrow's marine systems. Right now.

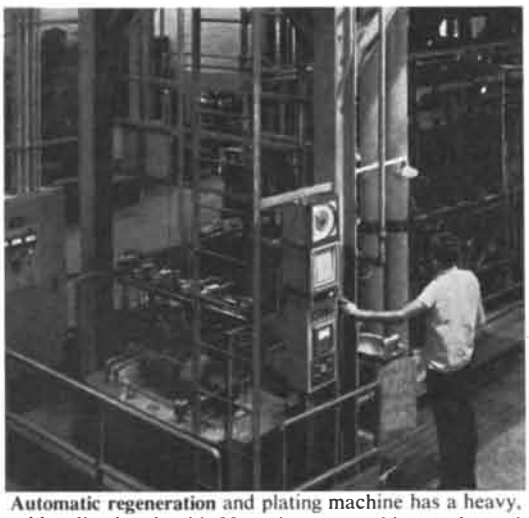
Today the world has plenty of seaports. And tomorrow there will be plenty of ports at sea.

Newport News Shipbuilding TENNECO **Where tomorrow is. Now.**

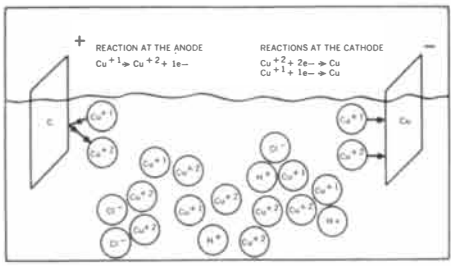
A Tenneco Company Newport News, Virginia 23607



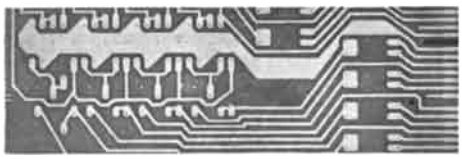
WESTERN ELECTRIC REPORTS



Automatic regeneration and plating machine has a heavy, rubber-lined tank with 22 stationary graphite anodes and 57 rod-like copper cathodes moving at the rate of 90 transfers an hour.



The action at the cathode. Electrochemical reversal of the etching reaction effecting etchant regeneration and copper recovery.



Typical printed wiring board consists of copper (only 0.0028 inch thick) laminated to a phenolic-resin panel. With the new process, unwanted metal is etched away with cupric chloride.

Creating an entirely new way to etch printed circuits.

One of the most common methods of printed circuit manufacturing is by batch-etching with ferric chloride. However, while batch-etching produces circuits of high quality, it also has some processing disadvantages.

For instance, it takes more and more etching time as the etchant is used. Then, to replace the spent etchant means considerable downtime. And the etching of 100,000 circuit boards produces 2000 pounds of copper in a non-recoverable form.

Engineers at our Columbus, Ohio plant set out to discover a better way to etch that would eliminate all of these inherent problems.

Their new process is the first closed-loop, spray-etching system that electrolytically reverses the chemical reaction of etching. It continuously recycles cupric chloride and has reduced the cost of etching wiring boards by over 90%.

Virtually all the problems of the old method have been overcome. No more machine downtime is required to change etchant. No more costly ferric chloride is needed. Etchant strength does not diminish. The etching rate is now constant and faster than the average ferric chloride rate. There's no more waste of etched copper. It is now recovered, about 20 pounds per hour, and resold.

Conclusion: The first completely closed-loop cupric chloride etching system in the printed circuit industry is a major innovation that has improved efficiency and quality, eliminated downtime and decreased costs by more than 90%. Furthermore, it has helped conserve a valuable natural resource.

Western Electric

We make things that bring people closer.

Established 1845 **SCIENTIFIC AMERICAN** January 1973 Volume 228 Number 1

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $10 PER YEAR, U.S., POSSESSIONS AND CANADA; $12.50 PER YEAR, ALL OTHER COUNTRIES. EUROPEAN SUBSCRIPTION COPIES DISTRIBUTED C/O INTERNATIONAL DISTRIBUTORS, VLAGSTRAAT 12B, BORGERHOUT 2200, BELGIUM.

THE COVER

The photograph on the cover shows the central panel of a large mosaic on the floor of a synagogue dating from A.D. 520. The synagogue was excavated in the 1920's at Beth Alpha in Palestine. In spite of the biblical injunction against the making of graven images, the central circle shows the pagan sun-god Helios driving a chariot. The four horses show only their faces, front and forelegs, and two wheels of the chariot are shown frontally at the bottom. Helios has an amazingly complicated halo, first an inner zigzag, next a white band, then a gold band and then seven red rays with gold trapezoids between them. Around the inner circle are the 12 signs of the zodiac, each named in Hebrew. In the corners are the four seasons, but they are not set opposite the correct zodiacal signs. Spring in the upper left corner is beside the signs of summer; winter in the upper right corner is beside the signs of spring, and autumn in the lower right corner is beside the signs of winter. The misplacing of the seasons suggests that members of that Jewish congregation must have had little interest in astrological reckoning. The significance of pagan symbols in Jewish religious art is discussed by Michael E. Stone in his article "Judaism at the Time of Christ" (*page 80*). The photograph was provided by Frank J. Darmstaedter of the Photographic Archive of the Jewish Theological Seminary of America.

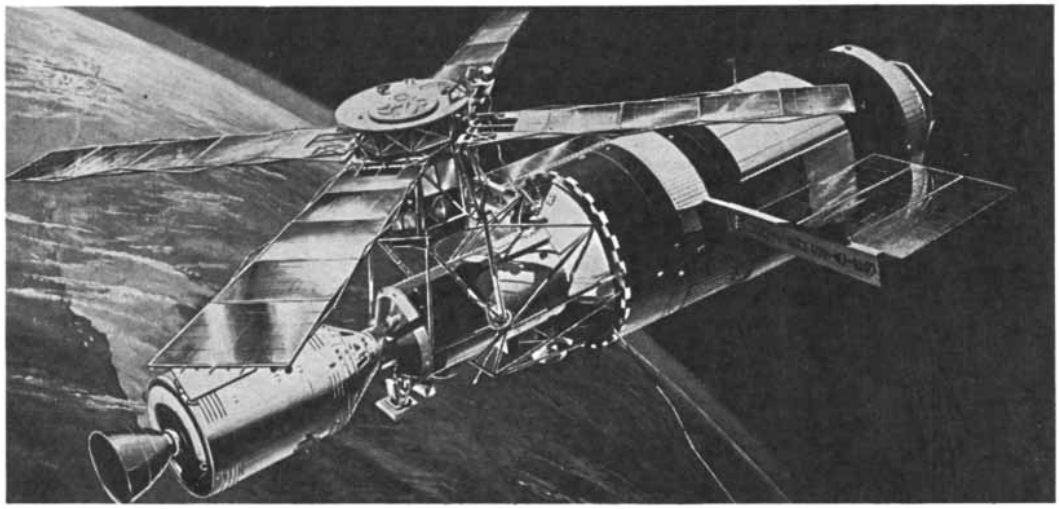
THE ILLUSTRATIONS

Cover photograph by Frank J. Darmstaedter

Page	Source
14–16	Andrew Tomko
17	Chicago Bridge & Iron Company
18–19	Andrew Tomko
20	Cominco, Ltd.
21	Andrew Tomko
23	David D. Porter, University of California Center for Health Sciences
24	Robin Ingle
25	Andrew A. Monjan, Gerald A. Cole and Neal Nathanson
26–27	Robin Ingle
28–29	Robin Ingle (*top*), Donald Henson (*bottom*)
30–31	Robin Ingle
32	Geza Teleki
34–35	Ilil Arbel
36–37	Geza Teleki
38–41	Ilil Arbel
42	Geza Teleki
48–52	Jet Propulsion Laboratory
53	Dan Todd (*top*), Jet Propulsion Laboratory (*bottom*)
54–55	Dan Todd
56–57	Jet Propulsion Laboratory
58	Jet Propulsion Laboratory
59	Jet Propulsion Laboratory (*top and bottom*), Dan Todd (*middle*)
60–67	Jet Propulsion Laboratory
68	Dan Todd
70	Frank S. Werblin
72	John E. Dowling (*top*), Bunji Tagawa (*bottom*)
73–79	Bunji Tagawa
81	Frank J. Darmstaedter, Jewish Theological Seminary of America
82–85	Enid Kotschnig
86	Frank J. Darmstaedter, Jewish Theological Seminary of America
89	Tom Prentiss
90–97	Allen Beechel
98	Tom Prentiss
101	Teisaku Kobayashi, Institute of Low Temperature Science, Sapporo, Japan
102–104	Jim Egleson
105–106	Charles and Nancy Knight
108–113	Ilil Arbel
116–120	Roger Hayward

Skylab, the world's loftiest photo studio.

Inside, Nikon cameras study invisible airglow, measure the sun's corona, and watch a spider at work.



Space photography has come a long way since its early days when just bringing back a picture seemed quite an achievement. The photographic tasks assigned to Project Skylab in particular are far more diverse and exacting than any that came before.

That means more sophisticated equipment. Versatile enough to handle the variety of assignments planned, yet easy enough to permit a non-expert to obtain the needed picture quality. In short, 35mm equipment. Which means Nikon, the most sophisticated 35mm system available.

Nikon cameras had already demonstrated their capabilities in previous missions, including Apollo 15, 16 and 17. These were essentially standard Nikon Photomic FTN models, modified to fit the conditions encountered in a vacuum. Special materials were used to prevent molecular changes (outgassing), new means of lubrications devised, some controls adapted for easier handling with bulky gloves.

Thus when NASA decided on 35mm equipment for Skylab, the Nikon system was able to take it in stride. It was quite a challenge, though, considering that subject matter ranges from the infinite to the infinitesimal.

Experiment SO 63

Purpose: To record ultra-violet emission from the airglow layer of the upper atmosphere. Equipment: Two motor-equipped Nikon cameras controlled by a specially designed Nikon interval timer. One camera is fitted with a multiple-layer coated Auto-Nikkor 55mm f1.2 lens for recording visible light, the other with a newly designed 55mm f2 Ultra-Violet Nikkor lens (the first of its kind and, like all Nikkor lenses, made from Nikon optical glass). The latter is mounted on a track enabling it to follow the curvature of the earth and providing automatic, simultaneous triggering of the conventional-lens camera at a predetermined point. Photos from both will be matched in evaluating the results of this experiment.

Experiment TO 25

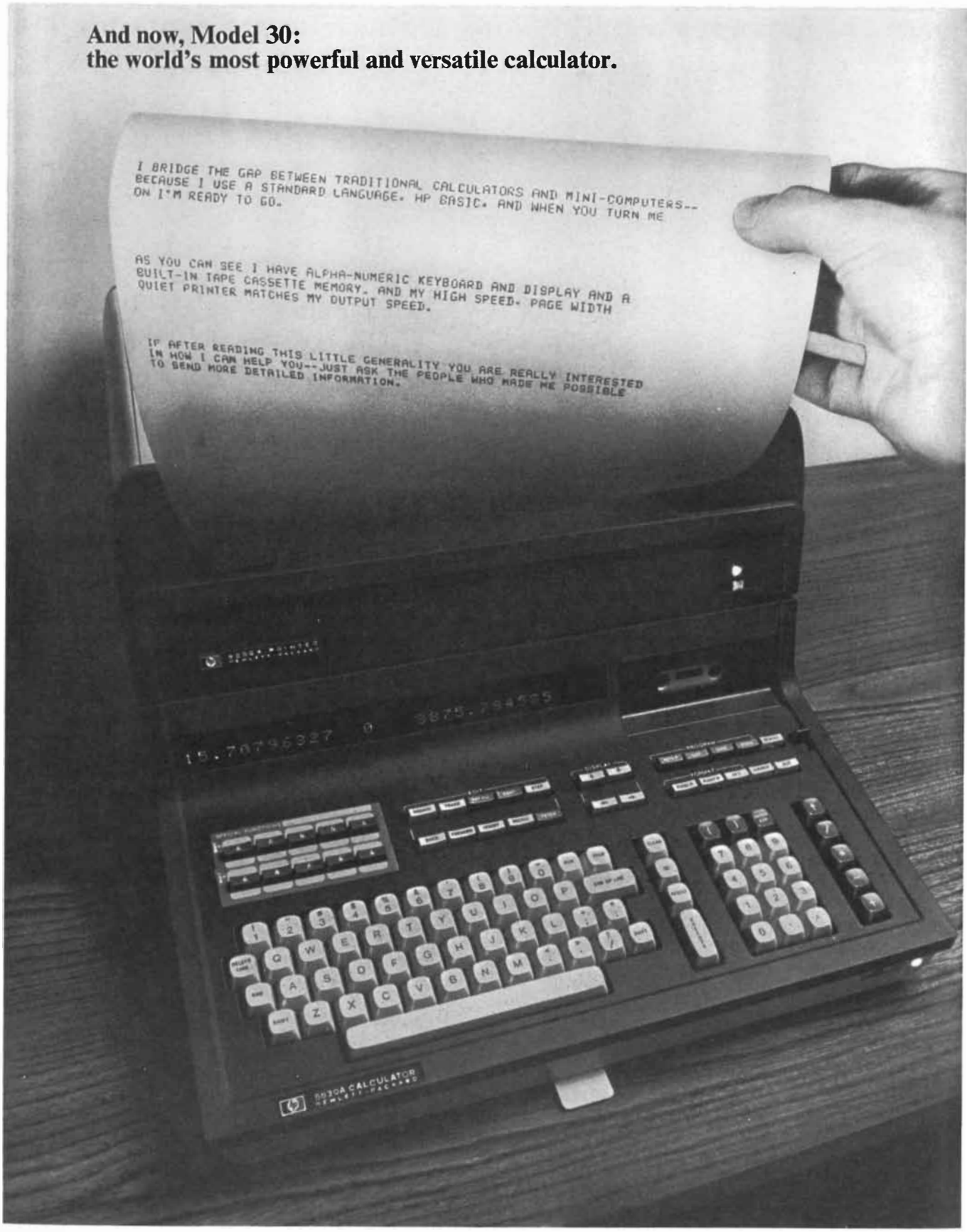
Purpose: Coronagraph contamination study, monitoring the presence of particulate matter near the spacecraft and measuring the solar corona. Equipment: Nikon Photomic FTN camera with multiple-layer coated Auto-Nikkor 55mm f1.2 lens, in fixed position. (Originally, NASA had considered a larger camera for this assignment but found 35mm equipment more suitable). A special reticle was designed for this camera's finder screen, containing degrees, x and y axes, and digital computer markings. Absolute finder accuracy is vital because the sun's image must be kept within a central 1mm circle. Since the camera is fixed, the observing astronaut will give directions in computer language for computer-controlled changes in the spacecraft's position and attitude.

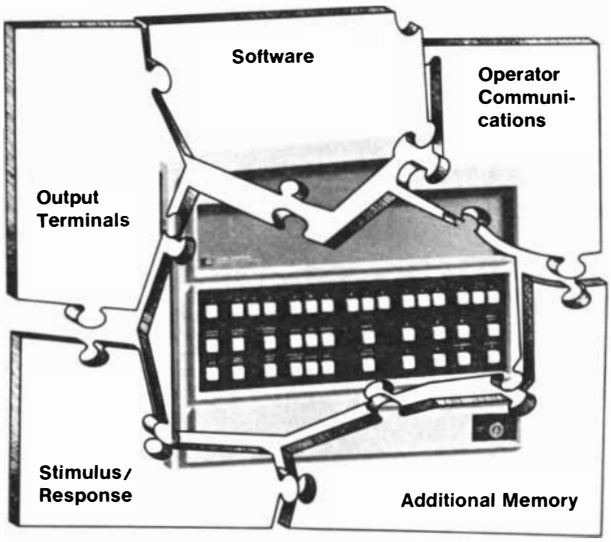
Other Nikon Skylab photography will be more like the kind you might do yourself. Included are closeup studies of a spider web spun in space as well as photos of various activities aboard the spacecraft.

In fact, the reasons Nikon works so well in outer space are the same reasons that have made Nikon the first name in 35mm photography on earth. Matchless versatility, repeatability and reliability. Ingenious design that keeps it ahead technologically while defying obsolescence. Ruggedness bordering on the incredible. Above all, picture quality that literally put 35mm photography on the map.

It's a camera worth looking into, for pleasure as well as for scientific pursuits. At Nikon dealers everywhere. (Ask also about the Nikon School of Photography.) Or, write for literature folio 19. Nikon Inc., Dept. SA, Garden City, N.Y. 11530. Subsid. of Ehrenreich Photo Optical Industries, Inc.
(In Canada: Anglophoto Ltd., P.Q.).

Some things are changing for the better.





Making the computer fit your problems.

The problem with many computer systems is that you have to make too many trade-offs between what the system can do and what you want to have done.

We believe you should have freedom to tackle problems your way, and not be forced to accept someone else's methodology. So Hewlett-Packard computers and systems are designed to help you be the master of how they're configured and used.

With HP's versatile 2100 computer, you can assemble a system that's right for your job. You determine how much and what kind of memory it should have, how you want to talk to it, and how you want it to provide your answers. The software to focus all this capability on your problem is equally flexible.

Want to hook up instruments? You choose from more than 75 standard HP stimulus/response instruments that plug directly into our computer.

But most important of all, we begin by giving you the training you need to understand and run what you have. If you need special assistance, systems analysts in our Data Centers are available for consultation.

Finally, every element of an HP computer system is fully supported from 172 sales and service offices in 65 countries. After the warranty, we can continue to maintain your system under a customer service agreement, or show you how to do it yourself.

Basically, it's all up to you. And most of our customers wouldn't have it any other way. Information on the HP 2100 will be sent upon request.

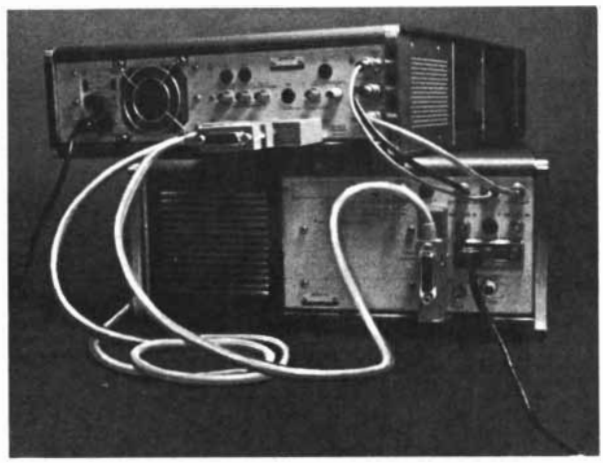
The new non-instrument that's making measurements easier.

Automated measurements have long been seen as a way to improve productivity, both in scientific research and in industry. But establishing local centralized control of measuring instruments hasn't been easy. Although some progress has been made over the years, each instrument typically has its own provision for external control. Assembling several instruments into an automated system often takes a system expert and a lot of costly interface equipment in addition to the instruments themselves.

This picture is now changing fast. Engineers throughout Hewlett-Packard have agreed to design their products using a consistent interface system, so that all instruments, no matter what their function, can be controlled by connecting them to a common interconnecting cable. Thanks to low-cost integrated circuits, even inexpensive instruments can have the new interface, and the cost of an automatic system need be little more than the sum of the costs of the instrument in it.

Instruments interconnected through the new interface communicate on a "take turn" basis: (one talks while the others listen). The benefits:

- Low-cost systems can be assembled with no special interface equipment.
- System management can range from simply one instrument controlling another, up through control by calculators and computers.
- System operation is simplified since control can be passed from one device to another.
- The system is flexible in speed, language, and size. Messages and data can be transmitted at up to one megabyte rates.



Now one common cable system is all that is required to interconnect instruments digitally.

This interface is going to enhance the usability of many of our new products and substantially reduce the rigors of making complex measurements. To obtain a more complete description write for the October 1972 issue of the Hewlett-Packard Journal.

For more information on the products described in these pages, fill out the coupon or write to: Hewlett-Packard, 1504 Page Mill Road, Palo Alto, California 94304.

HEWLETT hp PACKARD

HP sales, service and support in 172 cities in 65 countries.

00302

Please send me information on the following:
( ) HP 30 Calculator
( ) HP 2100 Computer
( ) Interface System

Name_____

Title_____

Company_____

Address_____

City_____ State_____ Zip_____

LETTERS

Sirs:

Scientists habitually exaggerate the extent to which their results were foreseen, and I fear that this tendency may be reinforced by statements made by Chris D. Zafiratos in his article "The Texture of the Nuclear Surface" [SCIENTIFIC AMERICAN, October, 1972]. He says: "The concept of an atomic nucleus was first proposed by Ernest Rutherford in 1911 and was verified by the experiments involving the scattering of alpha particles (helium nuclei) by Hans Geiger and Ernest Marsden in 1913."

In fact, Geiger and Marsden's early alpha-scattering results preceded Rutherford's nuclear model and were the inspiration for it. These early results (that alpha particles could be scattered through very large angles) were a complete surprise to Rutherford. His remark that it was like firing 16-inch shells at sheets of tissue paper and having some of them bounce back has become a classic in physics. Of course, Geiger and Marsden also made accurate comparisons with the nuclear scattering theory after it had been developed.

Zafiratos also leaves the impression that Niels Bohr arrived quite independently at his quantized nuclear model of the atom in 1913. Actually Bohr conceived the idea in 1912 while working in Rutherford's laboratory in Manchester. He wrote to his brother in July, 1912, "I have perhaps found out a little about the structure of the atom...."

None of the above detracts in any way from the work of Rutherford or Bohr, or for that matter from the excellent article by Zafiratos.

R. E. BELL

Rutherford Professor of Physics
McGill University
Montreal

Sirs:

It is surprising to find that F. G. Heath, in his article "Origins of the Binary Code" [SCIENTIFIC AMERICAN, August, 1972] has nothing to say about the important contributions of G. W. von Leibniz to the development of this system. Heath refers to Leibniz' writings on logical algebra but seems to have missed his clear description of the binary number system.

This Leibniz published in *Histoire-Mémoires de l'Académie Royale des Sciences,* Paris, Année 1703 (1705). In the *Mémoires,* pp. 85–89, is "*Explication de l'arithmétique binaire qui se sert des seuls caractères 0 et 1; avec des remarques sur son utilité, et sur ce qu'elle donne le sens des anciens figures Chinoises de Fohy.*" In the *Histoire,* pp. 58–63, is an unsigned piece, "*Nouvelle arithmétique binaire.*"

Leibniz had written on this subject earlier; on January 2, 1697, he had addressed a letter to Duke Rudolf August v. Braunschweig-Lüneburg-Wolfenbüttel describing the binary system. This was first published in 1968 (Leibniz, *Zwei Briefe über das binäre Zahlensystem,* edited by R. Loosen, F. Vonessen and J. Gebser).

Moritz Cantor, the historian of mathematics, had described Leibniz' work in *Mathematische Beiträge zum Kulturleben der Völker* (1861). In none of these works is there any mention of the Francis Bacon alphabetic binary system. Leibniz was under the impression that the system was invented *ca.* 3000 B.C. by one Fo Hi. Cantor says, however, that the Fo Hi system was not mathematical (like the Bacon system).

JOHN S. KEBABIAN

Scarsdale, N.Y.

Sirs:

Let me say straight away that my own interest has always been in coding: the discovery of digit sequences that are particularly suited for certain engineering purposes. This aspect seems to be halfway between ciphers and arithmetic processes. My article was certainly not intended to cover binary arithmetic.

However, I certainly found Mr. Kebabian's letter most interesting, and binary-code enthusiasts should read at least the French paper he mentions. As regards Bacon v. Leibniz, Bacon's cipher has always been well known, and I still feel that he should be given the credit. He had discovered the binary code as a cipher by 1605 and had published it in 1623, exactly as needed for arithmetic. Since his works were read widely by scientists in the following century, it is difficult to believe that Leibniz, 75 years later, could have been working without some influence from Bacon.

For instance, as late as 1833 Gauss and Weber were recorded as constructing a telegraph system that used the Bacon code. My own view of Leibniz is that he deserves a great deal of credit for almost discovering Boolean algebra long before Boole, which is why I stressed that aspect of his work in my article. A computer designer can recognize the value of Leibniz' work in this area directly from the original text.

It seems fairly certain that, as is mentioned in Mr. Kebabian's letter, the Chinese used binary-code sequences, at least up to 64. I am informed that the sequences can be found in the ancient Chinese book *I Ching,* which has been published in English, and that the binary sequences were used as a shorthand in fortune-telling, since the method used sequential coin-flipping with six tosses. To put the interpretation table in the same form as Bacon's table for easy look-up would have been a very natural thing to do.

F. G. HEATH

Professor of Computer Engineering
Department of Electrical
and Electronic Engineering
Heriot-Watt University
Edinburgh

Scientific American, January, 1973; Vol. 228, No. 1. 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.

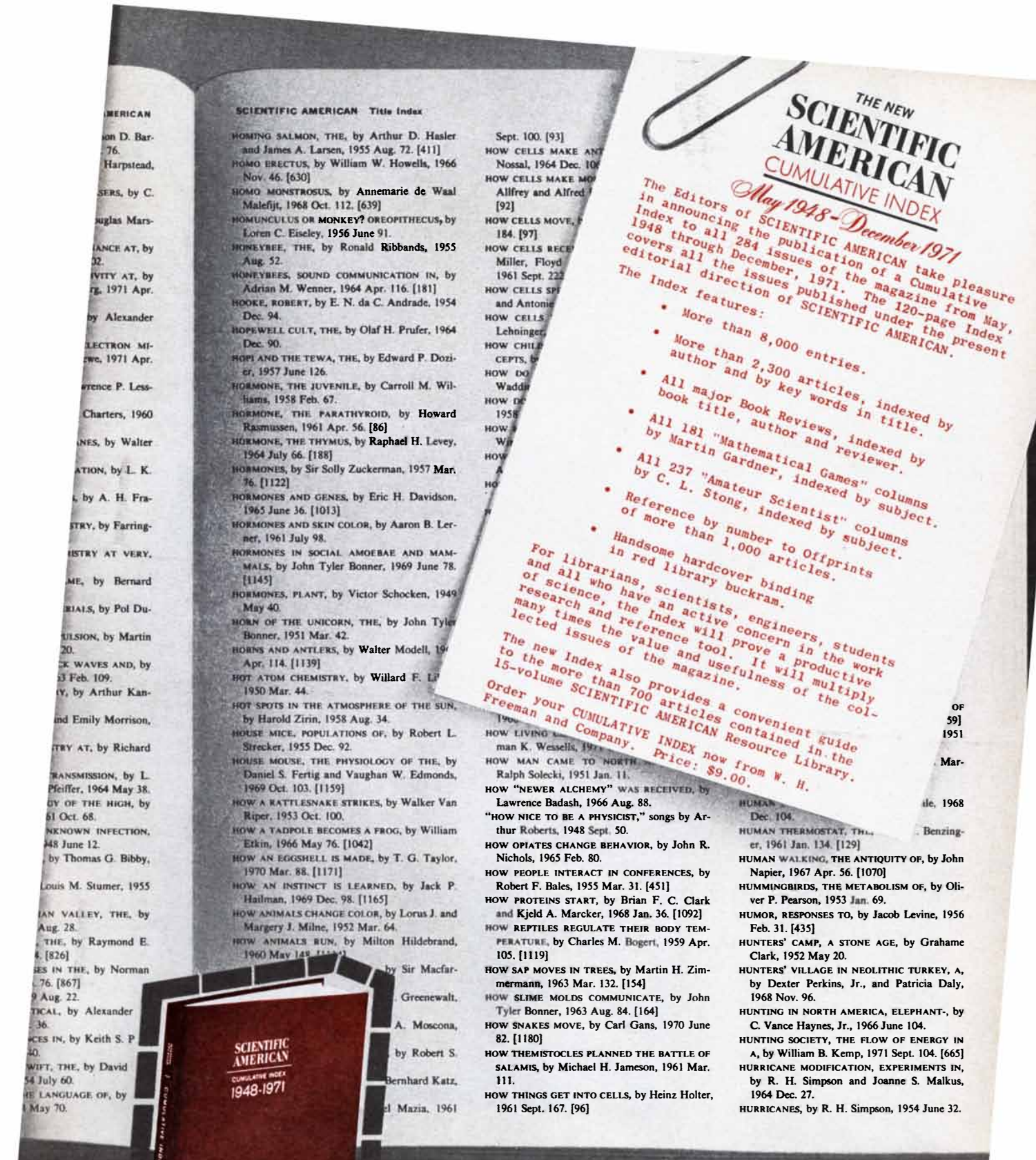
☐ Please enter change of address.

☐ 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

JANUARY, 1923: "The traffic surveys for the new Detroit-Windsor Bridge, which is to connect Detroit, Mich., and Windsor, Ont., have developed some startling figures for the consideration of the builders of highways and bridges. The total number of automobiles in the United States in 1922 was 10,448,632, an increase of more than 1,000 per cent since 1912. The increase of 1925 is quite sure to represent a total of automobiles of more than 15,500,000, and in less than seven years the number will be doubled! The highways have become too narrow to accommodate the ever increasing number of automobiles. The automobile becomes almost useless for the handling of downtown business traffic, and it is only a question of a short time until we shall come to the two-story street, one-way streets, and streets on which automobile traffic shall be absolutely prohibited, at least during certain business hours. The roadways on our bridges and in tunnels, if they can be made to carry automobile traffic safely, must be made for not less than three lanes of traffic in each direction for city bridges, and for a total of three or four lanes for country bridges. What the future will bring forth with such wizards as Ford to keep on reducing the cost and price of automobiles no one can predict, but it is safe to say that within the next decade all of the American continent will at last be on wheels."

"The coal strike, with resulting scarcity and high prices of fuel, has turned the attention of thousands of persons to the possibilities of fuel oil as a source of domestic heat. We do not refer to the single-room kerosene stove. All of us are using this as an emergency measure; all of us understand full well that it is nothing more than this. But what about the conversion of our hot-air or steam or hot-water plants to the use of oil fuel? This is not an emergency measure; it is a permanent substitution of oil for coal in the domestic cellar. Already great quantities of fuel oil are used in industry, and its field of usefulness is pretty well defined and understood. The use of oil for fuel in the home is comparatively new, however, since it has only been within the past few years that burners which were safe and economical have been developed."

"After a lapse of several years there is renewed interest in talking pictures—motion pictures accompanied by more or less realistic sounds. Obviously the pictures and the sounds must keep in perfect step throughout; otherwise the results are sometimes ludicrous and at all times unconvincing. All the earlier attempts made use of cylinder- or disk-type phonographic records, but these methods leave much to be desired. It is pretty well admitted today that any successful system of talking motion pictures must combine the sound record and pictures on a single film; any other plan only leads to trouble. What appears to be the most promising method is to record the sounds photographically at the same time and on the same film as the corresponding motion picture. The film moves intermittently—frame by frame—through the motion-picture mechanism, and continuously before and after. While it is moving steadily the sound record is made. For exhibition purposes the film goes through the projector mechanism frame by frame and steadily through the sound-reproducing device. A beam of light is passed through the sound record, and the varying degree of shadow, falling on a light-sensitive cell, causes a current of fluctuating strength to pass through a circuit that includes loud-speaking devices."

"Fräulein Gretel Steiff, who brought joy to so many million nurseries, has died. She is worthy of mention in these or any other columns. The invention by her of the 'Teddy Bear Doll' will go down in history with the Kiddie Car and other things that make life happy for children."

SCIENTIFIC AMERICAN

JANUARY, 1873: "Year by year the laws of health are better known and observed. In nearly all civilized countries the average duration of life is steadily increasing. In Geneva accurate registers have been kept of the average span of life since 1560, when it was 22 years six months. In 1833 it was 40 years and five months. Thus in less than 300 years the average duration of life is nearly doubled, and the same holds true for Paris and England. Science marches steadily onward and upward, always victorious. 'But oh,' asks some one, 'what is to be done with such a mass of people as will accumulate on the earth?' Why, the fact is that the earth is not 1/20th peopled now. Talk about the overcrowded population of Europe. England itself is not 1/4th populated at the present time, and there is not the least call for emigration. Palestine may yet be what it was once represented as being, 'a land flowing with milk and honey,' the great valleys of the Tigris and Euphrates may be regenerated and be repeopled by a greater population than Nineveh or Babylon ever knew."

"Each year since the commencement of the first term of President Grant the Secretary of the Navy has earnestly endeavored to impress on Congress and on the country the vital necessity of preserving the efficiency of our diminutive navy. We must certainly have a number of cruising vessels to do the work that falls to our navy in time of peace, and this work could be done by ships of comparatively light armament, of full power and of good speed under steam. We believe that there is not a vessel in our navy that possesses all the requisites of such a class of ships. It is to make good our deficiency here, we presume, that the Secretary of the Navy proposes to build 10 new vessels. We question seriously, however, the policy of building a new iron-clad fleet to compete with that of England, of Prussia or even of Spain. We may find that the expenditure of many millions in attempting to rival other nations may be saved by the comparatively inexpensive operations of a well-organized torpedo corps, and by the application of the wonderful ingenuity of our inventors to the perfection of floating and sub-aqueous torpedoes and torpedo ships."

"Surely the world is moving. One of the last and perhaps the greatest improvements in surgery is a new method of cutting away formidable tumors in some of the body cavities by means of electricity. A wire is passed round the base, the battery is set in motion and, presto, the tumor is separated without loss of blood, the vessels being seared, as it were, with a hot iron. Better still, under the lulling effects of chloroform or ether the patient experiences neither pain nor haemorrhage."

THE AUTHORS

DEREK P. GREGORY ("The Hydrogen Economy") is assistant director of engineering research at the Institute of Gas Technology in Chicago. He took up that work in 1970 after four years as research manager of Energy Conversion Ltd. in England, where he did research on fuel cells and lightweight batteries and, he writes, "joined the quest for an economic electric automobile" until he decided that it "was a lost cause." Now he is involved in "long-range planning and experimental studies concerning the gas industry's role in the long-term future, the conservation of energy resources and the energy-supply situation." Gregory received his bachelor's and doctor's degrees in chemistry at the University of Southampton and then worked in England for the Atomic Energy Authority of the United Kingdom and in the U.S. for Pratt & Whitney before returning to England in 1966. "Away from work," he writes, "I enjoy sailing and sculpture."

ABNER LOUIS NOTKINS and HILARY KOPROWSKI ("How the Immune Response to a Virus Can Cause Disease") are respectively chief of the virology section of the National Institute of Dental Research and director of the Wistar Institute of Anatomy and Biology. Notkins received his bachelor's degree from Yale College in 1953 and his M.D. from New York University in 1958. After internship and residency at the Johns Hopkins Hospital he joined the National Institutes of Health as a research associate in the National Cancer Institute, moving to the National Institute of Dental Research in 1961. His wife is a practicing architect and a member of the faculty of architecture at the University of Maryland. Their outside interests include exploring new and old buildings and sailing on Chesapeake Bay. Koprowski, who was born in Poland, received his M.D. at the University of Warsaw in 1939. He came to the U.S. in 1944, after four years of research on yellow fever in Brazil. He spent 14 years with the Lederle Laboratories, joining the Wistar Institute in 1957. The authors wish to acknowledge the assistance of their colleague Margaret Engel with the illustrations accompanying their article.

GEZA TELEKI ("The Omnivorous Chimpanzee") is completing work for his Ph.D. in primatology at the University of Georgia. He obtained his bachelor's degree in anthropology at George Washington University in 1967 and his master's degree in physical anthropology at Pennsylvania State University in 1970. Since his graduation from college he has worked at the Smithsonian Institution and Penn State and has spent a year in Tanzania as field assistant to Jane van Lawick-Goodall, director of the Chimpanzee Research Centre. Teleki was born in Hungary and writes that his languages include "English, Hungarian, German and basic Swahili."

BRUCE C. MURRAY ("Mars from Mariner 9") is professor of planetary science in the division of geological and planetary sciences at the California Institute of Technology. His degrees, all in geology, are from the Massachusetts Institute of Technology: bachelor's in 1953, master's in 1954 and Ph.D. in 1955. For three years after leaving M.I.T. he was an oil geologist. He went to Cal Tech in 1960 after two years in the Air Force, which assigned him to work as a geophysicist. Murray was a coinvestigator with the space projects that involved photographing the surface of Mars with television cameras aboard *Mariner 4* in 1964, *Mariner 6* and *Mariner 7* in 1969 and *Mariner 9* in 1971. He is leader of the television team for the Mariner Venus-Mercury mission scheduled for launch late this year. From 1966 to 1969 he was a member of the space science and technology panel of the President's Science Advisory Committee.

FRANK S. WERBLIN ("The Control of Sensitivity in the Retina") is associate professor in the electronics research laboratory of the college of engineering at the University of California at Berkeley. Noting that he received his master's degree at the Massachusetts Institute of Technology in 1962, four years after his graduation from that institution, and his Ph.D. from Johns Hopkins University six years after the master's degree, he writes: "The long span of time between the bachelor's and the doctoral degree was spent working in the electronics industry, from White Sands Proving Grounds on missile systems to Maimonides Hospital in Brooklyn on a biomedical engineering project. It was while working on an artificial bladder at Maimonides that I decided to learn something about physiology. Working as an engineer on an artificial bladder for a surgeon is much like having a job as a plumber's assistant. I went to Johns Hopkins thinking I would learn bladder physiology and solve the problem myself, but there I became fascinated with neurophysiology and began working on problems in vision." Adding that he and his wife recently bought "a big old house that we are renovating ourselves," he says: "This leads to a kind of schizophrenic existence. During the week I am concerned with using very fine microelectrodes to penetrate the retinal structures and to record from cells of microscopic dimension. On weekends I interact with bricks, mortar and wallboard that are often too heavy for me to carry alone." The aim of the weekend work, he says, "is simply to get four walls to be vertical and at right angles to each other."

MICHAEL E. STONE ("Judaism at the Time of Christ") is senior lecturer at the Hebrew University of Jerusalem and chairman of the department of Armenian and Iranian studies. Born in England, he obtained his bachelor's degree at the University of Melbourne in 1959 (graduating with first-class honors in Greek and Semitics) and his Ph.D. from Harvard University in 1965. He taught for a year at the University of California at Santa Barbara before going to Hebrew University. "I am in particular devoted to the study of Armenian manuscripts," he writes, "partly in search for otherwise unknown Jewish and early Christian writings and partly for the general interest the subject holds for me. I have pioneered the use of computers for the study of Armenian literary texts, particularly in the preparation of concordances."

M. YA. AZBEL', M. I. KAGANOV and I. M. LIFSHITZ ("Conduction Electrons in Metals") are physicists affiliated with the Academy of Sciences of the U.S.S.R. Azbel' is with the Academy's Institute of Theoretical Physics and Kaganov and Lifshitz are with the Institute of Physical Problems.

CHARLES and NANCY KNIGHT ("Snow Crystals") are at the National Center for Atmospheric Research. Charles Knight, who obtained his Ph.D. from the University of Chicago in 1959, has had several periods of residence and work on drifting ice stations in the Arctic; he notes that one of them, ARLIS I (Arctic Research Laboratory Ice Station I) "was the first attempt at an austere station, and it did succeed in being austere." His wife has also worked on Arctic research projects and was doing so when they met. They were the authors of "Hailstones" in the April 1971 issue of SCIENTIFIC AMERICAN.

IF YOU FEEL LOST IN A WINE STORE, HERE'S HOW TO FIND YOURSELF.

The difficult thing about buying wine is that there are so many wines to buy.

On the one hand, we're fortunate that there are so many good wines from which to choose.

On the other, the over 500 different kinds of wine, and the hundreds of different brands of each, domestic and imported, make the task of learning about wine almost impossible. Most people throw up their hands when confronted with so many choices and just grab a bottle because they like the label.

We at Inglenook Vineyards realize it's impossible to make a wine expert of you in one ad. But we can give you some pointers and a basic starting point, so that you'll have a better idea of how to go about buying wine.

TIP NO. 1: START SOMEWHERE.

First off, you need a starting point. You should become thoroughly familiar with four or five wines. Then when you try new wines, you'll have a basis for comparison.

We've arbitrarily listed five wines below which we feel are a good place to start. They're not the final answer, of course, because there is none. But they do represent five totally different, easy to distinguish tastes.

FIVE BASIC TASTES TO START WITH.

Name	Color	Taste
Pinot Chardonnay	White	Dry
Sylvaner Riesling	White	Semi-Dry
Charbono	Red	Dry
Gamay Beaujolais	Red	Semi-Dry
Gamay Rosé	Rosé	Dry

TIP NO. 2: BE HUMBLE.

When you go to buy these wines, don't be afraid to admit you're not the world's foremost expert on wine. There are so many wines, and vintages, and so forth that it's safe to say that *nobody* is the world's foremost expert on wine. So get all the help you can.

There is usually one person in every wine store who knows more than you do—the manager. Get his opinions on vintages, vintners, and the like. But make sure he drinks and enjoys wine regularly himself. If this is the case, you've chosen the right store, and found someone who can save you years of fiddling around in the dark.

TIP NO. 3: READ THE LABEL.

The better wines are generally produced and bottled solely at the winery.

This insures almost total control of a wine's production by the vintner. So look for the words "Produced and Bottled by." In French, these words read, "mis en bouteilles au chateau" and in German the term is "Original Abfullung." Otherwise the wine could have been made in one place, blended in another, and aged in still another.

Even better are the words, "Estate Bottled." This means that the grapes were grown under control of the vintner and in the vicinity of the vintner's chateau or estate. Only a few American wines bear this marking, and finding one in a wine store is a good way to find yourself.

Which leads us to Inglenook.

TIP NO. 4: SPEND MONEY.

Unfortunately, the age old saying "you get what you pay for" is generally true when it comes to wine.

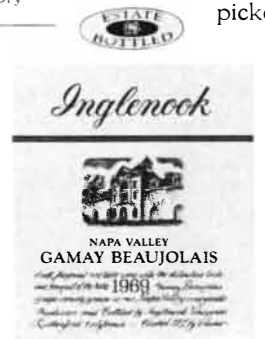
Meticulous tending of grapes and vines costs money, vintaging costs more money, and estate bottling costs even more.

And even though we all know someone who picked up a bottle of 99¢ imported wine and raved about it for years, you have to write it off to chance. For it just isn't possible to consistently produce great wine at a low cost.

Inglenook Estate Bottled wine is the most expensive wine made in America.

It comes from the Napa Valley, it's all vintage dated, and it's all produced and bottled by us.

So the next time you get that old lost feeling when buying wine, let Inglenook be your guide.



INGLENOOK

We make the most expensive wine in America.

This ad is one of a series. If you'd like copies of the other ads, send your name and address to The Cellarmaster, Box J, Inglenook Vineyards, Rutherford, CA 94573.

Established 1845 SCIENTIFIC AMERICAN January 1973 Volume 228 Number 1

The Hydrogen Economy

A case is made for an energy regime in which all energy sources would be used to produce hydrogen, which could then be distributed as a nonpolluting multipurpose fuel

by Derek P. Gregory

The basic dilemma represented by what has been termed the "world energy crisis" can be simply stated: At the very time that the world economy in general and the economies of the industrialized countries in particular are becoming increasingly dependent on the consumption of energy, there is a growing realization that the main sources of this energy—the earth's nonrenewable fossil-fuel reserves—will inevitably be exhausted, and that in any event the natural environment of the earth cannot readily assimilate the by-products of fossil-fuel consumption at much higher rates than it does at present without suffering unacceptable levels of pollution.

What is not generally recognized is that the eventual solution of the energy problem depends not only on developing alternative sources of energy but also on devising new methods of energy conversion. There is, after all, plenty of "raw" energy around, but either it is not in a form convenient for immediate use or it is not in a location close enough to where it is needed. Most of the research-and-development effort in progress in the U.S. on the energy problem is devoted to finding ways to convert chemical energy (derived from fossil fuels), nuclear energy (derived from fission or fusion reactions) and solar energy (derived directly from the sun) into electrical energy.

At present nuclear-fission plants supply about 1.6 percent of the electricity consumed in the U.S. (Of the remainder, fossil-fuel plants supply about 82 percent and hydroelectric plants about 16 percent.) Assuming that the development of economically feasible "breeder" reactors will soon eliminate any short-term concern about the resource limitation of nuclear energy, then by the year 2000 nuclear plants may be supplying as much as half of the nation's electricity.

If this projection is correct, and if the "energy gap" of the future is to be filled with nuclear power made available to the consumer in the form of electricity, then the U.S. will have gone a long way toward becoming an "all-electric economy." This trend can be detected already: the demand for electricity is currently growing in the U.S. at a much higher rate than the overall energy demand [*see illustration on next page*]. It has been estimated that whereas the overall U.S. energy consumption will double by the year 2000, the demand for electricity will increase about eightfold, raising the electrical share of total energy consumption from about 10 percent to more than 40 percent.

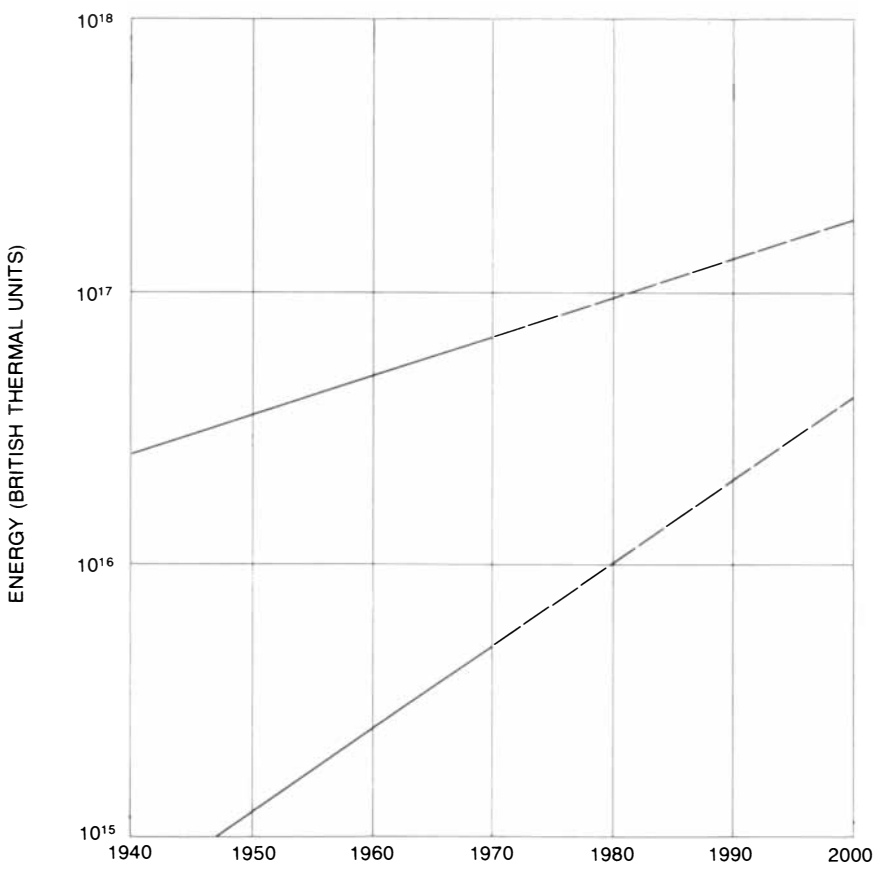
The question naturally arises: How desirable is this trend toward a predominantly electrical economy? Specifically, are there any other forms of energy that can be delivered to the point of use more cheaply and less obtrusively than electrical energy can? Consider such major energy-consumption categories as transportation, space heating and heavy industrial processes, all of which are primarily supplied today with fossil-fuel energy, mainly for reasons of economy and portability. As the fossil fuels run out, they will become more expensive, making the direct use of nuclear electrical energy relatively more economical. In this situation a case can be made for utilizing the nuclear-energy sources indirectly to produce a synthetic secondary fuel that would be delivered more cheaply and would be easier to use than electricity in many large-scale applications. In this article I shall discuss the merits of what I consider to be the leading candidate for such a secondary fuel: hydrogen gas.

In many respects hydrogen is the ideal fuel. Although it is not a "natural" fuel, it can be readily synthesized from coal, oil or natural gas. More important, it can be produced simply by splitting molecules of water with an input of electrical energy derived from an energy source such as a nuclear reactor. Perhaps the greatest advantage of hydrogen fuel, however, at least from an environmental standpoint, is the fact that when hydrogen burns, its only combustion product is water! None of the traditional fossil-fuel pollutants—carbon monoxide (CO), carbon dioxide (CO_2), sulfur dioxide (SO_2), hydrocarbons, particulates, photochemical oxidants and so on—can be produced in a hydrogen flame, and the small amount of nitrogen oxide (NO) that is formed from the air entering the

flame can be controlled. Moreover, assuming that the energy options are restricted to the use of effectively "unlimited" materials such as air and water, hydrogen is by far the most readily synthesized fuel.

In principle, then, one can envision an energy economy in which hydrogen is manufactured from water and electrical energy, is stored until it is needed, is transmitted to its point of use and there is burned as a fuel to produce electricity, heat or mechanical energy [*see illustration on opposite page*]. Such a hypothetical model is not without its problems and disadvantages, but on balance the benefits appear to be so great that I believe at the same time that we are moving toward an "electric economy" we should also be moving toward a "hydrogen economy."

Just as the food and beverage industry has found it uneconomical to collect and reuse empty containers, so the present energy industry cannot afford to collect and recycle used "energy containers": the by-products of the combustion necessary to produce the energy. The drawback in both cases is that the "no deposit, no return" system throws the burden of recovery and recycling onto the environment. Apart from the obvious harmful effect on the earth's atmosphere, this kind of energy cycle suffers from the further disadvantage of having an extremely slow step of several million years' duration for the re-formation of fossil fuels from atmospheric carbon dioxide [*see illustration on page 16*]. That is the basic reason we are running out of fossil-fuel reserves. In the hydrogen cycle, in contrast, only water is deposited into the atmosphere, where it rapidly equilibrates with the abundant and mobile water supply on the earth's crust. At another location the water is reconverted to hydrogen. The system is characterized by negligible delay and does not disturb the environment, yet it relies on the environment to carry out the "return empty" function. Assuming the availability of an abundant supply of nuclear or solar energy, this system can be operated as rapidly as the demand requires without depleting any natural resources.



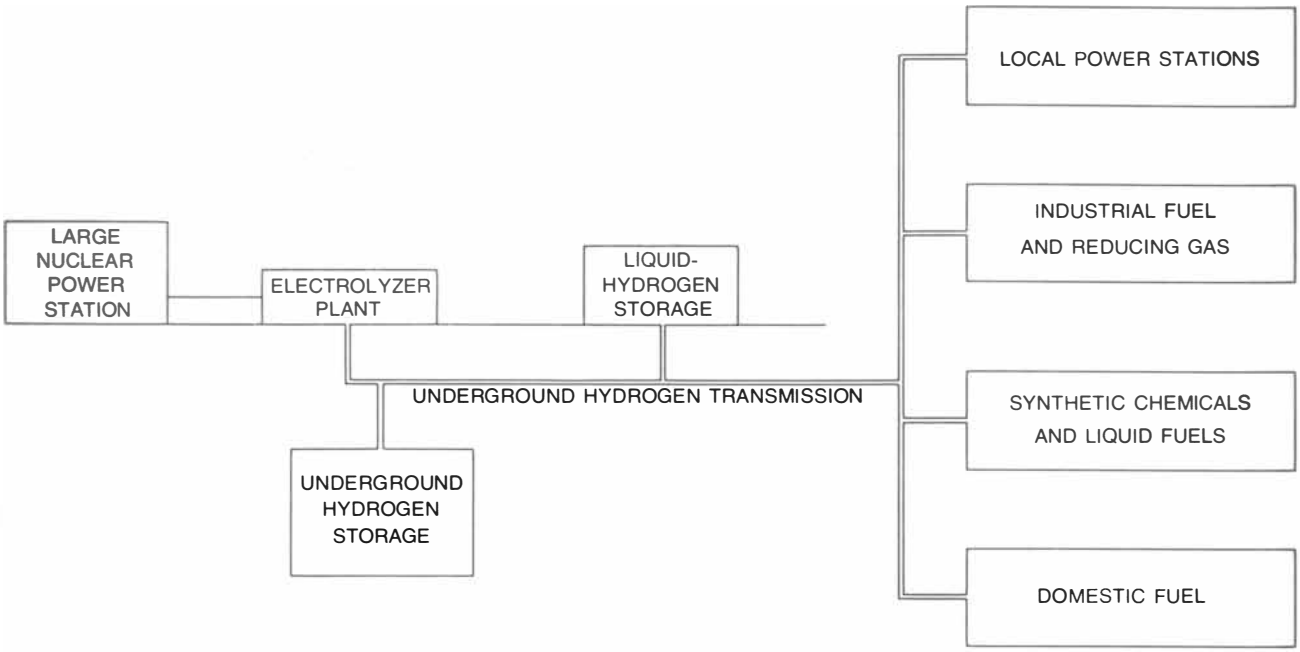
ACCELERATING TREND toward an "all electric" economy is evident in this graph, which shows that the demand for electricity (*bottom line*) is growing in the U.S. at a much higher rate than the overall energy demand (*top line*). Assuming that the trend continues, the U.S. is heading for a predominantly electrical economy sometime in the 21st century. The data are from the U.S. Department of Commerce and the Edison Electric Institute.

The idea of using hydrogen as a synthetic fuel is far from new. In 1933 Rudolf A. Erren, a German inventor working in England, suggested the large-scale manufacture of hydrogen from off-peak electricity. He had done extensive work on modifying internal-combustion engines to run on hydrogen, and the main object of his suggestion was to eliminate automobile-exhaust pollution and to relieve pressure on the importation of oil into Britain. (It is interesting to note that 40 years later the U.S. is concerned with the same two problems: automobile pollution and an increasing dependence on oil imports.)

Others have suggested using hydrogen as a fuel or as a means of storing energy. F. T. Bacon, a pioneer in the development of fuel cells in England since the 1930's, has always had as his ultimate objective the development of a hydrogen-energy storage system using reversible electrolyzer fuel cells. More recently the U.S. Atomic Energy Commission sponsored a series of studies during the 1960's of "nuplexes"—nuclear-agricultural-industrial complexes that derive all their energy from a single nuclear reactor. The AEC studies included the concept of water electrolysis to provide hydrogen as a precursor to the manufacture of fertilizers and chemicals. Within the past two years several articles have appeared in engineering and scientific journals proposing active studies of the production, transmission, storage and utilization of hydrogen in both combustion appliances and engines. Such studies are in progress at several universities and industrial research laboratories in the U.S. and abroad, including my own institution, the Institute of Gas Technology in Chicago, where our work is sponsored by the American Gas Association.

The difficulty of transporting hydrogen has historically prevented its use as a fuel. Clearly some better method than compressing it in steel cylinders has to be found. Storage and transportation as liquid hydrogen are already in use; metal hydrides and synthetic organic or inorganic hydrides have also been considered and have promise. There is no reason, however, why hydrogen should not be distributed in the same way that natural gas is distributed today: by underground pipelines that reach most industries and more than 80 percent of the homes in this country.

Before weighing the merits of the hydrogen-economy concept, it is instructive to consider the alternative: the all-electric economy. Suppose for a mo-



HYDROGEN ENERGY ECONOMY would operate with hydrogen as a synthetic secondary fuel produced from water in large nuclear or solar power stations (*left*). The hydrogen would be fed into a nationwide network of underground transmission lines (*center*), which would incorporate facilities for storing the energy, either in the form of hydrogen gas underground or in the form of liquid hydrogen aboveground. The hydrogen would then be distributed as it is needed to energy consumers for use either as a direct heating fuel, as a raw material for various chemical processes or as a source of energy for the local generation of electricity (*right*).

ment that one does not consider synthesizing a secondary chemical fuel; then one must face the prospect of generating and transmitting very large quantities of electricity. To meet the rising demand for electricity in the U.S. new generating stations are already being constructed in sizes larger than ever before. A few years ago a 500-megawatt power station was considered a giant. Today 1,000-megawatt stations are typical, and the electrical industry is contemplating 10,000-megawatt installations for the future.

In spite of the intensive efforts of their designers, the efficiency of steam-driven electric-power stations is still fairly low: about 40 percent for a modern fossil-fuel plant and 33 percent for a nuclear plant [see "The Conversion of Energy," by Claude M. Summers; SCIENTIFIC AMERICAN, September, 1971]. As a result the waste heat released from these large plants, or clusters of plants, is considerable. Accordingly they must be located near large bodies of water where ample cooling is available or in open country where cooling to the atmosphere will have no adverse local effects. Concern over the safety of nuclear reactors is also having a strong influence on the location of such plants. Because of these constraints the huge power stations of the future are likely to be built at distances of 50 miles or more from the load centers. Power stations located on offshore platforms floating in the ocean are already planned for the U.S. East Coast.

Power must be moved from the generating stations to the load centers. High-voltage overhead cables are expensive, in terms of both equipment costs and the land they occupy, and they are vulnerable to storm damage. Moreover, the electrical industry is encountering considerable resistance to the continued stringing of overhead power-transmission lines in many areas. Underground cables for carrying bulk power cost at least nine times (and sometimes up to 20 times) as much as overhead lines and thus are far too expensive to be used over long distances. Underground transmission is used only where the expense is justified by other considerations, such as aesthetic appearance or very expensive right-of-way. Much work is being done to develop cryogenic superconducting cables, which would allow large currents to be carried underground at a reasonable cost. At present, however, the technology is still at an early stage of development.

Some form of electrical storage would be of great value to the electrical industry, because power stations work most efficiently when operated at constant output at their full rated load. Since consumer demand varies widely both seasonally and during the day, however, the generating rate must be adjusted continuously. The only practical way available today to store large quantities of electrical energy is the pumped-storage plant, a reversible hydroelectric station; unfortunately only a limited number of sites are geographically suitable for such systems.

Thus it appears that several of the problems faced by the electrical industry—the siting of power stations, the expense of underground transmission and the lack of storage—are being amplified by factors that lead to larger and more remote power stations. The hydrogen-economy concept could help to alleviate these problems.

Hydrogen can be transmitted and distributed by pipeline in much the same way that natural gas is handled today. The movement of fuel by pipeline is one of the cheapest methods of energy transmission; hydrogen pipelining would be no exception. A gas-delivery system is usually located underground and is therefore inconspicuous. It also occupies less land area than an electric-power line. Hydrogen can also be stored in huge quantities by the very same techniques used for natural gas today.

Let us take a look at the existing gas-transmission network in the U.S. In 1970 a total mileage of 252,000 miles of trunk pipeline was in operation, carrying a total of 22.4 trillion cubic feet of gas during the year [*see illustration on pages 18 and 19*]. Such a pipeline system is

needed because natural-gas sources are located in certain parts of the country, whereas markets for the gas exist in other areas.

In the hydrogen economy hydrogen gas would be produced from large nuclear-energy (or solar-energy) plants located in places that provide optimum cooling and other environmental facilities. Even coal-fueled hydrogen generators, located close to the mine mouths, could be integrated into this power-generation network. A pipeline transmission system would grow up to link these locations to the cities in a way analogous to the growth of the natural-gas transmission system.

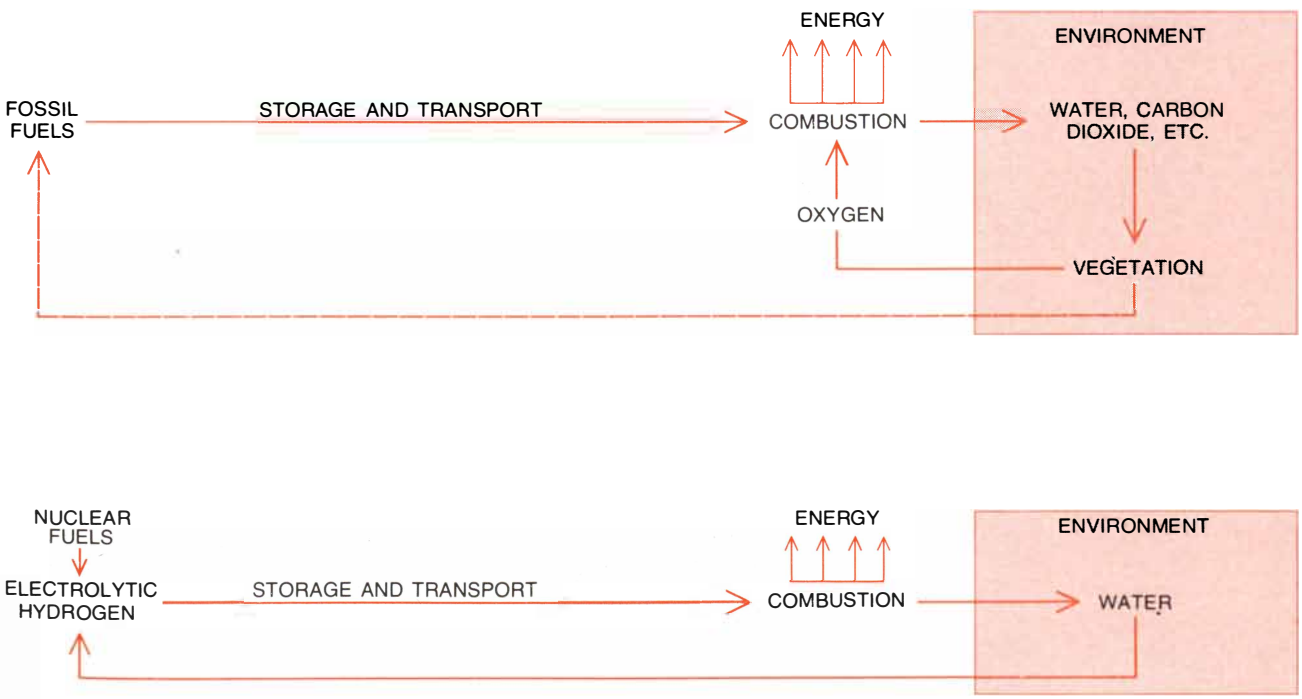
The technology for the construction and operation of natural-gas pipelines has been well developed and proved. A typical trunk pipeline, 600 to 1,000 miles long, consists of a welded steel pipe up to 48 inches in diameter that is buried underground with appropriate protection against mechanical failure and/or electrochemical corrosion. Gas is pumped along the line by gas-driven compressors spaced along the line typically at 100-mile intervals, using some of the gas in the line as their fuel. Typical line pressures are 600 to 800 pounds per square inch, but some systems operate at more than 1,000 pounds per square inch. A typical 36-inch pipeline has a capacity of 37,500 billion British thermal units (B.t.u.) per hour, or in electrical equivalent units 11,000 megawatts, roughly 10 times as much as a single-circuit 500-kilovolt overhead transmission line.

Natural gas is not the only gas to be moved in bulk pipelines, although no other gas is moved on such a scale. Carbon dioxide, carbon monoxide, hydrogen and oxygen are all delivered in bulk by pipeline. So far industry has had no incentive to pipeline hydrogen in huge quantities over great distances, but where it now pipelines hydrogen over short distances it uses conventional natural-gas pipeline materials and pressures. There is no technical reason why hydrogen cannot be pipelined over any distance required.

Because of the lower heating value of hydrogen (325 B.t.u. per cubic foot compared with about 1,000 B.t.u. per cubic foot for natural gas) three times the volume of hydrogen must be moved in order to deliver the same energy. Hydrogen's density and viscosity are so much lower, however, that the same pipe can handle three times the flow rate of hydrogen, although a somewhat larger compressor energy is required. Thus where existing pipelines happen to be suitably located, they could be converted to hydrogen with the same energy-carrying capacity.

In the hydrogen economy it will be possible to store vast quantities of hydrogen to even out the daily and seasonal variations in load. Natural gas is stored today in two ways: in underground gas fields and as a cryogenic liquid. At 337 locations in the U.S. natural gas is stored in underground porous-rock formations with a total capacity of 5,681 billion cubic feet. Whether hydrogen can be stored in underground porous rock can be finally ascertained only by future field trials. At present, however, 30 billion cubic feet of helium, a low-density gas with leakage characteristics similar to those of hydrogen, is stored quite satisfactorily in an underground reservoir near Amarillo, Tex.

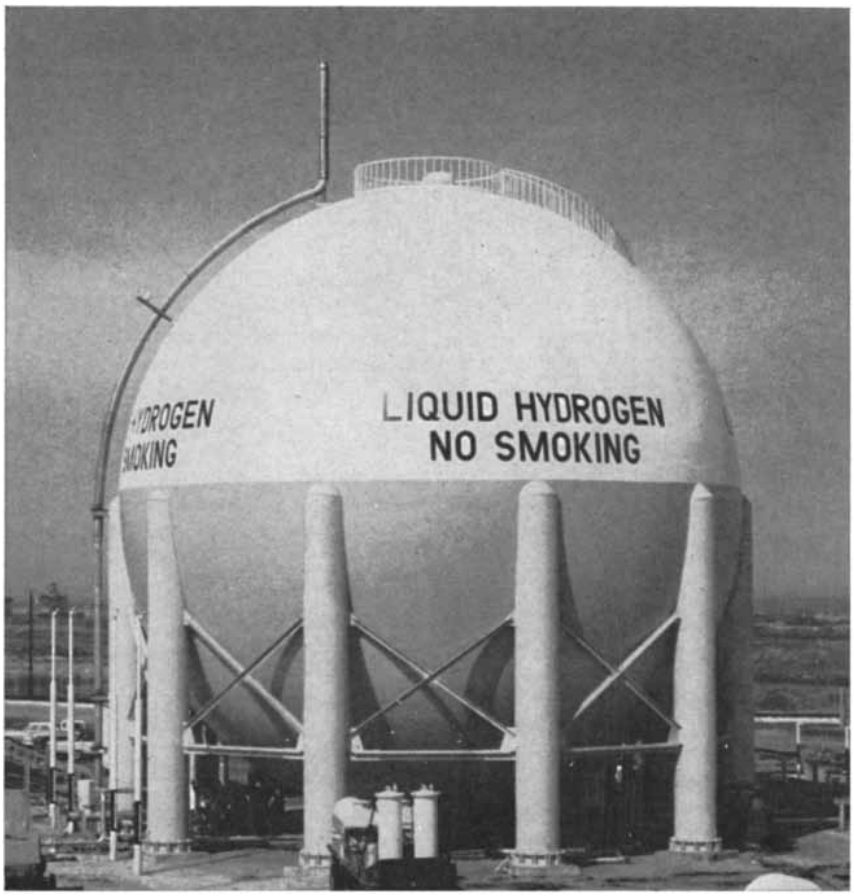
Cryogenic storage of natural gas is a rapidly growing technique; at 76 locations in the U.S. "peak shaving" operations involving liquefied natural gas are in use or under construction. There is no technical reason why a similar peak-shaving technique cannot be employed with liquid hydrogen. Liquid hydrogen used to be considered a hazardous laboratory curiosity, but it is already being used as a convenient means of storing



ENVIRONMENTAL EFFECTS of the present fossil-fuel energy cycle and the proposed hydrogen-fuel energy cycle are compared here. When fossil fuels are burned to release their stored energy (*top*), the environment is relied on to accommodate the combustion by-products. The re-formation of the fossil fuels from atmospheric carbon dioxide takes millions of years (*broken line*). On the other hand, when hydrogen is burned as a fuel (*bottom*), the only combustion product is water, which is easily assimilated by the environment. The fuel cycle is completed rapidly without depleting limited resources or accumulating harmful waste products.

and transporting hydrogen over long distances. Liquid hydrogen is regularly shipped around the U.S. in railroad tank cars and road trailers. The technology for the liquefaction and tankage of hydrogen has already been developed, mainly for the space industry. Indeed, the largest liquid-hydrogen storage tank is at the John F. Kennedy Space Center; it has a capacity of 900,000 gallons, equivalent to 37.7 billion B.t.u. or 11 million kilowatt-hours [*see illustration at right*]. Although the energy content of this tank is only about 4 percent of the energy content of a typical liquid-natural-gas peak-shaving plant, its energy capacity is 73 percent of the capacity of the world's largest pumped-storage hydroelectric plant, located at Ludington, Mich.

The cryogenic approach to energy storage has the advantage of being applicable in any location, no matter what the geography or geology, factors that limit both underground gas storage and pumped hydroelectric storage.



ENERGY STORAGE in the form of liquefied hydrogen is already a routine practice in the space industry. This vacuum-insulated cryogenic tank at the John F. Kennedy Space Center, for example, contains 900,000 gallons of liquid hydrogen for fueling the Apollo rockets. It is the largest facility of its kind in existence. In terms of energy its contents are equivalent to 37.7 British thermal units (B.t.u.) of heat or 11 million kilowatt-hours of electricity.

The simplest way to manufacture hydrogen using nuclear energy is by electrolysis, a process in which a direct electric current is passed through a conductive water solution, causing it to decompose directly into its elementary constituents: hydrogen and oxygen. Complete separation of the two gases is achieved, since they are evolved separately at the two electrodes. Salts or alkalis, which have to be added to the water to increase conductivity, are not consumed; thus the only input material required is pure water.

A number of large-scale electrolytic hydrogen plants are operated today in locations where hydrogen is needed (for example in the manufacture of ammonia and fertilizers) and where cheap electric power (usually hydroelectric power) is available. One of the largest commercial electrolyzer plants in the world is operated by Cominco, Ltd., in British Columbia [*see illustration on page 20*]. This plant consumes about 90 megawatts of power and produces about 36 tons of hydrogen per day for synthesis into ammonia. The by-product oxygen is used in metallurgical processes. Similar large plants are located in Norway and Egypt. Many smaller plants exist where hydrogen is produced from unattended equipment.

The theoretical power required to produce hydrogen from water is 79 kilowatt-hours per 1,000 cubic feet of hydrogen gas. In practice the large industrial plants are only about 60 percent efficient; a typical power-consumption figure is 150 kilowatt-hours per 1,000 cubic feet of hydrogen. This power requirement represents a major part of the plant's operating cost. Thus there is a considerable incentive—indeed, a real need—to increase the operating efficiency of such plants if one is to consider using electrolytic hydrogen as a fuel.

The fuel cell, the subject of intensive research and development as part of the space program over the past 15 years, is really an electrolyzer cell operating in reverse. The simplest fuel cell to build and operate is one that operates on hydrogen and oxygen, yielding water and electric power as its products. Hydrogen-oxygen fuel cells were selected and developed for both the Gemini and the Apollo programs because of their high efficiency, which reduces the amount of fuel needed aboard the spacecraft to supply its electric power. Much effort has gone into developing fuel cells with high efficiencies. This same technology can be applied to increase the efficiency of the reverse process: electrolysis. Electrolytic cells are operating in aerospace laboratories today with an efficiency of more than 85 percent.

Increasing the electrolyzer efficiency alone has relatively little merit as long as the present power-station efficiency in converting nuclear heat to electric power is only about 33 percent. This efficiency loss can, however, also be circumvented. For example, Cesare Marchetti at the Euratom laboratories in Italy has designed a chemical process for the thermal splitting of water to hydrogen and oxygen directly using the heat energy produced by a nuclear reactor. If water is to be split into its elements directly, it must be heated to very high temperatures—about 2,500 degrees Celsius—to achieve dissociation. Not only are such temperatures not available from nuclear reactors but also the gases cannot conveniently be separated from each other before they recombine. It is possible to conceive of a two-stage reaction in which a metal, say, reacts with steam at

a reasonable temperature to produce hydrogen and a metal oxide. The hydrogen is easily separated from the metal oxide, which in turn could be decomposed to oxygen and the metal by the application of heat. Unfortunately there does not appear to be any suitable metal that undergoes such a series of reactions at temperatures low enough to be compatible with nuclear reactors, whose construction materials limit operating temperatures to about 1,000 degrees C.

Marchetti's concept, therefore, is a far more complex reaction sequence involving calcium bromide ($CaBr_2$), water (H_2O) and mercury (Hg), in which, except for the hydrogen and oxygen, all the reactants are recycled. Each of the reactions proceeds at temperatures below 730 degrees C., which can be achieved in a nuclear reactor. Although the process appears to be feasible, development work is still required to try to bring the overall efficiency up and the cost down to practical limits.

The quantities of hydrogen that the hydrogen economy would require are immense. For example, if we were to produce today an amount of hydrogen equivalent to the total production of natural gas in the U.S., we would have to provide during one year the same fuel value as 22.5 trillion cubic feet of gas, or 22.5 quadrillion (10^{15}) B.t.u. of energy. This corresponds to about 70 trillion cubic feet of hydrogen, which, if we could produce it at a steady rate all year round from nuclear electrolytic plants, would require an electrical input of more than a million megawatts. The present total electrical generating capacity in the U.S. is 360,000 megawatts, so that we are envisioning a fourfold increase in generating capacity, which would require the construction of more than 1,000 new 1,000-megawatt power stations. That is in addition to the rapidly increasing demand for electric power for other uses. During the past five years, in contrast, the electrical generating capacity in the U.S. has grown by "only" 105,000 megawatts.

Such a formidable task of increasing capacity, however, does not follow solely from our turning to a hydrogen economy. As our huge consumption of fossil fuels declines in future years, we must provide at least an equivalent alternative energy source. Such numbers give a taste of the energy revolution that must take place within the next half-century.

At present the cheapest bulk hydrogen is made from natural gas. Clearly since hydrogen from such a source cannot be cheaper than the starting material, it cannot therefore be expected to replace natural gas as a fuel. Electrolytic hydrogen is even more expensive, unless very cheap electric power is available. Today's electricity prices are based on supplying a fluctuating load, but the capability of hydrogen storage would even out the load and might reduce the price of electricity somewhat.

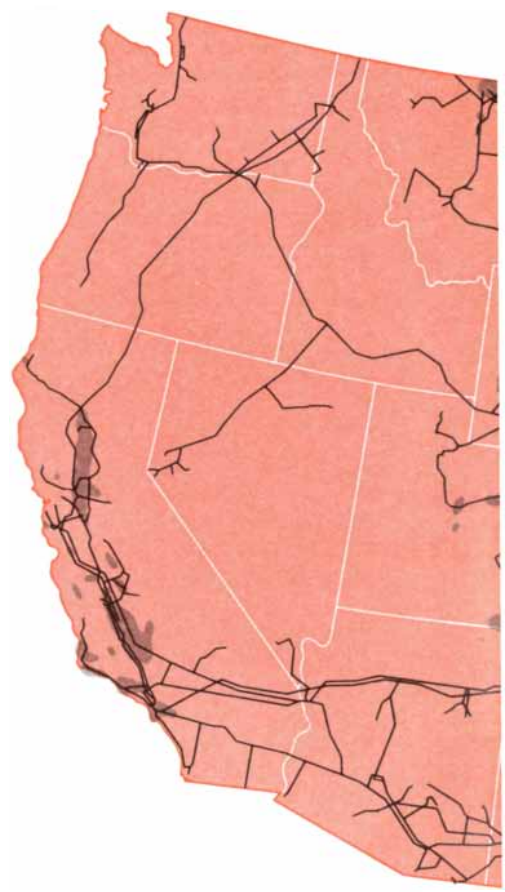
Although the cost of hydrogen produced from electricity must always be higher than the cost of the electricity, it is the lower transmission and distribution cost of hydrogen compared with electricity that makes it advantageous to the user. The latest economic figures published by the gas and electrical industries can be used to derive the production, transmission and distribution shares of average prices, charged to all types of customers, for gas and electricity, and these data can be compared in turn with corresponding figures for hydrogen made by electrolysis [*see illustration on page 21*]. The figures for hydrogen are derived from the hypothetical assumption that all the electricity generated in the U.S. in 1970 was converted to hydrogen, which was sent through the existing natural-gas transmission network (for an average distance of 1,000 miles) and was delivered to customers as a gaseous fuel. The electrolysis charge of 56 cents per million B.t.u. is derived from AEC estimates of the cost of building advanced electrolyzer cells. The hydrogen transmission and distribution costs are based on natural-gas costs, adjusted to take account of the different physical properties and safety factors for handling hydrogen.

Two things are obvious from such a comparison. One is that today it is far cheaper for the average customer to buy energy in the form of natural gas than it is in the form of electricity. The other is that it should already be possible to sell hydrogen energy to the gas user at a lower price than he now pays for electricity. Clearly, however, this hydrogen will find no markets while natural gas is as cheap as it is.

Looking to the future, we see that natural-gas prices, together with all fossil-fuel prices, will increase rapidly. These rises are brought about by their short supply, by the influence of pollution regulations and by such social pressures as land conservation and employee welfare applied to the mining industry. In contrast, the price of nuclear energy, although apparently rising fast now, can be expected to stabilize somewhat in the breeder-reactor era because there will then be no severe supply limit.

It is not possible at this time to forecast accurately what the cost of hydrogen energy is likely to be, but one can certainly look forward to considerably increased prices for all forms of energy. Even so, in the long run delivered hydrogen will be cheaper than delivered natural gas and very probably also cheaper than delivered electricity.

When hydrogen becomes as universally available as natural gas is today, it will easily perform all the functions of natural gas and others besides. Hydrogen can be used in the home for cooking and heating and in industry for heating; in addition it can serve as a chemical raw material in many industries, including the fertilizer, foodstuffs, petro-



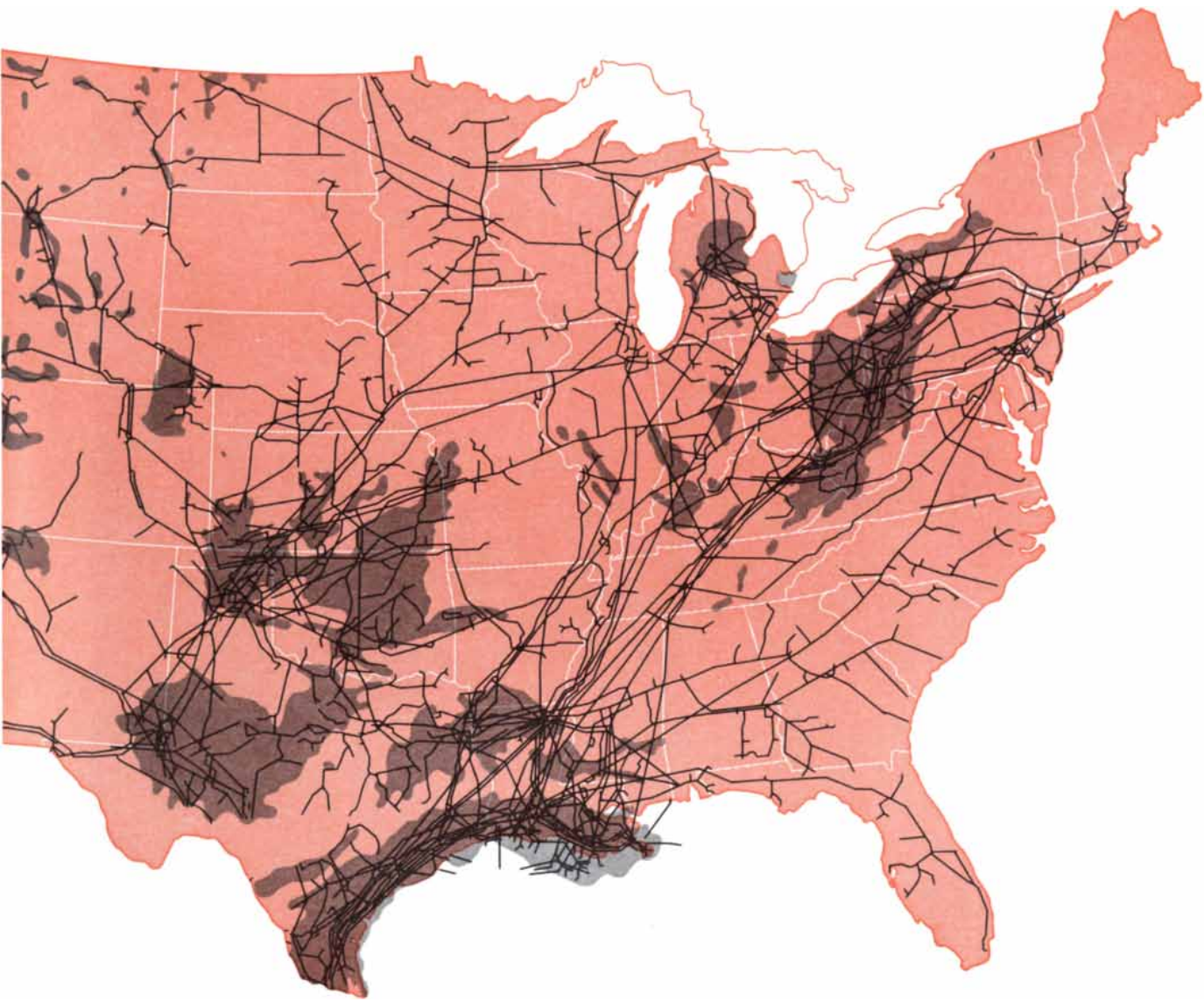
TRUNK PIPELINES extending for 252,000 miles (*black lines*) already exist in the U.S. for transmission of natural gas from areas

chemical and metallurgical industries. Hydrogen can also be used to generate electricity in local power stations.

The combustion properties of hydrogen are considerably different from those of natural gas. Hydrogen burns with a faster, hotter flame, and mixtures of hydrogen with air are flammable over wider limits of mixture. These factors mean that burners of hydrogen must be designed differently from those of natural gas and that modification of every burner will be necessary on changeover. Such widespread modification is not without precedent. A similar operation was carried out when the U.S. changed from manufactured gas (about 50 percent hydrogen) to natural gas; several European countries have recently undertaken the same conversion.

Hydrogen, because it burns without noxious exhaust products, can be used in an unvented appliance without hazard. Hence it is possible to conceive of a home heating furnace operating without a flue, thereby saving the cost of a chimney and adding as much as 30 percent to the efficiency of a gas-fired home heating system. More radical changes are possible, moreover, because without the need for a flue the concept of central heating itself is no longer necessary. Each room can have its heat supplied by unflued peripheral heating devices operating on hydrogen independently of one another. Indeed, the vented water vapor would provide beneficial humidification. Another radical change is the potential use of catalytic heaters. Since hydrogen is an ideal fuel for catalytic combustion, true "flameless" gas heating is possible, with the catalytic bed being maintained at any desired temperature, even as low as 100 degrees C. This prospect promises to revolutionize domestic heating and cooking techniques in the future. With such low temperatures it is virtually impossible to produce nitrogen oxides, thus eliminating the only possible pollutant from a hydrogen system.

Hydrogen is also the ideal fuel for fuel cells. The technological problems that have faced the development of practical, commercially economical fuel cells for

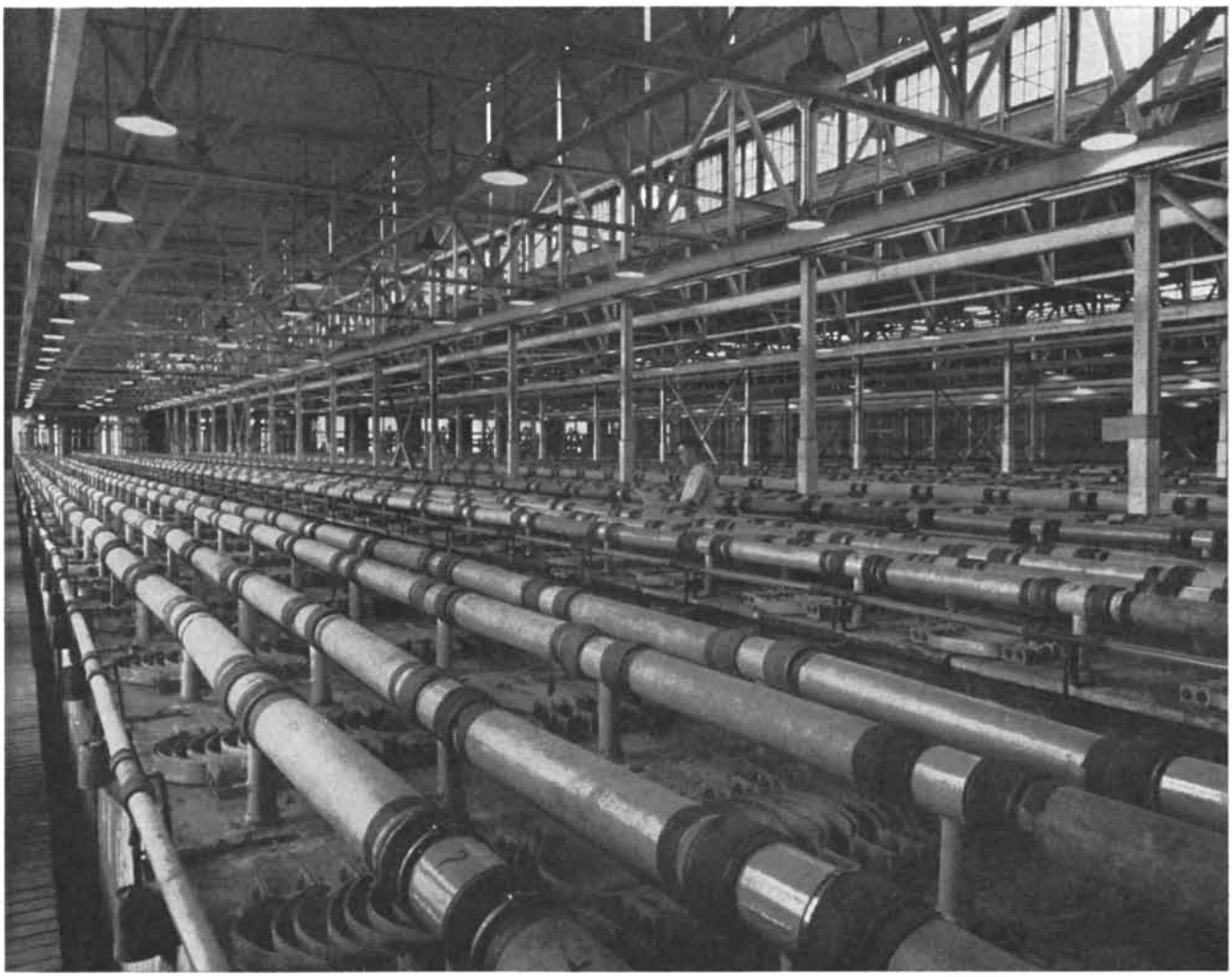


where the gas is produced (*gray*) to areas where it is consumed. The system, which is constructed almost entirely of welded steel pipe, carries approximately 61.4 billion cubic feet (or 1.5 million tons) of natural gas per day. Similar networks of underground hydrogen-gas pipelines would enable the giant nuclear (or solar) power stations of the future to be located far from the load centers.

more than a decade are very much reduced if hydrogen can be used as fuel. Fuel-cell electricity generators operating on hydrogen should be at least 70 percent efficient and can realistically be expected to find a place in the home, in commercial and industrial buildings and in industry. Larger, urban electrical generating stations could be fuel-cell systems or could be hydrogen-fueled steam stations. An earlier concept of operating a closed-cycle steam-turbine system on a hydrogen-oxygen fuel supply could become practical through the use of rocket-engine technology. Workers at the Massachusetts Institute of Technology have proposed such a system for submarines; it has been reported that an overall efficiency of 55 percent can be anticipated from it.

Hydrogen is an excellent fuel for gas-turbine engines and has been proposed as a fuel for supersonic jet transports. For this kind of use fuel storage and tankage as liquid hydrogen are practical. Although the large volume required may make its use less attractive for subsonic aircraft, the very considerable saving in weight over an equivalent fuel load of kerosene gives hydrogen a distinct advantage. Conventional internal-combustion engines will also operate on hydrogen if they are suitably modified or redesigned. R. J. Schoeppel of Oklahoma State University and others have shown that if hydrogen is injected into the engine through a valve in a manner similar to the way fuel is injected into a diesel engine, the preignition characteristics of hydrogen are overcome. Others, including Marc Newkirk of the International Materials Corporation and Morris Klein of the Pollution Free Power Corporation, have reported satisfactory operation of conventional automobile engines on hydrogen using carburetor and manifold modifications. Meanwhile William J. D. Escher of Escher Technology Associates has proposed a radically different approach to automobile engine design, using a steam system fueled by both hydrogen and oxygen. The use of liquid hydrogen as a routine private-automobile fuel is questionable on the ground of safety, although it is probably applicable to fleet users, such as bus lines and taxicab fleets.

Richard H. Wiswall, Jr., and James J. Reilly of the Brookhaven National Laboratory have proposed the use of metallic hydrides to store hydrogen as a fuel for vehicles. A magnesium-alloy hydride will store hydrogen energy as efficiently (on a weight basis) as a tank of liquid hydrogen, but some technical problems must still be overcome. At present there seems to be no single, obvious way in which automobiles can be operated on hydrogen fuel, but considerable work is



LARGE ELECTROLYZER PLANT for the production of hydrogen by the electrical decomposition of water is operated by Cominco, Ltd., in British Columbia. The 3,200 electrolytic cells, which cover more than two acres, consume about 90 megawatts of power and produce about 36 tons of hydrogen per day for synthesis into ammonia. By-product oxygen is used in metallurgical processes.

going on to investigate the various options available. If one has to synthesize a suitable liquid fuel for automobiles and aircraft, the starting material for the fuel must be hydrogen in any case.

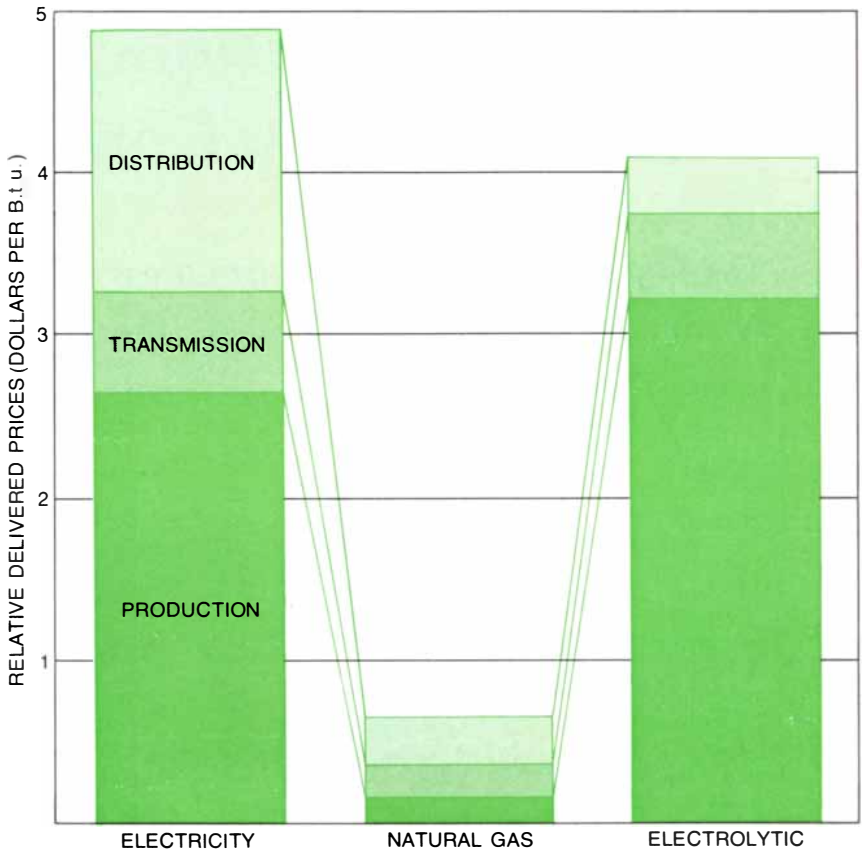
One of the main criticisms of the hydrogen-economy concept is that hydrogen is too dangerous for use in this way. Undoubtedly hydrogen is a hazardous material and must be handled with all due precautions. If it is handled properly, however, in equipment designed to ensure its safety, anyone should be able to use it without hazard.

In the days of manufactured gas (gas made from coal), which consisted of up to 50 percent hydrogen and contained about 7 percent carbon monoxide, people managed to live with the fire and explosion hazards of hydrogen as well as the toxic hazards of carbon monoxide. Of course, it takes only one major disaster to alert everyone to a hazard. The most famous hydrogen accident, the *Hindenburg* airship disaster of 1937, is still remembered with awe. Indeed, the almost universal fear of hydrogen has been described as the "*Hindenburg* syndrome." Spectacular as it was, however, that fire was almost over within two minutes, and of the 97 persons on board, 62 survived.

Very strict codes are enforced for the use of natural gas today; even stricter ones are applied to industry for the use of hydrogen. Most of these codes are realistically based on reducing the chances of accidents. Just as we have designed apparatus and procedures to enable us to fill our automobile tanks with gasoline and carry the resulting 20-gallon "fire bomb" at speeds of up to 70 miles per hour along a crowded highway and park it overnight right inside our homes, we can surely devise safe practices for handling hydrogen.

Hydrogen cannot be detected by the senses, so that a leak of pure hydrogen is particularly hazardous. Odorants are routinely used to make natural-gas leaks obvious, however, and no doubt the same can be done with hydrogen. Hydrogen flames are also almost invisible and are therefore dangerous on this score. Hence an illuminant may have to be added to the gas to make the flame visible. The flammability limits of hydrogen mixed with air are very wide, from 4 to 75 percent. It is the lower limit, almost the same as that for methane (5 percent in air), that causes the fire hazard with a gas leak. On the benefit side, however, since hydrogen is so much lighter than air and diffuses away at a far greater rate than methane, a hydrogen leak could actually be less hazardous than a natural-gas leak. The most significant hazardous property of hydrogen is the extremely low energy required to ignite a flammable mixture: only a tenth of the energy required to ignite a gasoline-air mixture or a methane-air mixture and well within the energy levels of a spark of static electricity (a probable cause of the *Hindenburg* fire, which occurred just after a thunderstorm). Thus safety practices will have to be based on the assumption that if a hydrogen fire can occur, it will! Huge quantities of hydrogen are handled in industry quite safely and without accident precisely because proper precautions are taken.

To recapitulate briefly, our recoverable fossil-fuel supplies will sooner or later become exhausted; we are already feeling the effects of the limited supply by having to pay more for fossil-based energy. Within the next 50 years we must be prepared to pay considerably more for energy from all sources, particularly for fossil fuels. One way of handling nuclear and other energy sources is to use them to convert water to hydrogen in large central plants and then to use hydrogen as a clean, nonpolluting fuel. Technically this is already feasible; only relatively simple developments have to be made, not approaching the magnitude of the technical tasks of developing the alternative energy sources—breeder reactors and solar engines—themselves. Economics and safety are the two obstacles to developing such a hydrogen economy. A combination of technical development and the expected adjustment in relative energy prices can justify the economics, and proper practices and design can ensure safety. If and when we move into a hydrogen economy, the world will undoubtedly be a far cleaner place to live in than it is today.



RELATIVE DELIVERED PRICES of various forms of energy are broken down in this bar chart into the shares represented by production (*solid color*), transmission (*intermediate color*) and distribution (*light color*). The comparison reveals that at present it is much cheaper to buy energy in the form of natural gas than in the form of electricity. Moreover, the breakdown shows that although the cost of hydrogen produced from electricity must always be higher than the cost of the electricity, the lower transmission and distribution costs of hydrogen already make it possible to sell hydrogen energy to the gas user at a delivered price lower than what he now pays for electricity. It is expected that natural-gas prices, together with all fossil-fuel prices, will increase rapidly in the future.

How the Immune Response to a Virus Can Cause Disease

The body's defense mechanism may not always be beneficial. In many cases the very process that should combat a virus is itself a cause of the damage associated with a viral disease

by Abner Louis Notkins and Hilary Koprowski

In biology as in physics it is a truism that the deeper one goes in exploring elementary questions, the more one encounters paradoxical and puzzling phenomena. Such is the case in the investigation of the well-known animal defense mechanism called the immune response, whereby the body fights off infections and other invasions by foreign matter. The classic concept of this mechanism is quite simple: in response to invasion by the foreign substance, or antigen, the body produces specific antibodies that bind to it, thus neutralizing the invader so that it does not harm the invaded organism. Investigators are now learning, however, that this is far from the entire story, that the mechanism of immunity is much more complex than had been supposed.

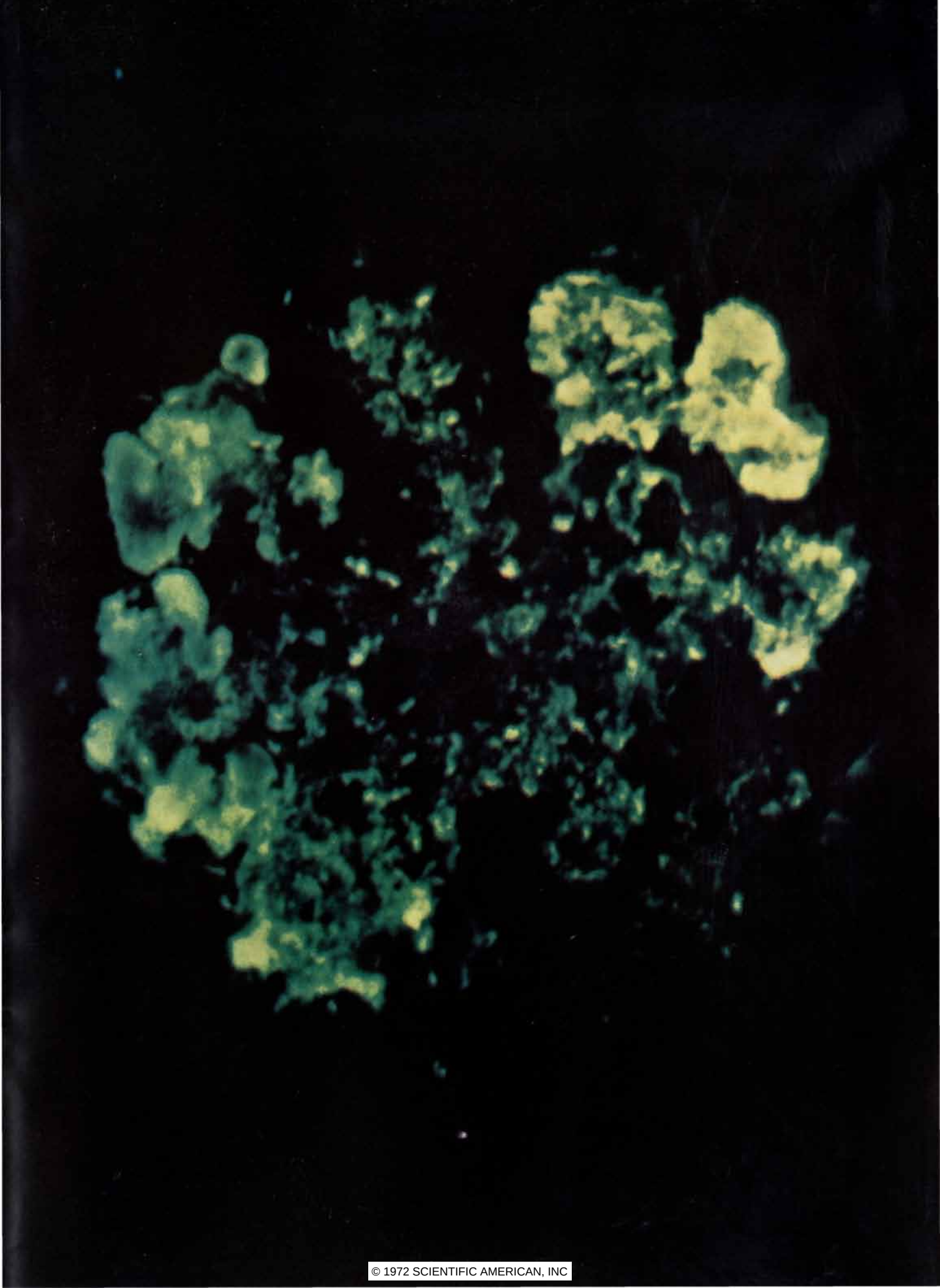
This is true in particular in virus infections. In the simple case of direct attack by a virus (for example in poliomyelitis) the virus invades a cell and uses the cell's material to replicate, and soon the new crop of viruses bursts the cell and emerges to go on to infect other cells. The timely appearance of antibodies may prevent the spread of infection and the appearance of symptoms. In infections caused by other viruses, however, there is growing evidence that the cells are damaged not directly by the replicating virus but by a specific immune response that produces the symptoms of the disease. The complexities of the immune response to viruses are under exploration in a number of laboratories, including our own at the National Institutes of Health (Notkins) and at the Wistar Institute of Anatomy and Biology (Koprowski). Gradually an account of the immunity mechanism's diverse operations is being pieced together, and what follows is a review of the emerging picture.

That the immunity system might sometimes be responsible for injurious effects was first suggested more than 60 years ago by Clemens von Pirquet, an Austrian pediatrician who was at one time also a professor at the Johns Hopkins School of Medicine. Von Pirquet noted that in "serum sickness," a disease that can follow injection of foreign blood serum, the patient's blood contained foreign proteins and antibody against them. He speculated that the combination of antibody with foreign proteins (antigen) perhaps produced a toxic substance that gave rise to the symptoms of the sickness: hives, rash, pain in the joints, shortness of breath and, in severe cases, death. He also conjectured that an interaction of antibodies with the viruses of such diseases as smallpox and measles might cause the skin eruptions characteristic of these diseases.

Von Pirquet's speculations that immune response to viruses might cause disease were not followed up at the time, but in the 1950's Wallace P. Rowe, a virus investigator at the National Institutes of Health, came on proof of the hypothesis in an ingenious series of experiments. Rowe was studying the pathology produced by a virus known as lymphocytic choriomeningitis (LCM), which infects rodents and occasionally man and causes an inflammation of the membranes surrounding the brain (meningitis). He observed that although in infected mice the virus multiplied rapidly in many organs, the animals at first showed no sign of illness. On the sixth day, however, after the mice had begun to show an immune response to the virus, they developed meningitis and died. Was the disease caused by their immunological response to the virus rather than by the virus itself? Reasoning that if he inhibited the immune response, he might be able to prevent the disease, Rowe treated mice with X rays in doses known to suppress the immune response. He then infected both the treated mice and untreated control mice with LCM virus. The irradiated animals did not develop meningitis, although the virus replicated in their tissues just as rapidly as it did in the control mice, which died.

Later a group of investigators at the Johns Hopkins School of Hygiene and Public Health (Donald H. Gilden, Gerald A. Cole, Andrew A. Monjan and Neal Nathanson) took Rowe's experiments a step further. It was known by this time that the immune system responds to foreign substances in at least two ways, one mediated by antibody and the other mediated by a specific group of the cells known as lymphocytes. These "immune lymphocytes" recognize antigens on the surface of foreign cells

IMMUNE COMPLEXES are formed when an antiviral antibody combines with a virus and binds complement. The complexes are detected by a technique in which an antibody to complement is labeled with a fluorescent substance and incubated with tissue; if complexes are present, the antibody binds to them and fluoresces under ultraviolet radiation. The photomicrograph on the opposite page, made by David D. Porter of the University of California Center for Health Sciences, demonstrates the presence of complexes in the kidneys of mink that were infected with Aleutian virus and developed glomerulonephritis.

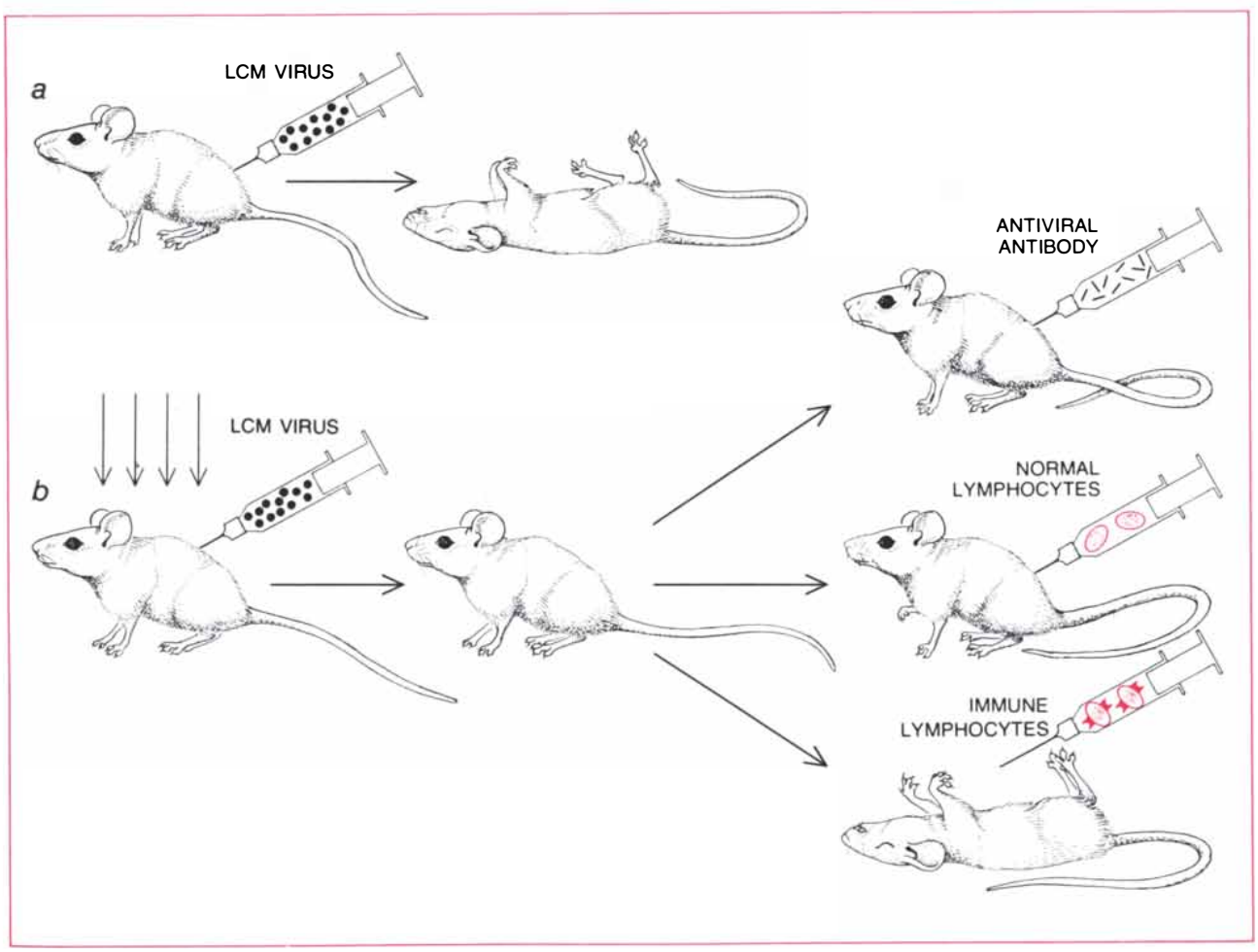


and thereby destroy tissue such as tumors or skin grafts [see "Markers of Biological Individuality," by Ralph A. Reisfeld and Barry D. Kahan; SCIENTIFIC AMERICAN, June, 1972]. Which of these factors was responsible for causing LCM disease in Rowe's experiments: antibodies or lymphocytes? The Johns Hopkins group used drugs to suppress the immunological response in mice, infected them with LCM virus and then divided the animals into three groups. One group received injections of anti-LCM antibody, the second was given anti-LCM lymphocytes and the third normal lymphocytes. The animals receiving the antibody or normal lymphocytes remained well but those given the immune lymphocytes developed the symptoms of LCM disease and died. Evidently in the case of LCM it was the combination of immune lymphocytes and the virus that produced the disease.

Extending their observations, the Johns Hopkins group found that in young rats LCM infection was not fatal but did damage the cerebellum, causing ataxia (inability to coordinate body movements). If the immune response was suppressed at the time of infection, however, the animals remained free of symptoms and cerebellar damage did not occur, even though the virus continued to replicate in the brain. As in the case of mice, development of the rats' disease was thus shown to be immunologically mediated. An interesting suggestion from these experiments is that perhaps other neurological disorders may arise from the immune response to viruses.

It was now time to look into the reasons why lymphocytes destroyed infected cells when the virus itself did not. In order to study this problem Duard L. Walker and his co-workers at the University of Wisconsin Medical School turned from experiments in animals to experiments in tissue culture. It was known that on infecting a cell some viruses induce the formation of viral antigens on the cell's surface. Walker reasoned that if these antigens are recognized by lymphocytes from animals immunized with the same virus, the lymphocytes might attack and destroy cells carrying the label of infection. To test this hypothesis he infected tissue-culture cells with mumps virus, which induced new antigens on the surface of the cells but did not destroy them. When he introduced into the infected cultures lymphocytes taken from animals that had been immunized with that virus, the lymphocytes did indeed destroy the infected cells. On the other hand, lymphocytes from animals that had not been immunized with the mumps virus did not attack the infected cells.

Other investigators soon obtained the same kind of result in tissue-culture experiments with LCM virus and the measles virus. A number of groups are now looking into the possibility that the interaction between immune lymphocytes and viral antigens formed on the surface of infected cells may account for some



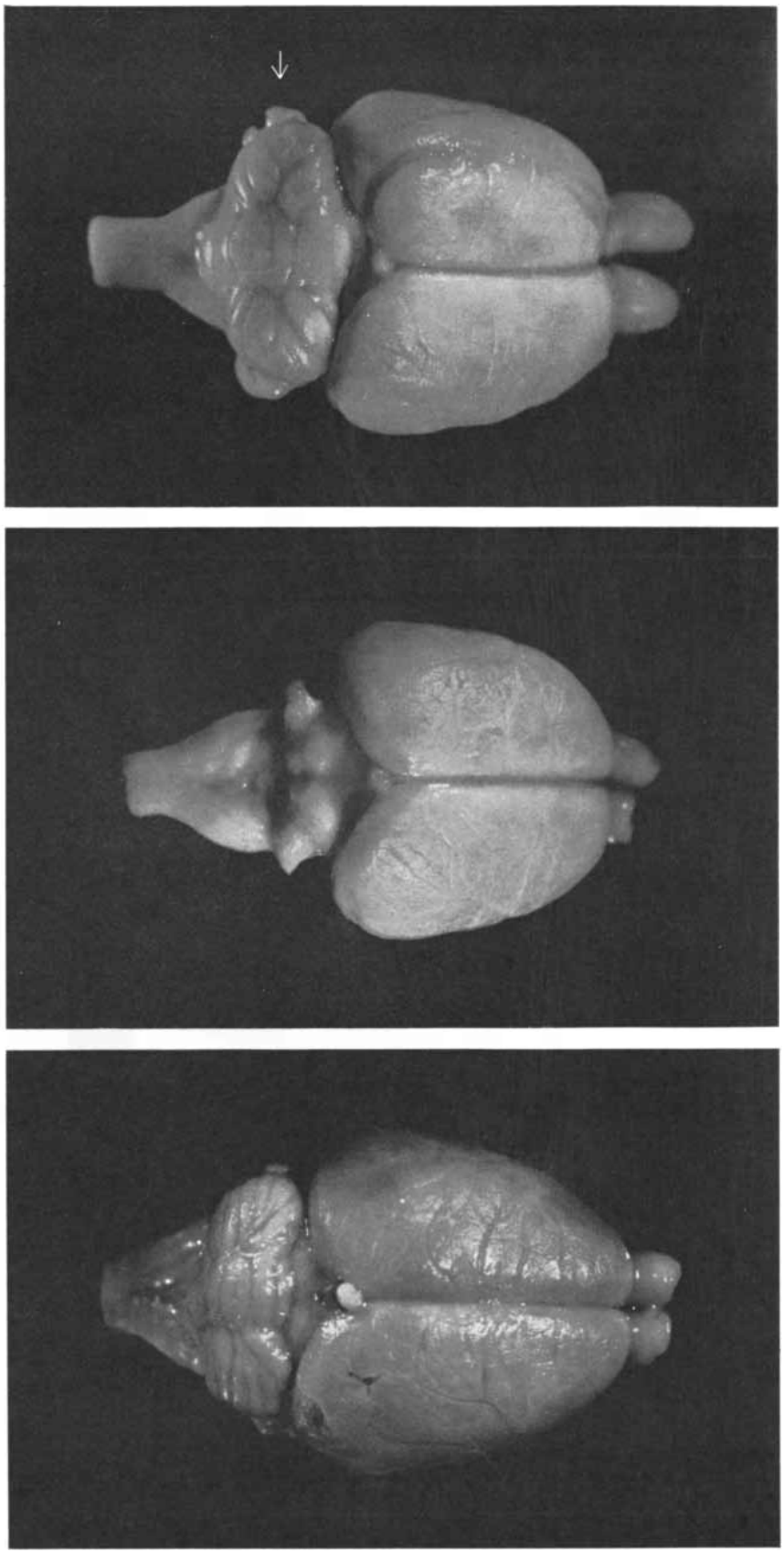
LCM VIRUS kills adult mice (*a*). If the immune response is suppressed by radiation or drugs (*b*), the mouse lives but develops a chronic infection. When immunological competence is restored by injecting lymphocytes from other animals immunized with LCM, test animal dies. Injection of anti-LCM antibody or of normal lymphocytes rather than anti-LCM lymphocytes does not cause death.

of the symptoms associated with viral diseases of man, including hepatitis.

Gary Rosenberg and Paul Farber in Notkins' laboratory at the National Institutes of Health undertook an even more detailed analysis of the behavior of lymphocytes in response to viral antigens. They used herpes simplex virus (HSV), which produces the familiar cold sores in man. They found that when lymphocytes from animals immunized with this virus were incubated in a test tube with HSV antigen, the antigen "turned on" the lymphocytes to replicate their DNA and divide. This reaction began within hours after exposure to the antigen and was quite specific: the anti-HSV lymphocytes were turned on only by the HSV antigen; they did not react at all to antigens of other viruses. In a follow-up study at the Wistar Institute with the rabies virus, Tadeusz J. Wiktor and Koprowski found that lymphocytes from rabbits immunized with that virus could be turned on not only by the complete virus but also by its subunits.

Further research showed that when lymphocytes are stimulated by exposure to viral antigens, they release potent chemical messengers, or mediators, that exhibit a variety of biological properties and are thought to be responsible for some of the inflammatory change and tissue injury associated with many viral infections. One of these mediators is known to attract inflammatory (white) cells and another can keep the inflammatory cells at the site of the infection. A third mediator, lymphotoxin, can destroy uninfected as well as infected cells, and a fourth is the now well-known substance interferon, which can inhibit the replication of viruses. Very likely a number of other mediators will be found to be released by the interaction of viruses with lymphocytes; mediators with at least a dozen different biological properties have been discovered in cultures of lymphocytes that are turned on by nonviral antigens.

If lymphocytes act as agents of tissue destruction and disease, might not antibodies also perform such a role? Mario Fernandas, Wiktor and Ernest Kuwert in Koprowski's laboratory began to explore the antibody phase of the immunity phenomenon. It had been known for some time that the attachment of antibody to antigens on the surface of cells could activate a group of proteins in the serum, known as complement, to break down cells. It was also known that under certain circumstances rabies virus could induce new antigens on the surface of tissue-culture cells without

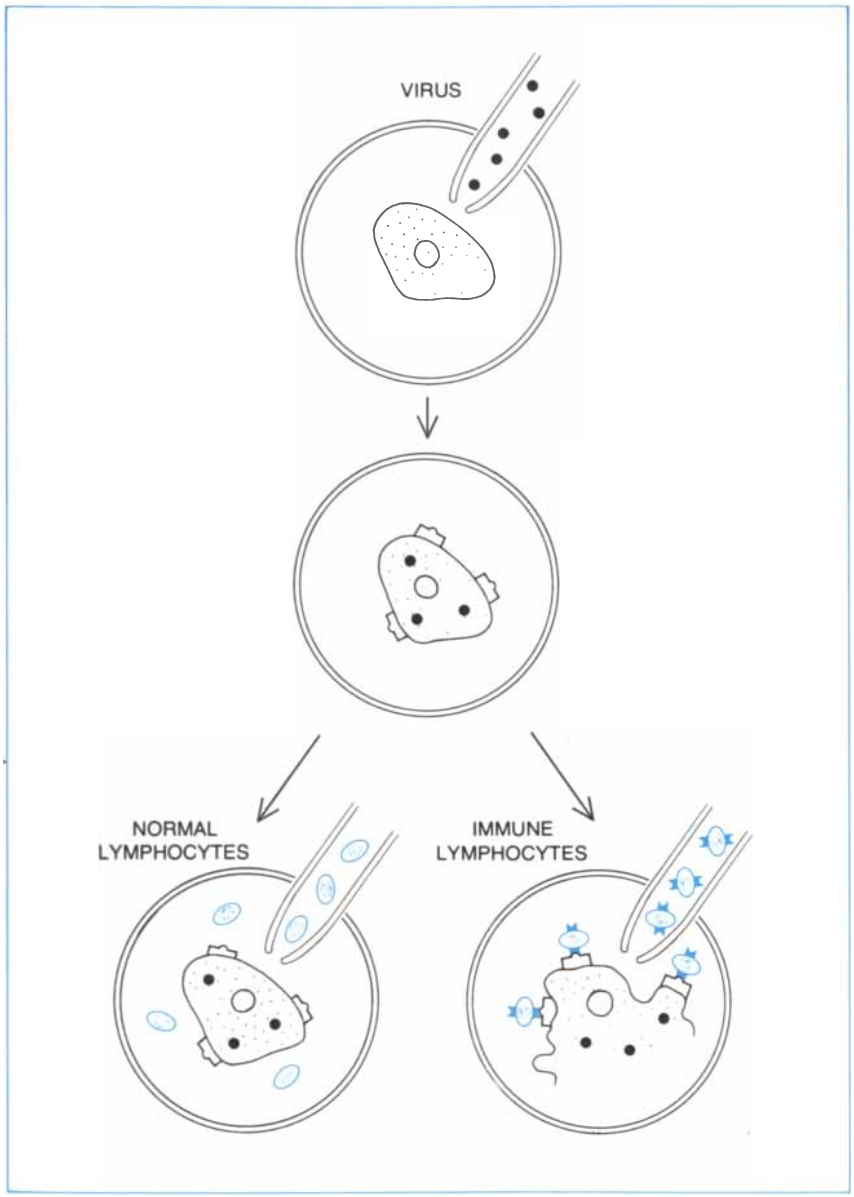


RAT BRAINS show the effects of an immune response to LCM infection demonstrated by Andrew A. Monjan, Gerald A. Cole and Neal Nathanson of the Johns Hopkins School of Hygiene and Public Health. The cerebellum is at the base of the brain (*top*). LCM infection produced severe cerebellar damage (*middle*), impairing the ability to coordinate movements. If the cellular immune response was suppressed, there was no brain damage (*bottom*).

destroying them. The Wistar Institute investigators added antirabies antibody or complement or both to the infected cultures. Neither the antibody nor the complement alone was injurious to the cells, but when they were added together the combination destroyed the rabies-infected cells.

Recent experiments suggest that antibody and complement also may contribute to the breakdown of cells infected with some of the well-known tissue-destroying viruses. Charles Wohlenberg, Arnold Brier and Joel Rosenthal of the National Institutes of Health laboratory showed that viral antigens on the surface of cells infected with influenza, measles, vaccinia and HSV make the cells vulnerable to destruction by specific antibody and complement long before these cells break down as the direct result of viral replication. It seems highly likely, therefore, that in the body the symptoms and other effects of these diseases are produced by a collaboration between the immunological process and the virus.

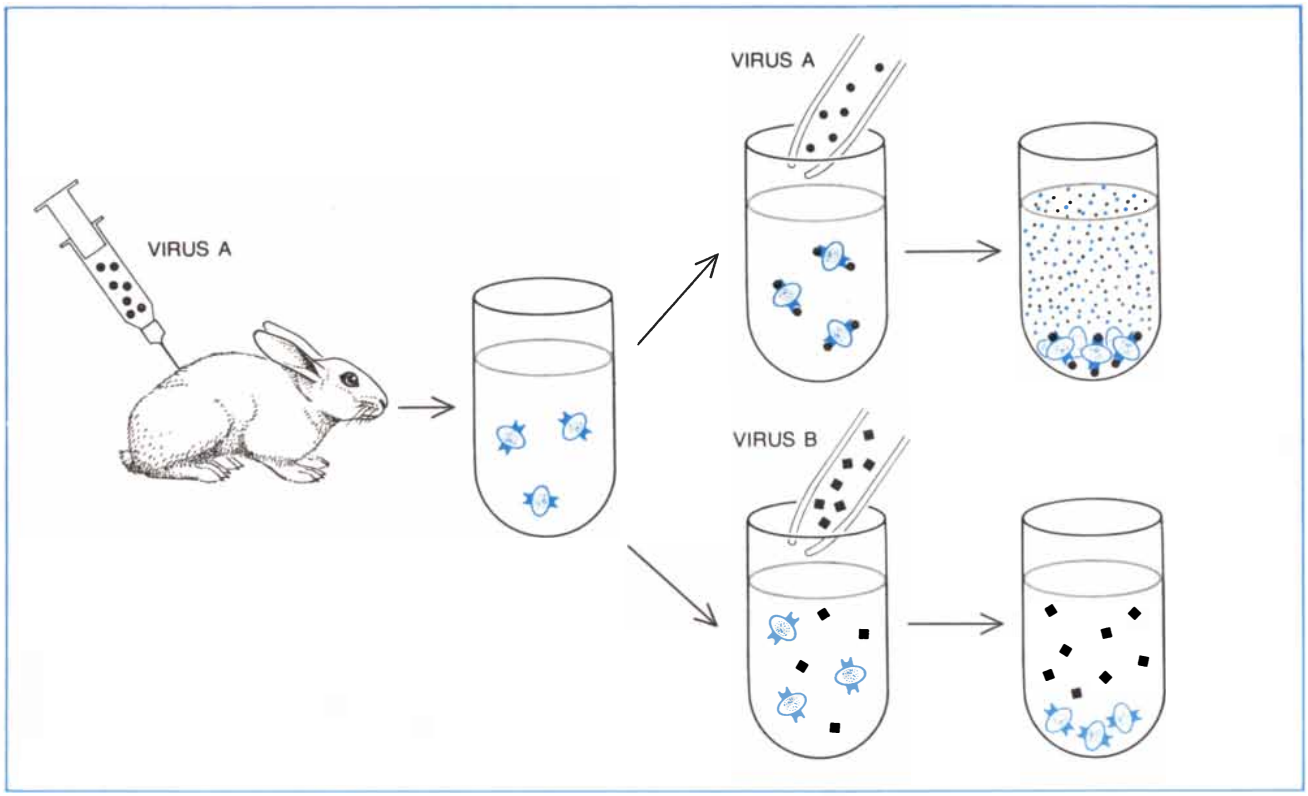
How does the interaction of complement with antibody bring about tissue damage and inflammation in the infected animal? Ralph Snyderman of the National Institutes of Health laboratory conducted an experiment that suggested a likely answer. He found that when complement and antibody to HSV were added to a culture containing cells infected with that virus, a mediator was released from one of the components of complement. This mediator was identified as one that had previously been shown to increase the permeability of blood vessels and to attract white cells. Although white cells are known to be important in the defense against certain infections, they contain potent tissue-destroying enzymes and so they may also act as agents of cell destruction in some viral infections. In fact, several groups of investigators have found that the injection of specific antibody into virus-infected animals has the effect of increasing the number of white cells and the amount of tissue damage in the infected organs. Now it appears that the sometimes fatal shock syndrome associated with dengue fever, a viral disease in Southeast Asia, may be mediated by antibody and complement. Scott B. Halstead of the University of Hawaii Medical School and Philip K. Russell of the Walter Reed Army Institute of Research, who originally proposed the idea, believe mediators released from complement may be one of the factors responsible for the increase in permeability of blood vessels and the consequent shock that marks this disease. Several workers are now attempting to gather proof for this hypothesis.



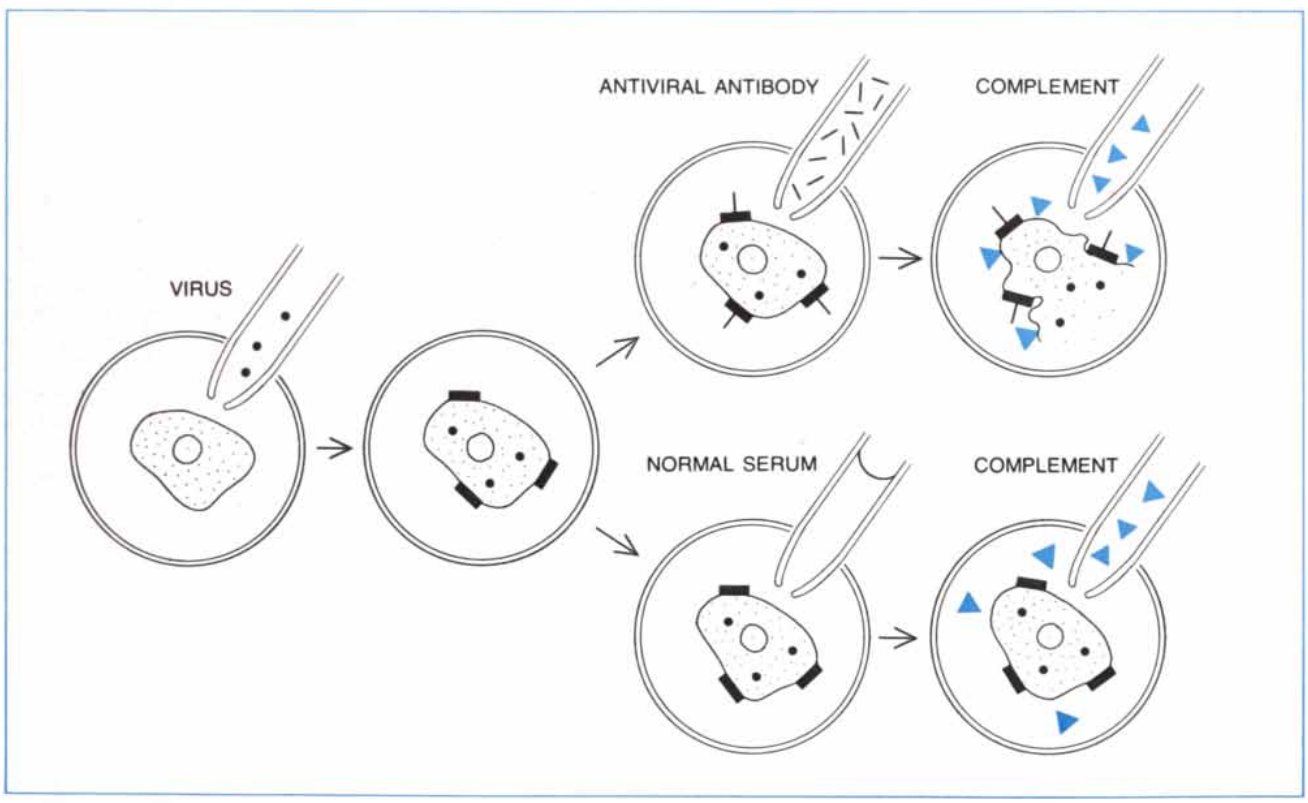
VIRUS-INFECTED CELLS are destroyed by an immune response as shown here. The infecting virus induces the formation of new antigens on the cell surface. Addition of lymphocytes to a culture of infected cells has no effect. Lymphocytes from an animal immunized with the same virus recognize the viral antigen, however, and cell destruction results.

Up to this point we have discussed primarily mechanisms by which the immune response to viral antigens on the surface of infected cells can cause tissue injury. Von Pirquet's studies in the early 1900's on serum sickness suggested a different mechanism, one involving the combination of antigen and antibody. Again, it was not until the 1950's that firm evidence began to come into view. Frederick G. Germuth, Jr., and his coworkers at the Johns Hopkins Hospital and Frank J. Dixon and his colleagues at the University of Pittsburgh School of Medicine found that they could produce serum sickness in rabbits by injecting combinations, prepared in the test tube, of foreign protein and the antibody to it. They also ascertained that some of these injected complexes, circulating in the rabbits' blood, became trapped in the capillaries of the kidneys and led to inflammation, loss of kidney functions and symptoms characteristic of the human disease known as glomerulonephritis.

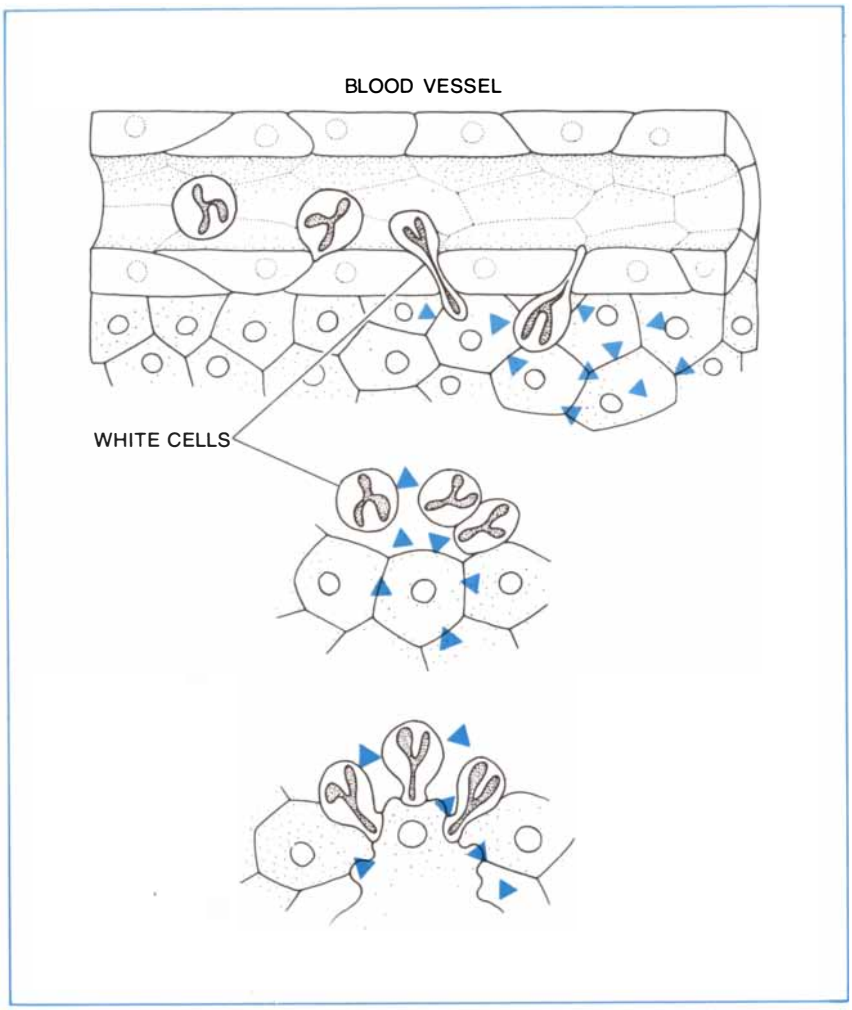
In the light of the evidence that the kidney disease was caused by an antigen-antibody complex, it was called "immune complex" disease. Although



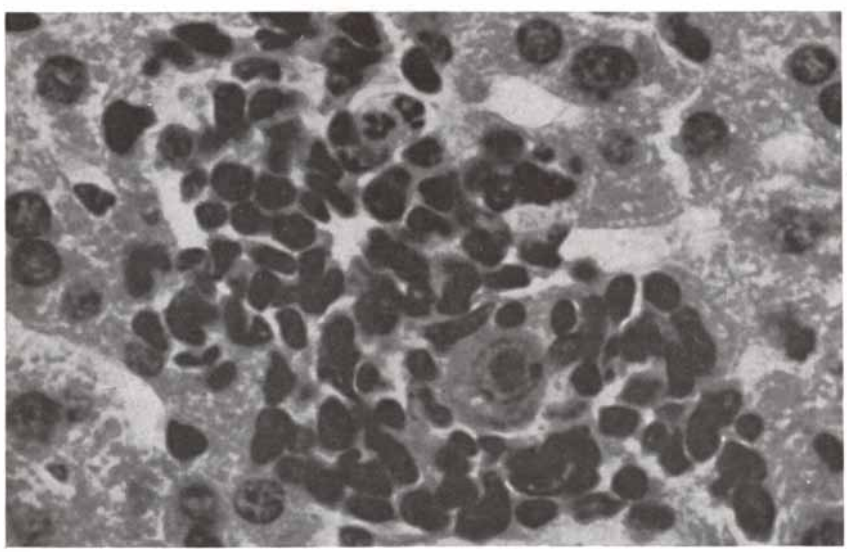
SPECIFICITY of the effect of immune lymphocytes was demonstrated by immunizing rabbits with a virus and then exposing their lymphocytes to the same virus and to another virus. The lymphocytes exposed to the same virus were stimulated to synthesize new DNA and to divide; the others were not. The stimulated lymphocytes produce mediators, some of which destroy normal tissue.



ANTIBODY-MEDIATED IMMUNE RESPONSE is demonstrated in a tissue culture. In the first step antiviral antibody recognizes and binds to virus-induced antigens on the surface of infected cells. Then complement interacts with the newly formed antigen-antibody complex and the cells are destroyed. Neither the antiviral antibody nor the complement alone destroys the infected cells.



SECONDARY EFFECT of the interaction of complement with antigen-antibody complexes is activation of certain components of complement (*color*) that increase the permeability of blood-vessel walls and attract white cells. Inflammation results, and enzymes released from the white cells can injure uninfected tissue, contributing to the symptoms of the infection.

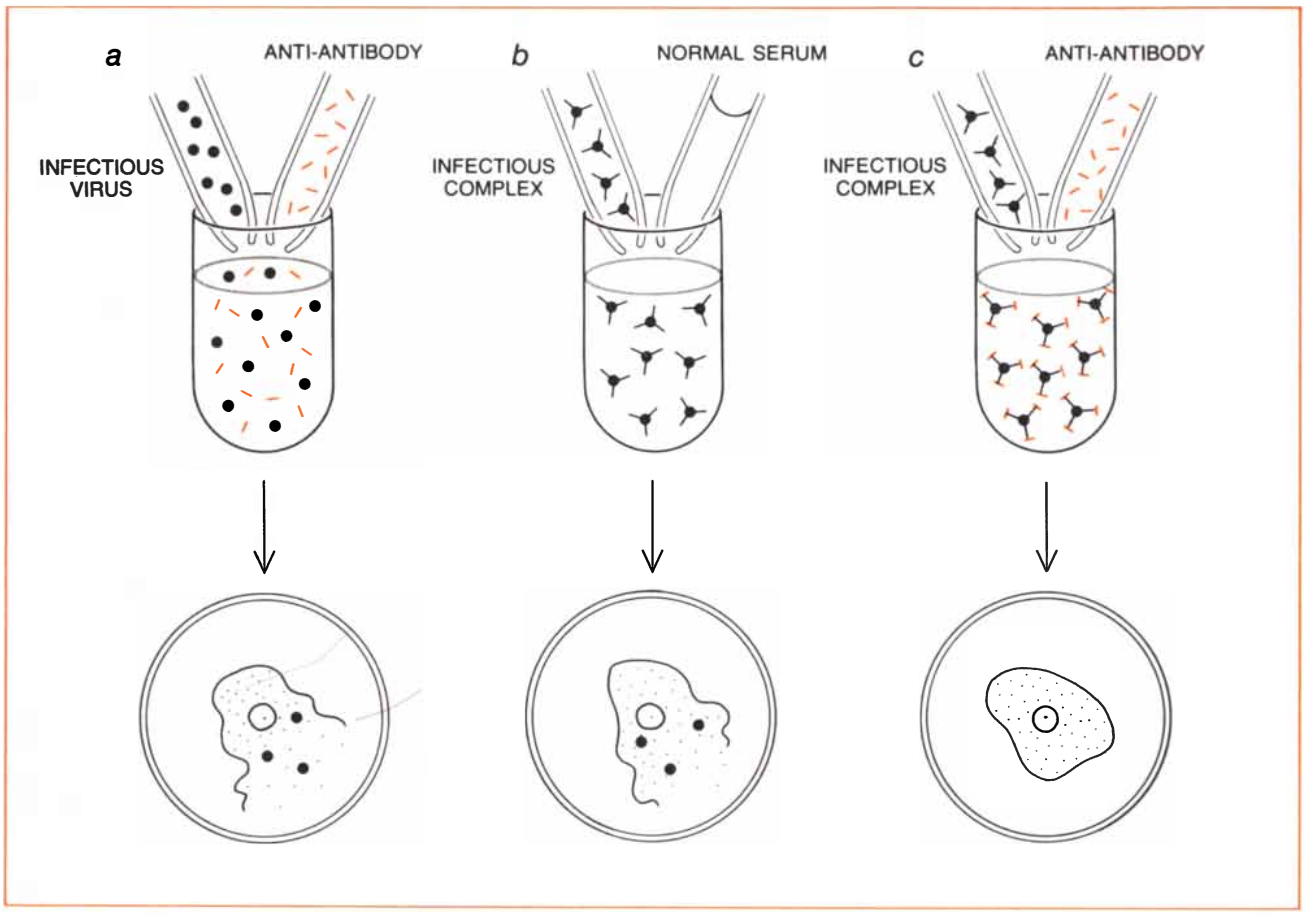


WHITE CELLS (*rounded, dark-staining cells*) are visible in a photograph of mouse-liver tissue made by Donald Henson of the National Institutes of Health. They were attracted to the site of a virus-induced lesion, presumably in the manner diagrammed at the top of the page: the large cell at the center of the clustered cells was infected with cytomegalovirus.

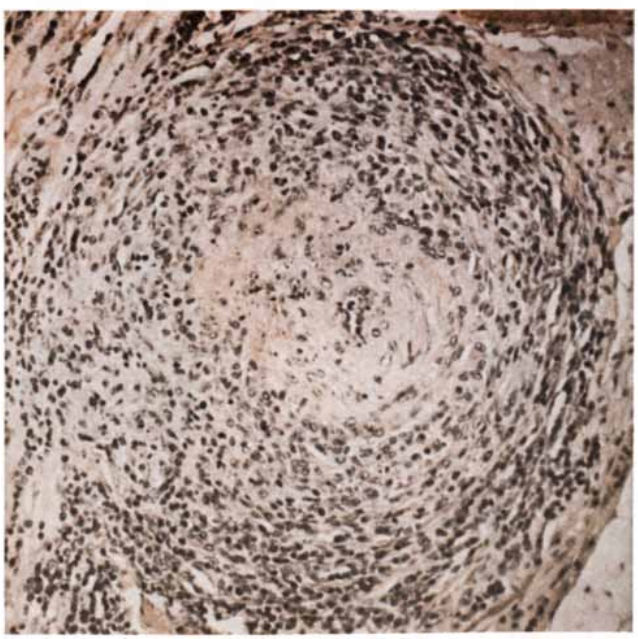
the immune-complex syndrome could readily be produced in animals in the laboratory, the fact that it could also occur in response to viral infections under natural conditions was not demonstrated until a decade later. Virus-antibody complexes were not easy to identify, and the problem became much more manageable when a new technique for doing so was developed.

For a number of years the National Institutes of Health laboratory had been studying an unusual virus called lactic dehydrogenase virus (LDV) because it elevates the level of that enzyme in the blood. Inoculation of mice with the virus produced large amounts of infectious virus in the blood and a chronic infection without any indication of an immune response. It was supposed that the animals were unable to make antibody against the virus, a situation known as immunological tolerance. The laboratory devised a highly sensitive technique for detecting virus-antibody combinations, however, and with this technique [*see top illustration on opposite page*] discovered that the infected mice were indeed making antibody to LDV. The reason it had not been recognized before was that although antibody had combined with the virus, the resulting complex remained infectious and therefore could not be distinguished from the virus itself.

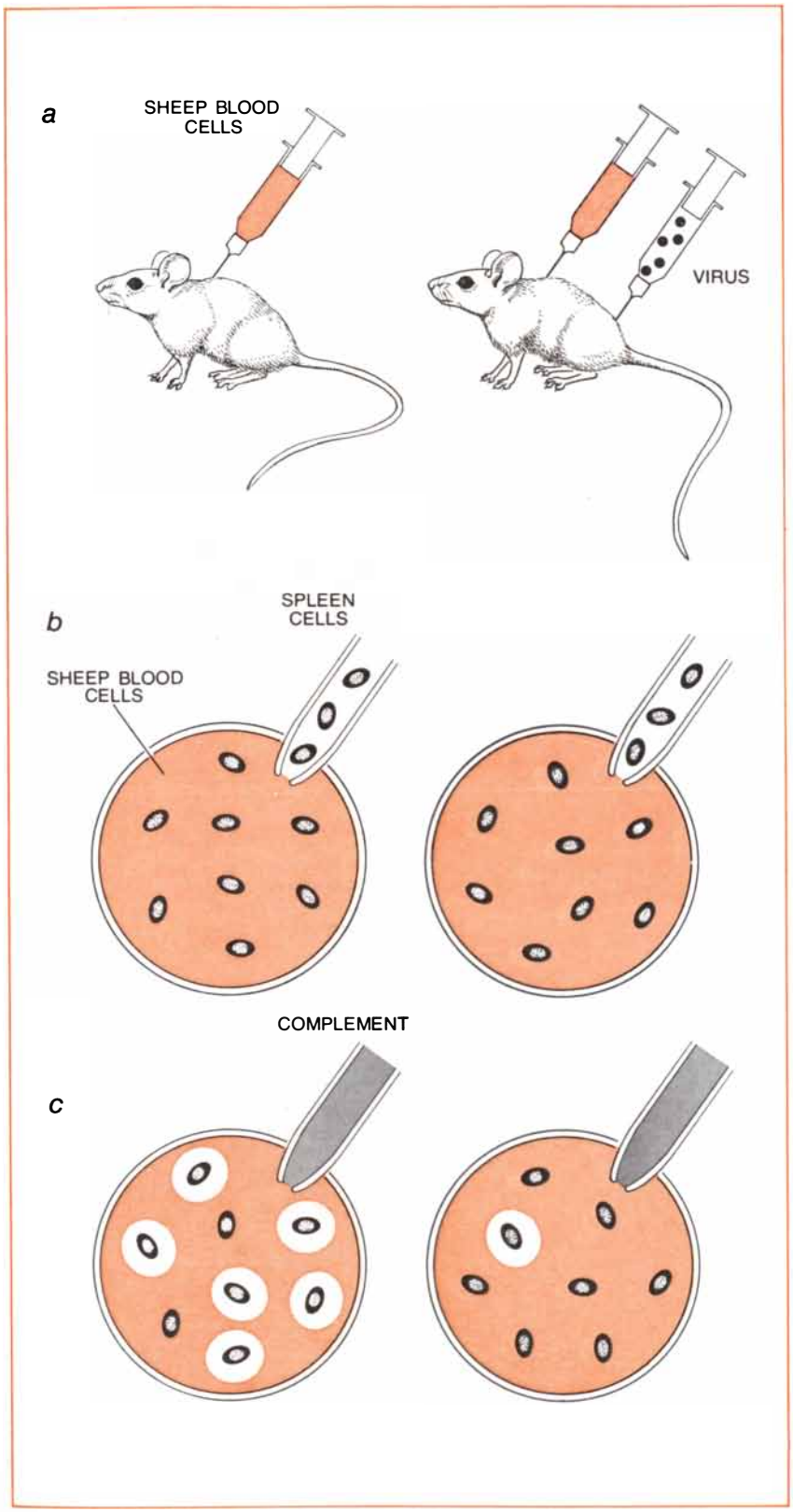
Other investigators proceeded to show that infectious virus-antibody complexes were in fact characteristic of several of the chronic viral infections. What is more, it soon became apparent that some of these chronic infections ended in glomerulonephritis. John E. Hotchin of the New York State Department of Health in Albany had observed several years earlier that infection with LCM virus did not kill newborn mice as it did adults; instead it established a chronic infection that eventually led to glomerulonephritis. How LCM virus produces kidney disease remained unclear until Michael Oldstone and Dixon, who was now working at the Scripps Clinic and Research Foundation in La Jolla, Calif., showed that LCM virus existed in the blood of chronically infected animals as an infectious virus-antibody complex and that the kidneys contained large amounts of LCM antigen, anti-LCM antibody and complement. Moreover, microscopic studies revealed the typical pattern seen when immune complexes are deposited in the kidneys. Similar observations quickly followed with other chronic viral infections, in some of which the inflammatory changes were not confined to kidneys but ap-



INFECTIOUS VIRUS-ANTIBODY COMPLEXES can be detected by a technique utilizing an antibody that recognizes an antiviral antibody as an antigen: an "anti-antibody." When the anti-antibody is incubated with untreated virus, the virus retains its infectivity (*a*). When virus with antiviral antibody on its surface (infectious virus-antibody complex) is treated with normal serum, the complex remains infectious (*b*). When the infectious complex is treated with the anti-antibody, however, the latter attaches itself to the antibody of the complex, neutralizing the complex (*c*). Thus the test distinguishes between virus and infectious complex.



ARTERITIS occurs in some viral immune-complex diseases. Photomicrographs of the coronary artery of mink infected with Aleutian virus, made by Porter, show mild arteritis (*left*) and, at higher magnification, severe arteritis with infiltration of white cells (*small dark-staining cells*) and obstruction of the vessel (*right*). Damaged arteries contained viral antigen, antibody and complement.



INFECTION of mice with leukemia virus depressed their ability to make antibody. Mice, some of which had been infected with virus, were immunized with sheep red blood cells (*a*). Spleen cells from the mice were spread over layers of sheep red blood cells in laboratory dishes (*b*) and complement was added (*c*). Only spleen cells in which antibody to the sheep cells had been induced were able, in the presence of complement, to destroy surrounding sheep cells. Counting the patches of dead sheep cells (*white*) showed spleens of infected mice contained many fewer antibody-producing cells than spleens of healthy mice.

peared in blood vessels and organs in other parts of the body.

The precise steps involved in the production of immune-complex disease by virus-antibody complexes are still only partly understood. Presumably when the complexes are trapped in the kidneys or on blood-vessel walls, they activate the components of complement; mediators are generated, inflammation results and the tissue-injuring enzymes are released from white cells. There are still many unanswered questions, however. Why does LCM virus produce a severe glomerulonephritis and LDV only a mild one? Why do some animals develop glomerulonephritis when they are exposed to a given virus whereas other animals do not? Are there genetic factors governing susceptibility to the virus or the immunological response to it?

Be that as it may, thousands of people develop glomerulonephritis each year, and so the finding that virus-antibody complexes produce the disease in animals has intensified research interest in such complexes and their possible involvement in human disease. Recently evidence has begun to accumulate that in man the hepatitis virus circulates in the blood as an immune complex and that some of the manifestations of this disease, including the associated high incidence of arthritis, may be due to these immune complexes. If it turns out that the immune response to viruses is actually responsible for glomerulonephritis or other immune-complex diseases in man, then controlling the adverse effects of the immune response may be essential for therapy. In animals many manifestations of these chronic infections, including glomerulonephritis, can be prevented or reduced by drugs that suppress the immune response.

The possibility that viruses and the immune response may also be involved in autoimmune diseases such as rheumatoid arthritis and lupus erythematosus is being investigated in a number of laboratories. An autoimmune disease is one in which the body treats its own tissue as an alien antigen and produces antibodies that attack the tissue. What causes the host suddenly to turn against its own tissues? It has long been suspected that viral infections may be a triggering factor. Several hypotheses about how a virus might bring the immunological system into play against the host's own cells have been suggested. The viral infection may unmask or release a potential antigen that normally is hidden within the cells, out of contact with the immune system. Or a viral antigen may

combine with an indigenous protein on the cell's surface and thus form a new "foreign" substance. Another possibility is that a viral infection may activate genes in the cell whose information is ordinarily repressed, thereby causing them to begin producing "new" substances that act as antigens. Wanda Baranska and Wojciech Sawicki at the Wistar Institute found support for this idea in experiments with mouse ova and embryos. They showed that an antigen that was present in the animals' earliest embryonic stage could not be detected in adults, but that when adult cells were transformed into tumor cells by the virus known as SV-40, the embryonic antigen reappeared on the cell surface. Apparently in these transformed mouse cells the genes were again able to redirect formation of the "embryonic antigen."

Still another possibility is that a viral infection may cause the cells of the immune system to behave abnormally and so produce antibodies against some of the host's own tissue. Although proof for this hypothesis is still lacking, there is considerable evidence, particularly from animal studies, that certain viral infections depress the function of the cells of the immune system. Again, it was von Pirquet who first observed that the reactivity to the tuberculin test (an immune response) was depressed in patients infected with measles virus. The effect of viruses on immune function received little attention, however, until 1963, when Robert A. Good and his colleagues at the University of Minnesota College of Medical Sciences showed that mice infected with a leukemia-producing virus were markedly depressed in their ability to make antibody against foreign substances. Other studies, notably by Walter Ceglowski and Herman Friedman at the Albert Einstein Medical Center in Philadelphia, showed that not only was the amount of antibody in the blood reduced but also the number of cells capable of making antibody was curtailed by as much as 99 percent [*see illustration on opposite page*]. Moreover, the immune response was depressed within a few days after infection, long before the animals developed visible signs of leukemia.

It soon became apparent that non-leukemia-producing viruses also could impair immune function. Richard J. Howard and Stephan E. Mergenhagen of the National Institutes of Health laboratory showed this in tests of the reactivity of LDV-infected mice to foreign skin grafts. These animals rejected grafts at a lower rate than uninfected animals. Moreover, there is evidence that the rejection of transplanted tumors also is slowed by viral infection of the immune system. In fact, it seems possible that the virus-induced depression of antibody-mediated and lymphocyte-mediated immunity may be a factor in the initiation and development of tumors and also may account for the chronic nature of certain viral infections.

MECHANISMS	PATHOLOGY
1 VIRUS-INDUCED ANTIGENS ON CELL SURFACE: a INTERACTION WITH IMMUNE LYMPHOCYTES b INTERACTION WITH ANTIVIRAL ANTIBODY AND COMPLEMENT	DESTRUCTION OF INFECTED CELLS
2 ACTIVATION OF MEDIATORS FROM IMMUNE LYMPHOCYTES OR COMPONENTS OF COMPLEMENT	INFLAMMATION, ALLERGIC REACTIONS, DESTRUCTION OF CELLS
3 FORMATION OF CIRCULATING VIRUS-ANTIBODY COMPLEXES	IMMUNE-COMPLEX DISEASE
4 IMMUNE RESPONSE TO HOST-CELL ANTIGENS ALTERED OR DEREPRESSED BY VIRUS	AUTOIMMUNE REACTIONS
5 INFECTION OF CELLS OF IMMUNE SYSTEM	INHIBITION OR ENHANCEMENT OF IMMUNE FUNCTION

MECHANISMS of the various immune-response disorders that are caused by viral infection and are discussed in the article are summarized, together with the associated pathology.

A curious twist in the already complicated story of immunity and viruses is the finding that lymphocytes, which are the body's major defense against tumors, can actually act as agents for the induction of tumors. It has been known for some time that an unusually high incidence of lymphomas (tumors of the lymphoid glands) occurs in animals or patients undergoing chronic stimulation of these glands as a result of autoimmune disorders or rejection of a graft such as a kidney transplant. It is also known that certain viruses, including the leukemia virus of mice and the mononucleosis virus of man, exist in lymphoid cells in a "latent" state. On the basis of these observations, groups headed by Martin S. Hirsch and Paul H. Black of the Harvard Medical School and Robert S. Schwartz of the Tufts University School of Medicine conducted experiments to see if stimulation of the lymphoid elements of the immune system might arouse the latent virus. In order to stimulate the immune system they exposed mice to foreign grafts, and they found that this activated the leukemia virus that had previously been latent in the lymphoid cells. The findings suggest that such activation of leukemia virus from immunologically stimulated lymphocytes may be responsible for the high incidence of lymphomas associated with autoimmune disorders and graft rejection.

All in all, it is now obvious that the interrelations encompassing viruses, immunity and disease are indeed complex. It appears that the immune response to viral infection can have both beneficial and deleterious effects on the host. On the one hand, it may be the chief or only weapon against the infection; on the other hand, it may be responsible for some of the noxious symptoms and even the fatal effects of the disease. Probably the immune response makes some contribution, large or small, to the pathologic picture in most viral infections. Although we must recognize that the immune system is not an unmixed blessing, it is encouraging to know that by learning more about it we may eventually find new approaches to the treatment of viral diseases.

THE OMNIVOROUS CHIMPANZEE

Observation of chimpanzees in the wild indicates that they not only eat plant foods but also hunt, kill and eat other mammals. Moreover, they display a well-developed pattern of sharing meat

by Geza Teleki

It is widely believed that apes and monkeys are vegetarians, and that man is alone among the primates in preying on other animals. The assumption has influenced a number of hypotheses about human evolution that were framed in the days when scarcely any of man's primate relatives had been studied in the wild. For example, it has been suggested that the pursuit of game and the consequent social sharing both of the hunt and of the kill were key factors in the divergence of the earliest hominids from the rest of the primate line.

Today, after some 40 years of field observations of ape and monkey behavior, it is quite clear that man is not the only primate that hunts and eats meat. Many other primates are omnivorous. One in particular—the chimpanzee —not only cooperates in the work of the chase but also engages in a remarkably socialized distribution of the prey after the kill. The chimpanzees whose predatory behavior has been most closely observed are semi-isolated residents of the Gombe National Park in western Tanzania. The area, formerly known as the Gombe Stream Chimpanzee Reserve, is where Jane van Lawick-Goodall began her notable long-term field study of chimpanzees in 1960. I myself spent 12 months watching the predatory behavior of these apes in 1968–1969.

Gombe Park covers some 30 square miles and has an estimated population of 150 chimpanzees. All belong to the eastern chimpanzee subspecies *Pan troglodytes schweinfurthii*. Goodall and her colleagues quickly came to know on sight some 50 individual apes that lived in a 10-square-mile zone centered on Kakombe Valley. This zone became the main study area, and its chimpanzees were made the subjects of daily records that now cover more than a decade of observation.

It was soon apparent to the observers that the chimpanzee population was organized in a social hierarchy headed by a single adult male that was senior to all the other males, adult and subadult. Most adult males and even some of the subadults, in turn, usually outranked the female chimpanzees, regardless of the females' age. The females' behavior in many situations showed, however, that there was an independent hierarchy among them as well. Between 1960 and 1970 the position of senior male, or "alpha," was occupied successively by two different adults. During my 12 months of observation the position was held by a chimpanzee Goodall had named Mike.

Sherwood L. Washburn and Irven DeVore of the University of California at Berkeley were among the first to observe primates eating meat. Studying olive baboons and yellow baboons in Kenya in 1959, they saw these ground-dwelling primates kill and eat newborn antelopes [see "The Social Life of Baboons," by S. L. Washburn and Irven DeVore; SCIENTIFIC AMERICAN, June, 1961]. Not long afterward Goodall found that the Gombe chimpanzees were omnivorous. Their diet included insects, lizards, birds' eggs and fledgling birds, young bushbucks and bushpigs, blue monkeys, redtail monkeys, colobus monkeys and infant and juvenile baboons. Meanwhile observations of primates in the wild around the world revealed that an omnivorous diet was far from unusual. By the end of the 1960's the list of primates that feed on animals larger than insects had grown to include two chimpanzee populations outside Gombe Park, two species of baboons in addition to the olive baboon and the yellow baboon, vervet monkeys in Africa, macaques in Japan and even woolly monkeys and capuchin monkeys in the New World. Only baboons and chimpanzees, however, are known to actively seek out and pursue their prey.

In the decade between 1960 and 1970 Goodall and others, myself included, were able to note that the 50-odd chimpanzees in the vicinity of the Kakombe Valley field station killed and ate no fewer than 95 individual mammals and attempted to capture another 37 that escaped. In 46 of the predatory incidents the kills were actually witnessed. Another 38 kills were known through examination of the chimpanzees' feces, and recognizable fragments of 11 additional prey animals were carried by the chimpanzees to within sight of an observer.

During my year at Gombe Park I witnessed 30 episodes of predation, 12 of them successful. Kills during the decade, including those I observed, averaged a little more than nine mammals per year. The fact that I witnessed 30 episodes and 12 kills in a 12-month period is less likely to mean that the period was an above-average one for predation than that in the other years some kills and many episodes probably went unnoticed. In terms of predation on baboons, however, the period of my observations was

MEAT-EATING CHIMPANZEES in Tanzania, seen in the photograph on the opposite page, are engaged in the socially structured sharing behavior that follows a successful hunt. The prey animal is a young baboon. Its captor is reacting negatively to the "requesting" behavior of the second chimpanzee, higher in the tree, by starting to move away. Most of the requesters that join a "sharing cluster" after a kill eventually receive a little meat.

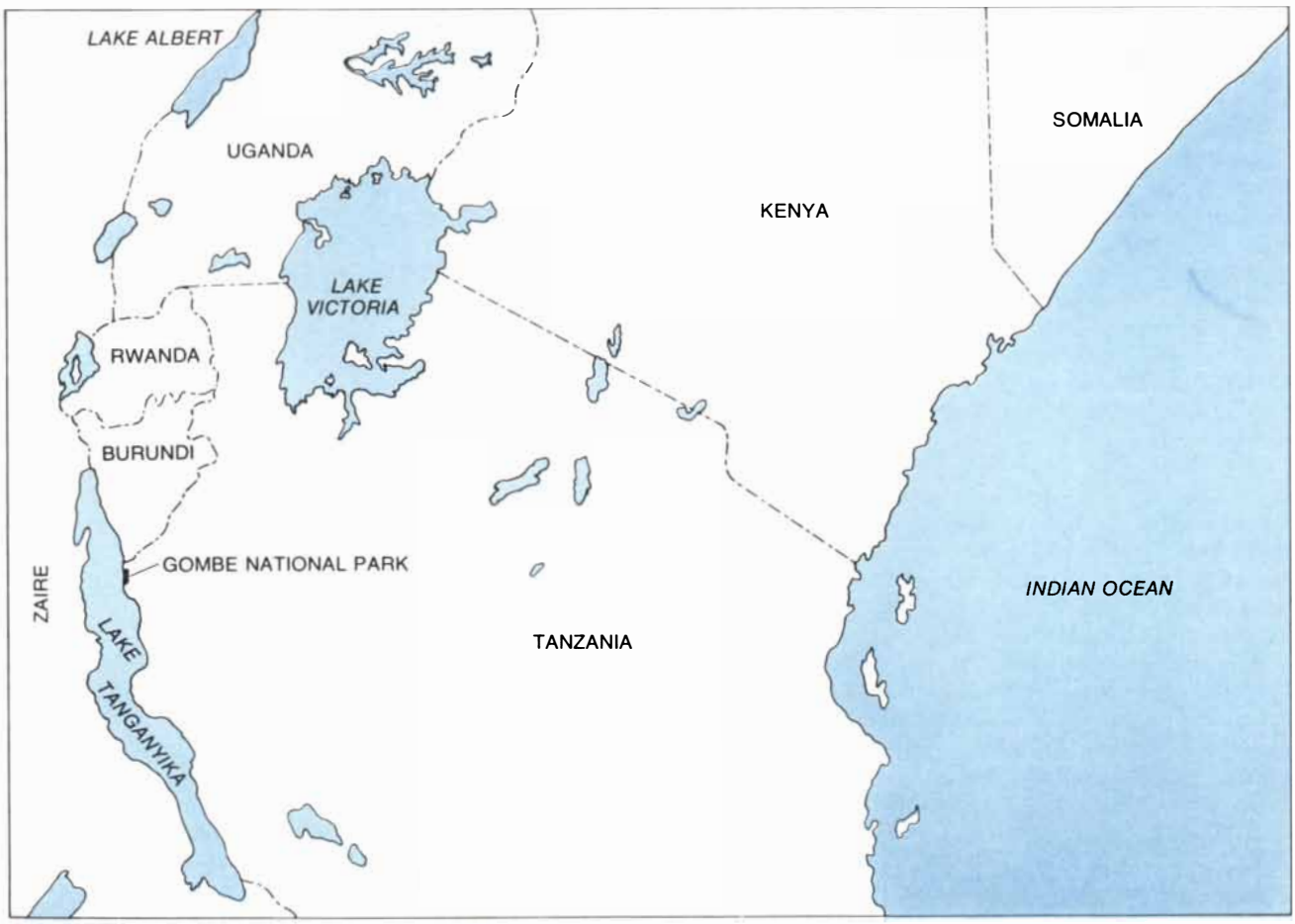
clearly not average. Goodall's records for the decade include identifications of 56 of the 95 prey animals. Primates were in the majority: 14 colobus monkeys, 21 baboons, one blue monkey and one redtail monkey. Of the 19 other mammals identified 10 were young bushpigs and the rest were young bushbucks. In contrast, of the kills I observed in 1968–1969 the prey animal on 10 out of 12 occasions was a baboon. Moreover, all 18 unsuccessful episodes I witnessed involved baboons as prey animals.

The prey species I have mentioned are very nearly the only mammals available to the Gombe chimpanzees, so that what might appear to be preference in reality demonstrates the diversity of the apes' predatory efforts. Indeed, the only limit to chimpanzee predation that I observed was the size of the prey animal. There is no evidence that chimpanzees capture or even pursue an animal that weighs more than about 20 pounds. For example, most captured baboons were infants or juveniles with an estimated weight of 10 pounds or less. Similarly, the bushbucks and bushpigs that the chimpanzees kill are either newborn or very young. Few of the adult mammals killed by the Gombe chimpanzees (colobus monkeys, blue monkeys and redtail monkeys) weighed as much as 20 pounds.

The prevalence of baboons as prey in Gombe Park must be due to the fact that baboons are the primates the chimpanzees most frequently encounter. Both species live mainly on the ground, both travel the same trails through grassland and forest and both eat many of the same foods and visit the same foraging areas. Each species may displace the other from special feeding sites, such as trees bearing fruit, and may even interact competitively over access to favored foods. At the same time baboons and chimpanzees engage in amicable interactions such as grooming and play. During my 12 months at Gombe I saw play groups made up of young chimpanzees and baboons at the field station almost daily. On two occasions I saw adult male chimpanzees suddenly pursue and capture a young baboon that until shortly before had participated in a mixed play group that was being tolerantly observed by female chimpanzees and baboons sitting a short distance away.

The coexistence of both social and predatory interactions between chimpanzees and baboons at Gombe may be anomalous. Elsewhere in Africa adult male baboons are reported to defend their troop against threatening carnivores, even to the extent of seriously injuring such formidable adversaries as large leopards. Indeed, I observed the baboons of one Gombe troop kill a 35-pound African wildcat that appeared to have threatened them.

The Gombe baboons' response to predatory chimpanzees appears to be far more ambivalent. I have seen an adult male baboon sit unchallenged shoulder to shoulder with three adult male chimpanzees that were dividing the carcass of a bushbuck, while other baboons and chimpanzees searched together through the underbrush for fallen pieces of meat. At the opposite extreme I have seen a single chimpanzee stand upright and run into the midst of



GOMBE NATIONAL PARK, a former game reserve in Tanzania where the hunting behavior of chimpanzees was studied, is on the shore of Lake Tanganyika. Jane van Lawick-Goodall's work in the area since 1960 has made the apes accustomed to human observers.

a baboon troop, where it selected, pursued and captured a young baboon while exhibiting complete indifference to the numerous adult male baboons that "mobbed" it, threatening it, slapping it and even leaping on its back. Nor was this an unusual event. The Gombe records show that predatory chimpanzees are rarely more than scratched by the baboons that undertake the defense of a troop member. A severe injury to a chimpanzee predator has never been reported.

Perhaps it is both the frequency and the variety of social interactions between the Gombe chimpanzees and baboons that are responsible for the ambivalence of the baboons' response. The use of the same trails and foraging areas, the mutual recognition of many communicative signals and the grooming and play among the young of both species may have resulted in the development of an interspecific "communal" atmosphere, so that the baboons respond to the chimpanzees much as they might respond to other baboons. That something of this kind is possible, at least in reverse, is apparent from the following observation. An adult baboon was fatally injured by another member of the troop; the fight was witnessed by several chimpanzees. Instead of considering the dead baboon fair prey, some of the chimpanzees inspected the carcass, touched it and finally groomed it. The same kind of curiosity has been observed when chimpanzees are confronted with a dead chimpanzee.

Insofar as it has been observed among the Gombe chimpanzees, hunting behavior is an exclusively adult activity and is almost exclusively a male one. Adult females have occasionally been observed pursuing and even capturing prey, but in every instance no adult males were in the vicinity at the time. The males may hunt alone, or two or more males may coordinate their actions. I once witnessed five males working together to surround three baboons that had taken shelter in trees; the movements of the chimpanzees were plainly cooperative.

A predatory episode that results in a kill consists of a sequence of three events, each marked by its own characteristic activities. The first of the three I shall call "pursuit," even though the distance covered is sometimes only a few feet and the time elapsed is seconds. The second event, "capture," is a brief period that ends with the initial dismemberment of the prey. The third and longest event, "consumption," involves highly structured activities. On one occasion I observed a consumption period that lasted nine hours and involved 15 chimpanzees.

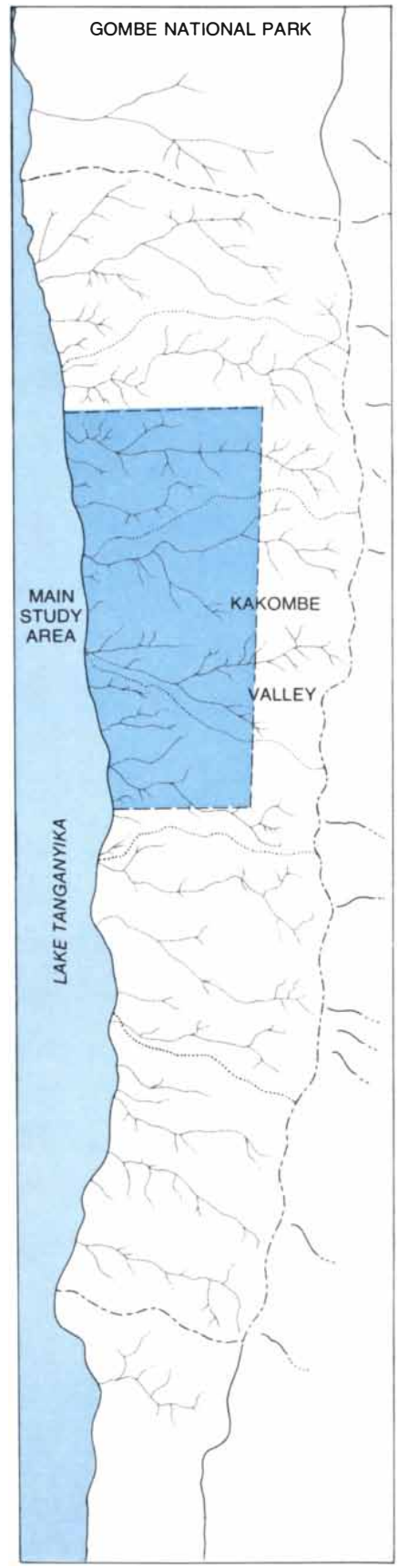
The mean duration of the 12 successful predatory episodes I witnessed was a little less than four hours; the shortest episode lasted an hour and 45 minutes. An unsuccessful episode is of course much briefer. The mean duration of the 18 I witnessed was 12 minutes. The Gombe chimpanzees spend more than 90 percent of the total time they devote to predation in sharing and eating the prey.

I did not always have the prey animal under observation at the start of a predatory episode, so that it was not always clear to me what had initiated the pursuit. When both the prey and the predator were in view, it was apparent that the chimpanzees perceived and often selected their prey before starting the pursuit. The indicative changes in the hunter's posture and expression were so subtle, however, that it is difficult to specify them.

Many episodes began when the male chimpanzees were relaxed, for example dozing or grooming or resting after they had eaten large quantities of fruit. When the chimpanzees were either interacting intensively among themselves or interacting aggressively with some other species of primate, predatory episodes were uncommon. When a baboon was the prey animal, one apparent stimulus to predation was the vocalization of infant or juvenile baboons. Associated with both aggression and distress, these sounds frequently occurred during the young baboons' play sessions or when a young baboon sought to return to its mother. In more than half of the 28 episodes involving baboons that I witnessed, the crying of young baboons preceded any evidence of predatory interest on the part of the chimpanzees.

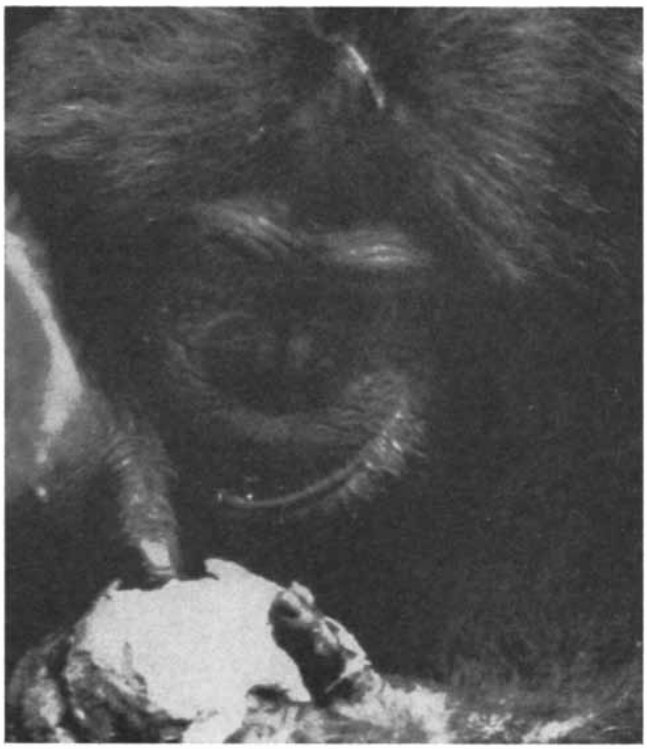
As I have indicated, one form of pursuit is simple seizure, an explosive act that lacks obvious preliminaries. The chimpanzee takes advantage of a fortuitous situation to make a sudden lunge and grab the prey animal. In this respect seizure is a kind of instantaneous capture that resembles "opportunistic" feeding on immobile prey (for example a fledgling bird or a "frozen" newborn antelope), the kind of meat-eating frequently noted among lower primates. I was able to witness this form of pursuit in 22 of the 30 predatory episodes I observed; seven involved simple seizure, and three of the lunges were successful.

Two other forms of pursuit are practiced by the Gombe chimpanzees. One



MAIN STUDY AREA at Gombe Park is a 10-square-mile zone (*color*) in and around Kakombe Valley. Of some 150 chimpanzees in the park, about 50 forage in this zone.

SUCCESSFUL HUNTER begins to dismantle the portion of the prey animal it has retained, the forequarters of a baboon. Almost all the Gombe chimpanzee hunters are adults and most of them are males. On the average one in every three hunts is successful.



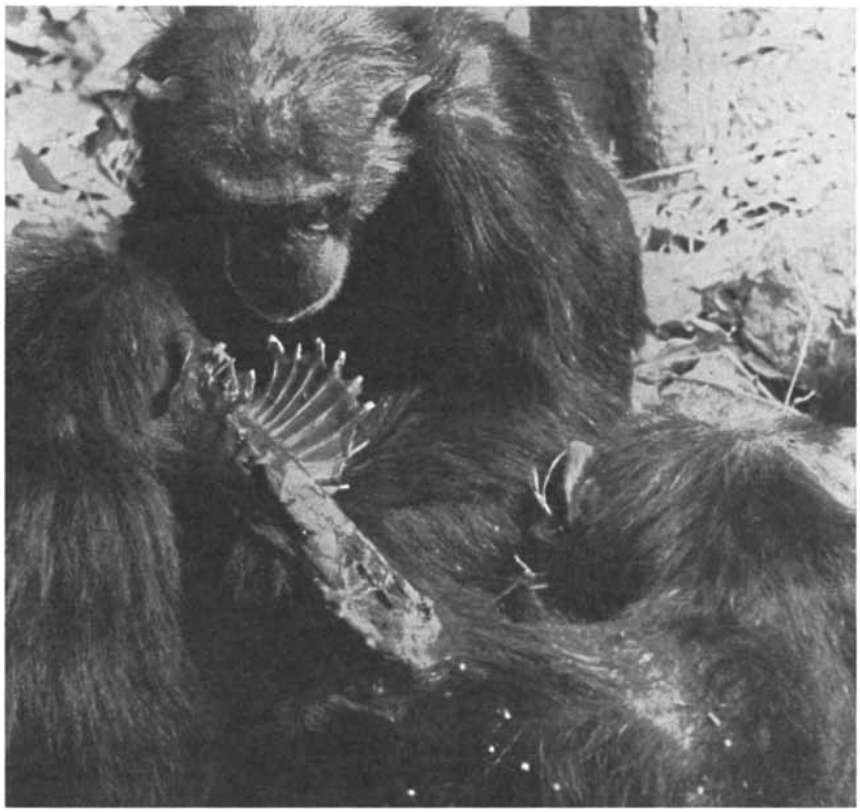
BREAKING INTO A SKULL, a Gombe chimpanzee uses its teeth to crush the frontal bone of a baboon (*left*) and gain access to the brain. With a finger (*right*) it then scoops out this prized portion of the prey. Other portions are often shared but the brain is not.

is chasing, which may involve a dash of 100 yards or more. The other is stalking, a cautious and painstaking process that can last more than an hour. Both have the appearance of being more premeditated and controlled than simple seizure. Moreover, on occasion both clearly involve a strategy and maneuvers aimed at isolating or cornering the selected prey. Eleven of the episodes that I observed involved chasing and four involved stalking. Six of the chases were successful but all the stalking efforts failed.

At times, particularly in the early morning and in the evening, the Gombe chimpanzees are quite vociferous. That is not the case when they are in pursuit of prey. Regardless of the time of day or the number of chimpanzees involved in the chase, all remain silent until the prey is captured or the attempt is broken off. This means, of course, that the hunters do not coordinate their efforts by means of vocal signals. Neither did I observe any obvious signaling gestures, although cooperation in movement and positioning was evident. It is noteworthy that the leading position during the chase frequently shifted from one hunter to another regardless of the chimpanzees' relative social rank. Evidently individual chimpanzees do not compete with one another for the most advantageous position during the pursuit of prey.

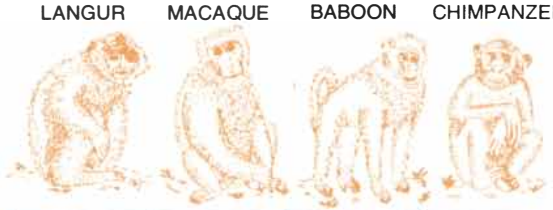
After the prey animal has been maneuvered within reach begins the central event in the predatory sequence: capture. It usually lasts less than five minutes and exhibits three consecutive stages: acquisition, killing and initial division. The first stage is very much like simple seizure; it consists of a final lunge and grab when the distance between the predator and the prey has been reduced to a yard or less. If the chase has been a cooperative venture, more than one chimpanzee may catch the prey. The instant of acquisition is usually signaled by a sudden outburst of vocalization; the cries not only end the silence of the hunt but their volume and pitch serve to draw other chimpanzees from distances of a mile or more.

The killing stage is normally brief. If the prey is in the grasp of a single chimpanzee, the chimpanzee may bite the back of the prey's neck or twist its neck in both hands. Alternatively the chimpanzee may stand upright, grasping the prey by its legs, and strike its head and body against the ground or a tree trunk. If the prey is caught by more than one



FRAGMENTATION OF PREY ANIMAL proceeds as members of a sharing cluster take pieces from a portion in the hands of the possessor, part of a colobus monkey's rib cage.

SUCCESSFUL REQUESTER (*left*) grasps between thumb and forefinger a sliver of meat proferred by its possessor (*lower right*). Another requester (*upper right*) watches closely.



	LANGUR	MACAQUE	BABOON	CHIMPANZEE
LEAVES, FRUITS, BERRIES, BLOSSOMS, BUDS, SHOOTS, BARK, PITH, GRASSES, RHIZOMES, BAMBOO	■	■	■	■
NUTS, SEEDS, PODS, GRAINS		■	■	■
ROOTS, BULBS		■	■	
MUSHROOMS			■	■
HONEY				■
INSECTS	■	■	■	■
SCORPIONS			■	
MOLLUSKS		■	■	
FISHES, CRABS		■	■	
REPTILES			■	■
BIRDS, EGGS			■	■
PRIMATES			■	■
ANTELOPES			■	■
RODENTS			■	
BUSH PIGS				■

OMNIVOROUS DIETS of chimpanzees and of three Old World monkeys include several of the same plant foods, but only one class of animals is eaten by all: insects. Of the four species only the baboons and chimpanzees hunt and eat mammals, including other primates.

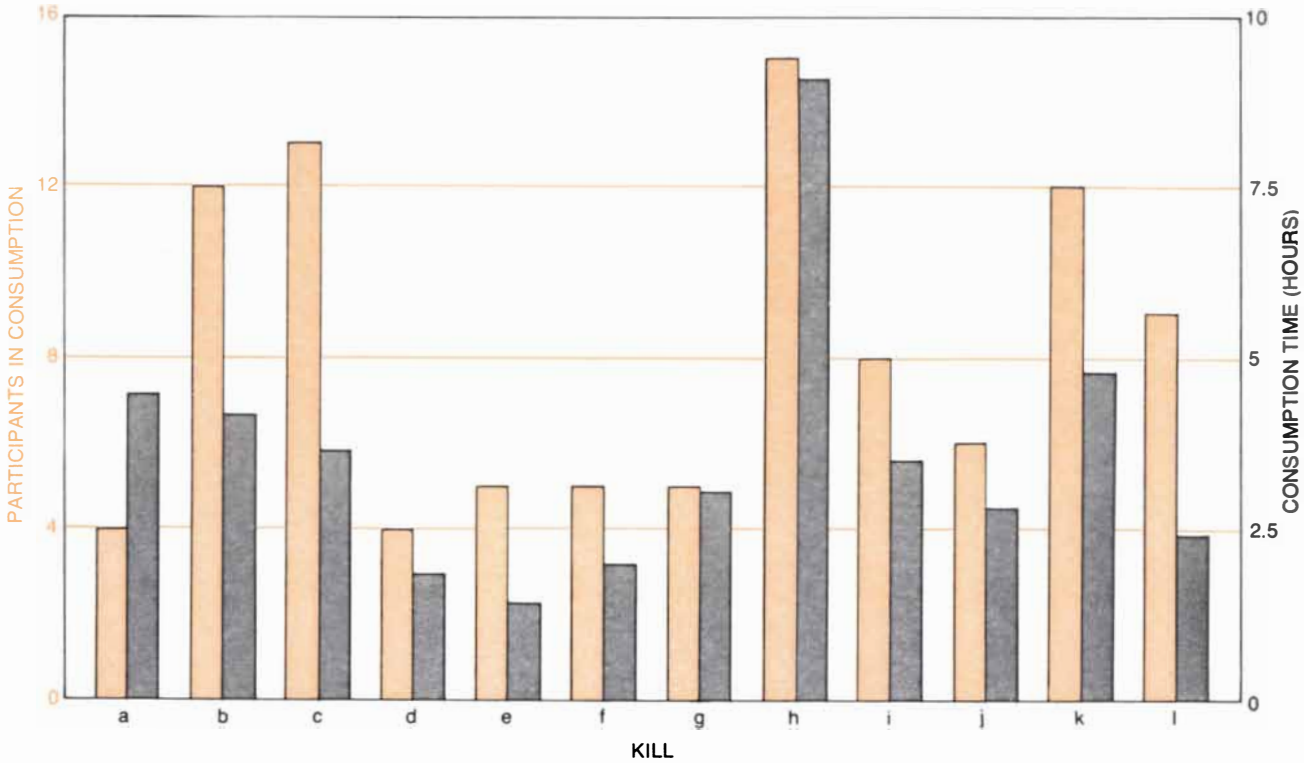
chimpanzee, it may be torn apart as each captor tugs on a different limb; in this way killing and dividing are accomplished simultaneously.

The final stage of capture, initial division of the carcass, is an activity that is unrelated to the "meat-sharing" that comes later. For at least a brief period after the killing the carcass appears to be "common property." If chimpanzees other than the captor or captors have arrived within reach of the prey, they are free to try to grab a part of the carcass without risk of retaliation from the similarly occupied captors. For example, I have seen six chimpanzees, four of them postkill arrivals at the scene, divide among them the arms, legs and trunk of a young baboon. Even under these conditions aggressive interactions are rare during the time of initial division. The few incidents I witnessed tended to be mild, and even male chimpanzees of high social status showed nearly complete tolerance for others.

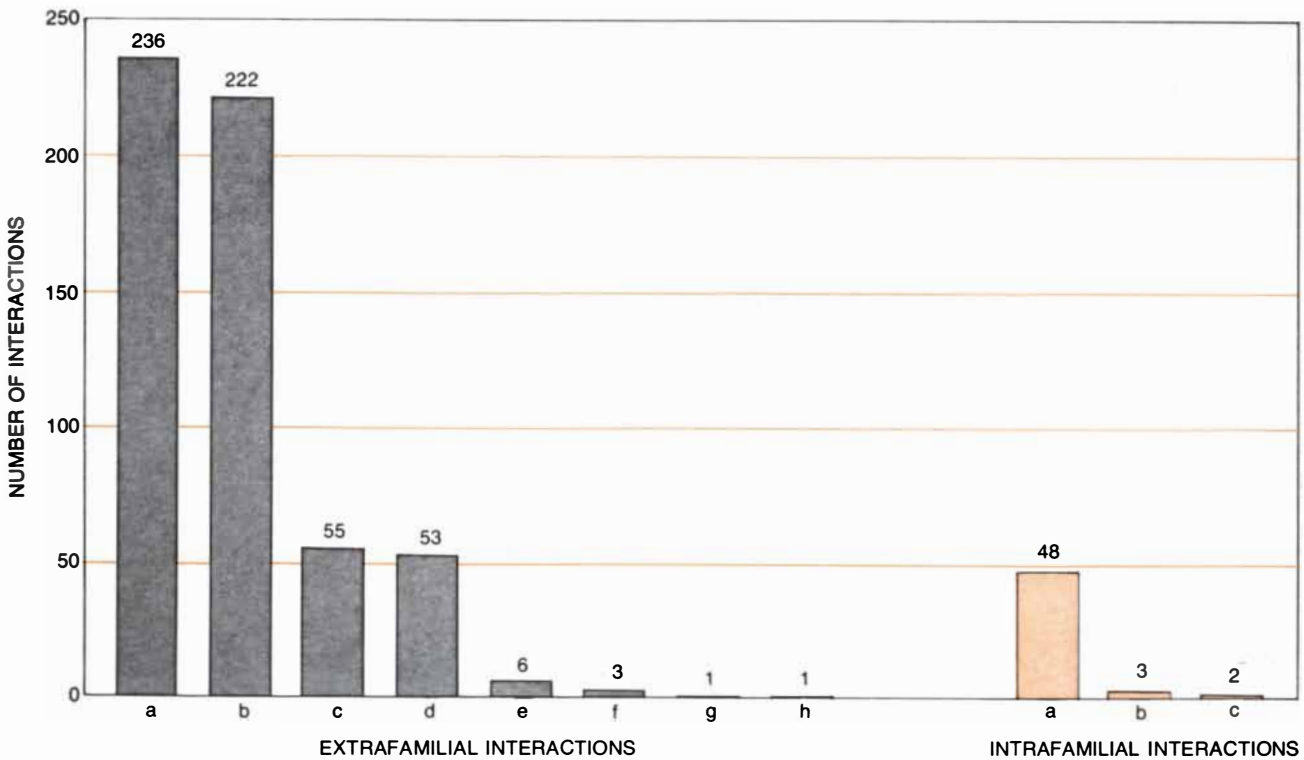
If the captor is able to hoard the prey for several minutes by moving away from the kill site, the common-property character of the prey animal lapses. Chimpanzees that have been attracted to the site are no longer likely to try to tear off a part of the carcass. In the event of a capture by more than one chimpanzee, the same "hoarders' rights" apply equally to the major portions of the carcass. The chimpanzees that share in the initial division then move off, usually no more than a dozen yards apart, each hoarding its piece of the carcass; those that did not participate in the initial division begin to congregate in "sharing clusters" around each possessor of a major fragment. Formation of the sharing clusters initiates the third, and socially the most significant, event in the predation sequence: the consumption period.

The Gombe chimpanzees do not move about their home range in fixed social units. The composition of their small bands is subject to constant change. As a result attendance during consumption periods varies a good deal from one predatory episode to another. It was rare, however, for fewer than five males and females of various ages to be within range of a kill site. At the other extreme, the number of chimpanzees gathered in different sharing clusters around a single kill once reached 15. Even when the number of chimpanzees in the clusters was this large, most of the participants usually obtained a piece of the prey.

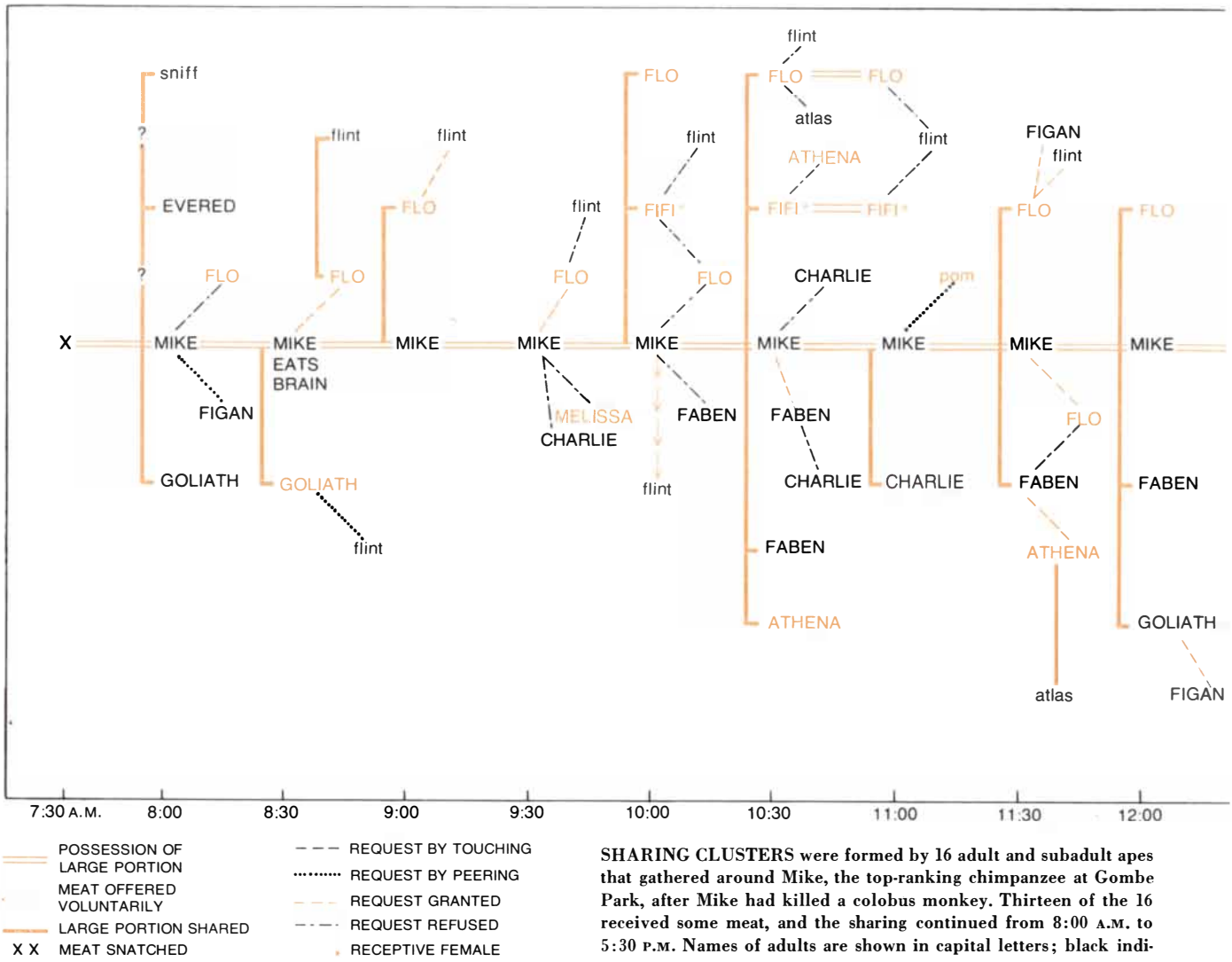
Considering the length of time devot-



CONSUMPTION OF PREY occupies most of the time devoted to predation. Gray bars show the duration of 12 predatory episodes; colored bars show the number of chimpanzees that joined sharing clusters after the kill. The shortest consumption period was one hour 40 minutes (*e*); five apes participated. The longest was more than nine hours (*h*); the prey animal was a colobus monkey and 15 apes participated. The prey animal in the 12th episode was a bushbuck. In the other 10 episodes the prey animals were baboons.



MEAT-SHARING INTERACTIONS in the sharing clusters are most frequent among the adult chimpanzees present. Female-to-male approaches, taking or requesting meat (*a*), are slightly more frequent than male-to-male approaches (*b*). Subadult approaches both to adult males (*c*) and to adult females (*d*), although far fewer, outnumber other nonfamily interactions: female-to-female (*e*), male-to-female (*f*), female-to-subadult (*g*) and subadult-to-subadult (*h*). No male-to-subadult approach occurred. Offspring-to-mother approaches (*a*) were the principal family actions (*color*). The mother-to-offspring (*b*) and intersibling (*c*) approaches were few.



SHARING CLUSTERS were formed by 16 adult and subadult apes that gathered around Mike, the top-ranking chimpanzee at Gombe Park, after Mike had killed a colobus monkey. Thirteen of the 16 received some meat, and the sharing continued from 8:00 A.M. to 5:30 P.M. Names of adults are shown in capital letters; black indi-

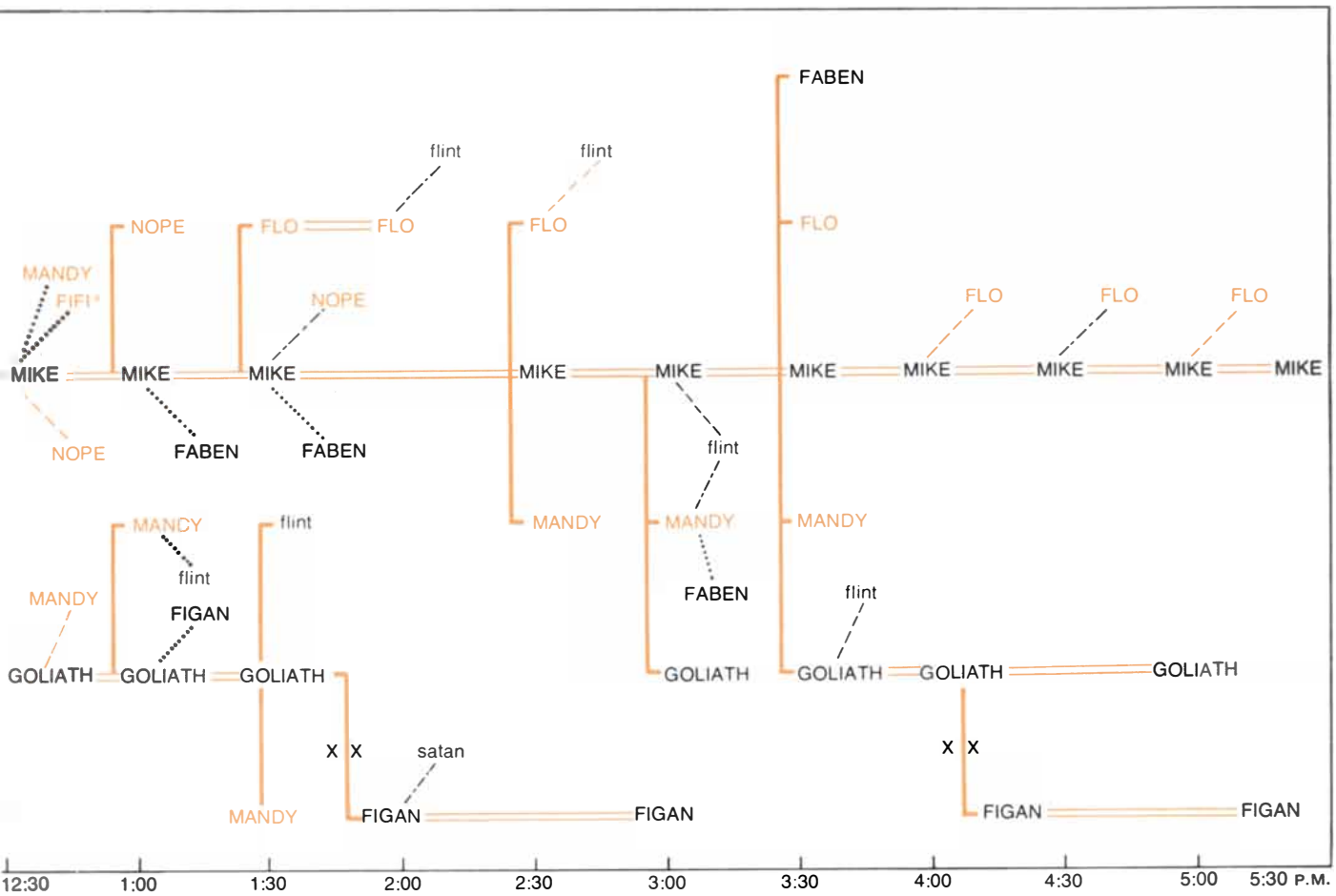
ed to consumption, the small size of the prey animals and the number of chimpanzees that congregate in sharing clusters, the conclusion is almost inescapable that social considerations and not merely nutritional ones underlie the Gombe apes' predatory behavior. For one thing, many predatory episodes are initiated soon after the chimpanzees have consumed large quantities of vegetable foods. For another, no chimpanzee at Gombe has ever been observed to capture and privately consume a mammal, however small, if other adult chimpanzees were present to form a sharing cluster. In the decade covered by the Gombe records exactly two unshared kills were observed. The events were simultaneous: two adult female chimpanzees happened to encounter two small bushpigs and each ate one of them. No other adult chimpanzees were present at the dual kill.

Before a sharing cluster disperses, all the prey animal's carcass will have been consumed: skin and hair, bones, bone marrow, eyeballs and even teeth. The brain is evidently the preferred portion. Although I observed frequent sharing of other parts of the carcass, not once in 12 months did I see one ape yield the brain or any part of it to another. A chimpanzee sometimes removes the brain tissue by pushing a finger into the natural opening where the skull joins the backbone. More often it opens a hole in the prey's forehead with its fingers and teeth and then scoops and sucks the cranial cavity clean.

When the Gombe chimpanzees are eating soft foods, they often put leaves in their mouth to form a kind of chewing wad. The procedure prolongs mastication and may also extend the savor of the food. The apes often use a leaf wad when they are eating meat, and they invariably do so when they are eating brains. After several minutes of chewing the eater usually discards the leaf wad, often giving it to one of the other chimpanzees in the cluster, so that brain tissue actually is shared to some extent, if only indirectly. Three or four chimpanzees in succession may chew on a wad before it is finally swallowed or shredded.

A wad of this kind was the only "implement" I saw a chimpanzee use during any episode of predation. A male that was feeding on a prey animal's brain pushed a leaf wad into the nearly empty skull cavity to soak up the remaining tissue and fluids. The tactic closely resembles other Gombe chimpanzees' repeated use of leaf "sponges" to soak up water from a natural bowl in the crotch of a tree, an action witnessed by Goodall some years ago.

The prey animal's brain may be eaten first or last. Otherwise when a single chimpanzee captor deals with its prey, the dismantling sequence usually begins with the removal and consumption of the viscera. Next the rib cage is cleaned and sectioned, the chimpanzee using its teeth, hands and feet to tear the skin and break the bones. The prey's limbs are consumed last. The process is thorough;



cates a male and color a female. Receipt or possession of meat is indicated by colored lines; a double line shows that the individual holds a sizable piece of the carcass. Unsuccessful requests for meat are shown in black. By keeping the largest portion Mike remained the center of attention all day. Twice Mike gave a big portion of the prey to Goliath, a mature male somewhat past its prime. Both times a younger adult male, Figan, subsequently rushed at Goliath and snatched away some or all of the portion. Mike's most favored sharing partner was Flo. This female successfully requested or was given meat 12 times; twice the portion received was a sizable one.

a careful inspection of a kill site following the chimpanzees' departure yields nothing but a few tiny fragments of the prey.

I have seen chimpanzees in a sharing cluster make use of three methods of getting meat. One that involves a minimum of interaction with the possessor of the meat is to retrieve dropped or discarded bits of the prey that fall to the ground. Such retrieval is usually the activity of subadult chimpanzees and adult females that evidently prefer not to approach the possessor directly.

The second method is simple taking: tearing off a portion of the prey or even seizing bits of meat from the possessor's hands or mouth. This direct approach is often used when the possessor is a female and the taker is one of its offspring. Male possessors are most likely to tolerate meat-taking by another adult male, particularly a sibling, by a female who is sexually receptive or by a female of high social status. The direct approach often finds taker and possessor calmly chewing on the same piece of meat, and I have seen as many as three sharing-cluster apes so engaged without discouragement from the possessor. Subadults rarely attempt the direct approach.

The third course of action, which involves specific behavioral patterns, I call requesting. Meat can be requested in the following ways. The requester can approach the possessor closely, face to face, and peer intently at the possessor or at the meat. Alternatively the requester can extend a hand and touch the possessor's chin or lips or touch the meat. The requester can also extend a hand, open and palm up, holding it under the possessor's chin. The requester may accompany each of the gestures with soft "whimper" or "hoo" sounds. Chimpanzees of virtually any age and of either sex request meat with this repertory of gestures. The youngest I observed doing so was an 18-month-old infant; the most important socially was "alpha" Mike, the top-ranking chimpanzee at Gombe Park.

If the possessor's response to a request is negative, the denial is indicated by ignoring the requester or turning away, by pulling the meat out of reach, by moving to a less accessible place, by vocalizing, by gesturing and occasionally by pushing the requester away. The possessor responds in the affirmative by allowing the requester to chew on the meat or tear off a portion, or by dropping a piece of meat into the requester's upturned hand. Of the 395 requests I observed, 114 were rewarded. Occasionally the possessor will detach a considerable portion of meat and with outstretched hand offer it to the requester. I saw this done only four times during my year at Gombe. On one of those occasions the possessor was holding an entire carcass. An adult male in the sharing cluster had been requesting meat persistently and the possessor at last divided the carcass in two and handed half to the requester.

When chimpanzees are sharing meat,

GESTURE OF REQUEST, arm outstretched with open hand palm upward, is used by a subadult chimpanzee to seek from its mother some of the leaf wad the mother is chewing.

their behavior is generally relaxed and uncompetitive. The pieces of the prey animal are consumed in a leisurely fashion and are evidently relished. I saw very few hostile interactions between apes with meat and apes with none. Individuals in a sharing cluster that had waited in vain for a long time would sometimes threaten or chase other chimpanzees that were also waiting, but they never made such a move against the chimpanzee with the meat.

High social rank is apparently no guarantee of success in requesting meat. Even though the possessor may be a relatively low-ranking adult, a high-ranking adult in the sharing cluster will approach the possessor with the same repertory of gestures that the other requesters use and will make no effort to assert superior social status aggressively. I once observed "alpha" Mike request meat for several hours from a subordinate male; Mike received nothing for his pains. Twice during my year of observing I saw meat seized from a possessor by a surprise dash-and-grab maneuver that may have had hostile overtones; the possessor's only retaliation was to scream and wave its arms. In the total of 43 hours of meat-sharing that I observed not once did two chimpanzees fight over possession of meat.

To what extent do the Gombe findings and similar observations elsewhere suggest a modification of present views concerning human evolution? The theme of hunting as a way of life has been central to many of the hypotheses that attempt to trace the evolution of human behavior and human social organization. For example, it is suggested that the evolution of erect posture left the evolving hominids' hands free to hold weapons and other tools and cleared the way for the pursuit of game on the open savanna. This development is often pictured as the crucial point in primate evolution: the time when "ape-men" (later to become "man-apes") expanded their range from the forests of Africa to the grasslands. As this environmental expansion took place, the hypotheses suggest, the evolving hominids acquired both the technology and the cooperative social organization, including the sharing of labor and food within some kind of nuclear family, that enabled them to survive as nomadic, omnivorous hunter-gatherers.

It seems possible to me that predation developed among primates long before the advent of ape-men. Many primate species have evidently shared a number of adaptive features for some millions of years. Might not the practice of preying on mammals, and on other primates in particular, be one of these features? To an omnivorous primate that enjoys complete mobility in a three-dimensional habitat other primates are perhaps the most readily accessible prey. Let us briefly test this concept against the record.

Most of the mammals the Gombe chimpanzees kill and eat are primates. Chimpanzees elsewhere in Tanzania and in Uganda have also been observed hunting and eating monkeys. In several other parts of Africa baboons are known to kill fellow primates. Turning to the distant past, the fossil remains of *Australopithecus* in South Africa have been found in association with the damaged skulls of baboons. The association at least suggests the possibility that *Australopithecus* occasionally preyed on the primates that shared its range.

Suppose that predation, cooperative hunting and socially structured food-sharing did become habitual among certain primates long before the first hominids arose. If that were the case, it would throw doubt on a number of current evolutionary hypotheses. For example, the sequence of erect posture, free hands and tool use as prerequisites to the emergence of hunting behavior would no longer appear to be valid. The same is true of the hypothesis that the open savanna is the habitat where hunting most probably developed. Moreover, whether or not predation is a far more ancient primate behavior pattern than has been supposed, at least one hypothesis must be abandoned altogether. As the actions of the Gombe chimpanzees demonstrate, socially organized hunting by primates is by no means confined to man.

The Gombe chimpanzees can be described in summary as omnivorous forager-predators that supplement a basically vegetarian diet in various ways, including the optional practice of hunting other mammals, with fellow primates being their most favored prey. The fact that the chimpanzees' meat-eating is optional means that in terms of meat consumption they do better than purely opportunistic omnivores that "collect" and eat an immobile prey animal now and then. At the same time the Gombe chimpanzees are free of the exclusive dependence on a meat diet that is characteristic of most carnivores.

The primary significance of the Gombe chimpanzees' predatory behavior is not, however, dietary. What is far more important is the behavior that accompanies the predatory episodes: cooperation in the chase and the sharing of the prey. The more we learn about primate behavior, the smaller the differences between human and nonhuman primates appear to be. The observations at Gombe do much to reinforce this conclusion.

We want to be useful ...and even interesting

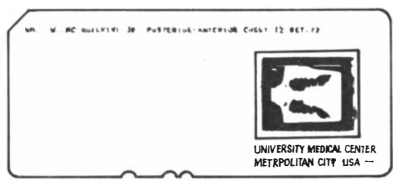
X-rays next year

What you see here is seldom seen. It is a widely distributed product of mass production: a 14 x 17 sheet of medical x-ray film.

Nobody buys this product simply to have it and enjoy it. All it does is capture and store information which radiologists interpret for other physicians. The radiologist's whole training and experience have been carefully tuned to its properties, but for storing information, which is the one and only product of diagnostic radiology, it is cumbersome and inefficient.

Current emphasis being what it is on efficiency in the delivery of medical care, Kodak intends about a year from now to market a group of devices and supplies which, where appropriate, would permit the large sheet of x-ray film to be discarded (for recovery of its rich store of scarce silver) the very same day it is exposed and processed.

One of the planned devices gets fed once a day a whole day's production of a radiology department's films. At the same rapid rate the device emits miniature film copies of the day's radiographs, each mounted in a window



cut in a standard tab card. The device is loaded by insertion of a light-tight box of the cards with the raw film already mounted in them; the automatic processing to archival keeping standards takes place right on the card, which is rather clever. Each patient's card, perhaps notched for a simple manual retrieval method, goes into his individual packet. This remains in the file room always. Copy cards duplicated from the primary cards are displayed on a reader that has powers an ordinary microfilm reader doesn't need. The microfilm radiographs carry densities to nearly 4.0. A piece of film on which portions of the image pass only 1/10,000 of the incident light needs a rather efficient illumination system and some good ideas on flare and glare prevention.

All this implies that microfilmed radiographs can convey to the diagnostician's mind no less intelligence than originally captured by the big film. Not "can," but do. And not mere opinion that they do but mathematically sound proof to withstand scrutiny by competent but disinterested critics. A comparison has indeed been done for two years by double-blind methods in a reputable school of medicine and is continuing. Since Kodak scarcely qualifies as disinterested in how it turned out, we consider it inappropriate to summarize the findings here or even give a specific citation in an advertisement. The kind of buying decisions required don't get made from advertisements.

For industrial x-ray, there is another way to go

Röntgen discovered x-rays on November 8, 1895. Early in 1896 a Yale professor made the first industrial radiograph in the U.S.A. The subject was a weld. He did it for the Ordnance Department of the U.S. Army. He used photographically sensitized paper, not plates or film. It took nearly 50 years for industrial radiography to catch hold on a large scale. When that finally happened, it was virtually all done on film.

Last year, under the trademark "Industrex," we introduced x-ray paper, a new kind of intensifying screen that brings the speed of the paper up to the speed that film can offer because x-ray film carries emulsion on both sides, and a processor for the paper that delivers the result in 10 seconds (handy to roll out to a loaded jetliner if there is a question about whether it should take off).

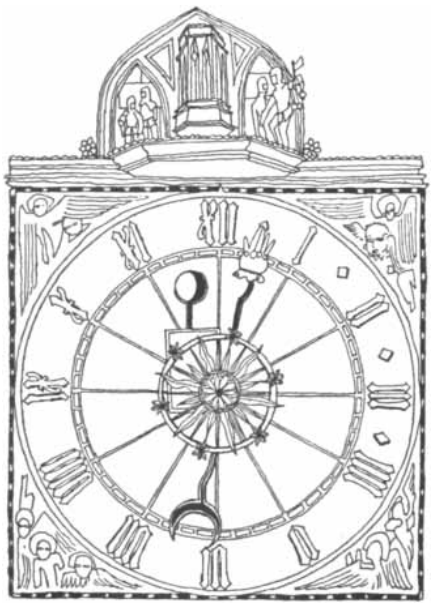
We price the paper at about 27% of the price of our industrial x-ray film. For users who merely want it to answer questions that come up, it seems like a very good deal indeed. Users who need x-ray for compliance with a code that specifies "film," as well as some code-making authorities themselves, are now thinking about that word "film."

For fuller information on either of the above subjects, write Kodak, Dept. 740, Rochester, N.Y. 14650.

Kodak

Already in business by 1895

SCIENCE AND THE CITIZEN



Secrets

Although the volume of secret Government research conducted in U.S. universities has declined sharply in the past decade, in part because of protests by students and faculties, a number of large institutions, chiefly state universities, continue to undertake classified projects. In the fiscal year 1972 the Department of Defense has at least 29 classified contracts with universities, not counting contracts for work done at off-campus facilities. Of the 29 contracts 12 are with two institutions: the University of Texas and the University of Michigan.

It cannot be ascertained what fraction of all Department of Defense contracts are represented by the 29 still in force, but the figure is probably less than 2 percent. In the preceding fiscal year (1971) the Department of Defense funded 3,700 university projects, of which 121, or 3 percent, were classified. Of the 3,700, 2,833 were for basic research, and all were unclassified according to Government policy. The 121 classified projects were all in the field of applied research. The 3,700 contracts and grants had a value of about $220 million.

These figures represent a significant decline from the fiscal year 1969, when classified agreements represented 10 percent of the $250 million in Department of Defense funds received by the universities. The $250 million, involving 5,000 contracts and grants, went to 250 academic institutions. That year some 60 institutions accepted classified agreements. The 1969 figures, in turn, were a decrease from 1960, when the Federal Government spent $449 million for research and development on U.S. campuses, with somewhere between 10 and 15 percent supporting secret work.

In the late 1960's both students and faculties began protesting against the willingness of universities to accept classified contracts. As a result some of the biggest recipients of such contracts (for example the University of Pennsylvania) announced that secret work would no longer be allowed on campus. A number of these institutions, however, retain off-campus facilities where secret work is done. The Massachusetts Institute of Technology, for example, administers the Lincoln Laboratory, and the University of Pennsylvania has a major interest in the nearby University City Science Research Institute, which it helped to establish in 1964.

Some universities maintain special libraries of classified documents for the convenience of faculty members who have Government clearance and work as consultants to Federal agencies. Some academic laboratories that do not normally undertake classified work may occasionally do so for compelling reasons. For example, at the University of California at Berkeley part of the chemistry building at the Lawrence Berkeley Laboratory has been a classified area for a few weeks when samples of debris from a foreign nuclear explosion are received. Although the research based on the samples can be unclassified, the composition of the samples themselves is a closely guarded secret.

The question of whether or not an academic institution should engage in secret Government research remains at issue. Faculties in engineering departments often favor classified work, are accustomed to it and believe it is essential if they are to stay abreast of developments in their field. Faculties in the natural and social sciences are often militantly opposed to accepting secret work. At the University of California at Berkeley, at the University of Michigan and at the State University of New York at Stony Brook administration officials or trustees have rejected faculty recommendations that classified work, or all work for the Department of Defense, be discontinued.

It is sometimes overlooked that projects sponsored by the Department of Defense and the Atomic Energy Commission, even those calling for basic research, have a clause allowing the project officer to classify any piece of work or discovery that may come out of the research if, in his judgment, classification is in the national interest. Although his decision can be appealed, it is clear that whoever makes the final decision will be influenced by the climate of international relations at the time.

The Ancestry of Corn

The exact ancestry of many important food plants (for example wheat, barley, rice, sorghum and soybeans in the Old World and beans, white potatoes, sweet potatoes and peanuts in the New World) is uncertain because there are too many likely candidates among wild species. The same uncertainty exists for quite the opposite reason with respect to the most important of all New World foods: corn. Not one of the New World wild grasses bears any close morphological resemblance to *Zea mays,* with its many-kerneled, leaf-wrapped ear. Yet corn must be descended from some New World grass. The question is: Which one?

An old answer has now been reproposed. After three years of breeding experiments, George W. Beadle, the geneticist who retired as president of the University of Chicago in 1968, reaffirms his belief of 40 years ago that the ancestor of corn is teosinte, a wild grass that still grows in Mexico, Guatemala and Honduras. If Beadle's view is accepted, it will restore teosinte to the eminence it enjoyed among 18th- and 19th-century botanists but lost during the first half of this century.

Although the seed spike of teosinte looks very little like an ear of corn, it seemed to early students of the question that teosinte was the most likely ancestor of corn on several grounds. Teosinte grows wild in the same region of the New World where corn may first have been domesticated. Corn has 10

chromosomes; so has teosinte. Moreover, the two plants are mutually fertile, and when they are hybridized, the 10 chromosomes of teosinte pair normally with the chromosomes of corn. Teosinte does not grow a cob, and the six or more triangular grains that make up each spike contain kernels smaller than those of corn. When the kernels are heated and "popped," however, they bear a close resemblance in appearance and taste to popcorn.

In spite of such arguments in its favor, teosinte began losing support as a potential ancestor of corn because of two lines of investigation that began in the late 1930's. One, pursued by several investigators, involved the frequency of reappearance of parental types in the second and later generations of hybrids between teosinte and corn. The results of these studies seemed to suggest that the genetic differences between the two plants were so great as to make untenable the view that teosinte had been directly transformed into corn through selection in the earliest stages of agriculture.

In a second approach Paul C. Mangelsdorf and Robert G. Reeves succeeded in hybridizing corn with another New World grass: *Tripsacum*. Because teosinte is intermediate between corn and *Tripsacum* they were led to hypothesize that (as Edgar Anderson of the Missouri Botanical Garden had suggested some years earlier) teosinte was more probably a descendant of corn than its ancestor. They proposed that some early form of wild corn had undergone chance hybridization with *Tripsacum*, thus giving rise to teosinte.

Mangelsdorf and his associates went on to seek a new candidate for the ancestral position. They found it in a primitive variety of domesticated corn: "pod corn." When they rendered pod corn increasingly primitive by successive inbreedings, they found that the plants ceased to form cobs and instead produced a number of loose kernels that broke off readily. Kernels that are easily dispersed would have been crucial to the survival of the first wild corn plants that evolved; if men did not strip the ears of domesticated corn and sow the kernels, the plant would become extinct in one generation. Mangelsdorf's achievement of a self-reproducing ancestral form of corn meant that by the 1950's there were only a few scholars who any longer espoused the cause of teosinte.

Among those few was Beadle, who had first studied corn-teosinte hybrids under R. A. Emerson at Cornell University late in the 1920's. Once free to resume active research in 1968, Beadle quickly took up this work of his youth. With the assistance of collaborators in Mexico he undertook to rear populations of some 50,000 second-generation and backcross corn-teosinte hybrids. Beadle's experiment was directed at measuring the essential genetic differences between the parent stocks. According to Mendelian principles, if there is a difference of only one gene between the parents, each original parental type will reappear in the plants of the second generation with a statistical frequency of one in four. The progression is geometric: with two independent gene differences the frequency of reappearance is one in 16, and with 10 gene differences the frequency is less than one in a million.

Beadle selected as one parental stock a particularly primitive Mexican variety of corn named Chapalote and as the other stock the most cornlike variety of Mexican teosinte, named Chalco. At the end of the third season of hybridization, he announced recently, the frequency of parental reappearance has proved to be about one in 500. This figure is somewhat lower than would be expected if there were five independently inherited gene differences between the parental stocks.

This suggests that the genetic distance between corn and teosinte is substantially less than had been supposed, and that at least in this respect teosinte can be reinstated as a possible ancestor of corn. Moreover, Mangelsdorf, while still firm in his opinion that teosinte is a descendant of corn and not its ancestor, has recently declared his abandonment of the hypothesis that teosinte is a hybrid of *Tripsacum* and corn. Mangelsdorf now proposes that teosinte is instead a mutated form of some ancient wild corn. Beadle comments: "If corn could have given rise to teosinte, the reverse is also possible. It is also much more probable, because teosinte is a highly successful wild plant. No known living form of corn is able to survive without the intervention of man."

Blowup

Last September the celestial X-ray source Cygnus X-3 flared up several times in a quite unprecedented way. The series of explosions was so spectacular that it was monitored by observatories all over the world. Curiously, in spite of the fact that Cygnus X-3 is an X-ray source the outbursts were detected in the radio regions of the spectrum and not at X-ray wavelengths. Observers were able to determine the approximate distance and diameter of the source and to propose a mechanism to account for the outbursts.

Cygnus X-3 had been known to be a highly variable but weak radio source. On September 2 a group led by P. C. Gregory of the University of Toronto observed the source with the 46-meter radio telescope at the Algonquin Radio Observatory in Ontario; they found that the radio emission had increased 1,000 times above the level recorded only two days earlier by R. M. Hjellming and B. Balick at the National Radio Astronomy Observatory in Green Bank, W.Va. The source continued to brighten at radio wavelengths for two hours after the outburst was discovered, and then it began to fade until it was at its usual level by September 12. Hjellming and Balick reported another explosion on September 19. By October 6 a total of four outbursts had been detected.

Cygnus X-3 was observed at X-ray wavelengths by the X-ray satellite *Uhuru* and by a steerable X-ray telescope aboard *Copernicus*, the fourth Orbiting Astronomical Observatory. The observations showed no corresponding increase in X-ray emission, and this negative result was corroborated by data from the spacecraft *Vela 5B*. The absence of an X-ray outburst can be understood on the basis of the mechanism proposed to account for the source. Cygnus X-3 appears to be a cloud of electrons and protons expanding at velocities close to the speed of light. The higher-energy particles that can generate X rays have lost energy by collision. Thus the emissions of the cloud are not X rays but the longer wavelengths of "synchrotron radiation" generated by the electrons as they spiral outward along lines of force in the cloud's magnetic field. On September 2 and 3 there were large fluctuations, about an hour apart, in the amount of radiation emitted by Cygnus X-3. Such fluctuations would be characteristic of emission from bursts of high-energy electrons periodically injected into the cloud. The major variations, however, generally took about a day, implying that the object is roughly one light-day across (some 2.5 times the diameter of the solar system).

Several methods were used in an effort to determine the distance of Cygnus X-3 during the outbursts. The best estimate was obtained by French astrono-

mers at the Observatoire de Paris-Meudon. They observed the spectral line of hydrogen at the radio wavelength of 21 centimeters and measured the amount of radiation that was absorbed by clouds of neutral hydrogen between the source and the earth. By this method they estimated that the distance to Cygnus X-3 is between 26,000 and 36,000 light-years, which places the source near the edge of the galaxy and behind large amounts of interstellar dust.

Four weeks after the first radio outburst a group at the Hale Observatories in California used the 200-inch telescope on Palomar Mountain to see if they could detect any infrared radiation at the radio position. On the night of October 2–3 an infrared-emitting source was discovered less than two seconds of arc away. It is probably associated with Cygnus X-3. There is no visible object brighter than the 17th magnitude at the infrared position, however, and all attempts to find a visible counterpart to the radio and X-ray source have been unsuccessful. The amount of interstellar dust between Cygnus X-3 and the solar system could obscure any radiation at visible wavelengths that the source may be emitting.

The outbursts of Cygnus X-3 have provided a unique opportunity for astronomers to examine the nature of a rapidly evolving radio source. In many respects the outbursts are analogous to the explosions that are observed in variable radio sources outside the galaxy. Observations of Cygnus X-3 at various wavelengths will for the first time provide information with which to test theories of expanding synchrotron sources. Moreover, since Cygnus X-3 is ejecting such high-energy particles, it must be regarded as a possible source of cosmic rays in the Milky Way.

The Population of the U.S.S.R.

At a time when many nations are concerned that their population is too large, the U.S.S.R. may be arriving at the conclusion that its population is too small. A survey of population trends in the U.S.S.R. has led David M. Heer, professor of sociology at the University of Southern California, to the opinion that the Soviet government may take steps to promote an increase in the national birthrate through moves to raise fertility in low-fertility areas. Writing in *Studies in Family Planning,* a publication of the Population Council, Heer says: "I believe that two features of the changing demographic situation may have become of special concern to the Soviet regime. The first of these was the decline in the Soviet rate of natural increase relative to that of other nations. ... The second major concern, in my opinion, is with the relatively high proportion of all births that occur to the Muslim nationalities in the Soviet Union and the relatively low proportion that occur to the Slavic groups."

Heer points out that in the period from 1955 to 1961 the rate of natural increase in the U.S.S.R., ranging from 16.6 per 1,000 to 18.1 per 1,000, was higher than that in the U.S. and substantially higher than that in most other developed nations. "Although lip service was paid to the idea that the rate of population growth was not sufficiently high, in fact the actual consequences of much Soviet legislation were then antinatalist." A liberalized abortion law adopted in 1955, for example, was of major importance in reducing the birthrate; Heer once estimated that in the late 1950's the U.S.S.R. had more abortions than live births. Another factor was that the system of family allowances, which at one time encouraged parents to have large numbers of children, was changed so that its value with respect to the average wage steadily declined. Two other policies that worked against a high birthrate were that Russian women were encouraged to participate in the labor force and that the Soviet government did not supply its citizens with enough housing space to encourage a large number of children. These policies, together with changes in the age and sex composition of the population, caused the natural rate of increase to decline to 11.1 per 1,000 in 1965 and to 9.2 per 1,000 in 1970. Up to now, Heer writes, "no formal government policy [on increasing the birthrate] has emerged," but he sees evidence that the issue is being keenly debated in government circles.

Here Comes the Haboob!

A standard article of folklore is the desert sandstorm, in which a wall of windblown sand marches across the desert, blotting out the sun and choking man and beast. Such storms do occur, and they are particularly well known in the area around Khartoum in the northern Sudan. The storm is called a haboob, from the Arabic for "violent wind." It now seems that there are storms in the southwestern U.S. that should also be called haboobs.

According to S. B. Idso, R. S. Ingram and J. M. Pritchard of the U.S. Water Conservation Laboratory and the National Weather Service in Phoenix, Ariz., a haboob is announced by a sudden change in the speed and direction of the wind. The air grows humid, the temperature drops and what looks like a solid wall of dust roars in over the countryside. The dust is raised from the arid land and driven along at as much as 45 miles per hour by an outflow of rain-cooled air from a towering cumulonimbus cloud.

The American haboobs are not so frequent as the Sudanese (two or three a year at Phoenix as compared with perhaps 24 a year at Khartoum) but they can be just as spectacular, the authors write in the *Bulletin of the American Meteorological Society.* They are usually caused by downdrafts from thunderstorms that originate in the Sierra Madre Occidental mountains in northern Mexico and move northwest, forming a squall line near Tucson. The line of storm cells, each with its own downdraft and churned-up dust cloud, moves northwest, the individual clouds merging to form a dust front that can extend as high as 8,000 feet. Within the dust cloud the visibility is reduced to a few hundred yards.

The authors report in detail on one such storm that occurred on July 16, 1971. It formed southeast of Tucson, passed through that city at 3:30 P.M. and moved into the valley of the Santa Cruz River, which flows northwest to Phoenix. Reports from several points along the path establish its mean speed of advance as 29 miles per hour. It reached Phoenix at about 7:00 P.M. As it moved up the valley the relative humidity jumped from 33 percent to 74 percent and the temperature dropped more than 23 degrees Fahrenheit; at each site the changes took place in only about seven minutes. Observers along the way photographed the storm's leading edge, revealing that it was composed of at least three "macrolobes": distinct arc-shaped fronts of cold air churning up surface soil and dust; each of these was composed, in turn, of smaller "microlobes." The storm took an average of 48 minutes to pass each point. Reports from aircraft put the maximum height of the dust cloud at 8,000 feet; radar indicated that some of the storm cells generating it reached 55,000 feet. It was, the authors conclude, "as good an example of a true haboob as those that occur in the Sudan."

Buick's new Opel Manta.
2,000,000,000 miles in Europe. $2,769 in America.

Manta. $2,768.50 for the plush Manta Luxus with all the standard features listed on the price sticker below. Give up a little plush and Manta is $2,578.50. Or go sporty with Manta Rallye—$2,757.50. All prices are Manufacturer's Suggested Retail Prices, including dealer new vehicle preparation charges. Optional equipment, State and local taxes and destination charges are additional.

Manta. The personal-size sport coupe proven in more than 2 billion miles of actual driving. Not on test tracks or proving grounds but on European roads.

Manta. With rack and pinion steering for twisting Alpine curves. A 1.9 liter engine for responsive autobahn driving. Disc brakes up front. And a grand-touring ride you'll just have to feel.

Manta. Luxurious. Economical. Comfortable. Contour bucket seats that recline.

Interiors that even include corduroy fabric in the Luxus model.

The Opel Manta, Manta Rallye, Manta Luxus.
Really ready for America.

BUICK MOTOR DIVISION
GENERAL MOTORS CORPORATION
MANUFACTURER'S SUGGESTED RETAIL PRICE

OPEL

HOMETOWN BUICK, INC.
HOMETOWN, U.S.A.

DELIVERED TO DEALER AT (DESTINATION) ARLINGTON, VIRGINIA

VEHICLE IDENTIFICATION NO. OPEL 00000 0 00000 FINAL ASSEMBLY POINT (PORT OF ENTRY) BALTIMORE, MARYLAND

ENGINE NO. 00 0000000 MODEL 57L 2DR MANTA LUXUS CPE

1 Suggested Retail Price of this Vehicle $2,768.50

THE FOLLOWING ITEMS ARE STANDARD EQUIPMENT ON THIS 1973 MODEL OPEL: 1.9 LITER ENGINE DESIGNED FOR NO-LEAD, LOW-LEAD FUEL, WITH HYDRAULIC VALVE LIFTERS AND AUTOMATIC CHOKE • POWER BRAKES WITH FRONT DISCS, SELF ADJUSTING • CARPETED TRUNK COMPARTMENT, PACKAGE SHELF AND FLOOR • CUSTOM CLOTH TRIM WITH VINYL EDGING, AND COLOR COORDINATED INTERIOR • HINGED REAR QUARTER WINDOWS • RUBBER BUMPER GUARDS AND STRIPS • ELECTRIC CLOCK • RALLYE TYPE WHEELS AND STEERING WHEEL • ADJUSTABLE FRONT SEAT BACKS.

THIS VEHICLE WAS MANUFACTURED IN [illegible] WITH ALL APPLICABLE FEDERAL MOTOR VEHICLE SAFETY AND EMISSION CONTROL STANDARDS 22.00

TOTAL AMOUNT $2,790.50

German precision imported by General Motors.
Sold and serviced by 2200 Buick dealers.

MARS FROM MARINER 9

The first spacecraft to go into orbit around another planet provides evidence that Mars is just beginning to heat up internally. Systems of channels and gullies suggest erosion by water or some other agent

by Bruce C. Murray

A little more than a year ago, in November, 1971, the complex robot spacecraft *Mariner 9* fired its braking rocket and was captured in orbit around Mars, thus becoming the first man-made satellite of another planet. From its orbital station, ranging between 1,650 and 17,100 kilometers (1,025 and 10,610 miles) above the surface, *Mariner 9* started sending back to the earth a steady stream of pictures and scientific information that was to continue for nearly a year. By the time its instruments had been turned off *Mariner 9* had provided about 100 times the amount of information accumulated by all previous flights to Mars. It had also decisively changed man's view of the planet that generations of astronomers and fiction writers had thought most closely resembled the earth. As a result of the *Mariner 9* mission it is now possible to make plausible conjectures about the geology of Mars, conjectures comparable, say, to those made about the moon in the early 1960's.

It will be recalled that the first close-up pictures of Mars, made in 1965 by *Mariner 4,* revealed a planetary surface whose principal features were large craters reminiscent of the bleak surface of the moon. Four years later the pictures sent back by *Mariner 6* and *Mariner 7* showed that the Martian surface was not uniformly cratered but had large areas of chaotic terrain unlike anything ever seen on either the earth or the moon. In addition a vast bowl-shaped area, long known to earthbound astronomers as the "desert" Hellas, turned out to be nearly devoid of features down to the resolving power of the Mariners' cameras. None of the pictures returned by the first three Mariners showed any evidence of volcanic activity, leading to the view that Mars was tectonically inactive.

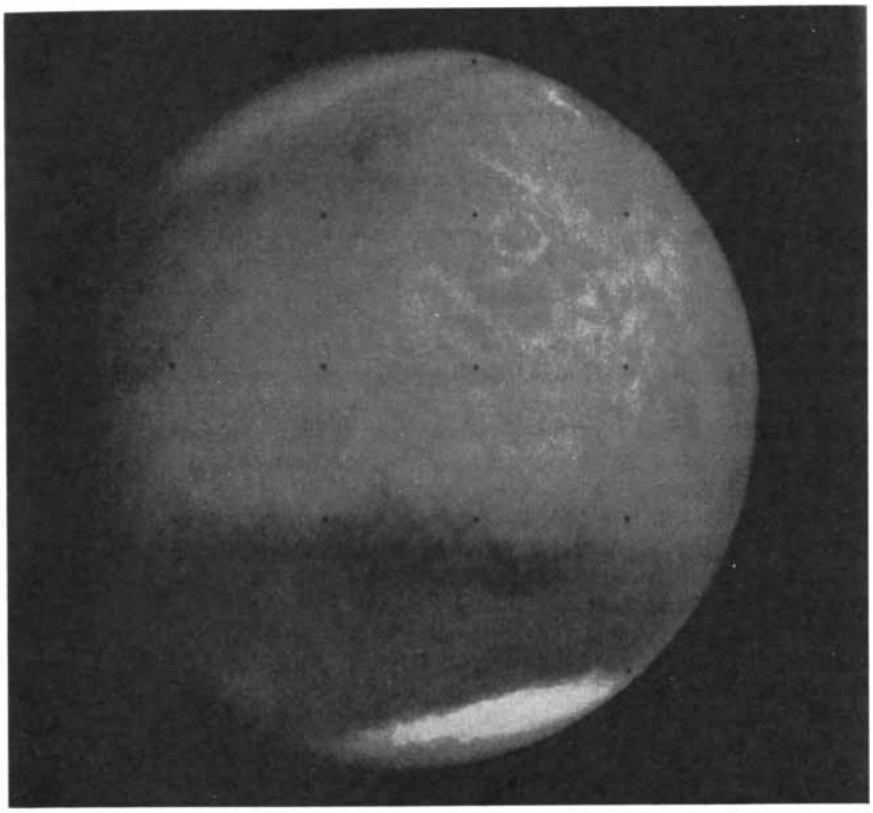
This view has had to be drastically revised in the light of the photographs sent back by *Mariner 9*. The new evidence emerged slowly as the clouds of dust that had shrouded the planet for weeks settled. It revealed, among other things, four large volcanic mountains larger than any such volcanic features on the earth. The *Mariner 9* pictures also show a vast system of canyons, tributary gullies and sinuous channels that look at first glance as if they had been created by flowing water. Elsewhere on the planet's surface there is no suggestion of water erosion. That is probably the major mystery presented by the highly successful mission of *Mariner 9*.

Designed and built by the Jet Propulsion Laboratory of the California Institute of Technology, as were the earlier Mariners, *Mariner 9* was crammed with instruments and electronic gear. After burning 900 pounds of retro-rocket fuel, which it had transported 287 million miles in 167 days, *Mariner 9* weighed 1,350 pounds when it finally went into orbit around Mars. The cameras and instruments it carried were designed by several groups of investigators from Government laboratories and more than a score of universities. The television team, to which I belonged, was headed by Harold Masursky of the U.S. Geological Survey and had nearly 30 members. Somewhat smaller groups were responsible for designing and analyzing data from the ultraviolet spectrometer, the infrared radiometer and the infrared interferometric spectrometer. Other groups had the task of analyzing the trajectory data (which have provided information about the gravitational anomalies of Mars) and the data provided by nearly 100 occultations of the spacecraft's radio signals (which have yielded new knowledge of the planet's atmosphere and surface).

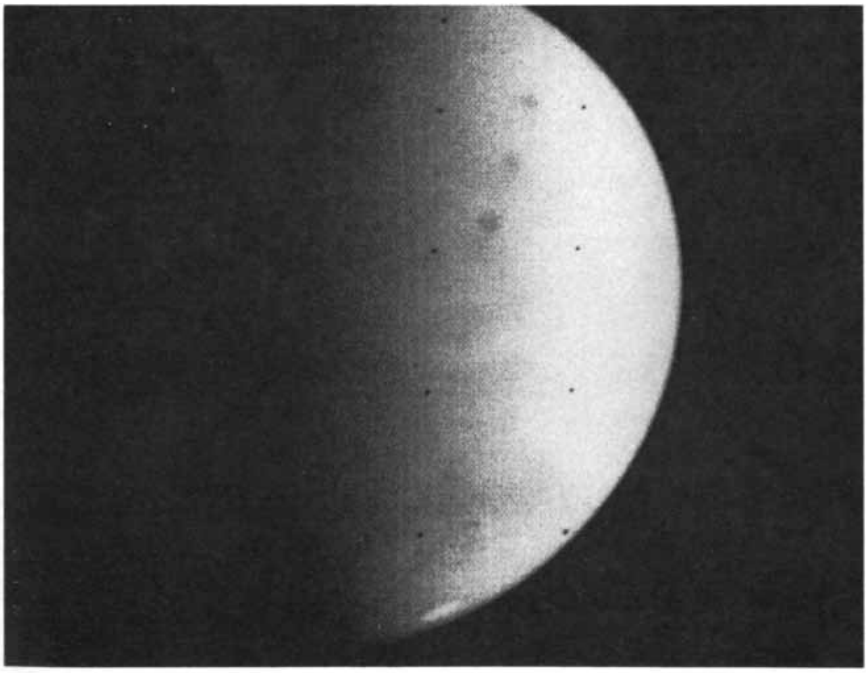
CHANGES ON SURFACE OF MARS are depicted in the photograph on the opposite page by a color-difference technique developed by J. A. Cutts of the Jet Propulsion Laboratory and J. J. Rennilson of the California Institute of Technology. It consists of two pictures of the same region in Depressio Hellespontica taken 37 days apart by the wide-angle camera on *Mariner 9*. Both pictures were originally taken through the same filter but at different angles and distances. The two images were rectified by a computer program at J.P.L.'s Image Processing Laboratory. To bring out changes that had taken place on the surface of the planet the earlier picture was printed in green and the later picture was printed in red. When the two pictures are superposed, surface areas that had become darker in the interval between the taking of the two pictures show up with a greenish cast whereas areas that had become lighter appear more reddish. The entire area in the composite pictures corresponds roughly to the smallest area resolvable with earth-based telescopes. Astronomers have observed seasonal changes in the markings of Depressio Hellespontica. Evidently the changes represent an integration of the detailed changes visible in the two *Mariner 9* photographs.

The Old Mars

There were strong reasons for the traditional belief in the resemblance between Mars and the earth. Mars rotates once every 24½ hours and its axis is tipped from the plane of its orbit by almost exactly the same amount as the axis of the earth, thus providing the same basis for the seasonal changes in the amount of solar radiation received by the planet's two hemispheres. Mars has white polar caps, originally thought



CLEAR PICTURE OF MARS was taken by *Mariner 7* at a distance of 395,249 kilometers as it approached the planet in August, 1969. Since it was then winter in Mars's southern hemisphere the south polar cap was at its maximum size. The north pole was shrouded in haze. The bull's-eye-shaped feature in the upper right quadrant, known as Nix Olympica ("Snows of Olympus"), was thought to be a huge crater. Subsequent closeup photographs taken by *Mariner 9* revealed it to be a gigantic volcanic mountain (*see illustration on page 58*).



MARS SHROUDED BY DUST was photographed by *Mariner 9* at a distance of 400,000 kilometers, a day and a half before reaching the planet on November 13, 1971. The greatest dust storm in more than a century had obliterated all surface features except for the south polar cap, rapidly shrinking with the approach of spring in the southern hemisphere, and four dark spots in the upper right quadrant. The spot nearest the shadow line is the top of the volcanic mountain Nix Olympica. The other three spots are also volcanic peaks.

to be composed of water, that alternate from one hemisphere to the other once every Martian year (687 earth days). The planet also exhibits dark and light markings that change on a seasonal basis.

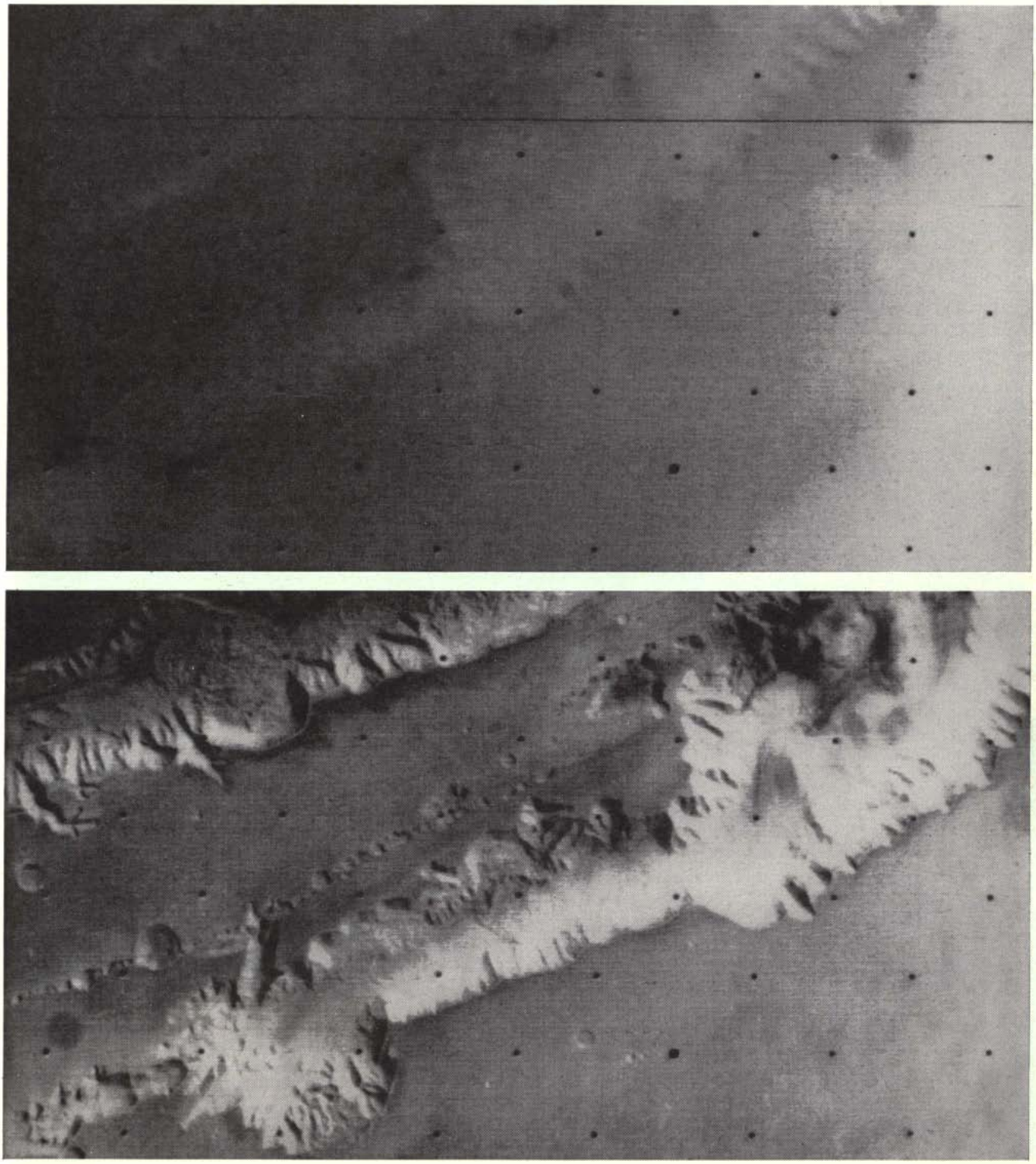
Early astronomers speculated that the dark markings might be vegetation. Later and more cautious workers still found it plausible that Mars had had an early history similar to the earth's, which implied the existence of oceans and an atmosphere with enough water vapor to precipitate and erode the surface. Because of Mars's small mass (a tenth the mass of the earth) and low gravity, such an aqueous atmosphere was assumed ultimately to have escaped, leaving the planet in its present arid state. This view of an earthlike Mars strongly influenced proposals for the biological exploration of the planet at the beginning of the space age. It seemed reasonable to suppose that life could have originated on Mars much as it had on the earth, presumably as the result of high concentrations of suitable precursor molecules in primitive oceans. Once life had appeared on Mars, microorganisms, at least, could very well have been able to adapt to changing environmental conditions and so could have survived for discovery and analysis by robot devices launched from the earth.

Such expectations were dampened by the findings of *Mariner 4*. Not only did Mars appear bleak and moonlike but also it was found to lack a magnetic field, which could have shielded its surface against energetic charged particles from the sun. Moreover, Mars's atmospheric pressure was found to be less than 1 percent of the earth's, lower by a factor of at least 10 than had previously been estimated. Since the force of gravity at the surface of Mars is more than a third the force of gravity at the surface of the earth, Mars should have easily been able to hold an atmosphere whose pressure at the surface was a tenth the pressure of the earth's atmosphere at the surface.

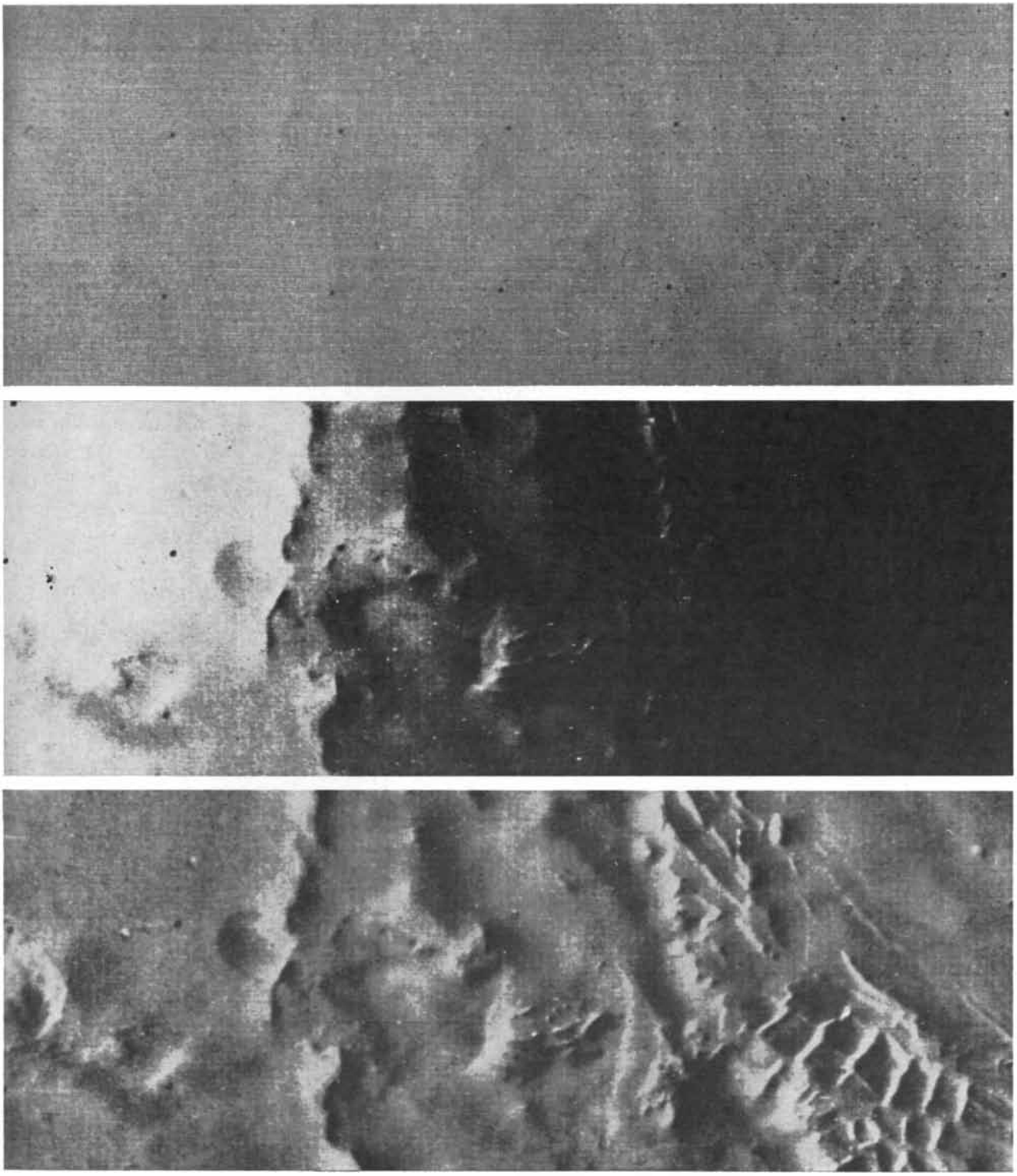
Mariner 6 and *Mariner 7* extended these observations. They confirmed that the polar caps are composed of very pure solid carbon dioxide—"dry ice" rather than water ice. The pictures revealing a chaotic terrain suggested that parts of Mars's surface had collapsed and that there had been a certain amount of internal activity. As a result some investigators speculated that the planet might just now be heating up, a circumstance suggested independently by thermal models of the interior. The preponderant view of the Mariner experimenters, however, was still that

Mars was basically more like the moon than like the earth. By then the light and dark markings on Mars seen through telescopes were generally attributed to some kind of atmospheric interaction with dust. Indications that the interaction was controlled by local topography were seen in the second set of Mariner photographs, but no general explanation was deduced. Even so some investigators held to the belief that the markings might instead reflect variations in soil moisture.

The 1971 mission to Mars was originally designed to employ two spacecraft, *Mariner 8* and *Mariner 9*, both of which were to be placed in orbit around the planet. The purpose of the two orbiters was to map most, if not all, of the planet's surface at a resolution high enough to reveal both external and in-



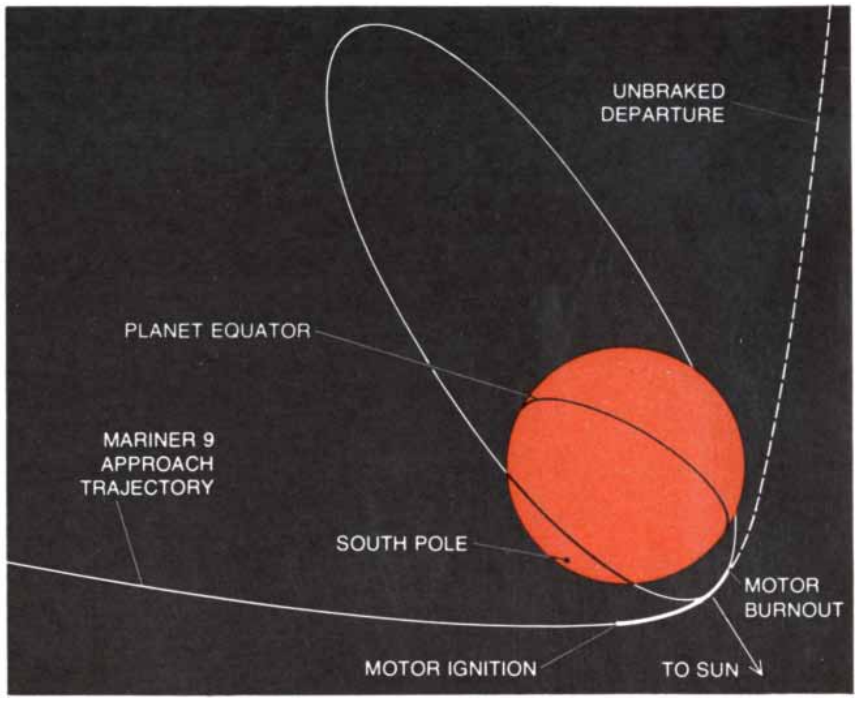
CLEARING OF DUST STORM is shown in these two views of a region in Coprates near Mars's equator. The picture at top, taken when *Mariner 9* had been in orbit 41 days, gives little or no hint of the rugged canyon that was revealed in the bottom picture, taken on *Mariner 9*'s 80th day in orbit. Coprates had long been recognized by astronomers as a feature that changes unpredictably in brightness. It now appears that the canyon looks brighter than the surrounding region toward the end of a dust storm, when the canyon atmosphere is still filled with light-reflecting particles and atmosphere above surrounding plateau has become largely dust-free.



RESULTS OF COMPUTER PROCESSING are demonstrated in these three displays of the same picture showing a region near Mars's south pole. The processing, done within the space of five minutes by J.P.L.'s "mission test computer," was carried out on every picture as it was received from *Mariner 9*. Each image was transmitted as a coded radio signal in which the brightness of each point in the picture was represented by a sequence of nine binary digits rather than as an intensity level (AM signal) or as a frequency tone (FM signal). Each picture frame consisted of 700 lines made up of 832 picture elements per line. The received signal was recorded on magnetic tape and simultaneously displayed on a cathode ray tube. In the initial presentation (*top*) the picture looks rather gray and featureless because the actual contrast on the surface of Mars is quite low. The first stage of computer processing yields the shading-corrected version (*middle*), which indicates the relative brightness in the scene along with some enhancement. The right-hand portion of the picture is dark because it is beyond the terminator and hidden in shadow. A second level of processing (*bottom*) suppresses large-scale differences in brightness in order to enhance preferentially the topographic detail in both the light and the dark areas. The picture is now seen to reveal a complex of transecting ridges two to five kilometers apart that the Mariner television group nicknamed Inca City. The ridges seem to be composed of a resistant material that has filled cross-cutting fractures and that later has been exposed by erosion of the surrounding material. *Mariner 9* took the picture at a distance of 2,937 kilometers.

ternal processes, to study transient phenomena on the surface and in the atmosphere and to provide reconnaissance over a long enough period (from nine months to a year) to observe seasonal changes in surface markings in the hope of clarifying their origins. When *Mariner 8* was lost during launching, the complementary missions of the two spacecraft had to be combined.

When *Mariner 9* reached Mars on November 13, 1971, the greatest dust storm in more than a century was raging on the planet, almost totally obscuring its surface. The first views from a distance of several hundred thousand miles revealed essentially no detail except a glimpse of the south polar cap [*see bottom illustration on page 50*]. The dust storm delighted the investigators who wanted to study the planet's atmosphere, since it promised to reveal how particles were transported by such a thin medium, but it was a disappointment to the investigators concerned with surface features. For example, there had been plans to take a sequence of far-encounter pictures, ultimately to be printed in color, showing the planet getting larger and larger, thereby providing a visual bridge between the level of detail seen through telescopes from the earth and the detail eventually visible in pictures taken from orbit around Mars. A limited effort to produce far-encounter pictures showing Mars in natural color had been made with images taken by *Mariner 7* through separate red, green and blue filters.

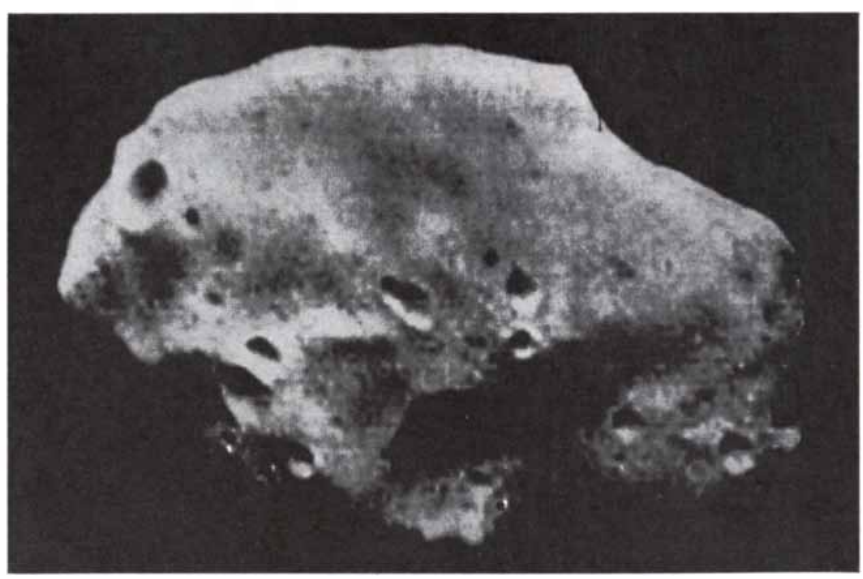


CAPTURE OF *MARINER 9* BY MARS took place on November 13, 1971, after the spacecraft's retro-rocket had generated a decelerating thrust of about 300 pounds for 15 minutes 15.6 seconds. The 287-million-mile voyage from the earth had taken 167 days. *Mariner 9* approached Mars from below and went into an orbit inclined at an angle of about 64 degrees to the planet's equator. The initial orbit ranged from 1,385 kilometers to 17,300 above the planet's surface. A subsequent correction reduced the orbit's high point to 17,100 kilometers and raised the low point to 1,650, achieving the desired orbital period of 11.98 hours.

The Great Volcanoes

The dust storm delayed the systematic mapping of the Martian surface for nearly three months. Even during the storm, however, four dark spots in the equatorial area were repeatedly seen in the early pictures taken from orbit. The spots clearly represented permanent surface features high enough to stick up through the dust. Presumably they looked dark simply because their surface reflectivity was lower than that of the bright, dusty atmosphere.

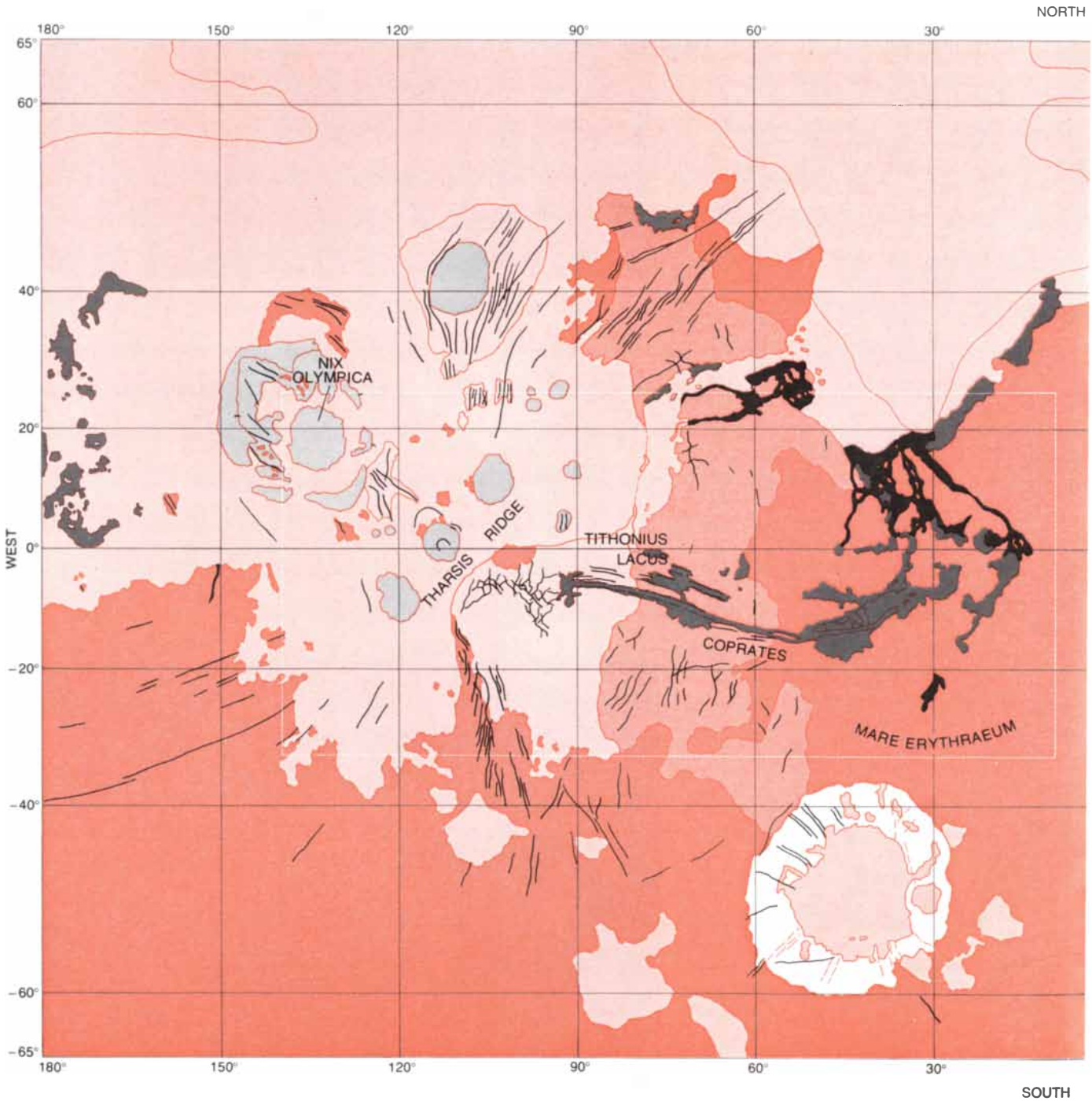
One of the four spots corresponded to the location of Nix Olympica ("Snows of Olympus"), so named because it was normally visible from the earth as a bright feature and also a variable one. When this dark spot was observed with the high-resolution, or narrow-angle, camera on *Mariner 9*, the image that emerged was breathtaking. What one saw was the characteristic pattern of coalesced craters that constitute a volcanic caldera. Such calderas are not uncommon on the earth, for example in the Hawaiian Islands. The Martian caldera, however, was 30 times larger in diameter than any in the Hawaiian chain. When the dust had settled, Nix Olympica was seen in full to be an enormous volcanic mountain more than 500 kilometers in diameter at the base, much larger than any similar feature on the earth; the caldera was only the summit [*see illustration on page 58*]. Atmospheric-pressure maps made later with the aid of the ultraviolet spectrometer and other techniques show that Nix Olympica is at least 15 kilometers high



HEAVILY CRATERED MARTIAN MOON, Phobos, was photographed by *Mariner 9* from a distance of 5,500 kilometers. The inner of the two Martian satellites, Phobos is about 25 kilometers long and 20 kilometers wide. Deimos, the outer satellite, is about half the size of Phobos. The picture of Phobos has been greatly improved by computer processing.

and possibly 30. For purposes of comparison, Mauna Loa, the tallest volcanic cone in the Hawaiian Islands, rises less than 10 kilometers from the floor of the Pacific. High-resolution photography revealed that the other three dark spots were also volcanoes, somewhat smaller than Nix Olympica, strung together to form a long volcanic ridge. Following the traditional name for that area, it is now called Tharsis Ridge [*see illustration on these two pages*].

The first recognizable features photographed by *Mariner* 9 presented a fascinating question: How can one explain why one entire hemisphere of the planet, the hemisphere observed by three earlier Mariners, shows scarcely any evidence of internal activity, whereas the first area to be investigated in detail on the opposite side of the planet has four enormous volcanoes? The explanation apparently is that Mars is just beginning to "boil" inside and produce surface igneous activity. Presumably this process is now well advanced in the Nix Olympica–Tharsis Ridge area but has

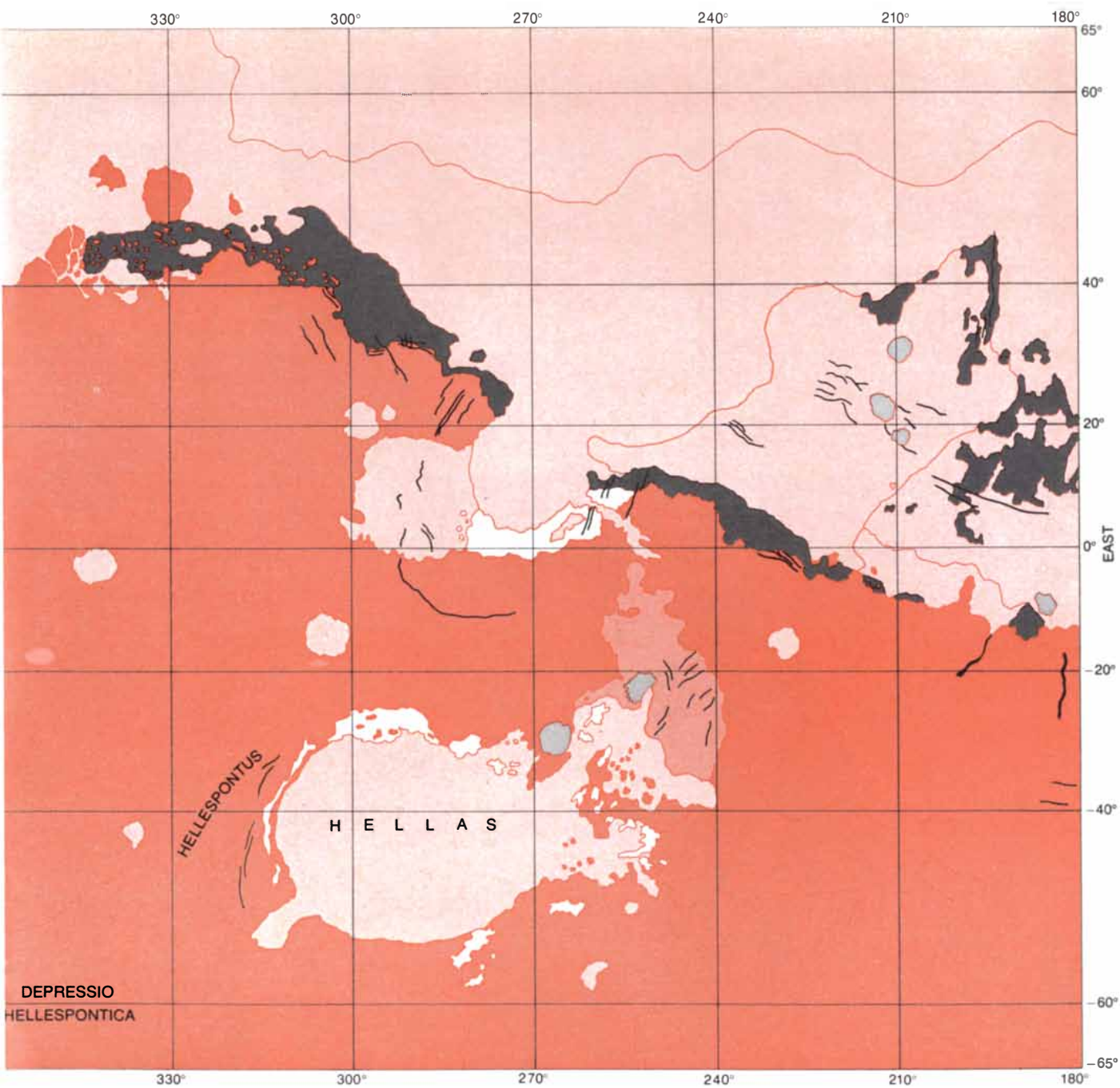


GEOLOGICAL MAP OF MARS is an effort to classify the features that compose the surface terrain. The white rectangle encloses the region depicted in the mosaic on the next two pages. The areas in pale color are smooth plains; areas in medium color are cratered plains; dark-colored areas are old cratered terrain. White areas are mountainous terrain. Areas in light gray are volcanic, embracing Nix Olympica and the three volcanoes that form Tharsis Ridge. Areas in dark gray identify terrain that has been modified in some fashion. Areas in black are channel deposits. Short hatch lines are inferred faults. A number of features referred to in

not yet spread to the planet as a whole. We may be witnessing on Mars a phase similar to one the earth probably went through early in its history, a phase whose record has been totally erased by subsequent igneous and sedimentary processes.

The rate at which a planet's interior heats up depends on a number of factors, chiefly the amount of radioactive material in its original accreted mass and the total mass, which determines the pressure in the interior and the degree of insulation. In very general terms, if Mars had the same original composition as the earth, one would expect it to heat up more slowly because it has only a tenth the mass of the earth. The sheer size of Nix Olympica suggests the possibility that deep convection currents are churning, a process that conceivably could lead some hundreds of millions of years hence to the kind of plate-tectonic phenomena responsible for the slow drift of continents on the earth.

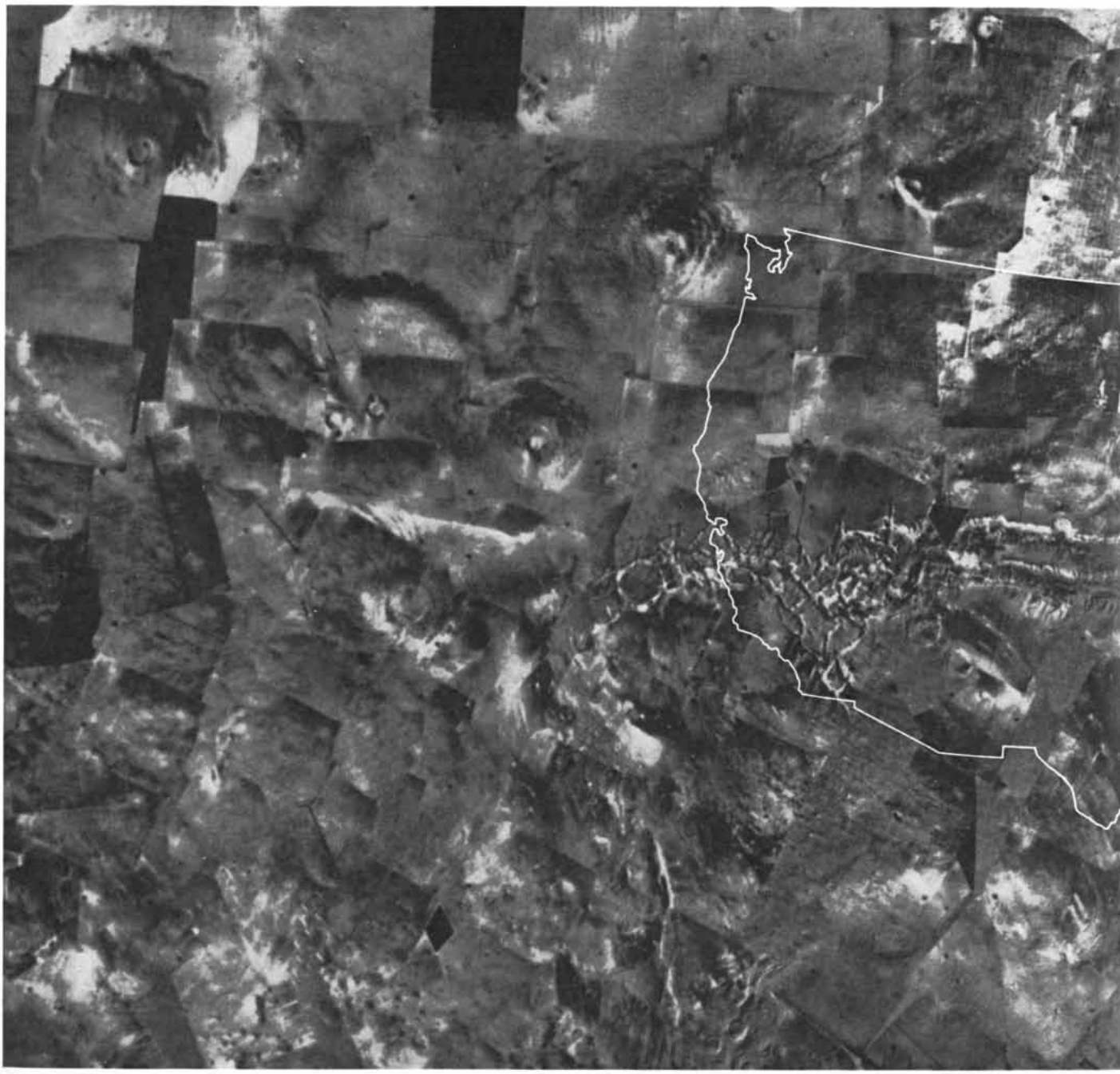
Immediately to the east of the volcanic province is a highly fractured area,



the text or depicted in photographs are labeled. A portion of Depressio Hellespontica is shown in the color photograph on page 48. A canyon in Coprates appears on page 51. A canyon in Tithonius Lacus appears at the top of page 59 and a crater in Hellespontus at the bottom of the same page. A sinuous valley in Mare Erythraeum is shown in the photograph on page 61. Hellas is a nearly featureless bowl more than 1,600 kilometers across. The map was prepared by Michael Carr, John F. McCauley, Daniel Milton and Don Wilhelms of the U.S. Geological Survey in cooperation with several members of television-experiment team of *Mariner 9* project.

and beyond that another extraordinary topographical feature was discovered: a series of huge canyons stretching east and west along the equator [*see illustration on these two pages*]. These canyons, 80 to 120 kilometers wide and five to six and a half kilometers deep, are much larger than any found on the earth. Again we must assume that their origin is due to fairly recent internal activity. Presumably large-scale east-west faulting has exposed underlying layers of the planet whose composition could conceivably trigger an erosion process of some kind.

One speculation is that deep permafrost is involved, associated perhaps with the arrival near the surface of juvenile water preceding and accompanying the rise of molten rock near the surface of the planet during the volcanic episode apparent to the west. Mars is everywhere below freezing just a short distance below the surface. Once permafrost was exposed to the atmosphere, its water content would sublimate, making available a loose, friable material sufficiently mobile to serve as an eroding agent in a mass-transport process. We must then ask where the material went. One possibility is that the winds of Mars have transported it as dust to other localities. (Although the Martian atmosphere is thin, its winds may blow at several hundred miles per hour.) Alternatively, the missing material may yet be discovered somewhere to the east of the canyons. Still a third possibility is that it may even have disappeared into the planet's



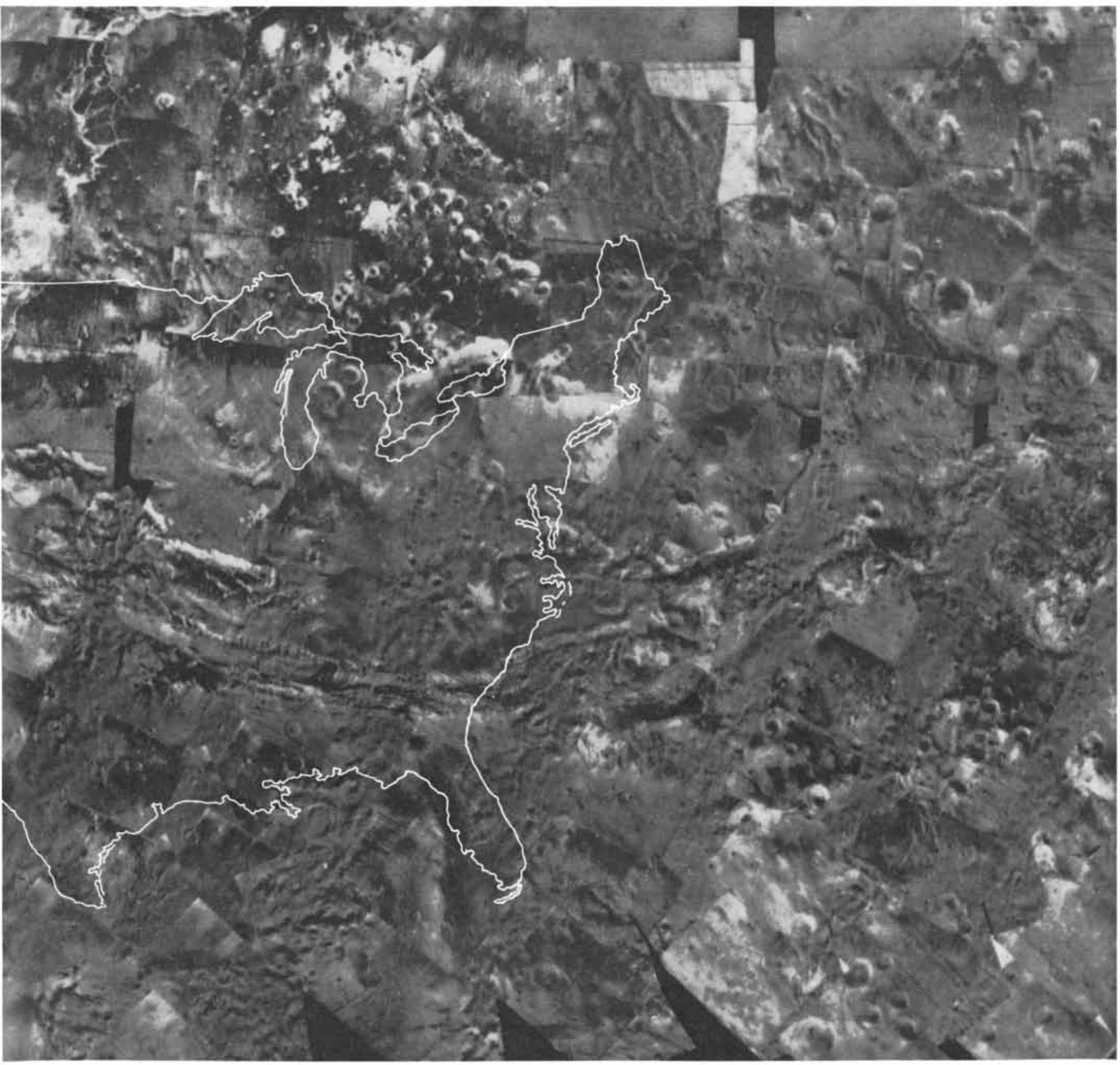
MOSAIC PANORAMA was made from several dozen *Mariner 9* wide-angle photographs specially computer-processed and matched by Raymond Batson of the U.S. Geological Survey at Flagstaff, Ariz. The mosaic depicts a region extending some 7,000 kilometers along the Martian equator. The area covered by the panorama is indicated on the geological map of Mars on the preceding two pages. The superposed outline map of the U.S. shows vividly the dimensions of the great canyons that parallel the equator for 4,000 kilo-

interior in a complex exchange process.

The largest of the canyons corresponds to a feature long known as Coprates, whose appearance sometimes changes with the seasons. By observing this canyon as the dust storm ended we were able to gain insight into its variable appearance. The canyon is so deep that considerable dust persists in the atmosphere between the canyon walls after the atmosphere above the surrounding region is comparatively dust-free [*see illustration on page 51*]. The dust-filled atmosphere makes the canyon look brighter than the surrounding landscape. Once the canyon atmosphere clears up, there is little contrast between the interior of the canyon and the surrounding area. Hence the variable "surface" markings associated with Coprates probably have nothing to do with the surface at all.

Similar atmospheric processes may well explain some of the other variable markings formerly attributed to seasonal changes on the surface. Other kinds of variation are not so simply explained, but evidently they always involve the interaction of dust, topography and atmosphere.

Like the volcanoes, the great canyons of Mars suggest a fairly recent episode in the history of the planet characterized by large-scale events. On the earth one often finds a reasonably steady state between processes of erosion and processes of restoration; thus one sees a range of surface morphologies from youthful to mature. In the case of the Martian can-

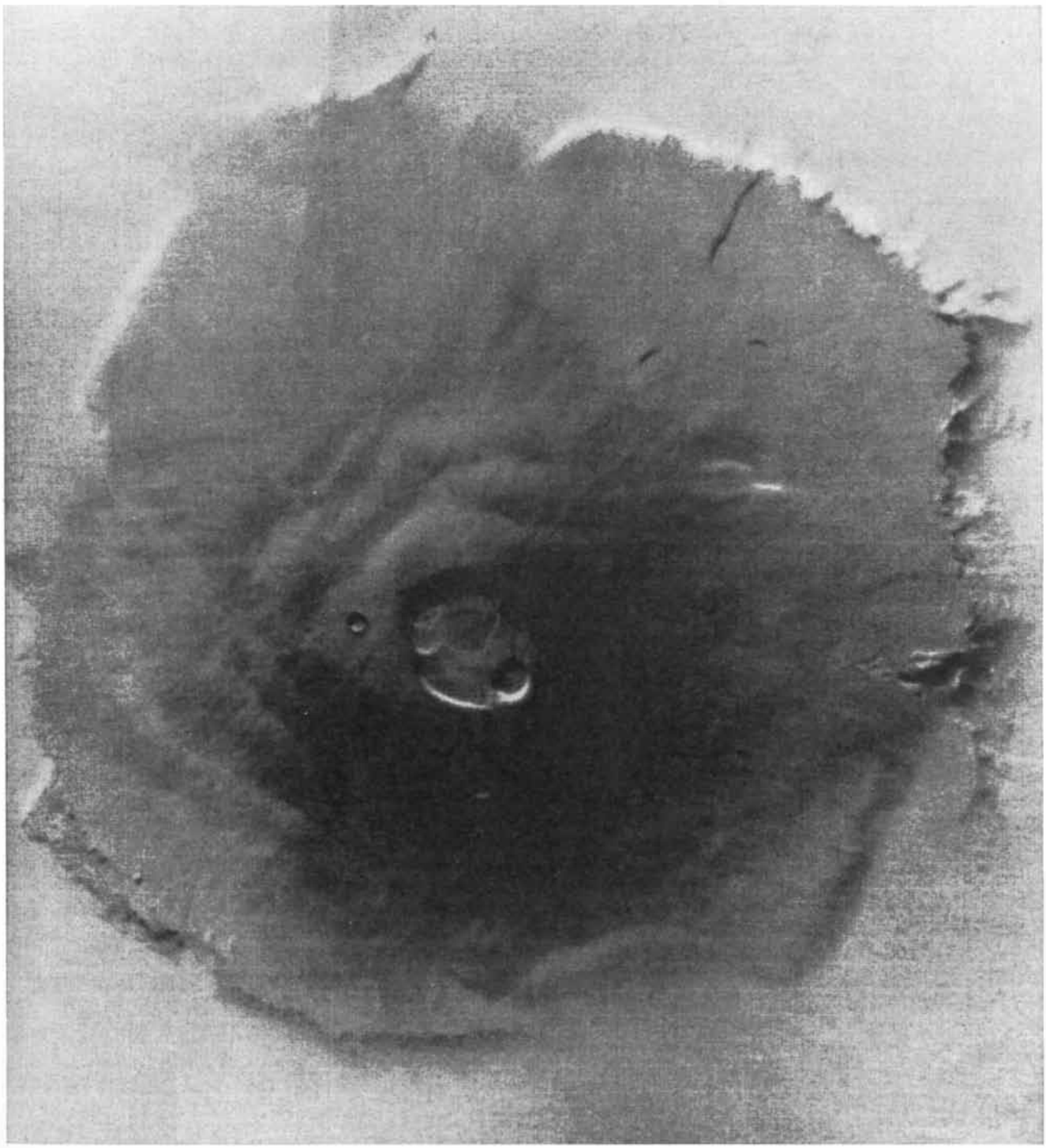


meters. Much larger than anything like them on the earth, the canyons average 100 kilometers in width and reach a depth of more than six kilometers. At the extreme left of the mosaic one can see the giant volcanic mountain Nix Olympica and immediately to the right the series of somewhat smaller volcanoes that form Tharsis Ridge. The sun was generally shining from the lower left when the pictures were taken, so that the shadows fall toward the upper right. For some, relief will be stronger if mosaic is turned upside down.

yons erosion does not seem to be balanced by a corresponding restoration; we do not see old degraded canyons with a mature form.

The Channels

The eastern extremity of the canyons joins a large area of chaotic terrain, a small portion of which was glimpsed by *Mariner 6*. The appearance of the chaotic terrain strongly suggests that it is the result of some kind of collapse and that the collapse is genetically related to the canyons to the west. Extending out from the chaotic terrain in a northwesterly direction are some extraordinary channels, which are also found in a number of other localities on the planet. It is hard to look at these channels without considering the possibility that they were cut by flowing water. Indeed, some of my colleagues think that is the only reasonable explanation.



CLOSE-UP OF NIX OLYMPICA is shown in a mosaic of *Mariner 9* photographs that have been specially computer-processed and matched. The picture is printed with north at the right, so that the sunlight seems to come from the top of the page; the feature is thus seen immediately as a cone and not as a depression. The volcanic mountain, rising from a great plain, is 500 kilometers across at the base, or much larger than similar volcanic mountains on the earth. Pressure mapping of the depth of the atmosphere indicates that Nix Olympica is 25 kilometers high. The main crater, a complex volcanic vent known as a caldera, is 65 kilometers in diameter.

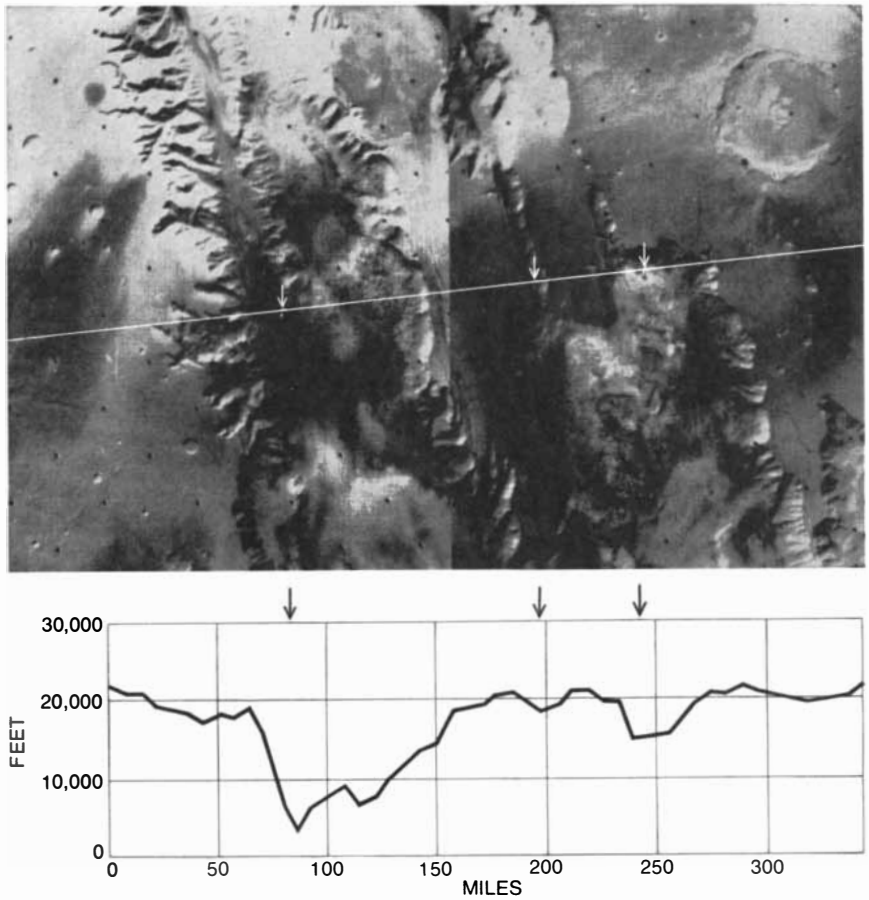
One can estimate the age of the channels by noting the size-to-frequency relation of impact craters on their floors. The channels are clearly younger than the crater-pocked terrain seen over much of the planet, yet they are by no means the youngest features of the Martian landscape.

The discovery of the channels has revived speculation that there may have been an earthlike epoch in the history of Mars. According to this view Mars may once have had a much denser atmosphere and water vapor in such abundance that rain could fall. Given rainfall, the channels could be easily explained. Less easily explained is why channels have survived in only a few areas and why older topography shows no evidence of water erosion. It would seem difficult to explain how the primitive Martian atmosphere, probably dry and reducing (in the chemical sense), could have evolved into a dense, wet one and then have been transformed again into the present thin, dry atmosphere consisting almost entirely of carbon dioxide. Moreover, the present atmosphere is strongly stabilized by the large amounts of solid carbon dioxide in the polar regions. If the channels were created by rainfall, it would seem that one must postulate two miracles in series: one to create the earthlike atmosphere for a relatively brief epoch and another to destroy it.

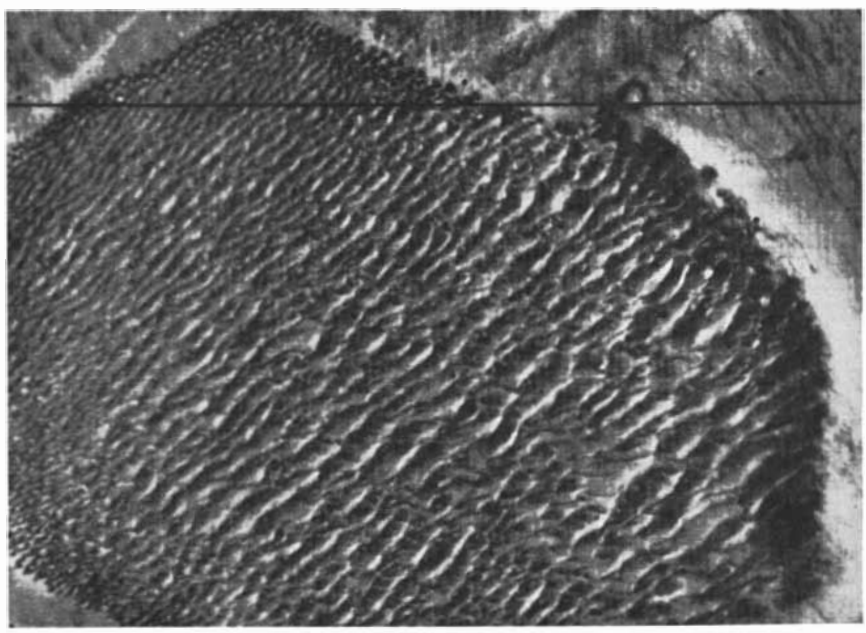
An alternative hypothesis presents at least as many difficulties. It is suggested that liquid water accumulated in underground reservoirs following entrapment and melting of permafrost. Hypothetically the reservoirs were abruptly breached, allowing the released water to create the channels. The observed channels are so large and deep, however, that a great volume of water must have been involved in their formation. Therefore it would seem even more difficult to ascribe the channels to a "one shot" open-cycle process than to a closed-cycle process such as rainfall.

The origin of the canyons and channels is one of the primary enigmas that has emerged from the *Mariner* 9 mission. Because of the importance of liquid water to life as we know it, the possible role of water in creating the canyons and channels has attracted particular interest.

Finally, there are a few areas on Mars to which the term "basins" seems appropriate. The most prominent is the large circular feature Hellas, more than 1,600 kilometers in diameter. Hellas has been observed from the earth for more than two centuries. Sometimes it rivals



CHASM IN TITHONIUS LACUS is in the extreme western part of the canyon system that runs east and west near Mars's equator. An ultraviolet spectrometer on *Mariner 9* measured the atmospheric pressure at the planet's surface along the track indicated by the white line. The pressure reading was then converted into the jagged depth profile that appears below the pair of photographs. The difference between the highest and lowest points is more than 6,000 meters, making the chasm four times as deep as the Grand Canyon in Arizona. Tithonius Lacus is immediately south of the Martian equator at about 85 degrees west longitude.



DUNE FIELD some 70 kilometers across was discovered by *Mariner 9*'s narrow-angle camera inside a crater 150 kilometers wide in the region known as Hellespontus. Ridges are one and a half kilometers apart. Dune fields as spectacular as this one appear rare on Mars.

the polar caps in brightness. *Mariner 7* demonstrated that Hellas is indeed a low-lying basin virtually devoid of features. It has been deduced from closeup photographs that the surface of Hellas has probably been smoothed by the influx of large amounts of dust carried into the basin by wind. *Mariner 9*, however, revealed that Hellas exhibited a few faint topographical features just as the planet-wide dust storm was ending. This suggests that variations in the brightness of Hellas may be due to frequent dust storms of a more local nature, a view originally adduced on meteorological grounds by Carl Sagan of Cornell University and his co-workers. Thus Hellas probably acts as a long-term collection basin for dust but may also serve as a source of dust when the Martian winds blow particularly hard. The observation by *Mariner 9* of small-scale dust storms and the recognition that they can alter the brightness of local areas give us further insight into some of the variable features that have been observed over the years from the earth.

One of the crowning achievements of *Mariner 7* was the study at high resolutions of the very large polar cap present during the southern winter of 1969. Measurements of reflectivity and temperature provided by an infrared spectrometer and an infrared radiometer on the spacecraft proved conclusively that the south polar cap was composed of very pure solid carbon dioxide, as had been predicted some years earlier. The photographs showed that the frost cover was thin (probably less than a few meters on the average) and that a variety of unusual surface features were also present in the vicinity of the south pole.

When *Mariner 9* reached Mars, it was late spring in the southern hemisphere, an ideal time for monitoring the wasting of the dry-ice cap and for examining in detail the unusual surface features that should have been further revealed. The disappearance of the south polar cap started out as expected but then clearly showed anomalous behavior. Curiously, the general outline of the shrunken cap persisted throughout the late summer, when the sublimation of the carbon dioxide should have been at a maximum [*see top illustration on page 63*]. This suggested to me that after the large annual cap of carbon dioxide has sublimated, it exposes a residual cap of ordinary water ice. Ordinary ice, of course, has a much lower evaporation rate than carbon dioxide, and traces of water vapor are present in the Martian atmosphere.

Mariner 9's pictures also disclosed a most peculiar terrain in the south polar area, which we named laminated terrain. Although its outline is not symmetrical, it covers much of the south polar region up to about 70 degrees south latitude. The laminated terrain is composed of very thin layers, alternately light and dark, whose gently sloping faces exhibit a certain amount of texture, or relief [*see bottom illustration on page 63*].

The thin laminas appear to be collected in units of 20 or 30 or more to constitute plates perhaps half a kilometer or more in thickness and up to 200 kilometers across. The plates have outward-facing slopes in which a banded structure can be seen. The laminar deposits have been found only in the polar regions, where carbon dioxide forms an annual deposit of frost. This suggests

BRAIDED CHANNEL adjacent to an impact crater 20 kilometers in diameter is an example of the type of feature that suggests there has been fluid erosion of some kind on Mars. If the agent was actually water, it is hard to understand why eroded terrain is confined to only a tiny fraction of the planet's total surface. The region shown is at six degrees south latitude 150 degrees west longitude.

that the laminations are associated in some way with the coming and going of volatile substances and that they may even retain some solid carbon dioxide or water ice. Since the laminations are marked by very few impact craters, one can deduce that they are a recent development in the history of Mars.

The North Pole

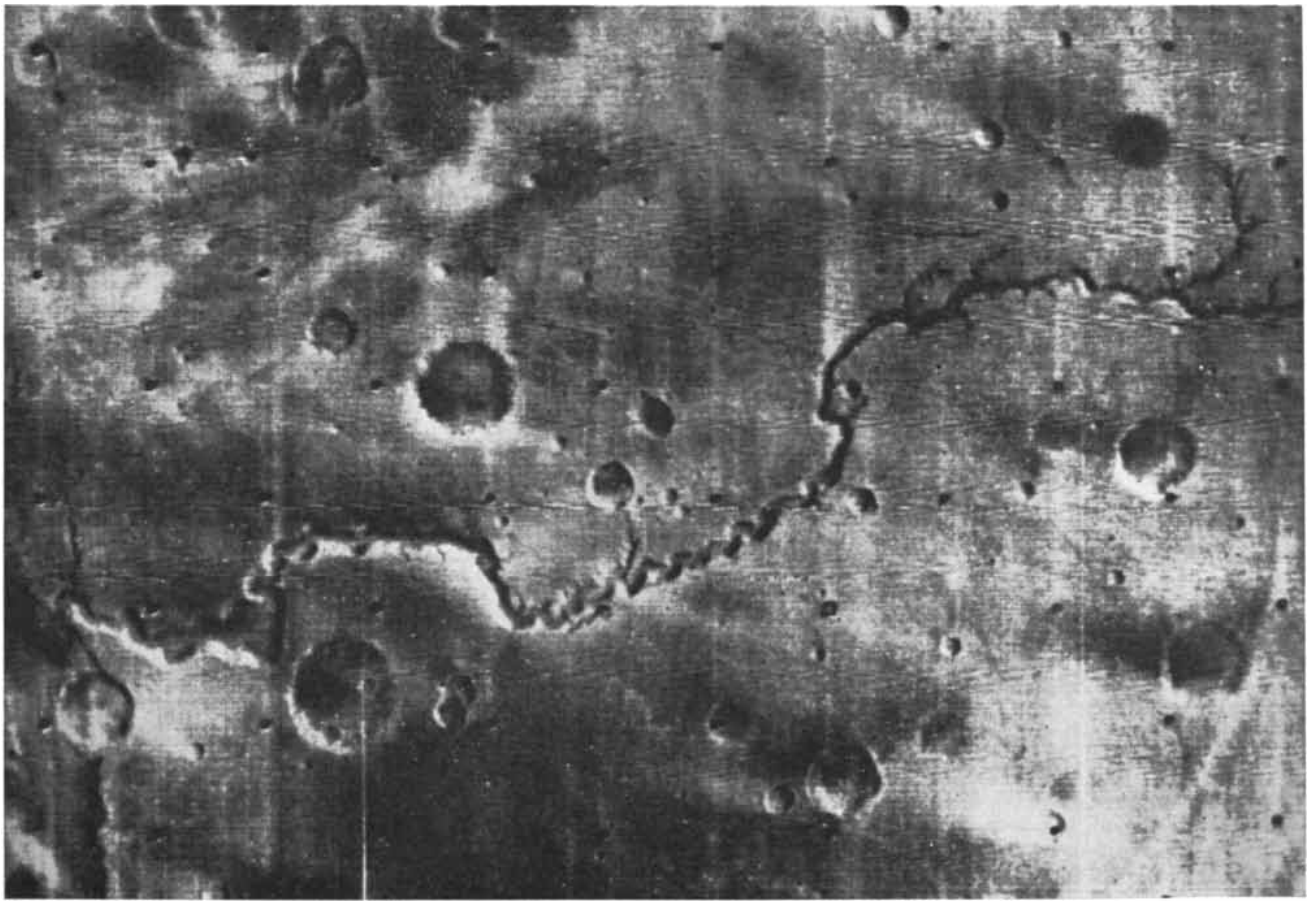
The north polar region of Mars finally became available for observation by *Mariner 9* rather late in the mission as a result of gradual changes in lighting, associated with the change in season and with the lifting of the haze that characteristically develops in the fall over each pole. The quasicircular structures characteristic of laminated terrain were found to be even more abundant around the north pole than around the south pole. One can see 20 or 30 individual plates arranged in a pattern reminiscent of fallen stacks of poker chips. The existence of the laminated terrain and circular-plate structures in the north polar regions as well as in the south polar ones indicates beyond any reasonable doubt that their formation must be associated in some way with the periodic deposition and evaporation of volatile material.

Michael C. Malin, a graduate student at the California Institute of Technology, and I have speculated that the distribution of the circular plates and their overlapping arrangement can be explained by changes in the tilt of Mars's axis. We posit that the rotational axis of the planet has been displaced over the past tens of millions of years as a result of convection currents deep in the mantle, currents that are probably associated with the production of volcanoes in the equatorial areas. As the spin axis has shifted, the laminated plates have formed concentrically around each successive position of the poles.

This speculation is at least consistent with information about Mars's gravity distribution deduced from changes in *Mariner 9*'s orbit. The planet exhibits gravitational anomalies suggestive of deep density differences of the kind that could be associated with deep convection. Moreover, there is a strong correlation between the gravitational anomalies and the location of the equatorial volcanoes.

The regular appearance of the laminas and of the plates themselves suggests that they are also associated in some way with periodic alternations of the climate of Mars. In collaboration with two other graduate students, William Ward and Sze Yeung, I have investigated the theoretical variations in the orbit of Mars over a period of time. We find that perturbations in the orbit caused by other planets, analyzed a number of years ago by Dirk Brouwer and G. M. Clemence, alter the orbit's eccentricity in a way that turns out to be quite favorable for our hypothesis. The eccentricity of Mars's orbit varies from .004, or nearly circular, to .141. Its present value is .09 [*see top illustration on page 68*].

The consequence of this variation in eccentricity is a variation in the yearly average amount of sunlight reaching the poles of the planet, together with a much stronger variation in the maximum solar flux when the planet is closest to the sun [*see bottom illustration on page 68*]. Although the variation in average radiant



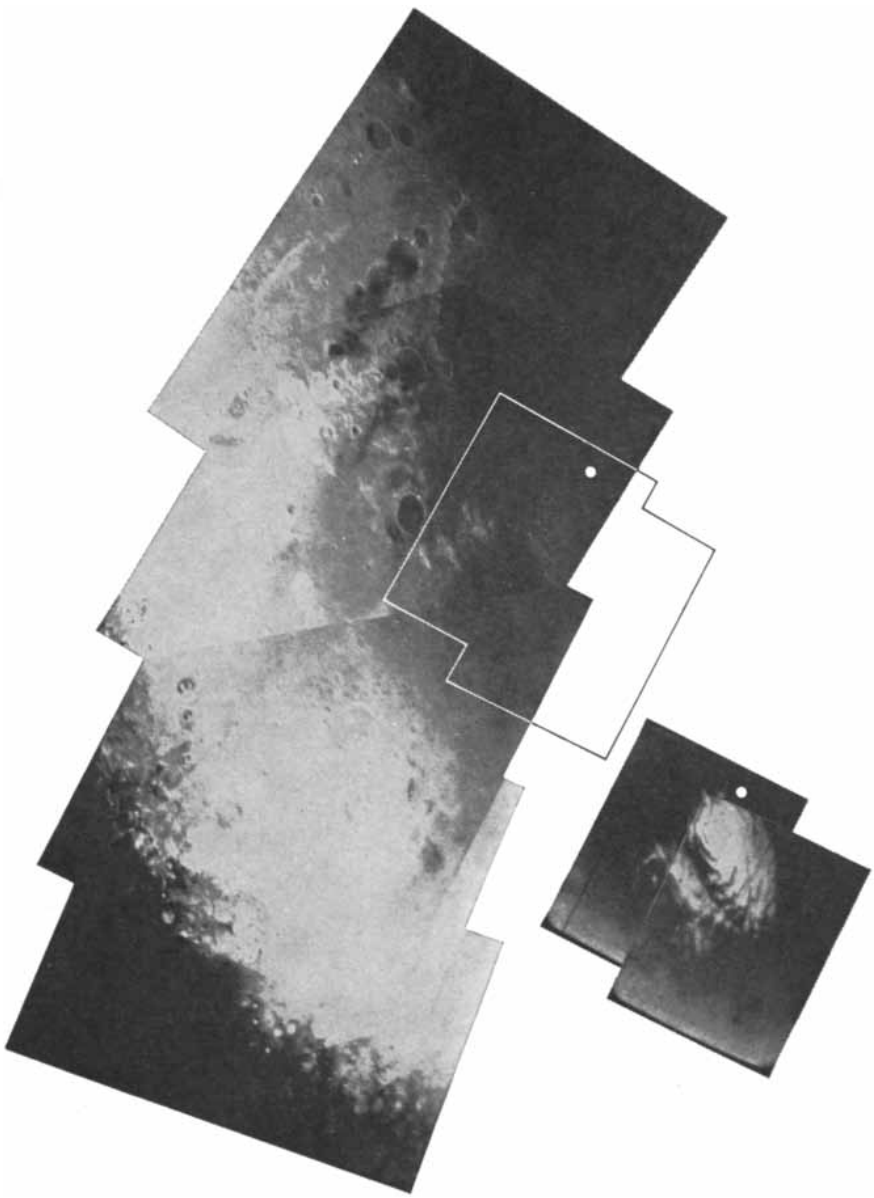
SINUOUS VALLEY 400 kilometers long and up to six kilometers wide is located at 29 degrees south latitude 40 degrees west longitude in the region known as Mare Erythraeum. The feature resembles the outline of a meandering river or one of the lunar rilles, which may have been created by flowing lava. The origin of such features on Mars has not been satisfactorily explained.

input at the poles is only a few percent, it is sufficient under some circumstances to cause a cyclical variation in the growth and sublimation of permanent carbon dioxide frost caps. Assuming that dust storms regularly deposit dust during the sublimation phases, thin laminas of the type observed could be produced. The plates themselves would therefore correspond to a periodicity of the order of two million years. Hence the laminated terrain seems to closely reflect the fluctuations in average radiant flux reaching the planet both in the short run (roughly 90,000 years) and in the long run (two million years). Inasmuch as a total of 20 or 30 plates are visible in the northern hemisphere the laminated terrain constitutes a record reaching back something like 100 million years. An alternative view regards the origin of the laminated terrain as being primarily erosional rather than constructional.

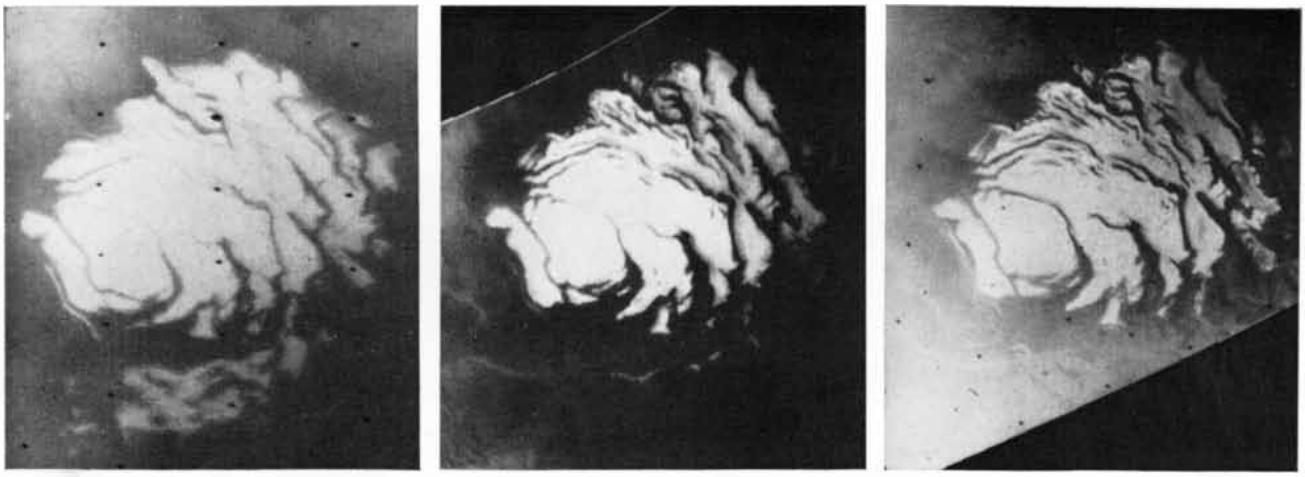
Evolution of the Atmosphere

If all the polar laminations accumulated in only a few hundred million years at most, representing no more than the past 5 percent of the history of Mars, what happened earlier? One encounters a basic difficulty in understanding Mars if one tries to apply the famous dictum of the 18th-century geologist James Hutton: "The present is the key to the past." Whether one looks at the volcanic terrain, the canyon lands, the channels or the polar laminations, all seem to record a remarkable degree of activity and change during the most recent part of Mars's geological past. I was led by these considerations to wonder if it is possible that the atmosphere of Mars as we know it may be a fairly recent acquisition. Malin and I are presenting this view as a "contentious speculation." It may be that Mars had no atmosphere at all, or only a very thin, unimportant atmosphere throughout the middle period of its history, lasting perhaps several billion years. Presumably there was an initial primitive atmosphere associated with the accretion of the planet, but this atmosphere may have been lost quite early, particularly if it consisted chiefly of hydrogen and methane.

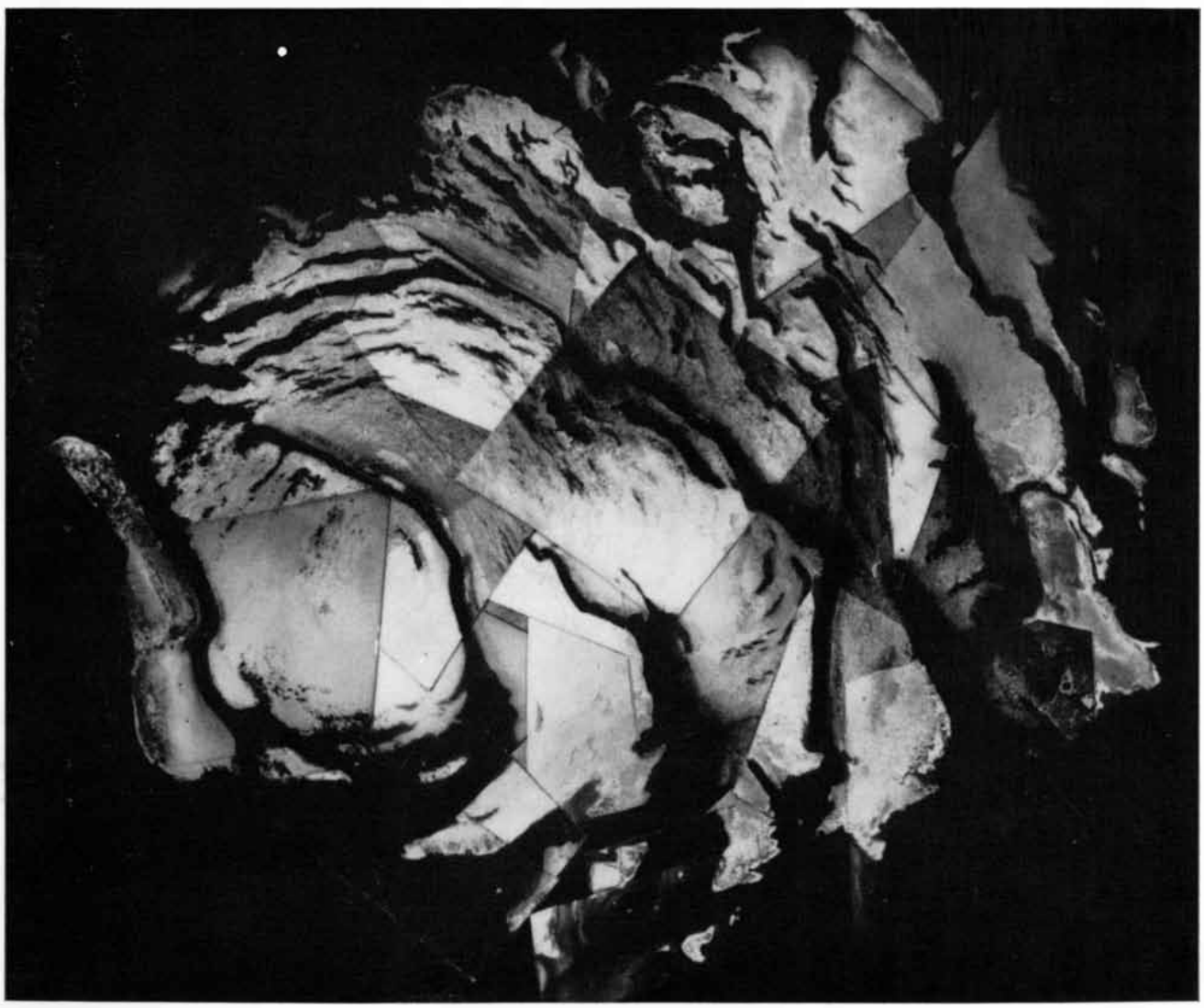
We think a significant fraction of the mass of the present Martian atmosphere was released during the formation of Nix Olympica and the other three volcanoes in the Tharsis Ridge area. The existence of widespread blankets of rock material and other evidence of somewhat earlier and more extensive volcanism and sedimentation suggest that rather large volumes of volatile substances have issued from the planet's interior in the later geological episodes. Thus it may be that as Mars matured enough to boil it simultaneously began to produce an enduring atmosphere. The atmosphere in turn has produced the laminated terrain and provided the wind-transport and erosion mechanisms to form the channels and the great canyons. On this hypothesis Mars is still far from reaching a steady state in which erosion would be balanced by modification, with a resulting development of a spectrum of morphological features. As part of our contentious speculation one might even imagine that in the early stages of the development of the present Martian atmosphere, before the polar cold traps for carbon dioxide were well established, enough water might have been brought to the surface to have flowed down the channels under peculiar, nonrecurring conditions. At least this possibility avoids the problem of positing two miracles in series and lets us settle for just one if ultimately liquid



SOUTH POLE IN DIFFERENT SEASONS can be compared in these views taken by *Mariner 7* in August, 1969 (*mosaic at left*), and by *Mariner 9* in November, 1971 (*right*). The area covered by the two *Mariner 9* pictures is indicated by the shape superposed on the *Mariner 7* mosaic. The geometric south pole is shown by a white dot in both views. In August, 1969, it was winter in the southern hemisphere of Mars and the frost cap of dry ice (solid carbon dioxide) was close to its maximum size. The right portion of the mosaic is dark because it is in shadow. When the *Mariner 9* pictures were taken 27 months later, the frost cap was shrinking rapidly with the approach of spring in Mars's southern hemisphere.



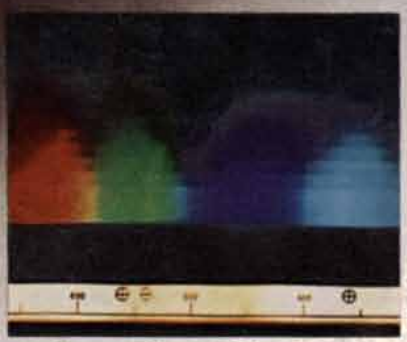
SHRINKING OF SOUTH POLAR CAP is depicted in these three views taken by *Mariner 9*. The original images have been enhanced, stereographically rectified and printed at the same scale. Viewed in the usual order, the pictures show the appearance of the frost cap after the spacecraft had been in orbit 14, 36 and 94 days. The first picture is still somewhat hazy because of the planet-wide dust storm. It can be seen, nevertheless, that the outline of the surface frost changed significantly in the three weeks between the first two views but then changed surprisingly little in the next eight weeks, when sublimation of the frost cap should have been at a maximum.



MOSAIC OF RESIDUAL SOUTH POLAR CAP was assembled (by Laurence A. Soderblom of the U.S. Geological Survey Center of Astrogeology in Flagstaff, Ariz.) from closeup pictures taken between the 58th and 94th day of orbital flight. The residual cap, which measures about 300 by 350 kilometers, is centered about 200 kilometers from true south pole (*white dot*) at about 45 degrees west longitude. Shades of gray have been optimized to bring out detail in frost-covered areas, hence frost-free surface appears black.

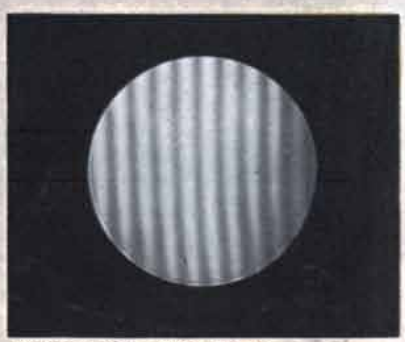
Polaroid®

How one great idea leads to another.

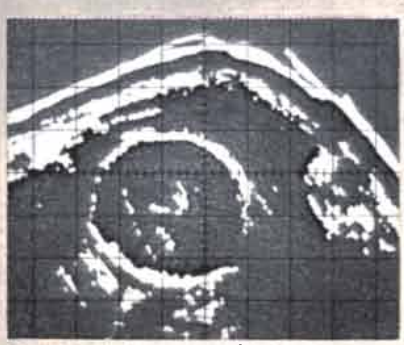


Spectrogram of xenon flash tube emission.

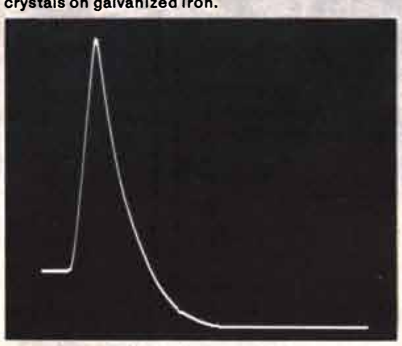
Scanning electron micrograph of zinc crystals on galvanized iron.



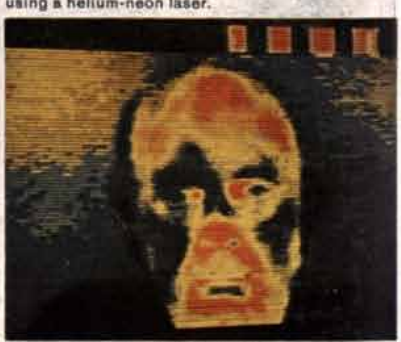
Interferogram of a mirror surface, using a helium-neon laser.



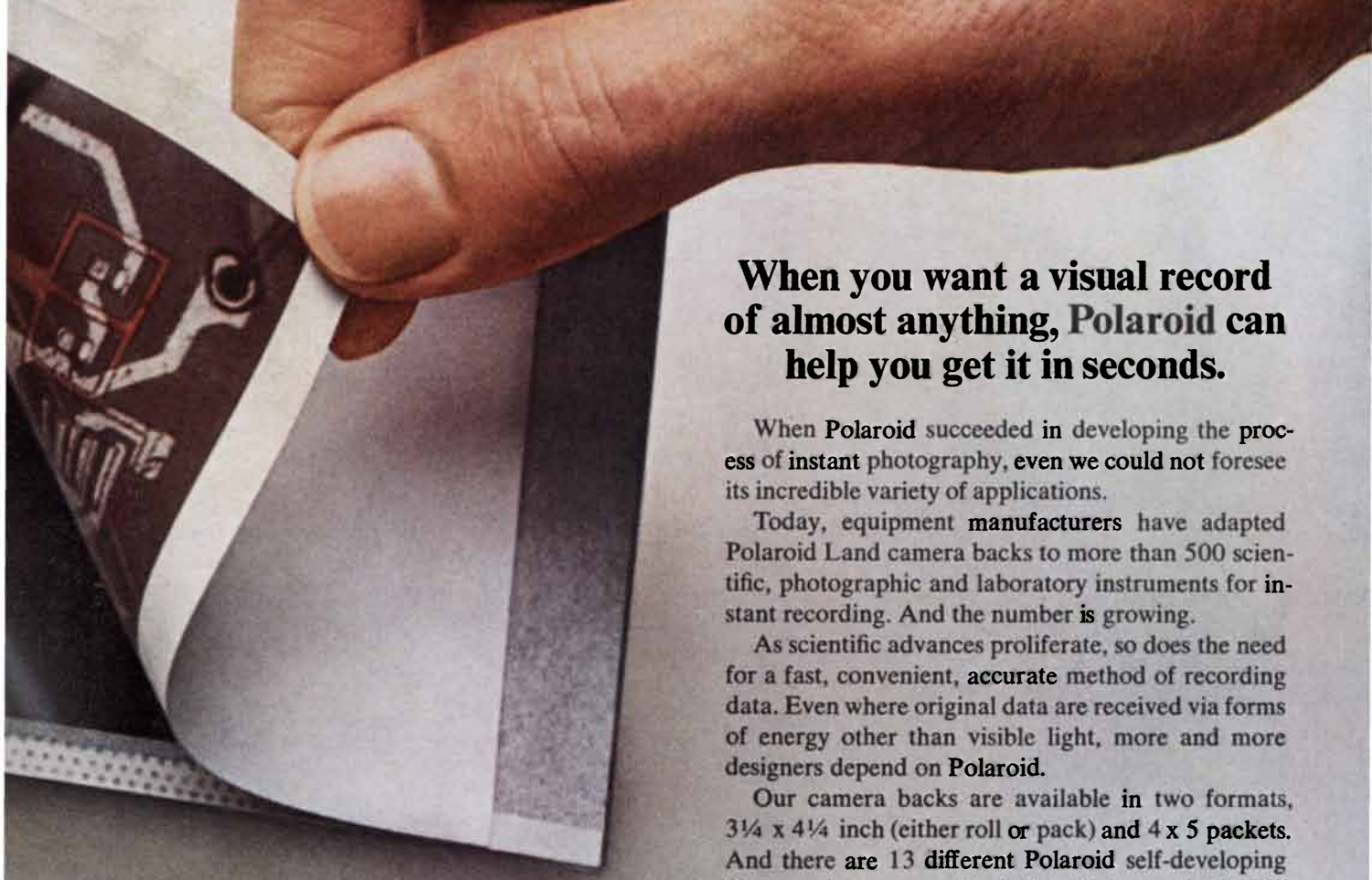
Recording of CRT display of fetal skull in utero made with reflected ultrasound waves.



Pulmonary function trace made with spirometer using fibre optics.



Thermograph showing blockage in carotid artery.



When you want a visual record of almost anything, Polaroid can help you get it in seconds.

When Polaroid succeeded in developing the process of instant photography, even we could not foresee its incredible variety of applications.

Today, equipment manufacturers have adapted Polaroid Land camera backs to more than 500 scientific, photographic and laboratory instruments for instant recording. And the number is growing.

As scientific advances proliferate, so does the need for a fast, convenient, accurate method of recording data. Even where original data are received via forms of energy other than visible light, more and more designers depend on Polaroid.

Our camera backs are available in two formats, 3¼ x 4¼ inch (either roll or pack) and 4 x 5 packets. And there are 13 different Polaroid self-developing films to choose from. Including high speed (10,000 ASA), high contrast, line and continuous tone transparency films, color, and one that produces a fully developed black-and-white negative with matching positive in seconds.

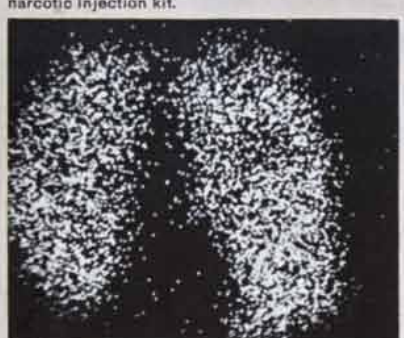
But Polaroid does more for original equipment manufacturers than sell camera backs and film. We have a wide range of experience, and are more than willing to share it with you. We can offer advice on making the right selections for your particular needs. And on how to adapt to your equipment in the most practicable way.

Perhaps you're working on a new development today. Or you might come up with one tomorrow. If making any kind of visual record is part of the plan, get in touch with us. Write to Polaroid Corporation, Dept. 57-235, 549 Technology Square, Cambridge, Massachusetts 02139.

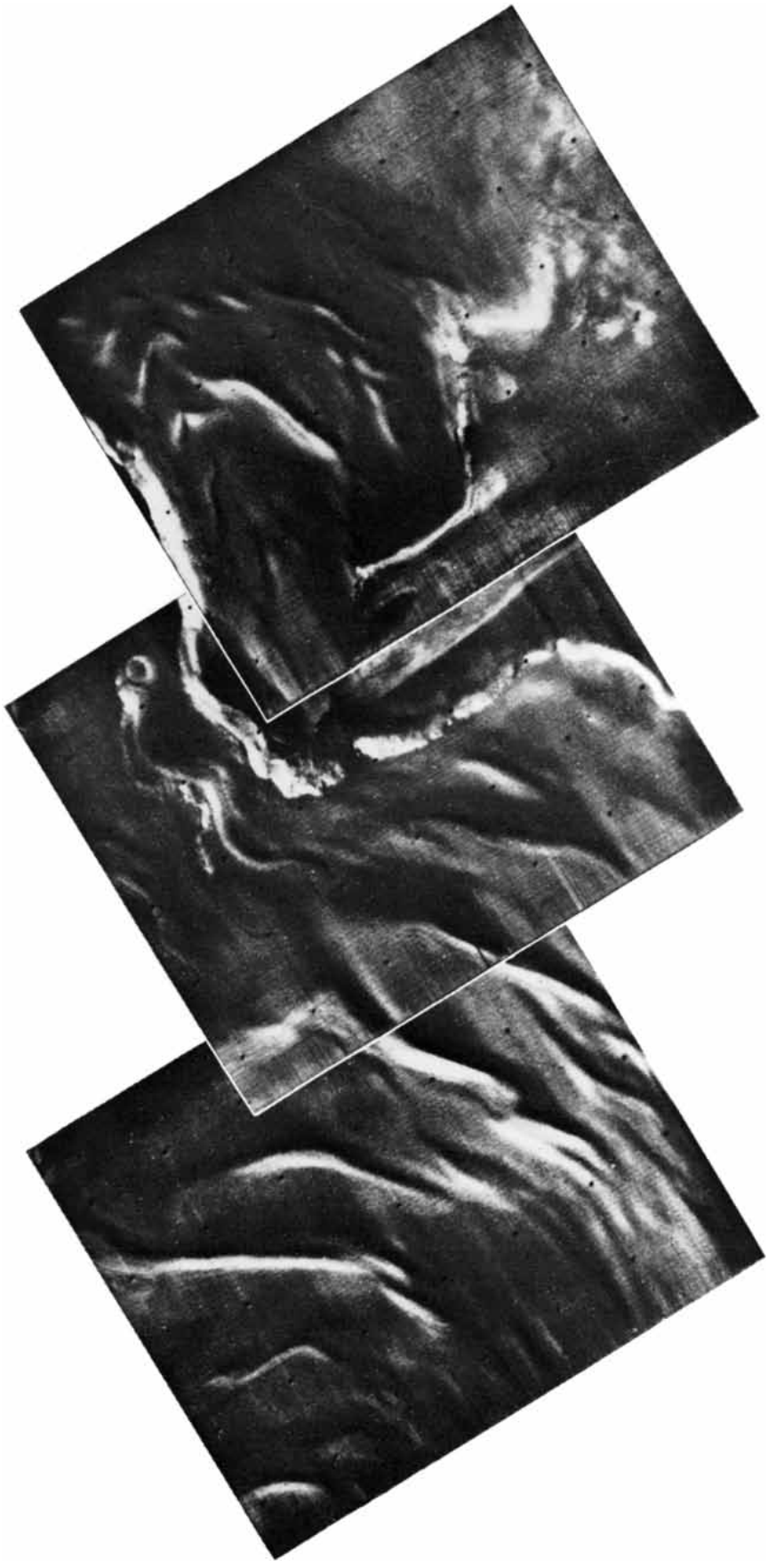
Our ideas may give you an idea.

Polaroid Land Camera Backs

Radiograph of prisoner's shoe revealing narcotic injection kit.



Lung scan made with scintillation camera using radioisotopes.



LAMINATED TERRAIN NEAR NORTH POLE resembles that in the south polar region. The author speculates that the distribution of circular plates and their overlapping arrangement may be evidence that the tilt of the axis of Mars has shifted over the past 100 million years. Such a shift could have resulted from convection currents deep in mantle of planet.

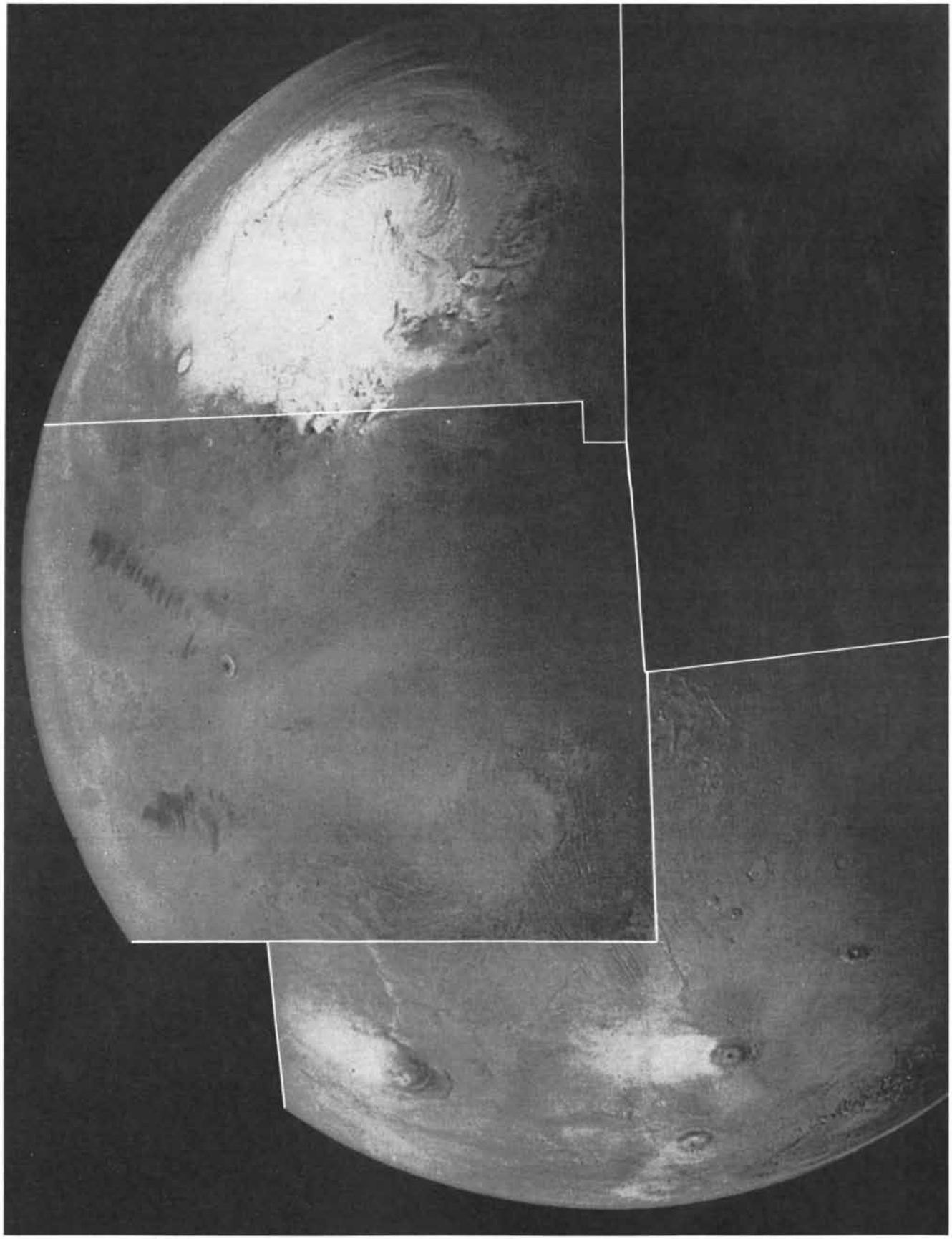
water is really required to explain the genesis of the channels.

The young-atmosphere hypothesis may also help to explain why "permanent dark areas" (for example the two-pronged feature Sinus Meridiani) should survive in the face of the frequent planet-wide dust storms. Again Mars somehow does not seem to us to be in a steady state, although others do not share our viewpoint. According to our hypothesis the dark markings may be the site of older surface materials not yet affected by the chemical weathering associated with the new atmosphere. In fact, there seems to be some correlation between the permanent dark markings and the terrain populated by the oldest craters.

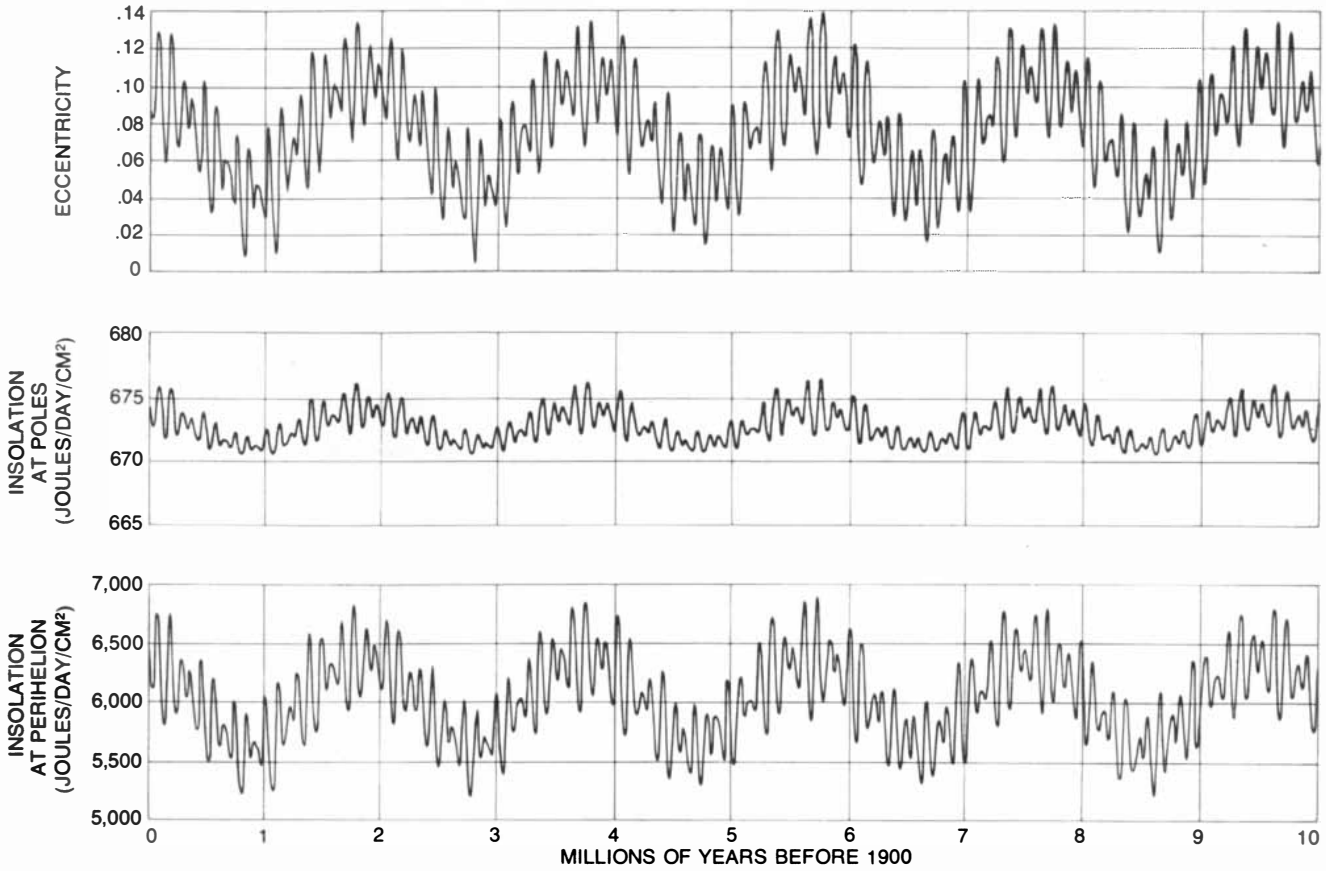
Other *Mariner 9* investigators, such as Sagan and W. K. Hartmann, have developed a quite different view of Mars's history. The nature of the old cratered terrains suggests to them that a long period of atmospheric erosion preceded the spectacular events of the more recent past. Thus the concept of an earthlike Mars is not by any means dead. Nonetheless, concepts of the geological history of Mars are changing rapidly in the light of *Mariner 9*'s highly successful mission. Perhaps ultimately some intermediate interpretation will fit the observations best.

Is There Life on Mars?

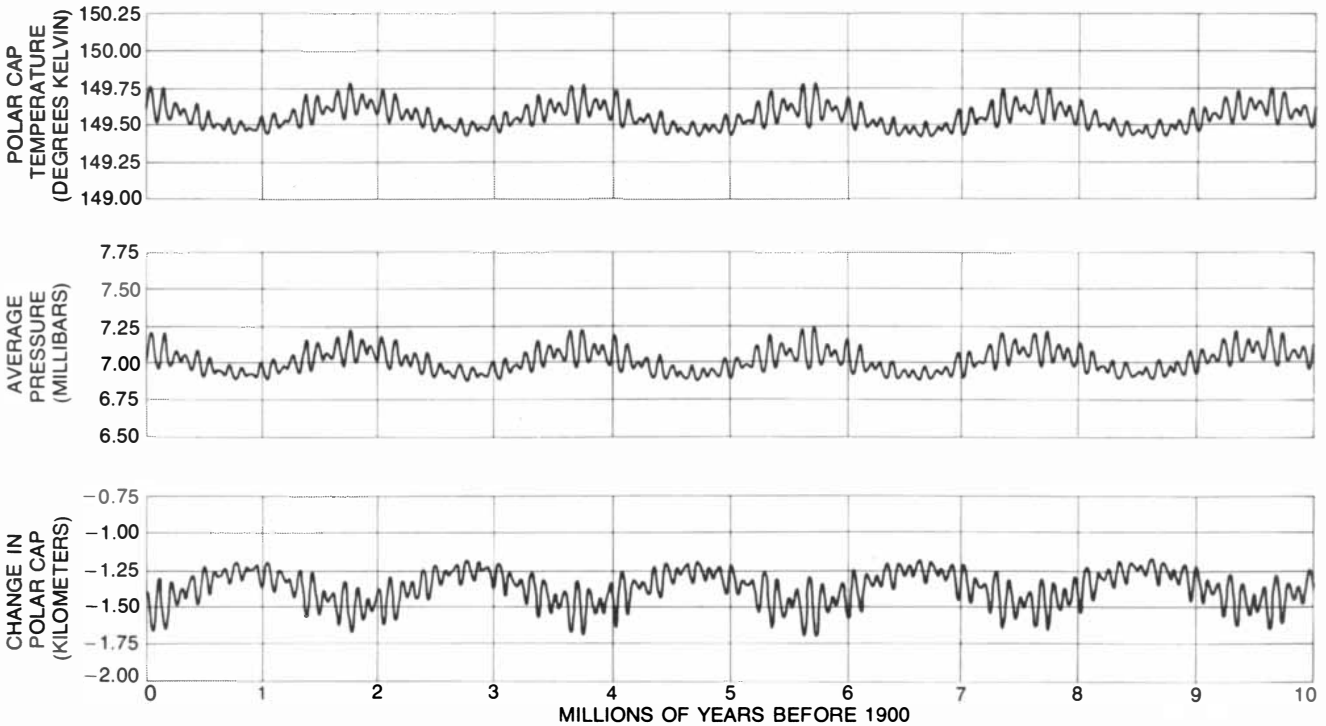
The present *Mariner 9* results suggest to me, however, a view very different from that of the early astronomers who thought that Mars was once earthlike and is now a dried-up fossil. I would argue that Mars is probably just now starting to become earthlike with the development of a durable atmosphere. "Just now" is hard to pin down quantitatively because the dates assigned to the craters on the basis of meteor flux rates are still highly uncertain. My guess now would be that the atmospheric "event," if it really happened, took place within the past quarter of Mars's history and certainly within the past half. If this contentious speculation should become widely accepted, it would necessarily imply pessimism about the possibility that past conditions were favorable for the appearance of simple forms of life on Mars. If Mars indeed was like the moon and lacked a significant atmosphere for much of its history, and if the maximum amount of water on the surface has been at most enough to create a few channels, it seems highly unlikely that there has ever been a sufficient accumulation of liquid water in the surface layers of Mars to allow the



NORTHERN HEMISPHERE OF MARS was photographed in three frames taken only 84 seconds apart by swiveling the wide-angle camera aboard *Mariner 9*. The bottom frame clearly shows Nix Olympica, the volcanoes of Tharsis Ridge and at the lower right the huge canyon that lies just below the equator. The pictures, taken on August 7, 1972, at an altitude of 13,700 kilometers, were among the last of the 7,273 produced by *Mariner 9*'s two cameras. Clouds of water ice or of carbon dioxide crystals obscured the planet north of the 50th parallel until the final weeks of the mission. The north polar cap is shrinking during the late Martian spring.



CHANGES IN ORBIT OF MARS over the past 10 million years may account in part for the peculiar circular plates and laminations observed in the planet's polar regions, according to a hypothesis developed by the author and his students. The eccentricity of Mars's orbit (*top curve*) has varied from .004, or nearly circular, to .141. This would lead to changes in the average amount of solar energy reaching the poles (*middle curve*) and in the peak energy reaching the planet when it was closest to the sun (*bottom curve*).



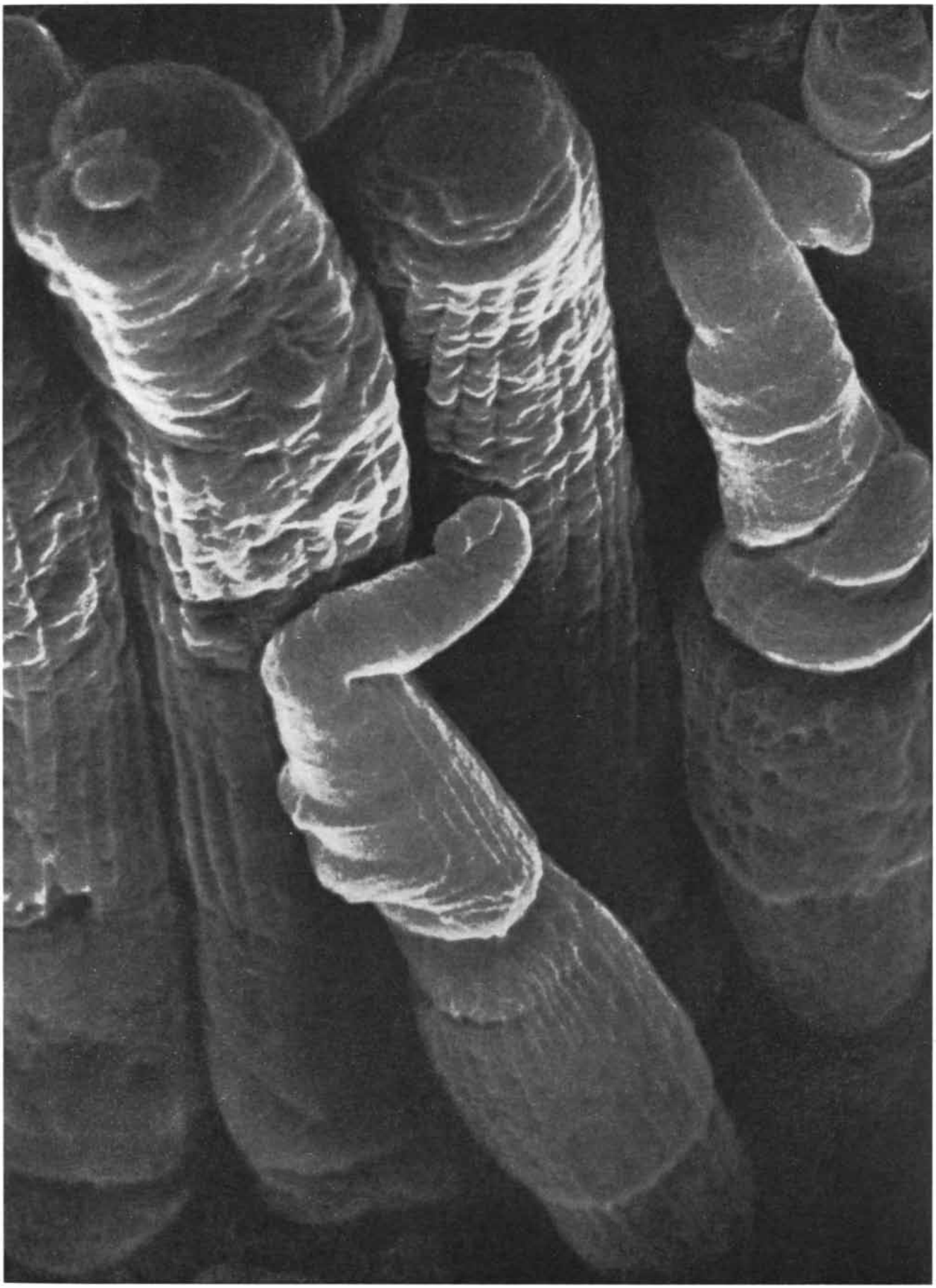
CHANGES IN SOLAR INPUT TO MARS, due to the changing eccentricity of the planet's orbit, could produce cyclical variations in the growth and sublimation of the carbon dioxide frost caps. The top curve shows the long-term changes in the temperature of solid carbon dioxide at the poles. The middle curve shows variations in average atmospheric pressure at the planet's surface that might have resulted from the variations in insolation. The bottom curve shows changes in the height of the permanent frost cap.

accidental development of life from prebiological organic materials. On the other hand, life-on-Mars enthusiasts argue otherwise and emphasize that if water has been available at all in surface layers, it would provide a favorable environment for the development of life. Obviously such a debate cannot be settled by the kinds of information collected by *Mariner 9*. The answer must wait for sophisticated chemical and mineralogical analysis of the surface soil itself.

Simultaneous with the flight of *Mariner 9* the U.S.S.R. undertook an ambitious mission whose objective was to land a capsule on Mars and conduct some analyses of the surface. Unfortunately the Russian lander *Mars 3* failed shortly after reaching the surface and apparently transmitted no useful information. I expect the U.S.S.R. to repeat this kind of mission late in 1973, and I look forward to seeing pictures from the surface and probably the results of some simple chemical analyses.

One hopes that by 1976 we shall be getting information back from a complex U.S. lander being developed in the Viking program and possibly from a second-generation Russian lander. The Viking capsule is being designed not only to look for organic compounds directly but also to perform some simple but important determinations of the basic inorganic composition of surface minerals. Such measurements will provide an important clue to the past chemical evolution of the surface minerals, including whether or not they have reacted chemically with water.

Will the Viking lander finally tell us if life now exists or ever did exist on Mars? I personally rather doubt it. I think the difficulty of obtaining an unambiguous yes or no is so great that it is probably beyond the grasp of even an investigation as ambitious as the Viking mission. My own view is that the final answer to the great search for life on Mars may have to await the return, probably by unmanned means, of a sample of Martian soil for sophisticated analysis in terrestrial laboratories. My guess is that the U.S.S.R., having demonstrated the ability to return unmanned samples from the moon, should be in a position to repeat this feat for Mars around 1980 (assuming that it continues to give unmanned space exploration the same priority that it has had in the past). The U.S. has no ambitious plans for exploring Mars beyond Viking. Thus *Mariner 9* will long be remembered as one of the high points in the American exploration of Mars.



RECEPTOR CELLS in retina of the mudpuppy, a large salamander, are seen end on in a scanning electron micrograph made by Edwin R. Lewis, Yehoshua Y. Zeevi and the author. The cylindrical rods and the pointed cones are enlarged about 5,000 diameters.

The Control of Sensitivity in the Retina

Interactions among nerve cells keep the response range of the system in register with ambient illumination, enabling the retina to form a high-contrast neural image over a broad range of light conditions

by Frank S. Werblin

A good photographer can make pictures with the same camera and the same roll of film on a sunny beach at noon and then while driving home at dusk, even though the average light intensities may differ by more than a million to one. By adjusting the aperture and the shutter speed to control the amount of light that reaches the film, he keeps the response range of the film in register with the available light. In the eye as in the camera there are "settings" that need to be made in order to adapt the visual system to cope with a wide range of light conditions. The difference is that the aperture of the eye, the pupil, plays only a small part in these operations; most of the adjustments are made within the retina, which, like the film, must encode the differences in light intensity in the scene.

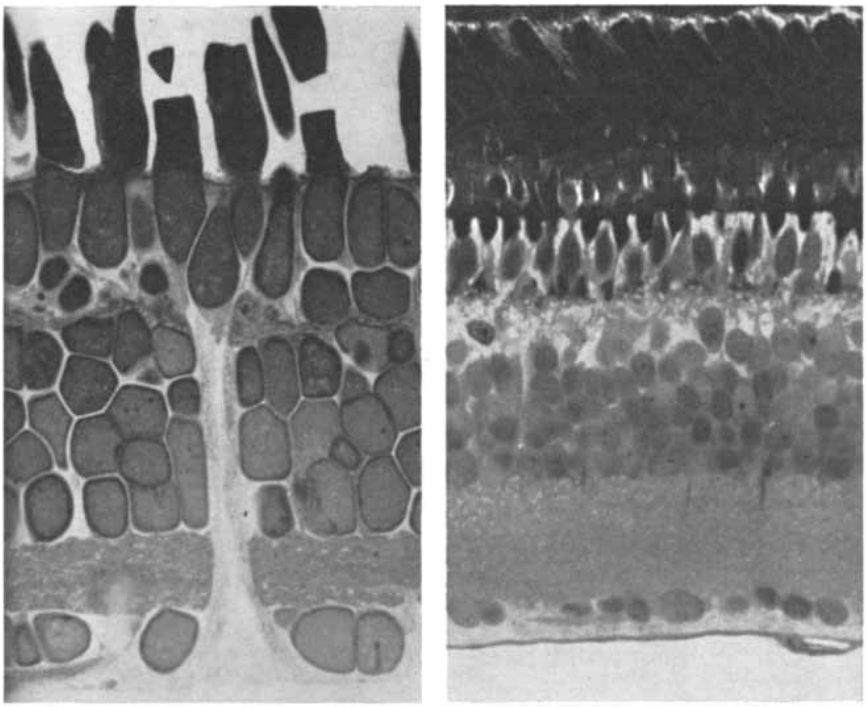
All visual information passes through the retina, which can be thought of as a special subsystem of the brain that has been brought out to the periphery to perform some essential early processing of the visual message. The retina communicates with higher centers of the brain through the optic nerve, which consists of about a million individual nerve fibers, each serving a different area of the visual field. Since the individual fibers of the optic nerve (and fibers in other parts of the nervous system as well) cannot accurately signal levels of activity over a range of more than about 100 to one, the retina must compress the very large range of intensities presented by the outside world into a narrower range that can be adequately handled by the optic-nerve fibers.

One possible method would be for the retina to generate for each point in visual space a signal that is proportional to the logarithm of the incident intensity. That is one of the solutions the camera-film system uses. Such a system, however, if it were used to span the entire range of intensities over which the visual system operates, would provide poor contrast discrimination because very little signaling capability would be available for each small increment in intensity. In order to solve this problem the system might use the full signaling capabilities of each of several different populations of neurons to cover different narrow ranges of intensities, for example one set of neurons for noon on the beach and another for dusk. Such a system would provide better contrast sensitivity, but the visual system would suffer from poor acuity because only a small population of neurons would be operational at any one time.

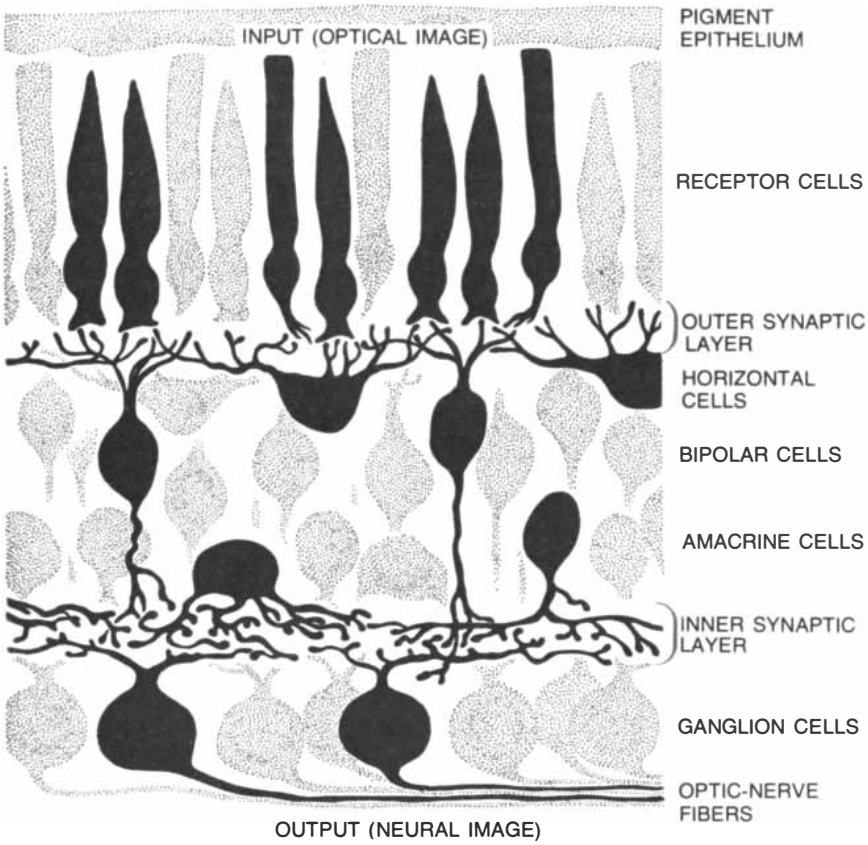
The retina actually performs operations that incorporate the best features of both of the foregoing solutions. As a result it is capable of signaling with high contrast sensitivity over a broad range of field intensities without sacrificing acuity. It is these operations I shall discuss, first describing the structure of the retina: the types of cell represented and the "wiring diagram," or the routes along which the cells communicate with one another.

In all vertebrates the retina is constructed with the same five basic types of cell [see "Retinal Processing of Visual Images," by Charles R. Michael; SCIENTIFIC AMERICAN, May, 1969]. I have worked with the retina of the mudpuppy (*Necturus maculosus*), which is remarkable for the size of its cells and is therefore particularly suitable for electrophysiological experiments at the level of individual cells. Information is carried in two different directions through the retina [*see bottom illustration on next page*]. It is carried in the input-output direction by a sequence of three cell types: receptor cells to bipolar cells to ganglion cells. Only the receptor cells contain photopigments and act as transducers, converting light energy into neural signals that are processed by succeeding retinal cells. The receptor cells drive the bipolar cells, which then pass the signal on to the ganglion cells. The ganglion cells generate the retinal output; their outgoing fibers comprise the optic nerve and carry information to higher centers. Information is also carried laterally, at right angles to the input-output pathway, by the horizontal and amacrine cells, which are strategically positioned to perform important operations that relate activity across different parts of the visual field. The actual connections among the cells can be inferred from information available in electron micrographs. Each neuron communicates with another at a synapse, or junction, probably by releasing a chemical transmitter substance that travels across the small space between the membranes of the two cells. The transmitter is presumably stored in and later released by synaptic vesicles: small packets within the cytoplasm that can be discerned in electron micrographs. One good indicator of the site and direction of synaptic transmission across cell membranes is therefore the presence of vesicles inside the membrane of the transmitting cell.

On this evidence the general scheme of the retinal wiring diagram is quite simple. Each lateral interneuron receives from the cell that precedes it in the input-output pathway, and it is capable of transmitting back to that cell, across to its neighbors or forward to the succeeding input-output cell. This means that horizontal cells can transmit to receptor cells, to other horizontal cells and to bipolar cells; amacrine cells can transmit to bipolar cells, to other ama-



RETINAL CELLS of the mudpuppy, *Necturus maculosus* (*left*), and of the frog (*right*) are enlarged 400 diameters in photomicrographs of sections of retinas made by John E. Dowling, now at Harvard University. The light-sensitive receptor cells are at the top and the ganglion cells, which form the output of the retina and send signals to the brain along the optic nerve, are at the bottom. The remarkably large size of the mudpuppy cells is evident.



STRUCTURE of the vertebrate retina is emphasized in a drawing based on a micrograph of a retina prepared by the Golgi method, which selectively stains a few cell bodies and their processes. The vertically oriented receptor, bipolar and ganglion cells constitute the input-output pathway. Laterally oriented horizontal and amacrine cells carry information across the retina at two distinct levels to relate activity in different parts of the visual field.

crine cells and to ganglion cells. The functional result of these anatomical relations is that each input-output pathway is strongly influenced by activity in neighboring pathways through the laterally oriented interneurons.

Although intercell communication is mediated by chemical transmitters, information is carried within each cell by means of changes in electrical activity. The following is a typical sequence of events: On receipt of the chemical transmitter the electrical properties of the receiving cell's membrane are altered; that leads eventually to a change in the electric potential across the membrane, which is related to the strength of the incoming chemical signal; the change in electric potential in turn affects the release of chemical transmitter from the neuron to the next cell. It is fortunate that the chemical events by which cells communicate with one another have an intermediate electrical correlate within each cell because changes in the electrical activity of cells are easily measured.

The measurement is made by passing a fine glass tube with a conducting center through the cell membrane and reading the potential across the membrane. The potential measured by this fine micropipette electrode gives an index of cellular activity, and it is usually recorded as a function of time following a flash of light. The electrophysiologist, then, has a fairly thorough description of the wiring diagram for the retina showing clearly all possible synaptic relations between cells, but he cannot study these synapses directly; with his electrodes he has access only to the electrical activity of the cells that results from synaptic interactions. He must infer the nature of synaptic function from his indirect electrical measurements.

In most animals the cells of the retina are too small to be penetrated without damage by existing micropipettes, and so Alexander Bortoff of the State University of New York Upstate Medical Center at Syracuse began in 1964 to take advantage of the large cells of the mudpuppy retina. He also showed that the cells could be filled with a stain passed through the recording electrode, so that they could be identified later. Working first with John E. Dowling at Johns Hopkins University and later in my laboratory at the University of California at Berkeley, I have followed up Bortoff's initial efforts.

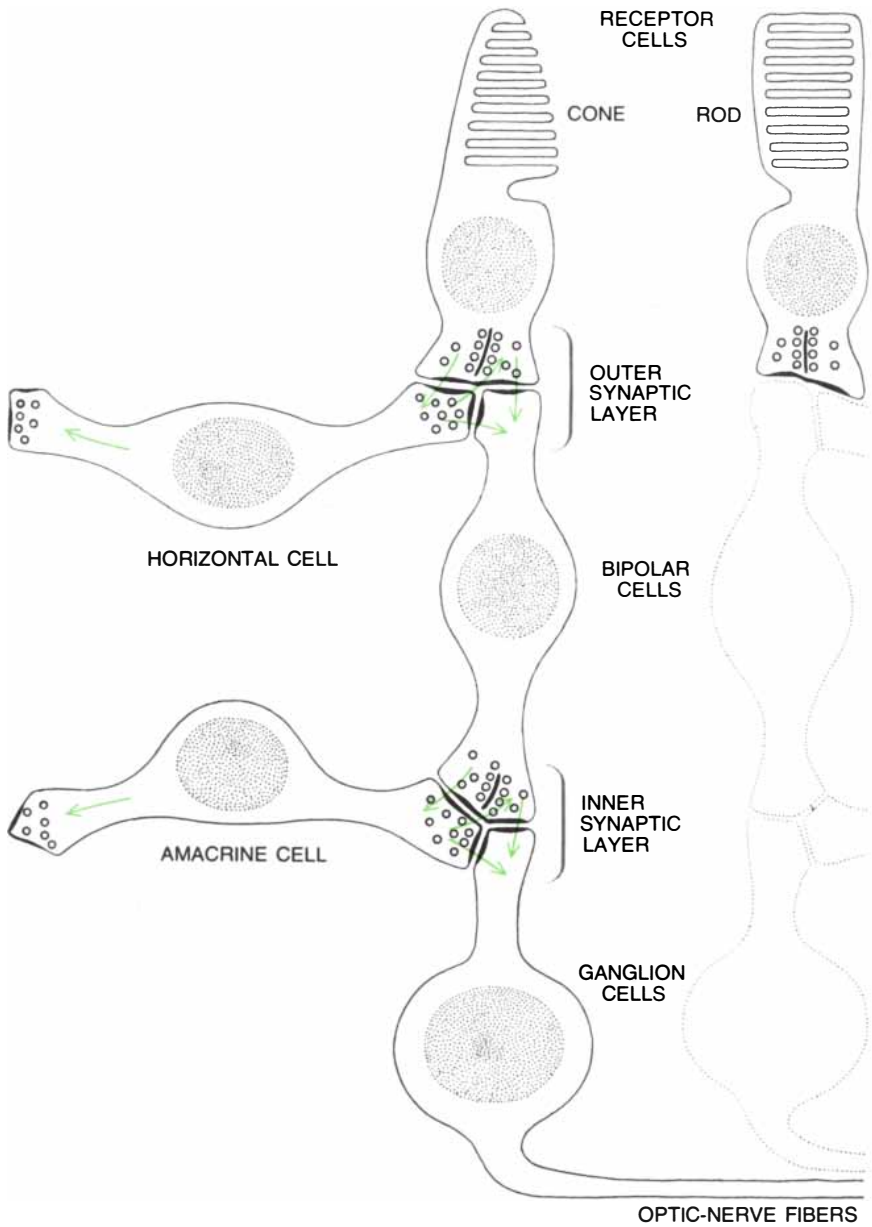
As a first step toward understanding the functional significance of synaptic interactions in the retina, I recorded

from each cell type with a micropipette while flashing spots of light of various configurations on the retina. The pipettes were filled with a stain that I could drive out of the tip into the cell after recording its electrical activity. By identifying the stained cell later I could relate the recorded electrical activity to a specific anatomical category. The most satisfying result of the initial study was that five basic types of electrical response were recorded from cells in the retina, and each response matched up with a particular cell type as determined by the stain-identification technique. From then on the type of cell could be determined simply on the basis of the characteristic form of the response without resorting to the tedious staining procedure.

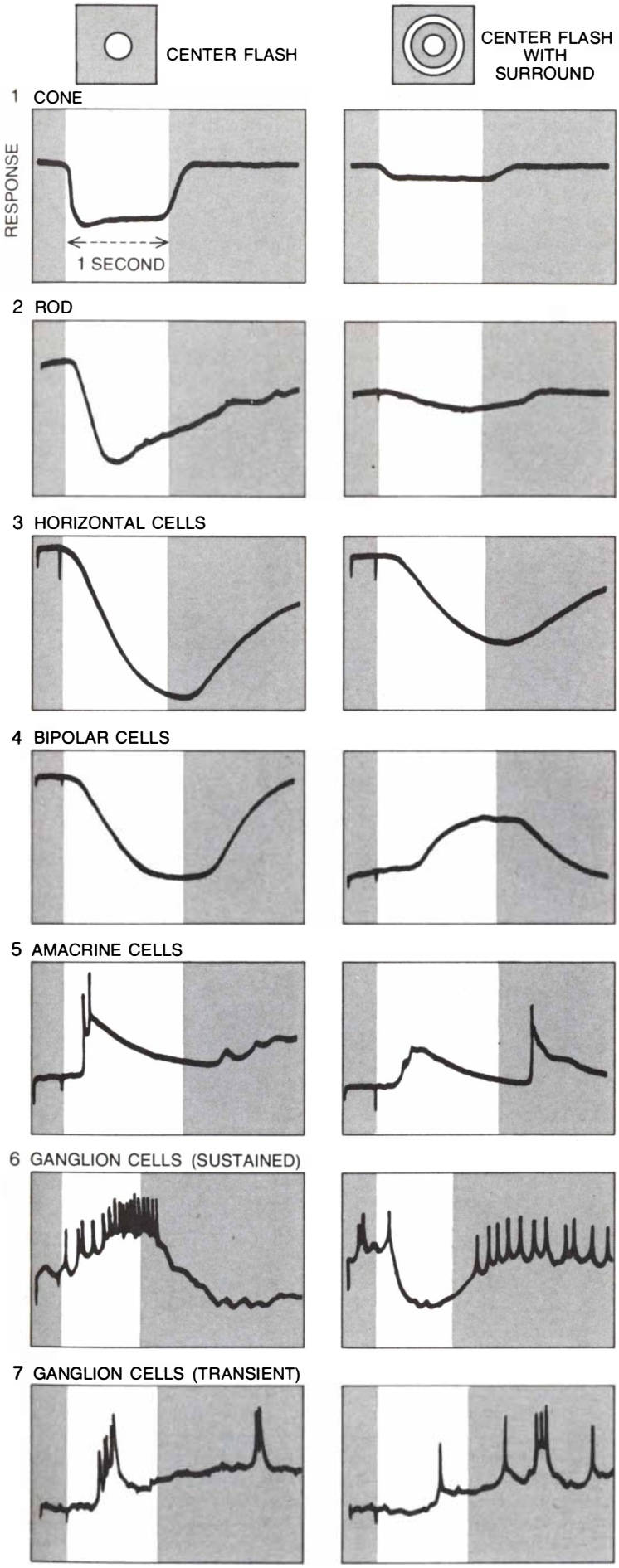
The flashing-light studies can be made to yield more detailed information about the retina if one activates separately the two distinct pathways that enter each synaptic complex. A bipolar cell can be driven either by the receptor cell immediately preceding it or by horizontal cells, depending on the form of the stimulus. A small spot of light illuminating only the receptors above the bipolar cell will selectively activate the receptor-to-bipolar synapse. Illumination by a surrounding ring of light will excite only more distant receptors, which communicate with the central bipolar cell only through horizontal cells, and so the ring preferentially activates the horizontal-to-bipolar (or horizontal-to-receptor) synapse. A similar set of stimulus conditions can be generated in order to activate the various synapses at the base of the bipolar cell.

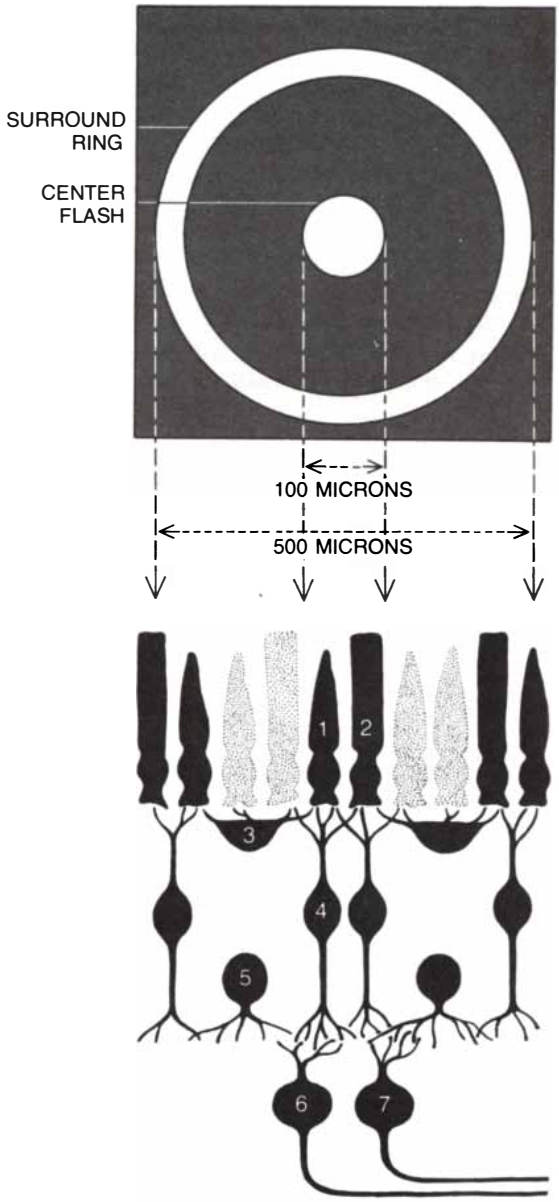
Let us first consider the general forms of the response to light and the ways in which these forms change for the different cell types as information proceeds from the retinal input to its output [*see illustration on next page*]. The receptor cells, the horizontal cells and (in some cases) the bipolar cells polarize in the negative-going direction, or hyperpolarize, with illumination. This form of behavior is most unusual for a neuron: most nerve cells polarize in the positive-going direction, or depolarize, when they are excited. Furthermore, nerve cells usually propagate nerve impulses along their length when they are excited, but there is no indication of such impulse activity in the three cells at the input end of the retina. The form of the retinal message changes dramatically after leaving the bipolar cell. The amacrine and ganglion cells behave more like classical neurons in that they depolarize and also seem to generate more typical nerve impulses when they are excited. The nerve impulse is an essential feature in neurons that need to communicate over great distances, for instance from the retina to the brain along the optic nerve, but it may not be required within the retina itself, where communication takes place over short distances.

The effect of interaction of the two pathways through the retina is most apparent in the recordings from the bipolar cell. It is responsive over a broad area of the retina—its receptive field—but illumination in different regions of the field has different effects. When the flashed spot of illumination falls only on receptor cells that are in direct contact with the bipolar cell, the cell hyperpolarizes; when neighboring receptor cells that have no direct connection with it are illuminated, the cell depolarizes. These neighboring receptor cells communicate with the bipolar cell only by means of horizontal cells carrying information laterally across the ret-



CONNECTIONS among cells were deduced by examining electron micrographs made by Dowling. Cells communicate by releasing a chemical transmitter substance at synapses, or junctions, with other cells; the substance is contained in synaptic vesicles (*small circles*). Messages can therefore be sent from cell terminals that have vesicles. All possible sites and directions of transmission of such messages are shown (*arrows*), as inferred from electron micrographs. Bipolar and ganglion cells simply "read out" and pass along signals formed by interactions of vertical and lateral neurons; lateral interneurons transmit back to cells that drive them, across to one another and on to succeeding input-output cells.





ELECTRICAL RESPONSES (*left*) **of retinal cells were determined by recording with a micropipette while flashing a spot and then also a ring of light on the retina, as shown above. The two kinds of receptor cell and the horizontal cell respond to light** (*unshaded portion of each record*) **by hyperpolarizing, that is, the membrane potential becomes more negative; the bipolar cell here hyperpolarizes when stimulated by a receptor cell but depolarizes, that is, the membrane potential becomes less negative, when it is driven indirectly by horizontal cells. Amacrine and ganglion cells depolarize and, unlike the other cells, generate more conventional nerve impulses. The "sustained" type of ganglion cell, like the bipolar cell, responds to the steady level of illumination; the "transient" type, like the amacrine cell, responds briefly to a changing stimulus.**

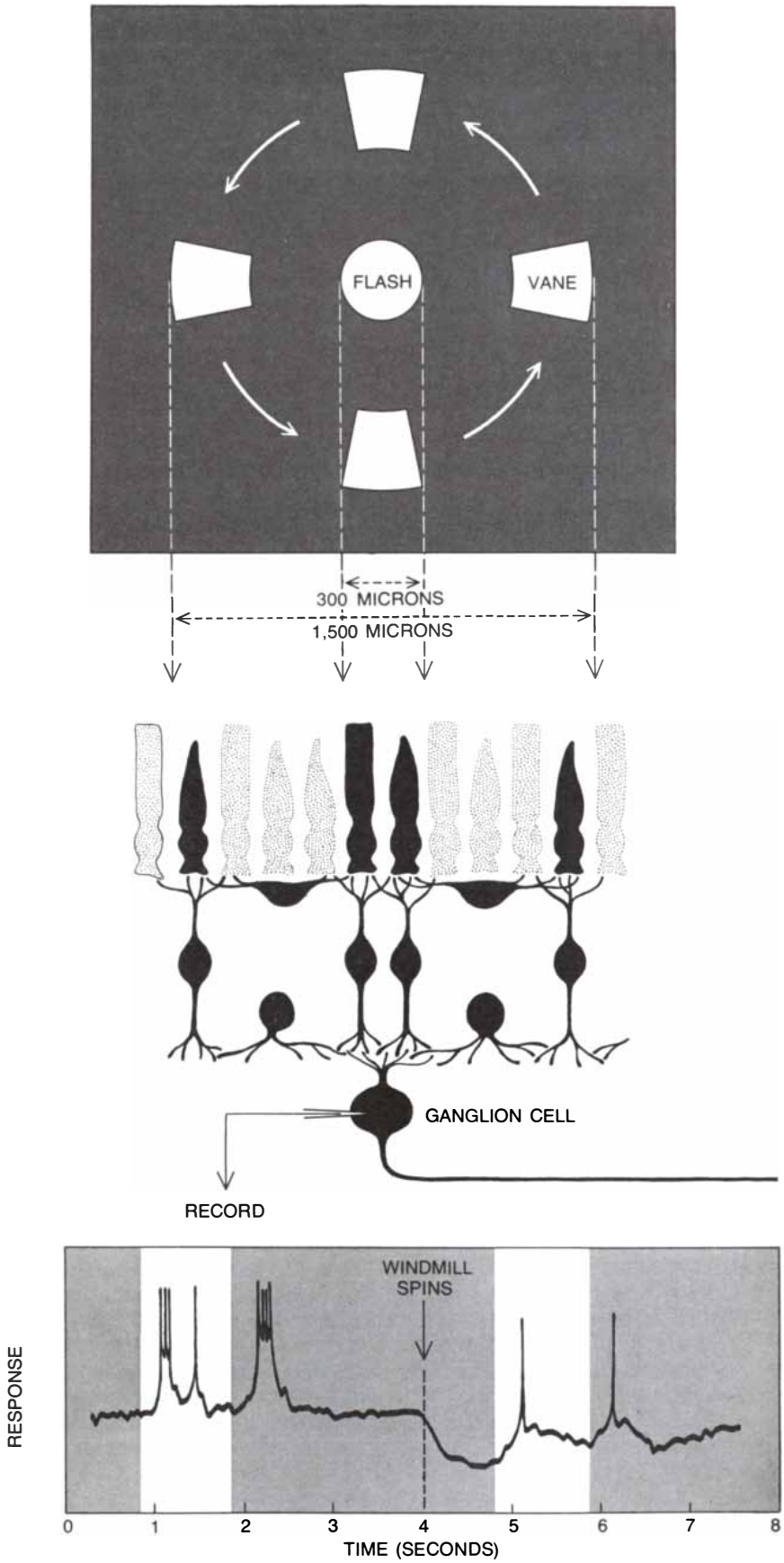
ina. The response of the bipolar cell is inverted when the surround is illuminated because an input from the horizontal cell and an input from the receptor cell through the synaptic structure at the base of the receptor cell affect bipolar cells in opposite ways. The receptive field for the bipolar cell is therefore said to be concentrically organized, with antagonistic center-surround components. (This form of lateral antagonism is a characteristic of most sensory systems, including the auditory and somatic systems and the visual system of invertebrate animals. The retina of vertebrates is unique in that the cells associated with lateral antagonism generate only slow potentials and show no sign of nerve-impulse activity.)

In our experiments the receptor cells themselves behaved quite differently. They were driven strongly by the spot of light that illuminated them directly, but their response was feeble when neighboring receptor cells were illuminated with the stimulating ring. Alan Baylor and his co-workers at the National Institutes of Health have shown that in the retina of the turtle under certain conditions the receptor cells can be driven in the opposite direction by horizontal cells, which is good evidence that the horizontal cells are feeding back to the receptor cells. In this sense receptor cells also have a broad receptive field with an antagonistic surround.

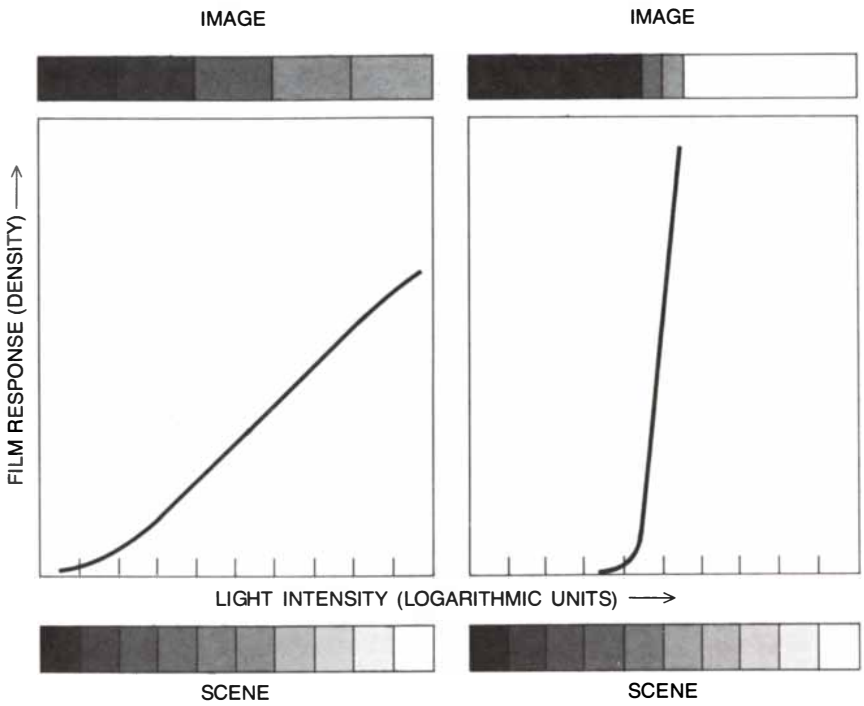
The horizontal cells extend across the retina, communicating with many receptor cells, and so it is not surprising that they are well driven by either the spot stimulus or the ring; even the ring falls on receptor cells to which the far reaches of the horizontal cells extend. Each form of receptive field is thus the direct manifestation of the structural organization of cells and synapses.

The retinal wiring diagram shows that bipolar cells drive both amacrine and ganglion cells, so that the antagonistic form of the bipolar-cell receptive field should also be represented in these subsequent cells. One type of ganglion cell, the sustained-response type, appears to be driven directly by the bipolar cells. Its activity (now in the form of nerve impulses) is elicited by the central spot but inhibited by the ring stimulus, following closely the general form of antagonistic activity in the bipolar cell. This sustained, concentric, antagonistic function measured at the ganglion cells seems to be one of the important components of the retinal output that is sent to the brain.

There is another form of retinal output, represented by another set of gan-



RESPONSE TO CHANGE by the transient ganglion cells is demonstrated by providing a "windmill" light, spinning in the ganglion cell's surround and continuously stimulating amacrine cells below the vanes. The amacrine activity propagates across the retina and impinges on the ganglion cell at the center. As a result the change-detecting response of the ganglion cell at the center of its field is reduced by change in surrounding region (*bottom*).



PHOTOGRAPHIC FILM can have low-contrast or high-contrast characteristics. Low-contrast film has a shallow operating curve spanning the full intensity range (*left*). High-contrast film has a steep curve spanning a narrow range (*right*). The bars represent the light intensities in a scene (*bottom*) and film response (*top*) as displayed in a positive print.

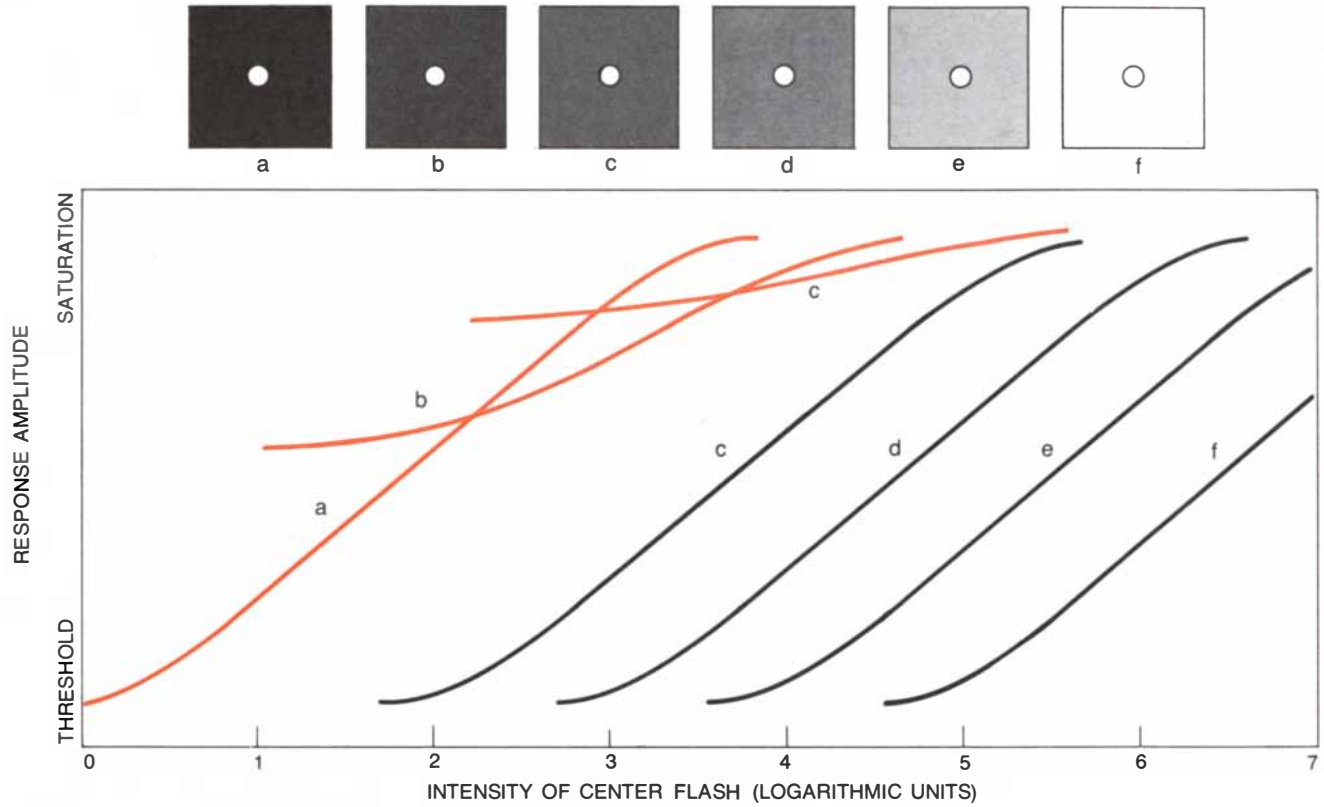
glion cells. These cells depolarize and fire nerve impulses only when the intensity or the configuration of illumination presented to the retina is changing, that is, when the bipolar cells that make synaptic contact with them are "turning on" or "turning off." Their response is affected by a lateral interneuron, the amacrine cell, that seems to have quite similar change-detecting response properties. In responding to change the amacrine cells appear to have an antagonistic effect on the change-sensitive ganglion cells, in a manner analogous to the effect of horizontal cells on bipolar cells. This was demonstrated by spinning a "windmill" light to introduce a constant rate of change in the surround of the input-output pathway to the ganglion cell. The spinning of the windmill decreased the response of the change-detecting ganglion cell at the center of the configuration [*see illustration on preceding page*].

The retina, then, appears to have two structurally similar systems of lateral interactions, one at the base of the receptor cells and the other at the base of the bipolar cells. At each site the lateral interneuron has the synaptic machinery necessary to modulate the signal that is being transmitted along the input-output pathway. Horizontal cells responding to sustained levels of illumination affect the magnitude of the sustained bipolar-cell signal; amacrine cells, responding to change, affect the responsiveness of the change-detecting ganglion cells. As a general principle of organization, the lateral interneurons receive input over a broad retinal area (the receptive-field surround) and form an averaged signal by which they modulate the local input-output pathways (receptive-field centers) of the bipolar and ganglion cells. Both the sustained-signal detectors and the change detectors have representatives among the ganglion cells that form the retinal output, so that messages about both the presence and the change of illumination are transmitted to the brain.

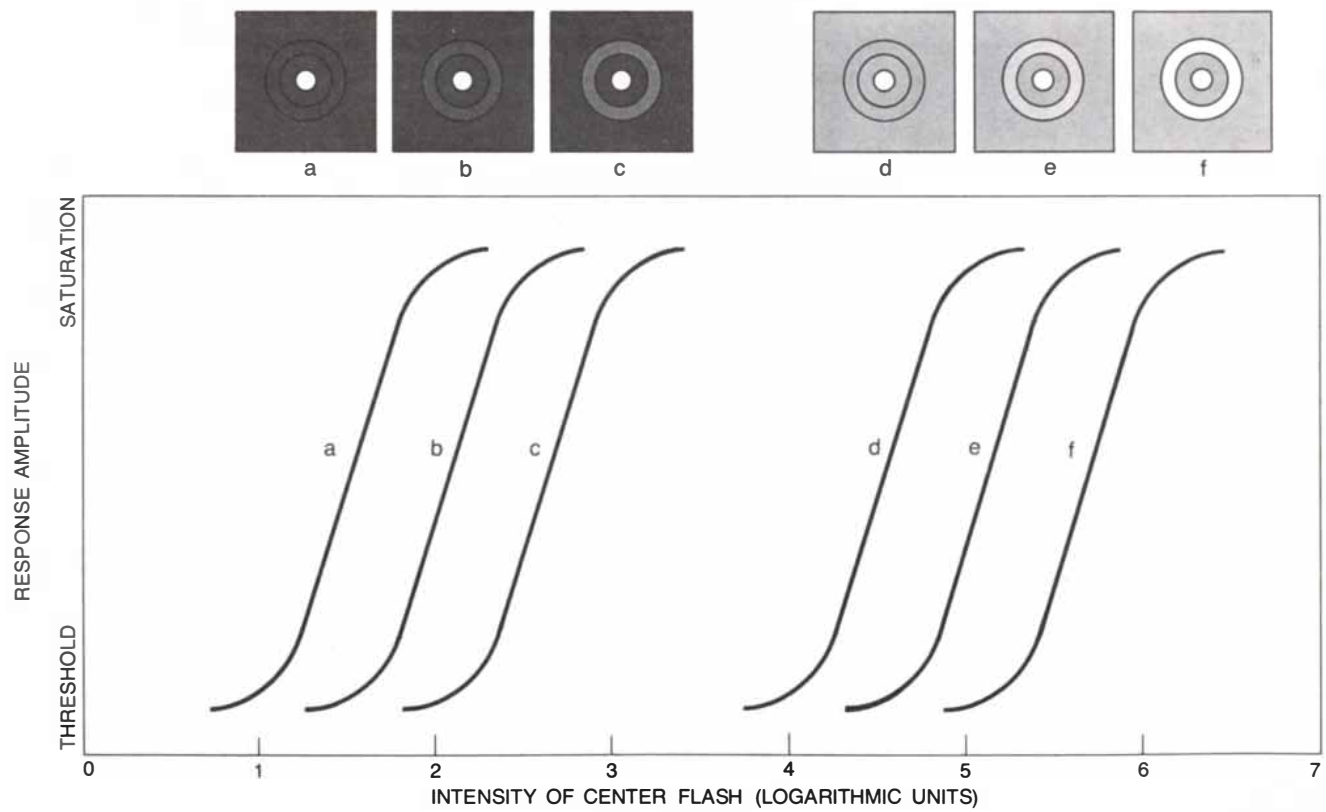
The two systems of modulation by lateral interneurons are two elements of the total process by which the retina adjusts its operating characteristics to prevailing light conditions. One aim of these adjustments is to generate a high-contrast visual signal, and so it will be useful to review the analogous procedures in photography before examining the retinal mechanisms. There is low-contrast film and there is high-contrast film. Each is defined by a characteristic operating curve that relates the intensity of the light striking it (in logarithmic units) to the response of the film: the density of the developed silver-salt grains in the emulsion. The operating curve of a low-contrast film spans about 10 logarithmic units (or 10^{10} units) of intensity, the same range that human vision spans. Its slope is correspondingly gradual: there is not much difference in response (film density) for a small difference in intensity. Low-contrast film is easy to use since all possible image intensities fall somewhere on the curve and therefore produce a response density on the film. The film's ability to separate different intensities—its contrast sensitivity—is minimal, however; the picture may be too "soft," or washed out. The operating curve of a high-contrast film, on the other hand, spans only about one logarithmic unit. It is correspondingly steep, and so the film gives good separation of different intensities and a sharply contrasting image. High-contrast film is harder to use. The photographer must move the significant intensities in the scene into accurate register with the film's response range. He does this by adjusting the camera's aperture and shutter speed. His adjustments take a few seconds to complete; then he exposes the film for a fraction of a second to make the picture.

With the flashing-light experimental setup we set out to establish the characteristics of each cell type's graded response. We determined the operating curve of a cell by flashing light at various intensities and measuring the graded amplitude of the response, from threshold to saturation. The resulting operating curve, as in a photographic film, represents the peak response of the cell, which is reached within a fraction of a second. We then changed the ambient conditions by increasing the background luminance in order to see how the operating curves were adjusted. By flashing the surround ring or spinning the windmill we activated the lateral pathways that modulate the visual message as it passes through the retina. What then happens to the operating curve represents the "setting" of the cell's response as a function of conditions in the visual field.

There are two kinds of receptor cell, the well-known rods and cones, and Richard Normann of our laboratory showed that they have different operating characteristics relating neural activity (rather than density) to the incoming light intensities. The rods are somewhat more sensitive, coming into operation at intensities about one logarithmic unit lower than the cones. At any one background level each receptor operates over a curve that spans about three logarithmic units and is therefore intermediate in slope between the



RECEPTOR-CELL RESPONSE RANGE is shown at six background-illumination levels. The rods (*colored curves*) are more sensitive than the cones in that they begin to respond at lower intensity levels, but they saturate quickly with increasing background illumination. Cones, however, appear not to saturate; their operating curves (*black*) shift along the intensity axis with increasing background illumination, so that they are optimally responsive over a narrow intensity range near each background level.



FIRST TRANSFORMATION of operating curves comes between receptor cells and bipolar cells. The response range of the bipolar cells follows that of the receptor cells with increasing background illumination but the narrower bipolar-cell curve is further fine-tuned within the receptor-cell range by input from the surround: an increase in the ring brightness shifts the curve to the right.

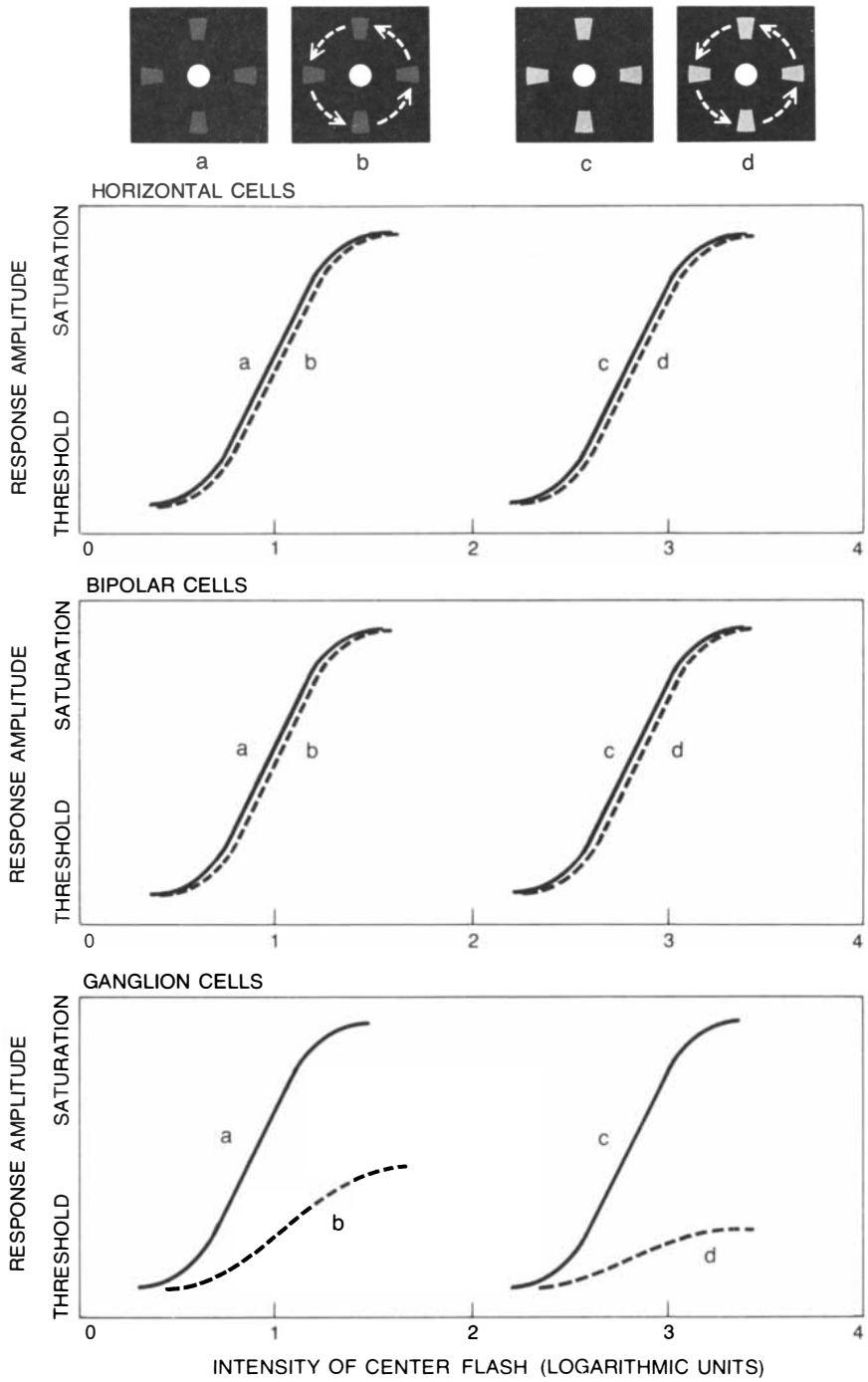
curves of low-contrast and high-contrast film. When the background level is raised, photochemical mechanisms that are still not fully understood come into play and somehow reduce the sensitivity of the receptor cells. These mechanisms move the cone operating curve along the intensity axis to keep it in register with ambient conditions; the shift takes place slowly, within about five seconds. The rod curves are compressed as the background level increases, saturating and becoming inoperative after only about four logarithmic units; the cone curves continue to shift along the intensity axis without compressing [*see top illustration on preceding page*].

This receptor-cell function is the first in a series of events that lead finally to a high-contrast output from the retina. Since each curve covers only about three logarithmic units of the 10 logarithmic units over which the visual system operates, the receptor cells can use their entire signaling capacity over this narrow range of intensity, thus making more millivolts of signal available per unit of intensity than if they had a shallow curve extending across the entire operating range.

In the camera, then, the light intensity is adjusted to fit the response range of the film. In the retina the response range of the cells is adjusted to fit the light intensity. The major "exposure setting" function is carried out by the receptor cells; fine-tuning and accentuation and modification of contrast are carried out at subsequent levels.

The receptor cells constitute the window through which all subsequent visual cells must view the world, so that the response range for all other retinal neurons always falls within the three logarithmic units spanned by the receptor cells. Horizontal cells, for example, have an operating curve that is similar to that of the receptor cells and shifts with it.

The bipolar cell's response curve is steeper than that of the receptor cells or horizontal cells. Its graded response goes from threshold to saturation within a little more than one logarithmic unit, much as the curve for a high-contrast film does. Since the bipolar-cell curve is responsive over such a narrow range of intensities it must be positioned even more accurately than the receptor-cell curve; so positioned it can signal, with a higher contrast function, the presence of important boundaries in the visual field. It is at this point that the antagonistic effect of the horizontal-cell input comes into operation. Remember that when the ring stimulus was flashed, the processes of the horizontal cells carried across the retina a signal related to the average light intensity in the immediate surround. The effect of the signal was to reduce the magnitude of the bipolar cell's response. The reduction is actually manifested as a shift of the bipolar-cell operating curve to the right—to a higher range of intensities—with increasing surround luminance. The shift is accomplished very quickly, within the fraction of a second that it takes for the horizontal cells to respond.



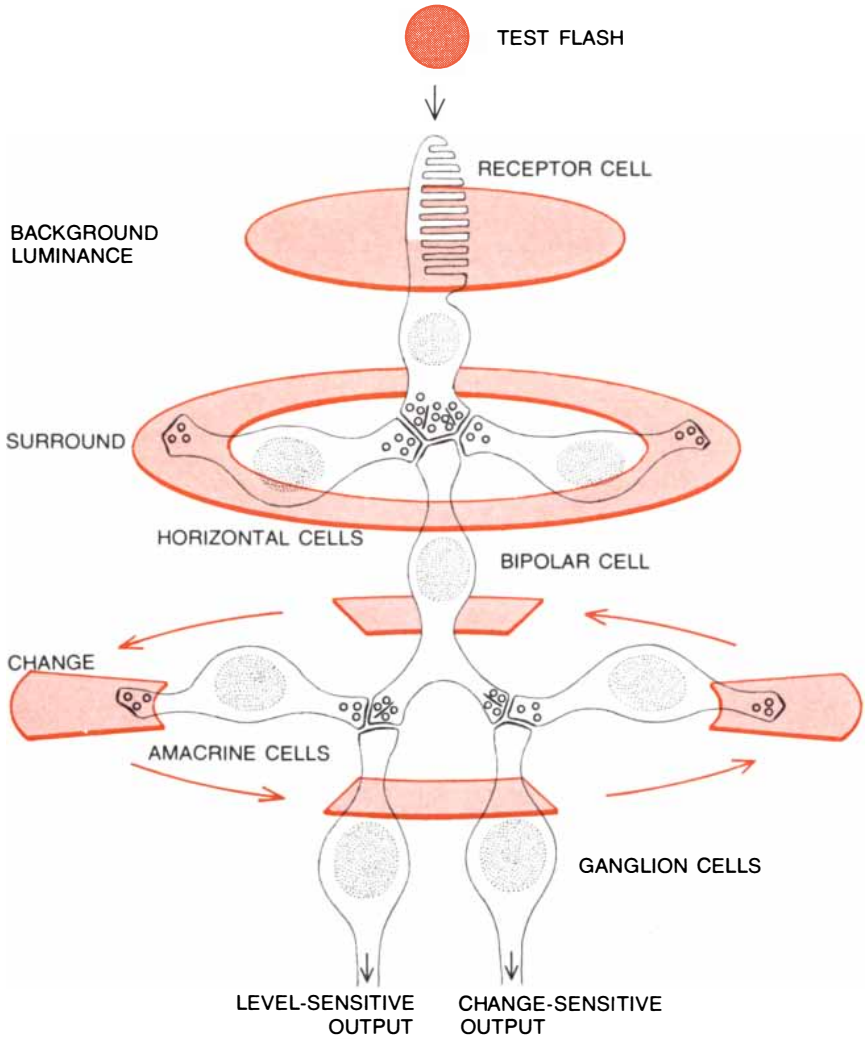
SECOND TRANSFORMATION comes between bipolar and ganglion cells. Bipolar-cell curves are shifted by increasing surround illumination (brighter windmill vanes). The curves for both sustained and transient ganglion cells driven by bipolar cells follow these shifts. When the windmill is spun (*broken curves*), the amacrine system suppresses the activity of the change-detecting transient ganglion cells: their operating curves are compressed.

By this mechanism the high-contrast operating curve of various groups of bipolar cells can be moved to different regions of the intensity axis in different parts of the visual field. Such a system outdoes any exposure-setting device on a camera. Imagine being able first to

control for general illumination in the visual field (through the exposure-setting function of the receptor cells) and then to fine-tune the exposure in local regions of the field as a specific function of the local intensity level! The system does have one disadvantage: since intensity discrimination in any part of the retina is controlled by the activity in the area immediately surrounding it, two separated points in the scene that actually reflect the same amount of light can appear to have different intensities if they are surrounded by "control regions" of different intensities.

The full mosaic of bipolar cells brings its high-contrast neural image to the inner retina. Here in the inner synaptic layer bipolar cells are connected to the amacrine and ganglion cells much as receptor cells are connected to bipolar cells and horizontal cells in the outer synaptic layer. Some ganglion cells simply pass along to the brain the sustained signals they receive from the bipolar cells; they are unaffected by the amacrine cells. Other ganglion cells, however, respond to change in the signal from the bipolar cells, which corresponds to the presence of movement or fluctuating intensities in the visual field. These change-detecting ganglion cells are embedded in a system of amacrine cells that also respond to change but that exert an antagonistic effect on the ganglion cells. When the amacrine system is activated, as by the moving windmill or other spatiotemporal change, it serves to reduce the effectiveness of the bipolar-to-ganglion signal and reduces the slope of the ganglion-cell operating curve. In other words, the change-detecting ganglion cells are activated by change at the center of their receptive field, but change in the surround activates the amacrine-cell system that acts to reduce ganglion-cell activity. This is another example of a concentric, antagonistic receptive field, but here both antagonistic components are driven by change. As a result change over broad regions acts to reduce the response to change occurring locally. Such a system has no analogy I know of with the technology of photography, although a similar technique is used in television. It is as if a camera system could switch automatically from a high-contrast film to a low-contrast film when it encountered a rapidly changing or very contrasty scene.

How can this system, which responds to local change but is antagonized by steady broad-field change, be an asset to the organism? The answer is that it provides for detection of the movement of small objects within the visual field but is not overwhelmed by the vast changes in contrast, covering broad retinal regions, that result, for example, from blinking or eye movements.



RETINAL SENSITIVITY is determined by three distinct properties of the visual scene, each activating a specific mechanism. Average background luminance affects photochemical processes in receptor cells. Luminance in surrounding regions affects interactions mediated by horizontal cells at the receptor-cell terminals. Spatiotemporal change in surrounding regions affects interactions mediated by amacrine cells at the bipolar-cell terminals. At any background ganglion cells carry signals that are related to luminance or change at local regions but are modified by corresponding activity in surrounding regions.

In summary, both ends of each neuron in the input-output pathway of the visual system are subject to some form of signal control. The first effect is at the outer end of the receptor cells, where photochemical processes somehow adjust the efficiency of the transducer so that the receptor cells operate best over different ranges of intensity depending on the average ambient intensity level. In the mudpuppy this operation takes many seconds to complete. Next, the narrow operating range of the high-contrast bipolar cell can be shifted, within the broader operating range of the receptor cells, according to the luminance level in the local surround of each bipolar cell. This shifting, mediated by horizontal cells carrying information across the retina, is accomplished within a fraction of a second and serves to "fine-tune" each bipolar-cell operating curve to the appropriate intensity range. Finally, the contrast sensitivity of the change-detecting output of the retina can be modulated by the presence of change in local surrounding regions; this is done through the lateral system of amacrine cells. With these mechanisms the retina can transmit a high-contrast, high-acuity message to the brain carrying information about the presence and movement of boundaries within the visual field, information that is essential to the brain's function of reconstructing the visual world.

Judaism at the Time of Christ

Traditionally Judaism is regarded as a unified set of beliefs. New evidence indicates that between 200 B.C. and A.D. 100 its expressions were remarkably diverse, and some were even pagan

by Michael E. Stone

A series of discoveries in recent decades has led to startling changes in modern conceptions of Judaism and Christianity in the period between 200 B.C. and A.D. 100. The work of scholars in several disciplines has uncovered unsuspected dimensions in the diversity of Jewish religious expression. The creativity of Judaism in the period before and after Christ may well be its outstanding feature.

With the conclusion of the great historical accounts in the Hebrew Bible about 400 B.C. there is a gap in our knowledge of Judaism for the next two centuries. Although a certain part of the literature of the Bible was edited and perhaps even written in these centuries, we know almost nothing about the development of Judaism during the period. It was, however, a time of great military and political change throughout the Middle East: Alexander the Great had conquered the Persian Empire, and the Hellenization of the ancient native cultures of the area was well under way. This Hellenistic age was one of unusual religious variety, of high theosophies and magic, of philosophies and mysteries. When the veil lifts on Judaism after 200 B.C., the world of the Jews differs greatly from the one reflected by late biblical literature. Beginning in 167 B.C. with the revolt of the Maccabeans against the Seleucid successors of Alexander, there is fairly plentiful literary evidence.

Judaism showed a fecundity of expression typical of other Hellenistic religions. Yet of all the multifarious forms of Jewish religious expression that arose, only two have survived and flourish today: the rabbinic form from which all modern Judaism descends and the form that is known as Christianity. A marginal survivor is Samaritanism, maintained by a tiny group of Samaritans in Shechem in central Palestine, who continue to worship on their holy Mount Gerizim, the site of their temple in ancient times. If we were dependent on only the writings transmitted within the Jewish tradition, our knowledge of Judaism around the time of Christ would be sparse indeed. Information about those groups that rabbinic Judaism opposed has been systematically ignored or suppressed in rabbinic writings. We would know nothing of the political fortunes of the Hasmonean dynasty, which followed the Maccabean revolt, nor of the rebels who were besieged by the Romans at Masada. Much of our knowledge of the Sadducees and the Essenes in the deserts of Palestine would have been lost. Information about them and about the Judaism of the Hellenistic Diaspora—the extensive settlement of Jews outside Palestine—has been transmitted to us outside the Jewish tradition. Because Christianity came from a different part of the spectrum of Judaism, its preferences and prejudices differed, and it preserved a very different body of Jewish writing. The documents provided by two remarkable archaeological finds—the Dead Sea scrolls of Qumran and the Gnostic manuscripts of Nag Hammadi in Egypt—are only now being assimilated.

Perhaps the first indication of the variety of Jewish religious expression in the period between 200 B.C. and A.D. 100 came with the translation into European languages of Jewish books that had been preserved only by the Ethiopian church and other Eastern Christian churches. Chief among these were the Book of Enoch and the Book of Jubilees, both translated from the Ethiopic in the 19th century. It was evident that the Book of Enoch served as a source for the Letter of Jude in the New Testament and for other early Christian writings. Compare, for example, the following two passages, the first from the Book of Enoch and the second from the Letter of Jude:

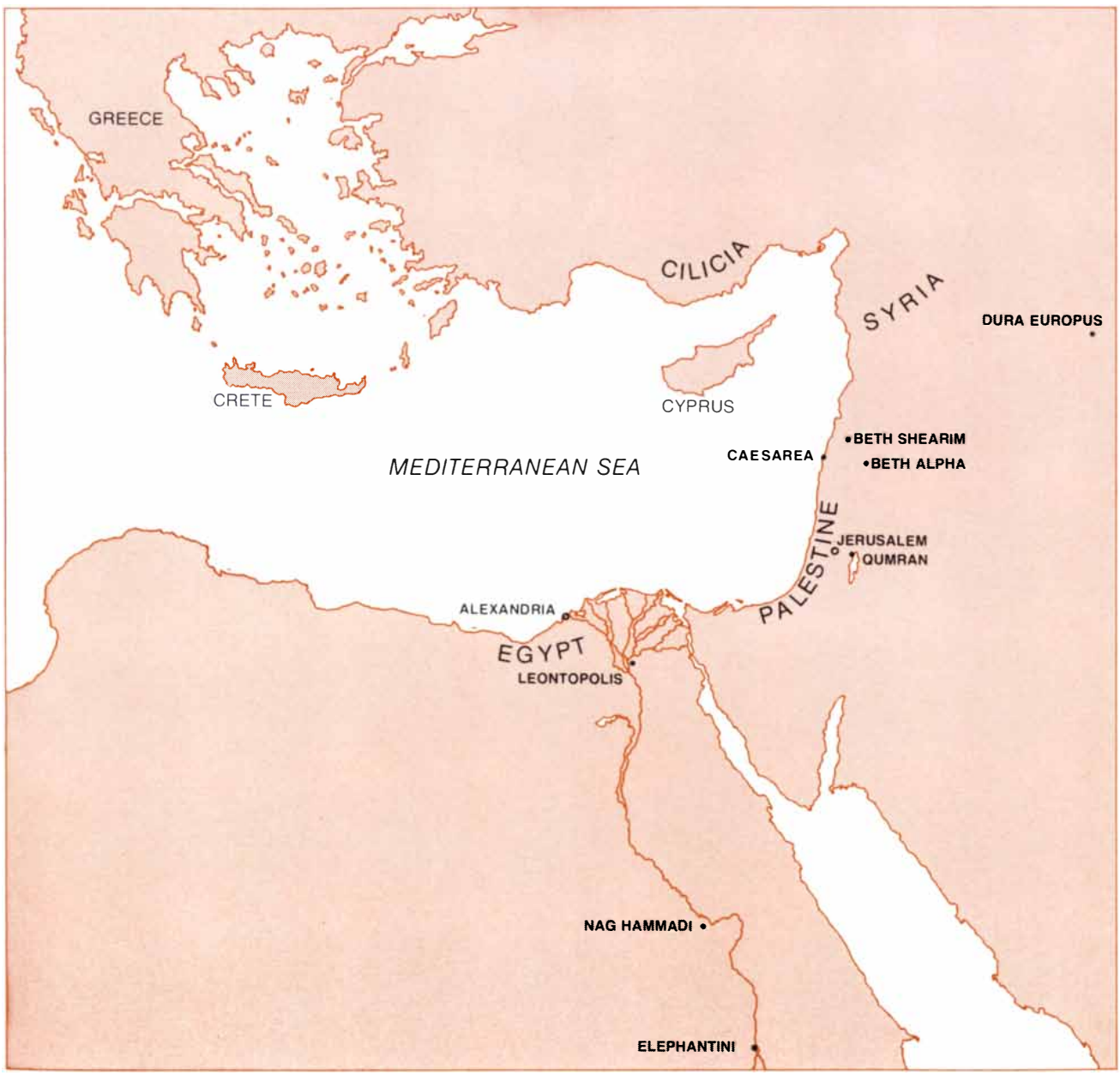
"And behold! He cometh with the myriads of His holy ones, to execute judgment upon all, and to destroy all the ungodly: And to convict all flesh of all the works of their ungodliness which they have ungodly committed, and of all the hard things ungodly sinners have spoken against them."

"It was of these also that Enoch in the seventh generation from Adam prophesied, saying, 'Behold, the Lord came with his holy myriads, to execute judgment on all, and to convict all the ungodly of all their deeds of ungodliness which they have committed in such an ungodly way, and of all the harsh things which ungodly sinners have spoken against him.' "

SYNAGOGUE FRESCOES dating from the third century found at Dura Europus in Syria are allegorical representations of Old Testament stories. The illustration on the opposite page is a detail from "The Cycle of Ezekiel" on the north wall of the synagogue. It shows Ezekiel in Persian costume standing beside three corpses. A Greek-style figure of Psyche is reaching for the head of the top corpse and three other Psyche figures fly down from above. It is thought that the Psyche figures represent the four winds, which were summoned by Ezekiel to breathe life into the corpses. The detail is from a photograph of the fresco provided by Frank J. Darmstaedter of the Jewish Theological Seminary of America.

The points of view in the Book of Enoch differ in many respects both from those in rabbinic sources and those in Christian stereotypes of Judaism. These differences have become the object of intensive scholarly research. Eleven fragmentary manuscripts of the Book of Enoch have also been found among the Dead Sea scrolls. J. T. Milik, who is interpreting and editing the manuscripts, reports that the earliest one was copied in the first part of the second century B.C. In his view some sections of Enoch are much older, and he even suggests that the final editors of the portions of Genesis in the Old Testament that deal with Enoch may have summarized an ancient source of the Book of Enoch (Genesis 6:1–4). One of the major points made by the section of Enoch that Milik regards as being the oldest is that evil is the result of the illicit union between angels (sons of God) and women. From this union issued not only forbidden knowledge but also demonic spirits.

The discovery of the Dead Sea scrolls helped to reawaken interest in Jewish and Christian nonscriptural writings. Whereas such writings, the Apocrypha and pseudepigrapha, had been known only in Greek or in translations made from the Greek, much of the material in the scrolls was in Hebrew or in Aramaic. The impact of the translation of these books on students of the period was profound. Both Jewish and Christian scholars had tended to evaluate evidence of varieties in Jewish religious expression in terms of their own orthodoxies. Jewish scholarship had underplayed the mystical, speculative and other nonrational aspects of Judaism. The result was a rather limited picture of the rabbinic Judaism that had emerged triumphant after the Temple in Jerusalem had been destroyed by the Romans in A.D. 70. This picture had its supporters among Christian scholars of rabbinic



GREEK INFLUENCE spread across the eastern Mediterranean lands following the conquests of Alexander the Great in the fourth century B.C. The Hellenization of Eastern cultures and religions, including Judaism, continued through the period of Roman rule and led to a rich variety of religious expression. Evidence for variety of Judaism has been found at several sites (*black dots*).

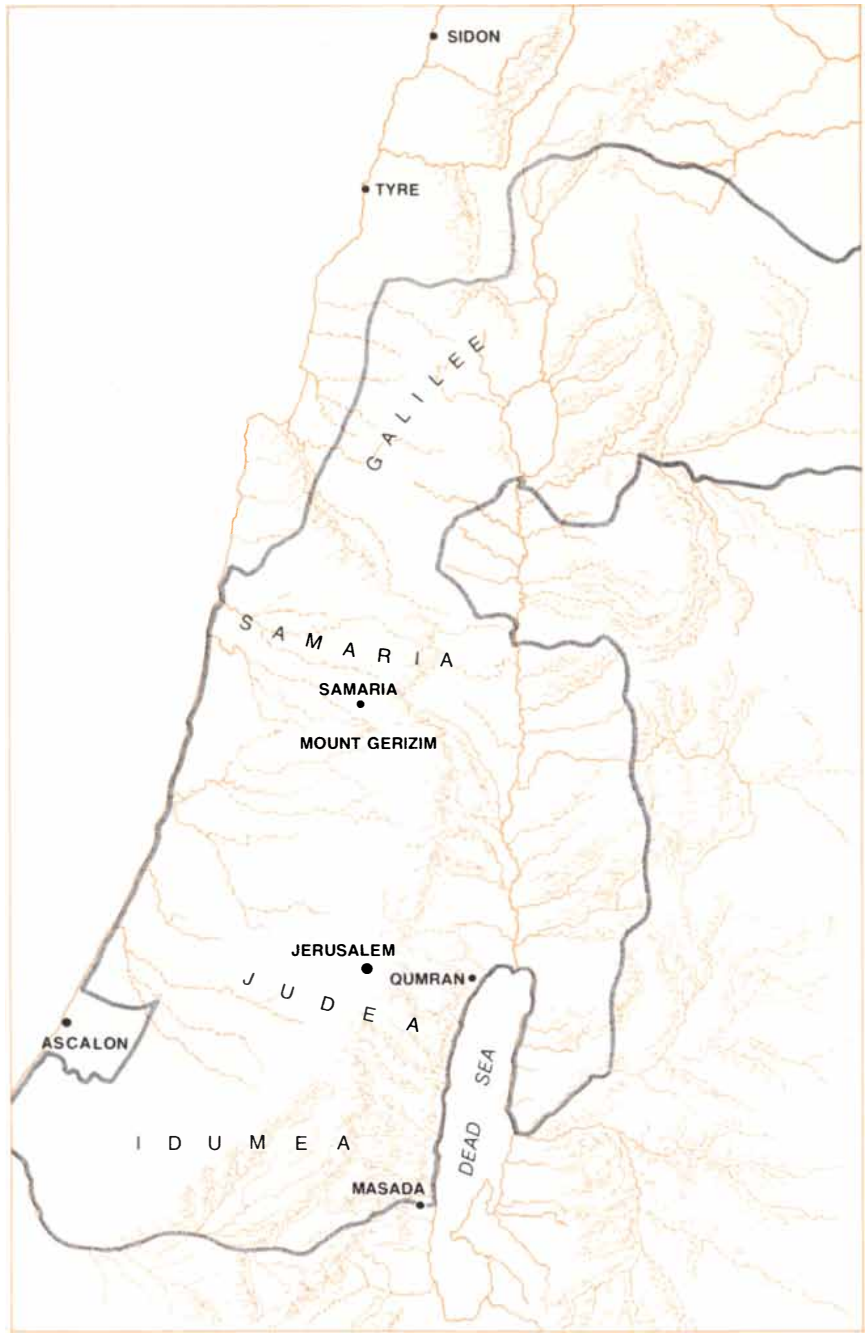
Judaism. On the whole, however, Christian scholars were much occupied with the Apocrypha and pseudepigrapha.

Their interest lay first and foremost in the search for the background of the New Testament. In this search the evidence of rabbinic Judaism was treated equivocally at best. The pseudepigraphical books, however, were viewed as "opposition literature" to a caricatured view of Pharisaic-rabbinic Judaism that was attributed to the "establishment" of the era before the destruction of the Temple. Thus the pseudepigrapha were regarded as being of central importance. With the finding of the Dead Sea scrolls features of the way of life of the Qumran sect that produced the scrolls became known, and suddenly a context was revealed in which the creation of at least part of the pseudepigraphical literature could be visualized [see "The New Covenanters of Qumran," by Shemaryahu Talmon; SCIENTIFIC AMERICAN, November, 1971].

The discovery of the Gnostic library near Nag Hammadi in Egypt in 1945 has had a similar impact. The Gnostics, a group of religious sects of the late Greco-Roman and early Christian eras, sought salvation through occult knowledge. Their movement posed a major challenge to orthodox Christianity in the second and third centuries after Christ. Until the discovery of the library most knowledge of Gnosticism was derived from Christian writers who had been involved in combating the heresy. As the Gnostic papyri are translated from the Coptic and published, it becomes increasingly evident that much of Gnosticism is probably of Jewish origin. A striking example is a treatise titled the Testimony of Truth. Embedded in the treatise are fragments of an older literature, including some that have many points of contact with orthodox Jewish interpretations of the Scriptures, although the theological stance is thoroughly Gnostic.

One particularly important fragment deals with the passage in Genesis in which the subtle serpent induces Eve to eat the forbidden fruit. The Gnostics regarded the serpent as a positive figure, the wisest of the animals and the instructor of Eve. Eve in turn instructed Adam in the true wisdom. In Aramaic the words for "Eve," "reveal," "serpent" and "beast" are very similar to one another, making possible a play on words that binds together these ideas. The Gnostic views in the passage are quite undeveloped, and the passage has no Christian elements at all. This shows that it is probably derived from pre-Christian Jewish Gnostic sources. The positive view of the serpent and of the knowledge gained from eating the fruit is a typical Gnostic twist to orthodox interpretations.

There is more evidence in these Gnostic manuscripts for a greater variety of religious expression and experimentation among Jews than most historians had suspected. For example, rabbinic Judaism propounded a view of man that incorporated a belief in free will and the ultimate moral responsibility of the individual. The members of the sect that produced the Dead Sea scrolls held the opposite view, that man's nature and his fate are predetermined. That view is well illustrated by a passage from the Manual of Discipline found at Qumran: "From the God of knowledge is everything that is and that will be, and before they were, he prepared their complete



KINGDOM OF HEROD (*gray borders*) **from 40 B.C. to 4 B.C. consisted of most of Palestine. Although Herod was king of the Jews, he remained a vassal of Rome. The Samaritans lived near Mount Gerizim and another unorthodox Jewish religious group lived at Qumran.**

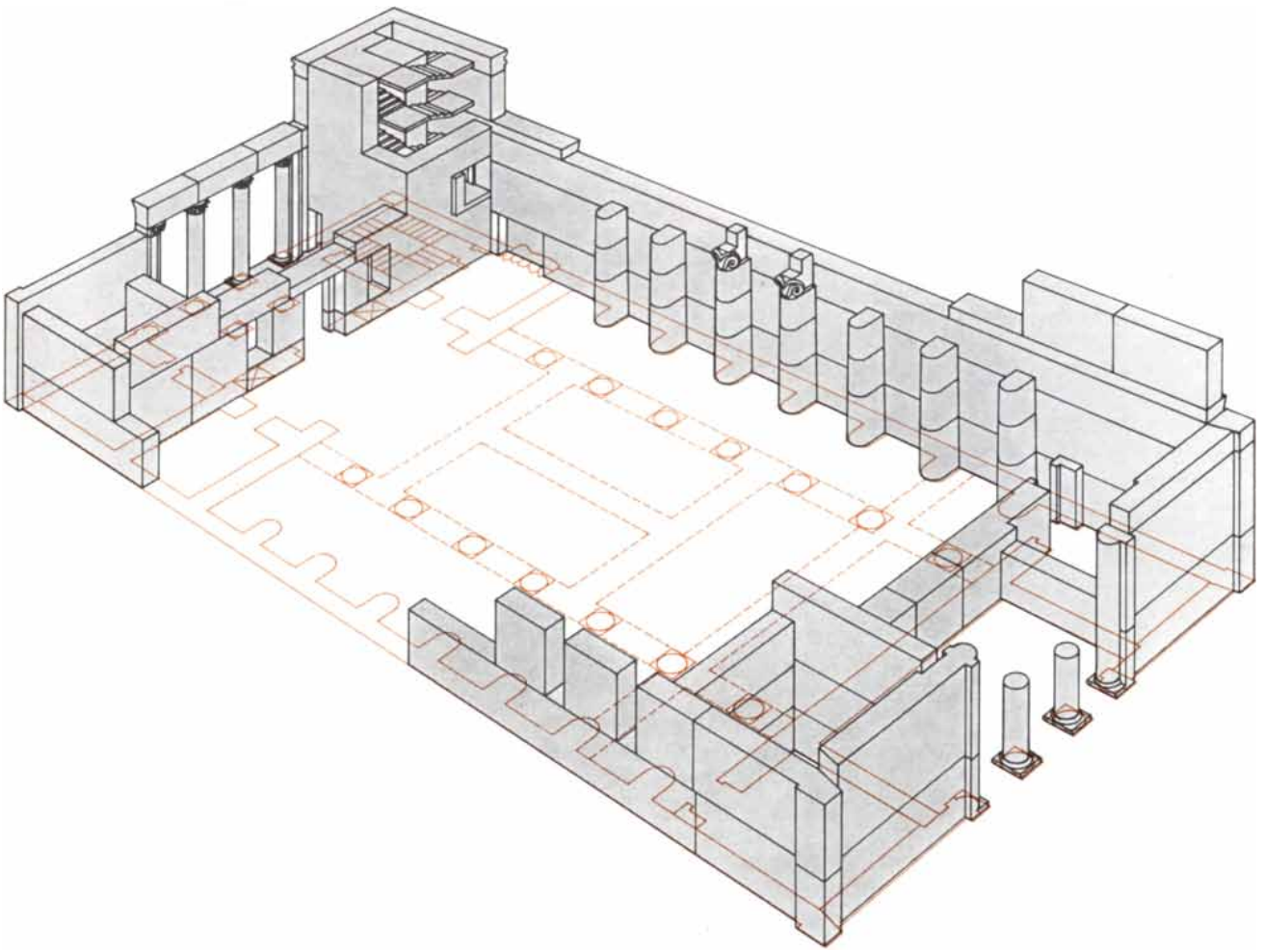
plan and when they are, they fulfill their task according to His glorious plan, and it cannot be changed." These opposing views can only be taken as indicating profound differences in the Judaism of the time.

A view of "original sin" from which some supramundane salvation was required is to be found in *The Fourth Book of Ezra,* a Jewish work written in the aftermath of the destruction of Jerusalem and the Temple by the Romans. Its writer cries in anguish: "O you, Adam, what have you done? For though it was you who sinned, the fall was not yours alone, but ours who are your descendants." Paul's resolution of a similar dilemma is in Christology; the author of *The Fourth Book of Ezra* is in the promised vindication of the righteous at the eschaton. The solutions are structurally very similar. The Gnostics pushed beyond positions such as these. For them the alienation of man from the world is absolute; the world and its creator are demonic. The purpose of man is to release a small morsel of the divine that has somehow become imprisoned in this world and to bring about its reunion with the Godhead, which is totally different from and apart from and unknowable by the world.

The pioneering work of Gershom G. Scholem in tracing the history of Jewish mysticism has highlighted the fact that even within rabbinic Judaism the mystical tradition was active. This tradition received little overt expression, however, in the Talmud and other rabbinic writings. Scholem has shown the lines of connection leading back from the earliest body of Jewish mystical writings, the "Chariot" mystical books of the middle of the first millennium after Christ, through certain hints and allusions in rabbinic sources, to certain of the pseudepigraphical books and the Dead Sea scrolls. New studies are reinforcing the typological similarity that can be observed between the Chariot texts and the Gnostic writings, and there are some indications of actual borrowings and influence.

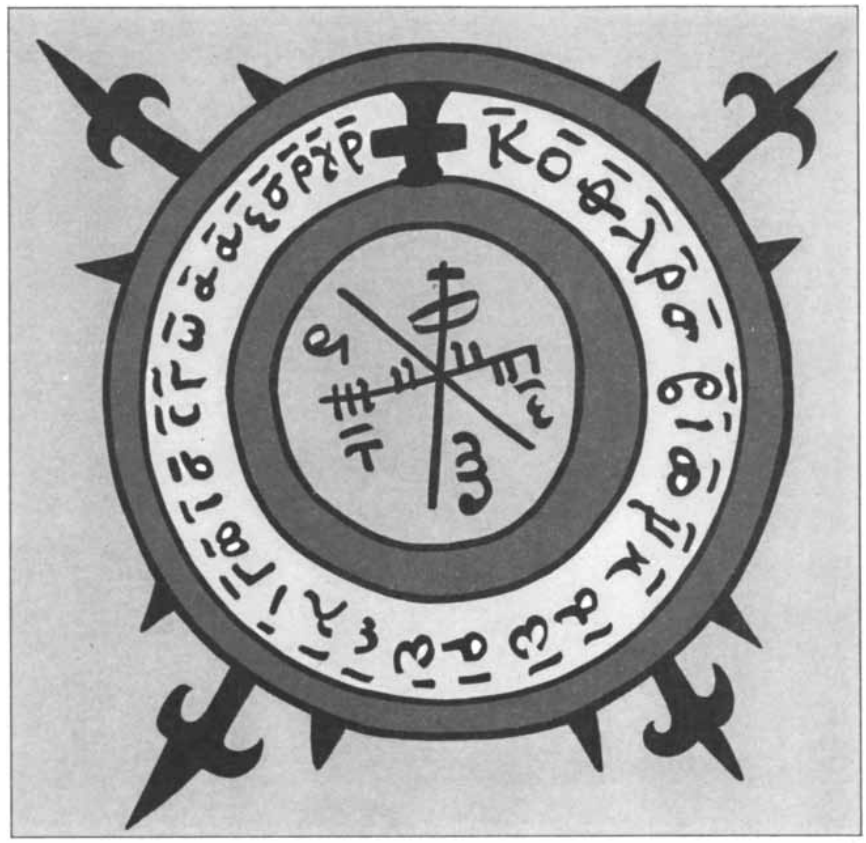
Scholem's work therefore serves to make quite explicit the continuity of the mystical tradition within Judaism from the period of the Second Temple (the Temple destroyed in A.D. 70, which had been built by Herod in 20 B.C.). Moreover, Scholem has shown that this tradition was not alien to rabbinic Judaism. On the contrary, it appears that certain important and central rabbis practiced mysticism. In the course of their speculations they employed a range of terminology that can be traced back as far as the Dead Sea scrolls. This is yet another factor that must be put in the hopper of our evaluation of the religious phenomena of the age.

As a result of their encounter with Greek philosophy many of the Hellenized Eastern cultures attempted to formulate their beliefs as orderly, coherent



TEMPLE AT ARAQ EL-EMIR was built by Hyrcanus, a member of the Tobiad family, early in the second century B.C. The reconstruction is based on findings from excavations of the site sponsored by the American Schools of Oriental Research. The existence of this and other Jewish temples outside of Jerusalem indicates that variety in the expression of Judaism had been permissible.

systems. The allegorical method used by Stoic philosophers was applied to biblical writings, so that the stories of patriarchs became allegories hinting at the true path to be followed. Where Jew encountered Greek, apologetics (often quite aggressive) and the writing of history also flourished. The large Jewish community in Alexandria looked to the stories of Joseph and Moses for legitimation of their position in Egypt. Indeed, Hellenized Jewish writers such as Artapanus attributed to Abraham, Joseph or Moses not only the invention of writing, astronomy and the administrative system by which Egypt was governed but also the founding of the Egyptian cults of animal worship. These cults were abhorred by Jews and Greeks alike.



MAGICAL SIGN associated with the name of Solomon was found in a magic book titled *The Testament of Solomon*. The invocation of Solomon's name has its origins in Jewish magic, where it was used in exorcism of demons who were the cause of human ills.

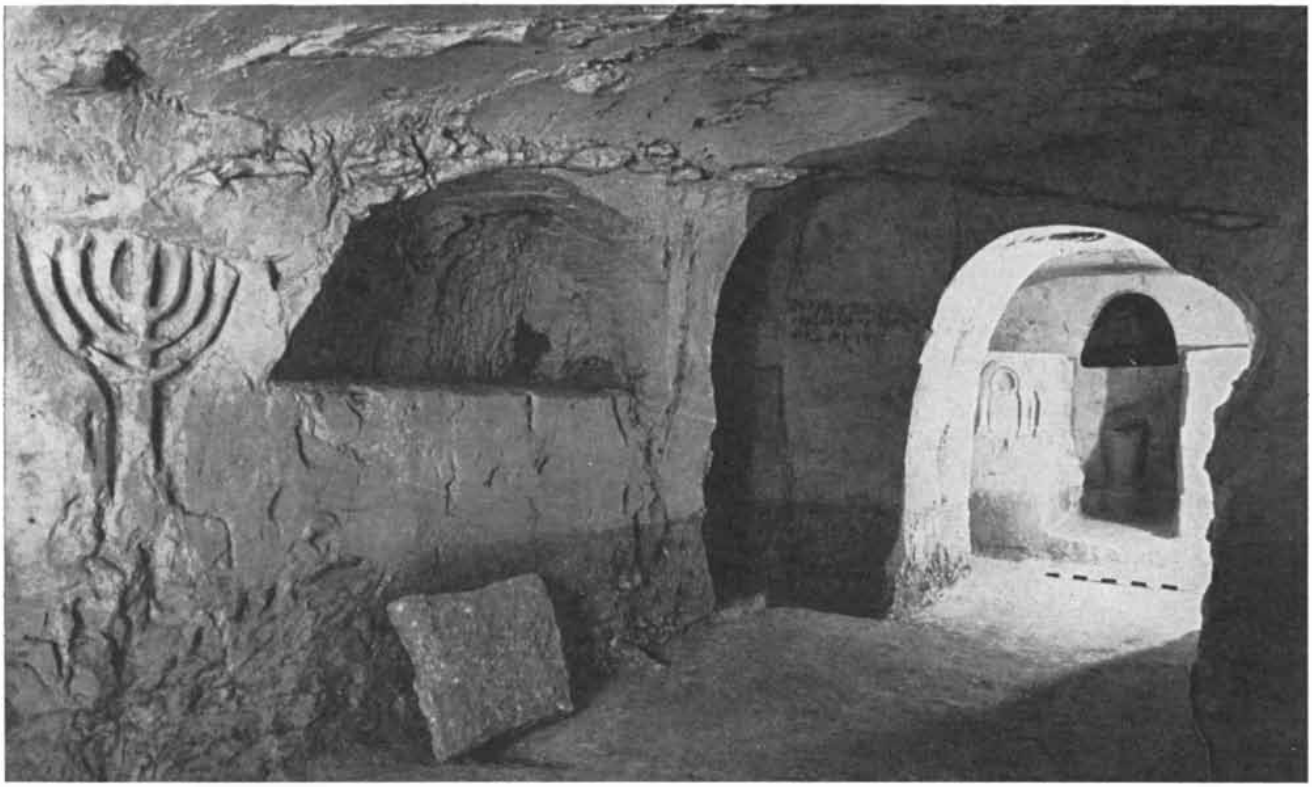
From the time of the reform of King Josiah in Jerusalem (621 B.C.), if not earlier, the predominant view was that the sacrifices to the God of Israel should be carried out only in the Temple in the holy city. According to biblical sources, efforts were made to do away with other temples throughout Palestine. That the Jerusalem Temple alone could serve as the site of Jewish sacrificial cult is also the view of rabbinic tradition. Nevertheless, there is evidence that throughout the entire period of the Second Temple there were temples and sacrifices outside Jerusalem and even outside Palestine. Papyri dating to the fifth century B.C. found at Elephantini in southern Egypt tell of a Jewish military colony with its own temple. The papyri further relate that the temple was destroyed in a pogrom of 408 B.C., and they incorporate a copy of a petition sent by the Jews of Elephantini to the Persian governor of Judea asking for his help in rebuilding their temple. The petition states that a similar letter had also been sent to the governor of Samaria, where another temple of the God of Israel existed at the time. The language of the petition indicates that the authors considered their request, and consequently the existence of their temple, as perfectly natural and legitimate. Yet the cult of the Elephantini temple appears to have been syncretistic, that is, it included deities other than the God of Israel.

It has been maintained that the Jews of Elephantini had a temple outside Jerusalem with a syncretistic cult because they had assimilated pagan practices during their many years of service as mercenaries in the Persian army. Moreover, their willingness to have a temple may have been due to the fact that their ancestors came from an area of the northern kingdom of Israel that had not been strongly affected by the reform of King Josiah. Although this is not a particularly persuasive hypothesis, it may nonetheless be true. What is indubitably true is that the temple of Elephantini is just one of a number of Jewish temples that were in existence throughout the period of the Second Temple.

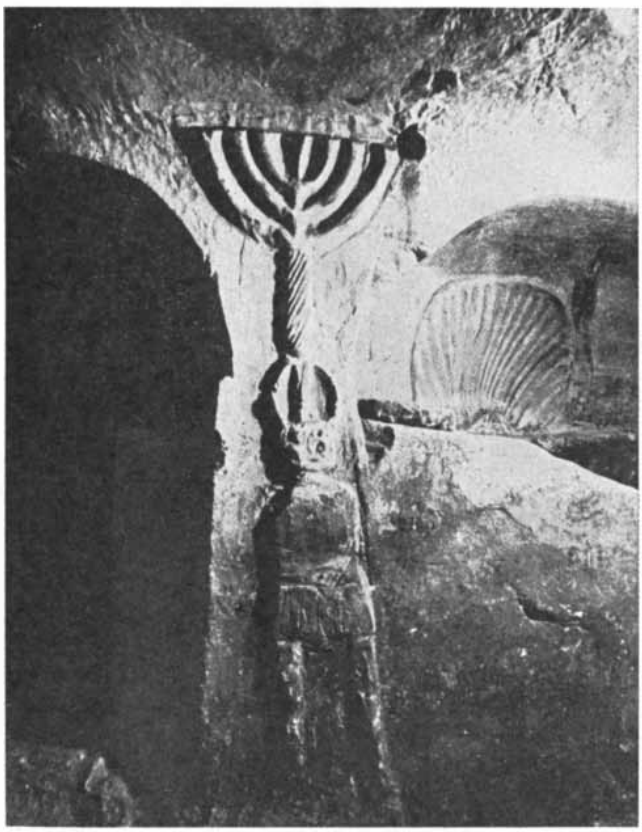
A temple dating to the second century B.C. was unearthed some 10 years ago at Araq el-Emir in Transjordan. The temple had apparently been built by Hyrcanus, who belonged to the Jewish baronial family of the Tobiads. The Tobiads had been particularly important during the preceding century. The events leading to Hyrcanus' withdrawal from Jerusalem to the other side of Jordan are related by the historian Josephus in *Antiquities of the Jews.* Josephus also writes of a temple built in Egypt shortly after this time. The high priest of the Temple in Jerusalem, Onias III, was deposed from office and went to Egypt. There with royal permission he built a temple at a place called Leontopolis. The temple remained in use until it was closed by the Romans shortly after they destroyed the Temple in Jerusalem in A.D. 70.

An account of another instance of a Jewish cult outside Jerusalem is also preserved in Josephus' *Antiquities.* He transcribes a proclamation of the governing body of the city of Sardis in Asia Minor, dating to the first century B.C. This document proclaims among other things that the council of the city permits the establishment of a Jewish cult, apparently a sacrificial cult.

Thus the evidence shows that Jews did in fact build temples and perform sacrificial rites devoted to the God of Israel in places other than the Temple of Jerusalem. Considering the widely held view that Jerusalem held exclusive claim to a temple, a claim established at least four centuries before Hyrcanus and Onias, the very fact that there were other temples should give us pause. Does this mean that the Jewish sources presenting this claim, primarily books in the Old Testament, are merely tendentious, that they represent as fact what was actually the wishful thinking of the



CATACOMBS AT BETH SHEARIM in central Palestine were used by Jews as a burial place from the second to the fourth century after Christ. The long, descending corridors were cut out of solid rock. The walls were decorated with carved religious symbols.



WARRIOR WITH A MENORAH, the candelabrum used in Jewish worship, was found carved on a wall of the catacombs at Beth Shearim. The symbolic meaning may be that the soldier is involved, perhaps in a religious sense, with the cause and hope of Judaism.

TORAH SHRINE at the Beth Shearim catacombs contains a number of religious symbols, including a scallop shell, a lion wearing a harness (*upper right*) and a lulab (*between columns at right*), a bundle of palm branches that was used at the Feast of Tabernacles.

authors? This conclusion would be rather extreme. What is beyond doubt is that the biblical sources do not present a complete view of what was permissible. There was far more variety in Judaism than they make apparent.

Although it is well known that every higher religion has its component of magic, the prevalence of magic in the Jewish religion of the Hellenistic-Roman era is notable. The Book of Enoch attributes the origin of magic to the forbidden teachings revealed by the fallen angels who had intercourse with the daughters of man. The view here, however, is a clear rejection of magical practice. In other groups there seems to have been a strong belief in demons and evil spirits and other aspects of the practice of magic. That is clear from various references in major writings of the period. A well-developed magical tradition was associated with King Solomon, who shackled the demons and forced them to aid in building the Temple. The name of Solomon was invoked frequently in the exorcism of demons, to whom human ills were attributed, and the name is also found in magical signs and amulets of the period.

Recently a significant document, *Sefer Harazim* (*The Book of Mysteries*), was discovered in medieval manuscripts. This book, a magical work written in Hebrew and dated in the third century, provides a link between pagan and Jewish magic. Typical of the pagan-Jewish mixture in the book is a prayer to the Greek sun-god Helios that is composed in the Greek tongue but written in Hebrew letters. Obviously the circles that produced this book viewed pagan elements as being admissible to their teachings and practice.

A comment should be made about Jewish painting and manuscript illumination in the late Hellenistic and Roman periods. From Jerusalem and other Jewish sites in Palestine there is virtually no Jewish representative art from the first century B.C. and the first century after Christ. The reason seems to be the interpretation of the Second Commandment, "You shall make no graven images." This view was not the only one. It contrasts with the situation in the centuries that followed, to which we must date a series of mosaic synagogue floors uncovered throughout the Land of Israel. These floors are covered with various representations, including in a few cases those of pagan deities. In the Jewish cemetery at Beth Shearim in the Jezreel valley many of the sarcophagi are adorned with representative art.

Furthermore, recent researches into illustrated Christian Old Testament manuscripts and the remarkable frescoes in a third-century synagogue at Dura Europus in Mesopotamia have shown that there existed, probably from pre-Christian days, illustrated Jewish biblical manuscripts in Greek. The iconographic and illustrative tradition of the Jewish biblical manuscripts was taken over by Christian art as it developed and thence entered the cultural traditions of Byzantium and Europe. The Jewish manuscripts drew their illustrative techniques and some of their scenes from contemporary Greek illustration of classical works. This tradition is in itself an interesting reflection of varied attitudes toward the Second Commandment. Moreover, it seems to raise broader questions about the early similarities and differences between Palestinian and Diaspora Jewry and on the relationship between Jewish Hellenistic culture and pagan Hellenistic culture.

In the final analysis the crucial issue is likely to be that of the "balance of power," of the relationship of these differing forms (or even categories) of religious expression to one another within the fabric of Judaism. It follows that the central problem will be the overall description of Judaism itself. For the historians of Christianity a better understanding of contemporary Judaism will illuminate the context not only of Jesus' preaching and of his disciples' activities but also of the spread of the new faith in the broader world. For the historian of Judaism, or of religion in general, the task of reevaluation will highlight how religious expression grew in a baroque, almost jungle-like fashion and how experimental and creative it was throughout this period. In the course of his labor the historian will be confronted on every side by the enormously varied and fruitful Jewish religion of the age that was so important for all subsequent generations.

The picture of Judaism I have given is by no means exhaustive, nor is it the result of a single major discovery or a new historical insight. Rather it is a synthesis of a series of discoveries that initially appeared to be unrelated to one another. As patterns of similarity and variation appear, our picture of Judaism in the Hellenistic and early Roman period must be modified. Since Judaism was one of the key elements in the whole of Western culture, our understanding of the relations between the different forms of Jewish religious expression is of major importance.

CONDUCTION ELECTRONS IN METALS

Inside crystals electrons and atoms act collectively to create waves described as quasiparticles. They help to determine whether a crystal is an insulator, a semiconductor, a conductor or a superconductor

by M. Ya. Azbel', M. I. Kaganov and I. M. Lifshitz

Modern technical civilization rests on the skillful exploitation of the electronic properties of metals. These properties determine not simply the electrical characteristics of metals but virtually the entire range of properties we associate with the metallic state. Although artisans and practical men have known for centuries how to manipulate specific features of the metallic state, it is only within the past few decades that anyone has been able to explain why different metals have different properties. The differences can be traced to differences in their energy structure, a concept introduced by quantum mechanics.

Just as the properties of various atoms differ because of differences in their energy structure, so can the differences among metals be traced to analogous differences in energy structure. In atoms and molecules the energy structure is determined by the energy levels occupied by individual particles: the electrons. Quantum mechanics shows that the energies of the electrons cannot assume arbitrary values but only strictly prescribed ones [*see illustration on page 91*]. In metals the energy structure is also determined by particles, but not by particles acting as they do in individual atoms. Inside the crystals that make up the metal, quanta of energy are associated with electrons and atoms acting collectively, generating waves that travel through the entire crystal. Such quanta, which behave in many respects like ordinary particles, are called quasiparticles. They are the elementary carriers of motion in a system of interacting atoms. Only in a gas do the constituent particles (atoms and molecules) serve both as structural units and as carriers of motion.

The nature of quasiparticles may seem less forbidding if we pause to consider how the energy structure of a metal can in principle be understood at all. How is it possible to construct a consistent theory of motion in view of the fact that each cubic centimeter of a metal contains some 10^{23} electrons, moving at high velocity and interacting not only with one another but also with the ions (atomic nuclei lacking a full complement of electrons) that constitute the regularly repeating centers of positive charge in the crystal lattice? It is clearly impossible to describe the motion of every electron. Fortunately it is also unnecessary, since we are interested only in the characteristics and properties of the metal crystal as a whole. In fact, the average behavior of an aggregate is always much easier to predict than the behavior of each of its members. This is the basis of all statistics, particularly of physical statistics (which is usually called statistical physics). When the statistical approach is adopted, an aggregate with a huge number of members is a distinct advantage; the greater the number of members is, the more reliable the predicted results are.

One of the fundamental conclusions of the application of quantum mechanics to liquids and solids in general, and to metals in particular, is that the energy spectrum of a macroscopic body (provided that the temperature of the body is near absolute zero) is similar to the energy spectrum of a gas of elementary particles. This means that the energy of the liquid or the solid is the sum of the energies of the individual simple motions of the particles. Moreover, the energy of these motions is quantized, which means that it can assume only certain distinct values. In metallic systems, however, we are concerned not with the simple motion of particles but with the collective motion of aggregates of particles.

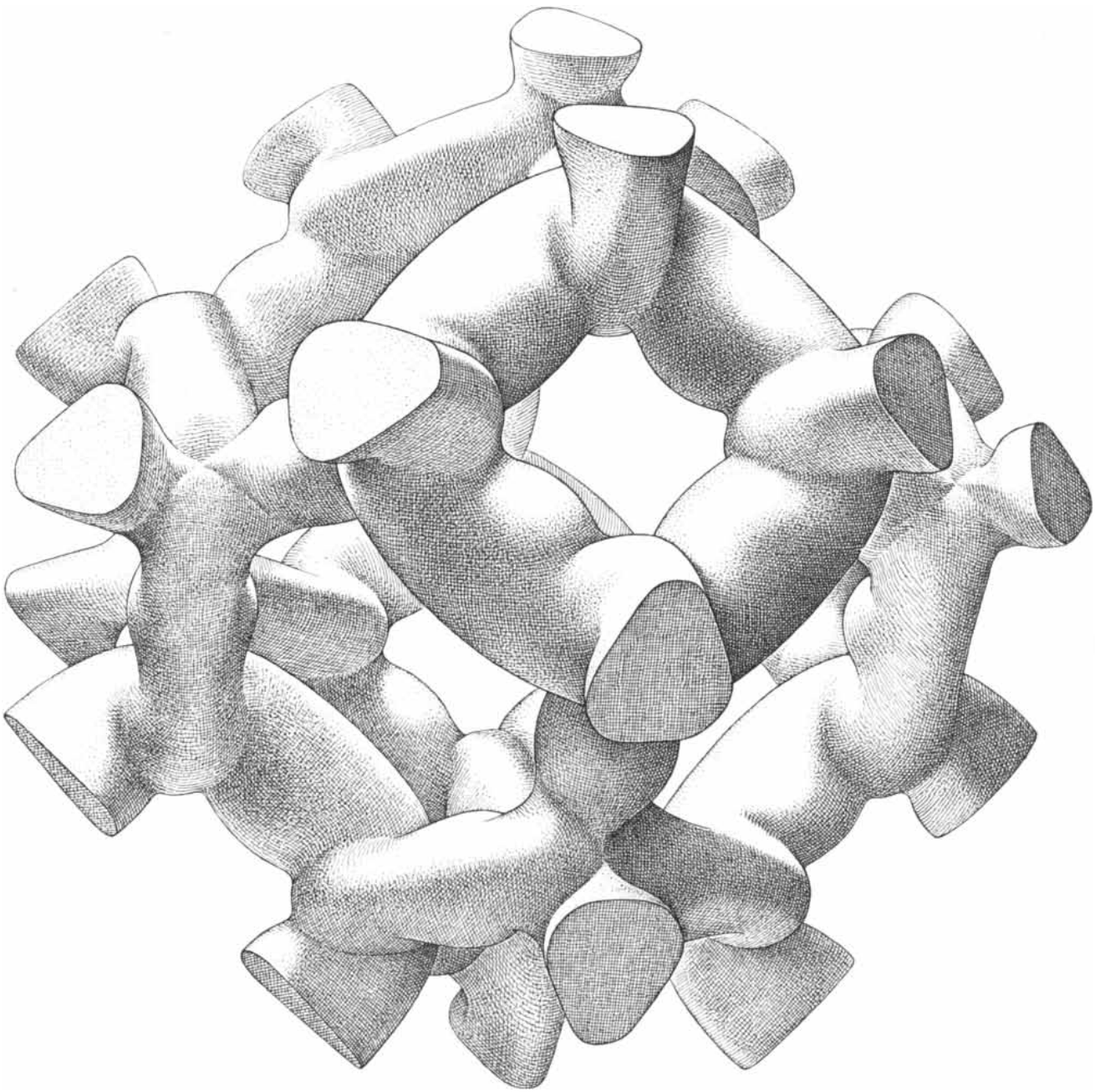
A fundamental property of a crystal is that its constituent atoms occupy identical positions in a periodic, symmetrical structure, the smallest repeating unit of which is called a cell [*see top illustration on page 90*]. In each unit cell similarly placed atoms have an identical environment and therefore interact identically with each of their neighbors. The atoms in a crystal can be regarded as oscillating circuits tuned to identical frequencies and capable of resonating with one another. Any excitation of one atom therefore causes an analogous excitation in neighboring atoms. The result is that excitations do not stay in one place but move as a wave through the entire crystal. The motions connected with these waves arise and are transmitted only in the form of individual packets, or quanta. These quanta are the quasiparticles.

In a crystal there can be many kinds of quasiparticles, generated by many kinds of collective motion. For example, waves of elastic crystal vibrations generate quanta of sound called phonons. Waves produced by oscillations of magnetic moment in ferromagnetic and antiferromagnetic materials (spin waves) generate quasiparticles called magnons. The quasiparticles that carry electric charge in metals are the conduction electrons. Although their charge is the same as the charge of the free electron, the charge carriers are typical quasiparticles with properties distinctly different from those of the electron itself. The diversity of the dynamic properties of quasiparticles is even greater than the variety of their types. In studying quasiparticles it is important to have a clear-cut understanding of the type of data that must be gathered and systematized.

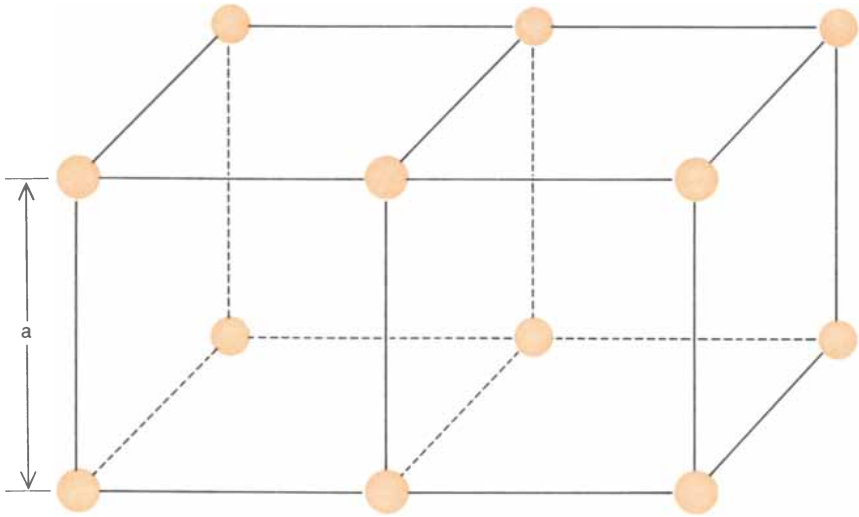
Thus we must draw some further distinctions between quasiparticles and ordinary particles.

The elementary particles of matter—electrons, protons, neutrons, photons, neutrinos and so on—exist in a vacuum: "empty space" that is homogeneous and isotropic (free of directional properties). Therefore in spite of their many differences in fundamental properties the elementary particles exhibit a high degree of uniformity in their dynamic properties. These properties are determined by the dependence of the particle's energy, E, on the particle's momentum, p. This dependence is expressed in the dispersion law, which for any particle possessing a mass, m, and moving at low velocity, is given by the formula $E = p^2/2m$ [*see illustration on page 92*]. The dispersion law is deeply connected with the geometry of the world, that is, with the homogeneity and isotropy of the vacuum.

The "vacuum" in which the quasiparticles exist is far more complex. The atoms embedded at specific sites in the lattice of a crystal make the space inhomogeneous and anisotropic (having directional properties). This greatly complicates the relation between the energy and the momentum of a quantum. Therefore the first task in clarifying the energy spectrum of a crystal is to determine the dispersion law for quasiparticles. One can try to calculate the dispersion law; the main forces acting between the particles (between electrons and ions and between electrons and electrons) are known, and the main laws governing the motion of atomic particles



FERMI SURFACE is an imaginary three-dimensional surface that represents the motions and energies of the electrons in a metal. Specifically a Fermi surface is a surface in momentum space below which all the electron states in a crystal are occupied at absolute zero and above which they are empty. This drawing represents a segment (the "third zone") of the Fermi surface of the metal lead.



ATOMS IN CRYSTAL are arranged in a three-dimensional strict order. There are many possible configurations; the simplest crystals have unit cells in the form of a cube. The distance between neighboring atoms (a) in a crystal is typically from one to five angstroms.

have been established. The calculations are so cumbersome, however, that various simplifying assumptions have to be introduced.

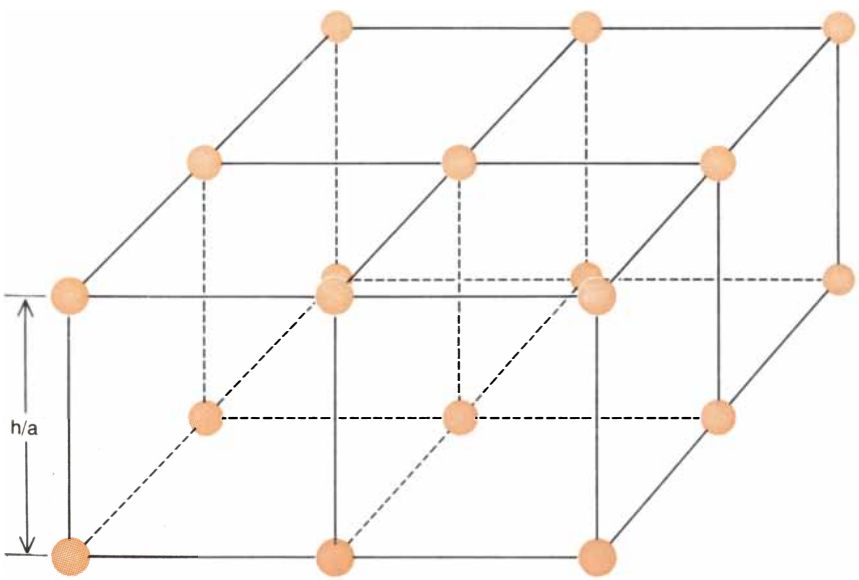
An approach that is generally more successful has been to allow theory to be guided by experimental results. Starting from the most general concepts of quantum mechanics, physicists have constructed a mathematical theory of those phenomena in crystals and properties of crystals that are sensitive to the crystal's energy structure and in particular to the dispersion law for quasiparticles. By comparing the predictions of the theory with experimental results it is possible to reconstruct the energy structure of crystals from the dispersion laws.

Let us consider in somewhat greater detail what is meant by the energy structure of a crystal. The term "structure" is particularly appropriate. The energy spectrum is described with the aid of geometric images. Textbooks on the electronic theory of metals have come to resemble catalogues of abstract art or portfolios of ultramodern architecture.

The geometric images describing the energy spectrum of crystals are somewhat difficult to understand since the constructions exist not in ordinary space but in momentum space, where each point corresponds not to a geometric coordinate of the particle but to its momentum. Thus we must translate temporarily to momentum space. To construct the energy structures we first introduce a kind of structural lumber. This is done by breaking up the space into identical cells with dimensions inversely proportional to the dimensions of the cells of the crystal whose properties we wish to describe [*see bottom illustration on this page*]. To be rigorous we should explain that the state of a quasiparticle in a crystal is characterized not by momentum but by a very similar quantity: "quasimomentum." Strictly speaking a crystal can be represented by a single cell, since all the quantities that depend on the quasimomentum are periodic functions of the quasimomenta in one cell. Experience has shown, however, that it is more convenient to use an infinite momentum space, in which one is able to take the periodicity into account directly.

The energy of a quasiparticle in a crystal is a complicated periodic function of its (quasi)momentum. It is customarily represented in the following manner. A surface is drawn through all the points of momentum space where the energy equals some selected value. Similar equal-energy surfaces are drawn for other values of energy. For an electron in free space the equal-energy surfaces are spheres whose radii increase in proportion to the square root of the energy. In a crystal the way the shape of the surface varies with energy determines the dispersion law for quasiparticles.

When the crystal is in the ground state (that is, at absolute zero), there are no regular vibrational motions and therefore no quasiparticles at all. They appear when energy is put into the crystal: when the crystal is heated, exposed to light or bombarded with particles. The statistical properties of the aggregate of quasiparticles are closely connected with the conditions for their appearance (creation) or disappearance (annihilation). Quasiparticles appear and disappear either singly or in pairs. Those that can appear singly are described by Bose-Einstein statistics and hence are



CONCEPT OF MOMENTUM SPACE must be used in discussing the energy structure of crystals. In momentum space the cells of a crystal have dimensions inversely proportional to those of the actual crystal. Thus a primitive cell of dimension a is transformed in momentum space into a cell with each side equal to h/a, where h is Planck's constant.

called bosons. Those that can appear only in pairs follow Fermi-Dirac statistics and are called fermions.

Phonons and magnons are typical examples of quasiparticles that are bosons. Because bosons can appear and disappear singly their dispersion law can be measured directly by applying the conservation laws of energy and momentum. The most widely used method of investigating the boson part of the energy spectrum of crystals is to bombard a crystal with neutrons and observe the relation between the momentum and energy transferred to the crystal and the momentum and energy lost by the neutrons (a phenomenon called inelastic scattering). The momentum and energy absorbed by the crystal go into the creation of bosons.

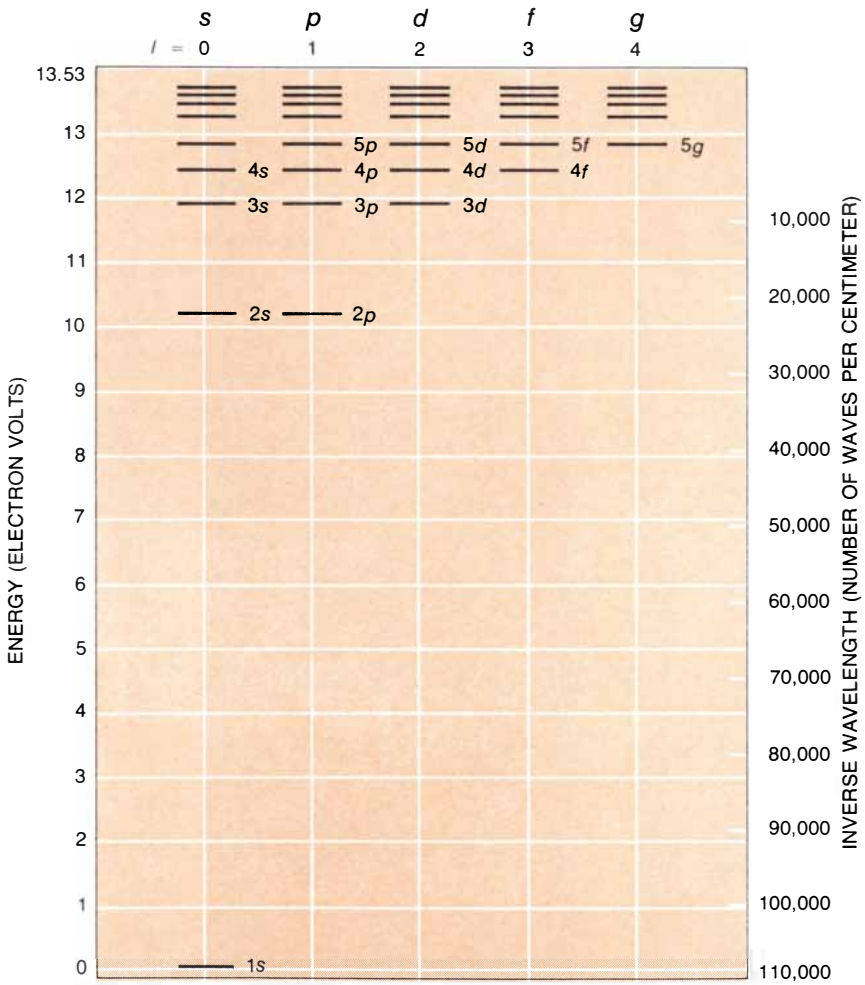
Fermions are created primarily by the motion of electrons within crystals. From this point on we shall restrict our discussion to these charged quasiparticles. The electronic properties of crystals are connected, however, not only with the motion of the quasiparticles but also with the nature of the ground state of the crystal: the "vacuum" in which the quasiparticles are produced and move. In quantum physics the ground state is not a state of absolute rest or motionlessness but only a singular "zero point" form of motion in which there are no quasiparticles. At a sufficiently low temperature all crystals, depending on the type of the electronic ground state, can be divided into four classes: dielectrics (insulators), semiconductors, metals of normal conductivity and superconductors.

Leaving aside the special case of superconductors, we shall try to describe, using a highly simplified model, the electronic ground state of the crystal and thereby explain why some substances (metals) conduct current near this state and others (dielectrics) do not. In all nonsuperconducting substances all the qualitative consequences of such a model remain valid in the actual substance and explain correctly the essence of the matter.

In free space the state of the electron is specified completely by its momentum; in a crystal that is not the case. In order to characterize the state of the electron in a crystal one must specify in addition to the momentum some integer called the zone number. For each zone number it is necessary to introduce a separate momentum space to construct equal-energy surfaces. If we ask what values of energy an electron can assume in a crystal, we find that only certain energy intervals are allowed and that these intervals are separated by forbidden bands [*see top illustration on page 93*]. Each allowed interval corresponds to a definite zone number and is named for the zone. The allowed energy bands of the crystal are a direct consequence of the allowed energy levels of the electron in the atom. The bands may overlap but they retain their individuality because each band possesses its own momentum space.

We now populate the momentum space of the *n*th band with true electrons, not quasiparticles, which are always present in the crystal. In the ground state the electrons naturally occupy the state with the lowest energy. As the Pauli exclusion principle states, however, no more than two electrons can be in each state. Therefore one cell of momentum space is filled completely by electrons whose number is just twice the number of cells in the real crystal. Filling this one cell of momentum space with electrons is equivalent to filling all of momentum space, meaning that all the states in the band are occupied. The total number of electrons in the crystal is, of course, the sum of all the electrons carried by all the atoms making up the crystal.

If all the allowed energy bands in a crystal in the ground state are filled or empty, the crystal is a dielectric—an insulator. If one or more allowed bands are only partially filled, the crystal is not only a conductor but also a metal



ELECTRONIC STRUCTURE of an atom is represented by the energy levels its electrons can occupy. According to quantum theory the levels can assume only discrete values. The electronic structure shown here is for the hydrogen atom. As the energy is raised the permitted levels lie closer and closer together. When the electron's energy exceeds 13.53 volts, however, the electron can no longer be held by the positive charge on the hydrogen's nucleus and escapes. As the angular momentum, *l*, of the electron increases in integral steps the minimum energy jumps upward one level at a time. In falling from the highest level to zero energy an electron emits ultraviolet radiation with a wavelength of 916 angstroms, which corresponds to not quite 110,000 waves per centimeter, as shown in the scale at right.

[*see bottom illustration on opposite page*]. The flow of current through a metallic conductor is actually a complicated process in which the electrons gain infinitesimal packets of momentum as a result of the force exerted by the electric field; they lose this momentum when they collide with the irregularities of the crystal lattice. The process is impossible, however, when the allowed bands are completely filled with electrons; the exclusion principle prevents the electrons from taking up a new position within the band, and the existence of forbidden bands prevents any interaction between the electrons of highest energy and the electric field. In metals, on the other hand, energy can be transported by electrons with near-maximum energy that are free to occupy empty states of still higher energy.

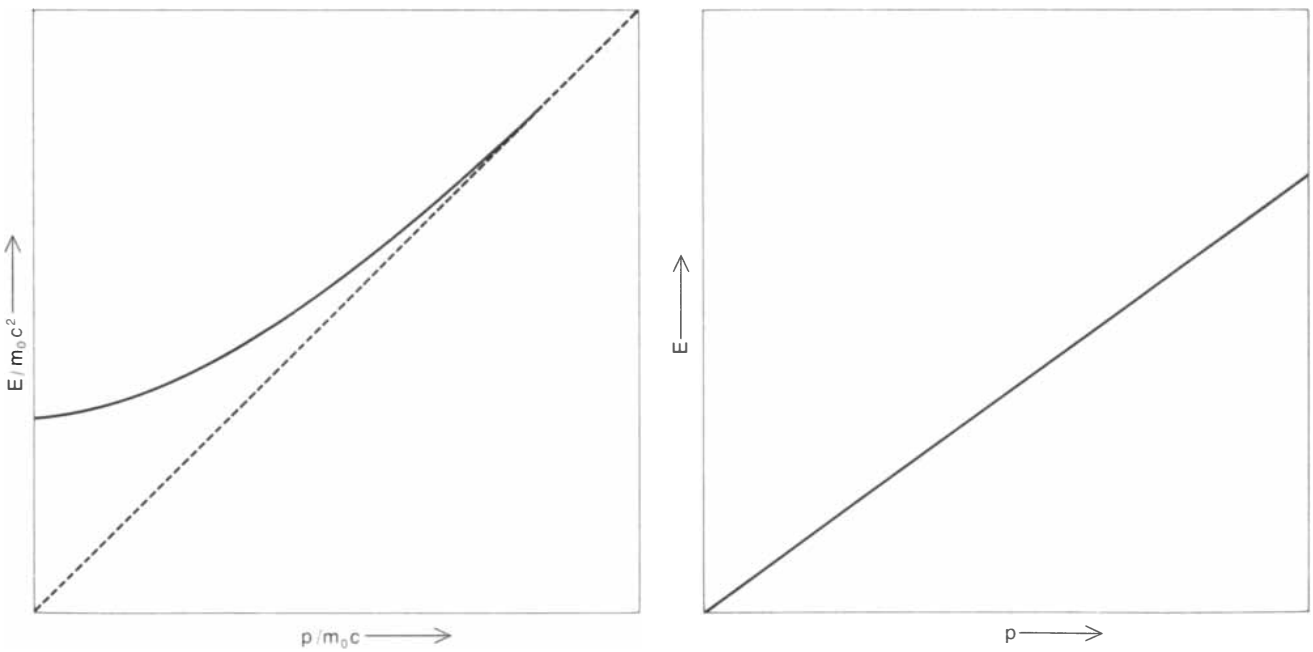
The energy below which all the electron states are occupied at absolute zero and above which they are empty is called the Fermi energy, and the corresponding surface in momentum space is called the Fermi surface. Thus a metal in the ground state can be represented by a quaintly shaped vessel everywhere filled with electrons. In the ground state there are no electrons outside the vessel.

Now that we have described the ground state of the metal we are ready to ask: How is the energy of electrons in metals increased? How are the quasiparticles created? The energy can be increased only by moving at least one electron from under the Fermi surface into the region of momentum space that is free of electrons. When this is done, a "hole," or empty space, is created in the previously filled energy band. The hole can be regarded as the antiquasiparticle of the newly "created" electron. By describing a hole as an antiparticle we emphasize the possibility of annihilation; when the electron returns "to its place," the metal returns to the ground state and two quasiparticles—an electron and a hole—vanish.

In dielectrics as the temperature is raised a few electrons acquire enough energy to escape from a filled band into an empty one, traversing a forbidden band in the process; as in metals, a hole is left behind. In certain crystals the forbidden band separating the filled band from the free one is quite narrow, so that at ordinary temperatures a substantial number of electrons are able to move into the free band and become charge carriers, again leaving holes behind. Such crystals are semiconductors, of which germanium and silicon, used in transistors, are the best-known examples. In metals that are good conductors even a significant rise in temperature causes little change in the distribution of the electrons in momentum space. Therefore the Fermi surface and its immediate vicinity—the region of space in which quasiparticles (electrons and holes) exist—become of fundamental importance.

To find the change in particle energy produced, say, by a rise in temperature, one multiplies the particle velocity by the change of momentum. An adequate investigation of the electronic energy spectrum, sufficient for a complete description of the properties of the metals, calls for a determination of the shape of the Fermi surface and of the electron velocities at points on the surface. Complicated Fermi surfaces are sometimes called "monsters." If the electron velocity is also represented on the Fermi surface, the resulting picture is that of a monster with projecting needles [*see illustration on page 94*].

Because electrons and holes, like all fermions, appear and disappear in pairs, one cannot accurately determine the dispersion law for the electrons in metals by such methods as the inelastic scattering of neutrons. The reason is that the equations that determine the change in momentum and energy of the scattering particle contain too many unknowns, namely the momenta and energies of both the electron and the hole. It was therefore necessary to develop other principles for determining the electron-dispersion laws, based mainly on the dis-



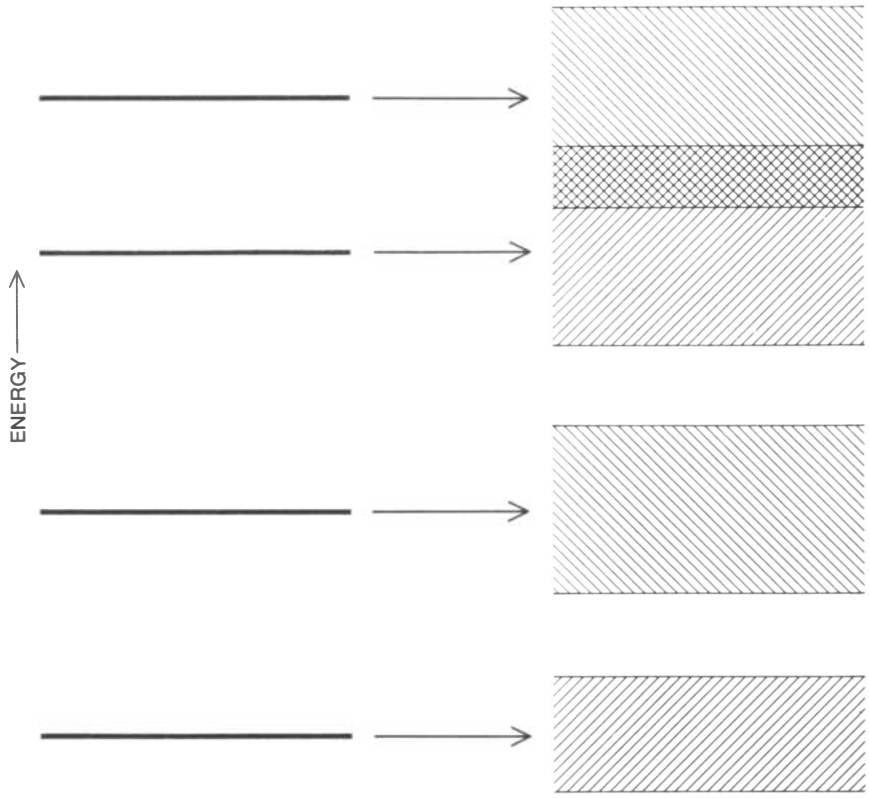
DYNAMIC PROPERTIES of a particle are a function of its energy, E, and its momentum, p, as expressed in the dispersion law. For a particle whose rest mass is greater than zero the dispersion law is written $E^2 = m_0^2c^4 + c^2p^2$, where m is the rest mass and c is the velocity of light. The equation produces the curve at left. When the particle's rest mass is zero, which is the case for the photon and the neutrino, the first term drops out and E simply equals cp, the momentum times the velocity of light, providing the curve at the right.

tinguishing features of the mechanics of the conduction electrons.

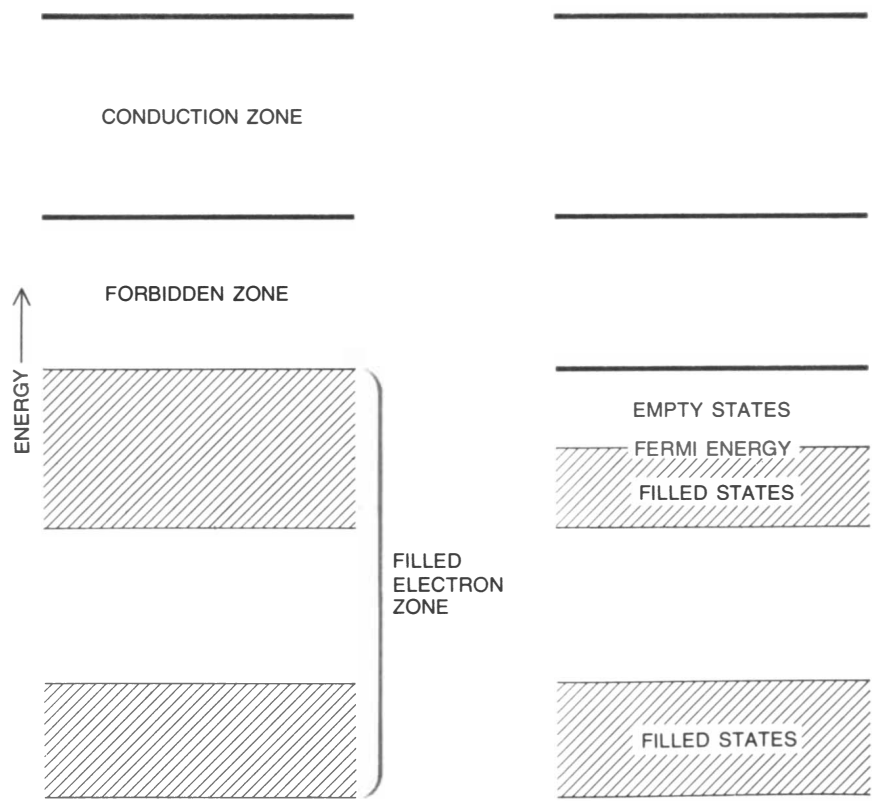
The mechanics of quasiparticles with a complex dispersion law is unique, and the motion of quasiparticles in external fields is usually quite unlike the motion of their "ancestors": electrons in free space. For example, from the fact that the energy is periodic as a function of the momentum it follows that the particle will execute oscillations in a constant and uniform electric field, rather than going off to infinity as one would normally expect [*see top illustration on page 96*]. Although this example demonstrates correctly the unique character of the quasiparticle motion, it is too idealized; we did not take into account the collisions among the electrons and the distortions of the crystal lattice, particularly the unavoidable distortions due to thermal motion of the component ions. Under real conditions in good conductors the influence of the electric field on the motion of the charge is small because it is impossible to produce a strong electric field; if one tried to produce such a field, the amount of heat released would melt the metal.

The unique motion of quasiparticles raises two closely related questions: How does this motion affect the electric properties of metals, and how is it possible to investigate in general the dynamics of the conduction electrons, that is, their dispersion law? The latter question can obviously be solved only in connection with the former. After all, it is impossible in principle to extract the quasiparticle from the crystal and investigate its properties, since it represents a collective excitation of the entire crystal. The natural way, as we have mentioned, is to establish theoretically the connection between the dynamics (dispersion law) of the conduction electrons and the various macroscopic (electric and magnetic) properties of the metal and then to use the measurement of these properties as indicators of the electronic spectrum. A similar approach to the solution of the problem of the electronic energy spectrum of metals was developed starting in 1950 by a group of theoretical physicists at the Kharkov Physical-Technical Institute of Low Temperatures in the U.S.S.R.

It turned out that the properties of a metal that are most sensitive to the dispersion law are the properties altered by a strong magnetic field. The trajectory of the conduction electron in a constant magnetic field is determined by the curves produced when the Fermi surface intersects the plane that is per-



ELECTRON ENERGY LEVELS IN CRYSTAL are broad bands (*right*), which can be separated by forbidden zones or can overlap. The allowed energy bands can be regarded as the sum of the allowed energy levels in the individual atoms in the crystal (*single lines at left*).



CONDUCTORS AND NONCONDUCTORS are distinguished by the way the bands in their energy structure are populated. In a nonconductor, or dielectric (*left*), the bands are either completely filled or completely empty when the crystal is in the ground state. In a metal, or conductor (*right*), one or more of the energy bands are only partially filled with electrons.

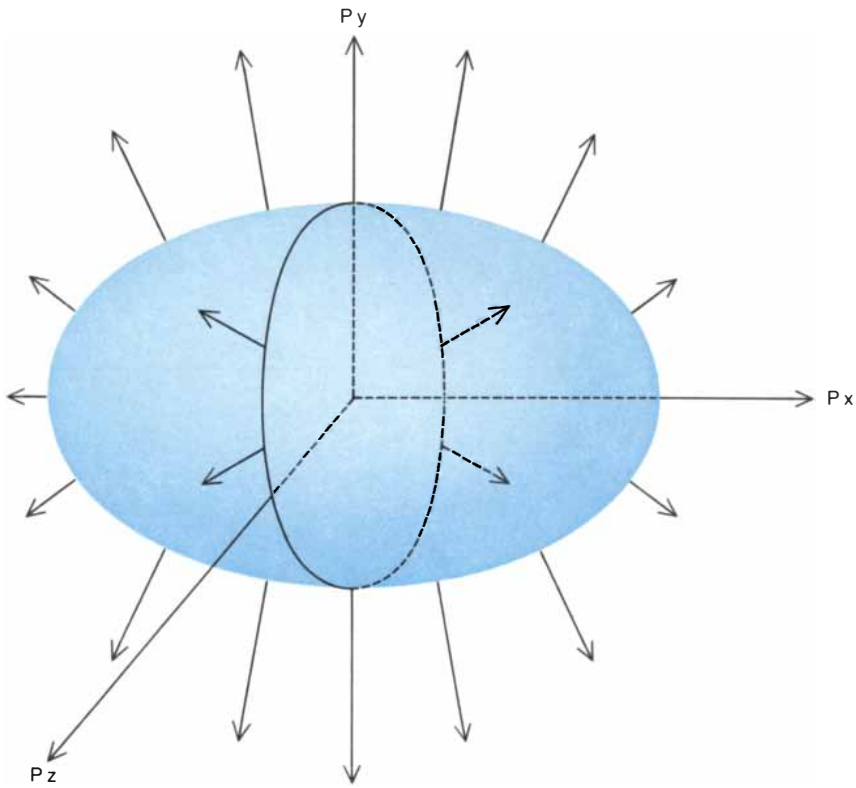
pendicular to the direction of the magnetic field [*see top illustration on opposite page*]; the direction of motion is determined by whether the area of the intersection increases or decreases with increasing energy. If the area of the intersection increases [*see curve "a" in bottom illustration on opposite page*], the quasiparticle moves in the same direction as the free electron would, and if the intersection area decreases [*curve "b"*], it moves in the opposite direction, that is, in the direction that would be taken by a positive charge.

In the case of a closed Fermi surface the character of quasiparticle motion is the same for all the quasiparticles on the surface and does not depend on the direction of the magnetic field. All quasiparticles move either like ordinary negative charges (if the surface "swells" with increasing energy) or like ordinary positive charges (if the surface "contracts" with increasing energy). In the case of a conduction electron a reversal in the direction of rotation, depending on the character of the Fermi surface, is customarily described not by changing the sign of the charge but by changing the sign of the electron mass. According to this convention an electron that is rotating counter to its expected, or normal, direction is assigned a negative mass. If the dispersion law is complicated, mass ceases to play a universal role. Various quantities with the dimension of mass appear in the various phenomena. Therefore the coefficient of proportionality between the force and the acceleration is not equal to the coefficient of proportionality between the momentum and the velocity. When an electron moves in a closed orbit, the period of its motion is expressed by the same formula that holds for a free charge, but the role of the free particle is assumed by what is called effective mass. The effective mass is determined by the rate with which the cross-sectional area changes with energy, and if the cross section decreases with increasing energy, this mass is negative.

Trajectories that intersect themselves are in a special class. When such trajectories are approached, the effective mass of the electron, meaning also the electron's period of revolution, tends to go to infinity. The particle's velocity along its orbit decreases on approaching the self-intersection point, so that the particle takes an infinite time to "reach" that point.

If the Fermi surface is not closed, the character of the motion of the charge depends strongly on the direction of the magnetic field. For closed cross sections the motion remains periodic; however, the direction of motion of the electron depends essentially both on the orientation of the magnetic field and on the position of the intersecting plane. If the orbit is open [*see right half of top illustration on opposite page*], the motion in the corresponding direction is in general aperiodic, and the charge goes off to infinity. Therefore when a conduction electron moves in a constant magnetic field, its trajectory is directly connected with the main characteristic of the quasiparticle: its dispersion law.

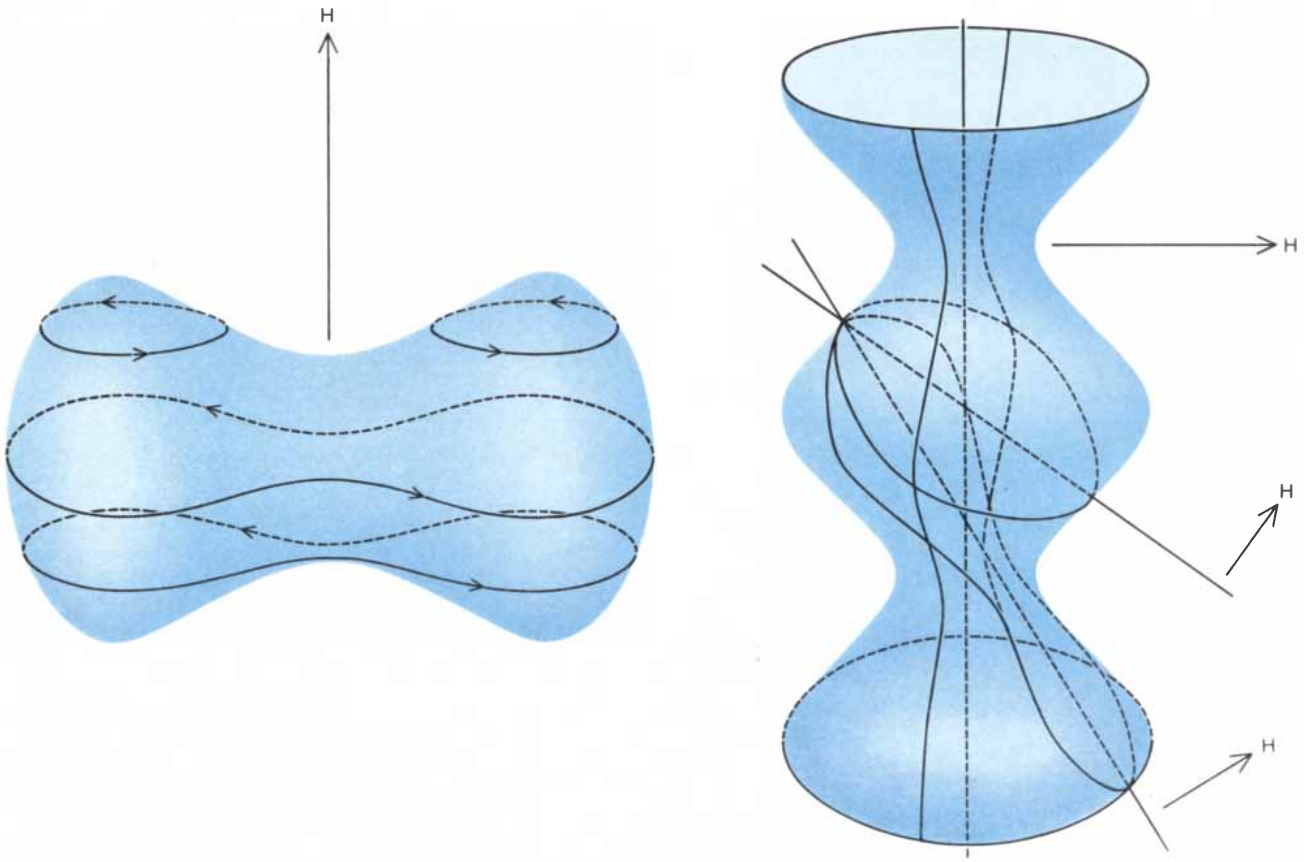
Of course, if the picture described is to be more than a simple idealization, it is necessary that the character of motion along the orbit have time to establish itself, that the quasiparticle be not "driven off the road" as a result of collisions. In other words, the collisions must be sufficiently rare. This means that the period of motion in the magnetic field should be appreciably shorter than the free-path time: the average time between successive collisions. Since the period of revolution of the electron is inversely proportional to the field, the field must be sufficiently strong. For the same reason it is desirable to have the free-path time as long as possible, meaning that the experiment must be conducted with samples of the highest possible purity and at temperatures as low as possible. (The lower the temperature is, the weaker the random thermal vibrations are that violate the ideal periodicity of the crystal.) All these conditions are readily satisfied in magnetic fields that have values of the order of 1,000 oersteds.



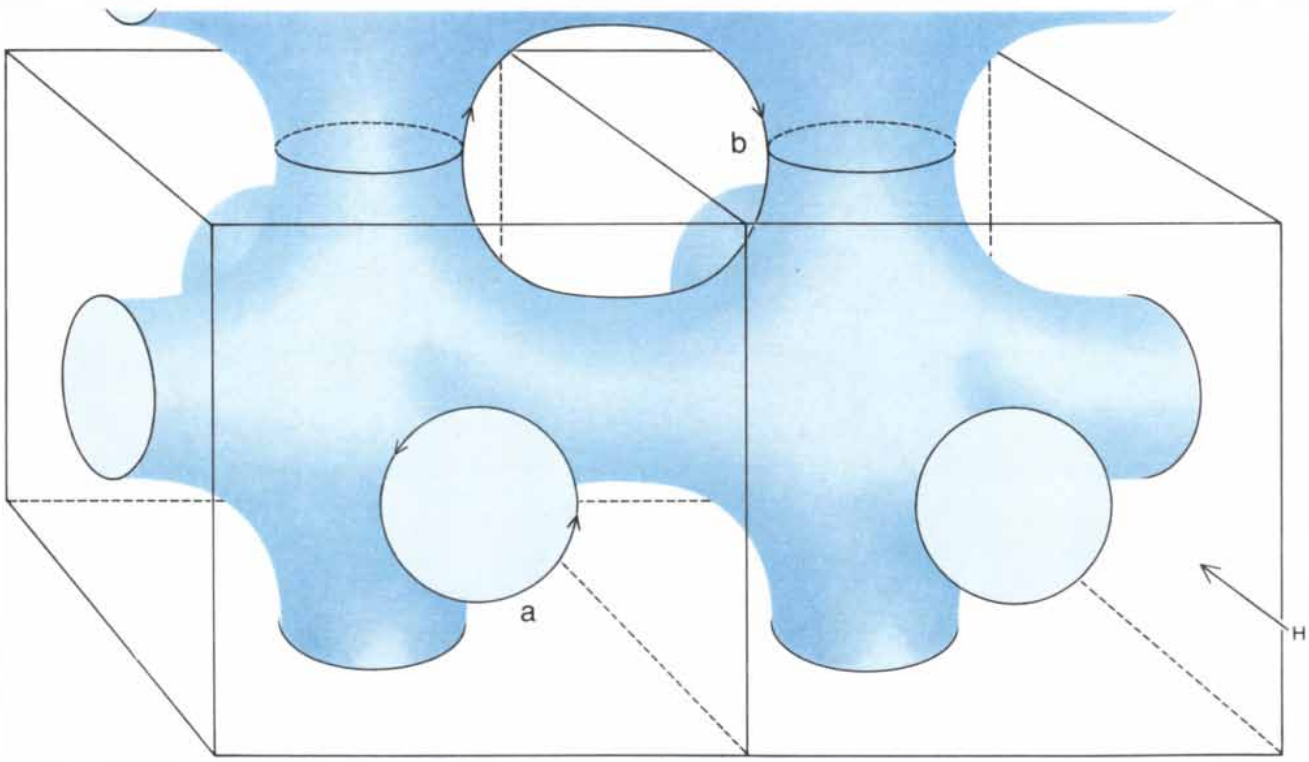
SIMPLE FERMI SURFACE is an ellipsoid. More complicated surfaces, referred to as "monsters," are shown on pages 89 and 98. In the ground state all of a crystal's electrons are contained within the Fermi envelope. The velocity vectors of electrons at the surface can be represented by arrows, or needles, to show how the "pressure" varies below the Fermi skin.

How can we probe the motion of the quasiparticle along its orbit in a real experiment? How can we "see" this motion in the crystal? The analogy with ordinary vision is appropriate. When we see some object, it means, on the one hand, that the object interacts with the rays of light so that the distribution of the light rays is altered and, on the other hand, that the interaction with the light is so weak that it practically does not affect the state of the object.

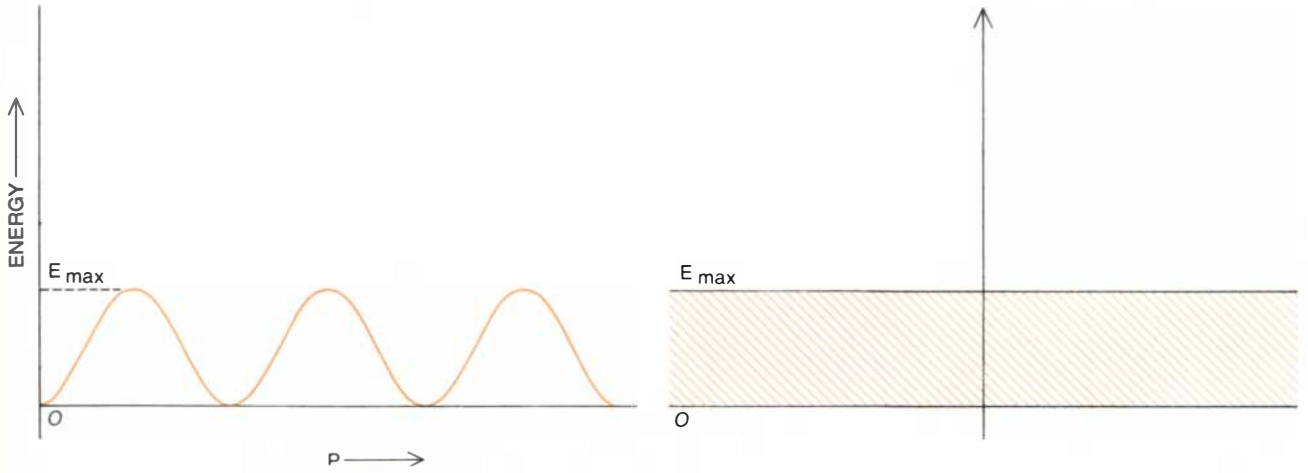
In the study of the conduction electrons one can use a variety of "light rays": ultrahigh-frequency radio waves,



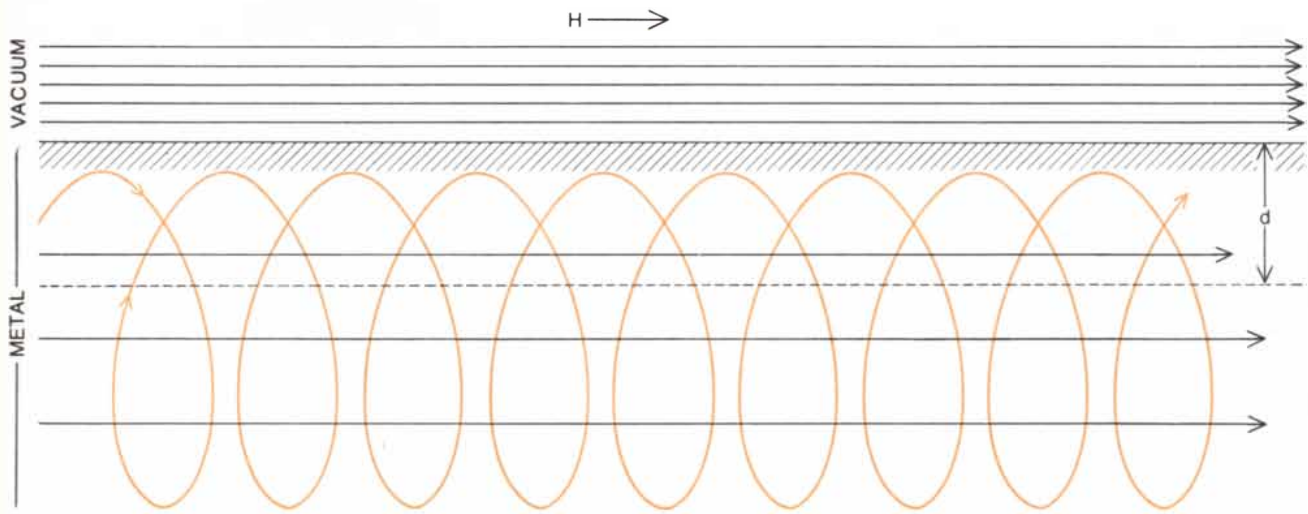
TRAJECTORY OF ELECTRON with a complicated dispersion law in a constant magnetic field, H, depends strongly on the form of the equal-energy surface, which is closely related to the Fermi surface. The trajectories of electrons are always perpendicular to the magnetic field. When the equal-energy surface is closed (*left*), the trajectories also remain closed no matter how the magnetic field is rotated. When the surface is open (*right*), the electrons can go off to infinity at certain orientations of the magnetic field.



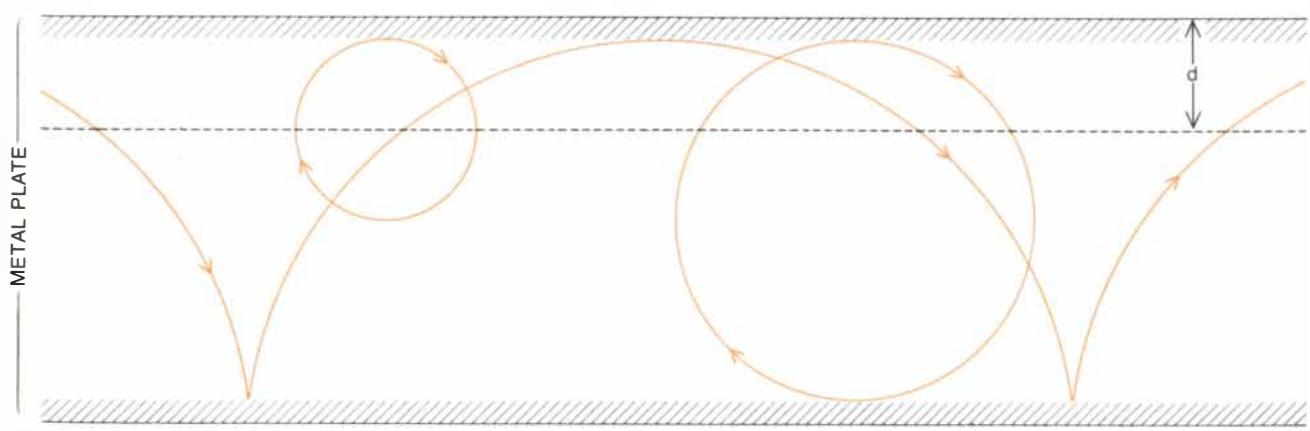
WHEN FERMI SURFACE IS A MONSTER, that is, complex, electron trajectories take several forms. With increasing energy the tubes that form the surface become thicker. As a result the area bounded by certain trajectories (a) increases whereas the area bounded by other trajectories (b) decreases. An electron on curve a moves in the same direction with respect to a magnetic field as a free electron would. On curve b, however, it moves in the opposite direction: the direction of motion of a positive charge.



ELECTRON PATH in a constant electric field (*left*) oscillates between zero and some upper value, E_{max}, that is a function of the particle's momentum and the force acting on it. The limits of the oscillation can also be represented as a band in space (*right*).



ELECTRON MOVING IN STRONG MAGNETIC FIELD parallel to the surface of a metal follows a helical path. During the free-path time it returns repeatedly to the skin layer, provided that skin thickness, *d*, is much smaller than the radius of the electron's orbit.



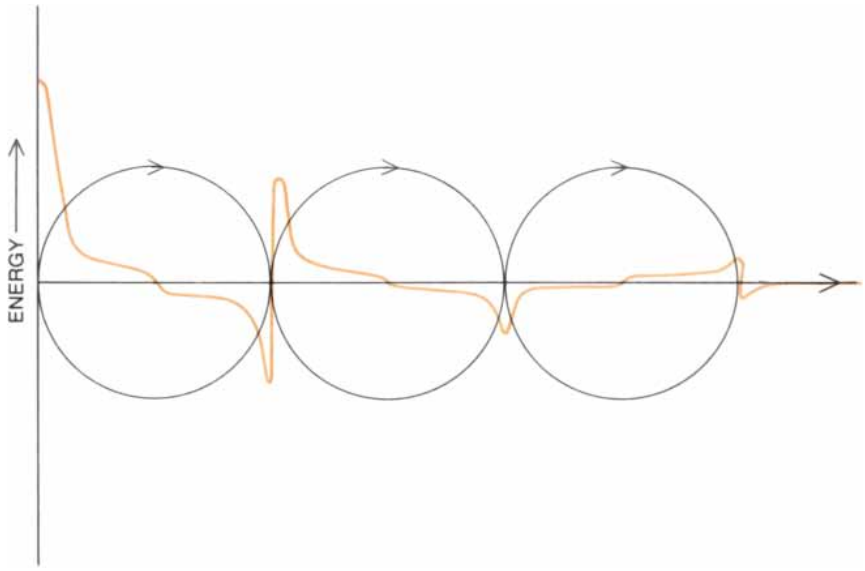
CYCLOTRON RESONANCE of an electron can be observed in a metal plate when the diameter of the orbit is smaller than the thickness of the plate. When the diameter exceeds the thickness of the plate, the electron is reflected from the internal surface and thus appears to travel in a series of bounces. In this illustration the magnetic field is perpendicular to the plane of the diagram.

ultrasound and even a stationary current flowing through the sample. Let us describe the most "graphic" methods: galvanomagnetic phenomena and cyclotron resonance.

Galvanomagnetic phenomena are associated with the flow of direct current through a conductor placed in a constant magnetic field. The current flowing in a given direction is proportional to the length of the path on which the charge acquires energy from the electric field. Since the magnetic field "twists" the charges, it allows them to move in a plane that is perpendicular to the field, but only until they turn for approximately half a cycle and not during the entire free path. As a result the electric resistance in a strong field turns out to be strongly dependent on the character of the electron orbits and consequently on the electron-dispersion law. If it were possible to produce "ideal" metals—metals whose electrons have an infinite free-path time—they would have infinite conductivity in the absence of a magnetic field. In a magnetic field this property is exhibited by all metals, but only if the magnetic field is parallel to the current. If the magnetic field is perpendicular to the current, certain ideal metals become ideal dielectrics with infinite resistance.

Other metals (for example metals with open Fermi surfaces) are sensitive not only to the relative orientation of the magnetic field and the current but also to the direction of the magnetic field with respect to the crystallographic axes of the crystal. There are magnetic-field directions in which the metal is an ideal conductor, and other directions in which the metal behaves like an ideal dielectric. Under real conditions this means that depending on the direction of the magnetic field, the resistance can change by hundreds of thousands or millions of times. By experimentally investigating the dependence of the resistance on the magnetic field we can not only establish the "direction of openness" of the charge trajectories but also clarify many other characteristics of the dispersion law. A comparison of the experimental results with the complete theory of galvanomagnetic phenomena, developed by the authors of this article together with V. G. Peschanskii, has yielded detailed information concerning the dispersion law of the conduction electron.

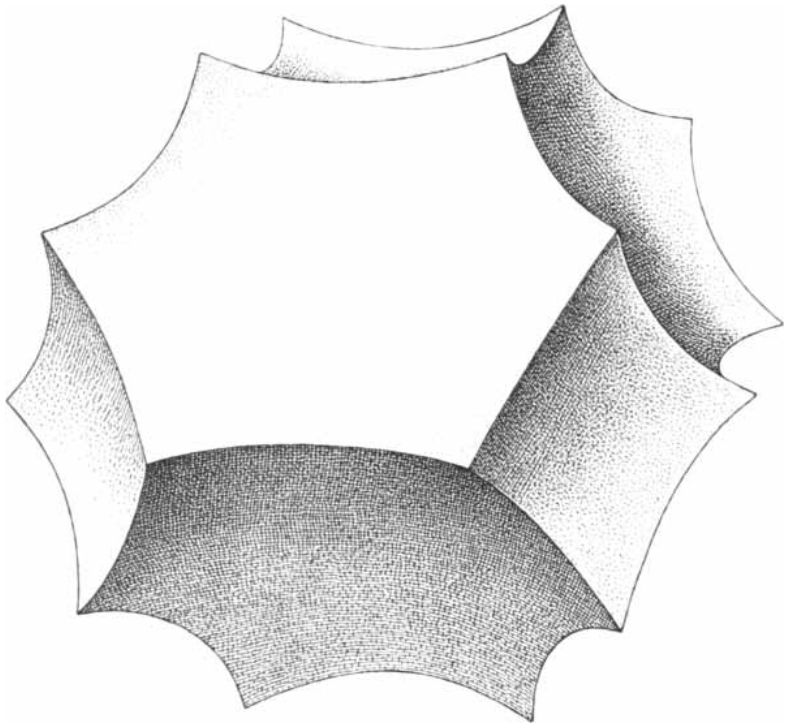
A very effective method of investigating Fermi surfaces is cyclotron resonance, which was predicted early in 1956 by one of the authors (Azbel') and E. A. Kaner. Let a metal be placed in a strong constant magnetic field that is parallel to its surface. Then if the charge trajectories are closed, the charge will move in closed curves in a plane that is perpendicular to the surface of the metal, completing a large number of revolutions during the free-path time [*see middle illustration on opposite page*]. Let a flux of radio waves be directed into the metal. As a result of the interaction of the radio waves and the conduction electrons the radio waves attenuate rapidly and penetrate to a small depth inside the metal (the skin-layer depth). In a very pure metal at the temperature of liquid helium the skin depth corresponding to radio waves of centimeter length is between 10^{-5} and 10^{-6} centimeter, whereas the electron orbit in a magnetic field of 10,000 oersteds is much larger (approximately 10^{-3} centimeter). If the period of revolution of the charge is a multiple of the period of the high-frequency field of the radio wave, the charge penetrating the skin layer will be accelerated with each revolution, the current will increase sharply and resonance will set in. The point is that the periods of revolution of different charges are different and depend on the position of the flat section of the equal-energy surface. Therefore not all the electrons take part in the resonance but only "selected" electrons located inside the sections where the period of revolution changes most slowly with changing section, so that the largest number of charges are under conditions close to resonance.



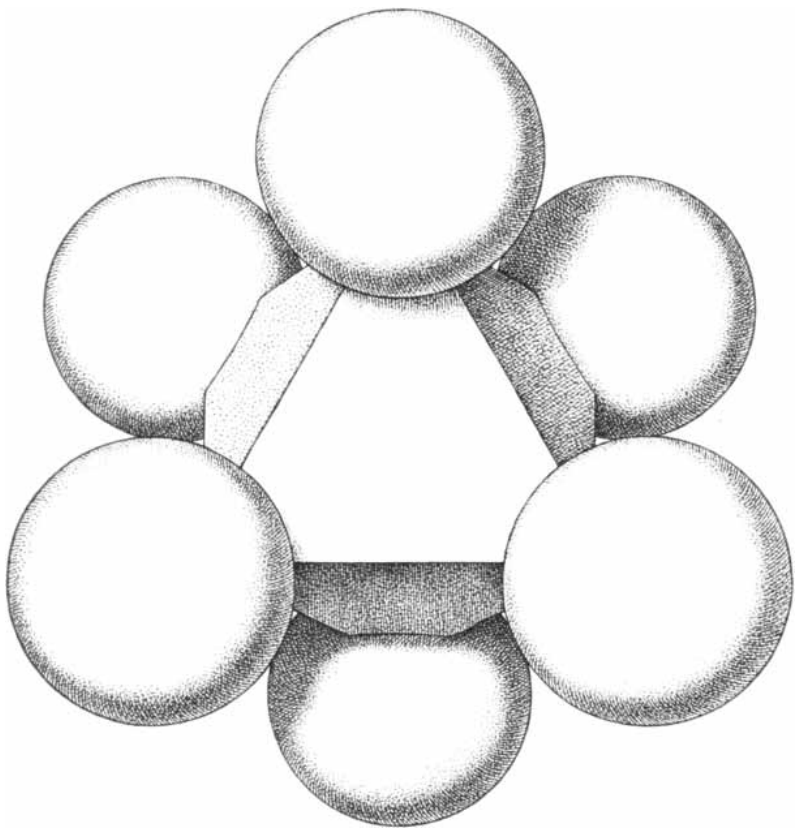
ELECTRONS ROTATING IN MAGNETIC FIELD parallel to the surface of a metal carry an alternating electric field into the interior of the sample. The explanation is that electrons leaving the skin layer carry electric current with them as they travel into the metal. The charges collect at the lower end of the orbit and produce a maximum current density, shown by the spikes where the first two orbits meet. The phenomenon is repeated with less strength at the next conjunction of orbits. The magnetic field is perpendicular to plane of the diagram.

Cyclotron resonance in metals has become one of the most widely used methods of investigating the electronic structure of metals. Resonance can be observed in the metallic plate only as long as the orbit corresponding to the "selected" charges taking part in the resonance fits inside the plate [*see bottom illustration on opposite page*]. As soon as this condition is violated the charges are scattered by the other surface of the plate; they are driven off the road, so to speak, and the resonance vanishes. This means that the dimensions of the orbit can be measured directly by determining the instant when the resonance vanishes. In this way it is possible to determine the geometry of the Fermi surface.

Cyclotron resonance is accompanied by a very interesting phenomenon. The charges leaving the skin layer carry electric current with them to the interior of the metal. They all gather together at the lower end of the orbit and produce there a maximum of current density. The alternating current generates an electric field that accelerates the next group of charges; these charges in turn carry away "their" current into the interior of the metal, and the cycle is repeated. As a result spikes of both field and current appear in the metal at fixed



"HOLE" PART OF FERMI SURFACE of indium resembles a truncated octahedron, a solid with 14 faces. The electronic part of the Fermi surface of indium is made up of overlapping spheres. If the spheres are dissolved, or removed, what remains is the shape illustrated here. It encloses a group of carriers, or quasiparticles, that behave as positive electrons, or holes.



ELECTRONIC PART OF FERMI SURFACE of tungsten is referred to by U.S. physicists as the jack because of its resemblance to the pronged metal object used in children's games. Like the hole part of the Fermi surface of indium, it is based on a truncated octahedron.

distances [*see illustration on preceding page*].

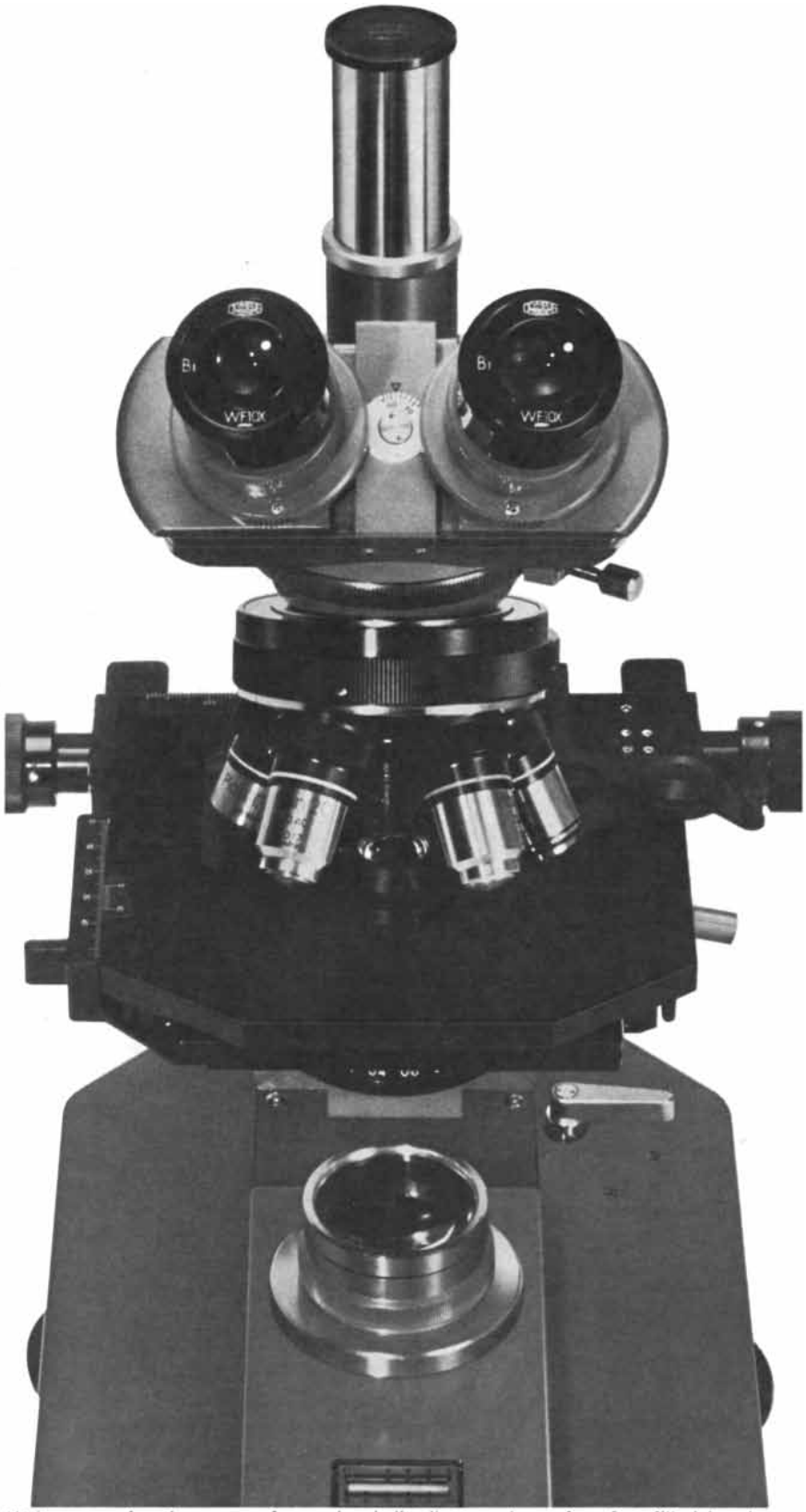
The fruitfulness of the quasiparticle concept as it is applied to metals is indicated by the fact that in the 10 years or so since the concepts were developed it has been possible to predict and explain a large number of physical phenomena and to exhaustively study the dispersion law of a large number of metals, and in particular to ascertain which Fermi surfaces are realized in which metals. This activity has resulted in a deeper understanding of the nature of the metallic state and in a clarification of the cause of the differences in behavior among metals. Investigations of the electronic energy spectrum have revealed unsuspected properties of metals.

Interest in the study of metals is not abating. Having learned the simplest properties of the conduction electrons and having determined the dispersion laws of the electrons of various metals, physicists forge ahead: they investigate the interaction of electrons with other quasiparticles, with one another and with various irregularities in the rigorous periodicity of the crystal structure. Attempts are made to completely calculate the observed properties. New features in the behavior of the conduction electrons, which require new theories and new experiments for their explanation, are being discovered.

Unfortunately all the concepts on which the progress in the understanding of the structure of the energy spectrum of solids is based pertain only to crystals. A new fundamental concept, the quasiparticle, has been introduced as the quantum of motion with a definite quasimomentum. The quasimomentum in turn is the consequence of the periodicity in the arrangement of atoms.

Solid-state physics is currently faced with the problem of constructing an effective theory of noncrystalline, amorphous substances, particularly disordered alloys and substances consisting of long-chain polymer molecules. Such structures include systems with a high degree of order of a noncrystalline type. These systems, which show a very complicated kind of order (as opposed to the "primitive" kind of order exhibited by the ions in crystals), are biological systems. There is no need to emphasize their importance and interest. The investigation of these systems has only begun, but one may hope that the ideas that were so fruitful in the quantum theory of the crystalline state will also turn out to be useful in some form in the new field.

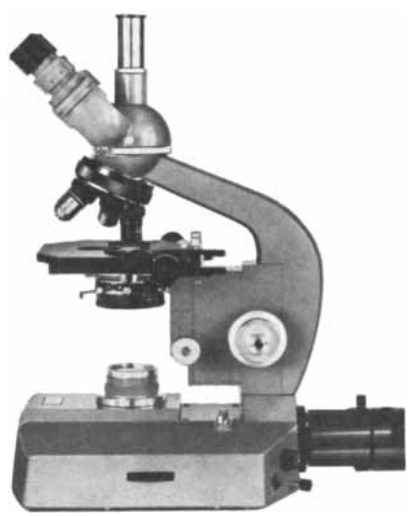
The microscope that works for you in multiple ways.



This is the FHT, in its basic form a valuable addition to the field of microscopy. When full play is given to its versatility it becomes a deluxe instrument—with enough interchangeable lenses to take you through practically any kind of research programs, a buit-in illuminator, focusing accuracy as fine as 0.0005mm and heads that can deliver a superwide field of view or a full photomichrographic system. And the magnification range extends from 28× to 2000×.

There are 8 models of the FHT, each one with a different eyepiece/objective/head/condenser arrangement, one just right for your application.

With the standard nosepiece mounted plan achromatic objectives you get super corrected components which make color and spherical aberration and field curvature almost non-existent, resulting in ideal images for both microscopy and photomicrography. Images are further guaranteed clarity by an N.A. 1.40 achromatic/aplanatic condenser. Added to this top microscope value is a full range of Olympus accessories—apochromatic objectives, photographic apparatus, exposure meter, phase contrast attachment, micro viewing screen, a dark field condenser.



Write to us for the name of your local distributor, plus a free fact-filled brochure.

OLYMPUS TOKYO *Seeing beyond man's vision*

OLYMPUS

OLYMPUS OPTICAL CO., LTD.
43-2, Hatagaya 2-chome, Shibuya-ku, Tokyo, Japan
OLYMPUS OPTICAL CO., (EUROPA) GmbH.
2 Hamburg 1, Steindamm 105, Germany
OLYMPUS CORPORATION OF AMERICA
2 Nevada Drive, New Hyde Park, N.Y. 11040, U.S.A.

SNOW CRYSTALS

The classic finely branched hexagonal crystal is only one of an almost infinite variety of shapes. Each form depends on the history of the snow crystal as it grows in a cloud

by Charles and Nancy Knight

The familiar six-sided figure of a snow crystal reflects the arrangement of water molecules in the structure of ice. A close look at natural snow crystals reveals, however, that beyond the basic hexagonal shape the crystals are far from uniform. If fresh snow crystals are collected and examined under the microscope, there is a good chance that many of the shapes will not be at all familiar. Some are flat plates; others are long needles; others are so complex and apparently irregular that they nearly defy description. The habit, or specific appearance, of the crystal depends on the conditions under which it was formed. By catching snow crystals in their natural environment and examining them it is possible to gain knowledge of these conditions.

Since snow is formed in clouds, let us first consider some of the general properties of clouds. Clouds form when air cools below the temperature at which it is saturated with water vapor. The cooling is usually caused by the air's expanding as it ascends from a region of higher pressure to one of lower pressure. There are three common types of ascent. The first is convection: relatively warm, light and moist air rises in plumes into colder, denser and drier air. The rate of ascent can be many meters per second. The second type of ascent is caused by large-scale atmospheric motions. Here the rate of ascent is slow, typically a few centimeters per second. The air can rise, however, over large areas for as long as a day. The third type of ascent is seen when horizontal winds encounter mountains. These three phenomena can act singly or in any combination.

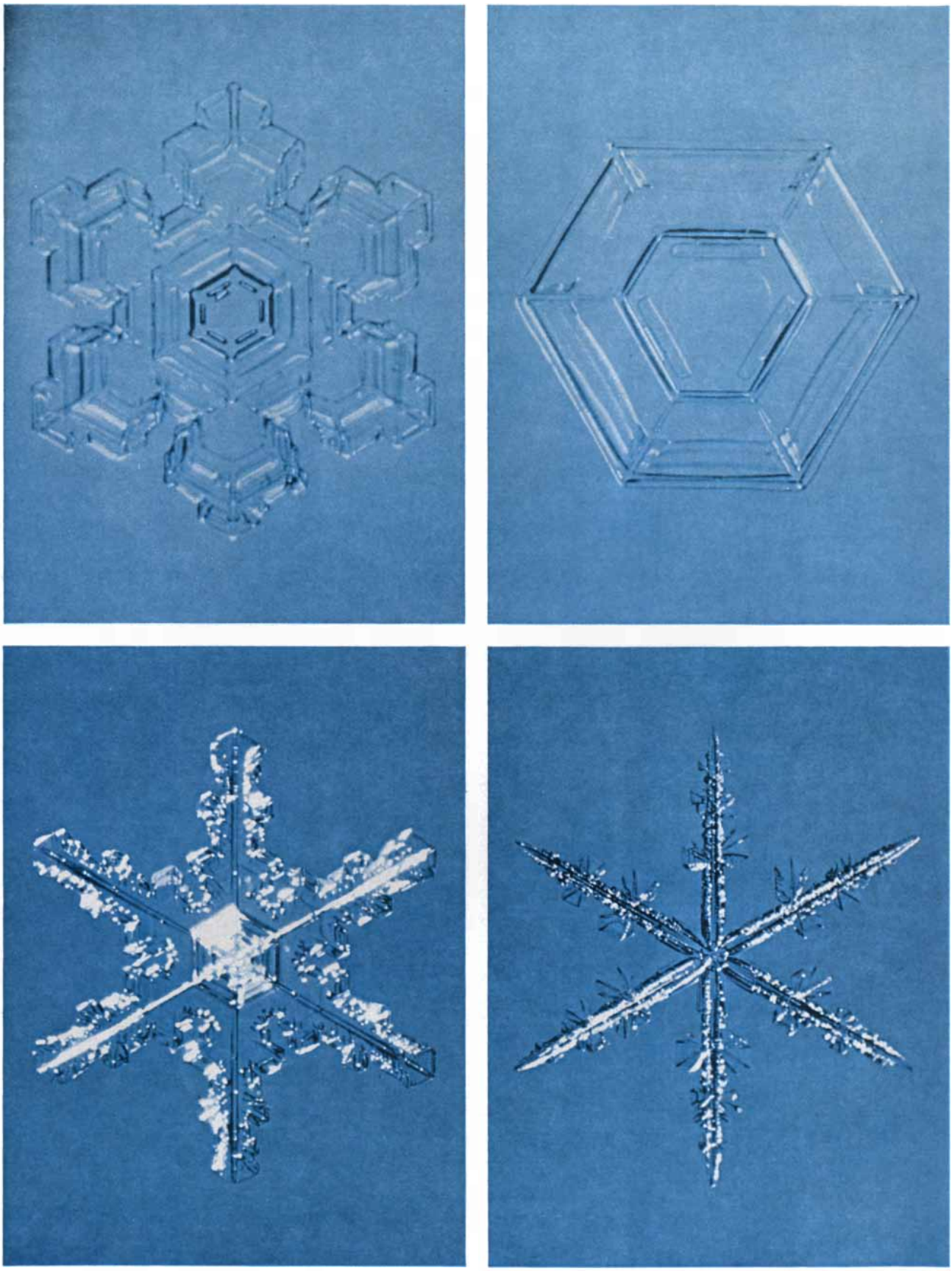
When the temperature at which the air reaches saturation is above the freezing point of zero degrees Celsius, a cloud of water droplets forms. The base of the cloud is flat, and the height of the base is the water-condensation level. When the temperature at which the air becomes saturated is below zero degrees C., as is typical in clouds that produce snow, there are two condensation levels, one for ice and one for supercooled water: water that remains liquid below zero degrees C. This situation arises because supercooled water always has a higher vapor pressure than ice at the same temperature. (Vapor pressure is the equilibrium pressure between a substance's vapor phase and its liquid or solid phase.) Thus when the two condensation levels coexist, the ice-condensation level is always the lowest: the first level that the air reaches as it ascends.

Since the rising and cooling air first reaches saturation with respect to ice, one might expect to find flat-bottomed ice clouds with their bases at that level. This, however, does not normally happen. Cold air can remain supersaturated with respect to ice without crystals forming for the same reason that supercooled water need not freeze. In both cases foreign particles are needed as nuclei for the ice to crystallize around, and such particles are comparatively scarce in nature. Therefore a cloud usually does not form at the first condensation level. Water vapor also needs solid nuclei to condense on and form drops, but the kinds of particles required are much commoner, and nature almost always has an abundant supply of them. The ascending air can pass the ice-condensation level without forming an ice cloud. Once the air reaches the water-condensation level, however, a cloud does form, composed of tiny drops of supercooled water.

Although nature is generally deficient in freezing nuclei, those nuclei that are present seem to be more effective in freezing supercooled water drops than in forming ice directly from a vapor. Some of the supercooled drops in the cloud contain or collide with freezing nuclei, and these drops freeze into ice. At this point the difference in the vapor pressures of ice and supercooled water becomes important. When supercooled water drops and ice particles coexist, vapor molecules around the water diffuse toward the ice. The water drop starts to evaporate, and the ice grows by direct condensation from the vapor.

The crystals grow as they fall, but since they are falling in updrafts they may actually be rising with respect to the ground, particularly in convective clouds. The clouds change with time, and the growth history of a single snow crystal can follow almost any course. Probably the commonest single habit is the planar dendrite, a flat crystal with delicate branches that is often regarded as the typical snowflake. Some of the prettiest of these snow crystals come from gentle updrafts where most of the crystal growth occurs as the crystals fall below the water-condensation level. Light snows on calm, cold nights are frequently of this type. In heavier snows the individual particles usually clump into large, fluffy snowflakes. The light, large aggregates consist of dendrites, whereas the somewhat heavier and faster-falling flakes can be clusters of bullet-shaped crystals.

The hexagonal form of snow crystals has long stimulated reflection. Johannes Kepler wrote an essay on the shapes of snow crystals in the 16th century. We now understand the hexagonal shape in terms of the arrangement of the molecules within the crystals. The speed with which water molecules find bonding sites on the crystal's surface depends on



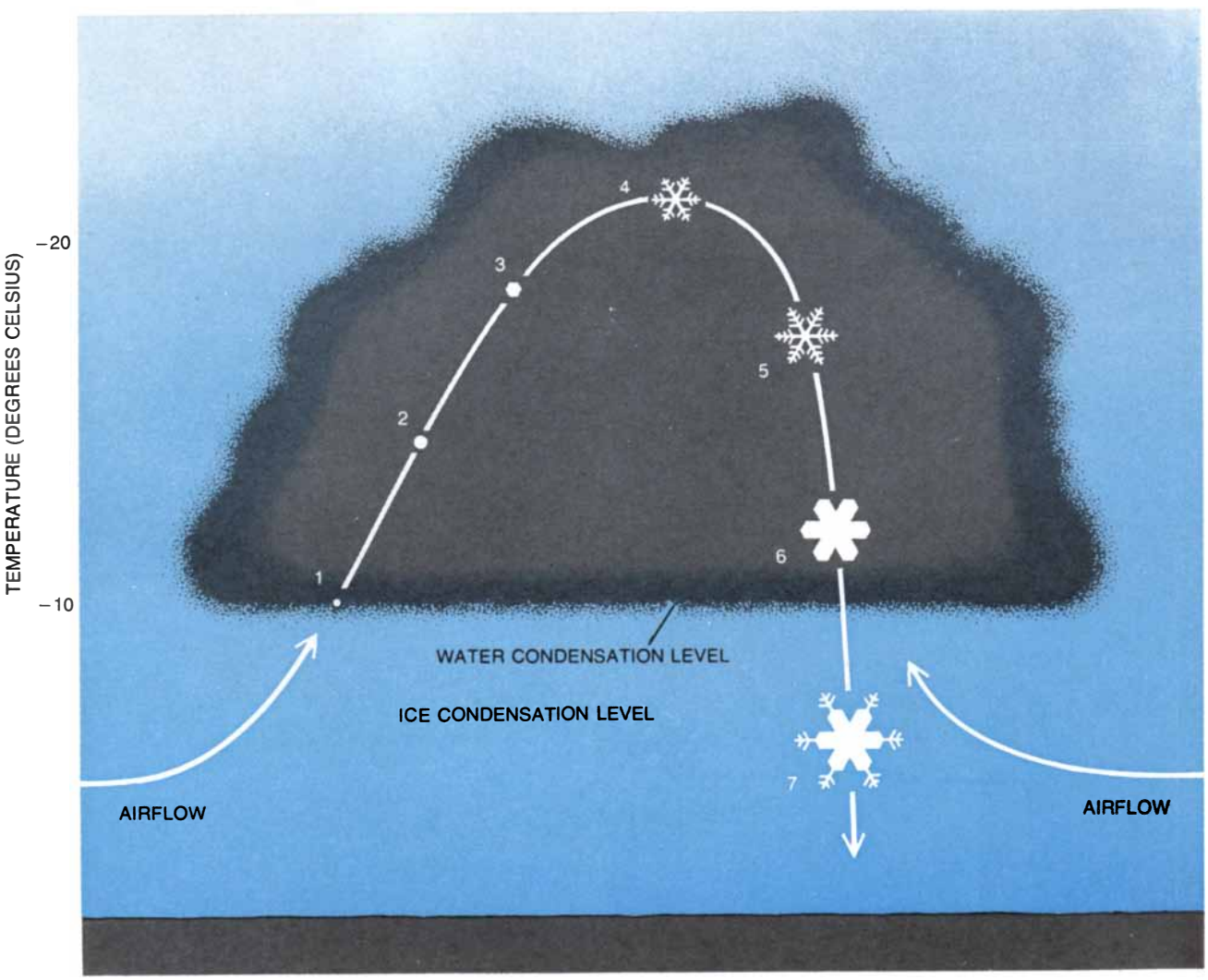
PLANAR, OR FLAT, SNOW CRYSTALS show the familiar hexagonal form. The two crystals at the top are plates; the two at the bottom, dendrites. These forms are not replicas but original flakes. They were caught by Teisaku Kobayashi of the Institute of Low Temperature Science at Sapporo in Japan and were photographed on a microscope slide in both transmitted and reflected light.

the structure of the surface. The structure of the surface in turn depends on the surface's orientation with respect to the hexagonal lattice of ice molecules of which it consists [*see bottom illustration on opposite page*]. Although the details of the process of crystal growth are difficult to study because the scale is so small, it is clear that some orientations of the crystal's surface can take molecules into the crystal lattice faster than others. The process gives rise to crystal shapes whose external hexagonal symmetry reflects the internal symmetry of the molecular bonding.

The simplest method of studying snow crystals is to replicate them by letting them fall into a thin layer of a dilute solution of plastic and solvent. The solvent evaporates rapidly, leaving a thin plastic cast of the snow crystal. (The ice itself evaporates through the cast.) The technique works beautifully with planar crystals, although it fails badly when the crystals are complex and three-dimensional. It nonetheless makes inspecting snow crystals quite convenient. One needs a low-power microscope to study the crystals, and working with a microscope is much easier at room temperature than in the cold.

Studying the more complicated snowflakes requires that the student suffer some discomfort. Snow can be captured and examined under a microscope in an unheated room or outdoors. Since snow evaporates rather rapidly even below zero degrees C., such examination must be fast in order to ensure that the details are faithfully recorded. One way to preserve the crystals without replicating them is to collect them in a cold liquid that does not dissolve ice. They can be studied and photographed at leisure, although the environment must remain cold. That is the method we have chosen in our work at the National Center for Atmospheric Research. Hexane is the liquid we use, but kerosene or unleaded gasoline would be equally satisfactory. Our procedure has the added advantage that it reduces the contrast between the crystal and the background, because the index of refraction of the liquid is not very different from that of ice. Thus the details of the more complicated snow particles are much clearer.

The traditional method of collecting snow crystals is to catch them on a piece of cardboard covered with black velvet, and then to pick them up with a dampened toothpick. This procedure can lead



SNOW CRYSTAL GROWS within a cloud that has been formed by convection: warm, moist air rising through a layer of cold, dry air. A droplet of water condenses (*1*) at the base of the cloud at the water-condensation level. It grows (*2*) as it rises in an updraft and eventually freezes (*3*) into an ice crystal. Water-vapor molecules in the cloud attach themselves to the lattice of the crystal, creating the branches of the familiar snow dendrite (*4*). The crystal is now falling (*5*) and starts to rime (*6*), or collide with relatively large water droplets. It falls out of the cloud altogether and continues to grow from vapor (*7*) until it drops below the ice-condensation level on its way to the ground. The growth of the snow crystal can follow almost any course of events. Temperatures shown are arbitrary.

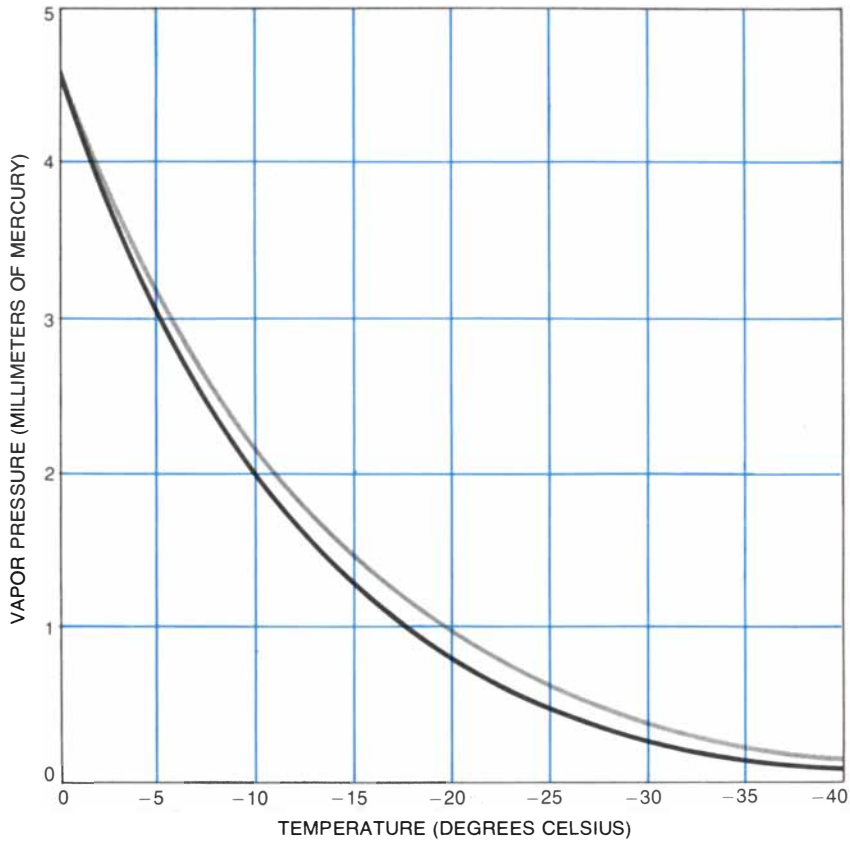
to a very distorted view of snow crystals, since one is inclined to select particularly large or symmetrical crystals for replication or examination. It is a good idea to occasionally look at everything that falls in a certain small area in 10 or 20 seconds to appreciate exactly what the storm cloud is producing.

The crystal lattice of ice guarantees that a snow crystal will be a hexagonal prism, but it does not determine whether the prism will be a thin hexagonal plate or a long hexagonal needle. Both crystal habits are found in nature. Experiments have shown that the habit of a snow crystal is almost entirely a function of temperature. As the temperature descends below the freezing point, the habit changes from planar to needle-like, back to planar and again to needle-like in four distinct regimes.

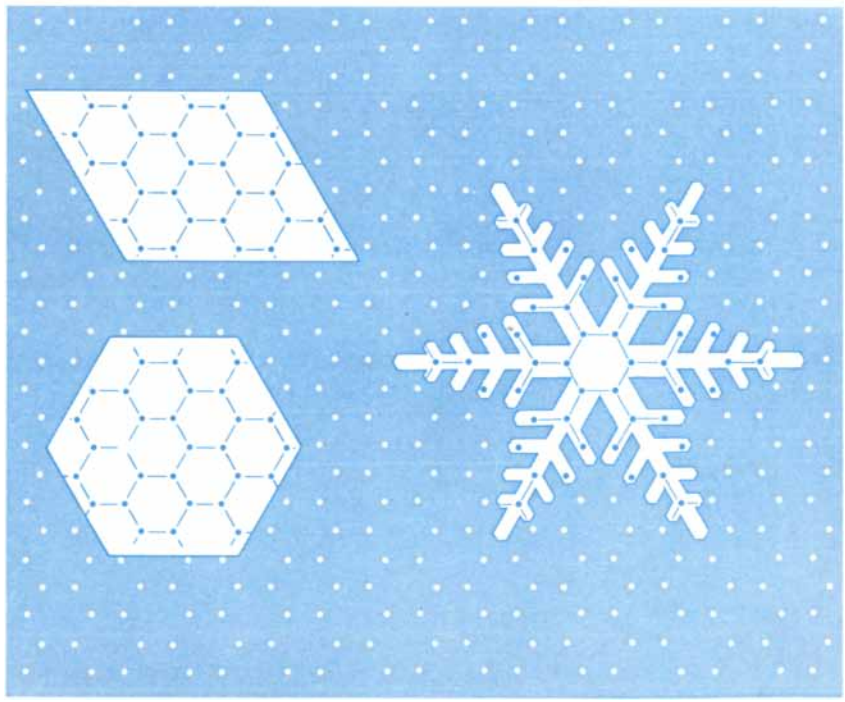
The amount of moisture in the air also affects the shape of the snow crystal. As the air becomes progressively more supersaturated the rate of growth increases and the crystal habits are more exaggerated: needles grow longer and plates become larger and thinner. The plates also become dendritic: adorned with fine branches. This branching is a result of the way the water molecules in the air diffuse to the plate. At high supersaturation the corners of the hexagonal plates are able to grow much faster than the sides because they can collect water molecules much more effectively. They develop into branches, which then split into more branches, giving rise to the familiar snow dendrites.

If the diffusion field of the moisture around growing snow crystals were perfectly symmetrical, one might expect perfectly symmetrical crystals. Very good symmetry of dendritic snow crystals is actually quite rare. This seems to be partly due to the way these crystals start and partly due to the fact that they are falling as they grow, giving a one-sidedness to the supply of material.

When a supercooled water droplet freezes, it can be transformed into a single crystal or into several different crystals, depending on the size of the drop and on its temperature when it is nucleated. A small drop with a diameter of from .01 to .05 millimeter becomes a single crystal if it freezes at a temperature above −20 degrees C.; it becomes two or more crystals if it freezes at a much lower temperature. Evidence of this fact is that clusters of bullet-shaped crystals are quite common. The crystals grow from a vapor in the lowest temperature regime, giving rise to the needle-like habit. The frozen droplet is com-



VAPOR PRESSURE of liquid water (*gray*) at any temperature below zero degrees Celsius is always higher than that of ice at the same temperature (*black*). Whenever a water droplet and an ice crystal are near each other, droplet evaporates and the ice crystal grows. Crystals grow most rapidly at −13 degrees C., where difference in vapor pressure is greatest.

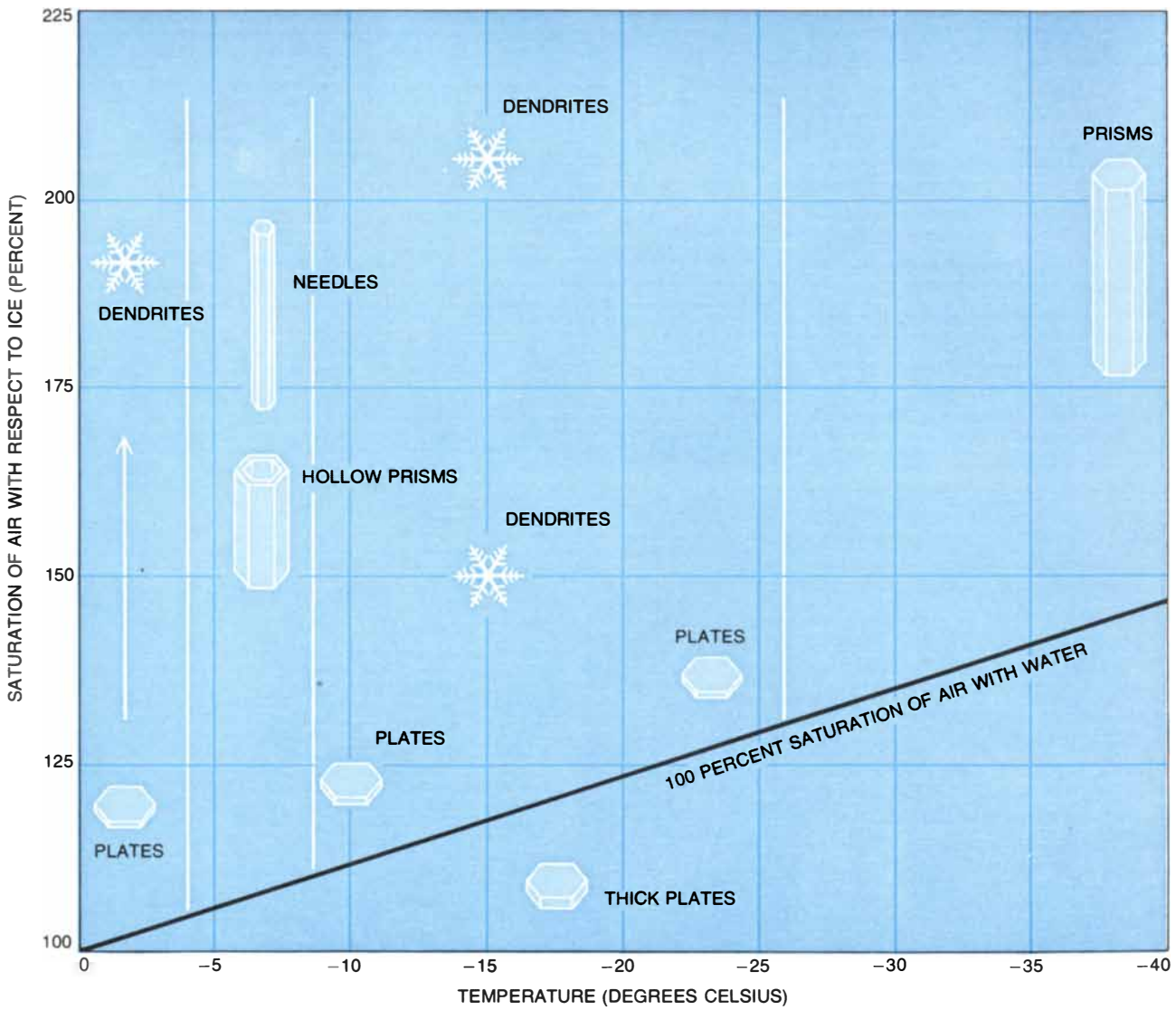


FACES OF A SNOW CRYSTAL (*white shapes*) can be oriented only in particular ways with respect to the hexagonal lattice of ice molecules (*dots*) of which it is composed. Some areas of the crystal's surface grow faster than others, giving rise to crystal shapes whose hexagonal symmetry reflects the internal symmetry of the way the molecules are bound.

posed of several crystals, so that subsequent growth in the vapor yields a cluster of bullet-shaped crystals and joined at their tips. They taper toward their junction because the space they have to grow in is limited. Sometimes the clusters break apart into single bullet-shaped crystals; sometimes they clump together, forming the complex snowflakes. When a single bullet cluster falls to a region of higher temperature and continues to grow, the outer end of each bullet develops into a plate or a dendrite. These three-dimensional snow particles are difficult to photograph and are often even difficult to recognize, particularly when they clump together during their fall.

Single straight-needle crystals are also common. When they fall into the habit regime centered at about −15 degrees C. (or ascend in an updraft if they grew between −4 degrees and −8 degrees), a plate or a dendrite grows at each end. When needles are capped by solid plates, the crystals are called tsuzumi crystals, after a Japanese drum that has that shape. Tsuzumi crystals can be found with almost any ratio between the length of the central column and the diameter of the end plate. Particularly when the central column is short, one end plate or dendrite is generally much larger than the other. Presumably this feature arises when the two end plates are so close together that their edges compete with one another for the supply of vapor.

A significant—perhaps even a major—proportion of the ordinary dendritic snow crystals are actually in a sense tsuzumi crystals, although they have an extremely short central column. In fact, the central "column" can be just the original water droplet frozen into a single crystal with a plate or a dendrite developed at each side along the axis of hexagonal symmetry. When the two end plates are less than .1 millimeter apart, the competition between them becomes very strong, leading to dendritic snow crystals with fewer than six branches. One end plate develops one, two or three of the six branches; the opposite end plate develops the other five, four or three branches. Then the two end plates may break apart, as they often do when they strike the velvet of the collecting pad. The result is a dendrite with fewer than six arms, although the angle be-



SNOW CRYSTAL HABIT, or form of growth, depends on the temperature and the amount of water vapor in the air from which the crystal grows. The habits are divided into four regimes; in two of the regimes the crystals grow as flat plates, and in the other two they grow as prisms and needles. High water-vapor content yields more exaggerated habits: dendrites instead of plates and long, thin needles instead of short, thick ones. Sometimes crystals are combinations when they fall through several regimes as they form.

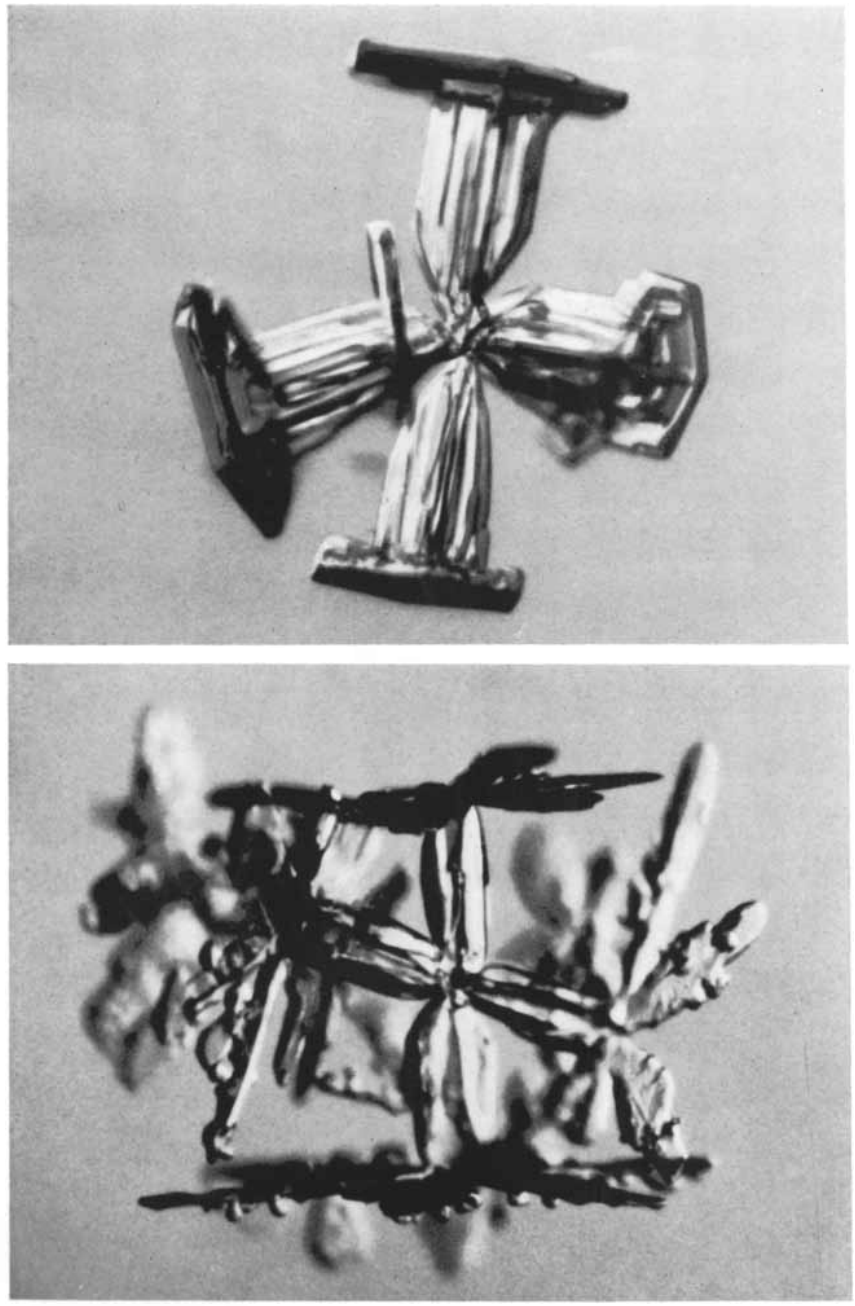
tween each pair of arms is always a multiple of 60 degrees. When such crystals fail to break apart, it takes careful scrutiny to determine that they actually have two layers.

When a snow crystal evolves within a cloud of supercooled water drops, it can grow not only by stealing vapor from around the drops but also by actually colliding with individual drops. The process is called riming, and it is an important complication in all types of snowflakes and snow crystals. As a crystal grows by diffusion from the droplets, its velocity increases as it falls through the cloud. At first the small droplets in the clouds are swept around the growing crystal along with the air, some of the drops evaporating away completely as they pass close to the crystal. If the crystal grows big enough, however, the larger drops can neither evaporate away nor bypass the crystal, and they collide with it.

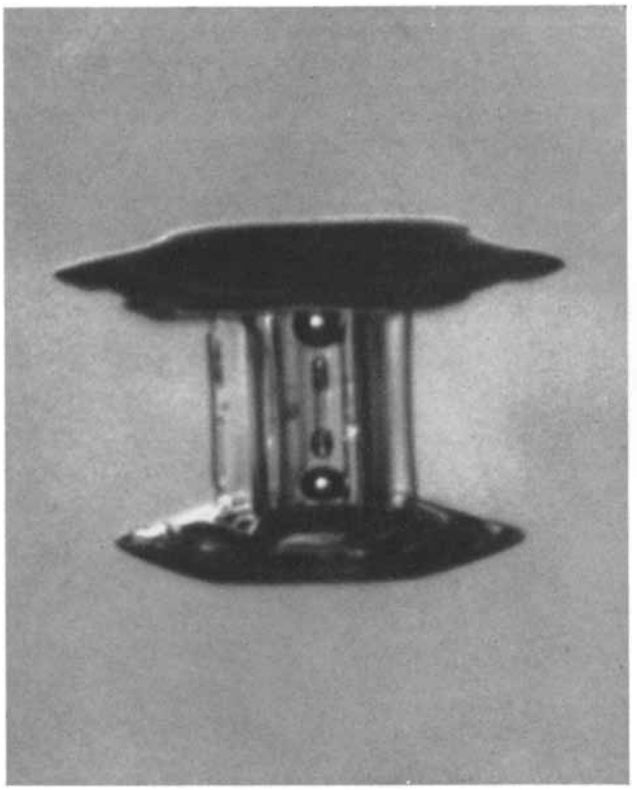
In any particular cloud a snow crystal must reach a certain size and attain a certain speed as it falls before it can be rimed appreciably. Growth by riming increases the crystal's speed of fall much faster than growth from vapor. When a crystal rimes, material is added mostly on its underside, thus increasing its weight without greatly increasing its air resistance. On the other hand, when the crystal grows from a vapor, material tends to be added at its sides, increasing its air resistance along with its weight. Therefore riming is an accelerating process: the more a crystal is rimed, the faster it rimes. It is not uncommon for rime deposits to have many times the mass of the original crystal.

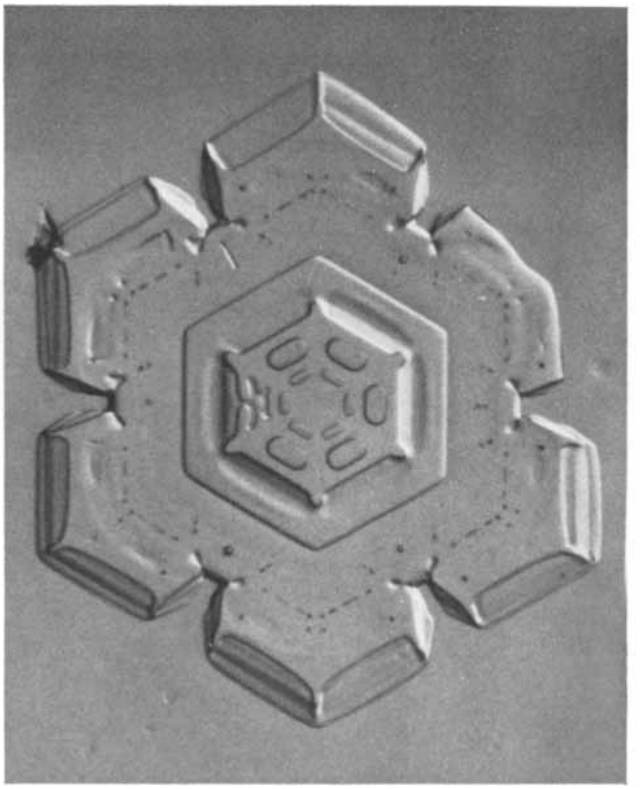
A snow crystal may rime and then continue to grow from a vapor if it falls into the very lowest regions of the cloud, where the water drops are too small to collide with it, or if it falls into the region below the cloud but above the ice-condensation level. When the riming creates new orientations of the crystal lattice, the result is quite picturesque. A crystal type known as the spatial dendrite originates in this fashion. The new crystals are not oriented at random with respect to the original crystal. They assume certain new orientations related to the old one in such a way that the lattice structures tend to fit one another across the interface. The production of new crystal orientations in riming is probably the same in principle as the freezing of water drops at low temperatures into several crystals.

When riming continues to such an extent that the original crystal is unrecognizable or nearly so, the resulting snow is called graupel. Clouds must be thick or updrafts fairly strong before the crystals remain in them long enough to grow into graupel. Since graupel particles have the largest mass and the highest speed of fall of any of the particles in a snowfall, they can be a stage in the formation of rain. Much rain is melted graupel. When the bases of the clouds are fairly high and there is a thick layer of warm, undersaturated air below them, the individual elements of the precipitation in the cloud must be large before they can fall to the ground as rain without evaporating on the way. The mechanism that produces graupel enables many clouds to rain, and it is this mechanism that rainmakers try to encourage by adding artificial freezing nuclei to clouds. Graupel particles are also the embryos for hail [see "Hailstones," by Charles and Nancy Knight; SCIENTIFIC

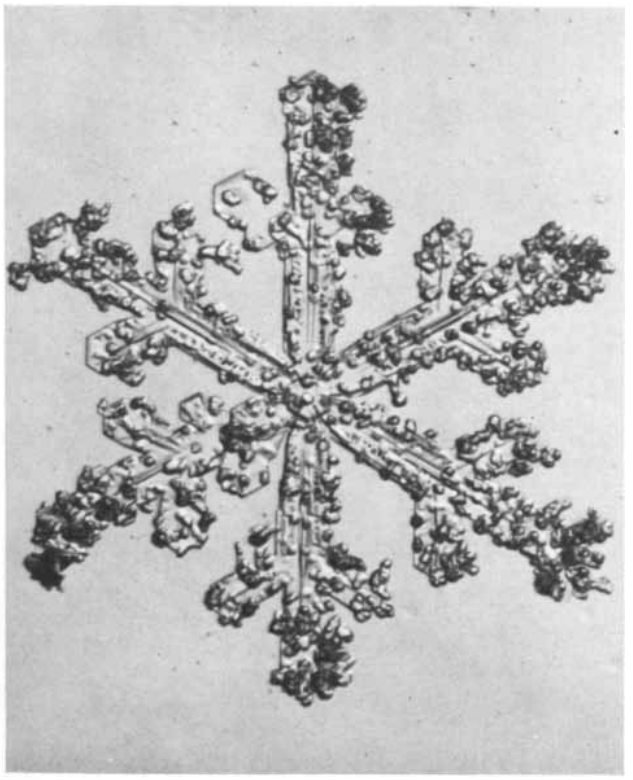


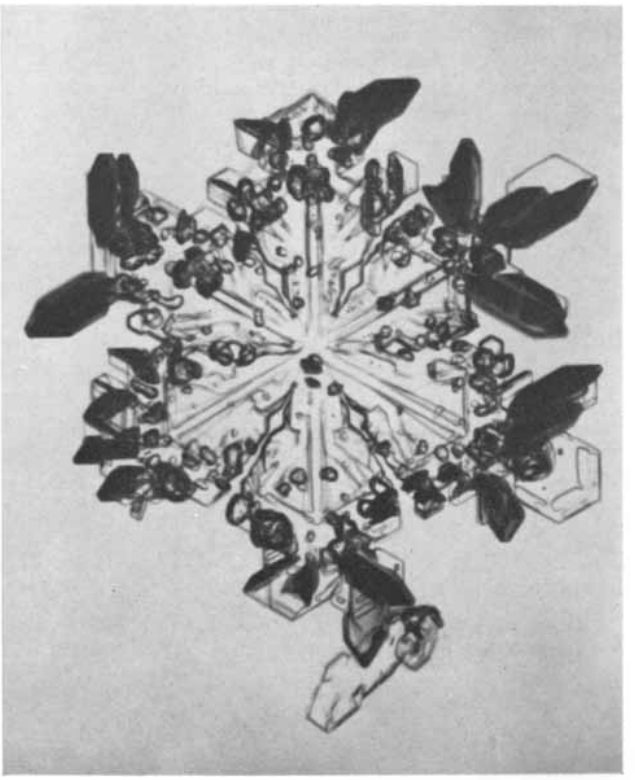
BULLET CLUSTERS originate when a tiny droplet freezes at very low temperature, forming several crystals shaped like bullets. The crystals are all joined at their tips, and the orientation of the ice lattice in each bullet is different from that of every other. If such crystals fall into higher temperature regimes, their ends grow plates (*top*) or dendrites (*bottom*).





TSUZUMI CRYSTALS, named after a Japanese drum of the same shape, start out as a central hexagonal column that grows a plate at each end. Sometimes the crystals are clearly of the tsuzumi shape (*left*); at other times they look almost like an ordinary planar crystal (*right*). Here the crystal has a very short, thick column with one end plate much larger than the other. The perfect inner hexagon is the smaller end plate, and the markings inside it are the details of the central column. If a central column grows two dendrites instead of two hexagonal plates, one dendrite may develop a few of the six branches and the opposite dendrite the remaining ones.





RIMED SNOW CRYSTALS have grown by capturing water droplets as well as vapor molecules. A lightly rimed dendrite (*left*) has a peculiar rounded appearance. New orientations of crystals are produced in many cases when a rimed crystal grows further from a vapor (*right*). The original crystal was a flat dendrite; the end result is a three-dimensional crystal known as a spatial dendrite.

AMERICAN, April, 1971]. Clouds have been seeded to suppress the formation of hail by furnishing nuclei for more graupel to freeze from supercooled water drops. The purpose is to make the resulting hailstones more numerous and smaller, so that they melt before they reach the ground.

A number of the phenomena related to snow crystals are imperfectly understood. Explaining why the growth habits of the crystals vary as they do is the classical problem. Other aspects, such as the formation of new crystal orientations in riming and the origin of certain types of graupel, are not yet completely explained. By far the most important problem, however, is one that has arisen rather recently. Over the past several years a number of workers have measured the concentration of ice nuclei with respect to the concentration of ice crystals in clouds that form in the same air. Particularly at the warmer subfreezing temperatures around −10 degrees C., they have found that there are literally thousands of times more ice crystals than their measurements of the concentration of ice nuclei had led them to predict. As a result of this discrepancy they have been led to postulate a process by which one snow crystal or ice particle, perhaps during riming, could give rise to many more ice crystals. Intensive laboratory study has so far failed to reveal such a process. The problem is unsolved: either the measurements of the concentration of ice nuclei are wrong or the crystal-multiplication process exists and up to now has simply eluded discovery. Once such processes are understood it may be possible to predict and to alter them.

Let us conclude with a calculation bearing on the proverbial problem of whether or not there have ever been two identical snow crystals. A typical snow crystal weighs about 10^{-6} gram. If the average amount of snow formed on the earth each year (including snow that melts or evaporates before it reaches the ground) is equivalent to a layer of liquid water three centimeters deep over the entire surface of the earth (probably an underestimate), and if the earth is three billion years old, then some 10^{35} snow crystals have formed in that time. This comes to some 10^{29} grams—about 50 times the mass of the earth. Each snow crystal, however, consists of some 10^{18} molecules of water. Considering the huge variety of ways that number of molecules can be arranged, it may very well be that there have never been two identical snow crystals.

MATHEMATICAL GAMES

Sim, Chomp and Race Track: new games for the intellect (and not for Lady Luck)

by Martin Gardner

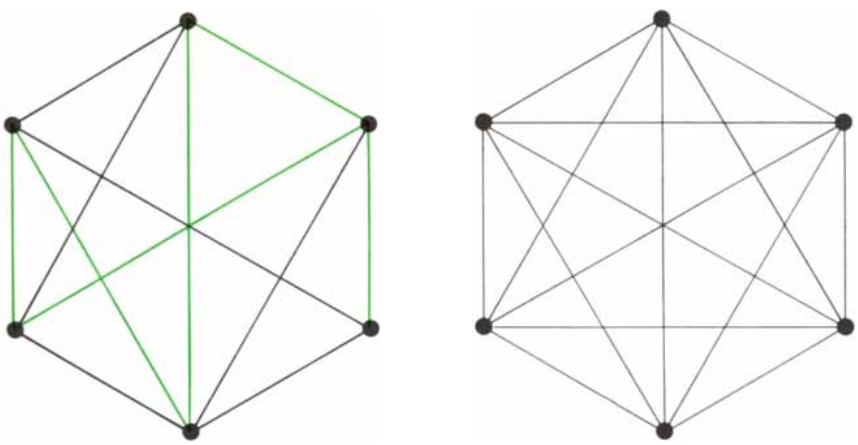
New mathematical games of a competitive type, demanding more intellectual skill than luck, continue to proliferate both in the U.S. and abroad. In Britain they have become so popular that a monthly periodical called *Games and Puzzles* was started in 1972 just to keep devotees informed. (Interested readers can contact the publisher at P.O. Box 4, London N6 4DF.) *Strategy and Tactics* (a bimonthly with offices at 44 East 23rd Street, New York, N.Y. 10010) is primarily concerned with games that simulate political or military conflicts, but a column in the publication by Sidney Sackson reports on new mathematical games of all kinds. Sackson's book *A Gamut of Games* (1969) has a bibliography of more than 200 of the best mathematical board games now on the market.

Simulation games are games that model some aspect of human conflict: war, population growth, pollution, marriage, sex, the stock market, elections, racism, gangsterism—almost anything at all. They are being used as teaching devices, and some notion of how widely can be gained from the fact that a 1970 catalogue, *The Guide to Simulation Games for Education and Training,* by David W. Zuckerman and Robert E. Horn, runs to 334 pages. (This valuable reference is available from Information Resources, Inc., 1675 Massachusetts Avenue, Cambridge, Mass. 02138.)

This month we take a look at three unusual new mathematical games. None requires a special board or equipment; all that is needed are pencil and paper (graph paper for the first game) and (for the third) a supply of counters.

Race Track, virtually unknown in this country, is a truly remarkable simulation of automobile racing. I do not know who invented it. It was called to my attention by Jurg Nievergelt, a computer scientist at the University of Illinois who picked it up on a recent trip to Switzerland.

The game is played on graph paper. A racetrack wide enough to accommodate a car for each player is drawn on the sheet. The track may be of any length or shape, but to make the game interesting it should be strongly curved [*see illustration on opposite page*]. Each contestant should have a pencil or pen of a different color. To line up the cars each player draws a tiny box just below a grid point on the starting line. In the example illustrated the track will take three cars, but to simplify things a race of two cars is shown. Lots can be drawn to decide the order of moving. In the sample game, provided by Nievergelt, Black moves first.



The game of Sim

You might suppose that a randomizing device now comes into play to determine how the cars move, but such is not the case. At each turn a player simply moves his car ahead along the track to a new grid point, subject to the following three rules:

1. The new grid point and the straight line segment joining it to the preceding grid point must lie entirely within the track.

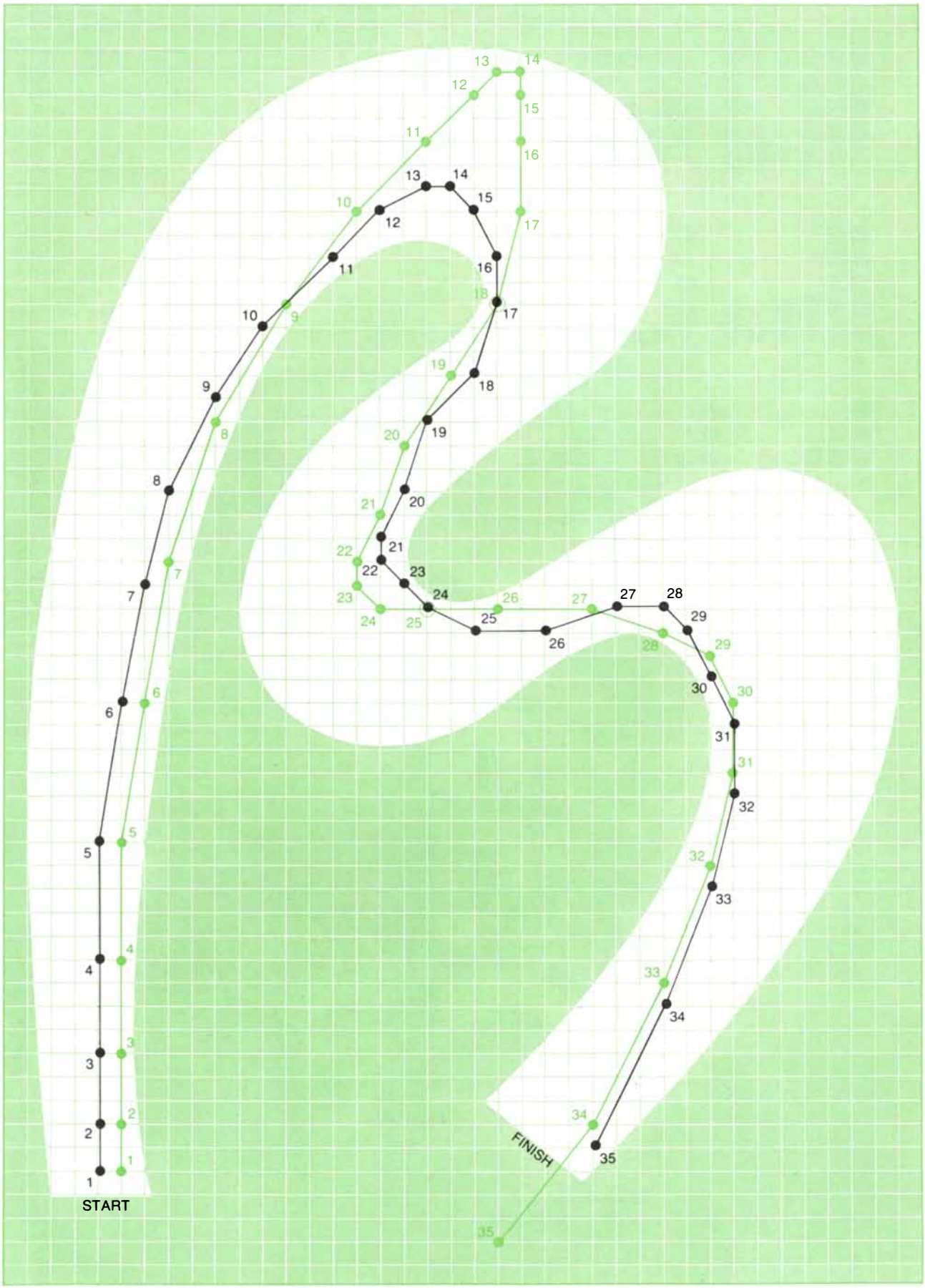
2. No two cars may simultaneously occupy the same grid point. In other words, no collisions are allowed. For instance, consider move 22. Green, the second player, would probably have preferred to go to the spot taken by Black on his 22nd move, but the no-collision rule prevented it.

3. Acceleration and deceleration are simulated in the following ingenious way. Assume that your previous move was k units vertically and m units horizontally and that your present move is k' vertically and m' horizontally. The absolute difference between k and k' must be either 0 or 1, and the absolute difference between m and m' must be either 0 or 1. In effect, a car can maintain its speed in either direction or it can change its speed by only one unit distance per move. The first move, following this rule, is one unit horizontally or vertically, or both.

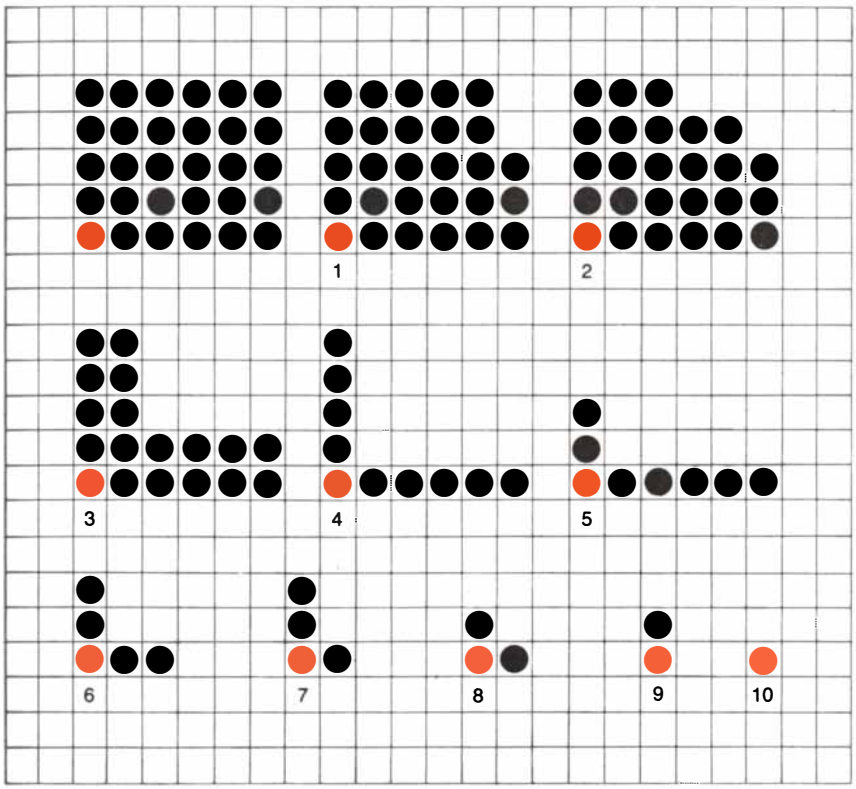
The first car to cross the finish line wins. A car that collides with another car or leaves the track is out of the race. In the sample game Green slows too late to make the first turn efficiently. He narrowly avoids a crash, and the bad turn forces him to fall behind in the middle of the race. He takes the last curve superbly, however, and he wins by crossing the finish line one move ahead of Black.

Nievergelt programmed Race Track for the University of Illinois's Plato IV computer-assisted instruction system, which uses a new type of graphic display called a plasma panel. Two or three people can play against one another or one person can play alone. The game became so popular that the authorities made it inaccessible for a week to prevent students from wasting too much time on it.

Our second pencil-and-paper game is called Sim after Gustavus J. Simmons, a physicist at the Sandia Corporation laboratories in Albuquerque, who invented it when he was working on his Ph.D. thesis on graph theory. He was not the first to think of it (the idea occurred independently to a number of mathematicians) but he was the first to publish it and to analyze it completely with a com-



The Race Track game



Chomp on a 5-by-6 field

puter program. In his note on "The Game of Sim" (*Journal of Recreational Mathematics*, Vol. 2, April, 1969, page 66) he says that one of his colleagues picked the name as short for SIMple SIMmons, and because the game resembles the familiar game of nim.

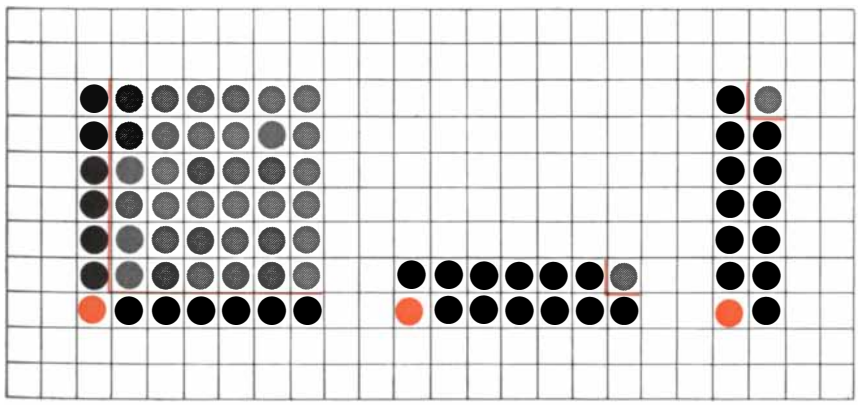
Six points are placed on a sheet of paper to mark the vertexes of a regular hexagon. There are 15 ways to draw straight lines connecting a pair of points, producing what is called the complete graph for six points [*see illustration on page 108*]. Two Sim players take turns drawing one of the 15 edges of the graph, each using a different color. The first player to be forced to form a triangle of his own color (only triangles whose vertexes are among the six starting points count) is the loser.

If only two colors are used for the edges of a chromatic graph, it is not hard to prove that six is the smallest number of points whose complete chromatic graph is certain to contain a triangle with sides all the same color. Simmons gives the proof as follows: "Consider any vertex in a completely filled-in game. Since five lines originate there, at least three must be the same color—say blue. No one of the three lines joining the end points of these lines can be blue if the player is not to form a blue triangle, but then the three interconnecting lines form a red triangle. Hence at least one monochromatic (all one color) triangle must exist, and a drawn game is impossible."

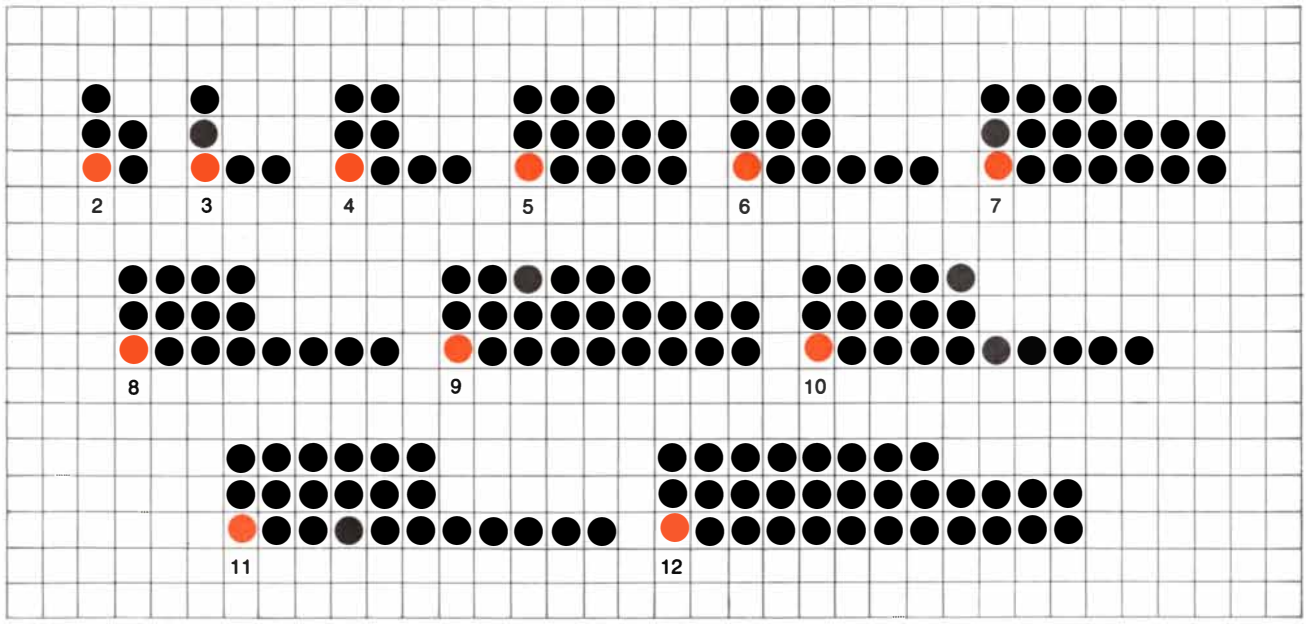
With a bit more work a stronger theorem can be established. There must be at least *two* monochromatic triangles. A detailed proof of this is given by Frank Harary, a University of Michigan graph theorist, in his paper "The Two-Triangle Case of the Acquaintance Graph" in *Mathematics Magazine* (Vol. 45, May, 1972, pages 130–135). Harary calls it an acquaintance graph because it provides the solution to an old brainteaser: Of any six people, prove that at least three are mutual acquaintances or at least three are mutual strangers. Harary not only proves that there are at least two such sets but also shows that if there are exactly two, they are of opposite types (colors on the graph) if and only if the two sets have just one person (point) in common.

Because Sim cannot be a draw, it follows that either the first or the second player can always win if he plays correctly. When Simmons wrote his note in 1969, he did not know which player had the win, and in actual play among equally skillful players wins are about equally divided. Later he made an exhaustive computer analysis showing that the second player could always win. Because of symmetry all first moves are alike. The computer results showed that the second player could respond by coloring any of the remaining 14 edges and still guarantee himself a win. (Actually, for symmetry reasons, there are only two fundamentally different second moves: one that connects with the first move and one that does not.) After the first player has made his second move exactly half of the remaining plays lead to a sure win for the second player and half to a sure loss, assuming of course that both sides play rationally. If 14 moves are made without a win, the last move, by the first player, will always produce two monochromatic triangles of his color. This 14-move pattern is unique in the sense that all such patterns are topologically the same. Can you find a way of coloring 14 edges of the Sim graph, seven in one color and seven in another, so that there is no monochromatic triangle on the field? A solution will be given next month.

The most interesting unanswered question about Sim is whether there is a relatively simple strategy by which the second player can win without having to memorize all the correct responses. Even if he has at hand a computer printout of the total game tree, it is of little practical use because of the enormous difficulty of locating on the



Winning first bites on square field, 2-by-n field and n-by-2 field



Winning first bites on 3-by-n fields

printout a position isomorphic to the one on the board. Simmons' computer results have been verified by programs written by Michael Beeler at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology and more recently by Jesse W. Croach, Jr., of West Grove, Pa., but no one has been able to extract from the game tree a useful mnemonic for the second player.

Sim can of course be played on other graphs. On complete graphs for three and four points the game is trivial, and for more than six points it becomes too complicated. The pentagonal five-point graph, however, is playable. Although draws are possible, I am not aware of any proof that a draw is inevitable if both sides make their best moves.

Our third game, which I call Chomp, is a nim-type game invented by David Gale, a mathematician and economist at the University of California at Berkeley. Gale is the inventor of Bridg-it, a popular topological board game still on the market. (It was first described in this department in October, 1958, and a winning strategy was first disclosed in July, 1961.) What follows is based entirely on results recently provided by Gale.

Chomp can be played with a supply of counters [*see top illustration on opposite page*] or with *O*'s or *X*'s on a sheet of paper. The counters are arranged in a rectangular formation. Two players take turns removing counters as follows. Any counter is selected. Imagine that this counter is inside the vertex of a right angle through the field, the base of the angle extending east below the counter's row and its other side extending vertically north along the left side of the counter's column. All counters inside the right angle are removed. This constitutes a move. It is as though the field were a cracker and a right-angled bite were taken from it by jaws approaching the cracker from the northeast.

The object of the game is to force your opponent to chomp the poison counter at the lower left corner of the array [*colored counter*]. The reverse form of Chomp—winning by taking this counter—is trivial because the first player can always win on his first move by swallowing the entire rectangle.

What is known about this game? First, we dispose of two special cases for which winning strategies have been found.

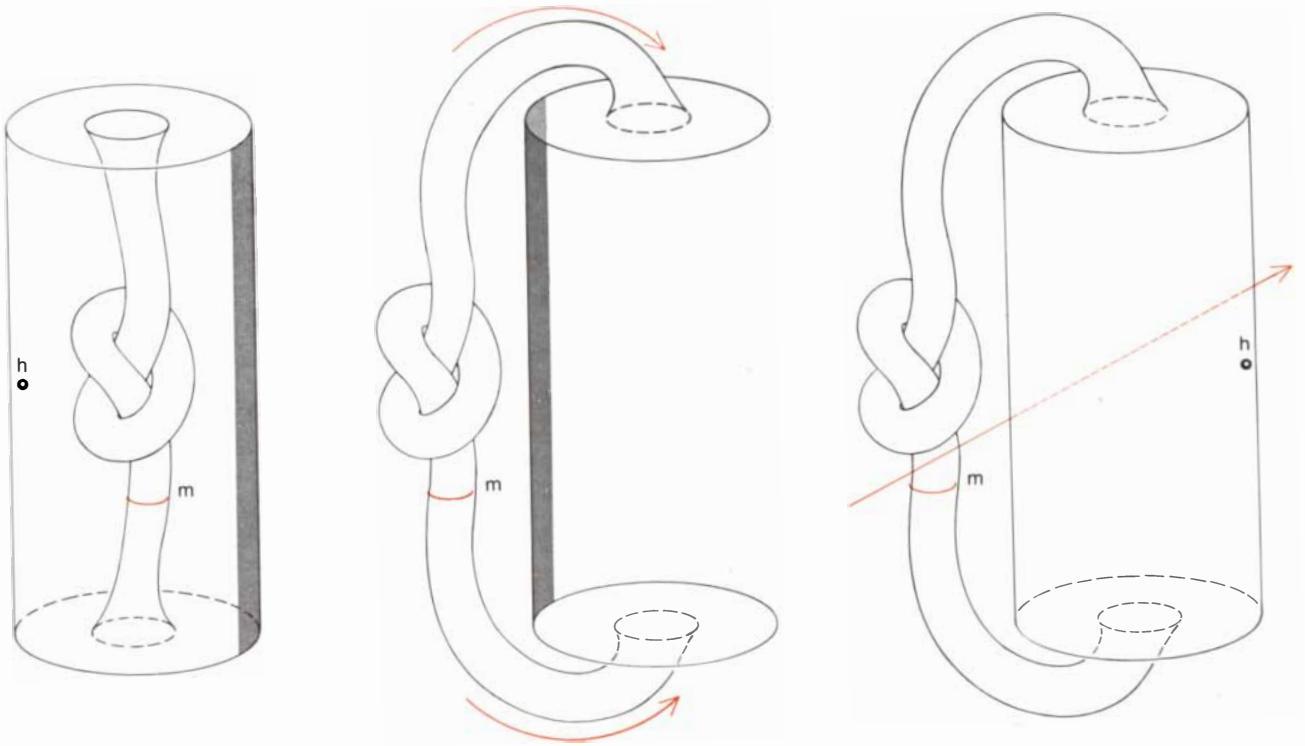
1. When the field is square, the first player wins by taking a square bite whose side is one less than that of the original square. This leaves one column and one row, with the poison piece at the vertex [*see bottom illustration on opposite page*]. From now on the first player "symmetrizes." Whatever his opponent takes from either line, he takes equally from the other. Eventually the second player must take the poison piece.

2. When the field is 2 by *n*, the first player can always win by taking the counter at top right [*see bottom illustration on opposite page*]. Removing that counter leaves a pattern in which the bottom row has one more counter than the top row. From now on the first player always plays to restore this situation. One can easily see that it can always be done and that it ensures a win. The same strategy applies to fields of width 2, except now the first player always makes sure that the left column has one more counter than the right column.

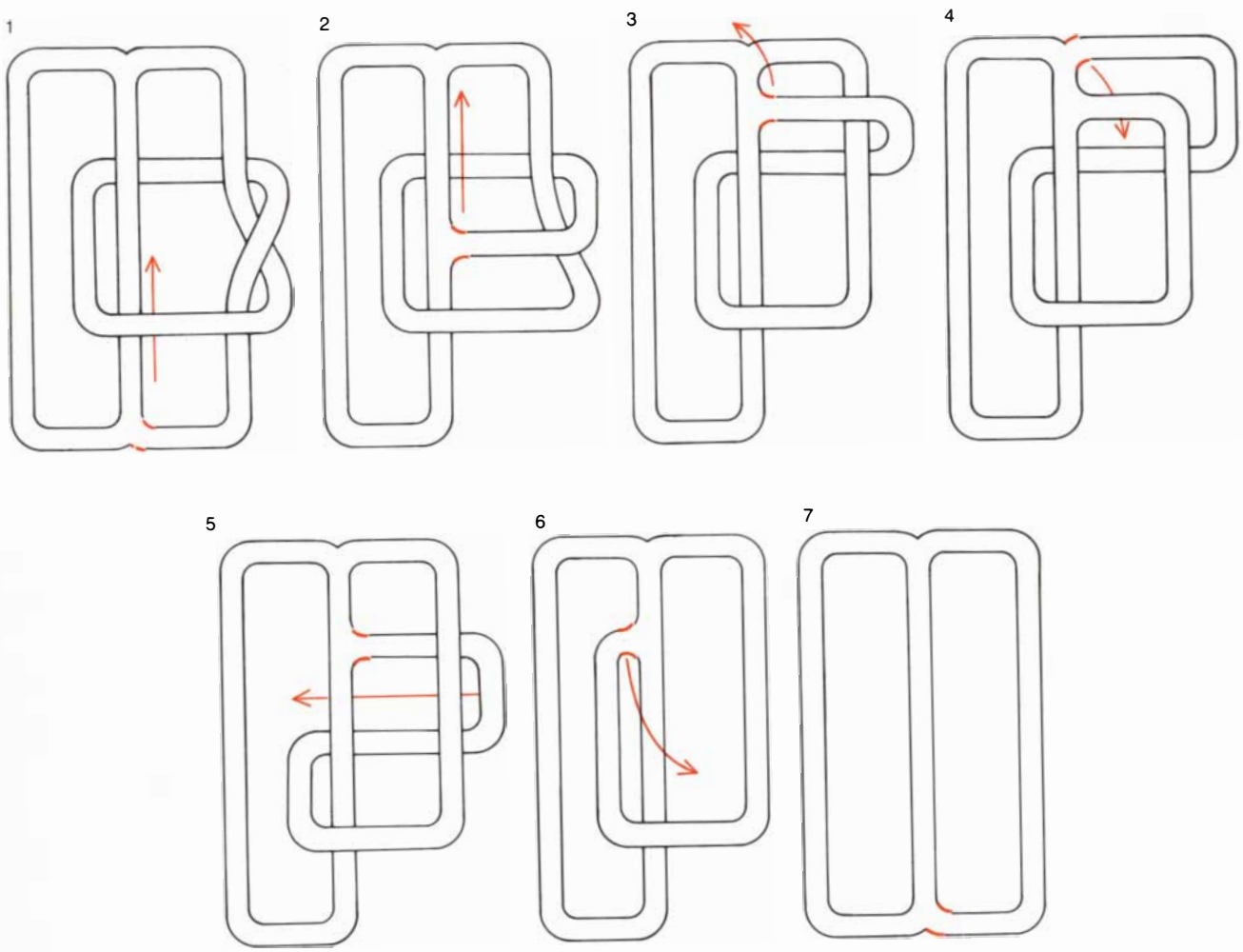
With the exception of these two trivial cases, no general strategy for Chomp is known. There is, however, and this is what makes Chomp so interesting, a simple proof that the first player can always win. Like similar proofs that apply to Bridg-it, Hex, generalized tick-tacktoe and many other games, the proof is nonconstructive in that it is of no use in finding a winning line of play. It only tells you that such a line exists. The proof hinges on taking the single counter at the upper right corner in the opening move. There are two possibilities: (1) It is a winning first move; (2) it is a losing first move. If it is a losing one, the second player can respond with a winning move. Put another way, he can take a bite that leaves a position that is a sure loss for the first player. But no matter how the second player bites, it leaves a position that the first player could have left if his first bite had been bigger. Therefore if the second player has a winning response to the opening move of taking the counter at top right, the first player could have won by a different opening move that left exactly the same pattern.

In short, either the first player can always win by taking the counter at top right or he can always win by some other first move.

"We normally think of nonconstructive proofs in mathematics as being proofs by contradiction," Gale writes. "Note that the above proof is not of that



Solution to torus-reversal problem



Unknotting a two-hole torus

type. We did not start by assuming that the game was a loss for the first player and then obtain a contradiction. We showed directly that there was a winning strategy for the first player. The word 'not' was never used in the argument. Of course we used implicitly the fact that any game of this kind is a win for either the first or the second player, but even the proof of this fact can be given by a simple inductive argument that does not use any law of the excluded middle."

This is essentially all that is known about Chomp except for some curious empirical results Gale obtained from a complete computer analysis of the 3-by-*n* game for all *n*'s equal to or less than 100. In every case it turned out that the winning first move is unique. The illustration on page 111 shows the winning moves for 3-high fields of widths two through 12. Rotating and reflecting these patterns give winning moves on 3-wide fields of height two through 12, because any *m*-by-*n* game is symmetrically the same as the *n*-by-*m* game.

A winning first move on a 3-high field must be one or two rows deep. (A 3-deep bite would leave a smaller rectangle and thus throw the win to the second player.) Roughly 58 percent of the winning first moves are two rows deep and 42 percent are one row deep. Note that the one-row moves either stay the same or increase in width as *n* increases, and the same is true of the two-row moves. A partial analysis of all 3-high fields with widths less than 171 showed that the sole exception to this rule occurs when *n* is 88. The winning first move on the 3-by-88 rectangle is 2 by 36, which is one unit less wide than the winning 2-by-37 move on the 3-by-87 field. "Phenomena like this," Gale writes, "lead one to believe that a simple formula for the winning strategy might be quite hard to come by."

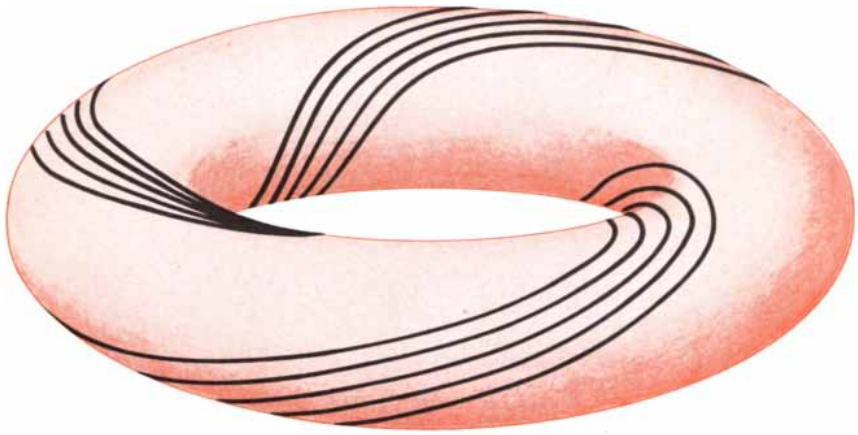
There are two outstanding unproved conjectures:

1. There is only one winning first move on all fields.

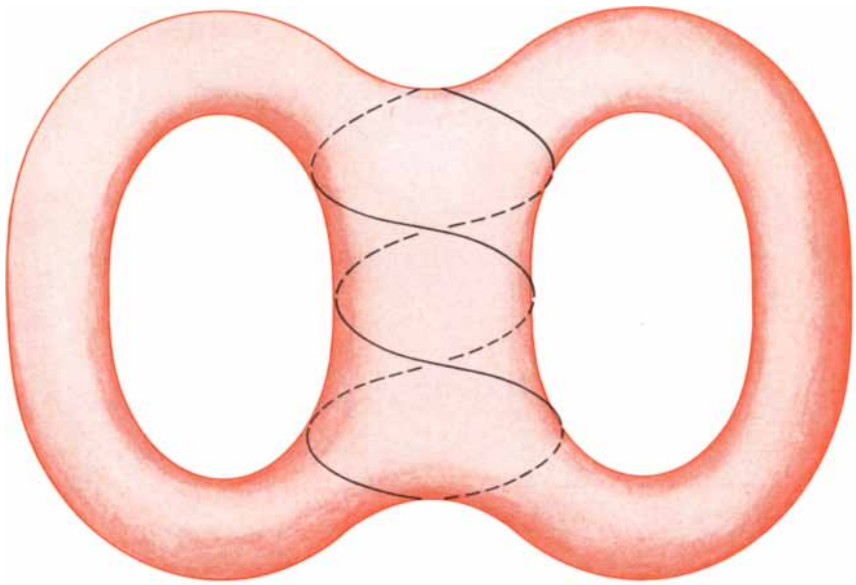
2. Taking the counter at the top right corner always loses except on 2-by-*n* (or *n*-by-2) fields.

The second conjecture has been established only for fields with widths or heights of 3. As a problem to be answered next month, readers are invited to discover the unique winning openings on 4-by-5 and 4-by-6 rectangles.

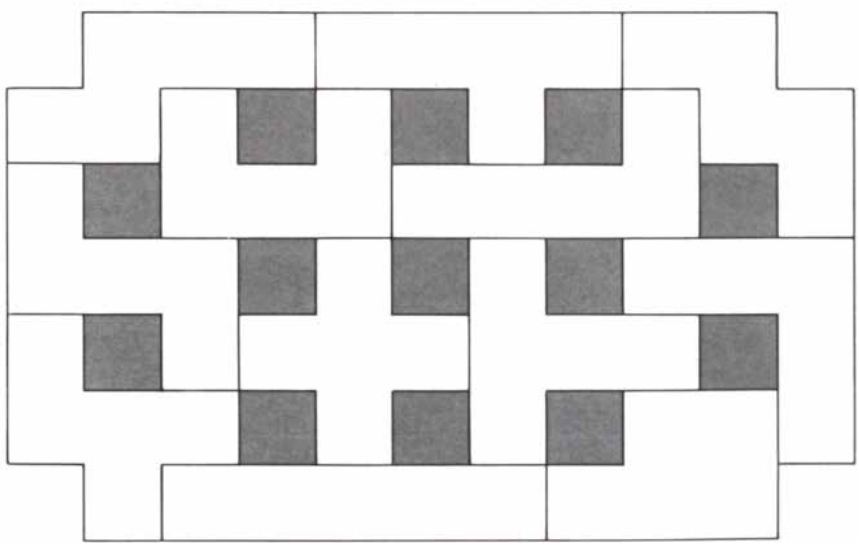
Here are answers to last month's problems. R. H. Bing shows how an internally knotted torus can be reversed through a hole to produce an externally



Knotted, nonintersecting curves on a torus



Rotating slice through a two-hole torus



Symmetrical solution to pentomino problem

knotted torus [*see top illustration on page 112*]. A small hole, *h*, is enlarged to cover almost the entire side of the cylinder, leaving only the shaded strip on the right. The top and bottom disks of the cylinder are flipped over and then the hole is shrunk to its original size.

As in reversing the unknotted torus through a hole, the deformation interchanges meridians and parallels. You might not at first think so because the colored circle, *m*, appears the same in all three pictures. The fact is, however, that initially it is a parallel circling the torus's elongated hole whereas after the reversal it has become a meridian. Moreover, after the reversal the torus's original hole is no longer through the knotted tube, which is now closed at both ends. As is indicated by the arrow, the hole is now surrounded by the knotted tube.

Piet Hein's two-hole torus, with an internal knot passing through an external one, is easily shown to be the same as a two-holer with only an external knot. Simply slide one end of the inside knot around the outside knot (in the manner explained last month) and back to its starting point. This unties the internal knot. Piet Hein's two-holer, with the external knot going through a hole, can be unknotted by the deformation shown in the bottom illustration on page 112.

Answers to the final three toroidal questions are:

1. An infinity of noncrossing closed curves, each knotted with the same handedness, can be drawn on a torus [*see top illustration on preceding page*]. If a torus surface is cut along any of these curves, the result is a two-sided, knotted band.

2. Two closed curves on a torus, knotted with opposite handedness, will intersect each other at least 12 times.

3. A rotating slice through a solid two-hole doughnut is used to produce a solid that is topologically equivalent to a solid, knotted torus [*see middle illustration on preceding page*]. Think of a short blade as moving downward and rotating one and a half turns as it descends. If the blade does not turn at all, the result is two solid toruses. A half-turn produces one solid, unknotted torus. One turn produces two solid, unknotted linked toruses. Readers may enjoy investigating the general case of *n* half-turns.

Wade E. Philpott of Lima, Ohio, was the only reader who sent all 13 solutions to the Diabolical cube, whose six pieces were shown in last September's column. Before this column appeared I had occasion to show the puzzle to John Horton Conway of the University of Cambridge. He mentally labeled the pieces with a

checkerboard coloring, then began testing the pieces rapidly, talking out loud and occasionally scribbling a note. It was like watching Bobby Fischer play blitzkrieg chess. About 15 minutes later he announced that there were just 13 solutions. To distinguish them, designate each piece of the Diabolical cube by the number of unit cubes it contains. There are three ways in which the two largest pieces, 6 and 7, can go:

1. Parallel and side by side. When properly placed, with the 5-piece wrapped around a projecting cube of 6, the 4-piece can go in three places. There are five solutions.

2. Parallel but on opposite sides of the cube. There are two solutions.

3. Perpendicular to each other. Crossing in one way yields four solutions, another way two, or six solutions in all.

Philpott also sent a proof that a pattern of 14 unit holes, each surrounded by eight cells, cannot be achieved with the 12 pentominoes. The proof establishes that at least 59 squares are needed to surround 14 holes. On all such patterns each pentomino will fit except the *P* and *W* pieces. Adding a 60th cell will accommodate only one of the two pieces, proving that the 60 cells of the pentomino set are not enough. Essentially the same proof had earlier been formulated by Joseph Madachy.

Christer Lindstedt of Göteborg, Sweden, had found a 13-hole solution before the problem appeared in September. Other solutions (no two alike) were found by Robert Bart, Bruce Beckwith, Neil E. Beckwith, Greg Buckingham, Andrew L. Clarke, H. I. da Costa, John D. Determan, Victor G. Feser, William J. Flora, William Grolz, Thomas M. Napier, Jack M. Welch, David N. Yetter and Thomas Zaslavsky, and jointly by three readers in Paris, L. J. Francois, J. R. Ponce de Leon Pina and J. F. Vincent.

Only Clarke (of Freshfield in England) found solutions with bilateral symmetry. He sent seven, one of which has two axes of symmetry [*see bottom illustration on page 113*].

Readers interested in Greco-Latin squares may wish to write Joseph Arkin, 197 Old Nyack Turnpike, Spring Valley, N.Y. 10977, for a remarkable construction. A self-taught mathematician working without computer aid, Arkin succeeded last year in constructing an order-10 cube on which three Latin cubes are mutually orthogonal, provided that all three are considered together. The construction leads to an order-10 magic cube on which all orthogonals and the four diagonals are magic.

THE AMATEUR SCIENTIST

An inexpensively modified transistor radio is used in several entertaining experiments

Conducted by C. L. Stong

A scheme whereby an inexpensive radio can be modified with a few accessories and put to work in experiments that fascinate children while leading them to an understanding of various natural phenomena has been devised by Adrian Popa of Newbury Park, Calif. He undertook the venture with his sons Brian and Mark, aged 12 and eight respectively. The radio cost about $4 and the outlay for accessories was trifling. Popa writes:

"Our project began when my sons asked how fast a hummingbird beats its wings when it hovers at our feeder. It occurred to me that perhaps an experiment could be devised that would enable the boys to time the wingbeats themselves. My goal was to keep the experiment and the related construction work at the level of youngsters while making the enterprise complex enough to stimulate original thinking.

"The transistor radio that was the principal instrument for the experiments is slightly larger than a pack of cigarettes. Sets of this kind include a sensitive amplifier that can boost almost any external signal. The amplified signals are transferred into sound by the self-contained loudspeaker.

"Signals from the external source can be fed into the amplifier by a pair of wires connected to the two outermost terminals of the three on the potentiometer that functions as the volume control of the set. (The volume control of some sets also serves to turn the set on and off. In this case the control has five terminals. The first and fifth terminals do the switching. Connect the wires to the second and fourth terminals.) The result of this simple modification is a portable amplifier with a self-contained power supply. The modification does not damage the radio. It can still be tuned to broadcast stations, which my boys often do during long experiments.

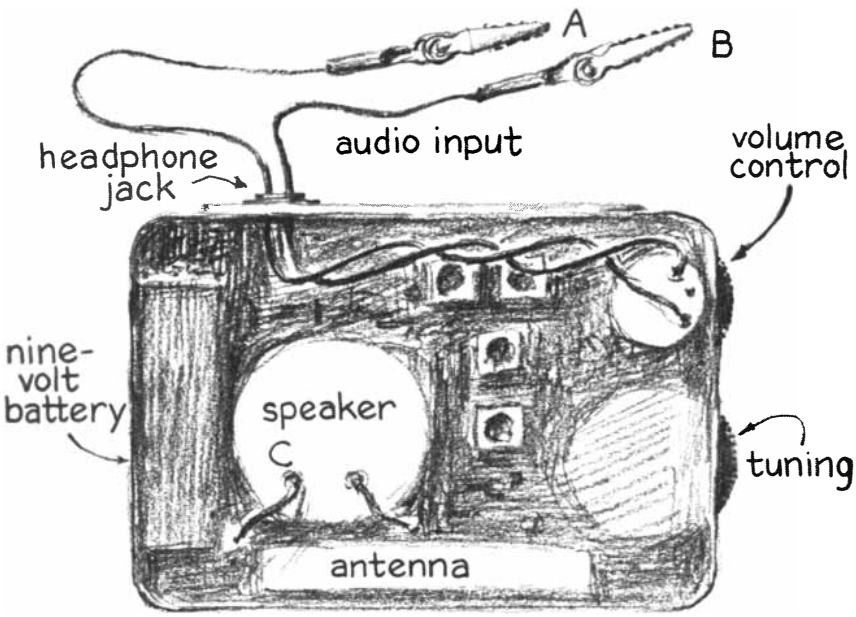
"External signals are amplified by first tuning the dial to a point at which no broadcast is heard. The additional wires are then connected to the signal source. We found that the simplest way to bring the wires out of the set was to thread them through the hole into which one normally would plug an earphone. We also learned that the set will not pick up objectionable noise if the wires are twisted together about two turns per inch. Noise can also be reduced by using shielded wire, but the additional reduction is not worth the effort. We install a pair of small alligator clips on the ends of the wires for convenience in connecting the amplifier to various signal sources.

"Our bird detector is a photovoltaic cell of the silicon type, commonly known as a solar battery. Selenium and cadmium photocells are more sensitive, but they are of the photoresistive type and require an external source of power. Their use would complicate the circuit.

"Cells of the solar-battery type are more sensitive than needed for these experiments. Indeed, they tend to become saturated in bright sunlight. We prevent saturation by covering the cell with a mask containing a small hole, which cuts off most of the incident light.

"We tested the apparatus by clipping the amplifier to the cell and exposing the cell to a fluorescent light. The loudspeaker emitted a hum at a pitch of 120 vibrations per second when the light was on. That is the characteristic rate at which these lamps flicker. The hum stopped when we covered the cell.

"What we had now was an optical receiver. We placed it on the floor near a window of our living room where a shadow is cast by hummingbirds that feed outside [*see bottom illustration on opposite page*]. We detected many interesting sounds that are made by the movements of the birds in addition to the low-pitched hum generated by their flapping wings. The apparatus also served as an excellent alarm that sounded when the birds fed. It enabled us to do other things while waiting for them to arrive. We have enjoyed many hours of bird watching without having to keep our eyes constantly glued on the feeder.



Adrian Popa's modified transistor radio

"Now that we had an optical receiver for detecting the movement of a bird's wings, we needed a clock to time the rate of the wingbeat. We decided to compare the rate with a tone of known pitch by using our ear, just as a piano tuner compares the tone of a piano string to that of a tuning fork. We knew we could generate a flickering light by punching spaced holes near the edge of a disk of cardboard and rotating the holes in front of a flashlight. We could convert the flickers into sound with the optical receiver. Moreover, we could learn the frequency at which the light flickered by multiplying the number of holes in the disk by the rate, in revolutions per second, at which the disk turns.

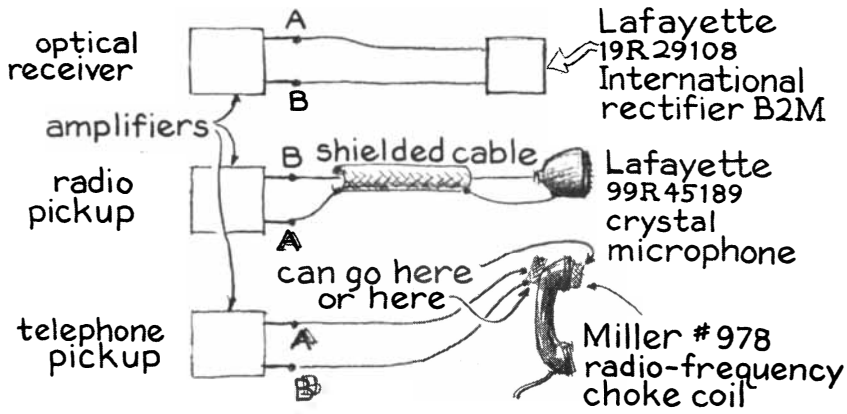
"We used a cardboard disk 14 inches in diameter. Several sets of holes were punched in the edge [*see top illustration on next page*]. We rotated the disk at known rates with a phonograph turntable. The wing rate was determined by shining a flashlight through the holes to the photocell at the same time that the bird signal was present.

"The boys observed that the wingbeat of a hovering bird was close to 41 vibrations per second, as generated by the perforated disk. The rate of 37.5 vibrations per second was a bit low and the 45-cycle rate was high. We were all astonished to learn that the hovering rate of big and little hummingbirds is the same. As the birds accelerated to leave the feeder we observed rates as high as 80 wingbeats per second, the value that is listed in our encyclopedia.

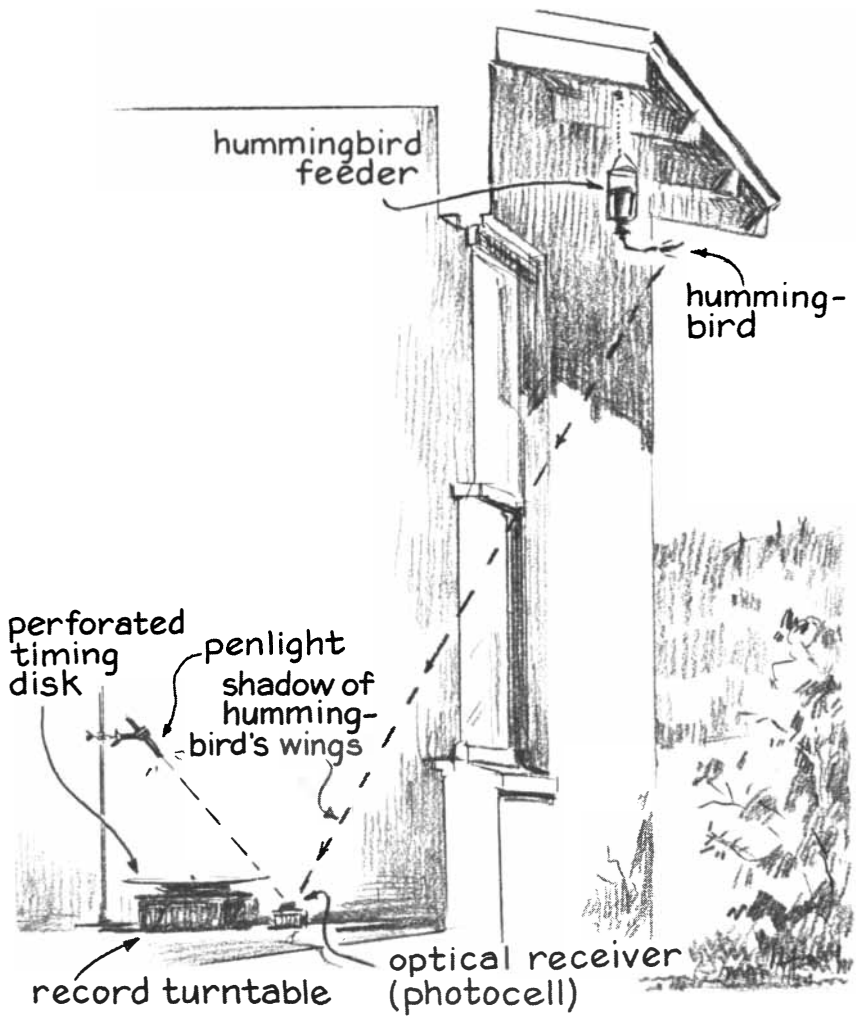
"The birds were not always cooperative. Often they would not hover long enough for us to make a good measurement. We solved this problem with a small tape recorder. We could replay the recorded sounds at leisure as many times as necessary to make an accurate comparison with the frequencies of our standard disk. We are now planning a series of experiments to measure the wing rates of insects. These experiments will require a more complex optical system, which we are now building.

"Our flashlight modulator, in the form of the perforated cardboard disk, led to a number of other experiments. We found, for example, that a series of interesting musical effects can be generated by punching evenly spaced long, short and odd-shaped holes in the edge of the disk. In another experiment we mounted a smaller perforated disk on the shaft of a toy motor. With this motorized light-chopper we could detect the modulated flashlight beam at a distance of several hundred feet. We used the system to send signals in Morse code. At night, by substituting the headlight of an automobile for the flashlight, we could detect the modulated beam at a distance of several hundred yards.

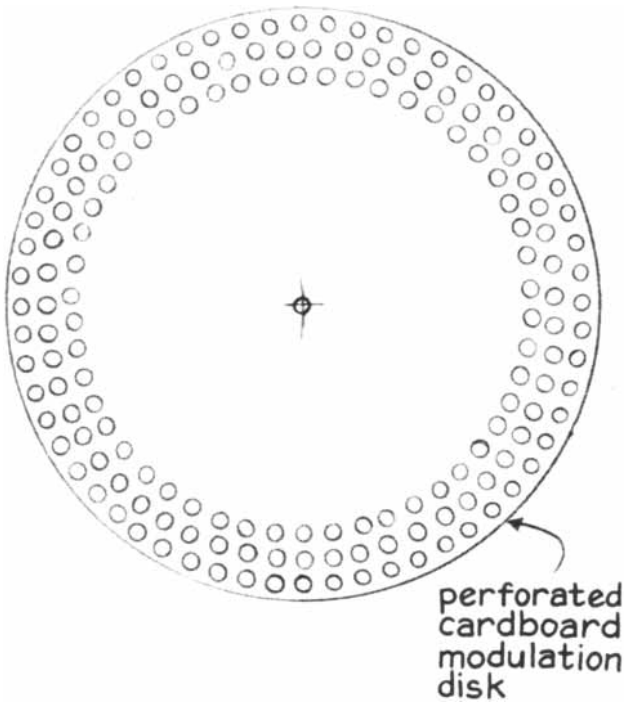
"We also modified the experiment to simulate the direction-finding apparatus of an airport instrument-landing system. Two sets of holes were punched side by side about an inch apart in a disk to generate simultaneously 15 and 30 cycles per second. The perforations were spaced so that one series of holes passed through one half of the flashlight beam and the other set passed through the other half [*see bottom illustration on next page*].



Arrangement of various applications



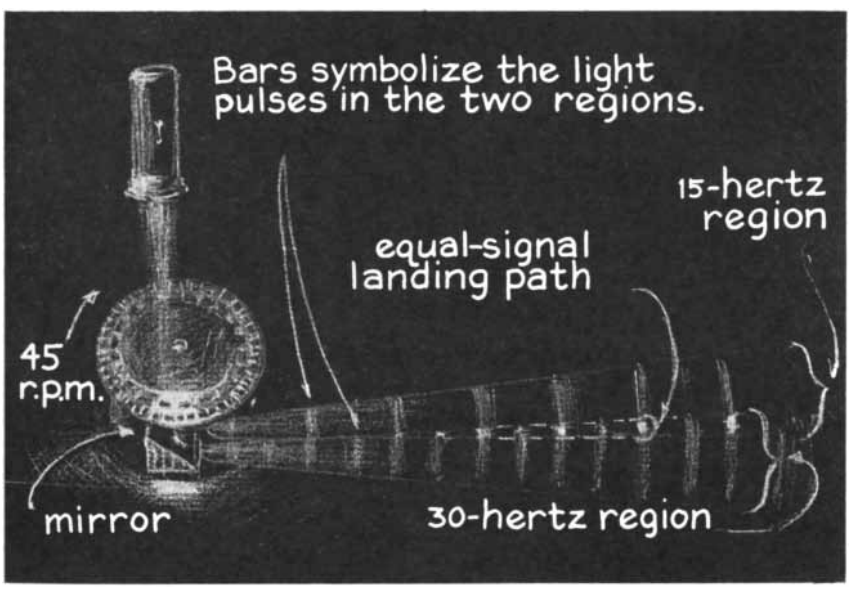
How the hummingbird's wingbeat is measured



record speed	number of holes around disk: 20	40	50	55	60
33⅓ r.p.m.	11.1	22.2	27.72	30.5	33.3
45 r.p.m.	15	30	37.5 low	41.25 close	45 high
78 r.p.m.	26	52	65	71.5	38

frequency (cycles per second)

A modulation disk



Apparatus for simulating a direction finder

"When the optical receiver is held in the beam at either side, one tone is louder than the other. Both tones are equally loud when the photocell is held exactly in the center of the beam. Blindfolded children and adults have used the system to find their way along the beam directly to the flashlight, just as airplane pilots follow an audio-modulated radio beam to the center line of a runway. After this demonstration even schoolchildren in the second grade can understand the concept of an all-weather instrument-landing system.

"We next altered the system to demonstrate the transmission of sound over a beam of light. We made a voice modulator with a one-pound coffee can plus a few other odds and ends that we found at home [*see upper illustration on opposite page*]. We cut the bottom out of the can and replaced it with a sheet of plastic kitchen wrap held in place by a rubber band. On the side of the can we soldered a quarter-inch nut with 20 threads per inch. The nut enabled us to mount the can on a camera tripod.

"Sound waves that enter the open end of the can cause the plastic diaphragm to vibrate. We reflected the beam of the flashlight off the diaphragm to our optical receiver. By talking into the can we could transmit speech of good fidelity up to about 50 feet.

"The plastic diaphragm turned out to be a poor reflector. Most of the light passed through the clear plastic. We tried a diaphragm of aluminum foil. The transmission range increased substantially but the fidelity was poor. I obtained Mylar film .00025 inch thick that was coated on one side with a film of aluminum. Both the quality of the reproduced sound and the range of this modulator were excellent.

"The modulator picked up voices 20 feet from the can and could transmit them several hundred feet on the light beam. It also made a fascinating passive modulator for 'spy' use, which the boys easily related to programs they had seen on television. I have substituted a laser beam for the flashlight in both the simulated system and the voice modulator for demonstrations at schools and colleges to illustrate my field of laser communications.

"The transistor radio can of course amplify signals from transducers other than photocells. For example, an inexpensive crystal microphone can be connected to the amplifier by a long shielded cable for monitoring the sounds made by a baby in a nursery. An intercommunicating system can be made with a pair of such microphones and radios. A small

coil of wire, such as a one-millihenry radio-frequency choke coil, picks up a substantial signal when it is placed near the receiver of a telephone. When the coil is connected to the input of the transistor amplifier, both sides of the telephone conversation are clearly reproduced.

"The amplifier can be converted into a tone generator by connecting one terminal of a capacitor to one terminal of the loudspeaker or to one terminal of the earphone jack and the other terminal of the capacitor to one terminal of the volume control [*see bottom illustration on this page*]. The capacitor feeds energy from the output of the amplifier back to its input—an action that generates continuous oscillations. The unit oscillates best when the capacitor is connected to a selected terminal of the potentiometer. Try both terminals of the potentiometer and make the connection to the one that works best. The capacitor can be of any value between 100 picofarads and one microfarad.

"The pitch of the tone varies inversely with the size of the capacitor. The pitch can also be varied through an appreciable range by adjusting the volume control. The oscillator can be used to generate an alarm signal. We have also employed it for practicing Morse code. The tone can be turned on and off by inserting a switch or a telegraph key in series with the capacitor.

"Another series of experiments that interested us involves the detection of military and airport radar stations. We pick up the radar signals by an antenna made of a paper clip [*see top illustration on next page*]. The signals are detected with a microwave diode of either Type 1N21 or Type 1N23. Both diodes can detect all the radar bands in common use, but the 1N21 is best for signals of 10-centimeter wavelength, called the *S* band, and the 1N23 is best for the three-centimeter wavelength, or *X* band.

"The diodes detect the sweep rate (the rate at which the beam of the radar scans the area) and the rate at which the beam pulses. Typically radar sweeps an area about twice per minute and the beam pulses about 400 times per second, a frequency approximately equivalent on the musical scale to *G* above middle *C*. The paper clip functions as an omnidirectional antenna: it picks up signals almost equally well from all directions. We have detected airport and Nike missile radars up to 10 miles away.

"The apparatus can be employed to demonstrate the principles on which highly directional antennas work. For example, the direction of a radar station from the experimenter can be established by putting a corner reflector on the paper-clip antenna. Cement aluminum foil to one side of a rectangular sheet of cardboard four inches wide and 12 inches long. Fold the strip to make a right angle. Support the paper-clip antenna inside the corner parallel to the fold. Brackets to support the antenna at this position can be made of cardboard.

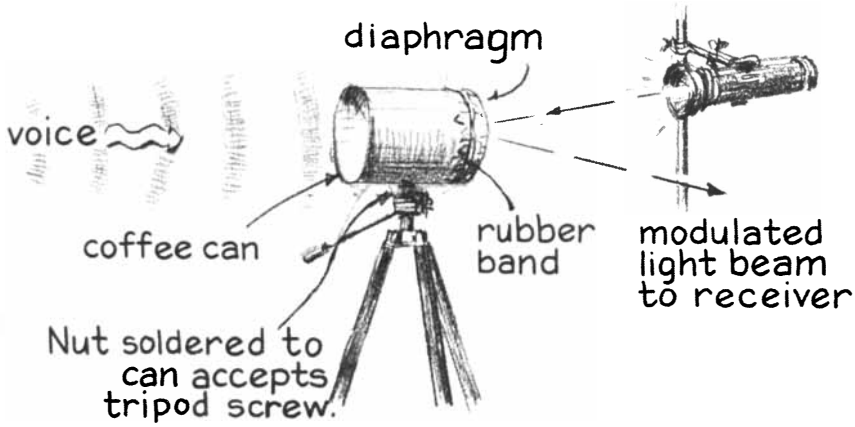
"The strength of the received signal will be at a maximum when the distance between the antenna and the corner fold is equal either to half of the length of the incoming radio wave or to one and a half wavelengths. The signal strength will approach zero when the separation is one wavelength. In the case of 10-centimeter waves the signal will be heard at maximum volume when the antenna is either two or six inches from the reflector. The signal will be heard loudest when the reflector is pointed directly toward the radar station.

"In a similar way one can determine with this apparatus the location of large metal objects that reflect radar waves, such as aircraft and water tanks. Of course, the set is not nearly so sensitive as commercial receivers that are designed to pick up radar signals, but it clearly demonstrates the optical nature of radio waves. The fact that radio beams are a function of antenna geometry suggests other experiments, such as making dish-shaped reflectors (like those of optical and radio telescopes) and miniature versions of conventional television antennas.

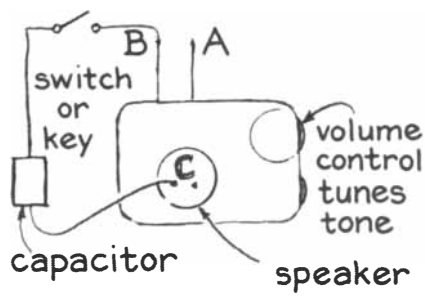
"An antenna of an entirely different type can be used for picking up the eerie signals, generated by lightning, that bounce back and forth between the Northern and Southern hemispheres of the earth. When these signals are transformed into sound, they are musical, but they differ in character according to the circumstances of their origin. Some are mere clicks. The click is heard when lightning strikes within 100 miles or so of the observer. If the stroke occurs 1,000 or so miles away, some of the emitted radio waves may reach the observer by bouncing off the ionosphere. They are heard as a short, high-pitched note known as a 'tweek' or a 'chink.'

"If the stroke occurs in the Northern Hemisphere at the middle or higher latitudes, a portion of the energy may be propagated along a line of the earth's magnetic field. The flux line guides the signal in an arching path to an altitude of some 8,000 miles; there it bends downward and returns to the surface in the Southern Hemisphere. Here the energy is reflected, returning along the same arching path to the general area where the stroke occurred.

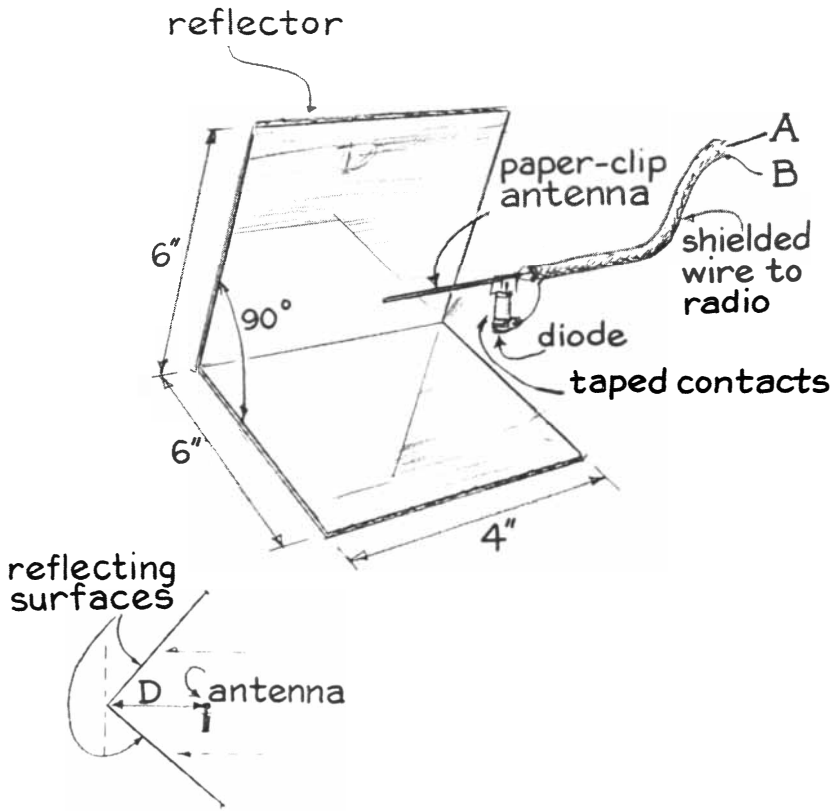
"In the course of this long excursion the band of frequencies constituting the signal becomes separated in time. The shorter waves travel through the ionized atmosphere at higher velocity than longer waves. The transformed signal is heard by the observer as a piercing tone of falling pitch, known as a 'whistler.' The returned energy may be reflected again and again. In effect, the 'click' generates a series of whistlers, each per-



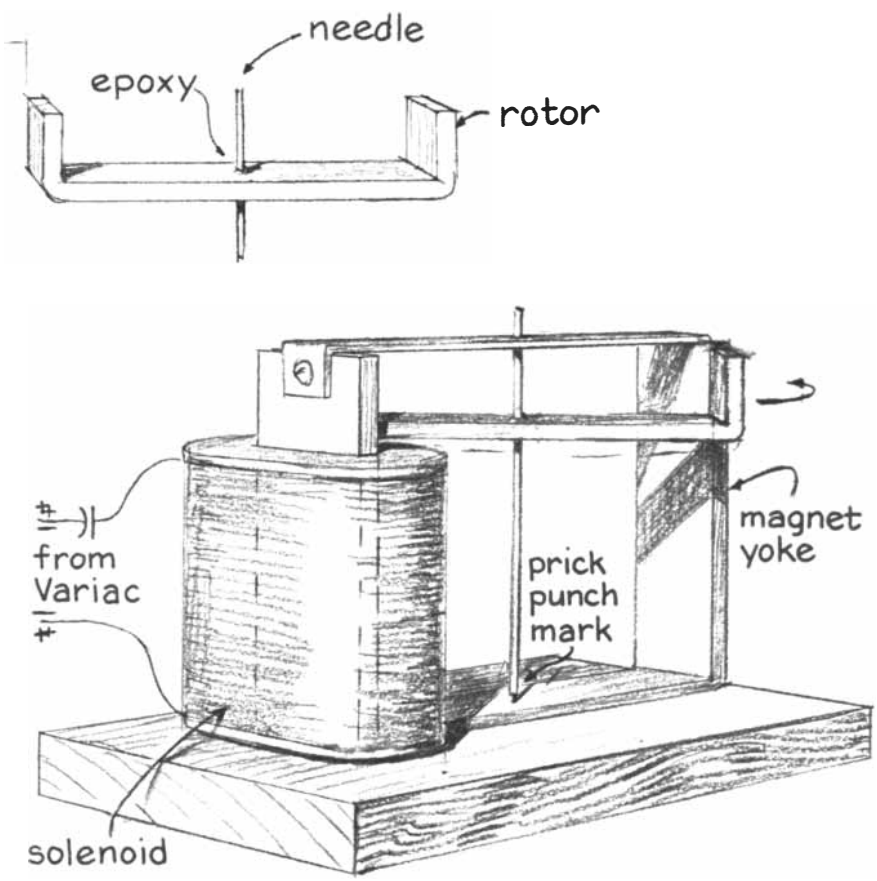
Transmitting sound on a beam of light



Arrangement for generating a tone



Directional antenna for picking up radar



Harry E. Stockman's parametric motor

sisting longer than its predecessor and with greater separation between tones of high and low pitch. The tones of a whistler as reproduced by the amplifier may range in frequency from about 6,000 to 500 vibrations per second.

"To hear these effects (as well as the 'dawn chorus,' an unexplained galaxy of birdlike sounds that appears to accompany magnetic storms) the experimenter must select a location remote from electrical power lines and other sources of man-made electrical noise. Attach to one lead of the amplifier a copper wire about 300 feet long. The wire serves as the antenna. Attach the distant end of the wire through an insulator to the highest available object, so that the antenna slants upward from the amplifier. Connect the other lead of the amplifier to a metal stake driven into moist soil. Also connect a one-millihenry choke coil in series with a one-microfarad capacitor of the Mylar type and attach the combination across the input of the amplifier. Whistlers are heard frequently in the middle and upper latitudes during seasons of maximum thunderstorm activity.

"In our more recent experiments we have moved on to more complex projects such as radio holography and Doppler radars. For these investigations we still use slightly modified transistor radios. We hope that amateurs will join us in the fun of devising still other ways to probe nature with these versatile sets."

A physicist investigating phenomena mathematically occasionally writes an equation that suggests a novel device of no obvious use. An example of such a device is an electric motor invented recently by Harry E. Stockman of Arlington, Mass. The machine consists essentially of a rotor in the form of a soft bar of iron supported at its middle by a vertical shaft that is free to turn between the poles of an electromagnet [*see bottom illustration at left*]. The magnetic field is provided by a solenoid that operates on alternating current. When the solenoid is energized by a 60-hertz current, the polarity of the magnet reverses 3,600 times per minute.

If the rotor consisted of a thin strip of hard steel instead of the soft iron bar, and if the strip were given an initial spin at the rate of 3,600 revolutions per minute, the steel rotor would become permanently magnetized, fall into lockstep with the alternating magnetic field and continue to rotate synchronously with the field. The synchronous motors of electric clocks operate on this principle. Stockman's motor runs asynchronously!

When the soft iron rotor is given a start, it continues to turn at a speed that varies roughly with the applied voltage. The tips of the rotor experience a strong force of attraction as they approach the poles of the magnet. The force decreases as the rotor moves away from the poles.

The variation of the force is explained by a similar variation in the inductance of the solenoid, reflecting the property of solenoids, which is to oppose electric current. The inductance of the coil is maximum when the poles of the magnet are bridged by the soft iron bar. The current and the magnetic force of attraction decrease.

Conversely, when the iron bar turns away from the poles, the inductance approaches the minimum and the exciting current and the force of attraction increase. In actuality the force of attraction does not become minimum at the instant the bar reaches its middle position between the poles of the magnet. The field cannot change instantly. Accordingly the bar is strongly attracted as its ends swing toward the poles of the magnet and less strongly attracted as they coast away.

Inductance is known as a parameter of electric circuits: a quantity that is normally constant but that can be varied by circumstance. In the case of this device the determining circumstance is the position of the rotating bar of iron. For that reason Stockman refers to his invention as a "parametric" motor.

The core of the electromagnet can be bent from a strip of flat iron 1/8 inch thick and one inch wide, making a flat-bottomed *U* with upright legs about 2½ inches long. The rotor can be bent from 1/2-inch stock. The ends of the rotor should be bent upward to form legs 1/2 inch long that leave an air gap of about 1/32 inch at each end when the rotor is placed between the poles of the magnet. The dimensions are not crucial.

The solenoid can be made of approximately 2,700 turns of No. 26 enameled copper wire. The assembled electromagnet will have an inductance of about .3 henry. Stockman connects a capacitor of eight microfarads in series with the solenoid and energizes the unit through a variable-voltage transformer. The capacitor is not strictly necessary, but the operation is greatly facilitated by taking advantage of resonance. The motor will operate on about five volts. The speed is nonsynchronous and roughly proportional to the applied voltage up to about 10 volts. At higher voltage the machine tends to lock in step with the frequency of the power source and to operate as a conventional synchronous motor.

GET THE BRAND NEW ONES--OR CATCH UP ON A FEW GREAT ONES YOU'VE MISSED

by starting your trial membership now in

THE LIBRARY OF SCIENCE

It's the easy—and economical—way to keep up with the very best in scientific reading and reference. More. Books to think about. Books for sheer enjoyment. All here. First. In the one scientific book club that covers *every* discipline in the expanding world of science.

Take any 3 books

(values to $49.50)

all for only $3.95

if you will join now, and agree to take only 3 more books (at member discount prices) within the next 12 months.

(retail prices shown)

63950. THE NATURE OF THE CHEMICAL BOND. *Linus Pauling.* Third Edition of this science classic with all chapters expanded to include changes in such matters as bond-energies and bond-lengths. **$17.50**

36530. BODY TIME: Physiological Rhythms and Social Stress. *Gay Gaer Luce.* All that is currently known about the internal physiological clockwork which governs our lives, determines our moods, affects our susceptibility to disease, our sense of pleasure and pain. **$6.95**

50291. A GOD WITHIN. *René Dubos.* Moving testament of faith in the destiny of man. For scientists, it creates a new, scientific theology of the earth which man must have if he is to survive. **$8.95**

64580. THE NEW PROMETHEANS. *Robert S. de Ropp.* The continuing battle between the creative and destructive forces of science with application to the scientific developments of the recent past. **$7.95**

34620. ASTRONOMY: Fundamentals and Frontiers. *Robert Jastrow and Malcolm Thompson.* The author of the best-selling *Red Giants and White Dwarfs* and Director of the Goddard Institute for Space Studies and an associate offer a refreshingly new approach to a subject which is enjoying an amazing upsurge of interest. **$12.50**

MEMBERSHIP APPLICATION

The Library of Science 2-33Y
Riverside, N.J. 08075

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.

3 books for $3.95. (write in numbers)

_____ _____ _____

Some expensive books count as 2 choices.

Name _____

Address _____

City _____ State _____ Zip _____

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

42150. CHAMBERS DICTIONARY OF ELECTRONICS AND NUCLEONICS. *Edited by L. E. C. Hughes, R. W. B. Stephens and L. D. Brown.* More than 9,000 individual entries, a comprehensive list of abbreviations, acronyms and symbols and over 100 pages of data tables. Skillfully arranged and cross-referenced. **$14.50**

82640. TECHNOLOGY IN THE ANCIENT WORLD. *Henry Hodges.* A much-needed technical survey of the tools and processes of early civilizations. Delightful and useful volume for the professional and nonprofessional. **$10.00**

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**

56120. INTRODUCTORY COMPUTER METHODS AND NUMERICAL ANALYSIS. *Ralph H. Pennington.* Latest revised edition of a highly successful work. Stresses recent changes in programming and application concepts. Requires only a basic mathematical maturity for the reader to get maximum value. **$10.95**

41940. THE DEEP WELL. *Carl Nylander.* Philosophical and archaeological explorations of "the well of the past" in Egypt, Iran, Canaan, Greece and elsewhere. A dedicated scientist who communicates his own enthusiasm and sense of wonder to the reader. **$7.95**

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. **$15.00**

39891. THE COMPUTER FROM PASCAL TO VON NEUMANN. *Herman H. Goldstine.* One of the world's leading mathematicians describes graphically the development of the computer. A highly personal account by one of the "fathers" of this twentieth-century miracle. **$12.50**

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

87410. WORDS FOR BIRDS: A Lexicon of North American Birds with Biographical Notes. *Edward S. Gruson.* Amusing, scholarly, magnificently researched—an ornithological, etymological, biographical and historical guide to 800 North American birds, including any number of bizarre stories about them. **$8.95**

32571. ALBERT EINSTEIN, CREATOR AND REBEL. *Banesh Hoffman with the collaboration of Helen Dukas.* New and engrossing best seller about the man whose work has been called "the greatest single stride science has ever made." **$8.95**

37570. CHEMISTRY FOR THE MILLION. *Richard Furnald Smith.* Uses historical framework to present a highly entertaining picture of what chemistry is (and how it got to be what it is), what chemists do, and the significance of chemistry for the modern world. **$7.95**

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**

50380. GLACIAL AND QUATERNARY GEOLOGY. *Richard Foster Flint.* Just-published masterwork on climatic change and glaciation over the last ten million years. Required reading for anyone interested in geology, geophysics, meterorology, botany, and conservation. *Counts as 2 of your 3 books.* **$24.95**

40110. THE CONDENSED CHEMICAL DICTIONARY. Eighth Edition. *Edited by Gessner G. Hawley.* Revised, expanded, updated and full of new features. Contains 18,000 entries, more than 1,000 pages. *Counts as 2 of your 3 books.* **$27.50**

32410. AMERICAN INSTITUTE OF PHYSICS HANDBOOK. *Coordinating Editor, Dwight E. Gray.* Newly published third edition contains latest research reports on such topics as acoustics, electricity and magnetism, optics, atomic and molecular physics. Nearly 140 contributors. *Counts as 3 of your 3 books.* **$49.50**

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

56540. INTRODUCTION TO MATRIX ANALYSIS. Second Edition. *Richard Bellman.* This second edition of the single, most widely praised work in the field stands head and shoulders above the competition. A brilliant modernization of a brilliant forerunner. **$14.75**

63860. THE NATURE AND GROWTH OF MODERN MATHEMATICS. *Edna E. Kramer.* Lively, ingenious new approach to the history of developing mathematical concepts. Emphasis throughout is on the clarification of 20th-century ideas. *Counts as 2 of your 3 books.* **$24.95**

64771. NO LONGER ON THE MAP. *Raymond H. Ramsay.* A delightful voyage to "places that never were" with a well-informed and entertaining pilot. Profusely illustrated with hundreds of old maps. **$10.95**

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**

BOOKS

George Ellery Hale: promoter of American astronomy and counsel to the Government

by Philip Morrison

THE LEGACY OF GEORGE ELLERY HALE: EVOLUTION OF ASTRONOMY AND SCIENTIFIC INSTITUTIONS, IN PICTURES AND DOCUMENTS, edited by Helen Wright, Joan N. Warnow and Charles Weiner. The MIT Press ($17.50). ASTRONOMY AND ASTROPHYSICS FOR THE 1970's, VOLUME 1: REPORT OF THE ASTRONOMY SURVEY COMMITTEE. National Academy of Sciences ($4.75). American science and the high technology with which it is so strongly coupled extend today from the computer bases across the Mekong to the Sea of Serenity. Yet it is easy to name important sites still unique in American science that were the conception of one man long ago. The 200-inch telescope on Palomar Mountain and the beautiful Washington headquarters of the National Academy of Sciences both display his name, officially for the one, less formally for the other. He was, of course, the astronomer-planner George Ellery Hale. *Floruit* from the birth of the University of Chicago to the eve of World War II; in his lifetime the driving force of astronomy in America, and in the 1917 war the most conscious of those who deployed American science for war. Helium, Edgewood Arsenal and the Liberty engine were the World War I radar and nuclear-weapons laboratories, and Hale's National Research Council was the Office of Scientific Research and Development of that simpler time. (The National Academy of Sciences itself was born out of Lincoln's need for official advice, "without compensation," on science and technology.)

Helen Wright published a full-scale biography of Hale only a few years ago. This large new volume, all three of its editors well known for concern with the history of modern science, displays Hale again, through direct reproduction of source materials plus a series of four expert essays evaluating the fields of Hale's interest in the light of later events. The editors have done a first-rate job; the book is evocative and readable, and its facsimiles and photographs are records more accessible and eloquent, if less complete, than a normal biography is. The materials preserve a celebration of Hale's centennial by the American Association for the Advancement of Science and the American Institute of Physics in 1968. (One is somewhat disturbed by the fact that the photographs and documents are not given as precise a provenance as a reader should expect from historians.)

It is fascinating to read the letter from Einstein to Hale in 1913, in which Einstein inquires whether by Hale's techniques star positions might be measured reasonably close to the sun without an eclipse. Hale offered no hope; he felt that the eclipse method, already being actively planned, was "very promising." Here you see the California Institute of Technology as Throop Polytechnic Institute, in an architect's drawing, still fairly recognizable; Alvan Clark proudly beside the Yerkes Observatory's 40-inch lens, five years in the grinding; the 55-foot tube of the 200-inch telescope lying on the shop floor at Westinghouse the day the last bolt was ceremonially fitted, with Einstein visibly present; Bernhard Schmidt before *his* grinding tables, and a National Academy meeting of a century ago, with Lewis Henry Morgan, Joseph and Mary Henry, a very young A. A. Michelson, Simon Newcomb, George Engelmann, Benjamin Silliman and other notables.

Hale found the National Academy a rather enfeebled body, without funds or headquarters. Prestige alone was what it offered, and Americans did not gain much prestige from a powerless and by no means ancient Academy. He saw in it a great future, lobbied the rich and appealed to the public in its name. He started its journal, the *Proceedings;* in 1916 he persuaded President Wilson to ask the Academy to set up the National Research Council "to encourage both pure and applied research for the ultimate end of the national security and welfare." The council's wartime success led to its perpetuation by executive order ("Hale wrote the initial draft") after 1918. The foundations then endowed the Academy, and the marble building was dedicated ("a temple of science") in 1924. Hale had chosen the architect and worked on the design.

By now the Academy, the center of a complex establishment, has become a substantial producer of plans, conferences, advice, compilations, coordination and interpretation. Its executive arm, the National Research Council, spends about $30 million per year, through committees of a truly formidable diversity, on everything from international studies of the coelacanth to the properties of bituminous paving. Government is of course its principal client. The output is words: a spate of reports flows constantly. The *Proceedings* continues; although it has been a decisive journal in the rise of molecular biology, it has not fulfilled Hale's hopes for "covering the whole range of science" with "important advances and the chief results." Nor are the myriad reports, almost always the work of committees and panels, very good reading. Some of them are important and most of them are useful, but they rarely transcend their hierarchical origins and their compromise opinions, expressed in the grayer tones of careful impersonality.

The thin volume on astronomy is a genuine exception. It has something exciting to say, and it manages to say it excitingly. Even its visual impact is worthy: the wonderful, if well-known, American Science and Engineering, Inc., and High Altitude Observatory montage of the sun seen by X rays and visible light is here, with the equally remarkable photograph of the "mice" galaxy pair and a not yet published computer drawing (made by the brothers Alar and Juri Toomre) that faithfully simulates the strange mouse tails by means of a computer-calculated galaxy collision.

To cite the report: "There has been at least one central theme of the last seven years, which makes it like the age of Galileo. It is the discovery of the ex-

istence... of a high-energy, explosive universe.... Two universes coexist—hot and cold. The 'hot' involves phenomena of explosion.... By 'cold' we mean ordinary stellar and interstellar matter, with temperatures from 50 million degrees down." Some 150 American astronomers, organized into panels covering some 10 specialties (radio, optical, statistical, theoretical and more), played some part. (Volume 2 will be a thick, detailed support of the recommendations containing the data from the panels.) This brochure devotes its pages to a sharp and reasoned case for expanding astronomy to employ fruitfully the now strong current of new Ph.D.'s toward this subject in its second Galilean age.

After all, the finest optical instrument now in all the world was designed and funded (it was Hale's work) some 40 years ago. The cost today of replacing Palomar with all its present modern instrumentation is estimated at $25 million. (The first cost was $6 million in hard dollars.) The design was a beauty. It was the first to use a big Pyrex mirror, a ribbed mirror structure, oil-pad bearings and the open-tube design of Mark Surrurier, which keeps the tube ends parallel even though the tube sags. New big telescopes are few and slow to come: no money.

The report makes four documented recommendations for highest priority, achieving a unanimity rare in a disparate profession. First of all, it asks for a Very Large Array, an aperture-synthesis radio telescope, equivalent to "a radio 'eye' 20 miles in diameter," now beginning in the Plains of San Augustin. The optical program needs advanced development of modern electronic auxiliaries, both for servo and data-treatment purposes and to transcend photography, as much as it needs new large telescopes for input. Infrared astronomy, where there is most flux and most ignorance, needs a doubled support, and the large orbiting packages now under construction for study of X rays and gamma rays from space in the late 1970's close the first short list.

Current astronomy expenses are about $70 million per year for ground-based work and three times as much from the National Aeronautics and Space Administration for space work. The Ph.D. flow is strong: good young astronomers are coming in plenty (half with physics degrees). The entire program—the four highest-priority items and seven others—would be made possible with an addition of about $80 million per year to Federal support of basic research in astronomy for the next decade. This would break down into about 60 percent in the space-based areas, with the rest on ground-based work, but including aircraft, balloons and sounding rockets. The decision lies beyond the astronomers, of course, or the National Academy. The public can perhaps observe that the U.S. Air Force costs taxpayers $80 million, not per year but about every 30 hours all year long.

Beyond the decade, two directions are pointed out. The one nearest at hand is the hope for construction and launch into earth orbit of a large space telescope, say 10 feet across, in the 1980's. Operated from the ground with precision over a very wide band of wavelengths, this enduring observatory in space would once and for all rend the veil the atmosphere casts on astronomy. It would cost $1 billion spent over a decade, an estimate well based on earlier studies. Beyond that, "in the relatively near future, we foresee... major facilities... and the operation of a project that will have as its goal the detection of intelligent life elsewhere." Hale would have applauded that remark; he counseled against making small plans.

There is no author credited, but Jesse Greenstein of Cal Tech was chairman, and Bruce Gregory, an astronomer on the staff of the National Research Council, was executive secretary of this expert 23-man committee. It was one panel that knew how to write.

The N.A.S. building has a large mural, based on the fire-bringing of Prometheus, and an eloquent citation (the cultured Hale surely approved it) from the Aeschylus tragedy, in which the Titan, by no means humbly, recounts all he has done for men:. "... hear the whole matter in a word: all human arts are from Prometheus." It does not seem to have impressed Hale and the energetic Academy builders of that expansive day that Prometheus was chained to his rock for those same high deeds. The trilogy is, like history, incomplete: we have never seen how Aeschylus handled *Prometheus Unbound*.

UNDERWATER SCIENCE: AN INTRODUCTION TO EXPERIMENTS BY DIVERS, edited by J. D. Woods and J. N. Lythgoe. Oxford University Press ($13.75). ARCHAEOLOGY UNDER WATER, by George F. Bass. Penguin Books ($1.95). Blood looks green in the deep, since water will not pass red light, and it is common for a "diver to complain that the water changed from a blue-green color near the surface to grey near the bottom." What changed, of course, was the intensity of the light—easily dropping a billionfold in 30 meters. The diver can reach that depth in a minute, but his eyes take about 25 minutes to adapt enough to make out objects on the bottom. In his switch to night vision he loses color vision. At night and in the dim depths all fish are gray. Preadaptation would help, but it is easily undone by an unshielded flashlight on the bottom.

The volume from which this story comes is an unusual collection of papers by eight British scientist-divers aimed at telling the lay reader and the nondiving scientist what their craft has accomplished in the past decade or so. The approach is practical but quantitative throughout, with much data and many references to the original work. The emphasis is on techniques and planning for further work, not on old results. Plants and fish, human perception and behavior, archaeology (an admirable brief survey), the forms of the bottom and the local undersea weather are all treated by experts. (Photography and the physiology of diving are only briefly treated.) There is a clear introductory chapter on the diver's apparatus and methods and another on the psychological performance of the diver at depth, including how to test what he can and cannot do.

It is hard for humans to work where fish dwell. They can do it, and "diving to normal depths is NOT dangerous," although it is a high-risk activity. Professionals, working in groups, well-trained, robust, equipped by navies or large firms, perform better than the scientist engaged in a solitary or a few-handed effort. One might think of the substitute for diving: "hire somebody else." "This is ... only sense when there is an exactly defined task to do, but this is not often true in research." Consider the famous wreck off Antikythera, which furnished its sponge-diver discoverers such treasures as the only classical computer and a bronze youth thought to be by Lysippos. The brave divers of Symē found the wreck by chance and raised objects by groping around until they touched something that could be hauled to the surface by rope. No maps were ever made, no archaeologist ever dived. That was in 1900, and the men worked in hard helmets at a depth of 180 feet, twice a day, five minutes at a time; "even so, one diver died and two were permanently paralyzed," as George Bass writes in his fine paperback. We shall never know much about that remarkable cargo. Today's aims are higher.

Nowadays it is self-contained diving, with modified versions of the Cousteau-Gagnan aqualung of the 1940's, that most diving scientists use. The naval divers have bigger, closed systems, rebreathing air that is cleansed of carbon dioxide. They avoid bubbles and work more quietly and longer, with controlled gas mixtures. Helium-oxygen atmospheres reduce the depth narcosis and "make a significant improvement in mental clarity at depth." Simpler versions of such devices are under development. The professionals nowadays use underwater houses, avoiding the severe time limits of decompression schedules, which demand long stops on the way up. The scientists have approached that: a Perspex hemisphere filled with air from the surface on four weighted legs is easy to set up and very useful as a sea-floor telephone booth, equipment store and quick refuge for a tired or confused diver.

Speech is difficult for the diver. This is not only the Donald Duck effect of helium mixtures but also the inability of the diver to speak clearly because of the acoustic load of the confining mask. He tries to modify his speech to compensate for the mask effect but achieves only a kind of diving dialect, hard to understand, severely distorted even for him by reflections at the air-water interface.

Ingenious surveying devices have been designed. One level used an inverted-*U* tube full of air that bubbled whenever the diver's end was placed at exactly the level of the reference end. Another scheme used a simple three-dimensional scale, the edges of a box frame marked out by rulers. Getting the box in any two views from a single camera allows accurate measurement of the position in three dimensions of anything near the frame, without accurate knowledge of where the camera was. One heroic diver simply used a standard theodolite in the clear Mediterranean, "dismantling it and soaking the component parts in kerosene overnight after each dive."

"Until recently, the diver's only contribution to micro-oceanography lay in his use as a laborer responsible for setting up current meters and thermometers...." The folklore of divers, however, was rich in accounts of the flows and temperature structures of the sea. J. D. Woods of the British Meteorological Office reports quite technically and with considerable charm his detailed studies made by divers using carefully designed dye tracers to study the interplay of laminar and turbulent flow on the ocean floor. They employed standardized packets of fluorescein dye moored at a fixed level. They photographed the sheet of steady flow on an "inversion," noting that here and there thin rafts of sheet turbulence appeared, causing rolls and clouds of dyestuff. The topic is austerely hydrodynamical, but the photographs are striking and the quantitative result is that "the most vivid support for the... theory comes not from the laboratory but from the ocean." Once again it is made plain how understanding and ingenuity can extend the range of experience beyond any expectations. The book is a good one for a browsing reader and an asset to anyone who might need to dive or work with divers.

The little Bass book, a splendid chronicle of the entire science of underwater archaeology by one of its most successful practitioners, is a fine personal mix of know-how, reliable history, adventure and prospect. It is a revision (1969) of a well-known book first published in 1966, available cheaply now for a larger audience: anyone who likes travel in time or in space or who admires the methods and the achievements of contemporary archaeology. From overturned canoes of the voyagers in white-water Minnesota and Ontario streams to the sacred well of Chichén Itzá to the ports of the Aegean or of Jamaica, the testimony of the spade has been confirmed by the witness of the suction hose and the diver's careful hand.

NEUROPOISONS: THEIR PATHOPHYSIOLOGICAL ACTIONS. VOLUME 1: POISONS OF ANIMAL ORIGIN, edited by Lance L. Simpson. Plenum Press ($22.50). As grimly fascinating as the flick of a striking cobra itself, this volume of related papers by experts from four continents brings the reader the viewpoints of both clinician and laboratory research worker toward their subjects. Six poisons are treated: two snake venoms, two marine poisons and two bacterial toxins—never mind the volume title! (The sequel, not yet scheduled, will give an account of poisons of plant origin.) Each poison is discussed from the distinct viewpoints mentioned in a knowing review by a specialist.

The toxin of the spore-forming anaerobic bacillus *Clostridium botulinum* (*botulus* is the Latin word for sausage, and the name was given to the fatal syndrome seen in southern Germany a century ago among those who ate a "regionally popular blood sausage") causes botulism. It is a rare disease: twoscore deaths worldwide each year would overestimate the toll. But it is dramatic, it may follow feasts of joy or of mourning and it is always potentially epidemic should some error enter into the commercial processing of food. It kills hundreds of thousands of animals each year—cattle, sheep, mink, wild waterfowl.

A fatal dose in humans is only a few micrograms of the toxin taken by mouth. This protein is "the most deadly of all biological poisons." Death results from failure of respiration or heartbeat. It is not surprising that the isolation and crystallization of the toxin were first reported at the close of World War II from the Army's laboratory at Fort Detrick, the U.S. center of biological warfare. Botulism is obviously a dangerous subject for research, and the "keystone of protection" is immunization of all workers and then scrupulous care in the handling and decontamination of all materials.

The "red tide," a great bloom of the brownish marine dinoflagellate *Gonyaulax* seen this fall on the New England coast, is known from time to time on most temperate shores the world around. Mussels, clams and other filter-feeding animals ingest the microorganisms, often concentrating the toxin in special organs. About half a milligram of the poison, called saxitoxin, is deadly to man.

Saxitoxin is a close kin to a better-known poison, tetrodotoxin, which is secreted in the ovaries of the puffer fish and strangely enough also by newts. The puffer fish's liver and the newt's skin also contain a high concentration. This poison is not a protein but rather an unusual linkage of three organic rings, a total of only 39 atoms: $C_{11}H_{17}N_3O_8$. The shellfish poison takes few lives, because a close watch is kept on shellfish foods by all public-health authorities for signs of the "red tide." Some Alaskan peoples posted beach sentries by night to watch for the telltale luminescence.

The puffer poison is not a normal public-health hazard, although it is a public-health problem because in Japan the gourmet savors the delicate white flesh of the fish, called *fugu,* in restaurants specially licensed to serve the dangerous food. The chef must pass a stringent examination to test his knowledge of the fish species, the seasonal variations and the cunning skills of eviscerating the fish without contaminating the delicious flesh. Some 80 to 100 Japanese have been martyrs to their antique delicacy each year for the 80 years since statistics have been kept. Most of the victims, to be sure, were served by friends, ama-

teurs preparing the favorite fish without government license. Death is generally by respiratory paralysis. Antidotes? None. Eight or nine hours of artificial respiration, however, saves many of the patients.

The Elapidae are a large family of venomous snakes, including the American coral snakes, the Asian cobras and kraits, the African mambas, the Australian death adder and many others. Their venoms are all neurotoxins: a mixture of small protein chains, inducing death by flaccid paralysis that in the end reaches the respiratory musculature. A lethal dose in man is a matter of milligrams. The entire mechanics of the bite is complex; it seems fair to conclude that death by snakebite in the Tropics is largely an occupational disease of the poor farmer working barefoot in his fields. The only therapy with any success is again the passive immune system, injecting after the bite sizable doses of serum from horses induced to make antivenin by repeated small doses of the specific venom, which is first inactivated by formalin.

The venom of rattlers, the most common American venomous snakes, is not at all similar. Most species of these crotalid snakes inject a protein-rich and complex dose that is not primarily neurotoxic but acts both locally and within internal organs to cause tissue damage, circulatory failures of several kinds (particularly internal hemorrhage and lowered blood pressure) and an entire variable complex of distressing symptoms. The tropical rattler seems alone in secreting a highly lethal neurotoxin, more like that of the cobra family. Again timely use of antivenin (within a few hours) is the method of choice; indeed, there is really no choice.

Since Hippocrates physicians have known the worst of all these intoxications: tetanus, "a triad of wounding, lockjaw, and death." Striated muscle becomes stiff first around the wound, and the end after many weeks is "generalized spastic rigidity" ending in respiratory failure. The bacterium *Clostridium tetani* has been known since the 1880's, and the crystalline toxin was produced after World War II. It is a typical globular protein, with a rolled-up coil of fewer than 1,000 amino acid links, "without any striking clue to explain its remarkable biologic activity." It acts rather like strychnine, blocking the inhibiting muscular control mechanisms of the wiring of spinal cord and brain stem. The human lethal dose is less than 100 micrograms of the pure protein, so that it "must be handled with great care," again depending on prior immunization and available serum.

Tetanus tells no mere dramatic tale of a few sudden deaths; its victims are counted in the tens of thousands per year, particularly in the tropical developing countries. It is a danger everywhere, most frequently to the young and often to the newborn, cut from the maternal cord without precautions, soon then to cry, refuse to suck and die in spasm in a couple of weeks. Immunization and wound cleanliness are essential; our army in World War II saw almost no tetanus, since active immunization was routine. In World War I, however, the wounded died of it commonly; no preventive treatment was used.

The nature of the effects of these toxins on the subtle linkages between nerves and between nerve and muscle is much discussed in this volume at a highly technical level. Not many conclusions are reached. It is known of course that a specific transmitter (tiny spherules of acetylcholine about 500 angstroms in diameter) seems to induce the nerve signal in the muscle fibers. Botulinus toxin seems clearly to stop that signal, probably by blocking the release of the signal substance. The plant toxin curare apparently blocks the access of the signal to the muscle receptor. The whole topic is yet to be substantially understood; the toxins are themselves powerful tools for carving our understanding.

Nowhere is there any discussion of the question that any reflective reader will ask: What accident of evolution has produced so fine a fit between locks and keys that seem never to have been made for each other? Are these toxins adaptive to bacteria and newt, as they plainly are to the fierce little blue krait underfoot?

PEOPLE'S REPUBLIC OF CHINA ATLAS, by the Central Intelligence Agency. U.S. Government Printing Office ($5.25). All of China is mapped in excellent shaded relief at a scale of 1:4,000,000 in this large but thin paperback bargain. The place-names are not very dense and the altitudes are not shown by tint, so that the maps are larger but perhaps no more useful than those of the best world atlas. On the other hand, besides interesting geographic accounts of each section of the great country and many interesting photographs, the cartographers have freely included economic and similar special maps (on rainfall, railroads, population, crops), providing frequent direct comparisons with a similar adjoining map of the U.S. The original and pleasing pedagogy extends to a map bearing the rubric "Kansas" placed across the Peking plain to suggest the climate of each region by a crude comparison with the U.S. There are a good many such useful and instructive figures. The book ends in a double-page spread with a colorful painted bird's-eye view of Peking (borrowed from a 1957 local publication of the prominent sights of that splendid city) that the would-be tourist or the mere dreamer can study for a long time. Do not miss the Ancient Astronomical Instruments Exhibition Hall, near the old observing platform on the east wall of the city. In that still forbidden city on the Potomac where CIA dwells there are manifestly artisans other than those who organized the attacks at the Bay of Pigs and on the Plain of Jars. These craftsmen-mappers have done a fine job, cool and apparently trustworthy, if not affectionate. (The transliteration scheme they use is not the present official one, nor are all the numbers up to date.) This may be the best buy any taxpayer has had yet from the CIA. (Rand McNally has put a similar, derivative edition into the stores at $4.95.)

NATURAL HISTORY OF INFECTIOUS DISEASE, by Sir MacFarlane Burnet and David O. White. Fourth Edition. Cambridge University Press ($3.95). Every decade since the first publication has closed with a new edition of this admirable book, which belongs on any short list of the biological classics of our time. This edition found a coauthor, a Melbourne colleague of Sir MacFarlane's. The senior author has spent some 20 years becoming a Nobel Laureate for his work on the clonal nature of the immune mechanism, but this volume retains the interest he began with: the view of infectious disease as an evolutionary interaction between agent, host and mediating species. It is the fashion today to call this attitude ecology; here is the pure quill, aged since 1940. The book includes a clear account of foundations, as before, with a sketch of the nature of the agents and of immunity. It goes on to describe epidemiology through time and space, control and antibiotics, and there are seven chapters each surveying a single disease and the history of its control. The two wholly new chapters are those on iatrogenic diseases, such as "hospital staphylococcus," and the subscience to come, "slow viruses." The prototype of the slow virus is probably kuru, that remarkable affliction of the people of the South Fore River of New

Guinea, apparently spread by ritual cannibalism of the brain tissue.

The story of herpes simplex—cold sores—is still here, augmented by tragedy. "Monkey herpes is well known because at least a dozen people have died of it." That virus is almost as lethal as rabies to man. It is as harmless in the normal host, however, as the lifelong recurrence of cold sores. Long ago the rule was formulated by Theobald Smith: The older the relation with the parasite, the less virulent the effect. The two viruses, like the two hosts, had a common ancestor long, long ago.

ILLUSTRATED EXPERIMENTS IN FLUID MECHANICS: THE NCFMF BOOK OF FILM NOTES, by the National Committee for Fluid Mechanics Films. Preface by Ascher H. Shapiro. The MIT Press ($3). Nowadays film scripts sell in all the bookstores. Here is the scientific viewer's counterpart of the script of *8½*. (Call it 21 + 133.) This group of university professors in fluid mechanics planned and produced between 1960 and 1969 nearly two dozen 16-millimeter narrated films, each a pedagogical account of a coherent set of phenomena in fluid flow, and 133 brief silent eight-millimeter loops, each displaying a single topic and one particular experiment. Two of the finer longer films are, for instance, one on surface tension and one on low-Reynolds-number flow. The surface-tension color film celebrates—sometimes in ultraslow motion, often in high magnification—the visual beauty of drops and bubbles, jets and soap films so scrupulously that the images carry a high appeal even to viewers who cannot profit by the simple formulas and clear argument of the engineering-college-level narration. The film on low-Reynolds-number flow shows not only the most remarkable reversible flow and the curious viscous propulsion of model sperm cells with helical tail "propellers" but also something of the narrator himself, the friendly, famed and redoubtable Sir Geoffrey Taylor. "Film in print" is no substitute for the image flowing on the screen, but these notes with their many stills from the films serve the student well, either before or after viewing, and may help to show those who do not know the films what they are missing. Students of fluid flow in this generation own a treasure in these films; the National Science Foundation spent its $3 million well; the men and women of the Educational Development Center whose talent made the films have here a corpus of work to be proud of.

INDEX OF ADVERTISERS

JANUARY 1973

BIBLIOGRAPHY

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

THE HYDROGEN ECONOMY

HYDROGEN FROM OFF-PEAK POWER: A POSSIBLE COMMERCIAL FUEL. R. A. Erren and W. Hastings Campbell in *The Chemical Trade Journal and Chemical Engineer,* Vol. 92, No. 2392, pages 238–239; March 24, 1933.

HYDROGEN: KEY TO THE ENERGY MARKET. G. De Beni and C. Marchetti in *Euro Spectra,* Vol. 9, No. 2, pages 46–50; June, 1970.

HYDROGEN, MASTER-KEY TO THE ENERGY MARKET. Cesare Marchetti in *Euro Spectra,* Vol. 10, No. 4, pages 117–130; December, 1971.

THE HYDROGEN ECONOMY. D. P. Gregory, D. Y. C. Ng and G. M. Long in *The Electrochemistry of Cleaner Environments,* edited by J. O'M. Bockris. Plenum Press, 1972.

HOW THE IMMUNE RESPONSE TO A VIRUS CAN CAUSE DISEASE

DAS LEBEN UND WIRKEN DES WIENER KLINIKERS CLEMENS FREIHERRN V. PIRQUET. Dissertation. E. Hoff. Verlag G. H. Nolte, 1937.

EFFECT OF VIRUS INFECTIONS ON THE FUNCTION OF THE IMMUNE SYSTEM. Abner Louis Notkins, Stephan E. Mergenhagen and Richard J. Howard in *Annual Review of Microbiology,* Vol. 24, pages 525–538; 1970.

DESTRUCTION OF VIRUS-INFECTED CELLS BY IMMUNOLOGICAL MECHANISMS. David D. Porter in *Annual Review of Microbiology,* Vol. 25, pages 283–290; 1971.

IMMUNE COMPLEX DISEASE IN CHRONIC VIRAL INFECTIONS. Michael B. A. Oldstone and Frank J. Dixon in *The Journal of Experimental Medicine,* Vol. 134, No. 3, Part 2, pages 32S–40S; September, 1971.

INFECTIOUS VIRUS-ANTIBODY COMPLEXES: INTERACTION WITH ANTI-IMMUNOGLOBULINS, COMPLEMENT, AND RHEUMATOID FACTOR. Abner Louis Notkins in *The Journal of Experimental Medicine,* Vol. 134, No. 3, Part 2, pages 41S–51S; September, 1971.

IMMUNOPATHOGENESIS OF ACUTE CENTRAL NERVOUS SYSTEM DISEASE PRODUCED BY LYMPHOCYTIC CHORIOMENINGITIS VIRUS. Donald H. Gilden, Gerald A. Cole, Andrew A. Monjan and Neal Nathanson in *The Journal of Experimental Medicine,* Vol. 135, No. 4, pages 860–889; April, 1972.

THE OMNIVOROUS CHIMPANZEE

BABOON ECOLOGY AND HUMAN EVOLUTION. Irven DeVore and S. L. Washburn in *African Ecology and Human Evolution,* edited by F. Clark Howell and François Bourlière. Wenner-Gren Foundation for Anthropological Research, Inc., 1963.

ONE OBSERVED CASE OF HUNTING BEHAVIOR AMONG WILD CHIMPANZEES LIVING IN THE SAVANNA WOODLAND OF WESTERN TANZANIA. M. Kawabe in *Primates,* Vol. 7, pages 393–396; 1966.

THE BEHAVIOUR OF FREE-LIVING CHIMPANZEES IN THE GOMBE STREAM RESERVE. Jane van Lawick-Goodall in *Animal Behaviour Monographs,* Vol. 1, Part 3, pages 161–311; 1968.

THE PREDATORY BEHAVIOR OF WILD CHIMPANZEES. G. Teleki. Bucknell University Press, in press.

MARS FROM MARINER 9

SCIENCE, Vol. 175, No. 4019; January, 21, 1972.

MARS: THE VIEW FROM MARINER 9. Carl Sagan in *Astronautics & Aeronautics,* Vol. 10, No. 9, pages 26–41; September, 1972.

ICARUS, Vol. 17, No. 2; October, 1972.

THE CONTROL OF SENSITIVITY IN THE RETINA

ORGANIZATION OF RETINA OF THE MUDPUPPY, *NECTURUS MACULOSUS,* I: SYNAPTIC STRUCTURE. John E. Dowling and Frank S. Werblin in *Journal of Neurophysiology,* Vol. 32, No. 3, pages 315–338; May, 1969.

ORGANIZATION OF THE RETINA OF THE MUDPUPPY, *NECTURUS MACULOSUS,* II: INTRACELLULAR RECORDING. Frank S. Werblin and John E. Dowling in *Journal of Neurophysiology,* Vol. 32, No. 3, pages 339–355; May, 1969.

ADAPTATION IN A VERTEBRATE RETINA: INTRACELLULAR RECORDING IN *NECTURUS.* Frank S. Werblin in *Journal of Neurophysiology,* Vol. 34, No. 2, pages 228–241; March, 1971.

LATERAL INTERACTIONS AT INNER PLEXIFORM LAYER OF VERTEBRATE RETINA: ANTAGONISTIC RESPONSES TO CHANGE. Frank S. Werblin in *Science,* Vol. 175, No. 4025, pages 1008–1010; March 3, 1972.

FUNCTIONAL ORGANIZATION OF A VERTEBRATE RETINA: SHARPENING UP IN SPACE AND INTENSITY. Frank S. Werblin in *Annals of the New York Academy of Sciences: Patterns of Integration from Biochemical to Behavioral Processes,* edited by George Hadu, Vol. 193; August, 1972.

JUDAISM AT THE TIME OF CHRIST

THE ANCIENT LIBRARY OF QUMRAN AND MODERN BIBLICAL STUDIES. Frank Moore Cross, Jr. Anchor Books, 1961.

MAJOR TRENDS IN JEWISH MYSTICISM. Gershom G. Scholem. Schocken Books Inc., 1961.

THE APOCRYPHA AND PSEUDEPIGRAPHA OF THE OLD TESTAMENT. Edited by R. H. Charles. Oxford University Press, 1963.

CONDUCTION ELECTRONS IN METALS

ELECTRONS AND PHONONS. J. M. Ziman. Oxford University Press, 1960.

ELECTRICAL CONDUCTION IN SOLIDS. H. Inokuchi. Routledge & Kegan, 1965.

INTRODUCTION TO SOLID STATE PHYSICS. Charles Kittel. John Wiley & Sons, Inc., 1971.

SNOW CRYSTALS

SNOW CRYSTALS: NATURAL AND ARTIFICIAL. Ukichiro Nakaya. Harvard University Press, 1954.

SNOW CRYSTALS. W. A. Bentley and W. J. Humphreys. Dover Publications, Inc., 1962.

THE SIX-CORNERED SNOWFLAKE. Johannes Kepler. Oxford University Press, 1966.

MATHEMATICAL GAMES

ON SETS OF ACQUAINTANCES AND STRANGERS AT ANY PARTY. A. W. Goodman in *The American Mathematical Monthly,* Vol. 66, No. 9, pages 778–783; November, 1959.

SIM—A NEW GAME IN TOWN. Arch Napier in *Empire* (supplement to *The Sunday Denver Post*), pages 38–40; May 24, 1970.

THE AMATEUR SCIENTIST

ELECTROMAGNETICS. John D. Kraus. McGraw-Hill Book Company, Inc., 1953.

THE RADIO AMATEUR'S HANDBOOK. Byron Goodman. American Radio Relay League, 1972.

Premise:
Pollution control begins with effective sensing and measurement.

Honeywell offers a practical way to detect harmful chemical vapors.

The job of cleaning up air pollution involves far more than control of smokestacks and auto engines. Many industrial processes are the source of other potentially toxic gases and vapors.

For example, the use of paint products results in concentrations of vapors such as toluene or benzene. Degreasing operations produce vapors of other hydrocarbons. Dry cleaning and refrigeration plants must control perchloroethylene and ammonia.

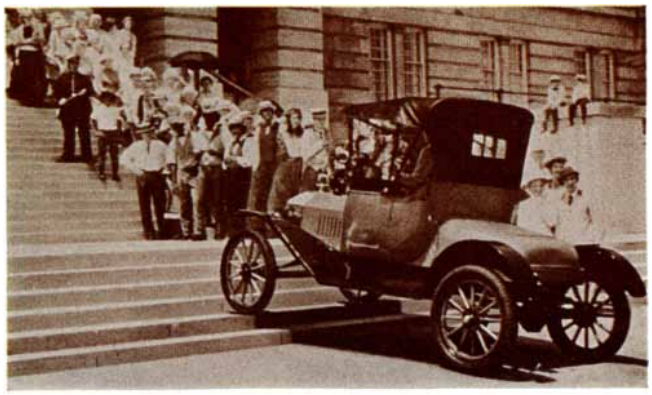
Charcoal filters, after-burners or scrubbers neutralize such pollutants. However, an effective control system requires dependable sensors to detect harmful vapor levels and provide for alarm control functions. Thus far, most common methods of sensing and measuring have been less than satisfactory or too costly.

Honeywell technology has developed simple, reliable detectors for many noxious gases and vapors. One such instrument, using the principle of ultra-violet absorption, can be calibrated to sense such pollutants as ammonia, benzene, hydrogen sulfide, perchloroethylene, trichloroethylene and sulfur dioxide. For a wide range of applications, Honeywell sensors can provide an economical, practical means of monitoring industrial atmosphere.

Honeywell fully recognizes that pollution control is a difficult and complex task. We'd like to help by sharing our specialized knowledge in the field of sensing and measurement. For details on the subject above, or other information, please write Mr. Richard Boyle, Director of Environmental Services, Honeywell, 2701 Fourth Ave. So., Minneapolis, Minn. 55408.

Honeywell

We built Ford Pinto to follow in the steps of the rugged old Model T.



Back in 1911, publicity stunts were the fashion in the car business.

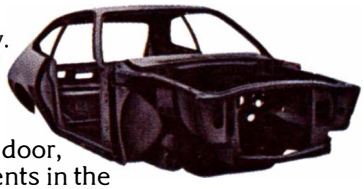
Like coaxing a Model T up the steps of the State Capitol in Nashville, Tenn., to impress people with its ruggedness.

But now it's sixty years later. And even though we want you to be just as impressed with Pinto's ruggedness—we figure we'd better give you some facts and figures. Not stunts.

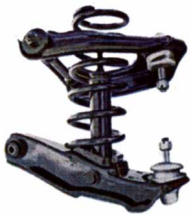
The Pinto engine. (Above.) Rugged and reliable. Improved and perfected in more than 10 years of driving in small Ford-built cars all over the world. Easy on gas, simple to maintain.

The Pinto transmission. A floor-mounted 4-speed fully synchronized transmission is standard on Pinto. It was designed to be "lubed for life". All it should need is inspection during routine dealer maintenance.

The Pinto body. Unitized into one piece of welded steel, with steel guard rails in the door, steel reinforcements in the roof. And a surprising amount of room inside.



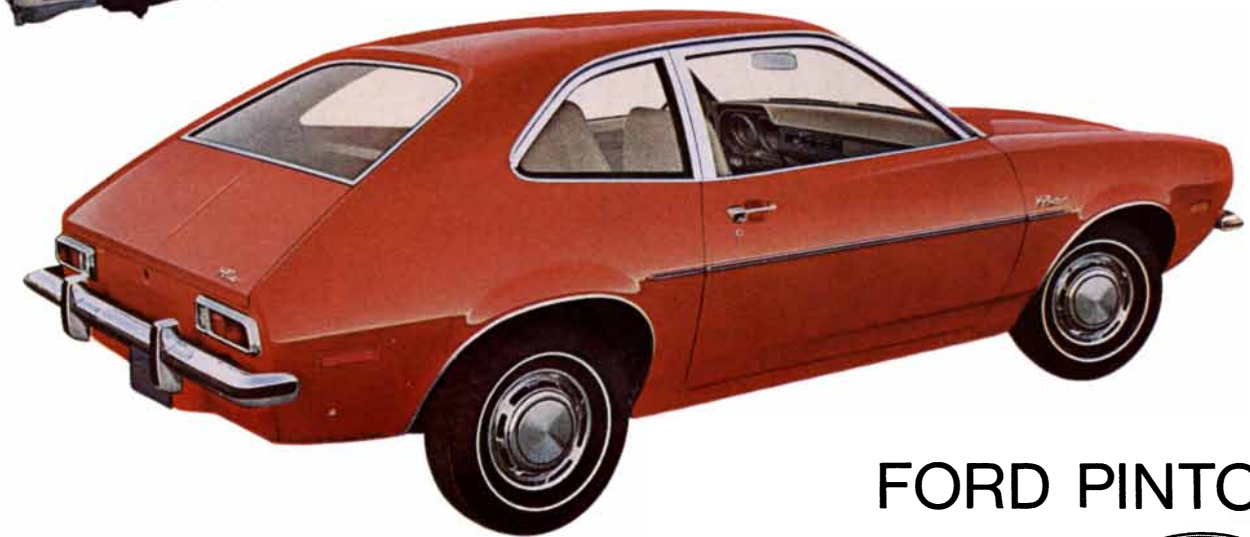
Extra-strength parts. Many Pinto components could be used in much bigger cars: the ball joints in the front suspension (below), the universal joint, starter motor, rear wheel bearings.



We built Ford Pinto to be a rugged, durable, basic car—a car that could follow in the footsteps of the legendary Model T.

See all the 1973 Pintos at your Ford Dealer's: 2-door sedan, 3-door Runabout, and the popular Pinto Wagon.

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



FORD PINTO

Shown here is a 1973 Pinto sedan with optional whitewall tires, accent and deluxe bumper groups.

FORD DIVISION Ford