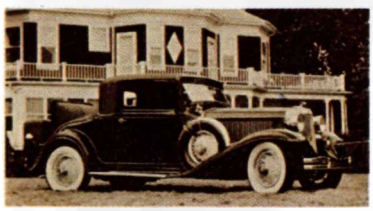
SCIENTIFIC AMERICAN

THE SOLAR CORONA

ONE DOLLAR

October 1973

Introducing the 1974 Chrysler. A totally new expression of an idea that has never changed.



From its beginning, the Chrysler has been designed with the idea that it should be an automobile of exceptional comfort and convenience.

An automobile of integrity.

And beauty.

And these are the very things we had in mind when we designed the new Chrysler New Yorker.

From its stunning grille, to its sleek side, to its broad, handsome rear deck and taillights, the car is truly new.

There's a quiet to the car that, quite naturally, you expect of such a fine car. And yet, the dimension of its quiet is something you really have to experience to believe.

Steel-belted radial tires, power disc brakes, power steering, TorqueFlite automatic transmission and an Electronic Ignition System are all standard. And all Chryslers run on regular gas.

The 1974 Chrysler New Yorker.

A totally new expression of an idea that has never changed.

Interior does not display standard safety belts.

Chrysler New Yorker

Extra care in engineering . . . it makes a difference.

CHRYSLER

Plymouth

Watch AFC Football, Championship Baseball Playoffs and the World Series on NBC-TV. Brought to you by Chrysler Corporation.

Conversation Pieces

Technically intriguing items from TRW, guaranteed to add luster to your conversation and amaze your friends.

Tornadoes, Rockets and Sonja Heine

So far, 1973 has been a banner year for tornadoes. By mid-year, more than 750 of these violent storms had swept down on the United States, killing 59 people and causing millions of dollars in property damage. Scientists expect the existing tornado record of 928 (set in 1967) to be easily shattered before the year is over.

Recently, tornado research has received help from an unexpected source — namely, studies made by TRW scientists of flow patterns in the propellant tanks in ICBM missiles. When you pump fuel out of a liquid rocket tank, much the same thing happens as when you pull the stopper out of your bathtub — a radial flow pattern develops (the particles move in spiral paths toward the center) and a vortex appears. To find out how swirling fluids behaved in propellant tanks, TRW scientists made some fundamental studies of the formation and behavior of vortexes. Further research has extended their analyses to the behavior of the large vortical patterns in the atmosphere we know as tornadoes, waterspouts, dust devils, and fire whirls.

A tornado begins with a thermal instability in the atmosphere, e.g., large mass of warm moist air under a layer of cold dry air. Under such conditions, violent updrafts may begin, around which the surrounding air begins to flow radially inward, in a swirling, spiral pattern. As particles get closer to the center of the flow pattern, their velocity increases. Some readers will recall the startling rotational speeds Sonja Heine achieved as she drew her extended arms closer to her body. Particles of air experience this same increase in rotational velocity as they get closer to the center of the system.

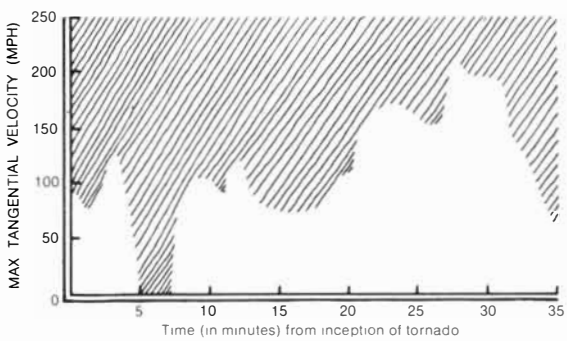
Ordinarily, turbulent diffusion opposes the swirling, and relaxes the disturbance — i.e., friction prevents Sonja from bringing her arms inward. However, in rare circumstances the radial inflow overwhelms turbulent diffusion, and a tornado develops. Actually, in a killer tornado much of the radial inflow is eventually confined to a layer near the ground, because at greater heights the increase of swirling ultimately creates a large centrifugal force that counteracts further radial inflow.

While dust and debris are being swept upward, the funnel of the tornado appears to descend. The latter occurs because the faster the air swirls, the more its temperature drops and the less moisture the air can contain. The resulting condensation of water vapor is seen as the funnel of the tornado, snaking down from the ominous cloud deck. Using these facts, TRW scientists have developed a formula which enables them to calculate the maximum velocity of winds in a tornado.*

TRW scientists have estimated the maximum wind speed in the funnel of a major tornado at around 225 m.p.h. Much of a tornado's destructiveness, however, stems not from the speed of the swirling wind, but from the radically low pressures inside the funnel. As a tornado engulfs a building, air trapped inside the building causes it to explode.

While much remains to be learned about large vortical storms, TRW's work with swirling liquid rocket propellants has lead to an important meteorological understanding of the behavior of destructive rotational storms.

**Maximum velocity, $V = (kgb)^{1/2}$, where b is the altitude of the cloud deck, k the fraction of the distance between cloud and ground the funnel cloud tip has descended, and g the acceleration of gravity.*



Using Weather Bureau data from the tornado of April 2, 1957, TRW scientists calculated the above time-history of estimated maximum wind speeds.

For further information, write on your company letterhead to:

TRW®
SYSTEMS GROUP

Attention: Marketing Communications, E2/9043
One Space Park Redondo Beach, California 90278

EVERYWHERE

you find people working with numbers, words, data and information, you'll find Olivetti. In offices, factories, banks, schools. In airports and hotels. In government offices. In every country of the world!

IN AMERICA

the Olivetti sales and service organization covers the country - with 80 district offices, a network of agents and dealers, and service centers coast to coast staffed with Olivetti-trained personnel. We also have teams of specialists whose sole job is to help you find solutions to specific problems and show you how to get the most out of our machines and systems.

FOR IN-HOUSE PROCESSING OF DATA

Olivetti leads the way. We created a world market for microcomputers by bringing programmed electronic calculation facilities right to the desk. Our famous Programma 101 was the world's first desk-top computer to use programs stored on magnetic cards. Since then we have exploited each new technological advance, developing new machines and new systems with ever increasing power and flexibility. And now we're proud to announce our latest achievement:

P652 SYSTEM FOR ADVANCED TECHNICAL AND SCIENTIFIC APPLICATIONS

allows you to process data for scientific research and technical projects on the spot, quickly and economically. The P652 has an exceptional storage and calculating capacity. Its keyboard is simple, functional; doesn't require any special operating skill. And it's an open-end system. Which means it can be extended at any time by adding peripherals using a variety of input/output media - from paper tape to magnetic cassettes.

olivetti

Established 1845 **SCIENTIFIC AMERICAN** October 1973 Volume 229 Number 4

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 MAILED C/O INTERNATIONAL DISTRIBUTORS, KOEPOORTBRUG 1, B 2000 ANTWERP, BELGIUM.

THE COVER

The colorful pattern on the cover is a spectroheliographic diagram of the sun in the extreme ultraviolet region of the spectrum. The diagram was constructed from measurements made by a spectroheliograph put aboard Orbiting Solar Observatory VI by the Harvard College Observatory Solar Satellite Project (see "The Solar Corona," page 68). The instrument isolates a single wavelength in the spectrum of the sun and the inner corona. Here the wavelength is 625 angstroms, the spectral line of magnesium stripped of nine electrons at 1.3 million degrees Kelvin. The deep red regions in the picture are the coolest and the white regions are the hottest. Spectroheliograms made in the spectral lines of elements of higher and higher ionization reveal the structure of hotter and hotter regions of the corona.

THE ILLUSTRATIONS

Cover courtesy of Harvard College Observatory Solar Satellite Project

Page	Source
17	Krauss-Maffei
18–24	Dan Todd
25	Milford Small
27	M. A. Epstein, University of Bristol Medical School (*top*); Guy de-Thé, International Agency for Research on Cancer (*bottom*)
28–29	Tom Prentiss
30	D. H. Wright, University of Birmingham
31–33	Tom Prentiss
34	W. Schiele, Deutsches Archäologisches Institut, Istanbul
36	Eric Mose
37	Manfred Korfmann
38–42	Eric Mose
53–64	Bunji Tagawa
68	Harvard College Observatory
70	Allen Beechel
71	High Altitude Observatory, Boulder, Colo. (*top*); Mees Solar Laboratory, University of Hawaii (*bottom*)
72	Allen Beechel
73	Harvard College Observatory, Imperial College of Science and Technology, York University and Science Research Council Astrophysics Research Unit at Culham Laboratory
74	Allen Beechel
75	American Science and Engineering
76	Allen Beechel
77	Naval Research Laboratory
78	Harvard College Observatory
80–93	Lorelle M. Raboni
95	Gabor Kiss
96–99	George Kelvin
100	Gerald Oster
101–102	George Kelvin
105	Stanford Linear Accelerator Center
106	Allen Beechel
107	Ilil Arbel
108–112	Allen Beechel
113–118	Ilil Arbel
118–119	Bill Lyon
120–125	Roger Hayward

SCIENCE/SCOPE

A "man-in-the-loop" RPV strike mission simulator has been built by Hughes in a company-funded research project to evaluate the problems of remotely-piloted vehicles. These include video bandwidth, jamming, and vehicle steering. The man-in-the-loop provides critical target recognition and sensor pointing for weapon delivery and also enables a narrower bandwidth to be used, thus minimizing jamming. The complete system makes use of a TV chain, flying-spot image scanner, digital scan conversion, digital computer, analog converters, hybrid interface equipment, and operator and experimenter control station consoles. It can simulate the critical portion of the RPV strike mission, from the time the vehicle, cruising at low altitude, "pops up" to high altitude to acquire the target, until it delivers the weapon.

A revolutionary reusable module to protect delicate electronic equipment in transit to and from maintenance or calibration facilities is one of a family of packaging, handling, transporting, and storage systems developed by Hughes as a systems approach called ISOPOD™. The module provides a dessicant basket, humidity indicator, relief valves, four-way forklift access, lift/tiedown rings, locking capability, and a 10-year design life. It requires no additional packaging, accommodates a wide variety of shapes, sizes, and weights, and isolates its contents from shock and vibration. Extensive tests and practical use show that the module virtually eliminates damage to equipment.

A high-resolution graphic display system incorporating the Conographic™ curve generator was introduced by Hughes at the recent WESCON show in San Francisco. The Conographic curve generator provides data compression of 10:1 or better as a result of its capability for generating conic curves, circles, and ellipses. Applications include seismic analysis, weather mapping, contour plotting, and computer-aided design and drafting. The new Conographic-12 system uses a Hughes 639 video memory, which provides a video display with 1029-line resolution. The video memory also provides selective erasure and makes it possible to view computer-generated graphics in high ambient light conditions.

Hughes Research Laboratories has two openings in R&D on advanced solid-state devices in the microwave and mm-wave region: 1) Extensive experience and demonstrated achievement in design and application, plus leadership ability. 2) Outstanding experience in fabrication and processing, including epitaxy, photolithography, etc.; this position requires use of advanced electron/ion beam techniques. U.S. citizenship required. Please send your resume to: Mr. A.J. Simone, Hughes Research Laboratories, 3011 S. Malibu Canyon Road, Malibu, CA 90265. An equal opportunity M/F employer.

New electronic products from Hughes include: a new type of environmental connectors which have a seal on each individual contact instead of on the connector itself, designed for airborne, deep space, shipboard, and undersea applications; they seal effectively at 250 psi (conventional connectors can fail at 5 psi)...an all solid-state millimeter-wave sweep generator which uses Hughes plug-in modules and external components to convert the Hewlett-Packard 8620B main frame into a compact, lightweight sweep test system in the 32-90 GHz frequency range.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY

Distributed Systems: maxi power at mini prices.

Minicomputers have gained great popularity in high-technology areas because they are often the most cost-effective way to do many dedicated tasks. But like everything else in this universe, they have their limitations. The user sometimes wishes he could add bulk storage and high-performance peripherals as cost-effectively as he acquired the original mini. For those who own or plan to acquire several minis, there is now a way to do this.

The concept, which we call distributed systems, consists of integrating all your HP 2100 dedicated minicomputer systems into a single network at whose center is another mini, a disc-based HP 2100S multi-programming, real-time executive.

In a distributed system, each user can gain the full power of the central computer as if he were the only user. Each can store large amounts of data on the central computer's disc, run programs at the central computer complex for his dedicated mini, and use its high-performance peripherals such as line printers and high-speed card readers. While each dedicated mini continues to acquire data, control experiments or monitor processes, the central computer can simultaneously analyze the data, generate reports for management, and be used for preparing programs.

A further level of computer power is also available with HP distributed systems, for the central computer can communicate directly with a host IBM 360/370, enabling HP 2100 networks to be an integral part of a hierarchal information distribution and processing system.

Key to it all are two Distributed Systems Packages: the HP 91701, about $6000 of hardware and software that establish and manage communications between the control 2100 and each satellite 2100 system; and the HP 91780 (about $5000) that performs a similar function for the central 2100 and host IBM computer.

If this idea interests you, ask for literature on Distributed Systems.

Increase surface chemistry data with an HP monochromatized ESCA Spectrometer.

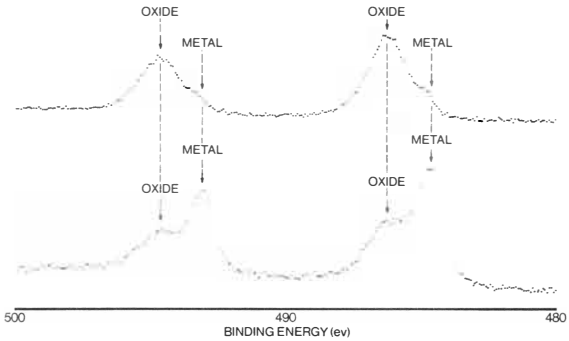
One of the most promising and powerful new analytical techniques, ESCA (Electron Spectroscopy for Chemical Analysis) now provides a chemical characterization of sample surfaces (10-20 angstrom deep). It has a broad range of applications: the effect of sunlight on polymers, the composition and uniformity of deposited thin films, the efficiency of washing processes, the molecular structure of complex organic molecules, and the effect of lubricants on metallic surfaces, as well as many other problems.

An especially important application for ESCA appears to be in characterizing catalysts. Catalysts are essential to the manufacture of materials, yet determining the properties of these compounds and finding methods to evaluate their efficiency are important problems that until recently have defied almost all analytical techniques.

Predicting the performance of fresh catalysts has been equally difficult. ESCA offers hope here; not only can it determine the oxidation states of species at the surface, but it can give information about surface conductivity, an important parameter for characterizing catalyst activity.

Although early investigators have already made important discoveries in surface chemistry and physics with ESCA, they have been unable to realize its full potential because of the design limitations of the first generation of instruments.

In all these spectrometers, the raw X-rays



An example of the kind of surface chemical information ESCA provides is shown in these spectra of tin. In the upper spectrum, tin (3d) ESCA peaks occur from the surface of a copper alloy sample. Mostly tin oxide with a small amount of metal is present. The lower spectrum shows tin (3d) peaks after the sample was surface "cleaned" with argon ion bombardment for five minutes. Note how the results indicate the partial removal of the tin oxide surface layer.

used to excite the sample produced satellite emission lines and a characteristic background that showed up in the ESCA spectrum and confused its interpretation. They were limited by an inability to use and control sample charging effects.

Our 5950A ESCA Spectrometer incorporates a unique spherically-bent quartz crystal monochromator that offers greatly improved performance above limitations of non-monochromatized X-ray photo-electron spectrometers. It allows chemists and physicists to take full advantage of the ESCA technique. System prices start at $125,000 (U.S. only). Ask for further information.

Complex statistics calculated at the touch of a few keys.

If you frequently figure means and standard deviations, correlations, histograms or curve fits, you can be freed from involved pencil and paper calculations. The HP 9805 gives printed or plotted answers to these stat problems at the touch of a single key.

You enter data once, then perform as many operations as you want. At a single keystroke, you get sample size, mean, standard deviation and correlation. Another keystroke and you get histogram cell statistics; curve fit coefficients, linear or parabolic; predict y from a known x; or compute t statistics including degrees of freedom. You never have to find intermediate results or re-enter data or program the machine.

The 9805 also performs the four basic arithmetic functions, %, $1/x$, $x/12$, $\log_e x$, $\log_{10} x$, e^x, exponents, and grand total accumulation. All calculations are performed to ten significant digits.

A built-in printer gives you a labelled record of all data entries and calculations. When you add the optional x-y plotter system, the plotting becomes literally automatic. You enter the data once, press the PLOT key and see your histograms or linear or parabolic curve fits take shape automatically. Change data and get a new plot, again by pressing a single key. (See one demonstration of the 9805, and we'll wager you'll never want to go back to the drudgery of hand-plotted data.)

You can also add a bright 10-digit LED display. For the big number-crunching jobs, you can add an internal Expanded Stat module for three additional curve fits (power, exponential, logarithmic), axis labeling, one-way analysis of variance, and normal curve overlays for histograms.

The basic stat version of the HP 9805 costs $1295. Send for details.

For more information on the products described in these pages, fill out the coupon or write to: Hewlett-Packard, 1509 Page Mill Road, Palo Alto, Calif. 94304.

Domestic USA prices only 00315

Please send me information on the following:
( ) HP Distributed Systems
( ) HP 5950A ESCA Spectrometer
( ) HP 9805 Calculator

Name_____
Title_____
Company_____
Address_____
City_____ State_____ Zip_____

HEWLETT hp PACKARD

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

LETTERS

Sirs:

In the article "An Advice-taking Chess Computer," by Albert L. Zobrist and Frederic R. Carlson, Jr. [SCIENTIFIC AMERICAN, June], mention is made of a £1,000 wager by international master David Levy that no computer-chess program will be able to beat him by 1978. Unless his chess-playing ability deteriorates radically, Levy seems to have a very safe bet.

Back in 1957 Professor H. A. Simon of Carnegie-Mellon University (one of the fathers of "artificial intelligence") made a number of famous predictions of what computers would be able to do within 10 years of that date. One of his predictions (none of the predictions has yet come true) was that a computer program would be chess champion of the world.

Professor Simon has since indicated that he thought then (in 1957) that far more effort would be devoted to problems of this type than has actually been expended on them. I do not know how much effort would be required to produce a chess-playing program by 1978 that can play as well as Levy can play now. There is serious doubt about whether it could be done in that time even if unlimited resources were available. Actually the resources are, and should be, quite limited. The few dozen people spending a few hundred thousand dollars a year that are working on this problem could not *produce* a program that plays at the 2,400 level (my guess as to Levy's chess-federation rating) by 1978 even if they now had some reasonable idea of how to go about it. There is no evidence that such ideas now exist. They are not going to get there by simply pushing the current technology in this area a bit further. It has taken 24 years to get from Claude Shannon's paper to a program that can establish a Class *C* rating. At the current rate of research, even with the much more powerful machines available, it will almost certainly take more than 24 years to get past Class *B*.

Note that best ratings now claimed for computer programs fall in the 1,400–1,600 Class *C* level. Class *B* is 1,600–1,800, Class *A* is 1,800–2,000 and expert is 2,000–2,200. The master level starts at 2,200. As a chess player and a computer scientist, I have been following the work in this area with great interest from the very beginning. Considering the methods now in use or seriously proposed, including those in the very interesting *Scientific American* article, I think that it is excessively optimistic to predict that a computer program will be able to play even at the Class *B* level by 1978.

SAUL ROSEN

Director
Computing Center
Purdue University
Lafayette, Ind.

Sirs:

As a preface, I should like to say that there are serious philosophical problems as to what constitutes "progress" in the field of artificial intelligence. One view, which can be traced back to an early paper by A. M. Turing (*Mind*, October, 1950), sets the goal of producing an intelligence in computers that is indistinguishable from human intelligence, and so it follows that progress toward that goal is measured by using human performance as a yardstick for computer performance. This view is ultimately correct, since the absence of progress would eventually indicate that Turing's goal cannot be achieved. As a working philosophy, however, there are several drawbacks to such an approach. First, no value is placed on understanding the nature of intelligence. Second, no allowance is made for the exploration of possible mechanisms of intelligence without immediate concern for their effectiveness. Our group at the University of Southern California has concentrated its effort on the formulation of mechanisms that would lead to intelligent chess play, and similar efforts are under way at Carnegie-Mellon University and elsewhere. Rosen claims that our ideas will not produce a Class *B* chess player. How can he be sure when we are unsure (but hopeful) and are investing several years of our time to test our ideas?

A previous critic of chess efforts, Hubert Dreyfus, was thrashed by a chess program at the Massachusetts Institute of Technology. Rosen may be setting himself up for the same fate. Although I have not indulged in any predictions or wagers, I feel that his predictions are excessively pessimistic.

ALBERT L. ZOBRIST

Information Sciences Institute
University of Southern California
Marina del Rey, Calif.

Sirs:

May I correct a statement in my article "Public Policy on Fertility Control" [SCIENTIFIC AMERICAN, July]?

On July 1 the Department of Health, Education, and Welfare reorganized its Health Services Administration and abolished the National Center for Family Planning Services. As a result administration of what remains of the Federal family-planning effort is now delegated to a new Bureau of Community Health Services that will have a cadre of full-time headquarters specialists working on family planning only one-tenth as large as formerly. The future of the Federal family-planning program is thus even more uncertain now than when the article was written this spring.

FREDERICK S. JAFFE

Director
Center for Family Planning
Development
Technical Assistance Division
of Planned Parenthood–
World Population
New York

Scientific American, October, 1973; Vol. 229, No. 4. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer; Arlene Wright, assistant treasurer.

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

Advertising correspondence should be addressed to Harry T. Morris, Advertising Director, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 25 cents.

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

☐ Please enter change of address.

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

NAME

NEW ADDRESS

OLD ADDRESS

Reclining seats. Steel-belted radials. Special suspension. All Grand Touring features. Without a Grand Touring price.

A Grand Touring car is made for long, hard driving without being hard on you. That's why Cutlass Salon offers the traditional GT things you see above.

Things like contoured lounge seats. Steel-belted radial tires. Special suspension and steering. Automatic transmission. Plus the traditional dependability of an Olds Rocket V8.

Cutlass Salon coupe or sedan. Driving one can be an authentic Grand Touring experience. Pricing one isn't.

from Oldsmobile

Cutlass Salon

Built in the Grand Touring tradition.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

OCTOBER, 1923: "The coast-to-coast flight is undoubtedly one of the most dramatic achievements of modern aeronautics. Lieutenants John A. Macready and Oakley G. Kelly of the Army Air Service, leaving Roosevelt Field on Long Island, reached Rockwell Field in San Diego after a nonstop flight of 26 hours 50 minutes, covering an airline distance of 2,600 miles. According to the pilots' own estimate the average speed maintained was 93½ miles per hour. In the initial stages of the flight the plane was greatly overloaded, and an altitude of only 1,500 feet could be maintained over Long Island, New York City and New Jersey. Reaching Indianapolis and Tucumcari in New Mexico, they flew by compass alone. They welcomed the hangars of Rockwell Field with relief, and came down in perfect condition, although neither flyer had slept during the entire trip. The lessons of the flight are most important. Apparently the airplane is now ready for commercial air transportation."

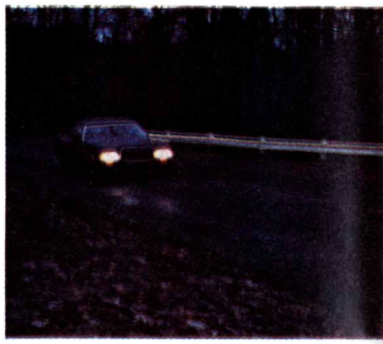
"Last year 14,000 lives were lost in this country in automobile accidents. The present registration shows that there are in the U.S. 12,000,000 vehicles, and the manufacturers estimate that the increase this year will be 3,000,000. The risk to safety lies in the fact that 90 per cent of these vehicles are congested on 10 per cent of our roads. Recommendations to reduce the ghastly slaughter come under four heads: first, to secure good designs for new roads; second, to promote adequate improvement of old roads; third, to insist on reconstruction of existing roads at places that have proved especially dangerous, and last, to improve the location of the center line on dangerous curves and elevations. Curves should be banked as a protection against skidding, and on all curves of more than four degrees the pavement should be widened. Another important recommendation is that a line about four inches in width should be painted in the center of the pavement on curves."

"About a year ago a group of scientists, in particular Drs. F. G. Banting and C. H. Best, working in the physiological laboratories of the University of Toronto, announced they had discovered a preparation that possessed the marvelous property of lowering the sugar content of the blood of dogs, when it was injected into them by means of a hypodermic syringe. This discovery was epoch-making in the history of medicine. The extract was made from the pancreas of animals, particularly the dog, the rabbit and the unborn calf. The process of making the extract has been applied on a commercial scale, and in this country there is at least one drug house that is ready to supply insulin in regular quantities. Insulin is not a permanent cure of diabetes; patients must keep on taking the drug constantly."

"The Bureau of Standards recently announced that the nearest approach to the absolute zero temperature that has yet been attained was recently achieved by Professor H. Kamerlingh Onnes of the University of Leyden. The record temperature of 272.18 below zero Centigrade, or, as the physicists express it, .82 degree absolute, was reached by the Dutch scientist in an unsuccessful attempt to solidify liquid helium. At that temperature the liquid helium showed absolutely no tendency to solidify. Every gas has been both liquefied and solidified except helium, which has never been reduced to the solid state."

SCIENTIFIC AMERICAN

OCTOBER, 1873: "The system of transporting oil by means of pipes laid over moderate distances has been in practice in the oil districts of Pennsylvania for several years. It remains yet to be determined whether the project can be carried out on a gigantic scale over more extended space. With the late discoveries in Butler County, Pa., the idea of transporting oil through iron pipes, from Titusville over the Alleghanies to Philadelphia on the sea board, a distance of 260 miles, is now exciting considerable attention. Mr. G. W. Platt, an engineer, gives detailed specifications for the construction of a huge conduit of this description. It is proposed to lay a cast-iron six-inch pipe, which at one locality of its route will be 3,000 feet above the sea level. Eight pumping stations will be established. Mr. Platt con-



The one-minute long distance rate. An inexpensive way to say you got home okay.

When you've something short to say, like "I'm home," we've got a way for you to say it for very little money.

It's our late-night and early-morning one-minute rate for long distance calls you dial direct without operator assistance.

You can use it any time after 11 at night all the way till 8 the next morning.

Any night of the week.

And it won't cost very much. Even coast-to-coast. New York to Los Angeles, for instance, is only 35¢, plus tax (additional minutes are 20¢ each).*

Obviously, most of your calls will take longer, so they'll cost more. But you'll still save money when you dial direct.

At AT&T and your local Bell Company, we want you to know all the ways to use your phone so you can save money. That includes using our late-night one-minute rate and dialing all long distance calls direct.

We hear you.

*Dial-it-yourself rates apply on all interstate calls within the continental U.S. (excluding Alaska) completed from a residence or business phone without operator assistance. They also apply on calls placed with an operator from a residence or business phone where direct dialing facilities are not available. For dial-it-yourself rates to Hawaii, check your operator. Dial-direct rates do not apply to person-to-person, coin, hotel-guest, credit-card, or collect calls, or to calls charged to another number, because an operator must assist on such calls.

These kids are playing in a box of tomorrow's glass bottles.

At one time or another, almost every child plays in a sand box.

What he plays in is not only the most common substance in the earth's crust, but the major substance that goes into the making of glass bottles and jars.

Glass is about 73% sand—so it makes an ideal environmental package. What's more, it's clear, pure, re-sealable and tasteless—so it also makes the kind of container that consumers prefer for foods and beverages.

Because glass is a natural product, it can be returned to the earth in virtually its original form. Because glass is a recyclable product, it can also go back to industry and be made into new containers. In addition, the glass industry today is continuing to develop new uses for waste glass such as road paving material, construction panels, bricks, insulation, terrazzo and reflective paints.

Nature invented glass. But man has perfected it. Given its natural ingredients, its packaging benefits and its environmental virtues, it's no wonder that consumer preference for glass containers continues to grow.

For a copy of "The glass container story," write: Glass Container Manufacturers Institute, Dept. H, 1800 K St., N.W. Washington, D.C. 20006.

Glass. It's a natural.

cludes that 23,000 barrels can be delivered every 24 hours, at 10 cents per barrel."

"The discovery of a human skeleton in one of the grottoes of Menton, a village on the south coast of France near Nice, has produced no small excitement in the scientific world. It appears from the recent investigations of M. Rivière that, at the upper portions of the caverns examined, remains of instruments and tools were found belonging to the prehistoric epoch that immediately preceded, in the west of Europe, the appearance of metals. While the fauna discovered in connection with the human relics indicate a very ancient paleontological epoch, the bone and stone instruments seem to point to a more recent period. The multiplicity of bone tools, needles, chisels and a *baton* of command, together with a peculiar necklace that closely resembles one found on the fossil man of Cro-Magnon, lead to the conclusion that the series of objects belongs to an age posterior to that of the bear, namely to that of the reindeer. It is believed, however, that the original owner of the skeleton existed during the latter age, and was buried in a cave formerly inhabited by men of the preceding epoch."

"The good news comes to us from Dundee, Scotland, of the safe arrival there of all the remaining survivors of the Hall arctic expedition, consisting of Captain Sidney O. Buddington and 12 others. The incidents and results of this latest polar exploration can be briefly summed up as follows: On the 29th of June, 1871, the steamer *Polaris,* Captain Charles F. Hall, sailed from New York on a voyage of arctic exploration. In August, 1871, she reached latitude 82°16′, the highest point ever attained by any vessel. Soon after the ship went into winter quarters Captain Hall was taken ill, and he died on November 8, 1871. Captain Buddington became master. In August, 1872, finding further progress northward impossible, Captain Buddington determined to return home, and the ship started for the south. She was now caught in the ice, and Captain Buddington caused a portion of the provisions and a part of the ship's company to be landed on the ice. On the night of October 15, the *Polaris* broke away from her icy moorings, leaving the hapless party of 19 persons on the ice. On the 30th of April, 1873, after 6½ months of dreary drifting, they were rescued and were safely landed at St. John's, Newfoundland."

THE AUTHORS

HENRY H. KOLM and RICHARD D. THORNTON ("Electromagnetic Flight") are respectively senior scientist at the Francis Bitter National Magnet Laboratory and professor of electrical engineering at the Massachusetts Institute of Technology. Kolm received his degrees from M.I.T. in physics: his bachelor's degree in 1950 and his Ph.D. in 1954. He is a founder of the National Magnet Laboratory and has a background in low-temperature and solid-state physics, cryogenic engineering, the generation of very intense magnetic fields (and their application to research and industry) and the development of superconducting-magnet systems. Thornton received his bachelor's degree in electrical engineering from Princeton University in 1951 and his Ph.D. in the same field from M.I.T. in 1957. Kolm and Thornton write that they have spent many hours cultivating their "feeling for vehicle dynamics," Kolm by flying and Thornton by sailing.

KEEN A. RAFFERTY, JR. ("Herpes Viruses and Cancer"), is professor of anatomy and head of the department of anatomy at the University of Illinois Medical Center in Chicago. Following his graduation from the University of New Mexico in 1950 he did graduate work in zoology, obtaining his Ph.D. from the University of Illinois in 1955. "My thesis topic," he writes, "involved the mechanism of formation of spontaneous tumors in frogs. In order to extend my experience to studies of tumors in mammals and birds I did three years of postdoctoral work at Yale University. From there I went to the Johns Hopkins University School of Medicine, where I taught anatomy for 12 years and did further research on tumors of frogs and mammals. Since I moved to Chicago in 1970 my interests have come to center almost entirely on human tumors."

MANFRED KORFMANN ("The Sling as a Weapon") is referent for prehistory and Near Eastern archaeology on the staff of the German Archaeological Institute in Istanbul. He received his Ph.D. in archaeology from the University of Frankfurt am Main in 1970. Before he began his university studies he spent a year in Jordan, teaching sports and German, and it was then that he became interested in Near Eastern studies. He notes that he has "some skill with both the sling and the bow."

D. E. KOSHLAND, JR. ("Protein Shape and Biological Control"), is professor of biochemistry and chairman of the department of biochemistry at the University of California at Berkeley. He received his bachelor's degree there in 1941 and his Ph.D. from the University of Chicago in 1949. He was associated with the Brookhaven National Laboratory and Rockefeller University until he returned to Berkeley in 1965. "As I grow older," he writes, "I note that my nonscience diversions—reading, book-collecting, tennis and politics—consume a smaller fraction of my time. After shedding a few crocodile tears, however, I realize this is the result of my own selection. Thus at the moment the puzzles in neurobiology intrigue me more than novels or tennis (possibly because my tennis gets poorer). My wife is professor of immunology and bacteriology at the University of California; we are close enough in fields to enjoy discussing science but far enough apart so that we are not tempted to boss each other's research."

JAY M. PASACHOFF ("The Solar Corona") is director of the Hopkins Observatory at Williams College and assistant professor of astronomy there. He was graduated from Harvard College in 1963, receiving his master's and doctor's degrees at Harvard also. He spent two years at the Hale Observatories and the California Institute of Technology before going to Williams last year. Pasachoff is interested in photography and also reports that he has found his return to New England good for his squash game. He is given to composing light verse, some in the form known as the double dactyl, which has achieved a certain notoriety. He gives the following example, titled "A Biographical Addendum":

Higgledy-piggledy,
Jay Myron Pasachoff,
Williams astronomer,
Dabbles in rhyme.

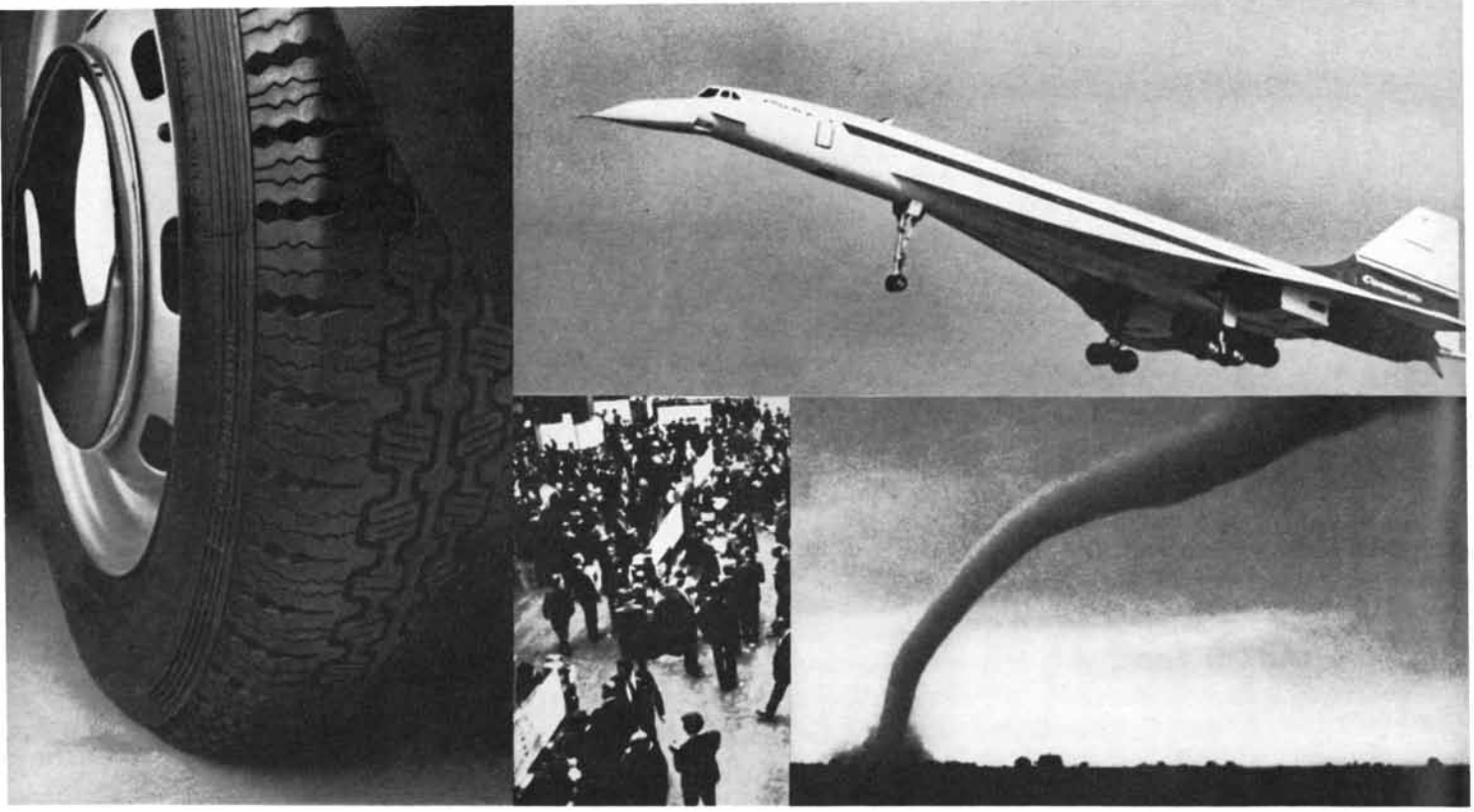
Solar eclipses and
Radiotelescopes
Keep him contented
The rest of the time.

OLLE BJÖRKMAN and JOSEPH BERRY ("High-Efficiency Photosynthesis") are with the department of plant biology of the Carnegie Institution of Washington in Stanford, Calif. Björk-

man, who was born in Sweden, obtained his Ph.D. (in genetics) from the University of Uppsala in 1960, having previously studied at the University of Stockholm. In addition to his work at the Carnegie Institution, which he joined in 1964, he is associate professor of biology at Stanford University. He describes his nonprofessional interests as hiking, boating and fishing. Berry received his bachelor's degree (in chemistry) and his master's degree (in soil science) at the University of California at Davis in 1963 and 1966 respectively and his Ph.D. (in botany) at the University of British Columbia in 1970.

GERALD OSTER ("Auditory Beats in the Brain") is professor of biophysics and research professor of obstetrics and gynecology at the Mount Sinai School of Medicine of the City University of New York. Oster's Ph.D. is from Cornell University; he has served on the staff of the Rockefeller Institute for Medical Research and was professor of polymer chemistry at the Polytechnic Institute of Brooklyn (now of New York) before going to Mount Sinai. He writes that his current interest is "mainly in the area of population control" and involves "trying to find simple tests (chemical or psychophysical) whereby a woman can by herself anticipate when she will ovulate."

ALAN M. LITKE and RICHARD WILSON ("Electron-Positron Collisions") are respectively physicist at the Lawrence Berkeley Laboratory and professor of physics at Harvard University. Litke was graduated from Johns Hopkins University in 1964 and obtained his master's degree and his Ph.D. at Harvard, where he remained as a research fellow in physics until going to Berkeley in May of this year. He writes: "I have often combined my scientific activities with my other interests. In 1965 I hitchhiked from Amsterdam to Athens on my way to work at the Weizmann Institute of Science in Israel. In 1968 I traveled around the world via Siberia and the colliding-beam research facilities of the Academic City outside Novosibirsk. This trip included diving for shellfish with the Ama divers of Japan and snorkeling off the Hawaiian island of Maui. The possibility of combining science and art has intrigued me for some time. I have collaborated with artists on sculptures that display scientific phenomena in an artistic setting." Wilson, who was born in England, obtained his degrees at the University of Oxford. He has been at Harvard since 1955.



The plot changes. The story stays the same.

On Monday morning, weather maps will be drawn on a CalComp system—those for the German weather service on a flatbed plotter—and those for the United Kingdom on a microfilm graphic system.

At the same time, a leading firm of London stockbrokers and many of their clients will begin their week with trend analysis graphs plotted on CalComp drum plotters.

During the week, Pirelli of Italy will use a CalComp plotter to design tire treads. The Concorde supersonic jet will fly, thanks in part to engineering drawings produced by a CalComp plotter. And map makers all over the world can draw maps with a speed and precision man cannot approach.

The plots change.

Wherever in the world a new need for computer graphics arises, CalComp will assuredly have the hardware and the software to meet the need.

We set the standards for our industry. We service what we sell around the world.

The plot changes. But our reputation for innovation and our concern for excellence—those stay the same.

For information on plotters, call your local CalComp office, or contact California Computer Products, Inc., SA-10-73, 2411 West La Palma Avenue, Anaheim, California 92801. (714) 821-2011.

CALCOMP

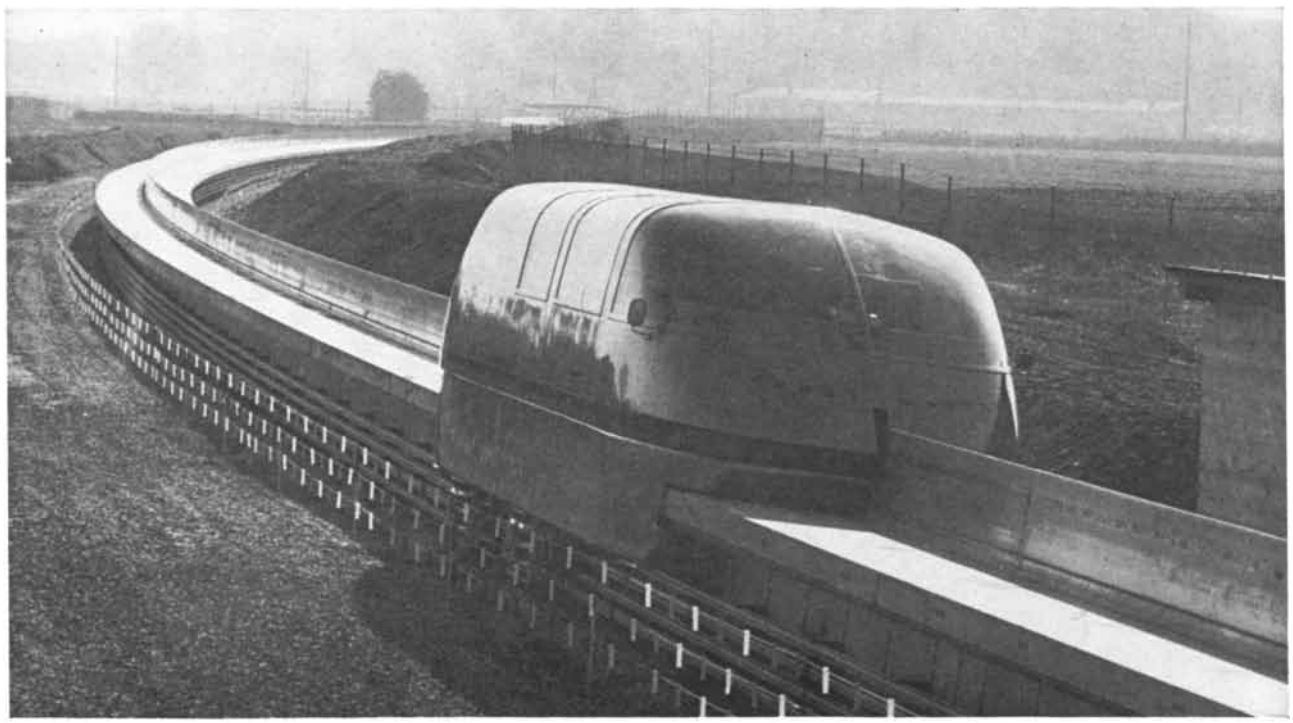
Established 1845 SCIENTIFIC AMERICAN October 1973 Volume 229 Number 4

Electromagnetic Flight

The future of high-speed ground transportation may well lie not with wheeled trains but with vehicles that "fly" a foot or so above a guideway, lifted and propelled by electromagnetic forces

by Henry H. Kolm and Richard D. Thornton

Public transportation employing vehicles with wheels appears to be incapable of speeds much above those attained now by technologically advanced railroads. Since a number of valuable purposes would be served by ground transportation at higher speeds, various alternatives to the wheel are being investigated in Germany, Japan, the United Kingdom, Canada and the U.S. Among the alternatives are air-cushioned transport, attractive magnetic levitation and repulsive magnetic levitation. In these studies it is becoming increasingly clear that the future in high-speed ground transportation belongs to vehicles that "fly" about a foot above metal guideways, being supported, guided and propelled by electromagnetic forces. With the technology available today, such systems could operate in the range



ELECTROMAGNETIC VEHICLE developed by a German consortium is operated on a test track. The vehicle runs about one centimeter above the guideway, employing the form of magnetic levitation termed attractive levitation. It is propelled by a linear induction motor that straddles the vertical aluminum reaction rail. Car shown is a test vehicle carrying passengers and instruments.

of speed from 200 to 300 miles per hour, serving certain corridors (Boston-Washington, San Diego–Los Angeles–San Francisco, Houston-Dallas and others) where the density of traffic is already sufficient to bear the cost. Eventually, as the technology improves and demand increases, such systems could span continents, operating in partly evacuated tunnels and moving at speeds limited in theory only by the inevitable curvature of their path. Electromagnetic flight holds the promise of being a juncture in transportation history comparable to the advent of the wheel, the automobile and the airplane.

Like other transportation systems, electromagnetic flight must evolve considerably before reaching a commercial scale. The course of its evolution depends as much on the socioeconomic climate and on national commitments as it does on scientific and technological progress. Public understanding of the issues and problems involved is vital. We therefore undertake here to review both the historical and the technical aspects of the subject.

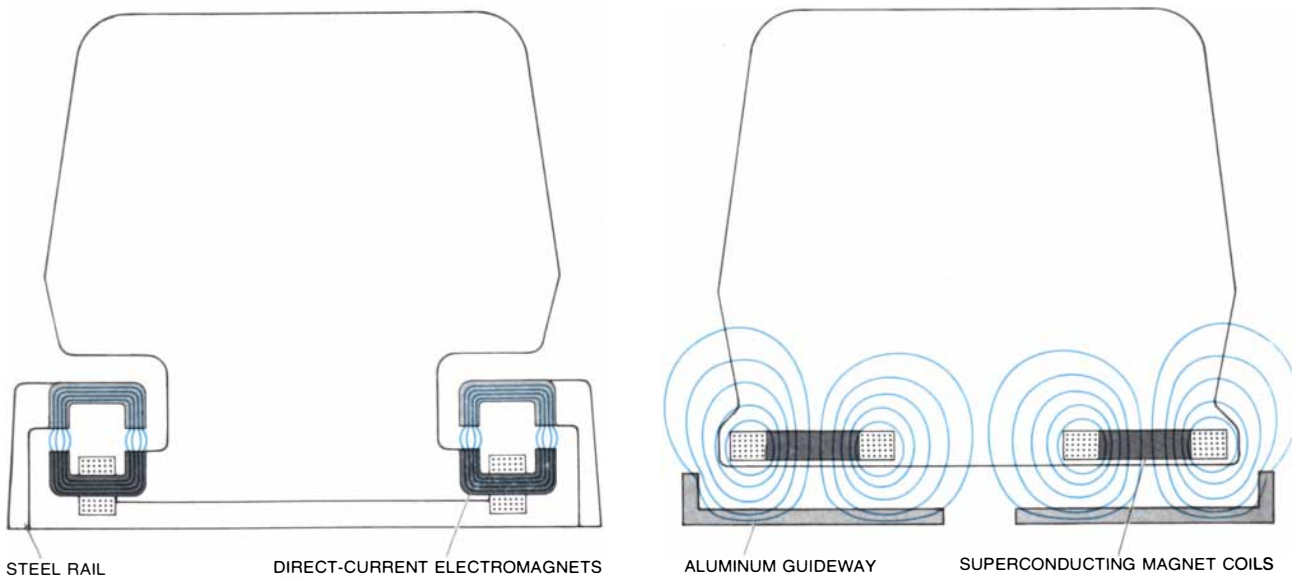
The first stage of the evolution began in the early 1960's with the realization by people concerned with ground transportation that fundamental innovations in technology were needed. Traffic on the highways and airways of several high-density corridors in the world was approaching saturation. The Japanese National Railways had just demonstrated the practical limits of conventional trains, finding that the traction and centrifugal strength of wheels and the frictional heating of pantograph current-collectors meant that such a train could not be operated regularly at a speed of more than about 200 miles per hour. In fact, the New Tokaida Line was unable to achieve day-to-day operation above 135 miles per hour, notwithstanding nightly realignment of the all-welded tracks and readjustment of the meticulously designed catenary cable. As a result of these findings many people started looking for alternatives to the wheel. From these efforts emerged what can be called the first generation of wheelless trains.

At the time the tracked air-cushion vehicle appeared to be the most logical alternative. Large untracked hovercraft had been built, and the technology of air cushions seemed to be well in hand. The British developed their Hovertrain, the French their Aerotrain and the U.S. Federal Railroad Administration its TACV. For various reasons, however, the development of air-cushion vehicles has not met with much success. The Hovertrain was officially terminated last February, construction of the Aerotrain has been postponed repeatedly and the TACV is now viewed primarily as a source of valuable information and by-products, notably the high-speed test facility built at Pueblo, Colo., as part of the program.

Magnetic levitation was also a well-known concept in 1965, the earliest proposal on record having been made in 1912 by the French engineer Émile Bachelet. Some of the early schemes were economically impractical, such as those that involved paving a guideway with permanent magnets to repel an oppositely magnetized vehicle. Others required supporting technology that existed but was unknown or at least untried in the context of railroad engineering.

Two entirely different approaches to magnetic levitation (which is often shortened to "maglev") are being pursued. One is attractive (or ferromagnetic) levitation, which involves vehicle-borne electromagnets that cling without contact to the underside of a suitably shaped steel rail or channel [*see illustration at left below*]. The system is inherently unstable because the attractive force increases as the magnet in the vehicle approaches the rail, requiring artificial stabilization by means of high-power, solid-state electronic feedback systems that regulate the current of the magnet so as to prevent contact.

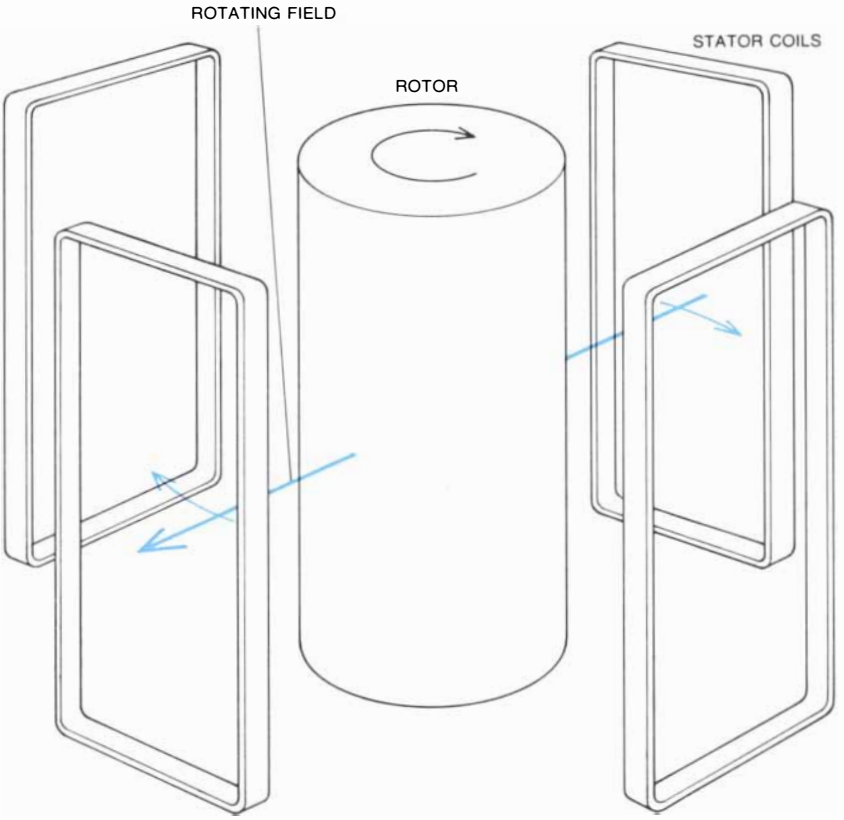
In such a system the amount of continuous power required to maintain lift remains quite reasonable (about two kilowatts per ton) up to clearance gaps of about two centimeters, although the peak power required to adjust the lift force increases rapidly as the amplitude and frequency of the required adjustments increase. A system employing attractive magnetic levitation therefore must have accurately aligned guide rails if high speeds are to be achieved. The exact trade-off between cost and speed



ATTRACTIVE AND REPULSIVE LEVITATION are portrayed. **Attractive levitation (*left*) is based on electromagnets that cling without contact to the underside of a steel rail. Repulsive levitation (*right*) is based on the repulsive force experienced by a magnet moving above a conducting, nonmagnetic surface. With superconducting magnets, which carry large persistent currents if kept at a low temperature, repulsive levitation can lift a vehicle about a foot above the guideway. Magnetic fields are depicted in color.**

is a controversial and largely undetermined issue. Although attractive magnetic levitation may not offer significantly higher speeds than wheeled systems, it does offer certain competitive advantages, including superior quality of ride, elimination of wear and greatly reduced noise.

The second approach to magnetic levitation is repulsive (or induced) levitation, which is based on the repulsion induced by a magnet traveling above a conducting surface. The scheme is appealing because it opens the way not only to wheelless trains but also to a fundamental departure from wheeled transportation. With repulsive levitation one can attain the mode of travel that has been called electromagnetic flight, employing vehicles called "magneplanes." Like aerodynamic flight, electromagnetic flight depends on a lift force derived from a vehicle's forward motion; the force is resilient enough to decouple the vehicle from variations in ground level, so that comfortable travel is possible without a secondary suspension system. Unlike the airplane, however, the magneplane can be flown in close proximity to a guideway because the lift force is strongest as the vehicle approaches the guideway. In this fashion the vehicle is compelled to follow the guideway.



GENERALIZED MOTOR for a vehicle in electromagnetic flight can be represented by a cylinder, which is the rotor, that is subjected to a rotating magnetic field perpendicular to its axis. The field is generated by a set of two or more coils, called the stator, that are supplied with two or more phases of alternating current. If the cylinder is a conductor, the motor is called an induction motor because it is driven by currents induced in the rotor. If the cylinder is itself a magnet, the motor is called a synchronous motor. Each of these types of motor can be "unwrapped" into a linear structure for electromagnetic flight.

Conventional electromagnets (and even permanent magnets) are capable of induced levitation at small clearances, but they are incapable of producing magnetic fields of the size and intensity required for flight. What has made electromagnetic flight possible is superconductivity, which is the virtual elimination of resistance in a conductor if it is sufficiently cooled. For the present purposes it will suffice to say that a 12-gauge copper wire, limited in normal use to a capacity of 20 amperes, will carry 10,000 amperes if it is made of currently available superconducting materials and kept near the temperature of liquid helium (4.2 degrees Kelvin, or 4.2 degrees Celsius above absolute zero). Such a temperature can be maintained by means of a refrigerator employing a closed cycle of helium as a kitchen refrigerator employs a Freon cycle.

The phenomenon of electromagnetic flight can be demonstrated by simple experiments. If one replaces the blade of a circular saw or the abrasive wheel of a bench grinder or a disk sander with an aluminum disk several millimeters thick, the spinning disk will be found to repel a suitable magnet held near its periphery. Electromagnetic flight can also be simulated without motion. Ordinary 60-hertz alternating current supplied to a wreath-shaped coil of copper wire will levitate the coil above an aluminum plate that is half an inch or more thick, although heating of the suspended coil will limit the duration of the flight.

The effect is best understood in terms of loops. Assume that a loop carrying an electric current is made to move past a somewhat larger conducting loop that is short-circuited [*see upper illustration on next page*]. As the magnetic flux from the traveling loop begins to link the stationary loop, it induces an electromotive force, resulting in an induced current that circulates in the stationary loop and tends to oppose the change in magnetic flux through that loop. The induced voltage is proportional to the rate at which the magnetic flux through the loop is changing, and the current is determined by the inductance and resistance of the loop.

When the traveling loop is directly above the stationary loop, the flux linkage is no longer changing, the induced electromotive force falls to zero and the induced current begins to decay owing to resistive loss in the stationary loop. As the traveling loop moves beyond the stationary loop, the flux linkage begins to decrease and thus induces a voltage pulse that is equal but opposite to the one induced during the approach. This pulse therefore suffices not only to cancel the induced current but also to induce a slight opposing current that is equal to the amount by which the originally induced current has decayed. This residual current persists for a fraction of a second after the vehicle passes and is called the eddy-current wake.

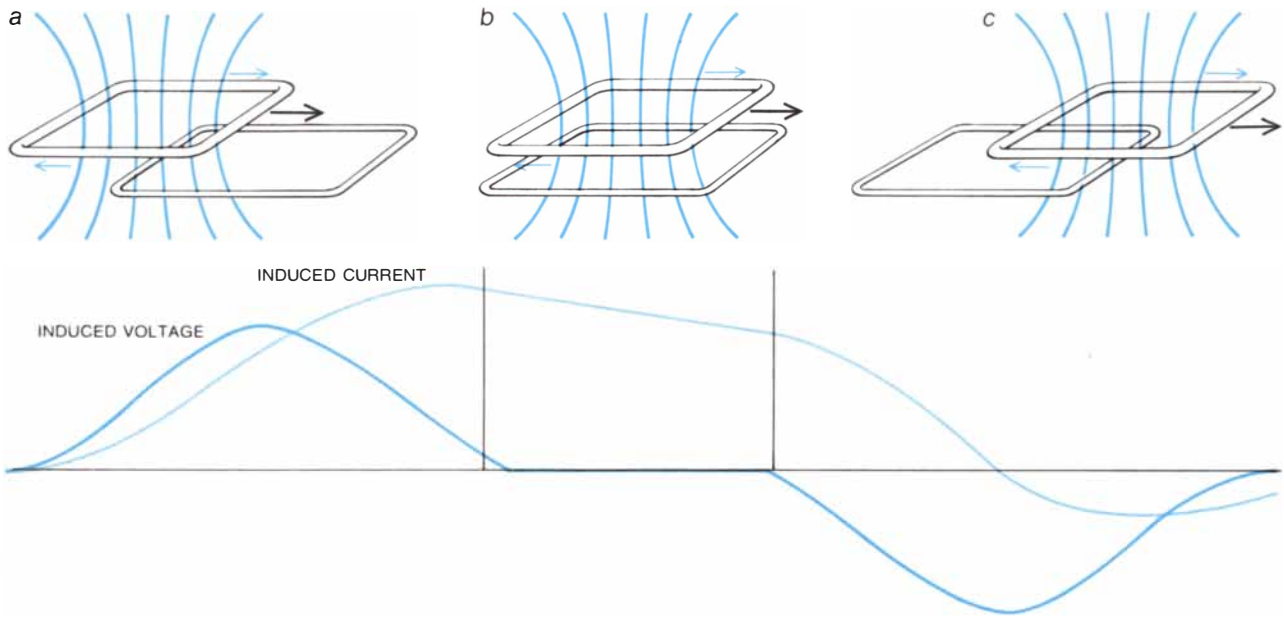
Both lift and drag are generated by the interaction of the two loops. Lift results from the fact that the induced current in the stationary loop flows in a direction opposite to the current in the approaching loop and therefore repels it. This repulsion alone would not produce drag, any more than drag is produced when a wheel is lifted by a cobblestone: all the energy lost in the approach is regained in the departure. Unfortunately, however, the electromagnetic cobblestone is not lossless. The cur-

rent in the stationary loop decays as the traveling loop passes, and therefore not all the approach energy is regained in the departure. The energy lost is just equal to the resistive dissipation in the stationary loop.

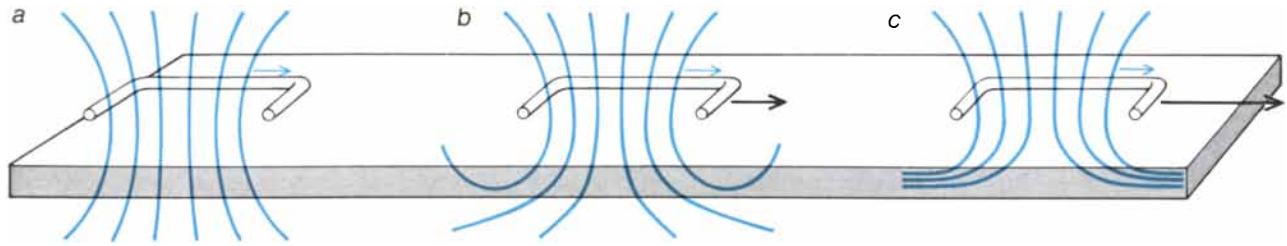
All the early investigators thought in terms of a guideway composed of discrete loops or coils, and there may in fact be certain advantages in such an arrangement. The guideway loops can be replaced, however, by a continuous conducting sheet without any fundamental change in the interaction. The conducting sheet (which we assume to be nonmagnetic) has no effect on the magnetic field of the current-carrying coil as long as the coil is stationary [*see lower illustration below*]. As the loop begins to move, currents are induced in the sheet; they oppose the local change in magnetic flux just as in the case of the discrete loop, except that these currents are distributed in the form of eddies and are called eddy currents. Their overall effect is to generate a magnetic field that opposes the field of the moving loop and keeps it from penetrating through the conducting sheet.

If the coil is moving rather slowly, some of the magnetic field still penetrates through the conducting sheet because the eddy currents have not become strong enough to keep it out entirely. As the traveling loop moves faster, however, it eventually reaches a speed at which none of the magnetic flux is allowed to penetrate. The sheet has become a magnetic shield. The moving loop now "sees" an image current of opposite direction, just as an electric charge facing a conducting sheet "sees" an image of opposite polarity. The image loop follows the real loop and repels it as if it were an equal distance behind the conducting sheet. The repulsive force becomes large as the moving loop approaches its image. This is the circumstance that makes electromagnetic flight susceptible to guidance.

With these principles in mind, one can understand the main characteristics of electromagnetic flight. If a current-carrying coil accelerates at a constant height above a conducting guideway, it experiences a lift force that is at



REPULSIVE LEVITATION is depicted schematically. As a current-carrying loop aboard the vehicle approaches a passive guideway loop (*a*), the increasing magnetic-flux linkage induces an electromotive force that drives a current around the guideway loop. The current tends to oppose the increasing flux linkage and thus repels the vehicle loop. When the vehicle loop is directly above the guideway loop (*b*), the magnetic flux is no longer increasing. The induced force falls to zero and the induced current begins to decay owing to resistance in the guideway loop. As the vehicle departs (*c*), the decreasing flux linkage induces an opposite electromotive force, which cancels the remaining current and induces a slight opposite and attractive current equal to the decay. Drag is caused by the decay and is therefore less the faster the vehicle travels. Decrease of drag with increasing velocity is rare in transportation.

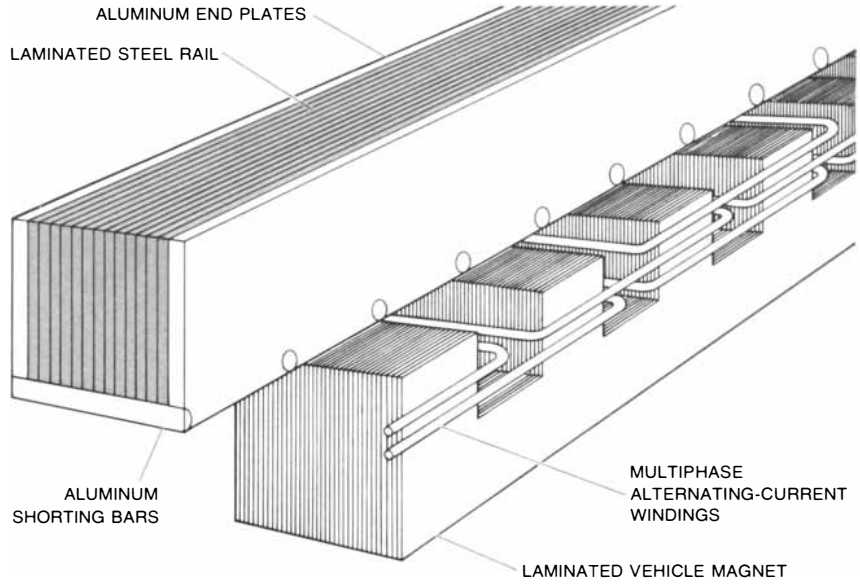


CONTINUOUS GUIDEWAY can also be employed in repulsive levitation. When the vehicle is stationary (*a*), the magnetic flux of the current-carrying vehicle loop is unaffected by the conducting, nonmagnetic guideway sheet. As the loop begins to move (*b*), eddy currents are induced in the sheet. They expel some of the magnetic flux and thus repel the moving loop. If the loop moves rapidly, most of the magnetic flux is expelled from the sheet, and the vehicle rides on a cushion of magnetic flux generated by an imaginary "image loop" of oppositely directed eddy currents that follows the real loop. The force increases sharply as the loop aboard the vehicle approaches the guideway sheet. It is this phenomenon that makes a vehicle in electromagnetic flight responsive to guidance.

first proportional to the square of velocity but eventually reaches a limiting value. The drag force is at first proportional to velocity but then passes through a peak value and decreases as the inverse of velocity. At high speeds the drag decreases more slowly because of the "skin effect," meaning that the induced eddy currents are prevented from penetrating into the metal and flow only in a thin layer called the "skin depth," so that the guideway appears to have higher resistivity. The fact that electromagnetic drag, unlike aerodynamic drag, decreases with increasing speed is a rather remarkable characteristic of electromagnetic flight.

For a practical guideway, takeoff becomes possible at 20 miles per hour, and the limit of lift is essentially reached at about 60 miles per hour. The lift is proportional to the product of the components of the magnetic field that are perpendicular to and parallel to the surface of the guideway, and the drag is proportional to the square of the component of the field that is perpendicular to the surface of the guideway. The lift-to-drag ratio is therefore proportional to the ratio of parallel components to perpendicular components of the magnetic field—a relation that gives some insight into the way the configuration of the magnets on a vehicle can be optimized.

If the magnetic field at the surface of a continuous-sheet guideway is 20 kilogauss (about the field strength at the pole face of a good magnet), the lift force is 60 pounds per square inch (about the pressure in a bus tire). This relation suggests that a bus could be levitated by magnet pads no larger than its footprint area—a circumstance that inspired the wheelless-train approach of the early investigators. Superconductivity was seen at first as simply a means for making such wheelless trains economically feasible. Its revolutionary possibilities went unrecognized for a surprisingly long time.



ATTRACTIVE-LEVITATION SCHEME developed by Rohr Industries, Inc., produces both levitation and propulsion. System employs multiphase alternating-current excitation and a laminated steel rail with aluminum shorting bars. Propulsion is by linear induction motor.

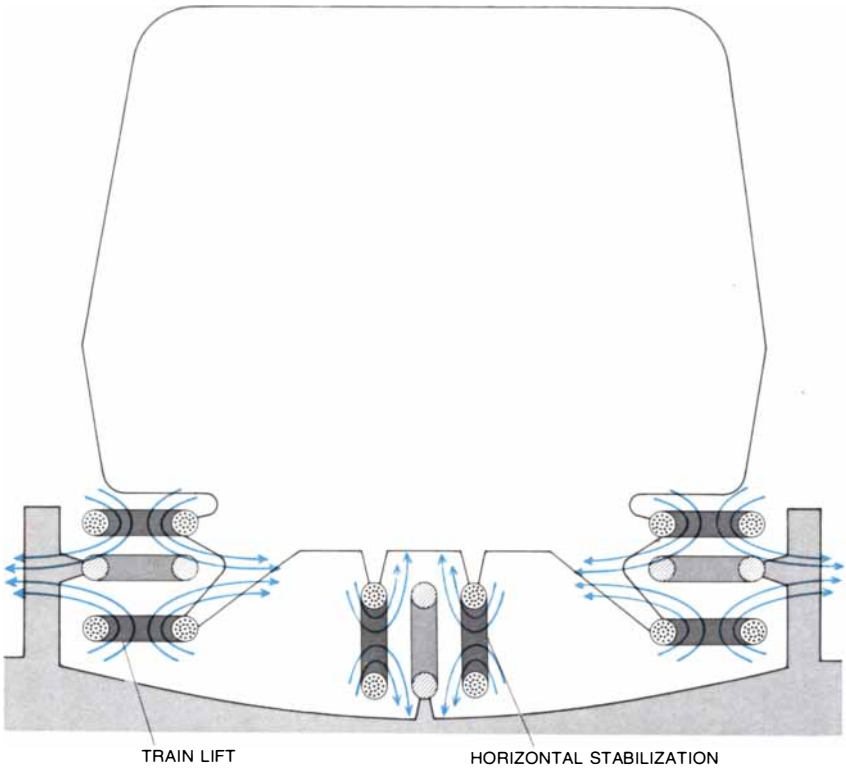
The evolution of electromagnetic flight is inextricably linked to the problem of propulsion. Two types of "linear motor" are being studied for this application. One is called the linear induction motor, the other the linear synchronous motor.

Many rotary electric motors can be represented by a cylinder subjected to a rotating magnetic field perpendicular to its axis [*see illustration on page 19*]. The rotating field is generated by a set of two or more coils supplied with two or more phases of alternating current; they are called the stator coils. If the cylinder is a conductor, such as an aluminum tube, the motor is called an induction motor because it is driven by eddy currents induced in the rotor by the rotating field of the stator. If the cylinder is itself a magnet (either a permanent magnet, an electromagnet or even a ferromagnetic structure), the motor is called a synchronous motor. An induction motor will supply torque even if its rotor is not in synchronism with the rotating field of the stator, although good efficiency is achieved only near synchronism. A synchronous motor, on the other hand, will supply no torque whatever once its rotor is pulled out of synchronism with the stator field. For this reason induction motors are used to start large synchronous motors and for applications involving loads and speeds that vary.

Both types of motor can be "unwrapped" into linear structures. The induction motor appeared to be the more obvious candidate for linearization because of its tolerance for variations of speed and because one of its two members can be simply a continuous strip of conducting material. Linear induction motors have received serious attention for many years, quite independently of magnetic levitation. Among other things, they serve to transport liquid metal and metal objects in conveyor lines consisting of a fixed array of magnets, representing linear stators, energized by alternating current. In transportation they were viewed as the only alternative to jet propulsion for tracked air-cushion vehicles; even earlier they were seen as a means of propelling wheeled vehicles above the speed at which wheel traction fails.

Propulsion by linear induction motors involves three serious problems. First, it requires that several megawatts of power be picked up from a live rail—a very costly operation that has yet to be solved for vehicles moving at high speeds. The alternative of generating power on board the vehicle is ecologically unacceptable, since it implies a 20,000-horsepower installation equivalent to a jet engine as a source of air and noise pollution, and the alternative of a fixed stator winding along the track is economically unthinkable. Second, to be reasonably efficient a linear induction motor must operate near synchronous speed, a performance that can be achieved only by carrying heavy frequency-conversion equipment on the vehicle in addition to the massive stator magnets. Third, the linear induction motor requires a clearance gap of about three-quarters of an inch between the stator and the reaction rail, which might be compatible with air-cushion transport or attractive levitation but would defeat the advantage of the much larger clearance made possible by repulsive levitation.

The linear synchronous motor was at first ignored because of its intolerance to variations of speed and its need for an expensive structure (either stator or rotor) along the entire track. The advent of superconductivity has changed the situation completely, so that the linear synchronous motor now appears to be the key to propulsion at high



NULL-FLUX SYSTEM was proposed by James Powell and Robert Danby of the Brookhaven National Laboratory. Two opposing coils on the vehicle carry a persistent superconducting current and straddle a single track loop. The system can also work with a single vehicle coil flanked by two track loops connected in opposition. This was the first scheme for repulsive levitation of high-speed transportation vehicles by superconducting magnets.

speeds. The superconducting levitation magnets provide magnetic fields that are not only very intense but also of large volume and can be efficiently coupled to fixed stator windings in the guideway, even at clearance gaps of about a foot. The increase in rotor current made possible by superconductivity opens the way to a reduction of the stator current and so to much simpler stator structures than were considered previously. As for the problem of synchronization, we have demonstrated at the Francis Bitter National Magnet Laboratory at the Massachusetts Institute of Technology that it can be solved easily by transmitting position information from the vehicle to wayside power-control units, which is much simpler than transmitting power from wayside units to a moving vehicle.

The development of magnetic-levitation systems has progressed quite far in Germany and Japan. In Germany two industrial teams (Krauss-Maffei and Messerschmitt-Bölkow-Blohm) have independently developed ferromagnetically levitated vehicles propelled by linear induction motors. By 1971 each consortium had completed a test track and an instrumented, passenger-carrying test vehicle. Both systems operate at a clearance of about one centimeter, with levitation magnets riding below an overhanging steel rail. A separate linear induction motor straddles a vertical aluminum reaction rail. Although the alignment requirements make for expensive construction, both systems represent serious competition for wheeled trains.

A third German consortium (Siemens, AEG-Telefunken and Brown, Boveri & Cie.) is pursuing a longer-range project involving repulsive superconducting levitation. The team has almost completed a circular test track 280 meters in diameter and is planning a 20-kilometer straight test section. All the German projects have had assistance from government ministries.

In Japan the Japanese National Railways (in collaboration with all the major electrical industries) decided after exhaustive study to adopt repulsive superconducting levitation with a small gap and with propulsion by linear induction motors. The plan was later modified by increasing the clearance gap to several inches and adopting linear-synchronous-motor propulsion. According to the schedule, a half-scale test vehicle will operate on a seven-kilometer test track by next year.

To complete our account of the first era of magnetic levitation we should mention a ferromagnetic attractive system being developed privately by Rohr Industries, Inc., of California. It differs radically from the two German systems in that the lift-magnet structure works on alternating current and is employed not only to lift but also to produce linear-induction-motor propulsion. The system has been demonstrated successfully at moderate speeds and may eventually provide the winning technology for a narrow-gap system.

The first proposal for the repulsive levitation of vehicles by superconducting magnets was published in 1966 by James Powell and Robert Danby of the Brookhaven National Laboratory. As physicists they were concerned with the design of magnets for confining, focusing and bending beams of high-energy charged particles. Although superconducting magnets were then being used by physicists, the broader scientific community had neither an interest in nor the money for applications of superconductors to transportation, and the transportation community viewed the proposal with more amusement than interest. Powell and Danby developed the concept in their spare time, with support (as they put it) from their wives.

Their first scheme involved a superconducting guideway, but they later refined what has come to be known as the "null flux" scheme, which is the basis of the system being implemented at full scale in Germany. Powell and Danby found that a significant reduction in drag can be achieved by arranging the guideway loops so as to achieve an extremely high field gradient, that is, a field intensity that changes drastically with changes in position. The feat is accomplished by having two opposing guideway loops flanking a single coil on the vehicle or, conversely, two opposing vehicle coils flanking a single guideway loop [*see illustration on this page*]. A null-flux suspension therefore provides not only a high lift-to-drag ratio but also strong restoring forces, or great stiffness. Unfortunately the system is so stiff that alignment of the guideway again becomes critical and a secondary suspension is essential. The secondary suspension creates added weight, aerodynamic drag and loss of power, which cancels much of the advantage in efficiency.

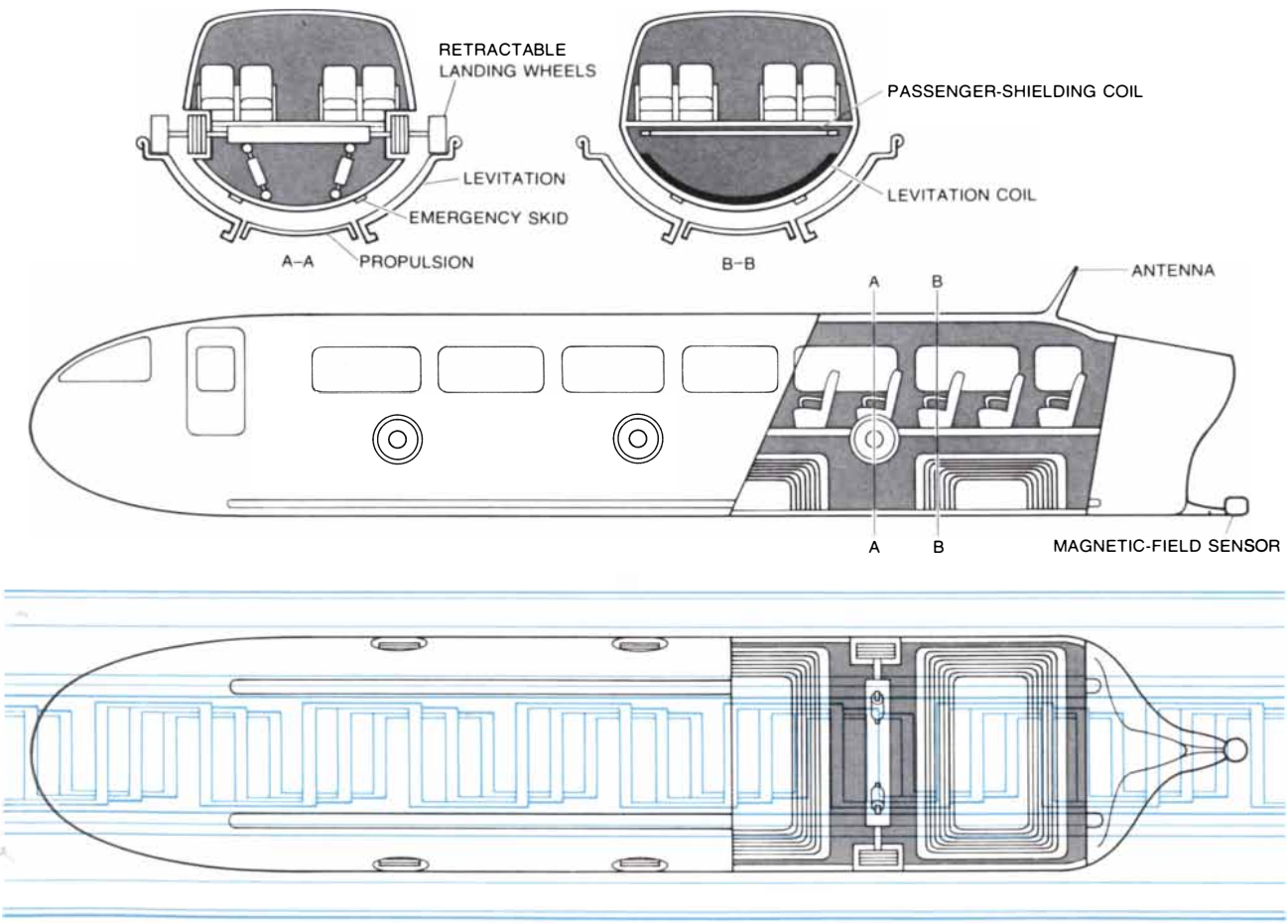
By 1969 several theoretical studies of inductive levitation had been published. Two of them are particularly noteworthy in the context of this article. C. A. Guderjahn, S. L. Wipf and several co-work-

ers, then at the Sandia Corporation, designed a guidance system for high-speed rocket sleds carried by two superconducting coils confined in *C*-shaped metal channels. This was the first proposal for a continuous-sheet guideway. The system provides very stiff suspension and high drag, owing to the fact that the magnetic field is predominantly perpendicular to the guideway sheets. Guderjahn and Wipf improved the lift-to-drag ratio, however, by adding a layer of steel to the upper leg of the channel so as to derive part of the lift by ferromagnetic attraction. A hybrid scheme of this kind has been adopted in the most recent variation developed by Powell and Danby and in one of the German projects. The second study was a detailed analysis of electromagnetic levitation and propulsion by Michael Tinkham of Harvard University and Paul L. Richards of the University of California at Berkeley.

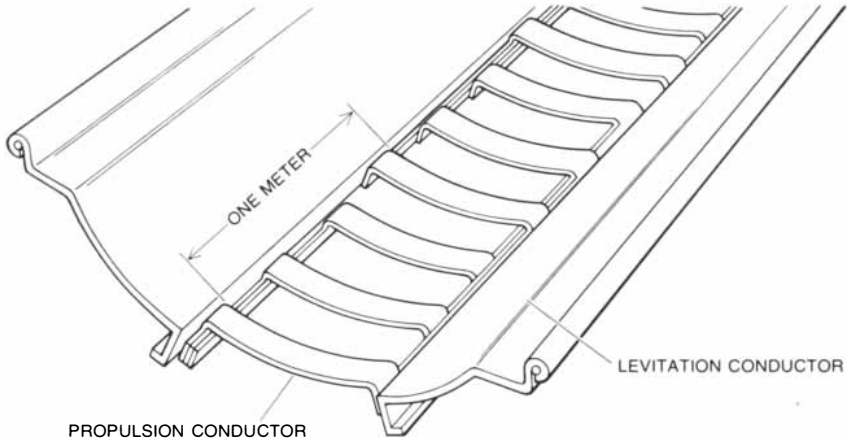
It was in 1969 that we initiated our work at M.I.T. on the magneplane system. We recognized that superconducting magnets offer the possibility of guided flight by lightweight vehicles, supported at an elevation of approximately one foot by resilient forces distributed evenly over a substantial area of the vehicle. This arrangement eliminates the most glaring absurdity of rail systems, in which massive girders in the vehicle concentrate the load onto two axle pads so that it can be carried at enormous pressure by two rails, whose concentrated load then has to be distributed over the ground by ties and ballast.

The magneplane is supported by saddle-shaped superconducting coils distributed essentially over its entire underside. It flies one foot above a semicircular aluminum trough, which allows the vehicle to assume a coordinated bank angle on curves in the manner of an airplane or a bicycle. (Self-banking is essential because even at a 45-degree bank, the steepest performed by an airplane, the tightest radius that can be negotiated at 300 miles per hour is one mile. If the bank angle were fixed by the guideway, even slight variations in speed would lead to unacceptable side forces on the passengers.) The weight of the vehicle is distributed over the guideway at about two pounds per square inch. The alignment tolerance is such that the guideway does not require a massive support bed and can be carried by light and flexible elevated structures resembling more the fuselage of an aircraft than a railroad bridge.

We also recognized that the propulsion system most compatible with electromagnetic flight is the linear synchronous motor, with stator coils along the guideway rather than aboard the vehicle. Other systems had eliminated active guideways categorically for economic reasons. In our system it turned out that the large, superconducting dipole magnets aboard the vehicle make it possible to have guideway coils that are considerably simpler than had been anticipated. The propulsion system consists of meandering conductors rather than coils, and



"MAGNEPLANE" VEHICLE developed by the authors at the Francis Bitter National Magnet Laboratory at the Massachusetts Institute of Technology is portrayed in cross sections (*top*), a side view (*middle*) and a bottom view, which also shows the guideway. The areas of the cross sections are shown by the lines *A—A* and *B—B*. The vehicle is designed to travel at speeds of from 200 to 300 miles per hour with levitation about a foot above the guideway by repulsive superconducting levitation and propulsion by a linear synchronous motor. At full scale the vehicle would be proportionately a considerable amount longer than the representation here.



MAGNEPLANE GUIDEWAY consists of meandering aluminum coils that serve as linear stators and are flanked by passive aluminum troughs that allow vehicle to roll into a coordinated bank angle in curves. Strong superconducting levitation coils on vehicle serve as rotor of a synchronous motor. Most of aluminum has both electrical and structural purposes.

we have managed to develop a guideway design in which most of the aluminum serves both structural and electrical purposes [*see illustration above*].

Detailed studies indicate that the cost of the guideway is comparable to that of any other projected high-speed system and that it is not in any way dominated by the fact that the propulsion system forms an integral part of the guideway structure. The linear synchronous motor thus offers a formidable combination of advantages: it eliminates the massive onboard equipment required for propulsion by the linear induction motor, along with the separate reaction rail, the narrow clearance gap and the need for a sliding power-pickup system. The reduced weight of the vehicle leads in turn to reduced costs for the guideway (a particularly significant factor in elevated construction) and reduced power requirements.

Synchronization of the linear synchronous motor is accomplished quite easily by transmitting position information from the vehicle to the wayside power-control units, which are positioned at intervals of about 10 miles. These units contain transformers and computers that actuate transformers and solid-state switches positioned at one-mile intervals along the guideway. The phase, frequency and amplitude of the multiphase propulsion current are automatically adjusted to maintain synchronous acceleration, deceleration and cruise. The system thus substitutes information transmission for power transmission.

Two additional advantages are inherent in the linear-synchronous-motor approach. First, positive control of each vehicle by wayside units connected to a central computer allows safe operation at the levels of traffic density and vehicle headway that are absolutely essential for the economic feasibility of any high-speed ground transportation system. Second, the system eliminates the shocks and lurches of wheeled systems and small-gap suspensions, thereby providing a comfortable ride that can be expected to make the system highly acceptable to passengers. The system does sustain slow oscillations (having a period of about one second) that are inherently underdamped and that cannot be damped completely by any known passive technique. We have found that the propulsion-control system provides a simple means for actively damping such oscillations. By adjusting the phase position of the propulsion on the basis of accelerometers carried on the vehicle, heave disturbances of considerable vertical force can be damped. Since the nonlinear repulsive force and the circular geometry couple heave with pitch, sway, roll and yaw, these other motions can also be reduced.

In sum, the magneplane can be likened to a magnetic surfboard riding the forward slope of a traveling magnetic wave whose speed of travel and height are constantly adjusted to keep the surfboard in smooth motion. The controlled "surf" also damps heave motions of the vehicle, thereby providing active control of the ride without the need for springs or shock absorbers, neither of which are very practical at high speeds.

The initial capacity of the guideway would be 6,000 passengers per hour in each direction. It could probably be doubled as the technology matures. The magneplane concept differs fundamentally from conventional railroad thinking: instead of providing 50 megawatts of wayside power capacity to propel a 1,000-passenger train every hour, it provides five megawatts of wayside power to propel a lightweight 100-passenger vehicle every minute. The achievement is made possible by off-line loading and automatic, positive control of all motions of the vehicle. Conventional railroads are basically linear; applied to the Northeast Corridor, for example, they can provide 10-stop trains at hourly or half-hourly intervals. Magneplanes, on the other hand, could serve dozens of stations connected by branch loops, with vehicles operating at one-minute intervals and none of them making more than one or two stops between Boston and Washington.

Initial development of the magneplane at our laboratory was supported by funds from the Sloan Foundation. The effort was later joined by the Avco Systems division of the Avco Corporation, the Raytheon Equipment Division and United Engineers and Constructors, Inc., with material support by the 3M Company and ALCOA. In this way a team of investigators with skills in numerous disciplines was brought to bear on the project. From M.I.T. the chief participants are Yukikazu Iwasa, Sanborn C. Brown, Joel H. Schultz and Robert C. Wallace. The results have had an influence on thinking about magnetic levitation and have also led to support by the new division for Research Applied to National Needs of the National Science Foundation. In addition to detailed studies of the need for a full-scale system, the work has progressed to the construction of a fully operational scale-model magneplane system that will soon be demonstrated publicly.

During this time the Federal Railroad Administration has been grappling with the problem of obtaining a factual base for a policy decision on the technology of high-speed ground transportation. Having committed some $30 million to the development of tracked air-cushion vehicles propelled by linear induction motors, the agency was reluctant to ask Congress to support a major program on magnetic levitation simply because Germany and Japan had done so. Between 1971 and early 1973 the agency provided money for analytical studies of magnetic levitation by the Ford Motor Company and the Stanford Research Institute; the studies confirmed the practical feasibility of both attractive and repulsive levitation and identified the essential requirements and problems in both areas.

The Federal Railroad Administration

did not view these studies as conclusive, however, and is now embarking on full-scale rocket-sled tests at 300 miles per hour on specially built one-kilometer tracks. The objective is to be able to choose by 1975 between what the agency regards as two alternative developments, neither of which can be clearly defined as a necessity. Unfortunately the special equipment constructed in a four-month design period will bear little resemblance to the passenger-carrying system that might eventually evolve from either approach to magnetic levitation, nor will the rocket tests provide meaningful information on propulsion, economics, quality of ride or acceptance by passengers.

It is both amusing and sobering to reflect on an analogous situation some 50 years ago, when airships and airplanes were the competing technologies for high-speed transportation. Airships were engaged in a profitable transatlantic passenger service, whereas airplanes had failed to show commercial potential, even after two decades of development. It is fortunate that the course of aviation was not determined by an irrevocable Government decision, no matter how wise it may have seemed at the time. It was determined in bicycle factories and barns by a handful of entrepreneurs motivated by an idea and a vision—not by the low bidder on a predetermined Government task.

With electromagnetic flight the only difference is that the entrepreneur requires a level of support amounting to a national commitment comparable to the space effort—a commitment that, like the space effort, cannot be justified beforehand on the basis of a clearly demonstrated need. Yet such a commitment is clearly of national concern, since the gap between what the U.S. is doing and what other countries are doing in transportation poses a serious threat to U.S. leadership in a world transportation market certain to be dominated by ground transportation for several decades to come.

Public opposition to the supersonic transport and to airport construction has made it clear that air travel has reached its limit in both speed and passenger capacity. The 7.5 square miles occupied by New York's John F. Kennedy International Airport would accommodate a 95-foot right-of-way from Washington to Boston (and newer airports are four times the size of Kennedy). The nation can no longer afford such inefficient use of land or the environmental impact of having aviation alone satisfy the society's increasing travel needs. The province of Ontario has already rejected a proposed highway system in favor of high-speed mass transportation, and a number of other countries that are not yet committed to elaborate highway or air systems will probably do likewise. Ecologically and economically acceptable high-speed mass transportation is one of the most worthwhile amenities that technology has to offer, and electromagnetic flight now appears to be the means for achieving it.

OPERATING MODEL of the magneplane is employed by the authors and their collaborators to study the motion of the vehicle and the functioning of the system for controlling propulsion. The vehicle and the guideway are 1/25 the size of a full-scale system.

HERPES VIRUSES AND CANCER

It has been known for some time that these ubiquitous viruses can cause cancer in experimental animals. There is now much evidence that a virus of this type is also implicated in some cancers of man

by Keen A. Rafferty, Jr.

The herpes viruses are among the most persistent and universal viruses found in man. They have the mysterious ability to appear and to disappear in individuals without any known route of infection or cause. One of these viruses, herpes simplex Type 1, causes "fever blisters" of the mouth and infects virtually the entire human population. Herpes simplex Type 2 is transmitted by sexual contact and causes sores on the genitals. Another common herpes virus is herpes zoster, which causes chicken pox in children and reappears as shingles in adults and also causes shingles in both children and adults. All these viruses belong to a group of viruses that have been shown to cause cancer in experimental animals. Although the herpes simplex and herpes zoster viruses apparently never cause cancer in man, there is good—albeit still inconclusive—evidence that some extremely malignant cancers in man either are caused by or are prominently associated with viruses of the herpes group. Because herpes viruses can sometimes be isolated from local infections or from tumor cells and grown in tissue culture, they may offer the best hope for proving the etiological role of viruses in human cancer. Moreover, they offer the possibility of developing vaccines to prevent certain forms of cancer, in much the same way that poliomyelitis is now prevented.

The word "virus" is from Latin and means poison or slime. It was first used in its modern sense by Louis Pasteur, who applied it to agents that were not bacteria but were nonetheless associated with disease. Although it was soon accepted that diseases could be caused by viruses, it has been difficult to obtain acceptance for the concept that cancer can be caused by viruses. This is not easy to understand in the face of unequivocal and long-standing evidence from animal experiments. In 1908 V. Ellerman and O. Bang showed that leukemia in chickens could be transmitted by injecting bacteria-free filtrates from infected chickens into healthy chickens. Three years later Peyton Rous showed that the same was true of chicken sarcoma, a malignant deep-tissue tumor. Rous received a Nobel prize for his work, but it is significant that the award was made 55 years after the original discovery.

One reason for the reluctance to accept the concept that cancer can be caused by viruses is that cancer is not contagious in the way that other viral infections such as the common cold and influenza are. Moreover, some human cancers are unmistakably associated with agents other than viruses. For example, it has been known for 200 years that chimney sweeps in England almost always developed testicular carcinoma and that the coal tar in chimney soot was the probable cause. In 1918 K. Yamagiwa and K. Ichikawa of Japan produced cancer growths by painting coal-tar derivatives on rabbit ears, thus showing that cancer could be caused by chemicals. X rays and the radiation from radioactive substances also emerged as causes of cancer. Women who were employed to paint radium on watch dials developed cancer of the lip because they pointed the tips of their brushes by putting them in their mouth. Radiologists who worked with X rays tended to develop cancer of the skin and other forms of cancer before strict safety practices were initiated. More recently resistance to the virus theory of cancer has stemmed from the difficulty of demonstrating the presence of viruses in tumors, particularly tumors of man.

There are two main types of viruses: those containing deoxyribonucleic acid (DNA) and those containing ribonucleic acid (RNA). In animal tumors caused by RNA viruses the virus may be reproduced continually and new viruses released without damaging the host cells. RNA viruses are believed to be responsible for some human leukemias, in which white or red blood cells divide prodigiously. With DNA viruses, however, replication of the virus and its liberation from the host cell always seem to be accompanied by the death of the cell. Thus if a DNA virus were to replicate in a tumor, it would lead to the destruction of the tumor cells and the virus itself would eventually disappear.

Herpes viruses are DNA viruses. There are two other kinds of DNA viruses that are known to cause tumors. They are the adenoviruses, which are responsible for colds and other respiratory diseases, and the papova (papilloma-polyoma-vacuolating) viruses, which may be responsible for a variety of tumors (including the wart, a benign growth that can be transferred from person to person by injection of the virus).

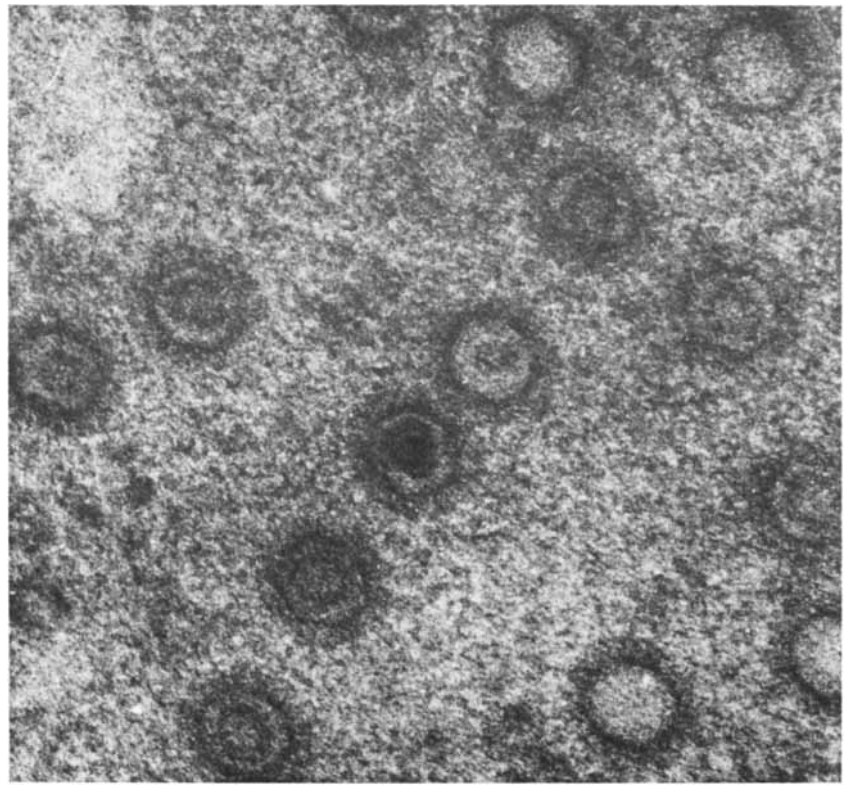
The realization that the herpes viruses might be involved in the formation of tumors emerged from the study of the curious epidemiology of certain tumors in frogs and chickens. The leopard frog (*Rana pipiens*) is particularly susceptible to a kidney carcinoma, implying that there are special causative factors. In 1932 Balduin Lucké of the University of Pennsylvania School of Medicine began studying such tumors, and he soon discovered that the nucleus of some of the tumor cells contained large, pale, pink-staining inclusions. These inclusions could easily be seen in microscopic section. It was then believed (and still is) that inclusions of this type (Cowdry Type *A*) are caused by viruses. Lucké's finding was therefore a clear indication that the frog kidney tumor, now known

as the Lucké tumor, was the result of viral activity. Unfortunately the appearance of the characteristic inclusions was sporadic, and there was no reliable way to directly detect the viruses themselves.

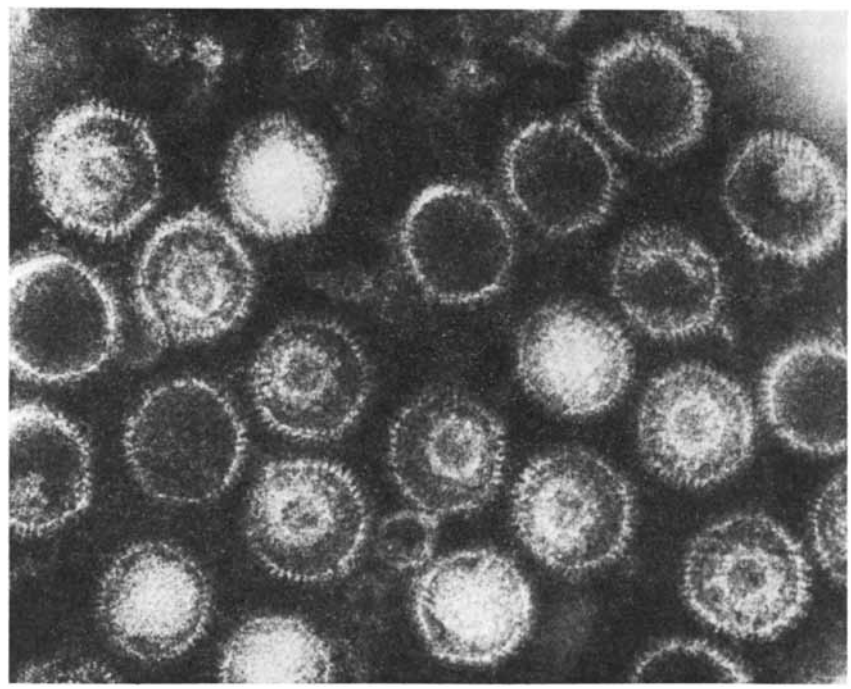
In 1956 Don W. Fawcett, who was then working at the Cornell University Medical College, made electron micrographs of Lucké tumor inclusions and found unmistakable evidence that they contained virus particles. At about the same time I began to look for similar inclusions in frog tumors. For several years, in spite of examining hundreds of laboratory-kept frogs, I was unable to find a single inclusion. Finally an accidental observation by Maria Roberts, a graduate student, suggested that the frog tumors have inclusions in the cell nucleus during the winter but not during the summer. She found inclusions in several frogs that had arrived in consecutive shipments during the winter and had been stored for a month in a cold room before being examined. On the other hand, all the tumors I had examined had developed spontaneously in frogs that had been in a warm laboratory for several months.

The obvious experiment was then carried out. Biopsies were made of the kidneys of laboratory frogs with spontaneous tumors. They confirmed that there were no intranuclear inclusions present. The frogs were then placed in a refrigerator for various periods of time, after which they were killed and the kidney tumors were examined. Inclusions appeared in the tumor cells, so that we were able to conclude that virus production had probably been initiated by prolonged low temperature. Further experiments indicated that the cells producing the virus died with the release of virus and that the virus appeared in the urine, a not unexpected finding in view of the location of the tumor [*see illustration on next two pages*].

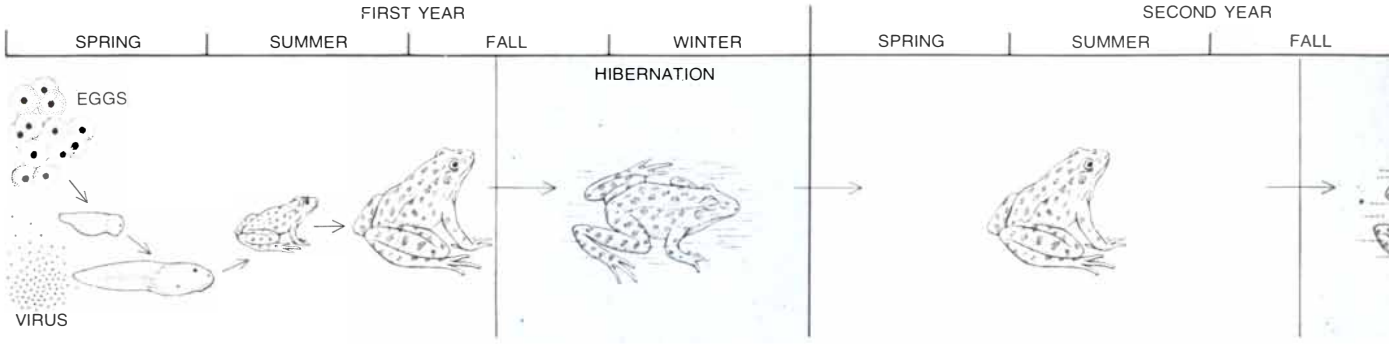
Next we were able to isolate a virus from the kidney tumors by growing it in tissue cultures of frog cells. Electron-microscope observations showed that it was a herpes virus. The tissue-culture virus did not, however, produce tumors when it was injected into tadpoles. This was in spite of the fact that it had been found by K. S. Tweedell of the University of Notre Dame that cell-free extracts of actual inclusion tumors did produce tumors. Later Maneth Gravell of the St. Jude Children's Research Hospital in Memphis showed that the tissue-culture virus and the most prevalent virus in the tumor tissue have different DNA's and



EPSTEIN-BARR VIRUS PARTICLES are seen enlarged about 160,000 diameters in this electron micrograph of a thin section of a human lymphoblast provided by M. A. Epstein of the University of Bristol Medical School. The particles are very similar in appearance to typical herpes virus particles. They have a hexagonal profile and either are empty or contain a nucleoid that is ring-shaped or dense in appearance. These virus particles were isolated from a tumor-cell tissue culture derived from a patient with Burkitt's lymphoma.



VIRUS PARTICLES that are virtually identical with the Epstein-Barr particles have been isolated from a tissue culture of human nasopharyngeal carcinoma tumor cells. Particles are shown enlarged about 160,000 diameters in this electron micrograph provided by Guy de-Thé of the World Health Organization's International Agency for Research on Cancer.



LIFE CYCLE OF THE LUCKÉ-TUMOR VIRUS proposed by the author is depicted. Adult leopard frogs (*Rana pipiens*) infected with the virus often develop kidney carcinomas. The process is thought to begin when virus in the spawning water infects eggs or tadpoles. In their first and second year there is no sign of tumor growth in young frogs. In the third or fourth summer some adult frogs begin to develop kidney tumors. These "summer tumors" do not produce virus. Depending on the locale and other factors frogs that develop tumors may die during the summer or may live until the winter hibernation. During hibernation tumor growth ceases

that, although they are both herpes viruses, one is probably not derived from the other. At this point we can only speculate that two viruses are present in the tumor, or that perhaps the virus was profoundly altered in the tissue culture (an unlikely possibility), or that the tissue-culture virus was present in the original tissue-culture cells. It is also possible that the amount of virus from the tissue culture injected into the tadpoles was insufficient to induce tumors in the animals.

Recently William Collard and his colleagues at the Saint Louis University School of Medicine and at the Tulane University School of Medicine have shown that RNA from the summer-phase frog tumors and DNA from the viral inclusions in winter-phase tumors will "hybridize," that is, when fragments of the threadlike molecules of the two nucleic acids are mixed together, some of them adhere to others. This indicates that stretches of the two molecules are intimately related to one another. Since the RNA could only have been formed by being copied from the viral DNA, these results provide additional evidence that some herpes virus genetic material is present in the apparently virus-free tumor. At any rate there is little doubt that the Lucké tumor is caused by a herpes virus, although it is not yet certain which herpes virus is involved. Quite recently Robert F. Naegele and Allan Granoff of the St. Jude Hospital completed the cycle by recovering the infectious agent from a tumor that had been induced by injection of a tumor extract, and using this in turn to induce still another tumor. Thus the famous "Koch postulates" have been satisfied, and the tumor is firmly placed among those transmitted by an infectious agent.

The early experiments on frog tumors served as an important forerunner of current work on the causes of human cancers. The results obtained alerted other investigators to the prospect of discovering human tumors of herpes virus origin.

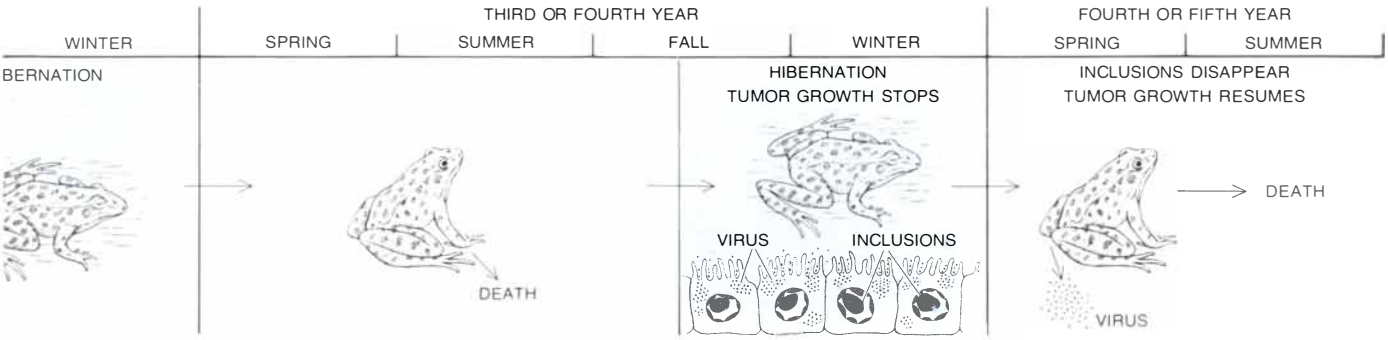
A malignant tumorous growth of the lymphoid system in chickens, known as Marek's disease, had led to the first unequivocal proof that a herpes virus is the cause of a cancer. Almost unknown a few years ago, except among a handful of academic veterinary pathologists, Marek's disease became the scourge of U.S. commercial chicken flocks in the late 1960's. At first the disease that afflicted the chickens was called polyneuritis or fowl paralysis. Later it was diagnosed as the disease that had been described in 1907 by Joseph Marek of Hungary. Marek had not recognized, however, that the disease is a malignant one. Cells of the lymphoid system become cancerous and invade the nerve cells, resulting in paralysis of the fowl. As the disease progresses other organs are invaded, and the bird eventually dies.

In 1967 several workers simultaneously demonstrated that when a cell-free filtrate from an infected chicken is injected into a healthy bird, a lymphoid tumor results. Although evidence of viruses was seen in some tumor tissue, such observations were rare. That brought up the question of how the virus was transmitted from one bird to another.

Keyvan Nazerian of the Department of Agriculture Regional Poultry Research Laboratory in East Lansing, Mich., using an antigen-antibody test, examined cells from various tissues from chickens with Marek's disease in an effort to determine how and where the virus was being released into the environment. The tumor cells seldom contained a virus, but when Nazerian examined skin sections, he found that cells of the feather follicles frequently tested positive for the presence of a virus. It was quickly shown that chicken feathers and dander (bits of sloughed-off dead skin) are highly infectious, and the route of transmission of the virus in commercial flocks seems clearly established. The virus was identified as a herpes virus. An interesting result of this work is the finding that herpes viruses are not necessarily delicate, as most viruses are. The virus causing Marek's disease is extremely resistant to drying and may remain infectious for a long time after it has been liberated in chicken dander.

An attenuated strain of the Marek's-disease virus and a related virus recovered from turkeys have both been used for live-virus vaccines, with the result that Marek's disease has been effectively eradicated in treated flocks. This suggests that cancer vaccines for man may be possible. A point worth noting is that in Marek's disease the tumor virus is not produced in the tumor cells. The virus transforms normal cells into tumor cells without necessarily producing viral progeny. Indeed, tumors that do produce new herpes viruses would probably be self-limiting, since virus production is thought to result in the death of the cell involved. It is an interesting fact that the replication of the herpes virus of Marek's disease takes place in nonmalignant skin cells that normally die in any case.

A few other animal tumors have been found to be caused by, or to be closely associated with, herpes viruses. One of them is pulmonary adenomatosis of sheep, a growth in the lungs that is responsible for substantial commercial losses among sheep raisers. A herpes virus of the cottontail rabbit was found by Harry C. Hinze of the University of Wisconsin Medical School to be capable of inducing malignant lymph tumors when



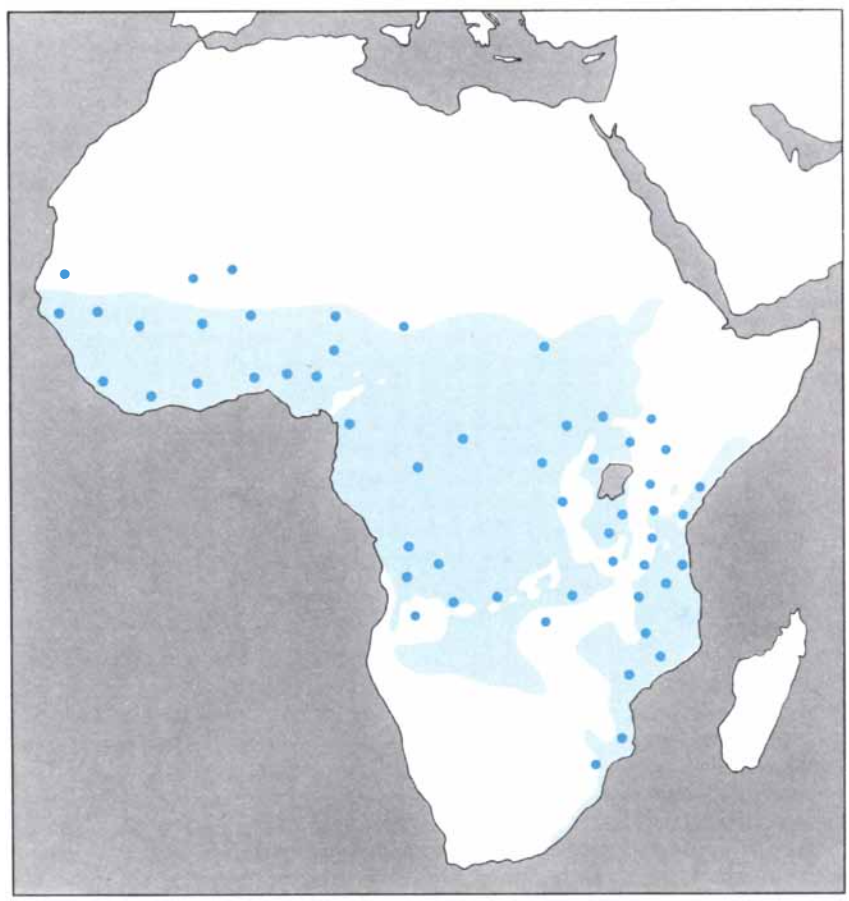
and inclusions appear in the nucleus of the tumor cells. Many of the tumor cells die, indicating that the virus has probably been activated by the low temperature and is replicating itself. The virus accumulates in the urine, and in the spring the tumor-bearing frog liberates large amounts of the virus into the spawning water with the urine it voids. In warm weather the inclusions in the tumor cells disappear, tumor growth resumes and the frog soon dies. Thus the Lucké tumor appears to be transmitted by virus that is produced in the kidney of tumor-bearing frogs that go into hibernation and then survive a winter after the tumor first develops.

it was injected into young rabbits. Both of these virus-tumor systems are being studied further.

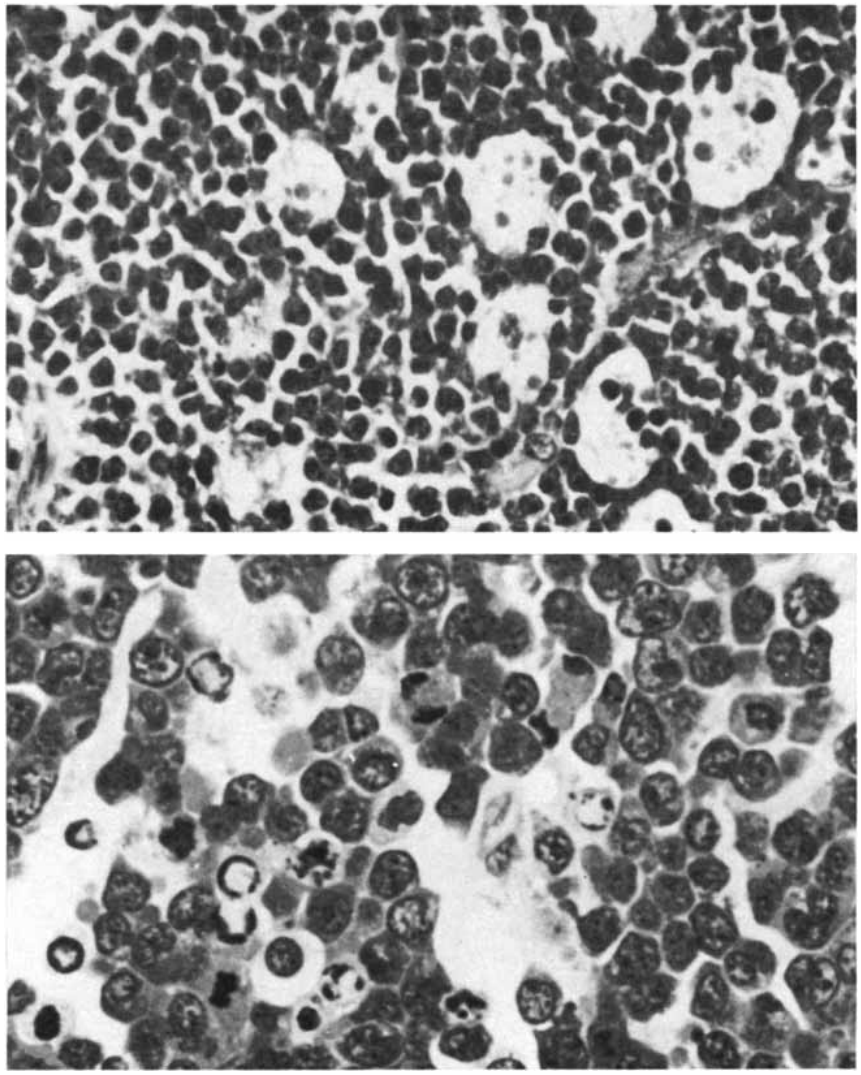
The herpes viruses that cause the Lucké tumor and the Marek's-disease tumor are both immunologically related to a herpes virus that is closely associated with a human cancer: Burkitt's lymphoma. In 1958 Denis Burkitt, a missionary surgeon in Uganda, reported that a large number of African children between the ages of four and 16 suffered from a tumor in the connective tissue of the jaw. These jaw sarcomas were thought to occur very rarely in Europe and in the U.S., but children of white missionaries stationed in Uganda sometimes acquired the disease. Burkitt also observed that all the cases came from areas in which malaria was endemic, or from areas in which temperature and rainfall conditions were such that malarial species of mosquitoes could be expected to be numerous [*see illustration at right*].

Intensive study of Burkitt's lymphoma began immediately after he published his report of this cancer that seemed to be transmitted in the same way as an infectious disease. New discoveries quickly followed. It was found that the tumor is rare in African children older than 16. One possible explanation is that exposure to some infectious agent is frequent in Uganda, so that everyone who is born in the area is likely to be exposed early in life. The adult European visitor to the area, on the other hand, may escape exposure until soon after his arrival. Moreover, a time-clustering phenomenon that had been suspected on the basis of hospital admissions was confirmed by M. Pike of the University of Oxford and R. H. Morrow, Jr., of Harvard University. For example, in Bwamba County in Uganda no cases of Burkitt's lymphoma had been reported before 1966; then seven cases were reported in 27 months and none has been reported since. Findings of this kind clearly suggest that the disease is transmitted by an infectious agent. Later K. Booth, Burkitt and their colleagues found that the lymphoma also occurs among the people of Papua and New Guinea. In both areas the conditions for malarial mosquitoes are ideal and malaria is endemic.

Although the lower jaw is most frequently the site of the tumor in Burkitt's lymphoma, tumors also arise at other sites: the upper jaw, the thyroid, the ovaries, the liver and the kidneys.



BURKITT'S LYMPHOMA IN AFRICA clusters (*colored dots*) **in those areas where climatological factors make malaria endemic. In such areas (*light color*) the altitude is less than 5,000 feet, the seasonal mean temperature is 60 degrees Fahrenheit or higher and rainfall is 20 inches per year or more. Burkitt's lymphoma is a human sarcoma of the jaw closely associated with a herpes virus. Data are taken from Denis Burkitt's original reports.**



BURKITT'S LYMPHOMA TUMOR in microscopic section (*top*) displays the characteristic "starry sky" effect of such tumors. The white "stars" are large histiocytes, cells that engulf debris from dying cells. Engulfed material is visible inside the histiocytes. A higher magnification of the tumor cells (*bottom*) reveals the empty-looking, circular structures (Type *A* intranuclear inclusions) that are characteristic of herpes virus infections. The photomicrographs were provided by D. H. Wright of the University of Birmingham.

D. H. Wright, who was then working at Makerere University in Uganda, made an intensive microscopic study of the tumor. He found that it originates in cells of the lymph nodes. Although the tumor is malignant in that it invades and destroys surrounding tissues (and ultimately causes the death of the patient unless he is treated), the cells of the circulating blood seem to remain normal, at least until the disease is in an advanced stage. This constitutes the difference between the leukemias, in which blood cells become abnormal, and the lymphomas.

Another discovery made by Wright was that microscopic sections of tumors from Burkitt's lymphoma contain histiocytes, or large, clear cells that engulf debris from dying cells of the tumor mass. It is this discovery that has helped other investigators to identify cases of Burkitt's lymphoma in parts of the world where it is rarer and less likely to arise in its most characteristic site: the lower jaw. There were some unexpected findings. With the new diagnostic aid reports began coming in that Burkitt's lymphoma occurs regularly in North and South America and in all Europe, including colder areas such as Scandinavia, where mosquito-transmitted malaria is virtually unknown. Thus it seems that if Burkitt's lymphoma is indeed transmitted as an infectious disease, it need not be transmitted by mosquitoes (and very likely never is). To keep matters in perspective it should be pointed out that even in Uganda, where the disease is perhaps commonest, the probability of developing it is one in 2,000. In the Temperate Zone the probability is far lower.

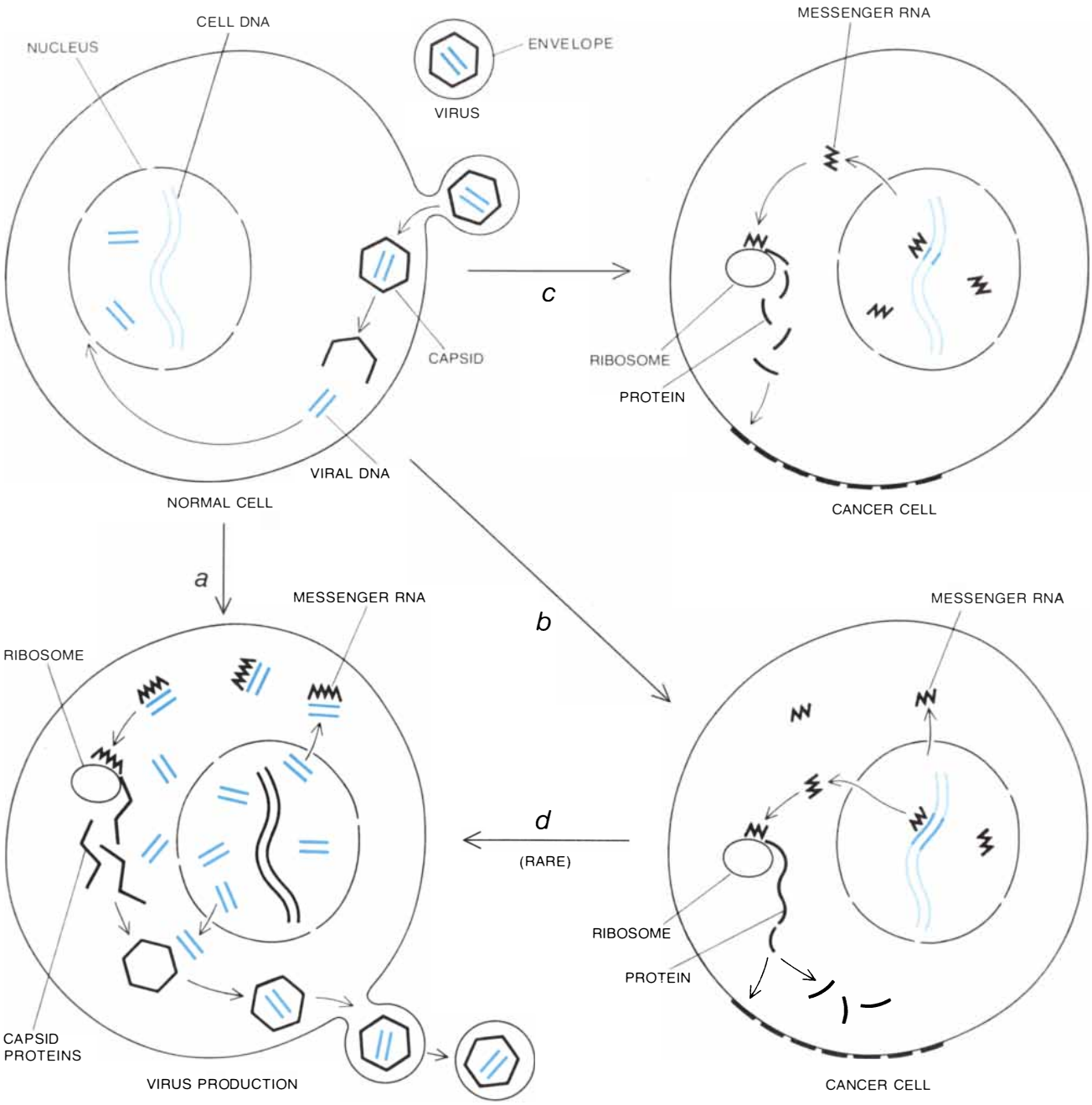
Once a virus was suspected of causing Burkitt's lymphoma intensive efforts were made to show that a virus is present in the tumor. This, however, proved to be difficult. Examination of tumor sections for virus particles with an electron microscope was fruitless, as were efforts to infect animals and tissue cultures with extracts of the tumor. Indeed, even efforts to grow the tumor cells themselves in tissue culture met with little success initially. Then M. A. Epstein and Y. M. Barr of the University of Bristol Medical School undertook to work with fresh tumor tissue flown in from Uganda, and similar experiments were begun by R. J. V. Pulvertaft, who was then working at the University of Ibadan in Nigeria. Epstein and Barr found that cells were growing in two of their cultures but that they grew in suspension in the culture medium rather than attached to the surface of the culture vessel, where tissue-culture cells normally grow. In retrospect, since lymphoma cells and other lymph-node cells periodically circulate in the blood, it might have been expected that the cells would grow in suspension. When the cells in the Epstein-Barr culture were examined with the electron microscope, herpes virus particles could be clearly seen. Pulvertaft was also able to demonstrate the presence of a herpes virus in his tissue cultures.

Epstein and Barr showed that the lymphoma cells could be propagated indefinitely in tissue culture, and that one of the cultures (designated EB-3) continually produces an infectious virus. Only about 2 percent of the cells die at any one time and liberate the virus. Other workers have shown that the virus, now called the Epstein-Barr virus or EBV, can infect normal blood cells in culture and convert them to production of the virus.

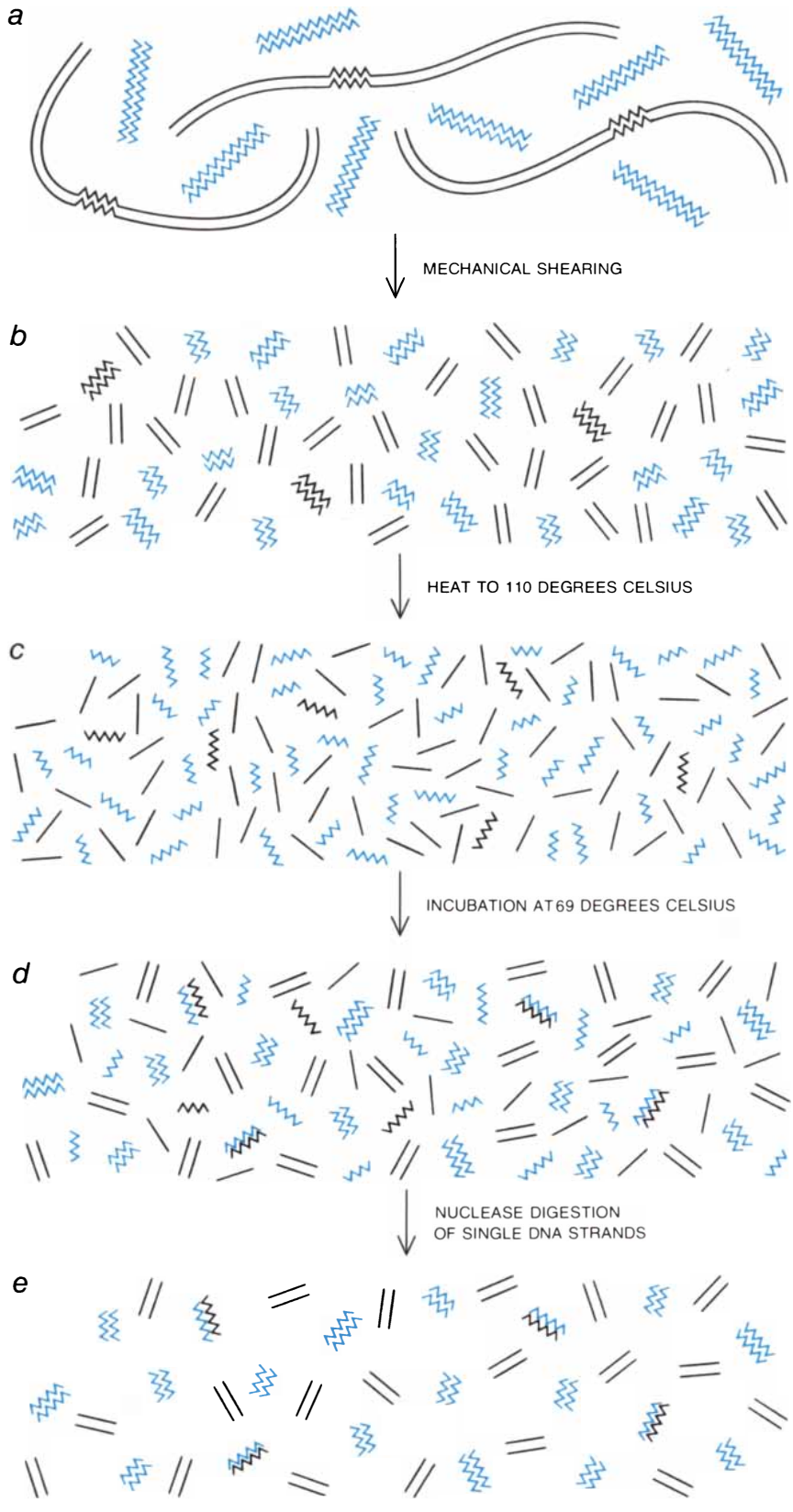
The identification of a herpes virus in Burkitt's lymphoma was significant because it is well known that some herpes viruses are quite widespread in the human population. In 1968 Werner and Gertrude Henle and V. Diehl of the University of Pennsylvania School of Medicine discovered that infectious mononucleosis, the "kissing disease" so common among American and European college students, not only is caused by a herpes virus but also cannot be distinguished from the Epstein-Barr virus, even with the most sophisticated nucleic acid hybridization or antigen-antibody tests. Therefore the two are presumably iden-

tical. This intriguing discovery was made fortuitously: a technician in the Henles' laboratory was found to have developed antibodies against the Epstein-Barr virus after she had contracted infectious mononucleosis. The Henles followed up the lead and showed that when cells from the blood of a person with infectious mononucleosis are grown in a suspension culture, they will produce a virus that is indistinguishable from the culture-grown Epstein-Barr virus.

The blood of patients with infectious mononucleosis contains dividing cells, which are never seen in the blood of uninfected people. The condition has, in fact, been described as a "self-terminating malignant disease." Microscopic examination of the dividing cells reveals that they bear a close resemblance to the proliferating white cells of acute lymphoblastic leukemia. Moreover, microscopic sections of lymph nodes from patients with infectious mononucleosis are



INVASION OF A CELL by a herpes virus normally leads to the production of new virus particles. The viral DNA replicates inside the cell's nucleus. Then some of the viral DNA moves into the cytoplasm. There it transcribes messenger RNA (mRNA), which attaches to ribosomes (*a*). Viral-coat proteins are produced and assembled into capsids, which combine with the DNA to produce a complete virus particle. The cell dies when the virus particles are released. In some instances (*b*) the complete viral DNA combines with the cell DNA. Messenger RNA is transcribed, and when this messenger RNA attaches to ribosomes, it produces proteins that are thought to migrate to the cell's surface. These protein patches on the cell's surface are believed to be involved in the transformation of the cell into a cancer cell. In rare circumstances the viral DNA in a cancer cell may revert to virus production (*d*). Another possible route for inducing malignancy (*c*) may occur when only part of the viral DNA combines with the cell's DNA. Because part of the viral DNA is lost, complete new viruses cannot be formed. Messenger RNA can be transcribed, however, and it may be responsible for the synthesis of proteins in the cell that are involved in the production or in the maintenance of a malignant state.



HYBRIDIZATION TECHNIQUE enables researchers to measure the amount of viral DNA (or RNA) present in cell DNA. Radioactive viral DNA (*colored segments*) is mixed with cell DNA (*a*). The molecules are sheared by forcing the mixture through a narrow aperture. The resulting short segments of DNA (*b*) are heated to unravel the double strands into single ones (*c*). When the mixture is cooled, some double strands re-form, enabling radioactive viral DNA to hybridize with any virus segments in the cell DNA segments (*d*). The mixture is treated with an enzyme that digests only single DNA strands (*e*). The radioactivity of the double strands is measured to determine how much DNA has hybridized.

remarkably similar in some respects to sections of Burkitt's-lymphoma tumors.

Do people who have had infectious mononucleosis have a greater probability than the general public of contracting cancer later in life? Epidemiological studies undertaken by Robert W. Miller of the National Cancer Institute and Gilbert W. Beebe of the National Academy of Sciences–National Research Council indicate that a history of infectious mononucleosis apparently is not associated with the incidence of cancer. This finding is consistent with the view of many that virtually everyone is infected with the Epstein-Barr virus, including those who never develop infectious mononucleosis.

There are a number of diseases of the blood system that seem to be prime candidates for association with, or even causation by, herpes viruses. Patients with Hodgkin's disease, a cancer of the lymphoid system, are almost always found to have high levels of antibody against the Epstein-Barr virus. Patients with acute lymphocytic leukemia also tend to show high levels of Epstein-Barr-virus antibody. In Hodgkin's disease there is a malignant proliferation of reticulum cells that, when they are viewed in the microscope, bear a striking resemblance to the histiocytic cells of infectious mononucleosis and the characteristic "starry sky" cells of Burkitt's lymphoma. Another cancer that is associated with the Epstein-Barr virus is nasopharyngeal carcinoma. Lloyd V. Old and his colleagues at the Sloan-Kettering Institute made this discovery while they were using serums from nasopharyngeal-carcinoma patients as controls in antigen-antibody studies of serums from patients with Burkitt's lymphoma.

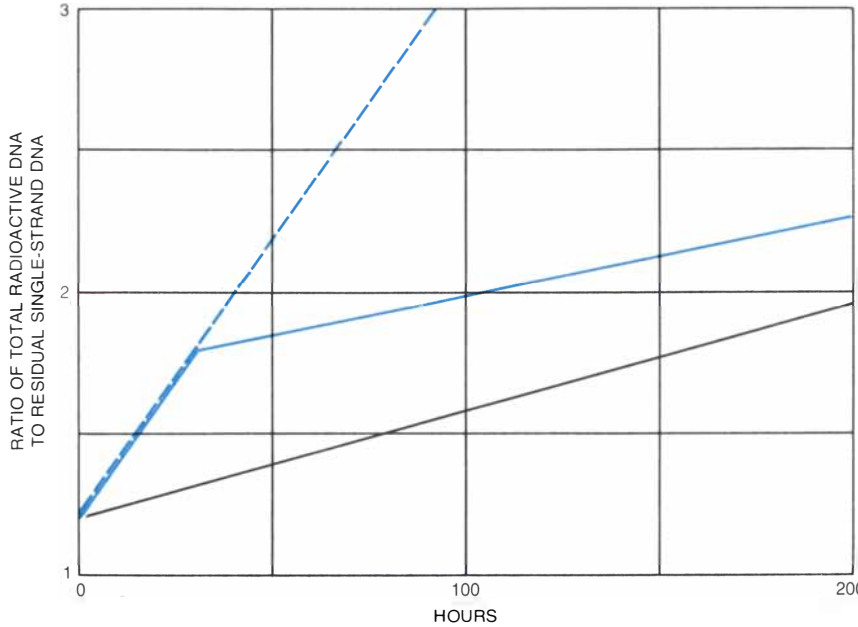
As the term indicates, nasopharyngeal carcinoma originates in the nasal cavity, the pharynx and the rear of the oral cavity. It is a true carcinoma in that the tumor consists of proliferating surface epithelial cells rather than lymphoid cells of deep tissues. Until this discovery was made only tumors of deep tissues had been associated with the Epstein-Barr virus. It is particularly interesting that a prominent structure of the upper pharynx is the tonsils, which consist of lymphoid cells covered by surface epithelium. This epithelium is perforated, and material in the mouth comes in contact with the tonsillar tissue. Clearly the lymphoid cells may become infected and transmit the infection to the epithelial cells. Furthermore, throat swabs from patients with infectious mononucleosis are a prime source of Epstein-Barr virus.

The incidence of nasopharyngeal carcinoma is extraordinarily high in certain populations. Virtually all the known cases are seen in Chinese from South China, Malaysia and Singapore. H. C. Ho of the Queen Elizabeth Hospital in Hong Kong has assembled data suggesting that the lifetime incidence in Hong Kong may be as high as one in 200. Among the "boat people" of Hong Kong the incidence is even higher. People of Chinese descent who are born in the U.S. show a lower incidence of the disease but still one that is higher than the incidence in the general population. As with Burkitt's lymphoma, nasopharyngeal carcinoma is found throughout the Western world at low incidence. Even though the data are still inconclusive, they do suggest that some cultural factor may be involved in addition to a possible viral one.

Guy de-Thé of the World Health Organization's International Agency for Research on Cancer in Lyon has grown cells from nasopharyngeal-carcinoma-tumor tissue in suspension culture. This culture produces virus particles that appear to be identical with those of Burkitt's lymphoma, but further studies are needed.

The genital form of herpes simplex (Type 2) is found in sores or reddened spots on the cervix and vaginal walls of infected women and on the penis and in the genital tract of men. Ysolina M. Centifanto and her colleagues at the University of Florida College of Medicine have found that 15 percent of men of all ages from 15 years on harbor Type 2 herpes simplex in the genital tract, for the most part in the prostate and vas deferens. The virus is seldom isolated from the female genital tract; women tend to have acute, short-lived and recurrent herpes virus genital infections. It may be that many men harbor the virus without developing the symptoms of infection but are capable of transmitting it.

André J. Nahmias of the Emory University School of Medicine has been able to establish a convincing association between the Type 2 herpes-simplex virus and cancer of the cervix, and Centifanto has described one case in which the herpes virus could be seen in the cells of tissue from cancer of the prostate. An intriguing bit of evidence comes from a study undertaken by Niza Frenkel and her colleagues at the University of Chicago, who were working with tumor tissue supplied by Nahmias. The study involved hybridization experiments with DNA from cervical-carcinoma cells [*see illustration on this page*]. They found that some DNA from Type 2 herpes-simplex virus is indeed present in these cells, but that 60 percent of the viral DNA molecule is missing. The missing portion could explain why neither the complete virus nor viral antigens can be isolated from the tumor tissue. The portion of the viral DNA that is present could become attached to the normal cell DNA and could be responsible for maintaining the cell in a malignant state. In addition the missing part would be required for the production of the complete infectious virus.



RATE OF HYBRIDIZATION is determined by measuring the amount of radioactivity in the final hybridization product over a period of time. The measurements are converted so that the vertical axis is a function of the increasing proportion of double-strand radioactive DNA present. If the cell DNA contained no viral DNA segments, the rate would be what is shown by the black curve. If the cell DNA contained complete copies of the viral DNA, the rate would rise constantly, as is shown by the broken colored curve. In a recent hybridization experiment with human cervical-carcinoma cells and herpes virus DNA, Niza Frenkel and her colleagues at the University of Chicago found that the rate rose but then dropped off, as shown by the solid colored curve. From these results they calculated that only about 40 percent of complete-virus DNA was present in cervical-carcinoma-cell DNA.

The recent discovery by Luis V. Melendez of the New England Regional Primate Research Center that the New World squirrel monkey harbors a herpes virus capable of inducing tumors in owl monkeys and marmosets is highly significant. For the first time it is possible to conduct cancer-transmission experiments on animals much more closely related to man than frogs and chickens. What is even more promising is that the resulting tumors in monkeys are similar to herpes-associated tumors in man. Unlike cultured cells infected with the Epstein-Barr virus, the cells harboring the squirrel-monkey virus (called herpes saimiri) grow on the surface of the culture vessel. This makes it easier to produce large amounts of the virus for investigation or for vaccine manufacture.

Quite recently two groups of investigators showed that the Epstein-Barr virus can induce lymphoid tumors in monkeys. One of these workers was Epstein himself, who used the virus to induce a tumor in an owl monkey. Thomas C. Shope, D. C. DeChairo and I. George Miller, Jr., of the Yale University School of Medicine induced malignant lymphomas in four cotton-top marmosets. Since such experiments cannot be performed in man, this is about as far as we can go in showing directly that a virus from human cancer is in fact a cancer-causing virus, which is a different matter from finding a virus in a tumor.

Thus the prospect is good that it will eventually be demonstrated beyond a reasonable doubt that herpes viruses can cause some forms of cancer in man. Nonetheless, an alternative explanation cannot be ignored. It is that herpes viruses are present in some human tumors merely as passengers and are not involved in the development of the tumor itself. If any human cancers are shown to be caused by a virus, there is hope that those forms of cancer can be eliminated, or at least much reduced in frequency, by the development of suitable vaccines or by finding ways to interrupt the primary routes of infection.

THE SLING AS A WEAPON

This manual missile launcher is not as well known as the archer's bow, yet it was used by lightly armed troops from India and Persia to Greece and Rome and even survived the advent of gunpowder

by Manfred Korfmann

Everyone knows that David killed Goliath with a sling, but what is the place of the sling in the history of technology? The fact is that slings were a regular weapon of warfare in Europe and the Near East at least from the Bronze Age until the 17th century of our era. Moreover, the sling has been the favored long-range weapon among numerous peoples, past and present, the world over. In Mesopotamia, in Persia and in Greece and Rome a slinger was considered a match for an archer. In that part of the world the sling had probably been known from the beginning of Neolithic times, some 10,000 years ago, and may even have been used toward the end of the Paleolithic.

David's victory over Goliath is often taken to be an allegory, but considering the nature of warfare in David's time, it might better be seen as an example both of the very real skill that slingers possessed and of the trust that they put in their weapon. A review of the encounter, as recorded in the First Book of Samuel, supports this view. David, it will be remembered, was the eighth son of Jesse. As the youngest son, he tended the family flocks. The occupation accounts for his familiarity with the sling; the weapon is used to this day by herdsmen protecting their animals. David had become harpist and armor-bearer to Saul, the king of Israel at a time when the Israelites were at war with the Philistines. The two armies had camped a short distance apart. Each day a Philistine champion, Goliath of Gath, "whose height *was* six cubits and a span," would come out of the enemy camp and offer to fight any Israelite champion and let the outcome of the single combat decide the war.

David came to the Israelite camp at a time when 40 days of Goliath's challenges had failed to rouse a single Israelite warrior. David volunteered to fight Goliath, but he refused to wear the armor or carry the arms that Saul pressed on him. At this point the narrative goes:

"And he took his [shepherd's] staff in his hand, and chose him five smooth stones out of the brook, and put them in a shepherd's bag which he had... and his sling *was* in his hand: and he drew near to the Philistine."

"And David put his hand in his bag, and took thence a stone, and slang *it*, and smote the Philistine in his forehead, that the stone sunk into his forehead; and he fell upon his face to the earth."

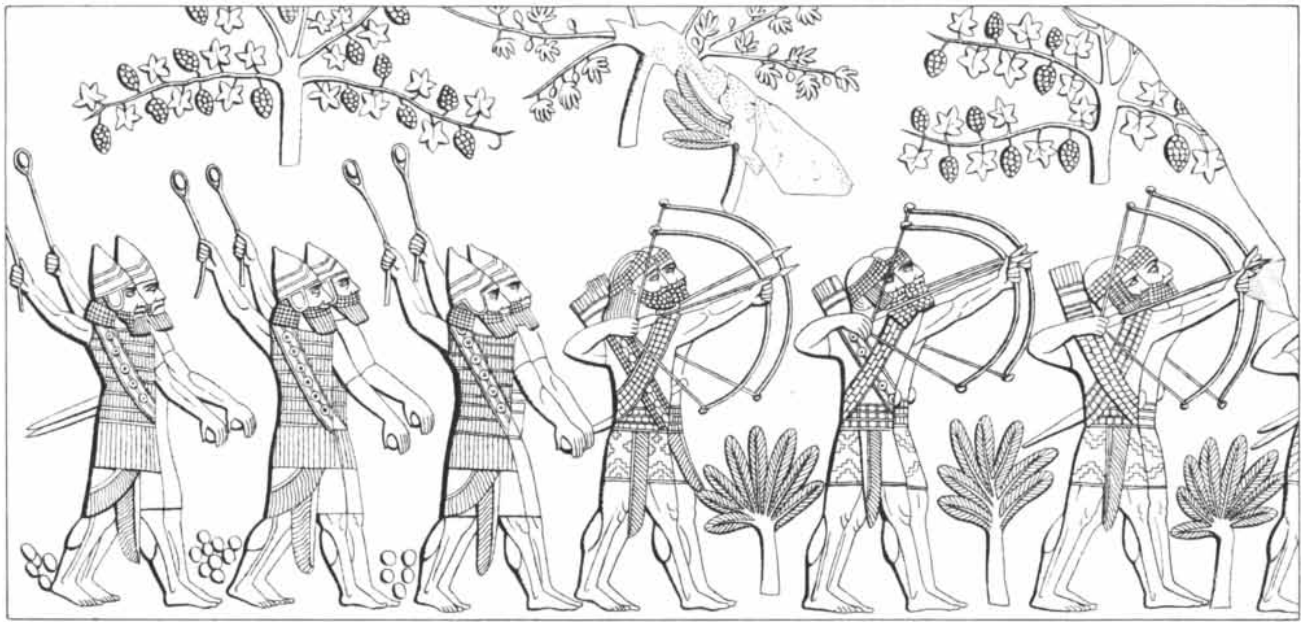
"Therefore David ran and stood upon the Philistine, and took his sword, and drew it out of the sheath thereof... and cut off his head therewith. And when the Philistines saw their champion was dead, they fled."

Goliath had been dressed in full armor for the encounter: a metal helmet, a coat of mail, metal greaves on his legs and a small shield slung on his back. He had a bearer who preceded him with a larger shield. The sword David used to cut off Goliath's head is not described, but Goliath's spear had a shaft "like a weaver's beam" and a heavy spearhead. This equipment is not unlike what was carried by a Greek hoplite, or heavy infantryman. It is meant for shock tactics; the spear is not for throwing but for thrusting or for defense against cavalry attack. Goliath's armor and arms, except possibly for the spear, would also have been appropriate for single combat with another warrior similarly armed. They were entirely unsuited, however, for the pursuit of an unarmored, swift-footed adversary, and so David would have been quite safe as long as he kept his distance.

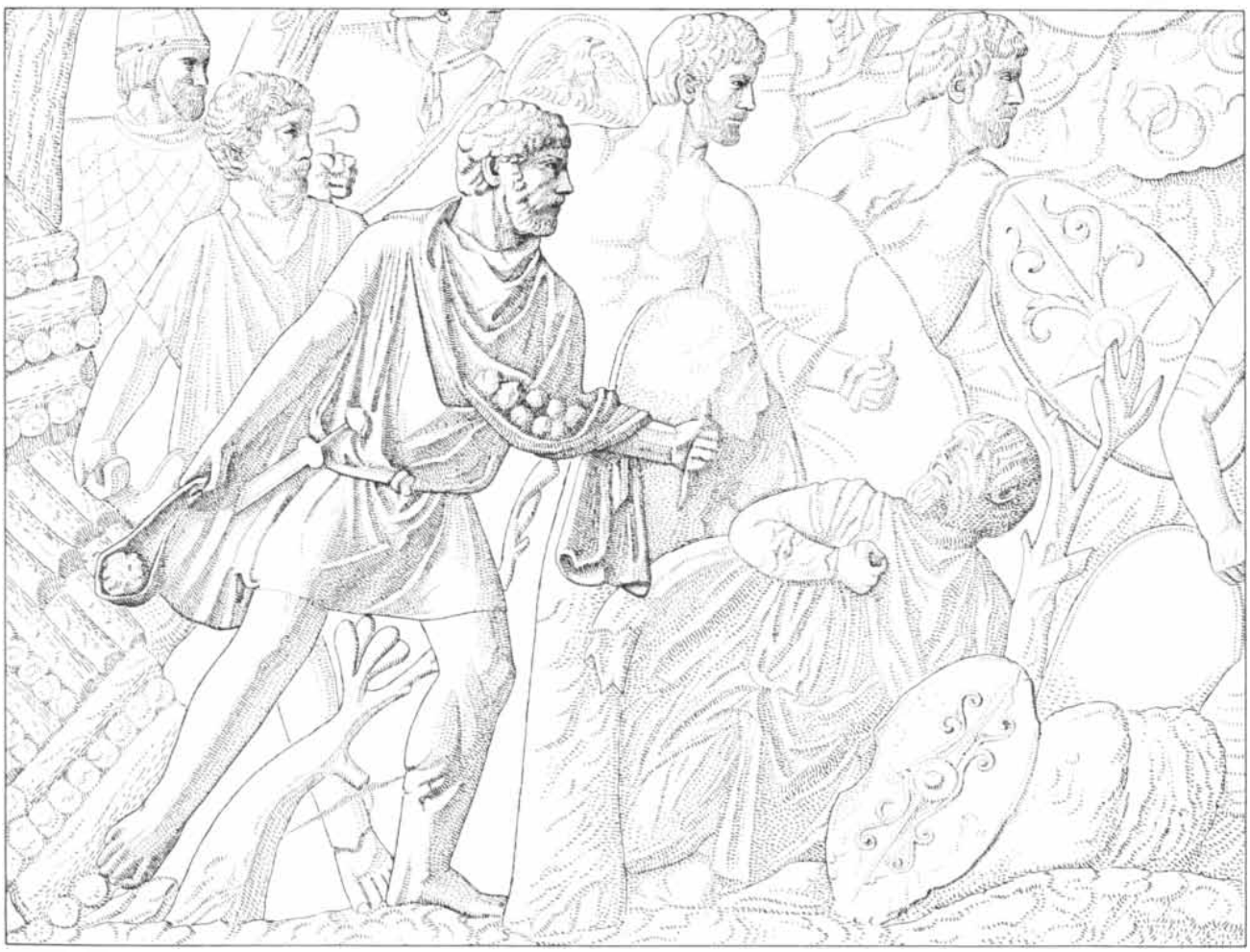
David would have had no intention of closing with his opponent; the sling is a long-range weapon. At the same time, whatever confidence David may have placed in God's help in preparing him for the encounter, he chose not one pebble for his sling but five. If his first shot had not struck Goliath in the face, the vital yet unprotected area that David had certainly aimed at, four more pebbles remained at his disposal. On balance it seems fair to credit David's victory not to divine intervention but to his skill as a slinger.

David is the best-known biblical slinger, but he is not the only one. The left-handed slingers of Benjamin (Judges) caused heavy casualties among the Israelites, and David's own elite corps "could use both the right hand and the left in *hurling* stones" (Chronicles). Why, then, has the sling remained almost unnoticed as a weapon of war? There is some hint of an answer in the *Iliad*. In Homer's account of the siege of Troy an allusion is made to the Locrians, warriors without armor who trusted in their bows and "well-twisted wool." The Greek word for "sling," however, appears only once in the entire epic. Even then the sling is mentioned not as a weapon but as an improvised bandage: one Trojan warrior binds up the wounded hand of another "with a strip of well-twisted wool..., the sling [used by] the shepherd of the host." It would appear that Greek peltasts, lightly armed troops whose ranks included slingers, javelin-

DAVID AND GOLIATH appear in the photograph on the opposite page as seen on the wall of a 10th-century Armenian church built on an island in Lake Van. Goliath is seen with his sword drawn. David, his long-range weapon loaded and ready, is standing much too close to Goliath in this portrayal; a cast of 250 yards was not unusual for the slingers of that time.



ASSYRIAN SLINGERS, swinging their slings parallel to their bodies, stand behind the archers in this drawing based on a relief from Nineveh showing one of the campaigns of Sennacherib (704–681 B.C.). Their place in battle suggests that they outranged archers.



ROMAN AUXILIARY in the Dacian wars, his sling at the ready, carries extra missiles in the fold of cloak flung over his shield arm. The figures in this drawing appear on a monument in Rome, the column erected to honor the Emperor Trajan's victory in the wars.

throwers and archers, received little recognition in the days when only hand-to-hand combat between armored warriors was considered honorable. Pictorial representations of hoplites (and even of archers and javelin-throwers) are common enough, but slingers are seldom seen.

Lightly armed troops nevertheless played a major role in warfare during classical Greek times. To them was given the initiation of battle. The shower of javelins, arrows and slingstones they let fly might open a breach in the opposing ranks; at the least it would expose weak points in the enemy formation that the advancing heavy infantry could exploit. Moreover, if the advance was unsuccessful, the lightly armed troops were available to cover the heavy infantry's retreat. An army that entered battle without peltasts was as good as defeated.

We have a detailed account of the fate of just such an army; it had been deprived overnight of almost all its lightly armed troops. This was the force of 10,000 Greek heavy infantrymen that was the backbone of a much larger army seeking to overthrow the king of Persia in 401 B.C. After the pretender to the throne who led them had fallen at the Battle of Cunaxa, the pretender's native forces fled and the Greeks were left alone. The Athenian Xenophon undertook to lead the 10,000 Greek infantrymen to safety, but on their first day's march they were so plagued by small numbers of enemy cavalrymen, archers and slingers that they could travel only some 25 stadia: less than three miles. That night Xenophon declared to his captains: "We need slingers ourselves at once, and horsemen also."

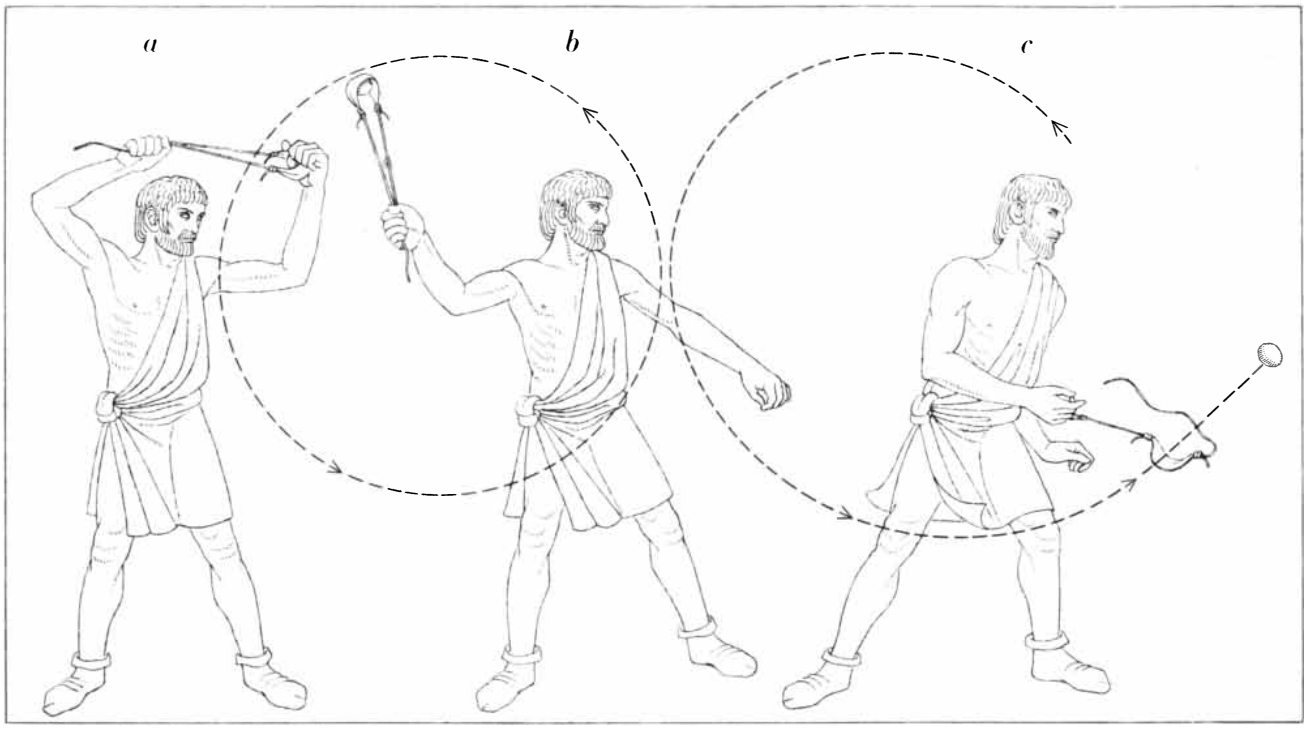
"Now, I am told," Xenophon continued, "that there are Rhodians in our army, that most of them understand the use of the sling, and that their missile carries no less than twice as far as those from the Persian slings. For the latter have only a short range because the stones that are used in them are as large as the hand can hold; the Rhodians, however, are versed also in the art of slinging leaden missiles."

STARTING POSITION, with sling poised above the head, is seen in this sculpture of a Neo-Hittite slinger from Tell Halaf in Syria, executed in the ninth or eighth century B.C.

Xenophon and his colleagues soon recruited from the ranks 200 slingers and a cavalry squadron of 50 men mounted on packhorses. The addition of these lightly armed forces to a body of 200 Cretan archers among the 10,000 enabled the Greeks thereafter to give a good account of themselves against the pursuing Persians. The Cretan archers' range was not the equal of the Persian archers', but the Rhodian slingers, Xenophon noted, "carried farther with their missiles than the Persians, farther even than the Persian bowmen." Considering that the archers of Persia were then regarded as the best in the world, his statement speaks well for the range of the Greek slingers.

Exactly how do the sling and the bow compare in range? The Roman military commentator Vegetius, writing in about A.D. 400, recommended target practice with the bow at a range of 180 meters. Even a modern sporting bow with a 45-pound pull can propel its arrow little more than 200 meters. With a long, light "flight arrow" and a bow with a 60-pound pull an archer trying for distance might achieve a range approaching 275 meters. By way of comparison I asked some young men in eastern Turkey to sling ordinary pebbles for me. In five out of 11 trials the pebbles struck beyond a mark placed 200 meters away, and the three best casts fell between 230 and 240 meters away. None of the young men appeared to be a skilled slinger; at least none had a sling in his possession at the time. Moreover, the missiles were pebbles selected at random rather than the carefully shaped stone, clay or lead missiles launched by slingers in Greek and Roman times. On the basis of Xenophon's comment alone it seems probable that a slinger casting lead missiles could attain a range in excess of 400 meters.

The sling I have been discussing is the flexible hand sling (*funda* in Latin). There is a second kind of sling: the staff sling (*fustibalus* in Latin). The hand sling can be no more elaborate than a strap three feet or so in length and an inch or



HAND SLINGER'S CAST begins (*a*) with the sling loaded and poised; the fixed end of the sling is looped around one finger of the slinging hand and the free end is held between thumb and forefinger. Three or four counterclockwise rotations of the sling (*b*), achieved primarily by the movement of wrist rather than arm, bring the missile up to maximum velocity. The missile is let go (*c*) when the slinger releases the free end of his sling; at the start of its parabolic flight it travels at more than 60 miles per hour.

so in width. At one end the strap is looped, knotted or tasseled, allowing it to be made fast to any of the four fingers of the slinger's throwing hand. The other end, which can be knotted to provide a grip, is held between the thumb and forefinger of the throwing hand. The thrower places the missile in a "pocket," sometimes artificially enlarged, at the end of the dangling loop. If the missile is made of stone or clay, it is usually the size of a small egg. A rotary motion of the wrist sets the sling whirling rapidly, either semihorizontally (around the slinger's head) or vertically (parallel to the body). After three or four revolutions the slinger releases the free end of the strap and the missile flies off on a tangent to the circle described by the sling.

The staff sling is inferior in range to the hand sling. At the same time it is easier to handle and can be employed to throw larger and heavier missiles. The sling, usually made of cord, has one end tied to a pole perhaps three feet long. The free end of the sling is temporarily attached to the end of the pole; either the end of the pole is notched so that the free end of the sling can be slipped into the notch or the free end of the sling is looped so that it can be slipped over the end of the pole. An enlarged pocket at the end of the dangling loop holds the missile. The pole, which at first is held parallel to the ground, is brought abruptly upright above the slinger's head; at the top of the swing the free end of the sling slips off and the missile flies free [*see illustration on opposite page*]. Used in both Greek and Roman times, the staff sling was a popular siege weapon through the Middle Ages. Even with the advent of gunpowder it served as a kind of grenade launcher well into the 17th century.

The longer the sling, within practical limitations, the greater the potential velocity of the missile. The natives of the Balearic Islands to the east of Spain were notable slingers. Indeed, Polybius, a Greek historian of the second century B.C., stated that the islands took their name from this fact, *ballein* being a Greek word meaning "to throw." However that may be, Balearic slingers served as lightly armed troops in many of the wars of classical times, perhaps most notably in the long conflict between Rome and Carthage. They always carried three lengths of sling: a long one for long range, a short one for short range and one of intermediate length for ranges in between.

Turning to the missiles themselves, it is clear that when they are nothing more than water-worn pebbles, the archaeologist cannot easily identify them as missiles. Only when a number of similar stones are found close together and show no evidence of being used for some other purpose (say hammering or rubbing), or best of all when they are of a kind that is foreign to the place where they are found, can such stones be considered potential evidence for the use of the sling at that place and at the time indicated. Fortunately, although vast numbers of naturally formed missiles may never be recognized, a good many ancient sling missiles were manufactured with great care. These artifacts are not always easy to perceive. Even when they have been recognized, they have often only led archaeologists to wonder what purpose such odd "clay eggs" may have served.

In the Near East the first manufactured sling missiles were spherical. They made their appearance shortly before the beginning of the sixth millennium B.C. Biconical projectiles were the next to appear [*see top illustration on page 40*]. A millennium later, around 4000 B.C., ovoid missiles also came into service. Three considerations, all concerned with the improvement of accuracy, seem to have worked together in the development and standardization of the missiles. The first objective was to make the weight of the missiles reasonably uniform, so that the slinger did not have to compensate for a different weight at every cast. The second objective was to provide a uniform, somewhat stream-

lined shape, in the interest not only of accuracy but also of velocity and distance. The third was to have the missile fit the pocket of the sling snugly, so that, as the Roman historian Livy remarked, "the missile may not fly out at random... but, seated firmly while being whirled, may be shot out as if from a bow-string."

In making these shaped sling missiles of stone the makers showed a natural preference for easily worked materials such as limestone. At a very early date, however, even earlier than the prepottery phase of the Neolithic, some men had recognized the advantages of clay as a material. Clay missiles are found in both prehistoric and historical sites all over the world. For example, clay missiles some 7,000 years old have been found at Hassuna, a site in Iraq, and similar missiles have been excavated elsewhere not just by the hundreds but by the thousands. It was not the absence of suitable stone for missiles that led to the use of clay; clay missiles are found at numerous sites where the supply of pebbles was plentiful.

The clay sling missiles are curious on two counts. First, in almost all instances they have been hardened by being dried in the sun rather than by baking. Second, they are surprisingly heavy for their size. The two facts are related. In order to attain maximum weight within rather narrow dimensional limits the makers of the clay missiles did not temper the material with chaff, as is done with pots and even bricks. The missiles were made of pure clay (or, very occasionally, of pebbles sheathed in clay) and are correspondingly dense. If these pure-clay missiles had been baked in a fire, the heat would have cracked them and made them worthless. This was why they were dried slowly in the sun.

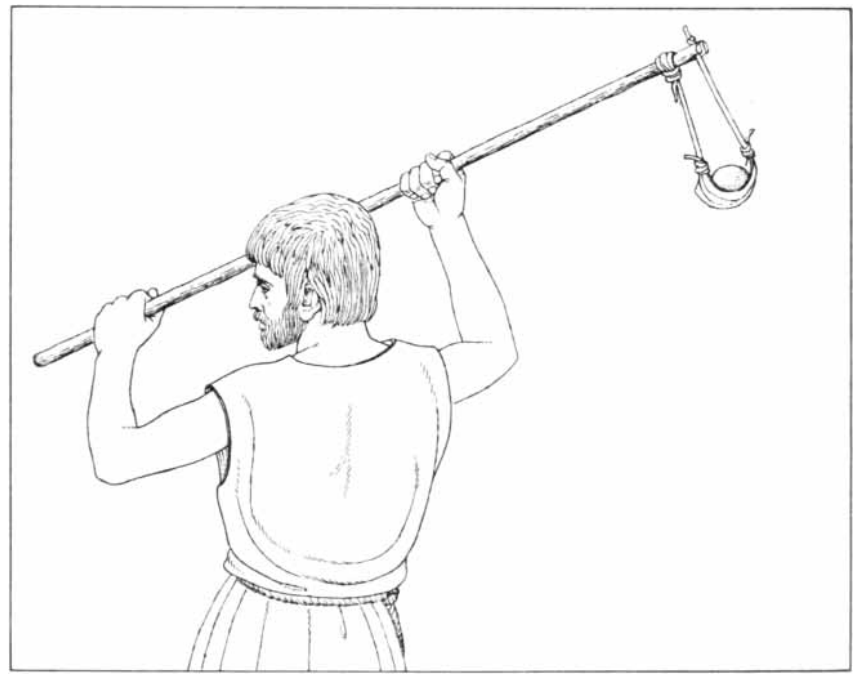
By classical Greek times, if not before, another kind of sling projectile had become common. The new missiles were made of lead. The Romans called them *glandes* because of their supposed resemblance to acorns. They were cast in molds and often carried inscriptions; the same few letters, scratched into the depressions of a mold, might serve to mark many hundreds of missiles. Often the inscriptions were routine: the name or number of the slinger's military formation, the name of the warring state or that of the commanding general. Not a few, however, are less formal. "Take this," reads one; "An Achaean blow," brags another; a third reads, "Your heart for Cerberus"; a fourth, "For Pompey's backside," and a fifth merely "Ouch."

As Xenophon's comparison of Rhodian and Persian sling missiles indicates, the standard missiles varied widely in size and weight. Measurement of a representative sample of biconical and ovoid stone missiles from sites in the Near East indicates how broad the range could be. The minimum weight was 13 grams; the maximum, 185 grams. With respect to volume the range was from about five cubic centimeters to about 65. (If the missiles had been perfectly round, the diameters corresponding to these extremes in volume would have been respectively about two centimeters and five centimeters.)

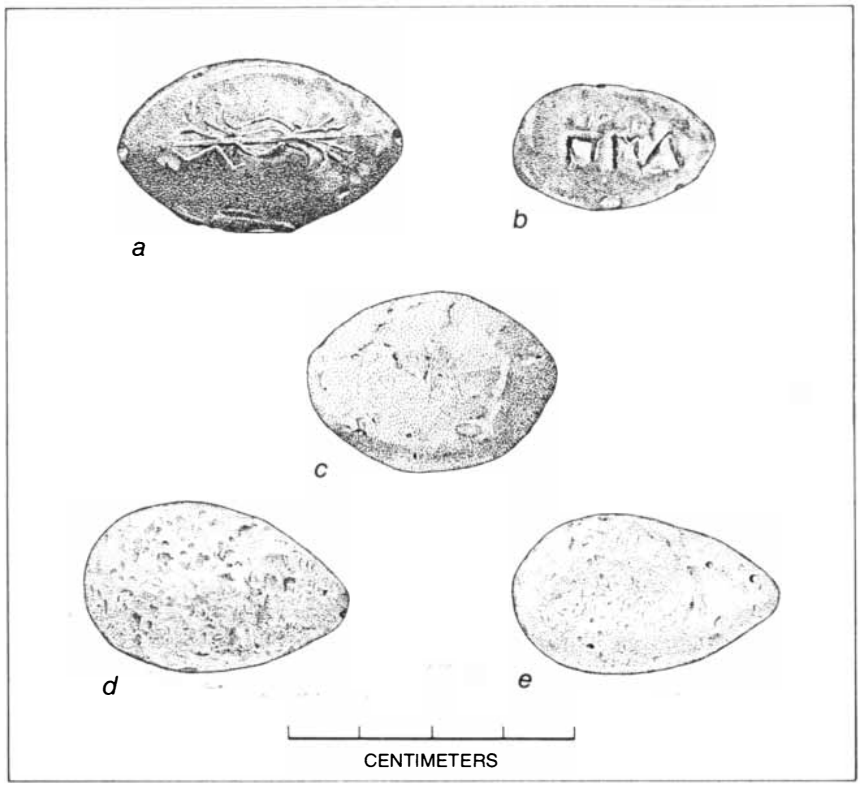
Taking sling missiles in general, whether they were made of stone, clay or lead, the range of weights is usually narrower than this. Few of the lighter missiles weigh less than 20 grams and few of the heavier ones more than 50. This was true, for example, in Roman times. In 1885 the German classical scholar K. Zangemeister published among other findings the weight of sling missiles from certain sites in Sicily and Italy. He found that the missiles with the lightest average weight (ranging from a minimum of 24 grams to a maximum of 46 grams) were from Sicily. The heaviest were from Asculum, a mainland site; their average weight was over 47 grams. Those from a second mainland site, Perusia, were intermediate in weight.

The missiles sometimes used by the Balearic slingers provide a notable exception to even the Near Eastern maximum of 185 grams. The Sicilian-born historian Diodorus Siculus, writing in the first century B.C., gave an account of the Battle of Eknomos, where Carthaginian forces, including 1,000 lightly armed Balearic slingers, defeated Agathocles of Syracuse. The slingers are given much of the credit for the victory. Diodorus stated that their stone missiles weighed one *mina* each. Now, the *mina* is variously calculated to be equivalent either to 330 grams or to 450 grams. If one conservatively chooses the lesser value (which roughly corresponds to the Roman pound and the Attic *mina*) and assumes that the Balearic missiles were made of limestone, each stone would have been 6.3 centimeters in diameter, or nearly the size of a tennis ball. That size and weight probably represent the outside limits for sling missiles made of stone.

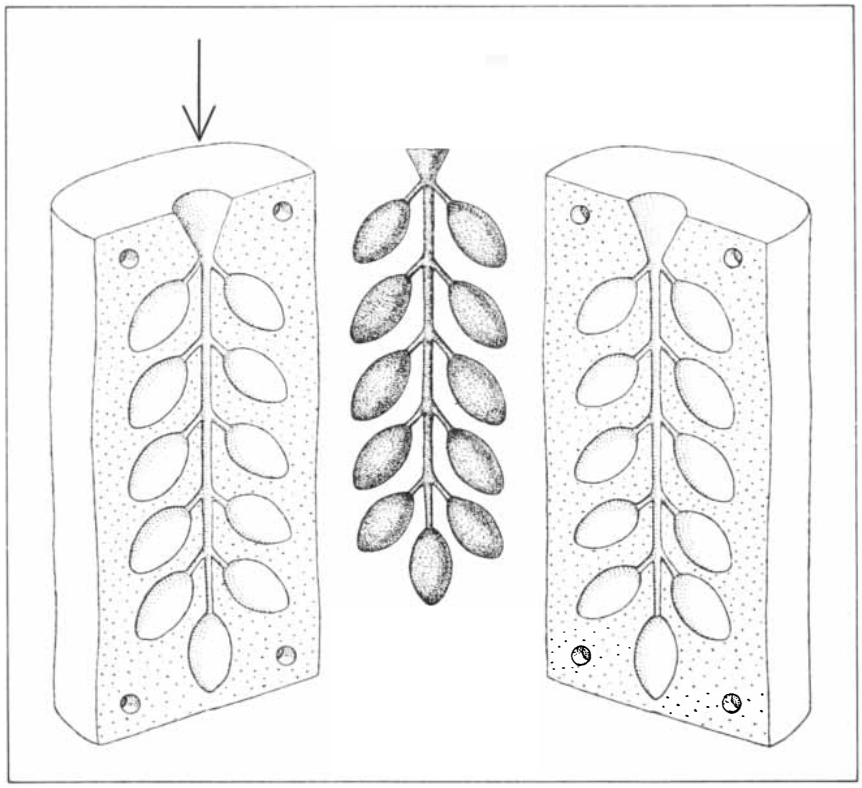
How accurate were the slingers of Greek and Roman times and how much damage could their missiles do? There is an abundance of documentary evidence on both points. Livy declared that Achaean slingers were the finest. He attributed this to the Achaeans' having practiced accuracy by slinging their missiles so that they would pass through a hole in a ring of "moderate circumfer-



STAFF SLINGER'S THROW is a quick snap from a horizontal to an upright position. The free end of the short sling is slipped into a notch at the end of the staff and will disengage at the top of the swing, releasing the missile. Staff slings do not have as great a range as hand slings but they will take heavier loads and were even used to throw grenades.



MAN-MADE MISSILES, as distinguished from pebbles, include large (*a*) and small (*b*) molded lead projectiles. The large one, of Greek or Roman origin, displays a thunderbolt design; the small one is among the hundreds excavated at Olynthus in Greece. The third missile (*c*) is a biconical type, made of sun-dried clay. The others (*d*, *e*) are ovoids of stone.



TERRA-COTTA MOLD of the "tree" variety was used to cast 11 lead missiles at a time. This reconstruction is based on a part of a mold that was discovered at Olynthus. Between the two halves of the porous mold is shown a tree casting before the detachment of its missiles.

ence" set up at a distance. As a consequence of this training, Livy wrote, the Achaeans "would wound not merely the heads of their enemies but any part of the face at which they might have aimed." One is immediately reminded of David's celebrated first cast.

The skill of the Balearic slingers was also attributed to special training. Diodorus wrote of them "that their mothers compel them, while still young boys, to use the sling continually; for there is set up before them as a target a piece of bread fastened to a stake, and the novice is not permitted to eat until he has hit the bread, whereupon he takes it from his mother with her permission and devours it." Precision in slinging is also reported of the Benjamites, the left-handed slingers mentioned in the Bible: they "could sling stones at a hair *breadth,* and not miss."

As for the effectiveness of the sling as a weapon, it is worth noting that the speed of a missile leaving a sling can easily exceed 100 kilometers per hour. If one assumes that a 25-gram missile had that velocity when it reached the target, the force of its impact would be equivalent to that of a golf ball falling from the top of a seven-story building. The energy of heavier missiles, of course, would be proportionately greater. Vegetius said that biconical sling missiles were more deadly than arrows against opponents clad in leather armor. Even if the missile did not penetrate the armor, Vegetius noted, it was capable of inflicting a fatal internal injury. If the opponent was unarmored, of course, the missile could easily penetrate the body. Celsus, perhaps the most judicious medical author of Greek and Roman times, included in his treatise *De Medicina* instructions for extracting lead and stone sling missiles from the bodies of wounded soldiers. His instructions followed by some centuries the observation by the Greek historian Thucydides that the slingers of a seacoast district in Epirus, the Acarnanians, so distressed invaders with a hail of missiles at long range that "it was not possible for [the invaders] to stir without armor."

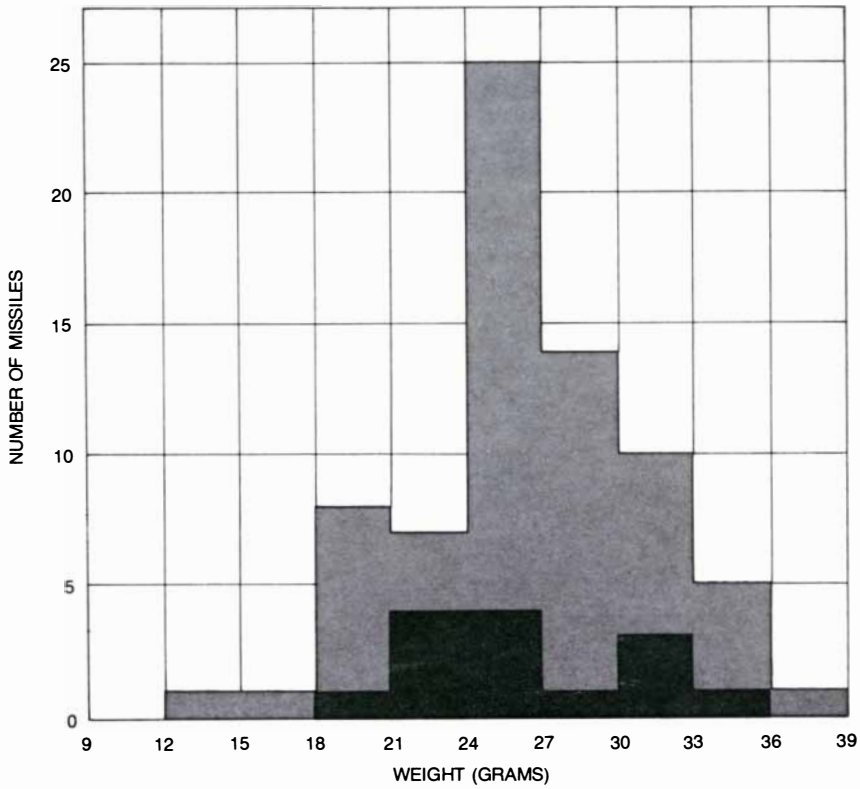
In more recent history we find the conquistadors' comments on the accuracy and effectiveness of Peruvian slingers. "Their chief weapon," wrote one Spanish observer, "is the sling. With it they throw a large stone with such force that it could kill a horse. Its effect is indeed only slightly less than that [of a Spanish firearm]; I have seen how a stone flung from a sling over a distance of 30 paces broke in two a sword that a man was holding in his hand."

In the 1930's some 500 lead sling missiles were uncovered by David M. Robinson in his excavation at Olynthus, an ancient city of northern Greece. More than 100 of the missiles bear inscriptions; some identify the missile as belonging either to the defenders of Olynthus or to the Macedonian troops who, under Philip, the father of Alexander the Great, captured the city in 348 B.C. Not all the inscriptions, however, provide such identifications.

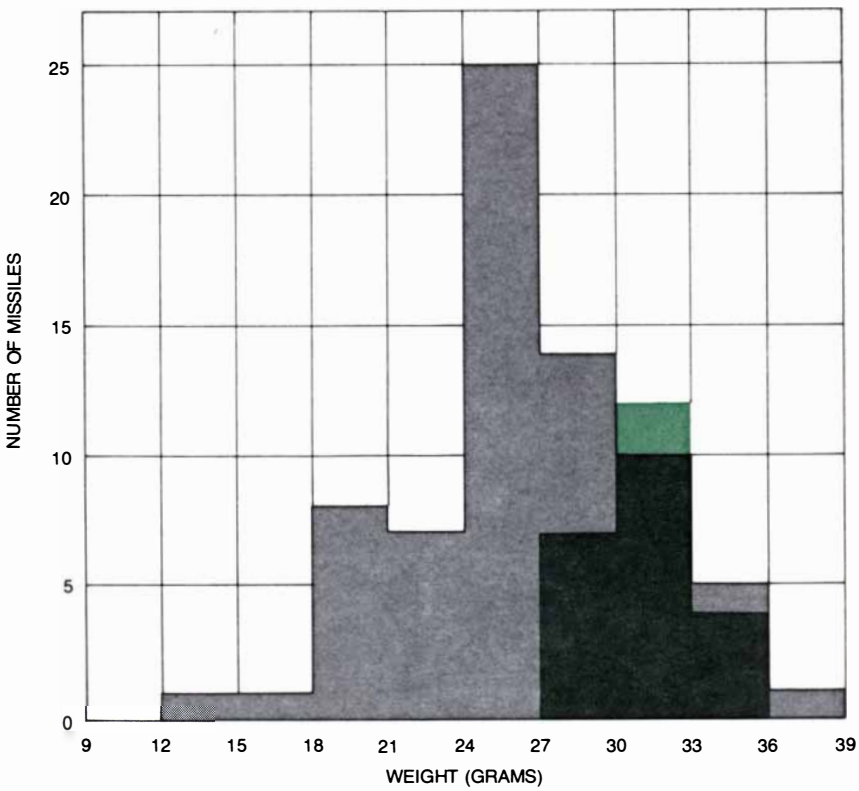
Robinson published detailed measurements of all the inscribed missiles. They ranged in weight from 18 to 35 grams. When the weights are compared with the identifying inscriptions, an interesting correlation emerges: most of the Macedonian missiles are at the heavy end of the range and most of the Olynthian ones run from light to medium [*see illustrations at right*]. This suggests that other missiles found at the site can be at least tentatively assigned to either the Macedonian or the Olynthian forces on the basis of weight. Moreover, on certain of the missiles that can be assigned to the Olynthians on the basis of their light weight there appear names: Potalos for one, and what may be either Timosthenes or Timostratos. Because it was often the custom to mark lead missiles with the names of the generals commanding the troops, it is quite possible that the two men named on the Olynthian missiles were leaders of the defending forces who are otherwise unknown to history. Similar studies of sling projectiles from other classical sites might yield equally unexpected historical information.

Slings continued in military service into the 17th century, but even by A.D. 400 the increasing employment of armor and fast cavalry was making the slinger obsolete. Vegetius recommended that slingers be trained to release their missile after a single swing rather than the customary three; the objective was obviously to increase the slingers' rate of fire. It was mainly the staff sling that survived the introduction of gunpowder and small arms, but I can vouch for the use of the hand sling as late as 1936. In that year, during the Spanish Loyalists' siege of the Alcázar, the fortress where the rebel garrison of Toledo had taken refuge, the besiegers lobbed grenades into the fortress with slings. A motion picture still exists showing one such slinger in action.

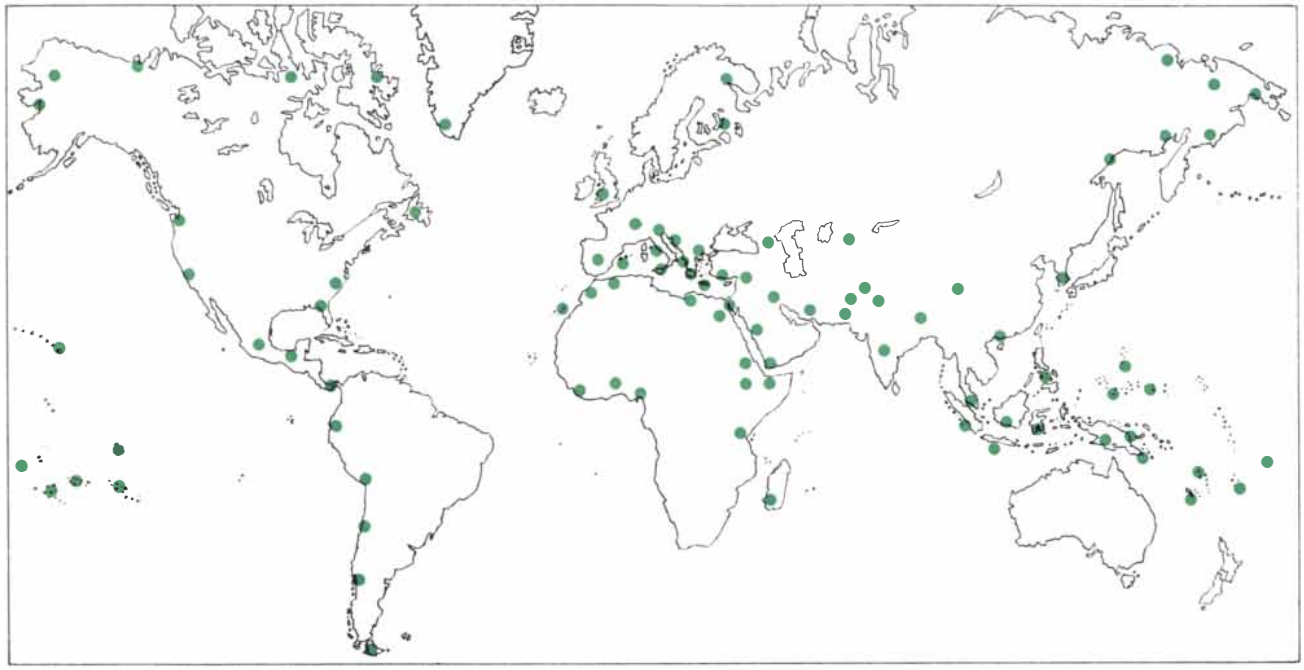
The British prehistorian V. Gordon Childe more than once toward the end of his life attempted to bring the significance of the sling as a weapon to the



DEFENDERS' MISSILES at the siege of Olynthus in 348 B.C. ranged from 19.5 to 33.4 grams in weight (*colored bars*). Nine of the 14, however, weighed less than 27 grams. When their weights are compared with the weight range of other sling missiles excavated at Olynthus (*gray bars*), most of the defenders' missiles cluster on the medium-to-light side of the scale.



ATTACKERS' MISSILES at the siege weighed more than the defenders'. Of the 23 identifiable as Macedonian, 16 weighed between 30 and 35.8 grams. Compared with the weight range of the other missiles (*gray bars*), the Macedonian missiles are on the heavy side.



WORLDWIDE DISTRIBUTION of the sling from the beginnings of history up to recent times (*colored dots*) suggests that there are few major areas of the globe where the weapon is unknown. The early importance of the sling in the Near East and in Europe adds weight to the argument that the art of slinging spread from this nuclear area. Unless the sling was independently invented in the New World, its presence there is also an argument for some kind of connection with the Old World via the Pacific or the subpolar zone.

attention of his colleagues. He met with little success, but I, at least, found his conjectures compelling. Indeed, taking Childe's work as my starting point, I have recently published a hypothetical reconstruction of the relative importance in prehistoric times of the sling and the bow in the Near East. I defined the area of my study as being bounded by the Bosporus on the west, the Indus on the east, the Caucasus on the north and Sinai on the south. In this large area, my evidence suggests, the two weapons were mutually exclusive for a period of several millenniums; that is to say, the peoples who used the one saw no good reason to take up the other.

This polarity of the sling and the bow is first discernible in the eighth millennium B.C.; it continues until the fourth millennium and even later in some parts of Asia. For example, the bow was used almost exclusively in Syria and Palestine before the rise of the city-states in those areas, but the inhabitants of other parts of the Near East preferred the sling. The bow was evidently unknown in the region until almost the end of the eighth millennium, but knowledge of the sling is some thousands of years older. One area of overlap between the two long-range weapons provides an exception to this polarity. In Asia Minor, at Çatal Hüyük and elsewhere about 6000 B.C., both the sling and the bow were used.

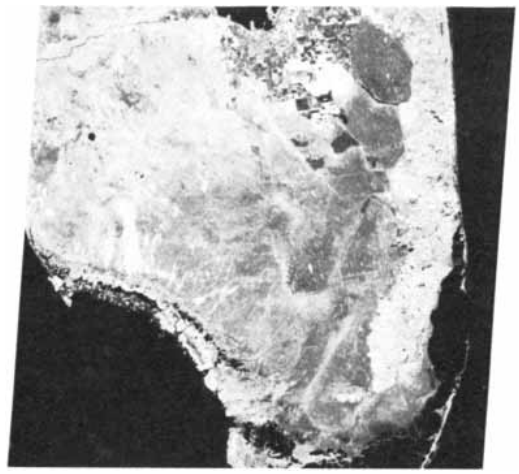
Support for my hypothesis consists of finds at more than 80 sites in the region, all of them fairly reliably dated, that show the presence of one or the other of the two weapons. The evidence is not confined to the Near East. The presence of the bow in Syria and Palestine represents a kind of Asian beachhead for this weapon; the staging area for the landing, so to speak, was clearly Africa. There the preference for shafted projectiles is easy to perceive in the Aterian projectile points of the Upper Paleolithic period, in the rock paintings of archers that are found in Africa and as far away as Spain, and in the thousands of small projectile points that have been found at sites throughout the Sahara. In all probability the Arabian peninsula can be added to this African bowmen's realm, although that region is still virtually terra incognita for the archaeologist. By the same token, at least by the beginning of Neolithic times the realm of the slinger in southwestern Asia had expanded to include not only the Balkans but also southeastern Europe in general.

The polarity in long-range weapons cannot be attributed to a lack of communication between the two realms; the peoples of both areas had ample contact with one another. For example, during the prepottery phase of the Neolithic in the eastern Mediterranean obsidian was regularly transported from parts of Asia Minor where the sling was used to as far south as Beida in southern Jordan, where the bow was supreme. Some explanation other than isolation must be sought. When the explanation is found, it may transcend the weapons themselves and thus lie outside the realm of archaeology proper, with its focus on material culture.

Perhaps future investigation of the separate realms of the sling and the bow will bring about a revival of "Kulturkreise" theory, which envisions the rise and expansion of "culture circles" in prehistoric times. In the present context a Kulturkreise hypothesis would envision one culture circle that expanded from Africa across Spain and western Europe, while another culture expanded from southwestern Asia across the Balkans and southern and eastern Europe. Of course, the bow and the sling would on this hypothesis merely be material-culture indicators of a far more complex constellation of social phenomena.

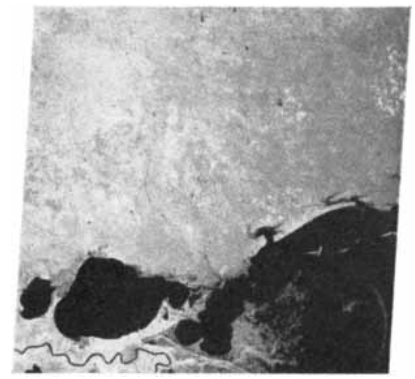
However that may be, it is clear that archaeological work in the future must pay more attention to the sling as a prehistoric weapon of major significance, not only in the Near East but also elsewhere in the world. Even with our present limited knowledge it is clear that the slinger and the archer were equals for thousands of years.

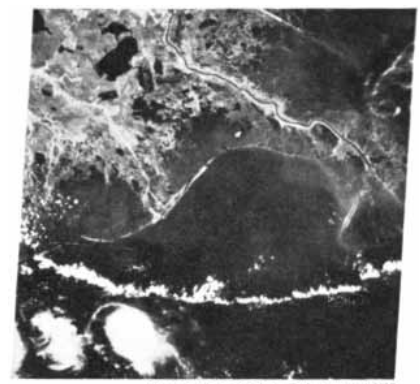
Taxpayers, you have invested in the space program. Now benefit from your investment.



Miami and environs, 22 March 1973

New Orleans and environs, 7 August 1972





Mississippi delta, 7 August 1972

The U.S. Geological Survey, which runs the EROS program (Earth Resources Observation Systems), points out an opportunity to serve the public interest by calling to your attention that imagery is now available of just about any region on earth, except within some 900 miles of the poles.

Everywhere else, NASA's ERTS (Earth Resources Technology Satellite) goes over every 18 days 567 miles high, looking down in four wavelength bands: 0.5-0.6 μm, 0.6-0.7 μm, 0.7-0.8 μm, and 0.8-1.1 μm. Anybody–grade school teacher, ranch broker, botanist, Cub Scout–with $1.75 to spend can have a 9″ x 9″ print of any place selected. The 9″ x 9″ will encompass 115 miles x 115 miles. Changes may be apparent from the previous pass.

For the sake of preserving the earth's glory, let's learn how to look through mankind's new microscope-in-reverse. Let's assume that if Aristotle and his friends had had even ordinary microscopes, we'd all be happier today.

San Francisco and environs, 4 April 1973

Details:

A quick way to get a look through that great new reverse-microscope in the sky is to phone 605-594-6511 between 7 a.m. and 7 p.m. Central Time. Tell what geographical area interests you and what you have in mind. You will be talking to a helpful, well-informed, and quite real person. His or her hand that isn't holding the telephone will be in contact with heavy computer power.

The computer will even know how much was under cloud each time ERTS passed over the region of interest. It might even report that the area was photographed in 1936 and that, if you'd rather have that black-and-white shot from 5,000 feet instead of infrared "false color" from 567 miles, you can so order. Or you might be told the same place was photographed by a Gemini astronaut with a hand-held camera or by completely automatic equipment from Skylab this year, and how would you like *that?*

The EROS price list reflects just the cost of reproducing the photographs. (The cost of the satellite is more widely shared.) A 40″ x 40″ color enlargement on paper, for example, is priced at $25. To avoid misunderstanding, you can take down the numbers of possibly interesting pictures and check them out on a microfilm browse reader at any of 19 locations in the U.S.A. before placing your order. Black-and-white copy negatives or color internegatives can be ordered if you wish to do your own printing. No copyright worries.

If you want to experiment with the color rendition of the bands to emphasize certain signatures, come in person any business day from 7:45 a.m. to 4:30 p.m. to EROS Data Center, 14 miles northeast of Sioux Falls, S.D. Scientific staff members are available there for consultation on instrumentation and interpretation problems.

Don't expect much, though, by way of direct interpretation. *That* the customer does, or consultants the customer may engage, assisted by a fast-growing literature. Aerial photography and the interpretation thereof are not leaving the private sector. On the contrary, those in the business welcome EROS as a tantalizer to swarms of new prospects.

EROS Data Center can also be queried by mail (ZIP 57198).

An article in *American Scientist 61*:175 can take you farther into this fascinating subject. Address reprint requests to Eastman Kodak Company, Dept. 55W, Rochester, N.Y. 14650. But get in touch with EROS, not Kodak, if you want a picture of the old home town from 567 miles up.

You have heard of self-fulfilling prophecy? Its converse, self-denying promise, may threaten if this advertising proves too successful. There must be a limit to how many inquiries a day can be handled in the manner promised here.

SOLID FORD

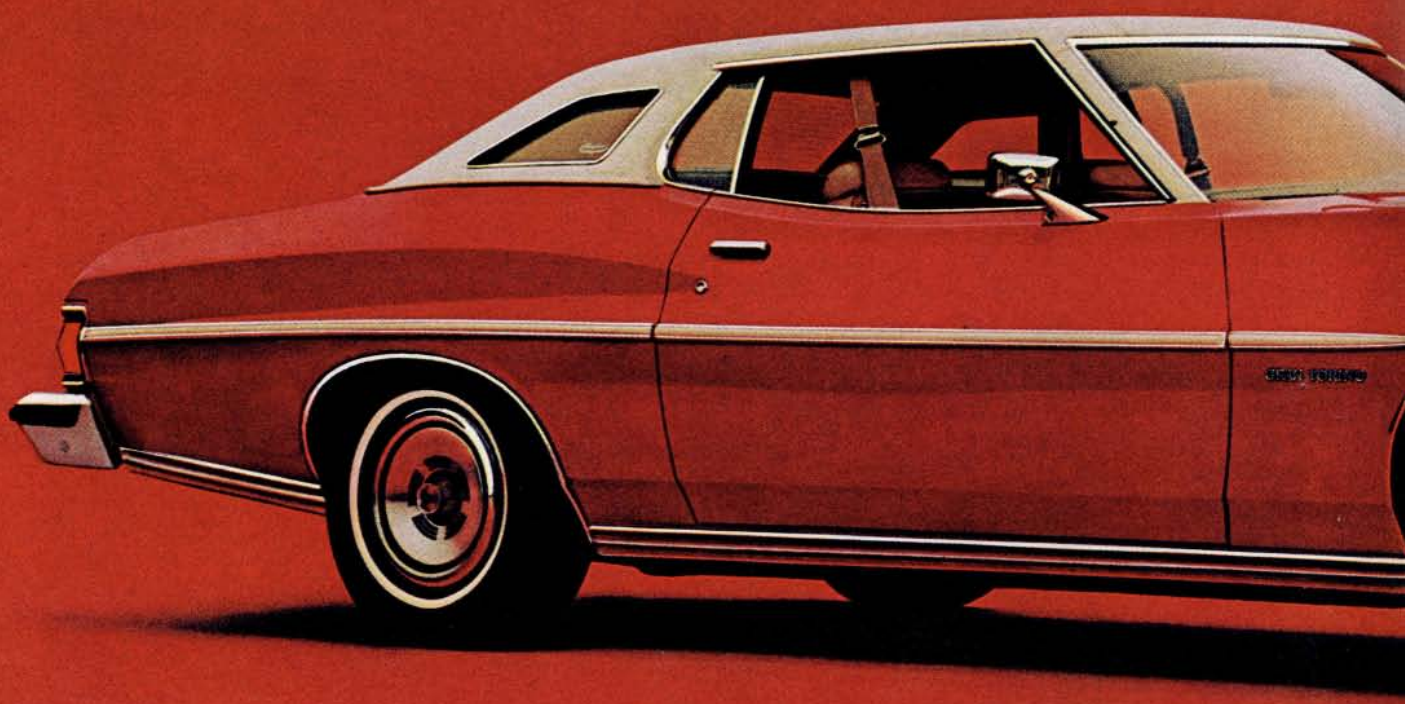
With a new young look and a smooth and steady ride.

There's more excitement than ever in the solid mid-size.

Outside, the Gran Torino's been restyled, with a gleaming new front end, a new rear end design and even an optional opera window. (Standard on Gran Torino Brougham.)

Inside, the Gran Torino Brougham's split bench seats (pictured at right) add a touch of luxury and convenience.

And while you can love Torino for its looks, you can trust it because it's solid. Torino's sus-



1974 Gran Torino Brougham shown with optional deluxe bumper group, electric rear window defroster and convenience group.

TORINO'74

pension, wide track and long wheelbase help give it a smooth and steady ride that spells quiet comfort and confidence on the road. As do Torino's optional steel-belted radial ply tires and power front disc brakes.

In fact, safety and security are part and parcel of Torino's design. And this year, Torino's Lifeguard-Design Safety Features include a new style interlock safety belt system that lets you really move in your seat with ease and comfort.

See the complete Torino line at your Ford Dealers.

The closer you look, the better we look.

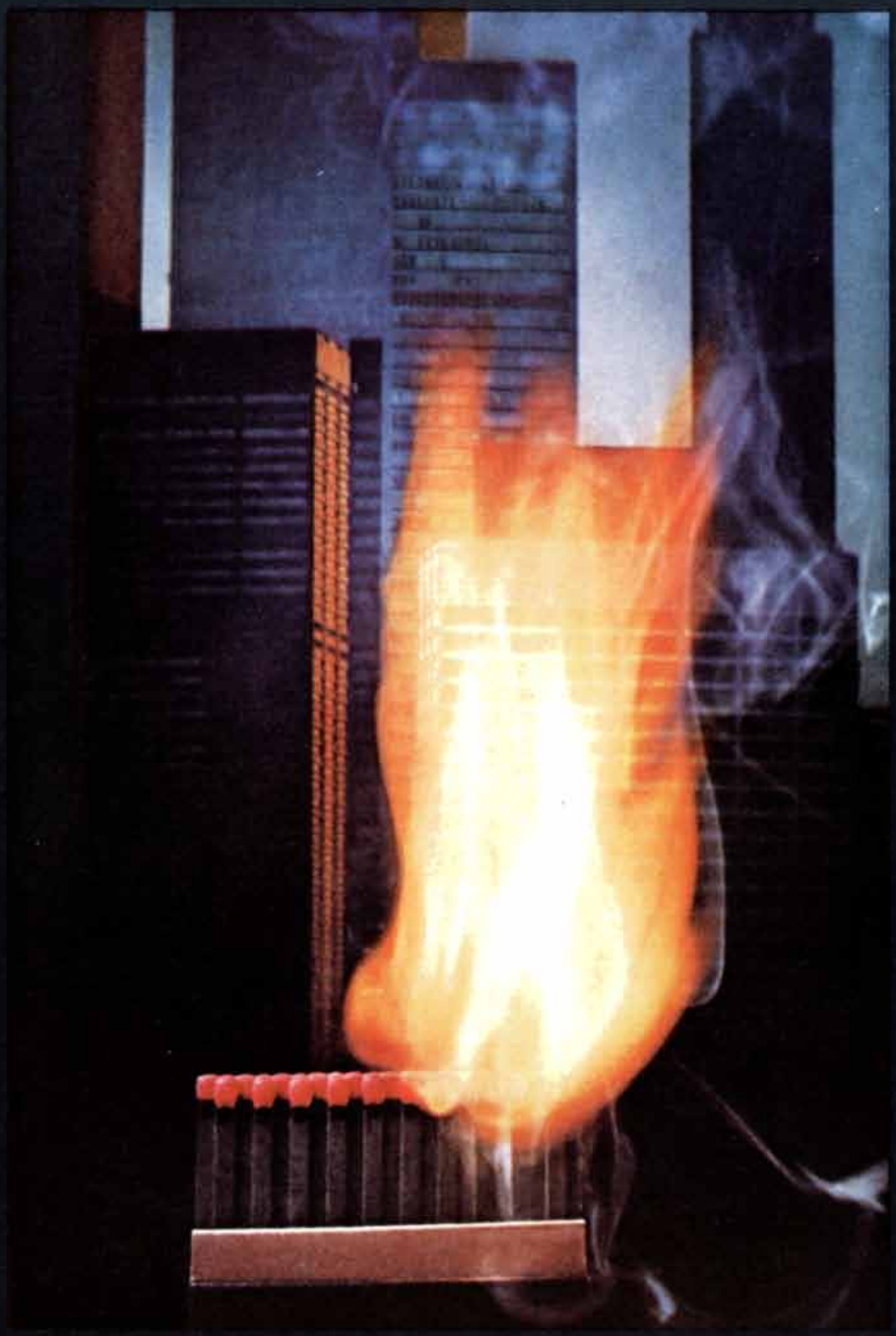
New optional fender skirts, power operated sun roof.

The solid mid-size.

FORD TORINO

FORD DIVISION Ford

Did you know there's an electronic system that can save lives in a burning building?



When fire breaks out in a modern office tower you need new ways to fight it. RCA has developed an electronic way. It's an important part of our new concept—Building Management System—to provide computerized electronic control for all major building functions.

In case of fire, BMS sensors quickly detect smoke and flames and report the earliest signs to a computer console.

Then, BMS automatically sounds fire alarms and regulates elevator service so firefighters can swiftly reach the blaze. Most important, BMS assists in getting people out of the building and controls air circulation both in the fire area and on adjacent floors—all with electronic accuracy.

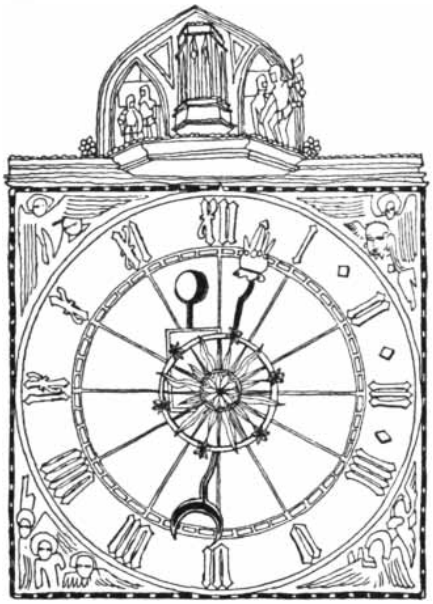
BMS also performs other duties electronically, such as monitoring elevator performance, adjusting air conditioning and heating to changes in the weather, and guarding against break-ins and theft. And, in doing all these things, it saves energy.

Electronics is creating ingenious new ways to enhance life. And RCA, which helped create the technology itself, is still pioneering the electronic way.

The electronic way

RCA

SCIENCE AND THE CITIZEN



The Other Bombs

The arms-limitation agreements between the U.S. and the U.S.S.R. deal with strategic nuclear weapons but not with tactical ones. The present extent of the U.S. tactical nuclear arsenal was recently described in a hearing of the Subcommittee on Military Applications of the Joint Committee on Atomic Energy. The chairman of the subcommittee, Senator Stuart Symington of Missouri, said: "The United States has now deployed in foreign countries, on our ships, and in this country, some tens of thousands of nuclear weapons with a total maximum possible yield of over several billions of tons of TNT—more than several hundred thousand times the yield of the bombs dropped on Hiroshima.... One cannot help but consider the implications incident to our defense and foreign policies if these facts were known by the appropriate committees of the Congress, as well as in more general fashion by the American people."

According to testimony at the hearing, U.S. tactical nuclear weapons are of four kinds. Three are for offensive or defensive use against troops, transport or other "targets of opportunity." The fourth is defensive and would create an obstacle in the path of an enemy advance.

Rockets with nuclear warheads are one of the three kinds of offensive-defensive weapons. Four configurations, two of them outdated, are currently available. The "Honest John" is an obsolete unguided rocket with a range of up to 40 kilometers. The "Sergeant" is an obsolescent, rack-mounted rocket that can carry either one of two warheads with different yields. The "Lance" is an up-to-date weapon with a range of up to 125 kilometers; it is mounted on a tracked vehicle and is accompanied by a tracked auxiliary transporter-loader. "Lance" rockets can carry any one of three warheads. The "Pershing" is a heavy rocket that can be transported by road; its range is up to 725 kilometers, and it too can carry any one of three warheads. The "several thousand" Honest Johns and Sergeants now ready for use will be phased out of Army service by 1980, but some Honest Johns will remain in the hands of U.S. allies after that date.

Artillery shells with a nuclear charge are the second class of tactical weapons. They are for use with the 684 six-inch (155-millimeter) guns and the 326 eight-inch guns currently in service with U.S. and NATO forces. Two kinds of shell are now stockpiled, and replacements for both are under active development. One is an eight-inch shell loaded with a uranium charge; it has been in stock for some 20 years. The other is a six-inch shell loaded with a "small" uranium charge that has been in stock for about 10 years. The range of the present six-inch shell is only 14 kilometers; its lighter successor will have a range of 25 kilometers.

Both the new six-inch shell, designated W-74, and the new eight-inch shell, the W-75, will have a plutonium charge rather than a uranium one. Their relatively small diameter requires that the fissionable material be used at less than the maximum possible explosive yield. Testimony before the subcommittee indicated that the plutonium shells to be supplied to U.S. and NATO forces, while "relatively few" in number, will cost some $900 million.

The third class of tactical weapons consists of aerial bombs. Five types are currently available. One, the B-28, is a low-yield bomb that is scheduled to be phased out. The testimony gave no details on two other bombs, the B-43 and B-57. The B-61, described as a "high-speed delivery" bomb, can carry warheads of different yields; its "high yield" warhead is equivalent in explosive power to "several hundred" kilotons of TNT. The W-72, or "Walleye," is a 1,125-pound glide bomb that is guided to its target by a television system. All five tactical bombs can be set to explode in the air or on the ground.

The defensive nuclear weapon is an ADM, or "atomic demolition munition." Two models, a "small" one and a "medium yield" one, are available. The first can be carried by one man; the second requires vehicle transport. The ADM is intended for the purpose of making huge craters. The tactical application of the weapon is to emplace the charges and then, when necessary, detonate them to block mountain passes and other restricted communication routes. The anticipated result would be either to deny the enemy the use of the route or to "channel" his forces until they formed assemblages large enough to be worth attacking with nuclear weapons. Holes for ADM's have been dug in some parts of Europe and the weapons have been placed in some of them.

The possibility of developing a fifth class of tactical weapons is under exploration at the laboratory level. These would be "neutron emitters," devices that could "deliver a neutron dose" strong enough to make individuals "very ill and... incapacitated... in a very short time." The doses would prove lethal within "a day or so." Speaking of this possible development, Harold Agnew, director of the Los Alamos Scientific Laboratory, noted that a "small but very elite group" of his colleagues is "working very aggressively, trying to influence the [Department of Defense] to consider using these... weapons...."

Details of the total number of tactical nuclear weapons and their exact explosive power were deleted from the testimony before the transcript was made public. A rough measure of their potential destructiveness is provided, however, by one witness's statement. He said that one of the aerial bombs, a comparatively low-yield weapon, had an explosive force equal to "a few hundred kilotons" of TNT. The explosive force of the uranium bomb exploded over Hiroshima was 14 kilotons.

The Craters of Venus

Venus—like the moon, Mars and to some extent the earth—is pocked with craters. This fact emerges from radar-astronomical studies, in which short radio waves penetrated the optically

opaque atmosphere of the planet, were reflected from the solid surface and were analyzed by computer on the earth to produce a relief map.

The map was made by a group at the Jet Propulsion Laboratory led by Richard M. Goldstein. It shows an area 910 miles across (about the size of Alaska) near the equator of Venus. In this area there are a dozen large craters ranging from 21 miles in diameter to 100 miles. Smaller craters may well be present but were not resolved.

To generate the map the JPL group employed two radio telescopes at the Goldstone Tracking Station in the Mojave Desert: a 210-foot dish and an 85-foot one. On June 20, 1972, the 210-foot dish beamed a series of radar signals at Venus. At that time Venus was 30 million miles from the earth. The return echoes were received by both telescopes, which are separated by 14 miles. This technique in effect provided stereoscopic reception and enabled the JPL group to resolve features as small as about six miles. The vertical resolution was some 650 feet. The information was fed into computers that constructed the map with a special program devised by Howard C. Rumsey, Jr.

The area studied is basically flat, varying no more than 3,300 feet in altitude. All the craters are very shallow; the one that is 100 miles across is only a quarter of a mile deep. The fact that the craters are so shallow is a clue to Venus' history. The craters could be the result of meteorite impacts before the planet had an atmosphere. Possibly they were subsequently filled with lava from the interior of the planet. Alternatively erosion on the surface could have filled the craters. "The shallower the craters are, the more dynamic the crust of a planet is," stated Harold Masursky of the U.S. Geological Survey. "Several of us are convinced that the surface of Venus will help us understand the earth's crust and the dynamics that lead to continental drift."

The JPL team plans to undertake more radar probes of Venus at about the time that the *Mariner* Venus-Mercury spacecraft (to be launched in November) flies by Venus in February, 1974. The radar information on the surface might then be correlated with features of the planet's atmosphere.

Nicholas Copernicus, M.D.

Nicholas Copernicus' fame as an astronomer has obscured the fact that he was also a physician. This little-known aspect of Copernicus' career is described by Marian Śliwiński, Minister of Health and Social Welfare of Poland, in *World Health*, a publication of the World Health Organization.

In 1491 Copernicus entered the Jagiellonian University of Cracow, where he studied philosophy, astronomy, mathematics and astrology. In the Middle Ages astrology, with its predictions about man and his fate, was closely related to medicine, and it aroused Copernicus' interest in the latter field. In 1501 he took up medical studies at the University of Padua. At that time such studies took three years to complete, two years for the bachelor's degree and one more for the doctorate. After that a training period of a year under an eminent physician was required.

Copernicus finished his medical studies at Padua in 1505 and probably obtained the degree of licentiate, which would have allowed him to practice medicine. It has never been possible to establish whether or not he received his doctorate in medicine, because the medical archives of the University of Padua for the period from 1503 through 1507 were destroyed. He returned to Poland in 1507 and became physician to the priestly chapter of Warmia. Strict conditions governed the practice of medicine under such circumstances, and it seems unlikely that he would have been allowed to practice if he had not received his doctorate.

Thereafter Copernicus devoted most of his time to astronomy, but he never gave up medicine and was always available to those who needed medical advice or treatment. He bought many medical books both for himself and for the library of the bishop of Warmia. Even as he was writing his great astronomical work *De revolutionibus orbium coelestium* he pursued a major medical project: an effort to develop mathematical models for medical research. Not enough historical evidence survives to make it possible to assess this work. Nonetheless, Śliwiński observes: "An astronomer of genius, justly famed for his heliocentric theory of the universe, Copernicus was also a physician ahead of his time. For this, too, he deserves to be remembered."

Experimental Cosmology

A remarkably simple experiment conducted by a physicist at Haverford College may help to resolve the important cosmological question of whether we live in an ever expanding universe or in an alternately expanding and contracting one. R. B. Partridge describes his empirical approach to that question and discusses some of the implications

of his finding in *Nature*. He bases his experiment on an assumption implicit in one of the leading cosmological theories: the Wheeler-Feynman absorber theory, put forward in 1945 by John A. Wheeler and Richard P. Feynman. According to that theory radiation along a "future light cone" (that is, radiation emitted in a diverging beam into free space) is completely absorbed. In contrast, all "open," ever expanding cosmological models assume less than complete absorption along a future light cone. If the universe actually conformed to one of the open models, among the effects one would expect would be a reduced power drain on a source radiating into an incomplete absorber. Partridge's approach consists in a sensitive experimental search for just that effect.

Before the search could even begin the following question had to be answered: Can one transmit an electromagnetic signal that, if it is absorbed at all, will be absorbed only at very great distances (or, equivalently, in the very distant future)? The answer, according to Partridge, seems to be yes. He chose to work with microwaves because the absorbing properties of "local" matter at microwave frequencies are known to be small. Taking into account absorption in the earth's atmosphere, in our galaxy and in intergalactic space, Partridge calculated that at least 95 percent of the microwave radiation leaving an antenna at a wavelength of three centimeters "will reach cosmological distances before being absorbed (if it is ever absorbed)."

The apparatus actually used for this purpose consisted of a diode microwave oscillator, the output of which was alternately switched between a local, or "here and now," absorber, which absorbs virtually all the radiation directed at it, and a large conical antenna, which was pointed at the zenith. If the absorption caused by local matter had differed from the absorption caused by very distant matter, the power inputs to the diode would have changed at the switching frequency.

Five separate sets of observations were made, each including many runs with the antenna covered and many with the antenna uncovered. Clear, cloudless days and nights in August and September were chosen; runs were not made when the central plane of the galaxy was near the zenith. The object was to discover "whether the power input to the diode changed when it was allowed to radiate into free space" (that is, with the antenna uncovered). Such an effect, "if significant, would have implied that it 'costs' more power to radiate into

free space than into a complete, local absorber."

No such effect was observed. Hence, Partridge deduces, "it seems either that the analysis in terms of the Wheeler-Feynman absorber theory is incorrect or that absorption along the future light cone is complete to better than one part in 10^8." This finding, he points out, "sets some constraints on possible future properties of the universe." It suggests that the universe is not ever-expanding but rather is "time-symmetric" (that is, alternately expanding and contracting).

Class of 1958

Fifteen years ago British medical statisticians undertook an ambitious project: they gathered data on virtually every child born in Britain over the period of a week and then periodically resurveyed the group. The effort was unusual in that most population surveys examine a representative sample of the population at a given time; such studies are sometimes called cross-sectional. The 1958 survey examined the population as a whole and over a long period; such studies are called longitudinal.

The longitudinal survey is described in a recent report by Ronald Davie, director of research of the National Children's Bureau in Britain. It began as the 1958 Perinatal Mortality Survey, focusing on the causes of stillbirths and deaths close to the time of birth. Some 17,000 infants, representing an estimated 98 percent of all the children born in England, Scotland and Wales during the week of March 3, were covered. In 1964 the project was transformed into the National Child Development Study (1958 Cohort), conducted by the National Children's Bureau with government support. The children were followed up at age seven and again at 11.

"Perhaps the most striking finding," Davie said, "was the extent of disadvantage in many directions suffered by particular groups of children. Nearly half of the children whose fathers had unskilled occupations were poor readers, compared with less than 10 percent of the children from professional families. . . . However, the dimensions of the problem reached out beyond educational attainment. Marked differences between children from families in different occupational groups were evident in the prevalence of speech difficulties, delayed sphincter-control, lower stature, poor physical coordination, a low level of immunization and the underuse of other medical and welfare services." The study has proved so informative for educators, health authorities and economic planners that the British government has appropriated £250,000 for a third follow-up of the children in the academic year 1973–1974.

Epidemic Addiction

Epidemics affect large portions of a population; they begin suddenly and spread quickly, and their duration can be limited both by natural factors and by measures of control. The word is usually applied to diseases spread by microorganisms, but Robert L. DuPont and Mark H. Greene have discerned the characteristics of an epidemic in heroin addiction.

Studying drug abuse in the District of Columbia, DuPont and Greene found that addiction is indeed spread from person to person, although the vector of contagion is not as easily isolated and identified as it is in microbial infections. They also found that temporal trends in addiction conform to the pattern associated with epidemics: a large and sudden increase in the number of cases, followed by an equally sudden decline. The results of their study are reported in *Science*.

Several indicators of the incidence and prevalence of the use of heroin were measured. These included statistics compiled by Washington's Narcotics Treatment Administration (NTA); questionnaires distributed to those addicts who reported for treatment; urine tests for the presence of opiates, conducted at the District of Columbia Jail; the number of deaths per month caused by opiate use or overdose, and police reports of arrests of addicts and seizures of drugs.

The incidence of addiction (the number of new addicts per year) was relatively low until about 1966; it then grew rapidly, reaching a peak in 1969. More than 20 percent of the 13,000 addicts being treated by the NTA reported that they first used heroin in 1969. By 1972 the incidence of addiction had returned to the rather low level of the 1950's. "The shape of the incidence curve suggests contact as the mode of transmission," the authors report. "This is consistent with previous data indicating that heroin addiction spreads by person-to-person contact."

On the other hand, the prevalence of addiction (the number of heroin users) did not reach a maximum until 1971. The decline since that year has been substantial. DuPont and Greene attribute the control of the "epidemic" to two actions by the municipal government: the establishment of the NTA, which

made treatment readily available, and sterner law enforcement, which made heroin scarce. They acknowledge the possible contribution of a third factor: a change in the attitude of the community toward opiate use. "Heroin is now 'out,' " they conclude. "No longer is the pusher seen as a glamorous individual, a fabulously successful businessman. He has become a parasite in the community. ... As the appalling consequences of heroin addiction have become apparent, previously susceptible teen-agers are no longer willing to take the risk of experimenting with heroin."

The Earth: Zoo or Petri Dish?

It seems likely that many, perhaps even most, stars have planets, and that many of the planets have surface conditions suitable for the maintenance of life. If this is so, the probability is high that man is not one of the most advanced organisms in the universe. Many of the planetary systems would be older than the sun's, and therefore might be expected to harbor civilizations millions of years ahead of our own.

If this argument is sound, it presents an enigma. If "they" are out there, why have we not heard from them? One possible explanation is proposed by John A. Ball in *Icarus: International Journal of Solar System Studies.* Ball suggests that extraterrestrial civilizations could communicate with us but that they choose not to.

"Occasionally," he observes, "we set aside wilderness areas, wildlife sanctuaries, or zoos in which other species (or other civilizations) are allowed to develop naturally.... The perfect zoo (or wilderness area or sanctuary) would be one in which the fauna inside do not interact with, and are unaware of, their zookeepers."

An extraterrestrial species, he speculates, might consider man an organism to be preserved in isolation and purity, a part of nature to be observed, as it were, through a one-way mirror. "I believe that the only way that we can understand the apparent non-interaction between 'them' and us is to hypothesize that they are deliberately avoiding interaction and that they have set aside the area in which we live as a zoo."

F. H. C. Crick and L. E. Orgel, also writing in *Icarus,* offer an alternative theory: that life on the earth began as a deliberate "infection" of microorganisms placed here by another civilization. Such a civilization, it is clear, would have regarded the earth not as a zoo but as a Petri dish.

Protein Shape and Biological Control

The processes of life are turned on and off by means of a universal control mechanism that depends on the ability of protein molecules to bend flexibly from one shape to another under external influences

by Daniel E. Koshland, Jr.

A living system must have both the capacity to act and the capacity to control its actions. We humans, for example, must be able to digest food, but we cannot be eating all the time. Hence we need both a positive control to turn on the process (a desire to eat when food is needed) and a negative control to turn off the process (a desire to stop eating when we have had enough). Similarly, the process of blood clotting must be turned on when we bleed from a wound but must be turned off afterward so that it does not lead to coronary thrombosis. We must also be able to turn muscles on in order to move and to turn them off in order to relax. In short, every biological system has built-in controls to initiate or accelerate a process under some conditions and to terminate or decelerate it under other conditions.

Considering the diversity of processes that must be regulated and the diversity of environmental conditions to which an organism must react, it might be expected that the controls in biological systems are enormously complex. That is true, and yet when one examines the processes more closely, it appears that the fundamental elements of control are remarkably simple and universal.

The fundamental control element in all living systems—from the smallest bacterium to man—is the protein molecule. Enzymes, the biological catalysts that control all the chemical processes of living systems, are proteins. Sensory receptors, which enable us to see, hear, taste and smell, are proteins. Antibodies, which provide immunity against infection, are proteins. Recent experiments have established that it is the ability of these proteins to change shape under external influences that provides the "on-off" controls that are so vital to the living system.

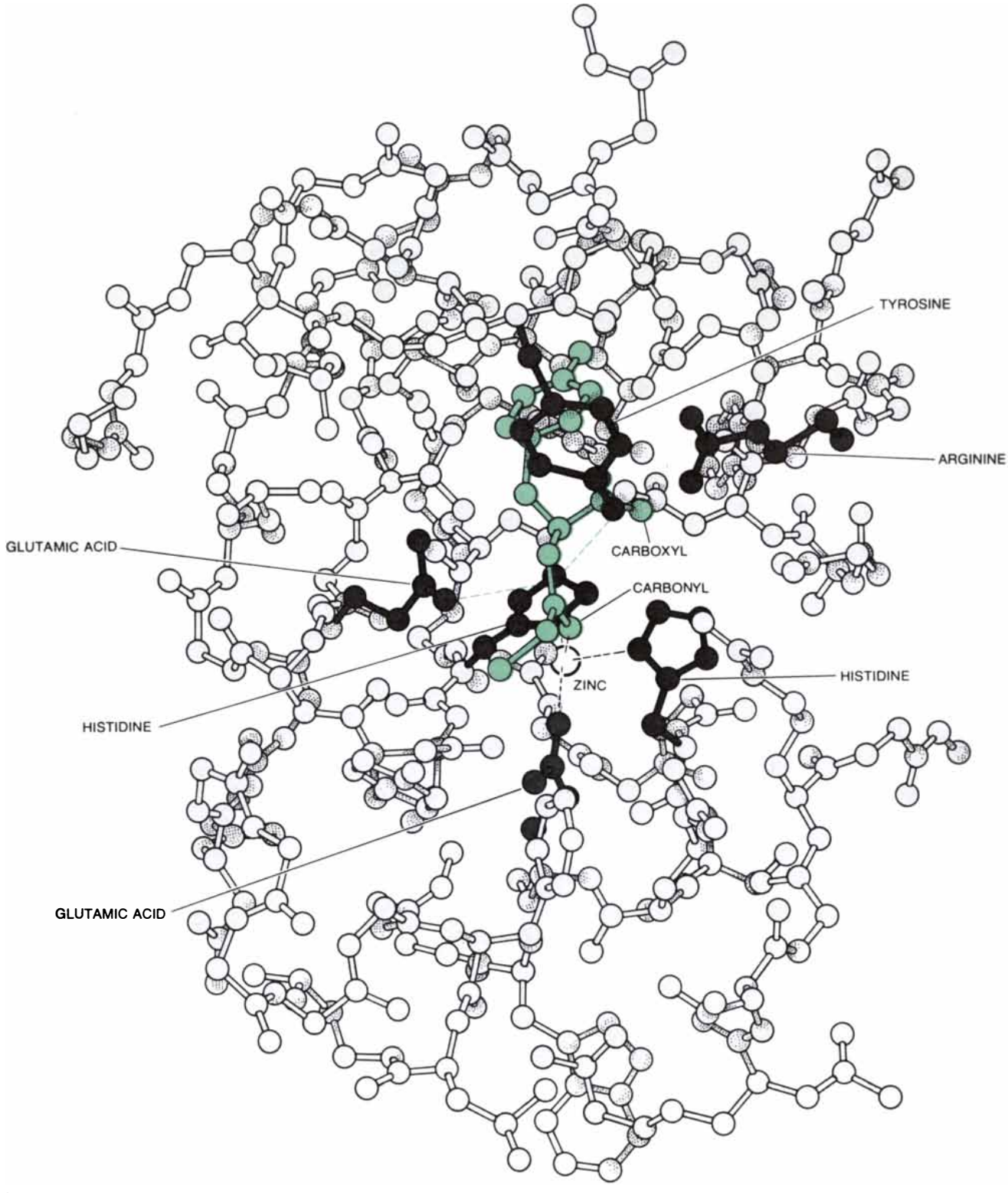
The concept of protein shape as a control mechanism arose from studies of enzymes, a development that is hardly surprising, since enzymes are the easiest to study of all the regulating proteins. It has been estimated that an average living cell contains some 3,000 different enzymes. Each of them catalyzes a distinct chemical reaction in which compounds called substrates are converted into other compounds called products. Fortunately for our understanding of biological systems many enzymes are quite sturdy molecules and can be extracted from a physiological system without destroying their biological properties. Hence they can be studied in the test tube and made to perform the same catalytic role there that they perform in the living organism. Moreover, one can subject them to the same environmental influences in the test tube that they experience in the living cell, thereby getting clues to their role in biological regulation. Finally, enzymes can be obtained in large enough amounts for their physical properties to be studied. However, even though our understanding of the role of shape in protein regulation began with enzymes, the principles of regulation worked out for enzymes appear to be universal and can be applied to other proteins that are more difficult to obtain in bulk.

Shape Changes in Enzyme Catalysis

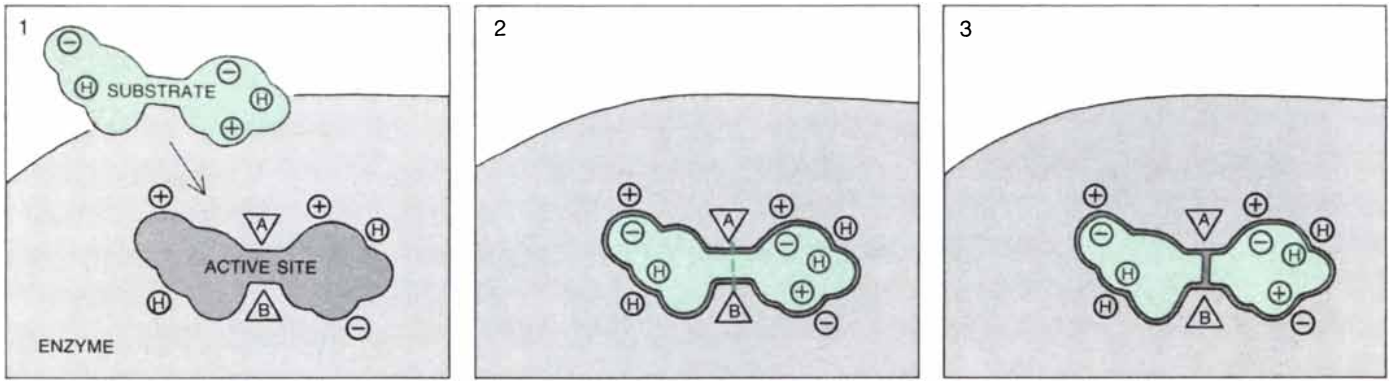
It has long been known that the basic mechanism by which enzymes catalyze chemical reactions begins with the binding of the substrate (or substrates) to the surface of the enzyme. The enzyme then polarizes the chemical bonds in the substrate, causing a reaction that leads to the formation of products on the surface. The release of these products from the surface regenerates the free enzyme and allows the cycle to be repeated.

Unlike catalysts made in the laboratory, enzymes have the special property called specificity, which means that only one chemical compound or a very few can react with a particular enzyme. This property can be explained by the template, or lock-and-key, hypothesis put forward in 1894 by Emil Fischer, which postulates that the enzyme is designed to allow only special compounds to fit on its surface, just as a key fits a lock or as two pieces of a jigsaw puzzle fit together. The complementary shapes allow one compound to fit and exclude other compounds that lack the correct size, shape or charge distribution [*see illustration on pages 54 and 55*]. Modern X-ray crystallography has revealed in detail precisely such a fit between enzyme and substrate [*see illustration on opposite page*].

Although this concept could explain much of the specificity data, some glaring discrepancies were found. For instance, certain oversized and undersized compounds were found to bind to the surface of the enzyme even though they failed to form products. Furthermore, it was difficult on the basis of the rigid-template theory to explain how sugars can compete with water in enzymatic reactions. It was also difficult to explain why substrates bind in a specific order in many enzymatic reactions. These facts and others like them led to the hypothesis that the enzyme does not exist initially in a shape complementary to that of the substrate but rather is induced to take the complementary shape in much the same way that a hand induces a change in the shape of a glove [*see illustration on pages 56 and 57*]. This "induced fit" theory assumes that the substrate plays a role in determining the final shape of the enzyme and that the enzyme is therefore flexible. Proof that proteins do in fact change their shape



PRECISE FIT between the active site on the surface of a large protein molecule and the specific substrate molecule with which the protein reacts is evident in this simplified three-dimensional drawing, based on X-ray-crystallographic data obtained by William N. Lipscomb, Jr., and his colleagues at Harvard University. The protein, rendered in shades of gray, is carboxypeptidase *A*, a digestive enzyme that (as its name implies) works by cutting the polypeptide chain of the substrate near its carboxyl end. The substrate, rendered in color, is carbobenzoxyalanylalanyl tyrosine. Approximately a fourth of the total number of atoms in the polypeptide chains that comprise the two molecules are represented in this view. The atoms shown are mostly carbon, with a small admixture of nitrogen and oxygen; all hydrogen atoms have been omitted. The six active-site side chains that specifically interact with the substrate are the darker gray. For example, a positively charged arginine side chain is shown attracting the negatively charged carboxyl group of the substrate. In addition certain hydrophobic, or oily, regions on the substrate are attracted to similar regions on the enzyme, strengthening the attraction between the two molecules. A zinc atom (*white*) forms an additional "coordination bond" involving a carbonyl group on the substrate and three other amino acid side chains (glutamic acid and two histidines) extending from inside the enzyme's bowl-shaped active site. The tyrosine side chain and the second glutamic acid side chain of the active site are catalytic groups that polarize the electrons in one of the substrate's chemical bonds, splitting that bond (*broken colored line*) and thereby dividing the substrate into the two parts that are the reaction products of this particular enzyme-substrate combination.



LOCK-AND-KEY MODEL of the mechanism by which enzymes catalyze chemical reactions, put forward in 1894 by Emil Fischer, postulates that the active site of the enzyme is a rigid, templatelike structure that allows only special compounds to fit, just as a key fits a lock. Such a compound, called a substrate, is attracted to the enzyme's active site by mutually attractive groups, such as the electrostatic charges on certain amino acid side chains (+, −), and by the coalescing tendency of adjacent hydrophobic side chains (*H*). In the lock-and-key model the catalytic groups (*A*, *B*) are poised in advance (*1*) to cause the reaction that ruptures the chemical bond

under the influence of small molecules was initially obtained by chemical studies showing differences in the reactivity to protein reagents of the amino acid side chains that are arrayed along the spine of the protein molecule. The hypothesis has since been verified with the aid of advanced physical techniques, most notably X-ray crystallography.

Once the concept of a flexible enzyme was entertained, the puzzling behavior of the special classes of compounds that were found to bind to the enzyme without forming products could be explained. One type of oversized molecule, for example, binds to the surface of the enzyme, but in doing so it distorts the protein into a shape that does not allow the catalytic groups to be properly aligned. Other compounds can bind to the enzyme too, but they do not have either sufficient size or the correct chemical characteristics to induce the proper alignment. Hence even though both types of compound bind, neither reacts with the enzyme. (Some molecules are of course too big to be bound even to a flexible enzyme; the examples cited here are chosen to show the differences between the rigid-template theory and the induced-fit theory.)

Shape Changes in Regulation

This finding of flexibility does not mean that all proteins must be flexible. Some may indeed be quite rigid, and these are explained very well by the lock-and-key hypothesis. Nor does it mean that enzymes that exhibit flexibility must do so with all chemical compounds. The finding does mean, however, that protein flexibility is a key feature of enzyme action. Indeed, the capacity to induce a change in shape has been found to be a widespread and vital feature of most, if not all, enzymes.

The concept of protein flexibility led to the deduction that small molecules not themselves involved in the chemical reaction could help to make a deficient molecule act as a substrate by altering the shape of the enzyme [*see illustration on page 58*]. In the case of a molecule that is too small to induce the proper alignment of catalytic groups, for example, certain molecules that are not consumed in the reaction can be added to produce a stable shape with the right alignment of catalytic groups. One way for this realignment to occur is for a second molecule to bind immediately adjacent to the deficient molecule, thus inducing the proper shape at the active site. This prediction of the flexible-enzyme theory has been confirmed in many cases (for example in the case of the carbohydrate-splitting enzyme hexokinase by Alberto Sols and his co-workers in Spain and in the case of the digestive enzyme trypsin by T. Inagami and T. Murachi in Japan). Molecules that bind far away from the active catalytic site can also induce a proper shape. In that case the induced change is transmitted through the protein like a row of falling dominoes until the active site is altered appropriately.

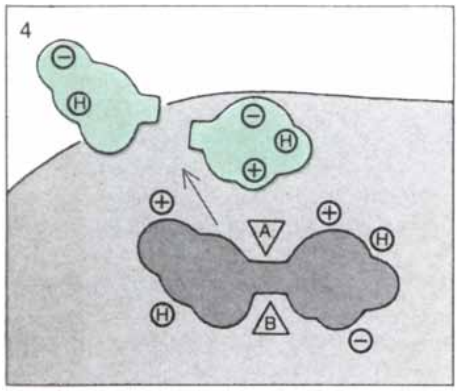
The reverse process can also occur. Flexible enzymes can be distorted out of the active shape by molecules called inhibitors. These molecules can cause a disruption of either the catalytic function or the binding function of the enzyme, in either case giving rise to an inactive shape [*see illustration on page 59*].

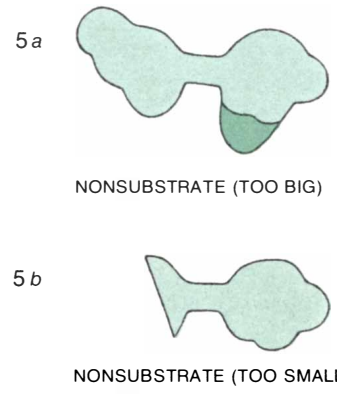
In short, a regulatory molecule that is not itself involved in the chemical reaction can control the activity of an enzyme by changing its shape. It can turn the enzyme on by inducing the correct shape or turn it off by inducing an incorrect shape. In biological systems one of the most important groups of such molecules is the hormones. Although hormones are secreted in small amounts, much too small to be important directly as foodstuffs or sources of energy, they have a tremendous influence on the regulatory processes of the cell. The manner in which hormones exert control is easily explained by the flexible-protein hypothesis. Since these molecules are not consumed, they can be used again and again to activate the enzyme molecules, unlike the substrate molecules, which are consumed. Therefore such regulators need be present only in very small amounts.

Sometimes, as in the case of adrenalin, the initial hormone induces the formation of a second molecule, cyclic AMP, which acts as a regulator for many enzymes by changing their shape [see "Cyclic AMP," by Ira Pastan; SCIENTIFIC AMERICAN, August, 1972]. In general the shape changes induced by the regulator molecules are similar to the shape changes induced by the substrate.

The Regulation of Pathways

In discussing the shapes of proteins it is difficult, tedious and often unnecessary to show the detailed parts of a large protein molecule. For convenience the different molecular shapes can be symbolized by geometric figures: squares, triangles, circles and so on [*see illustration on page 62*]. It is nonetheless important to remember that even though such shapes can be expressed by simple line drawings, a change from a circle to a square, say, designates subtle and com-





in the neck of the substrate (*2*), converting the substrate into the compounds called products (*3*). The release of these products from the surface (*4*) regenerates the free enzyme and allows the cycle to be repeated. According to this view, the precisely complementary shapes of the substrate and the enzyme that acts on it exclude reactions with compounds that are either too large to fit into the active site (*5a*) or too small to be attracted to the active site (*5b*).

plex changes in the orientation of the many amino acid side chains that constitute a protein. This procedure is analogous to using the symbol *C* for a carbon atom rather than writing out the complete quantum-mechanical description of the electrons, protons and neutrons that constitute the atom. The different shapes of proteins are referred to as conformations, because the protein changes fulfill a chemist's definition of a conformational change, that is, a change in the shape of a molecule caused by rotation around a single chemical bond. The terms "conformation" and "conformational change" are technical synonyms for the terms "shape" and "shape change" employed so far in this article.

The site on the surface of an enzyme at which the catalytic action takes place is called the active site. The binding site for the regulatory molecule is called the regulatory site or the allosteric site to distinguish it from the catalytic site. (The term "allosteric," meaning "the other site," was coined by the French biochemist Jacques Monod, a leader in this area of investigation, and it has gradually come to be used as a general term for regulatory proteins.) Regulatory molecules are also called effectors, modifiers or allosteric effectors.

Let us now consider how these ideas of protein shape help to explain some types of regulatory control. One of the most important decisions a living system must make involves how to process a food substance. We need energy in the form of molecules of adenosine triphosphate, or ATP, for all our bodily processes: to contract our muscles, to see with our eyes, to activate our nerves and to generate our structural materials. This energy comes from the combustion of food. A certain amount of ATP is constantly being used to maintain the system even in a resting state; the heart, for example, continues pumping even when we are asleep.

It is obviously desirable during periods of low energy demand to store energy for future needs. One of the ways of storing energy is in the form of glycogen, a chain of sugar molecules. Thus when a molecule of glucose, say, is ingested by a living system, it can be directed along alternate pathways: either it can be oxidized immediately to form carbon dioxide, giving off large amounts of ATP, or it can be stored in the form of glycogen to be released on future demand [*see illustration on page 60*]. The enzyme phosphofructokinase, which is involved in the first of these pathways, is turned on and off by variations in the level of ATP in the system. Another enzyme, glycogen synthetase, is involved as part of the second pathway in the regulatory control of the synthetic process; it can be turned on and off by the presence of regulatory compounds in the cell. A third enzyme, glycogen phosphorylase, catalyzes the reaction from glycogen to glucose-1-phosphate in the third pathway and is the enzyme used when the need arises for the retrieval of glucose from storage.

The way these enzymes work under the influence of supply and demand is simple and ingenious. When ATP levels are high, the phosphofructokinase enzyme of pathway 1 and the phosphorylase enzyme of pathway 3 are turned off and the glycogen synthetase of pathway 2 is activated. Glucose is therefore diverted to glycogen for storage. When an animal is frightened, however, it secretes adrenalin, which ultimately activates phosphorylase (pathway 3) and phosphofructokinase (pathway 1) and deactivates glycogen synthetase (pathway 2). In that case glycogen is removed from storage and converted into ATP energy to help the animal escape from the danger. In other words, when we are frightened, the hormone demands that our reserves of glycogen be made available as a source of energy. More than one demand can release glycogen from storage. Hunger and certain stimulators of muscular activity (such as calcium ions) activate phosphorylase and thus can also generate energy. In principle such influences act in the same way that adrenalin does.

This particular regulatory system can be used to illustrate two important points. First, enzymes must be available in every living system and yet they must not be equally active at all times. The alteration in shape under the influence of metabolites and hormones therefore provides a mechanism for turning these enzymes on and off under different external conditions (ranging, say, from starvation to satiety).

The second principle is more subtle. If all the enzymes involved in carbohydrate metabolism were to be activated simultaneously, one would simply have a short circuit going around and around pathways 1, 2 and 3, storing and burning glucose to no avail. It has usually been found for alternate pathways of this kind that a molecule that activates one pathway inhibits the other. In the case of glycogen storage, for example, the regulatory molecule inhibits the enzyme in pathway 2 and activates enzymes in pathways 1 and 3; the pathways for the oxidation of glucose and the removal of glucose from storage are accelerated and the pathway for the storage of glucose is blocked. In the absence of this regulatory molecule, equilibrium favors the inactive form of the enzyme in pathways 1 and 3 and the active form of the enzyme in pathway 2; in other words, the system favors glucose storage. Thus a short circuit is avoided by changing the shape of the enzymes reciprocally so that synthesis to glycogen is favored when ATP levels are high and degradation to carbon dioxide and ATP is favored when energy is needed.

Protein Structure

At this point it is worth considering in a little more detail how the design of proteins enables this regulation to proceed. The proteins that act as enzymes range in weight from approximately 10,000 daltons to many millions of daltons. (One dalton is roughly equal to the weight of one hydrogen atom.) The

higher figure is deceptive, however, because all such large proteins are made up of peptide subunits, which usually range between 15,000 and 100,000 daltons. For example, the enzyme phosphorylase, which is important in glycogen storage and degradation, is a dimer (a two-peptide polymer) composed of two identical subunits each with a molecular weight of 96,000 daltons. Similarly, aspartyl transcarbamylase, the first enzyme in the pathway leading to the synthesis of cytidine triphosphate, or CTP, is a dodecamer (a 12-peptide polymer) composed of six subunits of one kind (each with a molecular weight of 35,000 daltons) and six subunits of another kind (each with a molecular weight of 17,000 daltons). These subunits are attracted to one another by noncovalent forces: largely electrostatic attractions or hydrophobic bonds (a name used to describe the tendency of oily regions of a structure to be forced together in the same way that oil droplets tend to coalesce in water). Fortunately qualitative features do not change with the size of the protein, so that by studying the simpler proteins it is possible to understand the properties of proteins in general.

A peptide chain with a molecular weight of only 25,000 daltons is still large compared with the molecular weight of most substrates, which are usually compounds in the molecular-weight range of 100 to 1,000 daltons. Occasionally enzymes act on very large molecules such as DNA (deoxyribonucleic acid), cellulose or other proteins, but when they do, they usually bind only a small portion of these large molecules, so that the effective substrate size is still only about 1,000 daltons or less. This difference in relative size means that only a small portion of the enzyme's surface is actually involved in catalysis. The rest of the surface is available for binding the molecules that are involved in regulation and for the association of subunits with one another.

Cooperativity

The concept of protein flexibility provided an explanation for a long-known phenomenon that had been originally discovered by the Danish physiologist Christian Bohr in the 19th century. Bohr noted that the binding rate of oxygen to hemoglobin could be described by a sigmoid, or S-shaped, curve instead of the normal hyperbolic curve observed for the binding pattern of most enzymes [*see illustration on page 61*]. He correctly deduced that this unusual type of binding curve would result if the first molecule bound made it easier for the next molecule to bind, and so forth; hence he called the process cooperative. Since then hemoglobin has been the subject of intensive study by many prominent investigators, and much of our knowledge of the phenomenon of cooperativity results from examination of this vital protein. Cooperativity is not limited to hemoglobin, however. Sigmoid binding curves are also common and important features of the regulatory proteins.

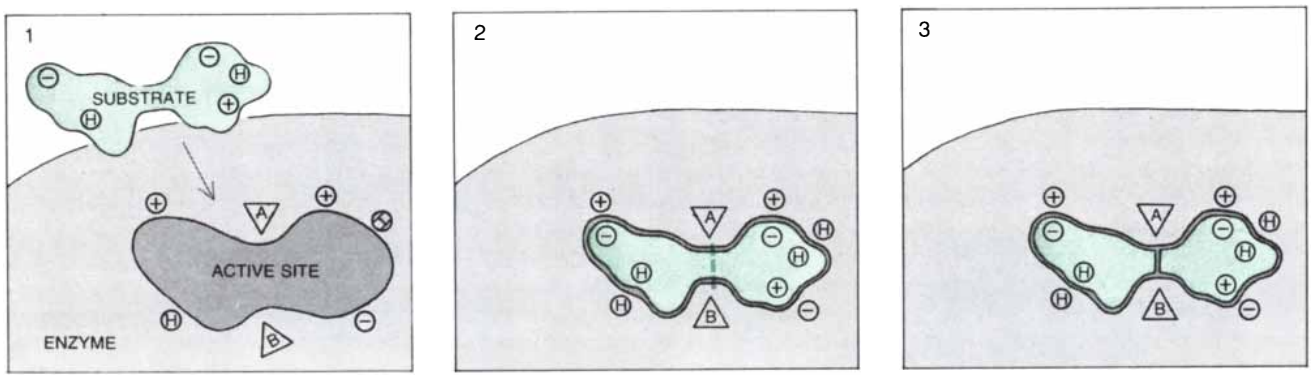
The appearance of a cooperative binding pattern in the case of a regulatory protein can be explained by means of the flexible-protein hypothesis, using the simplest multisubunit protein, a dimer made up of two identical subunits [*see illustration on page 63*]. A number of such dimer proteins exist in nature; the types of interaction they are involved in are similar to those of the more complex proteins. The binding of the substrate induces conformational changes that depend on the structure of both the substrate and the protein.

Three general types of conformational change are known. In one type the first molecule of substrate alters the subunit to which it is bound but does not alter the interactions between the subunits. The second subunit therefore binds substrate in just the same way and with the same affinity as the first. There is no cooperation between the subunit sites.

The next type of conformational change is quite different. Here the first molecule to be bound induces a conformational change in the first subunit, which induces shape changes in the second subunit. These changes, which are transmitted through the protein structure, change the active site in the second subunit so that it becomes more receptive to the substrate; hence the second molecule of substrate is bound more readily than the first. This phenomenon, called positive cooperativity, explains the sigmoid binding curve discovered by Bohr.

In the third case the first molecule induces a conformational change that makes the binding site in the second subunit less attractive to the substrate because of their incompatible geometries. The binding of the second molecule is therefore discriminated against in favor of the first and proceeds much less readily. This phenomenon is called negative cooperativity (cooperativity because of the subunit interactions, negative because the first molecule has a negative effect on the second).

The induced-fit hypothesis can readily explain the sigmoid curve of multisubunit proteins; a mathematical adaptation of the induced-fit approach was devised by George Némethy, David Filmer and me to do just that. Our solution is not the only possible explanation, but



INDUCED-FIT MODEL of enzyme action, developed largely by the author and his colleagues at the University of California at Berkeley, assumes that the enzyme does not exist initially in a shape complementary to that of the substrate (*1*) but rather is induced to take the complementary shape in much the same way that a hand induces a change in the shape of a glove (*2*). Once the substrate is bound, the catalytic groups of the enzyme are in position to cut the chemical bond in the substrate's neck, forming the reac-

there is substantial evidence that it is valid for a large number of enzymes exhibiting such sigmoid curves. Besides being able to explain a puzzling phenomenon in terms of the protein structure, the induced-fit approach predicted that a different type of interaction—negative cooperativity—should exist (in other words, that the induced conformational changes would make the second molecule bind less readily than the first). Such a phenomenon was not known in nature at the time, but its prediction from theory led to a determined search for it. In 1968 Abby Conway and I discovered an example of this strange phenomenon in an enzyme that participates in carbohydrate metabolism: glyceraldehyde 3-phosphate dehydrogenase. Negative cooperativity has since been found in many other enzymes.

The mathematical analysis of such a negative-cooperativity pattern explains in part why the phenomenon escaped detection. On superficial inspection a negative-cooperativity binding curve looks like a hyperbola. A more careful analysis shows that in reality the curve is not a true hyperbola. Such a curve can be explained only by assuming that the second molecule binds to the protein less readily than the first. Several examples of each type of cooperativity have now been established. Induced conformational changes can be caused by activators and inhibitors as well as by substrates, and therefore all three types of molecule can give rise to cooperative interactions.

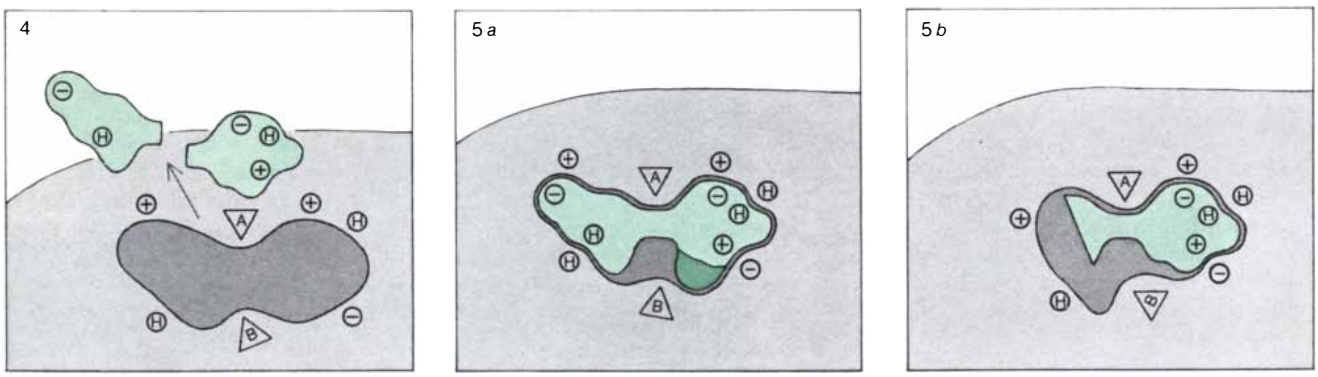
All these concepts apply to molecules with more than two subunits in the same way. In positive interactions the first molecule makes it easier for the second to bind, the second makes it easier for the third and so on. In negative interactions each interacting molecule makes it more difficult for the next molecule to interact. The small molecule (a substrate, an inhibitor or an activator) induces the change and the effect is then transmitted to the neighboring subunits. The greater the number of subunits is, the more dramatic the cooperativity can be, but the actual cooperativity pattern observed depends on the details of the individual protein structure.

Why should such cooperative phenomena exist in nature? Would it not be better for the protein to be designed correctly in the first place, so that these induced conformational changes need not alter the interactions between the subunits or the shape of the protein? One of the reasons protein flexibility is so important, of course, has already been mentioned: flexibility makes it possible for an enzyme to be regulated by molecules that are not themselves consumed in the enzymatic reaction.

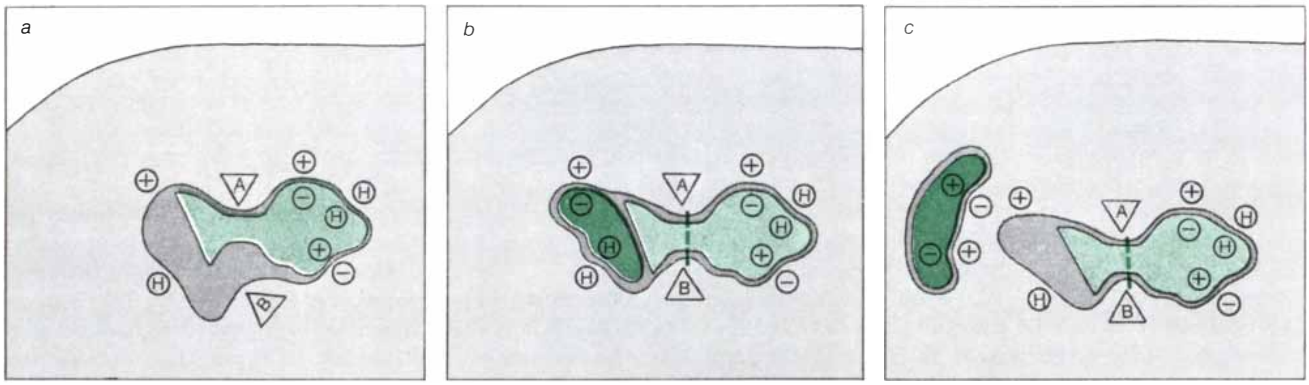
The concept of cooperativity, however, apparently provides another reason for protein flexibility. If one examines the binding curves for the three types of cooperativity and compares the change in concentration of a compound to the change in the activity, one finds that approximately an 81-fold change in the concentration of the substrate is needed to go from an activity level of 10 percent to one of 90 percent, assuming that the protein follows the hyperbolic binding curve of a normal noninteractive protein. In sharp contrast, only a ninefold change in concentration is needed if one assumes that the protein is designed with rather mild positive cooperativity. (In hemoglobin a threefold change in concentration will do the job, and in CTP synthetase a 1.5-fold change is sufficient.) In other words, a protein with positive cooperativity is much more sensitive to small fluctuations in the environment than a protein with a normal binding pattern. The sensitivity is increased for inhibitors and activators as well as for substrates. Positive cooperativity is thus an amplification device to make a small signal have a much larger regulatory effect. This increased sensitivity is extraordinarily important for the regulatory function of these proteins.

One example of the physiological importance of the phenomenon of cooperativity can be observed in patients who have a mutant hemoglobin that lacks the positive cooperativity of normal hemoglobin. Positive cooperativity enables hemoglobin molecules to absorb large amounts of oxygen in the lungs and to deposit large amounts of oxygen in the tissues, even though the pressure of oxygen does not vary much from one location to the other. This situation follows from the steepness of the sigmoid curve associated with positive cooperativity. The noncooperative protein is far less efficient in transporting oxygen, and patients possessing it are very sick. In a crude sense it might be said that the positively cooperative protein resembles a truck that takes on a full load of dirt at one location and empties all of it at another, whereas a protein without cooperativity cannot take on a full load to begin with and can only get rid of part of the load.

What, then, is the role of negative cooperativity? A look at its binding curve provides the answer to that question too. Here the protein is less sensitive to fluctuations in the environment than a noncooperative protein; in other words, a much greater change in the concentration of the substrate, the inhibitor or the activator is needed to go from an activity level of 10 percent to one of 90 percent. Some proteins should not be subject to fluctuations in the environ-



tion products (*3*), which leave the surface of the enzyme, returning the active site to its original noncomplementary shape (*4*). This concept of a flexible enzyme made it possible to explain the previously puzzling observation that certain oversized and undersized compounds were able to bind to the surface of the enzyme without forming products (*5a, 5b*). Even though both types of nonsubstrate compound bind, neither succeeds in inducing the proper alignment of the catalytic groups and hence neither reacts with the enzyme.



ACTIVATOR MOLECULES can, according to the induced-fit model of enzyme action, help to make a deficient molecule act as a substrate by altering the shape of the enzyme. For example, in the case of a molecule that is too small to induce the proper alignment of catalytic groups (*a*) a second molecule can bind immediately adjacent to the deficient molecule inside the active site, thereby inducing a stable shape with the proper alignment of catalytic groups (*b*). A molecule that binds at a site outside the active site can also induce the proper shape (*c*). In neither case is the activator molecule itself consumed in the ensuing chemical reaction.

ment, and the proteins that regulate such processes may use negative cooperativity to damp their sensitivity and make them less subject to environmental changes.

The Nature of the Shape Changes

How are conformational changes propagated through the protein molecule and how extensive are they? An understanding of these processes has flowed from two sources: chemical studies of the composition of proteins and X-ray studies of protein structure.

Some of the important structural data were obtained as a result of the pioneering work of M. F. Perutz and his coworkers on hemoglobin [see "The Hemoglobin Molecule," by M. F. Perutz; SCIENTIFIC AMERICAN, November, 1964]. Hemoglobin consists of four similar but not identical subunits arranged in a tetrahedral array. When oxygen is bound to the heme, or iron-containing, binding group of one of these subunits, there is a small shift in the position of the iron atom, which in turn causes a small shift of an adjacent histidine side chain. The histidine shift then compresses several amino acid side chains in a helical portion of the protein and squeezes a tyrosine side chain out of a small pocket. The movement of the tyrosine group dislocates the subunit to which it is attached and breaks the salt linkage between the subunits.

These are not the only regions that are affected by the binding of the oxygen atom, but they are illustrative of the type of change that can occur in a typical protein molecule. The point is that the protein is a tightly structured molecule with intimate close-packing relations between the atoms of the amino acid side chains. A shift of one such group by even a fraction of an angstrom causes realignments of other side chains, generating a "domino" effect that can extend through the entire protein molecule. The changes within the hemoglobin subunit cause a realignment of the subunits with respect to one another.

The importance of these protein shifts can be demonstrated with mutant hemoglobins isolated from sick patients. When the mutant hemoglobins were examined by Perutz and his colleagues, the amino acid side chains involved in the shape changes were found to be altered.

A further understanding of these processes has been obtained from experiments conducted in my laboratory at the University of California at Berkeley by Alexander Levitzki and William Stallcup. They worked with CTP synthetase, a key enzyme that converts uridine triphosphate, or UTP, to CTP in the metabolism of nucleic acids. This regulatory protein usually exists in the form of four identical subunits, but it can be studied as a dimer. Each subunit has binding sites for its three substrates: UTP, ATP and glutamine, and also for its regulatory effector, guanosine triphosphate, or GTP [*see illustration on page 64*]. The glutamine reacts at the glutamic site on the enzyme, forming what is called the glutamyl enzyme. In doing so it liberates ammonia, which binds to the ammonia site on the enzyme. The ammonia in turn reacts with UTP, which reacts with ATP. The close coupling of the chemical steps indicates that the sites are immediately adjacent to one another.

When the glutamyl enzyme is formed, the reaction produces a covalent bond between glutamine and a cysteine side chain on the surface of the protein. At the beginning the two cysteine groups of the two active sites are equally reactive. When glutamine reacts with the same side chain, however, a strange thing happens. The reaction of the substrate with one subunit turns off the cysteine side chain on the other subunit. The effect is only temporary; the glutamyl enzyme reacts further to regenerate free enzyme.

As it happens, this change in the shape of the enzyme can be "frozen" (somewhat like stopping a motion-picture film at a single frame) with the aid of an "affinity label," dioxoazonorleucine. The molecule of dioxoazonorleucine is enough like the molecule of glutamine so that it forms a covalent bond with the cysteine side chain but enough different so that it cannot react further. The result is that only one of the two initially identical subunits reacts with the dioxoazonorleucine. A reaction at one subunit turns off its neighbor, so that only one of the two potential subunits reacts at any one time, giving rise to what we have termed the "half of the sites" phenomenon.

This phenomenon is observed in a number of enzymes and indicates that in many cases the shape change can be transmitted over long molecular distances. On the basis of the size of the enzyme subunits the cysteine side chains in CTP synthetase are probably between 40 and 60 angstroms apart.

The most surprising part of the finding is that the immediately adjacent sites—the ammonia, UTP and ATP sites—are not altered at all when the cysteine group is modified. All their properties remain the same. Hence the formation of a bond apparently transmits a signal that has dramatic consequences as much as 60 angstroms away without perturb-

ing structures within four or five angstroms! When the reactivity of amino acid side chains to protein reagents is examined, many side chains are found to change position and many others to remain unchanged.

A similar pattern is evoked by GTP, the regulatory molecule. GTP activates the enzyme by altering the reactivity of the glutamine site on the same subunit, so that the glutamyl enzyme is formed more rapidly. The same shape change that activates the glutamyl site deactivates the GTP site of the neighboring subunit. Moreover, the shape changes induced by GTP do not alter the properties of the ammonia, ATP or UTP sites. Thus the GTP merely by binding to the surface of the protein can direct the alteration of some side chains and leave others unchanged. The same is true of many other activators, inhibitors and substrates.

The Scale of the Shape Change

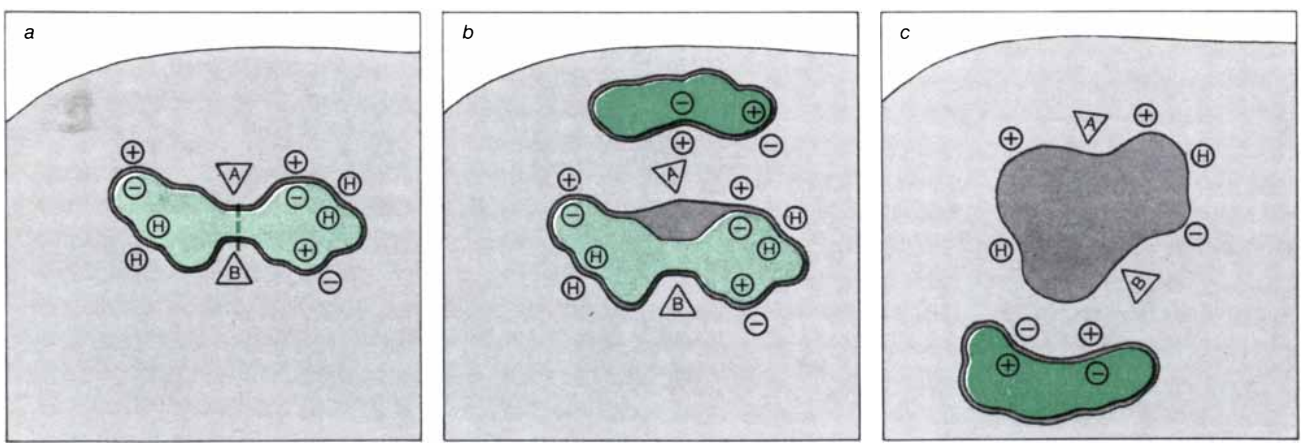
Hence it appears that induced conformational changes in proteins are not like the ever widening concentric ripples produced by throwing a pebble into a pond. They are more like a spider web in which the strands are devised to transmit a perturbation occurring in one corner of the web to another corner. The perturbation can be transmitted over long distances and can alter the positions of many strands, but a clever design can ensure that some strands will remain unchanged at the same time that others are shifting appreciably. The protein, like the spider web, is designed to transmit information in a focused manner to some regions and to leave others unchanged.

The schematic drawings used to illustrate the alteration of protein shapes in this article tend to exaggerate the relative movements necessary to achieve this control. Actually we do not know precisely how big the movement of catalytic or binding side chains must be to achieve on-off switching. There are strong suggestions, however, that the movements do not have to be very large. The length of a carbon-oxygen bond, for example, is about 1.3 angstroms. A catalytic group that needs to be positioned close enough to an oxygen atom to pull electrons out of it would therefore be ineffective as a catalyst if it were positioned next to the carbon atom. Hence a movement on the order of an angstrom or two would appear to be sufficient to make the difference between an effective catalyst and an ineffective one. The effective movement is probably somewhat less than that, particularly because most regulators do not completely turn off the enzyme but rather reduce its function in the direction of either catalysis or binding.

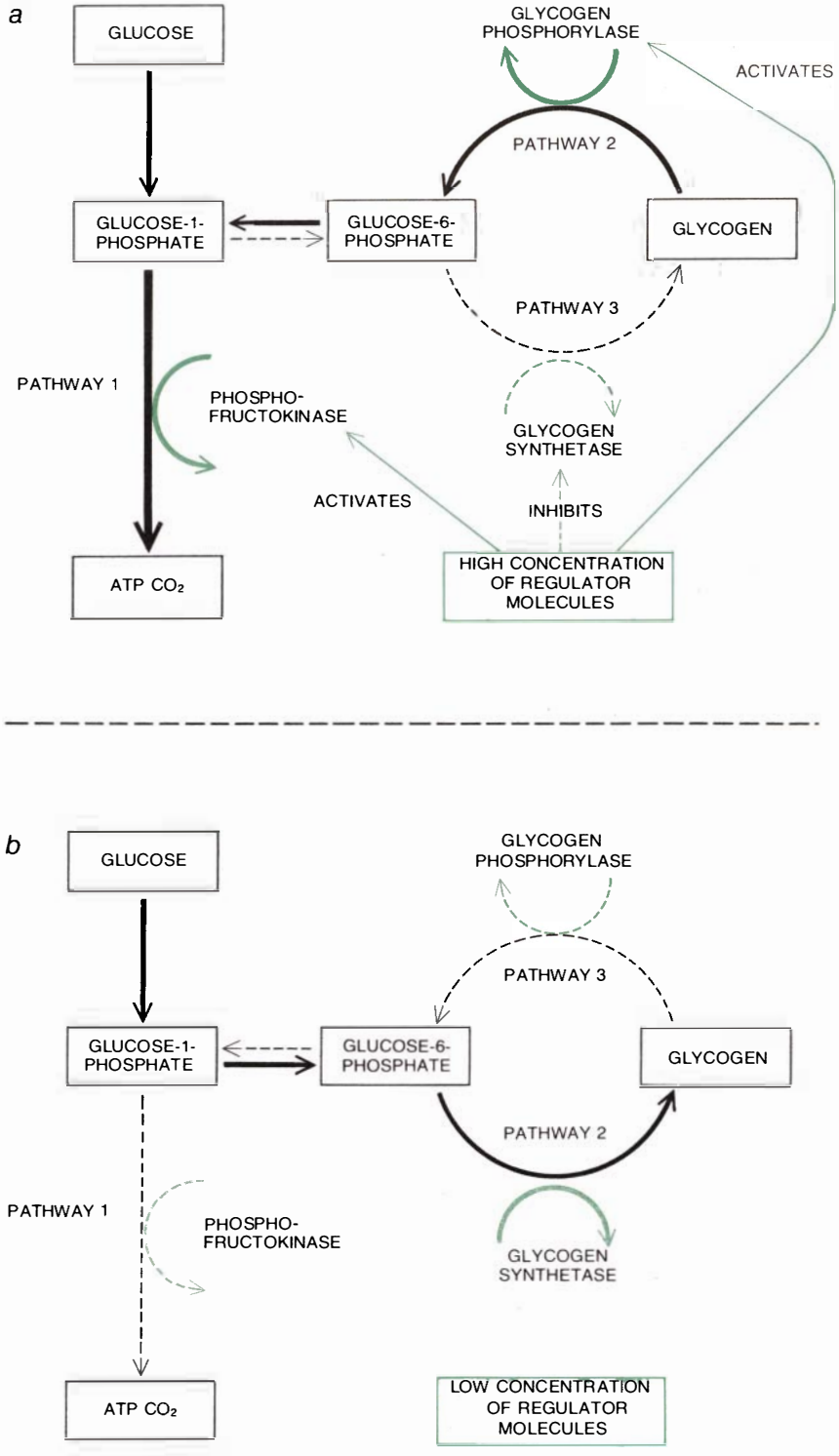
Movements of side chains by 10 or 12 angstroms have been observed in carboxypeptide molecules by William N. Lipscomb, Jr., and his colleagues at Harvard University, using the X-ray-crystallographic approach. It seems likely that the large movements observed in this type of protein are more than are necessary for regulation in general and that many of the atoms move much less or not at all. It has already been found that the movement of certain atoms such as the iron atom in hemoglobin and the sulfur atom in CTP synthetase may be less than an angstrom, and yet such a movement can trigger large conformational changes. Model experiments with simpler compounds suggest that an alteration of less than half an angstrom can also alter reactivity greatly. In short, the schematic drawings shown here should be taken to indicate that the movements are large enough to alter the function of a protein but that they do not represent the actual scale of conformational changes. Overall observations of protein molecules also indicate that no tremendously large shape changes occur in the subunit as a whole even though a few of the atoms move by several angstroms. By focusing on a few groups and using schematic pictures an erroneous impression could be obtained if one does not remember that one is looking at only a very small portion of the spider web. The gross anatomy of the web is unchanged; only a few strands shift with respect to one another, and yet these shifts are highly significant in terms of function.

Advantages of Induced Changes

One of the advantages of having enzymes that can be induced to change their shape in response to the proper stimulus can be seen in the example we have been considering: CTP synthetase. I have mentioned that the substrates ATP and UTP show strong positive cooperativity. In the case of ATP, for example, the first molecule of ATP alters the shape of the enzyme so that the subsequent molecules of ATP bind rapidly. Actually CTP synthetase is one of the most cooperative enzymes known (even more so than hemoglobin). The binding of the first molecule of substrate has a triggering effect that allows the rapid and complete binding of the subsequent molecules of ATP, so that only the free enzyme and a form of the enzyme with four ATP molecules bound are found in appreciable amounts. This positive cooperativity means that the enzyme is



INHIBITOR MOLECULES can, in the reverse process, distort a flexible enzyme out of the active shape by binding at sites outside the active site. In the case of a substrate that would otherwise react with the enzyme (*a*) such inhibitors can cause a disruption of either the catalytic function (*b*) or the binding function (*c*) of the enzyme, in effect "turning off" the enzyme-substrate reaction.



ALTERNATE BIOLOGICAL PATHWAYS are controlled by regulatory molecules through tbe alteration of enzyme shapes. In the human body, for example, ingested glucose is normally converted to glucose-6-phosphate, whicb can either be oxidized immediately to provide energy in the form of molecules of ATP (adenosine triphosphate) or stored in tbe form of glycogen (a chain of sugar molecules) to be released on future demand. Which pathways are chosen depends on the concentration of regulatory molecules. In times of energy demand high concentrations of these molecules convert the enzyme of pathway 1 (phosphofructokinase) and the enzyme of pathway 3 (glycogen phosphorylase) to active forms while at the same time converting the enzyme of pathway 2 (glycogen synthetase) to an inactive form, thereby channeling both the ingested glucose and the stored glycogen into the production of ATP (*top*). Low concentrations of the regulatory molecules enable the enzymes of pathways 1 and 3 to return to their inactive forms while enabling the enzyme of pathway 2 to become active again, thereby diverting the glucose into glycogen for storage (*bottom*). The reciprocal effect of the regulatory molecules on pathways 2 and 3 prevents their both being active at the same time and therefore rules out a futile short circuit in the system.

very susceptible to small fluctuations in the concentration of ATP and hence is highly sensitive to environmental influences on this molecule. In contrast, the regulatory molecule, GTP, induces conformational changes leading to negative cooperativity, which desensitizes the enzyme toward fluctuations of GTP in the environment.

What does all this mean in terms of enzyme action? It means that the enzyme has been programmed by its design through evolutionary time so that it can be responsive to small changes in ATP levels and at the same time be desensitized to rather wide fluctuations of GTP levels. A conceivable reason for this ability is that ATP is highly controlled in the biological system, since it is such a central compound for so many pathways. As a result enzymes such as CTP synthetase must respond readily to even small changes in the levels of ATP when they are regulated by the ATP level itself. On the other hand, GTP may fluctuate greatly from time to time in the organism's life cycle because it plays a role in different but less vital pathways. If a fairly constant level of CTP-synthetase activity is needed throughout these fluctuations, the negative cooperativity of GTP will ensure this desensitization to fluctuations in GTP levels. Of course, complete desensitization would allow the production of CTP in the absence of GTP, which would be a wasteful operation because both are needed for the synthesis of RNA. The protein ensures GTP control by requiring it to serve as the activator but eliminates excessive sensitivity through the device of negative cooperativity.

Another advantage of these induced conformational changes arises from the sequence of steps on an enzyme surface. High-energy intermediates are frequently formed in chemical syntheses, but if they are not isolated from water or other reactive substances, they decompose in side reactions that lower the yield of the reaction. Induced conformational changes can enable one step in the reaction to trigger the next step, which in turn triggers the third step and so on. In this way the high-energy intermediates exist for only brief intervals and are nestled in the protective harbor of the active site during the chemical changes. Wasteful side reactions are prevented, and the characteristically high yields of enzymatic reactions are achieved. In short, the induced-fit conformational change both explains the anomaly of a required order in the binding of substrates and provides a reason for it.

It therefore seems likely that flexible enzymes arose early in evolutionary time because of catalytic needs and certain specificity requirements. One of the greatest of these needs was the exclusion of water from some reactions. The cooperative property probably developed later because of the multisubunit structure of the enzyme; cooperativity survived because it had a useful function of its own: the amplification of some responses and the damping of other responses.

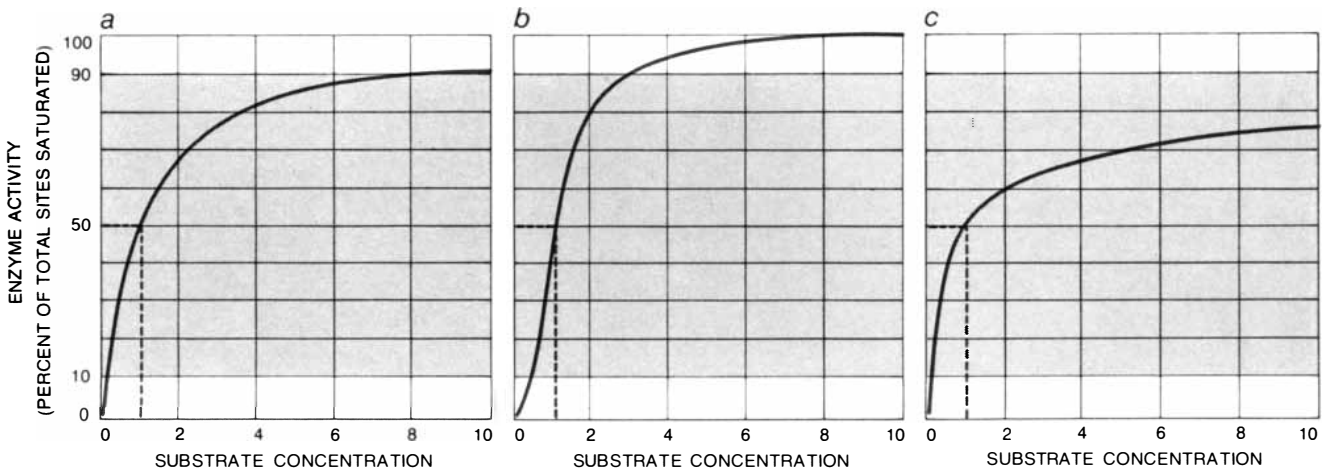
It is important to note that the parallelism between activators and inhibitors (positive effectors and negative effectors) and positive and negative cooperativity should not obscure their different and complementary roles. An activator is designed to turn an enzyme on and an inhibitor is designed to turn it off. Cooperativity is designed to increase or decrease the sensitivity of the enzyme to the environmental fluctuations of these regulators. An activator may show either positive cooperativity, negative cooperativity or noncooperativity, and the same is true of a substrate or an inhibitor. GTP, for example, is an activator of enzyme action but shows negative cooperativity in its binding pattern. In a simple sense one might say that activation or inhibition defines the key role of the regulatory molecules, whereas cooperativity provides the fine tuning of the system.

Finally, it is important to emphasize that although this account of the development of the induced-fit theory of enzyme action concentrates on the particular approach taken in our laboratory, many other workers in both the U.S. and Europe have made outstanding contributions to our understanding of the role of protein shape in biological control. Moreover, several possible alternative explanations for some of the regulatory properties of proteins have been put forward.

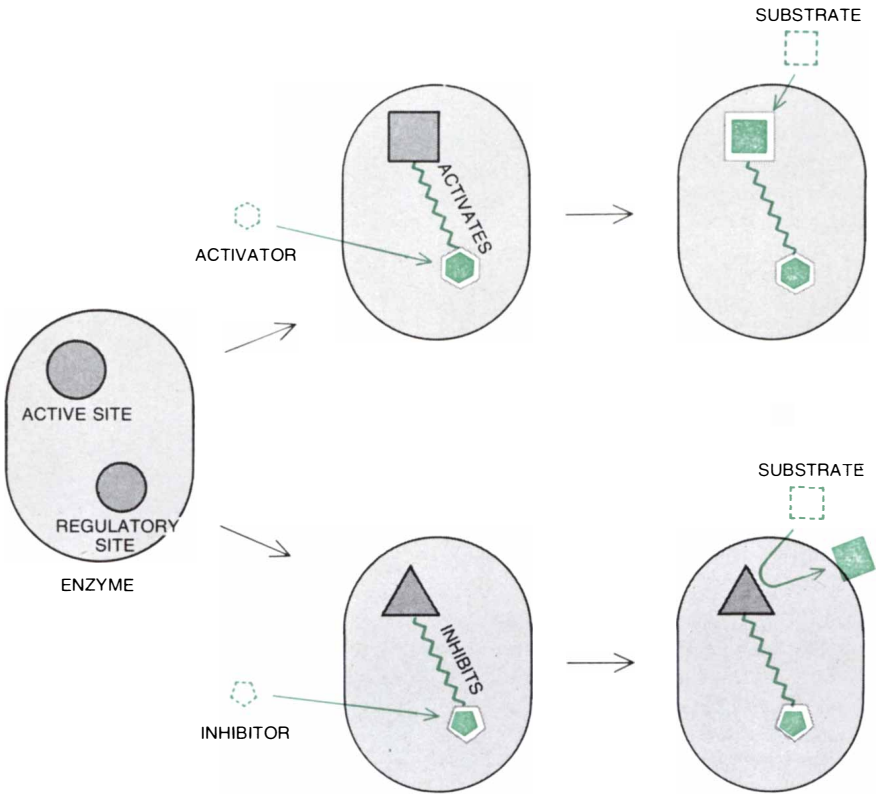
In particular Monod, Jeffries Wyman and Jean-Pierre Changeux at the Pasteur Institute have proposed an explanation of cooperativity that has quite different features from the induced-fit mechanism proposed above. It can be shown mathematically that their model can account for positive cooperativity and can also explain many features of activation and inhibition by effector molecules. Many features of their model appear to be present in the cooperative binding of oxygen to hemoglobin and in the properties of some enzymes; moreover, it has been of great value in clarifying the properties of regulatory proteins. Their approach cannot apply to all enzymes, however, because negative cooperativity requires an induced-fit model. In addition, Sidney A. Bernhard and his coworkers at the University of Oregon have postulated that in some cases an asymmetry of the type observed in insulin crystals by Dorothy Crowfoot Hodgkin and her colleagues at the University of Oxford may be important in regulatory proteins. In spite of the usefulness of these alternatives in describing some properties of proteins, it would appear that the main mechanism used throughout nature in the control of biological processes is the induced alteration of protein structure I have described.

Triggering Events

I have so far mentioned two types of event that can trigger conformational changes: (1) the reaction of a substrate to form a new covalent bond with an amino acid side chain and (2) the binding of a regulatory molecule (a substrate, an inhibitor or an activator) to the surface of a protein without the formation of a covalent bond. Both of these events occur in the reaction of substrates and in regulatory processes. In the preceding examples I used only the binding of effectors such as GTP to illustrate regulation, but regulation by the formation of covalent bonds can also occur at regulatory sites. Phosphorylation of the enzyme phosphorylase by ATP, for example, can lock the enzyme into a new structure that has different properties of reactivity and sensitivity to regulatory compounds. Earl W. Sutherland, Jr., and his colleagues at the Vanderbilt University School of Medicine showed that this covalent change is controlled by the hormone adrenalin through its mes-



THREE TYPES OF COOPERATIVITY observed in the binding patterns of different enzyme-substrate combinations are represented by these curves, which relate the change in the concentration of a substrate to the change in the level of enzyme activity induced by the substrate. Curve *A* depicts the normal hyperbolic binding pattern exhibited by most enzymes; in this case an 81-fold increase in concentration of the substrate is needed to go from an activity level of 10 percent to one of 90 percent. Curve *B* shows the sigmoid, or *S*-shaped, binding pattern associated with an enzyme that exhibits positive cooperativity (the tendency for the first molecule bound to make it easier for the next molecule to bind, and so forth); in this case only a ninefold change in substrate concentration is needed to go from an activity level of 10 percent to one of 90 percent. Curve *C* shows the binding pattern associated with an enzyme that exhibits negative cooperativity (the tendency for the first molecule bound to make it more difficult for the second molecule to be bound, and so forth); in this case the curve looks somewhat like a hyperbola, but in reality it is not. The curve approaches the final saturation state so slowly that it does not even reach the 90 percent activity level in this graph. Actually an increase in substrate concentration of 6,541-fold would be needed for such an enzyme to go from an activity level of 10 percent to one of 90 percent. To clarify relations between types of cooperativity all three curves are presented so that the situation at which half of the sites are occupied corresponds to a substrate concentration of 1. Similar curves can be obtained for activator and inhibitor molecules.



SIMPLIFIED SCHEME is presented here for illustrating the mechanism by which enzyme activity is controlled by the induced conformational changes brought about by small regulatory molecules that are not themselves involved in the primary enzymatic reaction. The complex polypeptide chain that comprises the actual enzyme is represented in this convention by the light gray area; the darker gray circles symbolize the shapes of the enzyme's active site and regulatory site in the absence of any bound molecules. The binding of the activator molecule (*colored hexagon*) induces a conformational change at the regulatory site (*symbolized by the change from a circle to a hexagon*) and also alters the enzyme's structure (*symbolized by the colored zigzag line*), which in turn changes the shape of the active site (*symbolized by the change from a circle to a square*). As a result the substrate (*colored square*) can now bind more easily than it can in the absence of the activator, since no further induced conformational change is needed for binding. In the reverse process the inhibitor molecule (*colored pentagon*) induces a different type of conformational change in the regulatory site (*symbolized by the change from a circle to a pentagon*), which is transmitted through the enzyme's structure, so that the resulting alteration at the active site (*symbolized by the change from a circle to a triangle*) makes the active site repel the substrate.

senger, cyclic AMP. Similar covalent changes are caused by the adenylation of the enzyme glutamine synthetase, as shown by Earl R. Stadtman, Helmut Holzer and their co-workers at the National Heart and Lung Institute. These covalent changes can be used to activate or to inhibit the enzyme. In such instances the reversal to the original enzyme is achieved by a second enzyme, which breaks the covalent bond of the regulatory molecule.

Why should nature use two devices, covalent changes and noncovalent binding, to induce the same kind of shape change for regulation? The answer is not known with certainty, but the suspicion arises that time is one factor that is involved. If an animal is frightened and needs large amounts of energy for a short period of time, it would be desirable to have an override mechanism that would ensure the high activity of some crucial enzymes until the crisis is past. The phosphorylation of phosphorylase induced by adrenalin would seem to provide such a mechanism. The formation of the covalent phosphoryl-enzyme bond converts the enzyme to a more active shape. When the crisis is over, the phosphate group can be removed by a second enzymatic process to regenerate the original, more placid enzymes, but for a time the normal instantaneous controls are eliminated to mobilize glucose for the crisis.

In a similar fashion, conformational changes that occur by the aggregation of subunits may also play an important role in regulation. In many instances it is found that a protein is active as a monomer but inactive as a dimer, or inactive as a dimer and active as a tetramer. The association of the polypeptide chains with one another causes a shape change with a resulting alteration in the activity of the protein. In essence, then, one subunit becomes the regulator of the other. If that is true, compounds that cause shape changes directly should also induce indirect changes in the polymeric structure of proteins. Indeed, some compounds induce changes in shape that cause them to dissociate. Such association-dissociation reactions appear to be important for many enzymes in metabolic pathways. The types of shape change in these associations and dissociations are exactly analogous to those induced by small molecules. It is also quite possible that certain hormones act as polypeptide regulators in much the same way that small molecules or subunits act in other instances.

Shape changes in proteins can occur very rapidly, in some cases in only a billionth of a second. Other shape changes occur in the millisecond range and have been shown to limit the rate of some enzymatic reactions. Still others take minutes to be effected; they are largely involved in association-dissociation reactions. This range of speeds serves the purposes of regulation well because some processes require an instantaneous correction in response to a stimulus, whereas a slower response is required to prevent an overshoot or to "lock in" a response until a crisis is passed. Furthermore, a slow conformational change has recently been suggested as a memory device in bacterial chemotaxis, in which a bacterium moves in response to a gradient in the concentration of a chemical substance. Such a change may have its counterpart in the neural systems of animals. Shape changes, in short, may occur over many time intervals from very fast to very slow, and each may have its usefulness.

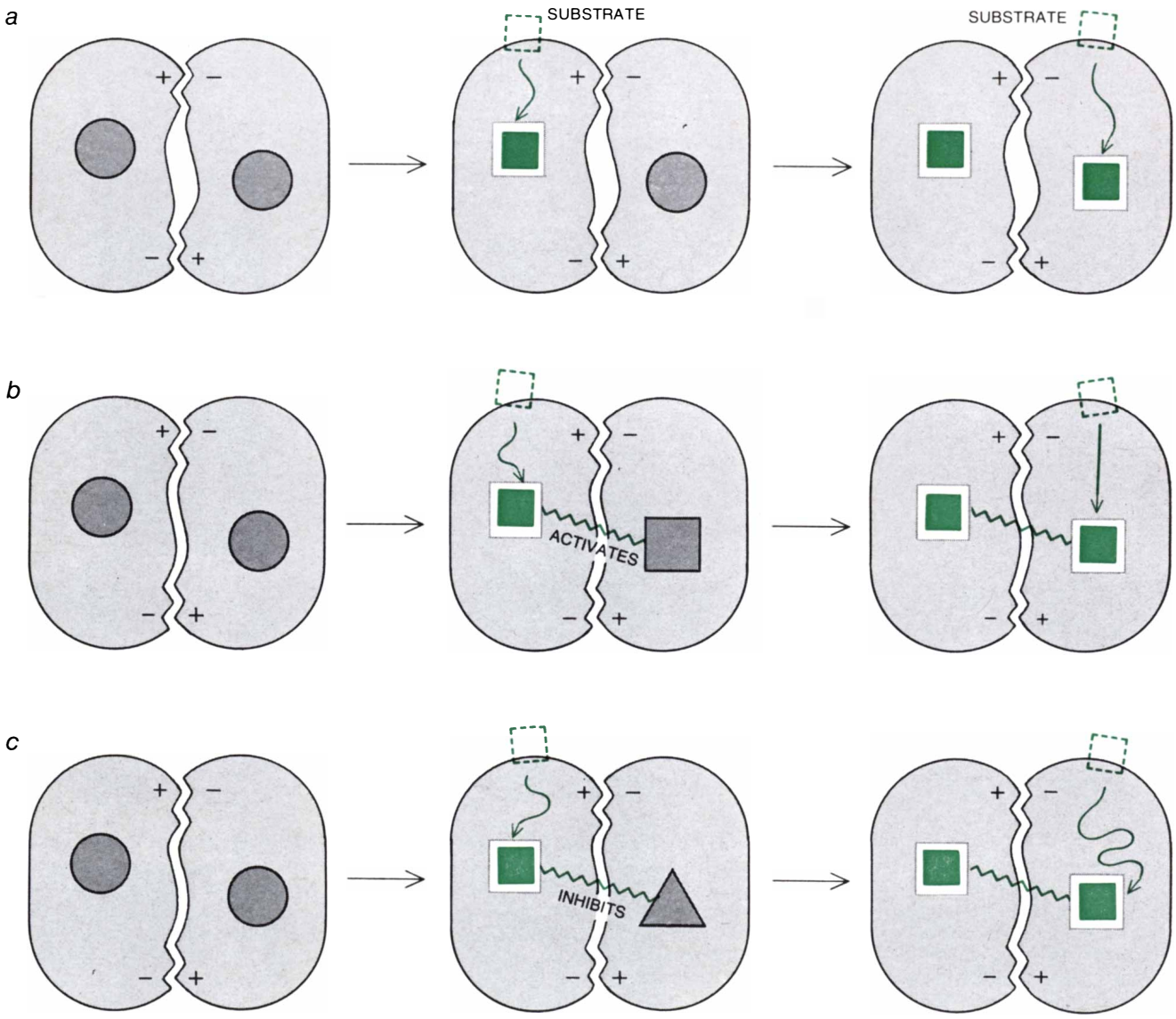
Application to Other Systems

Up to this point I have emphasized the regulatory control of enzymes because they are the regulatory proteins that have been studied the most intensively and they are readily available in the laboratory. As biochemistry has progressed, however, other molecules that have key regulatory roles in biological systems have been isolated. Receptor molecules involved in sensory systems have been shown to be similar to enzymes in terms of structure and binding properties. These molecules have the specificity characteristic of enzymes, and it is generally believed, but not absolute-

ly proved, that induced conformational changes are the signals that trigger the sensory impulse. When we smell or taste a compound, the compound induces a change in the shape of the receptor molecule, which triggers a response in our nervous system. Light induces a change in the shape of a protein in our eyes. Sensory phenomena are the highest form of regulation, and the brain is the ultimate regulatory apparatus of the most complex biological system: man.

Similarly, molecules that control protein synthesis have been isolated. Repressor molecules, for example, bind to DNA and prevent the reading of the genetic message unless they are removed by inducers. Again it is generally believed, although not yet absolutely proved, that the inducer causes a conformational change that peels the repressor molecule off the DNA, thereby initiating protein synthesis. Conformational changes are also believed to be crucial in promoter molecules, such as the cyclic-AMP-binding protein, which binds to DNA and aids in the initiation of protein synthesis. Such initiation and control of the reading of selected portions of the DNA allow different proteins to be made according to circumstances. A control of this type has been postulated to be a mechanism of differentiation, the reason nerve cells have one mixture of proteins and muscle cells have another. Differentiation, the key process that allows multicellular organisms to have specialized functions, is thus also dependent on changes in protein shape.

Antibody molecules, which protect us against invasion by foreign substances, induce a series of reactions in the apparatus called the complement-fixation system. That system is activated to destroy harmful cells and proteins by digesting them. Moreover, the presence of a foreign protein, such as a diphtheria toxin, can induce antibody-producing cells to multiply rapidly so that antibody to that specific harmful agent can be generated. This selective inducing of certain types of protective behavior is another regulatory device of living systems and is again believed to be caused by induced conformational changes of antibody molecules or antibody-receptor molecules. The transport of foodstuffs across cell walls is also thought to be triggered by induced changes in the

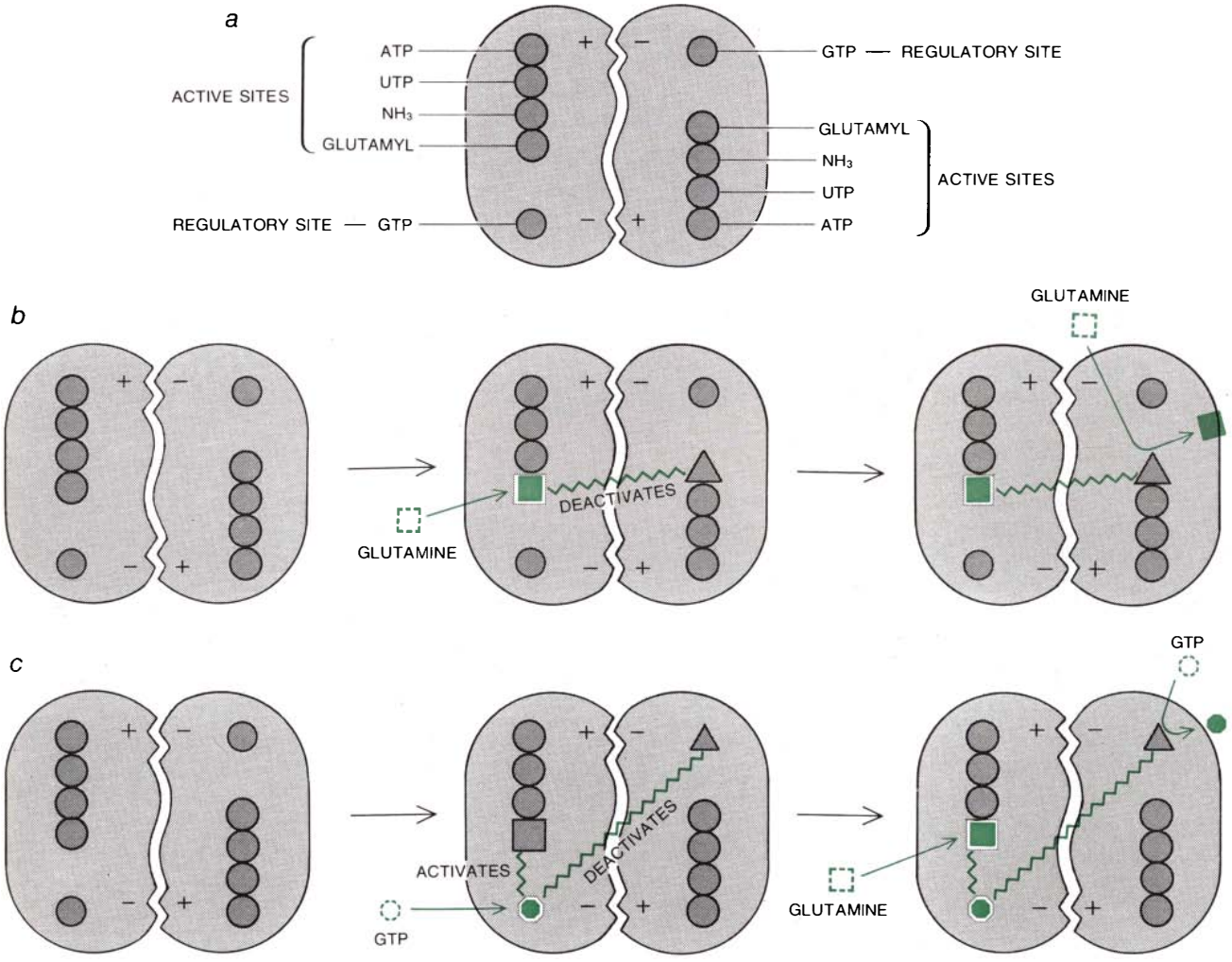


COOPERATIVITY IS EXPLAINED according to the induced-fit model of enzyme action in these diagrams, which use the simplified schematic convention introduced in the illustration on the opposite page. The type of protein represented here is the simplest known multisubunit enzyme: a dimer (that is, a two-peptide polymer) made up of two identical subunits. The three different enzymes shown exhibit respectively noncooperativity (*top row*), positive cooperativity (*middle row*) and negative cooperativity (*bottom row*).

shapes of proteins, as is the contraction of muscle.

There is no single regulatory process in the living system any more than there is a single type of control in a big city. The controls in the city are mediated by traffic lights, judges, payrolls, telephones, foremen, mayors and so on. The controls in biological systems are mediated by enzymes, antibodies, receptors, repressors and so on. Each of these protein molecules is designed differently and therefore each has a special history, giving rise to enormous diversity. Nevertheless, out of all this complexity comes a great simplicity. The flexibility of proteins and the shape changes they can undergo are the key features of regulatory control. Because the protein can exist in more than one shape and because the shape can be altered by external agents, the living system can respond to external stimuli and protect itself against environmental changes.

Nature thus places in the living cell the most powerful catalysts known and tames them to obey commands by subtle shifts in their structure, shifts measured in angstroms. These changes in shape can be imagined in a gross way as the alteration of a glove by the insertion of a hand, but on closer examination they seem much more like the delicate responses of a spider web designed with exquisite balance and interweaving. The delicate web we call a protein can be altered in shape by subtle perturbations, and through these perturbations functions can be turned on and off. As a result we feel like eating when we need food and lose our appetite when we are full. We can control the use of energy and the growth of specialized tissues, protect ourselves against hostile invading substances and develop a brain. All these essential regulations rest on the ability of a protein molecule to bend flexibly from one shape to another.



HIGHLY SPECIFIC NATURE of the changes induced in the structure and action of proteins as a result of the binding of selected substrate and regulator molecules is demonstrated in these schematic diagrams, which portray the behavior of the dimer of CTP synthetase, a key enzyme in the metabolism of nucleic acids. Experiments conducted in the author's laboratory have demonstrated the existence of specific binding sites on the surface of this enzyme for four substrate molecules and one regulator molecule (*a*). When the amino acid glutamine binds to the glutamyl site on one of the enzyme's subunits (*b*), it forms a glutamyl-enzyme intermediate. The conformational changes induced in this subunit in turn cause changes in the shape of the glutamyl site on the other subunit, making it less reactive. Since only one subunit of every dimer can engage in this reaction, the process is sometimes referred to as the "half of the sites" phenomenon. In contrast the binding of the regulatory molecule, GTP (guanosine triphosphate), to one of the subunits induces a conformational change that makes the glutamyl site of the subunit react more readily with glutamine at the same time that it causes the GTP site in the other subunit to assume a new shape that has less affinity for the second GTP molecule (*c*). In this case the design of the protein enables GTP to act as an activator of a catalytic reaction and an inhibitor of further GTP binding. The binding of both glutamine and GTP causes negligible changes in the shapes of the ATP, UTP or NH_3 (ammonia) sites.

Tomorrow morning, over 80 million people will be working to clean up this neighborhood.

There are over 80 million working people in the U.S. today.

And every day that they go to their jobs, they're helping to solve the problems that beset us. Urban decay, drug abuse, pollution, hunger, poverty, crime.

How? By paying taxes.

To some people, business is the cause of our nation's ills.

Why do we make this obvious point? To make another point: there are many people who simply do not realize that the original source of all funds for change for the better is business. Profitable, growing businesses, like ours, that provide jobs and income for working Americans.

To some people, business is the cause of our nation's ills. We say it's the cure.

It is the taxes of the American worker and American business that sustain government at all levels and, thus, pay for improvement programs our society needs.

Like low-cost public housing, medicare, training of the hard-core unemployed, veterans' education, the fights against drug abuse and cancer, school lunch programs—to name a few.

The betterment of our society depends on business.

Our nation receives its momentum and sustenance—directly or indirectly—from business. Growing business.

No growing business, no money to deal with the growing problems that plague us today.

Taxes from workers and business sustain government at all levels.

So we've got to work together. All 80 million of us.

International Telephone and Telegraph Corporation, 320 Park Avenue, New York, N.Y. 10022.

ITT

SERVING PEOPLE AND NATIONS EVERYWHERE

The way back.

Surely years ago when life was simpler, man looked around him with a sharp sense of wonder at the beauty and variety of the world.

We see that same feeling born again in every child who studies a butterfly resting on a petal, or catches his breath at the brilliance of a raindrop.

What happens to that sense of wonder?

Perhaps we grow up too fast, become too pre-occupied, too tired, or too sophisticated to hold on to it.

If this is true, we at Polaroid believe we have found a way back.

On the following pages, you will learn about a photographic system so radically new, so remarkably simple, it can lead you to find beauty in things you scarcely noticed before.

Polaroid inve

It can reveal the world to you as you have never seen it before.

The SX-70 camera, closed.

Forget everything you've ever known about photography, instant or otherwise.

The SX-70 Land camera is here.

Slim, graceful, balancing lightly in your hand, this package of more than 200 transistors, elegantly wrapped in top-grain leather, scarcely hints at the wonders it can perform.

One motion and it's open, ready at a finger's touch to propel into your hands picture after picture of a world you will feel you

nts The SX-70.

The SX-70 camera, open.

have only half-seen, half-felt until now.

As you read on, remember this: Remarkable as the SX-70 is, what is important is not what *it* can do, but what it enables *you* to do.

Because now you can almost wish the picture and have it, because the photographs are of such piercing beauty they seem more real than life, the SX-70 can sharpen your senses and set your imagination free.

In seconds, you see the faint outlines of your picture.

Less than two seconds after you touch the SX-70's red electric button, whoosh! The picture is ejected from the front of the camera, and begins to develop before your eyes, even in broad daylight. There is nothing to peel, nothing to throw away.

The photograph is hard, dry, and exceptionally durable. It will time its own development. As if from nothingness, the picture begins to emerge.

Even after you have a beautiful picture, it keeps getting better!

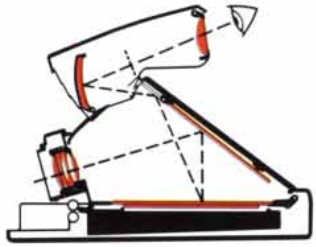
As the image blooms before your eyes, you realize this will be a photograph such as you have never seen, so startlingly real, you almost expect to hear your children speak to you.

Here is your picture, minutes later, fully developed.

It can reveal the world to you as you have never seen it before.

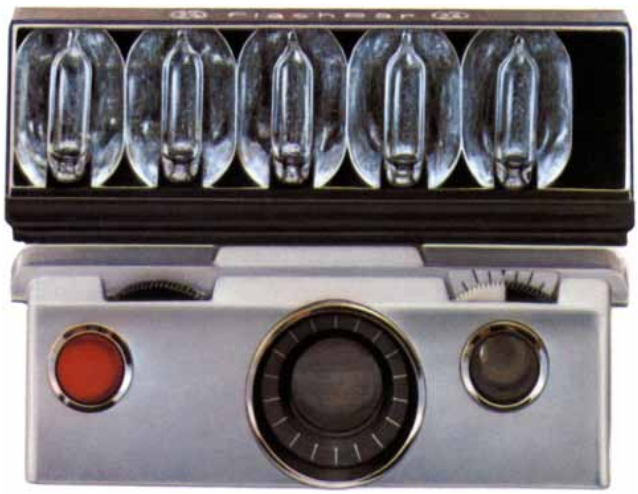
While the SX-70 moves more and more intimately into your life, you will soon learn to sense a picture, and own it. And you can own more and more little pieces of the world, as quickly as you can push the red electric button.

How The SX-70 makes

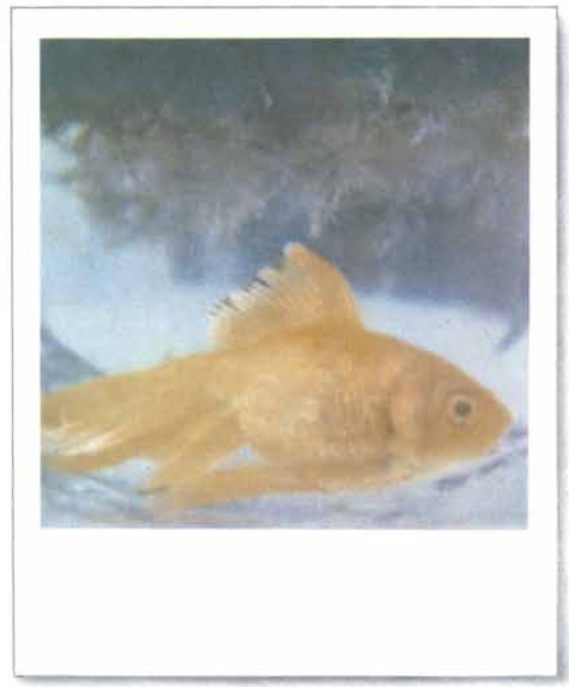


A remarkable computer-calculated complex of fixed and pivoting mirrors literally floats the image of the scene in midair to let your eye see what the camera lens sees.

A totally new brushed-chrome metal-plastic body material, a luxurious wrap of top-grain leather, and a form that reflects its remarkable capabilities, create an elegant, lightweight camera of striking design. Each slender 10-exposure film pack contains a wafer-thin battery, so you provide fresh power to operate the camera every time you load fresh film.



This rapid-fire, 10-shot GE FlashBar™ array lets you take flash shots from 10 inches to 20 feet or more away. And you can reshoot every 1.5 seconds to get a full action sequence. The SX-70 electronically picks the next flash to be fired. The aperture is set automatically as you focus, to let in the correct amount of light. Outdoors, an electric eye reads the light and programs the shutter for correct exposure.

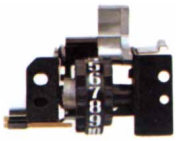


The film appears magically, outside the camera 1.5 seconds after you press the button. The picture, with an image area of 3⅛ x 3⅛ inches, is hard, dry, shiny and flat. There is nothing to peel apart, nothing to throw away. It times its own development. Automatically. You have nothing to do but watch and enjoy it.

the impossible possible.

Here are the makings for 30 flash pictures. Slip these in one pocket, the 24-ounce camera that's only 1″ x 4″ x 7″ in another, and you're ready for a wedding, a family reunion, a big party.

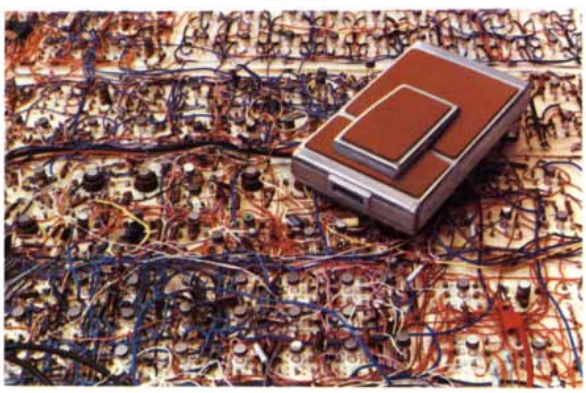
In minutes the picture develops fully before your eyes, *even in the brightest daylight*. An opaque chemical, spread inside the film, blocks the light during development. As this light barrier becomes transparent you see the image that has been formed. The pictures are exceptionally durable. You can handle them, stack them, put them in your pocket.



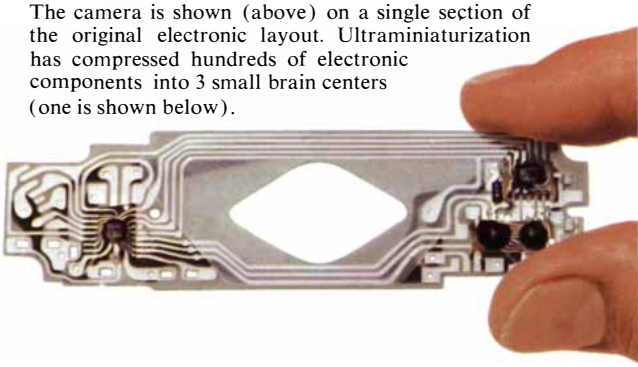
A tiny counter is located on the back of the camera. Insert a film pack and the counter reads "10". After each shot, that number decreases, to tell you how many pictures you have left.

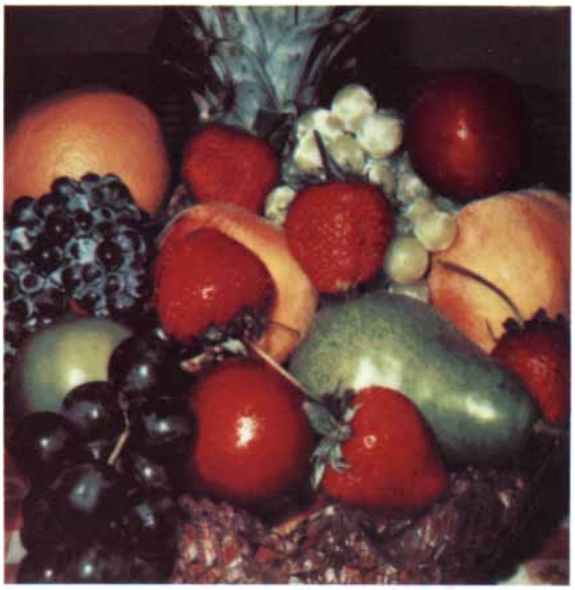
You can get close enough to count eyelashes with this computer-calculated 4-element lens system. No camera in the history of photography ever had one like it. As you focus, only the front element moves (never more than ¼ inch) yet you can focus on objects from 10.2 inches to miles away.

Once you press the electric shutter button, the rest is out of your hands. Literally. In 1.5 seconds, this remarkable 12,000 r.p.m. motor has propelled the picture out of the camera. You just keep composing pictures and pressing the button.



The SX-70 probably includes more new technology than any other consumer product in the world today. The camera is shown (above) on a single section of the original electronic layout. Ultraminiaturization has compressed hundreds of electronic components into 3 small brain centers (one is shown below).





How can a picture of something be more beautiful than the object itself?

There are sound technical reasons.

The SX-70 can search out beauty your eye alone can only guess at, because the camera can actually record far more detail than you can see without a magnifying glass.

Because the pictures are framed against a highly reflective chemical background, they have a remarkable luminous quality, as if lit from behind. The photographs seem almost three-dimensional.

The dyes themselves are totally new: a whole new spectrum of vibrant,

Pictures shown are actual size.

highly stable dyes that make possible prints of a brilliance and intensity that create a new standard for amateur photography. Notice the extraordinary reds in the rose and strawberries.

Incidentally, the picture of the rose was taken with an optional SX-70 close-up lens attachment which enables you to photograph small objects, actual size when you're only five inches away from your subject.

But to talk only of technical wonders is to miss what it's all about: the discovery of the world around you, in new ways that can stimulate and delight you.

The SX-70: why and when.

Perhaps once or twice in a lifetime, there comes an invention so radically new, it actually changes the way we live our lives.

Television was one.

We at Polaroid believe the SX-70 is another.

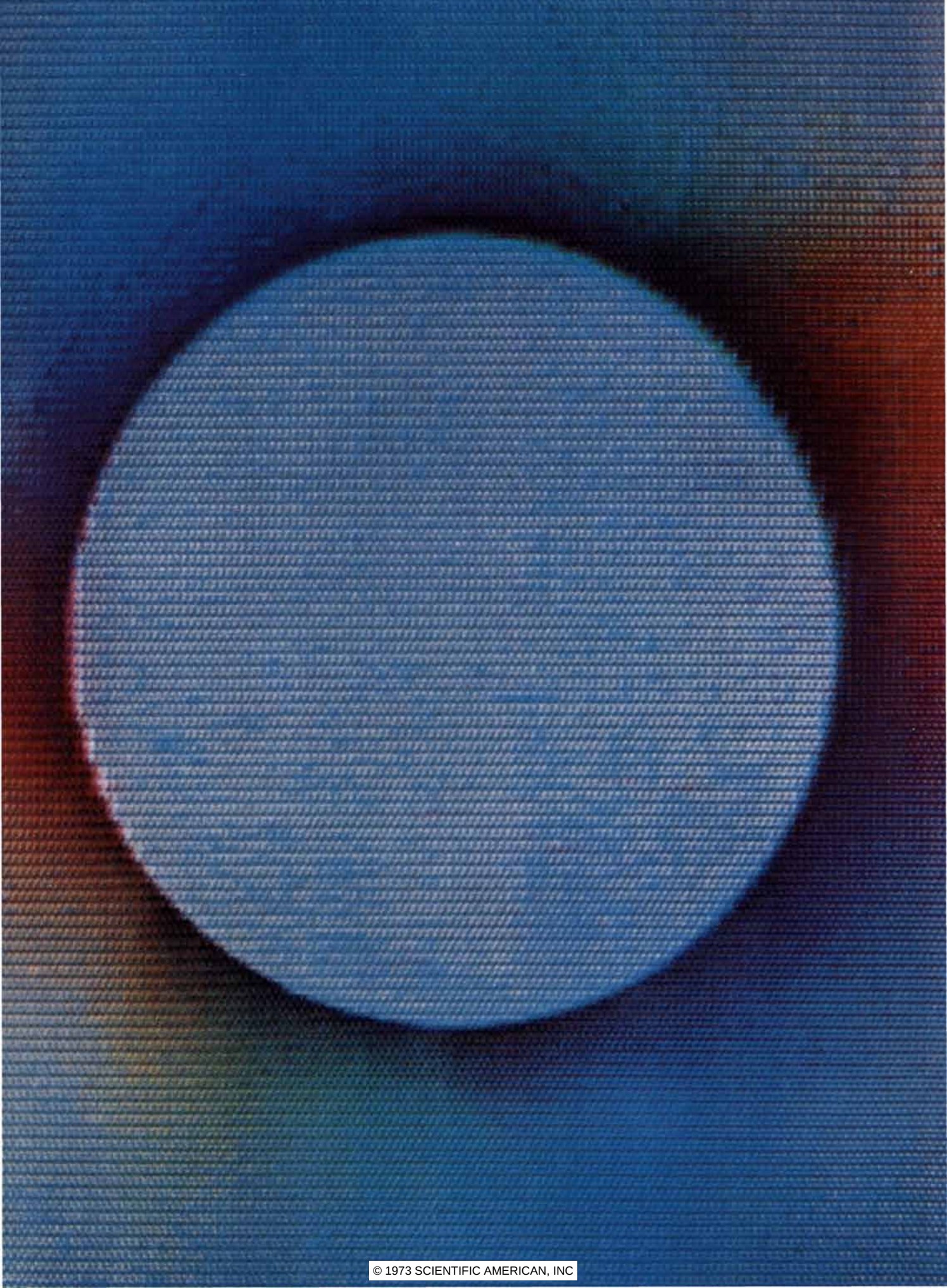
The virtual cascade of revolutions, mechanical, chemical, optical and electronic, that made the SX-70 possible had only one purpose: to free you from everything cumbersome and tedious about picture-taking, so that it could become at last, the simple creative act it should be.

Now that all you need to do is frame your picture, bring it into perfect focus and push a button, now that the picture is automatically delivered into your hand in less than two seconds, to time itself and develop into a photograph of a depth and brilliance unparalleled in amateur photography, what might once have seemed like a family duty or even just an interesting hobby, can become a spontaneous and constant pleasure in your daily life.

The SX-70, with a suggested list price of $180, is available now in limited quantity at your Polaroid Land camera dealer. We are increasing the supply as quickly as possible. Meanwhile, visit your dealer now to see a demonstration and to place an order for your own SX-70.

The SX-70 camera with optional leather carrying case.

Polaroid®



THE SOLAR CORONA

The physical nature of the sun's outer atmosphere is emerging from observations made not only on the ground during eclipses but also from rockets, from satellites and even from the moon

by Jay M. Pasachoff

The brilliant visible disk of the sun, the photosphere, obscures some of the sun's most interesting features. During a total eclipse of the sun, such as the unusually long one of this past summer, the photosphere is neatly masked and some of these features can then be seen. One of them is the chromosphere, the irregular level of the sun's atmosphere just above the photosphere. During a solar eclipse, for just a few seconds at the beginning and at the end of totality, the chromosphere is visible as a pinkish glow at the advancing edge of the moon. The red color results from the fact that the chromosphere emits strongly at the single wavelength in the hydrogen spectrum designated hydrogen-alpha. When even the glow of the chromosphere is extinguished by the moon, there emerges the pale splendor of the sun's outer atmosphere: the corona. Until this century the corona was as mysterious as it was awesome. In recent decades it has gradually been revealing its nature to telescopes and computers on the ground and to instruments in rockets, in satellites and even on the moon.

The corona is a million times fainter than the photosphere. Ordinarily it is concealed from observers on the earth by the everyday blue sky. In principle one should be able to see it simply by blocking out the sun's disk with one's thumb. In reality the light scattered in the sky near the sun is still too intense for the corona to be seen. It is fortunate for astronomers that they have a lunar thumb. The moon is 400 times smaller than the sun, but it is also 400 times closer to the earth. It therefore has almost exactly the same angular diameter in the sky as the solar photosphere: about half a degree.

The moon's orbit is nearly in the plane of the ecliptic (the plane of the earth's orbit around the sun). Each month it comes close to the sun in the sky; its far side (with respect to the earth) is lighted and its near side is dark. In this position it is called the new moon. The lunar orbit is tilted slightly from the ecliptic, so that the moon usually passes above or below the sun. Two to five times a year, however, the moon passes in front of the sun. Sometimes the conical shadow of the moon falls slightly off to the side of the earth and all that is seen is a partial eclipse. At other times the moon is at a relatively distant point in its elliptical orbit around the earth and consequently has a smaller angular diameter than the sun. During such an eclipse an observer sees a ring of sunlight around the dark moon; the eclipse is called annular, after the Latin *annulus*, meaning ring.

As long as the smallest fragment of photosphere remains in view so much light is scattered by the earth's atmosphere that the corona remains hidden. When the moon is close enough, however, and is properly centered over the disk of the sun, the lunar shadow sweeps across the earth in a band as much as 170 miles wide and thousands of miles long. From any point in this path of totality the corona is visible for a period that can be as short as a split second or as long as the theoretical maximum of seven minutes 40 seconds, depending on the conditions of the individual eclipse and where in the path the observer is standing.

Astronomers have had to travel all over the world to carry their equipment to a favorable spot in the path of totality. An instrumental malfunction at the crucial time can spoil the results of months of preparation. The weather is always a major factor; many expeditions have been ruined by rain or overcast, or even by a stray cloud obscuring the sun at the wrong moment. Eclipses have attracted expeditions most recently to southern Mexico on March 7, 1970, to Canada and Alaska on July 10, 1972, and to Africa on June 30 of this year. From the information gathered on many such expeditions a consistent picture of the corona has emerged that is now generally accepted.

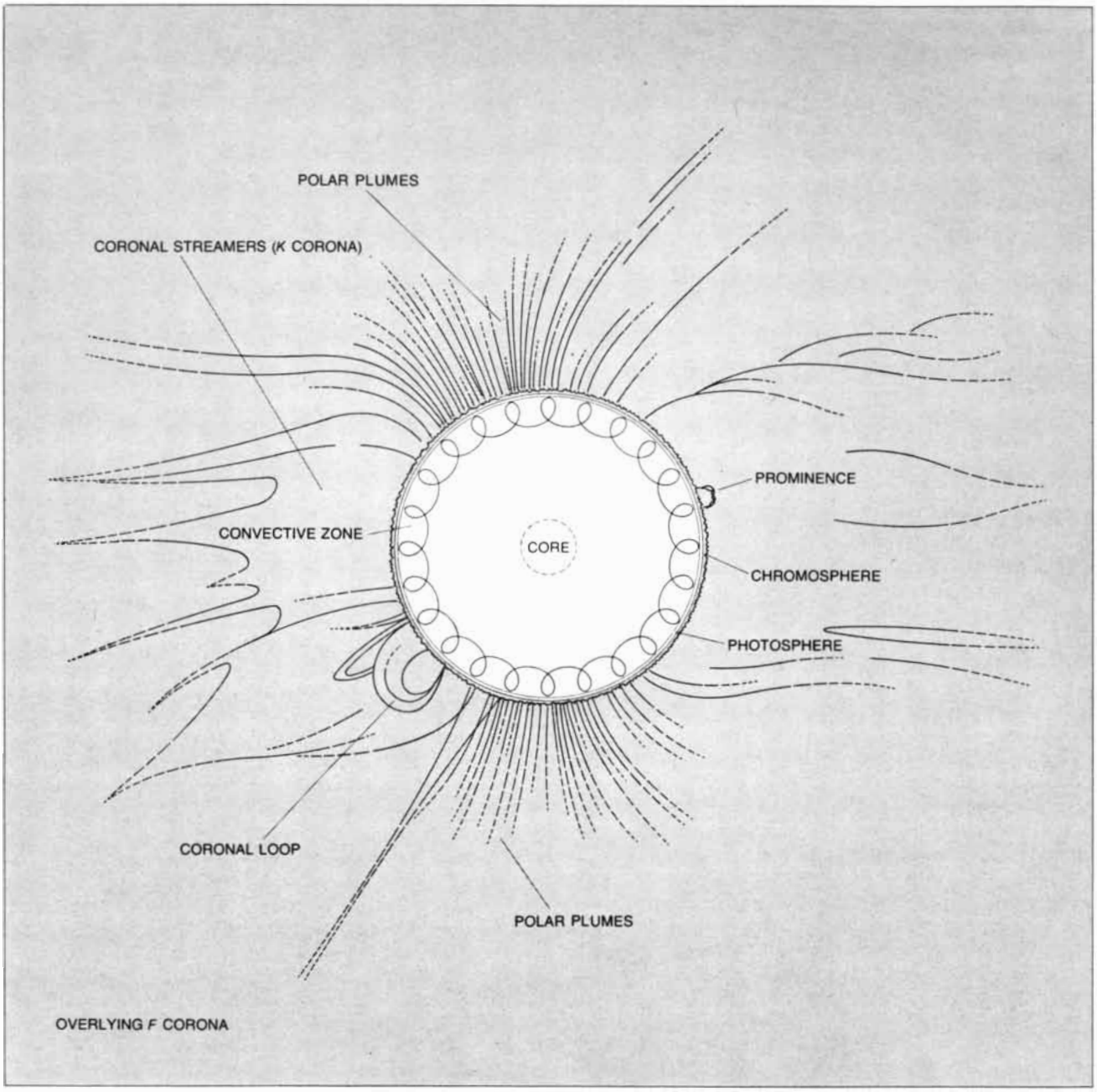
The corona that is seen during eclipses has three major components. The first, which is visible most clearly as the inner corona, extends about two solar radii out from the center of the sun: perhaps 1.4 million kilometers. It is made visible by photospheric light scattered by the electrons in the gas surrounding the sun. Its spectrum is continuous, that is, it does not show the dark absorption lines in its spectrum called Fraunhofer lines, which represent the wavelengths of light ab-

POLARIZATION OF THE CORONA is revealed on the opposite page in a color image synthesized by computer. Donald H. Menzel of the Harvard College Observatory and the author photographed the corona through polarizing filters during the total eclipse of the sun of March 7, 1970, at Miahuatlán in Mexico. The pictures were scanned photoelectrically and the information about the direction and amount of polarization from each one was stored in the computer memory. Lynn H. Quam and Robert B. Tucker of the Stanford Artificial Intelligence Laboratory used the computer to combine the information from the photographs and to generate a three-color display on a television screen. The hue corresponded to the direction of polarization and the saturation to the degree of polarization. The corona is polarized radially outward from the surface of the sun. On the synthesized image this is indicated by the fact that the colors of the corona are complementary at 90-degree intervals.

sorbed within the photosphere. This component of the corona is called the *K* corona, from the German *Kontinuum*.

At about two solar radii the *K* corona is less dominant, and radiation is added by sunlight scattered by the dust that is distributed through interplanetary space. In its spectrum this second component of the corona shows the dark Fraunhofer lines and is therefore called the *F* corona. The *K* corona continues out into space along with the *F* corona, although it gets fainter with increasing distance from the sun. It extends even past the orbit of the earth, although at such distances it is of course extremely thin. Thermal radiation from the dust has been observed in the far-infrared.

The third major component of the light of the corona is revealed by emission lines in the spectrum of highly ionized atoms near the sun: atoms stripped of as many as half of their constituent electrons. Although this radiation is very weak and makes a negligible contribution to the total intensity of the corona, it yields detailed information



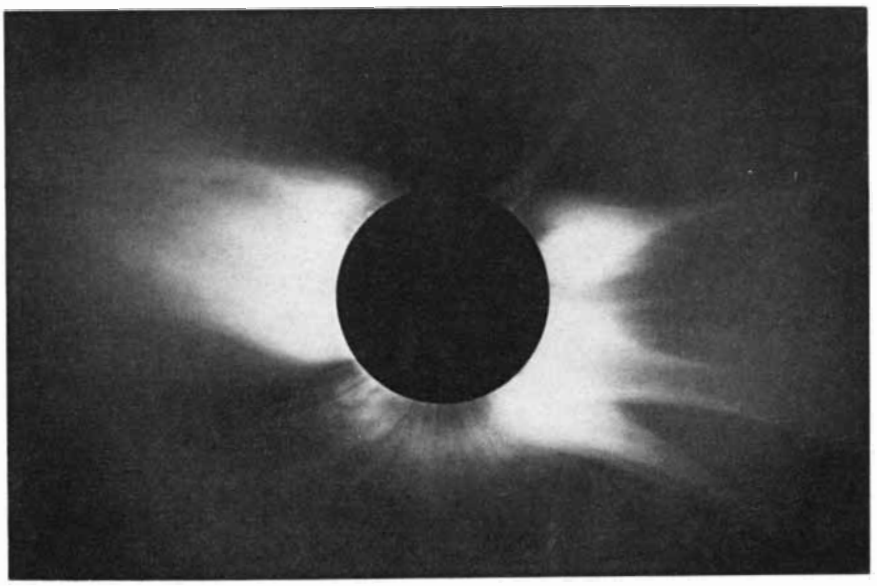
STRUCTURE OF THE SUN and the corona is summarized in this cross-sectional diagram. Energy from thermonuclear reactions in the core makes its way gradually to the sun's exterior layers. Most of the visible light received on the earth comes from the photosphere. Just under the photosphere is the convective zone. Shock waves from the convective zone carry energy up into the chromosphere and the corona. The temperature of the sun is at a minimum in the photosphere and lower chromosphere; the shock waves cause the temperature to rise through the upper chromosphere and the corona until it reaches some two million degrees Kelvin in the corona. From the corona the solar wind expands into interplanetary space. The visible corona has three major components. The first is the *K* corona: light scattered by the electrons in the gas surrounding the sun. The second is the *F* corona: light scattered by the interplanetary dust between the sun and the earth. The radiation of third component is emitted by highly ionized atoms near the sun.

about the physical conditions inside the corona.

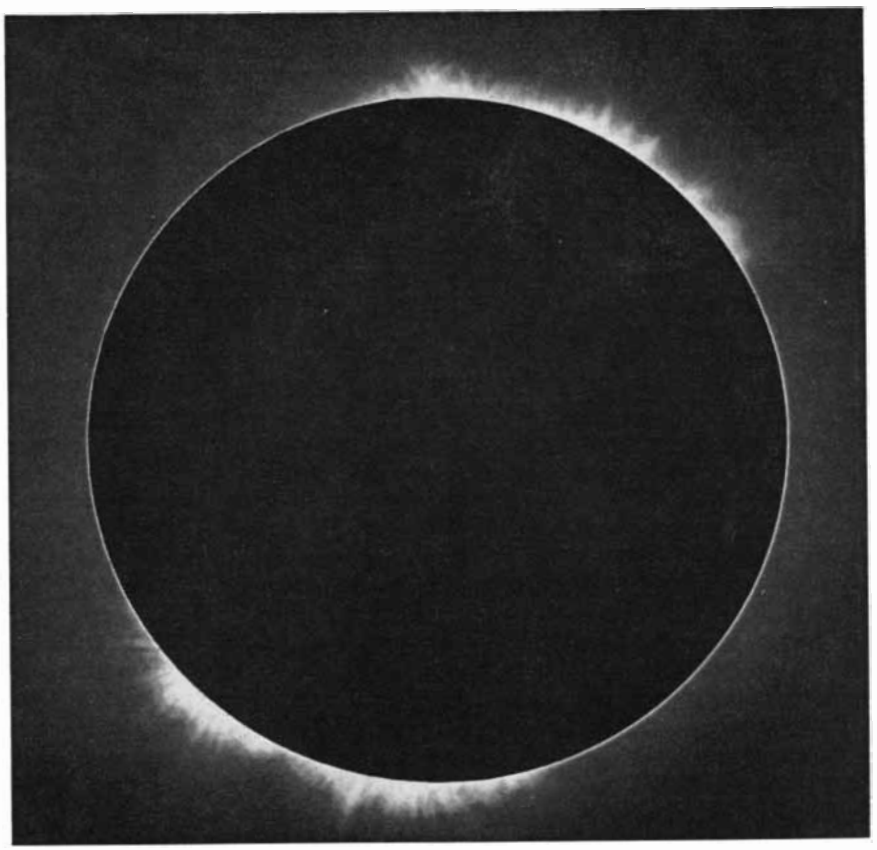
The corona has a temperature of about two million degrees Kelvin. Many lines of physical argument now lead to this result, although how the corona could have such a high temperature was one of the major questions in solar physics for many years. The continuous spectrum is a strong clue. Two million degrees is a kinetic temperature: it is an indication that the electrons are moving so fast that the wavelengths of photospheric light each one scatters is Doppler-shifted. The shifts are very large because of the high temperature and are completely random, so that they wash out all the absorption lines from the photospheric Fraunhofer spectrum. Historically this was the first reasoning that explained the continuous nature of the spectrum of the *K* corona. Only when the emission lines in the coronal spectrum were identified with the radiation from highly ionized atoms, however, were solar astronomers convinced that the coronal temperature was indeed high.

More than 60 chemical elements have been detected in the sun. Their abundances in the corona are now known to be essentially the same as the abundances in the photosphere. Two of the strongest lines in the visible region of the spectrum are the green line at a wavelength of 5,303 angstroms and the red line at 6,374 angstroms. The first line is emitted by iron atoms stripped of 13 of their 26 electrons by the extreme temperatures; it is designated Fe XIV. The second line is from iron atoms stripped of nine electrons; it is Fe X. (The un-ionized species of an element is labeled with the Roman numeral I, the first ionized state by II, and so on.) Two emission lines in the near-infrared are stronger: the lines of Fe XIII at 10,747 and 10,798 angstroms. These lines, however, lie in a relatively unexplored spectral region where photographic film is not sensitive. During the eclipse in June, J. Philip Schierer of Tektronix, Inc., and I used a new electronic spectrometer to observe these lines. Although my results are still preliminary, they do show the value of using devices such as silicon-diode Vidicons for the near-infrared.

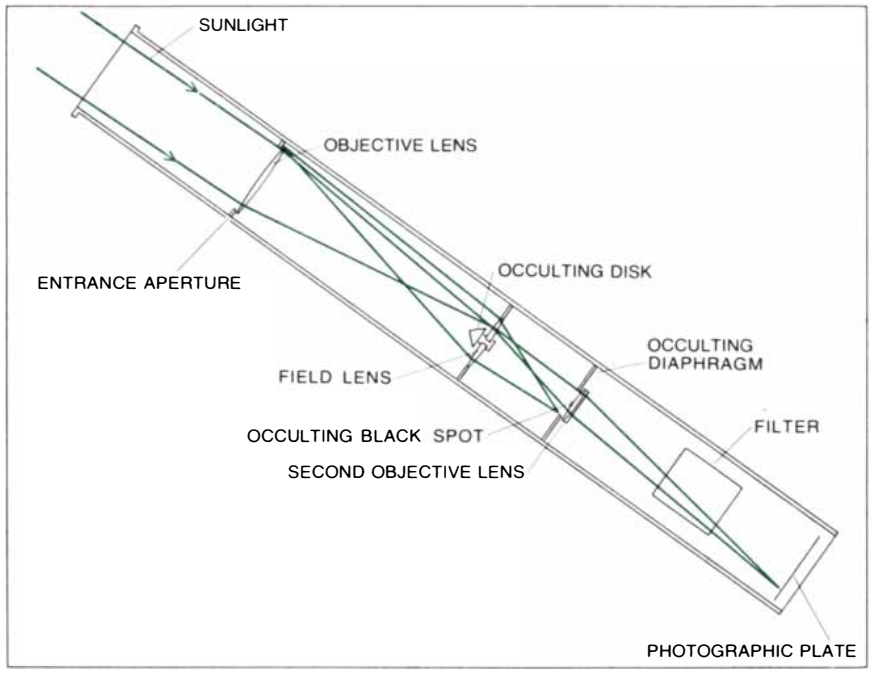
In order to release many electrons from their atomic bonds the gas must be very hot indeed. The emission lines also indicate the electron temperature of the corona in another way. They are much broader than the absorption lines from the photosphere, which are formed at an average temperature of a mere 6,000 de-



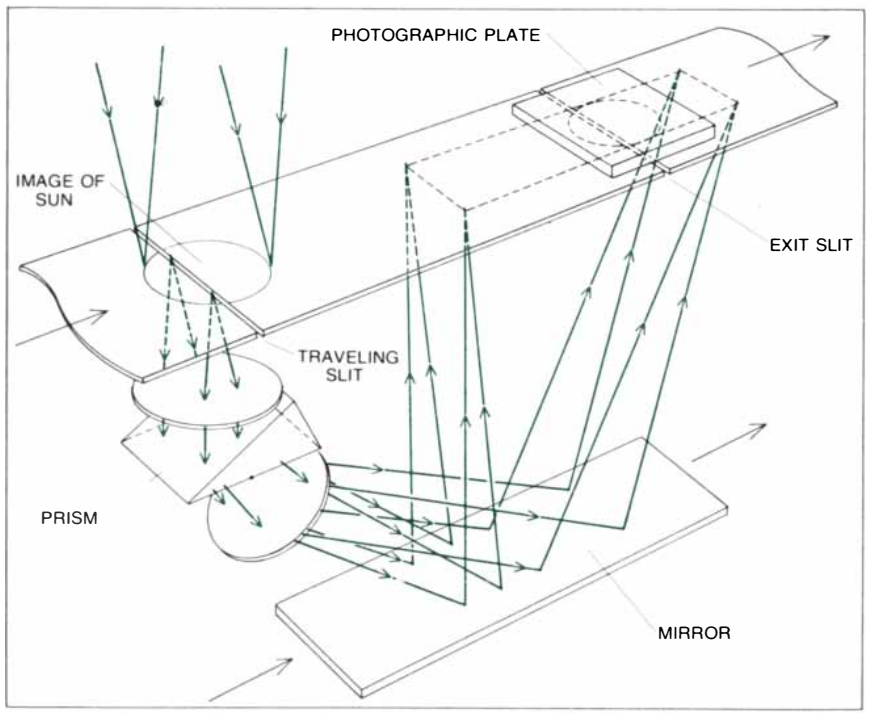
ENTIRE CORONA was photographed during the total eclipse of June 30 from Loiyengalani in Kenya by John W. Firor, Joseph H. Rush, Carl G. Lilliequist, Leon B. Lacey and Howard K. Hull of the High Altitude Observatory of Boulder, Colo. The photograph was taken in red light at a wavelength of 6,500 angstroms through a neutral-density filter whose density decreased radially outward from its center. Such a filter suppresses the bright inner corona in order to show the faint outer streamers in the same photograph. Intricate details of the corona are thus made visible. The circular halo near the ends of the streamers was caused by the filter's not evenly suppressing the light from the unusually bright sky during the eclipse.



INNER CORONA was photographed on March 8, 1970, the day after a total eclipse, with a coronagraph at the Mees Solar Laboratory of the University of Hawaii on the rim of Haleakala Crater on Maui. The picture is actually a composite of six images printed together to reduce the "noise" in each individual image. Photograph was made in green light of "forbidden" emission line of 13-times-ionized iron at a wavelength of 5,303 angstroms.



CORONAGRAPH allows the inner corona of the sun to be photographed and analyzed without an eclipse. Sunlight is focused on an occulting disk, which blocks out the bright photosphere. A diaphragm removes a bright ring of light diffracted by the periphery of the entrance aperture; a black spot at the center of the second objective lens eliminates the bright spot that forms in the center of the image from internal reflections within the main lens. A second objective lens focuses the image of the corona on a filter, which isolates the spectral region to be photographed. Light is coming from only one limb of sun and is thus off axis.



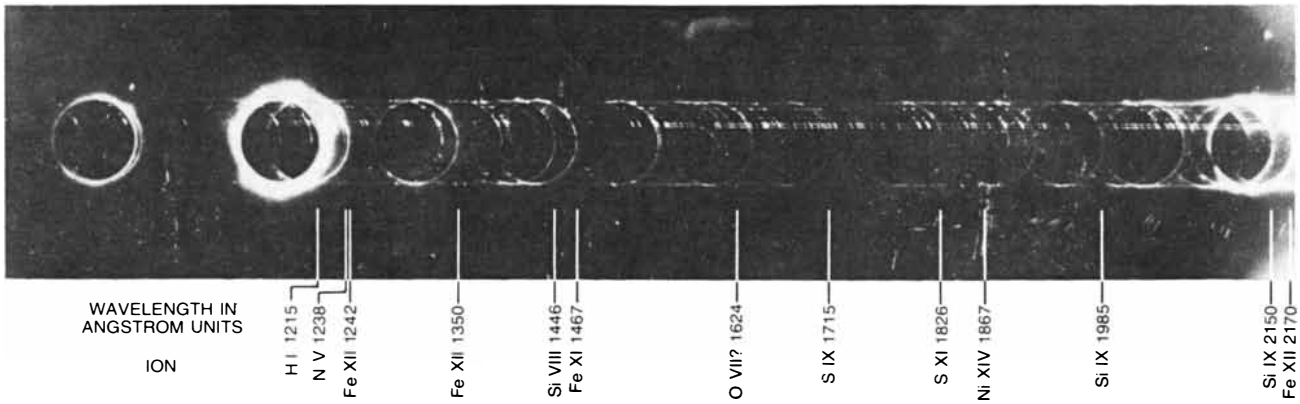
SPECTROHELIOGRAPH photographs the sun in one particular wavelength. Light from the sun enters a traveling slit and is dispersed into its spectrum of wavelengths by either a prism spectrograph (*as shown here*) or with a diffraction grating. One wavelength is isolated by a second slit. As the first slit scans the sun to build up an image, the second slit travels across the photographic plate with the same motion, so that the picture is built up in one narrow spectral region. In the case of spectroheliographs aboard the Orbiting Solar Observatory (OSO) satellites pictures such as the one on the cover of this issue of SCIENTIFIC AMERICAN are not directly photographed. Images are reconstructed from brightness measurements made photoelectrically, recorded on magnetic tape and telemetered to the ground.

grees. Here broader lines mean higher temperatures. All these lines are "forbidden" in ordinary conditions on the earth, meaning that the probability that radiation of those wavelengths will be emitted is very low. Solar forbidden lines result when the electron in an atom drops to a state of lower energy from a state of higher energy that is metastable, or long-lived. The density of the corona is so low that the atoms are not deexcited by collisions before they have a chance to radiate the energy.

Radio astronomy yields an independent picture of the corona. As the sun is examined at progressively longer wavelengths progressively higher layers come into view. Thus at centimeter wavelengths the radiation is emitted by the chromosphere, and at meter wavelengths it is emitted by the corona. Indeed, as the sun is scanned at longer and longer radio wavelengths the size of the image increases. Spatial structure in the corona is studied on a second-by-second basis by a radioheliograph composed of 96 linked antennas each 45 feet in diameter at Culgoora in Australia. The structure of the corona can also be studied by radar: by bouncing radio waves of different lengths off it. Such observations, together with theoretical calculations, are in agreement in showing that the corona has a temperature of two million degrees K. and has active regions of even higher temperature.

The earth's atmosphere does not pass electromagnetic waves that are shorter than about 3,000 angstroms, which is in the ultraviolet region of the spectrum. Therefore instruments have been sent above the atmosphere to gather short-wavelength information about the corona. Richard Tousey of the Naval Research Laboratory, who has been making spectra of the sun for many years with apparatus carried aloft in rockets, was the pioneer. In the spectral region of the shortest ultraviolet wavelengths and X-ray wavelengths one can even see the basic "permitted" spectral lines of the corona and not just the forbidden lines.

More recently satellites in orbit around the earth have provided stable platforms for observing ultraviolet radiation from the chromosphere and the corona. Leo Goldberg, who was then at the Harvard College Observatory, headed a group that flew an instrument that could function either as a spectrograph or as a spectroheliograph on two of the Orbiting Solar Observatories launched by the National Aeronautics and Space Administration: OSO-IV and OSO-VI. In the spectrograph mode the



SPECTRUM OF THE CORONA made with a slitless spectrograph aboard a rocket launched from Wallops Island, Va., during the total solar eclipse of March 7, 1970, showed 25 previously unknown spectral lines in the extreme ultraviolet. Thirteen of them are shown in this photograph. The spectrograph was not equipped with a slit because the image of the corona or chromosphere is so narrow that it acts as its own slit. Spectral lines from the chromosphere are short arcs, since they come from one limb of the sun at a time. Coronal lines are emitted higher above the sun's surface and show as complete circles. One particularly surprising result of this rocket flight was the discovery that one can see that there is un-ionized hydrogen within the corona. On the spectrum it shows up as the bright ring labeled *H* I. The experiment was a joint venture of the Harvard College Observatory, the Imperial College of Science and Technology, York University and the British Science Research Council's Astrophysics Research Unit at Culham Laboratory. Un-ionized species of an element is labeled with Roman numeral I, the first ionized state by II, and so on. *H* is hydrogen, *N* is nitrogen, *Fe* is iron, *Si* is silicon, *O* is oxygen, *S* is sulfur and *Ni* is nickel. Number after each species is the wavelength of the spectral line in angstroms.

instrument scanned the spectrum of the sun in the far-ultraviolet. The solar image was held steady while the angle of a diffraction grating was changed. The grating dispersed the radiation from the sun into its various wavelengths. In this way the different wavelengths of the spectrum could be made to fall in turn on a photomultiplier tube, which thus was able to record the intensity of the sun's radiation at each wavelength. In the spectroheliograph mode the instrument worked the opposite way. The spectroheliograms were made by holding the diffraction grating at one angle, so that only a single wavelength fell on the photomultiplier. The solar image was then scanned in a raster pattern to build up a picture of the sun in that one wavelength.

The resolution of the equipment on OSO-IV was one minute of arc. The resolution was improved to half a minute of arc on OSO-VI. Although that resolution does not match the resolution of one second of arc that can sometimes be achieved by instruments on the ground, much important information has been gained from the study of the far-ultraviolet region of the sun's spectrum from above the earth's atmosphere.

Each region of a particular temperature and density in the corona is characterized by emission from a particular ion. A sequence of spectroheliograms made at wavelengths corresponding to increasingly higher ionized states shows that greater heights within the corona are revealed at higher degrees of ionization. Thus it is possible to build up a three-dimensional model of the corona's structure.

An unusual and successful experiment was carried aloft in a rocket during the eclipse of March, 1970. It was prepared jointly by the Harvard College Observatory, the Imperial College of Science and Technology, York University and the British Science Research Council's Astrophysics Research Unit at Culham Laboratory. That group obtained spectra of the corona and of the chromosphere in the ultraviolet region from 977 to 2,200 angstroms and discovered 25 new coronal lines. Since that region of the spectrum is closed to observers on the ground, many of the spectral lines had not been analyzed. This line of coronal investigation is in its infancy.

One additional result of the March 1970 rocket flight was the surprising discovery that un-ionized hydrogen is visible within the corona. It was detected in the radiation of the ultraviolet spectral line designated Lyman alpha, which is emitted by a hydrogen atom when an electron drops to the ground state from the next higher energy level. The temperature of the corona is so extremely high that almost all of its hydrogen must be ionized. Nevertheless, the minute fraction that remains un-ionized—perhaps one atom in 10 million—can scatter enough radiation to be detected.

Another area of recent investigation has been the study of the sun at X-ray wavelengths between one and 100 angstroms, most notably from rockets and satellites by a group under Riccardo Giacconi at American Science and Engineering in Cambridge, Mass. Because X rays are too energetic and penetrating to be focused by ordinary telescope optics, Giacconi and his group have designed optics in which the rays are reflected at flat, grazing angles. With such apparatus they have succeeded in resolving detail to one second of arc. X rays are emitted mainly from "hot spots" in the corona high above active regions on the surface of the sun. In these hot spots a temperature of four million degrees K. is not uncommon, although some of the radiation is not simple thermal radiation and may reflect interactions of charged particles with the corona's magnetic field.

Like the evidence from the visible-light and radio regions of the spectrum, information from the ultraviolet and X-ray regions clearly indicates that the temperature of the corona is in the millions of degrees. It must be remembered, however, that this temperature does not imply that there is a great amount of energy present. The corona is very tenuous. There are only some 10^9 electrons per cubic centimeter at its base, and the number decreases rapidly with height. That density is 10 million times less than the electron density in the photosphere. Although each individual electron is moving at a high velocity and thus has a high kinetic energy, the fact that there are so few electrons means that the total energy of the corona is quite low.

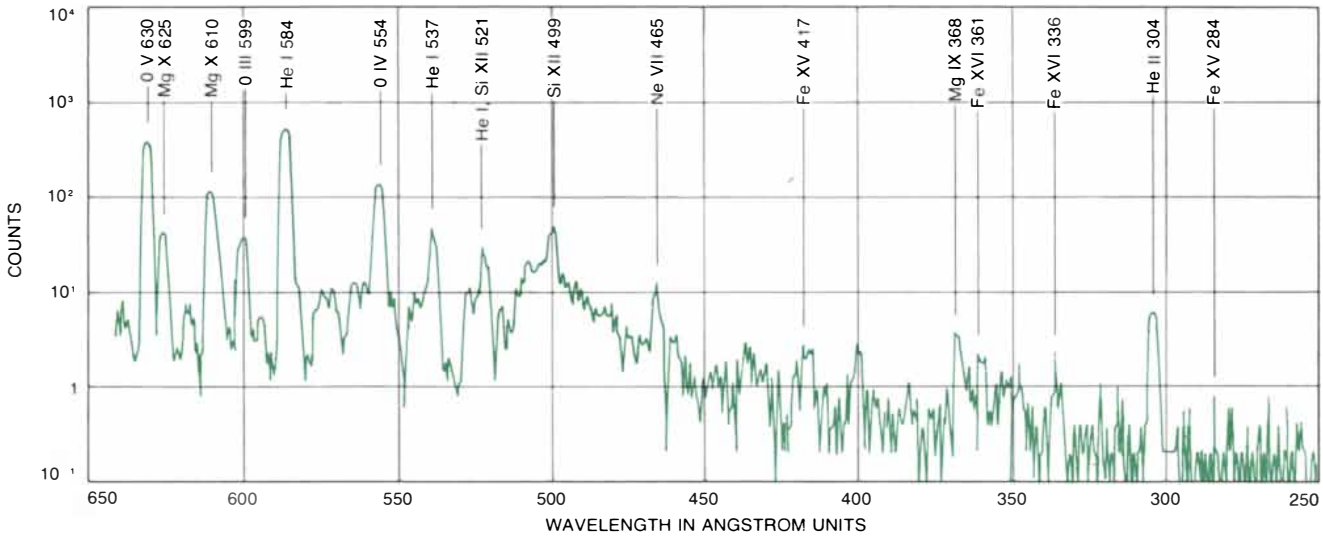
Nonetheless, the question is: Where do the individual electrons get their high kinetic energy? The energy wells up into

the corona as shock waves from lower levels of the sun, probably from below the photosphere in the convective zone [*see illustration on page 70*]. The temperature rises throughout the chromosphere and then jumps rapidly a million degrees in the few thousand kilometers immediately above the chromosphere. The exact mechanism by which the energy is transported to the corona and then dissipated is a matter of intensive research and some controversy.

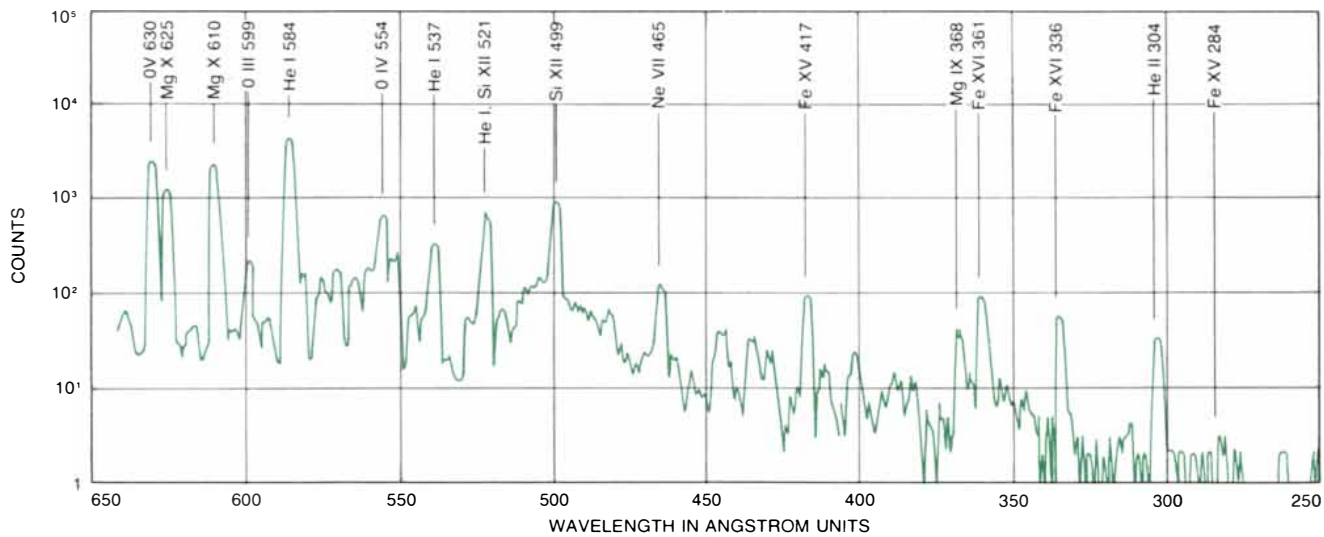
The emission lines from the highly ionized atoms make it possible to detect the corona even without a solar eclipse. Ground-based observations of these lines can be made from a few ideal sites. In principle the requirements for such observations are simple: the observer must have a very clean sky and very clean optics. In practice, however, these requirements are difficult to meet, and it was a major achievement when Bernard Lyot of France succeeded in constructing a coronagraph in the 1930's.

A coronagraph is usually a refracting telescope because microscopic pits in the mirrors of a reflecting telescope scatter too much light. At the focus of the primary lens of the telescope an image of the sun is formed on a small occulting disk that is exactly the size of the image of the photosphere. With the intense photospheric light blocked, there is hope of detecting the corona. The image of the corona is refocused and sent through filters that select the desired wavelengths for observation. Observations are usually made through filters that isolate the wavelengths around a coronal emission line, since such lines are brighter than the continuous spectrum.

Making a successful coronagraph is still something of an art. That fact is captured in an article on coronagraphs written by John W. Evans, director of the Sacramento Peak Observatory in Sunspot, N.M., in which he describes



SPECTRUM OF CORONA OVER QUIET REGION of the sun from 650 to 250 angstroms was measured from the Harvard College Observatory spectrometer aboard OSO-VI. From 504 to 470 angstroms the continuous spectrum of helium is visible as a gradual rise in intensity. *Mg* is magnesium, *He* is helium and *Ne* is neon. The principal investigators were Andrea K. Dupree, Martin C. E. Huber, Robert W. Noyes, William H. Parkinson, Edmond M. Reeves and George L. Withbroe of Harvard College Observatory.



SPECTRUM OF CORONA OVER ACTIVE REGION of the sun was measured from the Harvard spectrometer on OSO-VI in an area of sunspots and flares. The entire spectrum is about 10 times stronger than the quiet-region one at top. Also certain spectral lines, such as the line of nine-times-ionized magnesium at 625 angstroms, are enhanced with respect to the background continuous spectrum.

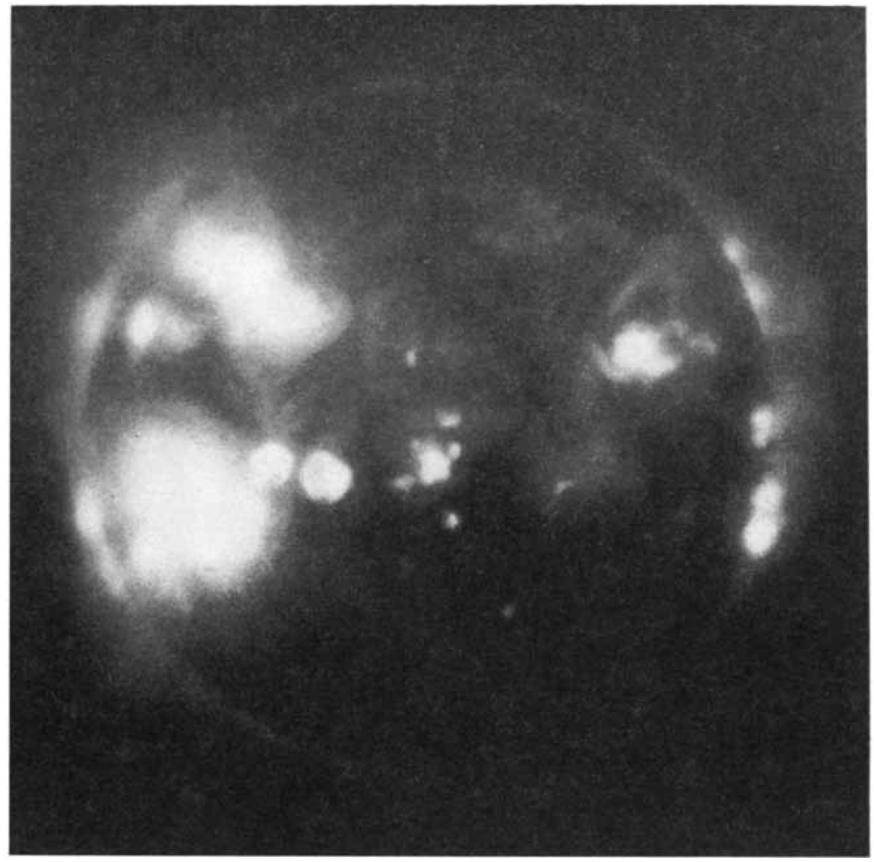
how to coat the primary lens with an almost infinitesimal amount of oil in order to minimize scattered light: "Touch the side of the nose with a forefinger and lightly touch the surface of the lens in two or three spots, leaving distinct finger marks. Now . . . spread the oil."

Sites suitable for coronagraph observations are rare. They must be on a high mountain above much of the earth's atmosphere and the sky must be exceptionally free of dust. One of the main sites in the U.S. is the Sacramento Peak Observatory, which is at an altitude of 2,811 meters (9,200 feet).

The Hawaiian Islands have become prime sites for coronagraphs because general meteorological conditions keep atmospheric dust below the high mountains. The University of Hawaii has developed an exceptional site at an altitude of 3,054 meters (10,000 feet) on the rim of Haleakala Crater on the island of Maui. There John T. Jefferies, Jack B. Zirker and Frank Q. Orrall form an active group of coronal investigators. The skies are so clean that the scattering of light by the atmosphere near the disk of the sun is commonly less than 10 millionths of the total intensity of the sun [*see illustration on next page*]. Observations of the corona can be carried out perhaps 275 days of the year. The High Altitude Observatory of Boulder, Colo., also has a coronal site in Hawaii, on the mountain of Mauna Loa on the main island. Other major coronagraph sites around the world are at Arosa in Switzerland, the Pic du Midi in France, Kislovodsk in the U.S.S.R. and Norikura in Japan.

When there is no eclipse, the *K* corona can be detected only with great difficulty. The electrons of the *K* corona polarize the photospheric light as they scatter it. Polarimeters have been built and used in connection with coronagraphs to observe the *K* corona. Such observations can yield information about the changes in the inner corona from one day to the next.

Certain forbidden spectral lines from highly ionized atoms have recently been detected at infrared wavelengths longer than those that can be observed from the ground. During the solar eclipse of November 12, 1966, Guido Munch, Gerry Neugebauer and Dan McCammon of the Hale Observatories detected two new lines. During the eclipse of March, 1970, Kenneth H. Olsen, Charles R. Anderson and John N. Stewart of the Los Alamos Scientific Laboratory detected nine more lines, all in the range of wavelengths from 10,000 to 30,000 angstroms. All these lines were observed from aircraft that flew above much of the earth's atmospheric water vapor, which absorbs the infrared wavelengths. Jet aircraft can also prolong eclipses by chasing the moon's shadow; the supersonic *Concorde* prolonged the totality of the June 30 eclipse to an unprecedented 74 minutes.



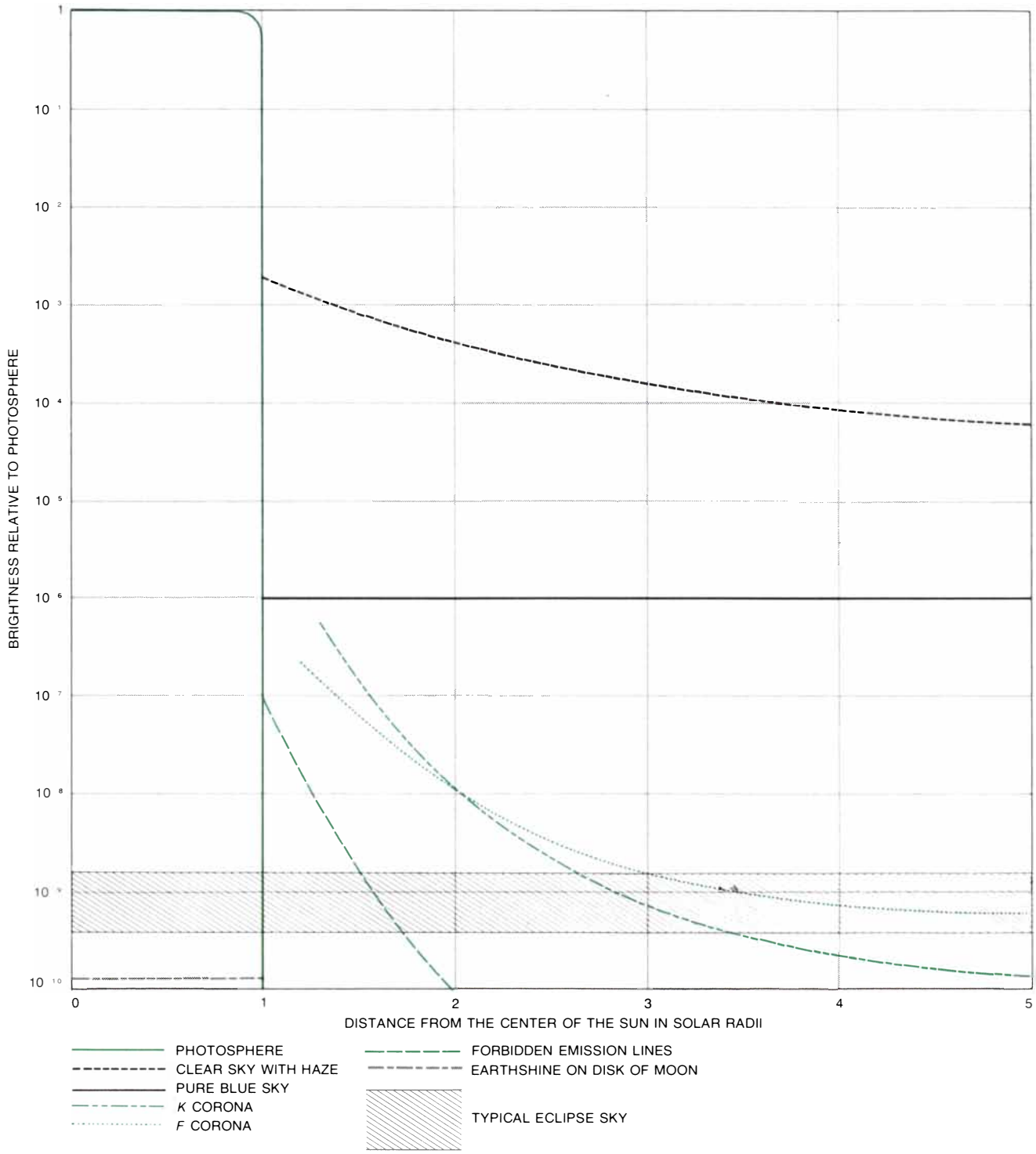
X-RAY PICTURE OF CORONA was made from a rocket launched by American Science and Engineering on the day of the March 7, 1970, total solar eclipse. (Rocket was outside the path of totality.) Wavelengths photographed ranged from three to 30 angstroms and from 44 to 55 angstroms. X-ray features corresponded to coronal features seen in white light.

The relationship of the corona's density and structure to its magnetic field can be studied both theoretically and observationally. Gordon Newkirk, Jr., and Martin D. Altschuler of the High Altitude Observatory have developed a method and a computer program for deriving the magnetic fields in space from magnetic regions observed on the surface of the sun. Their results can be displayed as diagrams of magnetic lines of force extending into space above the solar surface. The coronal structures presumably follow the magnetic lines of force.

Kenneth Schatten and Norman F. Ness of the Goddard Space Flight Center of NASA and John M. Wilcox of the University of California at Berkeley have devised an alternative method of predicting the structure of the corona. In order to make the calculations more realistic they assume that the source of the interplanetary magnetic field is an imaginary surface six-tenths of a solar radius above the photosphere. The magnetic field extends radially out from the surface and behaves according to certain simplified rules. Before the eclipses of September 22, 1968, and March, 1970, Schatten made predictions of what the coronal structure would be like. They seemed to agree reasonably well with the actual structure that eventually appeared. Gerald W. Pneuman and Roger A. Kopp of the High Altitude Observatory have developed a fuller theoretical treatment.

The solar telescopes that are now flying in space are revolutionizing observations of the corona. OSO-VII, launched in September, 1971, carried a coronagraph into orbit with an occulting disk at the end of a long boom. Coronagraphs orbited in space can be con-

structed differently from their ground-based counterparts and can reach much lower levels of intensity. The Apollo Telescope Mount of Skylab, the manned satellite observatory now in orbit, also carries a coronagraph. The Apollo Telescope Mount has television screens that display the detailed structure of the surface of the sun in the light of the hydrogen-alpha line and at X-ray wavelengths, and of the corona in white light. The major experiments of Skylab include a spectroheliometer from the Harvard College Observatory that is mapping the sun at wavelengths of from 300 to 1,350 angstroms with a resolution of five seconds of arc, a visible-light coronagraph from the High Altitude Observatory that provides pictures showing the structure of the corona between 1.5 and six solar radii at regular intervals, ultraviolet spectrographs from the Naval Research Laboratory and X-ray telescopes from American Science and Engineering and



RELATIVE INTENSITIES of the corona, of other parts of the sun and of the sky are shown with respect to the brilliant photosphere. The diagram is based on one made by H. C. van de Hulst. It is evident why the corona cannot be seen in the ordinary daylight sky and yet appears so clearly during a total eclipse of the sun. Curve for forbidden lines reflects total intensity of all such lines in the sun.

from the George C. Marshall Space Flight Center of NASA and the Aerospace Corporation.

Now that there is information about the corona flowing in daily from OSO-VII and Skylab, observations can be compared with theory much more frequently than they could if we had to rely exclusively on eclipses. It will be many years before we solar physicists will have synthesized the vast quantities of data we are now receiving from space. Nonetheless, we can be reasonably certain that current models of the corona will have to be drastically revised.

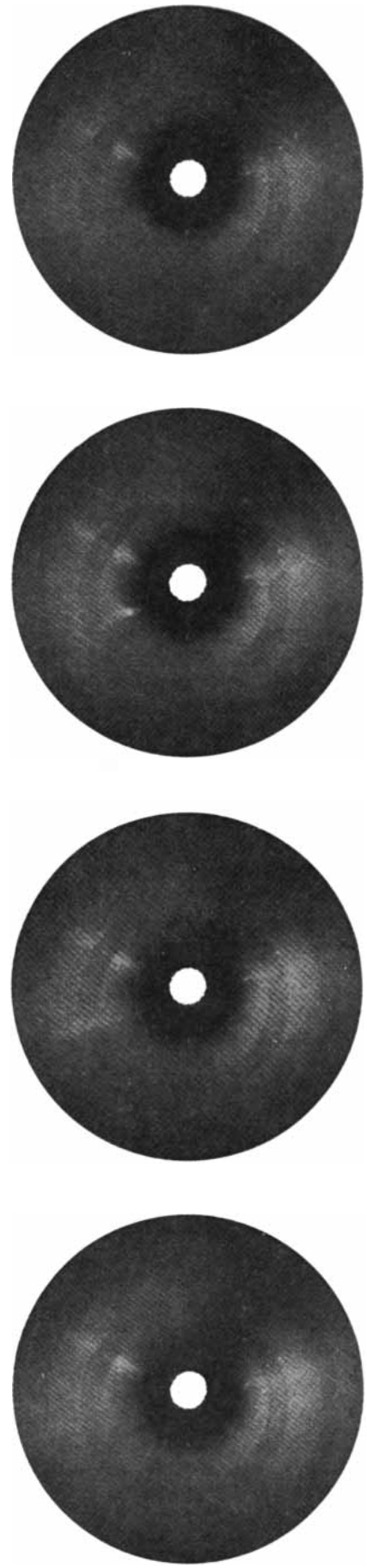
Observations of the corona always show a total intensity that includes not only a contribution from the *K* corona but also one from the *F* corona. How much the *F* corona contributes can be determined, because it differs from the *K* corona in two ways: in its spectra and in the polarization of its light. Donald H. Menzel and I have carried out a series of spectrographic and polarization observations during the past four total solar eclipses to investigate the separation of the *K* corona from the *F* corona. Whereas light scattered by electrons in the *K* corona is highly polarized, light scattered by the cold interplanetary dust of the *F* corona is basically unpolarized. The degree of polarization of the corona as a whole reaches a maximum of about 40 percent at a distance of 1.5 solar radii out from the center of the solar disk. Thereafter the amount of polarization decreases. That decrease can be interpreted as being the result of the superposition of the unpolarized *F* corona on the polarized *K* corona. The *K* corona itself reaches a polarization level of about 60 percent at about three solar radii.

In order to separate the polarized component from the unpolarized component, one must record the intensity of the corona during an eclipse through polarizers set at different angles. The method has a fundamental limitation, however, that has been pointed out by D. E. Blackwell of the University of Oxford. This is that the light of the *F* corona is not totally unpolarized; it has a residual polarization of a few percent. Thus results obtained for the separation of the *F* corona and the *K* corona beyond about five solar radii, where the total polarization is less than 10 percent, are inaccurate.

In order to separate the two components accurately one must examine the spectrum of the corona. As I have mentioned, the *F* corona simply scatters the Fraunhofer absorption lines from the solar photosphere. It does not affect the amount they absorb from the background level of the continuous spectrum. In the *K* corona, however, all the lines are washed out. To separate the two coronal components one must measure the intensity of the absorption lines in a spectrum made during an eclipse and compare those measurements with measurements of the lines in a spectrum of the photosphere made at another time with the same instrument. The degree to which the intensity of the spectral lines is reduced reflects the contribution of the *K* corona.

This method has been used for the hydrogen-alpha line, particularly by Blackwell and A. D. Petford. They derived a model for the two components of the corona that includes the distribution of temperature and the density of electrons near the sun out to a distance of 16 solar radii. Many of their observations were carried out from an aircraft at an altitude of 30,000 feet during the eclipse of July, 1963. The technique is a refinement of Blackwell's earlier work during the eclipse of June, 1954. During that first experiment he flew in an open cockpit at 30,000 feet. He stoically remarked that physiological difficulties made it inadvisable to attempt observations from a greater height, even though the aircraft could have climbed 13,000 feet higher.

One reason there is such interest in the outer corona is that there is a large gap between the corona we can detect during eclipses and the faint glow at the horizon known as the zodiacal light. The zodiacal light is visible at dark locations before sunrise and after sunset. It can be



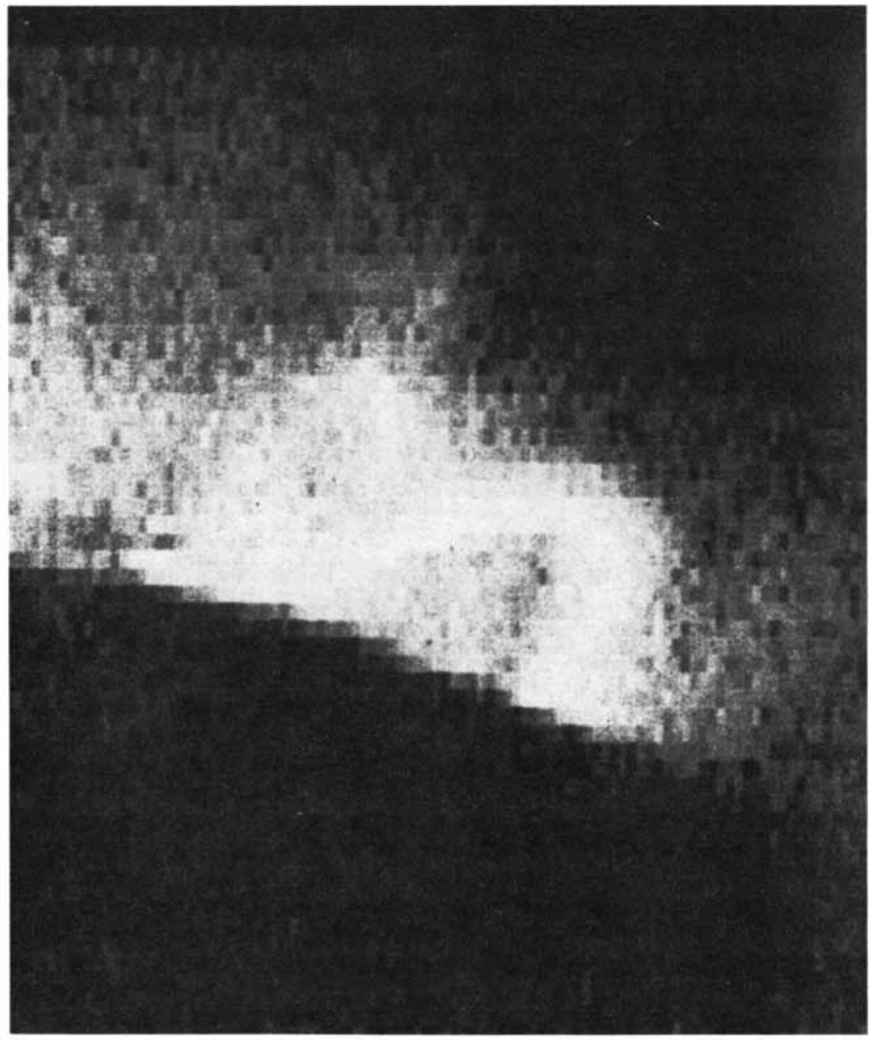
JETLIKE STREAMER formed quickly in the southeastern corona and disappeared rapidly within a period of less than five hours. It was photographed from a visible-light coronagraph aboard OSO-VII in four successive orbits on May 9, 1973. The top frame shows the sun at 5 hours 43 minutes Greenwich Mean Time, before the streamer appeared; the next frame shows the beginning of the streamer at 7 hours 15 minutes; the third frame shows the streamer nearly dissipated at 8 hours 47 minutes; the bottom frame shows the sun at 10 hours 17 minutes, after the streamer had disappeared. The time scale of the entire event is much shorter than was previously thought to be the case. The photographs were supplied by the Rocket Spectroscopy Branch of the Naval Research Laboratory in Washington, D.C.

detected out to about 80 solar radii. Presumably it is merely an extension of the *F* corona.

One set of data that helps to bridge the gap comes from the unmanned lunar vehicles of the Surveyor series. On the moon's airless surface there is obviously no atmospheric scattering of light. The corona simply rises each morning before the sun's disk and sets each evening after it. J. David Bohlin of the Naval Research Laboratory analyzed some of the data from *Surveyor 6* and *Surveyor 7*. He could follow one coronal streamer detected by *Surveyor 7* out to 22 solar radii.

Measurements at radio wavelengths can also be utilized to find the density of electrons in the outer corona. The corona tends to displace the apparent radio position of the Crab Nebula or quasars as they pass behind it. It also changes the apparent period of the radio signal from pulsars. Both types of change depend on the total density of electrons in the line of sight.

The structure of the corona varies strikingly from one eclipse to another. Its form has long been linked with the cycle of solar activity associated with the 11-year sunspot cycle. The shape of the corona is quite irregular when the number of sunspots is at a minimum (as



LOOPS OF IONIZED PARTICLES streaming away from the sun's surface more than 25,000 miles into space are shown in an electronically recorded image made by a scanning ultraviolet spectroheliometer aboard the manned satellite Skylab. The instrument is one of the solar physics experiments in the Apollo Telescope Mount and is the project of a group from the Harvard College Observatory led by Reeves. The Skylab spectroheliometer has a resolution some 30 times greater than that of similar instruments flown on the Orbiting Solar Observatories, revealing these distinct loops in the wavelength of 14-times-ionized iron at 417 angstroms. With the ultraviolet measurements very accurate temperatures can be determined for levels within the sun's atmosphere. Such measurements are made twice a day during both the manned and the unmanned periods of the Skylab mission.

it is now); it becomes much more symmetrical when the number is at a maximum. At sunspot minimum the strong solar activity is confined largely to the equatorial regions of the sun, and so at that time the corona mainly displays equatorial streamers. At sunspot maximum solar activity is much more general, and the corona extends farther in all directions.

The human eye sees the corona out to a few solar radii. Radial-density filters, which are now used at eclipses, show much detail. A radial-density filter is darker at the center than at the edges, and thus it allows the fainter outer corona to register within the limited dynamic range of film without overexposing the image of the brighter inner corona. Such photographs show with particular clarity that the spiky structure of the coronal streamers persists out to great distances from the sun.

These streamers attest to the fact that theoretical models of the corona's structure must have a number of components. Presumably a streamer is a region that has a slightly higher electron density than neighboring regions, so that it scatters more light. The newly discovered "holes" in the corona are regions of much lower density.

I believe the events of the current year will prove to have been extremely significant to our understanding of the corona. As the results from the June eclipse, from Skylab and from OSO-VII are fitted together, our ideas will change and our theories will improve. The picture that now seems to be emerging is that of a highly dynamic corona with large regions often changing within minutes or hours.

Calculations that have long been based on the assumption that the corona is a homogeneous, symmetrical cloud of gas must be thoroughly revised. We are now realizing that the complexity of the corona, and of the chromosphere as well, cannot be ignored if we want a model of the sun that is physically meaningful and not just a theoretical abstraction.

The lesson is an important one. The sun is only the nearest of billions of stars, concerning whose structure we are often content to accept general conceptions. Yet we know that other stars have chromospheres and coronas, and that many such stars are in a much more dynamic state than the sun is. The atmospheres of these stars must partake of the constant activity that now seems to characterize so much of the universe, which had long been regarded as being quite placid.

High-Efficiency Photosynthesis

Certain plants of hot and arid environments have evolved an unusual photosynthetic pathway with a high yield. This efficiency could well be turned to agricultural advantage

by Olle Björkman and Joseph Berry

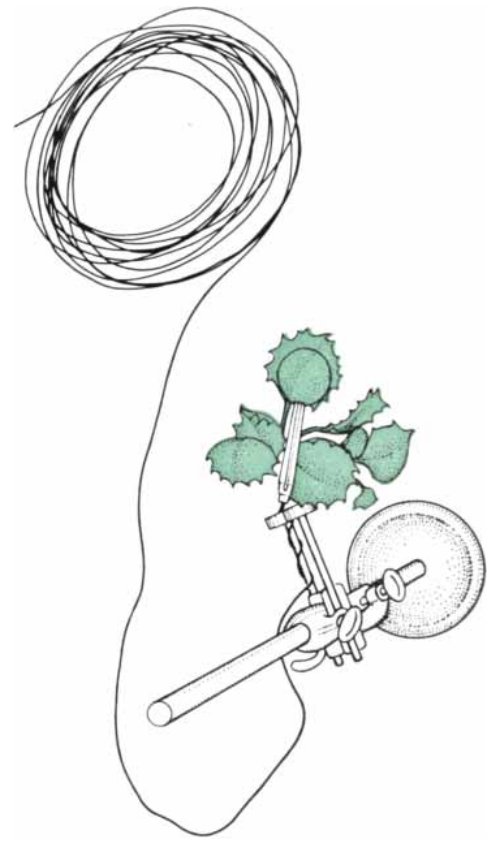
Plants make themselves out of water, carbon dioxide and minerals by the process of photosynthesis. It follows that any increase in the efficiency of photosynthesis could bring enormous benefits. Plant physiologists have recently learned that in certain plants the events of the photosynthetic process travel a unique dual pathway of carbon dioxide fixation, and that under certain conditions this pathway is far more efficient than the more familiar type of photosynthesis. The discovery raised immediate questions. Why did the pathway evolve? What kinds of plants have evolved it? Can the many plants that do not have it acquire it? The answers to the first two questions, at least, are fairly clear-cut.

Plant life in one form or another can be found in nearly every one of the very different natural environments that exist on the earth. Considering the restrictions imposed by such diverse habitats as arctic tundra, hot desert and shaded tropical rain forest, it seems remarkable that so many different kinds of plants are able to survive and reproduce in each. They do so, of course, because the plants themselves are functionally diverse and genetically adapted to the conditions that prevail in their respective habitats. Many of their evolutionary adaptations are subtly related to the plants' interactions with various components of the biological environment: bacteria, other plants and animal forms including protozoans, invertebrates (insects in particular) and the higher vertebrates.

Other plant adaptations are related to aspects of the physical environment, for example temperature, water supply and light intensity. These adaptations primarily involve the plants' various growth processes, the most fundamental being photosynthesis. The rates of some photosynthetic reactions depend on the intensity and quality of the light, the rates of others are mainly influenced by temperature, and the rates of still others are related to the amount of carbon dioxide in the air. As a result the rate of the overall photosynthetic process is greatly affected by the physical environment, perhaps more so than any other growth process. The success and productivity of a plant under environmental stress may therefore largely depend on efficient operation of the plant's photosynthetic machinery. If special adaptations that make for unusually efficient photosynthesis under conditions of stress have evolved, they are most likely to be found in plants that occupy extreme environments.

A striking example of one such adaptation is provided by a herbaceous perennial of the amaranth family, a group of herbs and shrubs that includes the familiar pigweed. The plant is *Tidestromia oblongifolia;* it grows in low, hot desert areas of the U.S. Southwest and is abundant on the floor of Death Valley in California. Death Valley is one of the harshest habitats on the earth; in summer it is the hottest natural environment anywhere in the Western Hemisphere. In Death Valley most plants grow only during the comparatively mild winter months, a period when most of the area's scanty rain (1.7 inches per year) falls. This is not true of *Tidestromia;* almost all its photosynthetic activity and growth occurs from May through August, the hottest and driest months of the year.

With the aid of a mobile laboratory we and our colleagues at the Carnegie Institution of Washington's Department of Plant Biology and at Stanford University recorded the photosynthetic activity of *Tidestromia* plants in Death Valley during the month of July. The plants were found to be photosynthesizing at high rates throughout the hours of daylight. They reached their maximum at noon, when both the heat load and the solar radiation were also at a maximum. (The ambient temperature at noon was about 122 degrees Fahrenheit.) The plants' maximum photosynthetic rate is



DESERT PLANTS that flourish in a hot, dry habitat were tested for photosynthetic performance under nearly natural conditions with the apparatus depicted here from

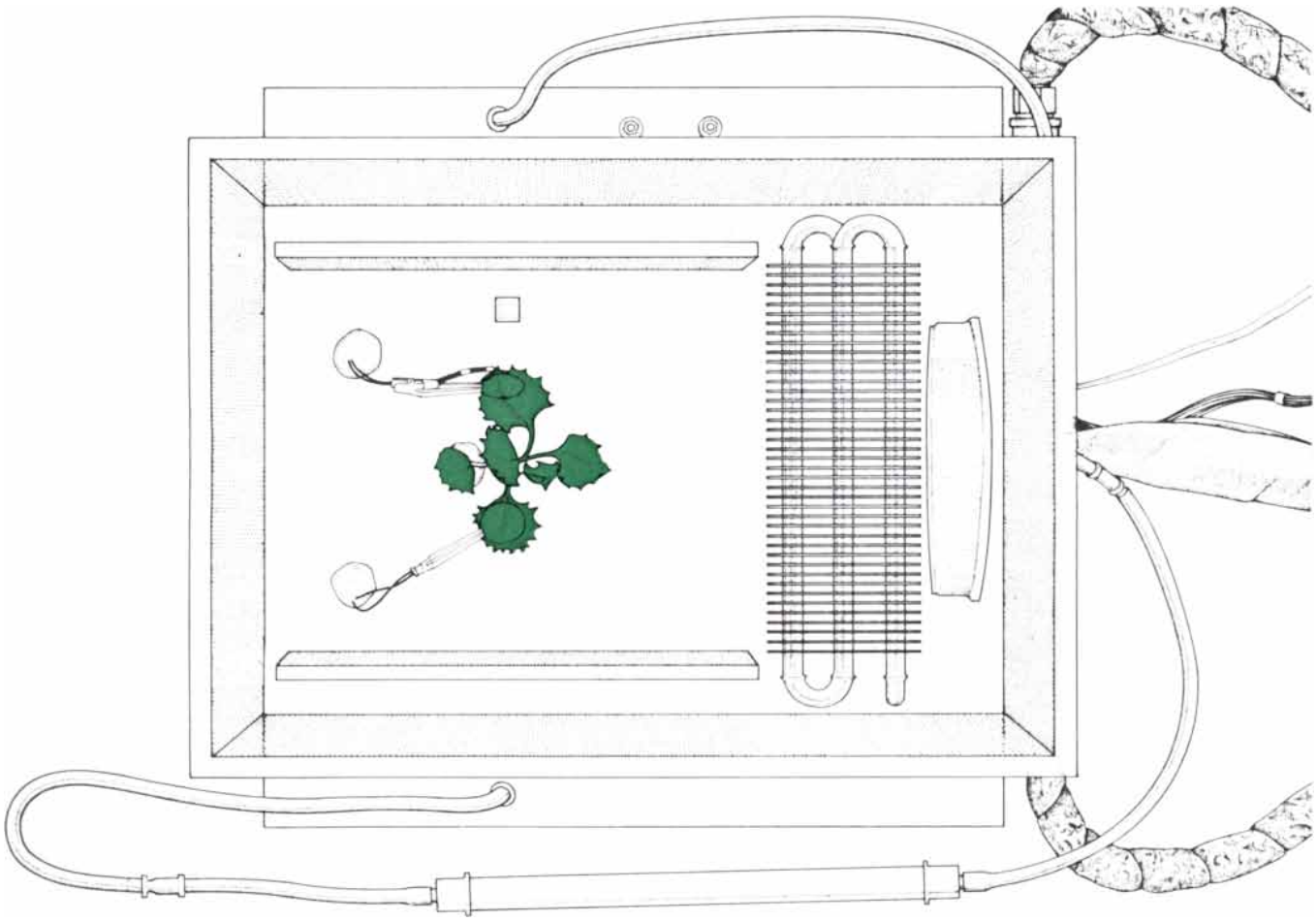
among the highest rates ever recorded in a natural habitat. It is comparable to the best rate of two notably productive agricultural plants: corn and sugarcane.

Experiments in which temperatures were controlled showed that the photosynthetic rate of *Tidestromia* is greatly affected by temperature changes. For example, the desert shrub does not equal a grass adapted to a Temperate Zone habitat in rate of photosynthesis when the temperature is below 20 degrees Celsius (68 degrees F.). When the temperature exceeds 30 degrees C. (86 degrees F.), however, the shrub's rate of photosynthesis exceeds that of the grass and continues to increase until it reaches a peak at 47 degrees C. (117 degrees F.), a temperature that would eventually be lethal to the Temperate Zone grass [*see illustration on page 87*]. Peak photosynthesis at temperatures above 47 degrees C. has been observed previously only in certain algae that have become adapted to a hot-springs habitat, which makes the desert shrub's performance remarkable indeed. Even if such temperatures were not lethal to plants adapted to more temperate environments, the photosynthetic mechanism in these plants would become inactivated and photosynthesis would come to a halt.

What is the water economy of a plant such as *Tidestromia,* which thrives at high temperatures in an environment with extremely dry air and a severely limited supply of water? Among plants in general only a small fraction of the water that is taken up by the roots serves as a reactant in photosynthesis. Most of it is transpired as water vapor by the leaves and released into the air. The plant is nonetheless protected against excessive loss through transpiration. Water vapor is not released by the leaves to the atmosphere at a constant rate. The leaves are covered with an impermeable waxy substance, and transpiration can be effected only through the stomata: pores in the leaf surface. The stomata vary in the size of their aperture; when they are closed, scarcely any water escapes from the moist cell walls of the leaf.

This phenomenon presents a puzzle. When the closed stomata protect a plant from undesirable water loss, they also prevent carbon dioxide in the atmosphere from entering the interior of the leaf and replacing the carbon dioxide consumed in the photosynthetic process. As a result the air within the leaf soon becomes depleted in carbon dioxide. Now, one factor that limits the rate of carbon dioxide uptake in photosynthesis is the carbon dioxide concentration within the leaf. This means that the depletion of the carbon dioxide is at first slowed, and that when the concentration of carbon dioxide has fallen to a certain level, photosynthesis comes to a halt.

Since transpiration and carbon dioxide uptake are linked, the success of a plant in a hot and arid environment depends to a large extent on how much carbon dioxide the plant can fix photosynthetically per unit of water lost in transpiration. The ratio is termed photosynthetic water-use efficiency. Efficiency, how-



above. The plants are *Atriplex hymenelytra*. Thermocouples attached to the control plant at left and the test plant in the chamber at right recorded leaf temperature. In the chamber the temperature was regulated by a circulating-water radiator and a fan to match the temperature of the plant outside. Air hoses delivered air of known humidity and carbon dioxide content. Under these conditions the test plant's transpiration of water and uptake of carbon dioxide could be measured by instruments in a mobile laboratory.

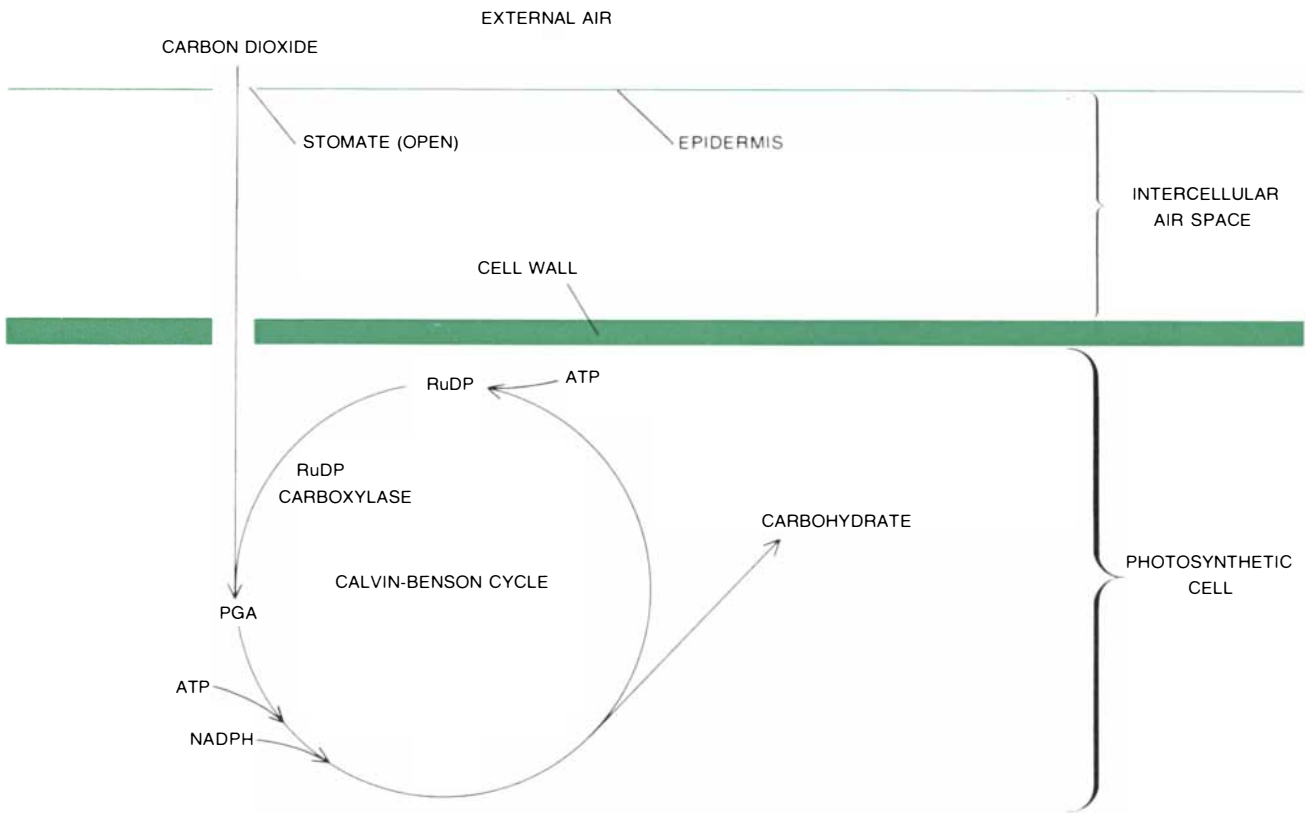
ever, is not the only factor to be taken into account. A reduction in stomatal aperture may increase a plant's ratio of carbon fixation with respect to water loss and thus increase the plant's water-use efficiency. As we have seen, however, this kind of water conservation inevitably leads to a slower absolute rate of photosynthesis.

The desert shrub *Tidestromia* has a much higher water-use efficiency than a Temperate Zone plant would have if it were moved to a desert habitat. At the same time the shrub is able to maintain a high absolute rate of photosynthesis. Laboratory investigations show that photosynthesis can proceed at an unusually high rate in the shrub even when the concentration of carbon dioxide within its leaves falls to a low level, as is the case when the stomata are partly closed. This enables the plant to combine a high efficiency of water use with a high absolute rate of photosynthesis. How is the feat accomplished?

Photosynthesis is a highly complex process that involves a large number of chemical reactions. In the simplest terms light energy is absorbed by the plant pigment chlorophyll and utilized to produce such high-energy intermediates as adenosine triphosphate (ATP) and strong reductants such as reduced nicotinamide adenine dinucleotide phosphate (NADPH) and also to oxidize water (a reaction that yields gaseous oxygen). The ATP and NADPH drive reactions that take up carbon dioxide from the atmosphere and reduce it to form carbohydrates, amino acids and other constituents of the plant.

Let us examine the latter part of the process in more detail. In most plants carbon dioxide from the air reacts with a compound that is generated within the leaf: ribulose-1,5-diphosphate, or RuDP, a phosphorylated sugar with five carbon atoms. When one carbon dioxide molecule reacts with a molecule of RuDP (the reaction is catalyzed by the enzyme RuDP carboxylase), two molecules are formed that have three carbon atoms each. The newly formed three-carbon substance is phosphoglyceric acid. Much of the phosphoglyceric acid formed within the leaf is converted into various end products (such as the sugar sucrose) in a series of energy-demanding reactions. Some phosphoglyceric acid, however, serves to regenerate molecules of RuDP so that they can once again act as carbon dioxide acceptors. The regeneration completes a loop that makes the process of fixing carbon dioxide self-sustaining. The overall process is called the Calvin-Benson cycle after its discoverers, Melvin Calvin and Andrew A. Benson.

In a few plants, the desert shrub *Tidestromia* among them, the initial fixing of atmospheric carbon dioxide is accomplished in a quite different way. In these plants the acceptor molecules within the leaf are phospho-enol-pyruvate (PEP). When one carbon dioxide molecule reacts with a molecule of PEP (the reaction is catalyzed by the enzyme PEP carboxylase), the end result is the formation of malic acid and aspartic acid. In each of these molecules there are four carbon atoms. The differences between the activities of PEP and RuDP and their associated enzymes are significant. First, PEP is substantially more reactive with carbon dioxide than RuDP is. Second, the activity of RuDP carboxylase is somewhat inhibited by oxy-



THREE-CARBON PHOTOSYNTHETIC PATHWAY is the usual one, so named because the initial product is a compound with three carbon atoms per molecule. In this simplified schematic diagram the carbon dioxide that the leaf has admitted from the surrounding air reacts with ribulose diphosphate (RuDP) in a reaction catalyzed by the enzyme RuDP carboxylase, forming two molecules of phosphoglyceric acid (PGA) with three carbon atoms each. In subsequent reactions some of the phosphoglyceric acid is converted to end products of photosynthesis and some of it is utilized to regenerate molecules of RuDP so that they can again serve as acceptors of carbon dioxide. The loop thus closed makes the process of fixing carbon dioxide a self-sustaining cycle driven by energy derived from light. End products of photosynthesis are carbohydrates, amino acids and other compounds that plant needs for growth.

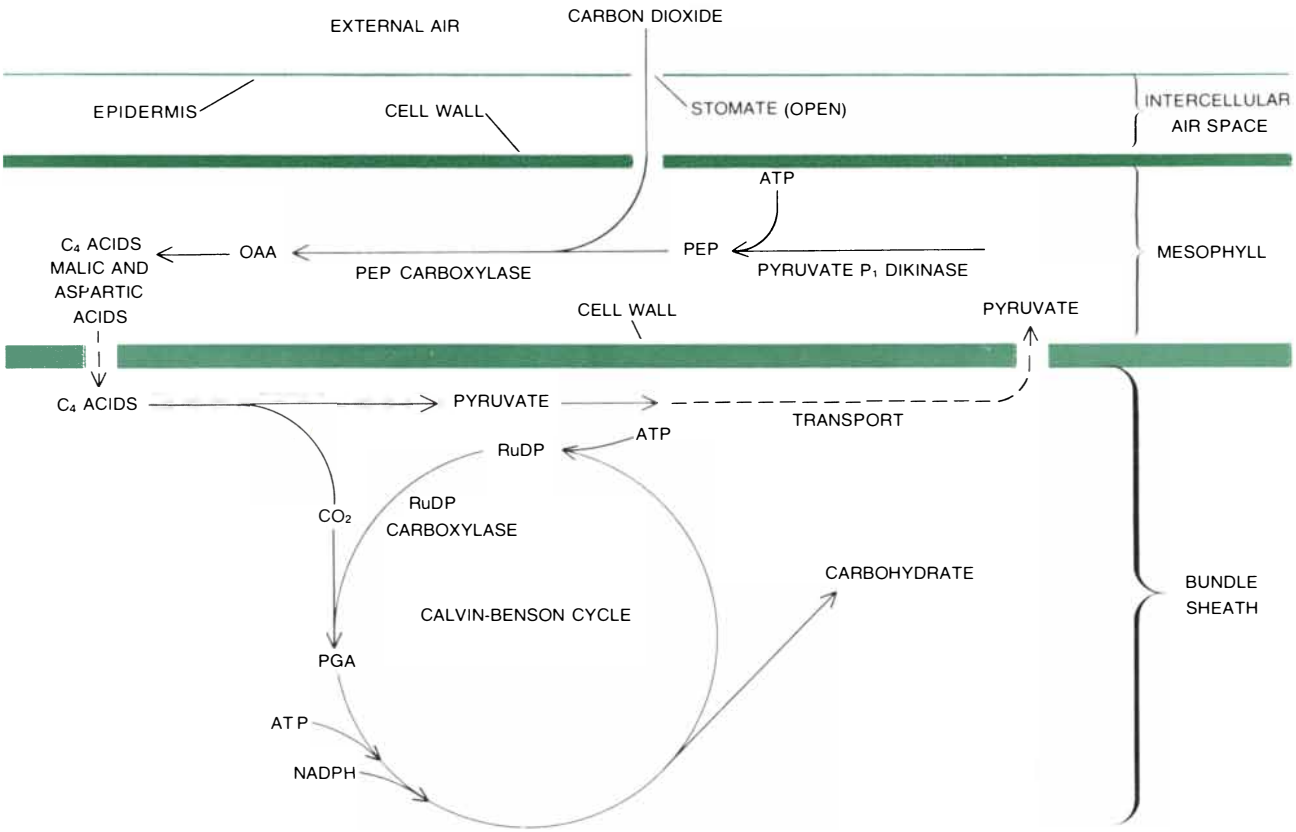
gen, whereas the activity of PEP carboxylase is not. This means that the fixation of carbon dioxide by the PEP system is more effective than the RuDP system when the proportion of carbon dioxide in the atmosphere is low and the proportion of oxygen is high.

The fixing of atmospheric carbon dioxide by PEP acceptor molecules is, however, only the beginning of the process in plants such as *Tidestromia.* Neither of the products of this pathway (malic acid and aspartic acid) can serve the functions equivalent to those of phosphoglyceric acid in the Calvin-Benson pathway; they cannot be converted into carbohydrates and other useful end products of photosynthesis without loss of carbon. What happens next is that the malic acid and the aspartic acid are broken down enzymatically; this process on the one hand releases the previously fixed carbon dioxide and on the other yields molecules of pyruvic acid with three carbon atoms. The next step closes the PEP loop and keeps the cycle self-sustaining; the pyruvic acid reacts with photosynthetically generated ATP to form additional molecules of PEP that are free to act as carbon dioxide acceptors.

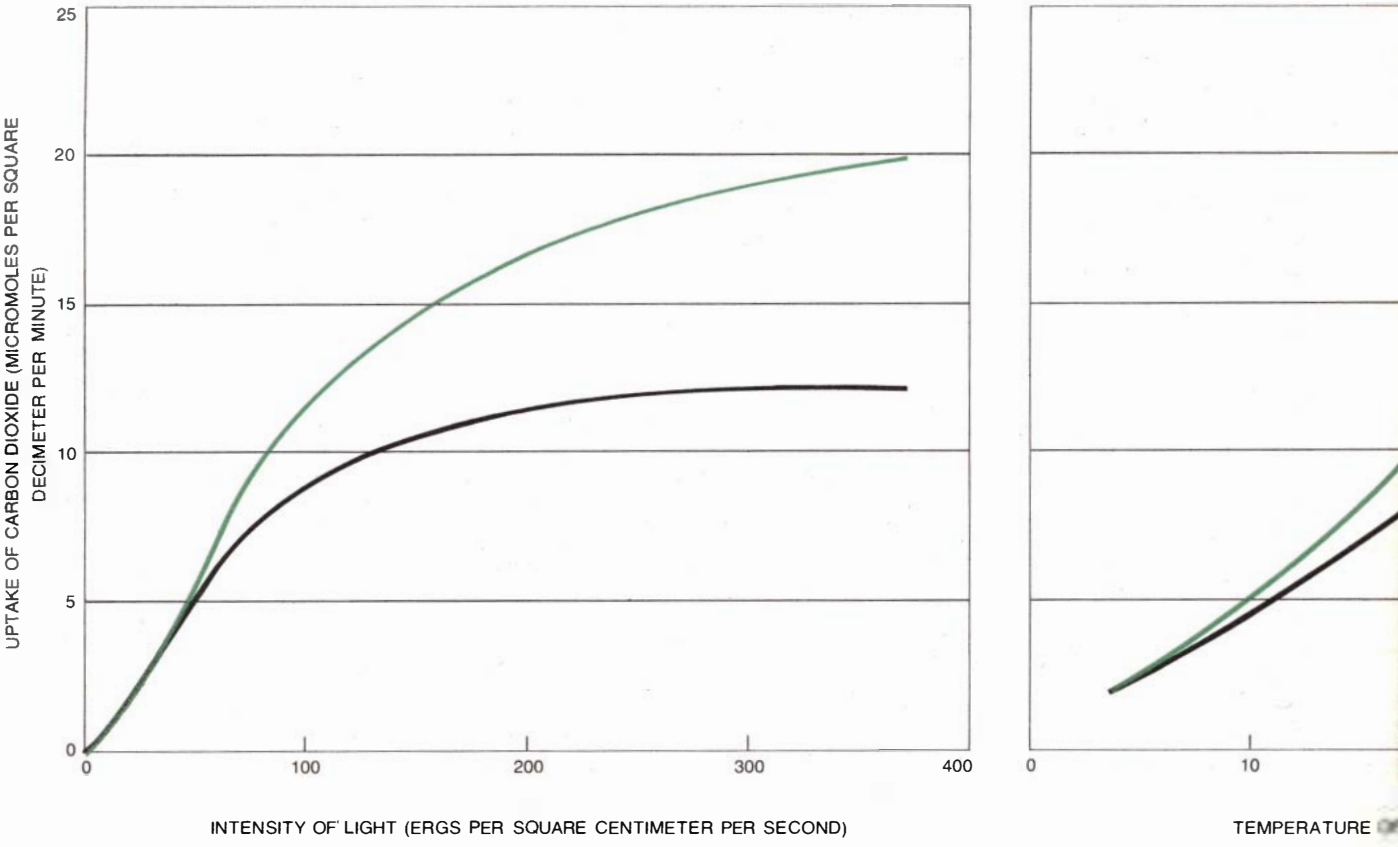
In essence the PEP system fixes atmospheric carbon dioxide only to release it again. This may seem to be a purposeless activity until one realizes that the fixation site is located at a significant physical distance from the release site. Moreover, once released, the carbon dioxide is immediately utilized by the plant's RuDP system and enters on the Calvin-Benson photosynthetic cycle. In plants such as *Tidestromia,* then, the PEP system provides an initial and very efficient means of fixing atmospheric carbon dioxide, even at very low concentrations. The system serves only one purpose: to feed carbon dioxide into the plant's Calvin-Benson photosynthetic system. Since the first product of the Calvin-Benson cycle is a substance with three carbon atoms per molecule, plants that accomplish photosynthesis only by means of this cycle are now commonly called C_3, or three-carbon, plants. This distinguishes them from plants, such as *Tidestromia,* that initially fix atmospheric carbon dioxide by the PEP system; these are known as C_4, or four-carbon, plants because the products of the PEP system, malic acid and aspartic acid, have four carbon atoms per molecule.

The first evidence for the existence of four-carbon plants was obtained in the early 1960's by Hugo Kortschak and his colleagues in the laboratory of the Hawaiian Sugar Planters' Association. They supplied carbon dioxide labeled with carbon 14 to photosynthesizing sugarcane plants and a few seconds later analyzed the compounds in their leaves. They found that the labeled carbon atoms were mainly concentrated in the four-carbon compound malic acid. The finding was a direct contradiction of the results of many similar experiments involving plants that ranged from primitive algae to the most advanced flowering species. In all these experiments the labeled carbon dioxide was first fixed in the three-carbon compound, phosphoglyceric acid.

In some respects the pathway Kortschak and his co-workers had discovered resembles the one used in carbon dioxide fixation by certain succulent plants. Even bacteria and animals have the PEP car-



FOUR-CARBON PHOTOSYNTHETIC PATHWAY occurs in some specially adapted desert plants. Carbon dioxide entering the leaf reacts with phospho-enol-pyruvate (PEP), a three-carbon compound, to form four-carbon oxaloacetic acid (OAA), from which malic acid and aspartic acid are formed. They are transported from the mesophyll cells, which are in the outer part of the leaf, to the inner bundle-sheath cells. There carbon dioxide is released from the four-carbon compounds and pyruvic acid, a three-carbon compound, is formed. The carbon dioxide is now fixed again in the usual cycle. The pyruvic acid returns to the mesophyll cells, where it acquires a phosphate group from adenosine triphosphate (ATP) to form PEP, thus regenerating the initial carbon dioxide acceptor molecule. This additional cycle for fixing carbon dioxide helps to increase the overall efficiency of utilizing the carbon dioxide.



PHOTOSYNTHETIC PERFORMANCE of the closely related plants *A. patula* (*black*) and *A. rosea* (*color*) differs markedly even when the plants are grown under identical controlled conditions. *A. patula* employs the three-carbon photosynthetic pathway, *A.*

boxylase enzyme, but its level of activity is much lower and its functions are different from those it serves in four-carbon plants. Soon after Kortschak's discovery two Australian plant physiologists, Hal Hatch and Roger Slack, confirmed the findings of his group and in a series of elegant experiments elucidated the main reactions of the four-carbon pathway. It was not long before the list of known four-carbon plants was expanded; it came to include not only sugarcane but also other crop plants of the grass family (such as corn and sorghum) and many nonagricultural grasses (including the summer-active weeds crabgrass and Bermuda grass).

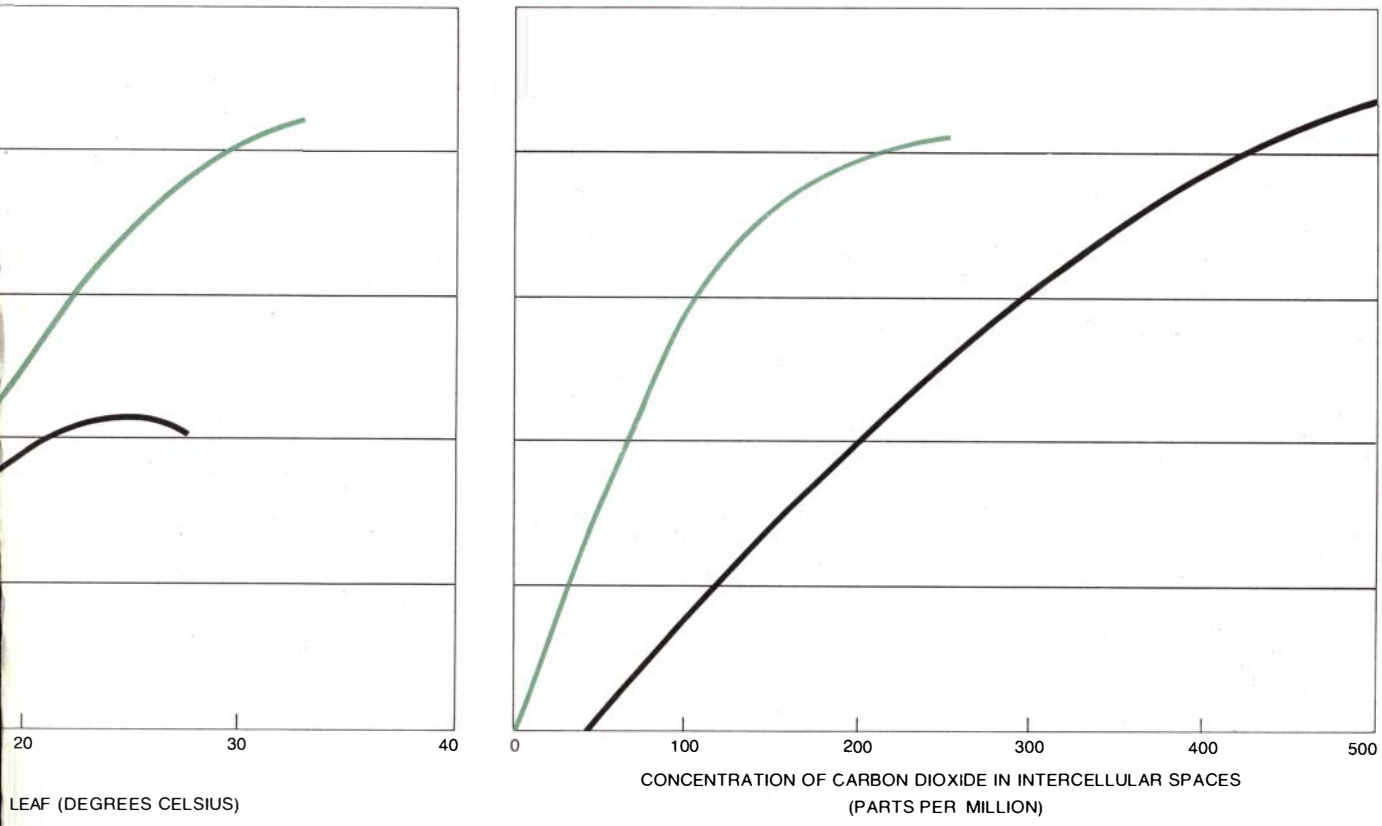
Today the four-carbon pathway is known to exist in nearly 100 genera in at least 10 plant families, both monocotyledonous and dicotyledonous, and the number of four-carbon species is counted in the hundreds. Not surprisingly the desert shrub *Tidestromia* proved to be a four-carbon plant. So did several other Death Valley plants of the saltbush genus *Atriplex*. This genus is one of at least 11 genera that include plant species of both the four-carbon and the three-carbon type. Considering that the four-carbon pathway can be present in some species of a genus and not in others, and that four-carbon plants are found among such entirely unrelated plant families as the grasses and the sunflowers, one conclusion seems inescapable: the four-carbon pathway must have evolved quite independently a number of times.

It was soon recognized that four-carbon plants, whatever their taxonomic relation to one another, share a number of additional characteristics that distinguish them from three-carbon plants. One of these characteristics, to which we shall return, is a specialized leaf anatomy that makes sense out of the unusual first step in four-carbon photosynthesis. Another is the presence of high concentrations of certain enzymes in the leaves of four-carbon plants that are found in three-carbon plant leaves only in much lower concentrations. As might be anticipated, these enzymes play an important role in four-carbon synthesis but are not involved in the three-carbon pathway.

A third characteristic, a by-product of the difference between the four-carbon and the three-carbon modes of carbon dioxide fixation, is a difference in the proportions of the two naturally occurring nonradioactive isotopes of carbon (carbon 12 and carbon 13) present in the carbon compounds of the two plant groups. This difference, incidentally, can be used to distinguish between the sugar of the sugar beet (a three-carbon plant) and the sugar of sugarcane (a four-carbon plant). The pure sucrose from the two plants is identical in all respects except for the ratio of the two carbon isotopes.

We come now to the question of why the four-carbon pathway evolved. Much of the evidence bearing on this question is to be found in observations of the four-carbon plants themselves, both in nature and in the laboratory. For example, if a four-carbon plant is placed in an enclosure so that the air can be sampled for analysis, the plant's photosynthetic activity soon almost entirely depletes the air of carbon dioxide. By way of comparison, a three-carbon plant in a similar enclosure will absorb carbon dioxide until the concentration falls from the normal level of 300 parts per million to about 50 parts per million. If the concentration goes any lower than 50 parts per million, the three-carbon plant no longer absorbs carbon dioxide but actually gives it off.

Consider, then, an experiment that places a four-carbon plant and a three-carbon one in the same enclosure. When the photosynthetic activity of both plants



***rosea* the four-carbon pathway. The differences increase with increased light intensity and leaf temperature and decreased carbon dioxide. Since plants are closely related, differences probably reflect differences in photosynthesis owing to differing pathways.**

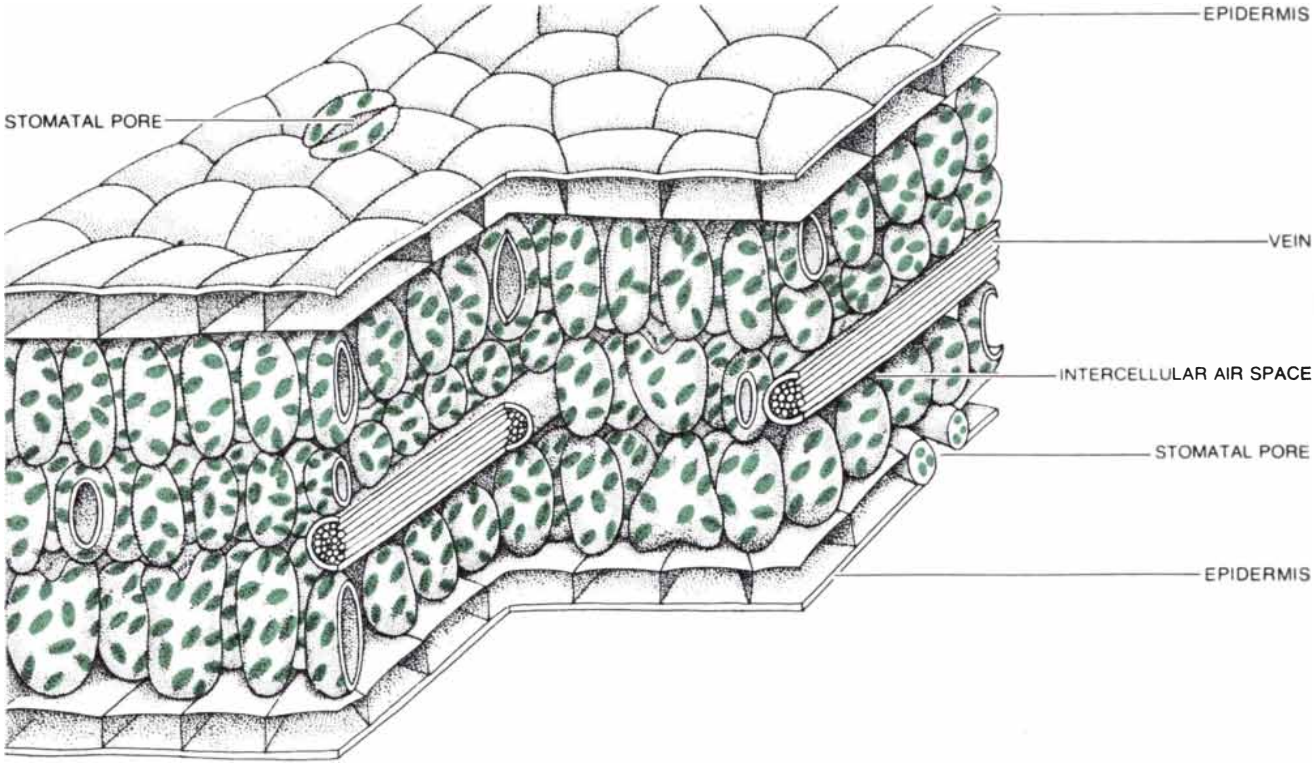
has reduced the concentration of carbon dioxide to below 50 parts per million, the three-carbon plant will begin to emit carbon dioxide. The four-carbon plant will now continue to absorb the scanty supply of carbon dioxide and keep on growing at the expense of the three-carbon plant until, depleted of its carbon resources, the three-carbon plant dies. The ability of four-carbon plants to absorb carbon dioxide from the air even when the concentration is as low as one or two parts per million has provided a useful means of screening various species of plants to determine whether or not they possess the four-carbon pathway.

The capacity to scavenge carbon dioxide and the lower ratio of carbon 13 to carbon 12 reflect certain important intrinsic characteristics of the four-carbon pathway. Under ordinary circumstances, however, neither of these features could in itself be expected to give any selective advantage to four-carbon plants. The net gain that the four-carbon pathway provides—a higher maximum rate of photosynthesis—can be realized only when the light is intense and the temperature is high. As an example, four-carbon plant genera such as *Tidestromia* are found primarily in areas where the plants' season of active growth coincides with maximum solar radiation, high temperatures and a limited supply of water. At the same time, however, a number of three-carbon plants also occur in these harsh habitats and a number of four-carbon plants successfully grow and reproduce in much milder environments.

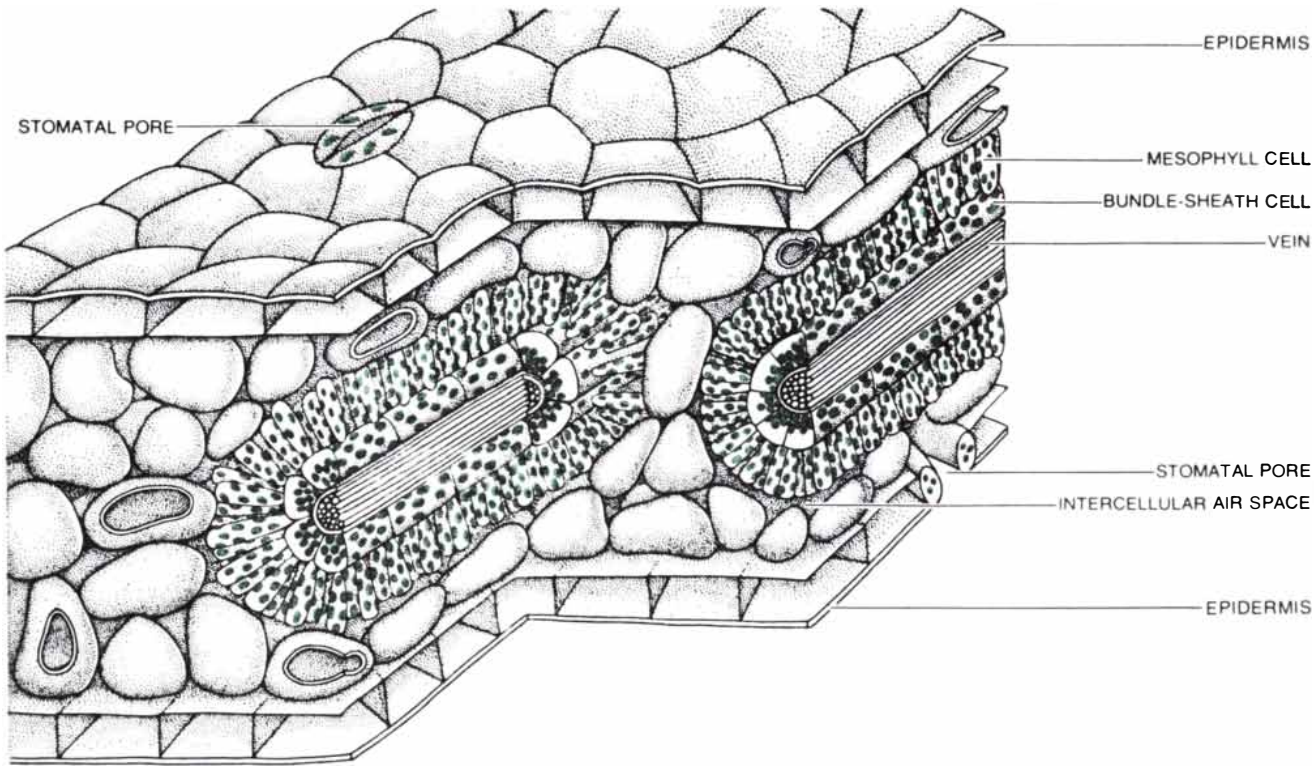
Ideally, to discover what the functional and adaptive significance of the four-carbon pathway is the experimenter should have at his disposal two groups of plants that are identical in genetic constitution except for the specific genes responsible for one group's possessing the four-carbon pathway. As it happens, a reasonably close equivalent to this ideal is provided by the three- and four-carbon species of *Atriplex*. Some of the species are genetically similar enough to hybridize and are capable of a high degree of chromosomal pairing. Moreover, two of them (the four-carbon species *A. rosea* and the three-carbon species *A. patula*) even coexist in the same environment and have similar patterns of summer growth.

One means of assessing the adaptive significance of the four-carbon pathway is to compare the photosynthetic efficiency of such a pair of plants when changes are made in the principal components of the physical environment. The plants' responses to the changes, taken together with a knowledge of how the same physical components vary in nature, will indicate the relative efficiency of the two plants in a wide range of potential habitats.

We have collected data of this kind by growing *A. rosea* and *A. patula* in the laboratory under identical controlled conditions, including near-optimal conditions for each species [*see illustration above*]. We compared their rate of photosynthesis (expressed in units based on the surface area of the plants' leaves) in a normal atmosphere at varying levels of light intensity and at various ambient temperatures. With respect to light intensity, we found that at 25 degrees C. (77 degrees F.) and a low illumination level the three- and four-carbon species photosynthesized at much the same rate. As the light intensity increased, so did both plants' rate of photosynthesis. At full sunlight intensities the photosynthetic rate of *A. rosea* was much higher than the rate of *A. patula*. We concluded that the four-carbon plant would be more efficient than the three-carbon one in open, sunny habitats, but that if the light were dimmed by the shade of other



LEAF STRUCTURE of the three-carbon plant *Atriplex patula* is portrayed. As in other typical leaves the cells, containing chlorophyll, which is shown in color, are of a single type, and they are found throughout the interior of the leaf. *A. patula* has a relative, *A. rosea*, that employs the four-carbon pathway for photosynthesis and has a different leaf structure, shown in the illustration below.

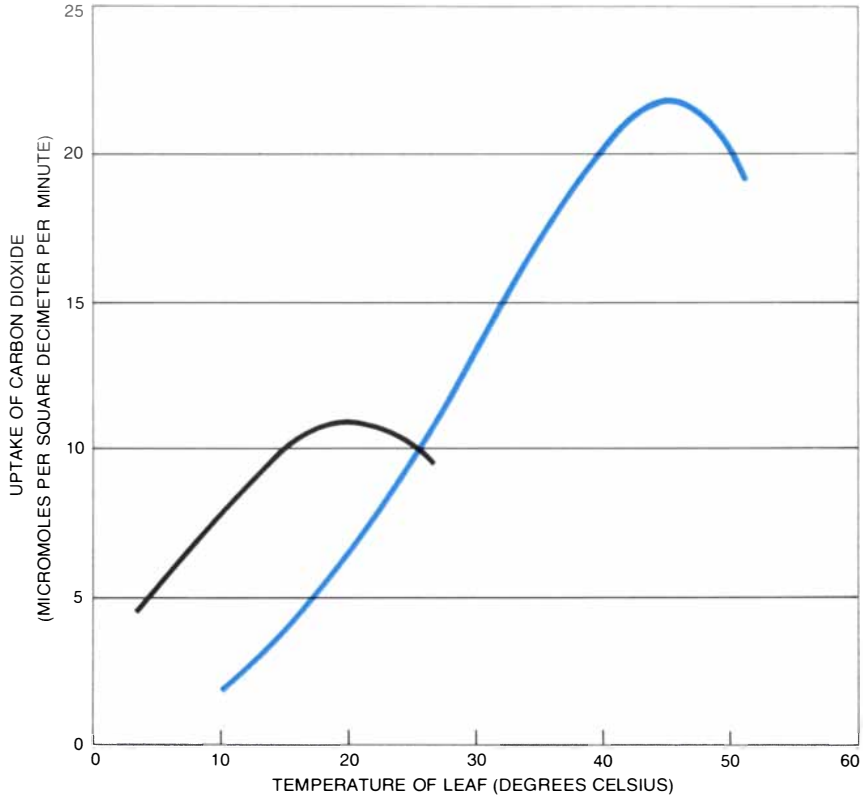


SPECIALIZED LEAF of *A. rosea* has nearly all its chlorophyll in two types of cells, which form concentric cylinders around the fine veins of the leaf. The cells of the outer cylinder are mesophyll cells; those of the inner cylinder are bundle-sheath cells.

plants or by clouds, the superiority of the one species over the other would be relatively insignificant.

We next varied the temperature while leaving the level of illumination constant. We found that the four-carbon plant's rate of photosynthesis was greater than the three-carbon plant's rate at 25 degrees C., but that as the temperature fell the difference in rate decreased and became insignificant at a temperature of five degrees C. (41 degrees F.). Raising the temperature above 25 degrees C. had the opposite effect: at 33 degrees C. (91 degrees F.) the photosynthetic rate of *A. rosea* was more than twice the rate of *A. patula*.

The results of the temperature and illumination experiments led us to two conclusions. First, the four-carbon species of *Atriplex* is a superior photosynthesizer under the combined circumstances of intense solar radiation and high temperature that are characteristic of an arid habitat on clear warm days. Second, this superiority of *A. rosea* over its genetically close relative, *A. patula*, means that the greater efficiency of the four-carbon plant is largely attributable to the plant's possession of the four-carbon pathway and not to other possible adaptations, unrelated to the pathway, that are also likely to have evolved under the pressure of natural selection.



PHOTOSYNTHETIC RATES of the grass *Deschampsia caespitosa* (*black*), which grows in cool, temperate environments, and the shrub *Tidestromia oblongifolia* (*color*), which grows in summer in Death Valley, are affected differently by temperature. *Tidestromia* fixes atmospheric carbon dioxide by four-carbon pathway, *Deschampsia* by three-carbon one.

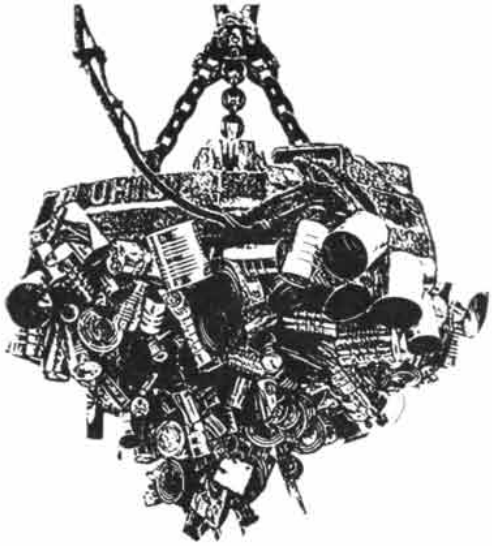
We have noted the relation between a plant's efficiency in utilizing water and the aperture of its leaf stomata, and also the inhibiting effect of closed stomata on carbon dioxide uptake. It is evident that one superiority of four-carbon plants is their ability to continue photosynthesis even when carbon dioxide is present in the interior leaf air spaces at very low concentrations. Assuming fully open leaf stomata and an atmosphere that contains the normal 300 parts of carbon dioxide per million, we calculate that in the interior of an *A. rosea* leaf the carbon dioxide concentration would be some 195 parts per million, whereas in the interior of an *A. patula* leaf it would be some 240 parts per million. One should not be misled, however, by the higher concentration in the leaf of the three-carbon plant. The 240 parts per million represents less than half the carbon dioxide concentration needed to saturate the photosynthetic pathway of the three-carbon plant, whereas 195 parts per million very nearly saturates the photosynthetic pathway of the four-carbon plant.

Of course, when the leaf stomata of both species are partly closed rather than fully open, the intercellular concentration of carbon dioxide in the leaf is reduced proportionately. As the concentration falls below 200 parts per million, *A. rosea* begins to surpass *A. patula* in photosynthetic efficiency. There is, however, another side to this coin. If the concentration of carbon dioxide is kept abnormally high rather than abnormally low, *A. patula* is then the equal, if not the superior, of *A. rosea* in photosynthetic efficiency.

Now, the rate of transpiration at any given stomatal opening (or, more precisely, stomatal conductance) is the same in both three- and four-carbon plants because the rate is determined solely by the difference in water-vapor concentration inside and outside the leaf. Why, then, is there a difference in the efficiency of water utilization between the two kinds of plant? The photosynthetic efficiency of the four-carbon pathway provides the answer. Since the four-carbon pathway can operate at very low intercellular concentrations of carbon dioxide, a four-carbon plant maintains a greater difference between the concentrations of carbon dioxide in the ambient atmosphere and in the leaf. At any given stomatal aperture, then, the diffusion of carbon dioxide into the leaf is a faster process for a four-carbon plant than it is for a three-carbon one. The result is a higher rate of photosynthesis for any given rate of water-vapor loss. For example, let us calculate the consequences of reducing the stomatal aperture (or conductance) of *A. rosea* leaves until the plant's rate of photosynthesis is no greater than the maximum rate for *A. patula*. Under these circumstances the four-carbon plant would transpire only a fifth as much water as the three-carbon one.

We have put calculations such as these to the test by means of comparative-growth experiments in the Carnegie Institution's experimental garden at Stanford University. The locale of the garden, the Santa Clara Valley, has a Mediterranean climate. Its summers are periods of warm temperatures and little or no rainfall, and unless plants in the garden are irrigated the available water supply is limited to whatever water is present in the soil at the start of the growing season.

We planted two test plots with seedlings of both *A. rosea* and *A. patula* in March and watered the seedlings uniformly until May 1. After that date we irrigated one plot but not the other. Both species in the irrigated plot grew at



St. Louis Converts Garbage Into Energy

In *St. Louis, enough electric power to serve 25,000 homes is generated by mixing garbage with coal in a conventional boiler. This reduces air pollution, saves landfill and conserves coal.

The city's garbage is first mined for steel cans. St. Louis is one of twenty cities reclaiming three billion steel cans annually by magnetic separation—the leading resource recovery method. Steel's unique magnetic property makes it possible.

What happens to the reclaimed steel cans? St. Louis sells them to one of our steel companies for remelting. In other cities the can scrap goes to the copper industry, detinning plants or ferroalloy producers as well as steel mills.

For more information, write Tinplate Producers, American Iron and Steel Institute, 150 East 42nd Street, New York, New York 10017.

Steel — the recycled material

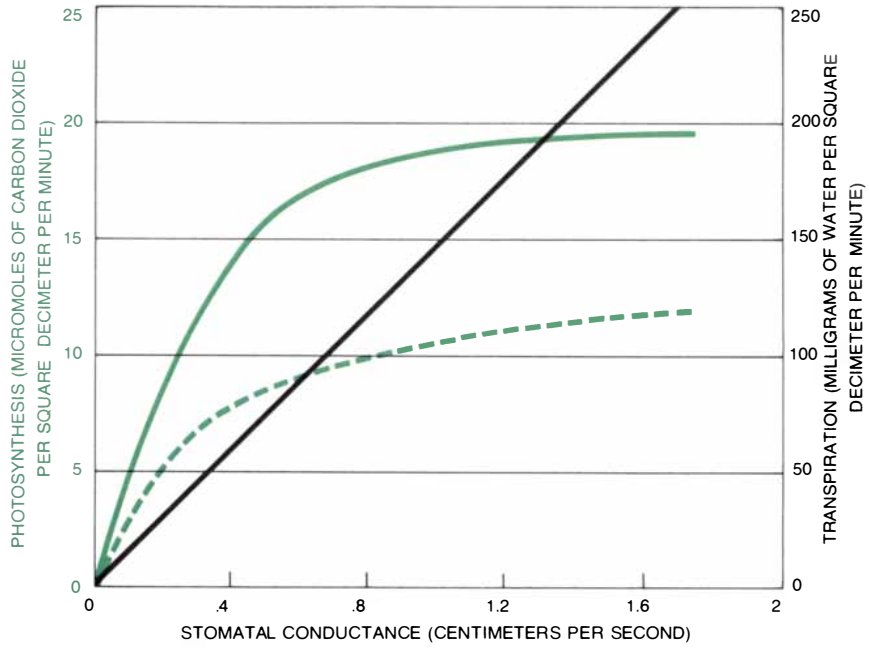
Tinplate Producers/American Iron and Steel Institute

*St. Louis Solid Waste Demonstration Project is funded and operated by U.S. Environmental Protection Agency, City of St. Louis, Union Electric Company, and American Iron and Steel Institute.

about the same rate, matured successfully and produced seed in abundance. In the dry plot the *A. rosea* seedlings grew at much the same rate as they did in the irrigated plot, but the growth of the *A. patula* seedlings slowed drastically when irrigation ceased. All the three-carbon plants died before they reached maturity. Meanwhile the four-carbon plants in the dry plot grew vigorously, matured and produced seed.

A further factor affecting the efficiency of metabolic processes in plants appears to have had its origin early in the course of evolutionary history. During most of Precambrian times, before photosynthetic plants had evolved, oxygen may have comprised only about 1 percent of the earth's atmosphere. Today, hundreds of millions of years later, the atmosphere contains some 21 percent oxygen as a result of photosynthesis by plants. The present high concentration of oxygen actually inhibits photosynthesis as far as three-carbon plants are concerned. The effect is relatively small when the temperature is low but it increases as the temperature rises. If a three-carbon plant is placed in an artificial atmosphere that contains the normal .03 percent concentration of carbon dioxide but an oxygen concentration of only 1.5 percent, its rate of carbon dioxide uptake is 40 percent higher than it is in a 21 percent oxygen atmosphere. Conversely, if the concentration of oxygen is raised above 21 percent or the concentration of carbon dioxide is lowered to less than the normal .03 percent, a three-carbon plant's rate of carbon dioxide uptake falls accordingly. The inhibiting effect is completely overcome, however, if the concentration of carbon dioxide is increased to a high enough level to saturate the photosynthetic process.

The inhibitory effect of atmospheric oxygen is not noticeable where four-carbon plants are concerned, even when the concentration of carbon dioxide falls below the normal .03 percent. Since an oxygen-poor atmosphere improves the rate of carbon fixation in a three-carbon plant, however, it seems logical to expect that the difference in photosynthetic performance between three- and four-carbon plants would become trivial in such an atmosphere. Experiments confirm this; when *A. rosea* and *A. patula* are placed in an enclosure with air that contains only 1.5 percent oxygen, the two species' rates of photosynthesis are much the same regardless of temperature and illumination. The same is true in a normal-oxygen atmosphere, provided that the concentration of the carbon dioxide has been raised to the saturation point.



PHOTOSYNTHESIS AND TRANSPIRATION of *A. rosea* and *A. patula* are shown. Transpiration of water from leaf (*black*) is same for both and decreases with stomatal conductance. Photosynthesis also decreases as stomata close. At any given stomatal conductance *A. rosea* (*solid color*) is superior to its three-carbon relative, *A. patula* (*broken color*), in carbon dioxide uptake. At a given rate of carbon dioxide uptake the four-carbon plant would have its stomata less open and so would tend to lose less water than three-carbon plant.

In summary, comparative studies of *A. rosea* and *A. patula*, including the findings with respect to the inhibitory effect of oxygen on photosynthesis in three-carbon plants, lead to the following conclusion. The superior photosynthetic performance of four-carbon plants under circumstances of intense radiation, high temperature and limited water supply is largely a result of the plants' greater efficiency in utilizing carbon dioxide at low concentrations. As we have seen, this efficiency is attributable to the four-carbon cycle of carbon fixation. The four-carbon cycle is not, however, the whole story. The leaf anatomy of four-carbon plants also plays a significant part. Let us briefly compare the leaves of *A. patula* and *A. rosea*.

A. patula has the leaf anatomy common to three-carbon plants. The cells that contain chloroplasts, the particles where the energy of sunlight is first absorbed by chlorophyll, are distributed throughout the leaf. The identical complement of photosynthetic enzymes is present in each cell containing chloroplasts. Each cell, following the Calvin-Benson cycle, independently fixes some of the carbon dioxide in the air that enters through the leaf stomata.

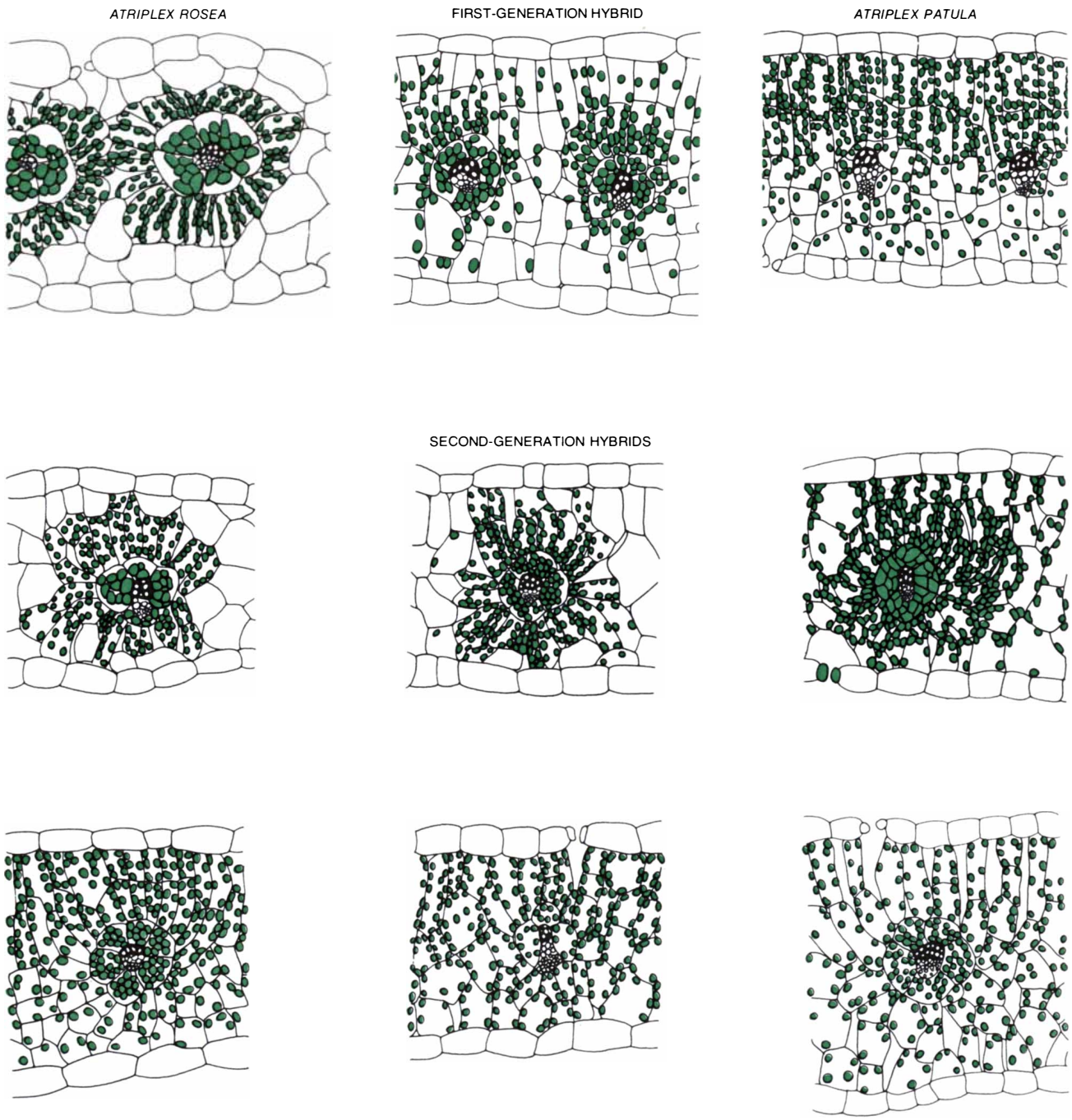
The cells surrounding the tiny leaf veins in the four-carbon plant *A. rosea* are arranged in a quite different way: they form a pair of concentric cylinders. Many cells in the leaves of four-carbon plants contain few chloroplasts and probably make no significant contribution to the photosynthetic process, but the cells in the two cylinders have large numbers of chloroplasts. The outer cylinder consists of what are called mesophyll cells. The cells of the inner cylinder, known as the bundle sheath, have thick walls.

We have found that the photosynthetic enzymes contained in the mesophyll cells are different from the enzymes in the bundle-sheath cells. This means that the two arrays of cells have different functions. For example, it is only in the mesophyll cells that atmospheric carbon dioxide is fixed by the four-carbon cycle. Similarly, it is only in the bundle-sheath cells that the products of the four-carbon cycle, malic acid and aspartic acid, are broken down to release the fixed carbon dioxide and simultaneously to form pyruvic acid.

Both the transport of the four-carbon compounds from the mesophyll sites of carbon fixation to the bundle-sheath sites where they are broken down and the return transport of pyruvic acid from bundle-sheath cells to mesophyll cells are processes that bring about a net transport of carbon dioxide from the outer cell array to the inner one. Once the car-

bon dioxide is present in the bundle-sheath cells it is fixed and subsequently converted into sugar and other end products of photosynthesis by the Calvin-Benson cycle. Because the carbon-fixing sites of the Calvin-Benson cycle are located inside the bundle sheath, they are cut off from direct contact with atmospheric carbon dioxide. If the inner cells were not fed carbon dioxide as a result of the four-carbon cycle in the outer cells, no usable photosynthetic end products would be formed and four-carbon plants could not grow.

The advantage of the four-carbon plants' two-cycle system of photosynthesis is that the plants can absorb low concentrations of atmospheric carbon dioxide very efficiently and feed the carbon dioxide into the Calvin-Benson cycle to obtain the photosynthetic end products needed for growth. The plants' specialized leaf anatomy provides the spatial compartmentalization required for the two separate sets of reactions and is therefore an essential component of photosynthesis by four-carbon plants. The two-cycle system necessarily requires a somewhat greater input of energy per molecule of carbon fixed, but that input is offset by a high rate of carbon fixation at the low concentrations of carbon dioxide that are characteristic of all inter-



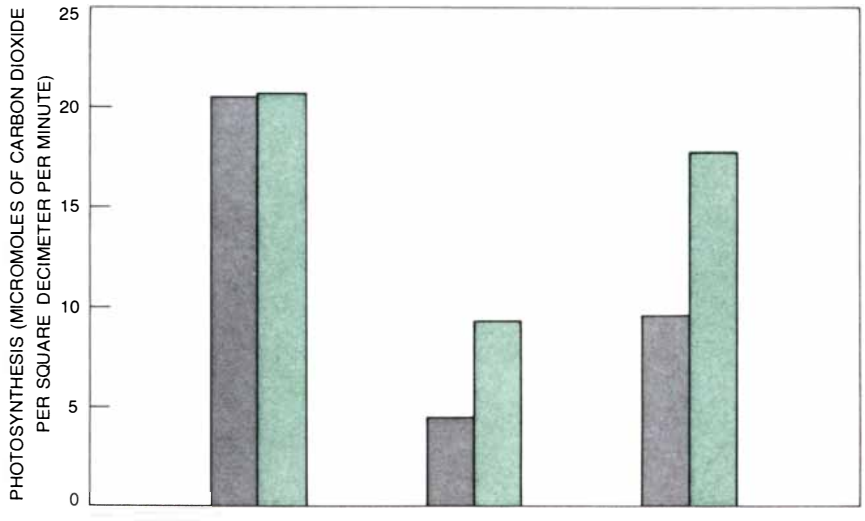
ANATOMY OF LEAVES is shown at top in cross section for *A. rosea*, *A. patula* and a hybrid obtained by crossing the two plants. In the first-generation hybrid the anatomy of the leaf combines traits from both parental types. In second-generation hybrids the anatomy varies greatly from plant to plant. Some have distinct mesophyll and bundle-sheath cells as in the four-carbon parent, *A. rosea*. Others lack this specialization of leaf structure as in the case of *A. patula*, the parent with three-carbon photosynthetic system.

cellular leaf spaces, both in four- and three-carbon plants.

Whether a plant possesses one or the other photosynthetic pathway is genetically determined. What, then, is the likelihood of being able to breed the four-carbon pathway into three-carbon plant stock? Hybridization experiments with *A. rosea* and *A. patula* have been conducted in our laboratory, and our colleague Malcolm A. Nobs has been able to hybridize *A. rosea* with other three-carbon species of *Atriplex*. These are the only crosses achieved so far, even though a number of three- and four-carbon species of the same genera are known. The results of the *Atriplex* experiments are illuminating.

The first-generation hybrids produced by fertilizing *A. rosea* plants with *A. patula* pollen are in many respects intermediate between the two parents, but more characteristic features appear in subsequent generations. With respect to leaf anatomy, for example, the leaves of first-generation hybrids are intermediate in morphology, and the leaves of some second-generation plants have morphological traits resembling those of either one parent or the other. A similar segregation of traits is apparent with respect to the plants' biochemical characteristics; an example is the considerable variation in the amount of PEP carboxylase present in the leaves of second-generation plants. At the same time single second-generation plants may resemble one parent in one characteristic and another parent in another. This is an effective demonstration that the characteristics exhibited by the offspring are independently heritable.

A few second-generation hybrids resemble their four-carbon parent both in leaf anatomy and in biochemical characteristics. One might expect that the hybrids would also possess a functional four-carbon-cycle system of photosynthesis. Among the several hundred such plants examined so far, however, this has not been the case. On the contrary, all the hybrids photosynthesize at much the same rate. Not one has been found to have a complete and integrated four-carbon pathway. Our results suggest that efficient dual-cycle photosynthesis requires that the necessary components not only be present but also be properly coordinated. Thus even though only a few genes may be involved in determining the genetic inheritance of each component of four-carbon photosynthesis, the requirement for a complete coordination of the anatomical and biochemical properties of the leaf could well make it impossible to introduce this pathway into plants lacking it. In the light of the evidence that the four-carbon pathway evolved independently a number of times, our results probably also reflect the massive difference in scale between the processes of nature and man's efforts at genetic manipulation in the laboratory. Sexual reproduction from generation to generation over thousands or even millions of years constantly brings into existence new and unique combinations of genetic material. It is scarcely surprising that the four-carbon pathway should appear more than once in nature even though the same accomplishment may prove impossible, or at least exceedingly difficult, to achieve by means of plant-breeding experiments.



AIR TESTS show that inheritance of efficient four-carbon photosynthesis is complex and has not been achieved by crossbreeding three-carbon and four-carbon relatives. Pairs of bars represent photosynthetic performance of *A. rosea*, a first-generation hybrid, and *A. patula* in normal air (*gray*) and air with a reduced content of oxygen (*color*). Oxygen inhibits photosynthesis in the three-carbon plants and hybrid but not in the four-carbon plants.

In our opinion there is at man's disposal a method for utilizing the four-carbon pathway that has far more promise than attempts to introduce the pathway into plants that lack it. Corn, sugarcane, sorghum and certain pasture grasses already have this photosynthetic pathway. This stock of plants for agricultural purposes can be increased by breeding certain wild species of four-carbon plants into agriculturally useful domestic species. The grain amaranth (a relative of *Tidestromia*), once a staple of early Central American Indian cultures, is an example of a four-carbon species that might be bred to yield a useful grain for animal feed or flour supplement.

As for existing four-carbon crop species, an important question concerns the way they are used. From our studies with wild species such as *Tidestromia* and the two species of *Atriplex* we feel confident that this mechanism is an adaptation to hot and arid environments. Specifically it is a component leading to increased efficiency of photosynthesis at high temperature and to increased potential for efficient utilization of water during growth. Are current agricultural practices taking full advantage of these potentials? In our opinion they probably are not. A primary objective of plant-breeding programs and of agricultural practice has been to obtain the highest possible yield per acre under optimal conditions. Although this objective may be valid for the richest farmland, it is of doubtful value in marginal regions where, for example, the cost or availability of a resource such as water limits the amount of land that can be cultivated.

Because of their special attributes domesticated four-carbon plants may well prove to be ideal crops in such areas if agricultural practices are adjusted to exploit their potential. Even in arid regions water is often used lavishly and hence inefficiently. Moreover, since for a given crop maximizing the absolute growth rate inevitably results in a decreased efficiency of water use, it is likely that current plant-breeding programs have produced varieties with comparatively low efficiency of water use. The same may even be true of certain high-yield varieties of four-carbon crop species. The development of agricultural practices and crop varieties appropriate to conditions where water is limited would result in a considerable increase in the productivity of arid land. Plants with four-carbon photosynthesis would play a central role in this development.

Auditory Beats in the Brain

Slow modulations called binaural beats are perceived when tones of different frequency are presented separately to each ear. The sensation may show how certain sounds are processed by the brain

by Gerald Oster

If two tuning forks of slightly different pitch are struck simultaneously, the resulting sound waxes and wanes periodically. The modulations are referred to as beats; their frequency is equal to the difference between the frequencies of the original tones. For example, a tuning fork with a characteristic pitch of 440 hertz, or cycles per second (*A* above middle *C* on the piano), and another of 434 hertz, if struck at the same time, will produce beats with a frequency of six hertz.

In modern investigations tuning forks are replaced by electronic oscillators, which can supply tones of precisely controlled pitch, purity and intensity. Beats are produced when the outputs of two oscillators tuned to slightly different frequencies are combined electrically and applied to a loudspeaker. Alternatively, the signals can be applied individually to separate speakers and the beats will still be heard. The result is the same whether the tones are combined electrically and then converted into sound, or converted into sound separately and then combined.

A quite different phenomenon results when stereophonic earphones are used and the signals are applied separately to each ear. Under the right circumstances beats can be perceived, but they are of an entirely different character. They are called binaural beats, and in many ways they are more interesting than ordinary beats, which in this discussion will be called monaural. Monaural beats can be heard with both ears, but one ear is sufficient to perceive them. Binaural beats require the combined action of both ears. They exist as a consequence of the interaction of perceptions within the brain, and they can be used to investigate some of the brain's processes.

The physical mechanism of monaural beats is a special case of wave interference. At any instant the amplitude of the resulting sound is equal to the algebraic sum of the amplitudes of the original tones. The signals are reinforced when they are in phase, that is, when the peaks and nulls of their waves coincide. Destructive interference diminishes the net amplitude when the waves are in opposition. The pure tones used in these experiments are described by sine waves; the resulting beats are slowly varying functions similar to, but not precisely conforming to, a sine wave.

A beat frequency of about six hertz, as in the example given above, would sound something like vibrato in music (although vibrato is frequency modulation rather than amplitude modulation). If the interval between frequencies is made smaller, very slow beats can be produced, down to about one per second, but at this speed the beats may be difficult to perceive. Rapid beats, up to about 30 hertz, are heard as roughness superimposed on the sound, rather like a Scotsman's burr. With still greater intervals beats are not heard; the two tones are perceived separately.

Beats are rarely encountered in nature because in nature sustained pure tones are rare. They abound, however, in mechanical devices. In an airplane, jet engines operating at slightly different speeds may produce a very strong-beat, often recognized only as a feeling "in the pit of the stomach." Acoustical engineers can filter out the whine of the engines, but the slow vibrations are difficult to suppress. Occupants of apartment houses may be annoyed by beats produced by machinery, such as two blowers running at different speeds, but they will have a hard time finding the source.

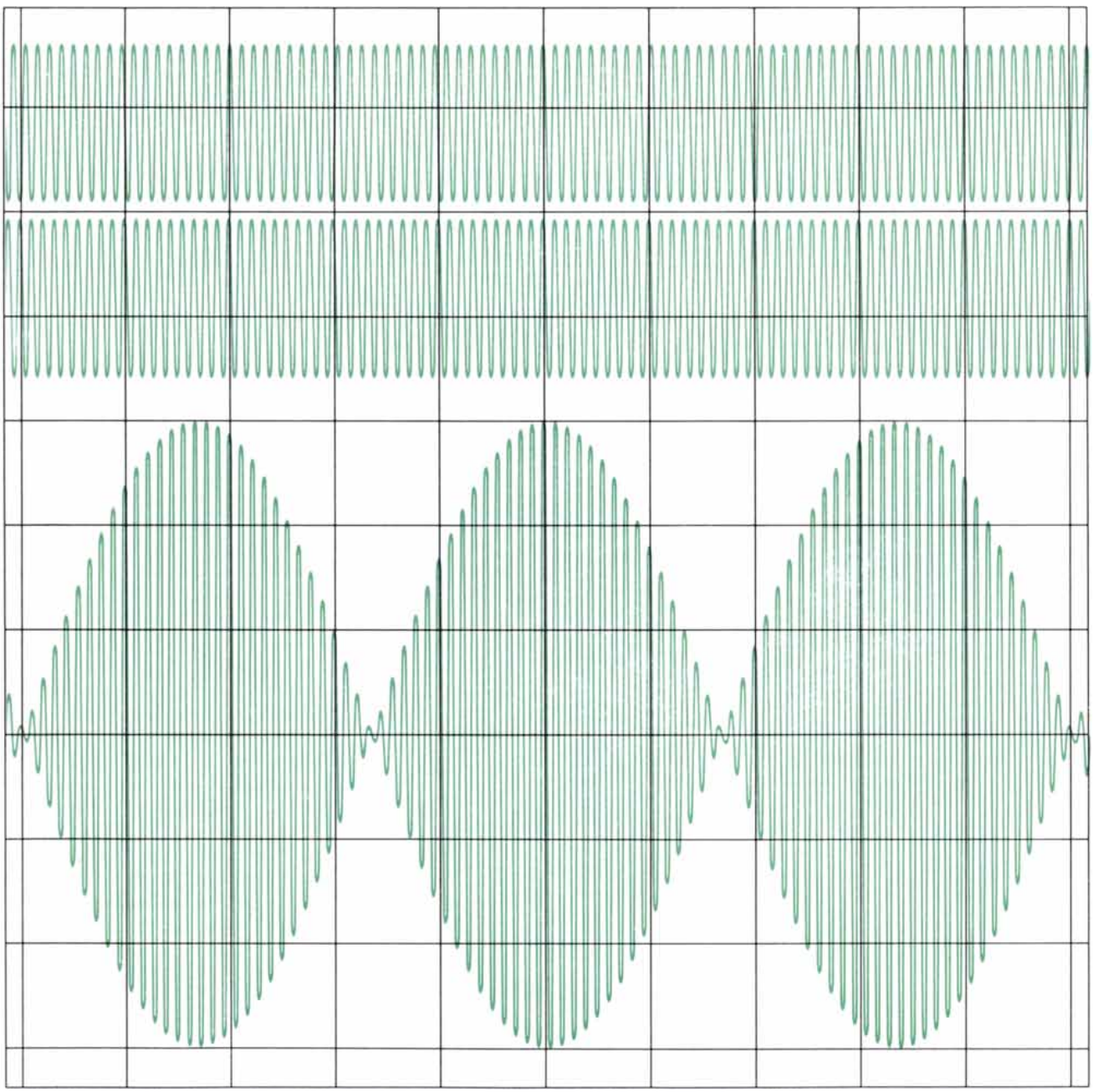
On the other hand, beats are used to advantage where frequencies must be determined precisely. Electrical engineers compare the output of a test oscillator with that of a standard oscillator by detecting the beats produced when their signals are combined. The tuning of pianos is another process that depends on beats. Typically the piano tuner will first listen for the beats produced by a tuning fork of 440 hertz and the *A* above middle *C*, and tighten or loosen the *A* wire until the beats slow to zero. He then strikes the *A* key and the *D* key below it and tunes the latter wire until 10 beats per second are heard. That frequency is produced by the interaction of the *A* string's second harmonic, or second multiple ($2 \times 440 = 880$), and the *D* string's third harmonic ($3 \times 290 = 870$). In this fashion, key by key, the piano is tuned; in theory it could be done even by someone who is tone-deaf.

Binaural beats were discovered in 1839 by a German experimenter named H. W. Dove, but as late as 1915 they were considered a trivial special case of monaural beats. It was argued that each ear was hearing sounds intended for the other. This extraneous result could be eliminated by placing the tuning forks in separate rooms, with the subject in a third room between them, and guiding the sounds through tubes to each ear. It was necessary to carefully seal each tube to the head, however, and another objection was raised: that sound presented to one ear could be conducted through the skull to the other. Bone conduction is well established, and indeed some hearing aids operate on this principle, although sound is attenuated a thousandfold from ear to ear.

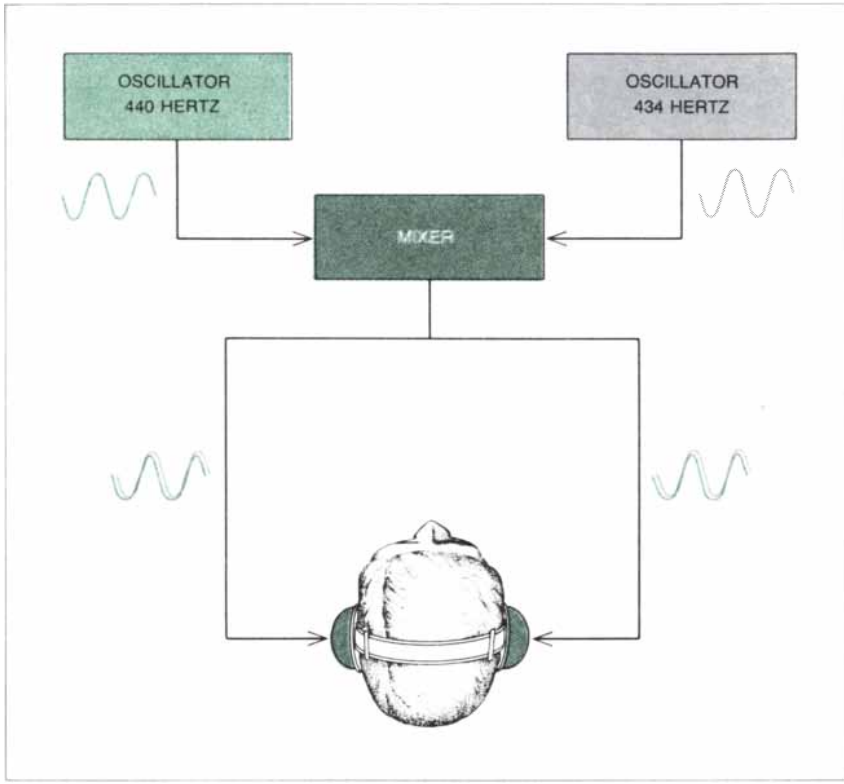
The possible contribution of bone conduction to the perception of binaural beats is eliminated, however, by the use of modern stereophonic earphones. Such earphones have padding, often liquid-filled, to insulate the head from the sound source, and are designed explicitly to prevent conduction effects. Indeed, stereophonic recordings played through earphones can sound unnatural because the instruments seem too isolated.

The difference most immediately apparent between monaural and binaural beats is that binaural beats can be heard only when the tones used to produce them are of low pitch. Binaural beats are best perceived when the carrier frequency is about 440 hertz; above that frequency they become less distinct and above about 1,000 hertz they vanish altogether. No person I have tested reports hearing beats for frequencies above 900 hertz. Experimental conditions, particularly the intensity of the sounds and the type of earphones used, can affect the results, however, and other investigators report detecting beats produced by tones up to almost 1,500 hertz. At the other end of the scale beats also become elusive. Below about 90 hertz the subject may confuse the beats with the tones used to produce them.

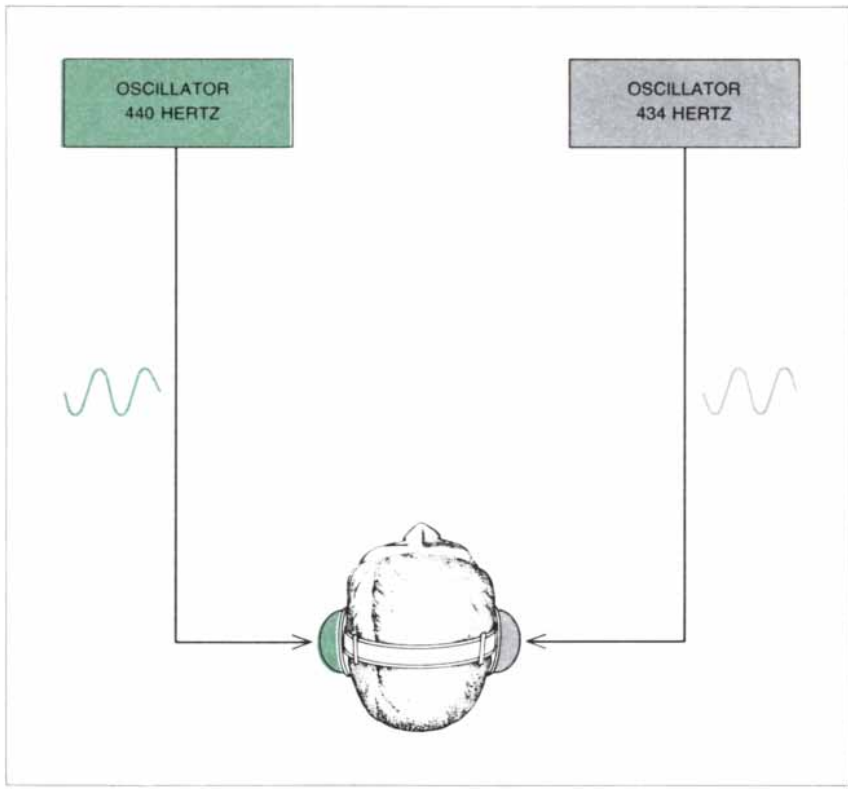
J. C. R. Licklider of the Massachusetts Institute of Technology developed a technique when he was working at Harvard University to measure a spectrum



SINE WAVES COMBINE TO PRODUCE BEATS in this illustration based on oscilloscope traces. The two waves at the top are of slightly different frequency; when they are combined, the resulting wave at the bottom varies slowly in amplitude. The variations are beats and would be perceived acoustically as modulations in loudness. If the two signals were presented separately to each ear, binaural beats would be heard. These differ in character from monaural, or ordinary, beats and are generated within the brain.



EXPERIMENTAL METHOD for generating monaural beats uses two electronic oscillators and a network, here called a mixer, to combine their outputs. Each ear hears a composite signal; the beats can be heard with one ear or both. With the oscillators tuned to the frequencies shown, six beats per second (440 hertz minus 434 hertz) would be perceived.



BINAURAL BEATS are produced when each oscillator is connected separately to one earphone. Again the beat frequency is six hertz, but in this mode the beats are less distinct. Whereas monaural beats are produced by the physical interference of two sound waves, binaural beats are a result of the interaction of auditory signals occurring within the brain.

of binaural beats [*see upper illustration on page 102*]. He adjusted the frequency of one oscillator until the interval was large enough so that the beats seemed "rough"; then he noted the frequency of the unchanged reference oscillator. Next he changed the setting of the reference oscillator and repeated the procedure. In this way the range of perception of each subject was recorded.

Another distinguishing characteristic of binaural beats is their muffled sound. Monaural beats produced with sounds of equal intensity pulse from loudness to silence, as their wave form would suggest. Binaural beats, on the other hand, are only a slight modulation of a loud background. I have tried to estimate the depth of the modulation, and it seems to be about three decibels, or about a tenth the loudness of a whisper. In order to help subjects recognize these relatively faint effects I usually present signals with monaural beats and then suddenly change to the binaural mode. With tones of about 440 hertz it usually takes two or three seconds for the subject to recognize the binaural beats.

To produce a monaural beat that varies from a maximum to complete silence the loudness of the two signals must be identical; if the signals are mismatched, the instantaneous amplitude of the algebraic sum will always be greater than zero. As the difference in intensity increases, the beats become less distinct. Binaural beats, on the other hand, have the same apparent strength regardless of the relative intensities of the two tones. In fact, E. Lehnhardt, a Berlin audiologist, discovered that binaural beats are perceived even if one of the signals is below the threshold of hearing.

J. J. Groen of the State University of Utrecht has studied this phenomenon. Working with tones of about 200 hertz, he found that beats were perceptible when one signal had a loudness of 40 decibels and the other a loudness of minus 20 decibels, a hundredth the loudness of barely audible sound. Evidently the brain is able to detect and process the signals even though one of them is too weak to impinge on consciousness. When the experiment is attempted monaurally, only the louder sound, without beats, is heard.

A perhaps related effect is the interaction of noise and binaural beats. Noise ordinarily masks sounds one wants to hear. For example, "static" sometimes overwhelms a weak radio signal. The perception of binaural beats, however, is enhanced by noise.

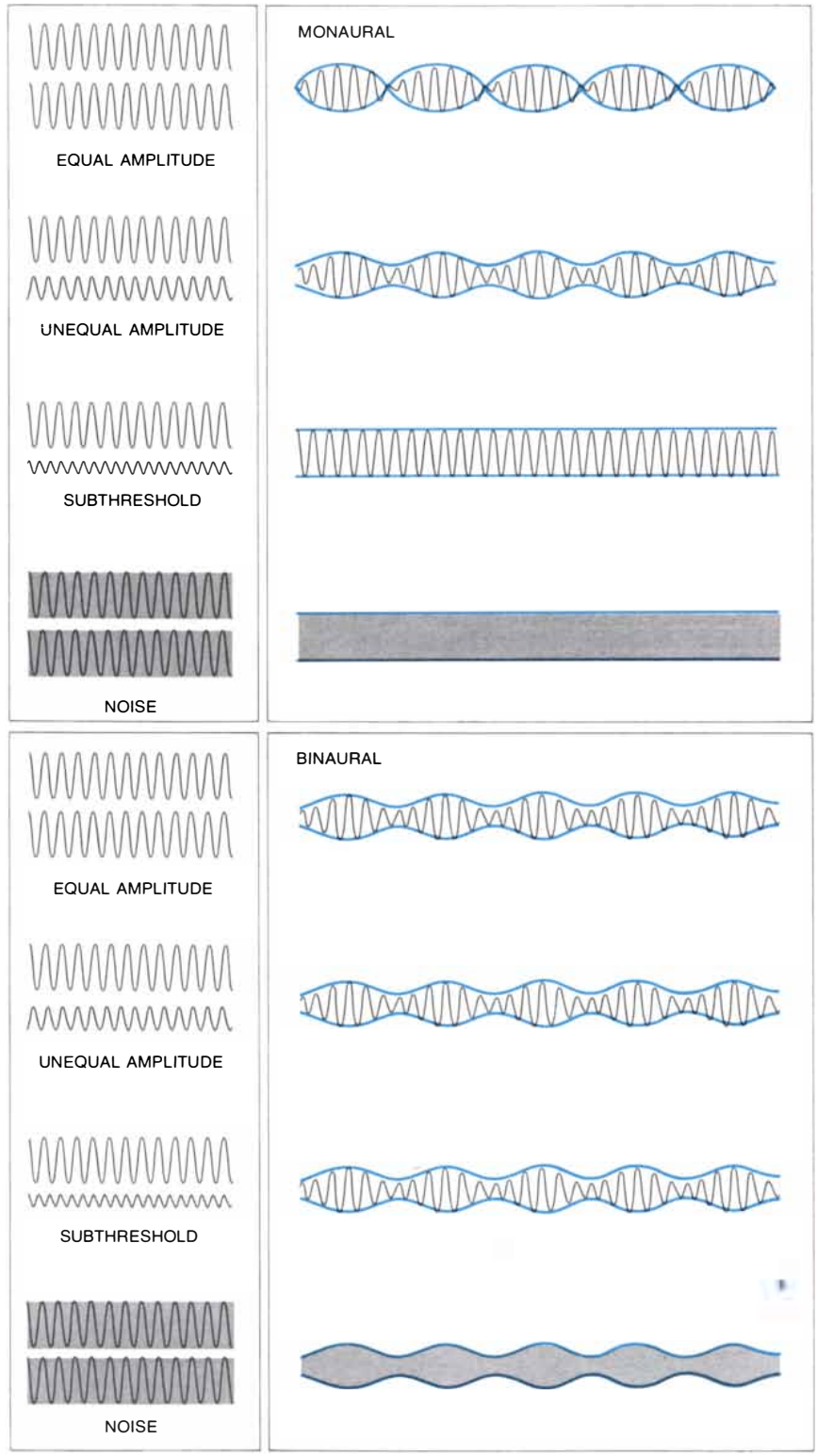
When two appropriate tones are pre-

sented to the ears so that binaural beats are produced, and are accompanied in each ear by noise just loud enough to obscure the tones, the beats become more distinct. In an analogous experiment with monaural signals only the noise will be heard. In the laboratory the source of noise is an electronic device that generates a random signal called white noise, which sounds something like the swish of the wind through swamp grass. When it is added to the signals at the proper loudness, the original tones cannot be heard, but the noise seems to be modulated by the beats. The enhancement of binaural beats by noise is explained by L. A. Jeffress and his colleagues at the University of Texas in terms of chance reinforcement. At any instant the amplitude of the noise will be more likely to be reinforced if the amplitudes of the signals are in coincidence. When the amplitudes of the signals are in opposition, destructive interference is more likely.

Listening to binaural beats produces the illusion that the sounds are located somewhere within the head. This in itself is hardly extraordinary: when music is played through stereophonic earphones, the orchestra seems to be somewhere in the head rather than "out there." It is intriguing, however, that when the beats are very infrequent, fewer than about three per second, they seem to move back and forth in the head. If the intensities of the two tones are different, the motion takes an elliptical path.

This apparent movement may be explained by the connection between binaural beats and the mechanism by which the brain senses the direction of sounds. For low-frequency signals, such as those used to produce binaural beats, sound is localized primarily by detecting the difference in phase between the sounds reaching the two ears [see "Auditory Localization," by Mark R. Rosenzweig; SCIENTIFIC AMERICAN, October, 1961]. Sounds of low frequency have wavelengths much longer than the diameter of the head; as a result the sound travels around the head by diffraction. Lord Rayleigh, the 19th-century English physicist, calculated that a tone of 256 hertz (middle *C*) striking the head from the side would reach the far ear with 90 percent of the intensity it had at the near ear. In other words, the head is not an obstruction to sounds of low pitch, and localization by the detection of relative intensity would be inefficient for those frequencies.

Localization by detection of phase dif-



PERCEPTION OF BEATS depends on the manner in which tones are presented to the ears. In these schematic representations the applied tones at the left can be assumed to be of low pitch and separated in frequency by a small interval. The four diagrams at the top represent the monaural condition. When signals of the same intensity (*equal amplitude*) are combined, the beats vary from loudness to silence. With signals of different loudness (*unequal amplitude*) the intensity of the beats is reduced. When one tone is below the threshold of hearing (*subthreshold*), no beats are perceived. If the tones are accompanied by noise just loud enough to obscure them, again no beats are heard. In the four diagrams at the bottom, representing the binaural condition, the wave forms at the left are the same as those above but are presented to each ear separately. Under these conditions beats are heard whether the signals are of equal or of unequal amplitude and even if one is subthreshold. If noise masks the tones, binaural beats are still heard, modulating the noise.

ferences is highly efficient, however. In an open area with no reflecting structures one can locate a low-pitched sound to within 10 degrees. To do so requires detecting a phase difference of less than one millisecond, a feat accomplished without difficulty by the mechanism of binaural hearing. The same phase difference is present in the tones that produce binaural beats, which is why slow beats seem to be in motion. A source of sound revolving around the head would produce a similar sensation.

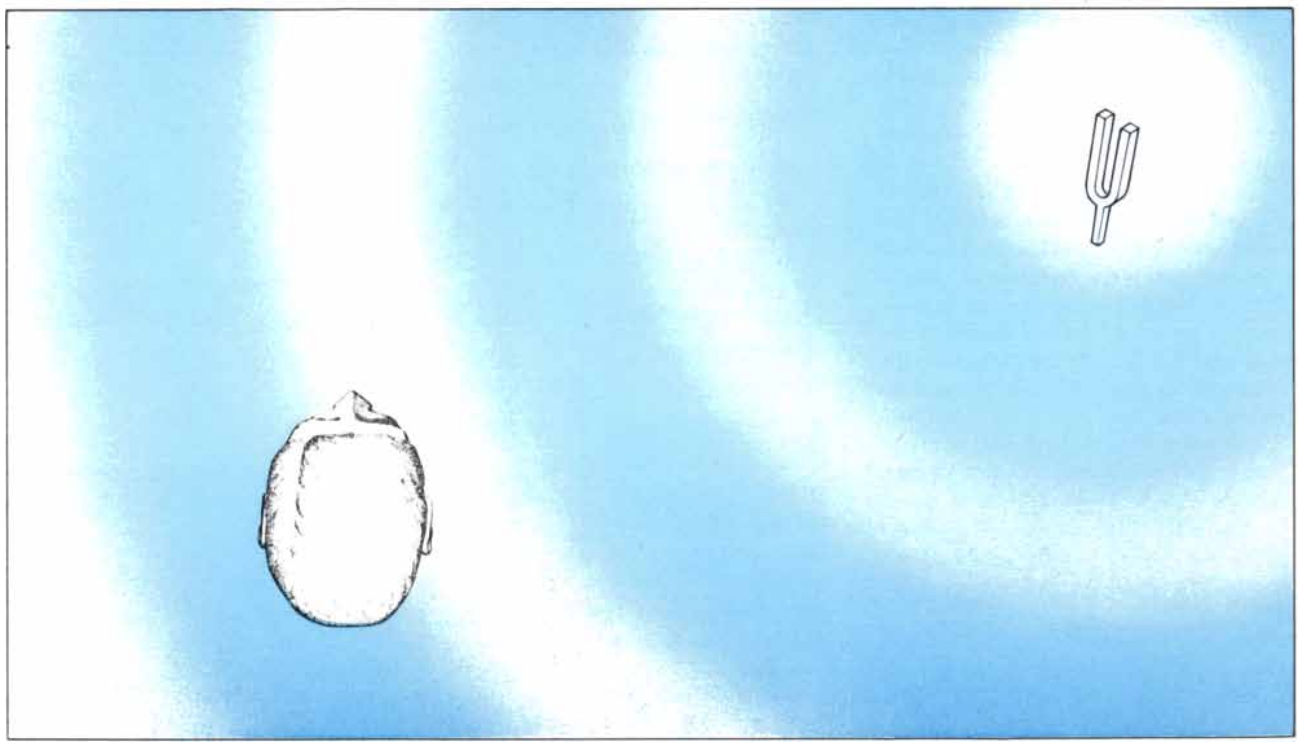
For sounds of higher pitch the wavelength is comparable to or smaller than the size of the head, and the head acts as a barrier, so that the ear in its shadow receives almost no sound. Above about 1,000 hertz sound localization is governed primarily by intensity rather than phase differences. It is significant that the ability to hear binaural beats also wanes when the tones presented approach 1,000 hertz. Direction-finding at the higher frequencies is less accurate than it is for low-pitched tones up to about 8,000 hertz, when the pinna (the external ear) becomes effective as an aid to localization.

The auditory mechanisms manifested in the perception of binaural beats aid human hearing in another way. It has often been observed that the ability to select and listen to a single conversation in a jumble of background noise is a remarkable and valuable human faculty. This phenomenon, sometimes called "the cocktail party effect," is dependent on binaural hearing. It is, in fact, an application of the enhancement of phase perception with noise also seen in the perception of binaural beats.

Hearing generally deteriorates with age. Yet I have found that older people are able to detect binaural beats and to locate sounds almost as well as the young. At 5,000 hertz the auditory acuity of a man of 60 is, on the average, 40 decibels below that of a man of 20, and the highest pitch he can hear, 8,000 hertz, is half that heard by the younger man. His acuity for low tones, however, is barely affected, and evidently his phase perception is also undiminished.

What is the neurological basis of binaural beats? The simplest explanation is that the number of nerve impulses from each ear and the route they travel to the brain are determined by the frequency of the incident sound, and that the two nerve signals interact somewhere in the brain.

One theory of the perception of pitch, called the telephone theory, was proposed by W. Rutherford in 1886. It postulated that the ear converts acoustic vibrations into electrical signals much as a microphone does, emitting one nerve impulse for each cycle of the tone. Single nerve fibers can respond to such stimuli only up to about 500 hertz, however, so that the telephone theory could describe the behavior of the ear only for the lowest frequencies. In 1865 Hermann von Helmholtz proposed the place theory, which ascribed pitch discrimination to the mechanical properties of the cochlea, or inner ear. The cochlea is a cone-shaped, fluid-filled vessel, rich in nerve endings and coiled like a snail shell. ("Cochlea" is Latin for snail.) The coiled tube of the cochlea is divided in half along its length by the basilar membrane, which vibrates in response to sound. Georg von Békésy found by direct visual observation that a sound of a certain frequency will make the basilar membrane bulge most noticeably in a certain place [see "The Ear," by Georg von Békésy; SCIENTIFIC AMERICAN, August, 1957]. This local stimulation, it is believed, excites receptor cells in the vicinity of the bulge and thus excites the nerve fibers connecting the receptor cells to the auditory area of the brain. Accord-



LOCATION OF THE SOURCE of a sound is determined for low-pitched tones by detecting the difference in phase between signals arriving at each ear. In this illustration a compression wave has reached the left ear while the right is near a maximum of rarefaction. By detecting such a phase difference the ears can find the direction of a low-frequency tone to an accuracy of about 10 degrees. At these frequencies little sound is blocked by the head; the wavelength is larger than the head and sound is diffracted around it.

ing to the place theory, the impulses transmitted by the auditory nerves reflect the intensity of the sound but not the frequency; what pitch is perceived is determined by the place on the cochlea where the nerve originates.

Above about 5,000 hertz the place theory seems to be adequate to describe pitch perception. At lower frequencies, however, the mechanical response of the basilar membrane is too unspecific to account for the precision with which the ear identifies tones. Furthermore, attempts to test the theory by excising in experimental animals those nerve fibers that should be the sole carriers of low-frequency tones have been unsuccessful.

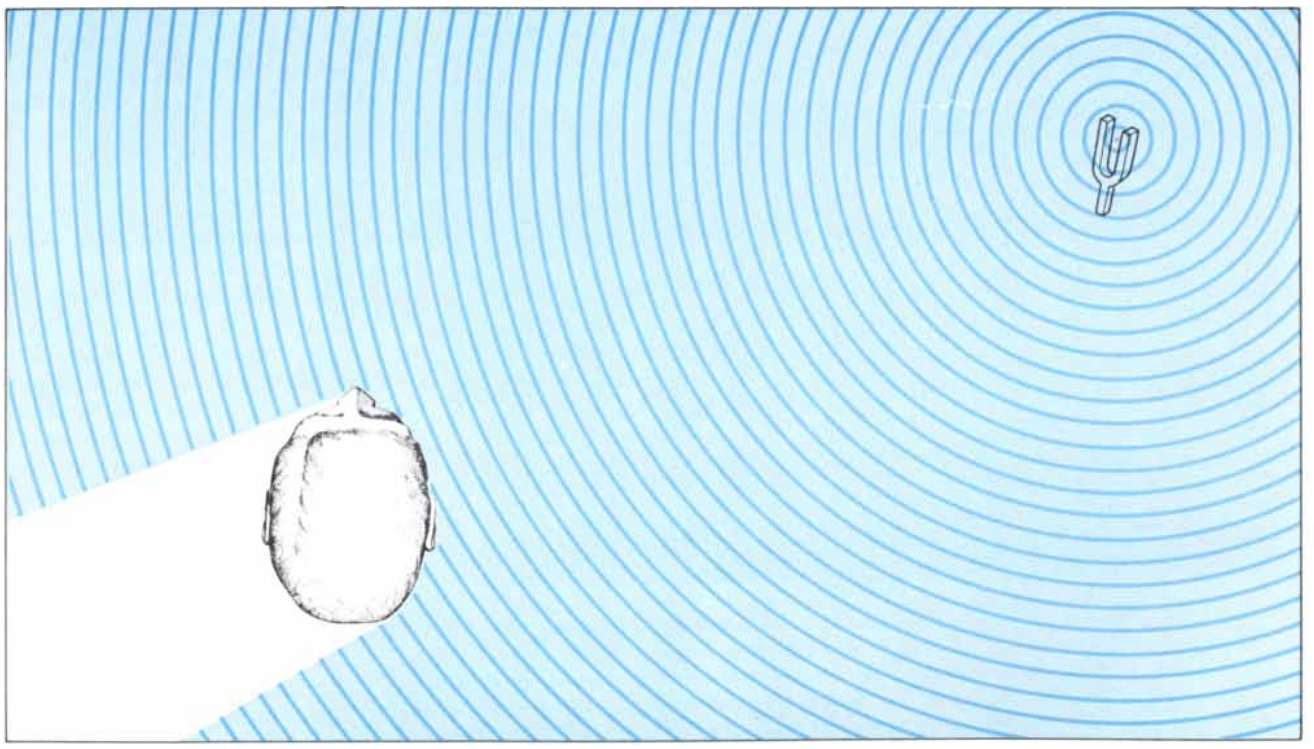
For the frequencies between 500 and 5,000 hertz Ernest Glen Wever of Princeton University in 1939 proposed the volley theory. Although individual nerve fibers cannot fire more than 500 times per second, a group of nerve cells could exceed this rate by firing in succession, Wever suggested, much as platoons in an infantry company could fire their weapons in successive volleys. Thus while some nerve cells are in their refractory period others are producing pulses. The fading of binaural beats at frequencies between 500 and 1,000 hertz suggests that the mechanism of the beats follows the telephone theory and, at the higher frequencies, follows the related volley theory.

Interaction of the signals from the two ears probably occurs at the brain center named the superior olivary nucleus. As the messages ascend the auditory pathways to be processed and interpreted at higher centers, this is the first center in the brain to receive signals from both ears [*see illustration on page 101*]. Actually there are two superior olivary nuclei; they are arrayed symmetrically on each side of the brain, and each is a terminus for nerve fibers from both ears. They have long been considered likely sites for the neural processing of low-frequency sound impulses.

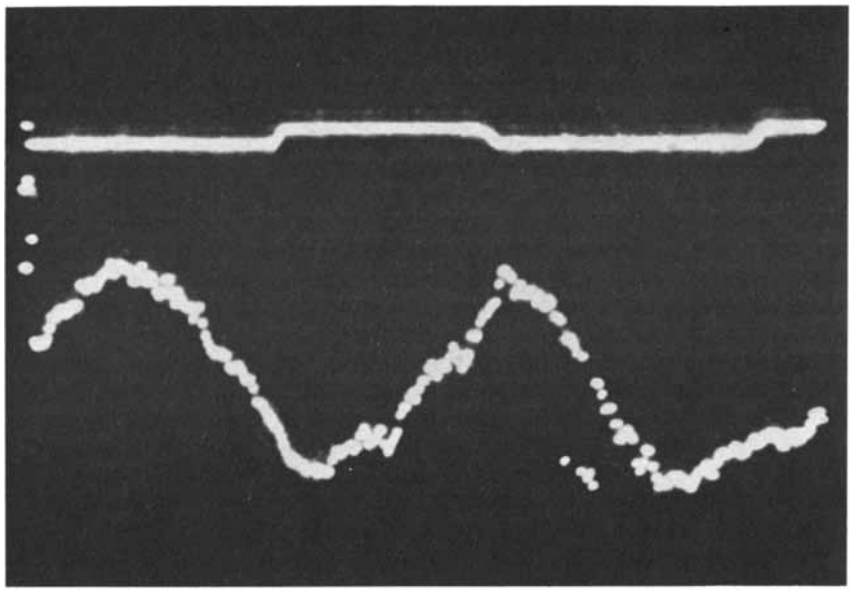
In experiments with cats Robert Galambos showed in 1959 that loud clicks stimulating both ears generate nerve impulses that meet in the superior olivary nucleus. When the clicks are simultaneous, the signals are reinforced at some site in the superior olivary nucleus. When a slight delay is introduced, however, the resulting signal is inhibited. Thus a small phase shift gives rise to a weaker perception of sound. It is presumably for this reason that one tends to turn toward the source of a sound and eliminate the phase difference. When one is listening through earphones, of course, turning the head has no effect on the phase of the signals.

Nerve potentials at the superior olivary nucleus of the cat have been measured directly. With human subjects it is possible to measure these signals by recording evoked potentials: small changes in the electrical properties of the scalp produced as a result of activity of the brain. Because they are objective indicators of certain brain functions evoked potentials have clinical applications. For example, in cases of possible hysterical blindness evoked potentials from the scalp above the occipital lobes can determine whether or not the brain is receiving visual information. Similarly, evoked potentials can be used to detect deafness in infants, which is otherwise quite difficult to diagnose. The potentials are very small (measured in microvolts) and are obscured by many random signals not associated with the stimulus. They can be measured on an oscilloscope, but special procedures must be followed.

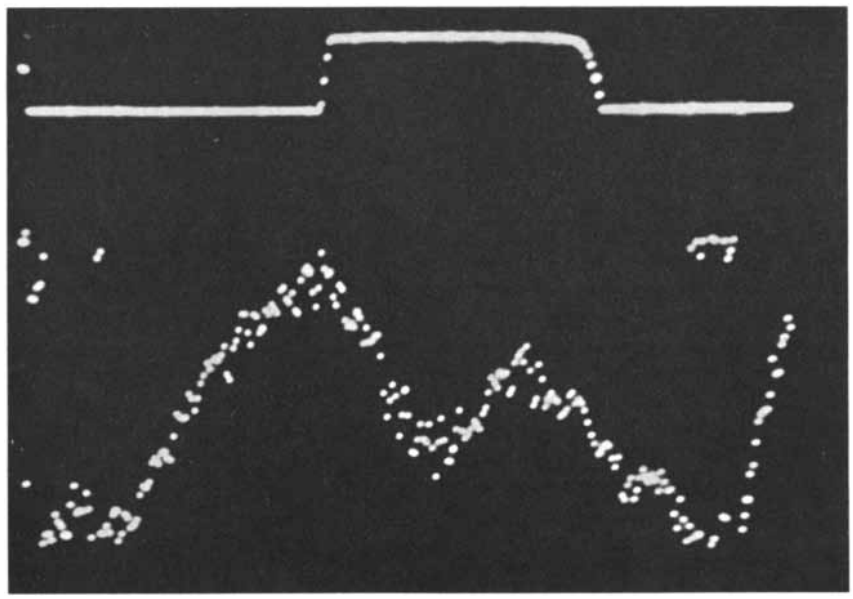
First, the horizontal sweep of the oscilloscope must be synchronized with the stimulus; this is done by using the stimulus current to trigger the start of



HIGH-FREQUENCY SOUND LOCALIZATION also requires binaural hearing, but differences in intensity rather than phase are detected. The wavelength of a high-pitched tone is smaller than the diameter of the head and a distinct sonic shadow is formed; thus one ear receives more sound than the other. This mode of sound localization is less accurate than phase detection except at very high frequencies. The transition takes place at about 1,000 hertz; at this frequency too the perception of binaural beats wanes.



EVOKED POTENTIALS for subjects listening to monaural beats are shown in a photograph of an oscilloscope screen made by the author with Adam Atkin and Neil Wotherspoon at the Mount Sinai School of Medicine. Tones of 300 hertz and 303 hertz were presented to each ear; an electrode attached to the scalp was used to measure electric potentials in the skin evoked by underlying electrical activity in the brain. By synchronizing the horizontal sweep of the oscilloscope with the beat frequency it was possible to correlate these small potentials (measured in microvolts) with the stimulus. The steplike wave form at the top of the screen is a signal used to time the oscilloscope sweep; the rise of each pulse corresponds to the moment of maximum loudness of the beat. The periodic wave below it records the evoked potentials. It consists of the average values for each point on the curve, determined by a small computer (a signal-averager) after many iterations of the procedure.



BINAURAL EXPERIMENT was conducted under the same conditions, except that the tones were presented separately to each ear. Evoked potentials were once again successfully recorded, but they differed from those detected under monaural conditions in amplitude, in wave form and in timing with respect to the stimulus. These differences suggest that binaural beats are processed in another way or at another site in the brain than monaural beats are. In the illustration the amplitude of the evoked potentials appears to be about the same as it is for monaural beats; it is actually much smaller. For clarity the vertical scale of the oscilloscope has been expanded, as can be seen by comparing the apparent amplitudes of the timing signals. In both illustrations bright areas not associated with the main wave form are extraneous signals produced by residual noise generated in the recording apparatus.

the sweep. In addition, a great many tracings must be made in order to obtain unambiguous data. A computer known as a signal-averager stores a series of tracings electronically, then on command adds the instantaneous potentials of all the tracings to produce a composite signal. Because the extraneous random potentials have no fixed phase relation to the stimulus they are progressively suppressed as the number of tracings increases.

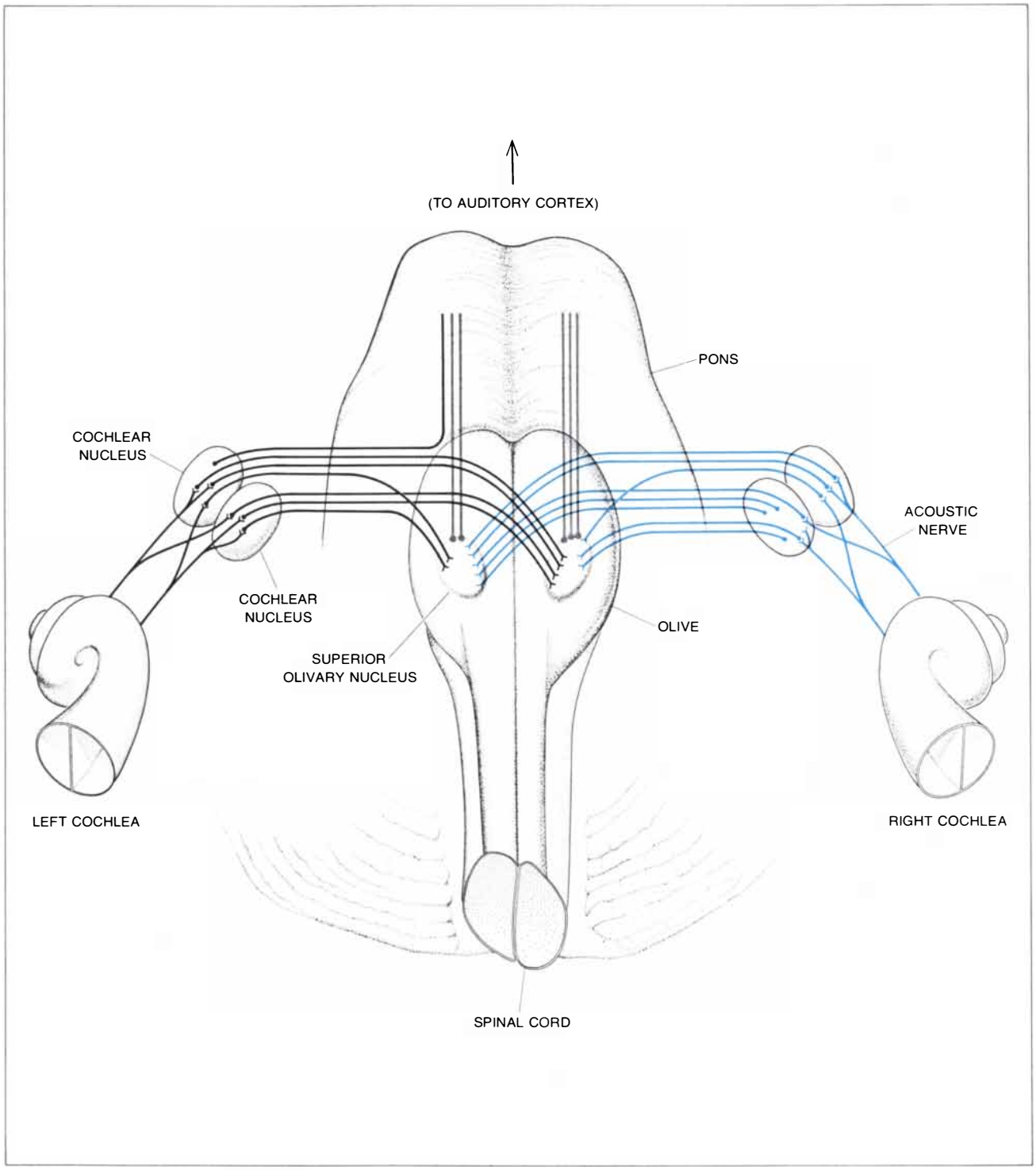
If binaural and monaural beats are indeed processed at different sites in the brain, it should be possible to detect this difference by measuring the evoked potentials. With my colleagues at the Mount Sinai School of Medicine, Adam Atkin and Neil Wotherspoon, I set out to test this hypothesis. Because the stimulus was a continuous tone rather than a brief click, it proved particularly difficult to obtain clear tracings. Eventually we learned that for effective results the subject must concentrate on the beats while in total darkness. This is a boring task, since the binaural beats are indistinct and many tracings must be averaged. Often the subject experiences auditory hallucinations imposing a spurious pattern on the sound, which spoils the results. Nevertheless, after many iterations of this procedure we were able to demonstrate that the evoked potentials produced by binaural and monaural beats differ qualitatively and quantitatively, indicating that they are processed differently [*see illustrations at left*].

Binaural beats may have clinical applications. With some of my students I examined a number of neurological patients and discovered that a few could not hear binaural beats. Among these patients a few could not localize sounds (produced by the examiner's snapping his fingers). It may be significant that some of those who could not hear the beats suffered from Parkinson's disease, a disorder of the central nervous system characterized by a lack of spontaneous muscular activity, an immobile facial expression and tremor. One patient, a violinist, was unable to hear binaural beats when he entered the hospital. As his treatment continued he began to perceive the beats produced by the very lowest tones, and gradually he progressed to higher frequencies. At the end of a week, when his condition was considered satisfactory, he could hear beats produced by tones up to about 650 hertz.

A sex-related variation in the ability

to hear binaural beats has also been discovered. J. V. Tobias of the Federal Aviation Administration in Oklahoma City studied the binaural-beat spectrum of a number of volunteers and found that the upper limit of the applied frequencies is higher for men than for women [*see lower illustration on next page*]. He went on to monitor the perceptions of three women over a period of six weeks and found that the spectrum extended to the highest tones at the beginning of menstruation, then declined before reaching a second peak 15 days after the onset of menstruation. The latter peak may correspond to the time of ovulation, when a woman is most fertile.

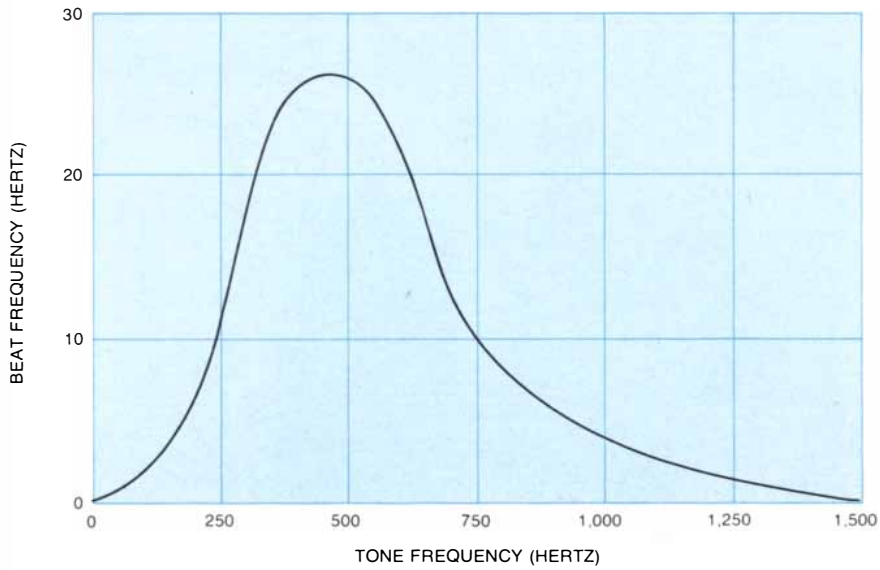
I have tested a few women of reproductive age, with results that tend to confirm Tobias' findings. It appears that some women do show marked variations



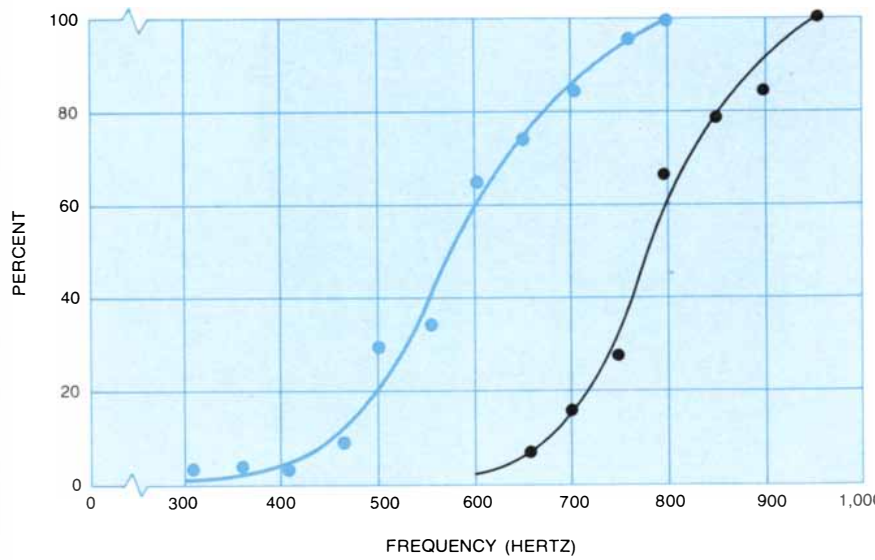
LOWER AUDITORY CENTERS of the brain are in the medulla oblongata, viewed here schematically from the back of the neck. Nerve impulses from the right (*color*) and left (*black*) ears first meet in the left or right superior olivary nucleus. These structures are part of the olive, an organ that in this view lies behind the brain stem. It is probable that binaural beats are detected here.

in the perception of binaural beats during the menstrual cycle. When the beats are not heard, women often hear two separate tones. Men, on the other hand, show no variation during the month. These results suggest that the binaural-beat spectrum may be influenced by the level of estrogen in the blood.

Binaural beats have been widely regarded as a mere curiosity. A recent textbook on hearing does not mention them at all. Yet the measurement of binaural beats can explain the processes by which sounds are located, a crucial aspect of perception. The enhancement of the beats by noise is a model of the mechanism by which auditory messages are sorted from a noisy background. That subthreshold sounds are effectively rendered audible by binaural beats suggests that there may be other stimuli processed by the brain of which we are not aware. Finally, it is possible that hormonally induced physiological or behavioral changes too subtle to detect by ordinary means may be made apparent by measuring the binaural-beat spectrum.

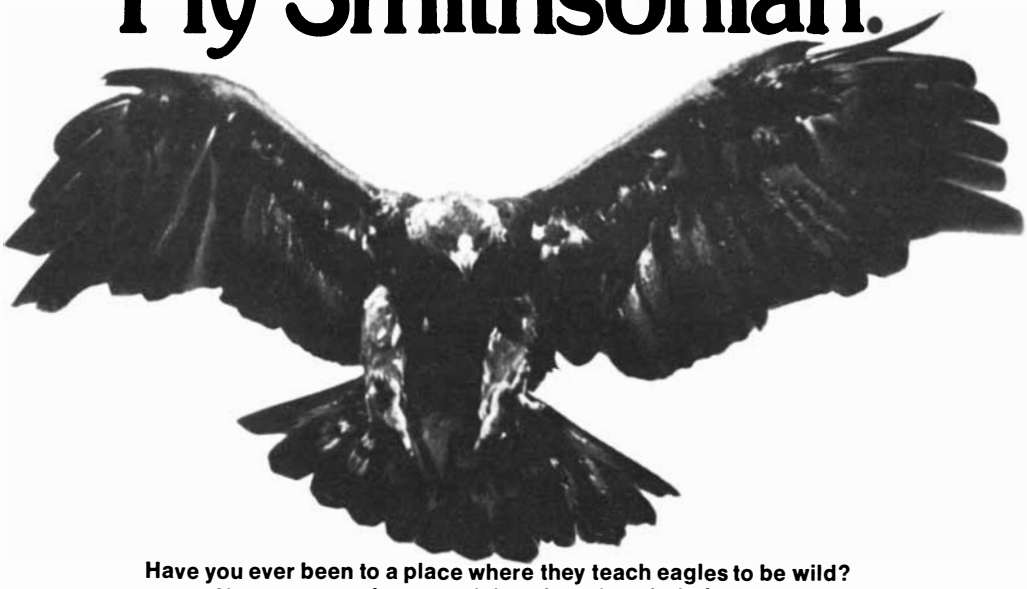


SPECTRUM OF BINAURAL BEATS was measured by J. C. R. Licklider, J. C. Webster and J. M. Hedlum. Rapid beats, up to about 26 per second, can be heard when the tones used to produce them are about 440 hertz. With tones of higher or lower pitch the maximum beat frequency declines. When the interval exceeds about 30 hertz, two tones are heard.



SEX-RELATED VARIATION in the perception of binaural beats is plotted from data compiled by J. V. Tobias. As the pitch of the tones used to produce beats increases, both men (*color*) and women (*black*) cease to perceive them. Women, however, lose the ability at lower frequencies. Some female subjects also report variations during the menstrual cycle.

Fly Smithsonian.



Have you ever been to a place where they teach eagles to be wild?
Have you ever journeyed deep into the mind of a man
who single-handedly revolutionized Man's entire perception of the universe?
Have you ever penetrated to the heart of the Cretan Maze, to confront the Minotaur?

Visit some of the world's most enlightening destinations.
Fly with SMITHSONIAN Magazine for the next 12 months... for less than one dollar a month.

We invite you to spread the wings of your imagination every month, and come with us into many surprising worlds of nature, art, history and science. And we invite you to bring your entire family along.

When did Man first set foot on the North American continent—and think about when Man is likely to first set foot on Mars. Is rainmaking becoming respectable? Why is it essential to save the prairie dog from extinction? What *really* happened the night that Paul Revere made his famous ride?

What happens when the gold-bedecked chieftains of the Akan pay a state visit to their new King? What is currently the rarest stamp in the world? Were dinosaurs the only warm-blooded reptiles that ever lived? What causes the brilliant colors of butterfly wings? How close are we to directly harnessing the power of the Sun?

If these questions intrigue you, you'll be intrigued and pleased with SMITHSONIAN. And if you are attracted and moved by beauty, you'll love SMITHSONIAN the more... because it is one of the handful of magazines in the world today that is almost a work of art in itself. Here every month you will find a dazzling display of beauties both ancient and modern—from the most recently discovered Pompeiian frescoes to the finest works of modern masters (many of them never seen outside private collections before).

What's more, as beautiful, informative, intelligent and entertaining as SMITHSONIAN Magazine is, it's *only one* of the knowledge-enhancing benefits you and your family receive when you become a National Member of the Smithsonian Associates.

Much more than a magazine

In addition to SMITHSONIAN every month, your membership in the Smithsonian Associates entitles you and your family to:

■ **Special travel programs** organized for Members only—nearly 50 opportunities each year to visit historic sites, art centers and scientific installations (many of which are not open to the general public).

■ **Discounts on quality books** dealing with history, art, science and nature.

■ **Discounts on unusual gifts** from the museum shops of the Smithsonian Institution... fascinating arts and crafts from all corners of the globe.

■ **Special welcome** when you and your family visit the Smithsonian in Washington, D.C. including use of the Member's Lounge and the privilege of dining with the Smithsonian staff.

The contribution you are making to knowledge

Beyond the extensive, almost unlimited, resources and opportunities for learning that your Membership will bring to you and your family, a substantial portion of your annual $10 Membership Fee goes *to help increase the fund of Man's knowledge.*

It goes to help support the many research and educational programs of the Smithsonian Institution's private sector which depends on the private support of individuals for its activities.

Join today— return the coupon below

If you are interested in exploring many new regions of the world and of the mind, if you respond to the historic and the beautiful and the innovative, if you would like to help increase the range of Man's mind, join the Smithsonian today. All it takes is a few brief moments to fill out and mail the coupon below to spread the wings of your mind.

Mail this coupon to:

SMITHSONIAN Membership Service
P.O. Box 2606 • Greenwich, Conn. 06830

Please enroll me as a National Member of the Smithsonian Associates, and enter my subscription for the next year (12 issues) of SMITHSONIAN.

Bill me for the annual dues of $10.

Name *(please print clearly)*

Address

City

State Zip

MEMBERSHIP BENEFITS

- SMITHSONIAN Magazine — 12 issues
- Travel Program — Domestic and Foreign
- Book and Gift Discounts
- Reception Center at Smithsonian Institution in Washington, D.C.

P6-903

ELECTRON-POSITRON COLLISIONS

When the electron and its antiparticle collide, they can annihilate each other and give rise to radiation or other particles. Such processes are currently studied with opposed beams of electrons and positrons

by Alan M. Litke and Richard Wilson

What happens when a speeding particle of matter collides head on with a speeding particle of antimatter? The question is currently being investigated in several high-energy physics laboratories around the world, where experimenters are examining the various types of interaction that can take place between two oppositely directed, intersecting particle beams, one consisting of accelerated electrons and the other consisting of accelerated positrons, or antielectrons.

The answer, based on the experiments already completed, is a tantalizing blend of the expected and the unexpected. So far the experiments have tended to confirm the validity at high energies of the fundamental theory of electromagnetic phenomena known as quantum electrodynamics. In addition they have revealed that collisions in which electrons and positrons annihilate each other to form the strongly interacting nuclear particles called hadrons happen much more often than anticipated. The latter finding has important implications for our understanding of the elementary particles.

The electron is of course the familiar constituent of atoms that circulates about the much more massive atomic nucleus. It has a negative electric charge, a mass of 9.1×10^{-28} gram and a spin, or angular momentum, of 1/2 (in units of $h/2\pi$, where h, the fundamental physical quantity known as Planck's constant, is equal to 6.6×10^{-27} erg-second). The electron and the positron are antiparticles of each other. They have the same mass and spin, and the same magnitude, but opposite sign, of electric charge. Hence when a positron is placed in an electric field or when it is moving through a magnetic field, it will "feel" a force that is exactly opposite to the one felt by an electron.

The existence of an antimatter equivalent of the electron was originally postulated 45 years ago by P. A. M. Dirac. At the time Dirac was trying to develop a quantum-mechanical theory of the electron that would satisfy Einstein's special theory of relativity. The best previous formulation was inadequate because it did not satisfy the requirement of the special theory that space and time be treated symmetrically. Dirac found that his solution not only was successful in describing many known phenomena but also predicted states of the electron with negative energy. At first these hypothetical energy states were ignored. Later Dirac realized that they could be explained by assuming that under ordinary circumstances the states are filled with electrons.

If such a negative-energy state is not filled, however, then according to Dirac's equations a "hole" would be created. This hole would behave like a particle identical in every respect with the electron except that it would be positively charged. Dirac called his hypothetical particlelike holes positrons. If a positron is to be produced, he argued, an electron must in effect be removed from the negative-energy "sea," leaving behind a hole. It followed that a positron must always be produced in a pair with an electron.

Soon afterward Carl D. Anderson succeeded in producing electron-positron pairs. When he illuminated a lead plate in a cloud chamber with gamma rays from a radioactive source, pairs of tracks could be seen originating from the same point in the lead. In the presence of a magnetic field the tracks were observed to bend in opposite directions. The process clearly represented a conversion of radiant energy into matter. The energy of the electromagnetic field associated with the gamma ray had been converted into the mass energy and the energy of motion of an electron-positron pair.

It is now believed that every particle has a corresponding antiparticle (although some electrically neutral particles such as the photon and the neutral pion are their own antiparticle), and that any particle-antiparticle pair can be produced by a gamma ray of sufficient energy. Conversely, any particle, on meeting its antiparticle, can make a transition to the negative-energy state, annihilating both particles and producing radiant energy.

In the experiments we shall describe physicists have studied the mutual annihilation of the electron and the positron. If the initial particles are sufficiently energetic, the annihilation can give rise to particle-antiparticle pairs as well as radiant energy. These experiments are a modern kind of alchemy in which an electron-positron pair can be converted into another type of particle-antiparticle pair, such as a proton and an antiproton. This kind of reaction comes as no surprise to a physicist. It is a manifestation of the interchangeability of matter and energy, the concept first put forward by Einstein in 1907.

Such experiments call for a beam of electrons to be directed against a beam of positrons, so that the particles can collide at high energy. They could not be undertaken until a number of difficult technical problems were overcome. One of the first problems encountered was how to obtain enough positrons. It is easy enough to produce beams of electrons (the large number of television sets in the world attests to that), since they readily "boil" off a hot tungsten filament, but positrons, on this planet at least, are few and far between. Some positrons are generated in cosmic ray collisions, and some are emitted by certain radioactive

nuclei, but these sources are not strong enough to generate a positron beam.

The positron-production problem can be solved by the following stratagem. If one allows a beam of high-energy electrons to impinge on a target such as tungsten, then inside the target the electrons, accelerated by the electric field of the target nuclei, will radiate electromagnetic energy, and this radiation will consist in part of gamma rays. As the gamma rays travel through the target they can give rise to electron-positron pairs, and these particles can in turn produce more gamma rays. The result is an electromagnetic shower, a cascade phenomenon in which many electrons, positrons and gamma rays are created. The last step is to reject the electrons and to capture the positrons emerging from the target. Since the particles are oppositely charged, they can be separated by means of a magnetic-lens system.

In principle the resulting beam of positrons could now be allowed to impinge on any target containing electrons; there are, after all, some 10^{24} electrons in a cubic centimeter of solid matter. In that case, however, not only would one get a background effect from collisions of the positrons with the atomic nuclei in the target but also the effective energy of the positron-electron collisions would be quite low.

Until 1965 all high-energy physics experiments involved the bombardment of a stationary target with a beam of energetic particles. That technique, however, is wasteful of energy, as can be demonstrated with the aid of the following analogy. A collision between two automobiles going in the same general direction is not particularly energetic and little damage may be done; a collision between a moving automobile and a stationary one is more energetic and can be more serious; a head-on collision between two speeding automobiles is highly energetic and is usually catastrophic. The "useful" energy derived from any of these collisions (that is, the amount of mass energy and energy of motion of the colliding vehicles that can be converted into other forms of energy, such as heat) is described quantitatively by calculating the total initial energy perceived by an observer who sees the two automobiles coming with equal and opposite momenta. An observer who looks at a collision in this way is said to be in the center-of-mass frame of reference.

In the case of a collision between a stationary particle and a particle of high energy the useful energy may be only a small fraction of the initial energy (a fraction much smaller than the preceding analogy suggests). This is a consequence of the relativistic laws of the conservation of energy and momentum. If two particles with equal and opposite momenta collide head on, however, the useful energy is equal to the initial energy. For example, a 2,500-million-electron-volt positron colliding with a stationary electron produces a useful center-of-mass energy equal to only 2 percent of its initial energy, making available only 50 MeV for conversion into new particles. On the other hand, when positrons in a 2,500-MeV beam collide head on with electrons in a 2,500-MeV beam, the full energy of the particles in both beams—5,000 MeV—is available for the production of new particles.

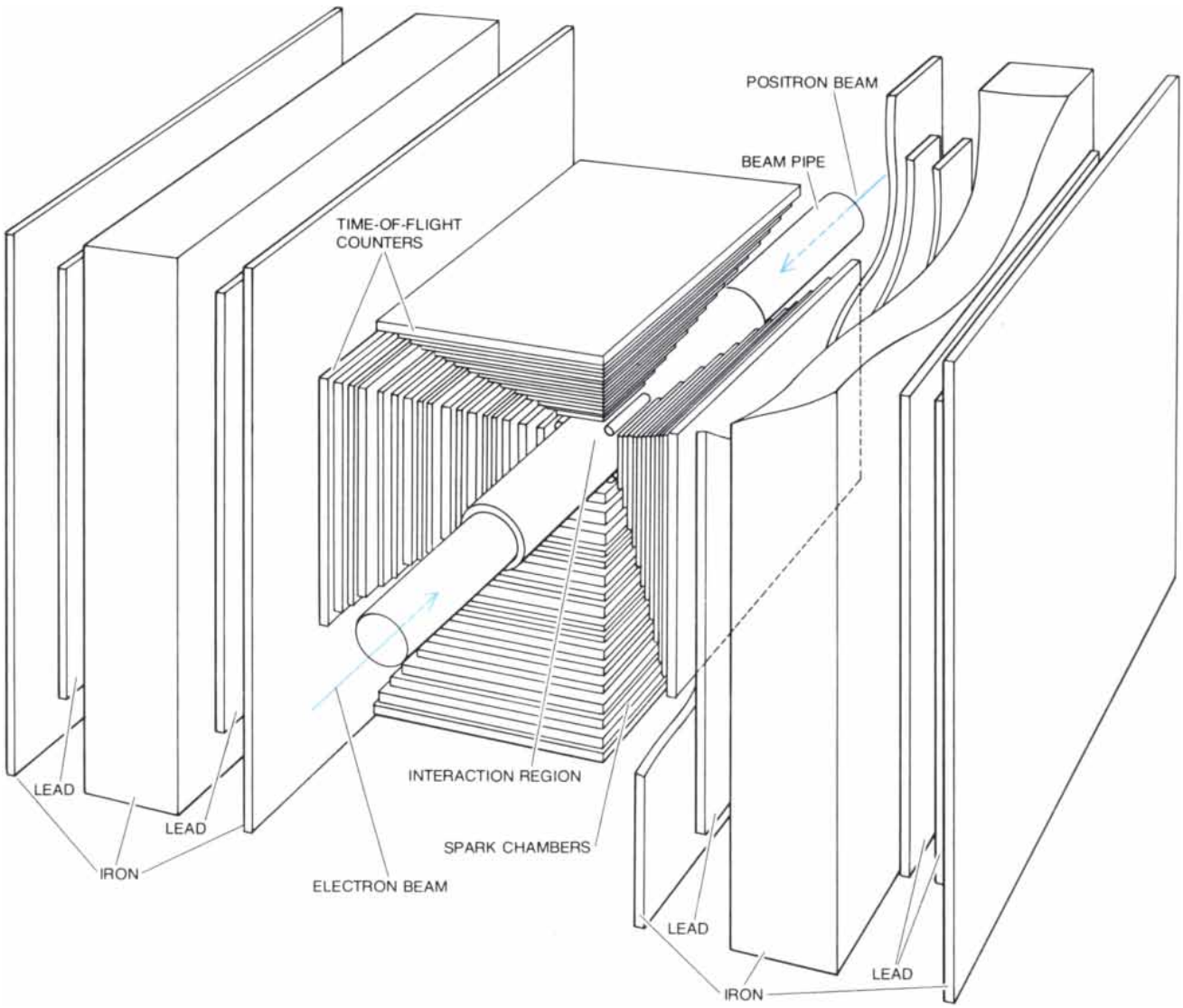
STORAGE RING at the Stanford Linear Accelerator Center is one of half a dozen such rings in the world used for experiments involving colliding beams of electrons and positrons. The ring, known as SPEAR, is visible at the lower right. It is one-seventh of a mile in circumference. Electrons and positrons are first accelerated in the two-mile linear accelerator seen in the distance. Then they are injected into the ring, where they circulate in opposite directions and collide with each other at points known as interaction regions.

Hence the most efficient way to obtain high energies in the center-of-mass frame of reference is to arrange matters so that one beam meets another head on. This objective, however, leads to another major technical difficulty: how to obtain collisions between the particles in the beams at a rate high enough to enable one to study the very rare annihilation reactions in which one is interested.

How rare are these processes and what collision rate does one need for their investigation? The probability that there will be a reaction between any two particles is expressed in terms of the cross section for the reaction. The reaction cross section can be thought of as the area in the vicinity of one of the particles (considered as a target) within which the forces responsible for the reaction will be felt by the other particle (considered as a projectile). For the reactions of interest here the cross sections are small indeed. They range from about 10^{-30} to 10^{-34} square centimeter.

A process generated with colliding beams will have a rate equal to the reaction cross section multiplied by a factor known as the luminosity. The luminosity is a characteristic of the machine in which the particles are made to collide. It is a measure of how frequently a particle in one beam passes close to a particle in the other beam.

In order to get a luminosity high enough to make it possible to study the rare annihilation reactions it is advantageous to give the particles in one beam repeated opportunities to collide with those in the other. That can be done by storing and accumulating the particles in a circular apparatus called a storage ring. Successive pulses of particles are injected into the ring and are made to circulate in a closed orbit for many revolutions. Two beams can circulate in separate rings that intersect at one or more points, or, in the case of particles of opposite electric charge, the beams can circulate in opposite directions in the same ring. Collisions between the particles in opposing beams take place at locations known as interaction regions, which must be at points where the orbits intersect. Electron-positron storage rings are now in operation or under construction at Novosibirsk in the U.S.S.R., Hamburg in



BYPASS ON-LINE DETECTOR (B.O.L.D.) of the Cambridge Electron Accelerator in Cambridge, Mass., was used by the authors and their colleagues to study what happens when particles of matter and antimatter collide. A beam of electrons at an energy of 2,000 million electron volts (MeV) entered the beam tube from the left, and a beam of positrons at the same energy entered it from the right. The region of their interaction was surrounded by detection equipment consisting of spark chambers, scintillation counters and metal plates arranged in four similar quadrants. B.O.L.D. has not been in use since June, when the accelerator was closed down.

Germany, Frascati in Italy, Orsay in France and Stanford University in the U.S. The first electron-positron storage ring in the U.S., the Cambridge Electron Accelerator in Cambridge, Mass., was shut down in June.

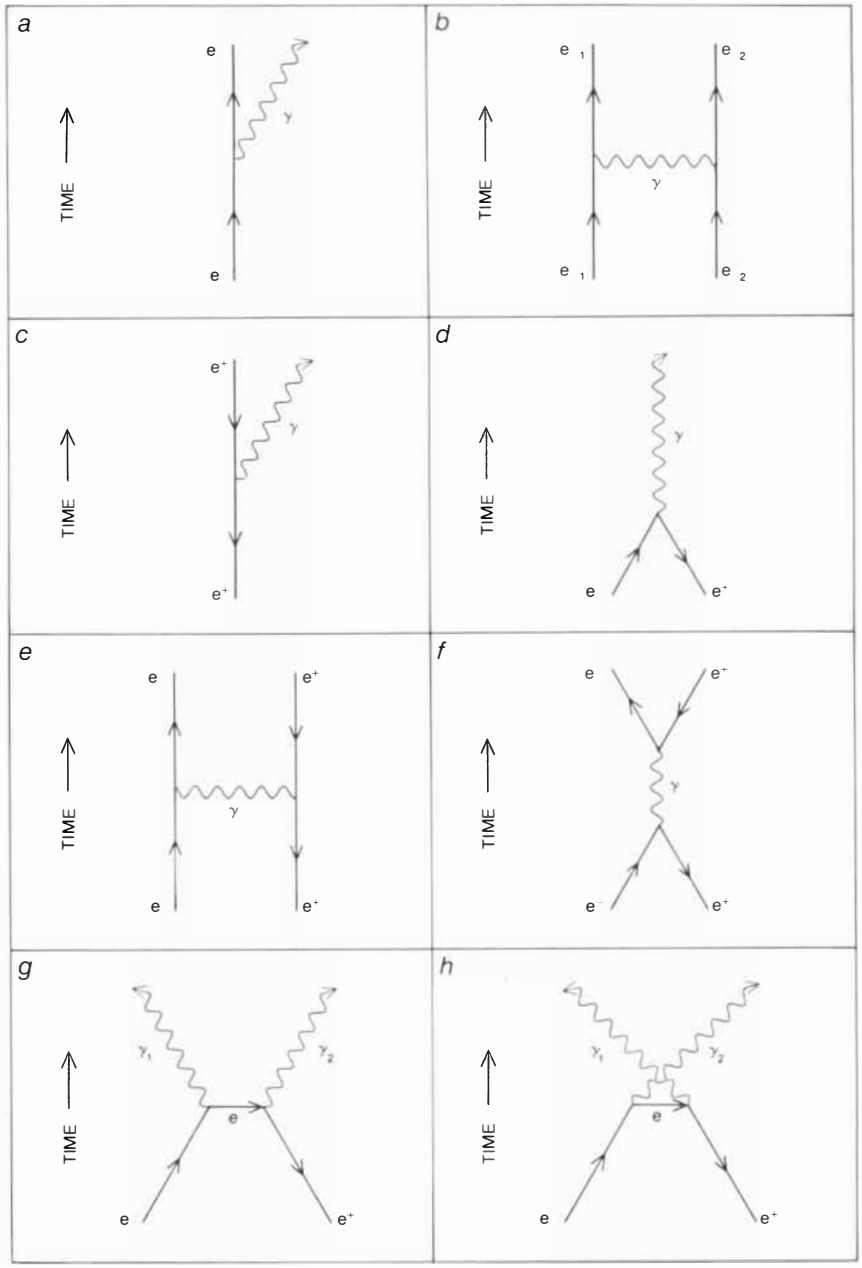
The major limitations on the luminosity obtainable in a storage ring are instabilities that arise as a result of the interaction of the two beams. In a storage ring only some orbits are stable. As the two opposing beams pass through each other at an interaction region, each beam exerts powerful electric and magnetic forces on the other beam that push the other's orbit toward an unstable region.

A method that can further increase the luminosity of a storage ring was suggested in 1965 by K. W. Robinson and G. A. Voss of the Cambridge group. They proposed an exceptionally strong-focusing magnetic-lens system to concentrate the beams in a small cross-sectional area at the interaction regions. When the beams are held to a small cross-sectional area, each particle in a beam is necessarily close to the particles in the second beam. This increases the probability of a collision and therefore the luminosity.

Unfortunately the defocusing effects of one beam on another also increase as the size of the beam is reduced. The defocusing forces are electromagnetic and hence follow Coulomb's law, which states that the force increases inversely as the square of the distance between the charges. If a decrease in the size of the beam is accompanied by, or is caused by, a proportionate increase in focusing, however, this defocusing effect is taken care of. The strong-focusing technique therefore enables one to achieve a smaller "beam envelope." This technique was first put to use in 1970 at the Cambridge Electron Accelerator. It has recently been used at SPEAR, the electron-positron storage ring at Stanford, to achieve the record luminosity of 10^{31} particles per square centimeter per second. With this luminosity, reactions with cross sections as low as 10^{-34} square centimeter can be produced at the rate of approximately one event every 17 minutes.

What are some of the experiments that can be performed with these machines? It is convenient to divide the experiments into four categories: (1) tests of the theory of quantum electrodynamics, (2) the search for new particles, (3) experiments with "equivalent photon" beams and (4) the production of strongly interacting particles.

Quantum electrodynamics is the the-



FEYNMAN DIAGRAMS, devised by Richard P. Feynman of the California Institute of Technology, show the various interactions of elementary particles, in this case electrons, positrons and photons (quanta of radiation). The emission of light by an atom, or the generation of radio waves by the electrons in an antenna wire, is based on a single reaction (*a*), where an electron (e^-) emits a photon (γ). This reaction can never be observed because it does not simultaneously conserve energy and momentum. It is therefore called a virtual reaction. It is possible, however, to combine a sequence of virtual reactions to obtain observable processes that do conserve both energy and momentum. For example, the photon emitted by one electron (e^-_1) can be absorbed by a second electron (e^-_2). The overall process (*b*) represented by this reaction is the scattering of one electron by another through electromagnetic forces; the overall reaction conserves both energy and momentum. The positron (e^+) can participate in similar interactions. It too can emit a virtual photon (*c*). In all the diagrams time proceeds from bottom to top; the positron differs from the electron only in that its directional arrow is pointed backward. An electron and a positron can annihilate each other and give rise to a photon (*d*). These virtual processes can be combined to obtain a representation of elastic scattering between a positron and an electron (*e and f*). In one case (*e*) the electron emits a virtual photon, which is absorbed by the positron. In the other case (*f*) the electron and the positron annihilate each other and give rise to a virtual photon, which is then converted into an electron-positron pair. One final interaction is the conversion of a positron-electron pair into two photons (*g and h*). In this case a virtual electron is exchanged. The reason there are two Feynman diagrams for this single process is that it is impossible to tell which gamma ray was emitted by which of the particles.

ory of electromagnetic phenomena for which Richard P. Feynman, Julian Schwinger and Sin-itiro Tomonaga received the 1965 Nobel prize in physics. It combines the classical laws of electric and magnetic forces with special relativity and quantum mechanics. It embraces phenomena as diverse as the chemical forces between molecules in the living cell, the generation of radio waves and the creation of electron-positron pairs by high-energy gamma rays. It is perhaps the most complete and best-verified physical theory both in the precision of its predictions and in their range. Clearly it is important for our basic understanding of nature to test quantum electrodynamics both with greater precision and in new physical domains.

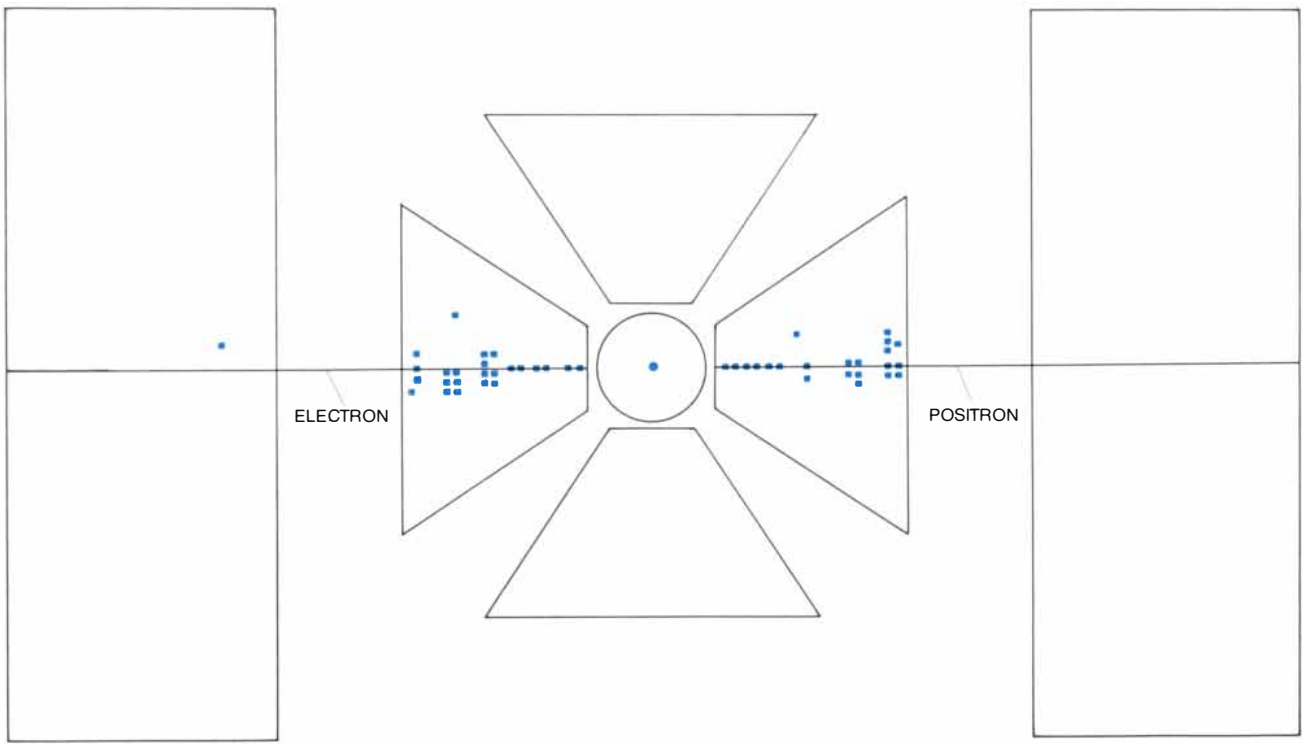
Electron-positron colliding-beam experiments provide a particularly advantageous way of testing the theory of quantum electrodynamics in the high-energy domain. For these tests three types of reaction have been studied: (1) reactions in which the electrons and the positrons appear simply to bounce off each other without losing any energy (a process known as elastic scattering), (2) reactions in which the electrons and the positrons annihilate each other to form muon-antimuon pairs and (3) reactions in which the electrons and the positrons annihilate each other to form pairs of gamma rays.

The elastic scattering of electrons by positrons depends on the internal structure of the particles as well as on the nature of the forces between them. The present assumptions are that the electron and the positron are pointlike charges (without any spatial extent) and that they interact only through long-range electric and magnetic forces. These assumptions are consistent with all the experimental results obtained so far, but it is worthwhile to check them at ever higher energies or, what is equivalent, at ever smaller distances.

In spite of the spectacular success of quantum electrodynamics, there are still some basic questions about the electron that remain unanswered. For example, the charge of the electron is often expressed in terms of a dimensionless quantity, α (known as the fine-structure constant), which is numerically equal to about 1/137. This mysterious number is one of the fundamental constants of nature, yet it cannot be derived from quantum electrodynamics or from any other theory. It is not even clear why electric charges always appear in nature as a whole-number multiple of the elementary charge of the electron. Similarly, the mass of the electron cannot be calculated in the framework of quantum electrodynamics or of any other theory; it is a property whose value must be determined by experiment. One may hope that some future theory of electrodynamics will be able to explain the origin of the electron's charge and mass, and it is with an eye to such a development that experimenters search for a structure of the electron at increasingly smaller distances.

Let us now turn to the experimental investigation of the electron-positron elastic-scattering reaction. As an electron approaches a positron (inside a storage ring, say) each particle exerts forces that cause the other particle to scatter, or deviate, from its initial direction of motion. The scattering angle of a particle is defined as the angle between its initial direction and its final direction.

In the colliding-beam experiments we are discussing, a beam of electrons and a beam of positrons of the same energy are allowed to collide. The number of electron-positron pairs per unit of time that emerge from the interaction region in a given range of scattering angles is counted. From this number and from the value of the average luminosity the scattering probability, or cross section, can be computed. The theory of quantum electrodynamics predicts the cross section for scattering as a function of the energy and the scattering angle. To test



CROSS-SECTIONAL VIEW of an event in B.O.L.D. shows a positron scattering elastically (without any loss in energy) off an electron. The two particles go off in opposite directions. This result is a consequence of the law of the conservation of momentum, which states that the total momentum of an isolated system cannot change with time. Although one particle is labeled the electron and the other the positron, there is no magnetic field in B.O.L.D. to detect charge of particles; thus one cannot determine which one is which.

the theory those predictions are compared with the experimental results.

Recently an experiment on electron-positron elastic scattering was conducted at the Cambridge Electron Accelerator by a group of physicists headed by Robert Little. (We are two members of that group.) A beam of 2,000-MeV electrons and a beam of 2,000-MeV positrons were made to collide for a total center-of-mass energy of 4,000 MeV. The interaction region was surrounded by detection equipment consisting of spark chambers, scintillation counters and metal plates arranged in four similar quadrants [*see illustration on page 106*]. In two months of data-taking 260 events were detected that could be identified as being attributable to the elastic-scattering process, with scattering angles between 50 and 130 degrees of arc. (In a typical event of this kind it is impossible to tell which particle is the electron and which is the positron, because the detector has no magnetic field.)

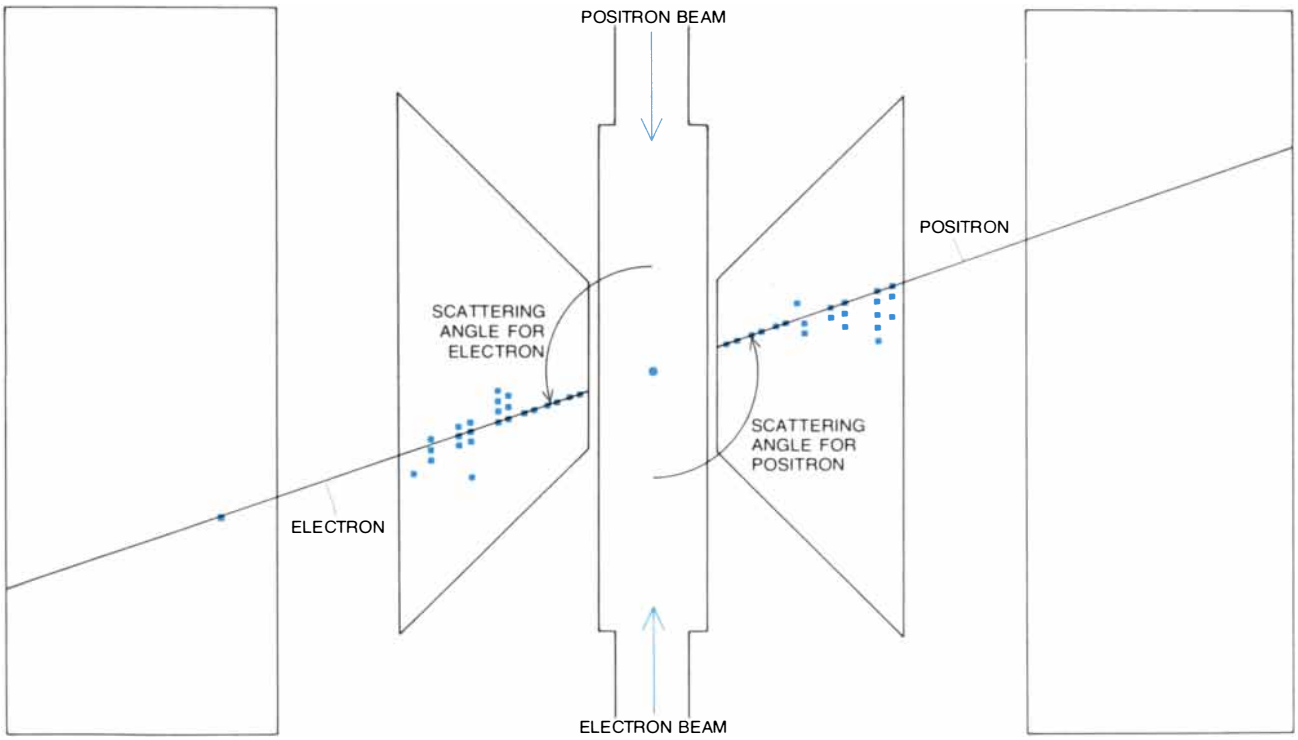
The two tracks recorded in the spark chambers for such an event appear to lie on a straight line [*see illustrations on these two pages*]. This result is a consequence of the law of the conservation of momentum, which states that the total momentum of an isolated system cannot change with time. In this particular colliding-beam reaction the total momentum before the collision is zero, since the momentum of the initial electron is exactly balanced by the opposing momentum of the initial positron. The conservation of momentum then implies that the total momentum after the collision must also be zero. That can only be the case if there is a balance between the momenta of the final particles. Consequently the electron and the positron must emerge from the scattering reaction traveling in opposite directions.

Electromagnetic showers can be initiated in the metal plates by the final particles. The showers are produced by the process we have described in connection with the production of positrons. The appearance of such showers in the spark-chamber record helps to identify the reaction unequivocally as the elastic scattering of electrons and positrons.

Before we present the results of our experiments we shall attempt to explain some of the ingredients that go into the theoretical predictions. In classical electrodynamics the force responsible for the scattering of one charged particle by another is the electric force described by Coulomb's law. In quantum electrodynamics this force is expressed in terms of the exchange of photons between the interacting particles.

The photon is the quantum of electromagnetic radiation. A beam of light or a beam of radio waves is made up of many photons; a gamma ray is a single high-energy photon. Photons are emitted whenever a charged particle is accelerated or decelerated. All interactions between electrons and photons—for example the emission (or absorption) of light by an atom or the generation (or reception) of radio waves by the electrons in an antenna wire—are based on a single reaction of the type termed a "virtual" reaction. The physicist's conventional representation of such a reaction is called a Feynman diagram [*see illustration on page 107*].

A virtual reaction can never actually be observed because it cannot simultaneously conserve energy and momentum. It is possible, however, to combine a sequence of virtual reactions to obtain real, observable processes (which must of course conserve both energy and momentum). Suppose, for example, the first electron emits a photon. The photon has zero mass and such a reaction can conserve momentum, but then it cannot conserve energy; according to the uncertainty principle this dual task is allowable only if the photon exists for a very short time. Alternatively (and equivalently) one can assume that both energy and momentum are conserved and assign a fictitious mass to the virtual pho-



PLAN VIEW of a positron scattering off an electron corresponds to the illustration on the opposite page but is perpendicular to the beam direction instead of across it. The upper and lower quadrants of B.O.L.D. are not shown. This view shows that not only is the positron scattered in a direction opposite to that of the electron but also the particles go off at an angle to their original direction of travel. This angle, known as the scattering angle, varied from 50 to 130 degrees of arc in the Cambridge experiments. Here it is 110 degrees.

ton. That photon is absorbed by the second electron in another virtual reaction. The overall process can be interpreted as the interaction of one electron with another electron through electromagnetic forces; here the reaction conserves energy as well as momentum.

The photon that is exchanged between the first electron and the second electron does not appear among the initial or final particles; it can never be directly observed. (That is why it is called virtual.) The exchanged virtual photon can, however, carry energy and momentum. The scattering of the two electrons is described by saying that these particles exchange a virtual photon that transfers momentum from one particle to the other.

The positron fits naturally into this scheme. Feynman showed that the virtual emission of a photon by a positron is completely analogous to the process we have described for the electron. His analysis also led to the prediction of a process altogether outside the domain of pre-quantum-mechanical physics: the virtual annihilation of an electron-positron pair to form a photon.

These virtual processes can be combined to obtain the quantum-electrodynamic representation of the elastic scattering of an electron by a positron in terms of two alternative Feynman diagrams. In one diagram the electron emits a virtual photon that is absorbed by the positron. In the other diagram the electron and positron annihilate themselves to form a virtual photon, which is converted into an electron-positron pair. By assigning numerical factors to each leg, vertex and virtual photon in the diagrams one can calculate the theoretical prediction for the reaction cross section.

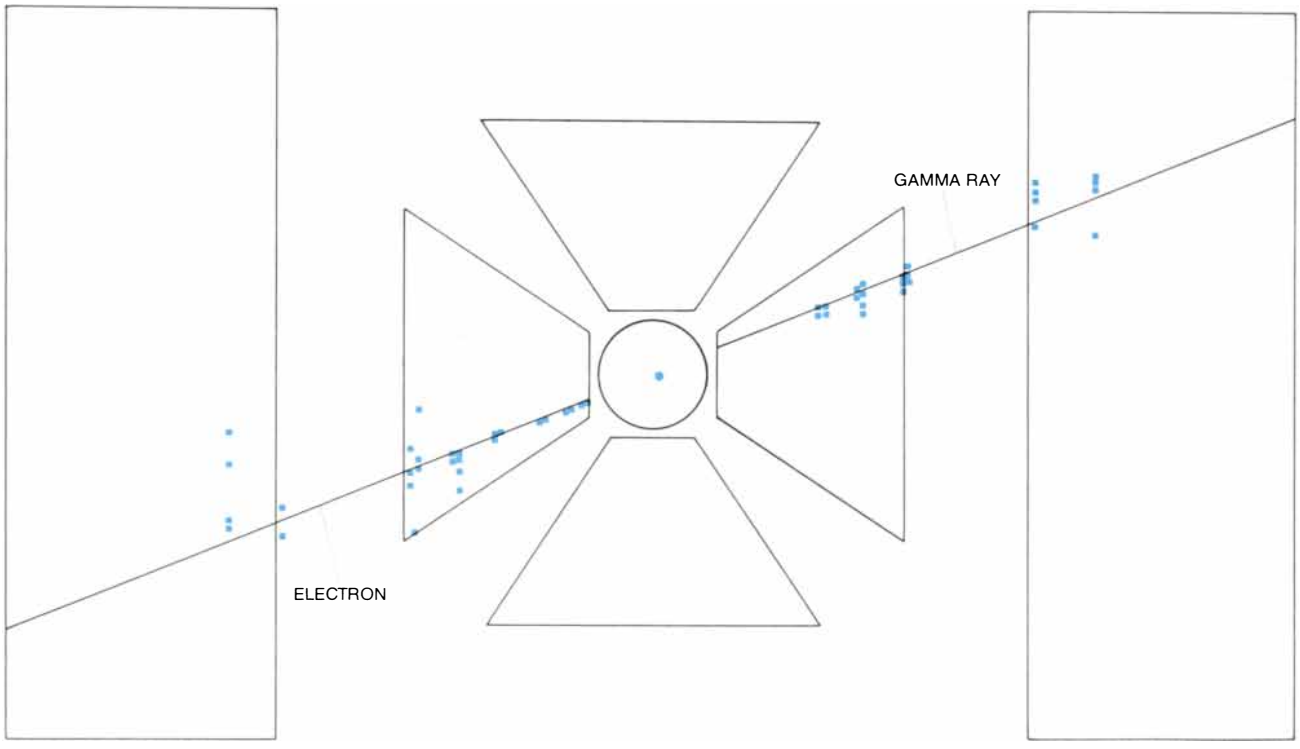
Let us now look at the experimental results obtained by our group at Cambridge. The total number of elastic-scattering events detected and the distribution of scattering angles of the final particles agree with the predictions of quantum electrodynamics. This agreement had also been found in other experiments at lower energies.

Another interesting particle produced in electron-positron colliding-beam experiments is the muon. This particle has the same charge and spin as the electron, but it has more than 200 times the mass of the electron. A free muon decays with a lifetime of two microseconds into an electron, a neutrino and an antineutrino. Muons, like electrons, do not interact with nuclei by the strong nuclear interaction. Also like the electron, the muon appears to be a pointlike charge. If the energy is high enough, muons can be produced in pairs by the annihilation reaction of an electron and a positron. This reaction also proceeds by the exchange of a virtual photon and can be represented by a Feynman diagram.

Muon-antimuon pairs can be easily identified in experiments. In passing through matter muons radiate less than electrons; they do not interact strongly with nuclei and hence can travel considerable distances. Since one muon is easily distinguished, a pair is even more easily distinguished. Muon-antimuon pairs have been produced at all the storage-ring laboratories, and the results agree with quantum electrodynamics. Particles such as muons, electrons and neutrinos, all of which have spin 1/2 and none of which respond to the strong nuclear force, are known collectively as leptons.

Another check of quantum electrodynamics is obtained by examining the reaction in which an electron and a positron annihilate each other to produce a pair of gamma rays. In this process a virtual electron is exchanged. Gamma rays are also easily identified. They produce electromagnetic showers in lead, but there is no charged-particle track leaving the interaction region. Again the experimental results agree with the theoretical predictions.

Electron-positron colliding beams provide a powerful tool in the continuing search for new particles capable of



EVENT CALLED VIRTUAL COMPTON SCATTERING has an electron and a gamma ray in the final state (*left*). A beam of positrons in the process of elastically colliding with a beam of electrons can be treated as a beam of "equivalent photons." In virtual Compton scattering the positron that produced the equivalent photon is degraded in energy and goes predominantly in the forward direc-

interacting with electrons or photons. A large number of such particles can be postulated. Proof of their existence would be helpful in formulating theoretical models of the leptons and of elementary-particle interactions. Many could in principle be produced in electron-positron collisions of sufficiently high energy, and they could be clearly identified if they in fact exist. An excited electron, an excited muon or a heavy lepton are hypothetical examples.

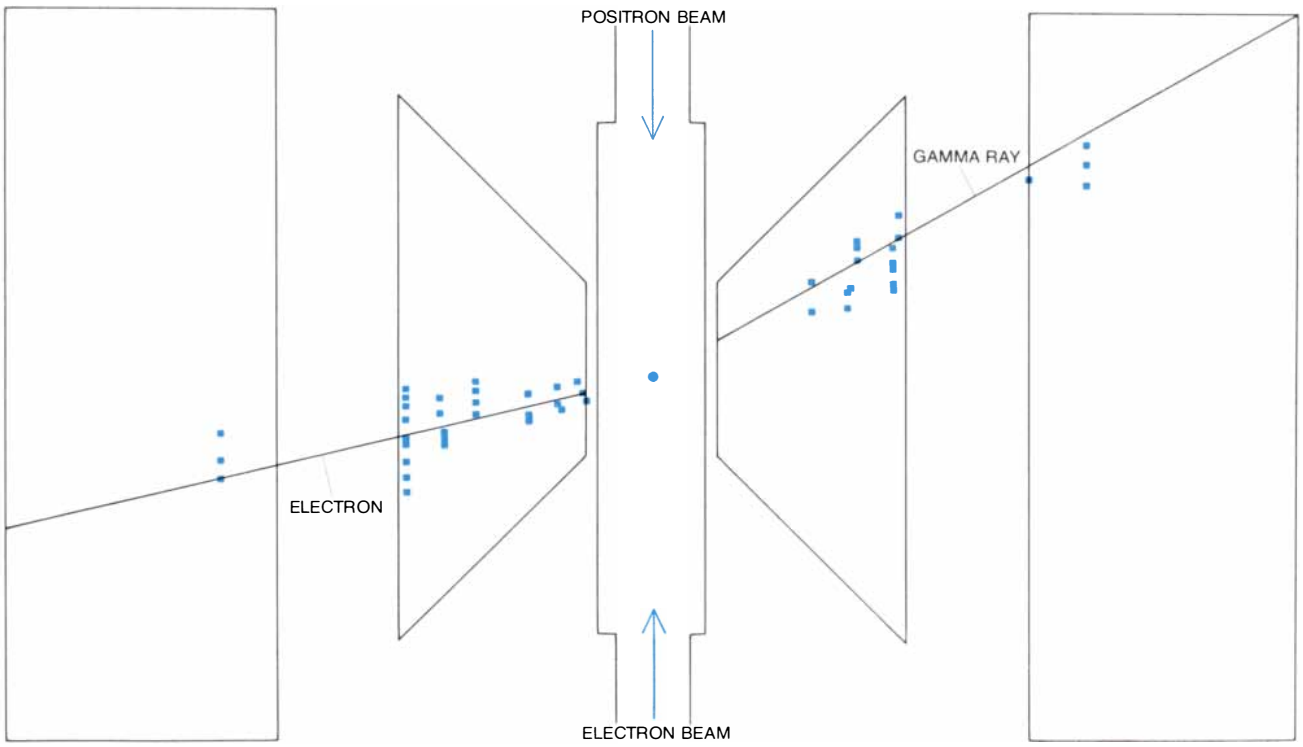
Atoms, nuclei and nucleons (protons and neutrons) can be excited from their ground, or lowest-energy, state to higher energy states by the absorption of a photon. It is not known whether electrons and muons also have excited states. If such states exist, they could be produced by electron-positron collisions and could be detected by their characteristic decays. If such excited states are discovered, it could indicate that the electron and muon are not "elementary" entities but have a composite structure.

The electron and the muon (along with their antiparticles) are the only known leptons with electric charge. Why there should be more than one and yet only two charged leptons is a long-standing puzzle in particle physics. One possibility is that there are more than two charged leptons but that only the electron and the muon have so far been discovered. Indeed, there may well exist a spectrum of leptons of increasing mass. The heavy leptons, if they exist, could be produced, like muons, in pairs in electron-positron annihilation reactions. They would decay into an electron (or a muon), a neutrino and an antineutrino, or into a neutrino and one strongly interacting particle or more. The decay modes would identify the particle. The discovery of a new lepton could shed light on the origin of the mass difference between the electron and the muon and more generally on the relation between the electron and the muon. At present there is no evidence for the existence of either kind of hypothetical particle.

Electron-electron or electron-positron colliding beams can also be used to study photon-photon collisions. This point has been emphasized in the past few years by theorists in the U.S.S.R., France and the U.S. The theory of these processes is based on the approximate equivalence of the electric and magnetic fields produced by a rapidly moving charged particle and the fields produced by a pulse of electromagnetic radiation. Any such pulse can be decomposed (by Fourier analysis) into a spectrum of photons. Then the interaction of a moving charged particle with matter can be approximately calculated in terms of the interaction with matter of the equivalent spectrum of photons. This "equivalent photon approximation" was formulated independently by C. F. von Weizsäcker and E. J. Williams in 1934. According to that approach a beam of positrons in the process of colliding with an electron beam can be treated as a beam of equivalent photons. One of the equivalent photons may scatter elastically off an electron, giving rise to an electron and a gamma ray in the final state. In addition the positron that generated the equivalent photon is degraded in energy and goes predominantly in the forward direction.

Events of this type have been observed in the spark-chamber records [*see illustration on these two pages*]. In the perpendicular view the electron and the gamma ray do not move in opposite directions along the same line, owing to the momentum carried off by the forward-going positron. Sixteen of these events have been recorded at Orsay and six at Cambridge at center-of-mass energies of about 700 and 3,000 MeV respectively; the results are in agreement with the predictions of quantum electrodynamics.

Both charged-particle beams can be treated as equivalent photon beams. This suggests the possibility of observing reactions in which equivalent photons col-



tion. In the view perpendicular to the colliding beams (*right*) the electron does not move in a direction opposite to that of the gamma ray, owing to the momentum carried off by the forward-going positron. Reaction could have taken place equally well if the equivalent photon had scattered off a positron instead of an electron. In B.O.L.D. there was no way to distinguish between the possibilities.

lide to produce electron-positron pairs, muon-antimuon pairs or particle-antiparticle pairs consisting of strongly interacting particles. The cross sections for the first two reactions can be calculated by the technique of quantum electrodynamics. Events of these types have been observed at Novosibirsk, Frascati and Cambridge, and the results agree with the theoretical calculations. The third reaction promises to be an important area for future experiments.

The most exciting field of research with electron-positron colliding beams at present is the production of the strongly interacting nuclear particles known as hadrons. These particles, because of their strong interactions, have a structure and a finite size, and their electromagnetic interactions cannot be described by a simple application of quantum electrodynamics.

Hadrons can be produced in pairs, as in the annihilation of an electron and a positron to yield a proton and an antiproton. The probability that a hadron pair will be produced in an electron-positron collision is expressed in terms of a set of functions called form factors, which are functions of the momentum transferred to the virtual photon. The form factors that describe the production of a hadron pair should, when they are smoothly extrapolated, also describe the elastic scattering of electrons from the hadron. The form factor is thus a measure of the distribution in space of the charge, or the magnetic moment, of the hadron.

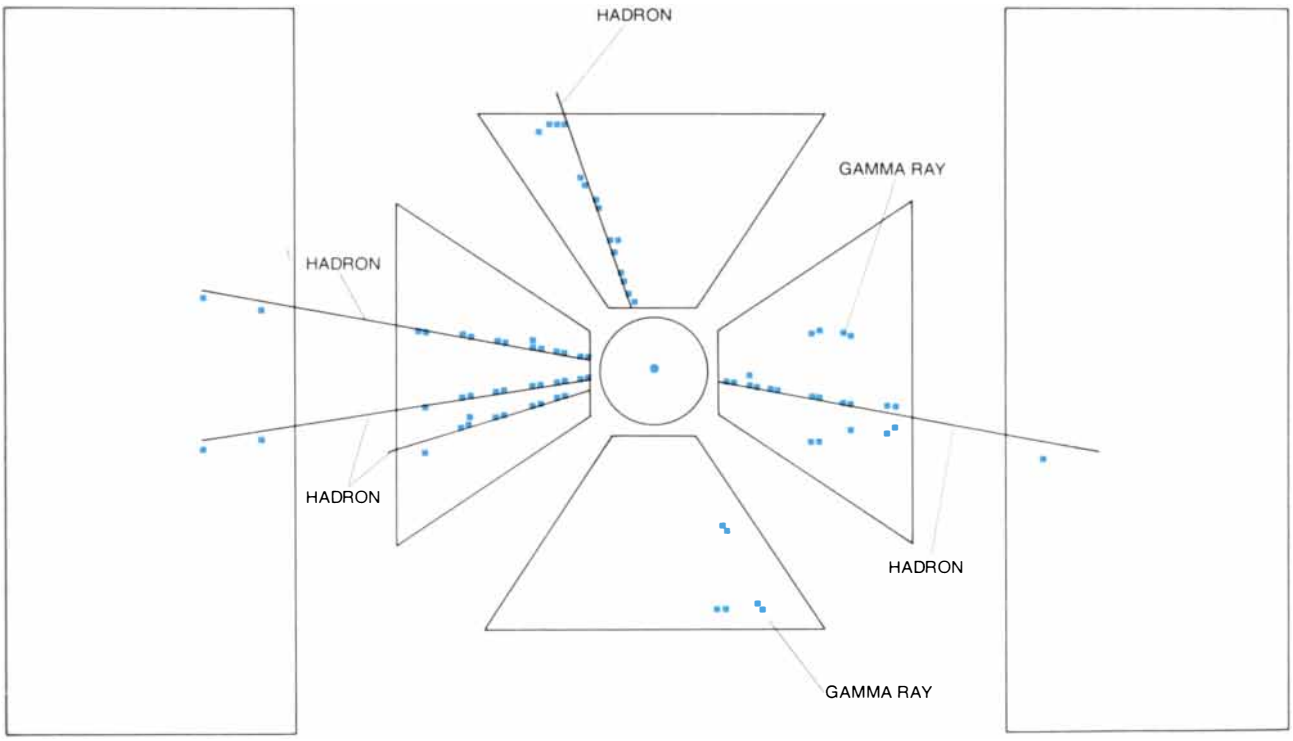
The hadron that has been the most intensively studied in electron-positron colliding-beam experiments is the pion. The pion is the lightest strongly interacting particle; it has a mass only 280 times that of the electron. If the pion were a point-like particle, the square of its form factor would equal 1. Deviations from 1 are caused by its strong interactions [*see lower illustration on opposite page*]. The most striking feature of the pion's form-factor graph is the bump at a center-of-mass energy of 760 MeV with a width of about 120 MeV. This is interpreted as a resonant state formed by the pion-antipion pair. The resonance can be considered to be a short-lived particle called the neutral rho meson, which decays in 10^{-23} second into a pion-antipion pair. Thus at center-of-mass energies equivalent to approximately the mass of the neutral rho meson (760 MeV) pion-pair production is dominated by the reaction in which an electron and a positron annihilate each other to produce a neutral rho meson, which in turn decays to produce the pion-antipion pair.

In addition to the neutral rho meson two other well-established resonances, the neutral omega meson and the neutral phi meson, have been produced in electron-positron collisions. It turns out that the colliding-beam experiments are the cleanest way of studying these three resonances.

Much interest has been generated by the observation, first at Frascati and then at Cambridge, that hadrons are produced more often than anticipated in electron-positron collisions at high energies. In these experiments events have been observed in which many charged particles are produced [*see illustration below*]. A great deal of effort is now going into the investigation of the detailed nature of these multihadron events.

Below a center-of-mass energy of about 1,100 MeV the production of hadrons is dominated by the rho, omega and phi resonances. The extrapolation to higher energy, before the recent experiments, was uncertain. The prevailing folklore led to the expectation that the ratio of hadron production to muon-antimuon pair production would be about 1 : 1. The measurements at Frascati and Cambridge show that the ratio is about 1 : 1 at a center-of-mass energy of 1,350 MeV but then rises to more than 4 : 1 at 4,000 MeV. This is a great surprise. The physics that underlies the large production cross section of hadrons is not yet understood.

One possibility is that additional reso-



HADRONS (strongly interacting nuclear particles) are produced in collisions between electrons and positrons at high energies. One event observed in B.O.L.D. had five charged particles in the final state. Behavior of the particles suggested that they were hadrons.

nances are responsible for the large cross section for hadron production. There is new evidence from Frascati for a new resonance, called the rho prime, at a mass of about 1,600 MeV. That resonance decays into four pions and has been seen also in photon-production experiments. The "quark" model of fundamental particles put forward a few years ago by Murray Gell-Mann leads one to expect other resonances as "companions" to the rho prime and part of the same "nonet." These resonances should also contribute to hadron production. Possibly other, more massive resonances exist that overlap and blend into a continuum, producing the high cross section.

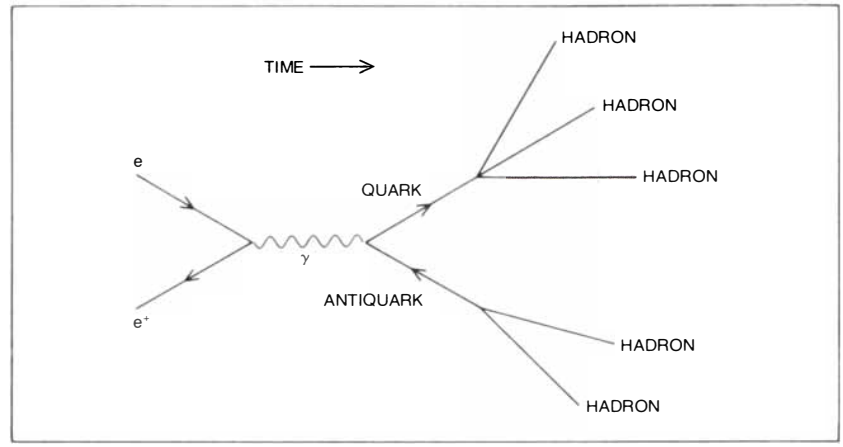
Another interpretation that provides a prediction against which we can compare our results is Feynman's "parton" model. Experiments on inelastic electron-proton scattering at Stanford are consistent with scattering from three pointlike constituents of the proton named partons. These constituents, to satisfy the data, must have spin 1/2; the data are also roughly consistent with the identification of the partons as quarks with fractional charges of $-1/3$, $-1/3$ and $2/3$. The success of the parton model in describing these experiments makes it of great interest to see if this model also applies to the annihilation reactions.

In the parton model one imagines that the hadron-production process proceeds as follows. An electron and a positron annihilate each other and produce a parton and an antiparton, which then are transformed into hadrons. The Feynman diagram for this process is similar to the diagram for electron-positron annihilation into muons.

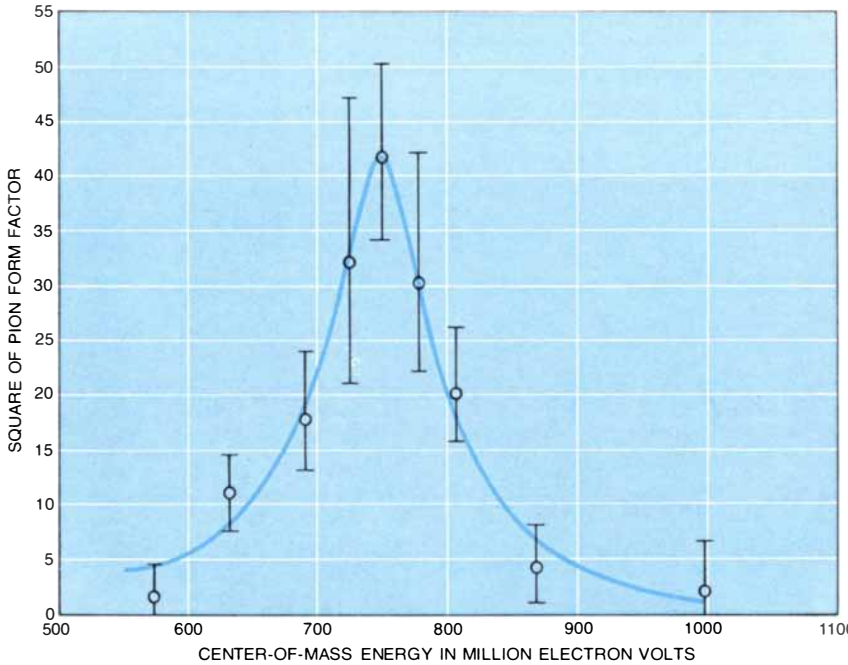
The cross section for hadron production can then be calculated just as the muon-antimuon production cross section is (at least if we neglect the effects of final-state interactions). The muon and the parton are both pointlike and have spin 1/2; they may differ, however, in electric charge. Therefore we reach a quantitative prediction: the ratio of hadron production to muon-pair production, at high energies, is equal to the sum of the squares of the parton charges. In the usual quark model this is 2 : 3. Sometimes a larger set of quarks, called colored quarks, are assumed. Then the ratio is 2 : 1. The experimental results appear to be in disagreement with both models.

Another prediction of the parton model is that the hadrons should come out of the reaction in back-to-back cones known as jets. There is no evidence as yet for any jetlike structures in the hadron events, but at the present energies such structures might not be very obvious. If jets exist, they should be easier to perceive at higher energies.

We see, therefore, that the existing experiments have only tantalized us and whetted our appetite for more and better experiments. Do heavy leptons exist? Are there more resonances? Will jet structures be found? Is the parton-model description valid at any energy? What is the reason behind the large cross section for hadron production? The answers to these questions are important for our understanding of the elementary particles. Physicists eagerly await the results of the next round of electron-positron colliding-beam experiments.



FEYNMAN DIAGRAM OF QUARKS, the theoretical entities that might be the ultimate subatomic particles, shows that such entities could account for the formation of hadrons in collisions between beams of positrons and electrons. A positron and an electron would annihilate to form a quark and an antiquark, which would then be transformed into hadrons.



PION IS NOT A POINTLIKE PARTICLE, as is shown in a diagram that plots the square of the pion's form factor against the energy of the center of mass of the system formed by the pion and its antiparticle. The form factor is a measure of the distribution in space of the charge of a hadron. If the pion were a pointlike particle, the square of its form factor would be 1. Deviations from 1 are caused by its strong interactions. Large hump at the center-of-mass energy of 760 MeV is interpreted as a state of resonance formed between the pion and its antiparticle. Data shown are from experiments done at Novosibirsk in the U.S.S.R.

MATHEMATICAL GAMES

"Look-see" diagrams that offer visual proof of complex algebraic formulas

by Martin Gardner

There is no more effective aid in gaining an understanding of certain algebraic identities than a good diagram. One should, of course, know how to manipulate algebraic symbols to obtain proofs, but in many cases a dull proof can be supplemented by a geometric analogue so simple and beautiful that the truth of a theorem is almost seen at a single glance.

Consider, for example, a basic summation identity: The sum of the first n positive integers is half of $n(n+1)$. In equation form:

$$1 + 2 + 3 + 4 \ldots + n = \frac{n(n+1)}{2}.$$

The first n consecutive positive integers can be depicted by dots in triangular formation [*see illustration below*]. Two such triangles fit together to form a rectangular array containing $n(n+1)$ dots. Because each triangle is half of the rectangle, we see at once that the formula for the number of dots in each triangle is half of $n(n+1)$.

This simple proof goes back to the ancient Greeks. They called any number of the form $\frac{1}{2}n(n+1)$ a triangular number, and any number of the form n^2 a square number because it could be represented by a square array of dots. The illustration at left at the top of the opposite page shows how square arrays prove that the sum of the first n *odd* integers is n^2. Think of the pattern as extending any desired distance to the right and down. Each reversed L-shaped strip contains the odd number of circles indicated at the top. It is obvious that each additional strip, that is, each new odd number in the series $1 + 3 + 5 \ldots$, enlarges the square by one unit on a side, and that the total number of dots in each square bounded by the nth odd number is n^2.

The Greeks also used square arrays to establish the identity $1 + 2 + 3 \ldots + n \ldots + 3 + 2 + 1 = n^2$. The case for $n = 5$ is shown in the illustration at right at the top of the opposite page. Is any explanation necessary?

Finding a formula for the sum of the squares of the first n integers takes a bit more doing. Consider the squares of the first five integers. As we have seen, any square can be represented as the sum of consecutive odd integers starting with 1 [*see bottom illustration on opposite page*]. In these arrays a row of nine dots occurs once, a row of seven dots twice, five-dot rows three times, three-dot rows four times and one-dot rows five times. The 15 rows can be stacked, beginning with the longest on the bottom, to form a skyscraper. By placing square arrays for 1^2, 2^2, 3^2, 4^2 and 5^2 on each side of the skyscraper, we can make a rectangle with a height equal to the sum of the first n integers. As we have seen, this sum is $\frac{1}{2}n(n+1)$. The width of the rectangle is $2n + 1$. The total number of dots in the rectangle is the product of height and width:

$$\frac{n(n+1)(2n+1)}{2}.$$

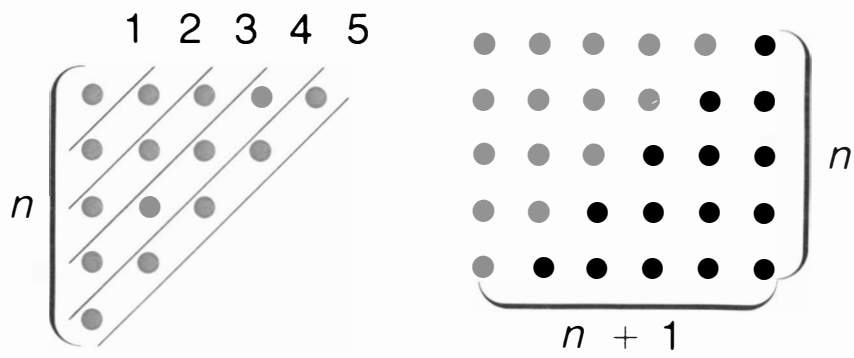
The skyscraper, which represents the sum of the squares of the first n numbers, is one-third of the rectangle. Dividing the above formula by 3, therefore, gives the formula for the skyscraper, which is the formula we seek:

$$\frac{n(n+1)(2n+1)}{6}.$$

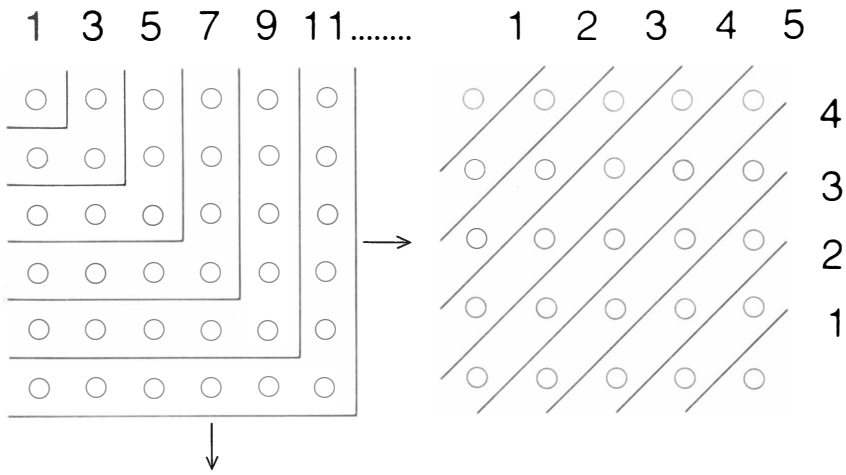
The formula should be familiar to all students of recreational mathematics. It gives the number of different squares, of all sizes, that can be found on a chessboard with n cells on a side. The standard 8-by-8 board, for example, contains $8(8+1)(16+1)/6 = 204$ different squares. It is not hard to see that the formula applies. An 8-by-8 square appears only once on the board. If a 7-by-7 square is placed on the board, it can be shifted to $2^2 = 4$ positions. A 6-by-6 square can be shifted to $3^2 = 9$ positions (eight on the border and one in the center), a 5-by-5 to $4^2 = 16$ positions, and so on.

The sum of the cubes of the first n integers is involved in a remarkable identity that astounds most students when they first encounter it. The sum of the first n cubes equals the square of the sum of the first n integers. In algebraic form: $1^3 + 2^3 + 3^3 \ldots + n^3 = (1 + 2 + 3 \ldots + n)^2$. An old diagram for it is shown in the top illustration on page 116. The square array of numbers, which extends down and right to infinity, is simply the multiplication table. Each number is the product of the number at the left of its row and the number at the top of its column. The table is divided into bent strips and the sum of the numbers in each nth strip is n^3. With a square of five bent strips the sum of all the numbers is $1^3 + 2^3 + 3^3 + 4^3 + 5^3$. Since this square is the multiplication table through 5, it is equally clear that the sum of all the numbers is $(1 + 2 + 3 + 4 + 5)(1 + 2 + 3 + 4 + 5)$, or $(1 + 2 + 3 + 4 + 5)^2$.

Unfortunately this geometric analogue is not as good a "look-see" proof as the preceding examples. It is not obvious, for instance, that the numbers in each nth bent strip have a sum of n^3. A more ele-



$1 + 2 + 3 \ldots + n = n(n+1)/2$

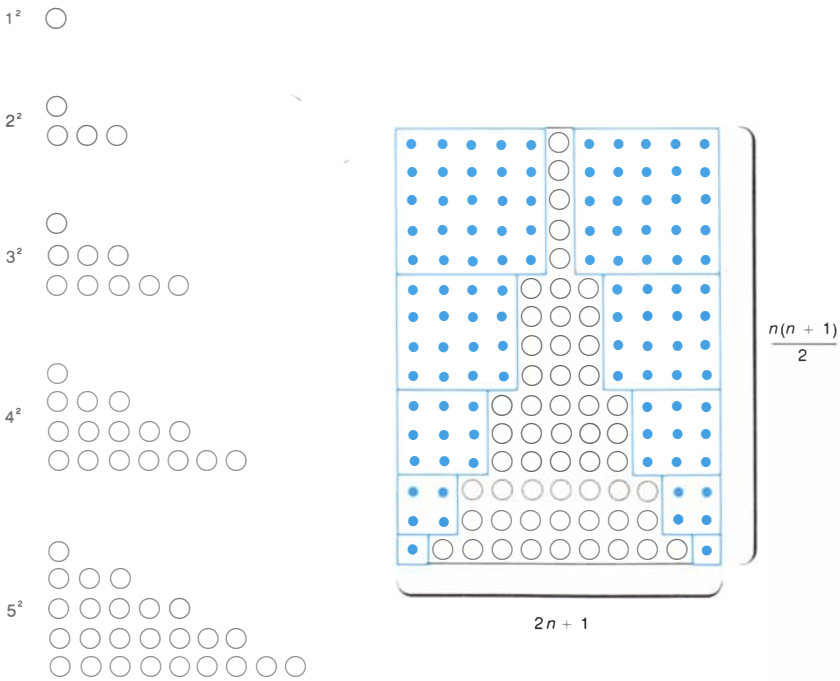


Sum of first n *odd integers is* n^2

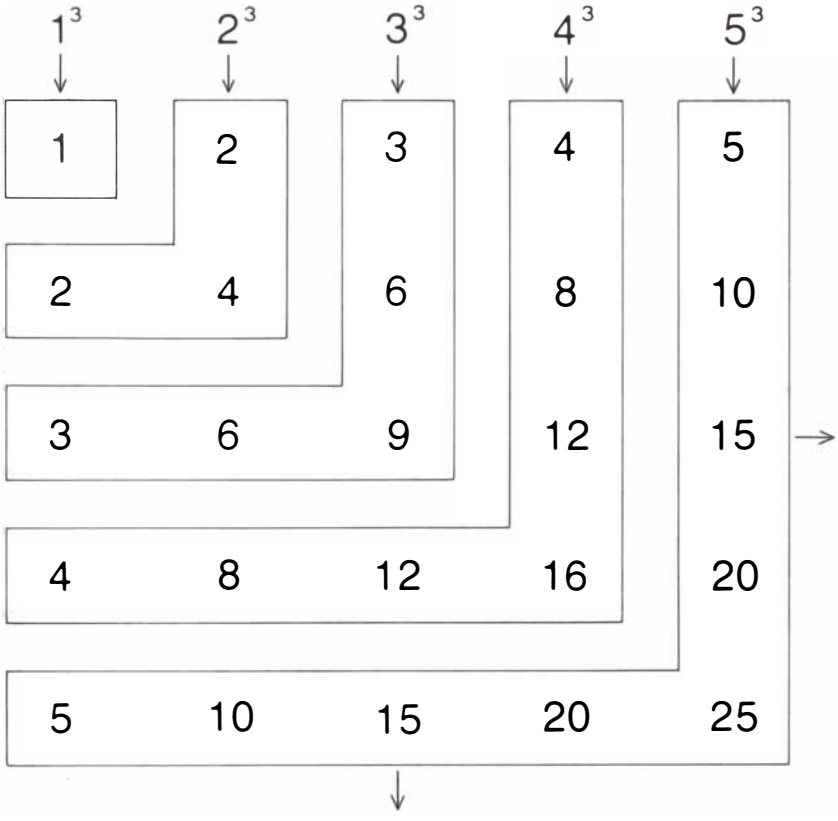
$1 + 2 + 3 + 4 + 5 + 4 + 3 + 2 + 1 = 5^2$

gant geometric analogue of the same identity was devised by Solomon W. Golomb and published in *The Mathematical Gazette* for May, 1965, pages 198–200. The isomorphism [*see bottom illustration on next page*] is easily explained. The large square has a side that equals the sum of the first eight integers, so that its area is $(1 + 2 + 3 + 4 + 5 + 6 + 7 + 8)^2$. This gives one side of the identity. For the other side note that the large square is made up of one square of side 1, two squares of side 2, three of side 3, four of side 4, and so on up to eight squares of side 8. For squares of even sides there is a square overlap, shown in solid color, but each overlap is adjacent to an empty square region, shown white, which is the same size. We can therefore take one of each pair of black overlapping squares and use it for plugging the hole next to it, and in this way eliminate all overlaps and holes. Now, $1 \times 1^2 = 1^3$, $2 \times 2^2 = 2^3$, $3 \times 3^2 = 3^3$ and so on. In other words, the total area is $1^3 + 2^3 + 3^3 + 4^3 + 5^3 + 6^3 + 7^3 + 8^3$, which is the other side of the identity.

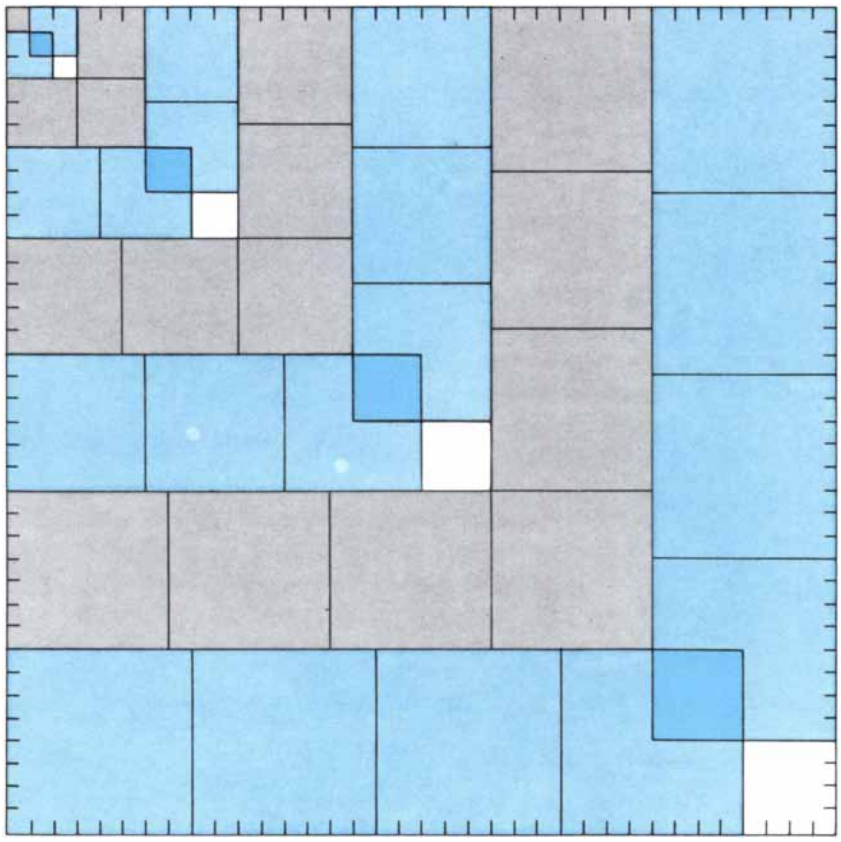
In the same article Golomb provides another proof, based on a suggestion by Warren Lushbaugh, for the same summation identity [*see top illustration on page 117*]. The squares shown have sides of 1, 2, 3, 4 and 5. There are no holes or



$1^2 + 2^2 + 3^2 \ldots + n^2 = n(n + 1)(2n + 1)/6$



$1^3 + 2^3 + 3^3 + 4^3 + 5^3 = (1 + 2 + 3 + 4 + 5)^2$



$1^3 + 2^3 + 3^3 \ldots + n^3 = (1 + 2 + 3 \ldots + n)^2$

overlaps. Each square of side n appears $4n$ times. We can write the identity

$$4(1 \times 1^2 + 2 \times 2^2 + 3 \times 3^2 \ldots + n \times n^2) = [2(1 + 2 + 3 \ldots + n)]^2,$$

which simplifies to the same identity as before. The sum of the first n integers is $\frac{1}{2}n(n + 1)$, and since the square of this equals the sum of the first n cubes, we can represent the sum of the first n cubes by the compact formula

$$\left[\frac{n(n + 1)}{2}\right]^2.$$

This too is a formula that puzzlists should know. It not only gives the number of different cubes, of all sizes, contained in a cubical chessboard of side n but also gives the number of different rectangles of all sizes (including squares) in a flat chessboard of side n. Thus the standard chessboard of side 8 contains 1,296 rectangles, and a three-dimensional board of side 8 contains 1,296 cubes. We can "see" how this counts the cubes by the same mechanical argument we applied to the squares of a chessboard. There is one largest cube of side 8. An order-7 cube goes in $2^3 = 8$ corners, an order-6 cube in $3^3 = 27$ spots, and so on.

The flat board provides still another geometric way of proving that the sum of the first n cubes equals the square of the sum of the first n integers. Robert G. Stein, in *Mathematics Magazine* for May, 1971, pages 161–162, explains how two counting arguments for the number of rectangles in a square chessboard give the two sides of the identity. See also Gene Murrow's more detailed solution of the rectangle-counting problem in his article "A Geometric Application of the 'Shepherd's Principle'" in *The Mathematics Teacher* for December, 1971, pages 756–758. (The Shepherd's Principle is: To count sheep, count the number of their legs and divide by 4.)

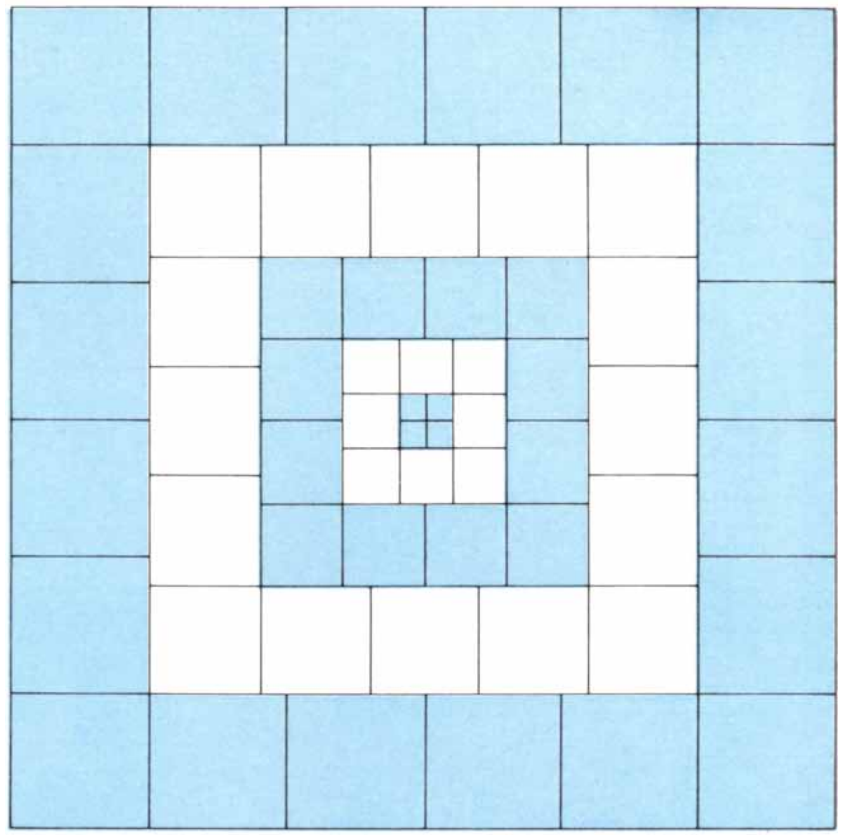
We turn now to another class of geometric analogues: dissections that illustrate simple identities involving squares that are the sums of other squares and cubes that are the sum of other cubes. Take, for instance, the familiar Pythagorean triplet: $3^2 + 4^2 = 5^2$. It is the only such triplet made up of consecutive integers. How can we cut an order-5 square into the fewest number of polyominoes that can be rearranged to make two squares of sides 3 and 4? Two solutions in four pieces, one with the 3-square intact, the other with the 4-square intact, are shown in the middle illustration on the opposite page. The

dissection cannot be done with fewer pieces. No polyomino can be longer than four units, therefore the 5-square must be divided by a cut joining left to right sides, and by another cut joining top to bottom—a procedure that must produce at least four pieces. Because there are many four-piece solutions, recreational geometers amuse themselves by adding other provisos. In the two solutions shown, the total cutting length (16 units) is minimal, and in the top solution all the polyominoes are rectangles.

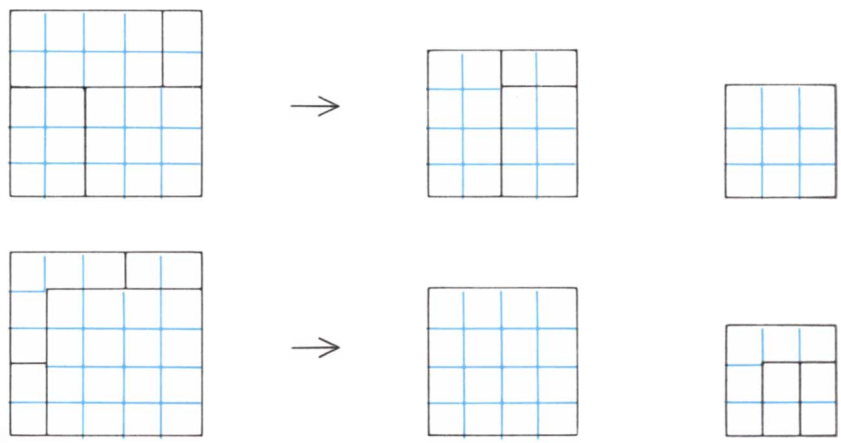
Henry Ernest Dudeney's puzzle books contain many dissection problems that illustrate other square identities. For example, the solution of Problem 357 in *536 Puzzles and Curious Problems* is an analogue of $2^2 + 3^2 + 6^2 = 7^2$. The pattern has six pieces and a cutting length of 27 [*see bottom illustration at right*]. Can the reader find a better solution with only five polyominoes?

Two cubes cannot have a cubical sum, but $w^3 + x^3 + y^3 = z^3$ has an infinity of integral solutions. The only solution in consecutive integers (indeed, the only solution with the first three terms consecutive) is $3^3 + 4^3 + 5^3 = 6^3$. This suggested to the British mathematician John Leech the following pretty problem: How can the order-6 cube be cut along integral lattice planes into a minimum number of polycubes (pieces formed by joining unit cubes) that will make separate cubes of sides 3, 4 and 5? E. H. Wheeler was the first to solve it. His eight-piece solution was published in *Eureka* (an annual publication of the Archimedean Mathematical Society of the University of Cambridge), Volume 14, 1951, page 23. In Wheeler's dissection, shown by the six cross sections in the left column in the top illustration on the next page, the 3-cube remains intact. (Cross sections of the intact cubes are shown in color for all three solutions.) A simpler eight-piece solution was later found by J. H. Thewlis of Argyll, Scotland, which Thomas H. O'Beirne of Glasgow further simplified as shown in the middle column in the illustration. The 4-cube remains intact, and only two polycubes are not rectangular blocks. A remarkable eight-piece dissection, with the 5-cube intact, found in 1970 by Emmet J. Duffy of Oak Park, Ill., is shown in the column at the right.

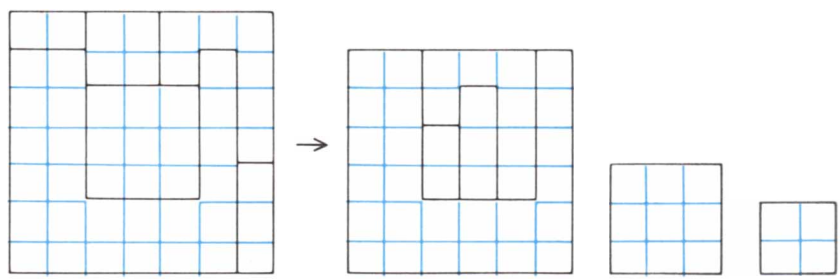
A few years ago O'Beirne asked himself: What is the minimum number of pieces for a solution of this problem in which all the polycubes are "blocks" (rectangular parallelepipeds)? It proved to be a difficult question. There must, of course, be at least eight polycubes, re-



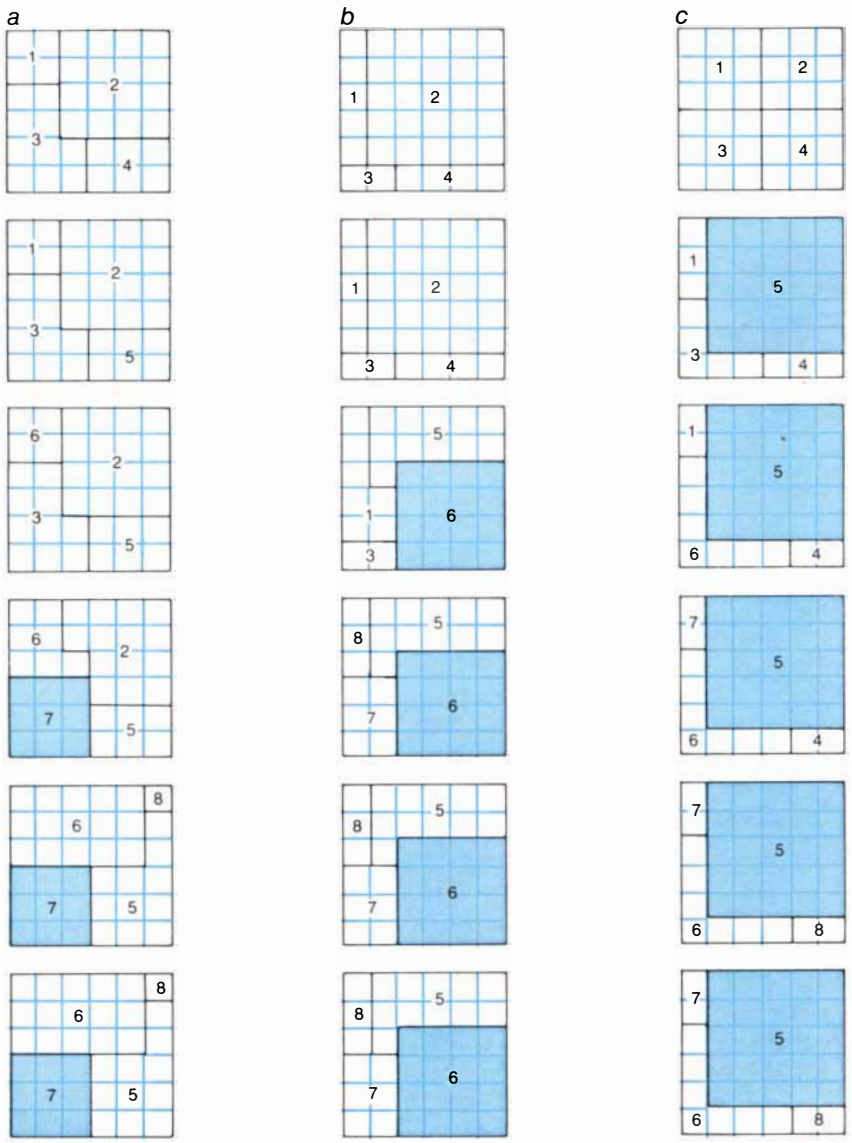
$1^3 + 2^3 + 3^3 \ldots + n^3 = (1 + 2 + 3 \ldots + n)^2$



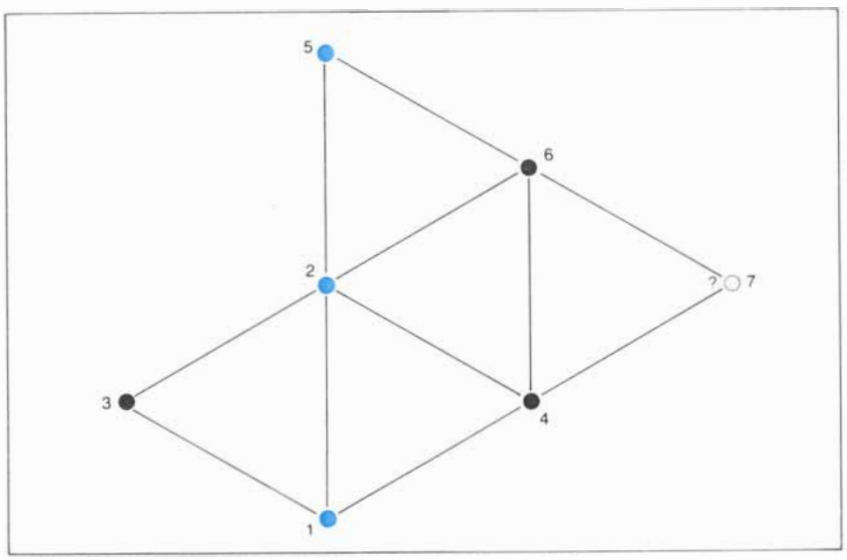
$5^2 = 3^2 + 4^2$



$7^2 = 2^2 + 3^2 + 6^2$



Order-6 cube's dissection by Wheeler (a), *O'Beirne* (b) *and Duffy* (c)



Solution to last month's problem

gardless of their shapes. No polycube can be six units long in any direction. This requires that the 6-cube be divided by at least three intersecting slices, one cutting all left-to-right rows, one cutting all front-to-back rows and one cutting all top-to-bottom columns. The procedure produces at least eight polycubes. By more complicated reasoning O'Beirne was able to show that an eight-piece dissection is not possible with all rectangular blocks.

Is it possible with nine blocks? Yes. In 1971 O'Beirne obtained a nine-block dissection that he believes is unique, although he has not been able to prove it. It is difficult to find the nine-block dissection without using a computer program. Next month I shall give O'Beirne's solution, and would welcome hearing from anyone who finds a nine-block solution with a different set of pieces, or a proof that there is no other solution.

In working on this polycube problem and others, it is helpful to have a few hundred plastic interlocking cubes of different colors. The Math Shop, 5 Bridge Street, Watertown, Mass. 02172, will supply postpaid (on prepaid orders) 100 plastic cubes (10 each of 10 colors) for $4, or 500 such cubes (50 each of 10 colors) for $20. The cubes have sides of about 5/8 inch. Similar plastic cubes, one centimeter on a side, are available in packages of 1,000 (100 each of 10 colors) for $14.

Last month's problem was to prove, using only seven points, that if the points of the plane are arbitrarily colored by two colors (say red and black), at least one set will contain the vertexes of a monochromatic triangle. The proof is: Consider any two red points. Call them 1 and 2, then add five more points in a triangular lattice pattern [*see bottom illustration at left*]. To avoid red equilateral triangles, 3 and 4 must be black. This in turn requires that 5 be red (otherwise triangle 3, 4, 5 is blue), which in turn requires that 6 be black (otherwise triangle 2, 5, 6 is red). Point 7 must be red or black. If red, triangle 1, 5, 7 is red; if black, triangle 4, 6, 7 is black.

Several hundred trenchant letters about Newcomb's decision paradox, July's topic, have been received, and more are still arriving. There have been so many letters that it is not possible for me even to acknowledge them, but Robert Nozick, who wrote the first paper about this bewildering problem, has agreed to read all the letters and comment on them in a guest column that will appear sometime early in 1974.

AN EASY WAY TO CHANGE JOBS!

'My system is simple, straight forward and quick – but it works. Do it right and I guarantee the highest earnings of your career!'

I've got something that can help you earn a great new living and perhaps even make you rich!

Now, I know this seems hard to believe, but I can also *prove* it to you!

I say this because I've got letters from *thousands* of satisfied customers, ... and a product which has been praised by 150 leading media!

What's more, at $10, it's an incredible bargain, and I'll even give it to you without asking you to risk *one penny!*

However, first let me tell you what I have.

I've got a *copyrighted job changing system* that you can use to move up in your field, or out to another field, but at significantly higher earnings.

It took myself and five other professionals two years and $250,000 to develop–*but it works!*

Furthermore, it doesn't require "genius" and it doesn't require "luck." All you have to do is put it into action.

The reason we developed it was because with 84 million employed, and 15 million circulating resumes each year, this area was ready for some revolutionary ideas.

We knew more people than ever owned prestige cars & yachts, summer homes and international retreats, as well as having securities, real estate holdings and lots of cash in the bank.

In short, many people in the U.S. are living good lives!

At the same time, however, the great majority have no excess cash, little job security, and are frequently restless, bored with their jobs, commuting long hours, and harrassed by inflation!

We asked ourselves how do people get to live the "good life"?

Well, we found that most successful people were there because they never wasted time in dead-end situations!

What these people did was to make crucial job changes, and *parlay* their higher earnings into small fortunes!

Take a look at the economics!

Do you realize that if you were to change jobs every 4 years, at an average annual increase of $4,000, and then put the increases in the bank at 6%,–that in 20 years you'd accumulate an extra *half million dollars!*

Getting raises is one thing, but getting significant increases because of job changes is a very important source for wealth!

The next question then, is how can you easily change jobs? This is where the unique system we've developed fits in.

Our system can work for anyone from $8,000 to $80,000. Do it right and you'll gain higher earnings, lifelong job security, but most of all, *everlasting* self confidence!

This is because once you've used it, you'll know you can *always* get a new job,–quickly and predictably.

Perhaps you're wondering why our system works? Well, it works because it's a *completely different approach,* based on totally new concepts.

But, also because it's simple, practical, and self-tailoring. You could start next week–and do it *without strain, confusion or worry.*

But, there is one catch! You won't be a success if you use old methods that have you sitting in employment agencies, laboring over application forms, answering blind ads, and scattering resumes.

To make more money without a hassle, you'll have to be willing to change. You'll also have to follow our system, have an open mind & have faith in yourself.

However, do this and a better life will be yours!

With our system, whatever you seek–a better job, a new career, higher pay, more satisfaction,–*I believe nothing can stop your success!*

Not age, sex, education, or even low earnings or past working history.

Personnel Magazine said we have a *"breakthrough."*

Business Week devoted a full page article and called it *"indispensable."*

The National Public Accountant even said it was *"capable of catapulting any average person into a position offering much greater rewards."*

However, your best proof of our system is that we've already received thousands of letters from grateful customers.

Letters like one from a gentleman in Arizona who wrote, *"In 4 weeks I changed jobs and raised my salary 33%! I wish I had it 10 years ago!"*

Another man from New York said *"I used one of your letters, sent 24 out, and got 13 interviews and 3 job offers!"*

Still another from California said *"In just 11 days I received an offer of $7,000 more!"*

I know this sounds almost too easy and I can't promise that you will do as well. But, then again you may do better!

Even the largest business magazine in the U.S., **Nation's Business**, said our system was *"incredibly effective."*

Now, if you're serious about wanting to move up, then I know that our system is something you've got to have!

In fact, I'm so convinced that you'll agree that it's worth *hundreds of times the cost,* that I'll make sure you have nothing to lose.

First of all, when your order arrives, we'll ship within 24 hours. No delays!

Secondly, you can examine our system for 15 days.

Third, if at the end of that time you are dissatisfied, return it, & I personally guarantee your 100% refund will be mailed in 3 working days–with no questions asked!

To let me prove everything I've said, and to take advantage of this nothing-to-lose offer, just fill in and mail me the coupon below.

Performance Dynamics Inc.
Attn: Mr. Robert Jameson, President
17 Grove Avenue
Verona, New Jersey 07044

Dear Mr. Jameson:

Your offer sounds great! Please rush me your professional job hunting system right away, but on one condition. I understand I may examine it for 15 days, & if at the end of that time I return it, you will mail my full refund within 3 working days, with no questions asked. On that basis, here's my ten dollars.

☐ Enclosed is my check or money order
☐ Chg. Bank Amer. ☐ Chg. Amer. Expr.
☐ Chg. Master Chg. ☐ Chg. Diners Club.

Acct. # Exp. Date

Name

Street

City State Zip

Please note: Postage is paid via parcel post. For faster delivery add Spec. Del. $1.00; U.S. Air $3.00; Fgn. Air $7.00. sa10

Mr. Jameson's ideas have been the subject of more than five hundred articles, ranging from 600 words in Business Week to 3,000 words in Chicago Today. This material has also been nationally advertised in leading media including The Wall Street Journal, Scientific American, Nation's Business, Signature, The New York Times, Newsweek International, The Los Angeles Times, American Scientist, Income Opportunities, Time, Specialty Salesman, Success Unlimited, Chemist, The Army Times, New York, The Chicago Tribune, True & others.

THE AMATEUR SCIENTIST

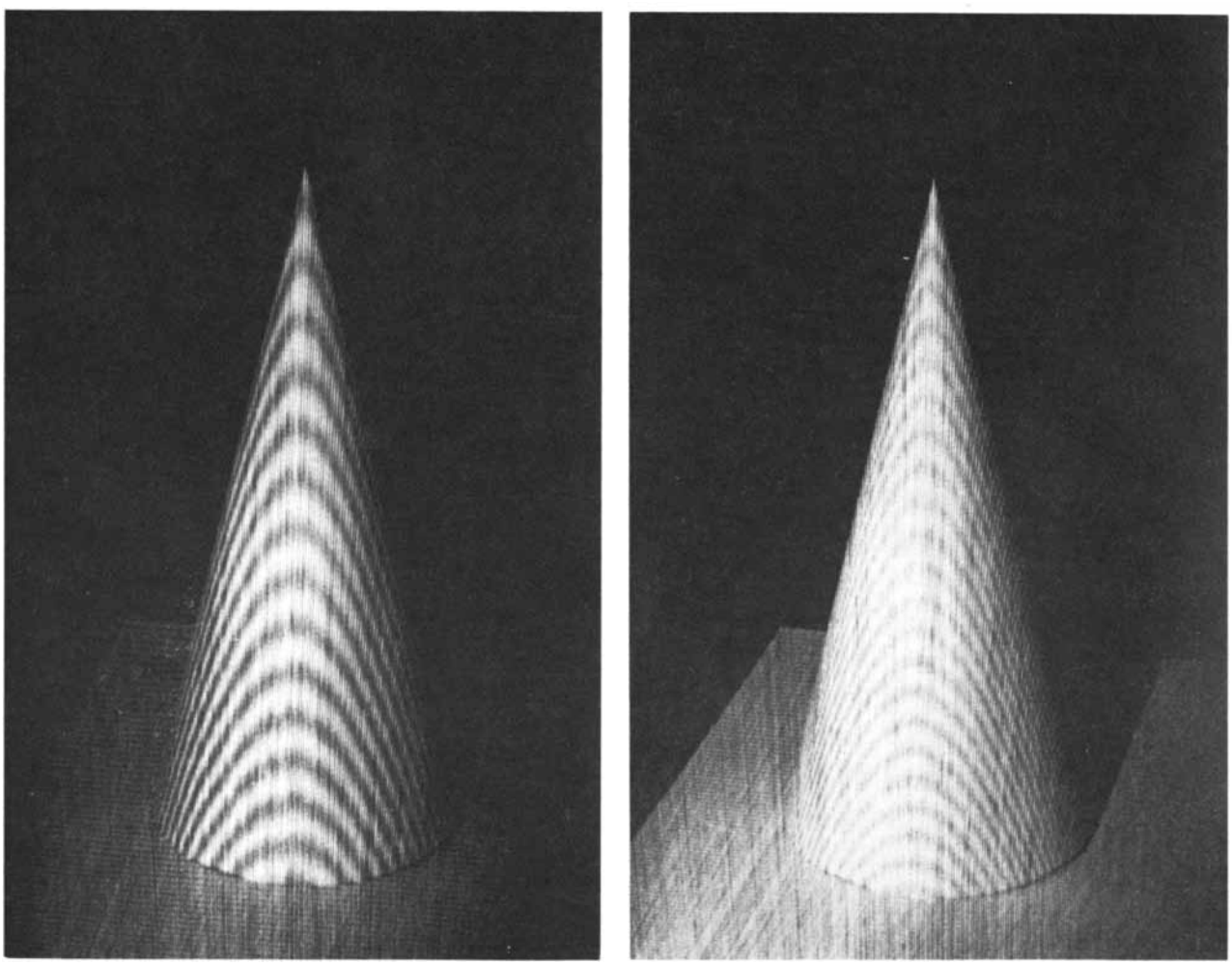
Moiré patterns draw contours, and a meter determines current in electrostatic circuits

Conducted by C. L. Stong

The array of dark and light bands known as a moiré pattern appears when light is viewed through overlapping window screens or similar structures consisting of repetitive elements. The repetitive elements need not be straight lines, such as the wires in window screening. Two sets of concentric circles that overlap generate moiré patterns in the form of radial lines. Concentric circles that overlap a grid of comparable spacing generate ellipses, parabolas or hyperbolas, depending on the angle of inclination between the plane of the concentric circles and the plane of the grid.

A moiré pattern can represent the solution of a mathematical equation. For example, the curves produced by superposing a grid (which can be regarded as the projection of a plane) and concentric circles (which can similarly be regarded as the projection of a cone) generate patterns that are solutions of equations that express the conic sections. Many other interesting properties of moiré patterns and some of their practical applications were investigated 10 years ago by Gerald Oster and Yasunori Nishijima [see "Moiré Patterns," by Gerald Oster and Yasunori Nishijima; SCIENTIFIC AMERICAN, May, 1963].

Among other techniques for generating moiré patterns, Oster and Nishijima suggested that the figures could be observed by looking through a periodic structure at its own shadow. For exam-

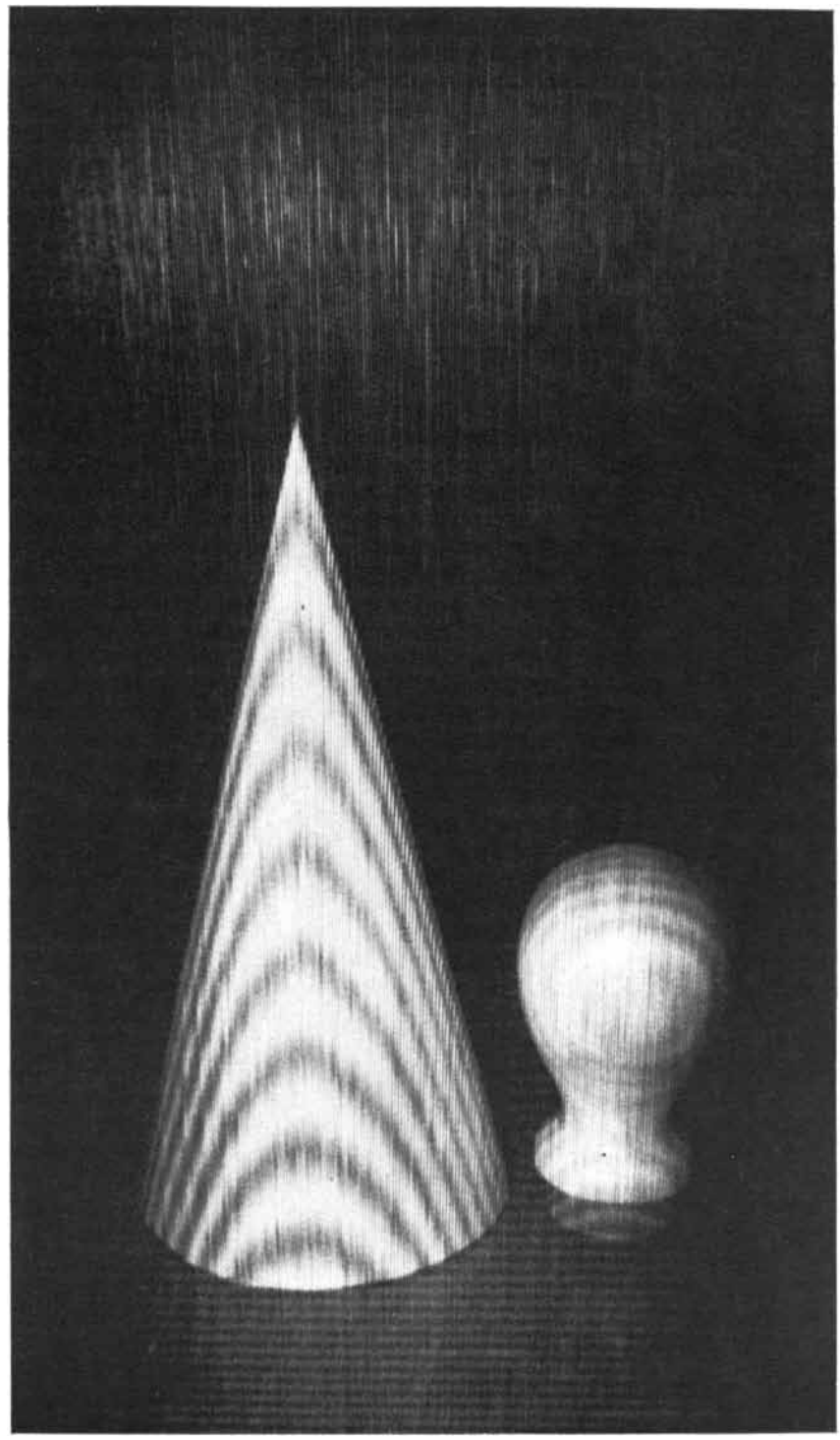


Change of moiré patterns resulting from shifting the position of the camera

ple, a pattern can be seen by casting the shadow of a window screen on a wall. Depending on the observer's point of view, some lines of the screen appear to be superposed exactly on top of some lines of the shadow. The pattern appears light in those regions. In other regions the lines of the screen fall between lines of the shadow. The pattern appears dark in those regions. The number of dark and light zones that are generated and the location of the zones within the field of view depend on the distance and shape of the object on which the shadow falls. For example, the shadow of a grid that falls on a flat surface generates a moiré pattern that also has the form of a grid. The spacing of the lines in this pattern varies inversely with the angle made by the plane with respect to the point of view. In contrast, the shadow of a grid that falls on a curved surface generates a pattern of dark and light bands that represent contour lines of the surface. Oster suggested that this effect, which is extraordinarily sensitive, might be used for making a contour map of the moon. A grid of known dimensions would be projected from the earth onto the moon's surface. The resulting shadow would be photographed through the grid. The moiré pattern would disclose surface features of the moon that escape not only telescopes but also cameras in close lunar orbit. This heroic scheme has not been tried for making lunar measurements, but a series of experiments that validate the concept has been made by Bill Lyon (3000 Fillmore Avenue, El Paso, Tex. 79930), who writes:

"My apparatus consists of a projection lamp, a grid of opaque lines, specimens of various shapes and a camera for recording the moiré patterns. The projection lamp is placed at an arbitrary distance from the grid. Specimens occupy an arbitrary position beyond the grid. The camera is put next to the lamp and at an arbitrary distance from it.

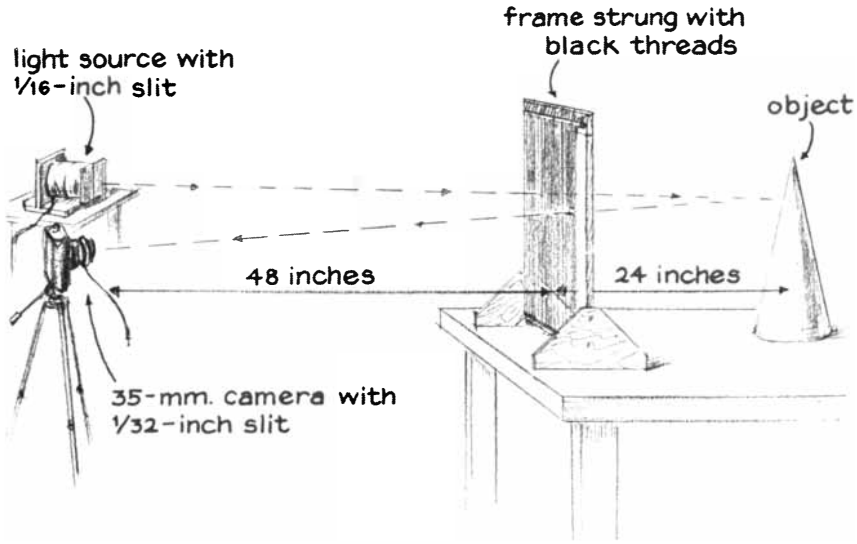
"In a typical experiment the lamp could be placed roughly four feet from the grid, the specimen two feet behind the grid and the camera a foot or less from the lamp [*see top illustration on next page*]. I was interested more in the general form of the patterns than in the sharpness or resolution of the bands that make them. For this reason I made a vertical slit for the lamp instead of employing a more effective but costly projection lens. The slit, which functions something like a cylindrical lens, casts reasonably sharp shadows of the grid on the specimen. The quality of the final image was further improved by providing a slit for the camera.



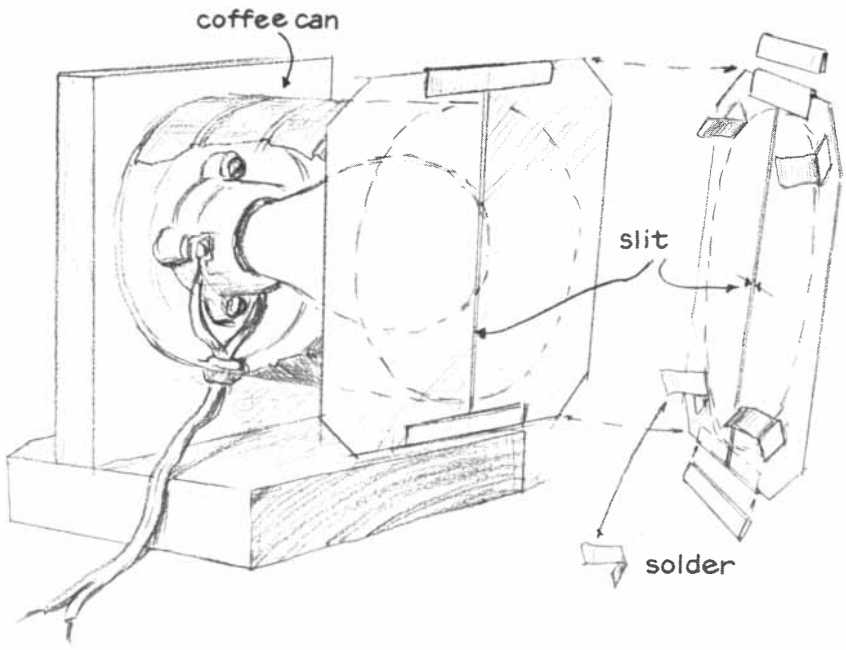
Moiré patterns generated by two three-dimensional objects

"The lamp housing was made from a discarded coffee can. A porcelain socket was mounted inside the can with a pair of machine screws. A pair of wood screws that passed through the bottom of the can fastened the lamp assembly to a wood base. The lamp cord was brought out of the can through a rubber grommet in the side. The grommet prevents the sharp edge of the metal from cutting through the insulation of the cord and short-circuiting the power line.

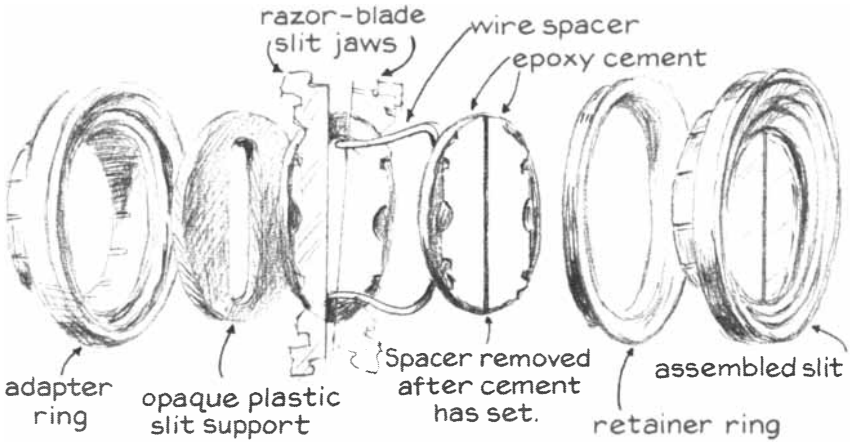
"The lid of the can was replaced by a sheet-metal top in the form of a slit. The sheet metal was obtained by cutting the top and bottom from a second tin can and flattening the side. Two rectangular pieces were cut from the strip, each somewhat wider than the radius of the can. They were soldered to clips of the



Bill Lyon's setup for shadow-generated moiré patterns



Details of the projection lamp



Arrangements for the camera slit

same material to form a slit approximately 1/16 inch in diameter. Four 90-degree brackets of sheet metal were then soldered in a circular array to the rear surface of the slit. The brackets serve as clamps for attaching the slit to the housing.

"The grid consists of black strings approximately 1/32 inch in diameter stretched in a diagonal array between opposite sides of a wood frame. Uniform spacing between the strings is achieved by supporting each strand, near its ends, in the thread of a long screw that is mounted at right angles on each side of the frame. The screws are prethreaded iron rod that is available in 36-inch lengths from dealers in hardware as stock from which steel bolts are cut. The dimensions of the grid are not crucial. It can be made two feet square or larger, depending on the space available to the experimenter.

"The frame, to which the screws are attached, can be made of one-by-two-inch lumber. I beveled the edges of the frame that are adjacent to the screws at an angle of about 45 degrees. Small brads spaced 1/8 inch apart in two staggered rows were driven into the beveled face. I was then ready to lace the frame by passing a loop of string around each brad and through the adjacent thread of the screw [*see illustration on opposite page*].

"The photographs were made with a 35-millimeter camera of the single-lens reflex type. Most of the photographs were made at a camera-grid distance ranging from two to three feet. The optimum distance must be determined experimentally. It will vary with the focal length of the lens. The mechanical slit for the camera can be made by the technique shown in the accompanying illustration [*at bottom on this page*].

"To photograph moiré patterns place an object six inches or more in height and diameter behind the grid and illuminate it through the grid with the projection lamp. Put the camera adjacent to the lamp, focus it on the specimen and set the lens at the minimum *f* number. Turn on the projection lamp. The moiré pattern cannot be seen by eye because of the faint illumination. Turn off the room light and make an experimental time exposure of approximately two minutes. A moiré pattern should be visible on the developed emulsion. Alter the exposure time as may be required for optimum photographic quality.

"As I have mentioned, the bands of the moiré pattern represent contour lines of the specimen. It has been suggested

that the scheme could be used for making sculptured portraits by means of the photographic process. It should be possible to rough out a model of a three-dimensional object by photographing its moiré pattern and using the bands as contour lines to guide a pantographic carving machine. Machines of this kind are currently used by die makers for sinking rough contours of models into blocks of steel. A skilled carver completes the die by hand. In my own case the experiments were of interest as aids in grasping the fundamentals of making and reading contour maps."

An amateur who builds a radio set, a small computer or a high-fidelity sound system routinely measures voltages, current and resistances with inexpensive meters. The measurements enable him to improve the performance of his apparatus and thus to increase the satisfactions of his avocation. In contrast, someone who undertakes to construct an electrostatic device such as a Van de Graaff generator, an electret motor or an ultrasonic microphone rarely makes measurements. The cost of instruments that work reliably in electrostatic circuits has been beyond the reach of most amateurs.

Recent advances in solid-state technology have altered the basic design of such meters in ways that reduce their cost dramatically. For example, a relatively inexpensive microammeter that is specifically intended for high-voltage electrostatic circuits has been designed expressly for amateur construction by R. H. Kaufmann (2208 Dean Street, Schenectady, N.Y. 12309), a retired electronics engineer. The instrument measures current from 25 millionths of an ampere to five billionths of an ampere in circuits that carry potentials of up to 500,000 volts. Kaufmann discusses the design and calibration of the meter:

"The meter was constructed for measuring minute currents in various parts of electrostatic generators and other devices, particularly currents that leak into the atmosphere by the phenomenon known as corona discharge. Essentially the instrument consists of a zero-centered meter of the d'Arsonval type that is actuated by a pair of matched field-effect transistors connected to operate as a differential amplifier [*see top illustration on next page*]. The sensitivity of the instrument can be adjusted by a four-position switch.

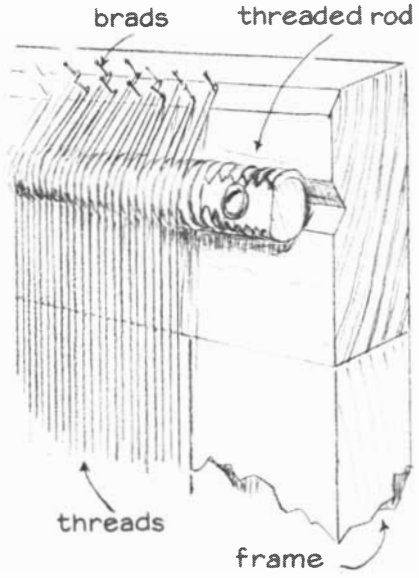
"Before undertaking the project I wrecked two ordinary microammeters during attempts to measure current in electrostatic circuits. The delicate coils that turned on finely wrought pivots were damaged by transient surges of current that accompanied the sparkover of high voltages. Accordingly, when I was designing the new meter, I shielded all working parts in a Faraday cage that consisted of a metal housing lined with copper screening. The plastic face of the meter was also covered with screening that was bonded electrically to the housing. Finally, all shielding was connected through a transfer switch to one or the other of the two input terminals. The terminals consist of 1/4-inch rods about four inches long that terminate in aluminum spheres 5/8 inch in diameter [*see bottom illustration on next page*].

"The zero-centered scale of the meter is graduated to indicate ±25 microamperes. The resistance of the meter, which is 1,200 ohms, must be taken into account when the amplifier is designed. To measure the resistance of the coil in any meter without damaging the instrument, divide 1.5 by the maximum current for which the meter is calibrated. The result is equal to the value in ohms of a resistor that, when it is connected in series with the meter and with a dry cell of 1.5 volts, will cause the pointer of the meter to move to the full-scale graduation. For example, a meter that is calibrated to measure maximum current of 25 microamperes will indicate full scale when its terminals are connected to a 1.5-volt battery in series with a resistor of 60,000 ohms ($1.5/25 \times 10^{-6} = 60{,}000$).

"With the meter thus indicating full scale, connect a variable resistor of, say, 5,000 ohms across the terminals. Gradually reduce the value of the variable resistor. The pointer of the meter will simultaneously drift toward zero. Continue to reduce the value of the variable resistor until the pointer indicates exactly half of the maximum reading. The value of the variable resistor at this adjustment is exactly equal to the resistance of the coil in the meter.

"The electrostatic microammeter works equally well with meter coils of any reasonable resistance, but if the value differs substantially from 1,200 ohms, the shunt resistors of the differential amplifier (R_S, R_T, R_U and R_V) must be altered from the values indicated by the accompanying circuit diagram. The procedure for determining the appropriate values will be explained.

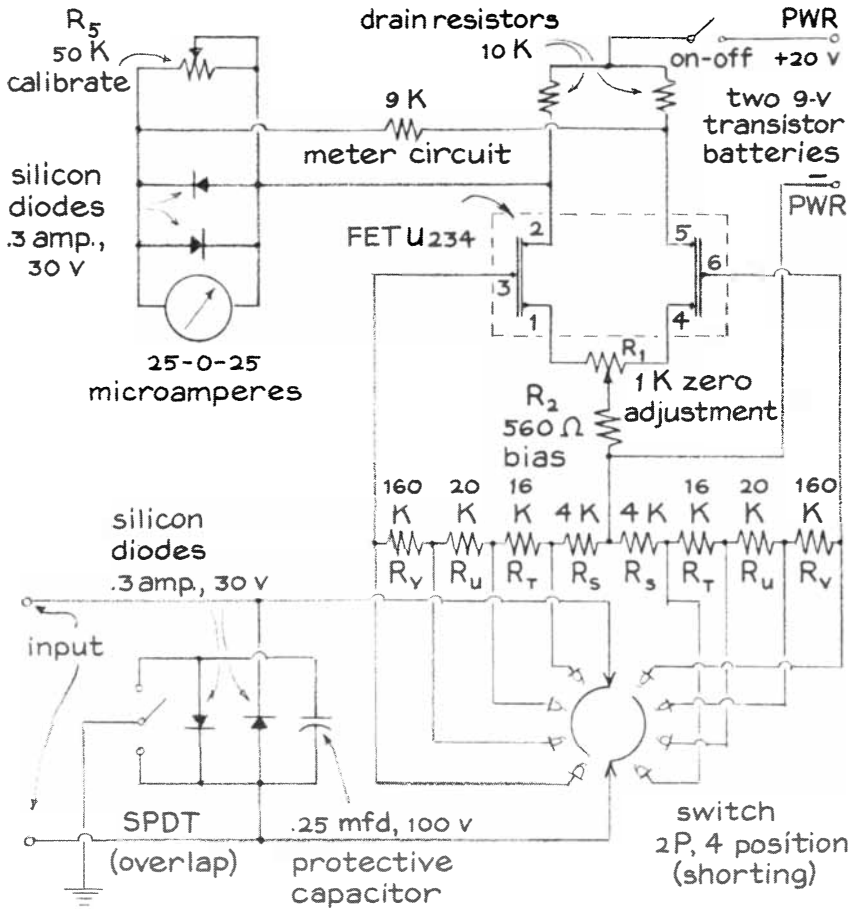
"The heart of the differential amplifier is a matched pair of alloy-junction, field-effect transistors potted in a single housing (No. U234). The devices are available from Semiconductor Specialists, Inc. (P.O. Box 66125, O'Hare International Airport, Chicago, Ill. 60666). The list price of the device is $3.90, but the distributor accepts only orders in the minimum amount of $5. (Orders from amateurs should be directed to the attention of Richard K. Dahlem.) The characteristics of the device vary somewhat from unit to unit. For this reason I found it necessary as well as interesting to measure the characteristics of each device I bought.



Lacing the grid

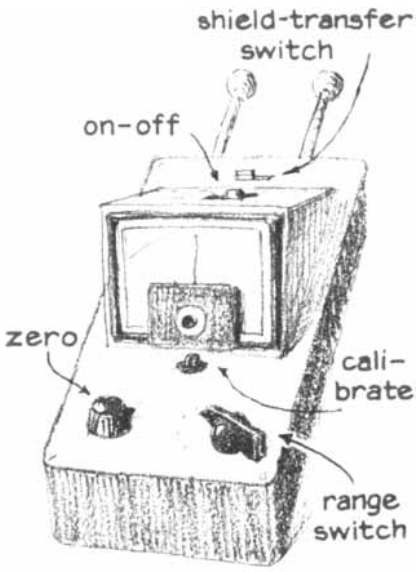
"Field-effect transistors include three terminals known as (1) the source, *S*, (2) the drain, *D*, and (3) the gate, *G*. To test the U234 I connected each of the two transistors sequentially in a simple circuit that included a 0–3 voltmeter and a 0–3 milliammeter [*see top illustration on page 125*]. A fixed potential of approximately 20 volts was applied across the source-drain (1–2) terminals of the transistor. Similarly, a potential that was gradually increased from zero to three volts by means of a five-kilohm potentiometer was applied between the source and the gate (1–3) of the transistor. This potential, designated V_{GS}, was tabulated. Simultaneously, corresponding current in the drain circuit, I_D, was observed and also tabulated. The resulting data were then plotted as a graph that displayed the drain current, I_D, against the corresponding gate-source voltage, V_{GS} [*see bottom illustration on page 125*].

"Note that the shape of the graph approaches a straight line within an interval that corresponds to a gate-source po-



Circuitry of R. H. Kaufmann's electrostatic microammeter

tential of between −1 volt and −2 volts. In other words, through this interval current in the drain circuit of the transistor varies in direct proportion to the gate-source voltage. The operation of the amplifier is said to be linear within this range.



Microammeter as completed

"The center of the linear portion of the graph corresponds, approximately, to a V_{GS} potential of 1.6 volts. At this potential the drain circuit of the device draws one milliampere. Excursions of V_{GS} from 1.6 volts cause directly proportional excursions of current, I_D, in the drain circuit. The potential of 1.6 volts is known as the operating point of the transistor. The operating point differs from one No. U234 device to another and must be determined experimentally for each device.

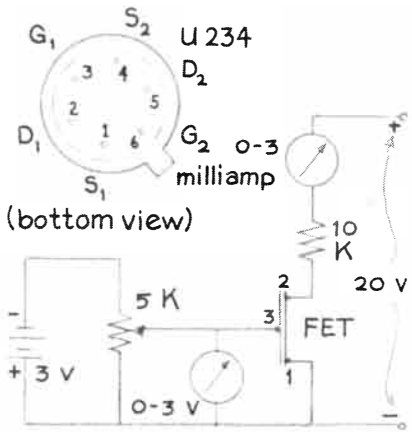
"Next I assumed a value of 1,000 ohms for the resistor identified as R_1. On the basis of this assumption and the data determined by the preceding test I calculated the value of resistor R_2. The arithmetical procedure is specified in simple, abbreviated form by the equation $R_2 = (V_{GS}/2\ \ I_D) - R_1/4$. After the corresponding numbers have been substituted the equation becomes $(1.6/.002) - (1{,}000/4) = 800 - 250 = 550 = R_2$. These resistors, R_1 and R_2, establish a bias voltage on the gates of the field-effect transistors of such value that the meter can be set to zero when no voltage is present across the input terminals of the instrument.

"To produce meter deflection from this steady-state zero position a differential voltage must be applied to the gates of the two transistors. The required voltage originates across the shunt resistors: R_S, R_T, R_U and R_V. The value of these resistors is determined by a relatively simple procedure. Before it can be undertaken, however, one must wire the remaining portions of the circuit, including the calibration potentiometer, the drain resistors, the meter and the protective diodes. The diodes are not essential. In effect, they act as fuses to protect the meter. Those who undertake the construction, however, are reminded that diodes are much less expensive to replace than the meter.

"Use a pair of matched 4,000-ohm resistors for R_S. Operate the four-position selector switch to connect the outboard ends of the resistors to the gates of the transistors. Set the calibration resistor, R_5, to its maximum value. Apply an adjustable source of low voltage to the input terminals and hence across the R_S resistors. Increase the voltage until the pointer stands at full-scale deflection. With any accurate microammeter measure the current in the R_S resistors. If 4,000 ohms is approximately correct for R_S, the voltage across the gates of the transistors (and across R_S) is equal to the current through the R_S pair multiplied by 8×10^3.

"Assume that the current amounts to 25×10^{-6} ampere, a typical value. The voltage across the transistors is equal to the product of the current multiplied by the resistance, or $25 \times 10^{-6} \times 2 \times 4 \times 10^3$, or .2, volt. (The transistors operate satisfactorily in this application with any potential across the gates between the limits of .5 and .1 volt.) Had the measured voltage fallen outside this limit, the value of R_S would have been increased or decreased from 4,000 ohms as necessary to develop .2 volt. With R_S equal to 4,000 ohms in this specific circuit, the meter indicates a full-scale deflection of 25 microamperes.

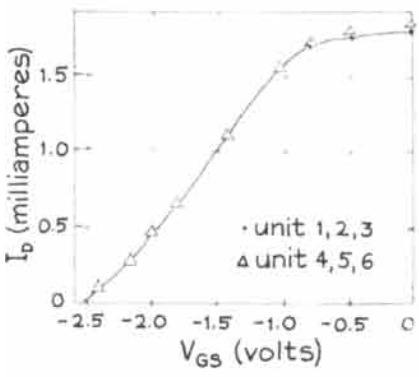
"Next I determine the value of the succeeding shunt resistors that control the gain of the amplifier and thus fix the scale of the meter. Assume that the next more sensitive full-scale value is to be five microamperes. The controlling resistor, R_T, is equal to $(.2/2\ I_L) - R_S$, in which I_L is the current $(5 \times 10^{-6}$ ampere) that is desired for full-scale deflection and $R_S = 4 \times 10^3$ has been previously determined. These values, when substituted in the equation, determine the value of R_T: $(.2/2 \times 5 \times 10^{-6}) - 4 \times 10^3 = 16{,}000$ ohms. Similarly, the value of R_U that would be required for



Circuit for testing transistors

the still greater sensitivity of 2.5-microampere full-scale deflection is found by dividing .2 by the product of two times 2.5 microamperes and subtracting the sum of R_T plus R_S. Thus $R_U = (.2/2 \times 2.5 \times 10^{-6}) - 4 \times 10^3 + 1.6 \times 10^4 = 20$ kilohms. Finally, full-scale deflection in response to an input of .5 microampere, which is equivalent to a sensitivity of 10 billionths of an ampere per scale division, can be achieved by making R_V equal to $.2/2 \times .5 \times 10^{-6}$ minus $R_S + R_T = R_U$, or $(.2/10^{-6}) - 4 \times 10^4 = 160$, kilohms.

"The diodes and the capacitor that function as protective devices must be of excellent quality to avoid the leakage of current that would impair the accuracy of the instrument. The forward conduction of the diodes should not exceed 10 billionths of an ampere at the normal working voltage of the instrument. Leakage through the capacitor should not exceed 10^{-10} ampere at an impressed potential of 10 volts. With the exception of the indicating meter and the matched pair of field-emission transistors, I constructed the instrument entirely of parts salvaged from surplus electronic assemblies."



Characteristics of transistors

How many of these best sellers belong in your scientific library?

Pick 3 that you don't have all for only $3.95 for yourself or for holiday gift-giving.

Choose from books on every scientific discipline including Mathematics · Astronomy · Cartography · Chemistry · Environmental Science · Physics · Genetics · and much more.

(retail prices shown)

50110. GENETICS OF THE EVOLUTIONARY PROCESS. *Theodosius Dobzhansky.* Definitive work on the modern synthetic theory of evolution by the world's foremost geneticist. **$12.50**

62801. MODERN COSMOLOGY. *D. W. Sciama.* "From the standpoint of readability, orderliness, continuity, evenness of level, and its ability to hold one's attention, the book is three standard deviations above the mean." *Physics Today* **$9.95**

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

63131. MOLECULES AND LIFE. *Joseph S. Fruton.* The interplay of chemical and biological ideas, with full documentation for the period from about 1800 to 1950. *Counts as 2 of your 3 books.* **$19.95**

55561. THE INSECT SOCIETIES. *Edward O. Wilson.* "For layman or specialist interested in a single, concise, lucid, and authoritative account of the most significant facts and theories about insect societies."—*Natural History. Counts as 2 of your 3 books.* **$20.00**

45471. ENVIRONMENTAL GEOSCIENCE. *Arthur N. Strahler and Alan H. Strahler.* Masterfully explores the physical and biological processes involved in interactions between Man and the environment of energy and matter which surrounds him. **$12.95**

56530. ISLAND AT THE CENTER OF THE WORLD. *Father Sebastian Englert.* Easter Island's mysterious giant statues, the islanders' puzzling past, their lost language—the island's priest for 35 years sheds new light on them all. **$15.00**

33430. THE ANALYSIS OF INFORMATION SYSTEMS. Revised Edition. *Charles T. Meadow.* An all-inclusive treatment of the four aspects of information processing and retrieval. **$14.95**

50281. GIANT REPTILES. *Sherman A. Minton, Jr., and Madge Ruthertord Minton.* A storehouse of information of the world's giant reptiles world. **$9.95**

70070. PRINCIPLES OF HOLOGRAPHY. *Howard M. Smith.* Comprehensively covers the history, theory, practice, and applications of this exciting technique. **$11.95**

61360. MATHEMATICAL STATISTICS WITH APPLICATIONS. *William Mendenhall and Richard L. Schaeffer.* Stresses inference as the objective of statistics and emphasizes the role of probability in inference making. **$12.95**

36141. BEYOND STONEHENGE. *Gerald S. Hawkins.* The author of *Stonehenge Decoded* assesses the astronomical significance of temples of the Nile, Mayan ruins, the linear drawings in the Peruvian desert, and the geometrical earthworks of the U.S. **$10.00**

84451. TOWARD A GENERAL SCIENCE OF VIABLE SYSTEMS. *Arthur S. Iberall.* Introduces a science that attempts to unify our understanding of all autonomous systems—natural, organic or man-made—proposing a continuum ranging from nuclear particles at one end of the scale to the cosmos at the other. **$14.50**

Take any 3 books
(values to $47.45)
for only $3.95
with membership in
The Library of Science
You simply agree to buy 3 more books—all at handsome discount—over the next 12 months.

36280. A BIOGRAPHICAL DICTIONARY OF SCIENTISTS. *Edited by Trevor I. Williams.* Over 1000 biographies of scientists in every field, from agriculture to zoology. Describes their lives, assesses their contributions and relates them to those of their predecessors, their contemporaries, and their successors. **$9.95**

44350. ELEMENTS OF CARTOGRAPHY. Third Edition. *Arthur H. Robinson and Randall D. Sale.* A wealth of new material, especially on techniques of air-mapping, and the most recent experiments in design and reproduction. **$12.75**

36760. BORANES IN ORGANIC CHEMISTRY. *Herbert C. Brown.* One of the world's leading researchers in organic synthesis reviews the work of a lifetime in a stimulating volume that will appeal to scientists of every stripe. *Counts as 2 of your 3 books.* **$24.50**

82880. TEXTBOOK OF POLYMER SCIENCE. Second Edition. *Fred W. Billmeyer.* A brilliant and best-selling presentation of the physics, chemistry and technology of polymers, the most important synthetic materials developed by industry. **$17.50**

How the club operates

The Book Club **News,** describing the coming Main Selection and Alternate Selections, will be sent to you 15 times a year at three to four week intervals. • If you wish to purchase the Main Selection, do nothing and it will be shipped to you automatically. • If you prefer one of the Alternates, or no book at all, simply indicate your decision on the reply form always enclosed with the **News** and mail it so we receive it by the date specified. • The **News** is mailed in time to allow you at least 10 days to decide if you want the coming Main Selection. If, because of late mail delivery of the **News,** you should ever receive a Main Selection without having had the 10-day consideration period, that Selection may be returned at Club expense. • After completing your trial membership, you will be entitled to take advantage of our fabulous bonus plan which offers savings up to 70% and more. • (Offer good in Continental United States and Canada only. Prices are slightly higher in Canada.)

Membership Application

The Library of Science
Riverside, New Jersey 08075

Please accept my application for membership and send me the three volumes indicated billing me only $3.95 for all three. I agree to purchase at least three additional Selections or Alternates during the first year I am a member, at special members' prices. Savings range up to 30% and occasionally even more. My membership is cancelable any time after I buy these three books. A shipping and handling charge is added to all shipments.

Send no money. Members are billed when books arrive.

Indicate by number the 3 books you want.

[ ]
[ ]
[ ]

Some expensive books count as 2 selections.

2-49Y

Name_____

Address_____

City & State_____ Zip_____

Book selections purchased for professional purposes may be a tax-deductible expense.

BOOKS

Mars, both real and romantic, and the principles of cosmetics

by Philip Morrison

MARS AND THE MIND OF MAN, by Ray Bradbury, Arthur C. Clarke, Bruce Murray, Carl Sagan and Walter Sullivan. Harper & Row ($7.95). Edgar Rice Burroughs, a writer of easy fiction "lacking refinement, with exquisite vulgarity," and Percival Lowell, the dreaming "Boston brahmin" who was "one of the worst draftsmen who ever sat down at the telescope," made our Mars, at least for the first two-thirds of this century. Burroughs dealt in no real Mars at all but used the planet metaphorically, as a place of jejune adventure and romance. However vulgar his prose in form and texture, the power of his imagination dominated his readers. His other creation, the inarticulate Tarzan of the Apes, is known worldwide perhaps to more people than any other literary character, the cinema bridging all barriers of language; moreover, Tarzan's identity flows from a recognized kinship older than that boasted by any epic hero. Lowell wrote and sketched his self-consistent optical illusion for decades; his canals too became bywords, although they exist no more than do the gem-studded cliffs of gold that Burroughs' John Carter saw on Mars.

Mariner 9 photographed the real Mars from a near orbit, after its arrival late in 1971 during a planet-wide storm until the orienting jets ran out of gas in October, 1972. About 50 remarkable full-page halftone reproductions chosen from those 7,500 television images are the burden of this unusual book. Here are the largest shield volcano, bigger than all Hawaii; the continental valley, more striking than the East African Rift; the braided stream channels in which now no water flows. (Some effort at orienting the page to make the relief appear correctly, craters as holes and not as lumps, is required of the reader.)

The brief text is unusual. The first half presents an edited encounter between the five men who met in public discussion on the eve of *Mariner 9*'s arrival, all brought to Pasadena by their interest in the event. Bradbury and Clarke are of course famous writers of science fiction, Sullivan is the science editor of *The New York Times* and Murray and Sagan are two scientist participants in the Mariner program. In their speculations Murray, a geologist, accepts the role of skeptic; Sagan, trained in biology and in the science of planets, takes a more open and romantic view of what *Mariner 9* and its successors may find.

The rest of the text, even more interesting, consists of afterthoughts, remarks written a year later in the light of *Mariner*'s new images, by each of the same five men, this time unfortunately without conversational exchanges.

There are issues still outstanding. *Mariner 9* showed no great feudal world of warring nations along the canals. But the view of Mars as cratered, dead and moonlike was also discarded. Mars is more like the earth than that; it has basalt plateaus and probably plate tectonics. Hot dry dust, instead of torrents of condensing rainwater, supplies the energy that drives those global hurricanes. Mars has glaciers and maybe has had ice ages. All the landscapes are windblown, and dunes stretch long across the page. Is Mars only now becoming watery, earthlike? Or do those dry stream beds bespeak a long, episodic, watery past of possible life?

We do not know. Maybe *Mars 4*, *Mars 5* and *Mars 6* will help; these space probes, carefully sterilized by their Russian designers, are due to arrive on the Martian surface next February. *Viking*, which the U.S. expects to launch in time to land on July 4, 1976, will make an explicit if preliminary search for microscopic life forms. Someday we shall know; we shall hear the winds and touch the dust.

That remains the goal these men see in Mars trips: "surprise . . . not blasphemous but celebratory," not "how the Apollo missions have given old ladies new plastic cooking pans made from nose-cone chemistries," but (Bradbury) "a gigantic whisper which has touched our multitudinous ears ever since, and that whisper says: Sweet man, dear blood, wild creature, rare device of the universe, fragile flesh—survive."

HARRY'S COSMETICOLOGY, FORMERLY THE PRINCIPLES AND PRACTICE OF MODERN COSMETICS, by Ralph G. Harry, revised by J. B. Wilkinson, in cooperation with P. Alexander, E. Green, B. A. Scott and D. L. Wedderburn. Chemical Publishing Co., Inc. ($35). Ralph G. Harry, F.R.I.C., F.R.S.M., A.R.P.S., wrought so well when in 1940 he published the first thin edition of this volume (now in its hefty sixth version) that his book "has now reached that degree of establishment recognized by its familiar name," like Bowditch or Spock. The puff is credible, because the entire volume speaks so fully and reasonably, even a little ironically, about its topic, generally found only at the pinnacle of ill-argued claims, that one is well disposed toward its simple praise of Harry.

The book opens for the reader an unexpected world. It is the productive, reasonable, working world of manufacturing chemists and clinicians that lies behind the vain models and clamorous commercials. The products must be devised, compounded, packed. If a face powder cannot make an ugly duckling into a glittering swan, it ought at least to "impart a smooth finish to the skin, masking minor visible imperfections and any shine. . . . The degree of opacity of the powder can vary from opaque and matte, as for example a clown's make-up to almost transparent. . . . Finally, it should serve as a vehicle for a pleasing odor to be disseminated by . . . perfume-laden particles over a warm and relatively large area. . . . No single substance possesses all the desired properties—covering power, slip, absorbency, adhesiveness and bloom—hence a modern face powder is a blend of several constituents, each one chosen for some specific quality." *Harry's Cosmeticology* describes the choice of materials, the parameter ranges, gives representative formulas and implies clearly that changing

fashion and advertising efforts accompany more or less systematic excursions through that complex space of all face powders. Similar descriptions and discussions are provided for the entire gamut of cosmetic products (excluding ordinary toilet soap).

A good face powder today is (and has been for a very long time) mainly the mineral talc. It must be of a good grade; the poorer stuff may be contaminated with tetanus spores, dangerously resistant to sterilization. It should be ground to submicron size to allow light scattering and hence covering power. Talc provides the "slip," "a characteristic smooth feeling on the skin." It clings moderately well to the face, but it does not absorb moisture nor does it always cover well enough. Hence a talc base will often be augmented with an opacifying pigment such as zinc oxide or titanium dioxide, and with rice starch to absorb fatty skin secretions and sweat, plus perfume and color—proprietary subtleties not included in Harry's general formulas. For people who wish some color and bloom to the face without appearing to be "made up," the right mix is four-fifths talc. For those who wish a definite opaque matte appearance, talc is cut to a third, and rice starch or particularly fine precipitated chalk is added to improve absorption. Between extremes lies a world of powders, packaged loose or compacted under high pressure with soapy binders.

The maker who wishes a recipe newer than the Great Wall may now try polystyrene microspheres, polyurethane-foam powders, microcrystalline cellulose or even starches modified by reactions with the hydroxyl radical to lessen swelling of the grains. "When any degree of bloom is required, there are few materials which can surpass starch." Powdered silk, which is silk fiber degummed, hydrolyzed and ground, "apart from any unique or desirable properties it might have, will provide opportunities for... rapturous copy writing." Silk so treated is a patented material.

So the book goes. The anatomy and function of skin, hair, teeth and nails are carefully recounted, emphasizing the normal, and not the pathological, variations of function. Then chapter by chapter the entire range of man's and woman's concern (older than *H. sapiens*) to modify surface is outlined in the same cool, understanding and analytic style. Creams and bath preparations, sun lotions and shampoos, insect repellants, rouges, dyes, antiseptics and humectants —the list is almost a synopsis of television commercials. Each function has its rationale and often its fads and its dangers. The tests that measure function are outlined; they are both direct and ingenious. Preserving, stabilizing, mixing, packaging and even spray-propelling products out of an aerosol container are treated. The work is international; indeed, the list of products and patents seems to transcend boundaries, even though national legislation is often decisive. The trend of the work is of course Anglo-American.

Our hair gains its color from granules injected into the soft cortical cells as they move up from the place of their formation at the root. Two main pigments color hair: the blackish melanin and the yellowish pheomelanin. (Red hair contains other pigments as well, with more iron.) The granules are colorless at first, but the copper-containing enzyme tyrosinase converts certain pigment precursors to a colored polymer form. Failure of this activity is the cause of graying. Albinism is usually the result of an inborn lack of the enzyme; the albino possesses normal melanic granules. Copper deficiency has been shown to induce graying in animals, but there is no consistent way to reverse the normal loss of tyrosinase activity with age. That the hair of Marie Antoinette, say, could have turned white overnight "would appear to be a case of poetic license"; the color in already grown hair is fixed. Perhaps she first lost her hair and the new-grown hair came in gray, or, even less romantic, perhaps she began her ordeal with dyed hair, which whitened after she lost access to her hairdresser.

All humans sweat, women rather less than men; Mongolians sweat less in the region of the larger sweat glands. The action of the most common antiperspirant, based on aluminum salts, is uncertain and complex. The chemistry of hair-waving is also complex. The use of heat and water for this purpose is as old as history; they soften the hair by modifying the hydrogen bonds, making the protein flexible. The hair will then hold any shape given it for a while. (It is untrue, although widely believed, that curly hair has an elliptical cross section and straight hair a circular one.) More reactive treatments, first invented about 1910, evidently affect other, stronger cross-links in addition to the hydrogen bonds.

Striped toothpaste (the color is extruded into the mouth of the tube) and self-heating aerosol shaving lather (the oxidant, hydrogen peroxide, is stored in a plastic bag within the pressure vessel and mixed in the exit opening with reducer in the soap) are clever new tricks of the package designer. There is much more in this compendious overview. Those who would represent the consumer need such information, and it ought to be accessible to young people in school and college libraries.

Recent progress in dentifrices is genuinely valuable; progress in packaging and in some hair products is of at least technical interest. In the field of pure cosmetics, however, there have been no major improvements lately. They are the hoped-for fruit of the study of physiology and biochemistry on normal skin, of those "day-to-day skin aberrations which are not sufficiently serious to merit medical attention." One will wait to see; meanwhile this book joins with the main body of science and technology an ancient art and a modern industry, in the U.S. alone reaching about the size of NASA.

FRICTION: AN INTRODUCTION TO TRIBOLOGY, by Frank Philip Bowden and David Tabor. Anchor Press of Doubleday & Company, Inc. ($5.95). About a fifth of all our gasoline is burned as a sacrifice to friction. At the same time the woven cloth on our back remains a fabric, rather than a slippery set of interlaced threads, because of friction. ("Try tying two dead eels together," suggest these Cambridge authors.)

"If any individual brought science into tribology it was surely Philip Bowden." The first draft of this fresh and nontechnical survey of the modern understanding of friction was written by Bowden and Tabor before Bowden's death a couple of years ago; the final simplified version has been done by his successor at the Cavendish.

The elementary analysis of friction goes back to Leonardo da Vinci; we all know it from our first steps into physics. The force that arises along the surface of contact between any two solids depends on the load across the surface and on the nature of the materials but not on the area of the surface. Even more, in many cases friction is proportional to load; we are given the coefficient of friction to describe that "law."

For a wet tire skidding on a wet road the friction force is one-fifth the load; for a dry tire and road the friction equals the load if the wheels are locked. (The rolling friction is 20 times less.) Like Ohm's law, this simple linear description of what is obviously, at the molecular level, a complex phenomenon is surprisingly accurate for a carefully chosen set of materials. The coefficient remains con-

stant to within a few percent in metal-to-metal contact, even if the load is increased by ten- to a hundredfold. Slide a steel ball over an acrylic plastic surface, though, and Leonardo's empirical approximation is found wanting: the "coefficient" is halved for a load increased sixteenfold. No more is friction proportional to load.

Why? Bowden put it crisply: "Putting two solids together is rather like turning Switzerland upside down and standing it on Austria—the area of intimate contact will be small." Those Alpine profiles are made visible nowadays in three dimensions by the scanning electron microscope, by viewing oblique sections with the optical microscope for vertical exaggeration or electronically by phonograph-pickup techniques. All these methods are exhibited here in photographs that complement the text.

The low-contact area is confirmed by electrical-resistivity measurements, allowing with some difficulty for the layers of oxide and adsorbed gases that dominate "clean" metal surfaces in the air ocean in which we engineering fishes must build. (At high loads the oxide films break down, the contact becomes a good conductor and the coefficient of friction can triple.) The load is borne on "the tips of the surface asperities," and the area of actual contact is tiny. The peaks deform permanently (they feel a high load pressure, of course) until the load is carried by a contact area big enough to support it. That value depends on the material; for mild steel it is some 10 tons per square centimeter.

Slide a heavy toolbox across a steel plate and the load is in fact supported by a contact area of deformed steel no bigger than a comma printed on this page. That area will adjust its size, independent of the size of the bodies, and stay pretty closely proportional to the load; the shape and the number of the contact peaks make little difference. For plastic materials the area adjustments are in fact viscous flows and hence time-dependent; the coefficient therefore varies with speed as well as load.

Friction force between clean metal surfaces is then mainly the force required to shear the friction junctions that form by the cold-welding of heavily loaded peaks. That shear force obeys the old laws of Leonardo, and it is plainly sensitive to grease films and to surface changes that affect the nature of the junctions. The transfer of metal, even under small load and slow motions at room temperature, rather hard to accept at first, is demonstrated by direct autoradiographs. A slider of radioactive lead leaves tiny fragments behind, and these can photograph themselves if an emulsion is left in contact with the track of the slider.

Wear always accompanies friction. Every junction must be sheared, but not every shear transfers material; the junction can fail either at a weld proper or at an oxide-layer film between metal surfaces. For steel on steel perhaps 1 percent of the junctions produce a wear fragment; for polyethylene on steel the "welds" are weak, and only one in 10 million transfers material.

There is a puzzle still. Why do materials not adhere at their friction junctions whenever we put one down on another? The answer is twofold: carry out the experiment with clean, soft metals, such as lead, and they do adhere; this was first described in 1734. At high temperatures the effect is even stronger: metal powders sinter together under pressure. Note, however, that adhesion can only be measured by relaxing the load. The big valleys around the little contact summits are stressed only lightly by the load; they will elastically return to their original forms, to change the shape of the interface and "snap [the weak junctions] apart one by one so that by the time we are ready to measure the adhesion there are practically no junctions left." (The "wringing" of precision gage blocks together, which makes them adhere quite strongly, exploits the adhesion of the grease film present; degreased blocks "wrung" will really adhere, and the expensive steel surfaces will be ruined.)

Bearing metals, ski surfaces, lubricants and their additives, brakes and tires and more topics still are treated in this easily read although imperfectly organized book. Rolling friction turns out to arise from the internal losses that accompany elastic deformation of the surfaces. It is small, and lubrication helps it very little, because few junctions are made and those are merely peeled apart, a process much easier than shearing. Noble metals, such as gold, show a friction coefficient several times higher than metals that form oxide junctions; in a good vacuum heat-cleaned metal surfaces will weld readily. Starting at the contact peaks, the junction grows as the peaks flow and deform; the whole of the surface may become an enormous friction junction. The friction bears little relation to the load; the coefficient may reach 50 or more; the surface is heavily damaged.

In the purity of the vacuum the contact surface is simple, but the laws of friction are more complex. The book is a narrative of the sophistication that commonplace experience today demands of the physicist, and of a victory he has won, which would surely have delighted Leonardo.

NEUROPHARMACOLOGY AND BEHAVIOR, by V. G. Longo. W. H. Freeman and Company ($6.95). HALLUCINOGENS AND SHAMANISM, edited by Michael J. Harner. Oxford University Press ($8.50). THE BOTANY AND CHEMISTRY OF HALLUCINOGENS, by Richard Evans Schultes and Albert Hofmann. With a foreword by Heinrich Klüver. Charles C Thomas, Publisher ($14.75). Strange voyages, real and imaginary, are the substance of these three diverse but interrelated volumes: Longo, a brief clinical text; Harner, an anthropological symposium; Schultes and Hofmann, organized around plant families.

In 1943 Albert Hofmann rode his bicycle home at lunchtime from the Basel laboratory where he had been preparing derivatives of lysergic acid. The familiar short trip seemed endless, complex, frightening. War "classified" his results for several years. Michael Harner, a New School anthropologist, in 1961 drank in the field a tea made from the boiled bark of a liana that grows in the upper Amazon; under the influence of its alkaloids, harmine and dimethyltryptamine (DMT) among others, he "met bird-headed people [and] enlisted the services of other spirit helpers . . . to fly through the far reaches of the Galaxy." Richard Schultes, economic botanist at Harvard, spent 13 happy years along the Amazon, collecting 24,000 plants, several languages and the habit of daily chewing coca. The psychologist Klüver, whose little 1928 book on mescal explicitly recruited the premedical student Schultes to his later career, sought out all over the U.S. people who had reported eidetic imagery, or "photographic" memory. By chance the noneidetic Klüver read that peyote might help him toward a share of the imagery; he first consumed mescal buttons in the laboratory at the University of Minnesota. Only Longo, a research pharmacologist with an atlas of electroencephalograph records to his credit, has stayed at home (in Rome) to prepare his book. (Although his book is succinct, it treats antipsychotics and tranquilizers as fully as hallucinogens.)

The fact is that among many hundreds of thousands of plant species a few hundred are important to human nutrition, and some 75 yield substances

whose effect is narcotic or hallucinogenic. Among plants used by man for purposes other than food, the opium poppy, the tobacco plant, coca and certain large-flowering tree daturas of the northern Andes are species known only as cultivated plants.

There is clear evidence that religion too lies along this path: the psilocybin of the little Mexican mushrooms is associated with an antique cult of the highland Maya, whose stone icons of the mushroom have been found in widespread Guatemalan digs. One paper associates it with "tender, generous, brotherly feelings." These effects are not particularly noted in laboratory studies, one author dryly remarks. The ecstatic chant of the shaman of northeastern Oaxaca is today inspired by the same psilocybin: "the psychoanalyst listens, whereas the shaman speaks," wrote Claude Lévi-Strauss. Eloquence heals; there, in absolute darkness, is poetic ritual: "Thirteen white lights, says. Thirteen mountains of points, says."

Almost 20 years after LSD, Hofmann tested by self-experiment another alkaloid: d-lysergic acid amide. Structurally close to his old drug, this one was less active by a factor of 50 to 100. It was no novel product of the Sandoz synthesists, however, nor was it derived in any way from ergot fungi. It was there in the seeds of certain Mexican morning glories, seeds described by the early Spanish chroniclers as "held in great veneration," the source of a magic Aztec potion, ololiuqui, used to "communicate with the gods and receive the secret things."

The chief grocery-store hallucinogen is an alkaloid in the spice nutmeg. There is a little evidence for its use in India, both ancient and modern, and it is commonly taken in American prisons. Distortion of time and space, floating and flying sensations, separation of the limbs from the body and occasionally imagery can result from ingestion of a teaspoonful or two of the spice, usually with plenty of unpleasant side effects. Its active principle is uncertain, but the best theories suggest one constituent that the body metabolizes into a compound related to a strictly contemporary folk drug, called STP by its Haight-Ashbury discoverers about 10 years ago. (The name was taken from that of the proprietary oil additive, said to increase the pep and power of automobile engines.) It took a couple of years before Government investigators could buy samples of STP, isolate the drug and test it. It was the synthetic 2-5-methoxy-4-methylamphetamine. The structure of all these substances is broadly related to that of mescaline.

The use of plants in the Americas has always been remarkably varied; we know how much the fields of the Old World gained once the ocean was crossed. But Europe has its hallucinogens too. Harner and Longo (the latter's clinical-textbook style is at this point suddenly illumined with four engravings of witch flight from Goya) both discuss the role of "flying ointment" and various other potions in the ritual and personal experiences of European witches and warlocks. Harner's Jívaro shamans used both *Datura*, yielding scopolamine, like the European plants, and the non-European harmine alkaloids. But the *Datura* is "too strong," they held, for shamanism itself. A shaman must operate socially while under the influence; the *Datura* is apt for the vision quest, the search for supernatural contact, but it induces sleep and muddies the more complex side of ritual behavior. Therefore in Europe there were two distinct phases within the practice of witchcraft: the Esbat, a ritual meeting of adepts without drugs, and the Sabbat, the "orgiastic gathering," which was in fact somnolent and solitary, simply a flight trance under the influence of some *Datura*-rich ointment, with its scopolamine and hyoscyamine.

How does it all happen? We are far from understanding the basis of these remarkable pathologies. Longo's text notes the similarities or antagonisms of the drugs to known transmitter substances in the central nervous system, and considers as well the evidence of electrophysiology. Simple explanations—synchronizing certain systems of brain waves, resemblance to some natural transmitter compound—seem to promise well, but they peter out. The remarkable complex within our head is not yet mapped territory. The more credulous of the 10 anthropologists and fieldworkers who appear in the Harner collection are even tempted by a wondrous anecdote or two to believe that the vine snuff or the mushrooms connect the user with another world, or even with distant earthly lands.

The reader of these related books sees a much more comprehensive unity: the drugs disclose an interior world. Shaman or anthropologist, novelist or organic chemist, witch or neurophysiologist or volunteer, the richness is internal. Let the visual field part and glide, the images turn into color negatives, or bear stripes and edges, and we can see grand snakes and rare tigers, unknown black

priests or dragon-headed rulers; our experience fills our head with imaginative illusions. No vision has yet been recounted more powerful than that of August Kekulé's, who, while daydreaming without drugs on the London horse bus a century ago, saw serpents join head to tail to form the benzene ring.

GRAVITY'S RAINBOW, by Thomas Pynchon. The Viking Press ($15). A V-2 missile links the two ends of the ballistic parabola that is gravity's rainbow. At one end is London, dogged and ingenious; the other end is by extension Germany itself. This bulky, intimate, particular and detached novel, a work of the first magnitude, is a tale of V-2.

There is a protagonist at each root of the parabola: in London, Slothrop, the young New England intelligence lieutenant, with a hidden childhood under psychological experiment; on the Continent, a mad and perverse S.S. officer, Captain Weissmann, commander of a V-2 battery. Slothrop leaves London to spend the year in fugue across shattered nations, in a search for the unique rocket, serial number 00/000, that Weissmann once requisitioned and launched with the S-Gerät, strangest of all the payloads of the Wehrmacht. The prose is intricate, turbulent, streaming without rest in a spate of idea, event and allusion.

Here we merely note that this is a novel of a new sort. James Joyce grounded his flowing work on the ledges of epic, myth and folklore. *Gravity's Rainbow* instead founds its intricacies and expertise of allusion on science and technology, adjusted to the wartime *ambiance*. There are a number of remarkable set-piece essays, one on organic chemistry and its origins, one on operant conditioning, another on rocket dynamics. Ackeret and Leibniz, the Poisson distribution and double integrals are references as natural to Pynchon as limericks and vintages are to other novelists. To sketch one example, an apt connection is made between the serial wood engravings of late medieval German art, the calculus of Leibniz, the successive frames of an Ufa film and the unfolding time trajectory of the rocket, striking unheard a theater in London. Not all readers will persist up the wordstream past the bizarre and explicit sexual couplings, diverse as molecules, into the analytic heart of the novel; those who do will gain a richness of thought and motive, for which they must pay in spiritual coin. The literary reviewers have treated the novel as seriously as it deserves; it is a brilliant book, but be warned, that glow is icy cold.

BIBLIOGRAPHY

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

ELECTROMAGNETIC FLIGHT

THE MAGNEPLANE: GUIDED ELECTROMAGNETIC FLIGHT. Henry H. Kolm and Richard D. Thornton in *Proceedings of the 1972 Applied Superconductivity Conference.* Institute of Electrical and Electronics Engineers, 1972.

PRELIMINARY DESIGN STUDIES OF MAGNETIC SUSPENSIONS FOR HIGH SPEED GROUND TRANSPORTATION. John R. Reitz, Robert H. Borcherts, L. C. Davis, T. K. Hunt, Dennis F. Wilkie and the Ford Motor Company research staff. Report No. FRA-RT-73-27, U.S. Government Printing Office, 1973.

GROUND TRANSPORTATION FOR THE EIGHTIES. Edited by R. D. Thornton, a special issue of *Proceedings of the IEEE,* Vol. 61, No. 5; May, 1973.

SUPER HIGH SPEED GROUND TRANSPORTATION. Takeshi Oku, Yoshihiro Kyotani and Toshio Sanuki, translated by Yukikazu Iwasa. The M.I.T. Press, in press.

HERPES VIRUSES AND CANCER

THE CHEMISTRY AND BIOLOGY OF VIRUSES: PERSPECTIVES IN VIROLOGY. Heinz Fraenkel-Conrat. Academic Press, 1969.

ONCOGENIC VIRUSES. Ludwik Gross. Pergamon Press, 1970.

ONCOGENESIS AND HERPESVIRUSES. Edited by P. M. Biggs, G. de-Thé and L. N. Payne. International Agency for Research on Cancer, World Health Organization, 1972.

THE SLING AS A WEAPON

GLANDES PLUMBEAE LATINE INSCRIPTAE. K. Zangemeister in *Ephemeris Epigraphica,* Vol. 6, pages 1–143; 1885.

EXCAVATIONS AT OLYNTHUS: PART X. David M. Robinson. Johns Hopkins University Press, 1941.

THE SIGNIFICANCE OF THE SLING FOR GREEK PREHISTORY. V. Gordon Childe in *Studies Presented to David Moore Robinson on His Seventieth Birthday: Vol. I,* edited by G. E. Mylonas. Eden Publishing House, 1951.

SCHLEUDER UND BOGEN IN SÜDWESTASIEN. VON DEN FRÜHESTEN BELEGEN BIS ZUM BEGINN DER HISTORISCHEN STADTSTAATEN. M. Korfmann in *Antiquitas 3,* Vol. 13, 1972.

PROTEIN SHAPE AND BIOLOGICAL CONTROL

INTRACELLULAR REGULATORY MECHANISMS. H. E. Umbarger in *Science,* Vol. 145, No. 3633, pages 674–679; August 14, 1964.

ON THE NATURE OF ALLOSTERIC TRANSITION: A PLAUSIBLE MODEL. J. Monod, J. Wyman and J. P. Changeux in *Journal of Molecular Biology,* Vol. 12, pages 88–118; 1965.

REGULATION OF ENZYME ACTIVITY. Daniel E. Atkinson in *Annual Review of Biochemistry,* Vol. 35, Part I, pages 85–124; 1966.

THE CATALYTIC AND REGULATORY PROPERTIES OF ENZYMES. D. E. Koshland, Jr., and K. E. Neet in *Annual Review of Biochemistry,* Vol. 37, pages 359–410; 1968.

THE MOLECULAR BASIS FOR ENZYME REGULATION. D. E. Koshland, Jr., in *The Enzymes: Vol. I,* edited by Paul D. Boyer. Academic Press, 1970.

THE SOLAR CORONA

OUR SUN. Donald H. Menzel. Harvard University Press, 1959.

THE FIERY SUN. Jay M. Pasachoff in *Natural History,* Vol. 81, No. 5, pages 48–53; May, 1972.

THE QUIET SUN. Edward G. Gibson. National Aeronautics and Space Administration, 1973.

HIGH-EFFICIENCY PHOTOSYNTHESIS

HYBRIDS BETWEEN *ATRIPLEX* SPECIES WITH AND WITHOUT β-CARBOXYLATION PHOTOSYNTHESIS. Olle Björkman, John E. Boynton, Malcolm A. Nobs and Robert W. Pearcy in *Carnegie Institution of Washington Year Book: Vol. 69.* January, 1971.

PHOTOSYNTHESIS AND PHOTORESPIRATION. Edited by M. D. Hatch, C. Barry Osmond and Ralph O. Slatyer. John Wiley & Sons, Inc., 1971.

PHOTOSYNTHETIC ADAPTATION TO HIGH TEMPERATURES: A FIELD STUDY IN DEATH VALLEY, CALIFORNIA. Olle Björkman, Robert W. Pearcy, A. Tyrone Harrison and Harold A. Mooney in *Science,* Vol. 175, No. 4023, pages 786–789; February 18, 1972.

COMPARATIVE STUDIES ON PHOTOSYNTHESIS IN HIGHER PLANTS. Olle Björkman in *Photophysiology: Vol. VIII,* edited by Arthur C. Giese. Academic Press, in press.

AUDITORY BEATS IN THE BRAIN

MASKING OF TONAL SIGNALS. Lloyd A. Jeffress, Hugh C. Blodgett, Thomas T. Sandel and Charles L. Wood III in *The Journal of the Acoustical Society of America,* Vol. 28, No. 3, pages 416–426; May, 1956.

BINAURAL INTERACTION IN THE SUPERIOR OLIVARY COMPLEX OF THE CAT: AN ANALYSIS OF FIELD POTENTIALS EVOKED BY BINAURAL-BEAT STIMULI. Joel S. Wernick and Arnold Starr in *Journal of Neurophysiology,* Vol. 31, No. 3, pages 428–441; May, 1968.

LIMITS FOR THE DETECTION OF BINAURAL BEATS. David R. Perrott and Michael A. Nelson in *The Journal of the Acoustical Society of America,* Vol. 46, No. 6, Part 2, pages 1477–1481; December, 1969.

ELECTRON-POSITRON COLLISIONS

ELECTRON-POSITRON COLLIDING BEAM EXPERIMENTS. N. Cabibbo and R. Gatto in *The Physical Review,* Vol. 124, second series, No. 5, pages 1577–1595; December 1, 1961.

THE FUNDAMENTAL PARTICLES. Clifford E. Swartz. Addison-Wesley Publishing Co., Inc., 1965.

PARTICLE PHYSICS: THE HIGH-ENERGY FRONTIER. Milton S. Livingston. McGraw-Hill Book Company, 1968.

ELECTRON-POSITRON INTERACTIONS. V. Silvestrini in *Proceedings of the Sixteenth International Conference on High Energy Physics: Vol. 4,* edited by J. D. Jackson and A. Roberts. National Accelerator Laboratory, 1972.

MATHEMATICAL GAMES

GEOMETRICAL ARITHMETIC. Helen A. Merrill in *Mathematical Excursions.* Dover, 1957.

VISION IN ELEMENTARY MATHEMATICS. Walter W. Sawyer. Penguin Books, 1964.

THE AMATEUR SCIENTIST

INTERFERENCE SYSTEMS OF CROSSED DIFFRACTION GRATING: THEORY OF MOIRÉ FRINGES. J. Guild. Oxford University Press, 1956.

ELECTROSTATICS: EXPLORING, CONTROLLING AND USING STATIC ELECTRICITY. A. D. Moore. Doubleday & Company, Inc., 1968.

THE '74 VEGA GT. EVERYBODY LOVES A WINNER.

In three short years Vega has won more automotive honors than any other American-built economy car, and has become one of the best selling cars made.

So what do we do? We change it.

But not without darn good reasons.

We reshaped the front end with the intention of making Vega tougher, not prettier.

With bigger, tougher bumpers. And cooling slots in place of a grille.

Instead of grasshopper legs to support the hatch, you now get telescopic hydropneumatic springless counterbalancing arms for easier opening.

On the lighter side, you also get nicer carpeting, and a bigger selection of available fabrics and colors and vinyl roofs and stripes and that sort of thing.

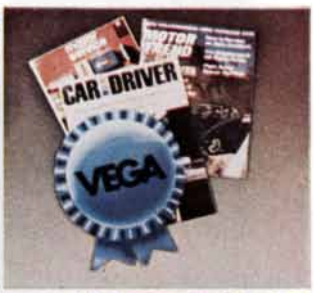
But, more than ever you want economy, of course, and we haven't overlooked that. We've improved the Vega overhead cam engine with modifications in the carburetor linkage, the inlet manifold and so on. And we've even increased the size of the fuel tank so you can go farther between stops.

See your Chevy dealer for a test drive.

See what it's really like.

Vega honors include: *Car & Driver* Readers' Choice, Best Economy Sedan, '71, '72, '73. *Motor Trend,* Car of the Year, '71. Economy Car of the Year, '73 Vega GT. *Motor Service & Service Station Management* mechanic survey, Easiest Subcompact to Service, '72.

SEE WHAT IT'S LIKE TO DRIVE A WINNER.



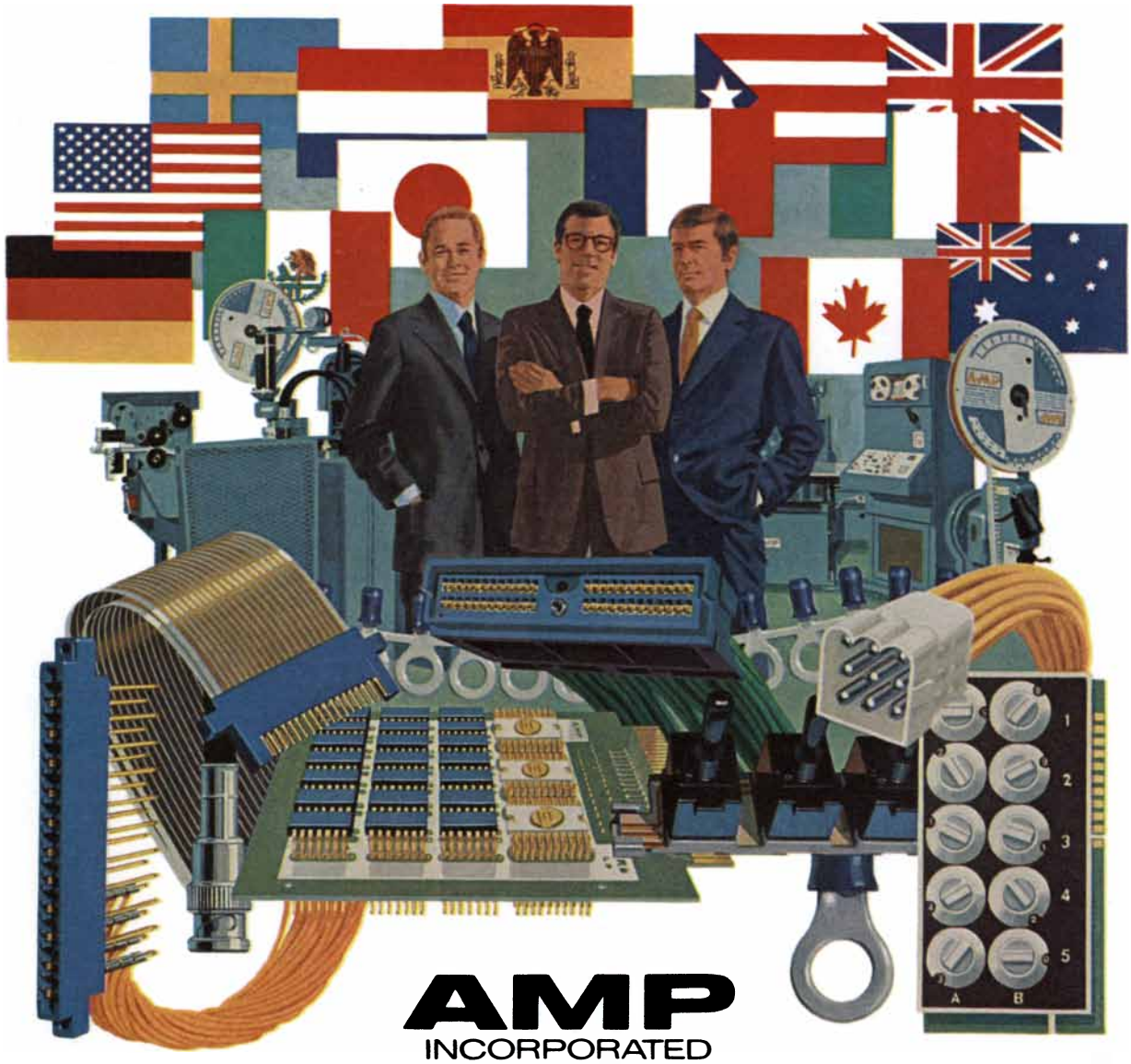
The world has a word for high-quality, electrical connectors. AMP.

It's the international symbol for high reliability and low applied cost... in aircraft, appliances, business machines, computers, home electronics, transportation, communications and electrical utilities.

The name AMP is known for something more than the outstanding performance of the products it represents... it means people, engineering innovation, manufacturing know-how and complete field service backup, worldwide.

Imagine what we can do for you... how our ideas, products and application equipment can help you. Write for our descriptive brochure, "We can help you put a smile on your bottom line."

AMP Incorporated, Harrisburg, Pa. 17105.



AMP
INCORPORATED

Manufacturing and direct sales facilities worldwide: Barcelona, Brussels, Buenos Aires, Frankfurt, London, Mexico City, Paris, Puerto Rico, Sao Paulo, s'Hertogenbosch (Holland), Sydney, Stockholm, Tokyo, Toronto, Turin, Vienna.