SCIENTIFIC AMERICAN

MATHEMATICS OF THE "I CHING"

ONE DOLLAR

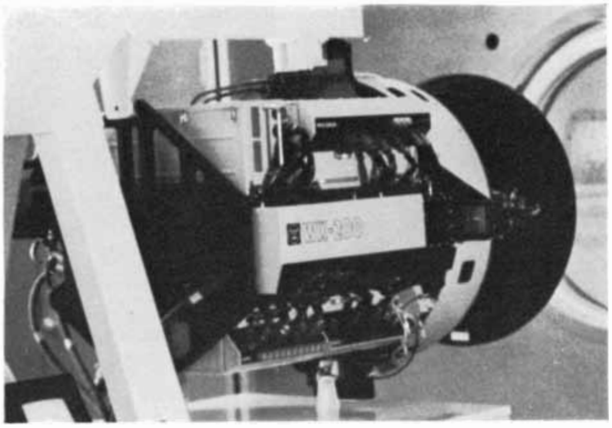
January 1974

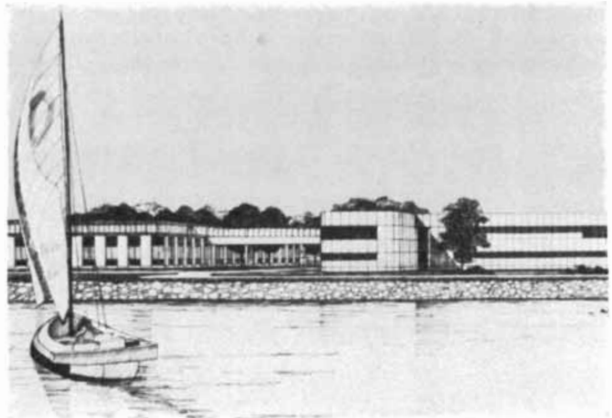
Somebody has to provide the technology so the problems get solved

Problem: **A developing nation needs a modern, national communications system.**

Solution: **Fly a balloon.**

We've just tested our first TCOM (Tethered Communications) station over the Bahamas. TCOM is an aerodynamic all-weather balloon connected to a ground station. It's also a unique communications/television relay system that can be installed in months at a fraction of the cost of coaxial cable installations or a synchronous satellite system. One balloon covers at least 125,000 sq. km. That means Liberia or Honduras.
Two would serve Ecuador or Senegal. In addition to radio and telephone service, TCOM can provide color TV, telephoto, and digital data transmission channels. Maybe better communications will clear up some of the world's misunderstandings. We've bet a lot of our own money on it.





Problem: **The high cost of avionics.**

Solution: **Low-cost WX system.**

Westinghouse has developed a family of modular airborne radars that can reduce total cost (both purchase and maintenance) by as much as four to one. For example, initial cost of our centerline design (WX 200) is $200K instead of the $600K of equivalent current models, plus it will deliver 100 hours MTBF (mean time between failures). It weighs 350 pounds instead of 700 and will fit most military aircraft instead of just one. So far, there's a substantial amount of our own money in this designed-to-cost product.

Problem: **The environmental problem.**

Solution: **Apply industrial know-how.**

Should Americans apply the proved, multi-disciplinary research methods of industry to the environment? We think so, and have announced plans for a multimillion-dollar Environmental Center that will be next door to our Ocean Research Center on Chesapeake Bay. The combined operation will comprise an impressive collection of scientists and engineers zeroing in full time on everything from agriculture to oil production, from water and sewage to power generation and the fish industry. Maybe not tomorrow, but there are solutions.

You can be sure if it's Westinghouse.
For our new brochure, write
W. R. Smanko, Marketing Manager,
Defense & Public Systems,
Westinghouse Electric Corporation,
Pittsburgh, Pa. 15222.

Westinghouse helps make it happen

Medical history is replete with controversy, so that it is refreshing to find Dr. Alexander R. Lucas reviewing with intelligence and detachment this book by the internationally known medical doctor, *Dr. David R. Hawkins* ***and that distinguished*** *Laureate Professor Linus Pauling* ***who has twice won the*** *Nobel Prize.* ***It is especially useful that Dr. Lucas emphasizes that Dr. Hawkins has used his treatment within a clearly defined medical setting, in which the patient is seen as a potentially responsible human being assailed by a tragic, but treatable illness.***

This is timely; for when a novel or unfamiliar approach to old and dreaded illnesses is being weighed by medicine, those who explore both its weaknesses and strengths and discuss them coolly do a valuable service to all of us. We know enough about medical history now to recognize that disagreement, sometimes of the most vigorous kind, has little or nothing to do with the merits or demerits of a particular theory or treatment. The protagonists in this kind of medical drama have, as Thomas S. Kuhn of Princeton has noted of science generally, no alternative but to assume adversary postures, for such are the rules of the game.

It would thus be a disservice to the tens of thousands of patients for doctors to dismiss this distinguished encyclopedic contribution without making their own assessment of its possible value in their practice. *(Orthomolecular Psychiatry) Pub. by W. H. Freeman Co., San Francisco, Cal.*

A. N. Spanel FOUNDER: INTERNATIONAL PLAYTEX CORPORATION

An Evaluation in The MAYO CLINIC *Proceedings*

by Alexander R. Lucas, M.D.

In "Science," during 1968, Pauling published a paper entitled "Orthomolecular Psychiatry." This paper engendered considerable interest among biologically oriented psychiatrists. It has been reprinted as the initial chapter in this book, ***which integrates the theoretical concepts promulgated by Pauling with the megavitamin treatment of schizophrenia initiated by Hoffer and Osmond in 1952.*** These concepts have caused much skeptical and antagonistic reaction among traditional medical practitioners, while stimulating publicity and an ever-increasing enthusiasm among lay groups. Some physicians have even relegated this treatment to the realm of quackery. ***This book not only deserves but demands to be read by anyone who in the future proposes to pass judgment on the concepts expounded.***

The term "orthomolecular psychiatry" was coined by Pauling to designate a form of treatment that will provide the brain with the optimum concentrations of substances normally present in the human body—in short, the right molecules in the right amount. ***This concept, which may at first seem foreign to psychiatrists dealing with interpersonal processes, is a fundamental one — a desirable state, the attainment of which no one can dispute. Underlying all mental processes and behavioral actions are neurophysiologic events determined by the chemical environment of the brain. That this concept has practical validity for psychiatry has already been proven in such conditions as pellagra and phenylketonuria, where it has been established that appropriate concentrations of vitamin B^3 and phenylalanine are necessary to prevent the most profound mental changes. Pauling has crossed interdisciplinary boundaries in applying the methods and discoveries of molecular biochemistry to clinical psychiatry; specialists should heed the ideas of brilliant thinkers from other fields, which often can illuminate new vistas in fresh perspective.***

The book includes chapters by more than two dozen contributors. It is organized into four sections dealing with theoretical concepts, clinical diagnosis of schizophrenia, treatment, and practical clinical approaches. ***The last section, written by Hawkins, is an admirably integrated overview of theory, diagnosis, and treatment based on the approaches used at the North Nassau Mental Health Center in Manhasset, New York. Hawkins' concepts, far from being fadism or quackery, seem scientifically based and truly eclectic, using not only megavitamins but also what is best of the traditional treatments. The idea promulgated by Hawkins and Pauling of obtaining a chemical profile of each patient is a rational approach to treatment, and one that I believe will come more and more into general use.***

My chief criticisms have to do with the imprecise and inconsistent use of the term schizophrenia throughout the book, and with some unwarranted statements about results of treatment. In places, "schizophrenia" is used too generally to imply severe psychotic illness, while in others it refers to relatively mild conditions involving sensory misperception. Megavitamin treatment seems to be advocated indiscriminately for all of these conditions and for others, such as alcoholism. Validity of diagnosis thus becomes difficult to evaluate, as do treatment results. Mild conditions and acute psychotic episodes often remit spontaneously regardless of treatment, or despite treatment.

In this book the reader will find the best and most comprehensive compilation on the subject. As an unexpected bonus there are authoritative reviews of the current status of biochemical and genetic knowledge of schizophrenia. The book should be read, the ideas critically assessed, and the treatment recommendations put to scientific scrutiny rather than being discarded offhand. I suspect it will maintain a permanent place in the long history of the search for a cure of mental illness.

Presented as a Public Service by A. N. SPANEL, *888 Seventh Ave., New York, N.Y. 10019*

"YOU AND EVERY MAN IN BUSINESS ARE TRUSTEES OF THIS NATION AND THE WORLD"®

Established 1845 **SCIENTIFIC AMERICAN** January 1974 Volume 230 Number 1

ARTICLES

20 ENERGY POLICY IN THE U.S., by David J. Rose
Recent steps to alleviate energy shortages leave the long-range problem unsolved.

30 ELECTROCHEMICAL MACHINING, by James P. Hoare and Mitchell A. LaBoda
Hard alloys can be cut into complex shapes in an electrolytic cell.

38 THE NERVOUS SYSTEM OF THE LEECH, by John G. Nicholls and David Van Essen
A repetitive array of small ganglia is used to study reflexes.

54 THE CELL CYCLE, by Daniel Mazia
What happens in the living cell between divisions? Four stages are now perceived.

69 THE AGE OF THE ELEMENTS, by David N. Schramm
The analysis of radioactive decay indicates when the oldest atoms were formed.

78 THE PERCEPTION OF DISORIENTED FIGURES, by Irvin Rock
Why, for example, does a square that is rotated 45 degrees look like a diamond?

87 THE PHYSICS OF THE BOWED STRING, by John C. Schelleng
The interaction of the string and a bow that sticks and then slips is investigated.

96 THE ASTROLABE, by J D. North
This instrument of the Middle Ages was used for observation and computation.

DEPARTMENTS

8 LETTERS

12 50 AND 100 YEARS AGO

16 THE AUTHORS

50 SCIENCE AND THE CITIZEN

108 MATHEMATICAL GAMES

114 THE AMATEUR SCIENTIST

123 BOOKS

130 BIBLIOGRAPHY

THE COVER

The design on the cover is based on a Chinese charm that displays the eight trigrams of the *I Ching* surrounding the *t'ai chi t'u,* the ancient Chinese symbol of yin and yang (see "Mathematical Games," page 108). This traditional arrangement, attributed to Fu Hsi, the legendary founder of China's first dynasty, places the trigrams so that opposite pairs are complementary both in a mathematical sense and in their symbolic meanings. The Ch'ien trigram at the top stands for heaven and father. Opposite it at the bottom is K'un, which symbolizes earth and mother. Similarly Tui (lake, youngest daughter) opposes Kên (mountain, youngest son), Sun (wind, eldest daughter) opposes Chên (thunder, eldest son), and Li (sun, middle daughter) opposes K'an (moon, middle son). If the trigrams are interpreted as binary numerals, the four trigrams starting at the top and moving down the left represent in sequence the numbers 0, 1, 2, 3. The four starting at the bottom and moving up the right represent in sequence the numbers 7, 6, 5, 4.

THE ILLUSTRATIONS

Cover illustration by Ilil Arbel

Page	Source
21–29	Allen Beechel
30	Dan Todd
31	Ben Rose
32–33	Dan Todd
34	Ben Rose
35–36	Dan Todd
37	James P. Hoare and Mitchell A. LaBoda
39	Tom Prentiss
40	John G. Nicholls and David Van Essen (*top left*), Tom Prentiss (*top right and bottom*)
41	John G. Nicholls and David Van Essen (*top*), Tom Prentiss (*bottom*)
42–48	Tom Prentiss
54	Keith R. Porter, David M. Prescott and Jearl F. Frye, University of Colorado
56–57	Enid Kotschnig
58–59	Keith R. Porter, David M. Prescott and Jearl F. Frye, University of Colorado
60	David R. Wolstenholme, University of Utah
62–63	Potu N. Rao and Robert T. Johnson, University of Colorado Medical Center
69	Hale Observatories
70–77	Gabor Kiss
79	Wide World Photos
80–85	Jerome Kuhl
87–90	Alan D. Iselin
91	Alan D. Iselin (*top*), John C. Schelleng (*bottom*)
92	Alan D. Iselin
93	John C. Schelleng (*top*), Alan D. Iselin (*bottom*)
94	John C. Schelleng
95	Ian M. Firth and J. Michael Buchanan, University of St. Andrews (*top*); John C. Schelleng (*bottom*)
96–97	George V. Kelvin
98	Milford Small
99–105	George V. Kelvin
109–113	Ilil Arbel
114–115	Eugene F. Ruperto
116–120	Jerome Kuhl

Qtam? Sno

Doesn't anybody in the computer business speak English?

The computer business has created its own language.

"Computerese."

It helps computer people communicate with each other about computers. But it doesn't help them communicate with other people.

We want to make computers easier for you to understand.

Because the difference between the right and wrong way of using them can mean a lot to the profitability of your company.

So we're going to dispense with computerese and talk a language businessmen can understand.

LETTERS

Sirs:

It is a delight to read Leon D. Harmon's article ["The Recognition of Faces," SCIENTIFIC AMERICAN, November, 1973], which inverts our own experience. For the feature film *Westworld* we fed motion-picture images into a computer scanning-playback device to produce block images with reduced information. We were faced with precisely the same problems of perception Dr. Harmon set out to study. In our case we needed the audience to read the images within the allotted screen time.

We could control block size, color and contrast, but we had to allow for an enormous range of audience distance from the screen. The final film contains 14 separate computer sequences totaling more than two and a half minutes. In many cases our decisions were based on considerations specific to moving images. For example, we felt safer with larger block size when there was cross-screen action, since any lateral movement was much more readable.

The technical work was done at Information International, Inc., in Los Angeles, under the supervision of John Whitney, Jr. Technical details are discussed in the November 1973 issue of *The American Cinematographer.*

MICHAEL CRICHTON

Culver City, Calif.

Sirs:

In the otherwise excellent article "High-Efficiency Photosynthesis," by Olle Björkman and Joseph Berry [SCIENTIFIC AMERICAN, October, 1973], two seriously misleading points are made in their "punch line" to support the thesis that C_4 plants should be selected for arid-lands agricultural productivity where heat and plant water stress are common: (1) They state that agricultural crops have been selected and bred for highest yield under optimal conditions. (2) They also state that maximizing absolute growth rate results in decreased water-use efficiency.

These two claims lead the reader to believe they are common generalizations. The opposite is more likely true. That is, most crops are bred and selected under the stresses encountered for a specified management in a given environment. Most often the most efficient use of water for maximizing crop yield results from supplying adequate water, so that there is no deficiency.

We would all agree that we should manage our lands and the environment as wisely as possible, but one can argue strongly that with increasing constraints on energy use, water use and environmental pollution the wisest use of our resources for growing food and fiber will be under the best possible management and environment, with the least acreage requirement.

E. R. LEMON

Cornell University
Ithaca, N.Y.

Sirs:

In his last paragraph Professor Lemon raises a fundamental philosophical point. What is the wisest allocation of our limited resources to agriculture? He argues for increasing the intensity of land use. We agree that the best use should be made of the land we now cultivate, but we also feel that increasing population will force us to increase the land area cultivated. This seems to require cultivation of land in the future that is now for one reason or another marginal or unsuitable for cultivation.

Our comments were made with the assumption that increasing the water-use efficiency of a crop plant could extend its profitable cultivation into regions where the quantity or cost of available water makes cultivation marginal with current varieties. It seems to us that varieties intended for such marginal conditions should be selected by criteria different from those used to select varieties intended for intensive cultivation. This is the intent of the first point to which Professor Lemon takes exception. We did not mean to belittle current plant-breeding objectives and approaches.

With respect to the second point, our intention was to suggest that because there is an antagonism between maximum photosynthetic rate and maximum water-use efficiency for any given plant, selection for either may exclude the other. We did not intend to imply that supplying a given plant with inadequate water would lead to greater water-use efficiency for growth. Once the supply of water is depleted any plant will experience water stress and growth will cease, perhaps resulting in crop failure. If and when such water stress develops depends on the amount of water originally available and the rate at which the crop uses it. Plants do differ in the rate and more important in the efficiency with which they utilize water. All else being equal, a crop plant with high water-use efficiency would obviously yield more growth for a given limited quantity of water than a plant with low water-use efficiency.

OLLE BJÖRKMAN

JOSEPH A. BERRY

Department of Plant Biology
Carnegie Institution of Washington
Stanford, Calif.

ADDENDUM

F. W. Goro's color photomicrographs on page 30 of the November issue, showing optical fibers in cross section, were made with a transmitted-light Mach-Zehnder interference microscope in collaboration with the Laboratory for Applied Microscopy of E. Leitz, Inc.

Scientific American, January, 1974; Vol. 230, No. 1. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer; Arlene Wright, assistant treasurer.

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

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

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

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

☐ Please enter change of address.

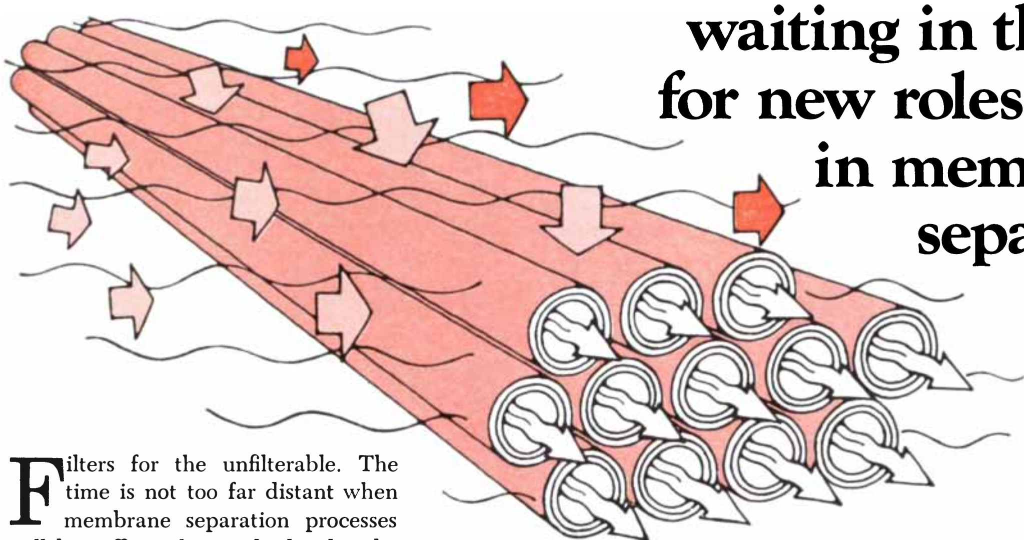
☐ Please enter subscription.
U.S. and Canada:
☐ 1 year, $10 ☐ 2 years, $18 ☐ 3 years, $25
All other countries:
☐ 1 year, $12.50 ☐ 2 years, $23 ☐ 3 years, $32

NAME

NEW ADDRESS

OLD ADDRESS

Hollow fibers: waiting in the wings for new roles in membrane separations.



Filters for the unfilterable. The time is not too far distant when membrane separation processes will be sufficiently standardized to be plugged in as a "unit operation" in computer simulations of processes. And hollow fibers may make it so.

According to Monsanto researchers' over-view – hollow fibers can do most everything flat membranes can do, but much better. The reasons are obvious indeed: the tubular geometry of hollow fibers, *ipso facto,* does away with the need for flat membrane support. The mini-tubes can withstand high pressures without collapse. (Monsanto's high pressure fibers for seawater desalting, for example, can take 1,800 psi.) And hollow fibers afford an exponentially greater membrane area for various kinds of permeations than is feasible with flat membranes. With the ratio of active membrane surface much, much higher per unit volume of pressure vessel, various kinds of super-filters can be more practically-sized. And last, the inherent efficiency of a continuous fiber spinning process generates "membranes" at a lower cost per square foot of active area. This means the most critical component of the super-filter costs less, promising a lower capital cost for the next generation of separation devices.

Sadly, however, membranes in mini-tube form have set up some hurdles of their own in the path of commercial development. The cylindrical form and the new nature of spun membrane walls have knocked out a considerable chunk of solid physical chemistry based on flat membranes. Some highly specialized kinds of separations (such as petroleum fractions through non-polar polyethylene membranes) may have to enter a brand new ball game when trialed through the same polymer in hollow fiber form. *IF* such specialized polymer membranes *can* be produced in hollow fiber form.

In short, hollow fibers are here, but only a still-limited number of types. Except for reverse osmosis applications, there are very few indeed "off-the-shelf" hollow fibers for uses such as gas diffusion separations, specific dialyses, and ultra filtrations. The reason could be *communications.* The researchers who know a lot about hollow fibers, how to make them and condition them to behave as needed are not the same researchers who are seeking better ways to separate xylene isomers, antigens, or amino acids. Such getting together might be the Square One start to enable hollow fiber separations to burgeon out and become as standardized as distillation columns. The field of applications is vast and the physical chemistry art of mini-tube separations is only well-begun.

Reduced to essentials, the variables for membrane separations that investigators must deal with are finite: molecular size of filtrate, media of exchange, transport driving force, morphology of the membrane wall (permeation, permeability, selectivity, steric & polar factors) – and the wall's mechanical and chemical characteristics. The last two variables are the provenance of the hollow fiber builder, but the specific job of separation establishes the parameters for the other variables. Not even a neophyte in the field will dispute the vast differences in such problems as diffusion separation of a rare atmospheric gas from air, stripping uric acid from a blood stream, separating hydrocarbon fractions, or straining out of solutions a specific antigen, hormone, enzyme or antibiotic. But when all the homework is done, the engineer *can* design the separation module.

Define the Job

Monsanto researchers are convinced that much can be done – when "needs" are defined, when performance parameters are set. Experience to date with reverse osmosis applications supports the belief.

The Office Of Saline Water retained Monsanto researchers to develop a high performance asymmetric hollow fiber system. That was in 1968. The problem was not easy, but with enough cut and try Monsanto developed a unique dry jet-wet spinning process. It produces a cellulose acetate hollow fiber with a readily permeable solute-rejecting outer skin and a highly porous inner wall. By subsequent annealing, the membrane wall can be radically "tailored" for handling various concentration levels of feed, to accomodate a wide range of liquid flux rates, and to withstand pressures up to 1000 psi. And, essential to the purpose of water conditioning and waste water recovery – the membrane wall can be made differentially-selective among mono-, di-, and trivalent ions in their various salt forms. The art of "conditioning" the fiber extends to optimizing the flux/selectivity trade-off – nature's rebellion against reversing osmotic flow.

The membrane separation "need" for reverse osmosis de-salting could hardly have been defined more clearly or exactly. And the rapid progress in desalination technology is an example of how "answers" follow needs . . . specifically defined and made known. Today, some fifty Monsanto-built units have been placed in field operation with various feeds. These range in size from 50 to 20,000 gal./day. Three nominally 100 gal./day OSW field test modules at Webster, S.D., have been in continuous operation for 40 months, delivering desalted water from 95-98% rejection of the mixed brine feed, with minimal maintenance.

Once in the swim (staff members prefer, throes) of a membrane separation problem, the way is opened for gleaning much valuable ancilliary information. As a basis for designing modules, Monsanto researchers made detailed analyses of hydraulic flow and stress distribution in hollow fibers. The results proved invaluable for engineering reverse osmosis modules for particular applications. In assessing pressure capability, the three causes for on-stream mechanical failure of hollow fibers were isolated as: elastic failure, catastrophic plastic collapse, and progressive yield or compaction. The critical mode for each was found to be a relationship between inside and outside diameter and the mechanical properties (elastic modulus, yield stress) of the membrane wall. As a result, Monsanto's manipulation of the cellulose acetate polymer and subsequent conditioning can give the fibers (and the modules) a valuable design tool: predictable performance.

Design the Unit

In membrane separations, engineering the modules must be the penultimate step. In the course of Monsanto's reverse osmosis studies for water purification, for example, it became apparent that modules could be substantially "optimized" for particular performance requirements by the "architecture" of the modules. Beyond porosity, selectivity, and mechanical strength of the hollow fibers – operating pressure, temperature, ion mix of the feed, packing density, and arrangement of the fiber bundle, all must be correlated. In short, "what the unit is supposed to do with what" sets up the final engineering challenge.

Monsanto has investigated and tested four different packing designs: double seal, parallel straight packing; single seal, looped; double-seal, radial flow; and woven fabric in a roll pack. The last proved to be a breakthrough in treatment of secondary waste water; reasonably spaced detergent cleaning greatly reduces fouling from oily slime and elusive organic contaminants that build up to impede the flux.

At this point in time for the purpose of water purification, Monsanto has developed three genera of hollow fibers:

Low pressure fibers that will make separations at 100 psi pressure. These have a high specificity for calcium and magnesium chlorides, sulfates, phosphates. They were developed for modules to purify low salinity hard feeds without the need for in-line auxiliary pumps.

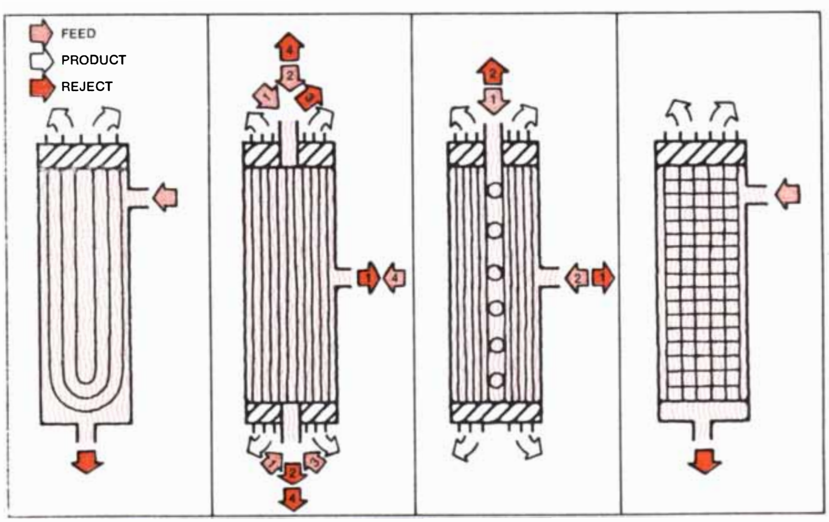
Medium pressure fibers, operable at 250-350 psi, for desalting brackish ground waters and mixed brine feeds.

High pressure fibers, operable up to 1800 psi, i.e. required for one-pass desalting of sea water.

Quite a bundle of knowledge about aqueous separations has been accumulated.

Beyond water purification, the know-how has begun to take root as a means of nickel and silver recovery in the metal plating industry and in pollution abatement measures. Yet the tap root membranes in fiber form did not exist five years ago.

Hollow fiber membrane separations – for isolating, concentrating, purifying, stripping or fractionating – are in early gestation. Sophisticated processing requirements, specialized needs, and high energy costs will foster new applications and stimulate growth. Water purification is well launched. The unique values of membrane separations have been demonstrated by successful stripping away excess citric acid from processed orange juice, by separating protein from whey, and in sophisticated dialyses in biochemistry and medicine. But the road ahead is longer by far.



Schematic of basic hollow fiber module designs: Feed (pink), product (white), and reject (red) flows are indicated with optional combinations indicated by numbers.

Monsanto welcomes your interest in feasibility studies relating to membrane separations. For further information, request a copy of: Hollow Fibers – Characteristics, Uses, Potential.

Address:
Dr. R. G. Schwartz
Monsanto Research Corporation
Station B, Box 8
Dayton, Ohio 45407
Tel.: (513) 268-3411

Monsanto
the science company.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JANUARY, 1924: "Pollution of coastal waters by petroleum oils can eventually be eliminated through the cooperation of the parties concerned, according to preliminary findings of a committee of representatives of the Bureau of Mines, the American Petroleum Institute and the American Steamship Owners' Association. The committee found pollution by oil to be present in some degree at most of the 35 places visited on the Atlantic and Gulf coasts. In the absence of oil-water separating apparatus aboard ship the use of barges for collecting oil-contaminated water and oily refuse offers the most immediately available means for coping with the situation after the vessel has arrived in port."

"In order to subject the moving-platform method of transportation to a full-size operative test Mr. H. S. Putnam has recently constructed a complete platform in Jersey City, N.J. The distinguishing characteristic of the electric system is the method used to propel the car. Instead of mounting a complete motor on the car and transmitting mechanical power therefrom to the wheels, the elements of a three-phase induction motor are separated, the part corresponding to the stationary element of the motor being straightened out and placed at intervals in the roadbed between the rails, whereas the part corresponding to the revolving element is spread out and mounted on the bottom of the car. The speed of the moving platform approximates nine miles per hour."

"The new 'balloon' tire that has just appeared on the market permits the use of pressures something less than half of those in common use today. It is made up specially with thin side walls that are not injured by the 1,000 flexions per mile of travel that the side walls receive. The work of perfecting the balloon tire was started in 1916, but the World War prevented much being done with the idea until 1920. Success came in 1922, and the new tires are now on the market. The sensation produced by riding on a car equipped with the new tires is that of floating along almost wholly in a horizontal plane. It is 'riding on mush.' "

"The English Channel tunnel project is kept alive by its promoters, although the British government persists in refusing to grant the necessary authority. The material to be encountered for the entire distance is very favorable, being a deep bed of chalk infiltrated with clay. With the boring machine designed for this work a heading 12 feet in diameter can be driven at the rate of 120 feet per day, and two machines started at opposite ends should meet in less than three years. It is proposed to complete this pilot tunnel and then to enlarge it to full section, so that the time for the completion of the concrete-lined tunnel is estimated at 4½ years. With present prices the cost is estimated at $145,000,000."

SCIENTIFIC AMERICAN

JANUARY, 1874: "Comet IV of 1874, discovered on the 23d of August by M. Prosper Henry at the Paris Observatory, presents some remarkable peculiarities. Its rapid changes of form, sudden elongation of tail and its brilliancy, which became so great as to render it visible to the naked eye for some time prior to its passage to its perihelion, are considered to be phenomena that may throw light on our hitherto indefinite knowledge regarding the constitution of comets. On the day of its discovery the body appeared in the telescope as a spherical nebulous mass, strongly condensed at the center and exhibiting no traces of a tail. There was little change until August 26, when a tail began to appear. On September 2 the tail had grown to two degrees and continued elongating. The nucleus remained nearly constant in size, although its brilliancy augmented until, on September 10, it became comparable to a star of the fourth magnitude. The spectrum of the comet is found to be composed of three brilliant and very distinct bands. The first is in the yellow portion, the second in the green and the third in the blue. There was no trace of a continuous spectrum in the intervals between these lines."

"Hitherto almost the only guide for interpreting the hieroglyphics with which the monuments of Egypt are covered has been the Rosetta stone, brought to England by the British army after the expedition of 1801 and now in the British Museum. But this is in every way inferior to the stone of Canopus. Half of the lines it contains are incomplete and of the remaining lines many are defaced or illegible, whereas the stone of Canopus is almost as perfect as on the day it left the sculptor's hand. It was accidentally discovered about five years ago when a wall of a temple fell and exposed the corner of a stone covered with Greek characters. In this state the inscription remained for some time; at length its value was perceived and it was removed to the Viceroy's museum. It is about seven feet high and bears three inscriptions, in hieroglyphic, in Greek and in the Hieratic (or Egyptian) characters. It is a copy of a decree made in the ninth year of the reign of Ptolemy III by the priests of Egypt. After setting forth the merits of their rulers and proclaiming extraordinary honors to be offered to them, the priests ordain that a general public festival shall be held for five days every year, and that every fourth year one additional day should be kept as a public festival in honor of the rulers. By these provisions the priests introduced the year of 365¼ days."

"Some time ago Struve discovered a small star near Procyon, which he regarded as being the probable cause of the irregularities of the movements of the latter body. Dr. Andrews has since repeated his calculations regarding the proper motion of Procyon, which appears to be circular, in a period of a little less than 40 years, around some invisible center. He does not now definitely conclude that this peculiar movement should be ascribed to Struve's star, but considers that the question will be decided next spring, if the new star is then visible. In such case Struve's star should be at a considerable distance from the common center of gravity, and a mass must be attributed to Procyon equal to 80 times that of our sun and to his companion a mass equal to 6⅔ of the same body."

"Gas engines are a class of prime movers from which much has been expected, but with which little success has been yet attained. The Otto and Langen engine, recently introduced abroad, acts much like a Cornish pumping engine. The explosion of the gas drives the heavily loaded piston rapidly to the top of the cylinder, and as it descends its weight exerts a useful power. It is economical, using less than half the gas required by the earlier engines, but it is more rattling and irregular in its action than even they were."

AMPEX has coined a new word of interest to people who use scientific data recorders

op•er•a•bil•i•ty—characteristic of a data recorder that lets you concentrate on the data and not the recorder; lets you get on with the business of capturing heart sounds or earth tremors, recording airport noise or smog level, or whatever, instead of worrying about the tape machine.

op•er•a•bil•i•ty—feature of a portable instrumentation machine which has all the important things you need to get at located on the front panel, including FM calibration, signal electronics modules, adjustments, and all operating controls.

op•er•a•bil•i•ty—trait of a scientific data recorder which has the lowest flutter and skew characteristics of any portable data recorder available.

op•er•a•bil•i•ty—the capability of a data recorder which enables it to function in the field as well as in the laboratory, without requiring special adapters and accessories every time you move it to a new location.

op•er•a•bil•i•ty—attribute of a tape machine which handles up to 32 separate data inputs, at speeds from 15/16 to 60 inches per second in either direction; covering IRIG bandwidths up to 300 kHz direct, 40 kHz FM; running on any kind of AC from 90 to 250 volts, from 47 to 420 Hz.

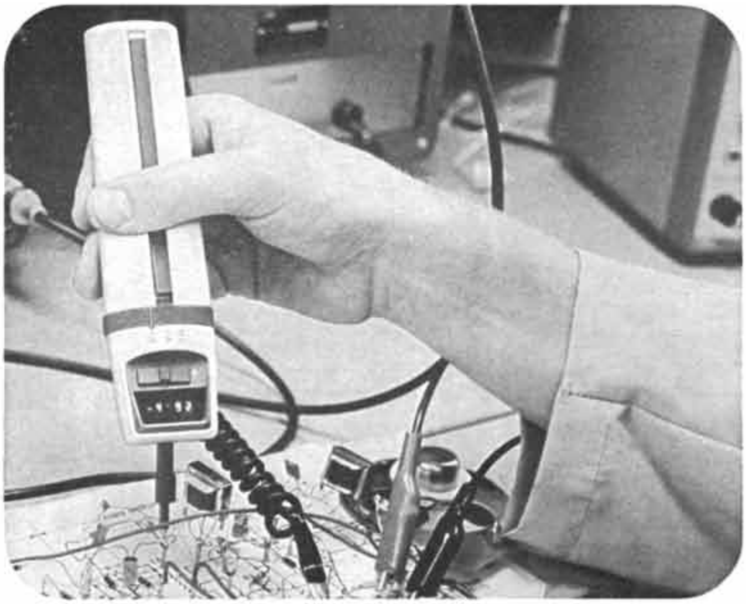
op•er•a•bil•i•ty—a word coined by Ampex engineers who designed Skylab's tape recorders, to describe their newest: the neatest, cleanest little package of mechanical and electronic innovations Ampex has ever put together.

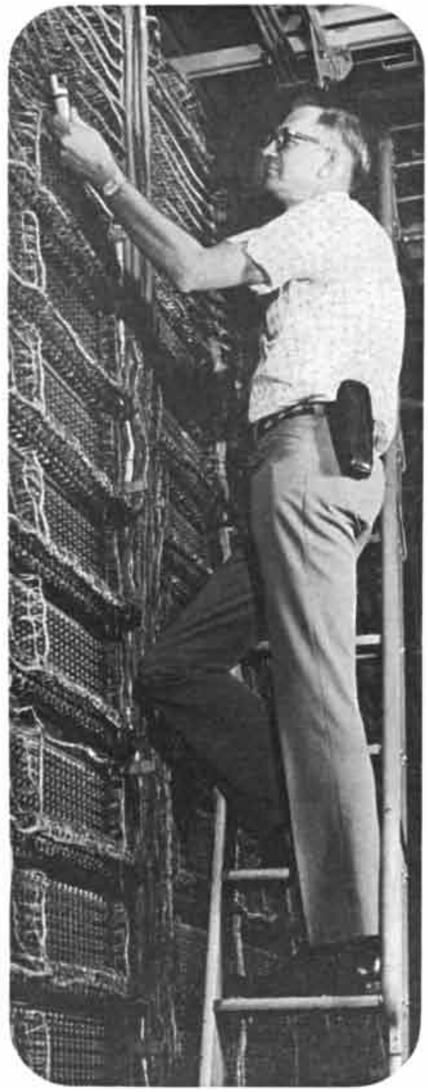
To further define the word *operability*, we've recorded eight pages of data into a new brochure entitled "Ampex PR-2200 Portable Instrumentation Recorder," which should be of special interest to scientists and research engineers. Send your name and address today for your copy.

Ampex PR-2200
portable instrumentation tape recorder

AMPEX

Ampex Corporation
Instrumentation Division
401 Broadway
Redwood City, CA 94063





A small tool with a big impact on field service

To designers and technology buffs, the fascinating aspect of the new HP 970A Digital Multimeter is how we managed to squeeze a complete 3½-digit autoranging DMM into a package that fits the palm of your hand. (The secret, briefly, is a unique thin-film IC that incorporates the equivalent of 3,000 transistors, and combines digital and analog circuitry, on the same hybrid substrate.)

To service technicians and engineers, those harried souls who keep electronic wizardry in good repair, the important news is how the 970A radically improves the measurement of volts and ohms. The battery-powered DMM goes wherever the work is to perform fast and accurate troubleshooting in hard-to-get-at places, and it does this so simply and easily that it's tough to make a measurement error.

To managers responsible for field service, and the customers they serve, the key benefit is the time that can be saved. Since half the cost of field service is labor, and the 970A speeds and simplifies a laborious task, its true value is often realized on invoices for service calls.

Whether a technician is toiling over a dishwasher, a television set, telephone switchgear or the most advanced computer, the 970A works the same way: he selects the desired function, attaches a clip lead to circuit common and touches the test point with the probe tip. A touch of the thumb on the DMM's bar switch,

THE AUTHORS

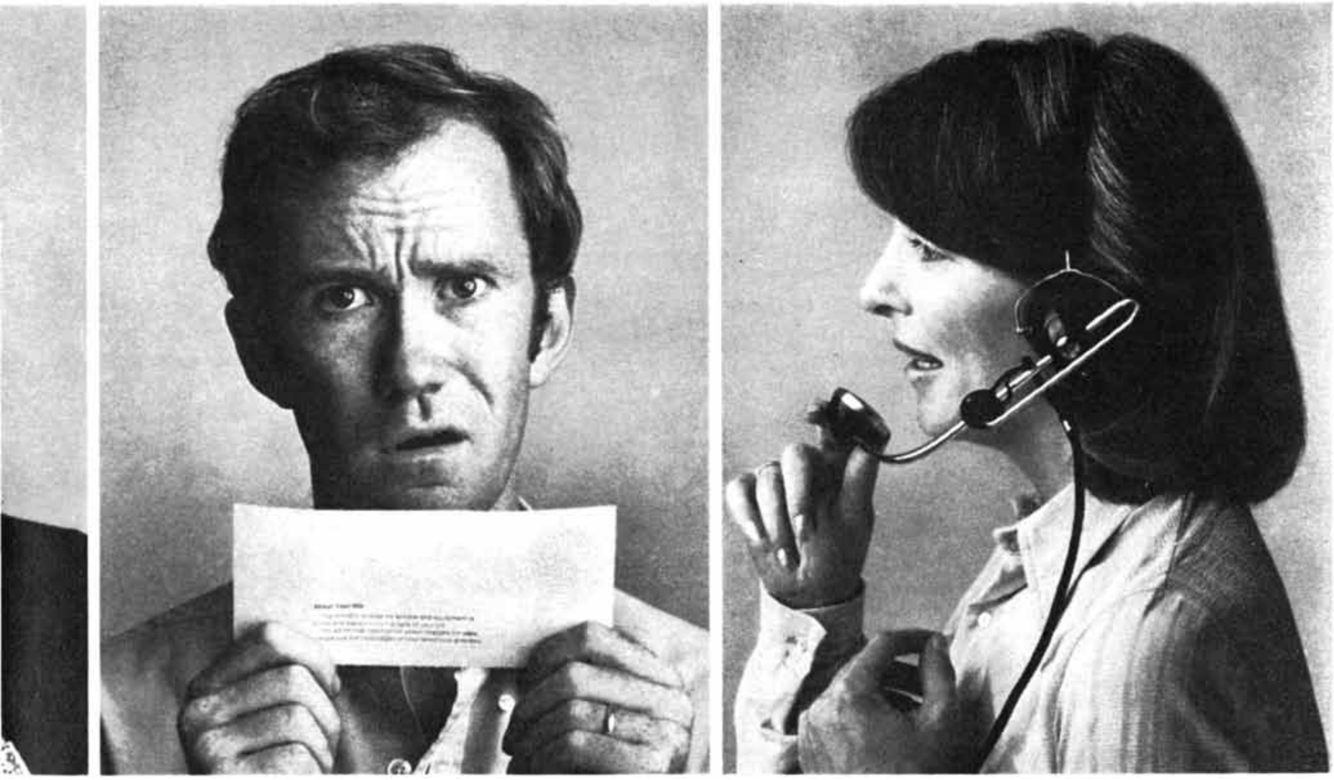
DAVID J. ROSE ("Energy Policy in the U.S.") is professor in the department of nuclear engineering at the Massachusetts Institute of Technology. Born in Canada, he served as an officer in the Royal Canadian Artillery in World War II and then completed his undergraduate work at the University of British Columbia, where he received his bachelor's degree in engineering physics in 1947. After obtaining his Ph.D. (in physics) from M.I.T. in 1950 he spent seven years at Bell Laboratories before joining the M.I.T. faculty. "Believing strongly that rational energy utilization and conservation are much underemphasized," he writes, "I drive an Aston Martin automobile with radial tires, thus (in that somewhat vintage model) getting 25 miles per gallon on long runs. But having also just blown up the engine, I have in addition become sensitive to certain cost/benefit aspects of the problem."

JAMES P. HOARE and MITCHELL A. LaBODA ("Electrochemical Machining") are in the electrochemistry department of the General Motors Corporation Research Laboratories. Hoare, who is a senior research chemist in the department, writes that "from childhood I have spent my life dividing my energies between science and music." He was graduated from Regis College in 1943, receiving his master's degree and his Ph.D. at Catholic University in 1948 and 1949 respectively. He taught chemistry and physics at Trinity College in Washington, D.C., for five years and then spent three years with the Naval Research Laboratories and three years with the Ford Scientific Laboratory before joining General Motors. He describes himself as a "rabid stamp collector" and adds that he also enjoys "playing tennis and bridge, swimming, bowling and dabbling in photography." LaBoda, who is a senior research engineer in the laboratory, was graduated from the University of Detroit with a degree in chemical engineering. He worked as an analytical chemist for the Hudson Motor Car Company and the Ford Motor Company before joining General Motors in 1952. He writes that his spare-time interests are "reading, wine making and bridge."

JOHN G. NICHOLLS and DAVID VAN ESSEN ("The Nervous System of the Leech") are respectively at Stanford University and the University of Oslo. Nicholls writes: "I have a medical degree and a Ph.D. from the University of London. I am now a professor at Stanford, by way of London, Oxford, Harvard, Yale, Harvard. What really interests me professionally is how the brain gets to be put together in the first place, cell by cell, and how its performance is modified by experience. In real life I take every opportunity I can to visit and read about Mexico and Central and South America. The special fascination of these places for me is the thread of continuity in their cultures from pre-Columbian times to the present, set in virgin scenery resembling nothing else on earth." Van Essen received his bachelor's degree from the California Institute of Technology in 1967 and his Ph.D. (in neurobiology) from Harvard University in 1971. He worked with Nicholls for a time and now is doing postdoctoral work in Oslo.

DANIEL MAZIA ("The Cell Cycle") is professor of zoology at the University of California at Berkeley. "I was born in Scranton, Pa.," he writes, "and spent my youth in Philadelphia, enjoying a good education and cultural opportunities in that grimy city, which didn't know it was blighted because urban blight hadn't yet been invented. The most important event in my career was my attachment to my teacher L. V. Heilbrunn at the University of Pennsylvania. It has taken me a long time to see how teachers are the guiding motive of a scientific career. I have had many students, some now famous scientists, and I wonder whether some of them will see their careers in the same way." Mazia received his Ph.D. from the University of Pennsylvania in 1937 and taught at the University of Missouri for 10 years before going to Berkeley in 1951. "I am interested in many things outside of work but am too overawed by the biographies of other *Scientific American* authors to mention them," he says. "I don't play the virginal, ride Lippizaners or practice conversational Erse."

DAVID N. SCHRAMM ("The Age of the Elements") is assistant professor of astronomy and physics at the University of Texas at Austin. Following his graduation from the Massachusetts Institute of Technology in 1967 with a degree in physics, he did graduate work at the California Institute of Technology, where he obtained his Ph.D. in physics in 1971. During his years as a student he also won a number of awards and championships as an amateur wrestler; among them was the National Greco-Roman Championship in 1971. While pursuing



his doctorate and working as a researcher at Cal Tech he served as assistant coach of the wrestling team.

IRVIN ROCK ("The Perception of Disoriented Figures") is professor in the Institute for Cognitive Studies of Rutgers University. He received his bachelor's and master's degrees in psychology at the City College of New York in 1947 and 1948 respectively and his Ph.D., also in psychology, from the New School for Social Research in 1952. He taught at the New School from 1950 to 1959 and then at Yeshiva University until 1967, when he went to Rutgers. Rock's most recent book is *Orientation and Form.*

JOHN C. SCHELLENG ("The Physics of the Bowed String") retired in 1957 as director of radio research at Bell Laboratories. He had been associated with the laboratories and its forerunner, the engineering department of the Western Electric Company, since 1918. Schelleng was graduated from Cornell University in 1915 and was a graduate student and instructor in physics there until 1918. He writes: "I joined the Bell System when the lowest radio frequencies were audible and stayed with it in its climb up the frequency ladder into the domain of gigahertz. After retirement I found myself on a low rung of the ladder with the frequencies of interest in music, but the principles proved to be much the same. As an inveterate amateur cellist I have had a passionate love affair with chamber music almost all my life, with more platonic and less lasting adventures with other hobbies, such as oil painting."

J. D. NORTH ("The Astrolabe") is visiting professor of the history of the exact sciences at the University of Aarhus in Denmark, where he serves on a leave of absence from the University of Oxford. His degrees are from Oxford and the University of London; his doctoral thesis on the history of 20th-century cosmology is also the subject of his book *The Measure of the Universe.* From 1964 to 1971 he prepared for publication an edition of the writing of Richard of Wallingford, a mathematician and astronomer of the 14th century. Richard of Wallingford built for the Abbey of St. Albans the oldest mechanical clock about which detailed information exists; North is directing the reconstruction of this extremely complex mechanical clock on the basis of a manuscript describing it that he discovered in the Bodleian Library at Oxford in 1964.

Established 1845 SCIENTIFIC AMERICAN January 1974 Volume 230 Number 1

Energy Policy in the U.S.

The President's appeal for U.S. energy self-sufficiency by 1980 cannot be regarded as realistic. The long-range options that are open to the nation are here considered in a "taxonomic" approach

by David J. Rose

Contemplating the energy crisis in their chilly homes this winter and facing an economic turndown stemming from fuel shortages, Americans increasingly wonder where it all went wrong. Had no one foreseen the problems that in retrospect receive so many glib explanations, that now require emergency correction because long-term guidance was lacking and that cannot be truly ameliorated for many years? It should have been obvious to the oil companies, the electric utilities, the automobile industry, Congress, the White House and the universities that without adequate energy an industrial society must throttle down.

The problem is large enough, once it is recognized holistically. The getting, refining, distributing and consuming of fuels account directly for about 10 percent of the nation's economic activity, or about $125 billion per year out of a gross national product approaching $1,300 billion. That is almost equal in dollar value to all of agriculture, food processing and food distribution, activities long recognized as requiring intellectual organization and balance, even to having their own department in the Federal Government. It might therefore seem that the development of a rational, long-range energy policy would be the first order of any nation's business. That the U.S. never had such a policy and is still without one can only be regarded as a major social failure.

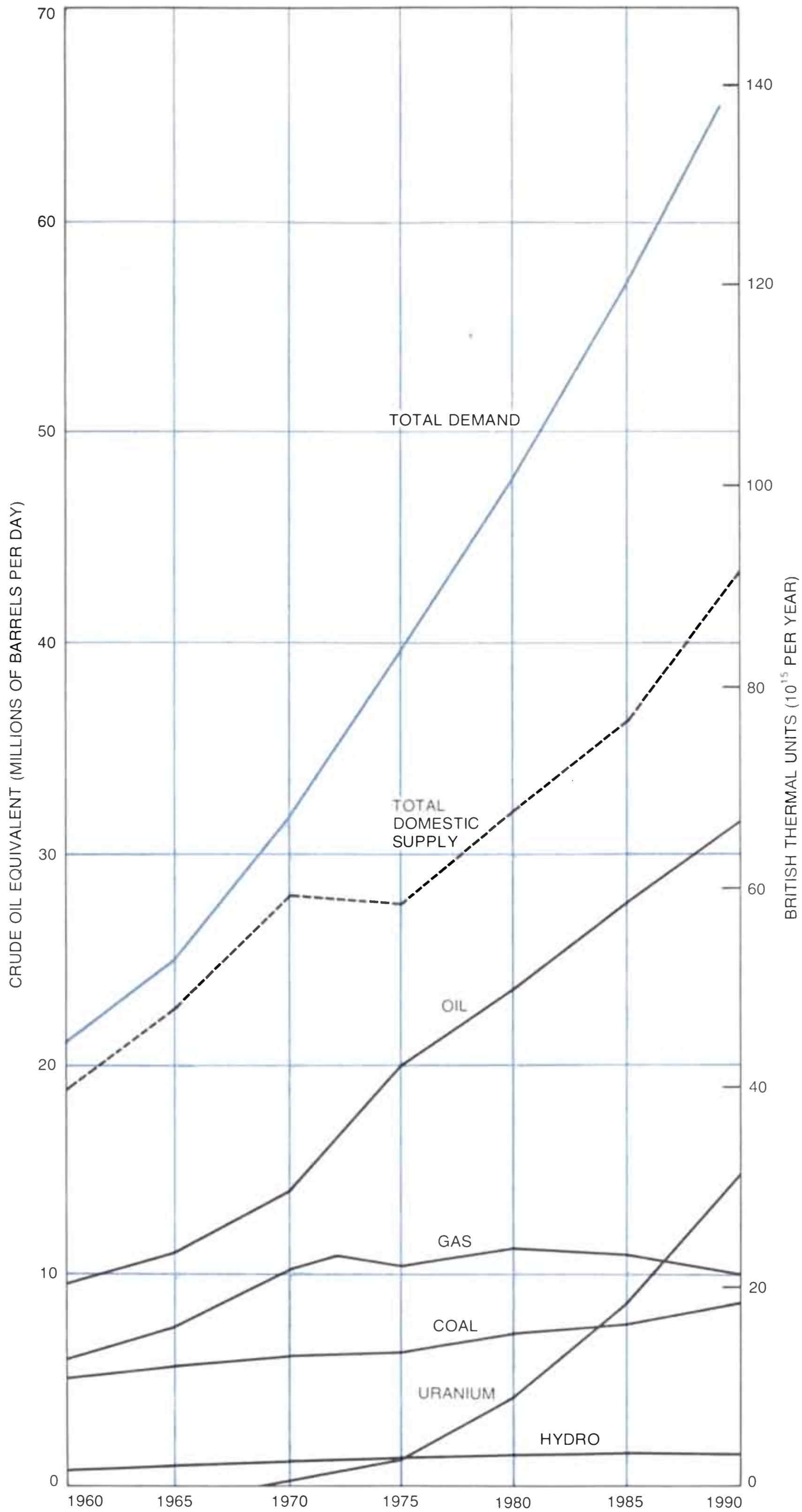
In fact, the energy crisis not only was predictable but also was in its general nature predicted. For one thing the petroleum industry is short of domestic refining capacity by about three million barrels per day. Its spokesmen give environmental restrictions on siting as a principal reason. The short-term demand for fuel, however, is well known to be highly inelastic; this means that a small shortage leads to large price increases. Thus not by any collusion but by a little benign neglect the petroleum industry could improve its lot substantially. Compounding the difficulty, and against advice from many sources, the industry has allowed tax credits and other incentives to increase its dependence on overseas refinery capacity. The automobile industry has paid virtually no attention to fuel conservation. The Federal Government has developed little capability to collect data on fuel demand and resources and has been content with petroleum-industry data. Few in decision-making positions in Government or industry paid attention to the scarcity of low-sulfur fuel as they promulgated environmental standards. Federal agencies responsible for developing nonpetroleum fossil fuels (particularly clean fuels from coal, which might have provided not only earlier relief but also competition to petroleum fuels) have been virtually starved while tax funds have been lavished on nuclear reactors. Few universities and public-information groups found it either interesting or rewarding to illuminate the issue. Our present difficulties were largely caused not by ignorance but by irresponsibility.

The President has announced a set of mandatory regulations, effective January 1, designed to reduce consumption of heating oil, gasoline and jet-aircraft fuels by 1.7 million barrels per day, or slightly less than 10 percent of last year's average daily demand. Even with this reduction, the President said, available supplies will still fall 7 percent short of the anticipated demand, so that "additional actions will be necessary." A predicted ultimate shortfall of 17 percent is largely attributed to the oil embargo imposed by the Arab states. As a long-range response to the Arab action the President proposed "Project Independence 1980," which he defined as "a series of plans and goals set to insure that by the end of this decade Americans will not have to rely on any source of energy beyond our own."

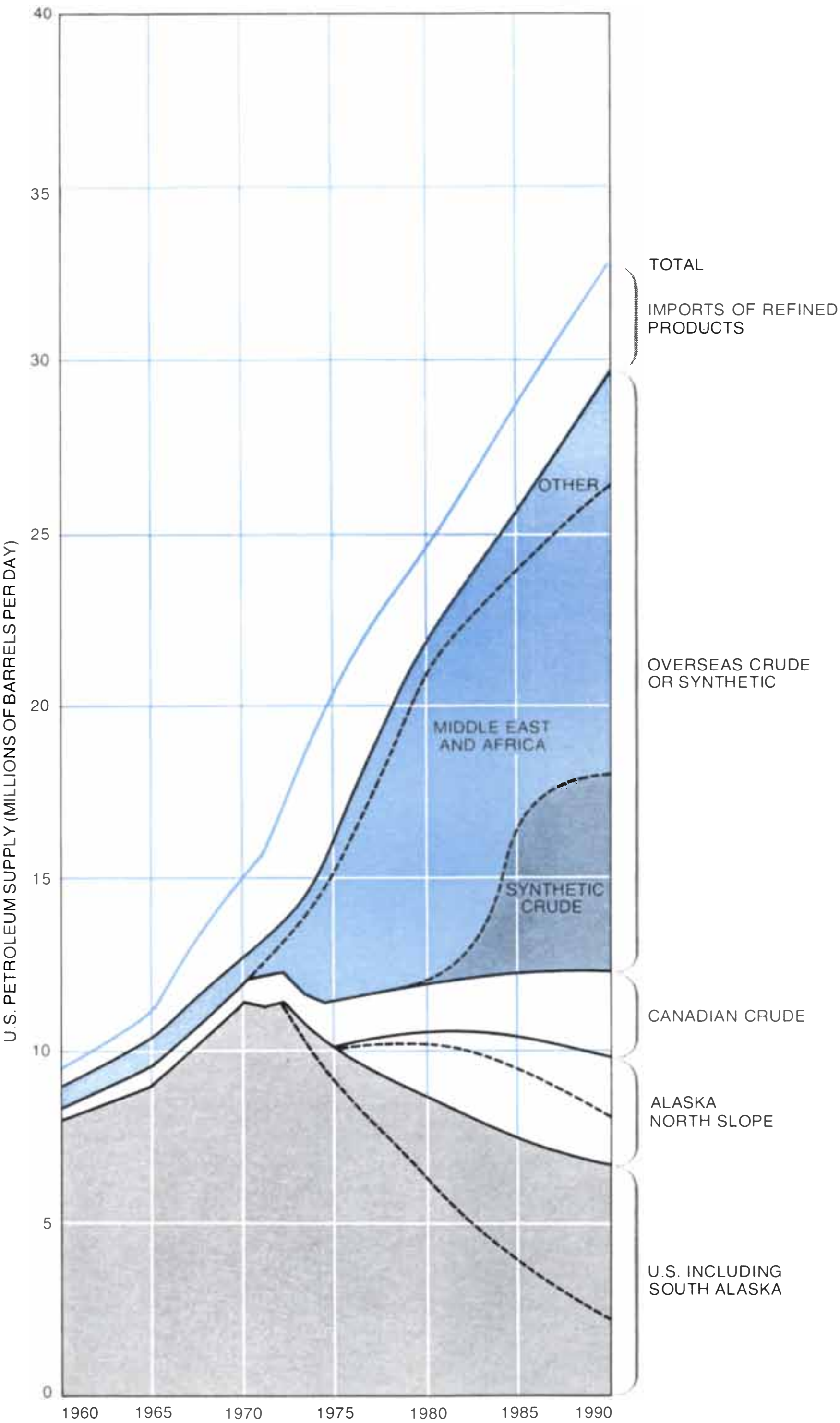
Such an ambition seems unachievable without the application of Draconian measures, and probably would be unwise besides. Lack of policy has in effect encouraged substantial foreign dependence. Estimates show that to achieve hemispheric (not domestic) self-sufficiency by 1980 means closing an energy gap of nine million barrels per day. Conceivably this might be achieved by combining the strict conservation of energy with the ruthless exploitation of all the energy resources tappable within the short span of six years. That would mean a sharp curtailment in the booming demand for oil (recently growing at 7 percent per year), the accelerated depletion of known oil fields, intensified drilling offshore in the hope of a major strike, the relaxation (if not the total abandonment) of environmental-quality standards, unrestricted strip-mining and a wholesale shift from oil and natural gas to coal, particularly for electric-power generation. Between now and 1980 it will be virtually impossible to build more nuclear-power plants than those already on the drawing boards. It is also unrealistic to expect any substantial production of synthetic crude oil from coal or oil shale by 1980. It is estimated that to achieve a capacity of five million barrels per day of synthetic oil from these sources would cost $50 billion and take eight to 10 years. Solar and geothermal energy can make no important contribution in the near future. And power from fusion reactors cannot be expected before the end of the century.

This preamble brings us to the point of asking what the energy problem is, instead of only what went wrong. If there is any excuse for the nation's being confronted with an energy crisis, it can be found in the sheer richness of the energy problem. The scale of the problem is too vast and its time horizons are too distant for it to fit the customary behavior of the institutions charged (or left) to deal with energy. For example, industrial rates of economic return lead to time horizons only five to 10 years hence, but the problems themselves have a much longer lifetime, and the rewards for solving many of them accrue only to the public at large and not to specific industries. Thus even if the U.S. could discern and immediately adopt the wisest set of actions to meet the present crisis, energy problems would still persist. Some measures and actions now proposed are part of a continuing series that, if sensible, will bring gradual progress and improvement but never total "solutions." Thirty years from now energy will still be a serious topic; only the details will change.

Analyzing alternative solutions to technical problems and weighing their consequences has an increasingly fashionable name: technology assessment. The term, although useful for naming the task to be performed, is hardly a recipe for how to go about the task. When the problem concerns a subject as multifaceted as energy, in which technology, economics, resource allocation and social goals all interact, it becomes extraordinarily difficult to balance costs and benefits and reach a national con-



U.S. ENERGY SHORTAGE was clearly predictable at least two years ago when the domestic supply of fuels began for the first time to fall sharply behind the rising total demand for energy. Domestic production of natural gas and crude oil reached an all-time high in 1972 and has been falling since. The oil and gas curves plotted here include a rising fraction of imports and, beginning in the late 1970's, limited quantities of synthetic gas and oil from coal and oil shale. The individual curves for hydroelectric power and fuels add up to yield total demand. The widening gap between domestic supply and total demand is accounted for almost entirely by the domestic shortage of oil, as is illustrated on the next page. (The two illustrations closely follow the projections made by the Shell Oil Company.)



U.S. DEPENDENCE ON OVERSEAS OIL cannot be eliminated in the foreseeable future except at what would seem to be prohibitive cost. It is estimated that the U.S. will have to import some 16 million barrels per day of oil from overseas in 1990, of which at least 14 million barrels will originate in the Eastern Hemisphere, chiefly the Middle East and Africa. Just to reduce this figure to eight million barrels per day the U.S. would have to build plants capable of producing six million barrels per day of synthetic crude oil from coal or shale, at an estimated cost exceeding $50 billion. This would be virtually as much oil as the U.S. is expected to pump from all its domestic wells in 1990. The broken lines in the projections for the U.S. and the Alaskan North Slope indicate how even their oil output will fall if there are no further discoveries. The curve for total supply is made up of barrels of varying B.t.u. content, depending on source, hence the total corresponds in B.t.u. content, but not exactly in barrels, to the oil curve in the illustration on the preceding page.

sensus. For one thing, many goals naturally oppose one another, such as cheap coal and minimum land disturbance. It is hard for a partisan of one view, no matter how conscientious, to assess opposing views. At the crudest level, to recognize the validity of an opposing view tends to weaken one's own. More intractable kinds of intellectual imbalance arise when advocates of a particular option attract a band of adherents who, in their overenthusiasm, convert the option into a crusade.

What is needed, among other things, is some overall taxonomy of energy: a listing of the options in a logical hierarchy, so that national debate leads to public illumination and eventually to more satisfactory choices. Unfortunately no unique taxomony exists, but any reasonable one is better than none; one hopes through study of the taxonomy to achieve some degree of insight. Then better decisions will follow. Here I attempt a taxonomy based primarily on technological issues, but it will soon become apparent where and how nontechnical issues enter also, and indeed often dominate the discussion. It will also become apparent that some currently popular ideas lack merit.

The technological discussion proceeds best from the particular to the general. One begins by comparing the simplest technical options (one component with respect to another, say). After that one compares alternative major devices, then alternate strategies for achieving major technical goals and so on, thereby constructing a succession of ever more complex intellectual assemblies. Each higher stage of assessment tends to bring in more and more nontechnological issues, such as environmental cost, resource use or social purpose.

Let us start, then, at a reasonably simple level and choose a topic: Comparison of various nuclear methods for generating nuclear power, arranging the options to produce a structure resembling a mobile, the kind that hangs from a hook on the ceiling [*see upper illustration on page 26*]. The mobile has two main segments: one labeled "Fission" and the other "Fusion." Of these two general routes to nuclear power the latter will probably not be available until after the end of the century. The fission branch of the mobile divides into two subbranches: converters (present technology) and breeders (future technology). Under converters there are two subclasses: light-water reactors and the more advanced high-temperature gas reactors. There are also two subclasses under breeders: the liquid-metal fast-breeder reactor that may be available in the 1980's, and beyond that, perhaps, the gas-cooled fast reactor.

The mobile analogy is useful because it presents specific technological "varieties" as options at the bottom of the structure; at the next higher level the options are between species of devices and at still higher levels the options are among genera, families, orders, classes and phyla. Thus the mobile establishes a taxonomic ordering of alternatives. In making assessments one gives least weight to the individual items at the bottom of the structure and increasingly more weight to options available as one moves upward. As in constructing actual mobiles, one must build the structure and balance the items from the bottom up.

What weights, in development dollars, are actually being given to elements in the nuclear-power mobile at the present time? Between fission and fusion the funding ratio is about five to one: $400 million for fission reactors to $80 million for fusion reactors. The ratio is roughly appropriate to the distant time horizon for fusion as well as to its remaining uncertainty. If and when fusion power becomes more certain, it will require more development funds than fission power did; fusion is technologically as far beyond fission as fission was beyond coal-burning.

The principal imbalances appear in the fission program itself. The gas-cooled reactor has been delayed for lack of development funds. The light-water devices were developed either with Federal money (as part of the nuclear-submarine program of the Westinghouse Electric Corporation) or with conscious acceptance of initial losses (such as those incurred by the General Electric Company in promoting the boiling-water reactor). The high-temperature gas reactor may actually be safer than the water-cooled reactors, more economical of uranium resources, more efficient (meaning that less waste heat is rejected to the environment) and perhaps even cheaper to build (although not all these advantages are confirmed). Its development lagged because its sponsor, the General Atomic Division of General Dynamics, could not afford to accept losses on the initial units. Now that General Atomic is part of the Gulf Oil Corporation that limitation has been removed; a first reactor nears operation and six more are on order.

A different and more serious imbalance applies to the breeder-reactor program. Its budget of $320 million in the fiscal year 1974 represents a fifth of the total U.S. research and development expenditure on energy and more than a third of the Federal effort. The breeder's chief advantage over present reactors is fuel economy. Whereas present reactors depend on fission of the rare uranium isotope U-235, the breeder can utilize U-238 (99.3 percent of all uranium) by converting it into fissionable plutonium. To be sure, uranium costs will rise appreciably by the end of the century if a breeder reactor is not developed, but since uranium costs are only a small fraction of the total cost, delivered electric-power costs would not rise more than a few percent. Thus a demonstration breeder reactor for the U.S. is less urgently needed than the Atomic Energy Commission, the White House and Congress have maintained. Europe and Japan, far poorer than the U.S. in reserves of fossil fuels and somewhat poorer in uranium, have more urgent reasons than we do to develop all forms of nuclear power, including power from breeder reactors.

In addition to pushing the breeder concept faster than the facts warrant, the Government has put virtually all its support behind the liquid-sodium-cooled version, allotting only $1 million per year to Gulf General Atomic's gas-cooled fast breeder. The relative promise of the two concepts is in no such disparate ratio. Still worse, concentration on only one technological device is risky and unwise.

A similar assessment can be made with respect to generating electric power from alternate energy sources. Again the options can be arranged in the form of a mobile [*see lower illustration on page 26*]. Some of the options that have been proposed can be dismissed out of hand. For example, it is easy to calculate that if a low dike were built around the entire U.S. to harness all the tides, the resulting electric power would only satisfy the needs of a city the size of Boston. To supply the U.S. electric needs by wind power would require windmills 100 meters high spaced a few kilometers apart over the entire country. Most of the suitable hydroelectric sites are already developed. (Hydroelectric generators now account for 10 percent of the U.S. electric-power supply.) It is clear that tides, winds and falling water are not solutions to the nation's energy problem.

The heat of the earth's interior is vast but normal flow of it to the surface is small. It has nonetheless been estimated that subterranean sources of steam and hot water have a potential for supplying about twice as much power as the U.S. currently obtains from hydroelectric sources.

Gaining in popularity is the notion of drilling holes to reach kilometer-size bodies of hot rock that lie anomalously near the surface. There are perhaps 1,000 such bodies in the geologically active western U.S., enough to satisfy the region's power needs for a very long time. The injection of water might both fracture the rock and jack up the strata to facilitate percolation. Steam would be withdrawn through separate exit holes. The idea is not unattractive, but there will be problems. Since hot water dissolves many minerals, it will be hard to keep cooler piping free of mineral deposits. Moreover, percolation channels tend to become enlarged where the flow is greatest, thus leading to large mass flow with poor heat transfer. Such difficulties are well known to the drillers of deep wells.

Solar energy is a different story. It is plentiful and free, but the problem is to collect it efficiently and economically. A million-kilowatt plant, equal in output to the largest conventional generating station, would require a collection area of 100 or more square kilometers, depending on the efficiency of conversion. That might seem to put such options beyond consideration, but a coal-burning plant, obtaining its fuel from strip-mining of coal seams half a meter thick, will cause the same area to be stripped in 25 years. With available technology a solar-energy power plant would cost between five and 10 times as much as a coal- or nuclear-power plant. Advocates believe the cost would fall sharply with suitable engineering development.

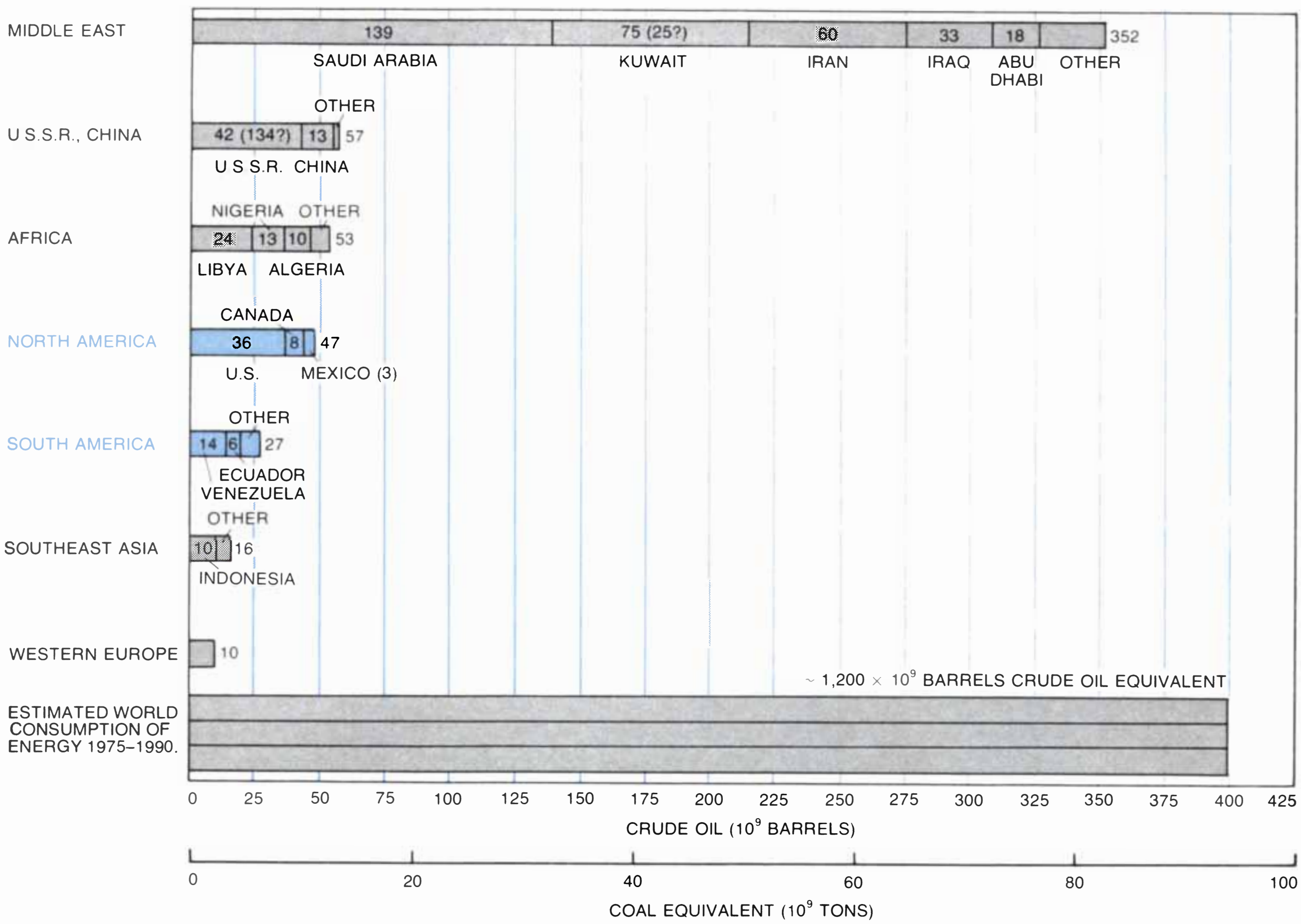
The idea of converting solar energy to electric power in space and beaming it down to the earth at microwave frequencies would provide energy around the clock, fair weather and foul. To be economically feasible the cost of available components would have to be reduced by a factor of about 100 and the cost of putting the components in orbit by a factor of about 10, over and above the economies promised by the space shuttle. Beyond that there is worry about the long-term effect of low-level microwave power on life near the receiving antennas, which would have to cover tens of square kilometers. Meanwhile terrestrial solar energy, including welcome applications of it for domestic space conditioning and water heating, enjoys for the first time some reasonable exploration: $12 million in Federal funds in the fiscal year 1974.

In assessing the available options for generating electric power a corrosive and ill-constructed debate has developed between some vocal advocates and critics of fossil fuels and nuclear fuels. The costs of generating electric power have been rising more sharply than the general price index for two reasons: the dramatically increasing cost of nuclear reactors and the need to use low-sulfur fuels in conventional power plants (or to add sulfur-recovery equipment). Except where low-sulfur coal is plentiful and cheap, nuclear electricity now tends to be cheaper than fossil-fuel electricity. Moreover, the disparity in price will probably increase if air-quality standards force more restrictions on fossil-fuel plants and as clean fossil fuel becomes steadily more expensive.

Several issues enter the discussion, some of them spurious. The clandestine and irresponsible use of nuclear-weapons material is quite unlikely to be prevented by this country's refraining from installing nuclear-power reactors. The core issue is environmental: Which type of power plant is actually, or potentially, more hazardous? It is my opinion that the environmental and epidemiological evidence strongly favors nuclear-power plants. The Atomic Energy Commission has spent more than $1 billion exploring the health and other environmental problems of nuclear energy. Although its record is not perfect and more remains to be done, a huge amount of information has been made public. The nuclear hazards are fairly well recognized and widely advertised. Principally they are associated with uranium mining and processing, with the normal operation of nuclear plants and fuel-reprocessing facilities, with long-term waste disposal and with accidents.

Conversely, the Department of the Interior, which has cognizance over coal and its technology, has spent hardly anything on the general environmental and epidemiological hazards of burning coal, leaving the problem largely to the Department of Health, Education, and Welfare and the Environmental Protection Agency. Thus the hazards of fossil fuels have been little studied or publicized. The data that do exist show that the total social cost of generating energy with fossil fuels has vastly exceeded the cost associated with nuclear fuels per unit of energy, at least with the environmental and work standards that applied through the 1960's.

For example, Lester Lave and Eugene



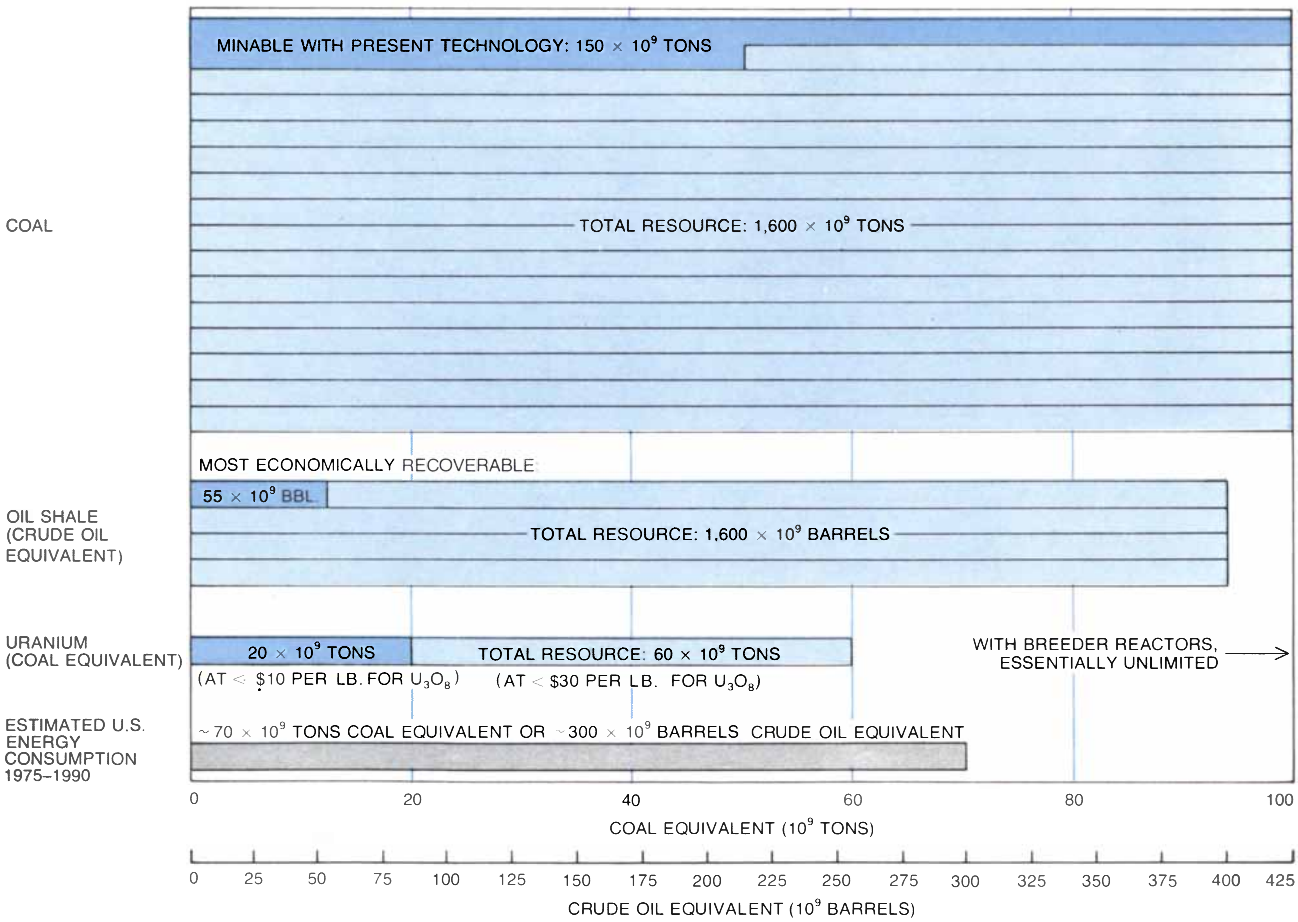
PROVED WORLD RESERVES OF CRUDE OIL total 562 billion barrels, distributed geographically as shown here. More than half of the proved reserve is concentrated in five Middle East states. The U.S.S.R. and China together possess about 10 percent of the total. The entire Western Hemisphere has 13 percent. The estimate for the U.S. includes 10 billion barrels from the North Slope of Alaska. The National Petroleum Council estimated that at the end of 1970 some 385 billion barrels of oil remained to be found on U.S. territory or immediately offshore. This in turn was believed to represent about half of all the oil ultimately discoverable. For the world as a whole N.P.C. estimates that proved reserves can be doubled in the next 15 years. For this reason the estimated total world consumption of energy between 1975 and 1990, equivalent to some 1,200 billion barrels, is not so alarming as might otherwise appear. In fact, natural crude oil will probably be supplying at least 60 percent of the world's total energy demand even in 1990.

Seskin of Carnegie-Mellon University and Thomas A. Hodgson of the Cornell Medical College present data implying that the pre-1968 health cost to New Yorkers from unrestricted coal burning in power plants was several thousand deaths per year, plus uncounted nonfatal disabilities of varying severity. Some 50,000 American coal miners are currently disabled with black-lung disease. To these social costs must be added the despoliation of land by strip-mining.

These social costs, which appear to be more than 100 times higher than the equivalent nuclear costs per unit of energy, will no doubt be reduced as environmental standards rise, but the cost of putting fossil fuels on an even environmental footing with nuclear fuels seems prohibitive. Meanwhile debate has concentrated on the more publicized nuclear hazards and has led indirectly to lowered air quality as a result of the construction or retention of fossil-fueled plants. We see here a clear case of unbalanced debate and consequent faulty decisions arising from an initial imbalance in available information. When the public is presented with a balanced picture of the consequences of burning fossil fuel, I feel sure there will be an accelerated movement toward nuclear power and much more caution about relaxing environmental standards during energy shortages.

We now pass to the next level in our hierarchical assessment, the allocation of primary energy resources among users [*see upper illustration on page 27*]. The electric-utility industry and transportation each take about 25 percent of the nation's fuel supply; another 20 percent is required for space heating and 30 percent is consumed by industry (which also takes about 40 percent of the electric power generated). Here at the level of dividing up the national energy budget it is again fair to ask whether the relative effort to develop better options matches the relative needs.

Again the answer is no, and coal once more serves as a good example. Fossil fuels, in spite of their drawbacks, will be needed for many years, not only for the generation of electric power but also for transportation, for home heating, for industrial purposes and so on. Coal, together with the oil shale of western Colorado and the tar sands of western Canada, is a unique North American reserve of fossil fuels. In a period of ever increasing prices and ever decreasing security the first two are the only resources capable of replacing imported fuels until better and more nearly inexhaustible resources can be rationally developed. The U.S. Geological Survey estimates that the U.S. possesses 1.6 trillion tons of



U.S. RESERVES OF COAL, OIL SHALE, URANIUM vastly exceed the world's proved reserves of crude oil, as depicted to the same scale on the opposite page. The U.S. coal resource of 1,600 billion tons is defined as half of the coal estimated to be present in beds as thin as 14 inches at depths of up to 4,000 feet. About 10 percent, or 150 billion tons, exists in beds comparable in thickness and depth to those being mined today. The principal deposit of oil shale is in the Green River Formation in Colorado, Utah and Wyoming. Some 55 billion barrels of shale oil are readily recoverable in seams more than 30 feet thick containing more than 30 gallons of oil per ton. The reserves of uranium are computed on the basis of use in present-day reactors. When breeder reactors are available, the reserve will become essentially limitless. Fuel for fusion, if it ever becomes practical, is likewise limitless. With an enormous effort the U.S. might reduce its energy consumption 10 to 15 percent below the figure represented by the bottom bar.

recoverable coal in beds at least 14 inches thick, lying no deeper than 4,000 feet. The total is equivalent to 500 times the total U.S. energy consumption last year and more than 20 times the energy the U.S. will consume between now and 1990 [*see illustration on preceding page*]. Of course, only a fraction of the coal reserve, perhaps no more than a third, is reasonably recoverable with existing technology at acceptable cost.

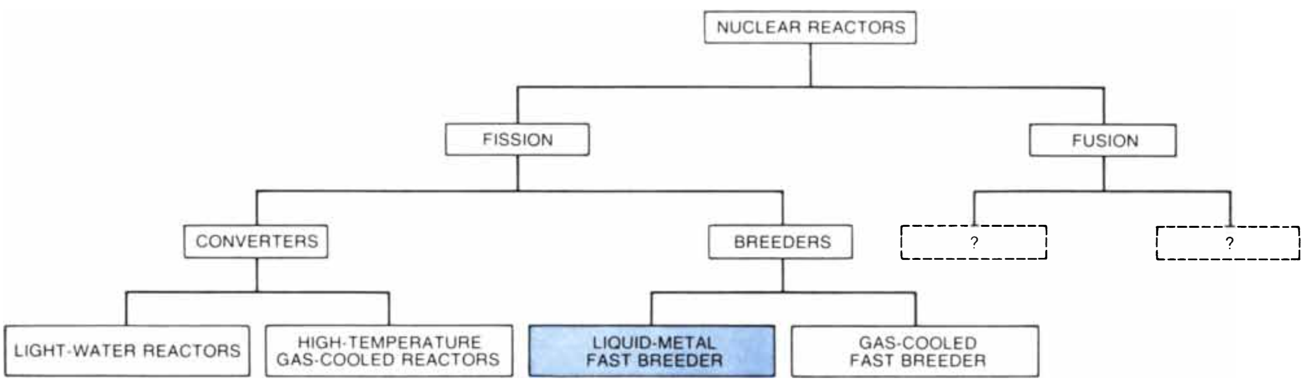
At present, however, coal provides only 18 percent of the nation's energy needs, a fraction that has dropped with time. (In 1900 coal supplied 70 percent of the nation's energy, and as recently as 1950 it supplied 36 percent.) Most present modes of coal extraction and use have been socially, environmentally and epidemiologically damaging. In this respect the technology of coal languishes. Oil shale may be an even worse environmental problem, but these and other difficulties are correctable, in my opinion, if the public has the will to demand correction.

Coal and oil shale (perhaps tar sands too, with the appropriate consideration of Canadian interests) now appear in their proper light: as raw materials for a synthetic-fuel industry that can limit economic and political threats from abroad. The domestic cost of producing low-sulfur crude oil and delivering it to the East Coast of the U.S. has been about $3.75 per barrel (42 gallons) until recently, but now has risen to just above $5. Until last fall approximately the $3.75 price was charged by overseas suppliers, mainly through the operation of the Organization of Petroleum Exporting Countries (OPEC), even though the actual cost of producing and shipping crude oil from the Middle East is only a fraction of that figure.

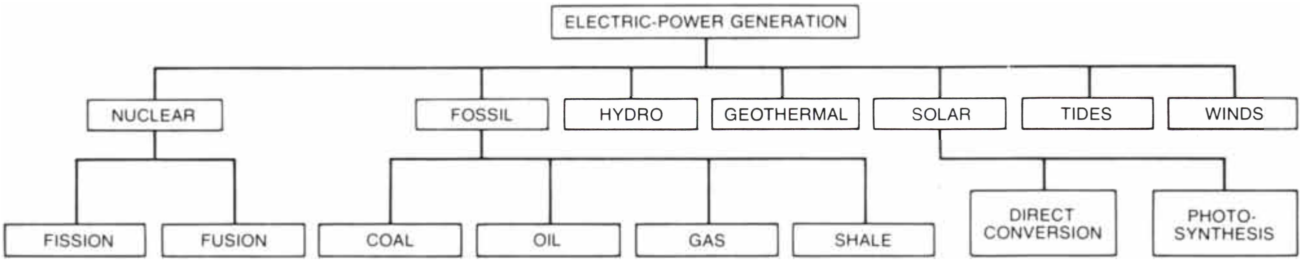
With the recent outbreak of war in the Middle East the Arab states raised prices substantially, and then several of them cut off supplies to the U.S. Saudi Arabia cut its total exports 20 percent and raised prices to recoup; Libya raised the posted price of her product from $4.604 to $8.925 a barrel. Such prices, of course, can scarcely be maintained under normal conditions. Nevertheless, within the next 10 years the U.S. can expect to be paying $8 or more per barrel (in 1973 dollars) for imported crude oil. Even if the price were to rise no higher, the annual U.S. bill for foreign oil in 1985 could reach $40 billion, if imports rise to the 15 million barrels per day given in some estimates [*see illustration on page 22*].

Even worse is the threat of eternal political blackmail, which can only be met with determined action by the U.S. One such possibility is establishment of a substantial synthetic-fuel industry. Between now and the early 1980's there would be enough time to develop environmentally acceptable methods for producing oil from coal, and perhaps from oil shale as well, at less than $7 a barrel. At the same time programs now under study will probably lead to the production of clean synthetic natural gas from coal at, say, $1.20 per 1,000 cubic feet, equivalent in energy cost to petroleum at $7 a barrel.

The implementation of such a strategy would be neither cheap nor simple, even if it is a good option. One can ask, "How much syncrude is enough?" and thereby



ASSESSMENT MOBILES display in simple, easily grasped form the various technical options available for doing a specific job or reaching a specific goal. Such mobiles are best constructed from the bottom up. A mobile for energy policy might be "assembled," as shown here, by considering first the devices available or potentially available for producing power from nuclear reactions. Light-water reactors have been generating commercial power for 15 years; the first high-temperature gas-cooled reactor is nearing completion. Both are simple converters, that is, they consume the rare uranium isotope U-235. Breeders, now under development, will produce more fissionable material than they consume. The U.S. has budgeted $320 million this fiscal year for development and construction of a liquid-metal fast-breeder demonstration plant. The gas-cooled fast breeder, on the other hand, will receive only $1 million in Federal funds. In the hope of demonstrating the technological feasibility of fusion reactors the U.S. will spend $80 million this fiscal year. Controlled fusion is unlikely to provide significant amounts of electric power before the end of the century.

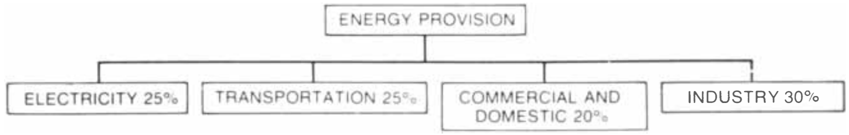


NEXT STEP IN ASSESSMENT leads to a larger mobile that includes sources of energy besides nuclear materials. The actual assessment considers the availability of the source, methods for exploiting it and cost. Tides and winds are readily shown to hold little promise. Geothermal and solar energy are potentially unlimited sources but the technology for exploiting them is rudimentary.

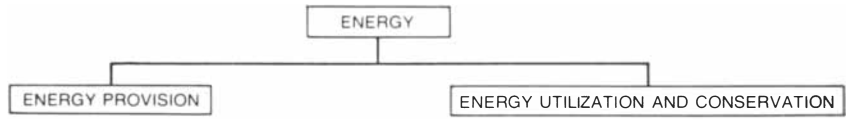
raise a host of other questions. To build the capacity needed to produce five million barrels per day (perhaps a third of the projected 1985 imports) would cost about $40 billion. Although that is no more than the U.S. spent in the 1960's on space ventures, its impact on the economy (particularly the construction industry) will be vastly different. Many difficult questions will have to be answered. What will be the impact on the engineering and skilled-labor market? How will other patterns of investment (investment in housing, for example) be affected? What will be the impact on the coal-mining industry? On water resources? What are the alternate strategies? At $7 per barrel how much more petroleum can be produced in the U.S. and its sea-bottom surround? More than 80 percent of the original U.S. petroleum reserve is still in the ground, and more is available with increased effort, but finding and extracting a substantially increased fraction would be very difficult.

International impacts are no less complex. Whether an increased U.S. production capacity would force OPEC prices to remain below $7 per barrel is hard to say, because the U.S. represents only a fraction of the world market. On the other hand, OPEC is not a monolithic structure, and the temptation for one OPEC country to abandon the consortium would be great. If that were sure to happen, the U.S. syncrude industry would have to have been built to stand largely idle. Underpriced by Arabian oil, it would be an economic and political weapon designed by Federal policy not to be used, in the same sense that the Department of Defense builds weapons systems not to be used. Other international questions relate to such matters as whether Western Europe and Japan will remain dependent on the Arab states, will turn to the U.S.S.R. (which is believed to have immense reserves waited to be proved in), or will turn even more to nuclear power. Questions such as these, which are difficult to frame, let alone answer, make the energy problem what it is.

Even if the U.S. does not elect to spend tens of billions on synthetic-oil facilities, coal-extraction technologies call for major improvement. It seems feasible to develop a fully automatic technology for mining coal underground. Where strip-mining remains the method of choice, the land can be properly restored; not doing so is a social problem, not a technological one. In many of the empty regions where coal is strip-mined land sells for not more than



ALLOCATION OF PRIMARY ENERGY RESOURCES is considered at the next higher level of assessment. Generation of electric power, transportation, industry and commercial and domestic consumers take roughly equal shares. In terms of end use the electric energy is divided between industrial, commercial and domestic use, with little for transportation.



HIGHEST TIER IN MOBILE balances energy provision with energy utilization and conservation. The U.S. neglected the right half of the mobile until the energy crisis arose.

$200 per acre, except where coal is present. To reclaim the land properly may cost $5,000 per acre in Appalachia and perhaps $10,000 or more per acre in some Western states, because of the arid climate and the need for long-term care. In contrast, the value of the coal in a two-foot seam in Appalachia may be $40,000 per acre and that in a 50-foot seam in the West may be $1 million per acre. Clearly the reclamation costs are small compared with the sale price of the coal but very large compared with the normal sale price of the land. A social decision, one not based entirely on microeconomics, must be made on the value of the land to the society as a whole.

The preceding assessment applies only to the next 50 or 100 years, certainly not longer. It would be compounding tragedy to convert any large fraction of the domestic coal reserves to liquid or gaseous fuels when coal hydrocarbons are so much more valuable as chemical raw materials. Eventually, as the U.S. and the world switch to energy technologies based on virtually inexhaustible resources—fission, fusion and solar power—coal will no longer be needed as a major fuel source.

I have dwelt on the coal question because it is an excellent example of issues that have deep roots both in technology and in society and that cannot be resolved without serious and simultaneous attention to the consequence of various options. I shall discuss only briefly other issues associated with the allocation of primary energy resources. The U.S. is currently spending about $1.5 billion a year for research and development on new energy options. The total is only about 1.5 percent of the sum that the energy "industry" contributes directly to the gross national product and is thus well below the average research and development expenditure in other technological sectors of the economy. The small scale of the budget for energy research and development is a strong clue to what has gone wrong. Of the $1.5 billion more than half goes toward the provision of electric power, where the options are in the best order. An imbalance arises not because electric-power developments receive too much attention but because other areas have received too little.

We now move to the highest level in this hierarchical assessment of energy: energy provision v. energy utilization and conservation [*see lower illustration above*]. All the energy provided (the prodigious sum of 1.9×10^{14} B.t.u. per day for the U.S., or 11 kilowatts continuously for every person) goes somewhere, and its rational use is only now receiving appreciable attention. The efficiency with which energy is consumed ranges from less than 5 percent for the ordinary incandescent lamp to perhaps 75 percent for a well-maintained home furnace. The automobile engine (particularly since the installation of emission controls) has an efficiency of less than 20 percent. Modern fossil-fuel power plants are more than twice as efficient. On the average probably less than 35 percent of all the B.t.u.'s consumed end up as comfort heat, useful work or visible light. Low as the figure is, it has probably quadrupled since 1900. Up to the present time the allotment of funds for developing more efficient energy converters has been paltry: perhaps a few percent of the total $1.5 billion budget for energy research and development.

The imbalance is beginning to be corrected. For several years a few Federal organizations have quietly and penuriously been studying how energy might be better utilized: the National Bureau of Standards, the Office of Emergency Preparedness and the Oak Ridge National Laboratory. Stimulated by real fuel shortages, their concern has become popular, reinforced by the enthusiasm of the environmentalists and conservationists. An Office of Energy Conservation was established in the Department of the Interior last year, with an annual budget of $6 million.

Some energy-conserving and dollar-saving options are straightforward, easy to apply and have now received attention: lower automobile speed limits, year-round Daylight Saving Time, a reduction in "cosmetic" lighting, lowered thermostats and so on. A less well-known energy-saving option is the use of radial tires, which can reduce gasoline consumption 5 to 10 percent (because of reduced friction between the tire and the road). Other options are easy to envision but more difficult to adopt: lighter automobiles, for example, and commercial buildings designed to achieve comfort with minimum energy consumption.

The study of energy conservation and rational energy utilization has languished for simple reasons. Until now energy has been so plentiful and so cheap that there has been little incentive for people to be frugal. It has not been easy for the individual to perceive that his bargain in energy entailed large costs to the society as a whole. And even when the fact is pointed out to him, the individual rightly sees that any personal sacrifice he might make has only a minuscule effect in ameliorating an environmental problem. The second big reason for the neglect of energy conservation is that industry is richly rewarded for selling energy and energy-consuming devices.

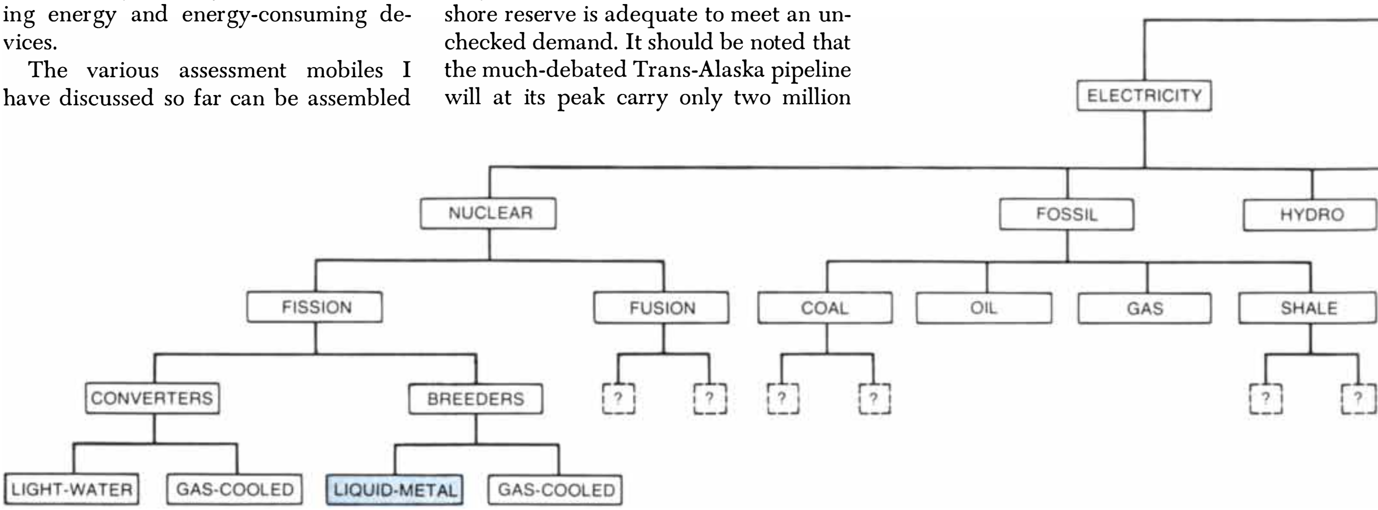
The various assessment mobiles I have discussed so far can be assembled into one large mobile that includes all the options laid out hierarchically [*see illustration on these two pages*]. In this arrangement one can see that utilization and conservation, on the right half of the mobile, merits as much attention and effort as all the options dangling in five tiers on the left half of the mobile. Yet one small line at the bottom left, the liquid-metal breeder reactor, represents an option that will receive $320 million in Federal and private development funds this year: 20 percent of the entire national budget devoted to energy options. What the mobile dramatizes is an embarrassment of blank space, where options are either absent or so poorly formulated that they have received little attention.

Several suggestions for ameliorating future energy difficulties have appeared in this discussion, coming mainly from reflection on new technical options, on conservation methods and on international economic relations. We can see several other strategies being implemented in some measure today. One approach involves a modified *laissez faire*, leaving parts of the problem to be solved in the marketplace; with increasing scarcity rising prices will bring supply and demand into some kind of balance and less energy will be wasted. To be sure, energy prices have been too low, but overreliance on this approach would work a hardship on many people.

Another strategy is to expand domestic exploration and production of petroleum fuels. The rocks of the continental shelves seem quite similar to the typical sedimentary rocks of the U.S. and Canada, so that they are believed to contain rich petroleum deposits. It is extremely risky to assume, however, that the offshore reserve is adequate to meet an unchecked demand. It should be noted that the much-debated Trans-Alaska pipeline will at its peak carry only two million barrels per day. That amount will not provide much relief from a predicted import requirement of 11 million barrels per day in 1980. Even where an oil field is known to exist it takes from five to seven years to bring it into production. And on the continental shelf off the U.S. eastern seaboard the first hole remains to be drilled; the existence of oil is simply conjectured.

Another strategy that is currently prominent is application of regulatory and other measures to limit energy use: taxes, allocations, rationing and outright bans. The strategy has the advantage that it can be designed for prompt effect. It is the only approach with this feature, and it is therefore invoked in this winter of fuel emergency. We should be aware, however, that such measures generally represent the failure of past policy (or the lack of any policy at all) and that the consequences of stringent regulatory actions are hard to predict. For example, the 1971–1973 controls to combat inflation were a substantial failure. The National Energy Emergency Act sent to Congress on November 21 of last year relies heavily on emergency controls, and it will have little to do with rational long-term energy policy. On the other hand, regulations designed for long-term constructive effect have their place in a coherent energy policy.

One may hope that from some combination of all these possibilities a better energy program will emerge, involving the Federal Government, state governments, industry and other sectors. Clearly the days are gone when parts of the problem could be denied or ignored, as for example when energy conservation and rational energy utilization were



COMPLETED ENERGY MOBILE is grossly unbalanced. The attention given to energy provision far outweighs the attention to utilization and conservation. For projects concerned with energy provision the U.S. is currently spending some $1.5 billion per year.

ignored by the President's office until the middle of last year.

To choose any particular combination of energy strategies without much thought and analysis would be to compound the difficulties. An adequate national energy policy must of course rise from basic decisions made by or on behalf of the country as a whole, decisions concerning national security, costs, the present and future quality of life and so on. When those decisions have been made, a planner can set down tentative desiderata—not yet firm decisions—about operational strategies. They would include the degree of dependence on foreign oil supplies, environmental standards and the balance between exploitation and conservation of natural resources. Only then can planners arrive at detailed strategies to achieve selected goals: the technology of nuclear power, new types of automobiles and so on. The process is obviously not straightforward; decisions at each level affect conditions both above and below.

These new intellectual and technological thrusts have their organizational counterparts, and the old arrangements for dealing with energy, at least at the Federal level, are being swept away. In general the old ones were not impressive. Federal energy policy consisted mainly of a strong nuclear program (coming from a strong Congressional Joint Committee on Atomic Energy, and a legacy from earlier military programs), a favorable attitude toward petroleum companies (direct U.S. tax credits for foreign royalty payments, depletion allowances and so on), the benign neglect of coal and several fairly successful regional programs (the Tennessee Valley Authority, for example). Technical activities were mainly in the Atomic Energy Commission; some were in the Department of the Interior and some were scattered in other agencies. Regulatory commissions were only partly effective. (For example, the Federal Power Commission set natural-gas prices so low as to simultaneously destroy exploration initiatives and create an insatiable demand.)

All of this is now changing rapidly. It is clear that more central coordination is required than the fragmented energy groups have managed heretofore. The optimum degree of coordination, however, is an important question. Too little can accomplish little; too much can throw the operations of agencies and groups with legitimate energy interests into disarray. At least there is need for the coordination of policy, and Governor John A. Love's former Energy Policy Office seemed to be doing that.

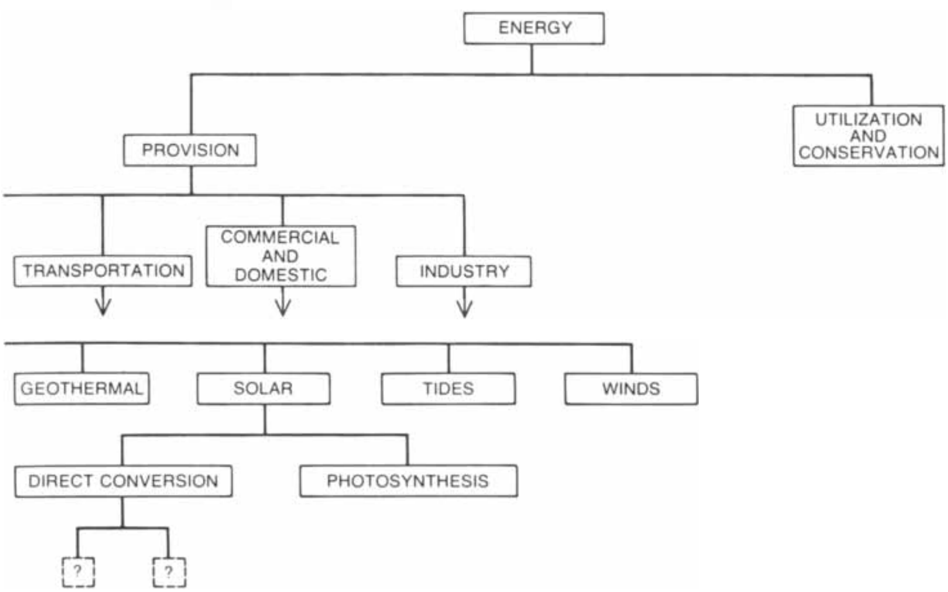
In jettisoning the Love office, the President sought a two-level approach. At the level of preparing technological options he proposed an Energy Research and Development Administration. This was to combine the energy-related activities of the various Federal agencies. The chairman of the Atomic Energy Commission, at the President's request, came up with a $10 billion, five-year program for the development of alternative energy technologies. These proposed expenditures include many projects already planned. At the policy and executive decision level the President proposed a Federal Energy Administration, to be headed by William E. Simon. Arrangements like this could free up to some extent the development of technological options, while giving the executive group freedom to pick and choose and to transcend narrow technical arguments. Congressional bills to implement these measures were introduced in the Congress by Senator Henry M. Jackson and Representative Chet Holifield.

Also at both the technological and policy levels Senator Jackson's own bill (S. 1283) aims at each of these technological and policy functions; through the allocation of $800 million per year of new funds for non-nuclear research, coordinating activities of the AEC, the Department of the Interior, the National Aeronautics and Space Administration and the National Science Foundation.

In the House, meanwhile, Representative Mike McCormack and his Subcommittee on Energy have considered not only a Department of Energy (not far from the President's ERDA-FEA proposal) but also the establishment of an integrated Department of Science and Technology. That is a still broader concept, requiring more public debate.

Whatever the administrative outcome, I personally favor an authority that will decide about rationing (I favor that too) and other short-term measures. Simultaneously such an authority should make a careful assessment of the long-term possibilities before committing the nation to some very expensive acts of commission or omission. The technology of energy conservation, the development (but not yet massive deployment) of new fuels from coal and oil shale, the full environmental costs of various fuel options all deserve careful study—not forgetting the important international repercussions of a change in the fuel-consuming habits of the U.S.

The role of energy is too pervasive and the interests of the participants are too manifold for any expectation that there will be a simple consensus. It can nonetheless be hoped that the national debate will give rise to a better common understanding of the problem. Then some steps can be taken in the direction of probable improvement, and from that new vantage the view ahead will be a little clearer than before.



About a fifth of this entire sum is devoted to work on a single mechanism, the liquid-metal fast breeder, which is represented by a single box at the extreme bottom of the mobile.

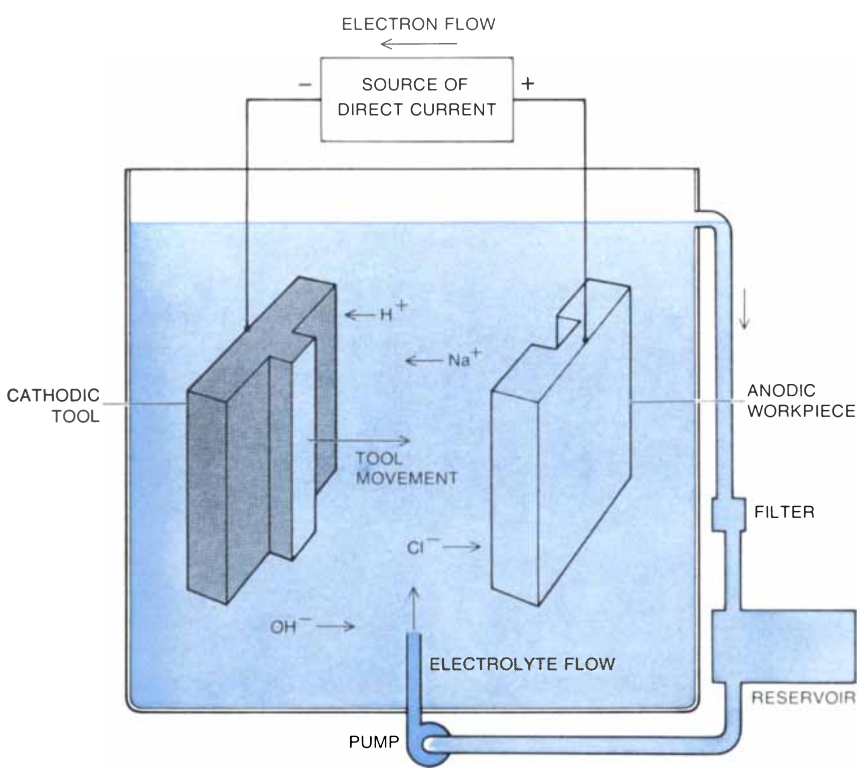
ELECTROCHEMICAL MACHINING

Hard alloys are difficult to cut into complex shapes by conventional machining. They can be worked to a smooth finish in an electrolytic cell where the hard metal is the anode and the tool is the cathode

by James P. Hoare and Mitchell A. LaBoda

The conventional approach to machining a piece of metal into a particular shape is to use a tool that is made of a harder material and is employed to chip or scoop out the softer metal of the workpiece. This approach runs into trouble when the piece to be machined is made of a high-strength alloy, because it may not be possible to find a tool material hard enough to do the job. Grinding techniques, which rely on a moving abrasive tool, can be brought into play, but they are slow and the range of shapes that can be fashioned with them is limited. A new and promising method of shaping extremely hard materials is electrochemical machining.

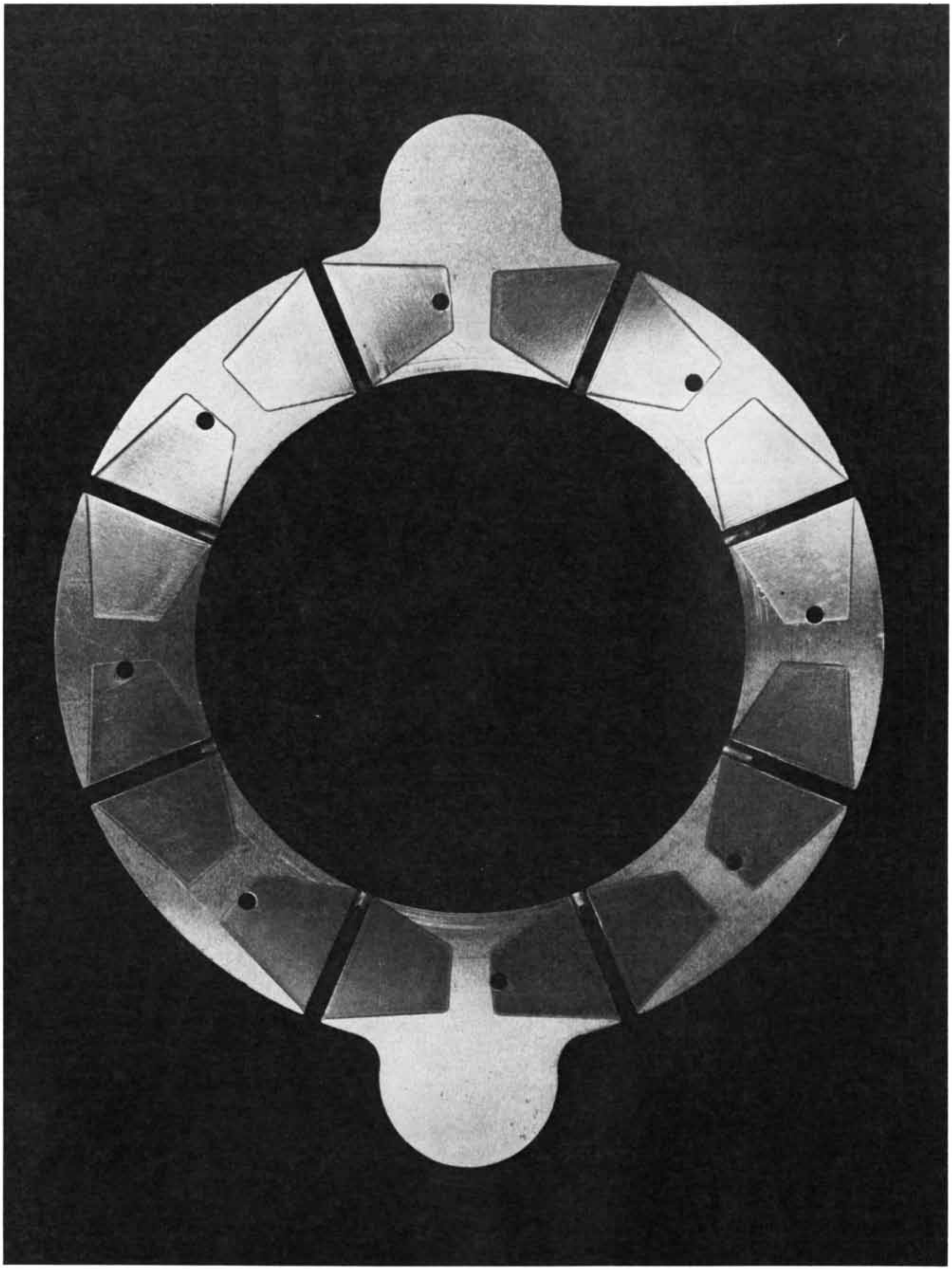
Electrochemical machining is essentially a process of controlled corrosion. The operation employs electricity to produce chemical changes. They take place in an electrolytic cell, which contains an electrically conducting liquid (the electrolyte) that is in contact with two metal conductors (the electrodes) connected through an external circuit [*see illustration on this page*]. The current that moves through the electrodes and the external circuit is carried by electrons, and the current that moves through the electrolyte is carried by ions, that is, atoms that have acquired a positive or negative charge by losing or gaining electrons. The electrode that carries electrons from the electrolyte into the external circuit is the anode, and the one that moves electrons into the electrolyte is the cathode. The key chemical changes occur at the electrode surfaces and are caused by the transfer of electrons between the electrodes and the ions.



ELECTROLYTIC CELL for electrochemical machining is depicted. The shaping tool, which is the cathode, and the piece being machined, which is the anode, are immersed in an electrolytic solution. The cathode and the anode are connected to a source of direct current and electrolyte is pumped between them as one electrode, in this case the cathode, is moved toward the other. Transfer of electrons between electrodes and ions in the electrolyte causes atom-by-atom removal of metal from the anode in a pattern established by the shape of the cathode. In a working cell the electrodes would be thousandths of an inch apart.

What happens in electrochemical machining is the atom-by-atom removal of metal by anodic dissolution. The piece of metal that is to be shaped (the workpiece) is made the anode and the metal tool giving rise to the shaping is made the cathode. They are connected to a source of low-voltage direct current. Then a strong electrolyte is pumped between the two electrodes. Without any physical contact of the tool and the workpiece, metal is removed from the workpiece in a pattern established by the shape of the tool as current flows between the electrodes. The rate at which metal is removed is not influenced by the hardness of the workpiece. Moreover, the process does not produce work hardening or induce stresses in the machined metal surface, which are both problems with conventional machining.

These properties of electrochemical machining give it an unparalleled advantage over conventional methods for machining the kind of metals commonly used in the aircraft and aerospace indus-



OIL PADS recessed only 250 millionths of an inch below the surface of this hydrostatic bearing were machined electrochemically. They are the hexagonal shapes spaced equally around the bearing surface. Pads as shallow and as smoothly finished as these could not be formed by grinding, which would be the conventional method of machining. Moreover, grinding or any other process involving direct contact between tool and workpiece would be costly and time-consuming, because the bearing is made of very hard steel.

tries, namely high-strength alloys designed to serve at high temperatures. Another advantage is that with a properly designed cathode the process can yield complex shapes that are almost impossible to achieve by grinding. An example is the cutting and finishing of a square blind hole (a hole that does not go all the way through the workpiece) in fully hardened steel. Electrochemical machining achieves not only a flat bottom but also exceptionally smooth machined surfaces (averaging .15 micrometer, or five microinches, in peak-to-valley roughness) with no scratches, burrs or grooves.

Machining time is often greatly reduced because the process eliminates the multiple operations required by conventional methods. For example, a job involving four or five conventional operations over a period of hours may require only one electrochemical machining step taking up a few minutes. Since no pressure is applied to the workpiece, thin foils and honeycombs can be machined without burring, warping, bowing or other distortions.

On the other hand, electrochemical machining may in certain circumstances be more expensive than conventional methods, since considerable skill and time are required to design and build the cathode and the cell to achieve the desired cut and to ensure a uniform flow of electrolyte. If many identical workpieces are to be machined, however, the cost of tooling can be recovered quickly because of the greater efficiency of the electrochemical machining operation and because there is virtually no tool wear. One other disadvantage of electrochemical machining is that in certain electrolytes the fatigue properties of the surface of some materials may be altered by chemical attack at grain boundaries (the boundaries of the tiny individual crystals of the metal).

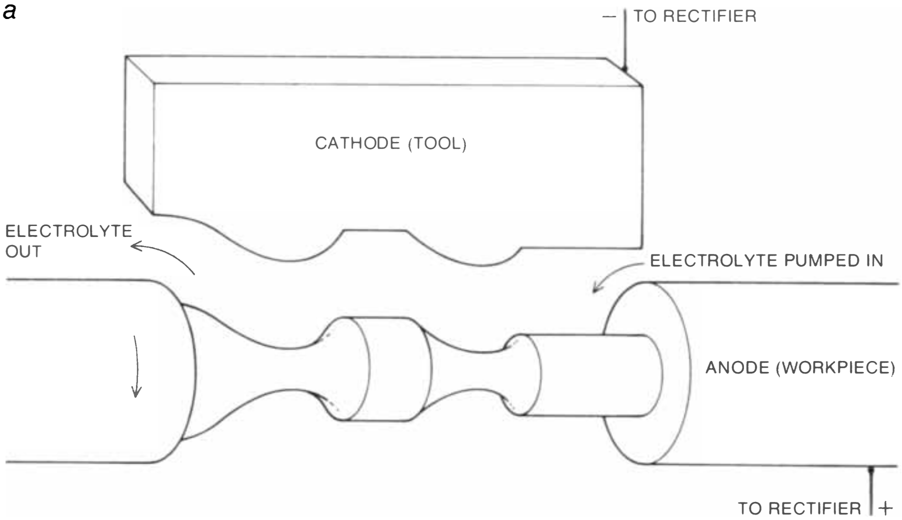
Employing electrochemistry in work on metal surfaces is hardly a new idea; electrochemical removal of thin layers from metal surfaces has been practiced for many years in the industrial processes of electropickling (in which the surface is cleaned) and electropolishing (in which the surface is cleaned and smoothed). Moreover, the concept of electrochemical machining was set forth in a British patent as long ago as 1929. The practice of electrochemical machining, however, is relatively new. One problem in the early efforts was with the electrolyte, which was usually sodium chloride. Electrochemical machining done with this electrolyte frequently yielded machined parts of insufficient precision.

Over a span of some 15 years our group at the General Motors Corporation Research Laboratories investigated the electrochemical machining properties of several hundred possible electrolytes. In 1967 we reported that a solution of sodium chlorate ($NaClO_3$) is a superior electrolyte for the electrochemical machining of ferrous metals from the standpoint of good dimensional control and excellent surface finish obtained at high rates of metal removal. In the light of these standards let us examine more closely how the process works.

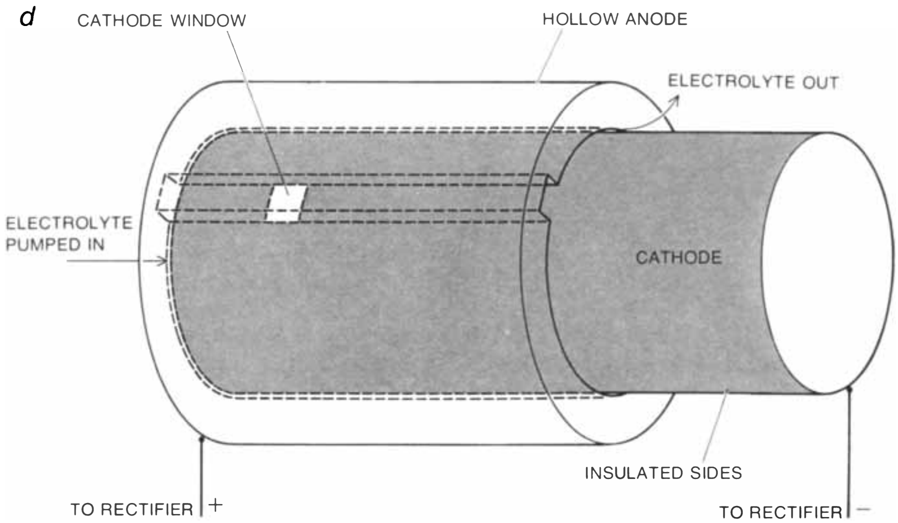
Suppose a sample of metal such as iron foil is put in a corrosive liquid such as a 10 percent solution of sodium chloride. Iron goes into solution as each atom of iron gives up two electrons to the metal phase and the resulting ferrous ion enters the solution phase. If this oxidation reaction (that is, this loss of electrons) were the only process taking place, a negative charge would build up on the surface of the metal until the electrostatic attraction between the negatively charged metal surface and the positive ferrous ions in solution overcame the driving force of the oxidation reaction. Metal dissolution would cease.

If corrosion is to proceed, there must be a sink for the electrons. Such a sink is provided by reduction processes, which come about in the following way. The sites on a metal surface are not equal-

a



d



SHAPING OPERATIONS that are possible with electrochemical machining include the six shown in these drawings. In lathing (*a*) the cathode is fixed in position while the rod-like anode is rotated below it, producing in the rod a shape established by the pattern of the sharp edge of the cathode. A cavity of complex shape can be machined in a block of metal in a single electrochemical operation called die sinking (*b*). External shaping (*c*) of a part

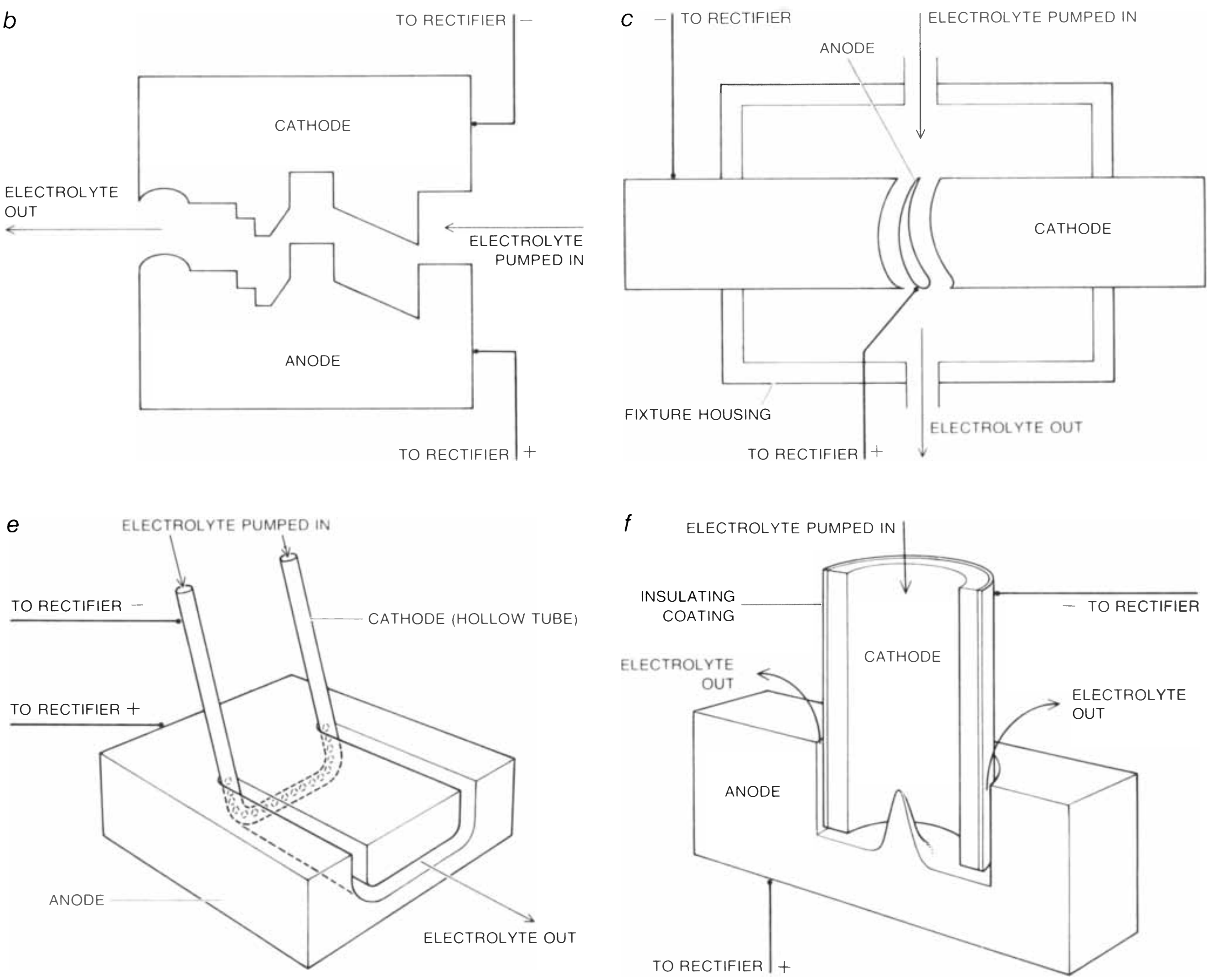
ly active electrically. As an oxidation process takes place at a high-energy site such as a peak, a reduction process (involving a gain of electrons) may occur at a low-energy site such as a valley. For example, molecular oxygen (O_2) dissolved in the liquid phase gains electrons and becomes water (H_2O) in an acid medium or a neutral one or becomes hydroxyl ions (OH^-) in an alkaline medium. Another example is the evolution of molecular hydrogen (H_2) by the discharge from the liquid phase of positive hydrogen ions (H^+) in an acid medium or of water molecules in a neutral or an alkaline medium.

Electrons flow through the metal from the oxidation sites to the reduction sites. The circuit is completed by the movement of ions through the liquid phase. This system is known as a local cell.

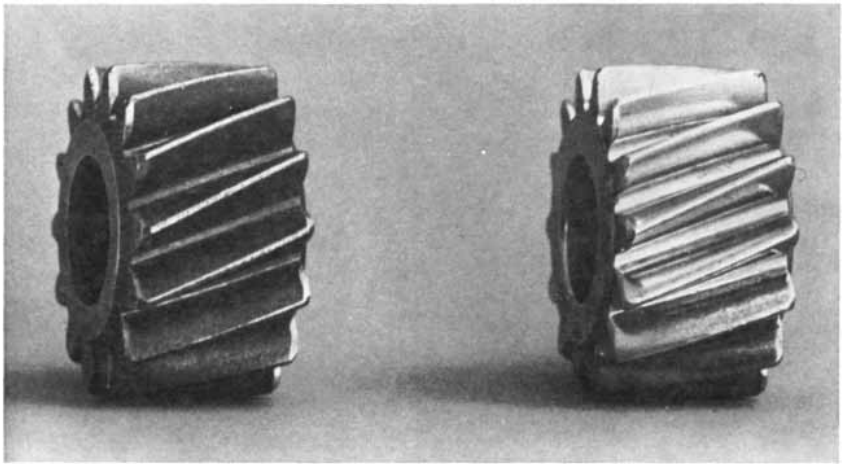
Since the surface of the metal has the same electrical potential at all points, the anodic and cathodic reactions have to proceed at this potential. It is called the mixed potential. To achieve anodic and cathodic reactions at the mixed potential the reactions in a local cell must be polarized from their respective equilibrium potentials to the mixed potential by the passage of current through the local cell. This polarizing current is called the corrosion or local-cell current. All wet corrosion of metal proceeds by a local-cell mechanism.

Most local-cell processes are slow because they are under cathodic control. They are also limited by the slow diffusion of reactants to the reduction (cathodic) sites on the corroding metal surface. If, however, one puts the corroding metal in contact with a "nobler" (less active) metal on which the cathodic reaction can proceed more easily, the corrosion current and hence the rate of dissolution of metal can be increased significantly. (This process is called bimetallic corrosion and must be guarded against in the design and construction industries. One often sees the result of bimetallic corrosion in the rapid and catastrophic destruction of a molding that is made of aluminum or magnesium and attached with iron or steel fasteners to the walls of a high-humidity area such as a bathroom.)

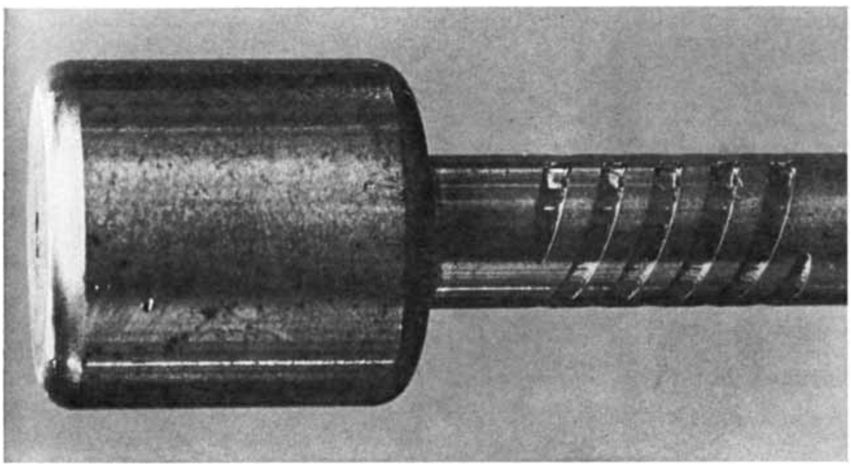
The corrosion current can be in-



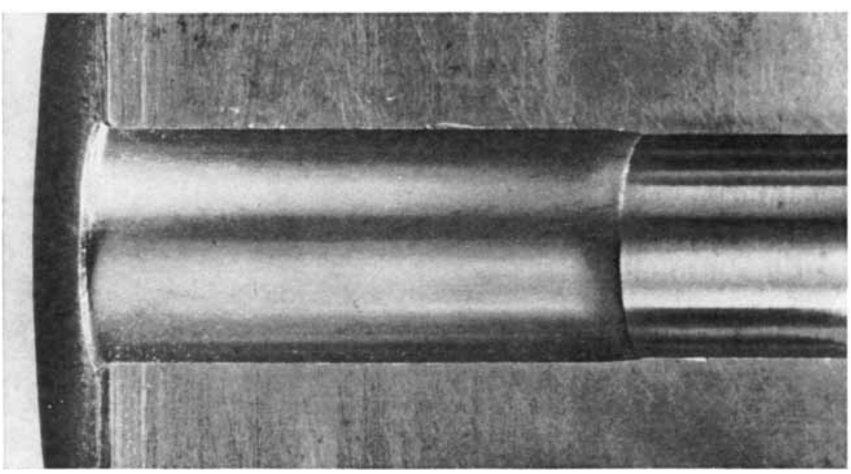
such as the blade of a turbine is done with a cathode shaped in the contours of the blade. With two cathodes both sides of the blade can be machined simultaneously. Internal grooving (*d*) involves a cathode in which only a small "window" is electrically active. If the tool is turned as it is passed through the inside of the workpiece, a helical pattern like the rifling of a gun barrel can be achieved. Wire cutting (*e*) removes a large piece of metal from the workpiece. The cathode is a heavy wire or a thin tube. Plunge cutting (*f*) employs a hollow tube as the cathode. Electrolyte is pumped through the cathode to cut a hole in a metal casting.



MACHINED GEARS were finished by two processes. Both gears were made by a gear-cutting machine, but the one at left was finished conventionally and the one at right was finished by electrochemical machining, a process that produced a much smoother surface.



LEAD SCREW designed for pumping oil or gas in a close-fitting seal was given its shallow threads by electrochemical machining. Grinding could cut the hard metal but could not make the triple-thread shape. Lathing could make the threads but could not cut the metal.



HOLE IN BUSHING was drilled originally by conventional processes. Then part of it was enlarged by electrochemical machining. The electrochemical process was able to produce a sharp shoulder, which in this sectioned bushing is the distinct line at right center of the hole. Electrochemical machining also produced a smoother bore than original drilling.

creased still more if an external potential is applied to the corroding metal, making it the anode of an electrochemical cell. The circuit is completed by a counterelectrode (the cathode) separated physically from the anode by an electrolytic solution. With this arrangement only the anodic process—that is, the oxidation process—operates at the corroding electrode. The rate of anodic corrosion can be increased or decreased by raising or lowering the applied electrical potential.

For electrochemical machining to be commercially competitive with conventional machining techniques, which remove metal at a rate of about two-tenths of a cubic centimeter per second, extremely high current density is required. Typically the current is between 50 and 500 amperes per square centimeter. In practice the upper limit to the density of current that can be achieved is fixed by the availability of high-capacity rectifiers, which convert alternating current to the direct current required by the system.

With the passage of such high currents the path the electrolyte follows between electrodes must be kept short to prevent an intolerably high voltage from developing. The gap between electrodes typically ranges from .005 to .13 centimeter. In other words, the tool and the workpiece are only a matter of thousandths of an inch apart.

As the electrochemical machining process continues, metal is removed from the anodic workpiece, widening the gap and therefore reducing both the current and the rate at which metal is removed. In order to compensate for these variations the distance between the anode and the cathode must be maintained by a mechanically driven feed system. One electrode is held stationary and the other is moved.

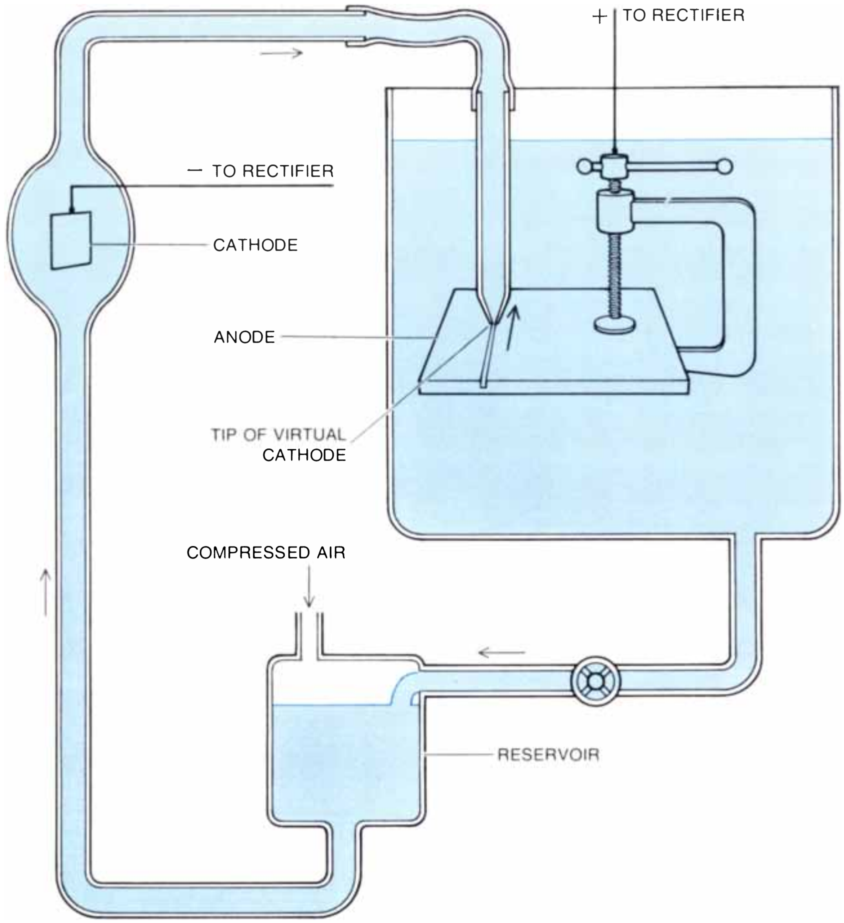
If the moving electrode advances too rapidly, it can short-circuit the anode and the cathode, causing catastrophic spark-erosion damage to the cathodic tool and melting, welding and mechanical tearing of the workpiece. To prevent such a costly accident protective sensing devices are typically employed to shut down the current within microseconds when the gap closes to less than a specified safe value. The devices usually work by detecting a rapid increase in current.

For a given rate of feed and a given voltage applied across the anode and the cathode the system rapidly comes automatically to an equilibrium gap distance. Consider a system where the rate of feed is causing the gap to diminish. As the gap narrows, the density of current increases because of a decrease in what electrical engineers call the "*IR* drop" (after Ohm's law, which states that the current I times the resistance R equals the electromotive force E, which is also called the potential or the voltage). The increased current, however, produces an increase in the rate at which metal is removed, thereby tending to increase the gap between electrodes. The opposing forces—feed rate tending to narrow the gap and current density tending to widen it—soon balance to achieve the equilibrium gap between the tool and the workpiece.

Since the electric current flowing through the system is large, the electrolyte tends to heat up. Indeed, if the electrolyte were allowed to remain stationary in the gap between the electrodes, the heat would soon vaporize the liquid and bring the process of metal removal to a halt. Therefore the electrolyte must be pumped through the gap at a high rate of flow.

Here, however, another problem arises. The increase in temperature from the time the electrolyte enters the gap to the time it leaves has been found to be about 45 degrees Celsius (113 degrees Fahrenheit). A rise this large can cause an increase of as much as 100 percent in the specific resistance of the electrolyte and 50 percent in its viscosity. Anything that affects the conductivity of the electrolyte, as these changes do, alters the current density and therefore affects the rate at which metal is removed. To prevent uneven removal of metal at hot spots in the gap the rise in temperature as the electrolyte moves through the gap should be kept below 10 degrees C. Therefore the flow line of the electrolytic system is ordinarily equipped with a heat exchanger.

The temperature rise can be reduced by increasing the rate of flow of the electrolyte. An excessive rate of flow, however, can lead to a pressure gradient along the gap that is high enough to bring about cavitation, that is, spontaneous formation of bubbles of vapor in the solution at points of negative pressure. Cavitation breaks up the solution and also produces a general roughening of the machined surface. At high rates of flow cavitation can be reduced and even eliminated by proper design of the components of the system. In any case the rate of flow of electrolyte must be uniform in order to achieve a uniform removal of metal.



JET-ETCHING ASSEMBLY is portrayed schematically. If the cathode is held still, the setup serves for drilling holes in a metal panel. If the cathode is moved, a groove is etched in the panel, as shown here. The apparatus forces a jet of electrolyte solution against the panel with compressed air. Splash container collects electrolyte, which is then recycled.

Another reason for keeping the rate of flow high is to remove corrosion products from the gap. If the electrolyte becomes contaminated with grains of metal or other substances, the rate of flow becomes uneven, the distribution of current becomes nonuniform and the surface finish is poor. To avoid this problem it is best to fit the flow line of the electrolyte with a filter.

Electrochemical machining is a highly versatile process, so that a considerable variety of machining operations can be accomplished with it [*see illustrations on pages 32 and 33*]. For example, external shaping can machine turbine blades for the jet engines of aircraft. The cathode tool is designed to fit the contours of the finished blade; with a pair of shaped cathodes both sides of the blade can be machined simultaneously.

Highly complex cavities can be machined quickly (and with acceptable surface finish) in fully hardened blocks of metal by a single electrochemical machining operation known as die or cavity sinking. The tool is fashioned in the shape of the desired cavity. The shaped cathode is moved toward the anodic block of metal while an electrolytic solution is pumped between the electrodes. As the tool approaches the surface of the workpiece, metal is removed from the high spots first and from lower spots with deeper penetration of the tool. When the tool has completely entered the workpiece, the current becomes relatively constant at all points.

A process known as plunge cutting is employed to cut holes in a metal casting. The cathode tool is a hollow tube through which the electrode is pumped; its shape is determined by the cross section of the desired hole. Since metal is to be removed only at the bottom of the hole, the sides of the tube are electrically insulated with a nonconducting film such as a coating of epoxy resin. Stray current invariably causes a certain amount of tapering of the hole, but the problem can be minimized by proper design of the cathode and a proper choice of electrolyte.

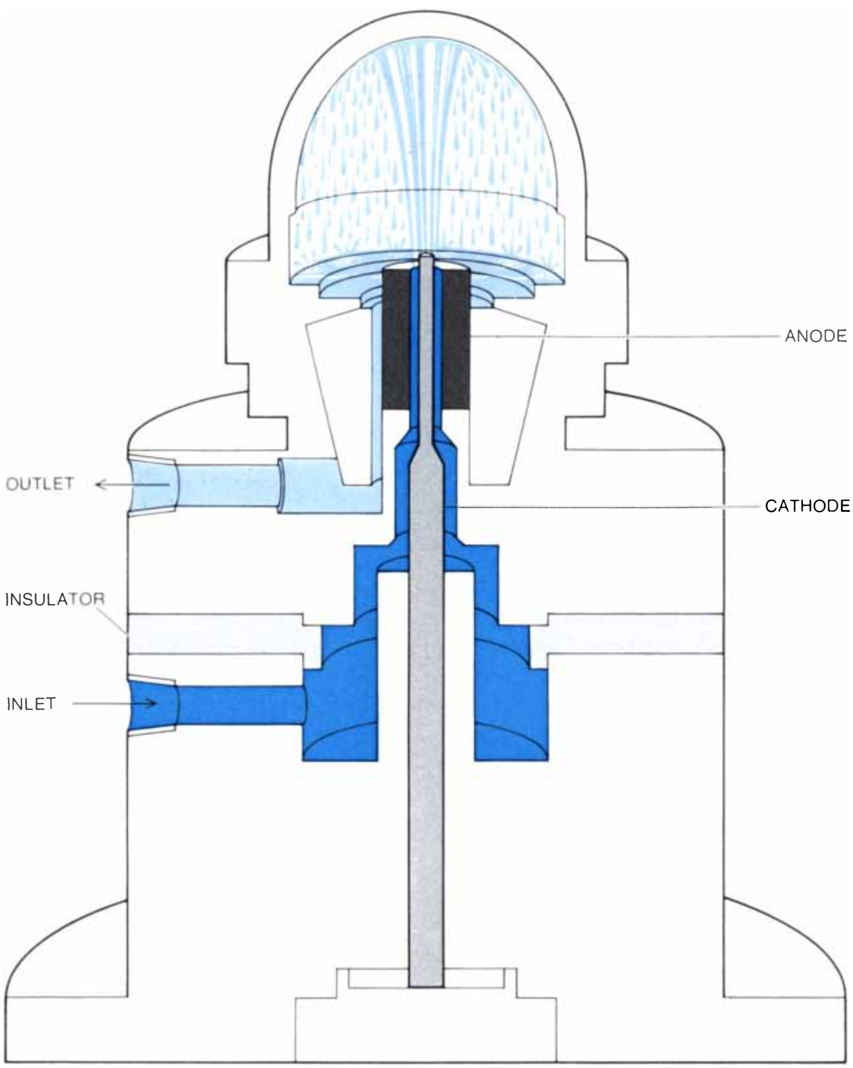
The nature of the process is such that the circumference of the bottom of the hole is machined more rapidly than the interior, so that at all times a small Vesuvius—a convex slug of metal—rises from the floor of the hole. It is possible for this slug to topple against the inside wall of the cathode, causing a short circuit. To avoid this hazard and also to prevent a loss of pressure in the electrolyte when the tool has worked its way completely through the casting, a false or backup workpiece can be attached to the bottom of the anode. The metal that will form the slug is held to the backup plate by a screw.

Electrochemical machining can also serve in lathing, trepanning, internal grooving and wire cutting. In lathing the rodlike workpiece is turned under a knife-edge cathode, which remains in a fixed position, to achieve symmetrical contouring of the rod. In trepanning a hollow cathode shaped according to the specified pattern lifts parts of complex shape and uniform thickness from a metal slab. Internal grooving is achieved by passing a thin or pointed cathode down the inside of a hollow cylinder that is connected as the anode. If the cylinder is rotated, helical grooving of the internal surface (similar to the rifling in a gun barrel) is achieved. Wire cutting is an operation that roughs out a large volume of metal. A heavy wire or a thin metal tube is passed through the workpiece. The electrolyte is streamed over the workpiece or through the tube and out of fine holes in the tube cathode.

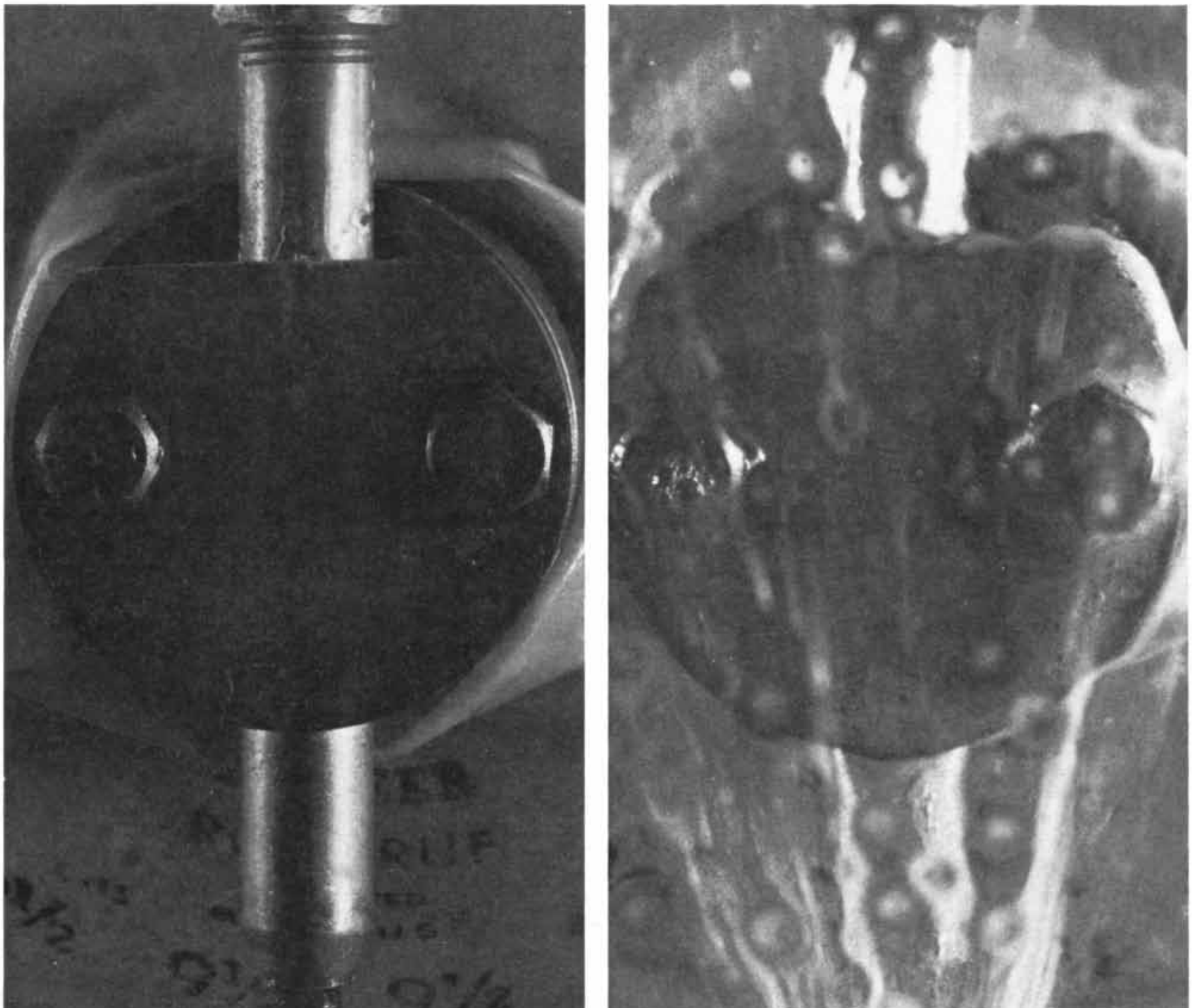
A new application made possible with the development of sodium chlorate as an electrolyte is embossing of both relatively fine and somewhat gross patterns on finely finished steel surfaces. For example, we have machined oil pads in bearings with excellent results in terms of geometry and depth control [*see illustration on page 31*]. In a typical operation bearing pads were electrochemically machined to a depth of .00062 centimeter (.00024 inch) with a variation of flatness at the bottom of .000062 centimeter (.000024 inch). Machining such designs by conventional methods is costly and time-consuming. Indeed, conventional techniques are confined to operations where dimensions are no shallower than .0013 centimeter (.0005 inch). A similar type of electrochemical machining can be employed to cut threads in a blank steel shaft in the production of lead screws. Once the proper alignment and spacing of the tool and the workpiece have been achieved, thousands of identical parts can be produced daily.

A limitation of electrochemical machining operations is that they cannot be employed to machine large areas, which would require unobtainable amounts of electricity. With a rectifier having a capacity of 10,000 amperes only 260 square centimeters (40 square inches) can be machined at a current density of 40 amperes per square centimeter (250 amperes per square inch). With a limited-area cathode that travels past the anode, however, metal can be removed from large areas without the need for large line currents. An example of this process is the electrolytic broach machine, in which the inside diameter of a workpiece is sized and finished by passing a cathode through the anode. We have been able to move a cathodic "window" through an anodic area 51 centimeters (20 inches) long in 20 seconds—a machining rate that is commercially practical. The outside diameter of a cylindrical casting can be sized and finished by passing a cathode of toroidal shape over it.

In the light of the applications we have described and of the development of the limited-area-cathode principle, we can predict a bright future for electrochemical machining. There is no doubt that electrochemical machining in its various forms can perform many manufacturing operations better and more economically than present methods do. We recommend that manufacturers develop their own body of knowledge and experience in this field so that the unusual capabilities of the process can be exploited.



ELECTROCHEMICAL MACHINE set up to size and finish a bushing is shown in cross section. The bushing is the anode. Electrolyte is pumped over the cathode and through the bushing. A splash shield covers the top of the machine, and the draining electrolyte goes through the outlet to a reservoir for recycling. This machine, which was built for experimental work in the authors' laboratory, is approximately 12 inches high and 10 inches wide.



HOLE IS DRILLED in a block of metal by an electrochemical machine in the authors' laboratory. At left the cathode tool is poised above the anode workpiece before drilling starts. The white coating on the tool is an insulating layer of epoxy resin that confines the cutting to the bottom edge of the tool. At right electrolyte fans out of the hole being cut in the workpiece. The electrolyte is pumped through center of cathode at high pressure. A gauge attached to the machine records the penetration of the cutting tool.

The Nervous System of the Leech

A repetitive array of ganglia with a few cells constitutes this simple central nervous system. Study of the cells' properties and connections explains certain behavioral reflexes and shows how they are modified

by John G. Nicholls and David Van Essen

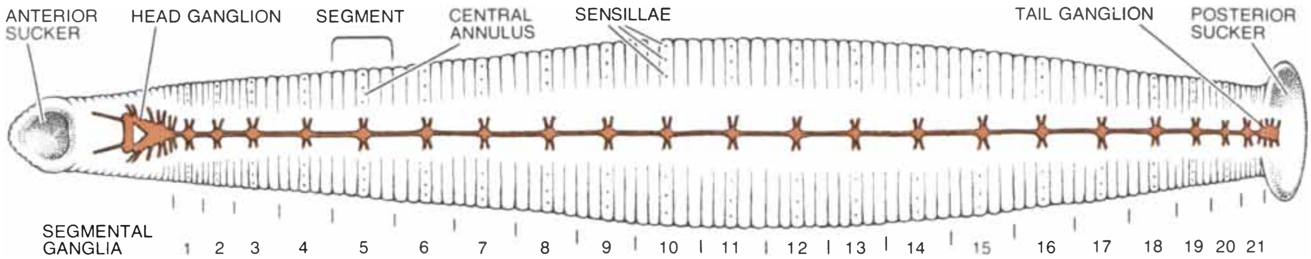
The signals with which nerve cells transmit information are essentially the same throughout the animal kingdom. As a result many problems in nerve physiology have been approached successfully through the study of animals with particularly large nerve cells that are amenable to experimentation. The mechanisms of the conduction of the nerve impulse and the active transport of ions into and out of nerve cells, for example, were first elucidated in the large peripheral nerves of the squid. The principles learned in these and other invertebrates were later found to be essentially identical with those that govern the same processes in the more complex nervous systems of vertebrates, including mammals and even man; rather than being a roundabout way, the approach by way of invertebrates proved to have been a shortcut. When neurophysiologists started to explore how interactions among large numbers of cells lead to coordinated behavior, it was therefore natural to look for experimental material among the invertebrates, including the leech. Why the leech?

If an investigator interested in studying general principles of neural organization were asked to design the ideal experimental animal, he would probably have no trouble prescribing its characteristics. The brain should contain relatively few cells, perhaps only a few hundred, so that the functional properties of each cell could be analyzed in detail. The individual cell bodies should be large enough to be impaled with microelectrodes and to be studied by biochemical techniques, yet the entire brain should also be small enough to be sectioned serially for electron microscopy. The synaptic connections between cells should be sufficiently intricate to produce complex behavior in the animal and yet simple enough to be traced electrically and anatomically, and they should be subject to modifications that alter the animal's behavior. Finally, the reproductive cycle of this all-purpose animal should be short enough to allow study of the genetic factors responsible for the blueprint that is followed during development.

Unfortunately there is no nervous system with all these qualities, some of which are mutually exclusive. And so one has to make do with different experimental preparations for studying various aspects of the nervous system's function. For example, the relatively small brain of the fruit fly *Drosophila* and of certain small nematode worms are being used to analyze the genetic factors that determine the formation of neural connections. Certain mollusks and crustaceans, which are not amenable to genetic studies because of their slow development, have particularly large neurons that are very useful for investigating the biochemistry of individual cells and the physiological connections that underlie reflex behavior. We have found that the leech provides a convenient system in which to explore a variety of problems, including neural specificity, synaptic transmission, integration, plasticity of behavior and regeneration of connections. The experiments described here were started in 1966 at the Yale University School of Medicine with Denis A. Baylor, who played a crucial role in formulating and defining the problems and techniques, and were then continued at Yale and at the Harvard Medical School by Ann Stuart, Dale Purves, Jan Jansen, Kenneth J. Muller, Damien Kuffler, Gunther S. Stent and us.

Our approach has been to analyze the nervous system cell by cell. The leech has only a small number of nerve cells performing any particular function. For example, one sensory cell innervates a large patch of skin that in the mammalian nervous system would be supplied by many cells. As a result one can selectively activate a single sensory nerve cell by natural stimulation and thereby set up a simple reflex. The chain of events that this stimulus to the animal sets in motion can then be followed sequentially through the nervous system, again by looking at individual elements. Much of the initial effort has been to identify the sensory and motor nerve cells that take part in certain reflexes and then study the behavior of these neurons and their electrical characteristics when they are active. Such preliminary but precise knowledge is essential before one can turn to the problem of how the performance of a neuron is modified by its previous history of activity and what the consequences are for signaling in the central nervous system. Another dividend of knowing the connections in detail is that one can study how individual nerve cells regenerate to form connections after their fibers have been severed by a lesion. This raises the important question of neural specificity, which in turn has a bearing on how one thinks about the genetic factors involved in laying down the nervous system during development.

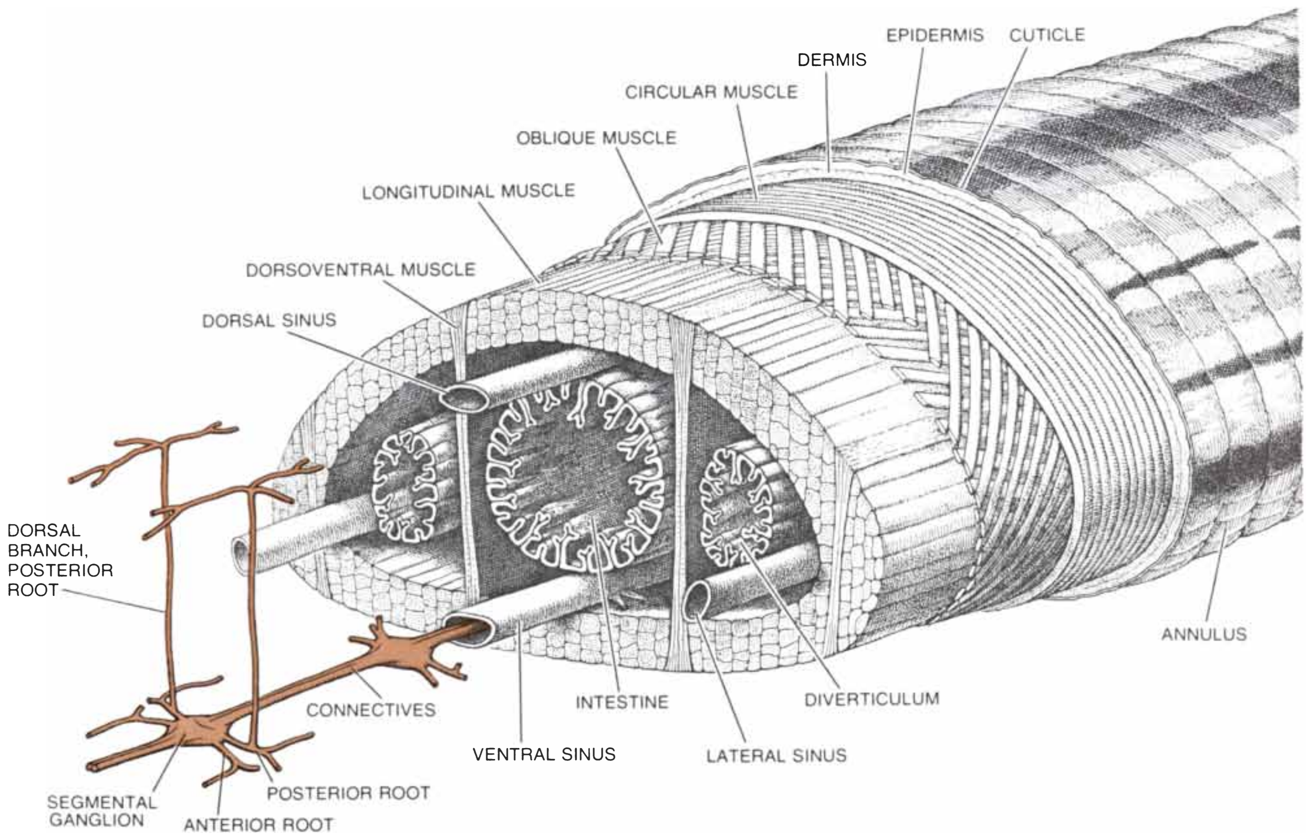
In the nerves of both higher and lower animals it is the cell membrane that, through its selective permeability to different ions, enables the nerve impulse to be set up and to propagate. Many experiments have shown that nerve cells in invertebrates are very similar to those in our own bodies with respect to both structure and signaling. In each instance the neuron produces electrical potentials about a tenth of a volt in amplitude, lasting for a small fraction of a second. Like the brief dots and dashes of Morse code, these are stereotyped signals that con-



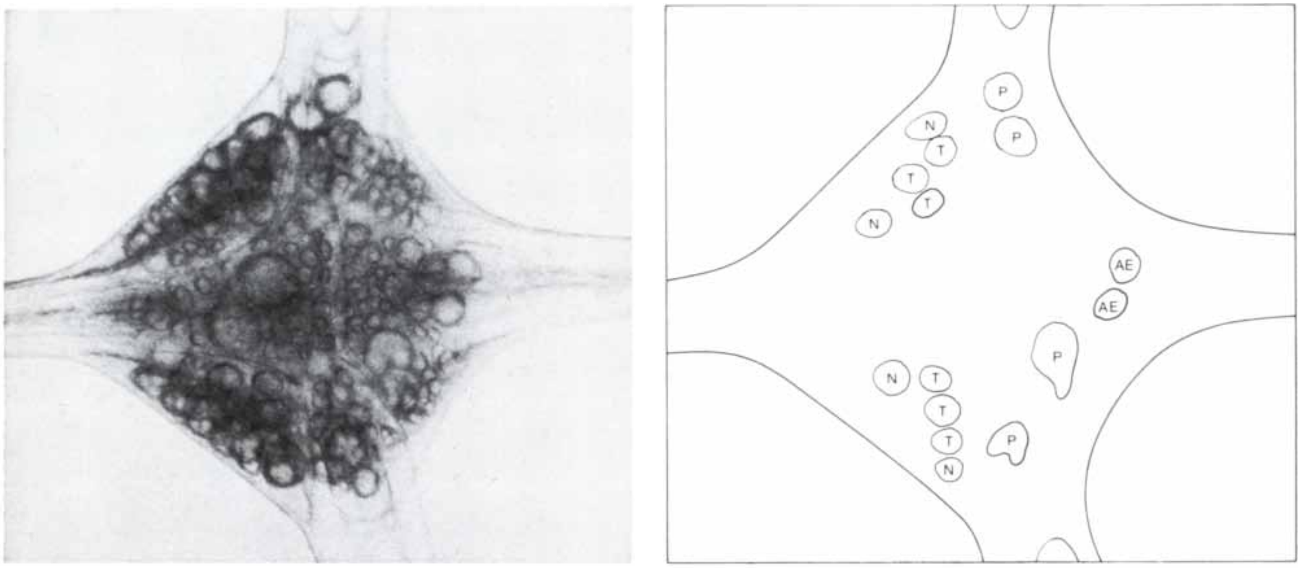
MEDICINAL LEECH ***Hirudo medicinalis*** **has a body consisting of a number of segments that are almost alike except at the two ends. The surface of most segments is divided into five circumferential rings, the annuli; the central annulus of each segment is marked by small sensory organs called sensillae. The nervous system (*color*), which lies along the ventral side of the animal shown in this view from below, consists of a series of stereotyped ganglia, one for each segment. There are two "brains," one at each end, made up of fused ganglia. The suckers are for locomotion and the one at the head also serves for attachment to blood donors.**

vey information from one end of the cell to the other. Information is passed on from one cell to the next at the synapses, where many incoming axon fibers converge on a postsynaptic target cell. In general, transmission across the synapse is effected by the release from the presynaptic terminals of a specific chemical substance: the transmitter. Some transmitters excite the next cell, causing it to emit impulses; others tend to inhibit it, damping its activity. Both excitatory and inhibitory synapses are found in all nervous systems, and their mode of operation is similar in principle. Cells can also interact through electrical synapses, where current flows directly from one cell to the next without the intervention of a chemical transmitter.

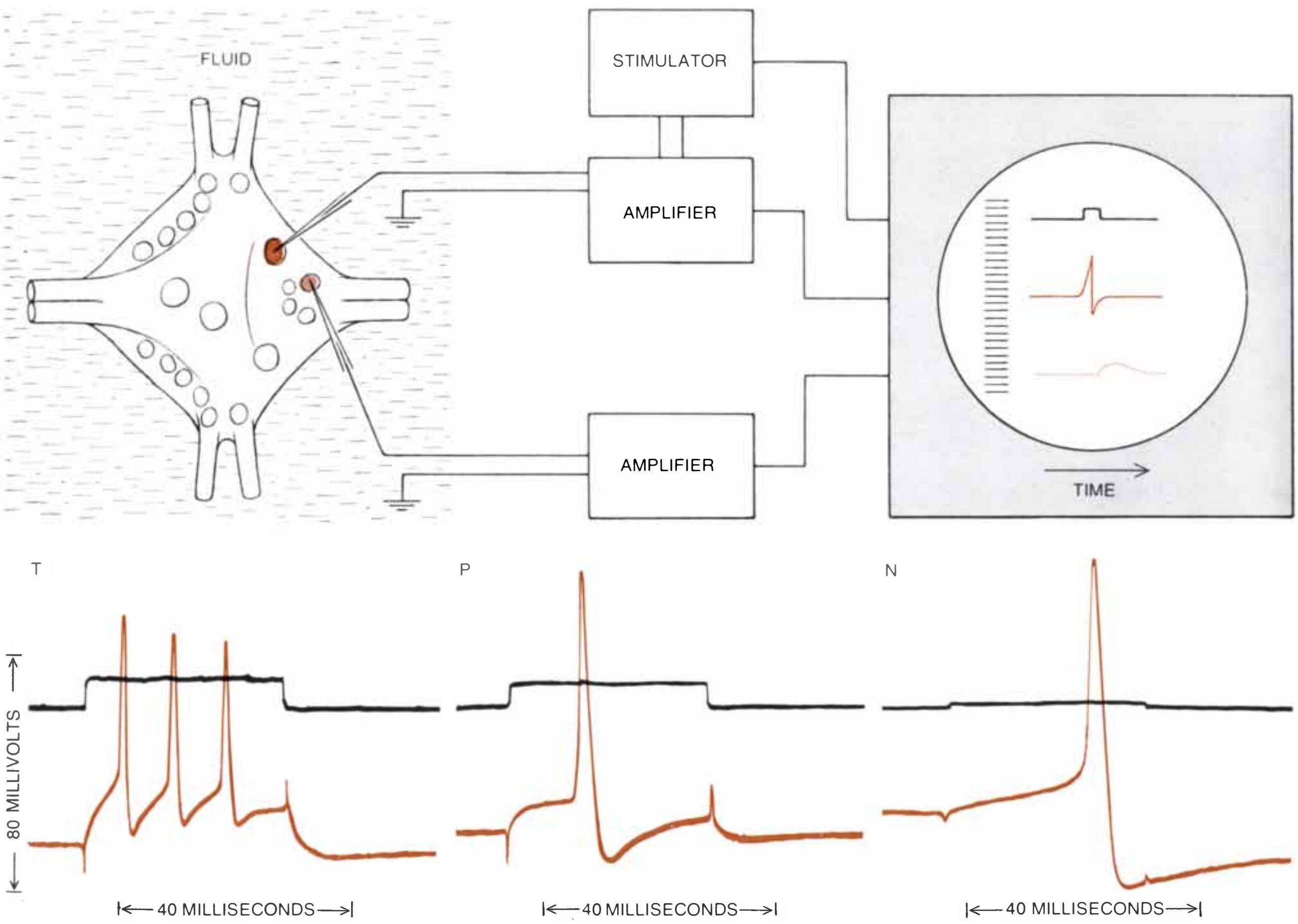
The principal advantages of the leech arise from the simplicity of its body plan, which is reflected in the structure of its nervous system. Both the body and the nervous system are made up of a number of very similar repeating units, and behavior consists of a relatively simple repertory of movements performed by well-defined groups of muscles. Each segment of the body is innervated by a ganglion. (Even the specialized "brains" at the head and tail are simply fused ganglia in which many characteristic features can still be recognized.) The ganglia are similar within one animal and also from one animal to the next. Each ganglion contains only about 350 nerve cells, which have characteristic shapes, sizes, positions and branching patterns. A ganglion innervates a well-defined territory of the body by way of its roots, which are paired bundles of axons, and it communicates with neighboring and distant parts of the nervous system through axon bundles called connectives. A segment and its ganglion can



LEECH'S NERVOUS SYSTEM (*color*) is enclosed in a blood vessel. The ganglia are linked by paired connectives, each serving one side of the ganglion; from each ganglion bundles of nerve fibers extend through two roots on each side to supply the skin and muscles. The body wall contains a set of layered muscle sheaths. Contraction of the circular muscles elongates the animal. The longitudinal and oblique muscles bend or shorten the body; the dorsoventral muscles flatten it. The diverticuli are large outpocketings of the intestine. An adult leech is generally about 1/4 inch wide and one to four inches long depending in part on time of its latest meal.



SEGMENTAL GANGLION is enlarged about 85 diameters in the photomicrograph at left and is mapped in the diagram at right. The outlines of individual cells are seen clearly; from them fibers extend in paired connectives to neighboring ganglia and, through the roots, to the skin, viscera and muscles. The seven sensory cells on each side of the ganglion include three touch cells (*T*), two pressure cells (*P*) and two nociceptive cells (*N*), which respond respectively to a light touch, stronger pressure and still stronger noxious stimuli. Two motor nerve cells, the annulus-erector cells (*AE*), are also labeled. This is the ventral, or bottom, surface of the ganglion; the "large longitudinal" motoneurons, which are also involved in some of the authors' experiments, are on the dorsal surface.



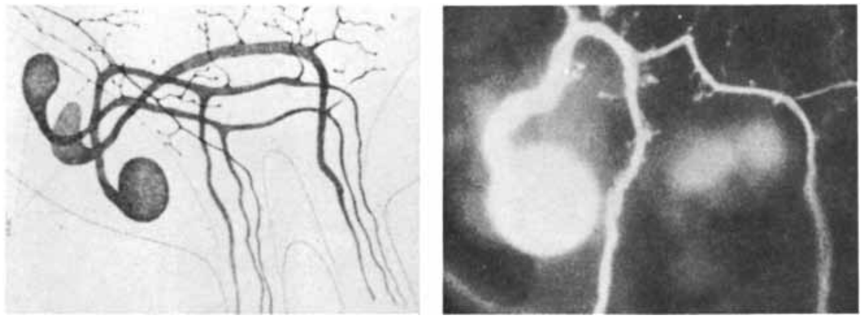
INDIVIDUAL CELLS are stimulated, and their potentials recorded, with fine glass electrodes. With the setup shown here a cell can be stimulated by injecting a current into it; the stimulated cell produces an impulse (*dark color*), which is recorded through the same electrode. The impulse may also cause a cell with which the stimulated one connects to produce a synaptic potential (*light color*), which can also be recorded. The three records (*bottom*) show the characteristic impulses of the three kinds of sensory cell: touch, pressure and nociceptive. The raised portion of the black traces shows the duration of the stimulating current. The colored traces are the impulses; each is an oscilloscope graph of the potential across the cell membrane in millivolts, plotted against time.

be studied alone or together with several adjacent segments.

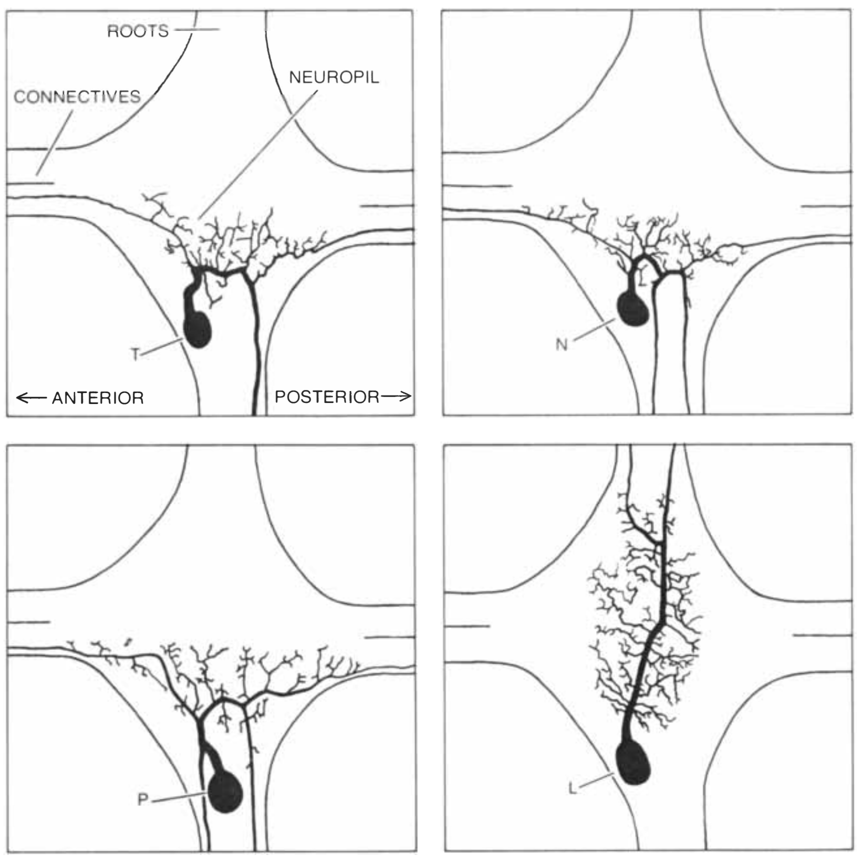
Many distinguished anatomists have studied the nervous system of the leech in great detail. In the beautiful drawings the Swedish anatomist Gustaf Magnus Retzius made more than 80 years ago he not only recognized individual cells but also grouped them into classes on the basis of morphological features. Later it was confirmed that his groupings had real functional validity, so that Retzius' drawings are still useful for physiological studies. The three cells in one Retzius drawing, for example, have distinctive electrical membrane properties and a specific functional role: they are the sensory cells that respond to a light touch applied to well-defined areas of skin [*see top illustration on this page*]. Two other kinds of sensory cells respond specifically to pressure and to noxious stimulation of the skin. Each type of cell—touch, pressure and "nociceptive"—can be recognized by morphological criteria, and the identification can be clinched by recording its membrane's electrical properties. The technique is to insert a fine electrode into a cell, stimulate the cell by passing a current into it and then record the electrical impulses the cell emits. These action potentials, in the case of a touch cell, are similar to but smaller and briefer than those in the pressure or the nociceptive cells. Nociceptive cells are also distinguished by a particularly large transient increase in membrane potential (called a hyperpolarization) that follows the impulse. In motor nerve cells the impulses and electrical properties are again different and quite distinctive. With practice one can look at a single action potential and usually be certain which kind of cell the electrode is in.

The branching patterns of a cell can be traced by injecting the fluorescent dye Procion yellow or by stimulating and then recording from its axons in roots and connectives. In addition to the major branches running through roots and connectives there are many such processes within the neuropil, the region in the center of the ganglion where synaptic connections are made. Although the structure is highly complex, the fingerprint of each cell is distinctive and characteristic. One can determine the pattern of the synaptic connections between cells by stimulating and then recording with intracellular electrodes. Although the synaptic potentials arise at a distance from the cell body, within the intricate neuropil, they are close enough to be recorded at the cell body.

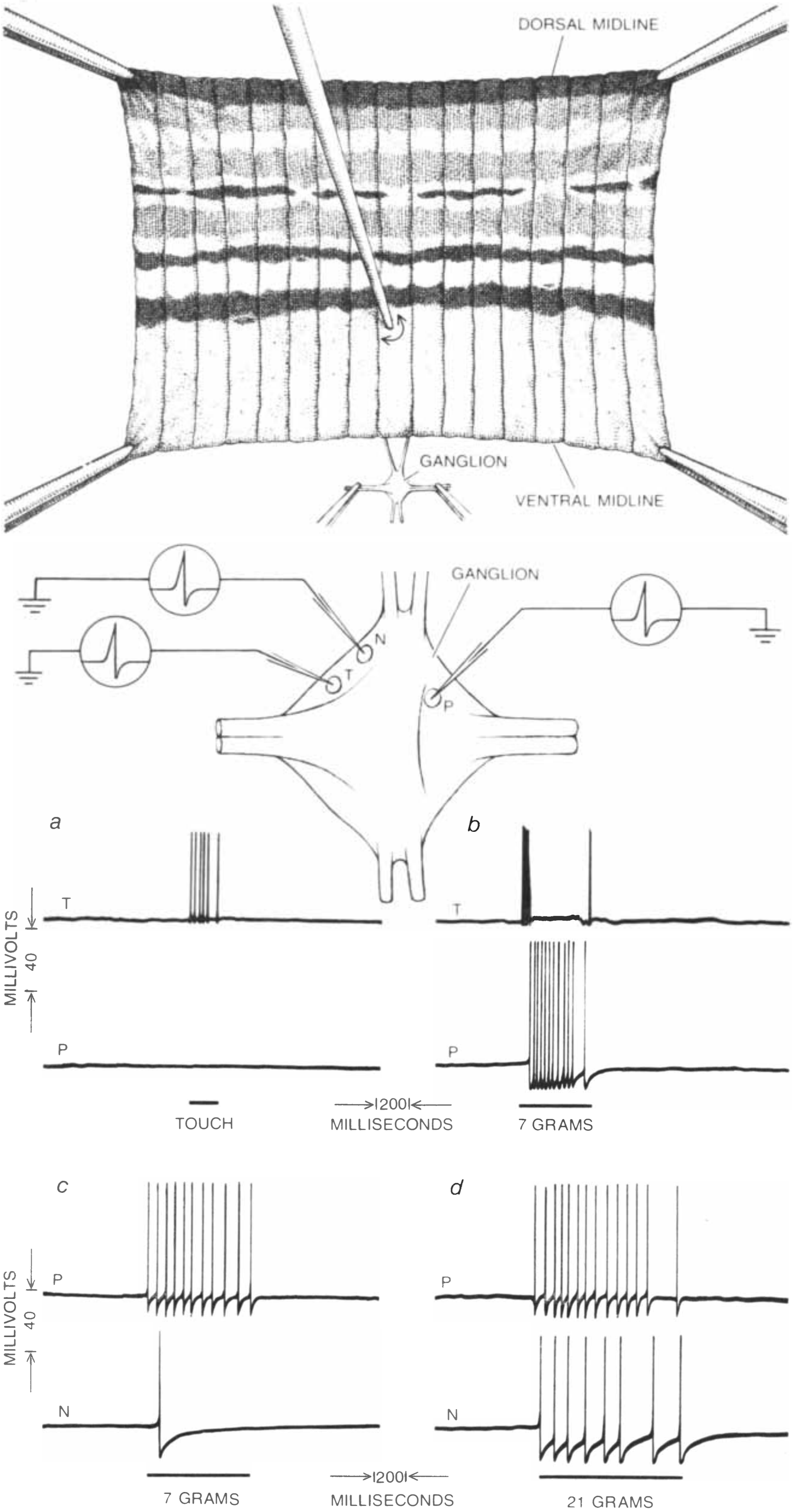
A convenient preparation for identifying the sensory and motor cells in a ganglion is made by removing part of the body wall of the leech and the ganglion or ganglia innervating it. Contractions of the musculature can still be elicited by sensory stimuli, and the preparation survives well in a physiological salt solution for more than 12 hours. In such a preparation one can study simple reflexes in considerable detail. For example, when one touches or presses the skin of the leech, it responds by shortening first the affected segment and then adjacent segments. This reflex resembles in several respects the familiar knee jerk in man, but in the leech one can hope to analyze all the nerve cells that play a



TOUCH CELLS are seen in a drawing made in 1891 by the Swedish anatomist Gustaf Magnus Retzius (*left*) and in a preparation made in the authors' laboratory (*right*). The Retzius drawing shows three touch cells and their branching processes, which run through the roots, through connectives and toward the center of the ganglion to form synapses with other cells. In the ganglion on the right one such touch cell was injected with the fluorescent dye Procion yellow. The similarity of its branching processes to those of one of the Retzius cells illustrates the high degree of precision in the neuronal geometry of the leech.



PROCESSES OF THREE SENSORY CELLS and of a motor nerve cell are shown. The drawings are based on serial sections of ganglia in which individual cells had been stained with Procion yellow. The touch (*T*), nociceptive (*N*) and pressure (*P*) cells send fine processes into the neuropil, the area in the center of the ganglion where they form synapses with other cells—including the large longitudinal motoneuron (*L*). The *L* cell sends its axon across the ganglion to innervate the longitudinal muscle on the opposite side.

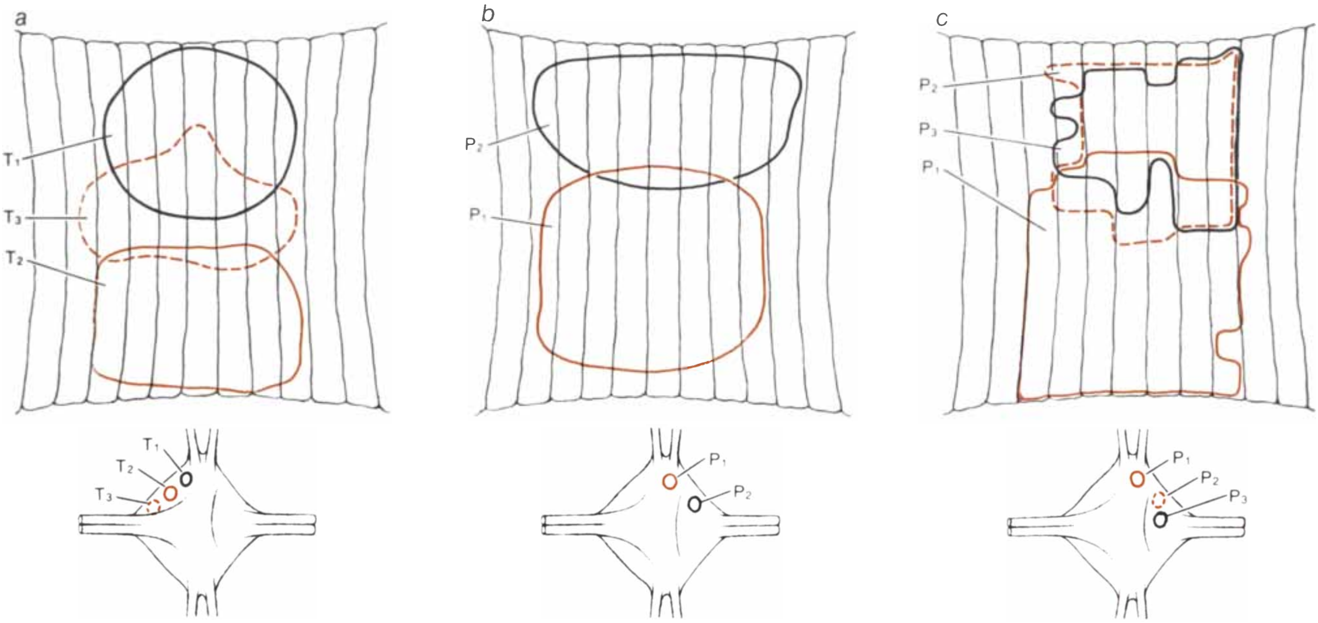


RESPONSES OF SENSORY CELLS to stimulation of the skin are compared. The experimental preparation is a segment of the body wall to which a ganglion remains attached by its roots. When simultaneous recordings were made from touch and pressure cells while the skin was touched lightly, the touch cell fired but the pressure cell did not (*a*). Pressing with a force of seven grams caused a maintained discharge in the *P* cell and brief, adapting discharges in the *T* cell (*b*). When pressure and nociceptive cells were compared, the seven-gram stimulus fired the *P* but not the *N* cell (*c*); a 21-gram stimulus fired both cells (*d*).

part first in the convergence of information from the skin to the central nervous system and then in its movement back out toward the muscles. First we shall describe the motor cells that innervate muscles and produce movements and then the sensory cells that converge on them and are activated by stimulation of the skin.

To determine whether a particular cell is a motoneuron, the first thing to do is to stimulate it through an intracellular electrode and see if some part of the body wall contracts. If movements occur, a number of additional tests must then be made to confirm that the axon of the stimulated cell synapses directly on the muscle fibers. (Many cells that have quite different functions can produce contractions indirectly by activating motoneurons.) So far 17 pairs of motoneurons that meet these criteria have been identified. The cells of each pair lie symmetrically on the two sides of the ganglion. These 34 cells probably constitute most of the motoneurons in the leech, which performs only a limited number of simple movements. These include shortening its body in response to cutaneous stimuli, swimming, twisting and walking like an inchworm by using its suckers. The main muscles that execute these movements are arranged in three layers. Directly under the skin there are circular fibers that contract to elongate the animal. Next come crisscrossed oblique fibers and, deeper still, powerful longitudinal fibers used for shortening and bending. In addition there are groups of muscles that flatten the animal and that erect bumps (like goose pimples) on the annuli, the circumferential rings on the animal's skin. Each motoneuron innervates only one of the muscles described above and, in each muscle, only a fraction of the muscle fibers.

We shall consider one motor cell in detail, the "large longitudinal" motoneuron. It is on this *L* cell that sensory cells of different modalities converge to produce shortening reflexes in the animal. The axon from an *L* cell runs to the body wall on the opposite side, where it innervates longitudinal muscle fibers of the segment on that side to produce a rapid shortening of the segment. The paired *L* cells on each side of the ganglion are coupled electrically, so that their impulses are synchronous and they act as a unit. This is a good example of the functional design of cell connections: the coupling coordinates the two sides and ensures that the body shortens evenly. The *L* cell is not used in swimming or



RECEPTIVE FIELDS on the skin of normal touch and pressure cells overlap slightly. The fields of each of three touch cells were mapped by touching the skin and marking positions from which responses were obtained (*a*). Cells were driven most effectively by touching at the center of their fields. A similar experiment mapped the fields of the two normal pressure cells (*b*). An abnormal ganglion with three pressure cells instead of two on a side was studied (*c*). Careful mapping revealed that the receptive fields of two of them (P_2 *and* P_3) overlapped almost completely; in this case it appears that two cells are doing the job normally done by one cell.

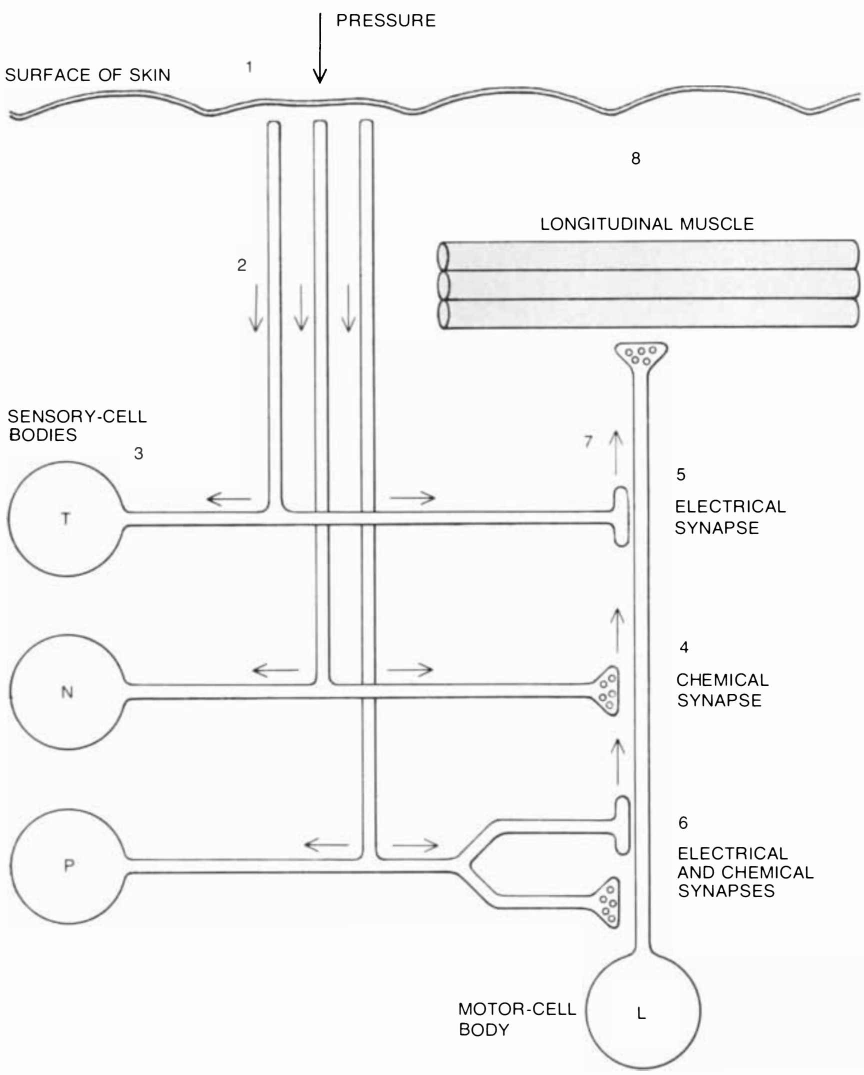
to bend the animal upward, sideways or downward; these movements are achieved by other motoneurons that innervate narrow bands of longitudinal muscle fibers on the back, sides or bottom of the animal.

The sensory input to the segmental ganglia can also be readily analyzed. The touch (*T*) cells fire in response to a light touch on the skin surface. Indentations of .05 millimeter or less are effective; indeed, *T* cells are so sensitive that they respond to the movement of fluid over the skin. The sensory discharge of a *T* cell in response to sustained indentation slows down rapidly and usually ceases within a fraction of a second, but a maintained discharge of impulses can be elicited if the tactile stimulus is moved back and forth within the receptive field: the area of skin over which the cell responds to touch. The pressure (*P*) cells respond only to a marked deformation of the skin and their discharge is maintained, often for 10 to 20 seconds or more, if the pressure is sustained. The frequency of the discharge is proportional to the extent of the indentation. Light touch is ineffective in driving these cells, whereas a pressure of several grams applied through a blunt stylus is enough to produce a high-frequency discharge. The nociceptive (*N*) cells require still stronger mechanical stimuli. The highest-frequency and best-maintained discharge is evoked by a radical deformation, such as pinching the skin with forceps or scratching it with a pin. The *N* cells, like the pressure cells, give a steady discharge, and they often continue to fire after the stimulus has been removed.

The three kinds of sensory cells—14 cells per ganglion—respond specifically and selectively to mechanical stimuli and are not activated by even large changes in the temperature, acidity or osmotic pressure of the bathing fluid. There is a striking parallel between these mechanosensory cells in the leech and those that innervate the human skin, which also distinguish among touch, deep pressure and noxious or "painful" stimuli. A number of experiments indicate that these cells in the leech are true sensory cells rather than second- or third-order neurons driven only indirectly by sensory neurons in the periphery, and that they are the principal cells conveying sensory information about touch, pressure and noxious mechanical stimuli to the segmental ganglia.

Each of the 14 sensory cells innervates a clearly defined area of skin and responds only to stimuli applied within one of these circumscribed receptive fields. The boundaries of a field can be identified by landmarks such as the segmentation and the coloring of the skin, so that one can predict which cells will fire when a particular area is touched, pressed or pinched. Each cell innervates a region somewhat larger than one segment (seven to nine annuli rather than five). Consequently two cells of the same modality in adjacent ganglia supply an area of skin at the boundary of a segment. On the other hand, there is little overlap of the receptive fields of the cells in one ganglion that serve the same modality. A touch cell, for example, innervates either the dorsal, the lateral or the ventral part of the skin [*see illustration above*]; the overlap is only in the millimeter range, or about the width of an annulus. A similar strict territorial distribution, with a comparable degree of overlap, holds for the pressure and nociceptive cells. Moreover, within the receptive field of a single cell the territory for each axonal branch is a distinct subdivision.

These observations demonstrate the high degree of specificity displayed by individual cells. To learn what factors determine the extent of the area innervated by a single cell it would be helpful to perform developmental or genetic experiments, but the reproductive cycle of the leech is too slow for that. Animals with clear-cut abnormalities of the nervous system are occasionally found, however, and one such leech has been studied in detail. Some of its ganglia contained one or two "extra" touch and pressure cells and extra annulus-erector, or *AE*, motor cells (which raise the bumps on the annuli). The total cell counts were clearly abnormal, the result of either a mutation or a developmental defect. We wondered how such an extra sensory cell would behave. Would it innervate the appropriate area of skin and respond to the appropriate stimulus? In one ganglion that contained three pressure cells instead of two on the right side, all three had the positions, membrane properties and impulse configuration characteristic of *P* cells. Two of them, however, sent their axons by way of the same root toward dorsal skin and responded to pressure in the same field; each cell behaved as though it were the only one. Similarly, in ganglia with four touch cells instead of three the field sizes and positions were normal but one field was doubly innervated. In some ganglia, however, there were two extra pressure cells (four on one side) and in these the arrangement of the peripheral fields became distorted. The results from this single animal suggest that when extra cells of a particular type are produced during development, they operate in parallel with a normal cell complement, but only up to a certain point. Furthermore, they demonstrate that a cell can receive a larger number of synaptic inputs than normal. Whereas the two *P* cells in normal ganglia synapse on the *AE* motoneuron, in this abnormal animal an *AE* cell could receive synaptic connections from three or four *P* cells. It is likely that animals will turn up with other intriguing abnormalities and so provide clues to the ground rules that determine how target cells become innervated during development.



SHORTENING REFLEX is generated and monitored as shown here. The skin is stimulated mechanically (*1*), causing the appropriate sensory cell to fire a nerve impulse, which is monitored by electrodes on the root (*2*). Alternatively the individual sensory cell can be stimulated with a microelectrode (*3*). In either case the sensory-cell impulses reach the large longitudinal motoneuron (*L*) by way of a chemical (*4*) or an electrical (*5*) synapse or both (*6*), eliciting a motor impulse that travels out along the motoneuron axon and is monitored by electrodes on the root (*7*). A strain gage measures the shortening of the muscle (*8*).

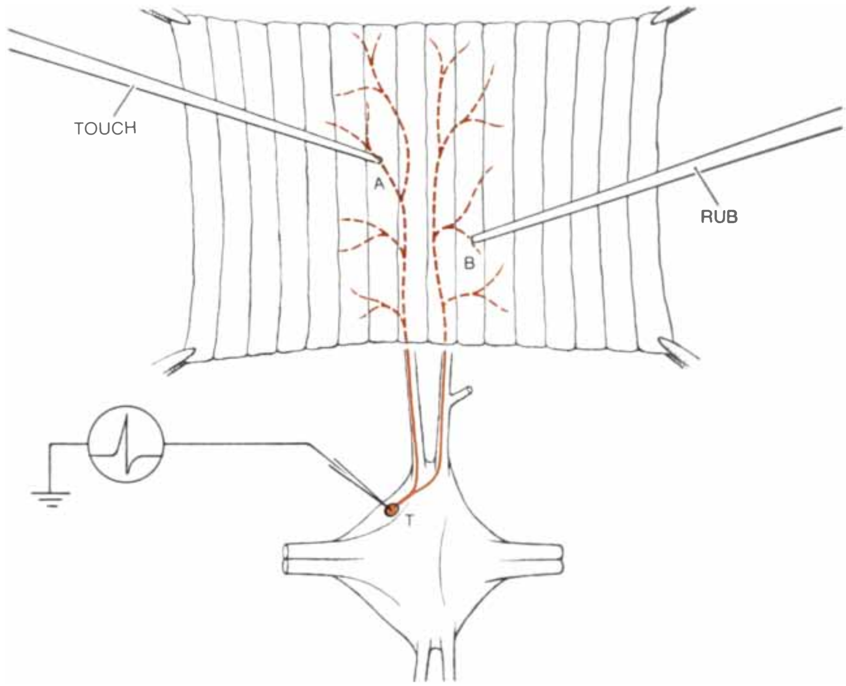
So far we have considered sensory and motor cells as independent units, but for a real understanding of the nervous system one needs to know how the cells interact to produce movement. Some clues to the wiring diagram are provided by simple reflexes such as the segment-shortening that follows stroking, pressing or pinching of the skin of a leech. The incoming limb for these reflexes is provided by the *T*, *P* and *N* sensory cells, converging on the *L* motoneuron, which innervates the longitudinal musculature of the body wall. Several lines of evidence indicate that the connections of *T*, *P* and *N* sensory cells with the *L* motoneuron are direct—that there are no unknown intermediary cells. This is an important point, because only if each constituent and its properties are known is it possible to pinpoint the sites at which any interesting behavioral modifications in signaling take place.

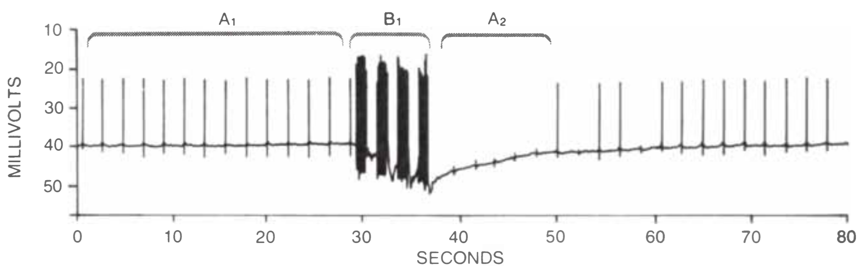
To demonstrate the behavioral reflexes these connections subserve, a length of body wall is removed from the animal together with its ganglion. Sensory and motor impulses traveling to and from the ganglion are monitored by external electrodes on the roots, and the tension generated by the longitudinal muscle is recorded with a strain gage. The activity in muscle fibers can also be registered by an intracellular electrode. Under these conditions a single sensory impulse usually causes little or no increase in tension in the longitudinal muscle fibers, but light, repetitive indentations of the dorsal skin give rise to bursts of impulses in the touch cell that consistently cause a shortening of the body wall. When the *L* cell is prevented from firing by the injection of a hyperpolarizing current (which decreases its excitability), the reflex contractions are blocked. This shows that the *L* cell is the only motor cell contributing to this particular reflex pathway. The results with reflexes initiated by the *N* and *P* cells are similar.

These three simple pathways exemplify the different forms of signaling encountered in the nervous systems of higher animals. The nociceptive cells act on the *L* motoneuron through conventional chemical synapses, the touch cells through electrical synapses and the pressure cells through a combination of both mechanisms. There are ways of distinguishing between the kinds of synaptic transmission. The characteristic properties of electrical synapses are that current spreads directly from cell to cell with no appreciable delay and that transmission is not blocked by high concentrations of magnesium ions. The connection of the touch cell with the *L* motor cell fulfills these criteria. Stimulation of the nociceptive cell, on the other hand, causes a synaptic potential to arise in the *L* cell with a delay of two to four milliseconds after the peak of the presynaptic action potential. This delay is typical of a chemical synapse. Furthermore, as one would expect at a chemical synapse, transmission is reversibly abolished by bathing the preparation in a solution containing an abnormally high concentration of magnesium ions, and this block is counteracted by increasing the calcium-ion concentration; this antagonistic action of the two ions is observed at other chemical synapses, where magnesium reduces and calcium increases the amount of transmitter liberated by the presynaptic terminals in response to an impulse. Finally, unlike the situation at the touch-cell synapse, no direct current spread can be detected between the nociceptive cell and the motoneuron. The third type of sensory neuron, the pressure cell, gives rise to synaptic potentials in the *L* motoneuron with characteristics of both chemical and electrical mechanisms, that is, one component of the potential is delayed and is reduced by magnesium ions.

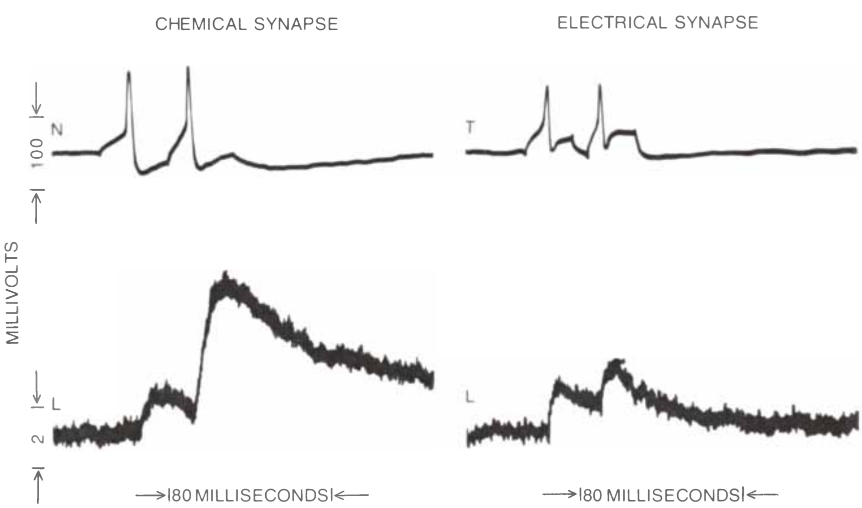
An important aspect of nerve function is the manner in which the properties of neurons and synapses change with prolonged use. It is a common feature in a variety of vertebrate and invertebrate nerve cells that the membrane potential increases following trains of impulses, and this in turn reduces their excitability. In leech cells this hyperpolarization is of large amplitude and long duration. For example, natural activation of a touch, pressure or nociceptive cell by just a few stimuli can increase the membrane potential from a resting value of 45 millivolts to 70, which then gradually returns to normal over a period of up to 15 minutes. The hyperpolarization is not confined to the cell body but also occurs in its peripheral axons and in fine processes within the neuropil.

Two distinct cellular mechanisms are responsible for generating the hyperpolarization. One is the activity of an ionic pump, which extrudes the positively charged sodium ions that enter a nerve cell during impulses and thus restores the ionic concentration gradients to their original level. (Similar transport mechanisms have been shown in other cells to be effected by an enzyme in the membrane.) The other mechanism is a prolonged increase in the potassium permeability of the membrane, which drives

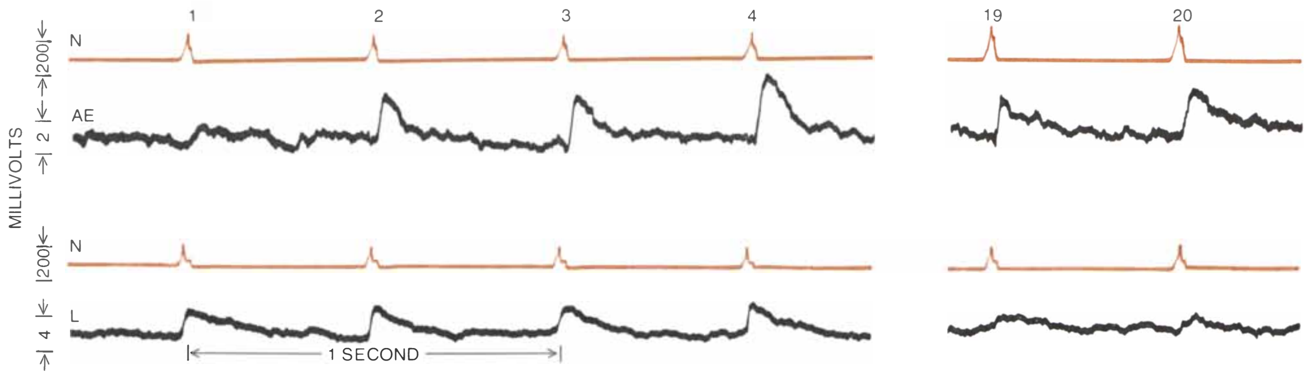




REPETITIVE STIMULATION decreases the sensitivity of a touch cell to mechanical stimuli. A stylus (*A*) applies brief mechanical stimuli, just above threshold, each of which initiates a single impulse that is recorded as shown (A_1). Another stylus rubs a different part of the same receptive field (*B*). The prolonged trains of impulses (B_1) produced by the rubbing hyperpolarize the touch cell, that is, they increase the resting potential, making the cell less sensitive, so that for a time it does not respond as before to the brief stimuli (A_2).



CHEMICAL AND ELECTRICAL SYNAPSES behave differently with repetitive activity. In the case of a nociceptive cell (chemical synapse), successive impulses liberate increasing amounts of transmitter and produce a larger potential in a motoneuron (*left*). This does not happen in the case of a touch cell (electrical synapse); the potentials it elicits in a motoneuron remain constant (*right*). In general electrical synapses are the more stable.



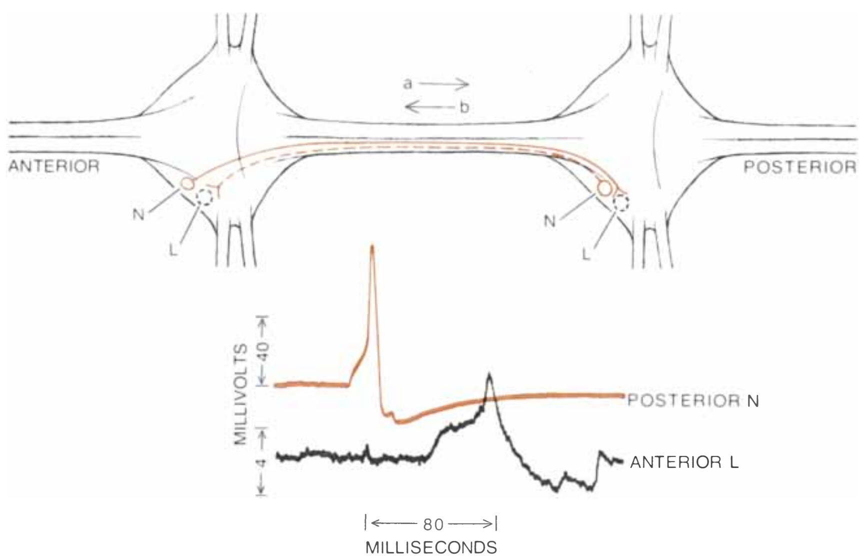
DIFFERENT MOTONEURONS behave differently with repetitive stimulation by the same kind of sensory cell. The traces show that potentials (*black*) elicited in an annulus-erector motoneuron (*AE*) by a nociceptive cell (*N*) increase to roughly two and a half times their original value and stay large; potentials elicited in a large longitudinal motoneuron (*L*) grow larger at first, then smaller.

the membrane potential in the hyperpolarizing direction.

The effects of the large residual change in membrane potential on the transmission of sensory information are most simply studied in touch cells because of the ease with which they can be activated. After an area of the skin of a leech has been stroked with a small stylus for a few seconds, stronger mechanical or electrical stimulation is required to initiate impulses. This decrease in excitability occurs throughout the entire receptive field of the touch cell [*see top illustration on preceding page*]. Wherever impulses travel in this cell, in toward the ganglion or out toward the skin, they leave behind a hyperpolarization. There is a striking similarity among changes at the periphery, along the axon and in the cell body, where the membrane potential is recorded directly with a microelectrode. At each site there is a decrease in excitability that parallels the hyperpolarization following impulses; both the hyperpolarization and the increase in the threshold are abolished by the drug strophanthidin, which is known to act selectively on ionic transport processes that carry sodium ions out of cells.

The next step is to see how repetitive firing affects the properties of synapses and reflexes. In general one would expect electrical synapses to remain rather stable in performance, whereas chemical synapses might be more easily changed. To test this idea we can stimulate a number of different kinds of sensory cells and examine the synaptic potentials arising in the single *L* cell with which they all connect. If a nociceptive cell, which acts through a chemical synapse, is activated for a few seconds, the synaptic potentials recorded in the *L* cell during the train first increase (facilitation) and then decrease (depression). In theory these two changes in performance could be caused by changes in the amount of transmitter released by the presynaptic (sensory) ending at each successive impulse or by changes in the chemosensitivity of the postsynaptic (motoneuron) membrane. A number of experiments suggest that it is the mechanism of transmitter release that is influenced by previous activity.

The effects of repetitive firing on chemical synaptic transmission are particularly well brought out by cooling. Leeches frequently live in cold ponds and they can be kept alive in refrigerators for long periods of time, yet in isolated preparations cooled to two degrees Celsius (36 degrees Fahrenheit) chemical transmission by the nociceptive cell to the *L* cell becomes almost ineffective within a few seconds as the synaptic potentials decline in size. If all synapses behaved like that, leeches would soon become paralyzed in the cold. The fact is that the synaptic connections to the *L* and the *AE* motor cells are affected in different ways by repetitive activity and by cold. The synapses on the *AE* cell show more facilitation and less depression during a train of stimuli than those



DIRECTION OF STIMULATION makes a difference when a sensory-cell impulse excites a motoneuron in an adjoining ganglion. The nociceptive cell in one ganglion excites the *L* cell on the same side of the next ganglion. Even in an intact nervous system (*above*) the potential elicited in the *L* cell is not exactly the same when the motoneuron is the posterior of the two cells (*a*) as it is when the motoneuron is toward the head of the animal (*b*); there is a small inhibitory component in the back-to-front signal. The directional effect is intensified by the cutting and regeneration of a connective (*right*). The appropriate cells interact again after regeneration. When a sensory cell is stimulated, it produces a normal excitatory potential in a motoneuron posterior to it (*a*) but an abnormal hyperpolarization, or inhibition, in a motoneuron anterior to it (*b*). This does not represent an error in reconnection: the same directional effect is seen all along the length of an animal that has been operated on. Even in ganglia at a distance from the lesion the effect is excitatory in the head-to-tail direction (*c*), whereas it is inhibitory in the tail-to-head direction (*d*).

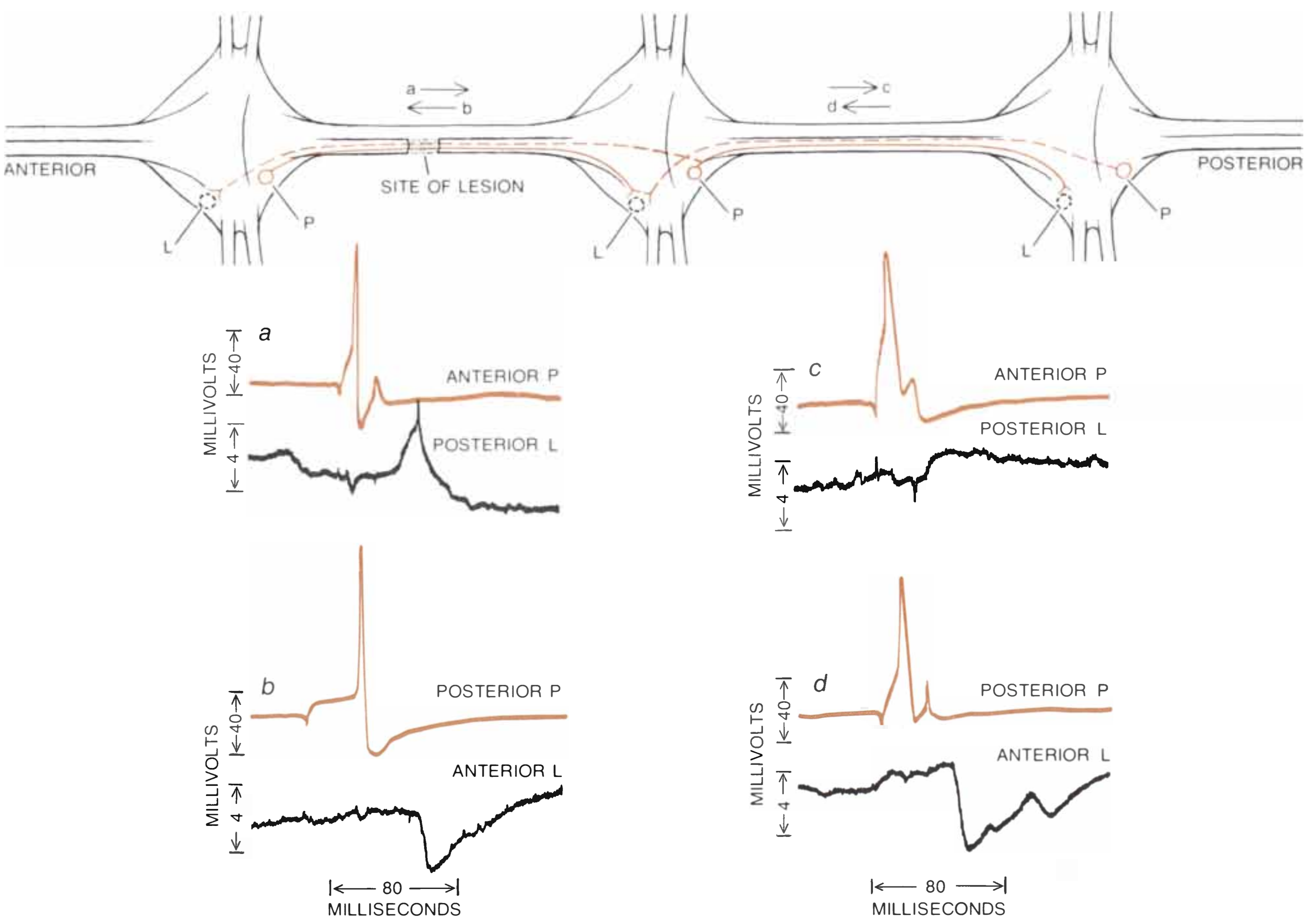
on the *L* motor cell and far less depression in the cold. These results suggest that the same sensory cell can act with progressively greater effectiveness on one target cell while the efficiency of its action on another is diminishing. Mechanical stimulation of the skin therefore initiates a dual sequence of reactions: first a shortening, which wears off, and then a maintained annular erection. Again, indirect evidence suggests that it is the release characteristics of the various endings of the sensory cell that are different.

The behavior of electrical synapses is quite different. Even with repeated stimulation of the touch neuron, which makes an electrical synapse on *L* cells, the synaptic potentials remain unchanged. The synapse also continues to function effectively in the cold. We can conclude that the electrical synapses provide a more reliable form of transmission than those operating by the release of chemical transmitters; they are less influenced by the history of activity or the environmental conditions.

Numerous experiments on invertebrates and vertebrates indicate that the synaptic connections of nerve cells are specific, in the sense that highly intricate connections are selectively established between certain cells but not between others. There is as yet little information about the processes that enable nerve cells to find their targets during either development or regeneration, nor is it clear how reproducibly any one cell forms its connections in different animals. So far experiments with vertebrate nervous systems have dealt with the reestablishment of synaptic connections between groups of nerve cells. A good system for examining regeneration between individual cells is provided by the connections of sensory cells in one ganglion with the *L* motor cell in the next. The connectives linking adjacent ganglia are up to a centimeter long and can be cut without difficulty in the animal. The particular cells we have examined before and after regeneration are the nociceptive and pressure cells. In control animals impulses in these *N* and *P* cells evoke synaptic potentials on the *L* motoneurons in the adjacent ganglia toward the front and the rear of the animal. These anterior and posterior interganglionic connections are direct and appear similar to those within a ganglion.

Leeches are anesthetized and one of the two parallel connectives is cut, the remaining connective being left intact to serve as a splint for regeneration. The lesion abolishes the synaptic connections of the sensory cells with motor cells in the two ganglia adjacent to the lesion, on the operated side of the leech. (Fibers do not cross to the connective on the other side.) Through-fibers running to distant ganglia and to the brains at the two ends of the leech are also interrupted by the operation. The leeches are examined for signs of regeneration and altered synaptic junctions from 45 to 260 days after the operation. In most of the leeches the cut connective shows obvious signs of regeneration, although occasionally it cannot grow back because of scar formation.

In leeches with a regenerated connective, synaptic connections are consistently re-formed between sensory cells in one ganglion and the *L* motor cell in the next, and the connections are specific: when we activated *P* or *N* cells in one ganglion while recording in succession from 10 or more neurons in the vicinity of the *L* motor cell in the neighboring ganglion, the usual result was that synaptic potentials were observed in the *L* cell but not in other cells. (Occasionally, in regenerated pairs of ganglia, one or two neurons close to the *L* cell might also



show synaptic potentials. We cannot yet tell if this represents a decline in the accuracy of regenerated connections.)

Although the appropriate cells interact once again after regeneration, certain properties of the regenerated synapses are consistently different from those seen in normal leeches. In particular the balance of excitation and inhibition is markedly altered. In one direction of stimulation the response resembles the normal one: when a sensory cell in the anterior ganglion of the pair is stimulated, an excitatory, depolarizing synaptic potential is recorded in the *L* cell in the posterior ganglion by way of the regenerated connective. On the other hand, when the direction of stimulation is reversed (from posterior sensory to anterior motor), an inhibitory, hyperpolarizing synaptic potential appears in the *L* motor cell. This asymmetry in the synaptic effects is invariably found whenever synaptic connections are reestablished.

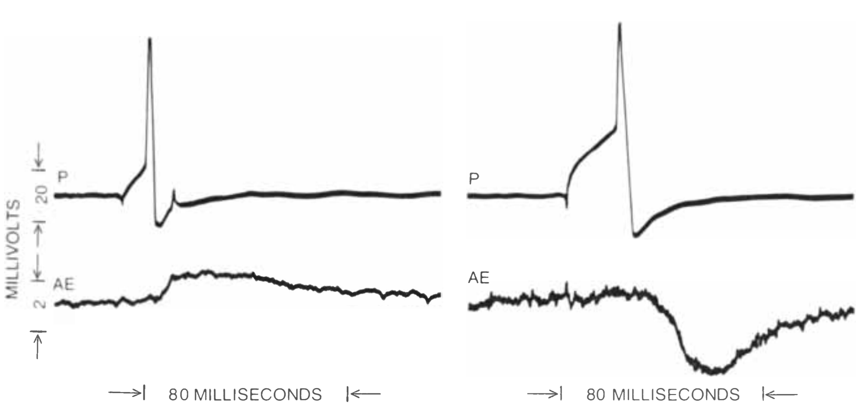
At first it seemed as though the changed synaptic interactions observed after regeneration might represent some form of error in reconnection. Quite unexpectedly, however, we have found that similar changes occur in the interactions of adjacent ganglia along the entire length of the operated leech. The explanation may lie in the fact that although the direct connections of these other ganglia are not directly severed by the operation, presumably much of the normal impulse traffic is chronically reduced by the interruption of through-fibers.

The similarity between these alterations in the synaptic connections of pairs of neurons in distant ganglia and those found at regenerated synapses prompted a reexamination of the connections in the normal leech. When repeated synaptic responses were averaged on a computer, a small but consistent asymmetry became apparent in the connections from sensory cells to *L* cells in the neighboring ganglia. Going from posterior sensory to anterior motor cell, a short-latency inhibitory component of the synaptic potential can be discerned, but in the opposite direction (anterior sensory to posterior motor) one sees only excitation. The existence of this asymmetry in control animals suggests that the marked changes in the synaptic connections after the operation represent an enhancement of a normal but ordinarily occult synaptic pathway rather than the formation of a new, abnormal pathway.

These results involving distant ganglia also raised the question of whether comparable changes might also occur within a single ganglion deprived of its normal input. To test this we cut all the connectives and peripheral nerves to a ganglion, leaving the ganglion isolated in an "organ culture" within the animal, with no possibility of regeneration. When the ganglion is removed from the leech after about a month, many of the normal synaptic connections are seen to be dramatically changed. If, for example, one examines the interactions of the *T* cell with the *L* motoneuron in one of these isolated ganglia, it is clear that the synaptic potential is produced by a combination of both chemical and electrical mechanisms; the potential is much larger and persists longer than the purely electrical synaptic potential seen in a *T*-cell synapse in normal ganglia. There is, in other words, an added chemical input, which could be brought about by the outgrowth of new fibers from an interposed neuron or could represent a direct change in the synaptic connections of *T* cells to *L* cells. Another striking change is seen in the interactions of *P* and *AE* cells in isolated ganglia: instead of an excitatory potential on the *AE* cell, one observes an inhibitory, hyperpolarizing potential when the *P* cell is stimulated. There is a behavioral correlate of these altered connections: if the skin is pressed, any annuli that happen to be erected are quickly flattened. This characteristic change is seen in ganglia all along the length of the animal in which one ganglion has been isolated.

The underlying causes of these changes in the interactions of sensory and motor cells after a lesion are not yet fully understood, but it is tempting to speculate that the lesion causes fibers that normally synapse on *L* and *AE* motoneurons to become ineffective. This partial denervation could conceivably lead to chemical supersensitivity, accessory sprouting of collateral nerve fibers or the formation of new synaptic contacts; we are now attempting to distinguish among these and other possibilities by a number of different techniques. What is clear from the results obtained by making lesions is that leech neurons can regenerate with a high degree of accuracy, and that the synapses within the central nervous system are capable of undergoing marked changes in effectiveness.

At first the analysis of a brain consisting of so few cells might seem to present a finite problem, and yet we have barely begun to study a number of essential questions. For example, we are only now starting to study the anatomy and chemistry of the nerve cells and synapses. At present little is known about the integrated behavior of the leech, although Stent and his colleagues at the University of California at Berkeley have made considerable progress in analyzing the neuronal mechanisms involved in swimming. For some problems it is simply a question of doing the work; for others technical advances are required. One hope for the future lies in the study of the development of the nerve cells. Recently we have begun to examine the structure of leech embryos in the hope of finding a miniature nervous system that can be reconstructed anatomically. Another promising approach is to maintain ganglia alive outside the animal in organ culture, where one can study neurons in a controlled environment over long periods of time. So far we have successfully maintained ganglia in this way for up to six weeks. In time the leech may help to elucidate many of the principles that apply in higher animals. In particular we hope that its nervous system will provide clues about the mechanisms by which use, disuse, abnormal use and lesions can bring about changes in the properties of synapses and thereby modify the performance of an animal.



ISOLATION OF A GANGLION alters the synaptic interactions within it. When a ganglion was maintained in the animal for 36 days after its roots and connectives were cut, an impulse in a *P* cell, which would normally excite the *AE* cell (*left*), instead inhibited it (*right*).

Mr. Sucy, is anyone using microfiche readers in teaching?

Jim, where can I apply for Federal funding of my research on picture-taking as a guidance tool?

What is the evidence, Jim, for increased retention with audiovisual aids?

Sucy, I need to set up a course in non-destructive testing. What do I do?

Does Kodak have any suggestions on how our professional society can reach its members with packaged training?

Where can I buy a Kodak visualmaker kit?

Service to education

James G. Sucy is the name. Manager, Education Markets Services, is the title. Address: Eastman Kodak Company, Rochester, N.Y. 14650. Phone: 716-724-4501. Function: to find the person in the best position to field any teacher's question that relates any Kodak product or service to the educational process. "Teacher" covers professor, educator, trainer, instructor–whatever you think it should cover. Patterns discerned in the questions will shape next year's and next decade's Kodak products and services.

Kodak

What type of programs do you have for military trainers?

Does Kodak have any workshops for teachers?

Can you help me plan a new projection room?

DO YOU HAVE ANY RECOMMENDATIONS ABOUT MAKING SLIDE-TAPE INSTRUCTIONAL UNITS?

Can you give us some information about using microfilm in our school library?

What services does Kodak have for junior high schools?

Do you know of any schools where a student of mine can learn to be a photographer?

Please send me some information about photography at the grade-school level.

HOW DOES KODAK ANALYZE AND PLAN TRAINING PROGRAMS?

I'd like to talk to someone about computer-assisted instruction for our industrial training.

Do you have any information about careers in photography?

Do you have any information to help me start a high school photo club?

Does Kodak have any films for loan about picture-taking?

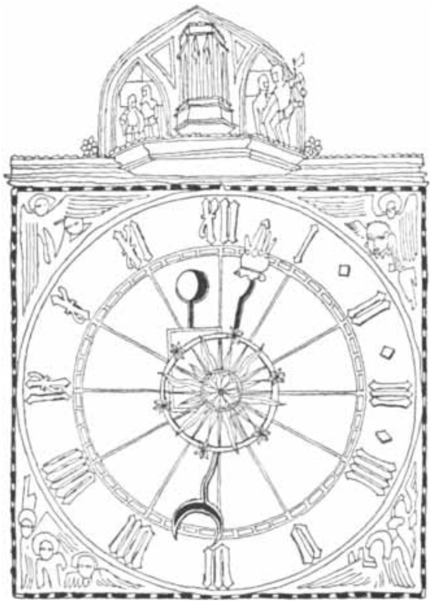
Would you send me information about the flexible classrooms in your new education center.

Should I use rear or front projection in our new building?

Our school of dentistry is individualizing. Do you have any information to assist us?

How can I make my own slides for my biology class?

SCIENCE AND THE CITIZEN



The Uranium Problem

Americans, Europeans and Japanese who are wondering why current shortages of oil were not better anticipated have now received a clear warning that there will be a severe uranium shortage around 1980 unless governments and industry take aggressive steps to find more of the nuclear fuel.

The warning is issued in a joint report by the Nuclear Energy Agency of the Organisation for Economic Co-operation and Development (OECD) and the International Atomic Energy Agency (IAEA). The report makes a comprehensive projection of the number and kind of nuclear reactors that will be built between now and 1990 in the U.S., Western Europe and Japan and weighs their fuel requirements against the known world reserves of uranium (outside the U.S.S.R. and China) and the projected capacity for extracting the ore and producing enriched reactor fuel by gaseous diffusion or gas centrifugation.

The report estimates that at the end of 1973 the U.S., Western Europe and Japan had an installed nuclear capacity of 50 gigawatts, or 50,000 megawatts. The U.S. share of the total was 58 percent. By 1990 the total capacity is expected to increase twentyfold to just over 1,000 GW, at which time the U.S., Western Europe and Japan will have respectively 51, 37 and 10 percent. It is estimated that the U.S. will obtain about 22 percent of its energy needs from uranium by 1990.

The OECD-IAEA report notes that there is more than enough uranium production and separative capacity to meet present needs. In fact, the uranium industry is depressed. As new reactors are completed, however, the situation will shift rapidly. Existing and proposed separative (that is, enrichment) capacity will be saturated by 1983. Between now and 1990 the Western nations and Japan will have to raise their separative capacity to 10 times its present level. This alone will require an investment of roughly $15 billion.

The central cause for concern, however, is an impending shortage of proved uranium reserves and production capacity. The U.S. has 26 percent of the world's "reasonably assured reserves"; the other leading countries are Canada (20 percent), Sweden (17.5 percent), South Africa (17 percent) and Australia (6.5 percent). As in all mining, an increased production rate requires an increase in proved reserves. Over the past eight years proved reserves have increased at an average rate of 65,000 tons per year, or roughly three times annual production. The report estimates that by 1980 annual production must rise to 50,000 tons, which would be "close to the ultimate or maximum production capacity which could be attained from presently known reserves and resources." Between 1980 and 1985 annual demand is expected to double again; "the presently known reserves would not support a production rate of anything like 100,000 [tons of uranium] per annum." To maintain the projected rates of production the discovery rate will have to increase from the recent figure of 65,000 tons per year to 230,000 tons per year in 1990.

"It is essential," concludes the report, "that urgent steps be taken to increase the rate of exploration for uranium so that an adequate forward reserve may be achieved.... Current uranium prices are generally not adequate to induce the necessary exploration, or the increased expansion of production capability needed. Some means should therefore be found to ensure that the production levels required are achieved so as to avoid shortage... in the 1980's. It is important that this matter be fully appreciated by governments, utilities and mining companies."

The report indicates that its predictions may be inadequate in one important respect: "The possibility that an accelerated nuclear programme might be instituted in many countries to overcome an energy shortage was not considered in this study."

What SNU?

Since the late 1950's it has been generally accepted that thermonuclear reactions in the core of the sun must result in the emission of a substantial number of high-energy neutrinos. These neutrinos should be detectable on the earth. Raymond Davis, Jr., of the Brookhaven National Laboratory set up an experiment to detect the solar neutrinos and to measure their flux. As of the summer of 1972 his results over the preceding five years indicated that only about a tenth of the number of neutrinos predicted by the theory of the sun's interior were actually reaching the earth. At a meeting of the International Astronomical Union held in Warsaw last fall, William A. Fowler of the California Institute of Technology summarized Davis' more recent results. The situation for the theory has grown even worse: it seems that the number of neutrinos reaching the earth may be as low as a thirtieth of the number predicted by the theory and may even be essentially zero.

Davis' experiment is heroic. His apparatus consists of a 100,000-gallon tank of tetrachloroethylene, a common dry-cleaning fluid, at the bottom of the Homestake Gold Mine in Lead, S.D. Twenty-two percent of the mass of tetrachloroethylene is chlorine 37. When a chlorine-37 nucleus interacts with a neutrino, it is transformed into a nucleus of the radioactive isotope argon 37. Such an interaction is extremely rare, however, since the neutrino has no detectable mass, no magnetic moment and no electric charge.

At intervals of some 50 days the tank is swept with helium to remove the argon 37, which is then introduced into a proportional counter that detects its radioactive decay. The results of the experiment are expressed in solar neutrino units, abbreviated SNU. One SNU is defined as the flux causing one neutrino-induced event per second per 10^{36} chlorine atoms. In Davis' tank, which contains some 10^{30} chlorine-37 atoms, one SNU corresponds to .2 argon-37 atom formed per day.

One recent theoretical calculation made by John N. Bahcall and his colleagues indicates that the flux of solar neutrinos should be approximately 5.6 SNU, corresponding to one event per day. Davis' results, however, indicate that the observed flux is only $.2 \pm 1.0$ SNU, or one event per month. On the basis of the customary scientific interpretation this means that the actual flux of solar neutrinos is probably no greater than 1.2 SNU and could be zero.

According to Fowler, there are only two reasonable explanations for the discrepancy between theory and observation. The first is that the neutrino decays on its way from the sun to the earth. The second is that every 250 million years or so new fuels are mixed into the solar interior; the sun heats up, expands and then cools to the point where nuclear reactions are shut down and the sun shines for a few million years on its store of gravitational and thermal energy.

If this second explanation is true, the energy radiated by the sun has presumably decreased for periods of about five million years at intervals of some 250 million years. Such behavior would have important consequences on the earth and might account for episodes of glaciation in the Quaternary period (the present), the late Paleozoic and the early Paleozoic. Fowler concluded his talk: "It [is] clear that the case of the missing solar neutrinos is still an exciting detective problem for astronomers, physicists, chemists and geologists. So, what SNU?"

The Sleep Peptide

Experiments dating back as far as 1913 have pointed toward the existence of a natural sleep potion: a factor in the body fluids of an animal that induces sleep when it is injected into another animal. Such a factor has now been isolated and partially characterized by a group at the University of Basel. They call it "sleep-factor delta" because it specifically promotes the production of delta waves, the long, slow brain waves characteristic of light sleep.

M. Monnier and G. A. Schoenenberger and their colleagues began by stimulating a sleep-producing region of the thalamus in the brain of rabbits and withdrawing venous blood from the brain of the sleeping animals. They also took blood from rabbits that had not been put to sleep, to serve as a control. The blood from the donor animals was dialyzed, or diffused through a membrane, to remove large molecules, and the purified fraction of the resulting "sleep dialysate" was injected into the brain of recipient rabbits. Within five to 10 minutes of the injection the electroencephalograph recorded an approximate doubling of delta-wave activity in rabbits given the active substance, whereas there was no appreciable change in the delta activity in rabbits given dialysate from the control donors. At the same time the motor activity of the rabbits administered the active substance was reduced by more than half, whereas that of the control recipients stayed about the same.

Characterization of the sleep factor depended on successive fractionations of the original dialysate, with each factor being checked for hypnogenic activity under double-blind conditions: the workers who tested the fractions did not know what fraction they were testing. First the dialysate was fractionated by gel-filtration through a chromatographic column. This removed the salts and other low-molecular-weight substances; only the desalted fraction, which included 5 percent of the original dialysate, exhibited the hypnogenic activity. That fraction was subsequently divided by two more stages of chromatography and by paper electrophoresis into successively smaller active fractions. Analysis of the final fraction, which was powerful enough to be effective in a dose of less than six billionths of a gram per kilogram of body weight, showed it to be a peptide containing seven or more different amino acids, with a molecular weight between 355 and 1,500 and most likely about 700. The findings "suggest strongly that this factor may be involved in the humoral regulation of sleep."

Supersorghum

Sorghum is the world's fourth most important food grain (after wheat, rice and corn), being the principal subsistence grain for more than 300 million people in underdeveloped countries. It has the advantage of growing on land that is too dry and too infertile for other grains. On the other hand, the most commonly grown strains have the disadvantage of being low in protein and in lysine, one of the amino acids essential in the human diet as a building block for protein. Now investigators at Purdue University have announced a discovery that promises to increase substantially the nutritional value of sorghum.

After analyzing some 9,000 strains of sorghum from many countries the Purdue group, which worked under a $1.7-million contract from the U.S. Agency for International Development (A.I.D.), found two strains from Ethiopia that are much richer than the commonly grown sorghum in protein and in lysine. Presumably the traits can be bred into the strains of sorghum that are grown widely for food. John A. Hannah, administrator of the A.I.D., said when the discovery of the Ethiopian strains was announced that improving the protein quality of sorghum "will amount to a gift of life, especially for children."

Eye in the Fingertip

Braille gives the blind access to many written materials, but Braille books are cumbersome and the transliteration of the written word into the Braille alphabet of embossed dots is time-consuming and expensive. A recent application of the technology of semiconductor integrated circuits promises to eliminate the need for this transliteration; it is a device that enables the blind to read ordinary printed material, and in some cases other forms of notation such as handwriting, sheet music and graphs.

The device is called the Optacon, for optical-to-tactile converter. Like Braille, it depends on the sensitivity of the fingertip to tactile patterns. Its essential components are a small optical sensor, or camera, with which the user scans a line of type, and an array of vibratory pins on which the user places his index finger. The image of one character at a time is reproduced by the vibrating pins.

The Optacon is the project of John G. Linvill and his associates at Stanford University and James Bliss, formerly of the Stanford Research Institute. Development was supported by a $1.8 million grant from the Office of Education of the Department of Health, Education, and Welfare. The reading aid is now being produced commercially by Telesensory Systems, Inc., of Palo Alto, Calif.

In experiments made in 1964 Linvill demonstrated that letter shapes presented tactually as a pattern of vibrating pins could be recognized. Research since then has been directed toward producing a device that provides accurate and efficient recognition of characters of various sizes and forms. It has also sought to reduce the size, weight, power consumption and cost of the device.

These goals have been approached through the use of integrated circuits. The "retina" of the optical system consists of 144 photosensitive transistors. They are arrayed in six rows of 24 on a single "chip," or substrate, of silicon. The 144 vibratory pins of the tactile screen reproduce this array; they are driven by piezoelectric crystals that, when stimulated, vibrate at 250 to 300 hertz. Each element of the retina has a corresponding element in the tactile screen. If the blackness of the print under a particular phototransistor exceeds a certain threshold, the corresponding pin is made to vibrate.

The present Optacon consists of a finger-size camera and a larger package containing most of the electronics and the tactile screen. The camera is moved with one hand and the images are sensed with the other. Experienced users can read with the Optacon at up to 90 words per minute, about half their speed reading Braille. In training programs it has been found that school-age children are the most successful subjects in learning to use the device.

Telesensory Systems has made about 500 Optacons. Many are being used by blind computer programmers, who are enabled by the machine to read printouts and manuals without assistance. The price is $3,450.

Linvill and his co-workers are now attempting to develop a much smaller reading aid, incorporating the optical and tactile transducers in a single package, that could be used with one hand.

The Water Hole

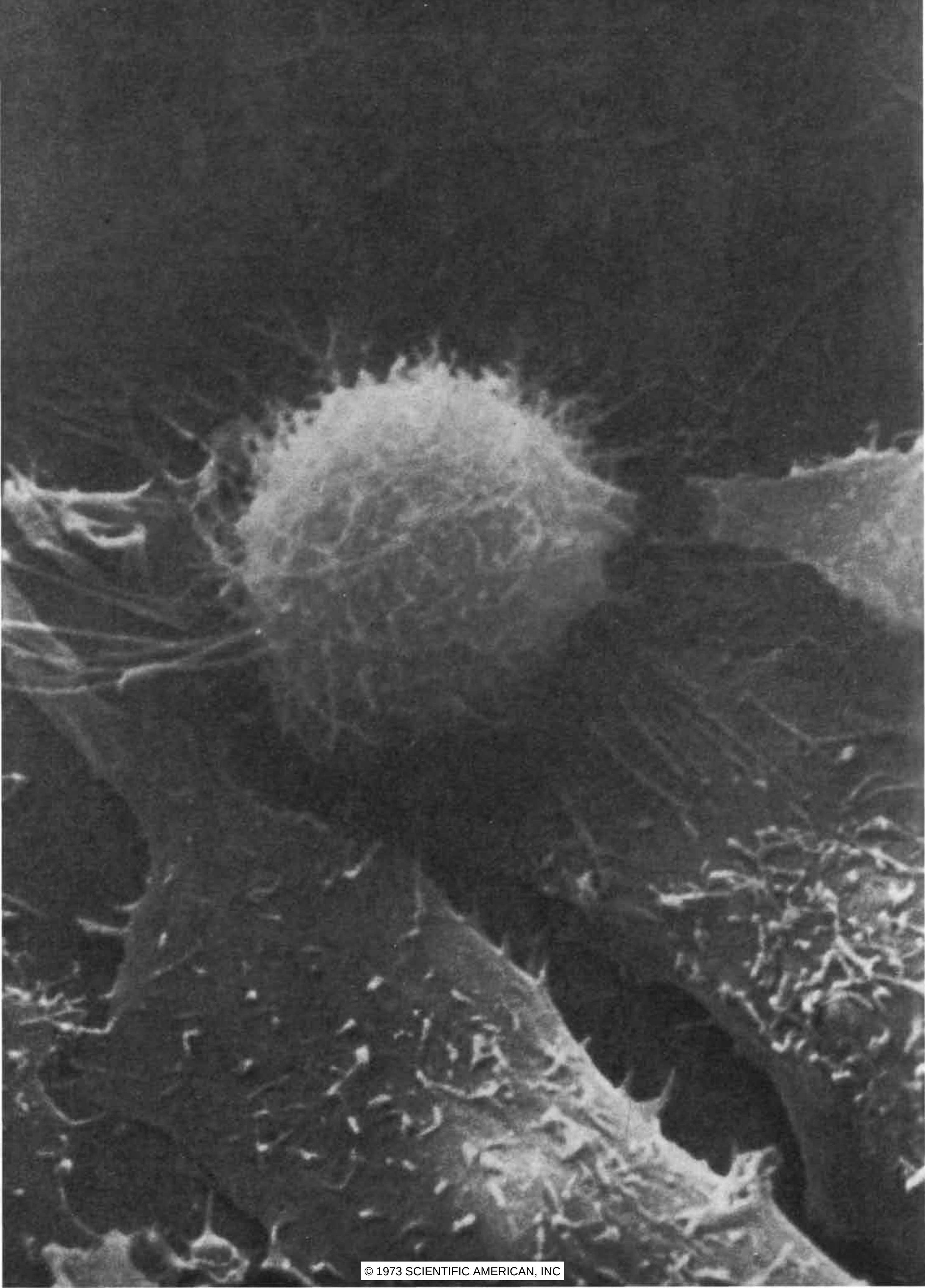
In the past few years there has been much speculation on the hypothetical problem of using a large radio telescope, such as the one at the Arecibo Observatory in Puerto Rico, to communicate with an extraterrestrial civilization. The prospects for such interstellar communication rest of course on many assumptions, one of the most crucial of which is that both the transmitting and the receiving civilizations will happen to choose the same microwave radio channel, presumably on the grounds that certain frequencies are more "natural" than others. From the beginning the consensus (at least among earthbound radio astronomers) has been that the most promising candidate for such a channel lies at 1,420.405 megahertz, the frequency at which radio waves are emitted (and absorbed) by a particular hyperfine energy-level transition of neutral hydrogen, the most abundant atom in the universe. The few serious attempts that have been made to intercept interstellar radio signals have in fact been confined to this frequency.

The question has now arisen of whether more attractive "natural" frequencies exist. In a recent issue of *Nature* two of the leading workers in this field, Frank D. Drake and Carl Sagan of the National Astronomy and Ionosphere Center of Cornell University, review the matter and come up with a proposal of their own for "two alternative natural frequencies for interstellar radio communication based on two quite different and, we feel, attractive arguments."

In the galactic plane, Drake and Sagan point out, "the 1,420 MHz line is noisy, which decreases its attractiveness." Present information places the minimum "sky noise" level from all sources at an "equivalent temperature" of about five degrees Kelvin, centered at a frequency of a few thousand megahertz. This minimum value, the astronomers note, is not far from both the hydrogen (H) transition frequency of 1,420.405 megahertz and the principal hydroxyl (OH) transition frequency of 1,667.358 megahertz; the minimum is thus "close to the frequencies associated with the components of the water molecule."

Although they enter "a caveat to avoid water chauvinism," Drake and Sagan argue that a strong case can be made "that extraterrestrial life will, in the main, also use water as a solvent system." The proximity of the principal microwave transitions of the neutral hydrogen atom and the hydroxyl radical to the minimum-sky-noise frequency had already prompted another student of the matter, Bernard M. Oliver of the Hewlett-Packard Company, to suggest that the preferred communications channel lies in this "water hole" between 1,420 and 1,667 megahertz. Even here, however, "there still remains a multitude of possible channels." Drake and Sagan therefore propose "a well defined 'natural' frequency in the water hole: a frequency selected by the center of mass of the water molecule itself." Their best estimate of this frequency is 1,652.418 megahertz.

An entirely different "natural" frequency, according to the Cornell workers, is "uniquely specified" by the intersection of two fundamental sources of noise: the black-body background radiation and the quantum noise of radiation. The frequency "hole" that corresponds to this intersection lies at about 56,000 megahertz, a frequency that cannot be specified more precisely because of uncertainty about the value of the black-body background temperature. Unfortunately this radio channel "is unavailable from the surface of the earth because of the strong molecular oxygen lines nearby." As a result, they conclude, interstellar communication on this channel will require radio telescopes in space, facilities that they believe "will probably be available soon (especially on a stellar evolutionary time scale)."



THE CELL CYCLE

What happens in the living cell between the time it is born in the division of another cell and the time it divides again? New methods of investigation reveal four phases in the cycle

by Daniel Mazia

Double or nothing. With few exceptions a living cell either reproduces or dies; the principle is so simple that no one has bothered to call it a principle. A cell is born in the division of a parent cell. It then doubles in every respect: in every part, in every kind of molecule, even in the amount of water it contains. Thereafter it divides with such equal justice that each new daughter cell is an identical copy of the parent. This doubling and halving, the cycle of growth and division, is known generally as the cell cycle.

Although these facts have been known for a long time, the concerted study of the cell cycle began only recently. Scarcely 20 years ago, when I published a lecture called "The Life History of the Cell," I thought I was saying something that needed saying when I concluded: "We are no longer satisfied with a good measurement or a good observation or the clever elucidation of some mechanism in the cell unless we can locate it on a time axis, and the time I refer to is not our time but that of the cell itself, as expressed in its life history." The past two decades have seen rapid progress along this very line. Although the first comprehensive book on the cell cycle, J. M. Mitchison's outstanding volume, was not published until 1971, he was able by then to cite 1,000 other cell-cycle studies that he had selected from an even larger literature.

The objective of all these studies has generally been to define as exactly as possible the events within a cell that characterize a given phase of its life history. Going further, the goal has been to discover just which events are the ones that drive the cell through time, moving it from one phase of its history to the next.

Growth and Division

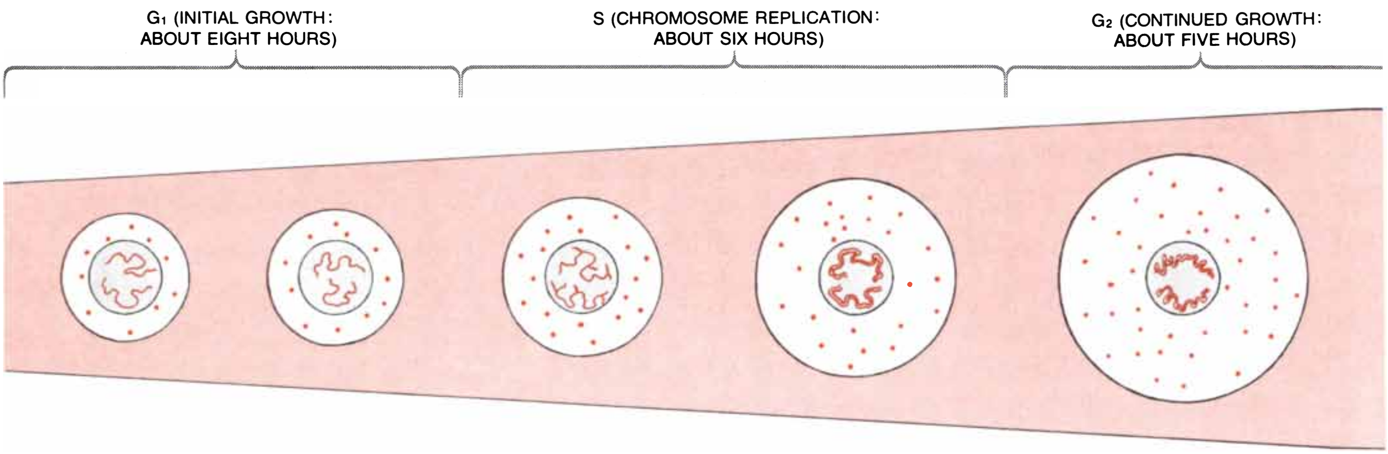
When we say that cells double in all their constituents and then divide in half, we are obviously describing the average case. When we observe individual cells, however, we find deviations from the average. For example, a number of kinds of cells have been weighed at various intervals throughout their period of growth. If the process of growth consisted in a steady production of all the constituents of the cell, including those required for further growth, we would expect a series of weight measurements to yield a smoothly accelerating growth curve, showing a rate of growth that increases in a regular way. In fact, such exponential growth curves have been found, but they are not always found. In some instances the growth rate is constant rather than accelerating. In others the growth rate is high at first and then decelerates. In still others the growth rate increases abruptly at some point. Of course, from the cell's point of view these variations might be matters of indifference as long as the cell has doubled in every way by the end of the growth period.

Is growth really an exact doubling? If it were, we could propose some enticing theories. Observation reveals, however, that the doubling and halving process is only approximate. When Dick Killander and Anders Zetterberg of the Royal Caroline Institute in Stockholm weighed sister cells at the time of division in mouse connective tissue, they found an average 10 percent difference in their weight. David M. Prescott, who was then working at the University of California at Berkeley (and later Killander and Zetterberg as well), followed the growth of unequal sister cells. They did not find the small one precisely doubling its lesser volume and the large one doubling its greater volume. Instead both cells tended to reach what I shall call an "adult" size before dividing, and it simply took the smaller cells longer to do so. Finally, the adult size was not absolutely fixed; Killander and Zetterberg found a variation of about 12 percent in the size of mouse connective-tissue cells that were ready to divide.

These findings showed that a mass of cells doubles and halves only in an average way and not in an exact way. This led us to a conclusion that sounds banal but has in fact provided the basis for more incisive research: evidently cells divide when they are "ready," and they are ready only when they have completed certain preparations for division. A further conclusion was that the preparations for division, whatever they may be, are hidden in the overall growth process.

If a cell fails to divide, it does not continue to grow; something limits cell size and halts the growth process. An old and not quite correct theory called attention to the cell nucleus as the limiting factor. In 1908 the German biologist Richard

DOUBLING CELL (*center on opposite page*) **has changed its surface configuration from flat to round as it enters the final phase of the cell cycle: mitosis, or cell division. The surface of the cell is covered by many long, thin projections; a number of these secure the cell to the substrate. The cells are from the ovary of a hamster. The adjacent flattened cells in this scanning electron micrograph have passed through much of the preceding (G_2) phase of the cell cycle. The micrograph was made by Keith R. Porter, David M. Prescott and Jearl F. Frye of the University of Colorado; the cells are seen magnified 8,300 diameters.**



PHASES OF THE CELL CYCLE are shown schematically, beginning at the far left with the first and longest of three "growing" phases: G_1. Space in the diagram is not proportional to time (*see time scale on top*). In a typical mammalian cell the first growing phase takes from six to eight hours, but it can take much longer. The G_1 phase ends and the second, or S, phase begins (*left*) when the cell begins to synthesize DNA. During the six hours of the *S* phase the DNA, the genetic material of the chromosomes, replicates. (The diagram shows the replication of only two chromosomes, whereas a human nucleus would have 46 chromosomes.)

von Hertwig proposed what I call the theory of the critical mass. In Hertwig's view a growing cell eventually reaches a size at which the ratio between the masses represented by cytoplasm on the one hand and the cell nucleus on the other becomes limiting. When that limiting ratio is reached, he proposed, some instability sets in and triggers cell division. The nucleus-cytoplasm relation theory, as Hertwig called it, does not hold up very well when mass alone is considered to be the factor that triggers cell division. As we have seen, individual cells may vary substantially in mass at the time of division. Moreover, cells can be made to divide before they have doubled in size. For example, Mitchison and I once performed an experiment that involved starving the cells of a kind of yeast that grows by cell division rather than by budding. We did this by depriving the yeast of nitrogen; the deprived cells could not grow because in the absence of nitrogen they could not make the proteins that growth requires. What they could still do, however, was convert some of the proteins in their possession into the kinds needed for cell division. Although starved, the yeast cells continued the cell cycle and divided into abnormally small daughter cells.

Even if the critical-mass theory is not tenable, the fact remains that the size of a cell appears to be limited by the capacity of a single cell nucleus to support growth. If in the laboratory we manufacture a cell containing two nuclei, or two sets of chromosomes in a single nucleus, the cell can grow to twice its normal adult size. One example of this behavior is the way mammalian cells in laboratory culture give rise to gigantic cells when exposed to an appropriate dose of ionizing radiation. The irradiated cells go through the normal cycle repeatedly but cannot divide; as the amount of genetic material contained in the cell nucleus increases, so does the size of the cell increase.

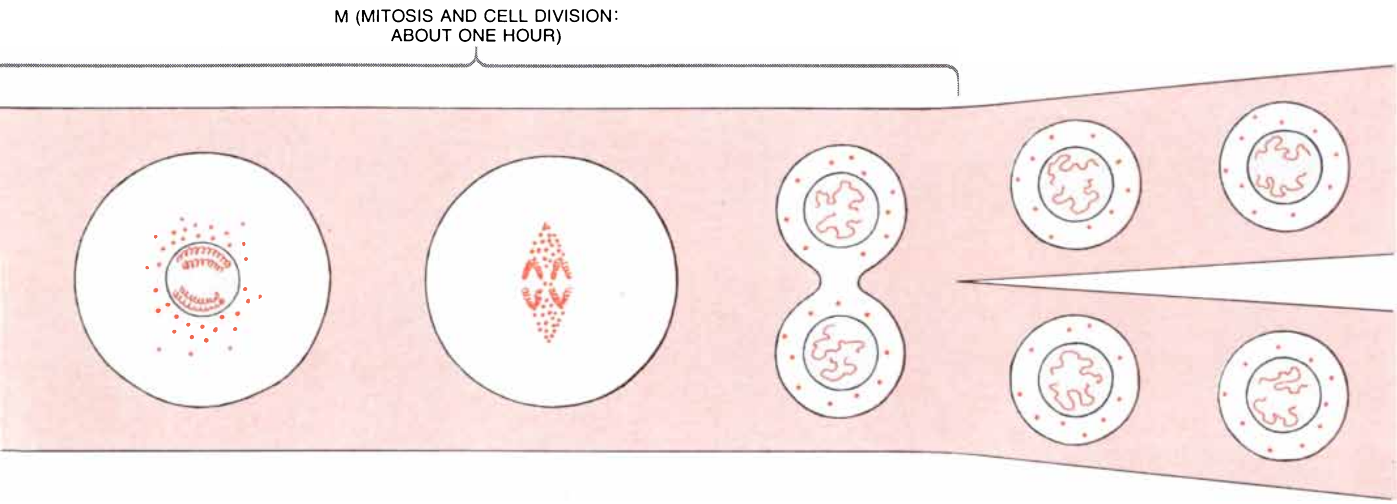
The Tempo of the Cell Cycle

Like other historians, the student of cell-cycle history finds it useful to recognize epochs. Furthermore, like other historians, he must distinguish between the markers provided by real events (for example a revolution or a decisive battle) and the arbitrary markers he invents for his own convenience (for example the Age of Exploration or the Renaissance). When he divides the cell cycle into two parts, a period of growth between cell divisions known generally as the interphase and a second period when division takes place, the student of the cell cycle is recognizing real events. The growing cell is different from the dividing cell in almost every way.

In terms of real time the interphase period is generally much longer than the period of division. In most plant and animal cells the entire cycle is accomplished in a day or less; a typical cycle takes about 20 hours. Of this time only an hour or so is devoted to cell division; the rest of the time is occupied by interphase growth. Some "hothouse" varieties of animal cells that have evolved under laboratory conditions may have a cycle as short as 10 hours, and some one-celled organisms such as yeasts and certain protozoans have an even faster cycle. (I exclude bacteria and other procaryotic organisms whose cells have no nuclei; their cycle can be completed in as little as 20 minutes.) The protozoan *Tetrahymena,* for example, is a favorite among students of the cell cycle because interphase and division together take no more than 150 to 180 minutes.

There is nothing arbitrary about the tempo of the cell cycle. Given optimum temperature and nutrition, the length of the cycle for a given kind of cell is always the same. Furthermore, the tempo is difficult to alter. Of course, under unfavorable conditions it will be slowed. If we wanted to speed it up and make cells grow faster, however, we would not know how to go about it. We have come to think of the duration of the cell cycle as the time required for the execution of a precise program that has been built into each kind of cell.

The cell cycle is a bicycle. One wheel is the wheel of reproduction; its responsibilities include not only the events that take place at division but also the definite preparations that precede division. The preparations are easily characterized: They are not vital to the survival of the cell at the time they take place, and they do not take place at all in cells that are destined not to divide. Perhaps the best example of such a preparatory activity is the replication of the genetic material in the chromosomes. A cell does not need to replicate in order to stay alive. Indeed, the replication has one meaning only: it expresses the intention of the cell to become two cells instead of one. The second wheel of the cell cycle is overall cell growth, the doubling of all the supplies and equipment that keep



When the replication of DNA is terminated, the cell enters the G_2 phase (*left of center*). In five more hours the cell has become twice its original size and is ready to enter the *M* phase (*right of center*). In this phase, which is the period of mitosis, the chromosomes have condensed into visible threads, the contents of the nucleus have mingled with the cytoplasm, and the molecules in the cytoplasm needed to move the chromosomes (*color*) have assembled into a mitotic apparatus. The chromosomes move apart; the parent cell, having doubled, is now halved by being pinched in two, and identical daughter cells begin the cycle all over again.

the cell viable and that will become the dowries of the daughter cells.

Nature tells us some of the reasons the cell-cycle program requires a certain minimum amount of time. Consider the eggs of animals: special kinds of cells that have been fattened up in advance and are full of the kinds of molecules that are usually produced only by growing cells. When the egg cell begins to divide, it divides again and again into smaller and still smaller cells. Its cell cycle, which requires no growth but only preparations for division and the act of division itself, may be completed in an hour or less. Short time intervals such as this tell us that the reproductive "wheel" of the cell cycle can rotate quite rapidly all by itself, but that the two wheels together travel less swiftly because the growth wheel requires time to produce the raw materials and machinery that the reproductive wheel must have.

Today's analysis of the individual phases that make up the cell cycle uses as its principal signposts the replication and division of deoxyribonucleic acid (DNA), the genetic material in the cell nucleus. This is logical; the mere numerical multiplication of cells can only be called true reproduction because the genetic material of the cells is exactly doubled and equally divided in the course of each cycle. Although the process of cell division had been studied intensively for more than half a century, not until about 1950 was it realized that the chromosomes, which carry the DNA, replicate during interphase and only separate during division. It can be proved that DNA doubles during only one part of interphase. That proof came in the 1950's with the development of refined techniques for autoradiography. Newly available radioactive isotopes made it possible to supply cells with radioactive ingredients out of which DNA would be made. The nuclei of those cells in which the DNA was being replicated, and only those cells, became radioactive and registered their presence on a photographic emulsion. Following cells through their cycles, we found that they began to double their DNA at a certain time in interphase and completed the doubling before the end of interphase.

The phases now used to express the timetable of the cell cycle were first defined in 1953 by Alma Howard and Stephen Pelc, who were working at Hammersmith Hospital in London. They called the period of DNA replication the *S* phase; this period provides the landmarks by which other phases are defined. Usually some time elapses between the "birth" of the cell at division and the beginning of DNA replication; that interval is called the G_1 phase. By the time the *S* phase is completed the cell is usually not ready to divide, and the interval between the end of *S* and the onset of cell division is called the G_2 phase. The completion of G_2 is marked by the beginning of the period of cell division variously known as the *M* phase (for mitosis) or the *D* phase (for division).

The duration of the G_1, *S*, G_2 and *M* phases is different in different kinds of cells, but the variations between individual cells of the same kind are small. For that matter, the variations between different kinds of cells are not huge. We can even, with due caution, speak of a typical cell cycle. For example, the timetable of a human cancer cell grown in culture is as follows: eight hours in the G_1 phase, six hours in the *S* phase, four and a half hours in the G_2 phase and one hour in the *M* phase. This does not differ very radically from the timetable of the cells in the root tip of the broad bean: G_1 four hours, *S* nine hours, G_2 three and a half hours and *M* 114 minutes. Even when there are variations between cell cycles, they show some uniformity. For example, the duration of all the phases may vary, but by far the greatest variation is found in the G_1 phase. When a cell cycle is long, most of the prolongation is in the G_1 phase; when a cell cycle is very short, as it is in egg cells, there is no measurable G_1 phase at all. Conversely, the duration of the *S* and G_2 phases can be remarkably constant. For example, in most mammalian cells the *S* phase lasts from six to eight hours and the G_2 phase lasts from three to five hours. The G_1 phase, however, may be as short as a few hours or may take days or even weeks.

The Events of Interphase

In its variability the G_1 phase is trying to tell us something important about the cell cycle, but we do not understand the message very clearly. Hidden somewhere in the cell's overall growth during G_1 must be some key processes that make it possible for the cell to enter the *S* phase and thereby commit itself to division in the future. In speaking of the events of the *S* phase it is an understatement to characterize them solely as "DNA synthesis" or "DNA doubling."

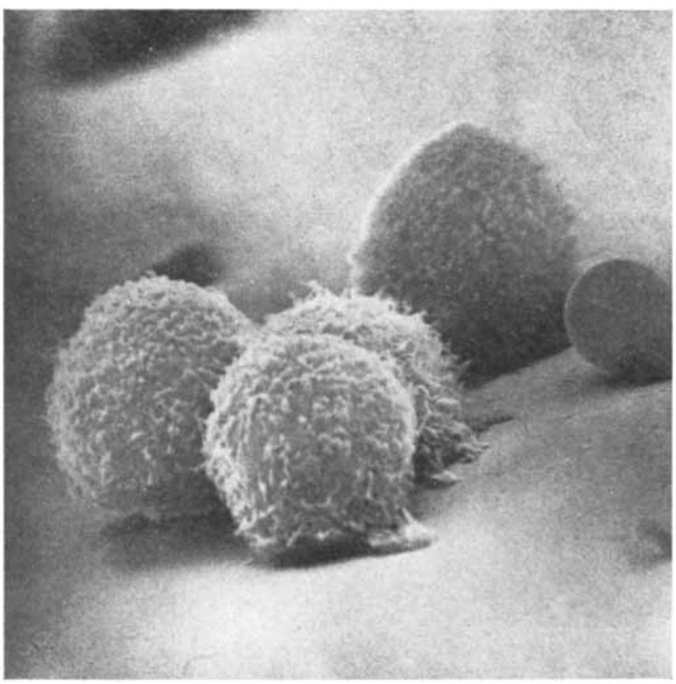
This, to be sure, is the phenomenon we measure, but the significance of the S phase is not in the mere doubling of the amount of the chemical DNA but in the exact replication of the chromosomes. Let us consider the DNA of a human cell at the beginning of the S phase. We can think of it as consisting of 46 threads with an aggregate length of six or more feet, packed into a nucleus that is less than four ten-thousandths of an inch in diameter. The process of replication requires making a copy of each of the 46 threads.

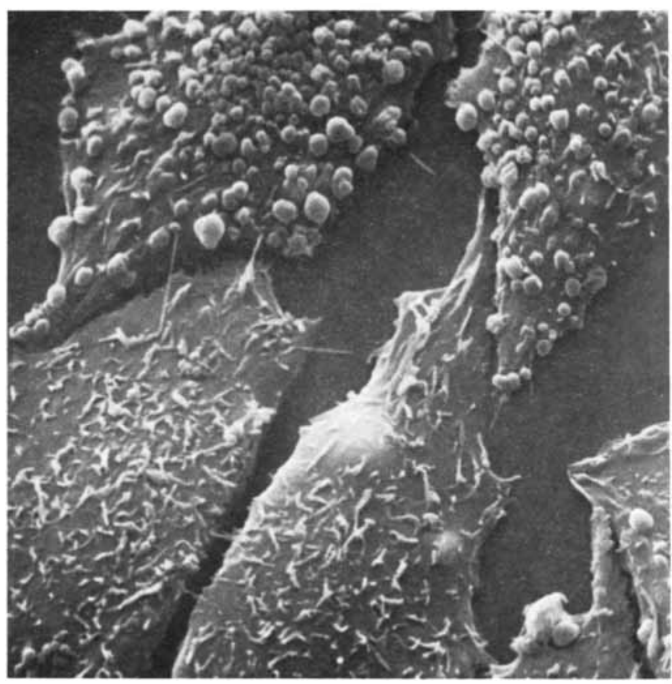
This process would be complicated enough if replication began at one end of each thread and traveled along it to the other end, but that is not how it works. Instead each thread replicates in segments according to a definite program. The segments do not replicate in tandem in any one chromosome, nor does any one chromosome complete its replication before another chromosome begins the process. Because the replication program is reasonably constant we are beginning to know which parts of which chromosome replicate first and which last. We also know that when the last segments are replicated, DNA synthesis shuts down and cannot résume until the next cycle.

Turning briefly to the G_2 phase, all we now know is that it embraces the final steps in the cell's preparations for division. If we arrest the synthesis of proteins by the cell during the G_2 phase, the cell will not divide. We know few other facts about what happens in G_2. The cell is nonetheless telling us in blunt language that it will enter division only when it is good and ready, and that it will not even try to divide if it is not fully prepared.

By following the increase in the amounts of enzymes within the cell it is possible to study the life history of the cell at the chemical level. The enzymes are the substances that define the biochemical capabilities of the cell at any given time. Just as we might visualize cell growth as a steadily accelerating process, so we might conceive of the cell's biochemical history as a record of steadily making more of everything. This is not so; different enzymes have different histories. Mitchison has found that enzymes fall into groups with three kinds of history. Some he calls continuous enzymes; they follow the exponential path and increase in a predictable way as the cell grows. Others Mitchison calls step enzymes; these increase abruptly at certain points in the cell cycle, different step enzymes increasing at different points. The third group Mitchison calls peak enzymes; they increase at certain points in the cycle and decrease thereafter, as if they were meant to carry out a particular job at a particular time and were discarded after the job was done. For example, the enzymes involved in making DNA are evidently peak enzymes that appear only during the S phase.

The existence within the cell of different programs for producing enzymes in the cell cycle leads us to broader questions. The very idea of programs turns our attention to the ultimate source of reproducibility in living things: the genes. It is a tenet of molecular genetics that a cell's production of any enzyme is the end result of a "readout" of a gene for that enzyme. The program for enzyme production during the cell cycle ought to tell us something about the program of the gene readout. Going a step further, we shall recall the consistency of the timetable of the phases of the cycle itself and ask whether genes determine the advance of the cell cycle from one phase to the next. Several geneticists who use one-celled organisms in their experiments have answered the question affirmatively by identifying specific genes that control specific steps in the cell cycle. When the genes are made unworkable by means of mutation, the cell simply "gets stuck" at some point in the cell cycle and cannot advance beyond that point. Leland H. Hartwell of the University of Washington has described a number of such gene mutations in yeasts; each of the mutant genes, when





SURFACE CHANGES in shape and texture are characteristic of hamster ovary cells grown in the laboratory. Daughter cells (*left*), just entering the G_1 phase, still have "hairy" surfaces; the projections may be remnants of the longer strands that are abundant in the M phase (*see photograph on page 54*). The newborn cells are just beginning to spread out on the substrate; the smaller smooth sphere is a fragment of cell cytoplasm and not a living cell. The flattened cells (*left of center*) have reached a late stage in the G_1 phase; their surfaces are covered with microvilli (small hairs) and blebs (blisterlike spheres). Several have established contact with

normal, had acted at a different point in the yeast cell cycle.

Changes in the cell as it goes through the cell cycle can be seen even in its superficial appearance. Working with the scanning electron microscope, Keith R. Porter, David Prescott and Jearl F. Frye of the University of Colorado have made portraits of one kind of cell in various phases of the cycle. The cells were Chinese hamster ovary cells grown on laboratory cultures. During the *M* phase the cells are spherical and are not strongly attached to the substrate they grow on. As the cells enter the G_1 phase they begin to flatten, and the cell surface exhibits bubblelike blisters and fingerlike projections known as microvilli. As the G_1 phase progresses, the margins of the cells become thin and active and present a ruffled appearance. During the *S* phase the appearance of the cell surface changes again: the cells flatten out and their "complexion" becomes smooth. Going through the G_2 phase the cells thicken once more: their surfaces again show microvilli and ruffles, although the blisters of the G_1 phase are much less abundant. Thus at least one kind of cell from one kind of mammal seems to reflect in its physiognomy the sequence of events in progress within it.

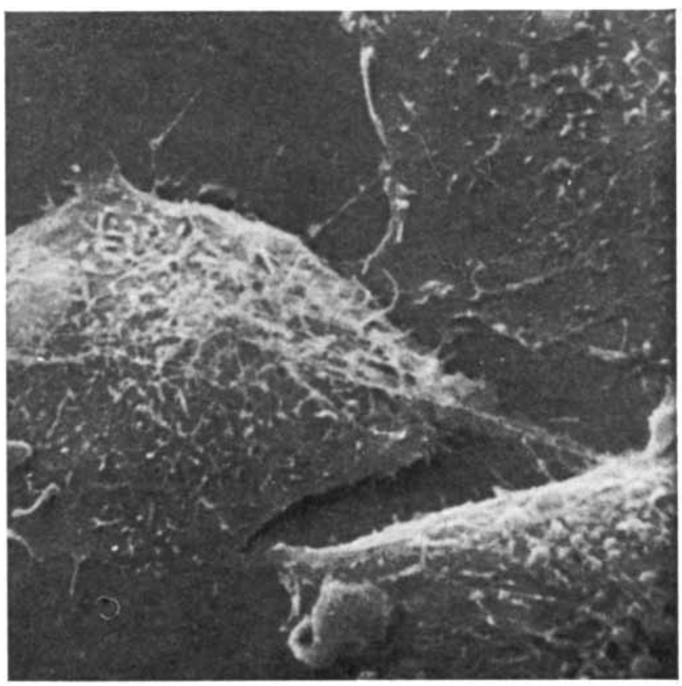
I have characterized the cell cycle as a progress through time; the cell changes radically in the course of the cycle and we are able to recognize distinct phases along the way. The phases, of course, could be of our own invention. The progress of the cell cycle could in fact be perfectly smooth, and the phases we have detected could be no more significant than signposts on a straight highway or numbers on the face of a clock. At the same time it is also possible that the phases we recognize are both real and distinct one from the other and that the progress from one phase to the next requires a definite event; a metaphorical whistle may have to be blown, a switch pulled or a signal given. The highway may be a toll road and the cell may be required to pay its way through a series of gates.

At present much of our evidence favors this second view and suggests that the transition from one phase of the cell cycle to the next calls for a distinct event. We have long been aware of the existence of what I have called "points of no return" in the cell cycle. For example, work in Erik Zeuthen's laboratory in Copenhagen has demonstrated that cell division in the protozoan *Tetrahymena* could be prevented by raising the temperature of the environment before a definite point in the cell cycle. Once that point was passed, however, exposure to high temperatures would not prevent division; evidently the protozoans had passed a point of no return. Moreover, the decisive point was passed some time before any sign of division appeared.

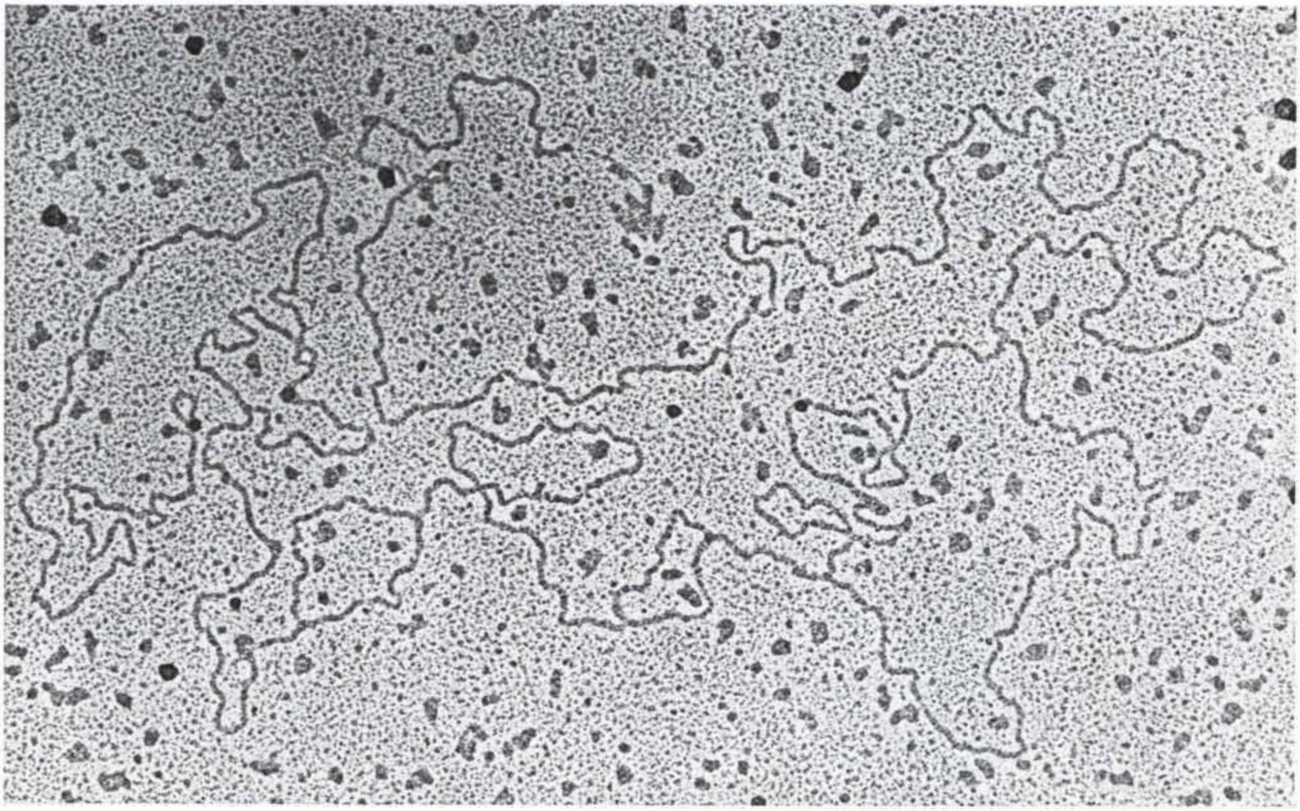
The Two Crucial Signals

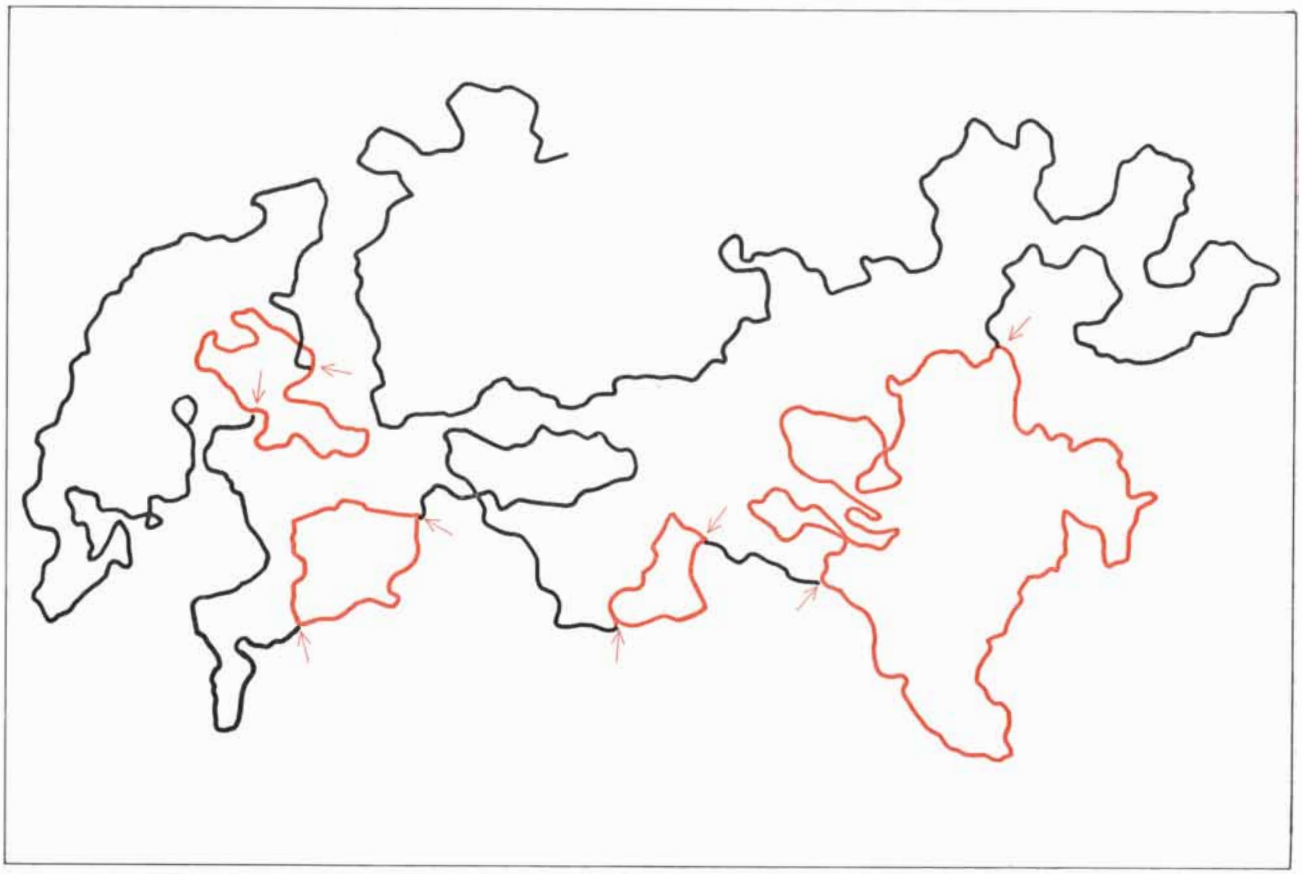
Of all the transitions in the cell cycle that we know the two most important are the transition from the G_1 phase to the *S*, when the replication of the chromosomes starts, and the transition from the G_2 phase to the *M*, when the chromosomes condense and mitosis begins. A discovery in the field of animal virology has provided a new and powerful means of studying the two crucial transitions. It appears that certain viruses, sometimes called Sendai viruses, alter the membranes of cells in such a way that two or more adjacent cells can fuse into one cell. The new cell contains the cytoplasm and the nuclei of all the cells within a common membrane. Even more remarkable, very different kinds of cells, and even cells from different species of animals, can be fused by the action of the virus and form hybrids that not only conduct normal processes but also reproduce as hybrids.

By fusing cells that are in different phases of the cell cycle we can obtain direct answers to the question of whether "switches" exist that move a cell from one phase of the cycle to the next. Working at the University of Colorado Medical Center in Denver, Potu N. Rao and



adjacent cells. The even flatter cells (*right of center*) have entered the *S* phase. Their surfaces are free of blebs and the microvilli are fewer and less prominent. On some parts of the cells' edges vertical "ruffles" (*light areas*) have appeared. Next (*right*) the cells have entered the G_2 phase. Ruffles are still present; the microvilli are more prominent and one cell (*foreground*) shows an array of blebs. The significance of the surface changes is not yet wholly understood, but their coincidence with the four cell-cycle phases suggests that cell membrane is active in the cycle. Like the micrograph on page 54, these are by Keith R. Porter and his colleagues.





CHROMOSOME REPLICATION, the key event in the *S* phase of the cell cycle, is seen in the micrograph (*top*) as it progresses in a fruit-fly egg. The seemingly tangled strand in the micrograph is the DNA from the replicating chromosomes. The accompanying map of the strand (*bottom*) locates the four areas where segments have begun the process of replication (*color*); the arrows point to the forks in the replicating segments. Both the micrograph and the map are the work of David R. Wolstenholme of the University of Utah.

Robert T. Johnson have merged cells in the S phase with cells in the G_1 phase and found that the nuclei of the G_1 cells began to make DNA long before they would normally have done so.

This result clearly demonstrated that a cell in the S phase contains something that triggers DNA synthesis, whereas this something is absent from a cell that has not completed its G_1 phase. Exactly the opposite result could easily have been predicted. For example, it is logical to propose that a cell remains in the G_1 phase because it contains something that keeps DNA synthesis turned off. Such a hypothetical inhibitor, if placed in contact with the nuclei of cells in the S phase, would presumably halt any further synthesis of DNA by the S-phase nuclei.

Further insight into phase-control switches comes from experiments involving the hybrids made by the fusion of kinds of cells whose normal cell-cycle schedules are different. For example, Jennifer Graves of the University of California at Berkeley grew hybrids of mouse and hamster cells. Hamster cells normally have a shorter G_1 phase than mouse cells, so that if each cell nucleus in the hybrid cells had followed its normal timetable, the hamster nucleus would have begun DNA synthesis some time earlier than the mouse nucleus. Actually both nuclei began synthesis at the same time. Moreover, the shift to the S phase occurred after the shorter G_1 interval of hamster cells. Evidently whatever turned on DNA synthesis in the hamster nuclei also turned it on in the mouse nuclei.

In a further experiment Graves followed the S phase to the end. Once the nuclei had entered the S phase, the nucleus of each species followed its own characteristic replication timetable and its own replication program. Those parts of the chromosomes that replicated first in the parent cell replicated first in the hybrids and those that normally replicated last in the parent were last in the hybrid. Evidently the signal for starting the S phase has no control over what happens once the phase has started. Instead the replication program characteristic of each species must somehow be contained in the nucleus of that species' cell.

These experiments with fused cells provide some answers about the role of switches in the cell cycle. They tell us that there is a something that is responsible for starting the replication of chromosomes. The something, moreover, pervades the cell; it can force any nucleus, ready or not, to enter the S phase. If the something is a special kind of molecule that can switch on replication, then its identification will be an important discovery indeed. It is also possible that the signal for starting chromosome replication is not a special molecule but is instead some change in the internal environment of the cell. This is a watery medium; its variables are the concentrations of ions and small molecules and its acidity or alkalinity. Assuming that the ingredients and enzymes needed for chromosome replication are all present, then a change in such simple factors as these could be responsible for switching on the replication process. "Simple" is not the word. We do know that most of the reactions and functions in cells are extremely sensitive to such variables but we do not know much about how they act or how the cell changes them.

An example of such an environmental signal comes from my own recent work with the unfertilized eggs of sea urchins. Egg cells are loaded with the ingredients and enzymes needed for replication. No replication occurs in the unfertilized egg, but within minutes after fertilization the replication process is turned on and running at high speed. I found that I could turn on the process in unfertilized eggs simply by adding a little ammonia to the seawater that bathed them. The ammonia easily penetrates the cell membrane and has the effect of making the cell interior more alkaline.

As I interpret this result, a change in the cells' internal environment (in this instance the environment of a cell that is otherwise ready to begin chromosome replication) switches on the replication process. The ammonia-treated egg nonetheless remains an unfertilized egg and can even be fertilized later, showing that the ammonia treatment is not a substitute for fertilization. Instead the ammonia treatment is a bypass, mimicking something that normally happens only after fertilization. Many other experiments show that the cell cycle is sensitive to simple factors present in the environment external to the cell, for example ions of hydrogen, potassium and calcium. Learning how factors in the external environment act through the cell membrane to influence the cell's internal environment is vital to an understanding of cell-cycle controls.

Virus-induced cell fusion has also been used to study the other key event in the cell cycle: entrance into the M phase. This is the period of cell division, characterized by condensation of the chromosomes. In condensation the chromosomes pack themselves into threads that are visible under the microscope. At Denver, Johnson and Rao have induced cells in the M phase to fuse with cells that are at other phases in the cell cycle. If an M-phase cell is fused with a cell in the G_2 phase, the G_2 nucleus gives rise to normal-looking condensed chromosomes. The chromosomes are double, as would be expected because they had undergone replication in the preceding S phase. When, however, an M-phase cell is fused with a cell in the G_1 phase, thereby forcing the chromosomes of the G_1 nucleus to condense before replication, the chromosomes are seen to be single rather than double.

The third possible combination, the fusion of an M-phase cell with a cell in the S phase, produces a result so startling that it has still to be digested. The chromosomes of the S-phase nucleus do condense, but they condense in small fragments. The effect is called "pulverization" [*see illustration on page 63*]. Evidently in the M phase there is a something that forces chromosomes to condense even when they are obviously not ready to do so. Whatever the something is, it appears to work on any kind of cell; hybrids have been produced that combine the cells of animals that are only remotely related, for example the cells of men and toads, and the experimental results are the same. If one cell contains condensed chromosomes, it can force premature chromosome condensation in the other cell.

The Cycle in Societies of Cells

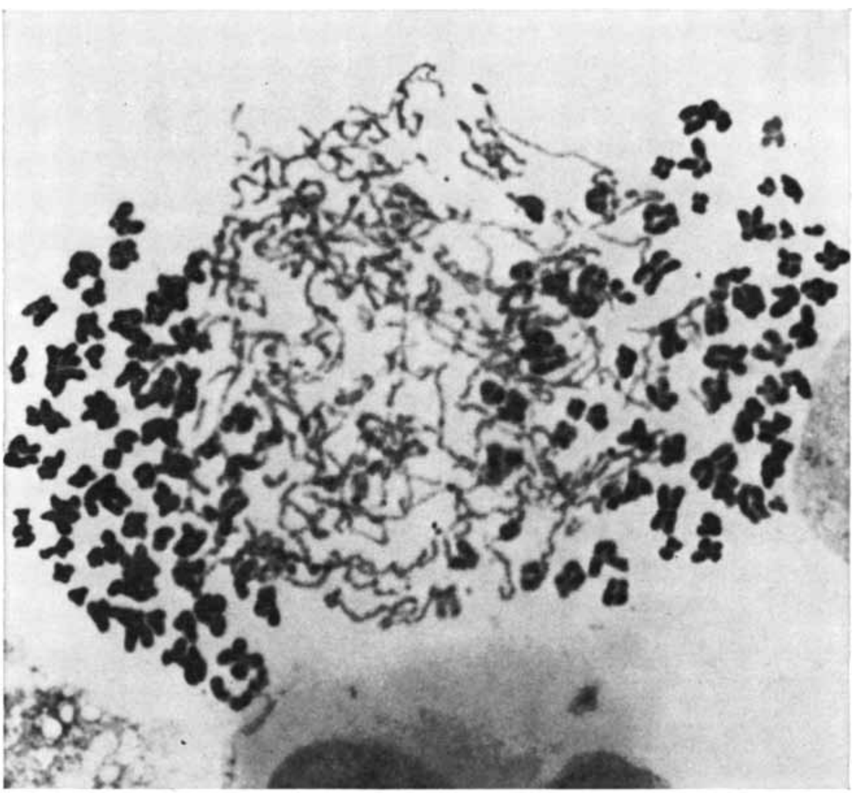
The governance of an organism such as a higher animal, which is a society of cells, dictates that some of the cells in the society will reproduce and that others will not. In general the cells of tissues that perform special services for the entire organism, such as cells of the nervous system and the muscular system, do not reproduce at all. In other tissues, such as skin, the blood-forming system and epithelial linings, the rate of production of new cells is nicely modulated to compensate for the continual loss of old cells. In still other instances, notably in immunity reactions and the healing of wounds, cell multiplication is turned on as a bodily response to external provocation. Only in instances of malignant cancer do body cells defy the governance of the organism and go through their cycles anarchically. There is nothing special about the cell cycle in cancer cells compared with other cells that reproduce; the tempo and the phases of the cycle are the same. It is only that cancer cells repeat the cycle without restraint. The private impulses of the individual cells have become a malignancy for the society of cells as a whole.

Our clues to the means by which reproduction in a society of cells is governed come from detailed analyses of the cell cycle. The chief clue is a very simple one: cells that do not divide never enter the S phase. Conversely, cells that do enter the S phase almost always complete that phase and go on to divide; replication of the chromosomes is evidently a commitment to ultimate division. Thus the control of cell division lies in the supervision of one decision: whether or not a cell will enter into replication. This is true but it is too simple. Cells that are not destined to divide are not cells stuck at the transition from the G_1 phase to the S phase, waiting indefinitely for the signal to begin replication. Some students of the cell cycle prefer to think of them as cells that have not entered the cycle at all. In this view two kinds of cells are distinguished: cycling and noncycling. In the body many kinds of specialized cells are noncycling cells whose progress toward division is permanently shut down. We think of the shutdown as the result of the action of the body's government. Many kinds of cells that have left the cycle can be made to reenter the cycle by putting them under culture in the laboratory, thus removing them from the governance of the body. Agents that cause cancer must cause noncycling cells to reenter the cycle.

At the same time the conversion of noncycling cells to cycling cells can be normal and important. An example is the lymphocyte in higher animals. In these small cells growth and division are shut down most of the time. In immunity reactions, however, lymphocytes begin to grow and divide, producing cells that contribute to antibody formation. Noncycling lymphocytes can also be transformed into cycling ones in the laboratory by exposing them to certain plant proteins known as lectins. When noncycling lymphocytes are thus artificially stimulated, nearly a day passes before they enter the S phase and begin chromosome replication. The transformation has been observed in some detail. For example, Lawrence A. Loeb and his collaborators at the Institute for Cancer Research in Philadelphia have learned that the stimulated lymphocytes need to produce the enzymes required for the replication process as a preliminary to entering the S phase. They are far from being stuck in the G_1 phase, ready to go into the S phase when signaled. When stimulated, it appears, a noncycling cell must first do the things that a cycling cell does in the G_1 phase.

The influence a cell society exercises over the cell cycles of its individual members is observable in quite simple social situations. One example is the "contact inhibition" that is studied in laboratory cultures of mammalian cells. The cells normally grow and multiply on a solid substrate, such as the bottom of a dish or flask. Their growth ceases, however, at the point where a certain population density is reached. This is when the substrate has become covered by a single layer of cells. As Harry Rubin of the University of California at Berkeley has shown, contact inhibition can be overcome by infecting the cell culture with a cancer virus. The infection transforms the cells in a number of ways. They may change in appearance. They are no longer confined to a single-layer conformation but will grow on top of one another. If put into a host animal, they will grow in its body as cancer cells.

As in the body, the control of cell reproduction in contact inhibition is achieved by shutting down the cell cycle before the cell enters the S phase and



CHROMOSOME CONDENSATION can be forced on cells in other phases of the cycle by inducing fusion with a cell in the *M* phase. Here prematurely condensed chromosomes of a cell in the G_1 phase appear as strands of chromatin (*center*). Not having replicated, the strands are single, unlike the doubled chromosomes of the *M* phase cell (*left and right*).

SECOND KIND OF FUSION, between an *M*-phase cell and a cell that has undergone S-phase chromosome replication and entered the G_2 phase, also brings about a premature

chromosome replication begins. Whatever the means may be, it is an effect of the cell society as a whole on its individual members. This can be demonstrated experimentally as follows. If a single-layer sheet of cells is punctured or otherwise wounded, the cell cycle resumes. This happens in all parts of the culture and not only among the cells along the edge of the wound, and reproduction persists until the empty space is filled. The cell cycle is then shut down.

We can hardly imagine that the cycle is shut down merely by the exchange of gentle caresses between adjacent cells. For one thing, different kinds of cells will grow to different degrees of crowding under different conditions before the cycle is halted. For another, contact inhibition can be overcome by making changes in the environment, for example by adding large amounts of blood serum. Moreover, some kinds of cells are not subject to contact inhibition at all.

The study of the cell cycle has led us to some understanding of the governance of cell reproduction as seen from inside the cell. We know that the key event will be found in the supervision of the sequence leading to the switching on of chromosome replication. At the same time studies of the cell cycle in whole organisms, or even in simple societies such as the contact-inhibitable cell cultures, direct us to a further question: How are inside events dictated from the outside?

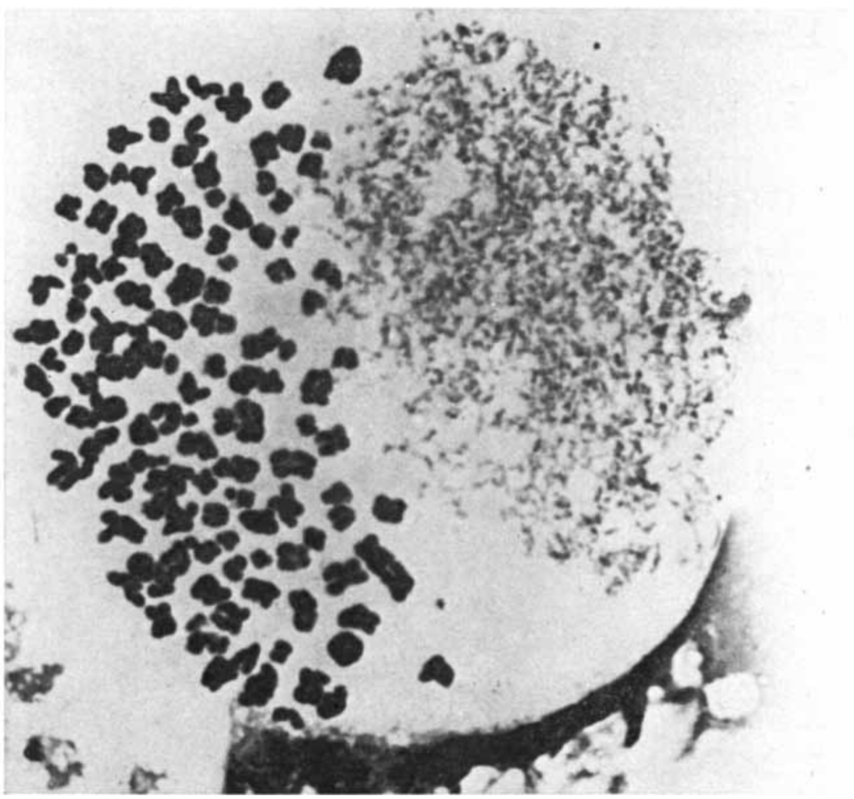
We can narrow the question in various ways. Does the society give its commands to the cell cycle by circulating hormonelike agents? Examples of such possible agents are the chalones investigated by W. S. Bullough of the University of London and the promine and retine being studied by Albert Szent-Gyorgyi at the Marine Biological Laboratory in Woods Hole, Mass. There is also much evidence that variations in the ordinary ions and small molecules of the cell's external environment can dictate cell cycles.

The Cycle and the Cell Surface

Any influences the outside may have on the inside are monitored by the cell membrane. The part played by the cell membrane in the governance of cell reproduction is now being examined with both interest and excitement. The cell biologists' passionate concern with the cell membrane can be summarized in a statement that sounds sarcastic: Whenever our ordinary explanations break down, we like to say that the final explanation will be found in the cell membrane. This is not really sarcasm. The cell membrane is not a wall or a skin or a sieve. It is an active and responsive part of the cell; it decides what is inside and what is outside and what the outside does to the inside. Cell membranes have "faces" that enable cells to recognize and influence one another. The membranes are also communications systems. Things outside a cell do not necessarily act on the cell interior by passing through the membrane; they may simply change the membrane in some way that causes the membrane, in turn, to make changes in the cell interior.

Today cell surfaces are being observed in a variety of ways; for example, we have seen how scanning electron microscopy is used to examine the outer aspect. Another example is the measurement of membrane potentials: the use of electrodes to determine the difference in voltage between the interior and exterior environments. One group of workers in the Department of Embryology of the Carnegie Institution of Washington has found that the membrane potential changes during the cell cycle, and furthermore that the cell cycle will be halt-

condensation of the G_2-phase cell's chromosomes. In this instance, because the cell's chromosomes have undergone replication, strands of chromatin (*center*) are doubled.



THIRD KIND OF FUSION, between a cell in the *M* phase and a cell in the *S* phase, has a peculiar outcome. The prematurely condensed chromatin does not form strands but instead is fragmented and pulverized (*right*). The three experiments were conducted by Potu N. Rao and Robert T. Johnson at the University of Colorado Medical Center in Denver.

ed by changes in the external environment that induce changes in membrane potential. At the University of Miami School of Medicine, Werner R. Loewenstein's group has made electrode studies of the junctions between cells in contact with each other. The channels that the junctions provide allow ions and small molecules to pass between the cells. They have found evidence suggesting that contact-inhibitable cells make junctions, whereas junctions are not formed between cells that are not.

The study of surface differences between cells that are contact-inhibitable and cells that are not is being hotly pursued because one key to the governance of the cell cycle may be found here. The surfaces of the two kinds of cells are different. Max M. Burger of Princeton University was the first to find that the membrane of contact-inhibitable cells reacts with those plant proteins that induce cycling in noncycling lymphocytes. The reaction is easy to observe; when the protein is present, the cell surfaces become sticky and the cells clump together. Cells that are not contact-inhibitable are unaffected by the presence of the protein. Cell-surface studies such as these are progressing rapidly. If the governance of the cell cycle can be revealed by the cell's surface, then there is every reason to look to the cell membrane for the links between the government and the governed.

Today any student of the cell cycle can expect to be asked just how the cell cycle relates to the cancer problem. If understanding plays some part in the solution of problems, it can be said that the cancer problem *is* the cell-cycle problem. Cells either grow and divide without external restraint or they do not. The many kinds of malignant growth that are called cancer have only one lethal attribute in common: all such cells pursue the cell cycle without restraint.

Studies of the cell cycle as viewed from the interior of the cell tell us what must be restrained: the entrance of the cell into the S phase and chromosome replication. The restraint must come from outside the cell. Cells alone do not switch their cycles on and off, but the governance of the cell cycle from the outside is seen in simple collectives, such as contact-inhibited cultures, as well as in the complex society represented by the whole body. Finally, studies of the cell membrane search for the means whereby outside things control inside events. Inside, outside or surface; any of the three could equally well give us the weapons needed for an attack on uncontrolled growth.

THE AGE OF THE ELEMENTS

Study of the formation and spontaneous decay of radioactive nuclei can reveal when certain of the elements were created. From this information one can infer the age of the universe

by David N. Schramm

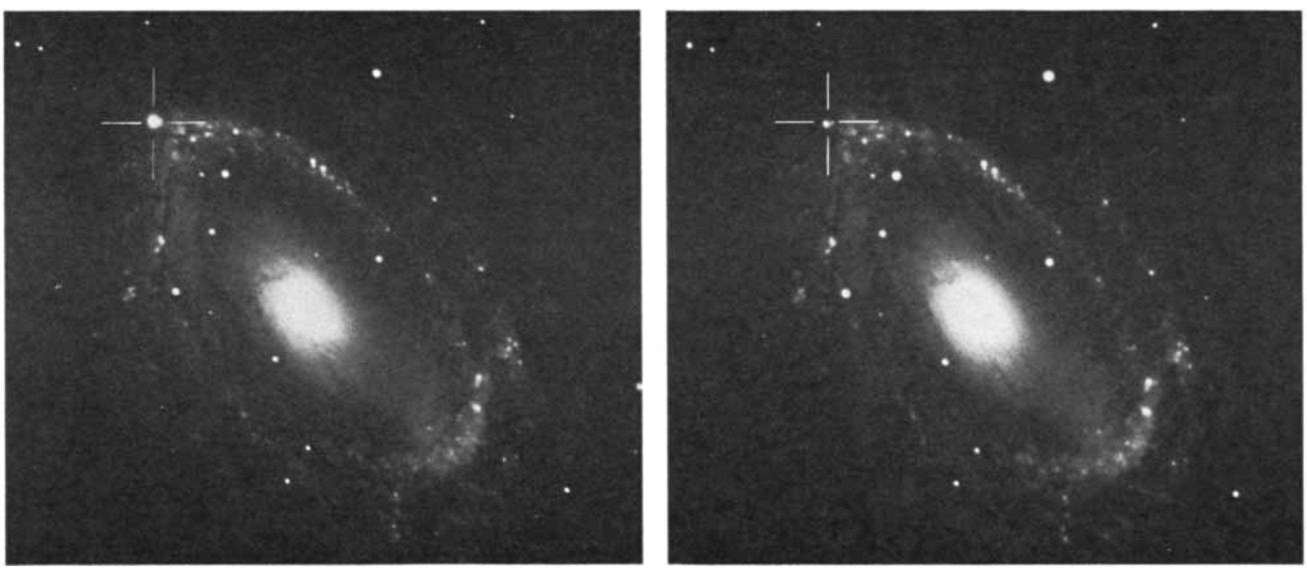
For as long as man has been able to think abstract thoughts he has wondered about the nature of the universe, including its origin and age. The age can be estimated by studying the process of radioactive decay, which has been going on for billions of years. During this period the mechanisms of the process have not changed; more important, the rate of decay has been constant since the elements were first formed. Today we can be confident that a sample of uranium 238, no matter what its origin, will gradually change into lead, and that this transmutation will occur at a rate such that half of the uranium atoms will have become lead in 4.5 billion years. There is no reason to believe that the nature or the rate of this process was any different in the very remote past, when the universe was new.

The history of these countless nuclear events is written in the chemical elements out of which the earth and the rest of the universe are made. By properly interpreting this history we can assign a date to the formation of those elements. From this date we can infer the age of the universe itself. The scientific discipline that is concerned with these techniques is called nucleocosmochronology.

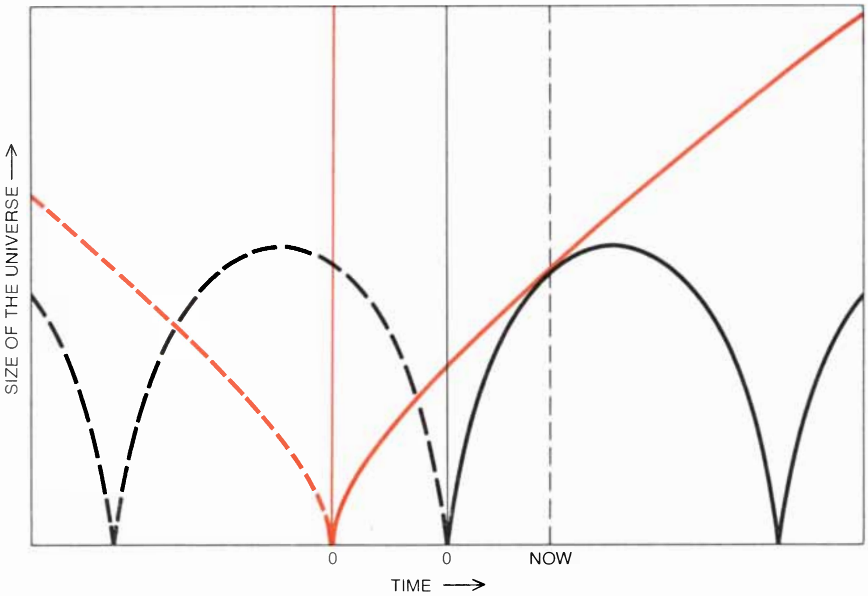
There are several other ways in which the age of the universe can be determined. One celebrated calculation was made by the Anglican archbishop James Ussher in 1658. By adding up the generations enumerated in the Bible he fixed the time of the Creation at 4004 B.C., which would make the present age of the universe 5,978 years. More recent measurements differ from this figure by six orders of magnitude. The uncertainty of the modern calculations, however, is thought to be a few billion years, whereas Archbishop Ussher considered his date accurate to within one year.

Besides nucleocosmochronology the two principal means of dating cosmological events are measurements of the expansion of the universe and observations of the stars in the globular clusters associated with many galaxies, including our own. Both techniques require precise measurements of astronomical phenomena.

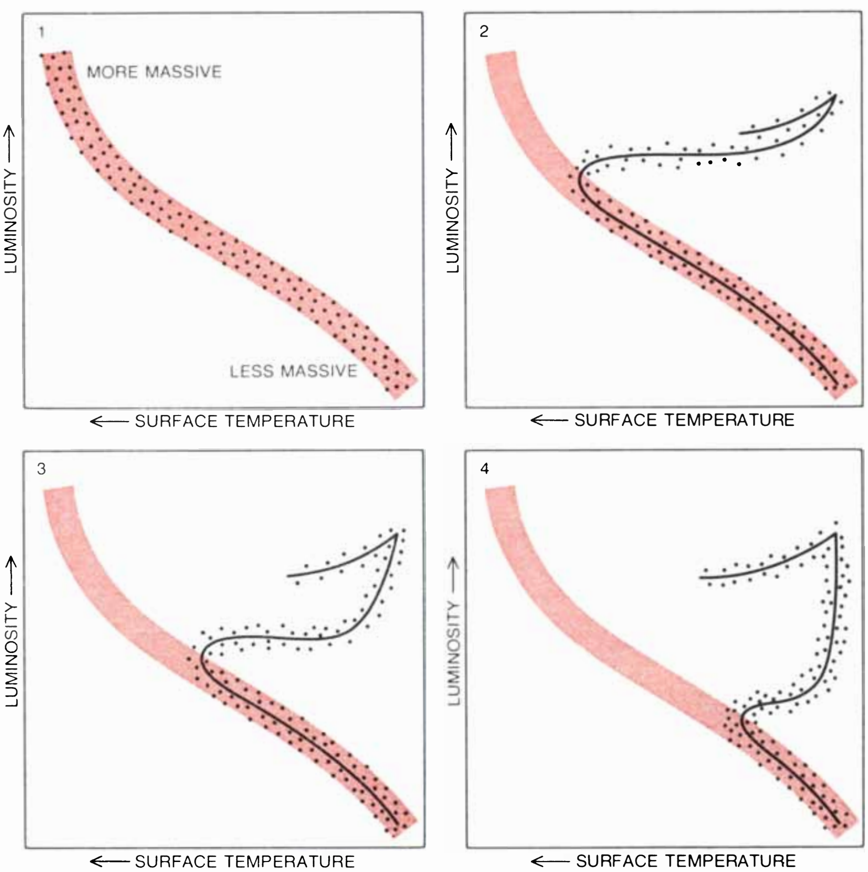
That the universe is expanding was first noted by Edwin P. Hubble in 1929. The expansion is now generally conceded to be explained by some version of the "big bang" theory of cosmology. In this theory the universe is considered to be



SUPERNOVA in a galaxy in the constellation Coma Berenices was at maximum brightness when the photograph at left was made in May, 1940. Eight months later, when the photograph at right was exposed, the supernova had greatly dimmed. All the elements heavier than bismuth are thought to be formed only in supernovas. Because these elements decay radioactively one can calculate when they formed and thus when the first supernovas occurred. The photographs were made with the 100-inch telescope on Mount Wilson.



EXPANSION OF THE UNIVERSE, as measured by the recession of distant astronomical objects, provides one method of dating the "big bang." The rate of expansion can be determined at the point marked "Now." This rate alone, however, cannot reveal whether the universe will continue to expand forever (*colored curve*) or will collapse again under its own gravitation (*black curve*). The maximum age given by this method ranges from 10 to 20 billion years; the accepted best value is about 18 billion years. The minimum age is half the maximum. Events before the big bang (*broken lines*) are in the unknowable past.



AGES OF STARS in globular clusters provide another measure of the age of the universe. In these Hertzsprung-Russell diagrams the diagonal band is the main sequence, which describes the characteristics of all stars when they are formed (*1*). With the passage of time stars leave the main sequence, starting with the most massive ones at the upper left (*2*), then proceeding down the curve (*3, 4*). By finding at what point on the curve stars leave the main sequence today one can estimate when the globular clusters were formed. Icko Iben, Jr., and Robert Rood recently calculated that the globular clusters are 13 ± 3 billion years old.

exploding from a very dense, hot, primordial state. The rival "steady state" theory, which postulates an infinite age, is refuted by the observation of the three-degree black-body background radiation [see "The Primeval Fireball," by P. J. E. Peebles and David T. Wilkinson; SCIENTIFIC AMERICAN, June, 1967].

The universe did not necessarily begin with the big bang. It is possible it was once a rarefied gas that collapsed, exploded and will now expand forever. It is also possible that it began in the condensed state and will return to it; in fact, it could repeat this cycle over and over again. *Our* universe, however, did begin with the primordial explosion, since we can obtain no information about events that occurred before it. The age of the universe, therefore, is the interval from the big bang to the present [*see top illustration at left*].

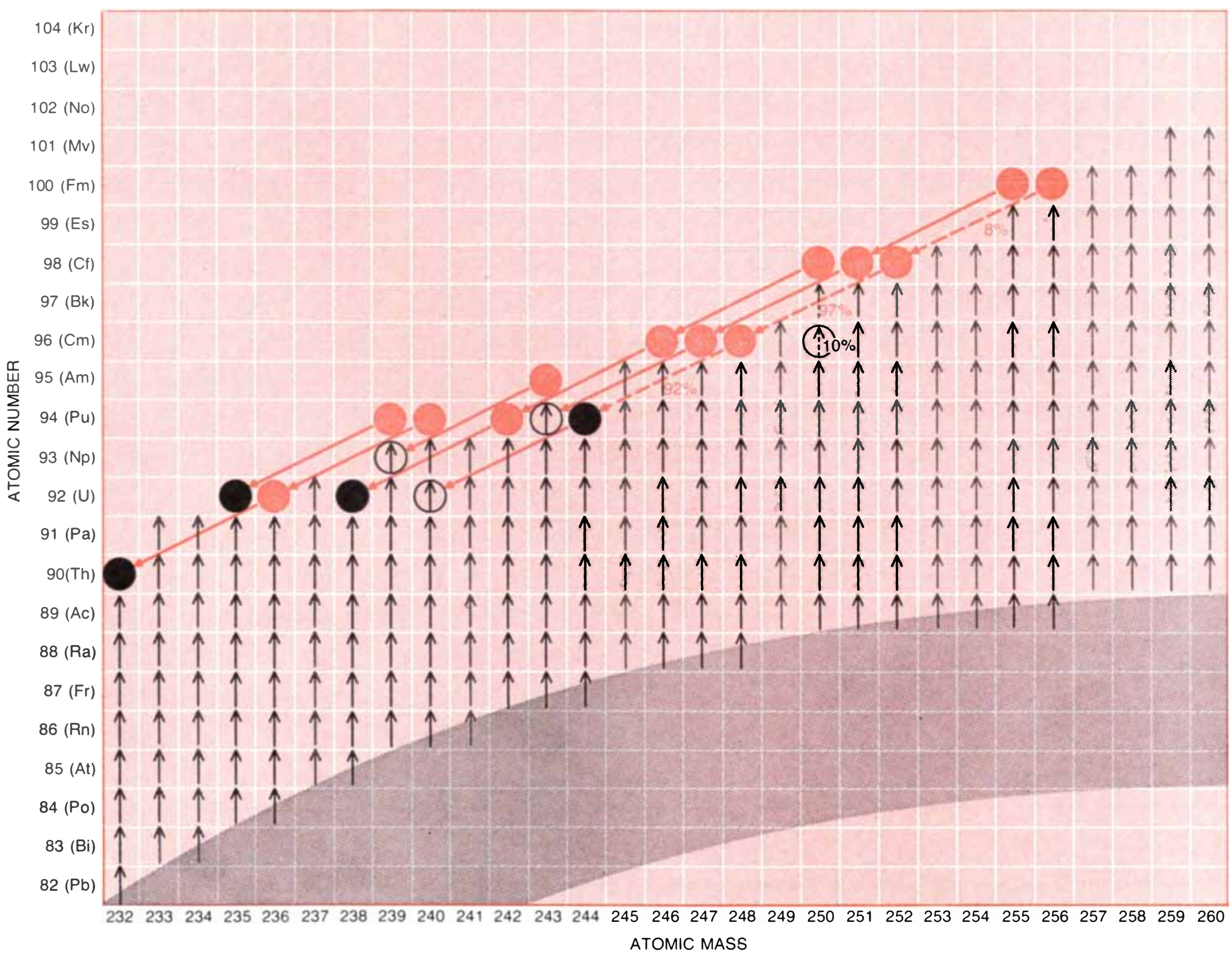
Allan R. Sandage, working with the 200-inch Hale telescope on Palomar Mountain, has recently remeasured the rate of expansion. His calculations indicate that if the rate has been constant since the big bang, then the age of the universe is 18 billion years. This is an upper limit of the age, sometimes called the "Hubble time"; it is the correct age only if the rate of expansion has not changed. Since gravitation tends to diminish the rate of expansion, the universe must be somewhat younger than the Hubble time. This deceleration has not been measured, but Sandage has determined its maximum value by noting that if it exceeded that value, it would cause effects that have not been observed. Thus his observations also supply a lower limit to the age of the universe: it probably cannot be younger than about half the Hubble time. The uncertainty of these measurements is considerable: the Hubble time could range from 10 to 20 billion years, and the minimum age as determined by this method could therefore be from five to 10 billion years.

The second astronomical method of dating the universe depends on the singular properties of the globular clusters [see "Globular-Cluster Stars," by Icko Iben, Jr.; SCIENTIFIC AMERICAN, July, 1970]. Calculations of stellar evolution enable us to estimate the age of a star if we know its mass, its luminosity, its composition and its surface temperature. The composition and surface temperature can be inferred from the star's spectrum; in most cases the mass and luminosity, however, can be determined only if we know the distance to the star. The globular clusters provide a unique set of specimens for these calculations. All the stars in a given cluster were formed at about the same time and out of the same material. Moreover, they are all at approximately the same distance from us.

When the surface temperature of a group of stars is plotted against the luminosity, the relatively young stars fueled by the nuclear fusion of hydrogen form the diagonal band called the main sequence; the more massive stars are in the bright-hot region, the less massive ones in the cool-dim region. This is the Hertzsprung-Russell, or H-R, diagram, named for Ejnar Hertzsprung and Henry Norris Russell. When the hydrogen in the core of a star is exhausted, it moves off the main sequence, soon becoming a red giant. Massive stars leave the main sequence much faster than less massive ones. In a population of stars that were all formed at about the same time one can therefore estimate the age of the stars merely by determining at what point on the H-R diagram the stars are no longer on the main sequence [*see bottom illustration on opposite page*].

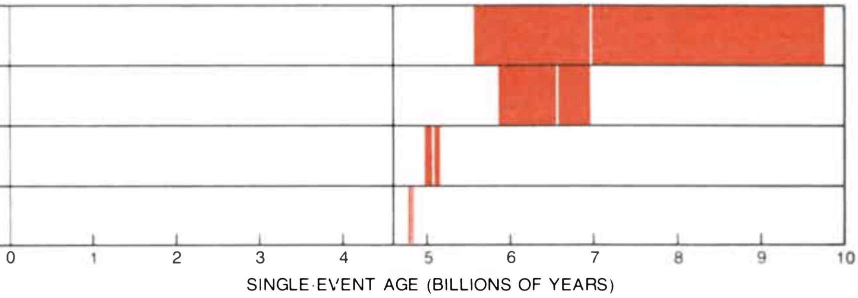
Icko Iben, Jr., and Robert Rood have recently calculated that the globular clusters were formed 13 ± 3 billion years ago, an age consistent with the nine to 18 billion years given by the rate of expansion of the universe. (Of course, the actual age of the universe must be greater than the age of the globular clusters, but it is thought that the clusters formed much less than a billion years after the big bang.)

As in the determination of the Hubble time, there is considerable uncertainty in these calculations. The principal sources of this uncertainty are in the theoretical calculation of stellar evolution and in the determination of stellar composition. Because of the inherent limitations of these methods a third, independent technique can provide an important means of con-



PRODUCTION OF FOUR ELEMENTS that serve as nucleochronometers is portrayed schematically. The chronometers (*solid black circles*) are thorium 232, uranium 235, uranium 238 and plutonium 244. The *r* process, or rapid process (*gray band*), produces nuclei with an excess of neutrons; these nuclei decay by beta emission until they reach a region where the ratio of protons to neutrons is stable. In beta decay the atomic mass does not change but the atomic number increases by one. The beta-decay paths that contribute to the production of nucleochronometers are shown as black arrows. Other nuclei also decay by beta emission; these paths are shown by gray arrows. In the beta-stable region alpha decay predominates; in this process the atomic mass decreases by four and the atomic number decreases by two. Alpha emitters are shown as colored circles; their decay paths are indicated by colored arrows. A few nuclei sometimes decay by spontaneous fission and are thus removed from this mass region. The alpha- or beta-decay paths of the remaining fraction are shown as broken arrows, and the percent of each nuclei that contributes to nucleochronometer production is indicated. Each of the chronometers consists of the nuclei produced at its own atomic mass and of the decay products of its progenitors of higher mass. Because each mass is produced in about equal abundance, the production rates can be estimated by adding up the number of progenitors. For example, U-238 is a product of all the nuclei of mass 238, 242 and 246, and of 10 percent of the nuclei of mass 250, so that its production rate is 3.1. One chronometer (Pu-244) is among the progenitors of another chronometer (Th-232).

NUCLEO-CHRONOMETER	PRODUCTION RATIO	ABUNDANCE RATIO
Th-232/U-238	1.9	2.4
U-235/U-238	1.5	.3
Pu-244/Th-232	.5	.0062
I-129/I-127	1.5	.0001



FOUR PERIODS in the history of the universe are measured by four pairs of nucleochronometers. The production ratios are estimated from theoretical knowledge of nucleosynthesis; the abundance ratios are calculated from the observed abundance of the elements in meteorites today but are given as the ratios present when the solar system condensed. From these data a "single-event age" can be derived for each of the pairs (*white lines*), although in some cases the uncertainty of the measurements is large (*colored bands*). Discrepancies between various single-event ages are caused by differences in the half-lives of the chronometers. Among long-lived chronometers, such as U-238, nuclei may survive from the earliest supernovas. Those that decay quickly, however, such as I-129, must be represented today only by nuclei created shortly before the solar system formed 4.6 billion years ago (*broken line*).

firmation. The use of radioactive nuclei to estimate the age of the universe is such a technique.

In the primordial fireball with which we assume the universe began only the lightest elements—hydrogen, helium and possibly some lithium—could have been produced. All the rest of the elements had to have been synthesized later in stars [see "The Origin of the Elements," by William A. Fowler; SCIENTIFIC AMERICAN, September, 1956].

The stars of the main sequence, including virtually all the stars formed shortly after the big bang, are fueled by the fusion of hydrogen into helium. As we have seen, it is the exhaustion of the hydrogen in the core that causes a star to leave the main sequence. When the star becomes a red giant, its outer envelope expands, but at the same time the core contracts gravitationally and grows hotter until a new fusion reaction is initiated, burning helium into carbon. Eventually the helium too is exhausted and the star is left with a carbon core.

The subsequent events depend on the mass of the star. A star with a mass less than about four times the mass of the sun will become a planetary nebula and leave behind a white dwarf. A star of from about four to about eight solar masses will probably explode as a supernova when fusion begins in the carbon core. In this case a dense core of neutrons may or may not remain after the explosion. This remnant, if it exists, is a neutron star rather than a white dwarf; the pulsars observed in some supernova remnants are thought to be neutron stars.

A star of more than about eight solar masses probably also becomes a supernova. When such a star explodes, it almost certainly produces a remnant, which may be a neutron star or perhaps a black hole. It is in the explosion of these massive stars that the elements from carbon to iron are formed. Iron is the last element in this sequence, however. Fusion reactions involving iron cannot fuel stellar processes; all such reactions absorb energy rather than release it. Iron and certain closely related elements thus represent the end product of the nuclear process from which stars derive their energy.

It should not seem surprising that all the heavier elements are formed in events that, compared to the number of stars, are rather rare. There have probably been fewer than a billion supernovas among the 100 billion stars in the galaxy. More than 98 percent of the universe is still hydrogen and helium, however. Thus if supernovas are a relatively rare phenomenon, so are the elements they produce.

The elements heavier than iron are formed primarily by the capture of neutrons by nuclei of iron and its neighboring elements. The capture occurs in two ways: by a slow process (called the *s* process) or a rapid one (the *r* process).

The *s* process takes place in the envelope of red-giant stars. In it neutrons are added to the nuclei one at a time over long periods, so that only relatively stable nuclei can be formed. Isotopes that decay quickly vanish before another neutron can be absorbed. For this reason the *s* process terminates with lead and bismuth; all heavier elements are radioactive to some degree, and those immediately following lead and bismuth are highly unstable.

The *r* process presumably takes place only in supernovas and probably operates in the region of the explosion just outside the neutron-star or black-hole remnant, where the neutron flux is intense. Under these circumstances neutrons can be absorbed by nuclei in rapid succession, so that regions of great nuclear instability are bridged. All the elements heavier than bismuth are believed to be formed in this way.

The nuclei initially produced by the *r* process are far too rich in neutrons to be stable or long-lived. As a result they immediately begin to seek more stable regions by beta decay. In this process a neutron emits an electron (and a neutrino) and becomes a proton; the result is that the atomic mass remains nearly constant but the atomic number increases by one. Beta decay continues until a stable ratio of neutrons to protons is reached. For example, lead 232, formed in the *r* process, with 26 more neutrons than the stable isotope lead 206, decays through eight beta emissions to thorium 232.

The elements that at present make the most suitable nucleochronometers are formed by the *r* process. Since the process presumably operates only in supernovas, a date establishing the origin of these elements in fact dates the supernovas. Most elements from carbon through iron are also formed in supernovas, so that this method truly yields the age of the elements. Moreover, from such calculations one can derive an age for the universe as a whole. Very massive stars that condensed when the galaxies formed must have become supernovas early, shortly after the big bang. A dating of these first supernovas would be approximately equivalent to the age of the galaxy and of the universe. Even if these stars did not evolve as quickly as is assumed, it is clear that the date of the oldest supernova provides a lower limit to the age of the universe.

In order to make these calculations it is not necessary to know the actual abundance of the elements today or at any time in the past. One need only know the ratio in which a suitable pair of elements is found today (the "abundance ratio") and the ratio in which they were found when first formed (the "production ratio").

The abundance ratios are determined by careful experimental measurement. Terrestrial rocks, although they are fundamentally composed of the same material as the rest of the solar system, usually do not provide suitable samples. In the 4.6 billion years since the solar system condensed, rocks on the earth have formed and re-formed many times. Chemical fractionation has altered their composition so that they are no longer representative of "average" solar-system material. Most of the samples are from meteorites, which have been undisturbed during the entire history of the solar system; recently lunar material has also been used.

The samples are first chemically prepared to isolate the appropriate elements. Because the quantities are typically minuscule, care must be taken to avoid contamination. The sample is then analyzed with a mass spectrometer, an instrument that separates elements and isotopes according to the ratio of mass to charge [*see illustration below*].

The production ratios must be calculated theoretically. In regions between the "magic numbers" of neutrons (2, 8, 20, 28, 50, 82, 126 and probably 184), which yield extraordinarily stable nuclei, the *r* process gives rise to an approximately equal abundance of each atomic mass. The fact that the initial abundances are roughly equal can be used to advantage in calculating the production ratios of the very heavy nuclei, since in this region more than one mass number contributes to each of the nucleochronometers. The multiple contributions are a result of alpha decay, in which a nucleus emits an alpha particle, or helium nucleus, decreasing its mass by four and its atomic number by two. Numerous heavier elements serve as progenitors, decaying to eventually become the nucleochronometers. One chronometer decays in such a way that it eventually contributes to the abundance of another chronometer. Because several atomic masses contribute to the abundance of each of these nucleochronometers the effect of variations from the average, roughly equal, abundance is reduced [*see illustration on page 71*].

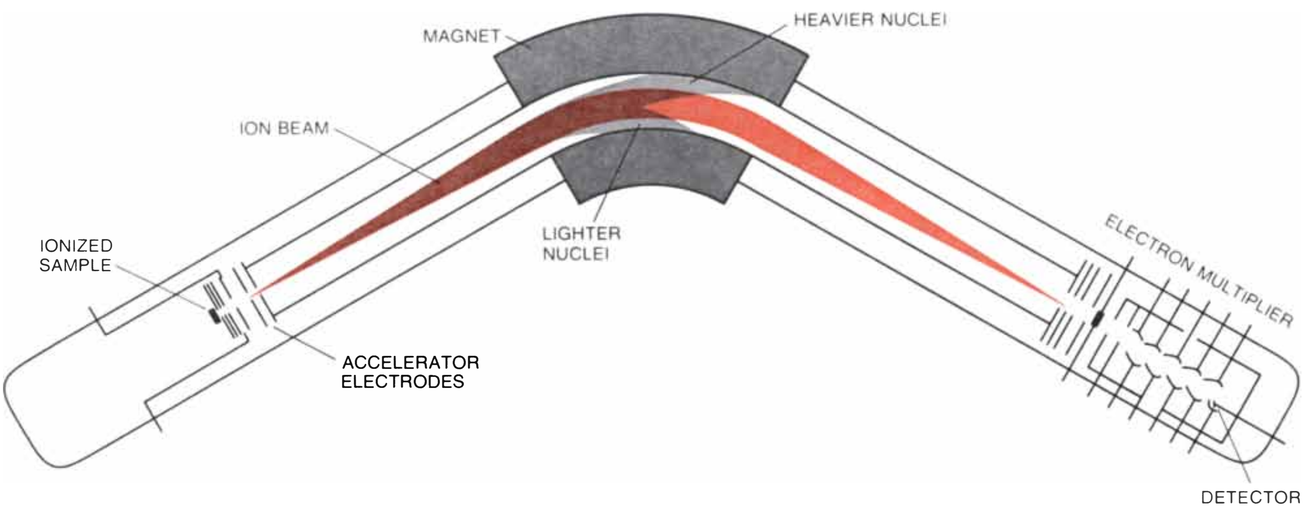
The chronometers of interest in this mass region are thorium 232, uranium 235, uranium 238 and plutonium 244. Their production rates can be calculated simply by adding up the number of their progenitors. For example, U-238 has progenitors of mass 238, 242 and 246 and receives 10 percent of the decay products of mass 250 (the rest of 250 spontaneously fissions). The total thus is 3.1. Pu-244 has as its progenitors all of atomic mass 244, 92 percent of 248, 89 percent of 252 and 7 percent of 256, for a total of 2.9. Th-232 is produced by nuclei of mass 232, 236 and 240, as well as all the progenitors that contribute to 244, for a total of 5.9. Finally, U-235 has these six progenitors: 235, 239, 243, 247, 251 and 255.

The production ratios of these elements can now be simply expressed. The ratio of Th-232 to U-238 is 5.9 divided by 3.1, or 1.9. The Pu-244 to Th-232 ratio is 2.9 divided by 5.9, or .5. In the case of U-235 to U-238 the ratio calculated by this method is 6 divided by 3.1, or 1.9; because nuclei with odd mass numbers (such as U-235) are more easily destroyed by neutron capture, however, this ratio is adjusted to 1.5.

A simple way to see how these ratios can be used to determine the age of the elements is to assume that all the elements were made in one event. This assumption is known to be wrong, but it provides an idealized model of the nuclear dating processes [*see illustration on next page*].

For example, with the chronometer pair Th-232 and U-238 it can be shown by experiment that the abundance ratio today is about 4.0. The calculated production ratio is 1.9; one can determine from the half-lives of the substances that the period required for such a change in ratio is seven billion years. That is the age of the hypothetical single event according to this chronometer pair.

In practice the calculation would not be made in quite this way. Rather than working directly with the present abundance ratio it is more convenient to derive from it the relative abundance at the time the solar system condensed 4.6 billion years ago. In this case the ratio then was 2.4 and the time from the single event to the formation of the solar system is 2.4 billion years. By adding the



MASS SPECTROMETER is the instrument with which the abundance ratios of the nucleochronometers are measured. A pure sample of the element under study is first isolated by chemical separation. The sample is then ionized in a vacuum, and the positively charged ions (which contain the nuclei) are accelerated by a series of electrodes. A magnetic field deflects the beam of ions, so that only those of a particular mass reach the detector electrode, where their presence is detected and their current amplified by the electron multiplier. Ions of greater mass and consequently greater momentum will not be deflected enough to reach the detector; those of smaller mass will be deflected too much. By varying the strength of the magnetic field the abundance of each isotope is measured in succession, producing a mass spectrum. The mass spectrometer thus discriminates between chemically identical substances.

age of the solar system one arrives at the same age of seven billion years. This procedure becomes important when shorter-lived chronometers are used.

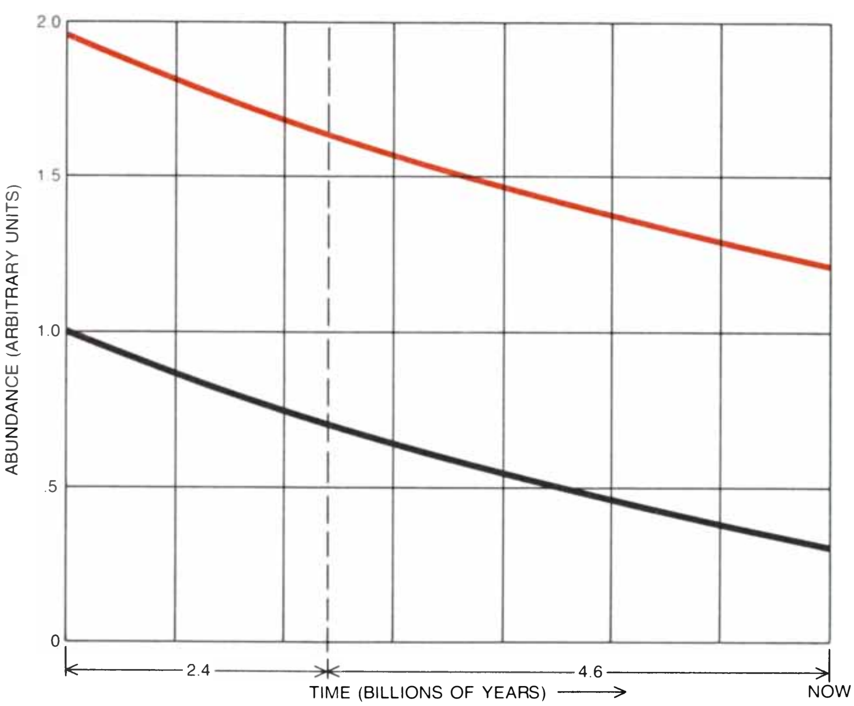
The one-event hypothesis is obviously implausible because it is most unlikely that all the supernovas of the past billions of years exploded at a single moment. The hypothesis is demonstrably wrong because the several nucleochronometer pairs each give a different date for the single event. Indeed, those chronometers with relatively short half-lives would have all but vanished if the only production event had occurred billions of years before the solar system formed. More complicated models are needed.

More than a decade ago William A. Fowler and Fred Hoyle used Th-232, U-235 and U-238 to determine the time scale of nucleosynthesis. With only those chronometers it was not possible to definitely rule out the single-event model. In the past few years, however, techniques have been devised for the use of two new nucleochronometers. They are iodine 129 and plutonium 244, and their observed presence in the solar system cannot be made consistent with a single nucleosynthetic event.

Measurements of the abundance of I-129 and Pu-244 require some subtlety. The half-lives of both substances are much less than the age of the solar system; whatever amount was present at that time, therefore, must by now have been reduced by radioactive decay to virtually zero. (The present abundance would not be zero, although it approaches zero asymptotically and may well be unmeasurable. Investigators at the Los Alamos Scientific Laboratory and the General Electric Company have recently detected traces of Pu-244 in nature.)

Although the plutonium and iodine chronometers themselves are not accessible to us, their decay products are. I-129 decays (by beta emission) to the stable xenon 129. Thus if a given quantity of I-129 was present in the materials that formed a meteorite 4.6 billion years ago, an equal quantity of Xe-129 should be detected in the meteorite today. Many other isotopes of xenon can also be expected. They can be separated by mass spectrometry, however, so that the presence long ago of I-129 in the meteorite would be indicated today by an excess of Xe-129.

John H. Reynolds and his co-workers at the University of California at Berkeley were the first to find this excess of Xe-129 and to prove that it is caused by the decay of I-129. By an ingenious method they were able to determine the ratio of I-129 to the stable isotope I-127 at the time the meteorite was formed, that is, when it first became cool enough to retain xenon gas. To obtain this ratio Reynolds irradiated the sample with neutrons, so that I-127 nuclei would capture a neutron and be converted to I-128, which decays in 25 minutes to Xe-128. Therefore the abundance of I-127 would be related to the excess of Xe-128. He then extracted the xenon by heating the sample in stages, observing the mass spectrum at each stage. At many temperatures only xenon of ordinary origin was driven out. The excess Xe-128 and Xe-129, when they did appear, were driven out at the same time, showing that the original deposits of the two isotopes of iodine were in the same place. It was therefore possible to exclude ordinary xenon from the decay products of iodine. From the ratio of the excess Xe-128 to Xe-129 Reynolds was then able to infer the ratio of the iodine isotopes at the formation of the solar system.

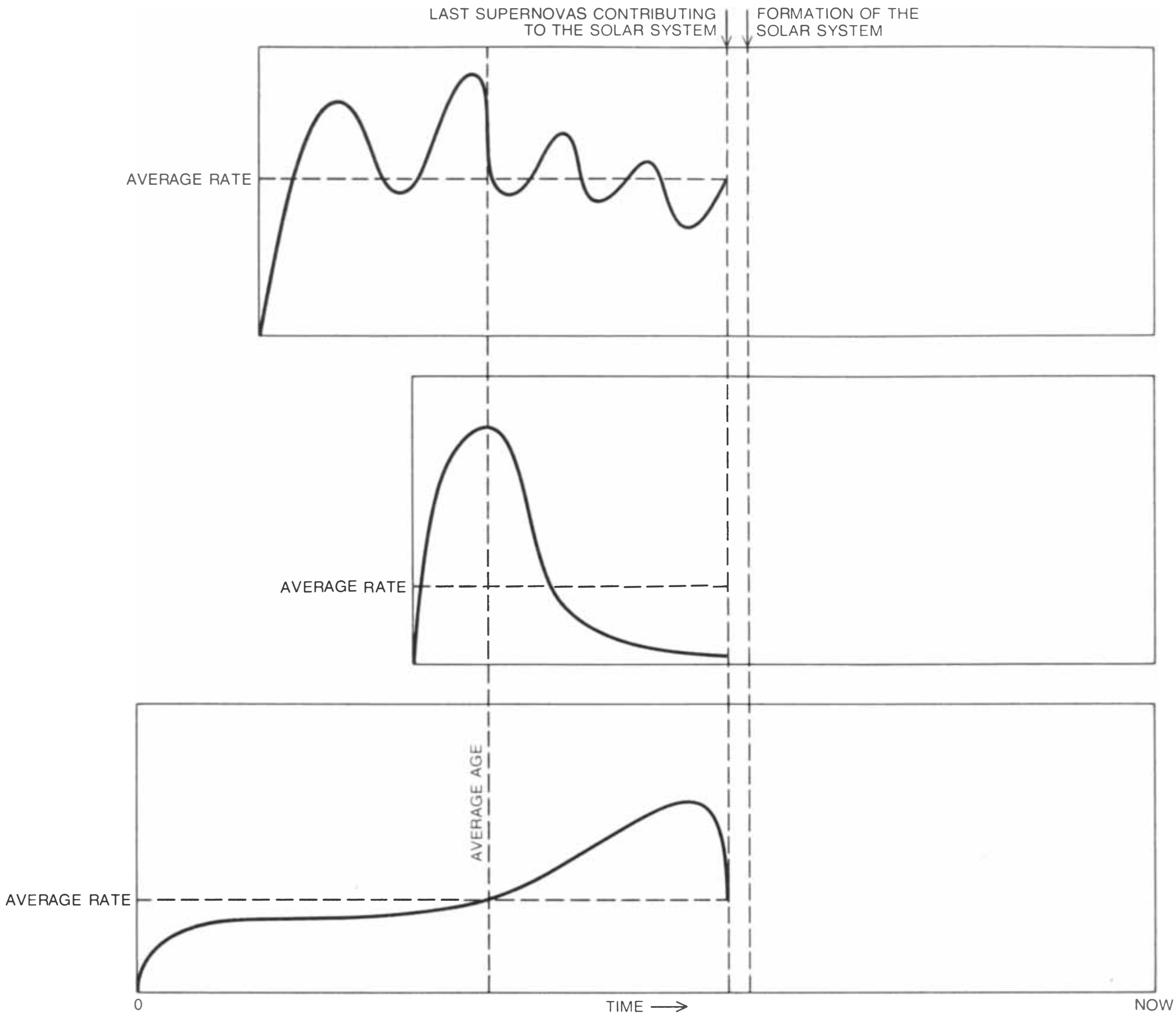


SINGLE-EVENT MODEL of nucleosynthesis is based on the decay of thorium 232 (*colored line*) and uranium 238 (*black line*). These nuclei are assumed to have been formed in a ratio of 1.9 : 1; their present abundance ratio is 4 : 1. From their known rates of decay it can be calculated that a single event seven billion years ago would have produced these ratios. When the solar system formed (*broken line*), the abundance ratio was 2.4 : 1. This model is known to be wrong, but it is an idealized demonstration of nuclear-dating methods.

The identification of Pu-244 as a chronometer also involved xenon isotopes. Pu-244 decays by alpha emission with a half-life of 82 million years. One in every 1,000 nuclei, however, instead of following this decay path spontaneously fissions. Among its fission products is xenon, but only the four heavy isotopes Xe-131, Xe-132, Xe-134 and Xe-136. These isotopes are also produced by the fission of other nuclei, but in proportions that are characteristic of each fissionable substance. It is therefore possible by examining the mass spectrum of the xenon isotopes to identify the fissioning nucleus.

Anomalous isotopic spectra of xenon, suggesting an unknown fission component, were discovered in meteorites by P. K. Kuroda of the University of Arkansas. Gerald J. Wasserburg and his coworkers at the California Institute of Technology found in 1969 that the excess heavy xenon was correlated with fossil fission tracks in meteorites. These tracks represent damage done to the crystal structure by the recoiling fragments of the nucleus. At that time the xenon isotope spectrum for Pu-244 fission was unknown, but this nucleus nevertheless became the only reasonable candidate. Several workers, including myself, began to investigate its possible chronometric applications. The issue was not settled until 1971, 10 years after Kuroda's discovery, when Reynolds' group was able to examine the isotopic spectrum of xenon produced by the fission of a known sample of Pu-244.

From the observation of Pu-244 fis-



HYPOTHETICAL MODELS OF NUCLEOSYNTHESIS postulate three possible distributions of supernovas during the early history of the universe. All three models assume the same "average age" of the elements but incorporate different "average rates" at which the elements were formed. If the number of supernovas has been nearly constant or if the fluctuations have been symmetrical (*top curve*), then the average age is approximately equal to half the total age of the nucleosynthetic process. If the number of supernovas was particularly high in an early period, however (*middle curve*), then the universe is younger than a constant-rate model would suggest. A low early rate (*bottom curve*) yields a somewhat greater overall age. This last model is inconsistent with observation.

sion products in a meteorite to the formulation of an abundance ratio for Pu-244 and Th-232 is a step fraught with uncertainties. The abundance value is to be representative of the entire solar system, yet possible chemical fractionation of the materials in the meteorite could substantially alter it. The value assigned is .0062, but it should be considered only a best estimate, not a definitive determination.

In the case of I-129 special measures must also be taken in estimating the production ratio. Iodine 129 (53 protons, 76 neutrons) is near the neutron magic number 82, and as a consequence the distribution of isotopes produced by neutron capture is far from uniform. It is necessary to determine the production rate by interpolating from the known production rates of stable isotopes in the immediate vicinity, such as tellurium 128 and 130.

This procedure is generally required to determine the production rates of all the lighter radioactive nuclei. One important exception is a chronology developed for the pair rhenium 187 and osmium 187, the longest-lived of the nucleochronometers, by Donald D. Clayton of Rice University. For these elements it has been shown that the production ratio is related to the ratio of the neutron-capture cross sections of the nuclei. The cross section is a measure of the readiness of the nucleus to absorb a neutron. Unfortunately the cross sections have not yet been experimentally measured, and the theoretical estimates being used provisionally introduce large uncertainties into the chronologies derived from this pair.

Using these several nucleochronometers, one can take a more sophisticated approach than the one-event model to discovering the age of the universe. When many events are assumed, an age can be assigned by determining when the sum of the remaining radioactive products from each supernova satisfies the observed abundance at the time the solar system formed. It can be assumed that the products of each supernova mixed in the interstellar gas as the galaxy rotated and that the solar system condensed from this gas. We do not know, however, how many supernovas exploded at different times throughout the history of the galaxy; an estimate can be supplied only by observations of other galaxies. Variations in the rate at which supernovas explode would obviously have an effect on the age indicated by the nucleochronometers. If, for example, there were an unusually large number of supernovas shortly before the solar system formed, then such long-lived chronometers as U-238 would indicate too young an age for the universe; the observed abundance would be not a result of the rate of decay but an artifact of the rate of production.

One approach to this problem is to devise a model for the relative number of supernovas over the history of the galaxy. The one-event model is one possibility, but there are many other candidates. The model is then tested to see if it can produce the observed abundance ratios of all the chronometer pairs.

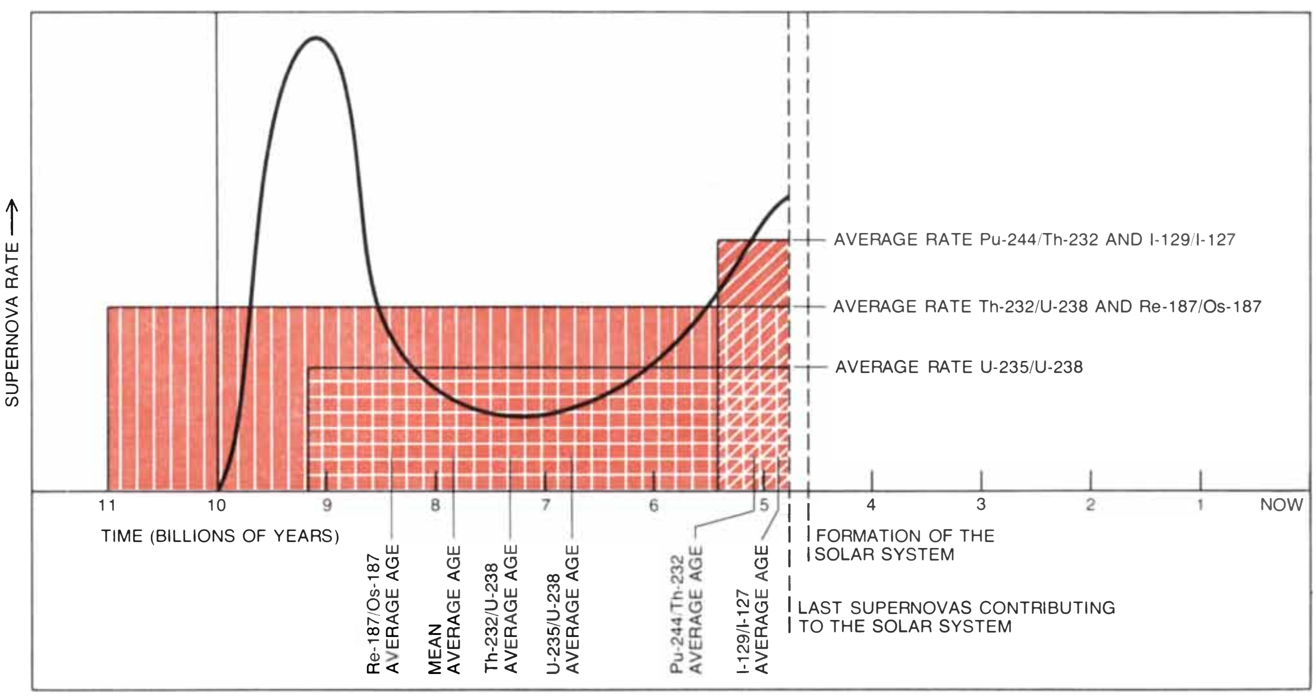
Wasserburg and I have developed an approach to nucleocosmochronology that is independent of models. By using the various time scales of the different chronometers to determine the relative number of supernovas in each period, the valid models are derived from the chronometers themselves. For example, any valid model must include some nucleosynthesis within a few hundred million years of the formation of the solar system or it could not explain the presence of I-129 and Pu-244 decay products. This condition alone precludes a single-event model in which the elements were formed seven billion years ago.

By deriving the model from the data we can estimate not only the time scale of the universe but also the rate of supernovas in the galaxy. (Actually we do not specify supernovas per se but merely the rate of the *r* process. As we have seen, however, it is believed the *r* process occurs only in supernovas.)

Since there have probably been almost a billion supernovas in the galaxy, it is easier to use statistical techniques than to sum the effects of each event. Any distribution of supernovas has an average rate and an average age. There is also a time after which supernovas could no longer contribute material to the solar system. Finally, the oldest nuclei must be older than the average age of the elements.

From the data provided by the nucleochronometers it is possible to find the average age and also to find how the supernova rate at certain times compares with the average rate of supernovas. These calculations are made possible by the differences in the lifetimes of the chronometers. A nucleus with any given half-life must have been formed no earlier than a few of its own half-lives before the solar system condensed or it would not have been present in measurable quantities in the solar-system materials. Thus the abundance of the nucleus yields the mean supernova rate averaged over a few half-lives of that nucleus. The average age of the events producing a particular chronometer pair can be shown to be the same as the age of a hypothetical single event producing that pair.

The very long-lived chronometers, such as Re-187, Th-232 and possibly U-238, persist as long as or longer than the entire duration of nucleosynthesis. The average rate of supernovas contributing to these nuclei is therefore the average supernova rate for the galaxy throughout its history. Moreover, the average, or single-event, age for these nuclei is the average age of the elements. Obviously the ages given by the Th-232 and U-238 pair and those given by Re-187 and Os-187 should be the same. The best estimate for the thorium-uranium pair is 2.4 plus 4.6, or seven, billion



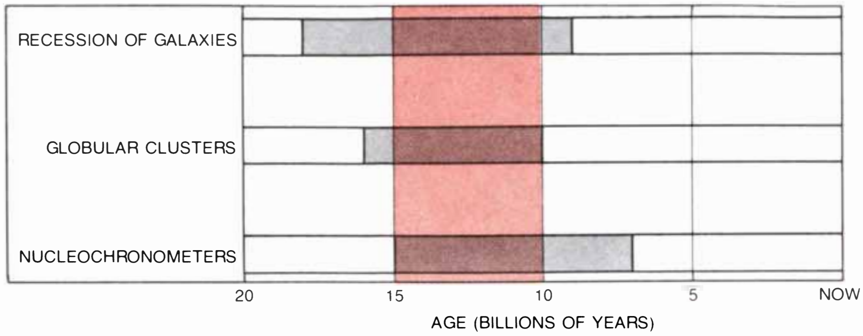
PROBABLE DISTRIBUTION OF SUPERNOVAS is constructed by comparing the information provided by the long-lived and short-lived nucleochronometers. The overall average rate is that given by Th-232/U-238 and Re-187/Os-187 (*vertical hatching*). Because this rate is higher than that indicated by the shorter-lived U-235/U-238 (*horizontal hatching*) it is deduced that the supernova rate must have been high during the early history of the universe and subsequently declined. High rates given by the very short-lived iodine and plutonium chronometers (*diagonal hatching*) suggest another peak in the number of supernovas just before the solar system formed. The resulting curve yields an age for the universe of about 10 billion years. Because of uncertainties in calculations, however, the possibility that the rate of supernovas has been constant cannot be excluded from consideration. A constant rate gives an age of about 11 billion years. The diagram is not drawn to scale; the fluctuations in the number of supernovas are greatly exaggerated.

years, whereas the rhenium-osmium age is 3.5 plus 4.6, or 8.1, billion years. Considering the uncertainties in these calculations, however, the inconsistency is not great.

Variations in the supernova rate can be revealed by the rates given by the shorter-lived chronometers with respect to those of the long-lived, or stable, nuclei. The shortest-lived chronometers convey information about the period just before the solar system formed; by comparing this rate to that given by the longest-lived elements we can find out if the rate of nucleosynthesis was higher or lower than average when the solar system condensed. Because there are two short-lived chronometers (I-129 and Pu-244) we can determine not only the relative rate at this time but also the time between the last nucleosynthetic event and the formation of the objects in the solar system. This interval has been found to be between 100 and 200 million years.

It is a curious fact that this period is comparable to the rotation time of the galaxy. The coincidence has been noted not only by me but also by H. Reeves of Paris and A. G. W. Cameron of Harvard University, who independently suggested that the nucleochronometers may be indirectly measuring the rotation of the galaxy. In the density-wave theory of spiral galaxies stars are thought more likely to form in any given region when the density is high; the period of variation of density would be determined by galactic rotation and would be about 100 million years.

Another possible explanation of this interval is that it represents the time it took the solar system to form after a final supernova triggered the beginning of its condensation. This theory meets with objections, however: all objects in the solar system are known to have solidified within a few million years of one another. Why would they solidify so "suddenly" 100 million years after the event that initiated their formation? It is also thought that the time required for a star such as the sun to form is on the order of 10 million rather than 100 million years. The galactic-rotation theory therefore seems more nearly satisfactory.



AGES OF THE UNIVERSE given by three dating methods are in fairly good agreement. The value given by measurements of the recession of galaxies is the accepted estimate, but it could range from five to 20 billion years. If only ages consistent with all three methods are accepted (*colored band*), the universe must be between 10 and 15 billion years old.

By combining the results obtained from the various chronometers one can reconstruct the probable rates of supernova production over the entire duration of nucleosynthesis [*see illustration on opposite page*]. The best values available today for the abundance and production ratios imply that the number of supernovas was above average in the early universe and that there may have been a smaller peak in the supernova rate shortly before the solar system was formed. The uncertainties are large enough, however, for a uniform rate of nucleosynthesis to satisfy the requirements of the data.

The second, later peak is very sensitive to the abundance of Pu-244. Unfortunately the value for this abundance is still quite uncertain.

The earlier peak is dependent on the production of U-235 with respect to the longer-lived nuclei. If the average rate of production of U-235 is low, then in an earlier period the supernova rate must have peaked strongly if the U-235 data are to agree with the overall average rate of nucleosynthesis as determined by the longer-lived chronometers. Although the magnitude of this peak is by no means certain, it is probable that the early supernova rate was greater than, or at least equal to, the average rate. A high early rate of supernovas suggests that the galaxy may have been brighter in the past. Perhaps there were so many exploding massive stars in the galactic nucleus that the galaxy was a quasar.

The calculated age of the universe as a whole is given by the longest-lived nuclei. The uncertainties of the data for these chronometers indicate that the age of the elements must be between seven and 15 billion years; the best estimate is approximately 10 billion years. Thus the nucleochronologies yield an age that is quite consistent with the ages calculated from the expansion of the universe and from the stellar populations of the globular clusters.

It should be possible in the future to diminish the uncertainty of these calculations considerably. A large improvement could be achieved by experimentally measuring the neutron-capture cross sections that would allow a better estimate of the production ratio of Re-187 and Os-187. Better determinations of the relative abundances of Pu-244, Th-232 and U-238 would also increase accuracy, as would improved calculations of *r*-process production.

In addition techniques for using other nucleochronometers may be developed. Two promising nuclei are samarium 146 (half-life 100 million years) and lead 205 (half-life 15 million years). These nuclei are not produced in the *r* process and they might therefore enable us to see if all nucleosynthetic processes yield similar chronologies.

It was once thought that aluminum 26, with a half-life of 740,000 years, could be used as a chronometer. Analysis of meteorites and moon rocks by me and my co-workers at the California Institute of Technology have shown, however, that no significant amounts of aluminum 26 were present when these objects formed. This result might be expected since the half-life is much less than the 100-million-year interval indicated by the I-129 and Pu-244 chronologies.

These nucleochronologies, coupled with the observed abundances of heavy elements in the stars and theories of star formation, have been used by various investigators in proposing detailed theories of the entire history of the galaxy. Further work on the correlation of nucleochronology with other astronomical information should yield important results in the near future.

It is impressive that three such different techniques as the observation of the recession of the galaxies, the measurement of globular-cluster ages and the calculation of nucleochronologies should yield consistent ages for the universe. If we accept only ages that agree with all three methods, then the universe is from 10 to 15 billion years old. To be able to measure any astronomical number, particularly such an important one, to an accuracy of 33 percent is deeply satisfying.

The Perception of Disoriented Figures

Many familiar things do not look the same when their orientation is changed. The reason appears to be that the perception of form embodies the automatic assignment of a top, a bottom and sides

by Irvin Rock

Many common experiences of everyday life that we take for granted present challenging scientific problems. In the field of visual perception one such problem is why things look different when they are upside down or tilted. Consider the inverted photograph on the opposite page. Although the face is familiar to most Americans, it is difficult to recognize when it is inverted. Even when one succeeds in identifying the face, it continues to look strange and the specific facial expression is hard to make out.

Consider also what happens when printed words and words written in longhand are turned upside down. With effort the printed words can be read, but it is all but impossible to read the longhand words [*see top illustration on page 80*]. Try it with a sample of your own handwriting. One obvious explanation of why it is hard to read inverted words is that we have acquired the habit of moving our eyes from left to right, and that when we look at inverted words our eyes tend to move in the wrong direction. This may be one source of the difficulty, but it can hardly be the major one. It is just as hard to read even a single inverted word when we look at it without moving our eyes at all. It is probable that the same factor interfering with the recognition of disoriented faces and other figures is also interfering with word recognition.

The partial rotation of even a simple figure can also prevent its recognition, provided that the observer is unaware of the rotation. A familiar figure viewed in a novel orientation no longer appears to have the same shape [*see bottom illustration on page 80*]. As Ernst Mach pointed out late in the 19th century, the appearance of a square is quite different when it is rotated 45 degrees. In fact, we call it a diamond.

Some may protest that a familiar shape looks different in a novel orientation for the simple reason that we rarely see it that way. But even a figure we have not seen before will look different in different orientations [*see top illustration on page 81*]. The fact is that orientation affects perceived shape, and that the failure to recognize a familiar figure when it is in a novel orientation is based on the change in its perceived shape.

On the other hand, a figure can be changed in various ways without any effect on its perceived shape. For example, a triangle can be altered in size, color and various other ways without any change in its perceived shape [*see middle illustration on page 81*]. Psychologists, drawing an analogy with a similar phenomenon in music, call such changes transpositions. A melody can be transposed to a new key, and although all the notes then are different, there is no change in the melody. In fact, we generally remain unaware of the transposition. Clearly the melody derives from the relation of the notes to one another, which is not altered when the melody is transposed. In much the same way a visual form is based primarily on how parts of a figure are related to one another geometrically. For example, one could describe a square as being a four-sided figure having parallel opposite sides, four right angles and four sides of equal length. These features remain unchanged when a square is transposed in size or position; that is why it continues to look like a square. We owe a debt to the Gestalt psychologists for emphasizing the importance in perception of relations rather than absolute features.

Since a transposition based on rotation also does not alter the internal geometric relations of a figure, then why does it look different in an altered orientation? At this point we should consider the meaning of the term orientation. What changes are introduced by altering orientation? One obvious change is that rotating a figure would result in a change in the orientation of its image on the retina of the eye. Perhaps, therefore, we should ask why different retinal orientations of the same figure should give rise to different perceived shapes. That might lead us into speculations about how the brain processes information about form, and why differently oriented projections of a retinal image should lead to different percepts of form.

Before we go further in this direction we should consider another meaning of the term orientation. The inverted and rotated figures in the illustrations for this article are in different orientations with respect to the vertical and horizontal directions in their environment. That part of the figure which is normally pointed upward in relation to gravity, to the sky or to the ceiling is now pointed downward or sideways on the page. Perhaps it is this kind of orientation that is responsible for altered perception of shape when a figure is disoriented.

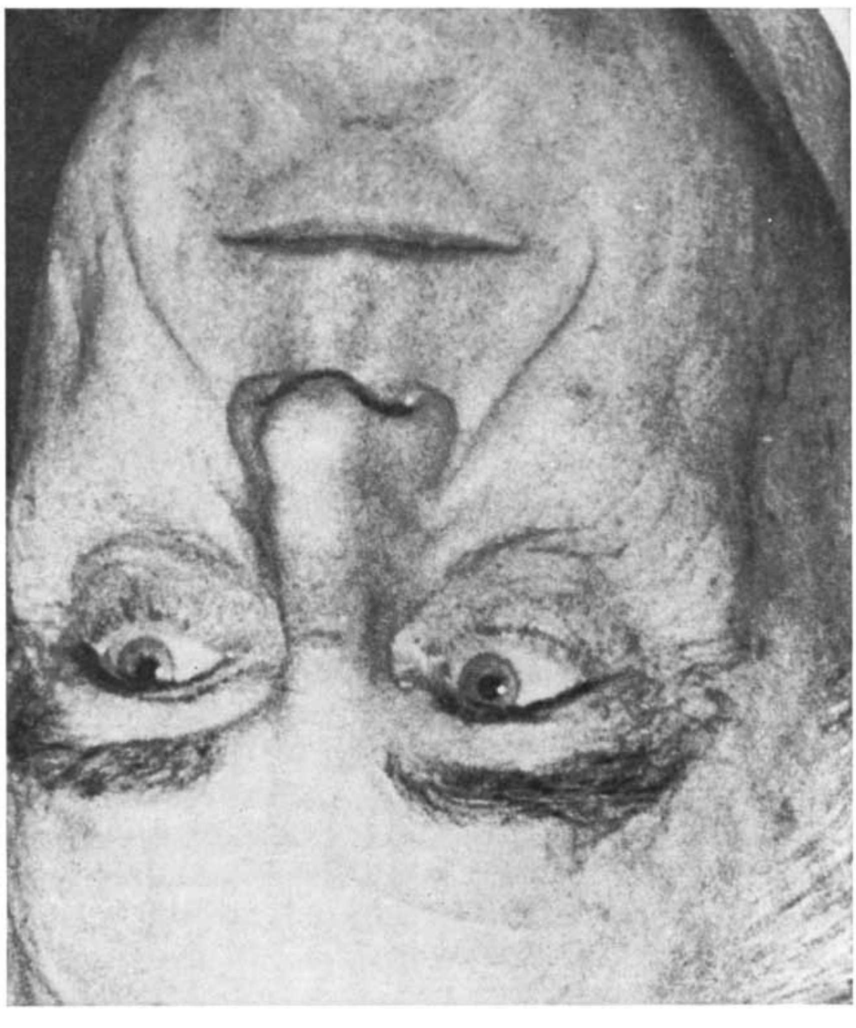
It is not difficult to separate the retinal and the environmental factors in an experiment. Cut out a paper square and tape it to the wall so that the bottom of the square is parallel to the floor. Compare the appearance of the square first with your head upright and then with your head tilted 45 degrees. You will see that the square continues to look like a square when your head is tilted. Yet when your head is tilted 45 degrees, the retinal image of the square is the same as the image of a diamond when the diamond is viewed with the head upright. Thus it is not the retinal image that is responsible for the altered appearance of a square when the square is rotated 45 degrees. The converse experiment points to the same conclusion. Rotate the square on the wall so that it becomes a diamond. The diamond viewed with your head tilted 45 degrees produces a retinal image of a square, but the diamond still looks like a diamond. Needless to say, in these simple demonstrations one continues to perceive correctly where the top, bottom and sides of the figures are even when one's posture changes. It is therefore the change of a figure's perceived orientation in the environment that affects its apparent shape and not the change of orientation of its retinal image.

These conclusions have been substantiated in experiments Walter I. Heimer and I and other colleagues have conducted with numerous subjects. In one series of experiments the subjects were shown unfamiliar figures. In the first part of the experiment a subject sat at a table and simply looked at several figures shown briefly in succession. Then some of the subjects were asked to tilt their head 90 degrees by turning it to the side and resting it on the table. In this position the subject viewed a series of figures. Most of the figures were new, but among them were some figures the subject had seen earlier. These figures were shown in either of two orientations: upright with respect to the room (as they had been in the first viewing) or rotated 90 degrees so that the "top" of the figure corresponded to the top of the subject's tilted head. The subject was asked to say whether or not he had seen each figure in the first session. He did not know that the orientation of the figures seen previously might be different. Other subjects viewed the test figures while sitting upright.

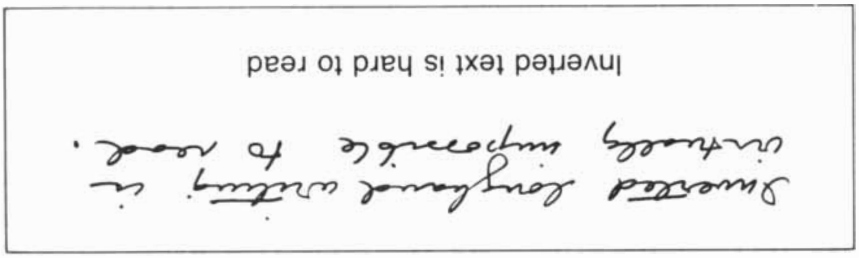
When we compared the scores of subjects who tilted their head with subjects who sat upright for the test, the results were clear. Tilted-head subjects recognized the environmentally upright (but retinally tilted) figures about as well as the upright observers did. They also failed to recognize the environmentally tilted (but retinally upright) figures about as often as the upright subjects did. In other words, the experiments confirmed that it is rotation with respect to the up-down and left-right coordinates in the environment that produces the change in the perceived shape of the figure. It is not rotation of the retinal image that produces the change, since altering the image's orientation does not adversely affect recognition and preserving it does not improve recognition.

In another experiment subjects viewed an ambiguous or reversible figure that could be perceived in one of two ways depending on its orientation. For example, when one figure that looked like a map of the U.S. was rotated 90 degrees, it looked like the profile of a bearded man. Subjects were asked to rest their head on the table when viewing the ambiguous figures. The question we asked ourselves was: Which "upright" would dominate, the retinal upright or the environmental upright? The results were decisive. About 80 percent of the subjects reported seeing only the aspect of the ambiguous figure that was environmentally upright, even though the alternative was upright on their retina [*see bottom illustration on page 82*].

Why does the orientation of a figure with respect to the directional coordinates of the environment have such a profound effect on the perceived shape of the figure? The answer I propose is that perceived shape is based on a cognitive process in which the characteristics of the figure are implicitly described by the perceptual system. For example, the colored figure at the left in the top illustration on page 81 could be described as a closed figure resting on a horizontal base with a protrusion on the figure's left side and an indentation on its right side. The colored figure to the right of it, although it is identical and only rotated 90 degrees, would be described quite differently, as being symmetrical with two bumps on the bottom and with left and right sides more or less straight and identical with each other. I am not suggesting that such a description is conscious or verbal; obviously we would be aware of the descriptive process if it were either. Furthermore, animals and infants who are nonverbal perceive shape much as we do. I am proposing that a process analogous to such a description does take place and that it is not only based on the internal geometry of a figure but also takes into account the location of the figure's top, bottom and sides. In such a description orienta-



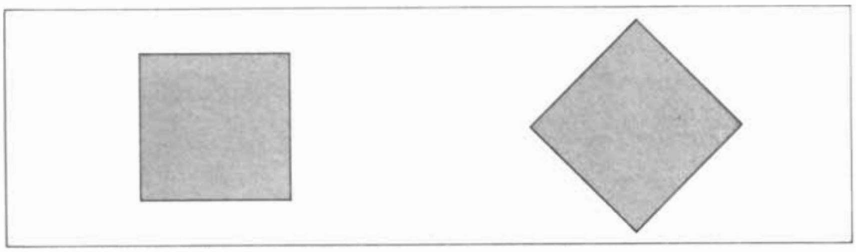
INVERTED PHOTOGRAPH of a famous American demonstrates how difficult it is to recognize a familiar face when it is presented upside down. Even after one succeeds in identifying the inverted face as that of Franklin D. Roosevelt, it continues to look strange.



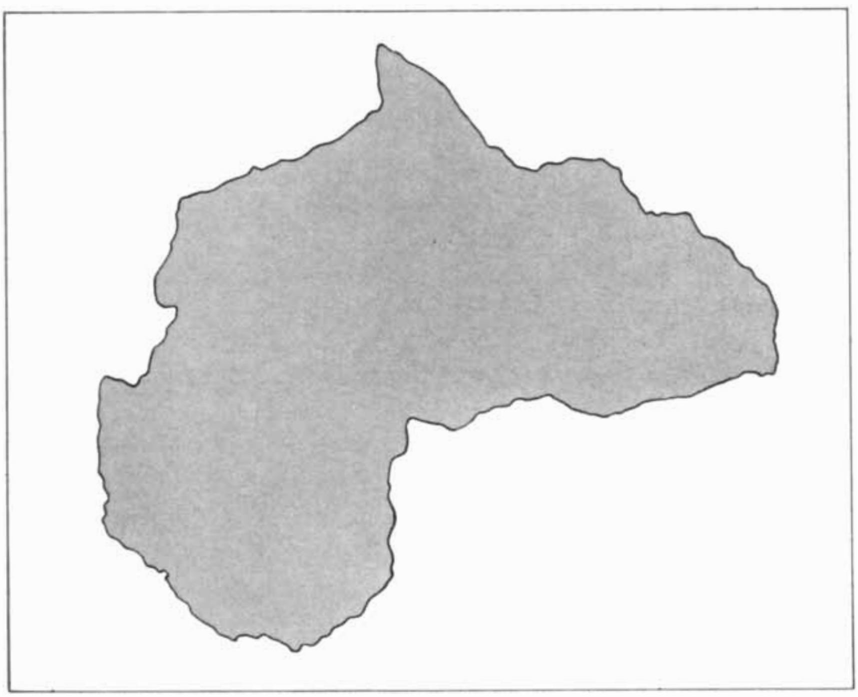
INVERTED WORDS are difficult to read when they are set in type, and words written in longhand are virtually impossible to decipher. The difficulty applies to one's own inverted handwriting in spite of a lifetime of experience reading it in the normal upright orientation.

tion is therefore a major factor in the shape that is finally perceived.

From experiments I have done in collaboration with Phyllis Olshansky it appears that certain shifts in orientation have a marked effect on perceived shape. In particular, creating symmetry around a vertical axis where no symmetry had existed before (or vice versa), shifting the long axis from vertical to horizontal (or vice versa) and changing the bottom of a figure from a broad horizontal base to a pointed angle (or vice versa) seemed to have a strong effect on perceived shape. Such changes of shape can result from only a moderate angular change of orientation, say 45 or 90 degrees. Interestingly enough, inversions or rotations of 180 degrees often have only a slight effect on perceived shape, perhaps because such changes will usually not alter perceived symmetry or the perceived orientation of the long axis of the figure.



SQUARE AND DIAMOND are two familiar shapes. The two figures shown here are identical; their appearance is so different, however, that we call one a square and the other a diamond. With the diamond the angles do not spontaneously appear as right angles.



"UNFAMILIAR" SHAPE shown here becomes a familiar shape when it is rotated clockwise 90 degrees. In a classroom experiment, when the rotated figure was drawn on the blackboard, it was not recognized as an outline of the continent of Africa until the teacher told the class at the end of the lecture that the figure was rotated out of its customary orientation.

There is one kind of orientation change that has virtually no effect on perceived shape: a mirror-image reversal. This is particularly true for the novel figures we used in our experiments. How can this be explained? It seems that although the "sides" of visual space are essentially interchangeable, the up-and-down directions in the environment are not. "Up" and "down" are distinctly different directions in the world we live in. Thus a figure can be said to have three main perceptual boundaries: top, bottom and sides. As a result the description of a figure will not be much affected by whether a certain feature is on the left side or the right. Young children and animals have great difficulty learning to discriminate between a figure and its mirror image, but they can easily distinguish between a figure and its inverted counterpart.

Related to this analysis is a fact observed by Mach and tested by Erich Goldmeier: A figure that is symmetrical around one axis will generally appear to be symmetrical only if that axis is vertical. Robin Leaman and I have demonstrated that it is the perceived vertical axis of the figure and not the vertical axis of the figure's retinal image that produces this effect. An observer who tilts his head will continue to perceive a figure as being symmetrical if that figure is symmetrical around an environmental vertical axis. This suggests that perceived symmetry results only when the two equivalent halves of a figure are located on the two equivalent sides of perceptual space.

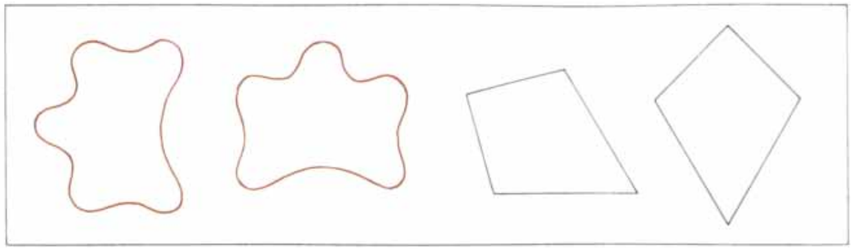
If, as I have suggested, the description of a figure is based on the location of its top, bottom and sides, the question arises: How are these directions assigned in a figure? One might suppose that the top of a figure is ordinarily the area uppermost in relation to the ceiling, the sky or the top of a page. In a dark room an observer may have to rely on his sense of gravity to inform him which way is up.

Numerous experiments by psychologists have confirmed that there are indeed two major sources of information for perceiving the vertical and the horizontal: gravity (as it is sensed by the vestibular apparatus in the inner ear, by the pressure of the ground on the body and by feedback from the muscles) and information from the scene itself. We have been able to demonstrate that either can affect the perceived shape of a figure. A luminous figure in a dark room will not be recognized readily when it is rotated to a new orientation even if the observer is tilted by exactly the same amount. Here the only source of information about directions in space is gravity. In a lighted room an observer will often fail to recognize a figure when he and the figure are upright but the room is tilted. The tilted room creates a strong impression of where the up-down axis should be, and this leads to an incorrect attribution of the top and bottom of the figure [see "The Perception of the Upright," by Herman A. Witkin; SCIENTIFIC AMERICAN, February, 1959].

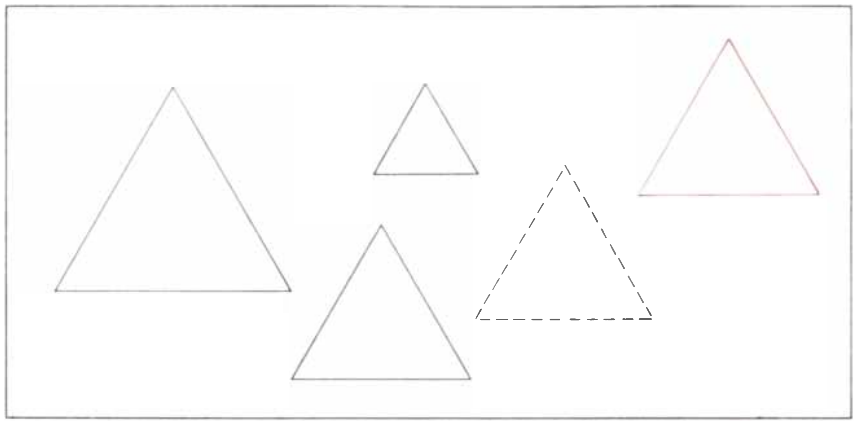
Merely informing an observer that a figure is tilted will often enable him to perceive the figure correctly. This may explain why some readers will not perceive certain of the rotated figures shown here as being strange or different. The converse situation, misinforming an observer about the figures, produces impressive results. If a subject is told that the top of a figure he is about to see is somewhere other than in the region uppermost in the environment, he is likely not to recognize the figure when it is presented with the orientation in which he first saw it. The figure is not disoriented and the observer incorrectly assigns the directions top, bottom and sides on the basis of instructions.

Since such knowledge about orientation will enable the observer to shift the directions he assigns to a figure, and since it is this assignment that affects the perception of shape, it is absolutely essential to employ naïve subjects in perception experiments involving orientation. That is, the subject must not realize that the experiment is concerned with figural orientation, so that he does not examine the figures with the intent of finding the regions that had been top, bottom and sides in previous viewings of it. There are, however, some figures that seem to have intrinsic orientation in that regardless of how they are presented a certain region will be perceived as the top [*see top illustration on next page*]. It is therefore difficult or impossible to adversely affect the recognition of such figures by disorienting them.

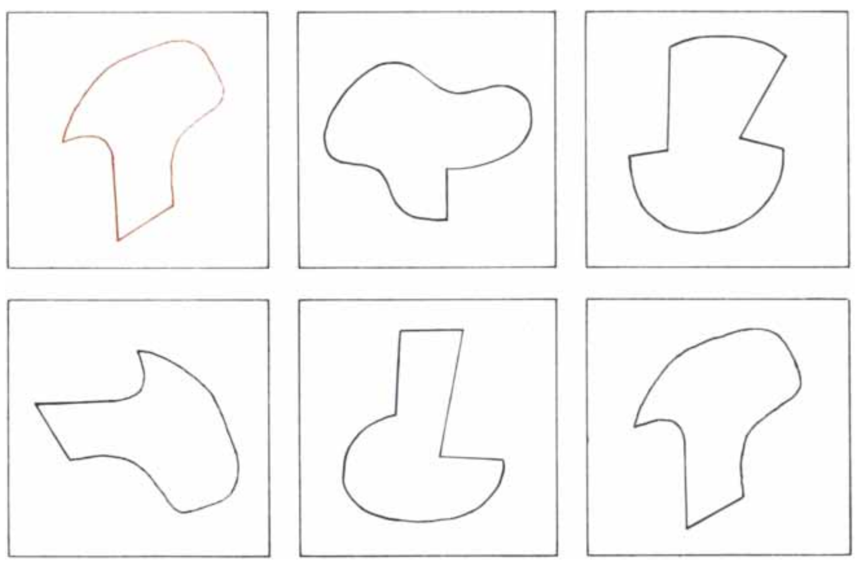
In the absence of other clues a subject will assign top-bottom coordinates according to his subjective or egocentric reference system. Consider a figure drawn on a circular sheet of paper that is lying on the ground. Neither gravity nor visual clues indicate where the top



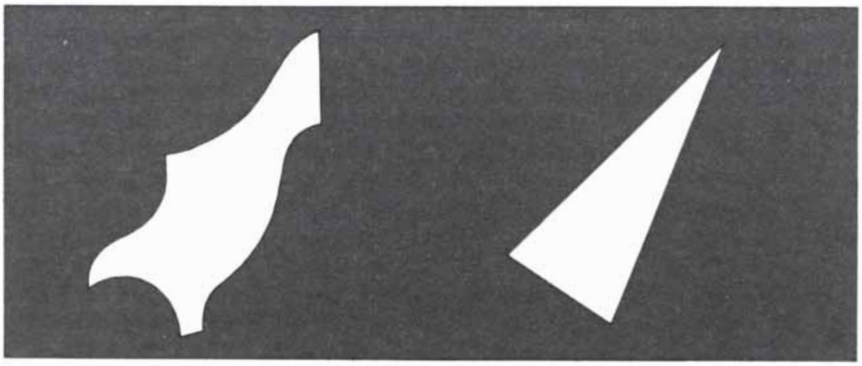
NOVEL OR UNFAMILIAR FIGURES look different in different orientations, provided that we view them naïvely and do not mentally rotate them. The reason may be the way in which a figure is "described" by the perceptual system. The colored figure at left could be described as a closed shape resting on a horizontal base with a protrusion on its left side and an indentation on its right side. The colored figure adjacent to it, although identical, would be described as a symmetrical shape resting on a curved base with a protrusion at the top. The first black figure could be described as a quadrilateral resting on a side. The black figure at right would be described as a diamondlike shape standing on end.



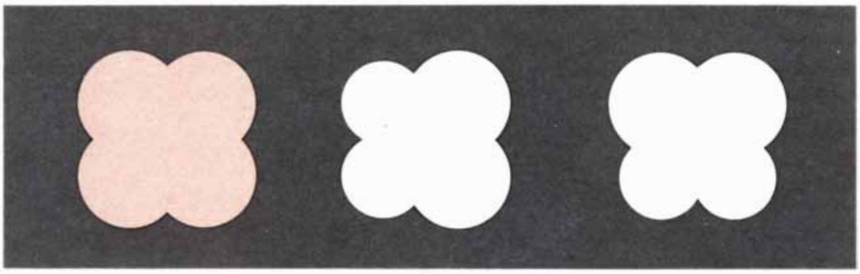
ALTERATION IN SIZE, color or type of contour does not change the perceived shape of a triangle. Even varying the location of the triangle's retinal image (by looking out of the corner of your eyes or fixating on different points) does not change perceived shape.



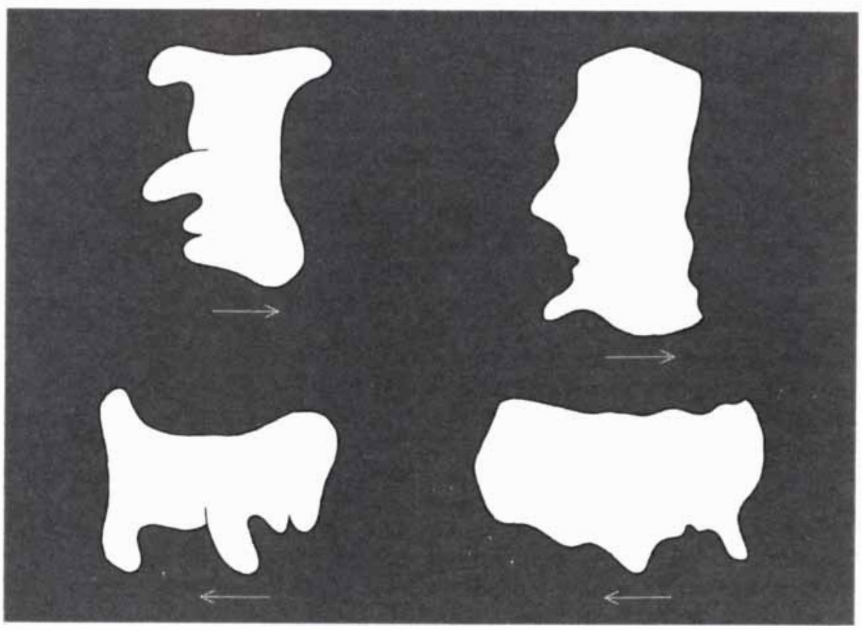
ROTATION OF RETINAL IMAGE by tilting the head 90 degrees does not appreciably affect recognition of a novel figure (*color*). Subjects first viewed several novel targets while sitting upright. Then they were shown a series of test figures (*black*) and were asked to identify those they had seen before. Some subjects tilted their head 90 degrees; others viewed the test figures with their head upright. Tilted-head subjects failed to recognize figures that were retinally "upright" (for example figure at bottom left) about as much as upright viewers did (to whom such figures were not retinally upright). Tilted-head subjects recognized environmentally upright figures (*bottom right*) as often as upright viewers did.



FIGURES WITH INTRINSIC ORIENTATION appear to have a natural vertical axis regardless of their physical orientation. A region at one end of the axis is perceived as top.



IMPRESSION OF SYMMETRY is spontaneous only when a figure is symmetrical around a vertical axis. Subjects were asked to indicate which of two figures (*middle and right*) was most like the target figure (*left*). The figure at right was selected most frequently, presumably because it is symmetrical around its vertical axis. If the page is tilted 90 degrees, the figure in the middle will now be selected as being more similar to the target figure. Now if the page is held vertically and the figures are viewed with the head tilted 90 degrees, the figure at right is likely to be seen as being the most similar. This suggests that it is not the symmetry around the egocentric vertical axis on the retina but rather the symmetry around the environmental axis of the figure that determines perceived symmetry.



AMBIGUOUS FIGURES can be perceived in different ways depending on the orientation assigned to them. Figure at left can look like the profile of a man's head with a chef's hat (*top left*) or, when rotated 90 degrees, like a dog (*bottom left*). Figure at right can look like the profile of a bearded man's head (*top right*) or like a map of the U.S. (*bottom right*). When subjects with their head tilted 90 degrees to one side viewed these ambiguous figures (*direction of subject's head is shown by arrow*), they preferentially recognized the figure that was upright in the environment instead of the figure that was upright on the retina.

and bottom are. Nevertheless, an observer will assign a top to that region of the figure which is uppermost with respect to his egocentric coordinate reference system. The vertical axis of the figure is seen as being aligned with the long axis of the observer's head and body. The upward direction corresponds to the position of his head. We have been able to demonstrate that such assignment of direction has the same effect on the recognition that other bases of assigning direction do. A figure first seen in one orientation on the circular sheet will generally not be recognized if its egocentric orientation is altered.

Now we come to an observation that seems to be at variance with much of what I have described. When a person lies on his side in bed to read, he does not hold the book upright (in the environmental sense) but tilts it. If the book is not tilted, the retinal image is disoriented and reading is quite difficult. Similarly, if a reader views printed matter or photographs of faces that are environmentally upright with his head between his legs, they will be just as difficult to recognize as they are when they are upside down and the viewer's head is upright. The upright pictures, however, are still perceived as being upright even when the viewer's head is inverted. Conversely, if the pictures are upside down in the environment and are viewed with the head inverted between the legs, there is no difficulty in recognizing them. Yet the observer perceives the pictures as being inverted. Therefore in these cases it is the orientation of the retinal image and not the environmental assignment of direction that seems to be responsible for recognition or failure of recognition.

Experiments with ambiguous figures conducted by Robert Thouless, G. Kanizsa and G. Tampieri support the notion that retinal orientation plays a role in recognition of a figure [*see illustration on page 85*]. Moreover, as George Steinfeld and I have demonstrated, the recognition of upright words and faces falls off in direct proportion to the degree of body tilt [*see illustration on opposite page*]. With such visual material recognition is an inverse function of the degree of disorientation of the retinal image. As we have seen, the relation between degree of disorientation and recognizability does not hold in cases where the assignment of direction has been altered. In such cases the greatest effect is not with a 180-degree change but with a 45- or 90-degree change.

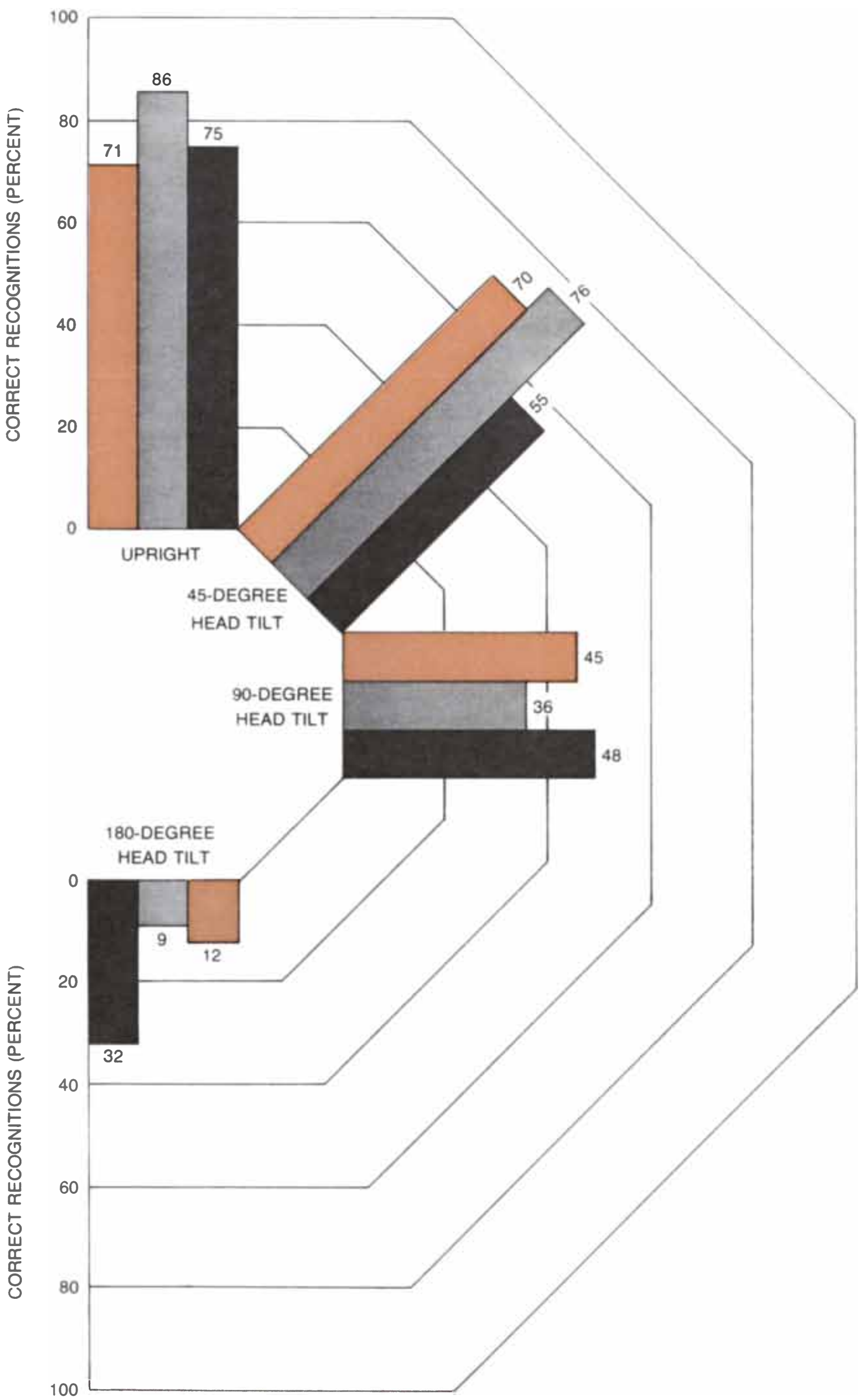
The results of all these experiments

have led me to conclude that there are two distinct factors involved in the perception of disoriented figures: an assignment-of-direction factor and a retinal factor. I believe that when we view a figure with our head tilted, we automatically compensate for the tilt in much the same way that we compensate for the size of distant objects. An object at a moderate distance from us does not appear small in spite of the fact that its retinal image is much smaller than it is when the object is close by. This effect usually is explained by saying that the information supplied by the retinal image is somehow corrected by allowing for the distance of the object from us. Similarly, when a vertical luminous line in a dark room is viewed by a tilted observer, it will still look vertical or almost vertical in spite of the fact that the retinal image in the observer's eye is tilted. Thus the tilt of the body must be taken into account by the perceptual system. The tilted retinal image is then corrected, with the result that the line is perceived as being vertical. Just as the correction for size at a distance is called size constancy, so can correction for the vertical be called orientation constancy.

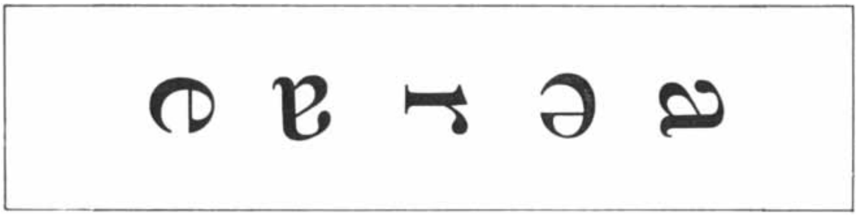
When we view an upright figure with our head tilted, before we have made any correction, we begin with the information provided by an image of the figure in a particular retinal orientation. The first thing that must happen is that the perceptual system processes the retinal image on the basis of an egocentrically assigned top, bottom and sides, perhaps because of a primitive sense of orientation derived from retinal orientation. For example, when we view an upright square with our head tilted, which yields a diamondlike retinal image, we may perceive a diamond for a fleeting moment before the correction goes into operation. Head orientation is then automatically taken into account to correct the perception. Thus the true top of the figure is seen to be one of the sides of the square rather than a corner. The figure is then "described" correctly as one whose sides are horizontal and vertical in the environment, in short as a "square." This correction is made quickly and usually without effort. In order to describe a figure the viewer probably must visualize or imagine it in terms of its true top, bottom and sides rather than in terms of its retinal top, bottom and sides.

If the figure is relatively simple, the correction is not too difficult to achieve. If we view an upright letter with our head tilted, we recognize it easily; it is of interest, however, that there is still something strange about it. I believe the dual aspect of the perception of orientation is responsible for this strangeness. There is an uncorrected perception of the letter based on its retinal-egocentric orientation and a corrected perception of it based on its environmental orientation. The first perception produces an unfamiliar shape, which accounts for the strange appearance of the letter in spite of its subsequent recognition. In our experiments many of the figures we employed were structurally speaking equivalent to letters, and in some cases we actually used letters from unfamiliar alphabets.

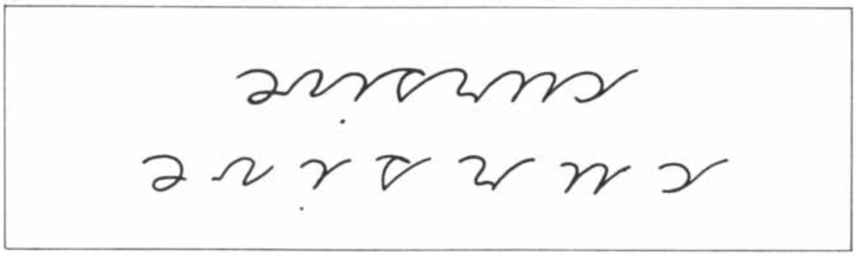
With a more complex figure, such as an inverted word or an upright word



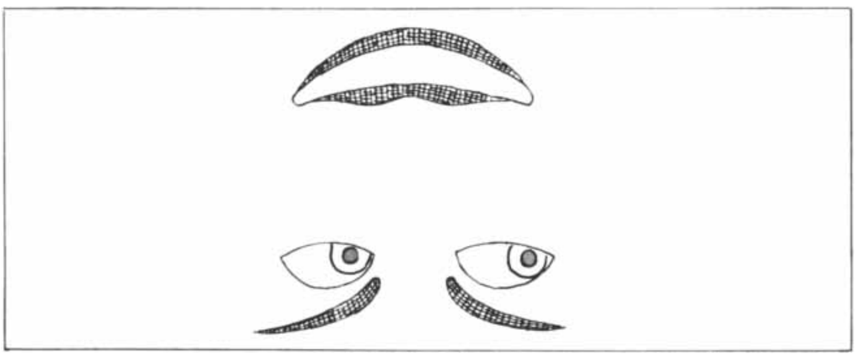
RECOGNITION OF CERTAIN KINDS OF VISUAL MATERIAL decreases almost in direct proportion to the degree of head tilt of the observer. In a series of experiments the number of correct recognitions of faces (*colored bars*), written words (*gray*) and fragmented figures (*black*) were recorded for various degrees of head tilt. Subject saw several examples of each type of test material in each of the head positions. For this visual material recognition is an inverse function of the degree of disorientation of the retinal image.



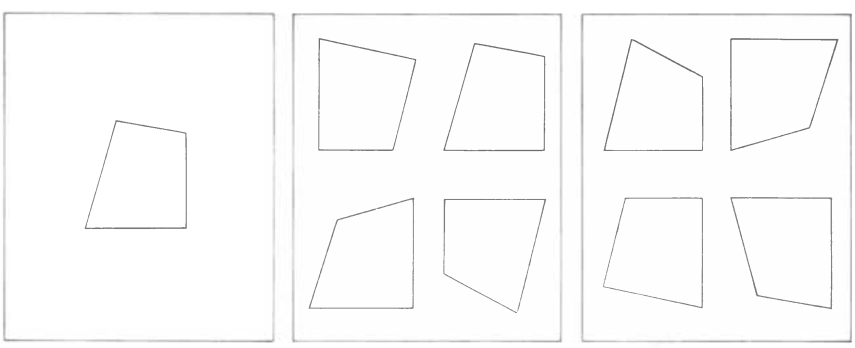
SINGLE LETTER that is tilted can be easily identified once it is realized how it is oriented. A strangeness in its appearance, however, remains because the percept arising from the uncorrected retinal image continues to exist simultaneously with the corrected percept.



INVERTED LONGHAND WRITING is difficult to decipher because many inverted units resemble written upright letters. For example, an inverted *u* will look like an *n* and an inverted *c* like an *s*. Moreover, the connection between letters leads to uncertainty about where a letter begins and ends. Several inverted units can be grouped together and misperceived as an upright letter. Separating the inverted letters makes them easier to decipher.



INVERTED FACIAL FEATURES are difficult to interpret because while attention is focused on correcting one feature other features remain uncorrected. For example, one might succeed in correcting the eyes shown here so that they are perceived as gazing downward and leftward, but at that very moment the mouth is uncorrected and expresses sorrow rather than pleasure. Conversely, one might correct the mouth and misperceive the eyes.



MULTIPLE ITEMS were found to have an adverse effect on recognition of even simple figures. Subjects sitting upright viewed the target (*left*). Then they were briefly shown test cards, some of which contained the target figure (*middle*) and some of which did not (*right*). The subjects were to indicate when they saw a figure that was identical with the target figure. Half of the test cards were viewed with the head upright and half with the head inverted. Recognition was poor when inverted subjects viewed the test cards. In other experiments with a single test figure head inversion did not significantly affect recognition.

viewed by an inverted observer, the corrective mechanism may be entirely overtaxed. Each letter of the word must be corrected separately, and the corrective mechanism apparently cannot cope simultaneously with multiple components. It is true that if an observer is given enough time, an inverted word can be deciphered, but it will never look the same as it does when it is upright. While one letter is being corrected the others continue to be perceived in their uncorrected form. There is a further difficulty: letter order is crucial for word recognition, and inverting a word reverses the normal left-to-right order.

The recognition of inverted longhand writing is even more difficult. When such writing is turned upside down, many of the inverted "units" strongly resemble normal upright longhand letters. Moreover, since the letters are connected, it is difficult to tell where one letter ends and another begins. Separating the letters of the inverted word makes recognition easier. Even so, it is all too easy to confuse a *u* and an *n*. This type of confusion is also encountered with certain printed letters, namely, *b* and *q*, *d* and *p* and *n* and *u*, although not as frequently. In other words, if a figure is recognized on the basis of its upright retinal-egocentric orientation, this may tend to stabilize the perception and block the correction process. The dominance of the retinally upright faces in the illustration on the opposite page probably is an effect of just this kind.

There may be a similar overtaxing of the corrective mechanism when we view an inverted face. It may be that the face contains a number of features each of which must be properly perceived if the whole is to be recognized [see "The Recognition of Faces," by Leon D. Harmon; SCIENTIFIC AMERICAN, November, 1973]. While attention is focused on correcting one feature, say the mouth, other features remain uncorrected and continue to be perceived on the basis of the image they form on the retina. Of course, the relation of features is also important in the recognition of a face, but here too there are a great number of such relations and the corrective mechanism may again be overtaxed.

Charles C. Bebber, Douglas Blewett and I conducted an experiment to test the hypothesis that it is the presence of multiple components that creates the difficulty of correcting figures. Subjects were briefly shown a quadrilateral figure and asked to study it. They viewed the target figure with their head upright. Then they were shown a series of test

AMBIGUOUS FACES are perceived differently when their images on the retina of the observer are inverted. If you hold the illustration upright and view it from between your legs with your head inverted, the alternative faces will be perceived even though they are upside down in terms of the environment. The same effect occurs when the illustration is inverted and viewed from an upright position. Such tests provide evidence that figures such as faces are recognized on the basis of their upright retinal orientation.

cards each of which had four quadrilateral figures. The test cards were viewed for one second, and the subjects were required to indicate if the target figure was on the card.

The subjects understood that they were to respond affirmatively only when they saw a figure that was identical with the target figure both in shape and in orientation. (Some of the test figures were similar to the target figure but were rotated by 180 degrees.) Half of the test cards were seen with the subject's head upright and half with the subject's head inverted. It was assumed that the subject would not be able to correct all four test figures in the brief time that was allowed him while he was viewing them with his head down. He had to perceive just as many units in the same brief time while he was viewing them with his head upright, but he did not have to correct any of the units. We expected that target figures would often not be recognized and that incorrect figures would be mistakenly identified as the target when the subjects viewed the test cards with their head inverted.

The results bore out our prediction. When multiple components have to be corrected, retinal disorientation has an adverse effect on recognition. The observer responded to twice as many test cards correctly when he was upright than he did when he was inverted.

As I have noted, when we look at figures that are difficult to recognize when they are retinally disoriented, the difficulty increases as the degree of disorientation increases. Why this happens may also be related to the nature of the correction process. I suggested that the observer must suppress the retinally (egocentrically) upright percept and substitute a corrected percept. To do this, however, he must visualize or imagine how the figure would look if it were rotated until it was upright with respect to himself or, what amounts to the same thing, how it would look if he rotated himself into alignment with the figure. The process of mental rotation requires visualizing the entire sequence of angular change, and therefore the greater the angular change, the greater the difficulty.

As every parent knows, children between the ages of two and five seem to be quite indifferent to how a picture is oriented. They often hold a book upside down and seem not at all disturbed by it. On the basis of such observations and the results of some early experiments, many psychologists concluded that the orientation of a figure did not enter into its recognition by young children. More recent laboratory experiments, however, do not confirm the fact that children recognize figures equally well in any orientation. They have as much difficulty as, or more difficulty than, adults in recognizing previously seen figures when the figure is shown in a new orientation. Why then do young children often spontaneously look at pictures upside down in everyday situations? Perhaps they have not yet learned to pay attention to orientation, and do not realize that their recognition would improve if they did so. When children learn to read after the age of six, they are forced to pay attention to orientation because certain letters differ only in their orientation.

In summary, the central fact we have learned about orientation is that the perceived shape of a figure is not simply a function of its internal geometry. The perceived shape is also very much a function of the up, down and side directions we assign to the figure. If there is a change in the assigned directions, the figure will take on a different perceptual shape. I have speculated that the change in perceived shape is based on a new "description" of the figure by the perceptual system. The directions assigned are based on information of various kinds about where the top, bottom and sides of a figure are and usually do not depend on the retinal orientation of the image of the figure. When the image is not retinally upright, a process of correction is necessary in order to arrive at the correct description, and this correction is difficult or impossible to achieve in the case of visual material that has multiple components.

All of this implies that form perception in general is based to a much greater extent on cognitive processes than any current theory maintains. A prevailing view among psychologists and sensory physiologists is that form perception can be reduced to the perception of contours and that contour perception in turn can be reduced to abrupt differences in light intensity that cause certain neural units in the retina and brain to fire. If this is true, then perceiving form results from the specific concatenation of perceived contours. Although the work I have described does not deny the possible importance of contour detection as a basis of form perception, it does suggest that such an explanation is far from sufficient, and that the perception of form depends on certain mental processes such as description and correction. These processes in turn are necessary to account for the further step of recognition of a figure. A physically unchanged retinal image often will not lead to recognition if there has been a shift in the assigned directions. Conversely, if there has been no shift in the assigned directions, even a very different retinal image will still allow recognition.

Waste Watcher Israel Proler—he collects steel cans by the millions

Israel Proler is one of America's new breed of waste watchers. The Chairman of Texas-based Proler International Corp., he is concerned about America's environmental shape. And he is doing something about it by reclaiming everything from junked autos to used tin cans.

Hundreds of millions of tin cans are magnetically reclaimed each year from municipal garbage and then processed in Proler plants. Along with thousands of tons of scrap from canmaking lines, the used cans are then shipped to Arizona and New Mexico. There they are used as "precipitation iron" in a chemical process that recovers copper from low-grade ore.

Israel Proler is on the prowl for used steel cans which can be economically handled by his plants. The nation's copper mines can use an estimated 10-billion annually.

Reclaimed steel cans also are recycled by the steel industry, detinning plants and ferroalloy producers.

Mr. Proler and other waste watchers are reclaiming 3 billion steel cans in 20 cities by magnetic separation—the leading resource recovery system. Steel's unique magnetic property makes it possible.

If you would like to know more about "waste watching," write to: 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

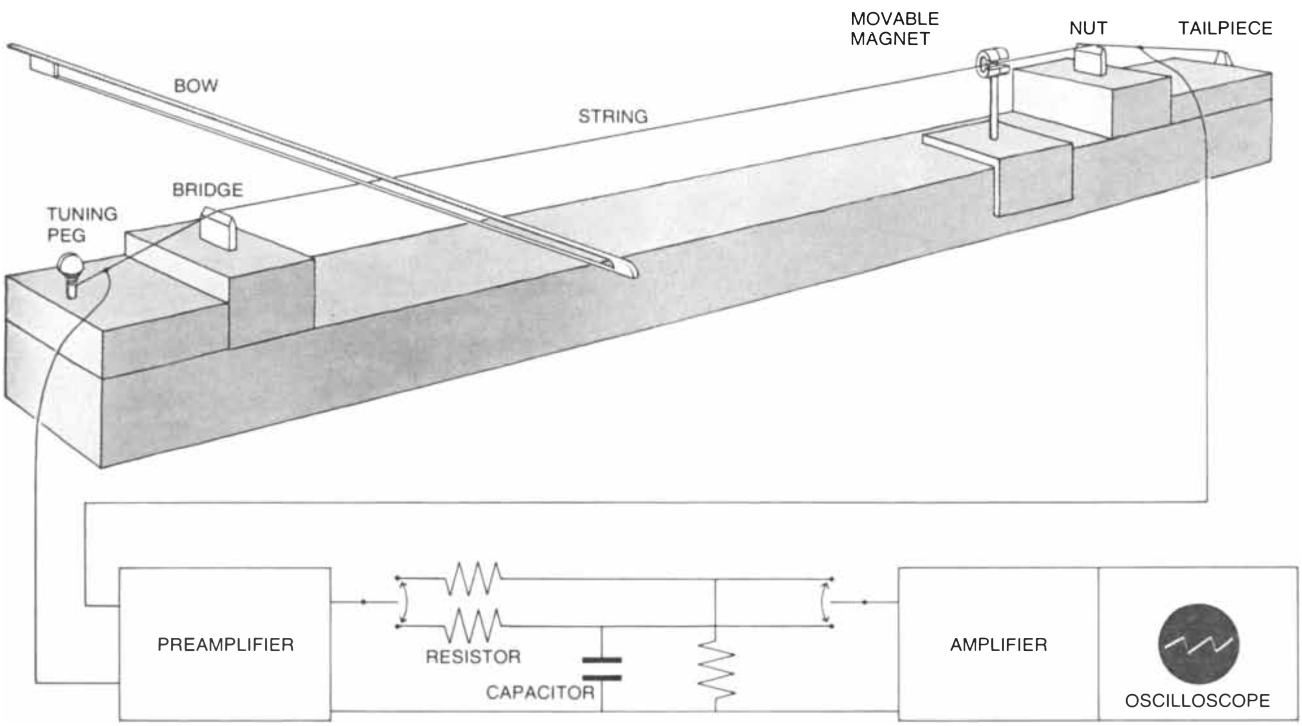
The Physics of the Bowed String

What actually happens when a violin string is bowed? Modern circuit concepts and an electromagnetic method of observing string motion have stimulated new interest in the question

by John C. Schelleng

The heart of the violin or any of its relatives, the center from which flows the acoustic pulse that is the very life of the music, is the bowed string. The string—its action under the fingers and the bow, its gratifying responsiveness and even the problems it forces the player to solve—plays a major role in establishing the musical identity of this family of instruments. Conceptually the string is the simplest of components, although its manufacture calls for meticulous care: it must be flexible, uniform and strong. In spite of this simplicity its action under the bow presents many unanswered questions. The elementary physics of its behavior can nonetheless be of considerable importance to the player.

Of the many papers published by Hermann von Helmholtz, ranging through physiology, anatomy, physics and the fine arts, there is one titled "On the Action of the Strings of a Violin" that appeared in the proceedings of the Glasgow Philosophical Society in 1860. Up to that time little was understood about what actually happens when a string is bowed. Helmholtz' procedure is a good example of how a well-conceived experiment combined with simple mathematics can illuminate a problem that could not at the time be solved with either approach alone. Today we would call his apparatus an oscilloscope; to him it was a "vibration microscope," an invention he credited to the French physicist Jules Antoine Lissajous. Through this instrument he looked at a grain of starch fastened to a black string, which he set in vibration by bowing. The objective lens of the microscope was mounted on a



MONOCHORD, a simple experimental arrangement used by the author to study the motion of a bowed string, consists of a single electrically conducting string mounted between two massive bridges on a firm base. The movement of the string through the magnetic field set up by a small movable magnet generates an output signal that can be amplified and displayed on an oscilloscope screen (*see circuit diagram at bottom*). With the two switches in the up position the system displays string velocity; with the two switches down the system displays string displacement. The string can be bowed by hand, by a pendulum-driven bow or rotary bow.

large tuning fork so as to vibrate slowly parallel to the length of the string. When both string and fork were set in motion at suitable rates, Helmholtz saw a "Lissajous figure," a form of oscillogram that displayed the position of the starch particle as it varied within the period of vibration of the fork. By similarly examining the motion at other points he experimentally acquired the basis for a mathematical description of the motion of the string as a whole.

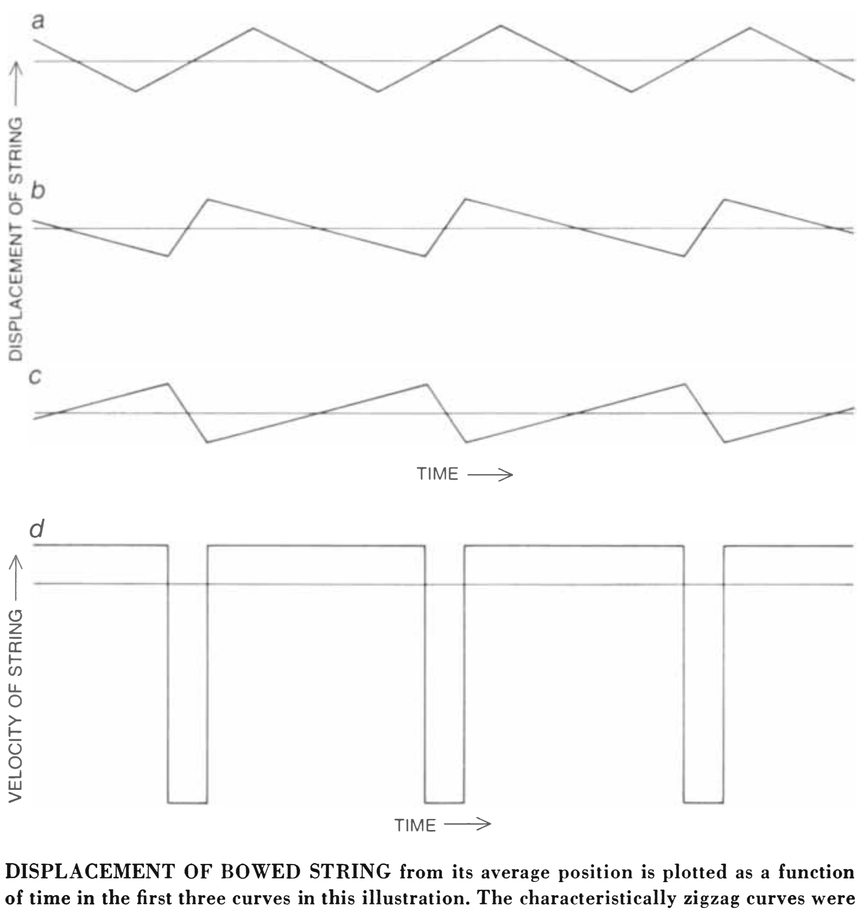
Helmholtz wrote that "during the greater part of each vibration the string is carried on by the bow. Then it suddenly detaches itself and rebounds, whereupon it is seized by other parts of the bow and again carried forward." Plotting the position of the bit of starch as a function of time, he found that every aspect of the picture as he found it except one could be represented by straight lines. During one period of vibration, almost regardless of where on the string he looked or where he bowed, the curve was a zigzag of two straight lines [*see illustration below*]. The two periods of time into which the vibration was broken were always in the same ratio as the two lengths into which the point of observation divided the string.

Something can be learned merely by looking at the lowest string of an instrument while it is being vigorously bowed [*see illustration on opposite page*]. In appearance it widens into a ribbon bounded end to end by two smooth curves. (Actually the position around which the string vibrates is moved a little to one side by the average force exerted by the bow in its direction of motion.) Helmholtz found mathematically that the boundaries are parabolas; because of their shallowness they are indistinguishable from arcs of a circle. It would be a mistake, however, to suppose that the string itself has this shape at any time. The string, Helmholtz found, has at any instant the shape it would have if it were pulled aside by the finger to some point on the arc: it is a straight line sharply bent at one point. The bend races around this edge once in every vibration; for the open *A* string of a violin, for example, it goes around 440 times in one second. If Helmholtz had been able to view the string with the aid of a stroboscopic lamp, the boundary would have disappeared and he would have seen the string as a sharply bent straight line. When the bow changes from "up bow" to "down bow," the motion of the bend around the edge changes from counterclockwise to clockwise.

The sideways velocity of the string at any point has two alternating values, unequal in magnitude and opposite in sign. As a result a typical zigzag displacement curve has a corresponding velocity curve that is rectangular in shape. The ratio of the two alternating velocities is the same as the ratio of the lengths into which the point of observation divides the string.

Two simple physical facts underlie the action of the bowed string. The first is that "sliding" friction is less than "static" friction and that change from one to the other is almost discontinuously abrupt. The second is that the flexible string in tension has a succession of natural modes of vibration whose frequencies are almost exact whole-number multiples of the lowest frequency; as a result the duration of a single vibration in the first, or lowest, mode precisely equals the duration of two vibrations in the second mode, three in the third and so on. Without outside compulsion the string is therefore by its very nature given to supporting a "periodic" wave, that is, a repetitive series of similar vibrations with a wave form dictated by the "stick-slip" process. The string allows the coexistence of a multitude of harmonics; the peculiarities in friction require it.

Helmholtz' shuttling discontinuity is the timekeeper that precisely triggers the capture and release of the string at the bow. There is a perennial explanation that views the string as a spring periodically pulled sideways to the breaking point of static friction. The spring recoils and is again captured. This view cannot explain the constancy in repetition rate over the wide range of bow forces, or "pressures," applied by the hand. The correct explanation must be given in dynamic terms; such an explanation is suggested by the constancy in time needed in the flexible string for the bend to travel twice its length. The timing is vividly illustrated by striking a long, taut clothesline near one end with a stick. A discontinuity is clearly seen hurtling to the far end, where it is reflected. On its return one feels through the stick (still resting on the line) an impulse much like the momentary frictional force on the string, which fails to hold at escape but succeeds at recapture.



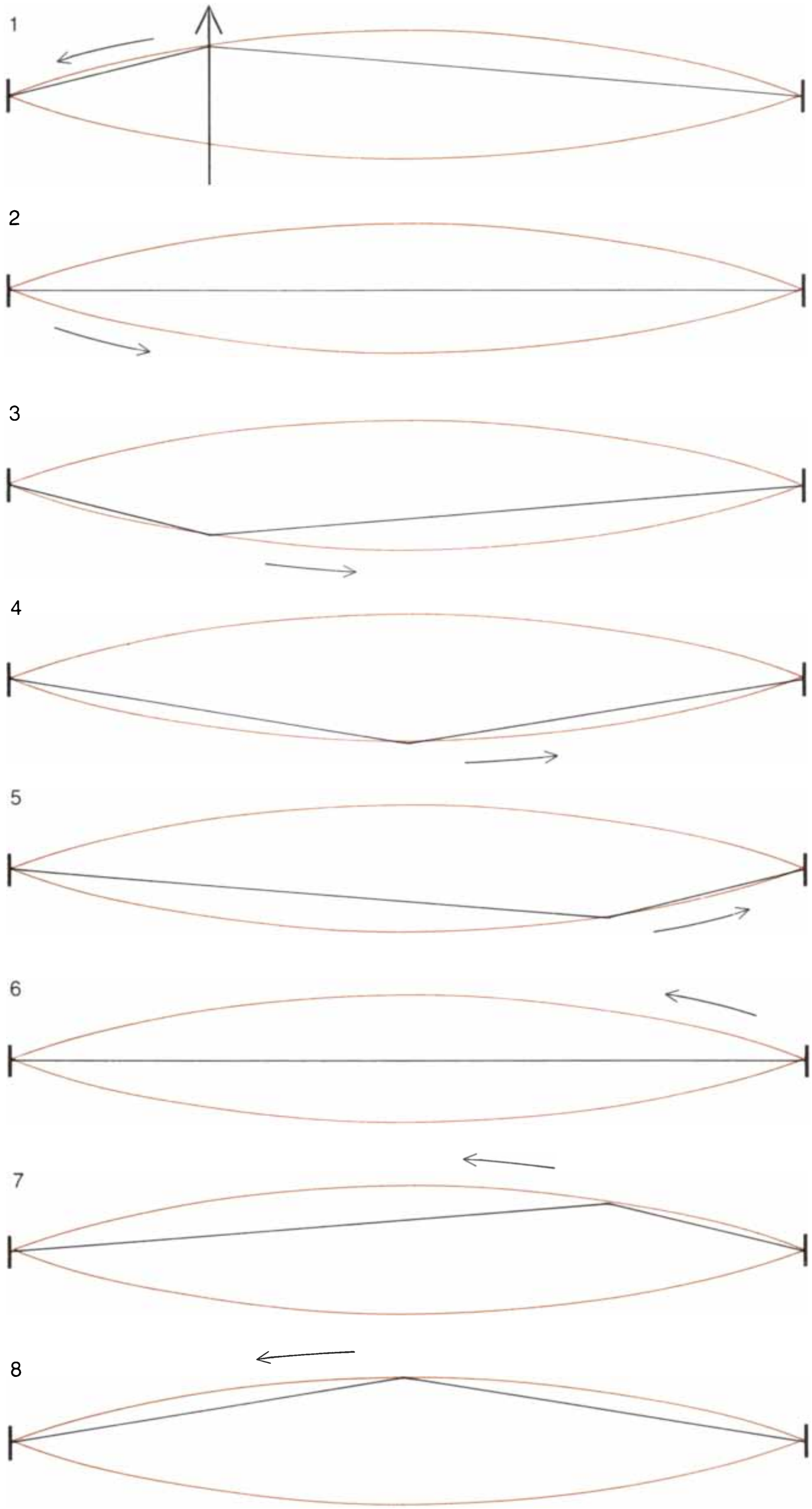
DISPLACEMENT OF BOWED STRING from its average position is plotted as a function of time in the first three curves in this illustration. The characteristically zigzag curves were obtained by bowing near one end of the string and observing at the middle (*a*), near the bridge end (*b*) and near the nut end (*c*). In each case the two periods of time into which the vibration is broken are in the same ratio as the two lengths into which the point of observation divided the string. Rectangular string-velocity curve (*d*) corresponds to curve *c*.

A simple experiment partially confirms this picture of the action of the bowed string [*see illustration on next page*]. An instrument is mounted with the strings horizontal. A light bow, suspended at its heavy end by a long thread, rests on one of the strings at a point near the bridge. A second bow sets the string in vigorous vibration. Before the hanging bow can begin to move slipping occurs at all times except at moments when the string reverses. Since friction in slipping is nearly independent of speed of slipping, the forces in the two directions of vibration are the same but the impulses imparted to the bow are proportional to the duration. The direction of the acceleration of the hanging bow will therefore indicate the direction of string motion during the longer duration.

The experiment shows, however, that the direction in which the hanging bow moves is the same as the direction in which the driving bow is moving. Therefore it is in the longer interval that the string moves with the driving bow; relative motion between driver and string is accordingly less than in the shorter interval and sticking is presumably occurring. If the hanging bow is now placed near the opposite end of the string, it will be found to move in a direction opposite to that of the driver.

Helmholtz believed that the velocity of the string while it is snapping back is constant. A half-century later C. V. Raman found that in most cases this is only approximately true. Raman's discovery came in the course of an ingenious study of the mechanical action of the violin involving both experiment and theory. With respect to the bowed string his point of departure was to describe the motion in terms of the progressive waves of transverse velocity that make up the standing waves of the Helmholtz system. The same wave can be described in terms of its lateral displacement or its lateral velocity. One advantage in emphasizing velocity is that these waves can be represented as straight lines.

The shape of the Raman wave in those cases that are of interest in music (Raman dealt with many that are not) again is a zigzag, but it differs from the displacement curves introduced above in that although the "zigs" are slow, the "zags" are instantaneous [*see top illustration on page 91*]. When such a wave is reflected from the immovable end of the string, it looks exactly as it did before except that its direction of propaga-



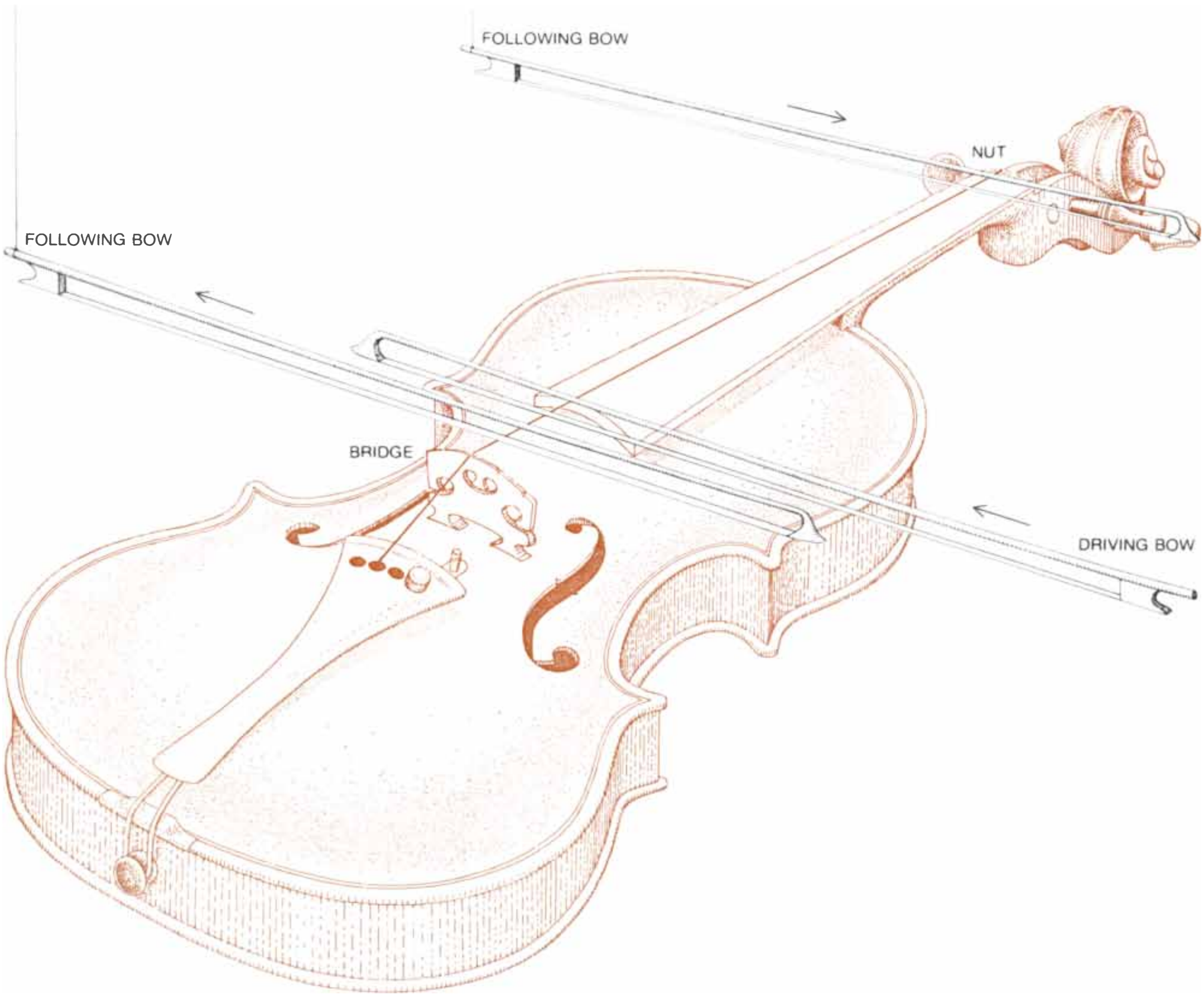
SHAPE OF BOWED STRING appears to widen into a ribbon bounded end to end by two smooth parabolic curves (*colored outlines*). As Hermann von Helmholtz found more than a century ago, however, the actual shape of the string at any instant is a straight line sharply bent at one point (*black line*). The bend races around the boundary once in every vibration. The direction of the bend's circulation in this particular series of diagrams corresponds to an upward motion of the bow; reverse the direction of the bow's motion and the bend reverses its direction of circulation. This peculiar motion is a form of standing wave.

tion is reversed. When vibration is in the fundamental mode, the length of the string is half the distance between zags.

This progressive velocity wave is of interest because, being incident on the bridge of the violin, it exerts a vibrational force whose shape is identical with its own. Insofar as the Helmholtz approximation holds, its harmonic structure therefore describes the tone quality of the string itself at the point in the instrument where the sound spectrum has not yet been influenced by resonances in or radiation from the body. The spectrum is remarkably simple: the amplitude of the nth harmonic is $1/n$ times the amplitude of its lowest, or fundamental, frequency. This simple relation is of considerable importance in investigating the sound spectrum of the violin as a whole.

Within the past few years, stimulated by circuit concepts and an electromagnetic method of observing string motion, there has been a renewal of interest in the physics of the bowed string both in this country and in Europe. More than half of the strings currently in use in stringed instruments are electrically conductive. If a small magnet is placed close to the string, the combination of a conductor moving in a magnetic field constitutes a magneto whose output can be displayed merely by inserting the string in the input circuit of a suitable amplifier connected to an oscilloscope. The electromotive force is proportional to the velocity of the string. The string can be mounted on the instrument proper or on a monochord: an experimental arrangement consisting of two massive bridges on a firm base with some means for providing tension in the string and a mount for the magnet [*see illustration on page 87*]. In my experiments two methods of bowing were used besides bowing by hand: a rotary bow developed by F. A. Saunders in his researches on violins and string action, and an ordinary bow driven by a 50-pound pendulum.

An electrical circuit connected to the monochord (or to the instrument) makes it possible to display the velocity or displacement of the string in the form of oscillograms. The first oscillogram in the bottom illustration on the opposite page, for example, displays velocity at the bow in a very flexible string. In the long period velocity lies above the zero line; that is also the velocity of the bow (if one ignores the slight ripples). In the short period of slipping there is a high negative velocity as the string whips



"FOLLOWING BOW" EXPERIMENT, performed in the course of the author's investigation, partially confirms Helmholtz' dynamic picture of the action of a bowed string. With the instrument mounted horizontally, a light bow, suspended at its heavy end by a long thread, rests on one of the strings at a point near the bridge. A second bow sets the string in vigorous vibration. In this situation the hanging bow is found (after a short period of slipping) to move in the same direction as the driving bow. The direction in which the hanging bow moves indicates the direction of string motion during the longer interval of each vibration. When the hanging bow is placed near the opposite end of the string, one finds that the "follower" moves in a direction opposite to that of the "driver."

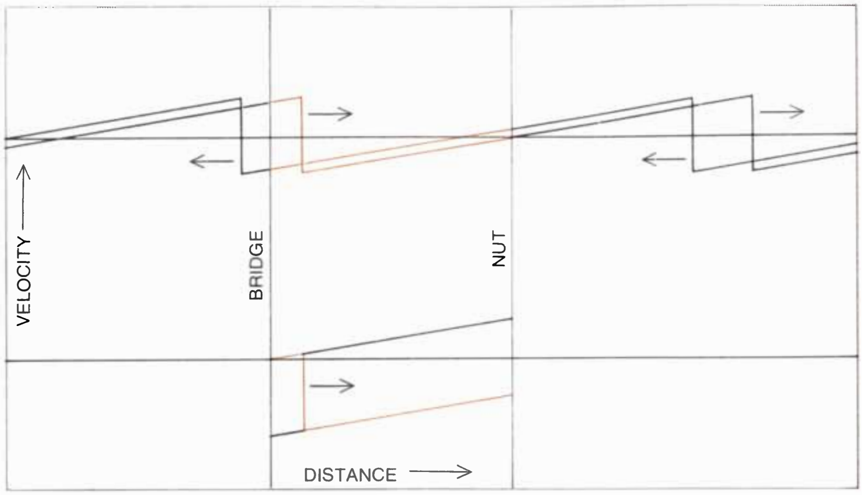
backward to take a new hold on the bow. The bow was near the bridge in this case and the shape of the curve is close to what the Helmholtz construction predicts. The zigzag in the second oscillogram shows the same vibration in terms of displacement instead of velocity.

Simplicity in instrumentation is not the only advantage in displaying velocity rather than displacement. In this way high-frequency detail that would not be suspected is brought out clearly.

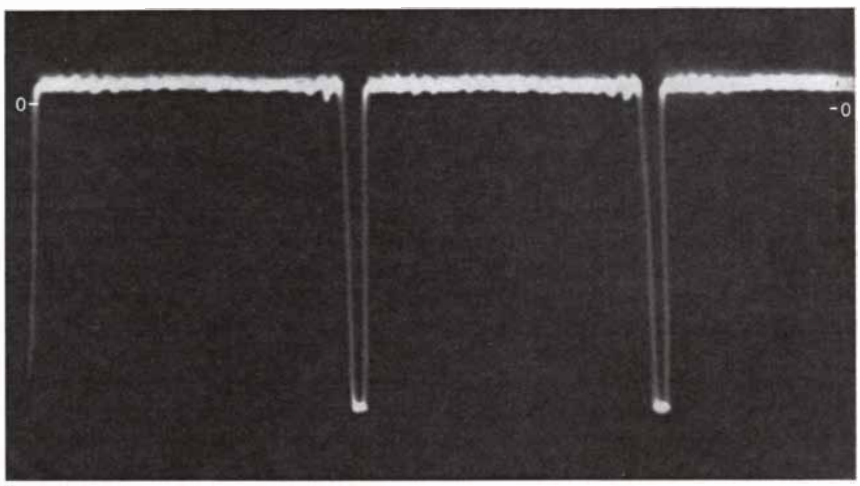
When the circuit is arranged to indicate displacement close to the bridge, it also indicates the transverse vibrational force applied by the string. The "sound of strings" alone, divorced from the modifying effect of the body of the instrument, can be produced by placing a magnet for each of the strings close to the bridge. The output from the four magnets in series is then passed through the integrator and is amplified and recorded. This arrangement sums up the forces applied by all strings. Playback therefore gives the effect of strings alone. (Recording is unnecessary if a dummy fiddle radiating no sound is used.)

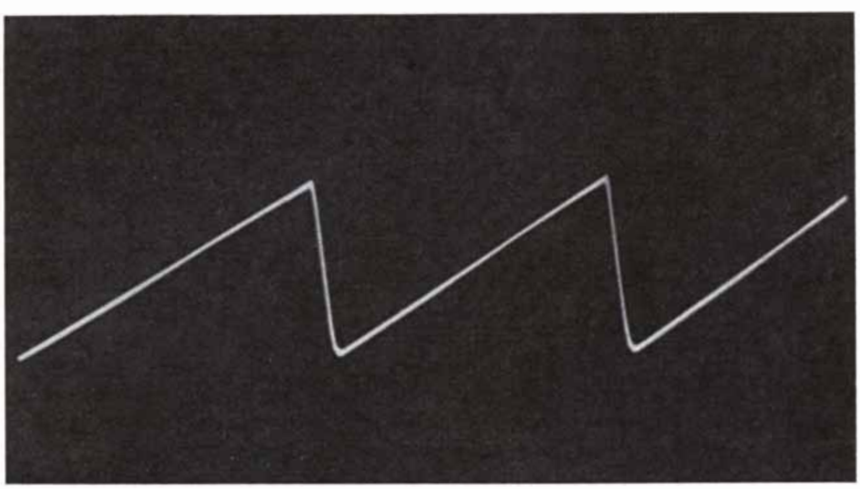
The result definitely resembles a bowed-string instrument, but an inferior one. If the system faithfully translates the force on the bridge into radiated sound pressure, the resulting sound spectrum with a given speed of bow will vary with frequency in an essentially inverse manner, the strongest effect being in the lowest tone. This "sound of strings" is not unlike the sound of the violin family in its upper octaves, but it differs drastically in the lower ones, where instead of the fiddle's achieving its strongest fundamental on the lowest tone the effect falls practically to zero because of the small size of the instrument in relation to the wavelength of such a tone in air. The harmonics then are to be credited with the fundamental tone heard subjectively. It is hardly necessary to emphasize the role that this characteristic has played in the evolution of the instrument.

In playing a bowed-string instrument there are certain limits to the speed of the bow, its distance from the bridge and the normal force applied; these limits must not be overstepped in a specific musical situation. For the experienced player this is usually a matter of choosing almost subconsciously from familiar patterns of action, but in extreme cases he is probably always aware of the limitations imposed on him. The ranges of these mechanical parameters are fortunately wide: the bow-to-bridge distance, for example, may vary from a

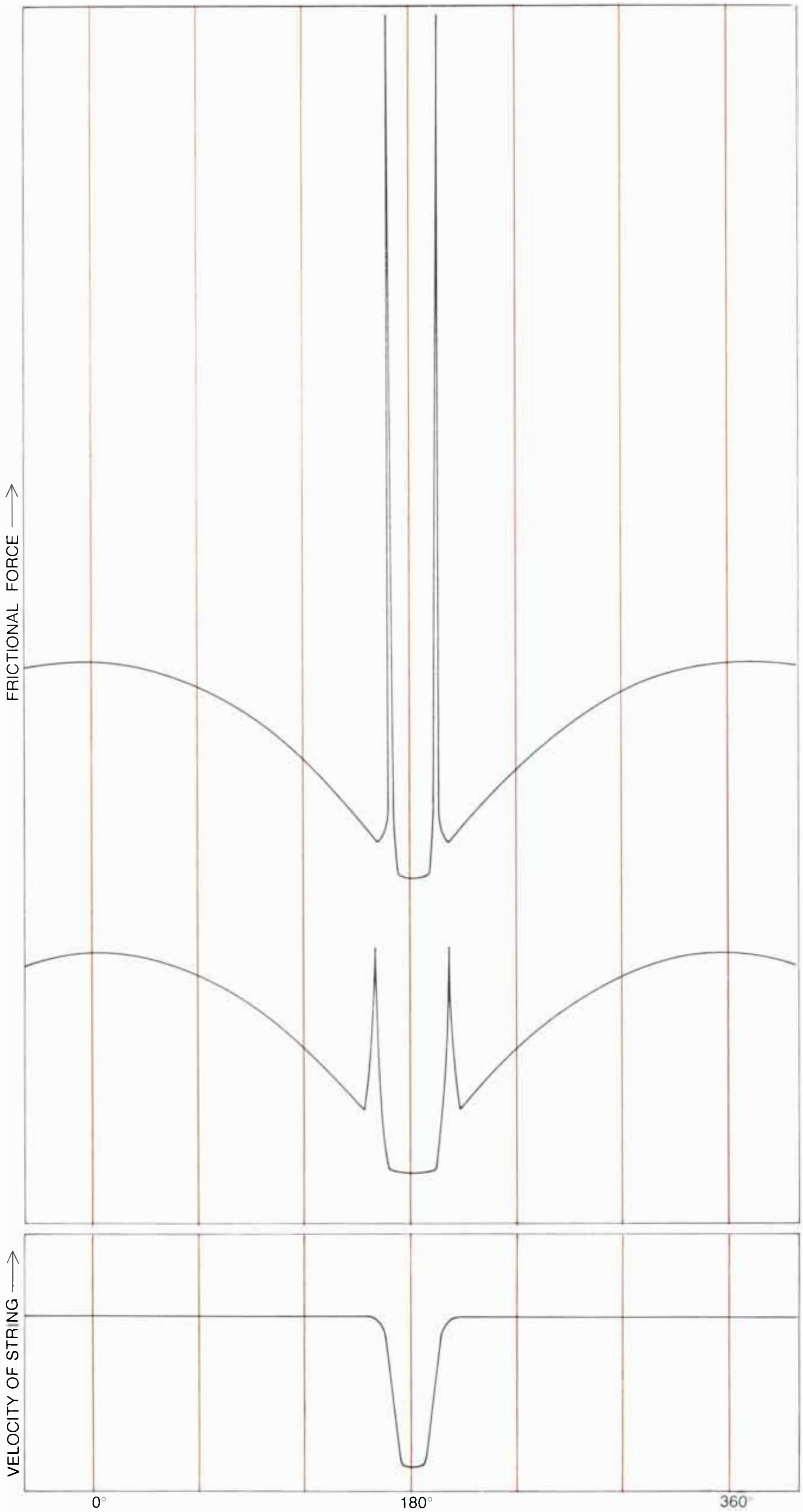


RAMAN WAVES were introduced by C. V. Raman to describe the motion of the bowed string. The shape of such a progressive transverse velocity wave differs from the corresponding Helmholtz standing wave of string displacement in that whereas the "zigs" are slow, the "zags" are instantaneous. When the oppositely moving Raman waves (*top*) are summed, the resulting wave (*bottom*) shows how the two velocities that exist alternately at any point on the string depend on the position of the discontinuity between "slipping" and "sticking."





MOTION OF VERY FLEXIBLE STRING at the bow is represented by these two oscillograms, which show string velocity (*top*) and string displacement (*bottom*) for the same vibration. The bow in this case was located about a twentieth of the length of the string away from the bridge. The shape of the curves is close to what Helmholtz picture predicts.



"RABBIT EARS" characterize the curves that represent the frictional force that is exerted by a bow in order to vibrate a string. One ear forms on the curve for frictional force when the discontinuity arrives from the nut, overcomes static friction and initiates slipping; the other ear forms when the discontinuity returns from the bridge and initiates sticking. The top curve shows the frictional force when the bow force has a typical value midway between the upper and lower allowable limits; the middle curve applies to the lower limit. For clarity ripples have been omitted from the wings of both of these curves (*left and right*). The curve at bottom is the string-velocity curve corresponding to the frictional-force curve at top.

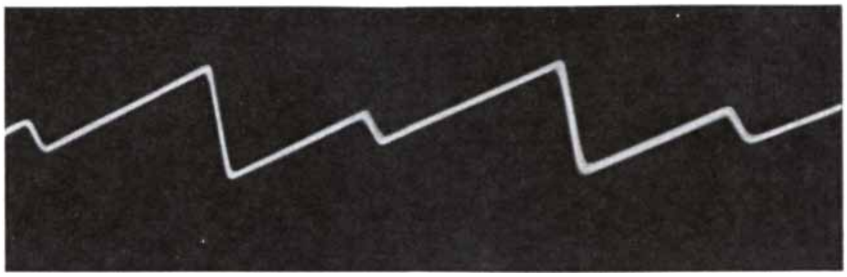
minimum value to five times that minimum; speed and force may range up to 100 times their minimum value. Given any two of these parameters, in order to assume an acceptable tone the third must fall within a range that depends on the physical constants of the string and the body of the instrument. For sustained tones these ranges are generous, although clearly all portions of a given range are not equally desirable. For example, position and speed being given, the highest permissible force may typically be 10 times the minimum. The first question is: What are the processes that determine the existence of these limits?

In order to explain the limitations on bow force it is helpful to consider how the frictional force at the point of contact of the bow and the string varies in time. Although one cannot at present display this force on an oscilloscope, it is possible to form a simplified qualitative picture on the basis of the physics involved. To do this one assumes (1) that the elementary laws of static and dynamic friction can be applied, (2) that the bridge acts as a high "mechanical" resistance (analogous to resistance in an electrical circuit) and (3) that the mass and tension of the string and its motion at the bow are known.

Assuming that the force on the string is always in the direction of the bow's motion, the points of maximum bow force occur at intervals of zero degrees, 360 degrees, 720 degrees and so on; the minimum points will then fall at 180 degrees, 540 degrees, 900 degrees and so on [*see illustration at left*]. The cyclic swinging of force between these two levels is what is needed in order to vibrate the instrument. In the small interval around 180 degrees slipping occurs. In transition from sticking to slipping there is a brief moment when the maximum of static frictional force required by bow force is exerted. Research has shown that "static friction" is really not quite static: velocity, although minuscule, is finite, and friction in reality changes continuously with velocity near zero velocity with a narrow maximum corresponding to static friction. The same curve is presumably followed in reverse in going from slipping to sticking. The "rabbit ears" evident in such frictional-force curves are the result. Such curves differ in detail in going from note to note because of complexity in the action of the body.

Consider the situation where the bow force has a typically intermediate value. From the zero-degree mark onward, with the string clinging to the bow, the force falls toward the minimum, which is dictated by dynamic friction. Then, at the moment when sticking seems most secure, the discontinuity arrives and overcomes static friction. The discontinuity needs to provide only the amount by which static friction exceeds dynamic friction. The discontinuity is capable of contributing more than is required here, perhaps much more. As the bow force is increased, however, the time will come when the discontinuity will lose out in this test of strength and the vibration will become an erratic squawk. Maximum bow force will have been exceeded.

A different kind of failure occurs when the bow force is decreased to a minimum. Here the "ears" of the frictional-force curve fall to the same level as the maximum force at the zero-degree mark, and with the lightest additional decrease static friction (as indicated by the "ears") becomes insufficient to hold the string near the zero-degree level. The result is an unstable string-displacement curve in which a new zigzag begins to form [*see top illustration on this page*]. If this new zigzag is allowed to develop, the fundamental tone will be replaced by the octave tone; in short, one will have failed to provide minimum bow force.

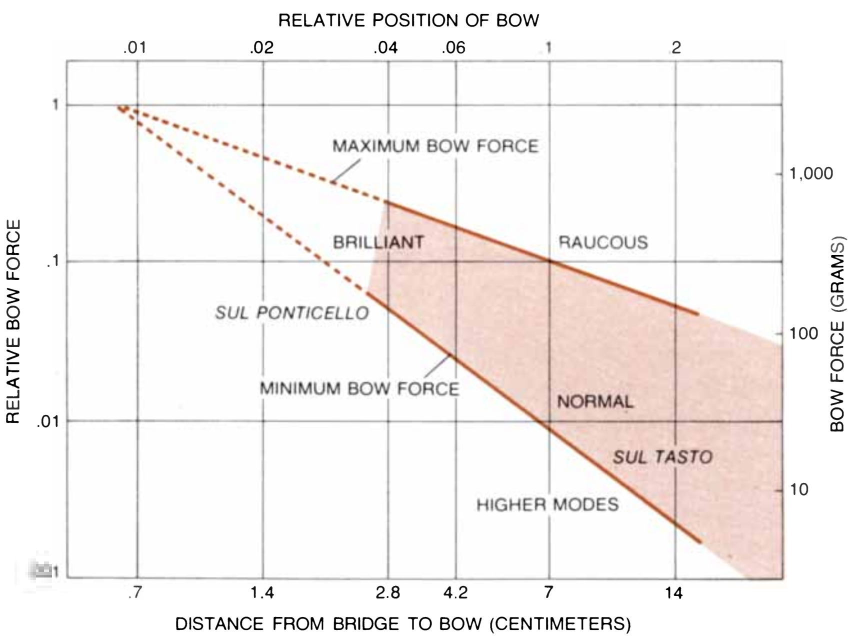


NEW ZIGZAG begins to form in the oscillographic displacement curve of a bowed string when the bow force is allowed to fall below the minimum bow force. If this unstable condition is allowed to develop, the fundamental tone will soon give way to the octave tone.

It is an important fact in the mechanics of playing that maximum bow force, which depends primarily on the string and on frictional coefficients, is inversely proportional to the first power of the distance of the bow from the bridge, whereas minimum bow force, which in addition depends on the body of the instrument, is (at least approximately) inversely proportional to the second power of the same distance. The quantities necessary for calculating these limits are known well enough to explain how the string reacts to bow forces. For sustained tones with a given bow velocity one can display the logarithmically linear trends of maximum and minimum bow force in terms of the relative distance of the bow from the bridge expressed as a fraction of the total length of string [*see bottom illustration on this page*]. The most important result is that the maximum bow force and the minimum bow force are equal when the bow is placed at a certain point very close to the bridge and that they diverge as the bow moves away from it. (Actually the curves near the intersection are to be regarded as extrapolations from the right, where normal conditions obtain.) It is this open space between the limits that gives to bow force the wide tolerance that makes fiddle-playing possible.

The forces toward the left between these lines are impractically high; normal playing is confined to the area toward the right. Farthest from the bridge the volume of the sound is least, the content of upper harmonics is minimal and the timbre has the gentle character that composers seek by designating *sul tasto:* "bow over the fingerboard." Exceed maximum bow force and the result is unmusical; fall short of the minimum and the solid fundamental tone is lost, leaving what is sometimes called a surface tone. The closer the bow is to the bridge, the less generous is the ratio between maximum and minimum bow force and the steadier is the hand that is needed. The experienced player prizes this domain for its nobility of tone; the beginner finds it prudent to play closer to the fingerboard. Closer still to the bridge bow force mounts prohibitively and the solidity of the fundamental tone disappears until little more than a swarm of high harmonics remains to suggest the fundamental tone; this is the eerie *sul ponticello* ("bow over the little bridge") of the composer. Within the normal-playing area the relative harmonic content increases—the tone becomes more brilliant—either as the bow moves toward the bridge or as the bow force increases toward its maximum.

Such a diagram is to be taken as quali-



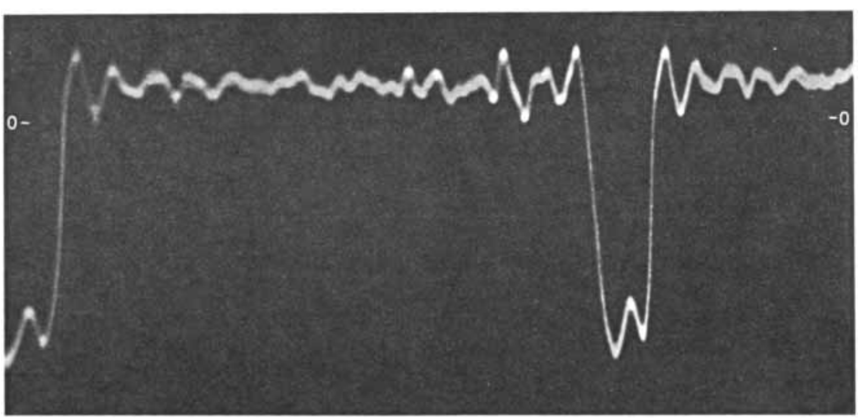
NORMAL-PLAYING RANGE for a bowed-string musical instrument is depicted for sustained tones at a constant bow velocity in this graph, which indicates the logarithmically linear trends of maximum and minimum bow force in terms of the relative distance of the bow from the bridge expressed as a fraction of the total length of the string. As the graph indicates, the maximum bow force and the minimum bow force tend toward equality (*upper left*) when the bow is placed at a certain point very close to the bridge and they diverge as the bow moves away from the bridge (*lower right*). The open space between these two limits, shown in color, accounts for the wide tolerance in permissible bow force. *Sul tasto* means "bow over the fingerboard"; *sul ponticello* means "bow over the little bridge." The second set of coordinates (*bottom scale; right scale*) suggests the normal-playing conditions for a typical *A* string of a cello bowed at a sustained velocity of 20 centimeters per second.

tative, in particular the curve for minimum force, which varies greatly from note to note because of the complexity in the response of the body of the instrument. Although the Helmholtz idealization is close enough to fact to provide a useful basis for many first-order calculations such as those described above, it is not completely trustworthy in other respects. In contradiction to what it implies, harmonic content increases with bow force, changing timbre and loudness. If loudness depended only on the "root mean square" vibrational force on the bridge, the effect would not be of much consequence, but when harmonics are radiated more efficiently than the fundamental tone or perceived more sensitively by the ear, the effect is of some importance. The fact remains that the player's major resource in controlling volume lies in the speed and placement of the bow. The implication that sound pressure is directly proportional to bow speed and inversely proportional to bow-to-bridge distance is not far wrong.

Bowed tones start in different ways, but perhaps most frequently the bow pulls the string sideways until its displacement can no longer be supported by static friction. Failure in friction, like release in plucking, sets up two oppositely traveling Helmholtz discontinuities, only one of which, the one toward the bridge, can become sustained. Until bow speed matches bow force, however, the condition is described as "raucous," and there may be many false starts before balance is realized. The art in such beginnings is to achieve the match in such a short time or at such a low sound level that an unpleasant effect is avoided. A noiseless beginning can be made by allowing the bow already in motion to make a "soft landing" on the string, thus entering the normal-playing zone by way of the zone labeled "Higher modes." Theoretically at least, bow forces and velocity can be balanced from the beginning.

In the foregoing discussion of the frictional force between the bow and the string a phenomenon of some interest was left unmentioned in the interest of simplicity, namely the role of the "ears" of the frictional-force curve in setting up reverberations between the bow and the ends of the string, some of which may persist for many periods of vibration. These effects are ignored in the classical discussion of the action of the bow but are prominent in oscillograms of string velocity. Consider a curve showing the motion of the string under the bow, where during the long interval of sticking the string might be expected to follow the unchanging speed of the bow [*see illustration below*]. It is in fact true that there is no slipping, but the string can nevertheless move by rolling on the bow except as prevented by the string's resistance to twist. The ripple in rolling implies a corresponding ripple in force exerted on the string.

A more comprehensive frictional-force curve would therefore show sharp fluctuations in force superposed on the smooth sections. One effect is to raise the minimum bow force somewhat. Away from the bow the ripples in velocity are much more pronounced. The motion would be completely suppressed at the bow if the string allowed no twisting, but it would still exist away from the bow.



ROLLING OF STRING on the bow during a long interval of sticking produces characteristic "ripples" in the string's velocity curve. To obtain this oscillogram an *A* string of a cello was bowed with 4.5 times the minimum bow force. The period immediately following the capture of the string by the bow (that is, the section of curve just to the right of each main pulse) shows mainly the decay of the pulse formed at that capture as it reverberates in the short section of string. The period before release shows mainly long-delayed reverberations in the long section, not only from the most recent release but also from earlier ones.

The term wolf note is commonly used to describe an unpleasant sound that appears consistently at an isolated frequency in a bowed-string instrument. Often its origin is obscure. There are many varieties of wolves but the most vicious of the species has its habitat in the cello (and sometimes in the violin or the viola) one octave and a few semitones above the lowest note. There is no mystery about its cause. The body of each instrument has a multitude of resonances, and the wolf tone (if one exists) arises at the most prominent of them. For the bowed string to behave properly its ends must be given a support whose rigidity is in keeping with the mass of the string. Fingering one of the heavier strings to the same frequency as the resonant frequency of the body therefore invites trouble. This nuisance manifests itself in different ways but the most characteristic way is the generation of two tones, both forced vibrations, close enough together to produce a harsh beating. Since the two tones straddle the resonance peak, they require less bow force than one tone would alone at the resonant frequency [*see top illustration on opposite page*].

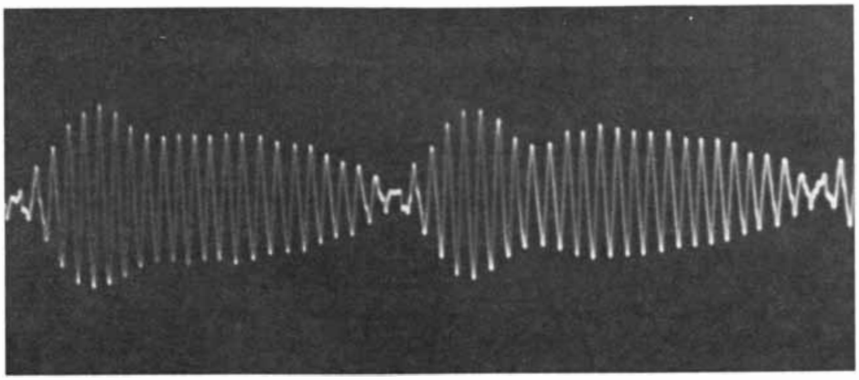
In the stiff strings of the piano, consisting as they do of thick steel wire, the frequencies of higher modes of vibration are not whole-number multiples of the frequency of the lowest mode but are somewhat sharp. That is not a defect: the "tang" it gives to the sound of the struck string is highly valued. What effect does stiffness produce in a bowed string? Clearly it is different from that in the piano. The mechanism of bowing produces a succession of almost identical vibrations. From a mathematical point of view this is another way of saying that the vibration is made up of harmonic components whose frequencies are exact whole-number multiples of the lowest frequency.

There can indeed be an effect in the bowed string. Although inharmonicity is perforce held at bay, freedom in vibration is restricted. One expects deterioration in tone quality through reduction in higher harmonics, difficulties in intonation and the need for abnormal bow forces. Before 1700, when wound strings became available, all violin strings consisted of gut, but the gut *G* string (the lowest) was unsatisfactory. The reason is not hard to find. With bowing the fundamental frequency is close to that of the lowest natural mode of vibration; in the gut *G*, however, seven times that frequency falls midway between the sixth and seventh vibrational modes and so is completely without the support of resonance. Regardless of the bow force, the seventh harmonic must be negligible. This difficulty in producing harmonics can be illustrated by a curve that shows velocity at the bow in a stiff string [*see bottom illustration at right*]. When such a string is bowed with minimum bow force, its behavior bears almost no resemblance to that of a flexible string.

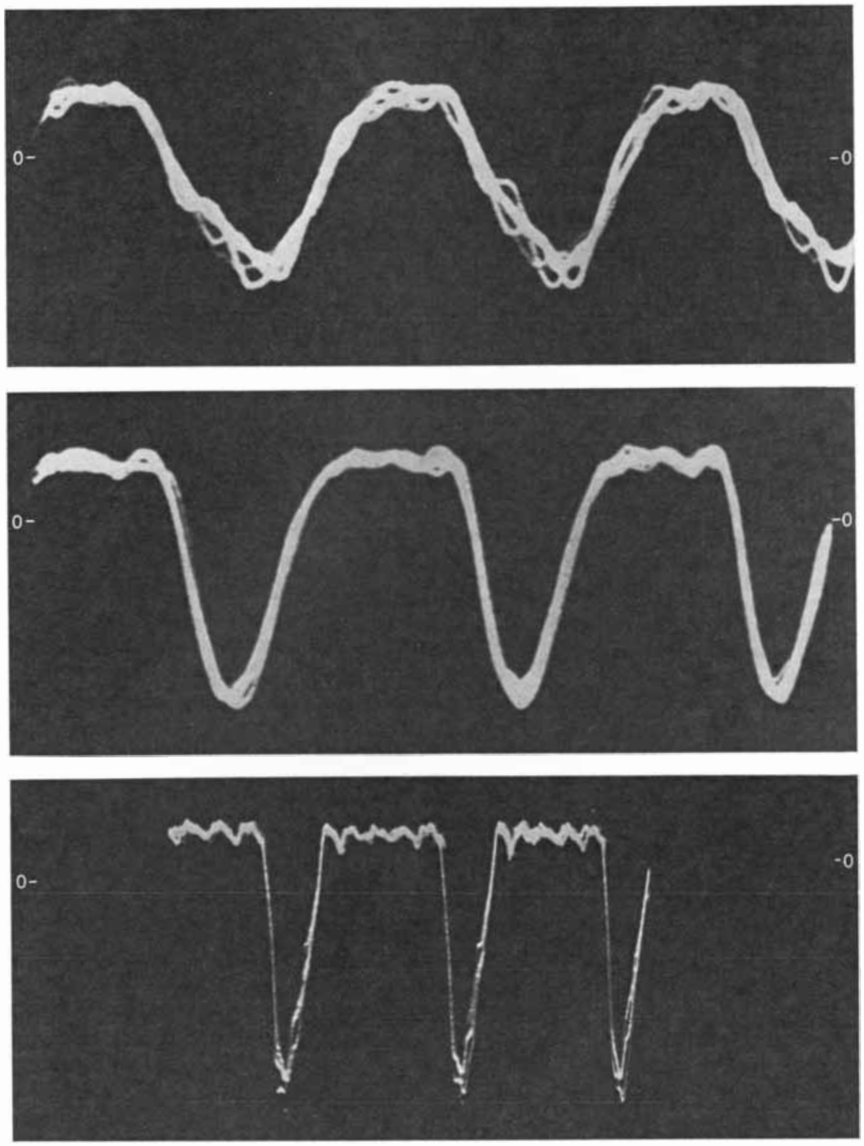
The sharpening of the *n*th mode caused by stiffness is directly proportional to the square of *n* and to a quantity called the coefficient of inharmonicity. If one changes a string mounted at a given position on a given instrument, keeping length and frequency unchanged, interchangeability requires that the same tension be maintained as well. Considering now a series of homogeneous strings of diverse materials, the coefficient of inharmonicity turns out to be proportional to the modulus of elasticity divided by the square of the density. For steel this ratio is about 50 percent greater than it is for gut; for aluminum the ratio is nearly five times greater than it is for gut. For silver, on the other hand, the ratio is about a third of the value for gut. A steel piano string that has the same pitch as the steel *E* string of the violin has an inharmonicity coefficient 20 times greater. If one ignores differences in tension in the four strings of a violin (actually it is considerably more in the highest string than in the others), inharmonicity for homogeneous strings of the same material, being inversely proportional to the fourth power of frequency, increases by a factor greater than 100 in going from the highest string to the lowest.

From calculations and measurements for several strings on the market, including some that are wound, it appears that when the coefficient of inharmonicity is equal to or less than .1, stiffness offers no disadvantage in bowing. In the steel *E* string of the violin the value is about .04. For one very good cello *C* string with metal winding on steel cable it is about the same.

As with many seemingly simple things, there is much about the bowed string that remains open to speculation. Does twisting lead to any important acoustic effect? How important is "negative resistance" throughout slipping and exactly how does rosin behave? How much do successive periods of vibration differ and is this slight wandering of any musical significance? The answers to such questions may be of little interest to the player, but the student of the bowed string would still like to know.



"WOLF NOTE," an unpleasant sound that may appear consistently at an isolated frequency in a bowed-string instrument (particularly in the cello), is produced by the "beating" of two or more tones generated by "forced" vibrations of a string clustering around the natural resonant frequencies of the body of the instrument. This oscillogram, which shows the string motion for a complicated wolf tone obtained from the *C* string of a cello, was supplied by Ian M. Firth and J. Michael Buchanan of the University of St. Andrews in Scotland.



EFFECT OF STIFFNESS in a bowed string is a deterioration in tone quality attributable in part to an increased difficulty in producing higher harmonics. This problem is demonstrated by these three oscillograms, which represent the velocity of a stiff string at the bow for three levels of bow force: minimum (*top*), intermediate (*middle*) and high (*bottom*).

THE ASTROLABE

This scientific instrument of the Middle Ages was used for both astronomical and terrestrial observations. It also served as an analogue computer, particularly for determining the local time

by J. D. North

The astrolabe was the most widely used astronomical instrument of the Middle Ages. It originated in antiquity and was still not uncommon in the 17th century. One purpose of the instrument was observational: it was employed for finding the angle of the sun, the moon, the planets or the stars above the horizon or from the zenith. It could also be used for determining the height of mountains and towers or the depths of wells and for surveying in general. Far more important, however, was the astrolabe's value as an auxiliary computing device. It enabled the astronomer to work out the position of the sun and principal stars with respect to the meridian as well as the horizon, to find his geographical latitude and the direction of true north (even by day, when the stars were not visible), and it allowed him to indulge in such prestigious and lucrative duties as the casting of horoscopes. Above all, in the days before re-

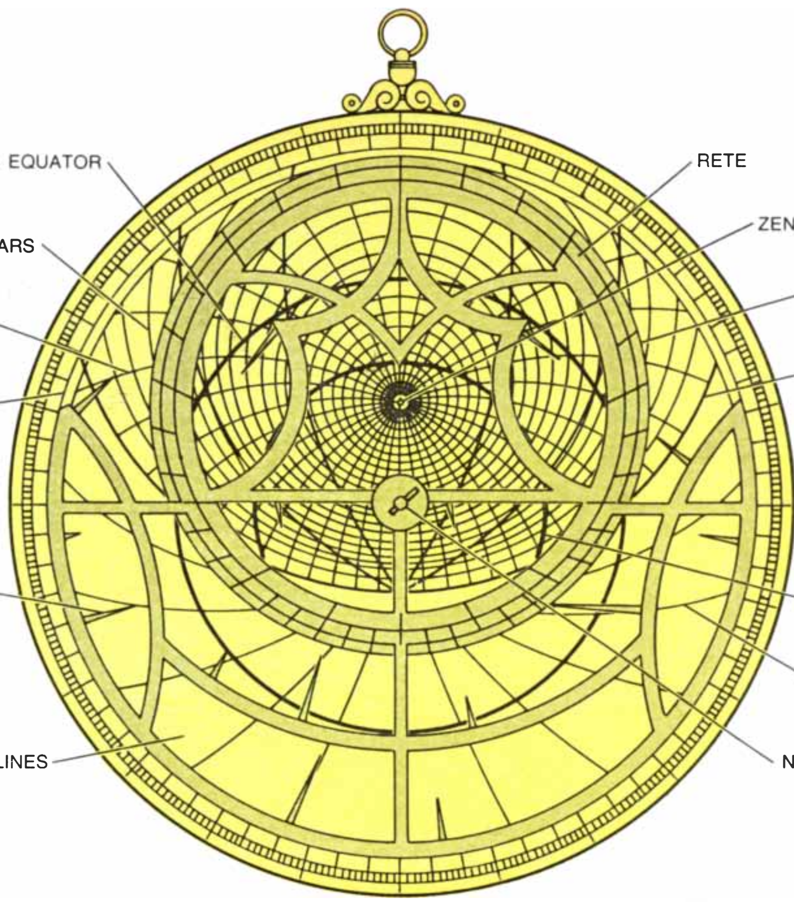


DIAGRAM OF THE FRONT OF AN ASTROLABE shows those parts that were central to its function as an instrument for calculation. The fretted network, known as the rete, is a reproduction of the heavens. The tiny pointers indicate the positions of the stars. The eccentric circle at the top is the ecliptic: the yearly path of the sun through the sky. The rete pivots around a pin that holds it to the plate behind it. The pin's position corresponded to the north celestial pole. The lines on the plate represent coordinate lines that are fixed with respect to an observer on the earth. The turning of the rete showed the daily motions of stars in relation to observer.

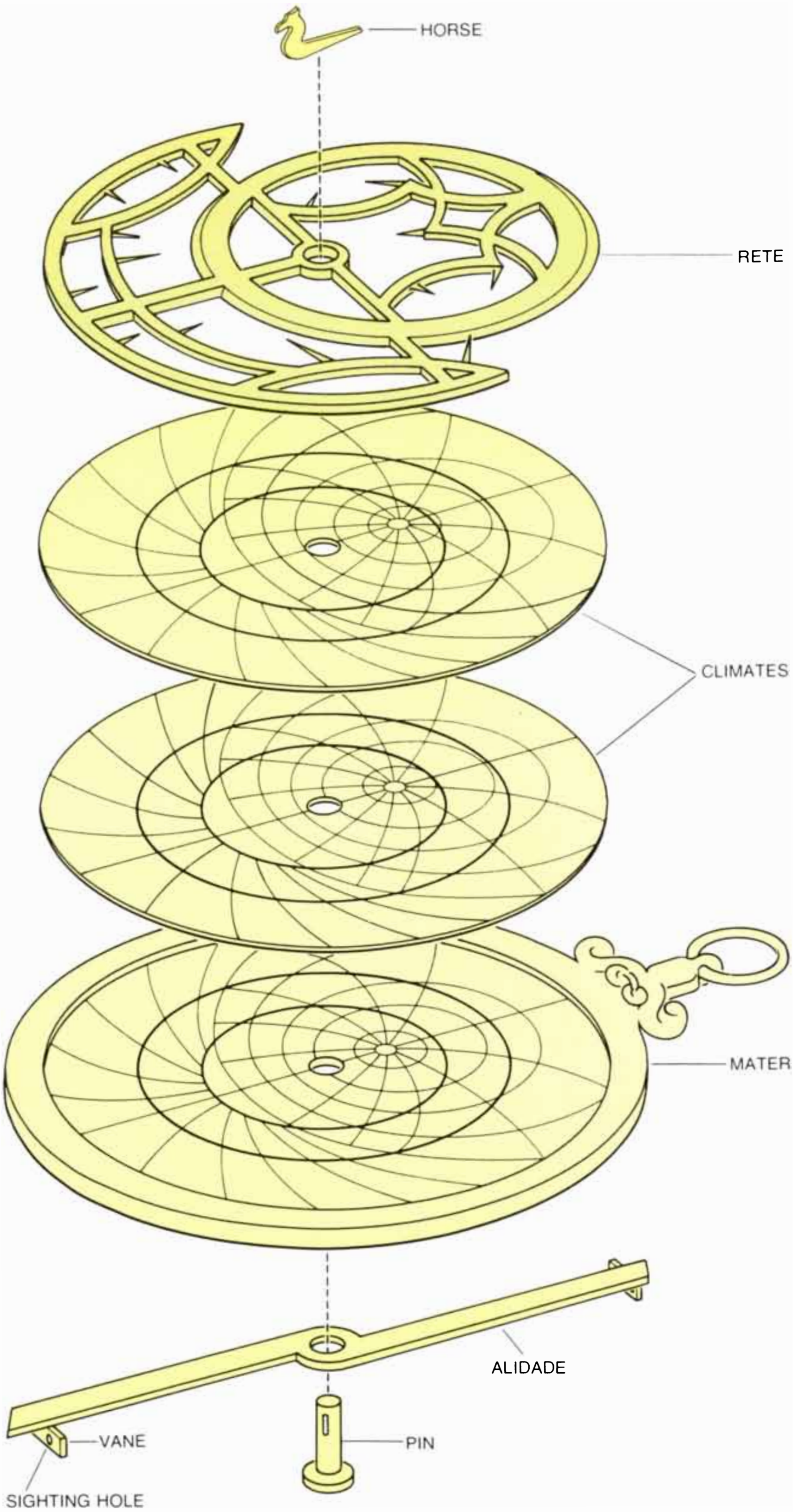
liable clocks were commonly available, the astrolabe provided its owner with a means of telling time by day or by night, as long as the sun or some recognizable star marked on the instrument was visible.

A more precise name for the instrument I am describing is the planispheric astrolabe. There are three other types of astrolabe: the linear astrolabe, the spherical astrolabe and the mariner's astrolabe. The linear astrolabe was an instrument that was difficult both to use and to understand, and it was rarely made. The spherical astrolabe was also rare; it was in the form of a globe, although it had much in common with the flat planispheric astrolabe. The mariner's astrolabe was a relatively late instrument; as far as is known it was first used only a little before the time of Columbus. It was a crude device, serving chiefly to find the altitude of the sun, moon or stars above the horizon, and it was used for much the same purpose as the sextant of later centuries. Basically it consisted of an alidade, or straight rule, pivoted centrally on a single pin on a circular plate. On each end of the alidade was a vane pierced with a hole. The mariner hung the instrument from his thumb and adjusted the alidade so that he could sight the celestial object through the holes in the vanes. He then read the altitude of the object on the scale of degrees around the edge. (In working with the sun he would have allowed one vane to cast a shadow on the other in order not to injure his eyes by direct observation.) The mariner's astrolabe was made of heavy brass so that it would hang steadily from its ring and shackle, and it was also pierced so that the wind would affect it as little as possible.

The planispheric astrolabe I shall henceforth call simply the astrolabe, since it was by far the commonest type. In order to fully understand even its simple uses, it is necessary to examine its outward form and trace how it acquired that form.

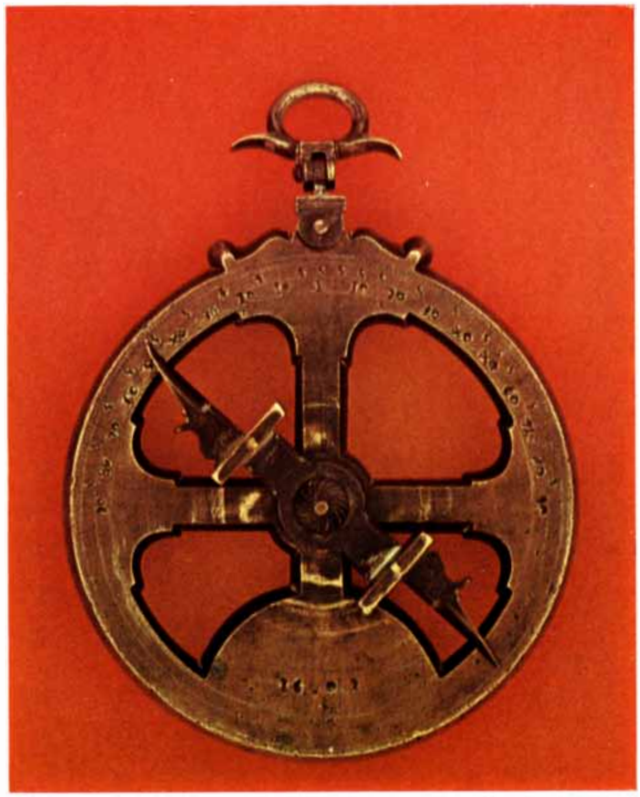
Both sides of the astrolabe bore valuable information. Generally speaking, the alidade was pivoted on the back. The back was a repository for information that could in principle have been stored elsewhere. It usually carried a number of scales and tables whose precise nature tended to change from century to century.

A scale that is found on almost all astrolabes is the calendar scale, which represented the days and months and correlated the position of the sun with

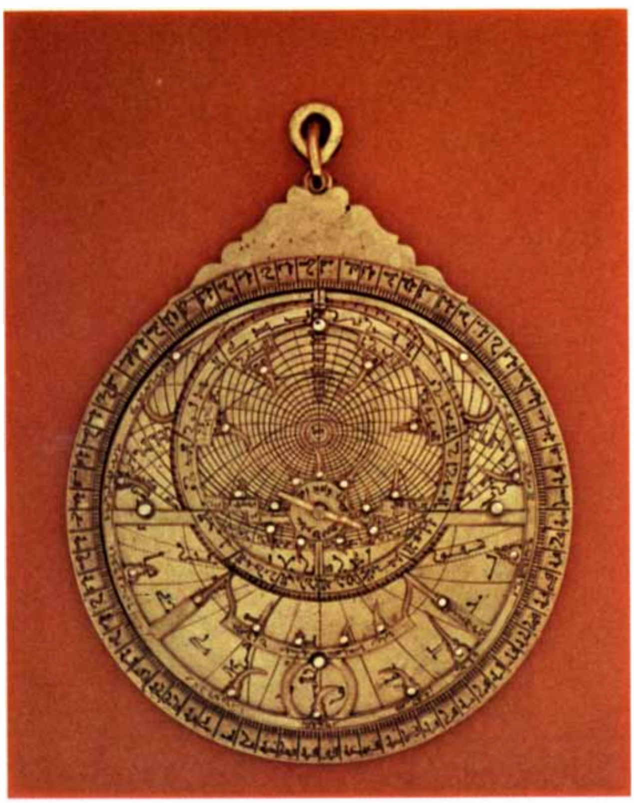


EXPLODED VIEW OF AN ASTROLABE shows the relationship of its various parts. The mater ("mother") is the main body of the astrolabe. The climates are plates engraved with coordinate lines for different latitudes, usually those to which the observer might travel. The alidade is a straight rule that was used for sighting celestial objects and finding their altitude. It was held to the back of the astrolabe (*see illustration on page 105*) and was free to rotate like the rete. The rete fits over all the climates, which are contained within the mater. The pin slides through the centers of all the plates and is secured by the horse, a wedge whose thicker end was traditionally in the form of a horse's head. Some astrolabes had no loose plates; in such instruments the mater was engraved as the one and only climate.

PERSIAN ASTROLABE was made in the 12th century. It is 5⅛ inches across and is of fairly simple design, as are most early astrolabes. Rete is regular and star pointers are straight and unadorned.



MARINER'S ASTROLABE was a crude device used chiefly to find the altitude of celestial bodies above the horizon. Alidade is on the front. This astrolabe is probably Spanish and is dated 1602.



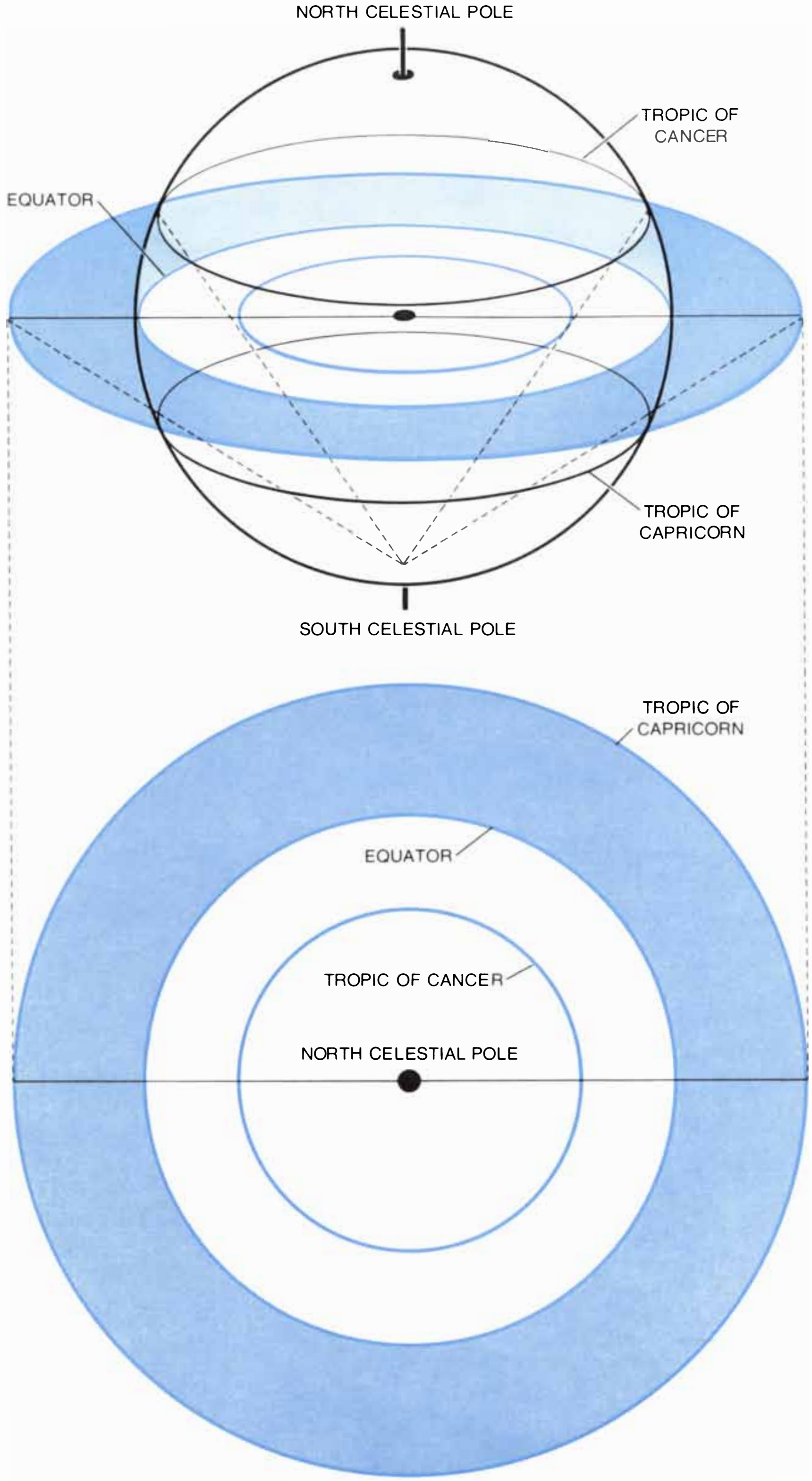
MOORISH ASTROLABE of the 13th century has raised knobs on the rete to assist the observer in rotating it. Both the front (*photograph at left*) and the back (*photograph at right*) of the instrument are shown to illustrate the placement of the rete and the alidade. All astrolabes on this page are in the collection at the National Museum of History and Technology of the Smithsonian Institution.

the date within the year. If the stars were visible by day, it would be easier to appreciate the apparent movement of the sun against the stellar background. This movement is of course a consequence of the earth's motion around the sun; as the earth proceeds along its orbit the sun appears to shift with respect to the stars. It is therefore often convenient to speak as though the earth were at rest in the center of a vast sphere on which all the celestial objects are situated. The stars and even the planets are at such immense distances in comparison with the size of the earth that the celestial sphere is a reasonable convention, as long as one is concerned only with the direction of the celestial objects from the observer.

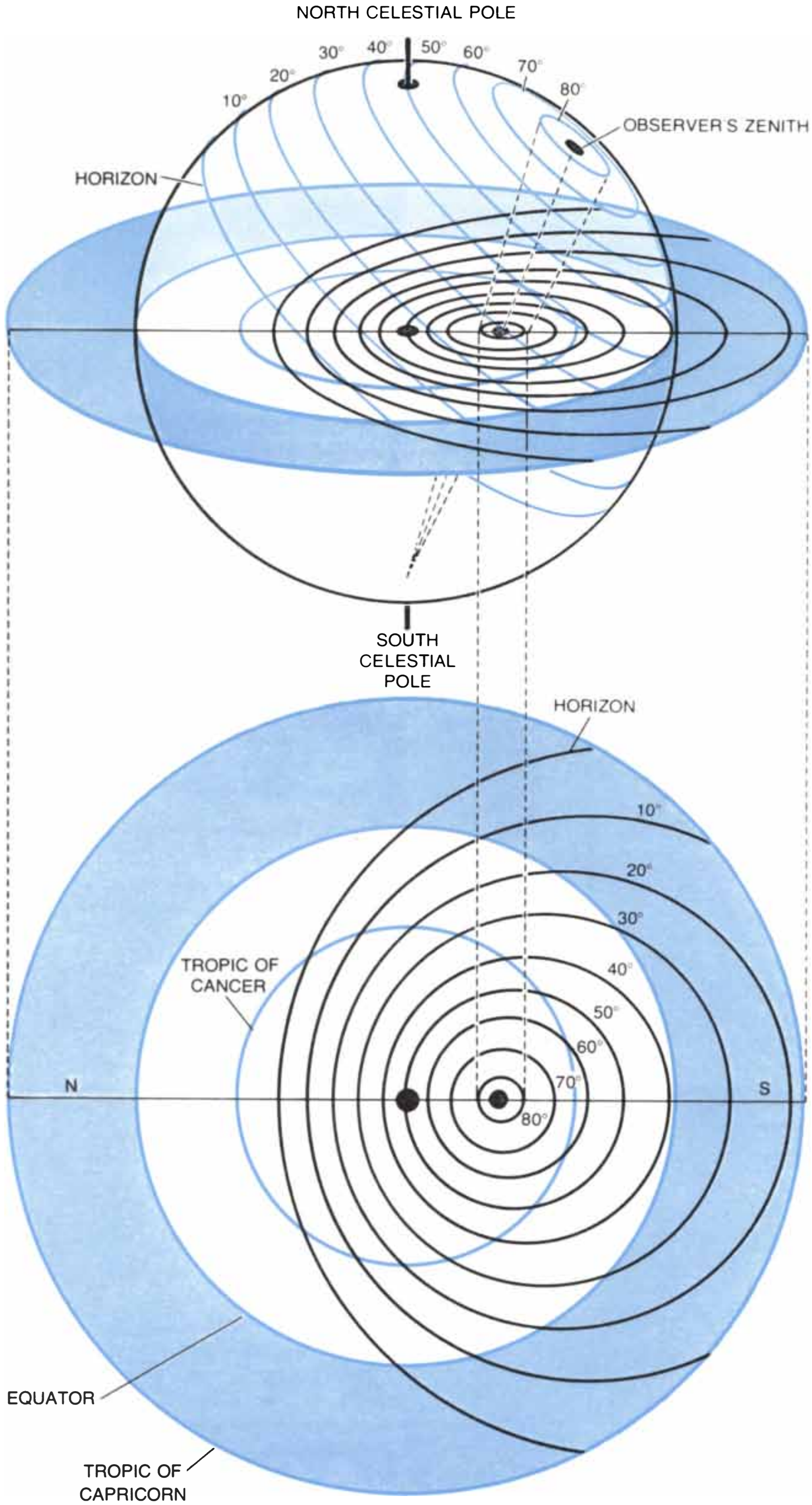
The path of the sun on the celestial sphere is the ecliptic, and the sun completes one circuit of the sky along this path in a year. The planets appear to travel in a band of sky several degrees on each side of the ecliptic; this band is the zodiac. It is possible to give the approximate position of the sun on the ecliptic (its place in the zodiac) for any date of the year. Leap years present a small problem, but it is not a very difficult one, since the accuracy required is only a relatively large fraction of a degree.

The calendar scale of the astrolabe has engraved on it the days and the months. There is also a zodiac scale, usually concentric with the scale of dates, which correlates the dates with the sun's position on the ecliptic. The sun's position can be given as a celestial longitude from zero degrees to 360 degrees, reckoned from some suitable point of origin. In the Middle Ages a variant of this system was used: the zodiac was divided into 12 signs. Each sign was 30 degrees in length and had been named after a prominent constellation. In actuality the constellations had long previously moved into neighboring signs as a consequence of the slow precession of the equinoxes, which in turn is due to a conical movement of the earth's axis. Partly because of precession, and partly because the time it takes the earth to go around the sun once is not exactly 365¼ days, there are small shifts in the sun's position for any particular date as the years pass. These shifts can be taken care of without much difficulty by the rules of the calendar. On an astrolabe, however, it could not easily be done, and a medieval calendar scale is likely to be 10 or 11 days out of register with one of today.

The front of the astrolabe is more important than the back. It has two principal parts. One, the rete, is a fretted plate,



STEREOGRAPHIC PROJECTION OF EQUATOR AND TROPICS shows how these circles on the celestial sphere (*top*) are projected onto the astrolabe plate (that is, onto the mater or one of the climates) or onto the rete. On most astrolabes the plane of the equator (or a plane parallel to it) is taken to be the plane of the projection. A line is extended from the south celestial pole to the desired point on the celestial sphere (in this case one of the tropics or the equator). The point where this line intersects the plane of the projection is the location of that celestial point on the map. A series of such points is charted to yield the coordinate lines. The equator and tropics are at right angles to the axis of the projection. As a result they turn out to be circles that are concentric and centered on the point representing the north celestial pole (*bottom*). Pin goes through the north celestial pole.



STEREOGRAPHIC PROJECTION OF ALMUCANTARS, or circles of equal altitude concentric with the observer's zenith and parallel to the horizon, makes circles on the plane of the projection. They do not, however, have a common center. In the illustration the observer's zenith is 40 degrees north of the equator. His horizon and almucantars are first shown as they appear on the celestial sphere (*top*). The stereographic projection has the property that circles on a sphere remain circles when they are projected onto a flat plane. In projection all the centers of the almucantars lie on the line (*NS*) that runs through both north pole and observer's zenith (*bottom*). Line is the projection of observer's meridian.

which like the rest of the astrolabe is usually made of brass. It overlies an unperforated circular plate. The rete (from the Latin word meaning "net") is a representation of the heavens. The tips of small pointers mark the positions of the brightest stars, an off-center circle represents the ecliptic, and there are also parts of three circles representing the celestial equator and the tropics of Cancer and Capricorn. Through the center of the rete is a pin around which it can rotate. The pin, which also holds the alidade on the back, is kept in place by a wedge passing through a hole in the point of the pin. The thicker end of the wedge was traditionally in the form of a horse's head, and thus the wedge was often called "the horse." If any durable transparent material had been readily available at the time, the rete would probably have been made of it; anyone today who wanted to build a simple astrolabe could use a sheet of plastic to make the star map.

The other principal part of the astrolabe is the plate under the rete. It is graduated with a series of circles and straight lines representing coordinate lines that are fixed with respect to a given observer. The center of the astrolabe, around which the rete turns, represents the north celestial pole, around which the stars appear to turn. Concentric with it are the Tropic of Cancer, the celestial equator and the Tropic of Capricorn. These circles can be represented both on the rete and on the plate below it. On the plate there is a line representing the observer's horizon and a point for his zenith. There is a set of almucantars, or circles of constant altitude, above the horizon and encircling the zenith. There are also lines of constant azimuth, which appear as arcs of circles radiating from the zenith and running down to the horizon.

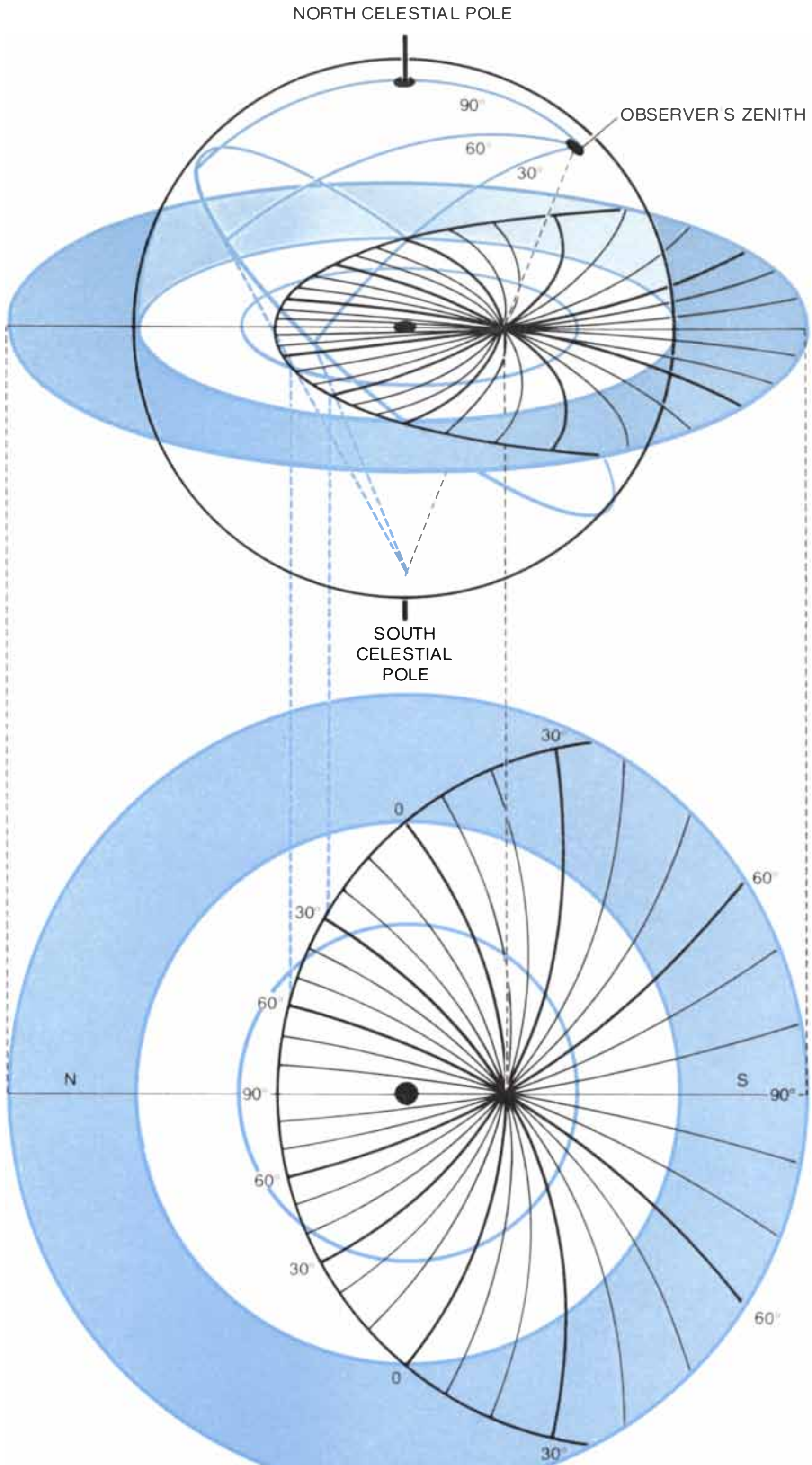
Clearly the distance separating the pole and the observer's zenith on the astrolabe plate depends on the geographic latitude of the observer. If he lived at the North Pole, the two points should coincide, whereas if he lived at the Equator, the two should be separated by whatever represents 90 degrees on the astrolabe plate. The necessity of having a different plate for every latitude at which the instrument was to be used was always a source of chagrin to the astrolabist. He would have a plate for his own latitude, and he might have as many others as he was likely to need on his travels.

Such plates were often called climates by an obvious extension of meaning. An astrolabe might have as many as four, five or even more climates, each plate being engraved on both sides and all being stacked in the mater, or main body, of the astrolabe. They fitted under the rete and were secured by the pin and horse [*see illustration on page 97*]. There were astrolabes that could be used at any latitude with a single plate, but they were not easy to use nor were they ever common.

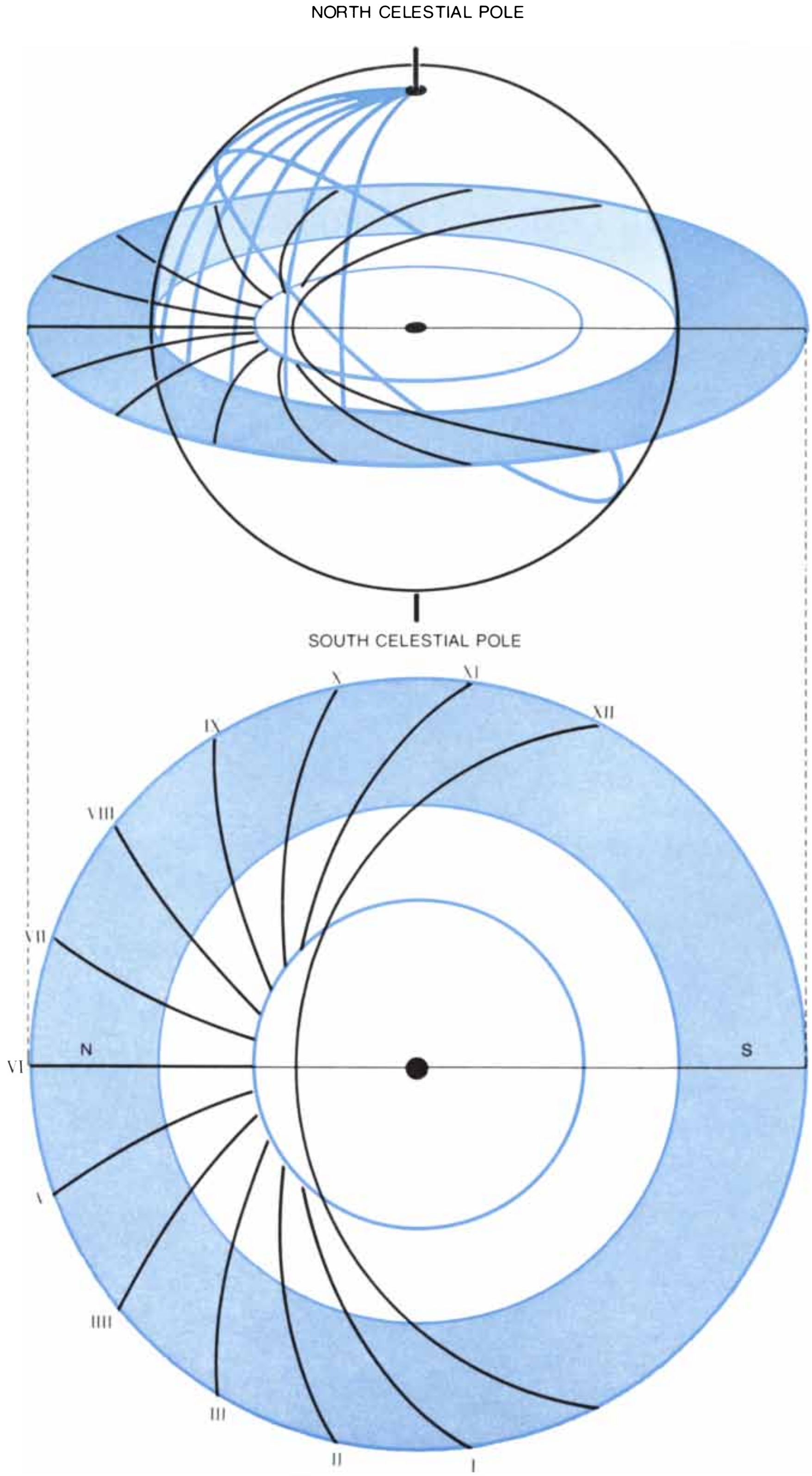
How are the stars and the coordinate lines on the celestial sphere mapped onto the rete and the climates? Suppose the observer was at the center of a large hemispherical dome on which the almucantars and the coordinate lines of constant azimuth were drawn at intervals of five or 10 degrees. Through this series of lines he would be able to see the stars of the night sky, which would move with respect to the lines because of the daily rotation of the earth. If the observer were to take a long-exposure photograph, the pinpoints of starlight would trace out arcs of concentric circles rotating around the north celestial pole. (In the true medieval manner we shall overlook the needs and prejudices of those living in the Southern Hemisphere.)

Just as it is possible to make a flat map of a terrestrial globe, so it is possible to map the two spheres introduced here: the fixed network of coordinate lines and the moving sphere of the sky. There are certain necessary practical requirements if the maps are going to be made of brass and are to serve at all times. If the two maps are to be arranged so that one pivots around a fixed point of the other, as with an astrolabe, then this point should be one of the poles, preferably the north pole if the instrument is to be used in the Northern Hemisphere. Furthermore, the projections of both maps should be alike for all positions of the rete and the plate with respect to each other; a map projection would be no good at all if it meant that the rete had to be distorted as it rotated.

The stereographic projection was admirably suited to the needs of the astrolabist. It has the property that circles on a sphere remain circles when they are projected onto a flat plane, and that the angles between intersecting circles on the sphere remain unchanged when they are projected. Although there are reasons for suspecting that other conventions were used in earlier times, the convention that was almost universally followed with small astrolabes was to project stereographically from the south



STEREOGRAPHIC PROJECTION OF LINES OF EQUAL AZIMUTH is a series of great circles that stretch from the horizon to the zenith. Hence they cut the horizon circle, and the almucantars (*not shown*), at right angles (*top*). Angles between intersecting circles on a sphere remain unchanged when they are stereographically projected onto a flat plane. Therefore on the astrolabe plate the lines of equal azimuth will be arcs of circles that cut the lines of the horizon, and the almucantars (*again not shown*), at right angles. Most astrolabes show only the lines of equal azimuth that would appear above observer's horizon (*bottom*).



STEREOGRAPHIC PROJECTION OF HOUR ANGLES **places the last set of coordinate lines on the plate of the astrolabe. The entire cycle of one day is divided into 24 hours. When the time was reckoned in unequal hours, as it is here, the period of daylight and the period of night, regardless of their duration, were both divided into 12 equal parts. Thus the hours of the day were not equal in length to the hours of the night. The hour lines were usually drawn only below the horizon line. Those portions of the concentric circles of the equator and the tropics are divided into 12 equal parts, beginning with the points of intersection with the west horizon. Corresponding points are connected then with smooth curves.**

pole of the celestial sphere onto the plane of the equator. A line was extended from the south pole to the desired object on the celestial sphere; the point where this line intersected the plane of the projection was the location of the celestial object on the map. A series of such points was mapped to yield the coordinate lines.

With this stereographic projection, the closer a southern star is to the south celestial pole, the farther it will be from the north celestial pole on the plane of the projection, that is, on the rete. The projection of the entire celestial sphere is infinite in extent. In practice the rete is almost invariably made only a little larger than is necessary to accommodate the Tropic of Capricorn. Stars on the rete are represented by the tips of brass pointers. In principle these could be bent after a time to allow for the precessional movement of the earth's axis (although such allowance is not worth bothering with over periods of half a century or less). The bending is more likely to happen by accident than by design, however, and the pointers were usually made as rigid as possible. On the rete the circles of the tropics and the equator are not much needed, since they also appear on the plate below, and so they simply serve largely as supports for the star pointers.

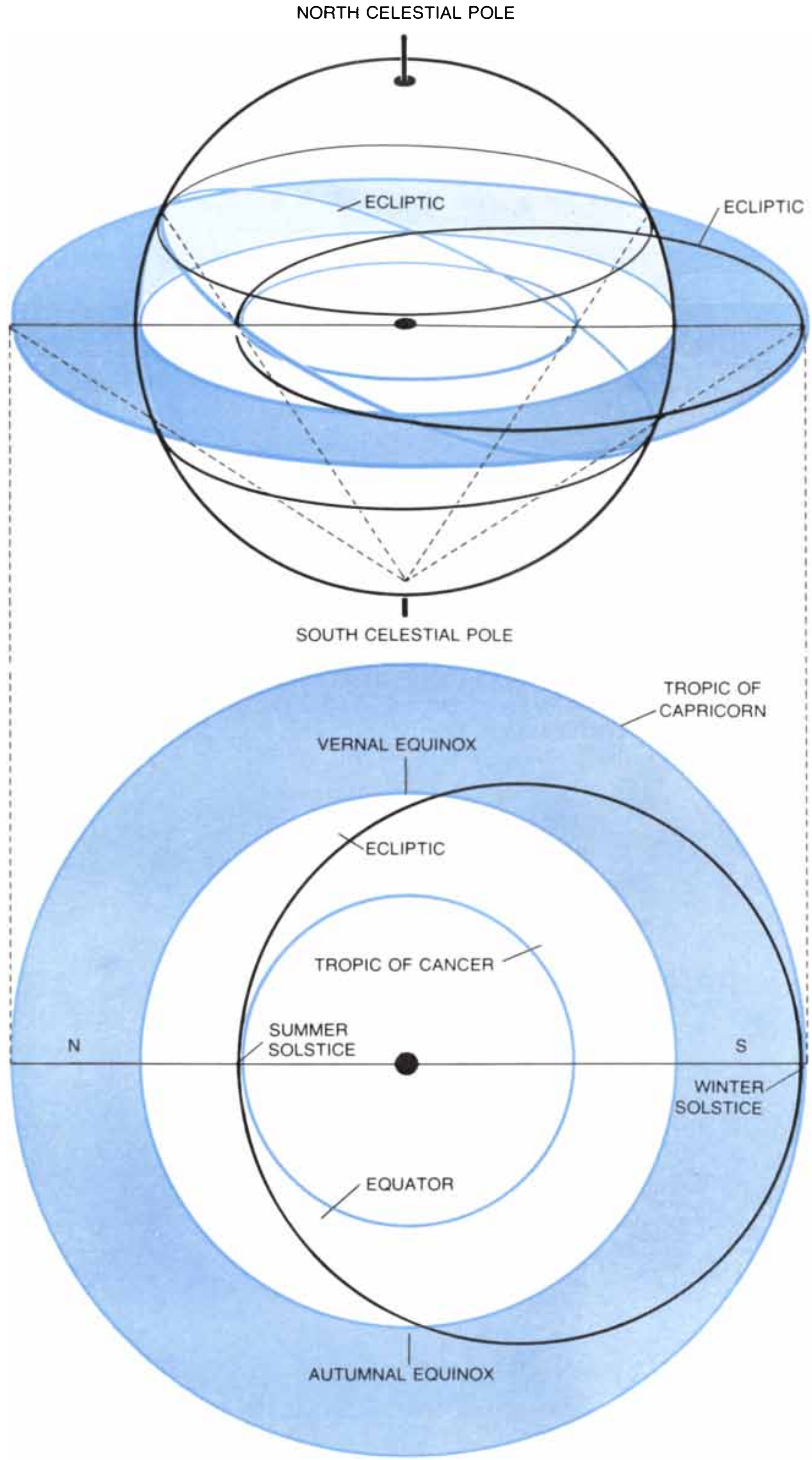
The equator and the tropics are at right angles to the axis of the projection. As a result they turn out in projection to be circles that are concentric with the rete and centered on the north pole (represented by the pin). Moreover, if any degree graduations were to be put on the equator of the celestial sphere, they would lie uniformly on the projected equator. Neither of these properties belongs to the most important circle on the rete, namely the ring that represents the ecliptic. The center of the ecliptic ring differs from the center of the equator and the tropics because the plane of the earth's equator is inclined at an angle of 23½ degrees with respect to the plane of the earth's orbit. Longitudes are measured along the ecliptic from the vernal equinox, one of the two points where the ecliptic crosses the equator. This is the beginning of the sign of Aries; when the sun is at the "first point of Aries," day and night are of equal duration.

At the vernal equinox the sun is passing from south of the equator to the north and is heading through Aries into the sign of Taurus on its progression around the ecliptic. When it reaches its most northerly point of the ecliptic at the summer solstice, 23½ degrees north of the equator, it leaves Gemini and passes into Cancer, hence the name of the tropic in the Northern Hemisphere at the latitude of +23½ degrees. As the sun continues its course along the ecliptic it eventually enters Libra as it again crosses the equator, although this time it is passing from north to south. This it does at the autumnal equinox, when again day and night are of equal duration. The sun reaches the winter solstice as it enters Capricorn 23½ degrees south of the equator, hence the name of the tropic in the Southern Hemisphere at a latitude of −23½ degrees. The sun's annual path as a whole is indicated by the outermost rim of the ecliptic ring on the rete.

How is the ecliptic ring on the rete constructed? All that is needed is to plot the points of the summer and winter solstices [*see illustration at right*]. Since in the stereographic projection circles remain circles on the map, these two points define the diameter of the ecliptic circle. The geometric center of the ecliptic will lie midway between the two points. The ecliptic circle, when constructed, will cross the equator at the points corresponding to the equinoxes. (It so happens that the geometric center of the ecliptic always falls at such a point that the angle made at the equinoxes from the center of the ecliptic to the center of the rete is twice 23½ degrees or, more precisely, twice whatever value is accepted for the angle the ecliptic makes with respect to the equatorial plane.)

The almucantars are drawn on the astrolabe plate below the rete in much the same way. The horizon of the observer is inclined to the celestial equator by 90 degrees minus the geographic latitude of the observer [*see illustration on page 100*]. To find the two points determining each of the almucantars, it must be remembered that the almucantars are no longer great circles in planes passing through the center of the earth; they are small circles parallel to the horizon. When they are drawn, the result is a series of circles around, but not concentric with, the observer's zenith. All their centers lie on the meridian.

The lines of equal azimuth are much more difficult to construct. They are a series of great circles stretching from the horizon to the zenith, and cutting the horizon circle and the almucantars at right angles. Since the stereographic projection leaves angles unchanged, the lines of equal azimuth on the astrolabe plate will be arcs of circles that retain this property. In general, astrolabes



STEREOGRAPHIC PROJECTION OF THE ECLIPTIC is used for the rete instead of for the astrolabe plate. The ecliptic is the apparent annual path of the sun on the celestial sphere as seen from the earth. The Equator of the earth is tipped at an angle of 23½ degrees from the plane of the ecliptic, so that this angle is preserved on the astrolabe rete. All that is needed to draw the ecliptic is to plot the point of the summer solstice on the Tropic of Cancer and the point of the winter solstice on the Tropic of Capricorn. These two points define the diameter of the ecliptic circle, whose center lies midway between. Ecliptic crosses equator at points corresponding to vernal equinox (first day of spring) and autumnal equinox (first day of fall). Ecliptic is divided into 12 signs of zodiac starting at the point representing vernal equinox. The lines dividing ecliptic radiate from the north celestial pole.

show only those parts of the lines of equal azimuth that appear above the observer's horizon [*see illustration on page 101*].

Before we turn to some uses of the astrolabe, what of its history? The theory of the stereographic projection can be traced back to one of the greatest of Greek astronomers, Hipparchus. He was born about 180 B.C. in Nicaea, not far from modern Istanbul, and he made observations from Rhodes and Alexandria. Unfortunately most of what we know about him comes from secondary sources. One of the most important of these sources is the Alexandrian astronomer Ptolemy, who was writing some four centuries later. Ptolemy was perhaps the greatest astronomer of the ancient world. His most important book, now known as the *Almagest,* makes no mention of the planispheric astrolabe. There are, however, references in his *Planisphaerium* to the "spider" of the "horoscopic instrument," suggesting that an instrument with something like the later form of the astrolabe was known in his day. The *Planisphaerium* is a treatise not on the astrolabe but on stereographic projection. It is known only from a Latin translation by Hermann of Carinthia (A.D. 1143).

Other scholars besides Ptolemy refer to the astrolabe, but many of the references are cryptic. The oldest surviving account of the instrument's construction and use was written in the sixth century by John Philoponos of Alexandria. A century later Severus Sebokht wrote on the subject in the Syriac language. After this time the instrument became moderately well known, judging from the many different treatises devoted to it in both the Islamic world and the Christian. Perhaps the first European treatise was one written by Hermann von Reichenau, or Hermann der Lahme (the Lame), a monk of Reichenau who died in 1054.

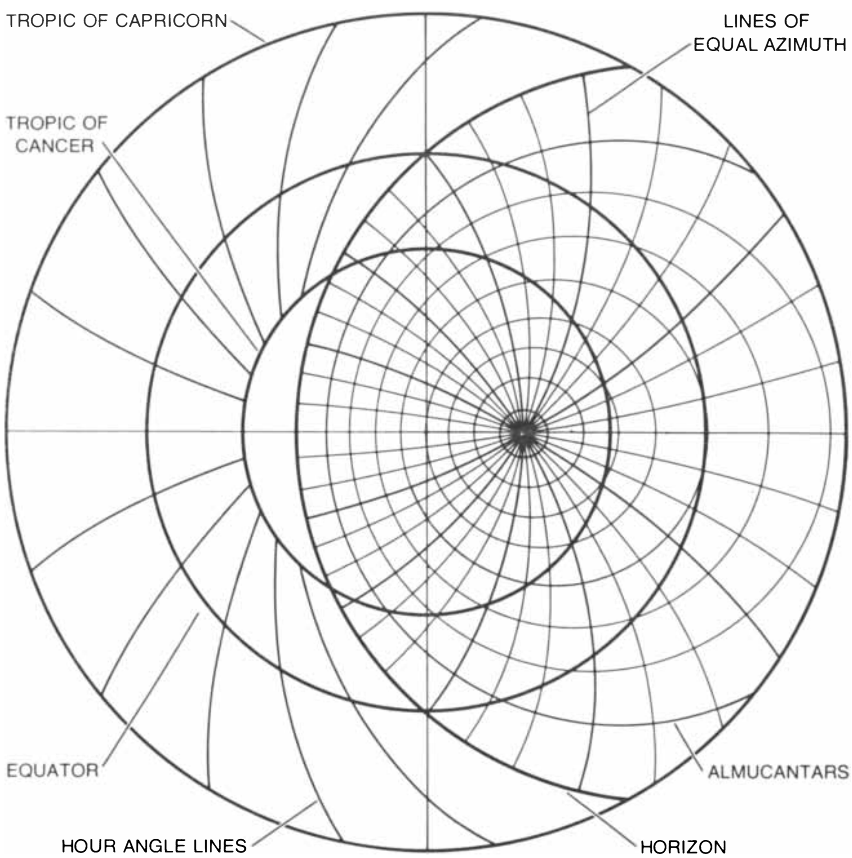
Much better known in medieval Europe was a work originally written in Arabic by Māshā'allāh, who is believed to have been an Egyptian Jew. It was translated into Latin by 1276, and it was the basis of the only good early treatise on the astrolabe in English, namely the one written a century later by none other than Geoffrey Chaucer.

His work, titled *A Treatise on the Astrolabe,* is dated about 1392. It survives in more than two dozen early manuscripts. In some of them it has the subtitle *Bread and Milk for Children.* The subtitle was probably provided by a scribe who was surprised at Chaucer's opening remarks in a work that would have been generally regarded as hardtack for adults. In a modern rendering the work begins:

"Little Lewis my son, I well perceive signs of your ability to learn the sciences of number and proportion, and I also have in mind your earnest request especially to learn the contents of the treatise on the astrolabe." Chaucer goes on to outline the contents of the treatise, which in fact seems never to have been completed. He explains the need for a work in English, and he mentions his debt to earlier astronomers. It is unfortunate that his English is about as difficult for the ordinary modern reader as Latin was for Lewis.

By the 16th century the advent of printing and the steady improvement in techniques of engraving for publication had given rise to a number of magnificent new treatises on the astrolabe. These were in turn partly responsible for some striking advances in the art of the instrument maker. Astrolabes became larger, more decorative and more finely and accurately engraved. Nevertheless, allowing for differences in the language of the inscription, there was little or nothing about the typical astrolabe of the early 17th century that would not have been immediately familiar to an astrolabist of a thousand years earlier. The oldest surviving dated instrument is believed to date from A.D. 927/8. This particular astrolabe also carries a signature that is difficult to decipher, but which could be an Arabic form of a Greek name (Bastulos or Nastulos).

Before the end of the 13th century the planispheric astrolabe was known and used from India in the east to Islamic Spain in the west, and from the Tropics to northern Britain and Scandinavia. Variations in the general style of decoration are usually characteristic of the country and period of origin. The star pointers of the earliest retes, for instance, are usually of a simple dagger shape engraved only with the name of the star. At the other extreme later Indo-Persian astrolabists would often work the rete into an intricate and highly symmetrical foliate pattern, a difficult thing to do with what is essentially a star map having an asymmetrical natural arrange-



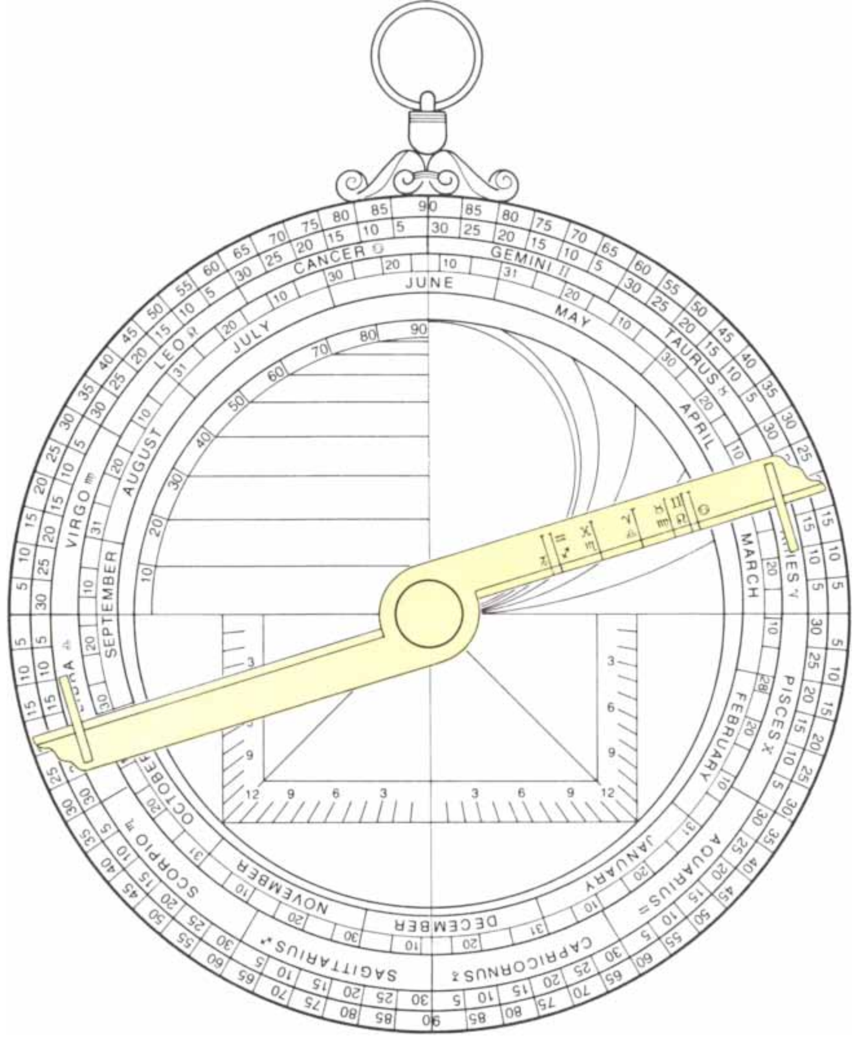
COMPLETE ASTROLABE PLATE shows all the coordinate lines as they appear with respect to one another on a climate. This illustration is a composite of all the stereographic projections shown individually in the series of illustrations on pages 99, 100, 101 and 102.

ment. Astrolabe makers throughout the eastern world often damascened their instruments with silver and gold. It is interesting to trace from their surviving signed work successive generations of the same family. The family might all, for example, have worked at a center such as Lahore, and thus perhaps have had connections with the Mogul court. Persian instruments tended to be extremely ornate, filled with fine ornamental engraving.

In the West the style of the rete is usually reminiscent of contemporaneous styles of church architecture. The style of written inscription is similar to the style of Western manuscripts in general, and is highly characteristic of the period in which it was done. There is good evidence that many astronomers the world over made their own instruments, although there was scarcely any important center of learning that did not at one time or another have its specialist workshops turning out instruments professionally. European instruments were rarely signed with the maker's name during the Middle Ages, a time when anonymity was considered no vice. By the 16th century European astrolabes were often signed.

The physical size of most astrolabes is between three and 18 inches, although much larger ones are found in a number of rather different forms as the dials of astronomical clocks. The use of the astrolabe as a clock dial goes back to classical antiquity, when the rete was made to rotate once daily by waterpower. After the invention of the purely mechanical escapement at the end of the 13th century, astronomical clocks were to be found in most large European cathedrals. In a typical arrangement the star map and the map of coordinate lines of the conventional astrolabe change places, the coordinate lines being made into the rete and the stars being painted on a plate behind it. Usually the stars were made to rotate and the rete was fixed, but sometimes these roles were reversed. A model of the sun is occasionally found on the ecliptic of the star map; it is moved along the ecliptic manually, or by a mechanism, so that it completes one circuit of the ecliptic in a year. In order to judge the time from such a dial one must be familiar with at least the basic principles of the use of the astrolabe.

The chief purpose of the astrolabe was for telling the time. First the altitude of the sun or of a star was found by employing it as an observing instrument.



BACK OF THE ASTROLABE carries the alidade and other information necessary to the observer. Around the rim of the example shown is a scale of degrees for measuring the altitude of a celestial body with the alidade. Immediately inside the rim the 12 signs of the zodiac are listed and divided into 30 degrees each. The scale ot the months and the days inside the zodiac scale correlates the position of the sun on the ecliptic with the correct date. It is not concentric with the other circles to allow for the sun's nonuniform motion along the ecliptic. The design of the interior portion of the astrolabe back varies widely with the individual instrument. Here the quadrant at upper left contains horizontal lines from the degree markings; their distances from the horizontal diameter of the astrolabe correspond to the sine of the altitude of an object above the horizon. The quadrant at upper right contains lines for computing the time in unequal hours directly, independently of the front of the astrolabe. These lines are used in conjunction with the graduations on the alidade. The two quadrants at bottom contain the "shadow squares." These could have been used in conjunction with a gnomon to get the cotangent or the tangent of the altitude of an object above the horizon. If they were accurately and completely divided, which they rarely were, they provided a means of measuring altitudes more precise than sighting with alidade.

Then, assuming that the observer knew where the sun or the star was on the rete, the rete was revolved until that point coincided with the almucantar for the appropriate altitude. (It is assumed that the observer knew which climate to choose for his latitude and on which side of the meridian line the object fell.) The refraction of the atmosphere, which changes the apparent position of objects in the sky, and which is greater the nearer they are to the horizon, was ignored. The sun's approximate position on the ecliptic for any day of the year is found from the calendar scale on the back of the astrolabe.

Once the rete is in the correct position, the observer can find his local time according to any one of several conventions. If the circumference of the astrolabe is marked in degrees, 15 degrees correspond to an hour. Noon will be when the sun is toward the top of the instrument, midnight when it is toward the bottom, 6:00 A.M. when it is to the left and 6:00 P.M. when it is to the right. Imagine now a great circle joining some object in the sky to the north celestial pole. The angle that this great circle makes with the meridian is the hour angle of the object. As a consequence of the stereographic projection, a rule lying on a line passing through the center of the astrolabe and the point of the instrument representing the object makes an angle with the vertical diameter (the meridian line) equal to the hour angle of the object. The hour angle is so named because it can provide one with a measure of time through its change as the earth rotates. It is usually quoted in hours, minutes and seconds rather than in degrees of arc.

A number of different kinds of time can be told from an astrolabe. The first is sidereal time, or time by the stars, which is defined as the hour angle of the first point of Aries. If the day is counted from zero hours to 24 hours beginning at midnight, 12 hours will have to be added to the count for sidereal time, because at the vernal equinox the first point of Aries (which is at that time the position of the sun) will cross the meridian at local noon.

A second kind of time is true solar time: the hour angle of the sun regardless of its position with respect to the stars. There is another and more familiar type of solar time called mean solar time, which postulates a "mean sun" moving around the equator (rather than the ecliptic) at a uniform rate throughout the year, and making one complete circuit in exactly a year like the true sun. The earth moves around the sun in an ellipse with the sun at one focus, and it travels faster in its orbit the closer it is to the sun. Therefore from the earth the true sun seems to speed up and slow down in its course around the ecliptic. Thus the true sun and the mean sun move not only along different paths but also at different rates. In order to convert from observed true solar time to the more useful mean solar time, one must apply a correction known as the equation of time. It is based on knowledge of the earth's motion in its orbit and it can be found in reference books. The correction for the equation of time was scarcely ever applied before the 17th century. In order to convert mean solar time to the local time at some standard location such as Greenwich, the observer needs to know his geographic longitude, and again that adjustment was seldom made.

A third kind of time is time measured in unequal hours. The ordinary man in the Middle Ages divided the period of daylight into 12 equal parts and the period of night into 12 equal parts regardless of the actual length of day and night. The length of the day-hours would obviously equal the length of the night-hours only when the sun was at one of the equinoxes. Many astrolabe plates include unequal-hour lines. In order to avoid their being confused with the almucantars, the unequal-hour lines were drawn only below the horizon line [*see illustration on page 102*].

Time in the Middle Ages was often reckoned from sunrise or sunset even when measured in ordinary equal hours. Many astrolabe plates show lines similar in appearance to the lines of unequal hours but which are in fact for measuring the time in equal hours from sunrise or sunset.

Although the astrolabe was primarily an instrument for determining the time, it was an extremely useful adjunct of the astrologer's art. To cast a horoscope for a particular moment of time, an astrologer needs to know the degree of the ecliptic that is on the eastern horizon ("the ascendent"), the degree of the ecliptic that is on the western horizon ("the descendent"), the degree of the ecliptic where it crosses the meridian ("the degree of mid-heaven") and the degree of the ecliptic where it crosses the northward continuation of the meridian, once called the midnight line ("lower mid-heaven"). These degrees are easily read off the ecliptic ring once the rete is correctly positioned for the moment of time that is of interest: perhaps a moment of conception, of birth, of death or of some other important event such as a coronation. Once the four key points of the horoscope have been found, the 12 astrological houses (which are not to be confused with the signs) can be ascertained and the planets can be assigned to them. There are, however, many systems by which the division can be made. These can be found in Chaucer's treatise on the astrolabe.

Like a modern electronic computer, the astrolabe in the Middle Ages was a source of astonishment and amusement, of annoyance and incomprehension. Imprecise as the astrolabe may have been in practice, it was undoubtedly useful, above all in judging the time. The instrument might have been used, more often than not, in the dark, but "dark" is hardly the word to describe the age in which it was so widely known and so well understood.

MATHEMATICAL GAMES

The combinatorial basis of the "I Ching," the Chinese book of divination and wisdom

by Martin Gardner

The *I Ching* (pronounced *ye jing*), or *Book of Changes,* is one of the world's oldest books and also one of the most enigmatic. For more than 2,000 years it has been used in the East as a book of divination and it still is studied with awesome reverence as a rich source of Confucian and Taoist wisdom. Tens of thousands of young people in the U.S. (particularly in California), caught up in the current occult explosion and eager to know more about Eastern mysticism and early Chinese history, are now consulting the *I Ching* as seriously as they consult the Ouija board or the tarot cards. C. G. Jung was convinced of the *I Ching*'s extraordinary power to foretell the future; he even asked it about the prospects of American sales of a new English translation of itself and got an optimistic answer. More recent pundits who are deep into occultism—Colin Wilson, for example—have written about their experiences with the *I Ching*'s terrifying oracular accuracy.

The early history of the *I Ching* is unknown. Most likely it began as early as the eighth century B.C. as a collection of peasant omen-texts, then slowly over the centuries these documents became combined with stick divination practices. A few centuries before Christ, near the end of the Chou dynasty, it acquired its present form and became one of the five great classics of the Confucian canon.

The combinatorial foundation of the *I Ching* consists of 64 hexagrams. They display every possible permutation of two types of line when taken six at a time. Each hexagram has a traditional Chinese name. The two kinds of line proclaim the basic duality of Chinese metaphysics: the broken line corresponds to yin, the unbroken line to yang. Taking the lines two at a time, there are $2^2 = 4$ ways to combine them into what are called digrams and $2^3 = 8$ ways to form trigrams. The trigrams, with their Chinese names and symbolic meanings, are shown in the illustration on page 110.

There are two ancient ways of displaying the eight trigrams in a circle. The oldest, known as the Fu Hsi arrangement after the mythical founder of China's first dynasty (the Hsia dynasty, 2205–1766 B.C.), is shown at the left in the bottom illustration on page 111 and also on the cover of this issue. Note that opposite pairs are complementary both in symbolic meaning and in the mathematical sense that each is obtained from the other by replacing yin lines with yang and yang lines with yin. This arrangement, usually surrounding the familiar yin-yang symbol, is still widely used throughout China, Japan and Korea as a good-luck charm to put over doorways and on jewelry. It is also called the "earlier heaven" or "primal arrangement." The King Wen arrangement (after the legendary father of the founder of the Chou dynasty), shown at the right in the bottom illustration on page 111 (also called the "later heaven" or "inner-world arrangement"), abandons the complementary positioning of the Fu Hsi sequence, so that the trigrams at the cardinal points of the compass symbolize the seasons in cyclic order. Starting at south (traditionally shown at the top) and going clockwise, the hexagrams at the cardinal points stand for summer, fall, winter and spring.

The oldest way of arranging the 64 hexagrams, which are known as the King Wen sequence, is the order in which they appear in the *I Ching* [*see illustration on opposite page*]. The rows are taken from right to left as the numbering indicates. Note that the hexagrams are paired in a singular way. Each odd-numbered hexagram is followed by a hexagram that is either its inverse or its complement. If the odd hexagram has twofold symmetry (is the same upside down), it is complemented to produce the next hexagram. If it lacks twofold symmetry, it is inverted.

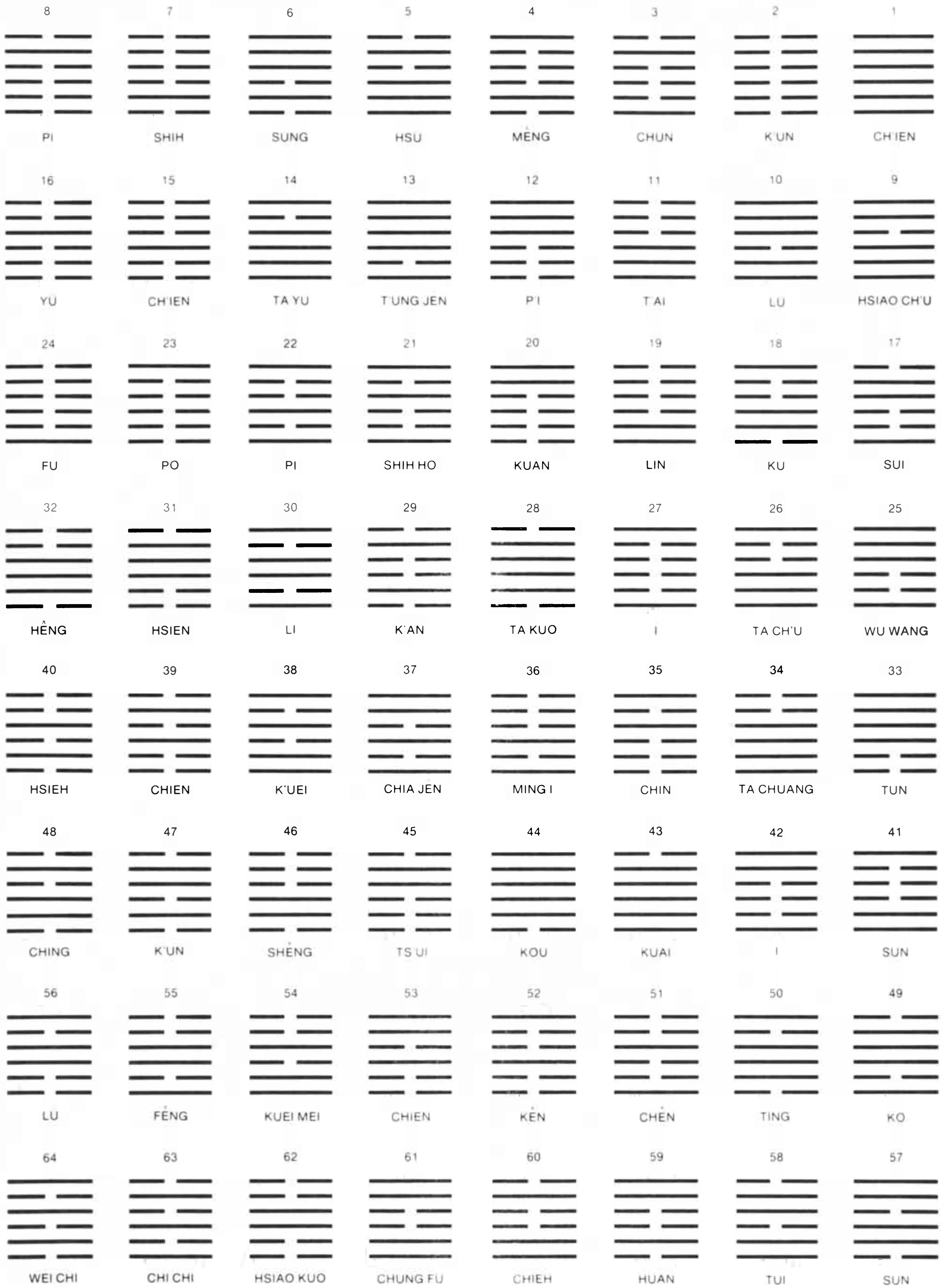
Is there any kind of mathematical order that determines the sequence in which the hexagram pairs follow one another? This is an unsolved problem. From time to time a student of the *I Ching* announces his discovery of a mathematical scheme underlying the arrangement of pairs, but on closer inspection it turns out that so many arbitrary assumptions are made that in effect the order must be assumed before it emerges from the analysis. As far as anyone knows, the pairs of the King Wen sequence are in random order, and there is no known basis for determining which member of a pair precedes the other.

Not until the 11th century did Chinese scholars discover a very simple and elegant way to order the hexagrams. This arrangement is attributed to Fu Hsi [*see top illustration on page 111*]. The white space at the bottom represents the *t'ai chi,* the state of the universe when it was "without form, and void" (as Genesis 1:2 puts it). This undifferentiated chaos divides into the yin (colored) and yang (black) halves of the row labeled 1. In row 2 we see the yin dividing into yin and yang, and similarly the yang. These binary divisions continue upward for six steps.

The chart now automatically gives all the polygrams of orders 1 through 6. Divide rows 1 and 2 vertically into four equal parts, replace the color in each part with broken (yin) lines and you have the four digrams. Rows 1, 2 and 3, divided vertically into eight equal parts, give the eight trigrams. Rows 1, 2, 3 and 4, in 16 parts, give the 16 tetragrams, rows 1, 2, 3, 4 and 5, in 32 parts, give the 32 pentagrams, and rows 1, 2, 3, 4, 5 and 6, in 64 parts, give the 64 hexagrams. The illustration on page 112 shows the hexagrams in their traditional Fu Hsi, or "natural," order. Taking them from right to left, starting at the bottom row, the hexagrams correspond to those provided by the Fu Hsi chart when read from left to right.

We are now ready to understand why Leibniz, who thought he had invented the binary system in the late 17th century, was so staggered when he first learned of the Fu Hsi sequence from Father Joachim Bouvet, a Jesuit missionary in China. Substitute 0 for each unbroken line, 1 for each broken line, then take the hexagrams in order, reading upward on each, and you get the sequence 000000, 000001, 000010, 000011, . . . , 111111. It is none other than the counting numbers from 0 through 63 expressed in binary notation!

Both Leibniz and Father Bouvet were convinced that Fu Hsi, smitten by divine inspiration, had discovered binary arithmetic, but there is not the slightest evidence for this. The 11th-century *I Ching* scholars had done no more than discover



King Wen arrangement of the 64 I Ching hexagrams a natural way to arrange the hexagrams. It was not until the time of Leibniz that the Fu Hsi sequence was recognized as being isomorphic with a useful arithmetical notation.

Since the powers of 2 turn up everywhere in mathematical and physical structures, it is not surprising that Chinese scholars have been able to apply the 64 hexagrams to almost everything, from crystal structures to the solar system and the cosmos. Z. D. Sung, in his amusing little book *The Symbols of Yi King* (Shanghai, The China Modern Education Company, 1934), tells how he was rotating a matchbox in his hand one day (to simulate the earth's rotation as it goes around the sun) when he suddenly perceived a natural way to generate the eight trigrams at the corners of a cube.

Let the three Cartesian coordinates of a unit cube, x, y, z, indicate the first, second and third digits of a three-digit binary number. Label the corner where the coordinates originate with 000. The other corners are labeled with three-digit binary numbers for 0 through 7, with 0 and 1 indicating the distance of the corner from the origin in each coordinate direction. The eight numbers correspond, of course, to the eight trigrams, with complementary trigrams at diametrically opposite corners of the cube [*see illustration on page 113*]. By a similar procedure corners of unit hypercubes generate the higher-order polygrams. The 64 hexagrams correspond to six-digit binary numbers at the corners of a six-dimensional hypercube.

Instead of plunging into higher dimensions, Sung divides the cube into 64 smaller cubes that he identifies with the 64 "moods" of the classical syllogism. (The major premise, the minor premise and the conclusion of a syllogism can each be of four different forms, giving 64 possible moods.) Sung was probably unaware that this had been done earlier by C. Howard Hinton in his 1904 book *The Fourth Dimension* (pages 90–106). Hinton takes a curious step into hyperspace. By considering the four "figures" of each syllogism (an ancient division based on the ordering of the subject, predicate and middle terms), he obtains 256 varieties that he identifies with the 256 cells of a 4-by-4-by-4-by-4 hypercube. Cells corresponding to traditionally valid syllogisms are colored black, then the hypercube is projected onto an ordinary 4-by-4-by-4 cube. The black cells are seen to be symmetrically disposed around one corner of the cube except for one cell that should be black but is not. This led Hinton to the discovery that the anomalous syllogism is valid after all if one applies a more liberal interpretation to syllogisms, one in which the predicate is quantified as well as the subject.

But we have strayed from the *I Ching*. The book (aside from its "Ten Wings," which are appendixes by Confucian metaphysicians) consists essentially of the 64 hexagrams, each followed by a brief explanation of the symbol and six "appended judgments." If the book is to be used as an oracle, one of the hexagrams must be randomly selected, and this must be done in such a way that the rules tell how to transform the chosen hexagram to a second one.

The oldest selection procedure, still followed by those who take the *I Ching* the most seriously, calls for 50 yarrow stalks, each one to two feet long. If yarrow stalks are not obtainable, 50 thin wooden sticks will serve. They should be kept in a lidded receptacle at a spot

TRIGRAM	NAME	IMAGES	TRAITS	FAMILY RELATIONS	PARTS OF BODY	ANIMALS
☰	CH IEN	HEAVEN COLD	STRONG FIRM LIGHT	FATHER	HEAD	HORSE
☷	K UN	EARTH HEAT	WEAK YIELDING DARK	MOTHER	BELLY	OX
☳	CHEN	THUNDER SPRING	ACTIVE MOVING AROUSING	FIRST SON	FOOT	DRAGON
☵	K AN	WATER MOON WINTER	DANGEROUS DIFFICULT ENVELOPING	SECOND SON	EAR	PIG
☶	KEN	MOUNTAIN	RESTING STUBBORN UNMOVING	YOUNGEST SON	HAND	DOG
☴	SUN	WIND WOOD	GENTLE PENETRATING FLEXIBLE	FIRST DAUGHTER	THIGH	BIRD
☲	LI	FIRE SUN LIGHTNING SUMMER	BEAUTIFUL DEPENDING CLINGING	SECOND DAUGHTER	EYE	PHEASANT
☱	TUI	LAKE MARSH RAIN AUTUMN	JOYFUL SATISFIED COMPLACENT	YOUNGEST DAUGHTER	MOUTH	SHEEP

The eight trigrams and some of their meanings not lower than a man's shoulders. The *I Ching*, carefully wrapped in clean silk, is kept alongside the sticks.

The book must never be consulted lightly. If you ask it something frivolous or in a skeptical mood, the book gives frivolous or meaningless answers. One should be completely relaxed, physically and mentally. It is essential to think of nothing, throughout the ceremony, other than the question being asked.

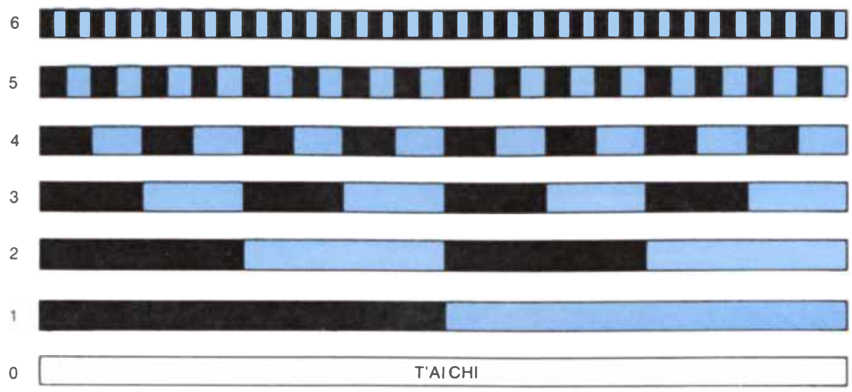
Let us assume that you are asking the *I Ching* a question and also casting the sticks. Your first step is to unwrap the book, spread the silk on a table and place the book on top. (The cloth protects the *I Ching* from impure surfaces.) An incense burner and the receptacle containing the sticks are placed beside the book. With your back to the south, make three kowtows, touching your forehead to the ground; then, still kneeling, pass the 50 sticks three times through the incense smoke by holding them horizontally and moving your hand in a clockwise circle. Return one stick to its container. It plays no further role in the ceremony.

Put the 49 sticks on the cloth, then with your right hand quickly divide them randomly into two piles. Call the left pile *A*, the other *B*. Take a stick from *B* and put it between the last two fingers of your left hand. With your right hand, push away four sticks at a time from pile *A* until one, two, three or four sticks remain. Place those sticks between your left second and third fingers. Next diminish pile *B* by pushing away four sticks at a time until one, two, three or four sticks remain. Place these between your left first and second fingers. (This last step can be shortened. Because the sum of the two remainders must be 0, modulo 4, the second remainder is easily calculated from the first.) Your left hand now holds either five or nine sticks. (The possible combinations are 1, 1, 3; 1, 2, 2; 1, 3, 1, and 1, 4, 4.) Put all these sticks to one side.

The remaining sticks are bunched together and exactly the same dividing procedure is repeated with them, beginning with the random division into two piles. At the finish your left hand will hold either four or eight sticks. (The possible combinations are 1, 1, 2; 1, 2, 1; 1, 3, 4, and 1, 4, 3.) Place them aside, next to the group put aside previously.

Bunch the remaining sticks and repeat the dividing procedure a third time. Your left hand again will hold either four or eight sticks. Put them aside, next to the two groups already there.

The number of sticks that now remain will be either 24, 28, 32 or 36. Count them by groups of four (that is, divide the total number by 4). The quotient will be 6, 7, 8 or 9. These four digits are the ritual numbers, which indicate the character of the bottom line of the hexagram. If the digit is even (6 or 8), the line is yin (broken); if it is odd (7 or 9), the line is yang (unbroken). But the ritual numbers tell you more. Seven and 8 mean that the line (whether yin or yang) is a stable line that cannot be altered. Six and 9 indicate a "moving" line that can be changed (for reasons soon to be explained) to its opposite.



How six yin-yang divisions generate the 64 hexagrams

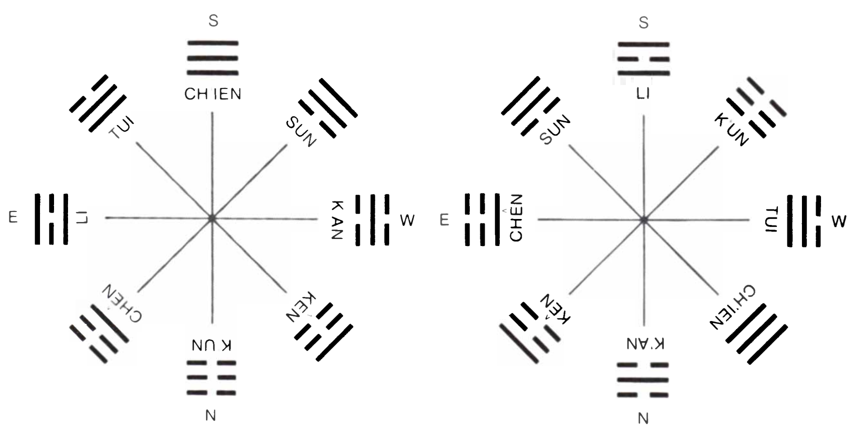
All 49 sticks are now bunched together and the entire ritual is repeated to obtain the hexagram's second line from the bottom. Four more repetitions give the remaining four lines. The entire ceremony, performed without haste, takes about 20 minutes.

Look up the chosen hexagram in the *I Ching* and study its accompanying text carefully. The text will answer your question and give counsel with reference to the present situation. If all six lines of the hexagram are stable, that is the end of the matter. But if one line or more are moving, change them to their opposites and look up the new hexagram. The commentary will pertain to what you can expect in the future if you follow the counsel of the first hexagram.

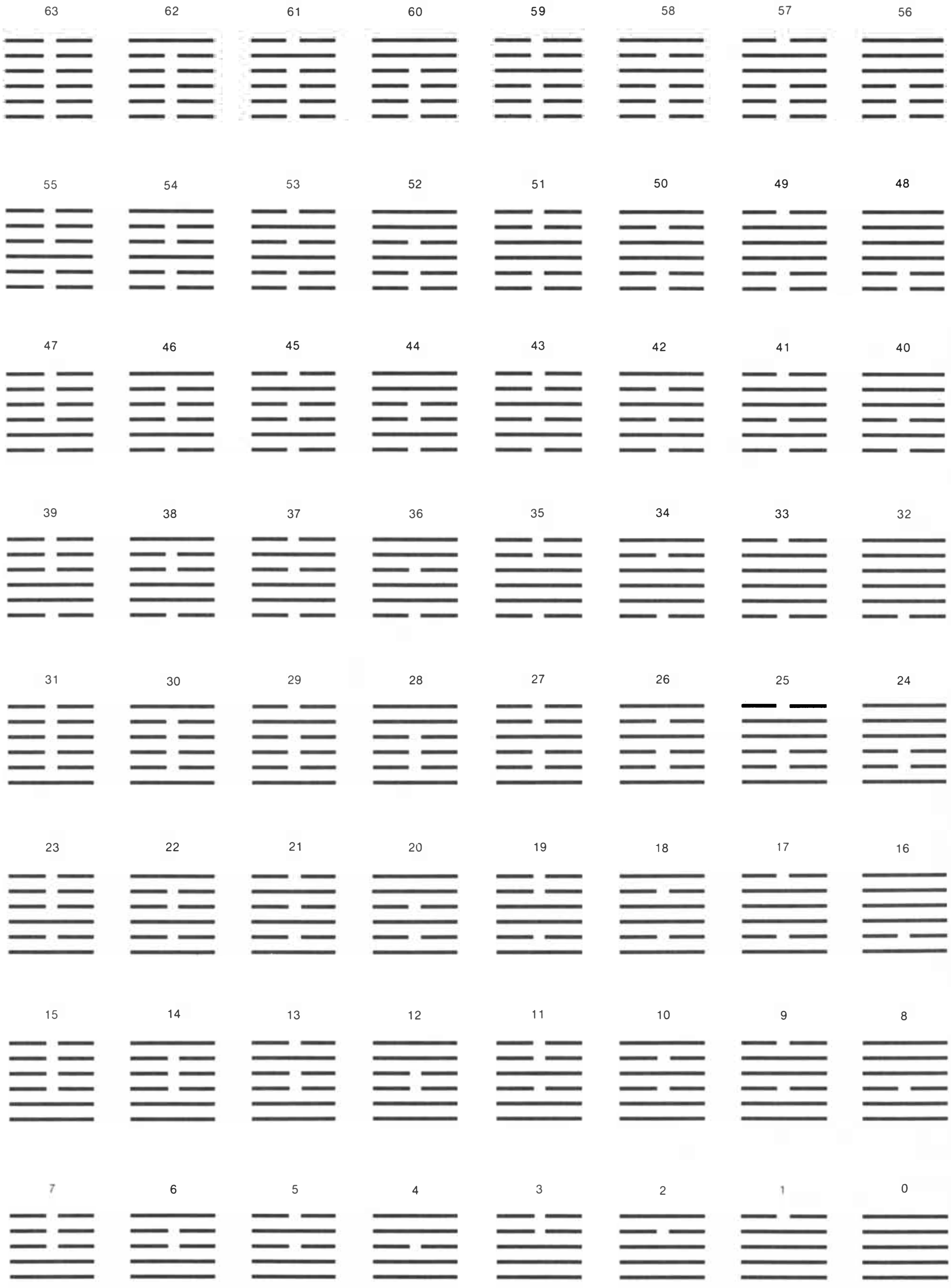
After the one or two hexagrams have been written down and the relevant passages in the *I Ching* have been read and meditated on, light another stick of incense, make three more kowtows of gratitude, put the sticks back in their box, rewrap the *I Ching* in its silk, then put book and sticks back in their usual high place.

Those too lazy to go through the ancient stick ritual can use a simpler method of casting that has been popular in China for several centuries. It calls for six identical coins, preferably old Chinese coins with square holes. They should be kept polished and should never be removed from their container except when the *I Ching* is being consulted. Observe the same beginning ritual followed for the sticks: kowtowing, kneeling, passing the coins through the incense and so on. Shake the coins in your cupped hands and let them drop simultaneously to the cloth. Having previously decided on which sides of the coins are yin and which yang, consult the following chart to determine whether the throw gives you 6, 7, 8 or 9.

Three yins = 6 (a moving yin line).



Fu Hsi arrangement of trigrams (left) *and King Wen arrangement* (right)



The Fu Hsi sequence that corresponds to binary numbers 0 through 63

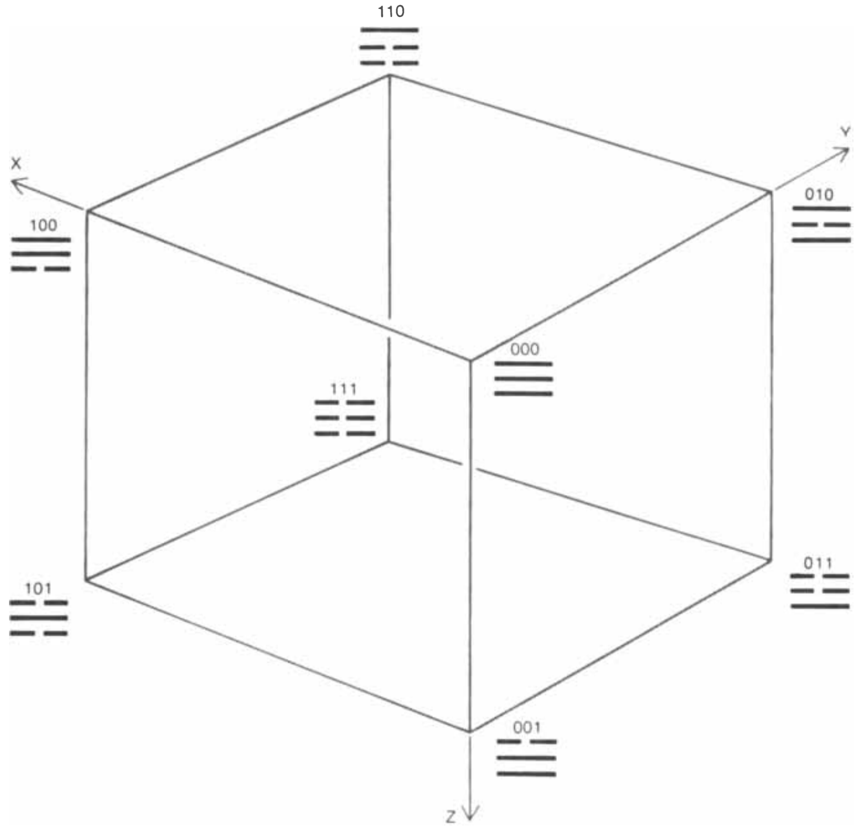
Two yins, one yang = 7(static yin line).
Two yangs, one yin = 8 (static yang line).
Three yangs = 9 (moving yang line).

(If one thinks of the yin side as 2 and the yang side as 3, the sum of the three values will be the desired ritual number.)

Working out the probabilities provided by the stick and coin procedures reveals a subtle difference between the two divination methods. As far as picking the initial hexagram is concerned, the methods are virtually the same, but the probabilities are not the same in choosing the second hexagram. It is not hard to show that in both procedures the probability of choosing a broken line at each of the six steps is 1/2: the same as that of choosing an unbroken line. (This assumes that each time the sticks are randomly divided into piles *A* and *B* and *A* is reduced to one, two, three or four sticks, the probabilities for each of the four outcomes are equal. This is not strictly true, but the deviations from equality are so slight that they have a negligible effect on the final results.) Thus any hexagram has the same probability of being selected as any other. The two procedures are also alike in giving a probability of 1/4 that a given line will be moving. Since there are six lines, 6/4, or 1½, lines of a hexagram, on the average, will be moving.

When coins are used, the probability that a broken line will change is the same (1/4) as the probability that an unbroken line will change, and similarly the probability is 3/4 that each type of line will remain stable. But when sticks are used, this is not the case. The probability that a broken line will change is 1/16 as compared to 3/16 for an unbroken line (or respective probabilities of 7/16 versus 5/16 that the lines will remain stable). In other words, when sticks are cast, it is three times more likely that an unbroken line will change than a broken one. It is true that any hexagram is as likely to be chosen first as any other, but the more broken lines a hexagram has, the more likely it is that it will appear as the second hexagram. Purists who object to coin-casting have sound mathematical support. Not only does the stick ritual discourage frivolous consultation but also its asymmetry produces a more interesting set of probabilities. We shall say nothing about such impious corruptions as the practice of obtaining the ritual numbers from dollar bills, license plates, telephone numbers and so on.

To readers who may wish to experiment with the *I Ching*, my first recommendation is the Richard Wilhelm and C. F. Baynes translation, rendered into English from the German. It is currently in print as a Princeton University Press hardcover. Two good paperback translations are also available: one by James Legge (Dover) and one by John Blofeld (Dutton).



Trigrams generated by a cube

The Wilhelm-Baynes volume includes the famous foreword by Jung in which he explains the oracular power of the *I Ching* by his theory of "synchronicity," a theory defended by Arthur Koestler in his recent book *The Roots of Coincidence*. According to Jung, the *I Ching*'s predictions, and relevant events that actually happen, are not causally linked in the Western scientific sense. They are "acausally" related in the Eastern metaphysical sense of being parts of a vast cosmic design that lies beyond the reach of science but is partially accessible to the subconscious mind of the person who casts the sticks. The 64 hexagrams and their meanings are Jungian archetypes, deeply engraved on the collective unconscious of humanity.

Tough-minded skeptics who test the *I Ching* realize at once why the book seems to work. The text is so ambiguous that, no matter what hexagrams are selected, it is always possible to interpret them so that they seem to apply to the question. Indeed, the scope for intuitive interpretation is so great that in China before Mao (I do not know how it is today) there was a large class of professional *I Ching* interpreters whose services were available for a fee on street corners, at fairs and in marketplaces. Surely one reason for the popularity of coin-casting was that it maximized the profits of these fortune-tellers by speeding up their readings.

And if the *I Ching*'s predictions fail to materialize? Well, perhaps the text was not correctly interpreted, or maybe you were not in the right frame of mind when you were tossing the sticks or coins. Besides, the future is not completely determined. The *I Ching*, like the stars of astrology, does no more than indicate probable trends.

Tender-minded believers in the occult, who have not yet consulted the *I Ching* and who long for powerful, mysterious magic, are hereby forewarned. This ancient book's advice can be far more shattering psychologically than the advice of any mere astrologer, palmist, crystal gazer or tea-leaf reader.

Last month's problem was to express 100 as the sum of three cubes, allowing each cube to be positive or negative. The three known solutions are $7^3 - 6^3 - 3^3$, $190^3 - 161^3 - 139^3$ and $1{,}870^3 - 1{,}797^3 - 903^3$.

THE AMATEUR SCIENTIST

Weather-satellite pictures are picked up in the home

Conducted by C. L. Stong

Pictures of regional cloud cover made by weather satellites have become familiar to television audiences in metropolitan areas of the U.S. as a regular feature of the evening news. The same pictures are captured elsewhere and at more frequent intervals by radio hams, amateur weather observers and a growing number of other enthusiasts. An apparatus for receiving satellite pictures of cloud cover can be assembled by an enterprising beginner for about the cost of a moderately priced television set. A typical installation is operated by Eugene F. Ruperto, Box 166, R.F.D. No. 1, West Alexander, Pa. 15376. Explaining how one of the more highly developed weather satellites broadcasts its signals and how the amateur can make an apparatus to receive them, Ruperto writes:

"After three years of operating a weather-satellite recorder I still get a thrill out of snapping a few switches in the hills southwest of Pittsburgh and receiving from space a picture of Hudson Bay or a tropical storm approaching the coast of Yucatán. Building the station has renewed my interest in the weather and also has introduced me to the study of satellites. Most interesting, however, is the final result: a cloud-cover picture of my immediate area from a spacecraft that passed my way only minutes before.

"My apparatus monitors two polar-orbiting weather satellites, ESSA-8 and NOAA-2. The first was launched by the Environmental Science Services Administration in 1968. It is now approaching the end of its useful life, although it con-

Satellite picture of Great Lakes cloud cover obtained by Eugene F. Ruperto's apparatus tinues to transmit clear signals. The second satellite, NOAA-2, was launched in 1972 after the agency had been renamed the National Oceanic and Atmospheric Administration.

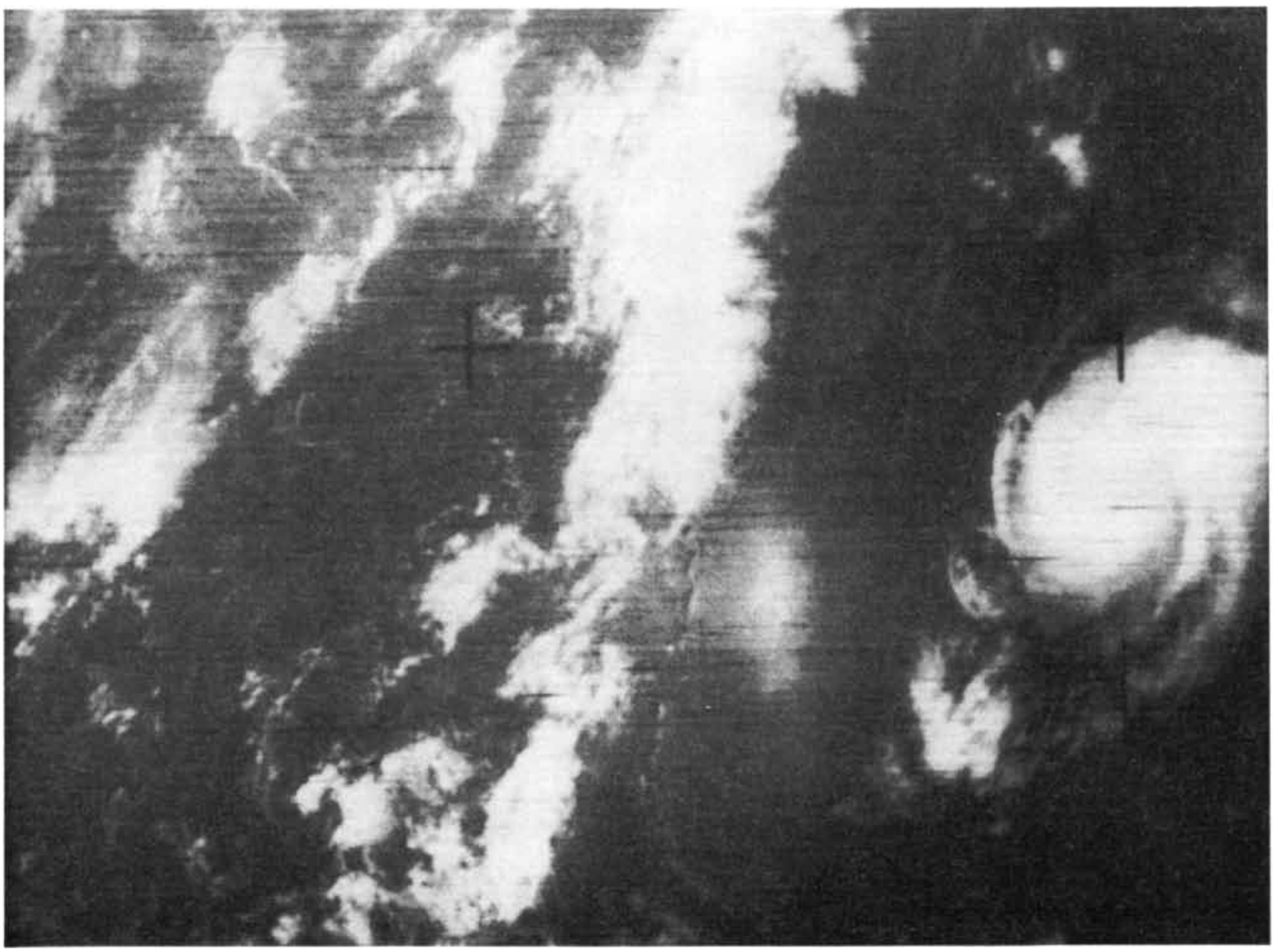
"NOAA-2 transmits continuous weather data on a frequency of 137.5 megahertz to all parts of the earth. Both ESSA-8 and NOAA-2 make a complete orbit every 115 minutes. The following discussion is based on the more technically advanced NOAA-2, which should continue to operate for some years. In its polar orbit it swings close to the North Pole, then moves south across the Equator and finally, having traversed Antarctica, returns north for the next orbit.

"The vehicle carries a reflecting telescope, the optical axis of which remains tangent to the earth's surface and pointed in the direction of the satellite's path. The objective mirror, an ellipsoid, is set at an angle of 45 degrees with respect to the optical axis of the instrument. The mirror rotates continuously at the rate of 48 revolutions per minute, plus or minus a few parts in a million. As the vehicle moves in orbit the rotating mirror scans the earth and space in the east-west direction. A cassegrainian optical system focuses radiation thus acquired through a beam splitter alternately to a bolometer and to a photodiode.

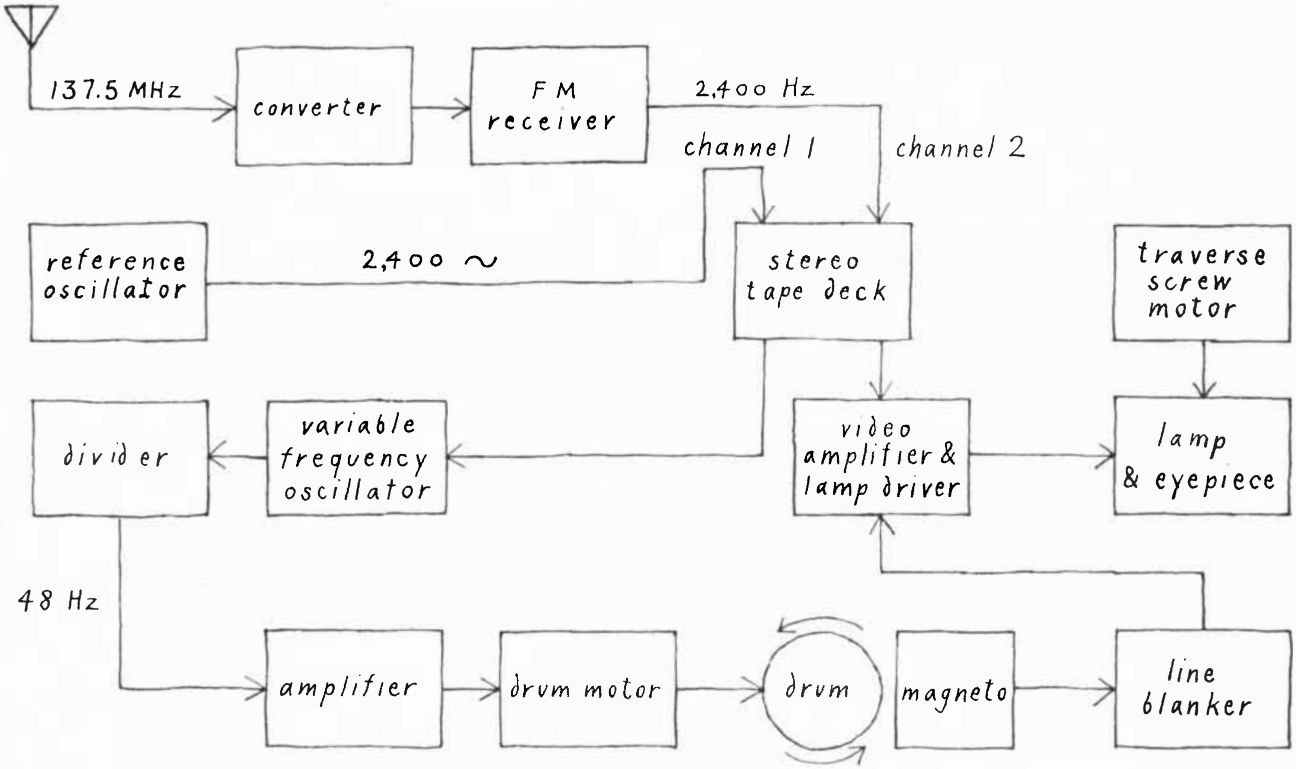
"The system functions as a two-channel scanner that examines the earth 48 times per minute in both infrared and visible light sequentially to disclose cloud cover on the unlighted and lighted hemispheres. Information collected during each scan is encoded and broadcast as a shortwave radio signal during an interval of 1.25 seconds. The first half of the interval, .625 second, begins with telemetry data and is followed by the infrared scan from 10.5 nanometers through 12.5 nanometers. The second half of the interval is occupied by telemetry data followed by the scan in visible light from .5 nanometer (green) through .7 nanometer (red).

"Each 1.25-second signal comprises a 'line' of picture data that is analogous to the line of a television picture. About 850 lines necessarily complete a weather-satellite picture, because the short radio waves follow a line-of-sight path that can reach any specific location on the earth's surface during a maximum interval of about 18 minutes. At my location, for example, signals are first picked up at about the time NOAA-2 approaches the latitude of Cape Farewell in Greenland, and they are lost as it approaches the Gulf coast of Yucatán.

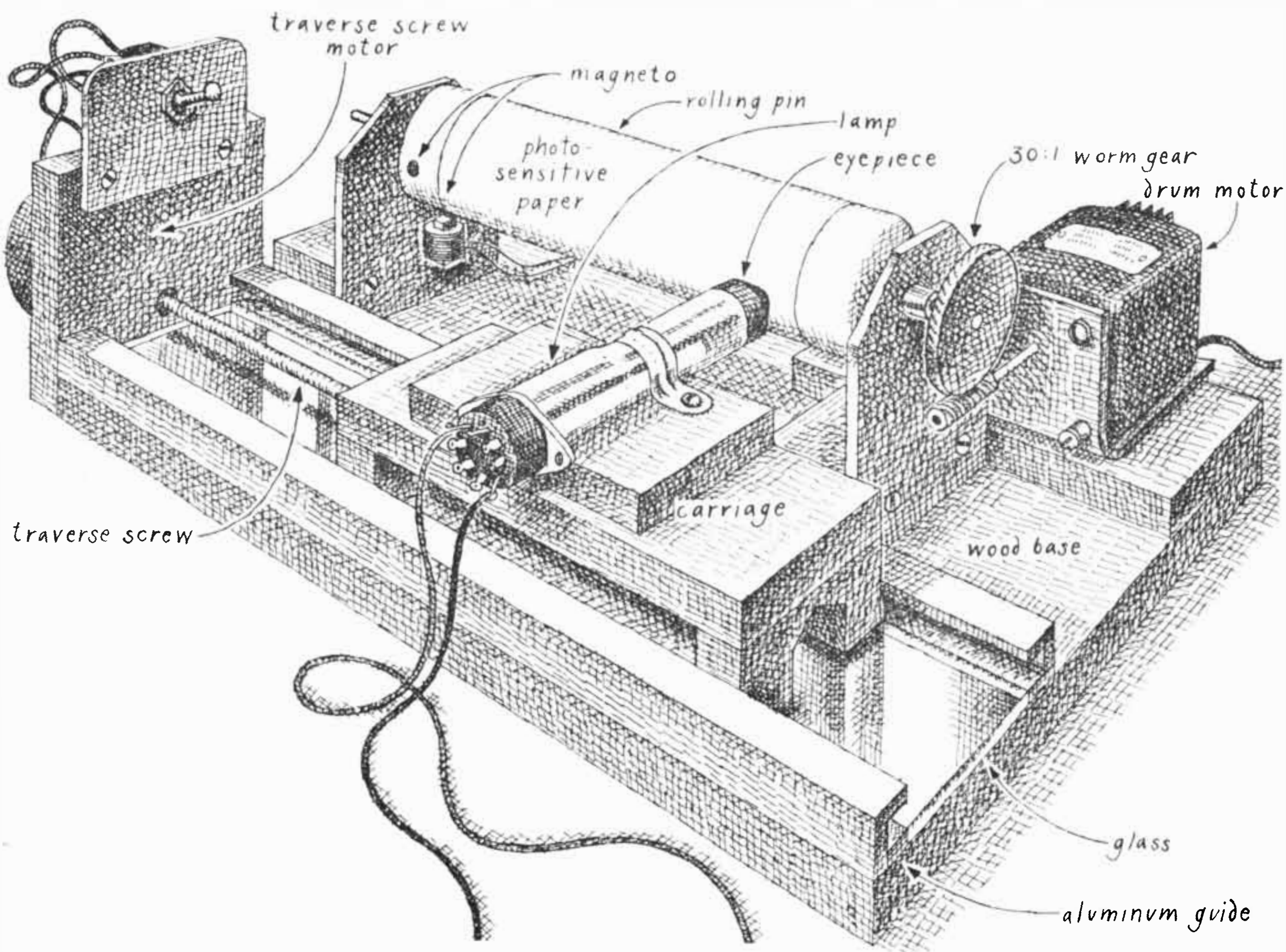
"My receiving system includes a surplus FM radio receiver and a homemade oscillator that generates a local signal of the same frequency as that of the satellite [*see top illustration on next page*]. Both signals are taped for subsequent playback and conversion into pictures by a homemade facsimile recorder. Essentially the recorder consists of a rotating drum of wood that carries a cylinder of photographic paper and a gas-discharge lamp that flickers in response to signals broadcast by the satellite. A motor-driven lead screw moves the lamp and lens system parallel to the rotating cylinder as a thin beam of light exposes the photosensitive paper line by line. The recording mechanism is housed in a lightproof cabinet when exposures are made. The radio and locally generated



Hurricane Ellen over Bermuda on August 20, 1973



Components of the weather-satellite receiving system



The facsimile recorder

signals are amplified and otherwise processed with several electronic devices.

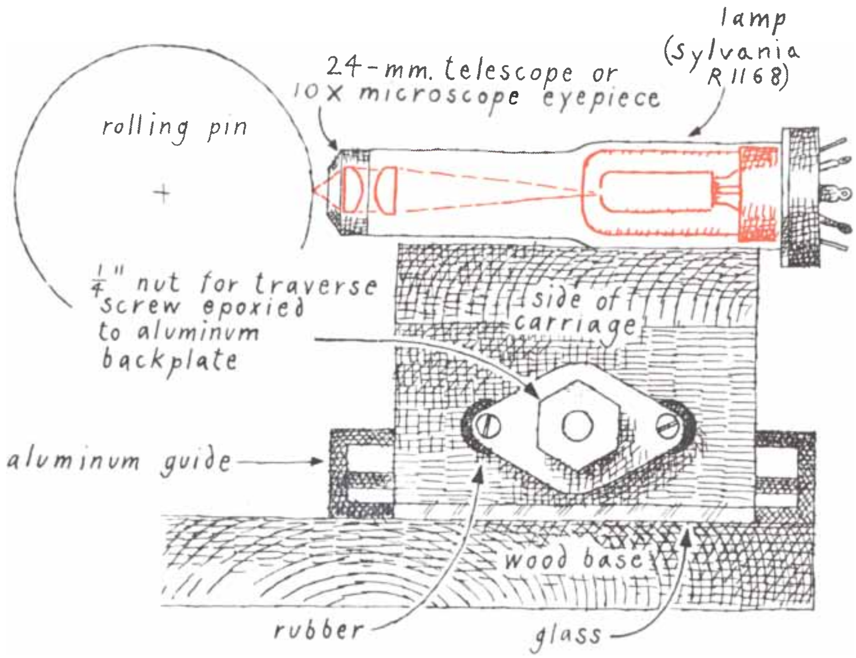
"Surplus radio receivers of the kind in my system can be bought from the Gregory Electronics Corporation (249 Route 46, Saddle Brook, N.J. 07662). My receiver was designed for operation at 49.54 megahertz. I inserted a converter between the antenna and the receiver to reduce the frequency of the incoming signals from 137.5 megahertz to 49.54 megahertz. Converters are available from Aerotron, Inc. (U.S. Highway 1, North Raleigh, N.C. 27608).

"Incoming signals can be reproduced directly in the form of photographs, but I prefer to record them on magnetic tape. This stratagem enables me to reproduce any number of original photographic copies and also to accumulate an inventory of signals for experiments that I do during periods when the satellite is out of range. In addition the tape affords a measure of protection against the loss of a picture if part of the system malfunctions.

"Signals from the satellite appear at the output of the receiver as an alternating current of varying voltage at a constant frequency of 2,400 hertz. The picture data are encoded as variations of the voltage. I refer to this current as the picture signal and feed it into one channel of the stereo tape recorder. Into the other channel I feed a similar current of 2,400 hertz at constant voltage that is generated by a crystal-controlled oscillator. I refer to this current as the reference signal.

"The picture and reference signals, together with alternating current from the 60-hertz power line, operate the facsimile recorder. The size of the recorder was determined by the dimensions of Kodak Polycontrast Rapid enlarging paper, which is available in standard eight-by-10-inch sheets. The drum that carries the paper is an 11-inch rolling pin of wood 2½ inches in diameter. I replaced the handles with an axial shaft of steel 1/4 inch in diameter and 13 inches long, which is fastened to the wood with epoxy cement. The improvised drum turns in bronze bearings of standard size fitted to brackets made of 1/8-inch sheet aluminum [*see bottom illustration on opposite page*].

"Photographic exposures are made by a point of light .006 inch in diameter. Rays from the lamp (a Sylvania Type R-1168 glow modulator tube) are focused on the rotating cylinder by a telescope eyepiece of one-inch focal length. When the eyepiece is in proper focus, it is within about a quarter of an inch of



Details of the lamp assembly

the paper. The lamp and the eyepiece are mounted in a 1¼-inch flared fitting (a short brass pipe) of the kind used to connect the trap under a kitchen sink. The optical assembly is clamped on the top of a wooden carriage that slides the length of the rotating drum on a rectangle of plate glass.

"The carriage is restrained laterally by a pair of aluminum rails and is propelled by means of a lead screw having 20 threads per inch. The screw is turned by a reversible Type CA Hurst synchronous motor. The motor operates from the 60-hertz power line. It is available from the Allied Electronics Corporation (2400 West Washington Boulevard, Chicago, Ill. 60612).

"At 12 revolutions per minute the lead screw advances the lamp carriage nine inches in 15 minutes. During this interval the satellite broadcasts about 720 lines of data. At a drum radius of 1¼ inches each of the 720 parallel lines of satellite data is 7.8 inches long. Hence the final image would be approximately square if all telemetry signals and both the infrared and the visible data were reproduced. I rarely reproduce all four images on a single sheet of paper, however.

"A small Alnico magnet is recessed in one end of the rolling pin and fastened by a dab of epoxy cement. Once during each revolution the magnet grazes the end of a solenoid that is rigidly fixed to the base. The solenoid consists of a laminated bar of transformer iron that carries about 3,000 turns of fine magnet wire. The magnet and the solenoid function as a magneto that generates a pulse of voltage at exactly the same point during each revolution of the drum. This point marks the edges of the picture, because the edges of the paper are aligned with the magnet at the time the sheet is wrapped around the drum and fastened to it with doubly coated adhesive tape.

"The drum is rotated through a worm reduction gear by a Bodine Type KYC reversible synchronous motor. The motor is available from the Minarik Electric Company (224 East Third Street, Los Angeles, Calif. 90013). Both the drum motor and the carriage motor are fitted with reversing switches. After a picture has been made I reverse both motors. The next picture can then be made without returning the carriage to its starting point.

"The speed of the drum motor is determined by the reference oscillator's frequency, which is fixed by a piezoelectric crystal of quartz. The crystal oscillator previously described in this department is suitable [see "The Amateur Scientist," SCIENTIFIC AMERICAN, May, 1973]. Incidentally, the pin numbers of the oscillator circuit that were specified in that article apply only to integrated circuits of the flat-pack type.

"The same article explains how to design digital circuits for dividing frequencies by any amount. Circuits of this kind are required by the satellite receiver. For example, I divide the 2.4-megahertz output of my crystal by 1,000 with a string of Type 7490 decade counters. This output can be momentarily lowered to 2,355 hertz with the device known as a phase-locked oscillator. Another string of integrated-circuit devices subsequently divides this frequency to 48 hertz. After appropriate amplification alternating current at this frequency energizes the drum motor. [Schematic diagrams of Ruperto's reference oscillator, phase-locked oscillator and other required devices for operating the drum motor will be forwarded without charge on receipt by this department of a self-addressed, stamped envelope.]

"As I have mentioned, the satellite transmits infrared data and visual data sequentially 48 times per minute. If the recording drum ran at this speed, both images would appear as slender strips divided by two patterns of telemetry signals. Each infrared pattern or picture of cloud cover can be made to occupy the full width of the photographic paper by running the drum at double speed: 96 revolutions per minute. Each full revolution then accommodates half of the data of each line, either infrared or visual. A photograph so made would be built up of alternate lines of both infrared and visual data.

"Occasionally I run my recorder at this speed and unscramble the images with the electronic device known as a line blanker. This device is essentially an electronic switch that automatically turns off the lamp when alternate lines of data appear. The line blanker is actuated by pulses from the magneto. A selector switch enables me to reproduce either the infrared image or a picture of cloud cover.

"The pulse of the magneto also enables the experimenter to adjust the phases of the synchronized drum and the satellite mirror so that lines of picture data begin and end at the edges of the photographic paper. I make this adjustment with a cathode ray oscilloscope. Signals from the satellite are fed to the vertical plates of the oscilloscope. The horizontal sweep of the oscilloscope is initiated by magneto pulses from the recording drum.

"If the drum and the satellite mirror are rotating in the desired lockstep, the oscilloscope displays at its left edge a somewhat ragged square wave of substantial amplitude followed by an undulating pattern of lesser amplitude. The undulating pattern represents the picture data. Unless the phase relation between the drum and the satellite mirror has been adjusted, however, the square wave usually appears at an intermediate point on the face of the oscilloscope, flanked by undulating patterns.

"When this configuration appears, I operate the push button of the phase-locked oscillator. This has the effect of lowering the speed of the drum motor. The square wave drifts toward the left of the oscilloscope display. When it reaches the extreme left edge, I release the push button. Thereafter the drum and the satellite mirror run in lockstep, and the lines of the picture begin and end at the edges of the paper. One can also 'phase' the drum and the mirror by interrupting the circuit of the drum motor momentarily, thus dispensing with the luxury of the phase-locked oscillator. Phasing in this way is a matter of hit or miss—mostly miss.

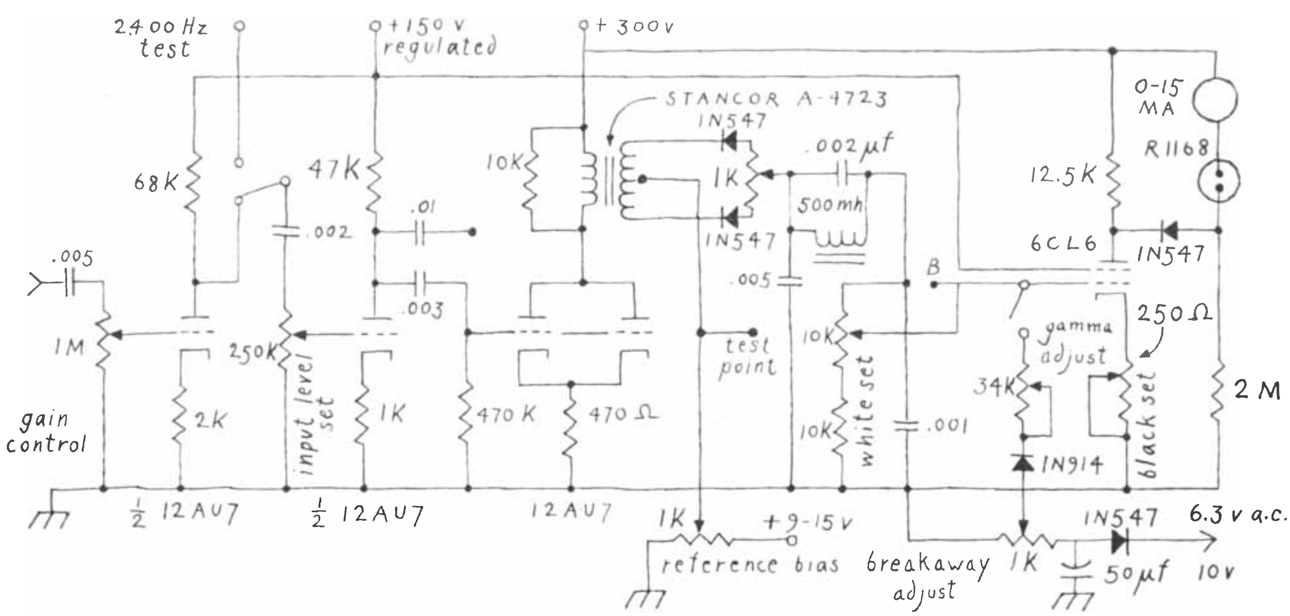
"A problem I encountered when I first tried to use the oscilloscope for making the phase adjustment was that the horizontal sweep of the instrument refused to work reliably at a frequency of less than two hertz. I solved the problem by again doubling the speed of the facsimile drum. At 192 revolutions per minute the drum magneto generates 3.2 pulses per second, a frequency to which the horizontal sweep responds nicely. This second speedup was accomplished by recording all signals at a tape speed of 3¾ inches per second and reproducing them at seven inches per second.

"Of course, the speed of the lead screw also had to be doubled. I accomplished this by changing the ratio of the gears that couple the motor to the lead screw. The final increase in speed had the incidental advantage of cutting in half the time required to make a picture.

"Before the system is turned on for the first time the contrast controls of the video amplifier should be adjusted. One control alters the density of the white areas of the image and the other one alters the dark areas. To make these adjustments, first record on magnetic tape a 2,400-hertz signal of constant amplitude. The signal can be taken from the output of the reference oscillator.

"Feed the reproduced signal into the input terminals of the video amplifier. Connect a vacuum-tube voltmeter or an equivalent meter that will measure peak-to-peak potential between test point *A* and the chassis. Adjust the amplitude of the input signal to the point at which 10 volts peak to peak appears at point *A*. Then adjust the white control to the point at which the meter in the lamp circuit indicates 1.5 milliamperes.

"Lower the input signal until the po-



Circuitry of the video amplifier

tential between point *A* and the chassis falls to two volts. Adjust the black control to the point at which the meter in the lamp circuit indicates 14 milliamperes. The contrast of the developed photographs depends on a number of factors including the characteristics of the photosensitive emulsion. For this reason the white and black controls may require a bit of fine adjustment by the time-honored technique of 'cut and try.'

"Perfectionists can improve the picture quality even more by fiddling with the gamma adjustment and the breakaway adjustment. Set each adjustment at the midpoint of its excursion. Thereafter alter each setting slightly in one direction or the other just before making a picture. Keep a record of the settings and the corresponding results. Eventually a combination will be found that produces the most pleasing pictures. I should point out, however, that the influence of these controls is small.

"How does one learn the time at which signals from weather satellites will appear and the direction from which they will come? The schedule is routinely announced by W1AW, the official station of the American Radio Relay League at Newington, Conn. This station broadcasts daily except holidays on all shortwave radio bands that have been assigned to amateurs from 1.8 to 145 megahertz. On some bands the broadcasts are picked up worldwide. The schedule includes a variety of bulletins of interest to amateurs. The schedule of W1AW is published monthly by the league's official journal, *QST*. The journal is available to members of the league and also is distributed by some newsstands and by dealers in amateur radio supplies.

"The bulletins give the exact time at which NOAA-2 will cross the Equator in that portion of its orbit where it intercepts the plane of the ecliptic in moving from south to north: the 'ascending node.' The location where it crosses the Equator is expressed in degrees of longitude west of Greenwich in England. The time of the crossing is stated in Greenwich Mean Time, which is reckoned on a 24-hour day beginning at zero hours (midnight). To convert Greenwich Mean Time into local U.S. time subtract from Greenwich Mean Time the number of hours that your standard time zone is west of Greenwich. The zone of Eastern Standard Time, where I live, is five hours west of and earlier than Greenwich. Central, Mountain and Pacific standard time zones are successively one hour earlier.

"To convert Greenwich 24-hour time

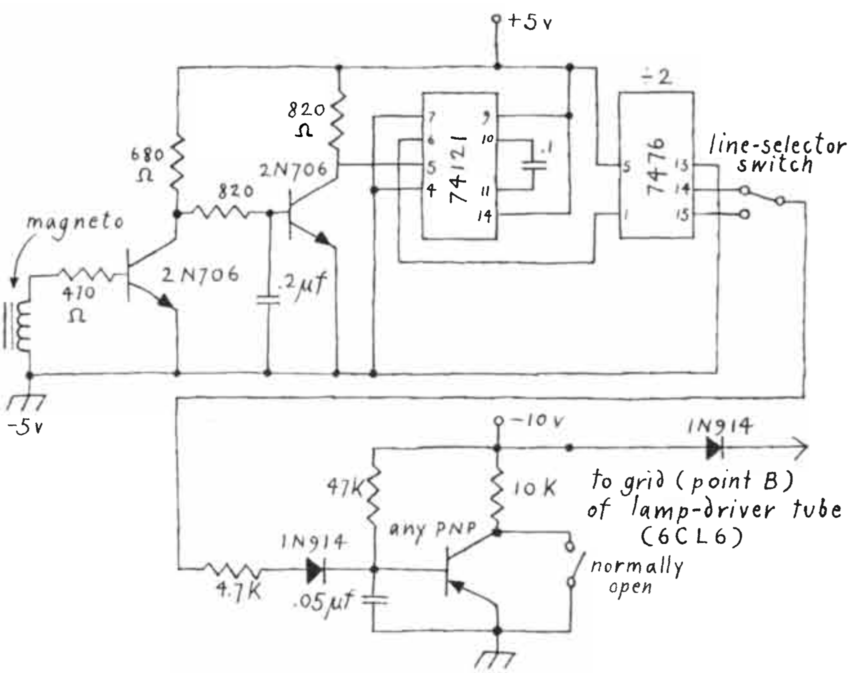
into 12-hour time, subtract 12 hours from Greenwich time if it exceeds 12 hours and label the difference 'P.M.' Greenwich time that does not exceed 12 hours is labeled 'A.M.' For example, assume that NOAA-2 is scheduled to cross the Equator at 20:34 GMT at 92 degrees west longitude. When might I expect to pick up its signals here in Pennsylvania? The crossing would come at 20:34 – 5:00, or 15:34 hours EST, which, as expressed by clocks, is 15:34 – 12:00, or 3:34 P.M.

"As I have mentioned, shortwave radio signals from NOAA-2, which orbits at an altitude of about 800 miles, reach my location when the vehicle is roughly between the latitudes of Greenland and the Gulf coast of Yucatán. The Gulf coast of Yucatán is at 20 degrees north latitude and 92 degrees west longitude. The satellite completes a 360-degree orbit in about 115 minutes, or at a rate of 3.1 degrees per minute. Therefore when the satellite is headed north, it will make the trip from the Equator to the Gulf coast of Yucatán in $20/3.1 = 6.5$ minutes. Assuming that NOAA-2 crossed the Equator at 15:34 GMT, as predicted by W1AW, I can expect to pick up its signals six minutes and 30 seconds later—at about 3:40:30 P.M. Eastern Standard Time.

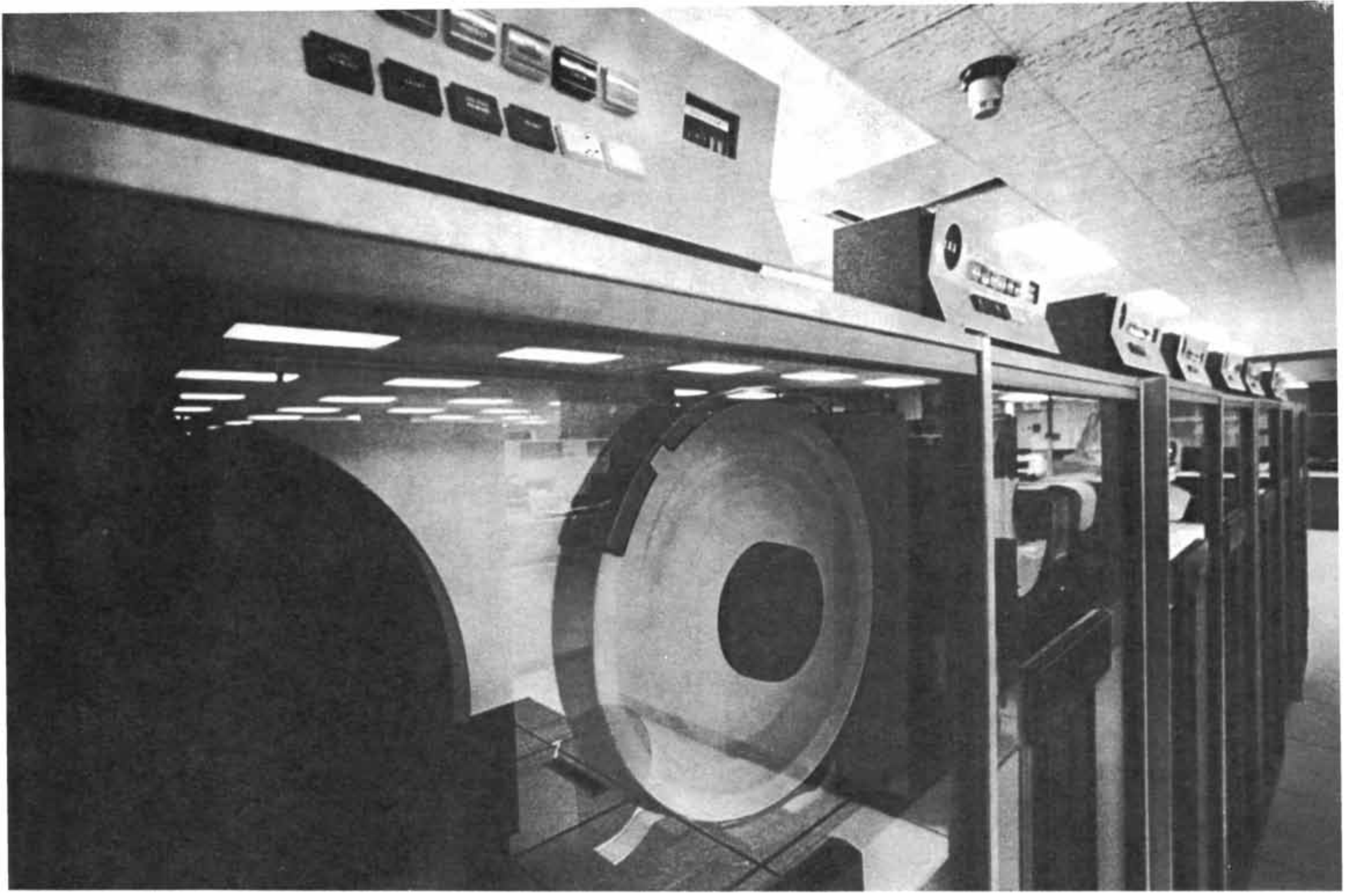
"The signals can also be picked up when the vehicle comes within line of sight of my location on its descending node, when it is traveling south. Thirty minutes or more must be added to the predicted times, depending on the position of the orbit and the node.

"After two passes have been recorded the amateur can easily predict following orbits for a few days. For example, if a pass comes at 13:25 GMT on Wednesday, you can expect to pick up the signal of NOAA-2 one hour later on the following day. By then the satellite will have drifted somewhat west of its Wednesday crossing. On Friday, two days later, it will have drifted still farther west—indeed so much farther that the signal will reach you on the succeeding orbit at 13:20, and on Saturday at 14:20. The performance is not quite clocklike, however. Simple prediction techniques do not work for more than a few days at a time. That is why I depend on the schedule broadcast by W1AW.

"The electronic devices in the weather-satellite receiver are relatively simple compared, say, with a color television set. They involve no critical adjustments. Even so, beginners will find that neighboring radio hams and television servicemen can pass along useful construction tips. Most of the circuits of my receiver were designed by Virgil Neher, a radio ham of La Verne, Calif. I am grateful to him and also to Wendell Anderson, whose article 'Amateur Reception of Weather Satellite Picture Transmission' in the November 1965 issue of *QST* first kindled my interest in this fascinating avocation."



Circuitry of the line blanker



If you make computers--or anything else that uses electrical connections--call AMP.

We can help you put a smile on your bottom line.

One way to make your bottom profit line look better is to increase the speed and efficiency of your assembly operations.

That's where we come in.

We make electrical and electronic connectors. Everything from simple, one-piece contacts to sophisticated interconnection systems. Plus a full line of automated machines to help you apply them easily. And economically.

Case in point: for a computer manufacturer, we developed a truly unique plating process (another example of AMP innovation) by impregnating the gold plated contacts with lubricating graphite. Result? We are now able to double the performance of our product and reduce the gold requirements.

So if you make computers—or anything else that uses electrical connections—give us a call at (717) 564-0101. Or write for our new brochure: "We can help you put a smile on your bottom line."

AMP Incorporated, Harrisburg, Pa. 17105.

Facts you should know about AMP.

- Over $400 million annually in sales.
- Over 10% spent on Research, Development and Engineering.
- Over 1,000 field service and sales engineers.
- Some 100 manufacturing plants, warehousing facilities and sales offices in USA and worldwide: Barcelona, Brussels, Buenos Aires, Frankfurt, London, Lucerne, Mexico City, Paris, Puerto Rico, Sao Paulo, s'Hertogenbosch (Holland), Sydney, Stockholm, Tokyo, Toronto, Turin, Vienna.

AMP INCORPORATED

Scientific best sellers...science classics... important scientific reference work- all at substantial discounts!

(Retail prices shown)

70240. PRE-COLUMBIAN CITIES. *Jorge E. Hardoy.* A distinguished authority on ancient and modern city planning synthesizes a wealth of archaeological detail in a remarkable study of Western hemisphere prehistoric civilizations. Implications for the contemporary urban crisis in America. *Counts as 2 of your 3 selections.* **$19.95**

88020. X-RAYING THE PHARAOHS. *James E. Harris and Kent R. Weeks.* Not since Howard Carter opened the tomb of Tutankhamen, a half-century ago, has science made archaeological discoveries to equal those revealed for the first time in this handsome new volume. Illustrated in color and black and white. **$10.00**

36280. A BIOGRAPHICAL DICTIONARY OF SCIENTISTS. *Edited by Trevor I. Williams.* Over 1000 biographies of scientists in every field, from agriculture to zoology. Describes the lives of these men but assesses their contributions and relates them to those of their predecessors, contemporaries, and successors. A unique and splendid reference. **$9.95**

36400. BLAKISTON'S GOULD MEDICAL DICTIONARY. Third Edition. Over 75,000 definitions in this newly revised third edition of the most famous and modern compendium of the terms used in all branches of medicine and related sciences. 35 contributing American specialists make it an invaluable reference tool for expert and general reader alike. **$18.50**

42261. DIGITAL COMPUTING AND NUMERICAL METHODS. *Brice Carnahan and James O. Wilkes.* A brilliant presentation of numerical methods with practical application to their implementation on digital computers. A "must" for every mathematician and engineer. **$13.95**

86070. WANDERING LANDS AND ANIMALS. *Edwin H. Colbert.* A boldly original epic history of our planet's drifting continents and the spread of animals through the ages. Replete with charts, tables, maps, illustrations and impressive glossary. Recounts the finding of *Lystrosaurus* in Antarctica, now regarded as keystone evidence to explain continental drift. **$12.50**

45471. ENVIRONMENTAL GEOSCIENCE. *Arthur N. Strahler and Alan H. Strahler.* The dean of American geomorphologists ably assisted by his son (a plant geographer at the University of Virginia), masterfully explores the physical and biological processes involved in interactions between Man and the environment of energy and matter which surrounds him. **$12.95**

46590. EVOLUTIONARY OPERATION. *George E. P. Box and Norman R. Draper.* Business, industry and technology are benefiting from the methods of Evolutionary Operation, a process improvement technique that utilizes the concepts of natural select on and the tools of mathematical statistics to maximize industrial efficiency. **$12.50**

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, or ganic or man-made—proposing a continuum ranging from nuclear particles at one end of the scale to the cosmos at the other. **$14.50**

62801. MODERN COSMOLOGY. *D. W. Sciama.* "This book is a masterpiece. Stylistically there is not a word wasted and not one needs to be added. 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. It is almost unbelievably up to date."—*Physics Today* **$9.95**

The LIBRARY OF SCIENCE

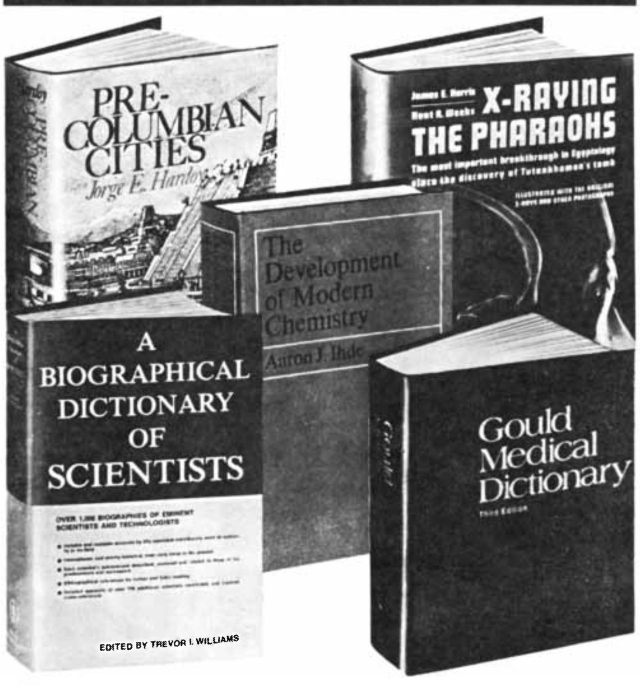
invites you to

Take any 3 books

(values as high as $56.40)

all for only $3.95

(You simply agree to buy 3 more books—all at handsome discount—over the next 12 months.)



The Library of Science
Riverside, New Jersey 08075 2-31A

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.

3 books for $3.95 **Indicate by number the 3 books you want.**

Some expensive books count as 2 selections.
Send no money. Members are billed when books arrive.

Name__________

Address__________

City______ State______ Zip______

Book selections purchased for professional purposes may be a tax-deductible expense.

38310. CLASSICAL ELECTRODYNAMICS. *John David Jackson.* Electromagnetic theory on the graduate level handled both conventionally and through consideration of the properties and interactions of charged particles. Ideal reference work for the physicist. **$17.95**

61360. MATHEMATICAL STATISTICS WITH APPLICATIONS. *William Mendenhall and Richard L. Scheaffer.* A uniquely practical work which stresses inference as the objective of statistics and emphasizes the role of probability in inference making. Covers statistical theory, hypothesis testing, linear models, more. **$12.95**

42070. THE DEVELOPMENT OF MODERN CHEMISTRY. *Aaron J. Ihde.* Revised, updated and reprinted, an 850-page history of chemistry filling the niche between short, popularized works and exhaustive, pedantic treatments. "Scholarly and well written."—*Scientific American* **$14.95**

63131. MOLECULES AND LIFE. *Joseph S. Fruton.* The interplay of chemical and biological ideas is presented with full documentation for the period from about 1800 to 1950, with some reference to both earlier and later developments. **$19.95**

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

78190. SEXUAL SELECTION AND THE DESCENT OF MAN. *Edited by Bernard Campbell.* Eleven experts in the fields of biology, genetics and anthropology examine the evidence for Darwin's theory of sexual selection, the strongest factor in human evolution. **$14.75**

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

38011. CITIES AND GEOLOGY. *Robert F. Legget.* The first geological work to take up the special problems of urban areas. A must for scientists, engineers, planners, and architects. **$15.50**

87210. WILDLIFE IN DANGER. *James Fisher, Noel Simon, and Jack Vincent. Preface by Joseph Wood Krutch.* Beautifully illustrated reference to all the species of animals and plants in danger of imminent extinction. "It will boil the blood of any sensible human being."—Peter Farb, author of *Man's Rise to Civilization.* **$12.95**

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.)

BOOKS

On Great Zimbabwe, the massive stone ruin that recalls a little-known African culture

by Philip Morrison

Great Zimbabwe, by P. S. Garlake. Illustrated. Stein and Day, Publishers ($20). A couple of hundred miles inland in southern Africa lies the watershed plateau between the Limpopo and the Zambesi rivers, "cool, well-watered, gently rolling," free from the tsetse fly, its fertile open woodlands dotted with spreading acacias. The bones of the land are granite hills, pocketed here and there with the metamorphic rocks that yield gold ores to the miner today as they did millenniums ago. Near one edge of the plateau, where streams gather to flow seaward across the low, fly-ridden grassy plains, lies Great Zimbabwe.

The ruin occupies a dual site. The Hill Ruin displays one set of high walls. On the edge of a cliff half a mile away across the valley is the Great Enclosure, a masonry wall surrounding an ellipse some 800 by 300 feet, "by far the largest single prehistoric structure in sub-Saharan Africa." More than half of all the stonework at Great Zimbabwe is in this wall, which attains a height of 32 feet and a thickness of 17 feet. Within the Outer Wall and around the site are scattered a complex of lesser walls and platforms and one striking, completely solid conical tower. All the stonework is dry-wall masonry, well-laid or roughly laid courses of local granite blocks. The rock exfoliates everywhere on the smooth-domed granite hills that dominate the site, and thin slabs peel off and slide down slopes to collect as scree. There they can be cracked and taken away. Natural fracture planes give all the broken pieces a quite regular shape: cuboids, with even a standard thickness, that lend themselves to construction "based on more or less regular horizontal layers of stone." So was Great Zimbabwe built (the word probably means "venerated houses" in Shona), and so also were built 150 or 200 smaller and cruder ruins dotted over the high granite country. No sockets, beams or posts are found; apparently the stone walls never bore roofing. The walls are all curved, and they meander smoothly over the plan. The entire courtyard was once paved with the common indigenous building material called *daga:* a puddled, clayey soil binding gravel. Pits outside the walls show where the clay came from. *Daga* plastered the walls up to seven feet, and all the dwellings within were round huts of *daga* joined by abutting walls of stone. The walls are freestanding now and often one end of a wall is a finished doorway and the other end a jumble of collapsed stone, senseless once its abutting hut was washed away. The walls do resemble brickwork, but the resemblance is wholly superficial. There is no evidence of familiarity with brick bonding; these walls are never bonded, whether to break vertical joints, to link the two faces of rubble-filled sections or to join walls at a corner. Indeed, there are no angular corners.

The spade has not found very much more. Huge middens are present on every slope. Spindle whorls, shards, iron hoes and similar domestic artifacts of the Iron Age are abundant. Metalworking in copper, bronze, gold and iron was practiced. Radiocarbon dates are few; rock is not to be dated. The oldest timbers found are durable poles that have been dated to the seventh century. Garlake holds that these samples, cut from trees that live for 500 years, provide only the upper limit to the age of the structures: they might well have been taken from earlier structures on the site and reused many times. Like all sites, Great Zimbabwe evolved; the few radiocarbon dates relevant to the footings of the great stone walls date from the 11th to the 15th centuries. A unique trade-goods cache, found buried in soil in an enclosure a few hundred feet from the Great Wall, can be dated more securely than anything else at Great Zimbabwe because it held Chinese celadon stoneware and an inscribed Persian bowl. The bowl was made in the 13th or 14th century. This hoard makes it plain that Great Zimbabwe then had far-flung trading contacts, the intermediaries being the Swahili coastal cities that flourished in those times. There is not much else, save for the remarkable and famous figurative carvings of soapstone (mostly removed), to add to the ordinary domestic goods. Those walls never enclosed a busy trading emporium; the very richness of the single hoard is evidence that a true center of trade would yield more. The celadon shards from China never give way to porcelain at this site in spite of the fact that in every trading post of the Portuguese, and in a number of other later ruins, the blue-and-white porcelain of the Ming is prominent. Great Zimbabwe was somehow preeminent when celadon came to the African coast but was in decline by the time the factories of the Ming sent porcelain, which was certainly by the late 15th century.

The enclosures were not mainly military, although they are protective. They have no slits or embrasures, even though the weapons of war in this land were arrow and spear. Walls do not defend other walls; nothing of the planned fortress is present. These structures were built and inhabited as a palace between A.D. 1000 and 1500 by a native people. They were built for "size and imposing grandeur, the product of two or three centuries of development of an indigenous stone-building technique, itself rooted in long traditions of using stone for field walls, building platforms and terraces." A small oligarchy—a royal and ecclesiastical court with sway over many vassals, if you will—grew out of an Iron Age subsistence economy, its growth based on political control of a vast region of gold and copper mining. The leaders dwelled there with a couple of thousand craftsmen, farmers and laborers to support the court. (Building the Outer Wall is estimated to have required 400 laborers working 50 days a year for four years.) The oligarchy's rule spread north and west over the plateau until, after some local decline in resources, the very strength of the wider society led the court to depart from Great Zimbabwe; the rulers moved north a couple of hundred miles into Matabeleland. Their new lands had no granite, so that all building in granite ceased. Tradition says that the old capital had "a severe shortage of salt"; we can read this as literally or as figuratively as we wish. Great Zimbabwe still stood, stripped of wealth and power; it remained for a while a provincial residence perhaps, and finally it became a religious monument of the old times.

When Carl Mauch, a young geologist, reached Great Zimbabwe in 1871, it was deserted and overgrown; the newly arrived people living there were led by an unimportant Karanga chief whose predecessor had taken the area away from the local people and turned the buildings into corrals. So was finally broken a tradition of sanctity that can probably be traced back 800 or 900 years. Mauch met a man of the ejected people whose father, a priest of Great Zimbabwe, had come every few years to sacrifice at the ruins. The last known sacrifice was in 1904. The structure remains, "a proud part of the Karanga and Shona past."

All of this describes only in part an enriching book by a Rhodesian trained in architecture and archaeology. Much of the careful, clear and handsome volume (there are many splendid photographs in color and in black and white), with a brief introduction by Sir Mortimer Wheeler, deals perforce with archaeology as a political battlefield. The early Portuguese never visited the site, but they spoke with Swahili traders who apparently had seen it. It was then within the Karanga kingdom of the Mwene Mutapa, ruling from the north, who held fief over the gold and the trading routes that had brought the Portuguese, as they had the Swahili before them. Since the Karanga had no buildings of stone, the Portuguese ascribed the storied places to Prester John, to Solomon or to the Queen of Sheba. These are all folk figures in Islam. The "aged Moors" cited in 1609 by one Portuguese missionary suggested such an origin: "the factory of Solomon." When Cecil Rhodes and his colonizing British South Africa Company came, they seized with delight on such theories; to admit the local origin of Great Zimbabwe would fault their claims to a land they insisted had been settled no earlier by the Bantu than by the Europeans. In 1891, within a year of occupation by the company, it engaged the antiquarian J. Theodore Bent, a man who had studied the origin of the Phoenicians but who was without formal archaeological training or experience, to study the site. It was his vivid and uncontrolled imagination that first related Great Zimbabwe to a Sabaean temple in Marib in Yemen. This structure dates from the seventh century B.C.; it has an elliptical dry-wall enclosure of about the same size and shape as that of Great Zimbabwe. "However, all comparisons end there." Its blocks are limestone, carefully cut and pecked, the faces linked with cross walls, the entire structure joined to a rectangular building that was pillared and roofed, inscribed and linear in decoration: "products of a tradition and technology quite unlike those of Great Zimbabwe." Bent and his successors stripped the ruins "in an obsessive search for the exotic." In 1895 the company granted a firm the right to dig for gold in all ruins, Great Zimbabwe excepted ("at Mr. Rhodes' express desire"). After a few years public reaction to this greedy and foolish enterprise brought it to a halt and a journalist, Richard N. Hall, wrote a book from the gold seekers' data that sought to build a coherent framework. Hall was made curator of Great Zimbabwe but himself stripped the ruins there lamentably, removing, as he said, "the filth and decadence of the Kaffir occupation." Outsiders complained again: "field work... worse than anything I have ever seen," wrote David Randall-MacIver after Hall was dismissed and MacIver was invited to investigate. He was a student and colleague of Flinders Petrie's and his skilled and honest, although hasty, work established that Great Zimbabwe was without alien influence. The opinion of the settlers has ever since been polarized against overseas experts. Gertrude Caton-Thompson followed MacIver around 1930, a generation later. She sought incontrovertible dating and carried out a model study. (Her team, by the way, was composed of women archaeologists; it is possible to see there the first excavation, made in her student years, by Kathleen Kenyon, who was later the famous excavator of ancient Jericho.) She too found almost no exotic objects, no easy dates. The Rhodesians Keith R. Robinson and R. F. H. Summers worked at the site in the late 1950's and early 1960's, and it was they who secured the radiocarbon dates (still rather vague) that are the foundations of the modern interpretation.

Garlake became Senior Inspector of Monuments in Rhodesia in 1964; he studied the digs, particularly the lesser sites, until 1970. Now he is at the University of Ife in Nigeria, and in secessionist Salisbury today "the Rhodesian museums no longer employ an archaeologist." Instead the official censor is at work. In the new guidebook "all theories relating to Zimbabwe will be presented absolutely impartially" and there will be no more search for unsettling evidence. It is distressing that in October, 1973, a national weekly in the U.S. published a piece lending strong endorsement to the most recent ascription of Sabaean origins to the walls at Great Zimbabwe. Yet truth is more enduring than granite; it seems to beckon the reader from even a single aerial photograph of the ruins, where, as Garlake writes, stone wall and natural boulder are so interdependent on the ground that "each seems almost a natural outgrowth and extension of the other." That organic way of building with cracked granite scree surely evolved among a people long accustomed to and wholly familiar with the gray granite of their own hills, not among alien brickmakers from a far continent. One may hope that before too long archaeologists and freedom alike will return to the land of Zimbabwe.

Boron, by A. G. Massey and J. Kane. Mills & Boon Limited (2.25 pounds). Lamp Phosphors, by H. L. Burrus. Mills & Boon Limited (1.50 pounds). Fibre Reinforcement, by J. A. Catherall. Mills & Boon Limited (1.50 pounds). The colorful little hard-cover booklets, each 60 to 80 pages long and illustrated with graphs, tables and photographs (the boron monograph even has color photographs: boron-deficient grapes and a greenish boron explosion), are members of a larger family. By now several dozen of these easily read, up-to-date and well-informed British studies have appeared, each treating a single new development in chemical, electrical or mechanical engineering, constituting a first nonmathematical survey intended for readers who have a technical background but may be new to the special subject. Of the entire flow this review considers a sample of three, one for each main branch of engineering.

The slightly dearer boron booklet treats the element and its compounds. You can see both the old 20-mule team and the newer open-pit mine, whose major mineral is kernite, a less-hydrated analogue of borax. That one pit furnishes the bulk of the world's boron. There follow the elemental uses: nuclear (the isotope of mass 10 eats slow neutrons), plant nutrition, strong lightweight fibers and the like. Most of the text treats the boranes, which are pyrophoric hydrides of boron. It was their preparation and study by Alfred Stock even before World War I that stimulated the widespread handling of chemical preparations in complex closed vacuum systems, with all the tricks now familiar to the modern chemist. No borane compound unknown to Stock was found until 1958! These compounds were the seat of much hope and effort as high-energy rocket fuels. Decaborane is about half as energy-rich as liquid hydrogen and half again better than kerosene; its organic derivatives are fine fuels, but the cost is $2,000 per gallon. Compounds based on icosahedral "cages" of boron atoms are also discussed. Borax is cheap and yet boron is dear, because only about a tenth of the weight of borax is in boron atoms. The glass industry remains the chief user of boron, mainly as dehydrated borax in commonplace bottle glass, where up to 2 percent of boric oxide pays its way in reducing flow temperatures and thus fuel costs. Heat-resistant chemical and ovenware borosilicate glass contains 10 or 15 times as much boron but is a specialized and relatively low-volume product compared with beer bottles. Sodium perborate is the cheapest and safest of all peroxy salts and is much used in detergents, particularly for very-hot-water washing, which is common in Europe and the U.K. Boron nitride, in the layered form of "white graphite" or as borazon, its cubic high-pressure form (akin to diamond), mimics the familiar allotropes of carbon. It is a fascinating stuff, and boron carbide itself, the next hardest substance after diamond and borazon, found large use as light armor for U.S. helicopters and troops.

Lamp phosphors now provide most of the world's light at night. The low-pressure mercury discharge in the usual fluorescent tube produces two main lines, both in the ultraviolet. The phosphor coating converts these to visible—even pleasantly colored—light. Long life, low cost, efficiency and control of color have been the goals in phosphor development, which has by now reached a sophisticated understanding of the quantum physics of phosphors. The most used substances are fluorapatites, the very stuff of bone, the microcrystalline particles of which are prepared with extreme care and are adjustable as to color with small amounts of activating antimony and manganese. A good lamp today has an overall energy-conversion rate of about 25 percent, with good color. "De luxe lamps" and color-television tubes are concerned more with color precision than with efficiency; red is what has always been lacking from the glowing surface. The remarkable europium-activated yttrium vanadate emits linelike narrow peaks in the red, a highly saturated color; the transitions take place in an inner electron shell, shielded from the fields of the nearby crystal ions.

Even the "anti-Stokes phenomenon" is cunningly used. A substance may absorb two photons step by step, to jump up two states. Its emission can then—anomalously—be at a higher frequency than either of the two photons it absorbed. Just this trick makes the now familiar light-emitting gallium arsenide diodes of the pocket calculators, which are intrinsically infrared emitters, glow red, green or blue—although with efficiencies as low as one part in 10,000.

The third monograph deals with composite fibrous materials: strong fibers dispersed in a yielding matrix. Wood consists of hollow cellulose fibers dispersed in a resin, lignin. The strongest of all man-made materials are the stiff, single-crystal filaments known as whiskers, particularly those of silicon carbide, graphite and other light covalently bonded compounds and elements. Glass, ceramic and metal filaments can also be used. An analysis of the strength of such composites is given under the simplifying assumption that bonding to the matrix is good. Fibers produced by a complex graphitization of a synthetic textile fiber have a stiffness twice that of piano wire. The notorious compressor-fan blades of the Rolls-Royce RB211 engine, made for the Lockheed Tristar, were the first major use of carbon-fiber-reinforced epoxy. The light, stiff blade saved a great deal of weight, since the design of the entire engine reflects the stresses on its outermost row of blades. Impact toughness was lacking, however; water drops eroded the blade surfaces, and the engine could not swallow without damage a bird large enough to meet the safety standards. Rolls Aeroengines fell to the receivers and Lockheed tottered. Here are photographs of a golf club and a squash racket made of this same light, strong material; they work admirably. The future will see slow but steady progress, the authors predict. "Spectacular advances can result in spectacular retreats."

These books and their kin are a valuable resource for the engineering library. Any reader with technical interests can profit from the ones close to his concern.

THE PSYCHOLOGY OF ANOMALOUS EXPERIENCE: A COGNITIVE APPROACH, by Graham Reed. Hutchinson & Co. Ltd. (2.50 pounds). Unusually readable, this small volume by a Canadian professor of psychology orders and illuminates a diverse and profound collection of psychological problems. That is not to say the problems are solved; indeed, the whole force of the book leads to a realization that we are at the beginning of insight into questions as subtle as those raised by absentmindedness, hallucination, the "tip of the tongue" experience, *déjà vu* and delusions. The reader will find no graphs, no equations, no biochemistry, hardly a number. He will find no puzzling over the emotional reactions to ancient and unconscious stresses of childhood. He will find no invocation of other worlds to which a sip or a smoke can bring partial entry. What he will see is a straightforward logical argument, based on direct experience and on the literature of experiments with patients and less disturbed people. Introspective verbal reports form key parts of the evidence. It all reads like the old psychology, but there is a difference. The text is explicitly based on the European writers, such as Karl Jaspers, who have stressed to the point of a systematic philosophy the distinction between the content of psychological experience and its form. "One man believes that he is being ruthlessly pursued by Goldfinger, whilst another believes that he is Portnoy's lost twin brother." This difference may give insight into psychic development, perhaps even into therapy, but it does not explain the phenomenon of delusion. Do the men hold their beliefs with unshakable conviction? Such a question may throw more light on the pattern.

Take hallucinations. These may appear to any of the senses, or even to internal sensations such as proprioception. One hears of the "cocaine bug"—an imagined insect crawling on the skin—or the insistence that the patient's liver has been eaten away. Induction of these states by drugs and by sensory deprivation are now well known. Here Reed is skeptical. Deprivation indeed reduces the noise level of normal perception against which imagery can appear. Images under sensory deprivation seem more related to the imagery of the drowsy and the newly awakened. In both models the strict criteria that might distinguish true hallucination are not generally applied. Hallucinations lie for this author between illusions on the one hand and unusual schemes of categorizing on the other. It is the absence of recognition of the subjectivity of the perception that makes the difference. The unreal perception is not itself the heart of the matter; it is the belief in the external correlate that marks the true hallucination. A circular case is cited. An old lady was very disturbed by the smell of the poison gas her evil neighbor pumped into her parlor. No one else smelled it, she was assured. Her enemy was so ingenious, she retorted, that his gas was odorless! Her experience was thus no percept at all, but a projection from internal ideas.

So too with delusions. Again the content may be varied, but the form remains more nearly fixed and is primary. In delusion the environment presents a world of new meanings, said Jaspers. As an oversimplified example, Reed suggests we consider a box on the table. Normally we perceive and classify it appropriately, in a useful hierarchy of relations. But suppose, with all our skills at perceiving boxes, colors, people and the rest, we have suffered a single "slippage" in our scheme. The box is now primarily classified as "Thing which may hurt me." It is now easy for the remaining, quite intact reasoning to suggest that the box is filled with high explosive or is radioactive or is merely a symbol of evil intent. It is not so much what we believe that is important to delusion but rather how we believe it, believe it strongly enough so that we search experience to find among familiar structures a novel fit for the novel meaning.

There is the ring of solid metal here. A natural-science reader might hesitate to accept this entirely non-neurological level of discourse as final, yet he can begin to see how a relatively local disorder in the function of mind, perhaps one day to be correlated with a local condition of the brain, might induce out of a life's experience a complex system of delusion. The crucial point of the book is that we do not function as recorders or cameras. We function by actively assembling a tiny sample of the richness of cues the world gives us into our working model of that world. We act as problem-solvers as much with respect to attention, perception, imagery or memory as when we weigh conclusions consciously. Anomalous experience immediately enters the continuum of constructive, problem-solving, interpreting processes. *Déjà vu* needs no reincarnation; "seeing" someone who is not present or feeling that one is being followed needs no spirits. We must expect discrepancies in our constructions; we find them within ourselves and with others, and surely with the fullness of the world.

Altogether this is a book for the thoughtful, lying within psychology at its frontier with philosophy. Its clear survey of fact and experience is in itself interesting and novel. The experience of the *Doppelgänger*—the encounter with a hallucinated image of oneself—is frequent among great poets and novelists. *Déjà vu,* its opposite *jamais vu* and the living double are related in *l'illusion des sosies.* (*Sosie* is French for "double.") In this bizarre phenomenon an individual

Anatomy of a Nikon Camera

Most Nikon photographers are content to enjoy their cameras for what they are, magnificent picture-taking instruments. But just what kind of technology is engineered into a Nikon that makes it perform so responsively, so precisely, so reliably year after year? The data is quite extensive and in some cases, confidential.

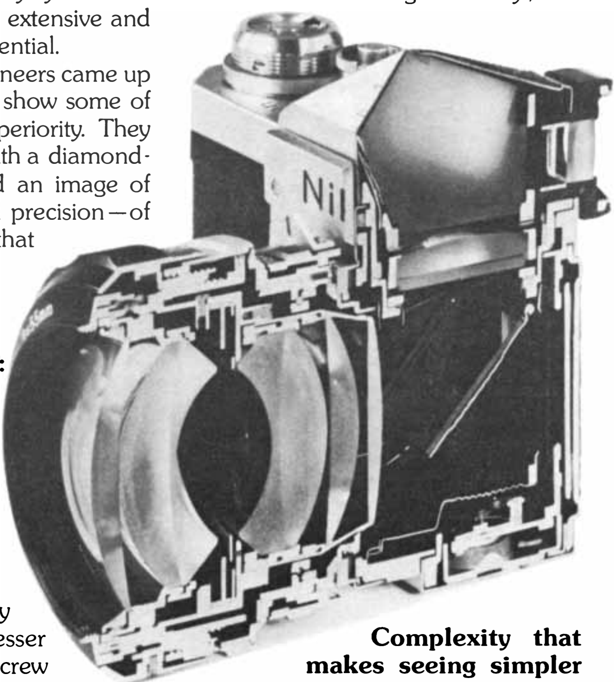
However, Nikon engineers came up with a different way to show some of the reasons for its superiority. They cut a Nikon in half, with a diamond-blade saw. It revealed an image of ingeniously purposeful precision—of wondrous complexity that belies the natural ease and speed with which the camera operates.

Precision threading: expensive, and important

Note the profusion of finely threaded components. The elements of every Nikkor lens, for example, are secured by threaded retaining rings, rather than by friction as they are in lesser lenses. Further, every screw is threaded into the part it is holding, not simply tightened against it. And, because screw holes are drilled before lens elements are mounted, you'll never find metal burrs inside a Nikkor lens. Precision of this sort is costly but the only real way to assure the permanent alignment accuracy required for such consistently sharp pictures, year after year.

The mount can make or break the lens.

One place a Nikkor lens has no threads is where it attaches to the camera body. Some cameras still have screw mounts which make changing lenses a slow process. Others use a breech-lock ring which can be mistaken for the aperture ring and allow the lens to come off accidentally. In contrast, the Nikon bayonet system requires only a quick 1/6 turn to remove a lens or lock it securely to the camera. The precision with which the body and lens are mated makes Nikon one of the very few cameras to accept an interchangeable f1.2 lens. (The extremely shallow depth-of-field of this lens demands meticulous mounting accuracy.)



Complexity that makes seeing simpler

Around the finder screen and prism you'll see a great deal of complex mechanics, quite unlike an ordinary single lens reflex. That's because, on the Nikon, both finder and finder screen are interchangeable. Naturally, this requires flanges and catches and releases which most cameras don't need. But as a result, you have your choice of any of 18 different finder screens and six different finders. You can tailor your choice of screen to the lens you're using, the subject you're shooting and the amount of light falling upon it. You can choose a finder that's perfect for ground level, action, available light or any other kind of photography. So you never have to compromise your view of the image. And that, after all, is really the great advantage of the single lens reflex system: to be able to frame and focus on any subject, near or far, quickly, easily and accurately.

Flatter film means sharper photographs

It's the function of the pressure plate to hold the film perfectly flat. Actually, it doesn't do that completely in *any* 35mm camera. Nikon designers have taken an extra step. The film winds onto the takeup spool in reverse, which counteracts the film's natural curl. The result is measurably flatter film and, as a direct result, sharper photographs. This reverse wind also pulls the film *around* the sprocketed roller instead of just across, so more sprockets are engaged and it becomes much less likely that you'll ever rip any sprocket holes out of the film, no matter how energetically you operate the winding lever or how fast a motor drive winds the film.

These are only a few of the design innovations harbored within Nikon FTN and F2 cameras. If space permitted, we could give you dozens of other insights into the care and ingenuity which is lavished on these exceptional cameras. Pick up a whole Nikon at any dealer, release the shutter, operate the winding lever, interchange the lens, the finder, the finder screen and you'll have some appreciation of the precision machining and assembly which also goes into every Nikon. (If, however, you'd like to see the actual cutaway camera, visit Nikon House in New York City, where it's on display, along with half Nikon movie and Nikkormat cameras.) Or write, and we'll send detailed literature. Nikon Inc., Dept. SA, Garden City, N.Y. 11530. Subsidiary of Ehrenreich Photo-Optical Industries, Inc EPOI (Canada: Anglophoto Ltd., P.Q.)

INDEX OF ADVERTISERS

JANUARY 1974

AMERICAN IRON & STEEL INSTITUTE ... 86
Agency: Hill and Knowlton, Inc.

AMERICAN TELEPHONE AND TELEGRAPH COMPANY, THE ... 16, 17
Agency: N. W. Ayer & Son, Inc.

AMP INCORPORATED ... 121
Agency: The Aitkin-Kynett Co., Inc.

AMPEX CORPORATION ... 13
Agency: Hal Lawrence Incorporated

ASAHI OPTICAL CO., LTD. (Int.) ... 131
Agency: Grey-Daiko Advertising, Inc.

ATLANTIC RICHFIELD COMPANY ... 65
Agency: Needham, Harper & Steers Advertising, Inc.

AUDIO DYNAMICS CORPORATION ... 119
Agency: Harvard, Peskin & Edrick, Inc.

BOOK OF THE MONTH CLUB ... 5
Agency: Wunderman, Ricotta & Kline, Inc.

CROWN PUBLISHERS ... 124, 129
Agency: Sussman & Sugar, Inc.

DATA GENERAL ... 6, 7
Agency: Scali, McCabe, Sloves, Inc.

EASTMAN KODAK COMPANY ... 49
Agency: Rumrill-Hoyt, Inc.

EDMUND SCIENTIFIC COMPANY ... 119
Agency: Chittick Advertising

EXXON CORPORATION ... 132, Inside Back Cover
Agency: McCaffrey and McCall, Inc.

GALLERY OF HOMES, INC. ... 124
Agency: The Ad Group

GAMES & PUZZLES ... 106

GENERAL MOTORS CORPORATION, GMC TRUCK & COACH DIVISION ... 68
Agency: McCann-Erickson, Inc.

HEWLETT-PACKARD ... 14, 15
Agency: Richardson Seigle Rolfs & McCoy, Inc.

JAPAN AIR LINES CO., LTD. ... 9
Agency: Ketchum, MacLeod & Grove Inc.

LEITZ, E., INC. ... 4
Agency: Ries Cappiello Colwell, Inc.

LIBRARY OF SCIENCE, THE ... 122
Agency: Henderson & Roll, Inc.

MERCEDES-BENZ AG (Int.) ... 132, Inside Back Cover
Agency: Heumann, Ogilvy & Mather G.m.b.H. & Co.

MERCEDES-BENZ OF NORTH AMERICA, INC. ... 131
Agency: Ogilvy & Mather Inc.

MICRO INSTRUMENTATION & TELEMETRY SYSTEMS, INC. ... 129
Agency: Mellekas & Associates

MONSANTO RESEARCH CORPORATION ... 10, 11
Agency: Advertising & Promotion Services

NATIONAL CAMERA, INC. ... 106
Agency: Langley Advertising Agency

NIKON, INC., ASSOCIATION OF EHRENREICH PHOTO OPTICAL INDUSTRIES ... 127
Agency: Gilbert, Felix & Sharf Inc.

NIPPON KOGAKU K. K. (Int.) ... 127
Agency: K & L Advertising Agency

OLYMPUS OPTICAL CO., LTD. (Int.) ... 53
Agency: Fuji Agency, Ltd.

OXFORD UNIVERSITY PRESS ... 120
Agency: Franklin Spier Incorporated

PERFORMANCE DYNAMICS INC. ... 107
Agency: McCormack Services

POLAROID CORPORATION ... 66, 67
Agency: Doyle Dane Bernbach Inc.

QUESTAR CORPORATION ... 64

RCA ... Back Cover
Agency: J. Walter Thompson Company

ROYAL ACADEMY OF SCIENCES, THE ... 125

SCANDINAVIAN AIRLINES SYSTEM (Int.) ... 68
Agency: Sky Market Promotion AB

SCIENTIFIC AND TECHNICAL RESEARCH COUNCIL OF TURKEY, THE ... 129

SOUTHERN CO., THE ... 19
Agency: Liller Neal Battle & Lindsey Inc.

SPANEL, A.N. ... 2

STANFORD COURT, THE ... 18
Agency: Spiro & Associates

UNITED AUDIO PRODUCTS, INC. ... 52
Agency: Ries Cappiello Colwell, Inc.

U. S. PIONEER ELECTRONICS CORP. ... 18
Agency: Philip Stogel Company Inc.

WALL STREET JOURNAL, THE ... 53
Agency: Batten, Barton, Durstine & Osborn, Inc.

WESTINGHOUSE ELECTRIC CORPORATION ... Inside Front Cover, 1
Agency: Ketchum, MacLeod & Grove Inc.

recognizes all the attributes of another—say a wife or a husband—but insists nevertheless that the person is an impostor or a double. Many more such remarkable states are cited, some bizarre in the extreme, some familiar in daily life. The entire work is a fascination.

AIRPLANES: FROM THE DAWN OF FLIGHT TO THE PRESENT DAY, by Enzo Angelucci. Color illustrations and design by Marco Rota. McGraw-Hill Book Company ($14.95). Three specific groups these days seek to know something about particular airplanes. The newspaper reader or television viewer, anxious enough about the headlines, has reason to learn the nature of the Skyhawk or the Sukhoi alluded to by the wire-service reporter who has seen it whizzing by. Then there is the amateur of design history, for whose benefit many volumes of hard-edged color drawings of aircraft types "in the markings of the Royal Nepal Air Force" are prepared in loving detail. This group comes close to the builders of scale models, and may indeed overlap with them. Authenticity and completeness are their goals, although it is form they model, not function. Finally, students and readers in general share a historical interest, almost an antiquarian one: How did these fast-evolving creatures of a man-made evolutionary process look long ago?

This big, handsome reference book, with a knowing text by an Italian publishing executive and aircraft buff, drawings in color by Marco Rota and black-and-white drawings by an entire team of artists working with him, is translated from an Italian edition of 1971. It effectively straddles the special needs of all the groups mentioned. More than 900 species of aircraft are pictured in small, neat drawings, usually viewed from a single convincing angle instead of in the rigidly perpendicular mechanical views of the usual technical artwork. The current-events follower can study the Sukhoi Su-7B ground-attack plane, or any other expensive modern weapon, up to the F-15 and the Panavia 200 Panther, which were prototypes in 1973. The nostalgic can see the Red Baron's Fokker triplane ("an insoluble problem in the wing structure") or the Blériot XI, first across the Channel. The text is a good mix of nostalgia, analysis and narrative detail. Every airplane drawn is given a tabular line describing its size, engine, builders, dates and performance. A few pages present a score each of charming postage-stamp-size drawings. The most interesting is one that shows in color the preadaptive history of aircraft, before anything but balloons flew, from Father de Lana's evacuated copper balls to Tom Edison's flying ship of 1880; these "dreams and visions between 1670 and 1880" have no entries in the factual tables. Only general aviation today is somewhat slighted; the military, the airlines and the big private jets and turboprops are here in all completeness, but the little Cessnas and their kin in the tens of thousands are represented on two agreeable pages only by 35 drawings of light planes from all over the world. (The few most famous balloons and dirigibles sneak in somewhere too.)

PORTABLE WORLD, edited by Barbara Guilfoyle. Photographs (unless otherwise noted) by Bob Hanson. Museum of Contemporary Crafts, New York ($3). Right at you from the front cover hops a cool young man on a one-cylinder internal-combustion pogo stick (a Hop Rod), his ears grasped by an FM headset radio, flashlight glasses flanking his eyes, fishing vest and army belt crammed, furnished with crimson Swiss knife, dog-repelling flashlight, emergency neck pendant containing a supply of insulin and syringe and red-trimmed Lectra-sox. These modern ingenuities do not overshadow the photographs within: a parfleche, a yurt or an old folding featherbed-in-a-chest. George Washington was the lucky owner of that leatherbound metal bed, a gift from an admirer. Photographs of objects as simple in form as the tiny mountain stoves and lanterns or as complex and elegant as the television set with 1½-inch screen or the pocket calculator, or conscious works of art such as the colorful abstract-appliqué nylon sleeping bag, fill the large pages of this slender exhibition catalogue devoted to the artifacts that accompany and nourish human movement. The exhibit, arranged by Barbara Bullock, comprised five categories: historical, body extensions, indoor, outdoor and shelter and vehicles. All these sections are sampled by the photographs; the list of the entire exhibit is even more enticing to a reader prepared to travel or to imagine travel. Anyone with admiration for design will be pleased by the solutions offered in times past and present to the challenges of portability, as old as the days when all humans gathered food over open plains or as new as Skylab. This is no technical work; the objects shown are identified carefully but described with only a few words. It is a lighthearted and yet serious celebration of mobile humanity, one paying no more attention to the practical than to the symbolic and decorative needs of the rolling stone.

BIBLIOGRAPHY

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

ENERGY POLICY IN THE U.S.

THE NATIONAL ENERGY OUTLOOK. Shell Oil Company, Houston, March, 1973.

ENERGY RESEARCH AND DEVELOPMENT: REPORT OF THE TASK FORCE ON ENERGY OF THE SUBCOMMITTEE ON SCIENCE, RESEARCH, AND DEVELOPMENT, DECEMBER, 1972. Committee on Science and Astronautics, U.S. House of Representatives. U.S. Government Printing Office, 1973.

ELECTROCHEMICAL MACHINING

ELECTROCHEMICAL MACHINING. Edited by A. E. De Barr and D. A. Oliver. American Elsevier Publishing Co., Inc., 1968.

AN INVESTIGATION OF THE DIFFERENCES BETWEEN NACL AND $NACLO_3$ AS ELECTROLYTES IN ELECTROCHEMICAL MACHINING. James P. Hoare, Mitchell A. LaBoda, Michael L. McMillan and Augustine J. Wallace, Jr., in *Journal of the Electrochemical Society,* Vol. 116, No. 2, pages 199–203; February, 1969.

PRACTICE AND THEORY OF ELECTROCHEMICAL MACHINING. John F. Wilson. John Wiley & Sons, Inc., 1971.

ANION EFFECTS ON THE DISSOLUTION OF STEEL IN ECM BINARY ELECTROLYTE SYSTEMS. James P. Hoare and Kao-Wen Mao in *Journal of the Electrochemical Society,* Vol. 120, No. 11, pages 1452–1456; November, 1973.

THE NERVOUS SYSTEM OF THE LEECH

SPECIFIC MODALITIES AND RECEPTIVE FIELDS OF SENSORY NEURONS IN THE CNS OF THE LEECH. J. G. Nicholls and D. A. Baylor in *Journal of Neurophysiology,* Vol. 31, No. 5, pages 740–756; September, 1968.

REGENERATION AND CHANGES IN SYNAPTIC CONNECTIONS BETWEEN INDIVIDUAL NERVE CELLS IN THE CENTRAL NERVOUS SYSTEM OF THE LEECH. J. K. S. Jansen and J. G. Nicholls in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 69, No. 3, pages 636–639; March, 1972.

A COMPARISON OF CHEMICAL AND ELECTRICAL SYNAPTIC TRANSMISSION BETWEEN SINGLE SENSORY CELLS AND A MOTONEURONE IN THE CENTRAL NERVOUS SYSTEM OF THE LEECH. J. G. Nicholls and D. Purves in *Journal of Physiology,* Vol. 225, No. 3, pages 637–656; September, 1972.

THE CONTRIBUTION OF MEMBRANE HYPERPOLARIZATION TO ADAPTATION AND CONDUCTION BLOCK IN SENSORY NEURONES OF THE LEECH. David C. Van Essen in *Journal of Physiology,* Vol. 230, No. 3, pages 509–534; May, 1973.

THE CELL CYCLE

THE CELL-CYCLE: GENE-ENZYME INTERACTIONS. Edited by George M. Padilla, Gary L. Whitson and Ivan L. Cameron. Academic Press, 1969.

THE BIOLOGY OF THE CELL CYCLE. J. M. Mitchison. Cambridge University Press, 1971.

THE MAMMALIAN CELL AS A MICROORGANISM: GENETIC AND BIOCHEMICAL STUDIES *IN VITRO*. THEODORE T. Puck. Holden-Day, Inc., 1972.

THE CELL CYCLE IN DEVELOPMENT AND DIFFERENTIATION. Edited by M. Balls and F. S. Billett. Cambridge University Press, 1973.

THE AGE OF THE ELEMENTS

NUCLEAR ASTROPHYSICS. William A. Fowler. American Philosophical Society, 1967.

PRINCIPLES OF STELLAR EVOLUTION AND NUCLEOSYNTHESIS. Donald D. Clayton. McGraw-Hill Book Company, 1968.

NUCLEOCHRONOLOGIES AND THE MEAN AGE OF THE ELEMENTS. David N. Schramm and G. J. Wasserburg in *The Astrophysical Journal,* Vol. 162, No. 1, part 1, pages 57–69; October, 1970.

THE PERCEPTION OF DISORIENTED FIGURES

RECOGNITION UNDER OBJECTIVE REVERSAL. George V. N. Dearborn in *The Psychological Review,* Vol. 6, No. 4, pages 395–406; July, 1899.

THE ANALYSIS OF SENSATIONS AND THE RELATION OF THE PHYSICAL TO THE PSYCHICAL. Ernst Mach, translated from the German by C. M. Williams. Dover Publications, Inc., 1959.

ORIENTATION AND SHAPE I AND II in *Human Spatial Orientation.* I. P. Howard and W. B. Templeton. John Wiley & Sons, Inc., 1966.

SIMILARITY IN VISUALLY PERCEIVED FORMS. Erich Goldmeier in *Psychological Issues,* Vol. 8, No. 1, Monograph 29; 1972.

ORIENTATION AND FORM. Irvin Rock. Academic Press, 1974.

THE PHYSICS OF THE BOWED STRING

ON THE MECHANICAL THEORY OF VIBRATIONS OF BOWED STRINGS AND OF MUSICAL INSTRUMENTS OF THE VIOLIN FAMILY. C. V. Raman. Indian Association for the Cultivation of Science, 1918.

THE MECHANICAL ACTION OF VIOLINS. F. A. Saunders in *The Journal of the Acoustical Society of America,* Vol. 9, No. 2, pages 81–98; October, 1937.

ON THE SENSATIONS OF TONE AS A PHYSIOLOGICAL BASIS FOR THE THEORY OF MUSIC. Hermann L. F. Helmholtz. Dover Publications, Inc., 1954.

THE BOWED STRING AND THE PLAYER. John C. Schelleng in *The Journal of the Acoustical Society of America,* Vol. 53, No. 1, pages 26–41; January, 1973.

THE ASTROLABE

A TREATISE ON THE ASTROLABE. Geoffrey Chaucer, edited by the Rev. Walter W. Skeat. Oxford University Press, 1872.

THE ASTROLABES OF THE WORLD. Robert T. Gunther. Oxford University Press, 1932.

TRAITÉ DE L'ASTROLABE. Henri Michel. Gauthier-Villars, Imprimeur-Éditeur, 1947.

THE PRINCIPLE AND USE OF THE ASTROLABE in *Oriens-Occidens.* Willy Hartner. G. Olms, Hildesheim, West Germany, 1968.

MATHEMATICAL GAMES

THE FUNDAMENTAL IDEAS OF CHINESE SCIENCE: (G) THE SYSTEM OF THE BOOK OF CHANGES in *Science and Civilisation in China: Vol. 2.* Joseph Needham. Cambridge University Press, 1956.

COMBINATORIAL ASPECTS OF THE HEXAGRAMS IN THE CHINESE BOOK OF CHANGES. F. van der Blij in *Scripta Mathematica,* Vol. 28, No. 1, pages 37–49; May, 1967.

THE AMATEUR SCIENTIST

EVOLUTION OF AN AMATEUR WEATHER-SATELLITE PICTURE STATION. Charles H. McKnight in *QST,* Vol. 52, No. 4, pages 28–34; April, 1968.

On the test track. Picture taken at 90 mph.

The 450 SE Sedan. This is the car that may finally prompt you to invest in a Mercedes-Benz.

It's only human. You see a Mercedes-Benz go by and think to yourself, "one of these days..."

Now comes the 450SE Sedan. An automobile so enlightened in its concept and craftsmanship that it may finally prompt you to invest in your first Mercedes-Benz.

"The best sedan in the world," said *Road & Track* magazine. *Autoweek's* editors consider it "certainly the most exciting sedan in the world...so far ahead of its time it leaves us stunned."

The 450SE is powered by a unique engine. One which gives you better than 25 percent more power per cubic inch *and* better fuel economy than any domestic luxury car.

The secret is a 4.5-liter overhead-cam V-8 with electronic fuel injection. There is simply nothing like it in any other sedan in the world.

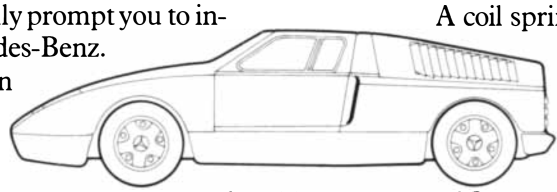
Suspension better than ever

Considering the Mercedes-Benz reputation for handling, that is quite a statement.

The front suspension of the 450SE is a direct lesson from our 180-mph experimental C-111 Coupe. Its zero steering offset gives this latest Mercedes-Benz remarkable straight-line stability. And this full 5-passenger sedan has a turning circle only 38.0 feet in diameter.

The fully independent rear suspension is exactly like that of the famous 450SL sports car. A coil spring and gas-pressurized shock absorber mate each wheel to the surface of the road.

Remarkable handling is the result. So remarkable, that the Mercedes-Benz engineers consider the 450SE their best handling sedan ever.



Front suspension adapted from experimental C-111.

The safest sedan in the world?

Mercedes-Benz would never say that. But the experts at *Autoweek* had no qualms. They called the 450SE "the safest car in the world available to the public."

No automobile manufacturer anywhere in the world is more concerned with safety. No one else has been more active in making it a part of his product for as long a time.

Mercedes-Benz has always aimed to build not just successful automobiles, but significant automobiles. That has been achieved in the 450SE Sedan. What awaits you now is the most remarkable driving experience of your life.

We'd like you

Today, nuclear power supplies 5% of America's electricity. Tomorrow, it could supply over 50%. We need to reach this goal as soon as possible.

Energy from the atom is one of our country's most promising sources of power.

Ten years from now, nuclear energy should be generating about one-third of our electrical power. By the year 2000, over half of it will probably be supplied by nuclear energy.

Where uranium is found.

Uranium ore is the raw material for nuclear fuel. So far, the commercial deposits of uranium that have been found in America are in the western states.

One of the largest deposits discovered in recent years is the site of Exxon's Highland Mine near Casper, Wyoming. Here, we are removing over 14,000 tons of uranium ore a week.

Once the ore is mined, we extract the uranium from the ore and turn it into a fine powder known as "yellow cake."

It is then converted into a gas and sent to plants run by the Atomic Energy Commission. Here, the fissionable, or energy-producing, component is concentrated. This process is called enrichment.

The fuel for a reactor.

We make the enriched uranium into fuel for reactors at Exxon Nuclear Company's plant in Richland, Washington.

First, we convert it into pellets about twice the size of an aspirin tablet. After numerous checks for precise weight, size, shape and uniformity, the pellets are sealed in zirconium or stainless steel tubes. The tubes are then grouped together in a precise geometric pattern to form a fuel bundle or assembly.

A single tiny pellet of enriched uranium can generate as much energy as 150 gallons of fuel oil. Thousands of these pellets go into a nuclear fuel bundle. A single bundle can generate enough electricity to run 4,000 average-sized homes.

This is Exxon's uranium mine in Wyoming. When the surface mining is completed, most of the area will be filled in, contoured and planted to match the rest of the countryside. Grass, wildflowers and shrubs that Exxon planted are already growing on much of the earth that has been excavated from the mine.

When these assemblies have passed several thousand checks and inspections, they are shipped to nuclear electric-power generating plants. It takes about 600 assemblies to make up the core or heat source of a nuclear reactor. And such a reactor can generate enough electricity to run a city the size of Boston.

Making nuclear power more plentiful.

Exxon is currently working on several projects to expand the availa-

Did you know your car could have built-in reflexes?

Some drivers will always be better than others. But thanks to advances in electronic technology, all drivers will be better some day because their autos will have built-in reflexes.

Drivers will have electronic devices to control skidding detect highway hazards, warn of malfunctions—all automatically.

Electronic improvement of the automobile is almost limitless. By 1980 auto manufacturers may be spending as much on solid state electronic systems as on engines. Even so, electronics should eventually lower the cost of running a car.

Today, RCA provides the auto industry with components for the new electronic ignition systems. Tomorrow, we will provide a variety of other advances—such as electronic devices to replace inefficient mechanical methods of fuel mixture, which will cut energy consumption and aid in emission control.

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