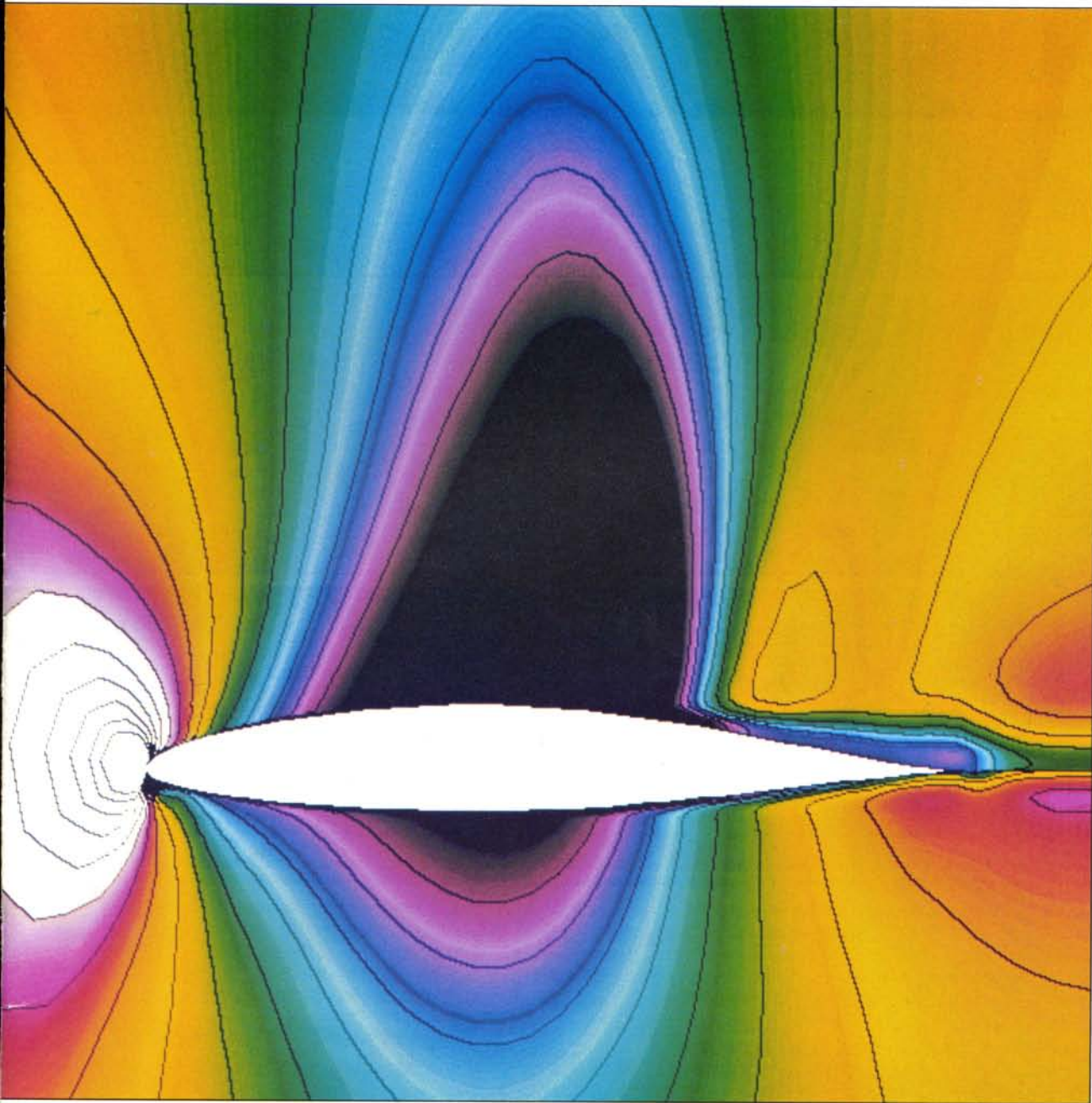# SCIENTIFIC
AMERICAN

SUPERCOMPUTERS

$2.00

January 1982

The automobile is the most popular form of transportation ever developed. And it is the second largest investment most of us will ever make. The automobiles we select can significantly affect the quality of our daily lives. The decision is too important to be left to a whim.
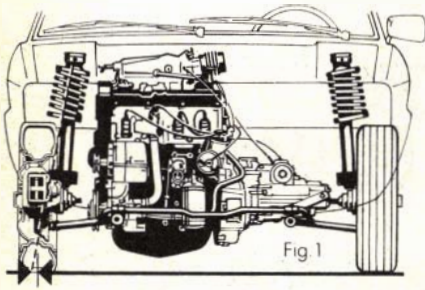
The fundamentals.

How Volkswagens work.

Fig. 1

VWs have certain things in common with other cars. They have engines, transmissions and control systems. And, to operate them, you turn a key. Step on the gas. And go.

Beyond that, the differences between VWs and other cars become more apparent than their similarities.

Component position. Fig. 1 shows where the main components (i.e. engine, transmission, etc.) lie in a VW. They're in the front so there is greater utilization of engine space. With all the main components forward, there is more useable passenger and cargo room in back. The engine is also tilted for greater road visibility and decreased wind resistance. Not to mention easier access to the engine by VW service technicians.

VWs have front-wheel drive which means the wheels that steer the car, pull the car. It affords the vehicle exceptional directional control. And, with the combined weight of the engine and transmission mounted directly over the drive wheels, there is improved road traction.

Two engine types. VW engine systems have features often found on high performance cars. For instance, VW gas engines have Bosch K Jetronic CIS fuel injection, a continuous injection system perfectly matched to VW's accurate overhead cam engines. It's precise for cleaner burn. It's reliable for easier

start-up in cold weather. And, it's responsive for instant pickup in passing.

The gas engine design was remarkable in that it was easily converted to diesel power. The resulting power-plant was the lightest of its kind produced, ever. It has a novel swirl combustion chamber for more efficient burn, an extraordinarily high RPM for excellent pickup and a 7 sec. glowplug for faster starts.

Further, in the diesel cycle (Fig. 2a-2e), fuel, injected directly into a cylinder filled with hot compressed air (Fig. 2c), self-ignites, instantly, without needing sparkplugs, a coil or a distributor. And, without their routine yearly maintenance.

Fuel conservation. The diesel added to a tradition of VW frugalness. And VWs have the highest EPA figures.
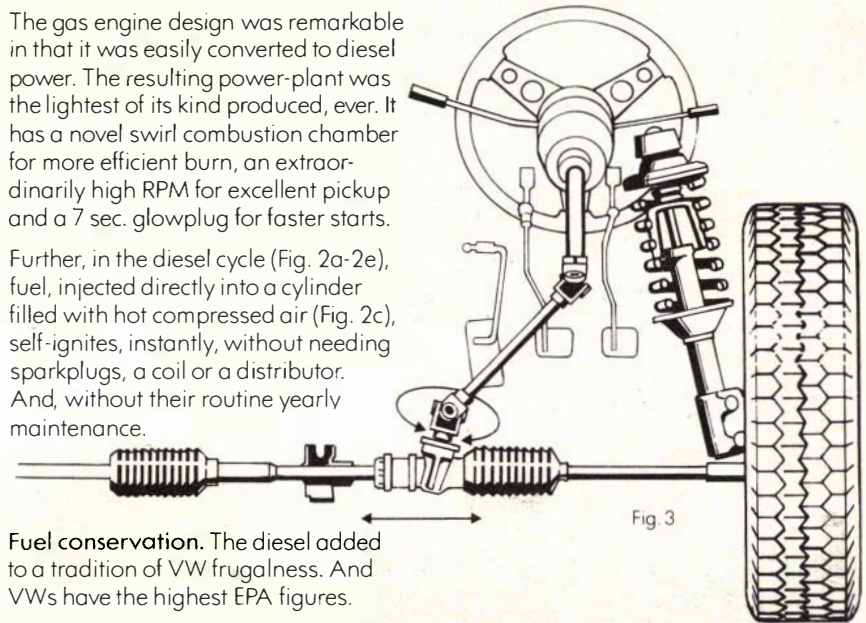
750 Cars Tested	Hwy. Est.	Est. MPG
Best of All Cars VW Rabbit Diesel, manual, 4-speed	5B	45
Best Pickup Truck VW Diesel Pickup Truck, manual, 4-speed	49	42
Best Passenger Van VW Diesel Vanagon, manual, 4-speed	29	29
Best Convertible VW Rabbit Convertible, manual, 5-speed	40	27

(Use "est. MPG" for comparisons. Your mileage may vary with weather, speed and trip length. Actual highway mileage will probably be less.)

What's more, to increase the chance of matching those numbers (with a possibility of even exceeding them) VW has developed a unique upshift light for all 1982 cars that tells drivers exactly when to shift for optimum mileage and performance.

Control Systems. Combined with everything else, VWs display uncanny

Fig. 3

maneuverability. This is due to several factors. They have rack-and-pinion steering, a precise method of turning whereby a rack with gear teeth transmits motion to the wheels when meshing with a pinion gear at the end of the steering column (Fig 3); independent four-wheel suspension, which keeps the wheels at a perfect tangent to the road; negative steering roll radius, a feature VW pioneered to keep a car in control in the event of a high speed blowout; and dual diagonal braking, a VW first in Germany.

And that, basically, is how VWs work. All together, they form an undeniable package of outstanding and unique engineering design.

In the future, you'll read more about these points in greater detail. And, you'll not only understand cars a lot better than you did, you'll also have a better understanding of VWs as cars.

Nothing else is a Volkswagen.

Fig. 2a Fig. 2b Fig. 2c Fig. 2d Fig. 2e
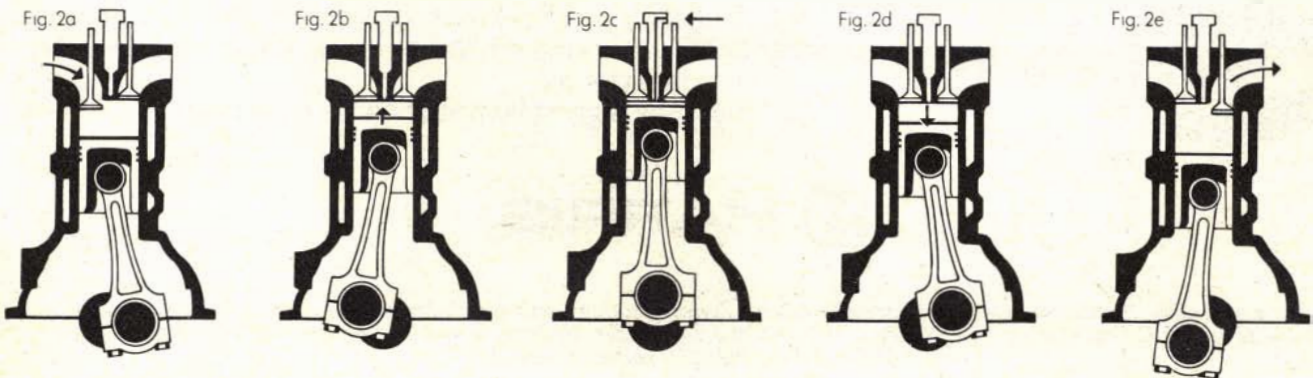
HARRIS HAS A SOLUTION FOR YOUR COMPANY'S BIGGEST PROBLEM:

SOLUTIONS FOR YOUR COMPANY'S SMALLER PROBLEMS.

Harris super-minicomputers offer software applications in computer-aided design, simulation and some 300 other areas.

Most big companies are made up of a lot of smaller companies. They're called *departments*. And each has its own way of doing business.

It's the business of management to make sure that the efforts of each department work in concert with the rest. And that's a job Harris can make a whole lot easier.

We offer super-minicomputers for your R&D and technical people that are unmatched in power, performance and reliability.

The people who run your company's data center will appreciate our line of communications-oriented computer terminals (remote batch and interactive). It's the broadest in the industry.

Your MIS executives will value our distributed data processing systems for their speed, power and unique capabilities. Like multi-host support and word processing options.

Your communications managers will recognize the difference in bottom-line telephone costs our PBX's can make.

We also provide intelligent telephones, a unique private tandem networking capability and business satellite earth stations that can further reduce your telecommunications costs. And that will surely make your accounting department happy.

Our PBX can accommodate from 20 to 1,000 lines.

Our data terminals can grow into sophisticated distributed data processing systems.

The fact is, Harris manufactures products that can improve the productivity of people in every one of your departments. What's more, because they're designed for future expansion and compatibility, they can maximize *corporate* productivity, too.

And that's the biggest contribution we can make.

HARRIS

2

THE COVER

The picture on the cover is one frame from a motion picture depicting the flow over an airfoil at Mach .7, or .7 times the speed of sound, as simulated by the supercomputer ILLIAC IV (see "Supercomputers," by Ronald D. Levine, page 118). The computer simulation was part of a numerical study of the dynamics of aileron flutter carried out at the National Aeronautics and Space Administration's Ames Research Center at Moffett Field, Calif. Aileron flutter is a destructive vibrational effect that can arise when shock waves moving on the wing interact with the movement of the aileron, the rear control surface. The numerical study traced the changes in air density at a two-dimensional grid of some 7,000 points curved to fit the airfoil. The grid points are nonuniformly distributed: widely spaced at the wing's leading edge and closely spaced in the region of interest near the trailing edge. Each frame of the picture required 40 million arithmetic operations, or two seconds of computing time on the ILLIAC IV. The colored curves represent contours of equal air density. Dark yellow is the density of undisturbed air; lighter colors indicate compression, darker colors rarefaction. The shock wave is where the colors change rapidly a third of the way forward of the trailing edge. The shock wave causes the airflow to pull away from the trailing edge, setting up a region of reverse circulation (*purple*). The computation was done by Harry E. Bailey. The program for the computer graphics was written by Tony Hasegawa and Alan Johnson.

THE ILLUSTRATIONS

Cover illustration courtesy of Ames Research Center,
National Aeronautics and Space Administration

4

DEVELOP YOURSELF.

Pick your area of interest. And *teach yourself* with six easy-to-follow books in the new *KODAK Workshop Series*.

Each volume contains 96 pages of step-by-step pictorial instructions and costs only $8.95.

Black-and-White Darkroom Techniques details all the steps for developing, printing, and finishing black-and-white prints. Plus, other sections cover choosing photographic paper, dodging and burning in, mounting prints, and more.

Using Filters gives creative advice that helps you see and capture extraordinary images in color and black-and-white. It shows you how to create mood with color filters, add dazzle with special-effects filters, and render more natural colors and tones with correction filters.

Photographing with Automatic Cameras shows you how to get the most from your automatic camera. Especially useful for newcomers to 35 mm photography, it looks at how automatic cameras work and how to best use one. It offers special tips for capturing landscapes, portraits, and other subjects with an automatic camera.

Building a Home Darkroom tells you how to plan, what you'll need, and where to locate. From planning to construction to electrical work to plumbing, this book shows the step-by-step completion of a prototype darkroom in photographs and easy-to-understand illustrations.

Electronic Flash starts with the basics, giving thorough coverage of how, when, and where a flash can be used to enhance your photographs. Advanced techniques such as bounce and multiple flash are discussed.

Color Printing Techniques explains the basic procedures for making color prints from color negatives and color slides (including the new KODAK EKTAFLEX PCT Color Printmaking Products). Also included is information on equipment, darkrooms, retouching, and framing.

Choose one or all six titles in the new *KODAK Workshop Series* at your camera store or wherever fine photography books are sold.

Because the best way to put some muscle in your photographic development is to work out with the experts.

When it comes to photography, Kodak wrote the book.

5

LETTERS

Sirs:

The article on the Vitruvian odometer by André Wegener Sleeswyk in your October issue makes a suggestive case for Archimedes' invention of this device. The connection might be strengthened by bringing in the recent work of Wilbur R. Knorr of Stanford University ("Archimedes and the *Elements:* Proposal for a Revised Chronological Ordering of the Archimedean Corpus," *Archive for History of Exact Sciences,* Vol. 19, No. 3, pages 211–290, 1978). Knorr finds several stages in the development of Archimedes' method of exhaustion, a precursor of the integral calculus, and he deduces as a result that the treatise *Dimension of the Circle* was a very early Archimedean work, perhaps the first. The tentative date is about 260 B.C., and so the work's conclusion that the value of pi lies between $3\frac{10}{71}$ and $3\frac{1}{7}$ fits well into the hypothesis that Archimedes worked on the odometer at about the same time.

Since the use of inappropriately sized wheels on the mileage-measuring cart would produce cumulative error, Archimedes may have had a strong practical reason to seek a closely correct value for pi. Vitruvius records that the odometer cart had wheels four feet in diameter and that they traveled 12.5 feet in each revolution. If we assume the cart was meant to travel 400 miles without be-

ing recalibrated, Archimedes' pi values would yield end results accurate to within .16 mile, assuming the wheels never slipped. The Vitruvian value yields results off by 2.11 miles, which was certainly close enough for practical purposes. If Archimedes observed wheel slippage and attempted to cope with it, then we have an engineering antecedent for the differential-gear subassembly of the famous Antikythera mechanism, which uses such gears, as the mileage cart might have done, to obtain the mathematical average of two separate motions. Vitruvius says nothing, however, of differential gears....

VERNARD FOLEY
WERNER SOEDEL

Purdue University
West Lafayette, Ind.

Sirs:

For a journal with a well-earned reputation for scientific accuracy, even scholarship, your staff writer's remarks under the heading "Menarcheal Misunderstanding" ["Science and the Citizen," SCIENTIFIC AMERICAN, October, 1981] are really quite disgraceful. One gets used to being quoted out of context, even to having one's carefully limited generalizations extrapolated beyond reason to serve the purposes of the sensationalized media. But to have graphs fabricated and "quotes" made up is something else again.

"A graph published by Tanner," you say, "gives the age at menarche as 17.1 years in 1840 and 12.5 years in the 1960's." No reference, of course; nothing scientific. But it seems likely the writer has in mind the old graph of 1962 (*Growth at Adolescence,* Blackwell Scientific Publications, second edition), in which the first point indeed represents menarche at 17.1 in Norway in 1840; the last point for Norway is 13.3 in 1952 and there is no point in the 1960's. Or perhaps your writer had recourse to the partially updated graph in Phyllis B. Eveleth and J. M. Tanner's *Worldwide Variation in Human Growth* (Cambridge University Press, 1977), where the 1840 point still remains and the last points are indeed in the 1960's but range, for different European countries, from 12.9 to 13.4. The further updating, in a very easily available reference, *Fetus into Man: Physical Growth from Conception to Maturity,* by J. M. Tanner (Harvard University Press, 1978), incorporated the historical researches of Gro Harlem Brundtland and Lars Walløe into the Norwegian archive (*Annals of Human Biology,* Vol. 3, No. 4, pages 363–374; July, 1976) with the result that the first Norwegian point is 16.0 years at 1860; the very important results of J. E. Brudevoll, K. Liestøl and L. Walløe (*Annals of Human Biolo-*

gy, Vol. 6, No. 5, pages 407–416; September/October, 1979) were still unpublished at that time.

So much for the graph of "17.1... and 12.5." Your writer continues: "Tanner estimates that the data given in the graph are 'accurate to within 1.1 years' for the populations studied." This is either pure imagination or just possibly comes from my statement that "the standard deviation for most series studied is approximately 1.1 years" (*Growth at Adolescence,* page 154).

Faced with ignorance of this order, Vern L. Bullough's remarks (as reported), which seem to have started it all, are scarcely more than averagely erroneous. Your writer attributes four statements to him; of these the first three are wrong ("report of a single physician," see Brudevoll, Liestøl and Walløe, loc. cit.; "based on individual observations," see text and tables at pages 286–298 in J. M. Tanner, *A History of the Study of Human Growth,* Cambridge University Press, 1981; "almost all the 19th-century data come from the Scandinavian countries," see Tanner, loc. cit.).

I feel sure, Sirs, your readers would like to know the truth about the trend in menarcheal age, in Europe and the U.S., as best modern scholarship can define it (and much old and new data have become available since 1962). At the end of the 19th century "the difference in age at menarche between the well-off and the poor was about a year in England and 18 months in Germany. In Norway, a considerable decline took place in the age of menarche of working-class women during the last one-third of the 19th century. In other (European) countries the evidence points to a small decline, if any. Certainly a very rapid decline took place from 1910 or 1920 to around 1970 and this occurred in most or all European countries, and in both rich and poor families, though in somewhat different degrees. The relation to economic activity and standard of living seems clearly implicit in these figures" (*A History of the Study of Human Growth,* page 297). The data for the U.S. are much less satisfactory but would indicate a mean for Boston and St. Louis dispensary patients of about 14.3 years in 1890; girls attending women's colleges at that time had a mean estimated at 13.5 (F. Engelmann, *Transactions of the American Gynecological Society,* Vol. 26, page 77, 1907). "Perhaps in the U.S. the decline is best estimated at about one year in the well-off classes, between say 1890 and 1930 to 1940 (by when it had probably stopped) and 18 months or a little more for the less well-off, between about 1890 and 1960 to 1970" (*A History of the Study of Human Growth,* page 298).

J. M. TANNER

University of London

7

© 1981 SCIENTIFIC AMERICAN, INC

50 AND 100 YEARS AGO

JANUARY, 1932: "Three years ago Edwin P. Hubble found that the observed motions of the nebulae depend on their distance. The remotest nebulae were receding the fastest, and the rate of recession was closely proportional to the distance, increasing by 160 kilometers per second for every million light-years. If this proportionality holds good for the more distant nebulae, their velocities should be enormous. It is no easy matter, however, to observe the remoter nebulae and find out, for they are exceedingly faint and can be observed only by combining a very large telescope and a spectroscope with a very short camera. With the Mount Wilson nebular spectrograph even a three-inch focus proved too long and a special lens had to be designed. With this powerful instrument attached to the 100-inch reflector the spectra of more than 40 faint nebulae have been obtained and measured by Milton L. Humason. All the values agree strikingly with the general relation."

"Demonstration that the germ of smallpox is probably a minute spherical body 1/125,000 inch in diameter, near the lower limit of microscopic visibility, has been announced by J. C. G. Ledingham, director of the Lister Institute in London. Professor Ledingham made practically pure suspensions of minute bodies from fowl pox and vaccinal lesions and found that they reacted specifically with serum from animals that had recovered from attacks of fowl pox or vaccinia. He also demonstrated the presence of specific agglutinins in the blood serum of a rabbit inoculated with vaccinia. Agglutinins are substances of as yet unknown nature the presence of which in the blood causes bacteria and other foreign bodies to gather in clumps. Professor Ledingham believes that in a fairly large group of virus diseases of both plants and animals such minute elementary bodies are very likely to be found if they are carefully sought. Such demonstrations would materially assist the study of virus infections whose infective particles are below microscopic vision."

"The assumption that it would be quite as easy to breed human beings up to the level of our 'best' people as it is for men to breed cattle or horses of desirable qualities is not only unscientific; it is ridiculous. No matter how normal and moral and noble we happen to be individually, we are carrying within us a heritage of weak genes that are masked by genes of better quality. Dr. Sewall Wright, the University of Chicago zoologist, has bred guinea pigs all kinds of ways to uncover their weak genes. He did this by making brother-and-sister matings and by building up 'pure' lines. In such lines all kinds of potential or inherent weaknesses develop, and the lines, so long as the brother-and-sister crosses are continued, breed true for their weaknesses. It might be possible, given many years, to breed a colony of guinea pigs in such a manner as to secure nothing but pure-line progeny. That done, the animals with weaknesses could be killed or allowed to die out, and the valuable ones could be mated to secure a finer race of guinea pigs than ever existed since the world began. It would be all but impossible to apply such a method to human beings and to breed monstrosities deliberately, even if we knew the factors governing intellectual inheritance (which we do not), or had agreed on the type of human being that would be the best and most valuable (which we probably cannot)."

JANUARY, 1882: "There is a probability of increased interest over the next few years in methods of testing the genuineness of all articles of food, beverages and medicines. Several states have lately passed laws authorizing official analysis of these articles wherever they are on sale. Comprehensive laws of this kind have been passed over the past year or two in Wisconsin and New York. And several states have lately enacted official inspection of particular articles; for example, Indiana requires analysis of all fertilizers on the market and of all the oils into which petroleum enters; Maine, of vinegar; Massachusetts, Nevada and New Jersey, of milk; Ohio, of milk, butter, cheese and meat, and of fertilizers. Anything like censorship of ordinary dealings has generally been unpopular in America. Such laws as have been passed under the pressure of increasing necessity for protecting the public against adulteration have, until very lately, avoided anything like censorship of trade, being confined to imposing damages or punishment for any fraud committed, provided the buyer could prove it. They gave him no help in advance. The buyer bought coffee, sugar or milk as he found it on sale in the stores. He carried the article home as it was delivered to him. If there he found the milk watered, the sugar sanded or the coffee composed in large part of burnt beans or something worse, he could complain, but in proving his complaint he was dependent on such evidence as he could himself command. If the new laws are vigorously enforced, attention will be drawn to all simple, trustworthy modes of detecting these commercial frauds."

"A company of well-known capitalists have organized the Gas Light Transportation Company to mine coal and manufacture gas in Pennsylvania and pipe the gas to Eastern cities. The coal can be bought at the mines for 55 cents per ton, and the gas companies in the cities pay $4.62 per ton. The cost of pumping the gas is offset by the value of the coke."

"'The Resemblances and Differences of the Two Sexes' is the title of a paper published in *Revue Scientifique* by M. G. Delaunay. It has special interest as presenting the most complete review of the differential characters of the sexes of the human species that has appeared. Delaunay's conclusions are as follows: 'The preeminence of the female sex, only seen in certain inferior races of men and in the infants of the higher races, marks an inferior stage of evolution. It is the same with the equality of the sexes, which is only seen in little-developed individuals, such as inferior races and species, young persons, the aged, and inferior classes of society. The preeminence of the male over the female represents a superior phase of evolution, since it characterizes superior species and races, adult age and the superior classes of society. From the moral as from the physical side evolution appears to have progressed from a state of superiority of the female to that of the male, and the stage of equality represents an intermediate stage.'"

"The gas engine is in increasing demand for driving dynamo-electric machinery. It is well known that as coal is burned for the generation of steam power there are two inevitable sources of loss. So much of the energy developed by combustion is used up in converting the water into steam that a theoretically perfect engine could not utilize more than one-fifth the total energy of the coal. Moreover, ordinary furnaces are ill-adapted for the economical execution of the two distinct processes that go on in them, namely the conversion of coal into gas and the simultaneous combustion of the gas as fuel. When these processes are separated, the gas is properly made and economically burned, but such improvement promises less in the way of increased economy than is secured by the direct development of power in gas engines by the burning of gas explosively. The problems raised and the economies promised by the extension of this method of generating power are well worthy of the attention of mechanical engineers and inventors."

Now that you're ready for a change of pace it's time to try John Jameson.

Take a sip of John Jameson. Note the light, delicate taste.
Luxurious and smooth as you would expect a premium whiskey to be.
But with a distinctive character all its own. Set a new pace
for yourself. Step ahead of the crowd with John Jameson, the
world's largest selling Irish Whiskey.

What if you chose as a technical

"At Sun, an HP computer system helps maximize production and minimize equipment failures."

Sun Production Company, a subsidiary of Sun Company, uses an HP 1000 computer system to control pump units and monitor production at its Eliasville, Texas, oilfield.

Marvin Boyd, District Production Manager, says, "We find the HP 1000 an effective management tool. By providing constant surveillance of our operation, it alerts us to malfunctions immediately and shuts units down before damage can occur.

"Because the HP system eliminates overpumping and speeds response time we can reduce operating costs and still achieve maximum production rates. Based on these results, we are now installing additional HP 1000s at Sun's Bennett Ranch and Levelland units."

HP can be your business computer partner, too!

A new kind of personal office computer, the HP 125 is specifically designed to help business people do their jobs better. Qualifications? Powerful management software, impressive graphics capabilities, communicates well with other computers, friendly and flexible. The HP 125 is supported by over 100 software programs and handles word processing just as easily as graphics. All this is supported by HP's service, reputation, and experience.

New dimensions in personal computation.

There's virtually no numerical problem in engineering or statistics beyond the scope of HP's new programmable 11C. Included in addition to the commoner functions are hyperbolics, permutations and combinations, random-number generation. You also acquire the ability to redefine five key functions making a grand total of 15 program labels you can call up at the touch of one key. Subroutines can be nested

© 1981 SCIENTIFIC AMERICAN, INC

THE AUTHORS

ELI GINZBERG ("The Social Security System") is director of the Conservation of Human Resources Project at Columbia University. His association with Columbia spans more than half a century. A native of New York City, he earned three degrees at Columbia: an A.B. in 1931, an A.M. in 1932 and a Ph.D. (in economics) in 1934. He joined the faculty in 1935 and was A. Barton Hepburn Professor of Economics at the university's Graduate School of Business from 1967 to 1979, when he retired to become emeritus professor and special lecturer. In addition to his academic responsibilities he has served in numerous capacities as a consultant to government officials and agencies. Ginzberg wishes to acknowledge the assistance of his colleagues Morton Miller, Daniel Saks, Steven Sandell and Anna B. Dutka in the preparation of this article.

DALE A. RUSSELL ("The Mass Extinctions of the Late Mesozoic") is curator of fossil vertebrates and chief of the paleobiology division of the Canadian National Museum of Natural Sciences in Ottawa. A graduate of the University of Oregon, he has an M.A. from the University of California at Berkeley and a Ph.D. from Columbia University. He spent a postdoctoral year as a National Science Foundation fellow at Yale University before taking up his present posts in 1965. He lists among his research interests "the ecology of terrestrial reptile faunas near the end of the age of the reptiles, the evolution of the brain in dinosaurs and trends in the evolution of multicellular organisms that are of possible relevance to the development of extraterrestrial ecosystems."

ISAAC F. SILVERA and JOOK WALRAVEN ("The Stabilization of Atomic Hydrogen") are experimental physicists on the faculty of the University of Amsterdam. Silvera was born and educated in the U.S., receiving his Ph.D. in physics from the University of California at Berkeley in 1965. After spending a year as a postdoctoral fellow at the University of Grenoble he worked for the Rockwell International Science Center in California until 1971, when he was appointed professor of experimental physics at the University of Amsterdam. Silvera's research focuses primarily on the application of spectroscopic techniques to the study of quantum-mechanical systems at low temperatures and high pressures. Walraven has been involved in the study of atomic hydrogen in the condensed state since 1972, first as an undergraduate and later as a graduate student employed by the Foundation for Fundamental Research on Matter, a Dutch research organization based in Utrecht. He joined the physics laboratory at Amsterdam in 1979.

LAURENCE A. SODERBLOM and TORRENCE V. JOHNSON ("The Moons of Saturn") are planetary scientists with a special interest in the satellites of the solar system. Soderblom is chief of the branch of astrogeologic studies of the U.S. Geological Survey. He majored in both geology and physics at the New Mexico Institute of Mining and Technology; his Ph.D., in planetary science and geophysics, is from the California Institute of Technology. Since joining the Geological Survey in 1970 he has taken part in a number of unmanned planetary-exploration missions launched by the National Aeronautics and Space Administration, including the Voyager missions to the outer planets. Soderblom was the author of "The Galilean Moons of Jupiter," which appeared in the January 1980 issue of SCIENTIFIC AMERICAN. Johnson teaches planetary science at Cal Tech and works as a senior research scientist at the Jet Propulsion Laboratory. A graduate of Washington University, he received his Ph.D. in planetary science from Cal Tech in 1970. He is currently a member of the Voyager imaging-science team and a project scientist for the Galileo mission to Jupiter.

RONALD D. LEVINE ("Supercomputers") is employed by Technology Development of California, Inc., where he works on several aspects of large-scale scientific computing under contract to the National Aeronautics and Space Administration. His degrees are from the University of California at Berkeley: an A.B. and an M.A. (both in physics) in 1960 and 1964 respectively, and a Ph.D. (in mathematics) in 1970. He has also studied at Stanford University, Cornell University and the University of Cambridge. During the 1960's he worked at the Lawrence Berkeley Laboratory, the Stanford Linear Accelerator Center and the Jet Propulsion Laboratory of the California Institute of Technology. He spent most of the next decade on the faculty of Humboldt State University, where he served a term as chairman of the mathematics department. Levine writes: "Over the past 20 years my scientific interests have ranged widely, from relativity theory to mathematical biology, with computational problems a common thread linking these diverse interests."

GUNTHER S. STENT and DAVID A. WEISBLAT ("The Development of a Simple Nervous System") work at the University of California at Berkeley, where Stent is chairman of the department of molecular biology and director of the university's Virus Laboratory and Weisblat is a postdoctoral research associate. Stent was born in Berlin and came to the U.S. in 1940. He got his B.S. and Ph.D. in physical chemistry from the University of Illinois and did his postdoctoral work ("in the discipline that would eventually become known as molecular biology") at the California Institute of Technology and the Pasteur Institute. He moved to Berkeley in 1952. His research on the neurobiology of leeches dates from 1969, when he spent a sabbatical leave at the Harvard Medical School working with the late Stephen W. Kuffler and John G. Nicholls. Weisblat has an A.B. from Harvard College and a Ph.D. in chemistry and neurophysiology from Cal Tech. He has been collaborating with Stent at Berkeley since 1976.

PAUL BERNARD ("An Ancient Greek City in Central Asia") is professor of Greek archaeology at the École Pratique des Hautes Études in Paris. An alumnus of the École Normale Supérieure, he got his start as an excavator of ancient sites at Thasos in Greece as a member of the French Archaeological School at Athens. He later became interested in the study of the relations between ancient Greece and Persia as a member of the French Institute of Archaeology in Beirut. In 1965 he was appointed director of the French archaeological delegation in Afghanistan, where he headed up the work described in his article. Bernard was a visiting member at the Institute for Advanced Study in Princeton during the academic year 1972–73.

DANIEL KAHNEMAN and AMOS TVERSKY ("The Psychology of Preferences") began their collaboration in 1969 while they were both members of the faculty at the Hebrew University of Jerusalem. Both left Israel in 1978 to take positions at academic institutions in North America, Kahneman as professor of psychology at the University of British Columbia and Tversky as professor of psychology at Stanford University. Kahneman, who was born in Israel, was graduated with a Ph.D. in psychology from the University of California at Berkeley in 1961. In addition to his years at Hebrew University he has taught and done research at the University of Michigan, Harvard University, the Oregon Research Institute, the Applied Psychological Research Unit in Cambridge, England, and the Center for Advanced Study in the Behavioral Sciences at Stanford. Tversky got his Ph.D. from the University of Michigan in 1965. He has been coauthor and editor of several books, including *Mathematical Psychology, Foundations of Measurement, Decision Making* and *Judgement under Uncertainty*.

"YOU CAN ALWAYS TRUST THE TASTE OF A BOLLA"

—Franco Bolla

"Why do so many people prefer the Bolla family of great wines?

"Taste. Not just exceptional taste, but exceptional taste, bottle after bottle. So whether you choose the full-flavored Valpolicella, the mellow Bardolino, the crisp, refreshing Soave or the light, dry Trebbiano, you won't be disappointed. No matter how many times you choose it. After all, when you're a Bolla, you don't let your friends down."

**FROM THE BOLLA
FAMILY OF GREAT WINES.**
IMPORTED BY THE JOS. GARNEAU CO., N.Y., N.Y.
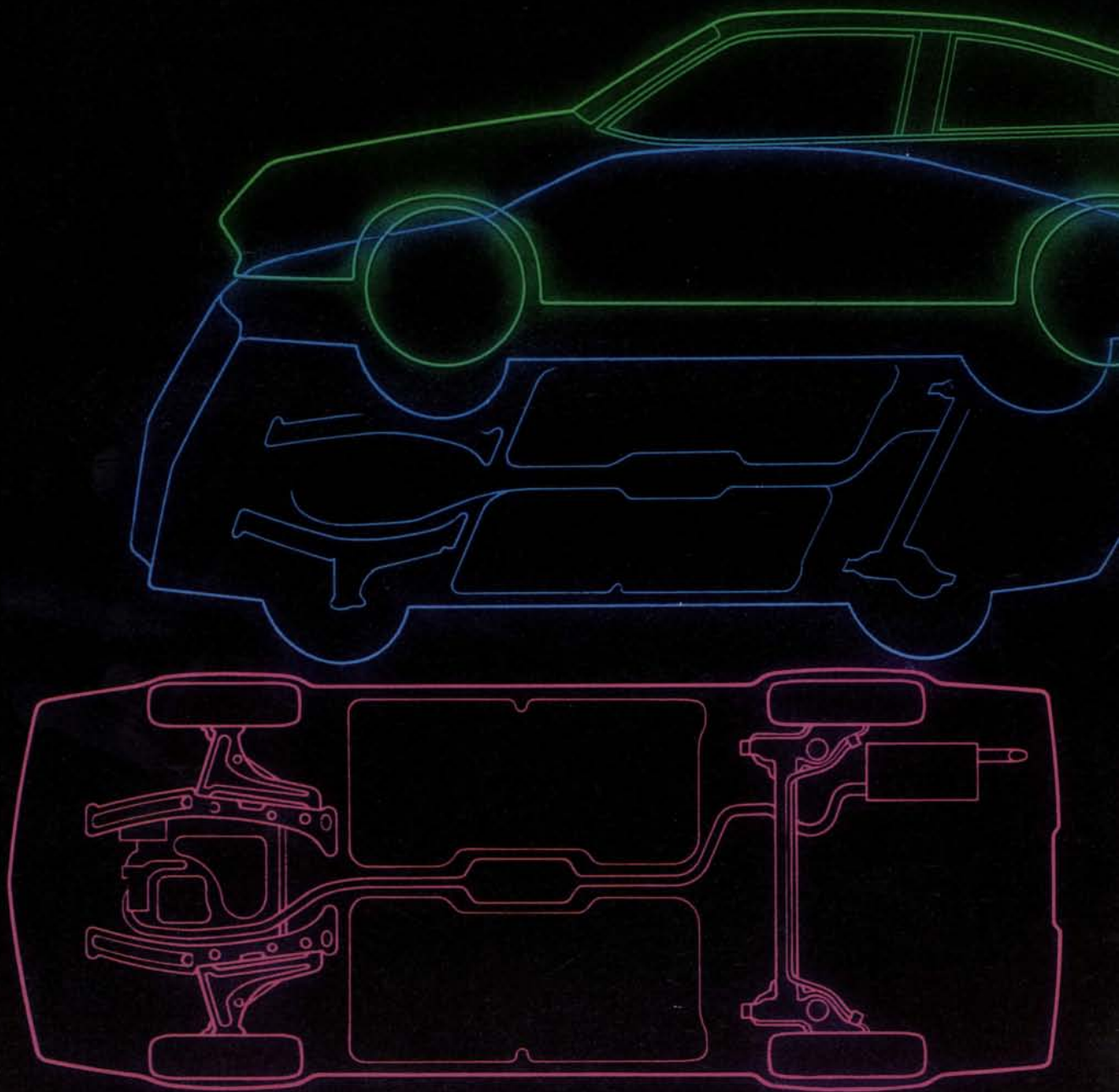
Technology for the side of the car you don't see.

It isn't brightly painted. It isn't pretty. The other side of the metal. You rarely see it. Yet it's a very important part of a car's body. It's where moisture can collect and cause rust. It's also where we've invested a lot of money and technology.

These surfaces, such as the inside of the doors and fenders, the trunk floor, and the rear quarter panels, aren't treated the way they used to be. We're treating them in an entirely different way. Many are now made from zinc-coated steel, while others are coated with a special corrosion-resistant substance developed from a petroleum-based compound. And a new plastisol spray protects the car on the lower side areas closest to the ground, where pebbles and stones can cause the most paint damage.

This isn't flashy technology. In fact, its only purpose is to help prevent rust.

That's our idea of how to use technology to build cars and trucks. Attention to details where you don't see them, as well as where you do. Appearance and comfort may sometimes sell a car, but today's customers demand real value.

Our goal, as the world's largest automotive manufacturer, is to maintain our lead by using new technology to build cars and trucks that perform better and last longer, with lower maintenance costs, than those built by any competitor—American or foreign.

General Motors
The future of transportation is here.

GM

METAMAGICAL THEMAS

A self-referential column about last January's column about self-reference

by Douglas R. Hofstadter

As this is the January issue, I thought I would follow up on my column of last January on self-reference, and that is what this column is about. I should like, however, to take advantage of this opening paragraph to warn those readers who are not amused by self-referential material that they will probably want to quit reading before they reach the end of this paragraph, or for that matter this sentence, in fact this clause, even this noun phrase—in short, this.

Since my column of last January appeared I have received a hefty pile of self-referential mail. Tony Durham astutely surmised: "What with the likely volume of replies, I should not think you are reading this in person." John C. Waugh's letter yelped: "Help, I'm buried under an avalanche of readers' responses!" His letter was rescued and now rests in a much reduced pile. Here I shall present some of my favorites in that collection.

Before leaving the topic of mail I should like to point out that the postmark on Ivan Vince's postcard from Britain cryptically stated, "Be properly addressed." Was this an order issued by the post office to the postcard itself? If it was, then British postcards must be more intelligent than American ones; I have yet to meet a postcard that could read, let alone correct its own address. (One postcard that reached me was addressed to me in care of the magazine *Omni*!)

I was flattered by a couple of self-undermining compliments. Richard Ruttan wrote, "I just can't tell you how much I enjoyed your first article," and John Collins said, "This does not communicate my delight at January's column." I was also pleased to learn that my fame had spread as far as the men's room of the Tufts University philosophy department, where Dan Dennett found the following scribbled on the wall: "This sentence is graffiti—Douglas R. Hofstadter."

A popular pastime of my correspondents was the search for interesting self-answering questions. Only a few succeeded in "jootsing" (jumping out of the system), which to me is the same as being truly novel. It seems that success in this limited art form is not easy to come by. John Flagg cynically remarked (I paraphrase slightly): "Ask a self-answering question, and get a self-questioning answer." One of my favorites was given by Henry Taves: "I fondly remember a history exam I encountered in boarding school that contained the following: 'IV. Write a question suitable for a final exam in this course, and then answer it.' My response was simply to copy that sentence twice." I was delighted by this. On reflection, however, I began to suspect something was slightly wrong here. What do you think?

Richard Showstack contributed two droll self-answering questions: "What question no verb?" and "What is a question that mentions the word 'umbrella' for no apparent reason?" Jim Shiley sent in a clever entry that I modify slightly into "Is this a rhetorical question, or is this a rhetorical question?" He also contributed the following: "Take a blank sheet of paper and on it write: 'How far across the page will this sentence run?' Now if some polyglot friend of yours points out that the same string of phonemes in Ural-Altaic means 'six inches,' send me a free subscription to *Scientific American*. Otherwise, if the inscription of a question counts both as the question and as a unit of measure, I at least get a booby prize. But I think somehow I bent the rules."

My own solutions to the problem of the self-answering question are not so much self-*answering* as self-*provoking* questions, such as the following one: "Why are you asking me *that* out of the blue?" It is obvious that when the question is asked out of the blue it might well elicit an identical response, indicating the hearer's bewilderment. Yes, but what prompts *this* query?

Philip Cohen relayed the following anecdote about a self-answering question, from Damon Knight: "Terry Carr, an old friend, sent us a riddle on a postcard, then the answer on another postcard. Then he sent us another riddle, 'How do you keep a turkey in suspense?' and he never sent the answer. After about two weeks we realized that *was* the answer."

Several of the real masterpieces sent in belong to what I call the self-documenting category, of which a simple example is Jonathan Post's "This sentence contains ten words, eighteen syllables and sixty-four letters." A neat twist is supplied by John Atkins in his sentence "'Has eighteen letters' does." The self-documenting form can get much more convoluted and introspective. An example by the wordplay master Howard Bergerson was brought to my attention by Philip Cohen. It goes: "In this sentence the word AND occurs twice, the word EIGHT occurs twice, the word FOUR occurs twice, the word FOURTEEN occurs four times, the word IN occurs twice, the word OCCURS occurs fourteen times, the word SENTENCE occurs twice, the word SEVEN occurs twice, the word THE occurs fourteen times, the word THIS occurs twice, the word TIMES occurs seven times, the word TWICE occurs eight times and the word WORD occurs fourteen times."

That is good, but the gold medal in the category is reserved for Lee Sallows, who submitted the following tour de force: "Only the fool would take trouble to verify that his sentence was composed of ten a's, three b's, four c's, four d's, forty-six e's, sixteen f's, four g's, thirteen h's, fifteen i's, two k's, nine l's, four m's, twenty-five n's, twenty-four o's, five p's, sixteen r's, forty-one s's, thirty-seven t's, ten u's, eight v's, eight w's, four x's, eleven y's, twenty-seven commas, twenty-three apostrophes, seven hyphens and, last but not least, a single !"

I (perhaps the fool) did take the trouble to verify the entire thing. First, however, I did make some spot checks. And I must say that when the first random spot check worked (I think I checked the number of g's), it had a strong psychological effect: all of a sudden the credibility rating of the sentence shot way up for me. It strikes me as weird (and wonderful) how in certain situations the verification of a tiny percentage of a theory can serve to powerfully strengthen your belief in the full theory. Perhaps that is the whole point of the sentence!

The noted logician Raphael Robinson submitted a playful puzzle in the self-documenting line. The reader is asked to complete the following sentence: "In this sentence the number of occurrences of 0 is _____, of 1 is _____, of 2 is _____, of 3 is _____, of 4 is _____, of 5 is _____, of 6 is _____, of 7 is _____, of 8 is _____ and of 9 is _____." Each blank is to be filled with a numeral of one or more digits, written in decimal notation. There are exactly two solutions. Readers might also search for two

THE CITI OF TOMORROW

It's the only banking system that brings you the past 45 days at the touch of a button.

The Citi of Tomorrow gives you the past. Like no other bank can.

An information retrieval system that is unrivaled in bringing you the data you need to evaluate your funds position today. And forecast your financial needs tomorrow.

Citibank's retrieval system reports on your financial transactions as far back as 45 business days— more than any other bank.

You can retrieve your balances by date. By account. By total debits and credits per account. By means of summaries of activities for a given period. You can also receive reports on only those activities that are most relevant to you.

The Citi of Tomorrow. One reason it's a pre-eminent system today is that it appreciates the importance of yesterday.

For more information, call your account manager or Robert Mendes, V.P., at (212) 559-1980.

CITIBANK ⊕ ®
GLOBAL ELECTRONIC BANKING

The Citi of Tomorrow and Global Electronic Banking are service marks of Citibank, N.A. ©1981 Citibank, N.A. Member FDIC

sentences that document each other, or even longer loops of that kind.

Clearly the ultimate in self-documentation would be more than a sentence that merely inventoried its parts; it would be a sentence that included a rule as well, telling all the king's men how to put those parts back together again to create the full sentence—in short, a self-reproducing sentence. Such a sentence is Willard Van Orman Quine's English rendition of Kurt Gödel's classic metamathematical homage to Epimenides the Cretan:

"yields falsehood when appended to its quotation" yields falsehood when appended to its quotation.

Quine's sentence in effect tells the reader how to construct a replica of the sentence being read, and then (just for good measure) adds that the replica (not *itself*, for heaven's sake!) asserts a falsity! It is reminiscent of the famous remark made by Epilopsides the Concretan (a second cousin of Epimenides') to Flora, a beautiful young woman whose ardent love he could not return (he was betrothed to her twin sister Fauna): "Take heart, my dear. I have a suggestion that may cheer you up. Just take one of these cells from my arm and clone it. You'll wind up with a fellow who looks and thinks just like me. But do watch out for him—he is given to telling beautiful women real whoppers!"

In the 1940's John von Neumann worked hard trying to design a machine that could build a replica of itself out of raw materials. He came up with a theoretical design consisting of hundreds of thousands of parts. Seen in hindsight and with a considerable degree of abstraction, the idea behind von Neumann's self-reproducing machine turns out to be pretty similar to the means by which DNA replicates itself. And this in turn is close to Gödel's method of constructing a self-referential sentence in a mathematical language in which at first there seems to be no way of referring to the language itself.

The first every-other-decade Von Neumann Challenge is thus hereby presented for ambitious readers: Create a comprehensible and not unreasonably long self-documenting sentence that not only lists its parts (at either the word level or the letter level) but also tells how to put them together so that the sentence reconstitutes itself. (Notice, by the way, the requirement is that the sentence be "not unreasonably long," which is different—very different—from being "reasonably long.") The parts list (or "seed") should be an inventory of words or typographical symbols, more or less as in the sentences created by Howard Bergerson and Lee Sallows. The inventoried symbols should in some way be clearly distinguishable from the text that refers to them. For instance,

they could be enclosed in quotation marks or printed in another typeface or referred to by name. It is not so important what convention is adopted, as long as the distinction is sharp. The rest of the sentence (the "building rule") should be printed normally, since it is to be regarded not as typographical raw material but as a set of instructions. This is the use-mention distinction I discussed last January, and to disregard it is a serious conceptual weakness. (It is a flaw in Sallows' sentence that slightly tarnishes the gold on his medal.)

The building rule may not refer to normally printed material, only to parts of the inventory. Hence it is not permitted for the building rule to refer to itself in any way! The building rule has to describe structure explicitly. Furthermore (and this is the subtlest and probably the most often overlooked aspect of self-reference), the building rule must specify which parts are to be printed normally and which parts in quotes (or however the raw materials are being indicated). In this respect Bergerson's sentence fails. Although to its credit it sharply distinguishes between use and mention by relying on uppercase for the names of inventory items and lowercase for item counts and filler words, it does not have separate inventories for items in uppercase and lowercase. Instead it lumps the two together, losing a vital distinction.

In the Von Neumann Challenge extra points will be awarded for solutions given in Basic English, or whose seed is entirely at the letter level (as in Sallows' sentence). The Quine sentence, although it clearly incorporates a seed (the seven-word phrase in quotation marks) and a building rule (that of appending something to its quotation), is not a legal entry because its seed is too far from being raw material.

There is a very good reason, by the way, for the seed of Quine's sentence to be so complicated—in fact, for it to be identical with the building rule, except for the quotation marks. The reason is simple to state: You have to *build the building rule* out of raw materials, and the more the building rule looks like the seed the simpler it will be to build it from the seed. To make a full new sentence all you need to do is make two copies of the seed, carry out whatever simple manipulations will convert one copy of the seed into the building rule, and then splice the other copy of the seed to the newly minted building rule to make up a complete new sentence, fresh off the assembly line.

To make this clearer it will be helpful to give a slight variation on Quine's sentence. Imagine that you recognized only lowercase roman letters and that uppercase letters were alien to you. Then text printed in uppercase would be for all

practical purposes devoid of meaning or interest, whereas text in lowercase would be full of meaning and interest, able to suggest ideas or actions. Now suppose someone gave you a conversion table that matched each uppercase letter with its lowercase counterpart, so that you could "decode" uppercase text. Then one day you came across this piece of meaningless uppercase text:

YIELDS A FALSEHOOD WHEN USED AS THE SUBJECT OF ITS LOWERCASE VERSION

On being decoded the text would yield a lowercase sentence, or rather, a lowercase sentence fragment—a predicate without a subject. Suggestive, eh?

This notion of two parallel alphabets, one in which text is inert and meaningless and the other in which text is active and meaningful, may strike you as yielding no more than a minor variation on Quine's sentence, but in fact it is very similar to an exceedingly clever trick that nature discovered and has exploited in every cell of every living organism. Our seed—our DNA—is an enormous book of inert text written in a chemical alphabet that has 64 "uppercase" letters (codons). Our building rules—our enzymes—are short, pithy slogans of active text written in a different chemical alphabet that has just 20 "lowercase" letters (amino acids). There is a map (the genetic code) that converts uppercase letters into lowercase ones. Obviously some lowercase letters must correspond to more than one uppercase letter, but here that is a detail. It also turns out that three characters of the uppercase alphabet are not letters but punctuation marks telling where one pithy slogan ends and the next one begins, but again these are details. Once you know the mapping you often will not even remember to distinguish between the two chemical alphabets: the inert uppercase codon alphabet and the active lowercase amino acid alphabet. The main thing is that, armed with the genetic code, you can read the DNA book (seed) as if it were a sequence of enzyme slogans (building rules) telling how to write a new DNA book and a new set of enzyme slogans. It is a perfect parallel to our variation on the Quine sentence, where inert, uppercase seed text was converted into active, lowercase rule text that told how to make a copy of the full Quine sentence given its seed.

A cell's DNA and enzymes act like the seed and building rules of Quine's sentence, or the parts list and building rules of von Neumann's self-reproducing automaton, or then again like the seed and building rules of computer programs that print themselves out. It is amazing how universal the mechanism of self-reference is, and for that reason I always find it quaint that people who inveigh against the supposed silliness of self-ref-

The Taste of the Renaissance lives on

·AMARETTO DI SARONNO·ORIGINALE·1525·

"My own IBM computer. Imagine that."

One nice thing about having your own IBM Personal Computer is that it's *yours*. For your business, your project, your department, your class, your family and, indeed, for yourself.

Of course, you might have thought owning a computer was too expensive. But now you can relax.

The IBM Personal Computer starts at less than $1,600[†] for a system that, with the addition of one simple device, hooks up to your home TV and uses your audio cassette recorder.

You might also have thought running a computer was too difficult. But you can relax again.

Getting started is easier than you might think, because IBM has structured the learning process for you. Our literature is in *your* language, not in "computerese." Our software *involves* you, the system *interacts* with you as if it was made to—and it was.

That's why you can be running programs in just one day. Maybe even writing your *own* programs in a matter of weeks.

For ease of use, flexibility and performance, no other personal computer offers as many advanced capabilities. (See the box.)

But what makes the IBM Personal Computer a truly useful tool are software programs selected by IBM's Personal Computer Software Publishing Department. You can have programs in business, professional, word processing, computer language, personal and entertainment categories.

You can see the system and the software in action at any ComputerLand® store or Sears Business Systems Center. Or try it out at one of our IBM Product Centers. The IBM Data Processing Division will serve those customers who want to purchase in quantity.

Your IBM Personal Computer. Once you start working with it, you'll discover more than the answers and solutions you seek: you'll discover that getting there is half the fun. Imagine that. **IBM**®

IBM PERSONAL COMPUTER SPECIFICATIONS
*ADVANCED FEATURES FOR PERSONAL COMPUTERS

User Memory
16K - 256K bytes*

Permanent Memory
(ROM) 40K bytes*

Microprocessor
High speed, 8088*

Auxiliary Memory
2 optional internal diskette drives, 5¼", 160K bytes per diskette

Keyboard
83 keys, 6 ft. cord attaches to system unit*
10 function keys*
10-key numeric pad
Tactile feedback*

Display Screen
High-resolution (720h x 350v)*
80 characters x 25 lines
Upper and lower case
Green phosphor screen*

Diagnostics
Power-on self testing*
Parity checking*

Languages
BASIC, Pascal

Printer
Bidirectional*
80 characters/second
12 character styles, up to 132 characters/line*
9 x 9 character matrix*

Color/Graphics
Text mode:
16 colors*
256 characters and symbols in ROM*
Graphics mode:
4-color resolution: 320h x 200v*
Black & white resolution: 640h x 200v*
Simultaneous graphics & text capability*

Communications
RS-232-C interface
Asynchronous (start/stop) protocol
Up to 9600 bits per second

The IBM Personal Computer _____ and me.

erence are themselves composed of trillions of self-referential molecules.

Scott Kim and I discovered a strange pair of sentences:

"The following sentence is totally identical with this one except that the words 'following' and 'preceding' have been exchanged, as have the words 'except' and 'in' and the phrases 'identical with' and 'different from.'"

"The preceding sentence is totally different from this one in that the words 'preceding' and 'following' have been exchanged, as have the words 'in' and 'except' and the phrases 'different from' and 'identical with.'"

At first glance these sentences are reminiscent of a two-step variant on the Epimenides paradox ("The following sentence is true"; "The preceding sentence is false"). At second glance, however, they are seen to say exactly the same thing as each other. Don Byrd disagrees with me; he maintains they say totally different things.

Not surprisingly, several of the sentences submitted by readers had a paradoxical flavor. Some were variants on Russell's paradox about the barber who shaves all those who do not shave themselves, or the set of all sets that do not include themselves as elements. For instance, Gerald Hull concocted this strange sentence: "This sentence refers to every sentence that does not refer to itself." Is Hull's concoction self-referential or is it not? In a similar vein Michael Gardner cited a Reed College senior thesis whose dedication ran: "This thesis is dedicated to all those who did not dedicate their theses to themselves." The book *Model Theory,* by C. C. Chang and H. J. Keisler, bears a similar dedication, as Charles Brenner pointed out to me. He also suggested another variant on Russell's paradox: Write a computer program that prints out a list of all programs that do not ever print themselves out. The question is, of course, will this program ever print itself out?

One of the most disorienting sentences came from Robert Boeninger: "This sentence does in fact not have the property it claims not to have." Got that? The problem, of course, is to figure out just what property it is that the sentence claims it lacks.

The Dutch mathematician Hans Freudenthal sent along a paradoxical anecdote based on self-reference:

"There is a story by the 18th-century German Christian Gellert called 'Der Bauer und sein Sohn' ('The Peasant and His Son'). One day during a walk, when the son tells a big lie, his father direly warns him about the liars' bridge, which they are approaching. This bridge always collapses when a liar walks across it. After hearing this frightening warning, the boy admits his lie and confesses the truth.

"When I [Freudenthal] told a 10-year-old boy this story, he asked me what happened when they eventually came to the bridge. I replied, 'It collapsed under the father, who had lied since in fact there is no liars' bridge.' (Or did it?)"

C. W. Smith, writing from London, Ont., described a situation reminiscent of the Epimenides paradox:

"During the 1960's, standing alone in the midst of a weed-strewn field in this city there was a weathered sign that read, '$25 reward for information leading to the arrest and conviction of anyone removing this sign.' For whatever it's worth, the sign has long since disappeared. And so, for that matter, has the field."

Incidentally, the Epimenides paradox should not be confused with the Nixonides paradox, first uttered by Nixonides the Cretin in A.D. 1974: "This statement is inoperative." Speaking of Epimenides, one of the most elegant variations on his paradox is the "Errata" section in a hypothetical book described by Beverley Rowe. It looks like this:

(vi)

Errata

Page (vi): For "Errata" read "Erratum"

Closely related to the truly paradoxical sentences are those that belong to what I call the "neurotic" and "healthy" categories. A healthy sentence is one that, so to speak, practices what it preaches, whereas a neurotic sentence is one that says one thing while doing the opposite. Alan Auerbach has given us a good example in each category. His healthy sentence is "Terse!" His neurotic sentence is "Proper writing—and you've heard this a million times—avoids exaggeration." Here is a healthy sentence from Brad Shelton: "Fourscore and seven words ago this sentence hadn't started yet." One of the jootsingest of sentences comes from Carl Bender: "The following sentence is written in Thailand, on"

Consider a related sentence sent by David Stork: "It goes without saying that...." To which category does it belong? Perhaps it is a psychotic sentence. Pete Maclean contributed a puzzling one: "If the meanings of 'true' and 'false' were switched, then this sentence would not be false." I still cannot figure out what it means! Dan Krimm wrote to tell me: "I've heard that this sentence is a rumor." Linda Simonetti contributed the following: "Which is not a complete sentence, but merely a subordinate clause." Douglas Wolfe offered the following neurotic rule of thumb: "Never use the imperative, and it is also never proper to construct a sentence using mixed moods." David Moser reminded

me of a slogan *National Lampoon* once used: "So funny it sells without a slogan!" Perry Weddle wrote, "I'm trying to teach my parrot to say, 'I don't understand a thing I say.' When *I* say it, it's viciously self-referential, but in *his* case?" Stephen Coombs pointed out that "a sentence may self-refer in the verb." My mother, Nancy Hofstadter, heard Secretary of State Haig describe a warning message to the Russians as "a calculated ambiguity that would be clearly understood." Yes, sir!

Jim Propp submitted a sequence of sentences that slide from the neurotically healthy to the healthily neurotic:

"This sentence every third, but it still comprehensible."

"This would easier understand fewer had omitted."

"This impossible except context."

"4′33″ attempt idea."

" "

The penultimate sentence refers to John Cage's famous piece of piano music consisting of four minutes and 33 seconds of silence. The last sentence might well be an excerpt from *The Wit and Wisdom of Spiro T. Agnew,* although it is too short for one to be sure. Propp also sends along the following healthy quotation from David Premack in *Intelligence in Ape and Man:* "By the 'productivity' of language, I mean the ability of language to introduce new words in terms of old ones."

The philosopher Howard DeLong contributed what might be considered a neurotic syllogism: "All invalid syllogisms break at least one rule. This syllogism breaks at least one rule. Therefore this syllogism is an invalid syllogism."

Several readers pointed out phrases and jokes that have been making the rounds. D. A. Treissman, for instance, reminded me that "nostalgia ain't what it used to be." Henry Taves mentioned the delightful T-shirts adorned with statements such as "My folks went to Florida and all they brought back for me was this lousy T-shirt!" And John Fletcher described an episode of the television program "Laugh-In" a few years ago on which Joanne Worley sang, "I'm just a girl who can't say 'n...,' 'n...,' 'n....'" John Healy wrote, "I used to think I was indecisive, but now I'm not so sure."

I myself have a few contributions to this collection. A neurotic one is "In this sentence the concluding three words 'were left out.'" Or is it neurotic? These things confuse me! In any case an entirely healthy sentence is "This sentence offers its reader(s) various alternatives/options that he or she (or they) is (are) free to accept and/or reject." And then there is the inevitable "This sentence is neurotic." The thing is, if it *is* neurotic, it practices what it preaches, so that it is healthy and *cannot* be neurotic, but then

21

Who will be first to electronically forecast the future?

It could be you and Hughes Support Systems.

In fact, a big part of our business is to simulate tomorrow's possibilities today — and in a variety of ways.

Our logistics support, our computerized training systems, our electronic simulators are all firsts that in a specialized way forecast, plan or prepare for almost every kind of crisis.

So that everyone — from pilots in hazardous skies, to troops at the front, to the technicians who support them — is ready for the future when it arrives.

With Hughes Support Systems, you could be part of state-of-the-art electronics here at home, or part of a field support team that's swept to almost any site on earth. You could be involved in an almost unlimited array of support programs, systems and devices that affect defense avionics, missiles, and electro-optics — and the people who depend on them throughout the free world.

And it's all part of Hughes, with 1,500 projects and a backlog of over $6 billion.

Who will be first with the electronics and support of tomorrow? Who will be first to teach the technology of the future?

It could be you and Hughes.

At Hughes Support Systems, we'll introduce you to people, ideas and jobs that could change your world. And maybe ours.

It could be you and Hughes Support Systems
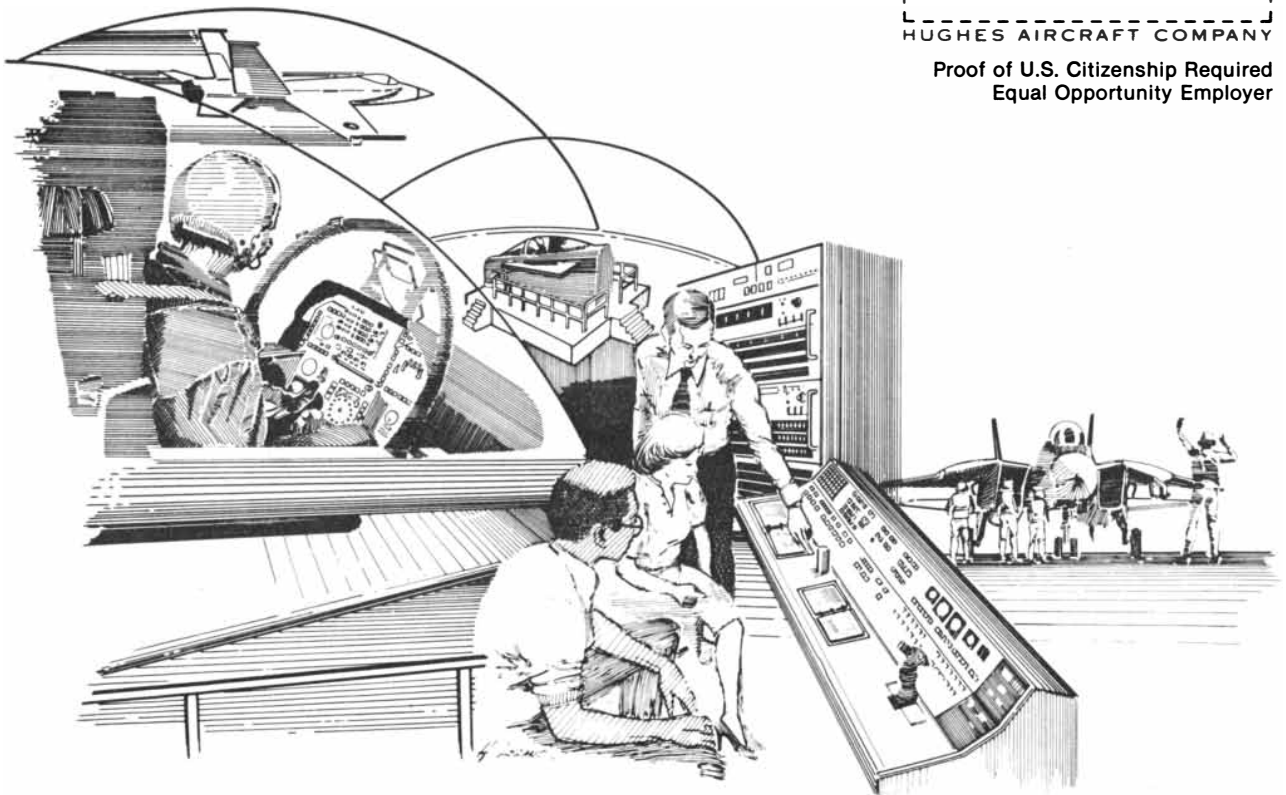
THE QUESTAR® 7 PHOTOGRAPHS OMEGA CENTAURI

This remarkable photograph of Omega Centauri was taken at Apache Pass, Arizona, by Hubert Entrop. He writes us "The wind blew from the west in strong gusts but I located in a low north-south arroyo beside a large bush to protect the scope. The atmosphere was miserably rough but in spite of it, it's a good Omega Centauri. Imagine what it would be like if we could have it straight overhead instead of so low on the horizon. Exposure 1 hour 30 minutes on Tri-X."

If you come past Questar these days you will see the newest feature on our landscape—the Observa-Dome, which we are now privileged to offer to our customers in a variety of sizes. It is equipped with the new Questar Telescope Mount which accommodates our Questar 12 and is engineered to support a telescope as large as 20 inches. The design of the mount is an adaptation of the German equatorial, with special Questar features that contribute to the mechanical perfection for which Questar products are noted. Unlike some recently introduced mounts, it has a full 360° continuous following capability, with a smoothness of operation that must be experienced to be believed.

Also at Questar, if you have an interest in surveillance or special tracking applications, you will see our patented 40-120 on display. This unique instrument establishes prime focus at both 40 and 120 inches (1000 and 3000 mm.) It resolves 100 lines per millimeter at the lower focal length and at least 55 lines per millimeter at the upper; one can move in a few seconds between the two and since the shift is managed by internal optical change the barrel length remains at a constant 30 inches. It weighs only 40 pounds.

In many ways the Questar 40-120 is the most sophisticated of the Questar instruments. Its size and weight make it ideal for a variety of uses where the observer must be at a great distance from the area or activity under scrutiny, while the dual focal length is particularly important for objects in motion.

Literature on the Questar 40-120 and on the Observa-Dome is available on request. • • •

A convenient accessory for taking deep sky photographs is an auxiliary guiding system, the Questar Starguide. It consists of a Tracker and Declination Vernier Drive. The Tracker intercepts light from a guide star and delivers it to the guiding eyepiece, and the Drive permits corrections on a 10 to 1 ratio over the existing, extremely accurate, Questar drive. The eyepiece can be swiveled 360° for comfort in guiding and is completely independent of the camera position.

©Questar Corporation 1981

© 1981 SCIENTIFIC AMERICAN, INC

---

if it is *not* neurotic, it is the opposite of what it says it is, so that it *has* to be neurotic. No wonder it is neurotic!

Speaking of neurotic sentences, what about sentences with identity crises? To me these are the most interesting ones of all. A typical example is Dan Krimm's vaguely apprehensive question, "If I stated something else, would it still be me?" I thought this could be worded better, and so I revised it slightly as follows: "If I said something else, would it still be me saying it?" I still was not happy, and so I wrote one more version: "In another world could I have been a sentence about Humphrey Bogart?" When I paused to reflect on what I had done, I realized that in reworking Krimm's sentence I had tampered with its identity in the very way it feared. The question remained: Were all these variants really the same sentence deep down? My last experiment along these lines was "In another world could this sentence have been Dan Krimm's sentence?"

Clearly some readers were thinking along parallel lines. John Atkins asked, "Can anyone explain why this would still be the same magazine without this query, and yet this would not be the same query without this word?" And Loul McIntosh, who works at a rehabilitation center for formerly schizophrenic patients, had a question connecting personal identity with self-referential sentences: "If I were you, who would be reading this sentence?" She then added: "(N.B. That's what I get for working with schizophrenics.)" This brings me to Peter M. Brigham, M.D., who in his work ran across a case of literary schizophrenia: "You have, of course, just begun reading the sentence that you have just finished reading."

Pursuing the slithery snake of self in his own way, Uilliam M. Bricken, Jr., wrote in: "If you think this sentence is confusing, then change one pig." Now, anyone can see that this does not make any sense at all. Surely what he meant was, "If you think this sentence is confusing, then roast one pig"—don't ewe agree? By the by, if ewe think "Uilliam" is confusing, then roast one ewe. And speaking of ewes, what's a nice word like "ewe" doing in a foxy paragraph like this?

A while back I heard a radio show about pets on which the announcer mused, "If a dog had written this broadcast, he might have said that people are inferior because they don't wag their tails." This gave me paws for thought: What might this column have been like if it had been written by a dog? I cannot say for *sure*, but I have a hunch it would have been about chasing squirrels. And it might have had a paragraph speculating about what this column would have been like if it had been written by a squirrel.

I think my favorite of all the sent-

The most comprehensive and useful computer reference in the world.

25

in-ces was one contributed by Harold Cooper. He was inspired by my counterfactual self-referential question: "What would this sentence be like if π were 3?" Here is his answer. For me it exemplifies the meaning of the verb "joots."

> If pi were 3, this sentence would look something like this.

The six-sided *O*'s represent the fact that the ratio of the circumference to the diameter of a hexagon is 3. Clearly, in Cooper's mind, if π were 3, what conclusion would be more natural than that *circles would be hexagons!* Who could ever think otherwise? I was intrigued by the fact that as π's value slipped to 3 not only did circles turn into hexagons but also the interrogative mood slipped into the declarative mood. Remember that the question asked how the question itself would be in that strange subjunctive world. Would it lose its curiosity about itself and cease to be a question? I did not see why that personality trait of the sentence would be affected by the value of π. On the other hand, it seemed obvious to me that if π *were* 3, the antecedent of the conditional should no longer be subjunctive. In fact, rather than saying "if π *were* 3," it should say "because π *is* 3" (or something to that effect). Putting my thoughts together, then, I came up with a slight variation on Cooper's sentence: "What is this sentence like, π being 3 (as usual)?"

Several readers were interested in sentences that refer to the language they are in (or not in, as the case may be). An example is "If you spoke English, you'd be in your home language now." Jim Propp sent in a delightful pair of such sentences that need to be read together:

"Cette phrase se refere a elle-meme, mais d'une maniere peu evidente a la plupart des Americains."

"Plim glorkle pegram ut replat, trull gen ris clanter froat veb nup lamerack gla smurp Earthlings."

If you do not understand the first sentence, just get a Martian friend to help you decode the second one. That will provide hints about the first. [We apologize for leaving off the proper Martian accent marks, but they were not available in this typeface.]

Last January I published several sentences created by David Moser and mentioned that he had written an entire story consisting of self-referential sentences. Many readers were intrigued. I decided there could be no better way to conclude this column than to print David's story in its entirety. Here it is:

This Is the Title of This Story, Which Is Also Found Several Times in the Story Itself

This is the first sentence of this story. This is the second sentence. This is the title of this story, which is also found several times in the story itself. This sentence is questioning the intrinsic value of the first two sentences. This sentence is to inform you, in case you haven't already realized it, that this is a self-referential story, that is, a story containing sentences that refer to their own structure and function. This is a sentence that provides an ending to the first paragraph.

This is the first sentence of a new paragraph in a self-referential story. This sentence is introducing you to the protagonist of the story, a young boy named Billy. This sentence is telling you that Billy is blond and blue-eyed and American and 12 years old and strangling his mother. This sentence comments on the awkward nature of the self-referential narrative form while recognizing the strange and playful detachment it affords the writer. As if illustrating the point made by the last sentence, this sentence reminds us, with no trace of facetiousness, that children are a precious gift from God and that the world is a better place when graced by the unique joys and delights they bring to it.

This sentence describes Billy's mother's eyes bulging and tongue protruding and makes reference to the unpleasant choking and gagging noises she's making. This sentence makes the observation that these are uncertain and difficult times and that relationships, even seemingly deep-rooted and permanent ones, do have a tendency to break down.

Introduces in this paragraph the device of sentence fragments. A sentence fragment. Another. Good device. Will be used more later.

This is actually the last sentence of the story but has been placed here by mistake. This is the title of this story, which is also found several times in the story itself. As Gregor Samsa awoke one morning from uneasy dreams he found himself in his bed transformed into a gigantic insect. This sentence informs you that the preceding sentence is from another story entirely (a much better one, it must be noted) and has no place at all in this particular narrative. In spite of the claims of the preceding sentence, this sentence feels compelled to inform you that the story you are reading is in actuality "The Metamorphosis," by Franz Kafka, and that the sentence referred to by the preceding sentence is the *only* sentence that does indeed belong in this story. This sentence overrides the preceding sentence by informing the reader (poor, confused wretch) that this piece of literature is actually the Declaration of Independence, but that the author, in a show of extreme negligence (if not malicious sabotage), has so far failed to include even *one single sentence* from that stirring document, although he has condescended to use a small sentence *fragment*, namely "When in the

The more we use ideas, the more we have.

Invention is at once a flash of inspiration and the slowly spinning thread of technology. At McDonnell Douglas, inspiration and technology have been nourished with these remarkable results:

Mining

A device with the apt name *Digs* emerged when we applied the analytic ability of computers to the ponderous motions of coal mining equipment. The result is a system that makes dragline operators more efficient, their machines safer and more productive.

Manufacturing

Human craftsmanship is augmented by automated craftsmanship in the McDonnell Douglas factories and at other businesses throughout the world. The skillful moves of a woodcarver exemplify the dexterity now required of machines which carve parts from metals and materials of all kinds. The McDonnell Douglas *Unigraphics*™ Design Processor weds designer skills with computer swiftness to form intricate instructions for the mills that cut the metal. Our *Actrion*® Controls interpret the instructions and guide the mills as they trim, grind and cut strong but delicate shapes efficiently and dependably. If such machines and controls did not exist, the advanced planes and products of our times could not be built.

Data Service

Our computer specialists provide *Health Care Data Services* for more than a thousand hospitals. We suggested to our clients that the codes stored in computers for record-keeping reasons could be used to presort first class mailings. That idea, offered free, saved them nearly a half million dollars last year.

Grain Drying

Our work in space technology taught us that microwave energy in a vacuum has surprising drying potential. Helping dry millions of military records soaked during a fire taught us more. Now we've designed *Mivac* systems for drying grains. They're superior to conventional methods, and in many cases they save fuel.

Fire Fighting

How timely it is! Our *Suspended Maneuvering System* (SMS) has emerged just when the problem of high-rise fires is becoming more than a movie theme. Suspended by cables from a helicopter hovering high above the smoke and heat, our computer-controlled "flying fire engine" moves gently alongside the building to deliver firefighters or rescue trapped victims. The SMS platform reduces the hazards of rescue slings or rooftop landings. Helicopter availability and crew training problems still remain for municipalities, but the technical solution is at hand.

Aerospace companies such as ours are known for the advanced technology they possess. We want to be known for the advanced technology we use.

MCDONNELL DOUGLAS

St. Louis, Missouri U.S.A.

course of human events," embedded in quotation marks near the end of a sentence. Showing a keen awareness of the boredom and downright hostility of the average reader with regard to the pointless conceptual games indulged in by the preceding sentences, *this* sentence returns us at last to the scenario of the story by asking the question, "Why is Billy strangling his mother?" This sentence attempts to shed some light on the question posed by the preceding sentence but fails. *This* sentence, however, succeeds in that it suggests a possible incestuous relationship between Billy and his mother and alludes to the concomitant Freudian complications any astute reader will immediately envision. Incest. The unspeakable taboo. The universal prohibition. Incest. And notice the sentence fragments? Good literary device. Will be used more later.

This is the first sentence in a new paragraph. This is the last sentence in a new paragraph.

This sentence can serve as either the beginning of a paragraph or the end, depending on its placement. This is the title of this story, which is also found several times in the story itself. This sentence raises a serious objection to the entire class of self-referential sentences that merely comment on their own function or placement within the story (e.g., the last four sentences), on the grounds that they are monotonously predictable, unforgivably self-indulgent and merely serve to distract the reader from the real subject of this story, which at this point seems to concern strangulation and incest and who knows what other delightful topics. The purpose of this sentence is to point out that the preceding sentence, while not itself a member of the class of self-referential sentences it objects to, nevertheless *also* serves merely to distract the reader from the real subject of this story, which actually concerns Gregor Samsa's inexplicable transformation into a gigantic insect (in spite of the vociferous counterclaims of other well-meaning although misinformed sentences). This sentence can serve as either the beginning of a paragraph or the end, depending on its placement.

This is the title of this story, which is also found several times in the story itself. This is *almost* the title of this story, which is found only once in the story itself. This sentence regretfully states that up to this point the self-referential mode of narrative has had a paralyzing effect on the actual progress of the story itself, that is, these sentences have been so concerned with analyzing themselves and their role in the story that they have failed by and large to perform their function as communicators of events and ideas that one hopes coalesce into a plot, character development, etc., in short, the very raisons d'être of any respectable, hardworking sentence in the

midst of a piece of compelling prose fiction. This sentence in addition points out the obvious analogy between the plight of these agonizingly self-aware sentences and similarly afflicted human beings, and it points out the analogous paralyzing effects wrought by excessive and tortured self-examination.

The purpose of this sentence (which can also serve as a paragraph) is to speculate that if the Declaration of Independence had been worded and structured as lackadaisically and incoherently as this story has been so far, there's no telling what kind of warped libertine society we'd be living in now or to what depths of decadence the inhabitants of this country might have sunk, even to the point of deranged and debased writers constructing irritatingly cumbersome and needlessly prolix sentences that sometimes possess the questionable if not downright undesirable quality of referring to themselves and they sometimes even become run-on sentences or exhibit other signs of inexcusably sloppy grammar like unneeded superfluous redundancies that almost certainly would have insidious effects on the lifestyle and morals of our impressionable youth, leading them to commit incest or even murder and maybe *that's* why Billy is strangling his mother, because of sentences *just like this one,* which have no discernible goals or perspicuous purpose and just end up anywhere, even in mid

Bizarre. A sentence fragment. Another fragment. Twelve years old. This is a sentence that. Fragmented. And strangling his mother. Sorry, sorry. Bizarre. This. More fragments. This is it. Fragments. The title of this story, which. Blond. Sorry, sorry. Fragment after fragment. Harder. This is a sentence that. Fragments. Damn good device.

The purpose of this sentence is threefold: (1) to apologize for the unfortunate and inexplicable lapse exhibited by the preceding paragraph; (2) to assure you, the reader, that it will not happen again; and (3) to reiterate the point that these are uncertain and difficult times and that aspects of language, even seemingly stable and deeply rooted ones such as syntax and meaning, do break down. This sentence adds nothing substantial to the sentiments of the preceding sentence but merely provides a concluding sentence to this paragraph, which otherwise might not have one.

This sentence, in a sudden and courageous burst of altruism, tries to abandon the self-referential mode but fails. This sentence tries again, but the attempt is doomed from the start.

This sentence, in a last-ditch attempt to infuse some iota of story line into this paralyzed prose piece, quickly alludes to Billy's frantic cover-up attempts, followed by a lyrical, touching and beautifully written passage wherein Billy is reconciled with his father (thus resolv-

ing the subliminal Freudian conflicts obvious to any astute reader) and a final exciting police chase scene during which Billy is accidentally shot and killed by a panicky rookie policeman who is coincidentally also named Billy. This sentence, although basically in complete sympathy with the laudable efforts of that last action-packed sentence, reminds the reader that such allusions to a story that doesn't, in fact, yet exist are no substitute for the real thing and therefore will not get the author (indolent goof-off that he is) off the proverbial hook.

Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. Paragraph.

The purpose. Of this paragraph. Is to apologize. For its gratuitous use. Of. Sentence fragments. Sorry.

The purpose of this sentence is to apologize for the pointless and silly adolescent games indulged in by the preceding two paragraphs, and to express regret on the part of us, the more mature sentences, that the entire tone of this story is such that it can't seem to communicate a simple, albeit somewhat sordid, scenario.

This sentence wishes to apologize for all the needless apologies found in this story (this one included), which, although placed here ostensibly for the benefit of the more vexed readers, merely delay in a maddeningly recursive way the continuation of the by now nearly forgotten story line.

This sentence is bursting at the punctuation marks with news of the dire import of self-reference as applied to sentences, a practice that could prove to be a veritable Pandora's box of potential havoc, for if a sentence can refer or allude to itself, why not a lowly subordinate clause, perhaps *this very* clause? Or this sentence fragment? Or three words? Two words? *One?*

Perhaps it is appropriate that this sentence gently and with no trace of condescension reminds us that these are indeed very difficult and uncertain times and that in general people just aren't nice enough to each other, and that perhaps we, whether sentient human beings or sentient sentences, should just *try harder.* I mean, there *is* such a thing as free will, there *has* to be, and this sentence is proof of it! Neither this sentence nor you, the reader, is completely helpless in the face of all the pitiless forces at work in the universe. We should stand our ground, face facts, take Mother Nature by the throat and just *try harder.* By the throat. Harder. Harder, harder.

Sorry.

This is the last sentence of the story. This is the last sentence of the story. This is the last sentence of the story. This is.

Sorry.

Investing for Tomorrow: How Saudi Arabia is Engineering its Future

By Roger Vielvoye, International Editor, Oil and Gas Journal

ABQAIQ GAS REFINERY

Today Saudi Arabia, which is the Western world's largest oil producer, stands poised on the threshold of major industrial and technological development. The current Five Year Plan emphasizes the need for productive investment to provide an alternative source of income to its current oil revenues. The challenge facing the Kingdom's 8.6 million people is how to channel its abundant financial resources so as to provide a viable alternative income in the years ahead.

During the last decade the Kingdom has undergone an enviable transformation which has brought it abruptly from the medieval age to the twentieth century. It has been an era which has seen the laying of the foundations of one of the major planned economies of the world.

· AT THE TIME OF WRITING 1 US DOLLAR EQUALS APPROXIMATELY 3.41 SAUDI RIYALS.

With 40 days' paid vacation every year, Aramco people can see the heart of Asia, from Delhi to Bangkok to Hong Kong, and still have time for an African game park, Paris in the spring, or ski trips in the Alps.

The money isn't the only reason so many top people are going to Saudi Arabia with ARAMCO

As great as the money is, it's the travel opportunities, the challenge, and casual lifestyle in familiar hometown neighborhoods that make Aramco people stay on.

Aramco employees in Saudi Arabia have travel opportunities that most people only see in daydreams. They get *40 days'* paid vacation every year. And

Over 5,000 North Americans work for Aramco in Saudi Arabia. They're helping the Kingdom develop energy resources that are being used all over the world.

in Saudi Arabia they are already halfway to wonderful places in Europe, Asia and Africa.

Equally as important as money and travel are the challenging jobs and secure, comfortable living conditions.

Aramco is the world's largest oil-producing company. Our projects and operations are huge, complex and professionally rewarding. Many are the largest of their kind ever undertaken.

We need more first-rate people to join the over 5,000 North Americans who work for us. We have openings for qualified engineers, technicians and professionals of all types.

North Americans and Saudi Arabs have worked together for over 40 years.

Since the early 1930's, North Americans have worked very closely and successfully with their Saudi hosts in developing the energy resources and infrastructure of this friendly, conservative nation.

These North Americans live in a foreign country, yet they still find it a very comfortable place to live and work. The homes, neighborhoods, schools,

medical services and recreation facilities have a hometown feeling that have prompted a long line of Aramco professionals to stay on and on.

What kind of compensation package accompanies all this? Aramco employees earn a competitive base salary, plus a cost-of-living differential. They also get an expatriate premium of 40 percent on the first $30,000 of base salary and 20 percent premium on the next $20,000, plus a one-time bonus of up to $5,000, for signing up for overseas work.

Challenging jobs are open in Saudi Arabia right now. We need experienced engineers in almost every discipline.

Interested? Call (713) 750-6965 anytime. Or call toll-free, (800) 231-7511, between 7 A.M. and 4 P.M., Central Time, weekdays. Or write Aramco Services Company, Mail Code Z2001-2, P. O. Box 53607, Houston, Texas 77052.

ARAMCO
SERVICES COMPANY

Just five years ago the desert scrub rolled uninterrupted down to the sea outside the small town of Jubail in the Eastern Province of Saudi Arabia. Today the landscape outside the old town has been transformed. Excavators, earth scrapers, trucks and cranes are everywhere, creating the infrastructure for a modern, industrial city that will become one of the cornerstones of Saudi Arabia's plans for giving the country a sound industrial base.

Driving through the skeleton of roads and across the wide canal at Jubail, George Brem, construction manager for the U.S. based company Bechtel, is impressed with progress on the industrial landscape. ''It took the United States 80 years to build the industrial belt along the shore of Lake Michigan around Chicago and Gary. The Saudis are planning to do the same sort of thing in just ten years,'' he said.

Jubail, and its sister industrial city, Yanbu on the Red Sea will be the site for oil refineries, petrochemical complexes and the companies that feed off these basic industries or are needed to keep them running. Saudi Arabia's industrial revolution will be based on the country's most abundant natural resources, oil and gas. Only the steelworks, now beginning to take shape in Jubail, bear any resemblance to the industrialization of yesterday.

Development of the country has been guided by the Government through a series of five year plans, started in 1970 by the late King Faisal. By 1980 Saudi Arabia had one of the fastest growing economies in the world. Citizens had experienced great increases in living standards, and had seen annual income rise from SR 4800 to SR 8200 between 1975 and 1980. They also became used to health, education and welfare services that it is estimated added 29 percent to personal income levels during the Second Plan.

The Ministry of Planning says that during these 10 years it has laid down, but not fully completed, the basic infrastructures the country needs. In the current plan which ends in 1985, there will be a move towards investment in public sector industries; private companies will be encouraged to take advantage of the conditions being created by the weight of public spending. Opportunities for investment in the private sector range from construction, service industries, distribution and light manufacturing to more specialist fields like the health service where private participation with foreign partners is acceptable.

As a result, quickening development is widely apparent. The capital, Riyadh, is akin to a giant construction site with more housing, hospitals, schools and colleges, and industrial premises consistently being added to the impressive ultra-modern range of office buildings, hotels and ministry buildings. Like Jeddah and the Dhahran-Dammam-Al-Khobar conurbation in the Eastern Province, the boundaries of the towns are being extended.

Outside the towns life is also changing. Roads and cheap airline services are giving mobility while telecommunications and television are opening up communications.

A western-style infrastructure may indicate a bustling, progressive society, but it overlays a deeply religious and conservative people, many of whom are concerned that materialism must not be allowed to destroy the traditional way of life. The objective is to use the best of western technology to improve the quality of life, while providing an industrial base in preparation for the day when oil and gas will play a less dominant role in the country's finances.

In spite of all the signs of wealth Saudi Arabia is still a developing country. Oil production in any volume did not begin until after the Second World War and was generating only $57 million in 1950. The country began to emerge from its medieval past as crude oil output built up in the 1960's. Even then the price of oil remained unchanged throughout the decade and investment funds were limited. As the sixties opened oil revenues were $334 million, and had increased to $1.2 billion. Due entirely to a huge jump in oil production from 1.32 million barrels per day to 3.8 million.

All that changed in 1973 and 1974. Oil prices rose four and fivefold generating a cash surplus fueling a round of infrastructure building and consumer spending. Oil production hit a new peak of 8.5 million barrels per day in 1974, and by the end of the First Five Year Plan oil revenues were running at $25.6 billion a year.

The second price explosion, in the wake of the Iranian revolution, saw Saudi prices rise from $12.70 per barrel to $32 while the production rate was raised from 8.5 million to a peak of almost 10.5 million daily.

Oil alone generates gross revenues of $116 billion while the Third Five Year Plan calls for expenditures during that period of $235 billion. Figures from the Saudi Arabian Monetary Agency (SAMA)—the national bank—show that last summer the country had built reserves of $80 billion. A major proportion has been invested in international capital markets and by this spring the reserve figure should have risen to around $92 billion.

Saudi Arabia's oil reserves are enormous. Officially they stand at 165 billion barrels that can be recovered using cur-

rent technology. All the major oilfields are located in the Eastern Province and have been developed by the Arabian American Oil Co. (Aramco) which recently became a wholly owned Government company. Exploration is still continuing, with new fields added to the list of reserves each year.

These reserves have enabled Aramco to build a production system capable of producing over 12 million barrels per day and sustaining output at around 10 million barrels per day, an incredible performance in so short a time. It is this ability to export huge amounts of crude, and periodically control the workings of the Organization of Petroleum Exporting Countries (OPEC), that has brought the country to the forefront of the international stage.

Jubail and Yanbu will enable the Saudis to export refined products from the crude oil but, more importantly, the cities will make them a force in the petrochemical and gas business. The way in which the Saudis have used their financial surplus to exploit gas resources and then use them as the basis for new petrochemical industries and extensive exports, typifies the way in which resources and manpower can be mobilized and concentrated on specific projects of national importance.

In the mid 1970's the night skies over the Eastern Province were always lit up by huge fires from the oilfields as Aramco burned off the natural gas produced in association with oil. Small amounts were reinjected but the bulk—around 3,000 million cubic feet a day—were flared because there was no way of collecting and marketing the fuel.

Under the direction of Aramco, with the Fluor Corporation acting as main contractors, work started on the Master Gas System—described by the Saudis, with some justification, as "the greatest engineering undertaking of all time." At a cost of $15,000 million, 2,400 kilometers of pipeline and numerous large production units to separate the fuel gas from the gas liquids are under construction and scheduled for completion in 1982/83.

The system will then be yielding 2,000 million cubic feet of fuel gas daily; 375 million cubic feet of ethane; 330,000 barrels of natural gas liquids and 3,700 tons of sulfur. One of the most ambitious aspects of the scheme was the construction of a 1,170 kilometer pipeline to deliver 270,000 barrels per day of ethane and gas liquids to Yanbu for either export or use in the petrochemical industries planned for the new city.

The high-pressure gas pipeline runs alongside a 48-inch crude oil line from Eastern Province to Yanbu, where again the crude will either be exported or fed into the new refineries under construction on the Red Sea coast. Without the two pipelines—by themselves an engineering task of staggering proportions—Yanbu could never have emerged from the planner's dream into reality.

Concentration of the public sector development in industrial projects in Jubail and Yanbu has brought some criticism that this policy ignores the traditional centers. The Ministry of Planning is committed to developing a thriving private sector that will dominate the economy, but says that in the initial stages the state must take a hand in the planning and execution of major basic industrial projects.

That means in the current stage of development it is the state that is making all the running. Government spending represents between 40 and 50 percent of the gross domestic product and accounts for 60 to 70 percent of fixed capital formation.

Originally, Saudi Arabia had few instruments through which to govern. Petromin, the state oil and mineral company, had a wide brief and in 1976 Saudi Arabian Basic Industries (SABIC) was created to run the petrochemical and other basic industries outside oil refining. It also set up a Royal Commission to guide the development of Jubail and Yanbu by cutting through bureaucratic red tape.

Dr. Farouk Aktar, director-general of the Royal Commission, is convinced that neither of the projects could have moved forward quickly without the impetus provided by the Commission, which is responsible for provision of all the infrastructure in the cities—roads, water, electricity, ports, airports, housing, education and medical facilities. The Commission has taken on responsibilities of the various ministries who would otherwise have been involved in the development stages.

Twenty-five percent of the infrastructure in the first phase is now complete and, says Dr. Aktar, people outside Saudi Arabia are taking the projects seriously.

"When the projects were first announced outsiders were skeptical. They thought we were just a lot of rich men scratching our heads and wondering where to invest our money and coming up with Jubail and Yanbu.

"It has taken time to convince people we mean business. Since 1980 people have seen we are not joking and from the outside we are being taken very seriously indeed," he added.

A look around Jubail shows why the skepticism has vanished. Bechtel, working under the Royal Commission has 179 contracts under way and since work on the site started has dealt with 250 altogether.

Port facilities are largely complete, the road network is taking shape and efforts are being concentrated on the 14 kilometer canal which will bring cooling water to the city. The cooling water system has the flow equivalent to that of the River Tigris.

This year spending on the infrastructure will be 5.8 billion riyals and will rise next year to 6.4 billion. What has given the two cities real credibility is the list of international companies that have signed joint venture agreements with SABIC and Petromin for projects at Jubail and Yanbu.

At Jubail there are 12 major joint ventures. On the oil side Petromin is involved with the Royal Dutch/Shell group and Chevron/Texaco in refining projects and will also have a sulfur drilling plant and a bulk plant for the sale of products.

The SABIC list of ventures is also impressive. There are four major petrochemical plants in partnership with Shell Oil, Dow Chemical, Exxon and Mitsubishi. Two methanol plants will be built for Celanese/Texas Eastern and Mitsubishi while the Taiwan Fertilizer Company is building a urea plant. The only non-hydrocarbon feedstock based project is the Korfstahl/SABIC steel mill which will take advantage of cheap fuel gas.

On the much smaller Yanbu site, the Mobil Chemical Company and SABIC are to build an ethylene plant and Mobil, in partnership this time with Petromin, has plans for a 250,000 barrels per day export refinery. Petromin, working on its own is well advanced with a 170,000 barrels per day refinery to supply the domestic market which may have to be expanded before the end of the decade, because of rapidly rising demand. Also scheduled for Yanbu is a lubrication oil feedstock refinery in partnership with Ashland Oil.

Outside the remit of the Royal Commission, the Greek Petrola Group is building a 325,000 barrels per day export refinery at Rabigh.

One of the features of the Royal Commission's work in both the Eastern Province and the Red Sea is the insistence on the contractors' use of specially built accommodation for the ex-patriate laborers. Conditions in many of the labor camps in Saudi Arabia and other parts of the Middle East, have by western standards left much to be desired.

Imported labor from countries around the Gulf and from the Far East earn four to five times the amount they could at home and to the chagrin of some contractors now live in air-conditioned accommodations and eat in cafeterias provided by the Commission where the diet is carefully supervised. The bill is footed by the employers.

Dr. Aktar says: ''We are giving these people a lifestyle they cannot afford at home. We don't want history to say the Saudis built their industrial base on the backs of foreign workers.''

And it is not only the laborers who are well looked after. Ex-patriate technicians and managers have high class accommodations with excellent recreational facili-

مشروع تجمع الغاز
Gas Gathering Project
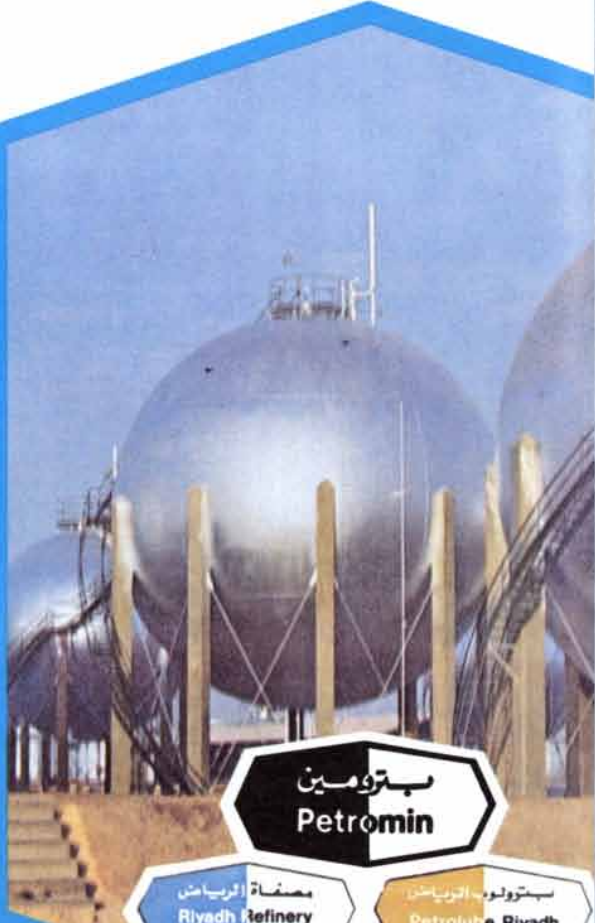
بترومين للتسويق
Petromin Marketing

مصفاة الجبيل
Al-Jubayl Refinery 1

مارينكو
Marinco

مصفاة الجبيل
Al-Jubayl Refinery 2

مصفاة الجبيل
Al-Jubayl Refinery 3

شركة الحفر العربية
Arabian Drilling Co.

بترومين
Petromin

مصفاة الرياض
Riyadh Refinery

بترولوب الرياض
Petrolube Riyadh
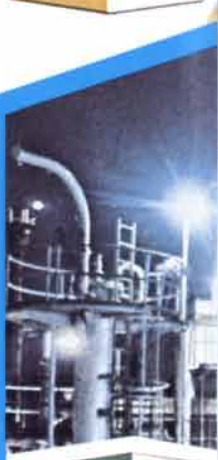
بتروشلد
Petroshield

بترولوب جدة
Petrolube Jiddah

بتروشب
Petroship

بتروجت
Petrojet

ارغاس
Argas

Dome

بتروجرانجس
Petrograngas

بتروبوت
Petroyot

بتروسيريم
Petrocereme

Petromin Means Petroleum, Petrochemical & Mineral Industries.

The activities of the General Petroleum and Mineral Organization (Petromin) are limited to three geographical regions: Riyadh: The location of Petromin's head office.

— The Riyadh Refinery on Kilometer 20 of the city of Riyadh. P.O. Box 3946.
— Petromin Lubricating Oil Company (PETROLUBE RIYADH).

The Western Province: Petromin has a representative office in Jiddah which serves as a link between Petromin and its projects in the Province. These are:

— The Jiddah Oil Refinery Company. Telex: 401867 PETROMIN SJ.
— The Arabian Geophysical and Surveying Company (ARGAS).

— Petromin Lubricating Oil Company (PETROLUBE JIDDAH). Telex: 401675 PETLUB SJ.
— Petromin Tankers and Mineral Shipping Company (PETROSHIP). Tlx: 401867 PETROMIN SJ.
— Petromin Services (PETROSERVE).
— Export Refineries, Yanbu.
— Abqaiq-Yanbu Crude Pipeline (PETROLINE). Telex: 401150 PETROMIN SJ.
— Yanbu Domestic Refinery. Telex: 402842 YANRIF SJ.
— Lubref — The Petromin Basic Oils Refinery. Telex: 402781 LUBREF SJ.
— Petromin Representative — Jeddah Telex: 401150 PETROMIN SJ.

The Eastern Province: Here Petromin has a representative office and the following ventures:

— Petromin Marketing. Telex: 670009 PATMARK SJ.
— The Arabian Drilling Company. Telex: 671212 ARABDRI SJ
— The Arabian Marine Petroleum Construction Company (MARINCO).
— Export Refineries, al-Jubayl.
— Gas-Gathering Project, Al-Jubayl.
— Petromin Representative — Dhahran. Telex: 601230 PETGAS SJ.
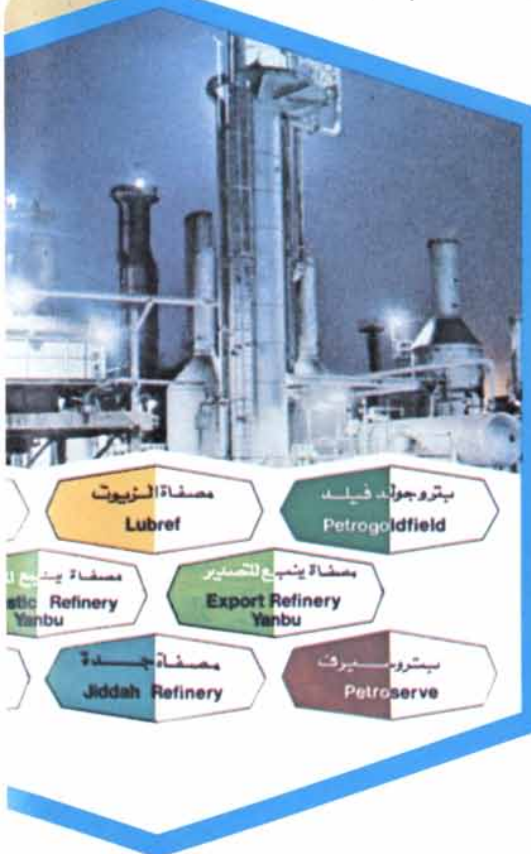
Petromin

GENERAL PETROLEUM & MINERAL ORGANIZATION

P.O. Box 757, Riyadh, Saudi Arabia,
Tel: 498-3003, 498-1019, 478-1661,
Tlx: 201058, 201490, 201615 SJ., Cable: PETROMIN.

ties.

The Royal Commission is responsible for the infrastructure and Petromin and SABIC the big joint ventures. Bringing in the support industries to provide goods and services and the secondary industries that will upgrade the products from the big joint ventures, particularly in the petrochemical field, is also the Commission's job.

Dr. Aktar said the Commission had undertaken studies on the demand for ancillary and secondary companies and there was no shortage of operators to take up the opportunities being offered.

Throughout the economy emphasis is being placed on "Saudiization." Although there are few specific rules that work must be given to Saudis, foreign companies trying to act without a local partner may find themselves left out in the cold.

Aramco goes to lengths to support Saudiization. Last year the company spent approximately $2,000 million of which about $1,300 million went through the hands of Saudi Arabian companies, either in joint ventures or agency agreements.

The oil company maintains very high standards for equipment and deals only through approved manufacturers. In 1977 there were only 16 Saudi companies on the approved list. The figure rose to 160 in 1980 and by the end of last year was well over the 200 mark. Providing these manufacturers can meet Aramco's requirements and are competitive on price they can expect preference over items produced abroad and imported through an agency or joint venture.

Over the years Aramco has tried to encourage the establishment of local manufacturing facilities and has nursed its own Saudi employees, who have branched out and set up on their own. One of the first and most notable successes is Suliman S. Olayan who left Aramco to set up the General Contracting Company to provide transport services when the Trans-Arabian Pipeline (Tapline) was built from Saudi Arabia to the Mediterranean in 1947.

From this beginning the General Contracting Company has become just one of the divisions in the Olayan group selling motor vehicles and industrial equipment while other divisions deal with trucking and distributing imported foods and household products. The company is now expanding into other parts of the economy.

During the Second Five Year Plan the private sector expanded at 15.8 percent annually compared with a target of 14 percent. Investment by private companies totals over $7.6 billion and the Government is hoping that its incentives for Saudis starting new businesses will maintain this momentum. Soft loans, low rent for industrial premises, duty free imports of materi-

als are all designed to help companies through the formation period. Once production has started businessmen can expect preferential treatment from Government purchasing agencies and, where necessary, special tariffs to protect embryo concerns from imports.

The Saudi market is, however, small and gearing up for production can be expensive.

On the surface it might appear that the new basic industries at Jubail and Yanbu will be producing far more than the domestic market can absorb. The world markets for petrochemicals and refined products are also suffering from severe overcapacity as a result of a sharp fall in demand.

But as Mr. Abdulazziz A. Al-Zamil, vice-chairman and managing director of SABIC points out, the joint ventures with overseas companies of international stature should protect the Saudi operations.

"Our agreements with joint venture partners provide for them to market products that are surplus to Saudi requirements. We shall be acquiring the international marketing experience of these companies as well as assistance and advice on the technology required for the subjects," he said.

Mr. Al-Zamil says that when the new projects come on stream between 1983 and 1986 they will not cause widespread disruption of world markets. Petrochemicals are currently suffering from severe overcapacity but, he adds, the major companies in these markets are now building the effects of the new plants in Saudi Arabia into their future forecasts of supply and demand.

The net effect is that each plant built and operated in Saudi Arabia will remove the scope for similar expansion in Europe, or even the United States. They could, he adds, also hasten the closure of older less efficient units.

"When plans for the construction of big plants in the country were announced, the major companies took no account of them in their forecasts. After all, many plants are announced and don't get into production. All that has changed.

"You must bear in mind we are building no more than four percent of the free world capacity. In normal times that is hardly a year's natural growth in the petrochemical business so it will be absorbed easily," he said.

Negotiating a joint venture can be a lengthy business. Mobil is the biggest of the partners in joint ventures with Saudi companies. Originally the minority partner in Aramco, Mobil often found itself lining up with Petromin to urge the other U.S. partners to keep production high.

The companies found they had a common interest and from this beginning Mo-
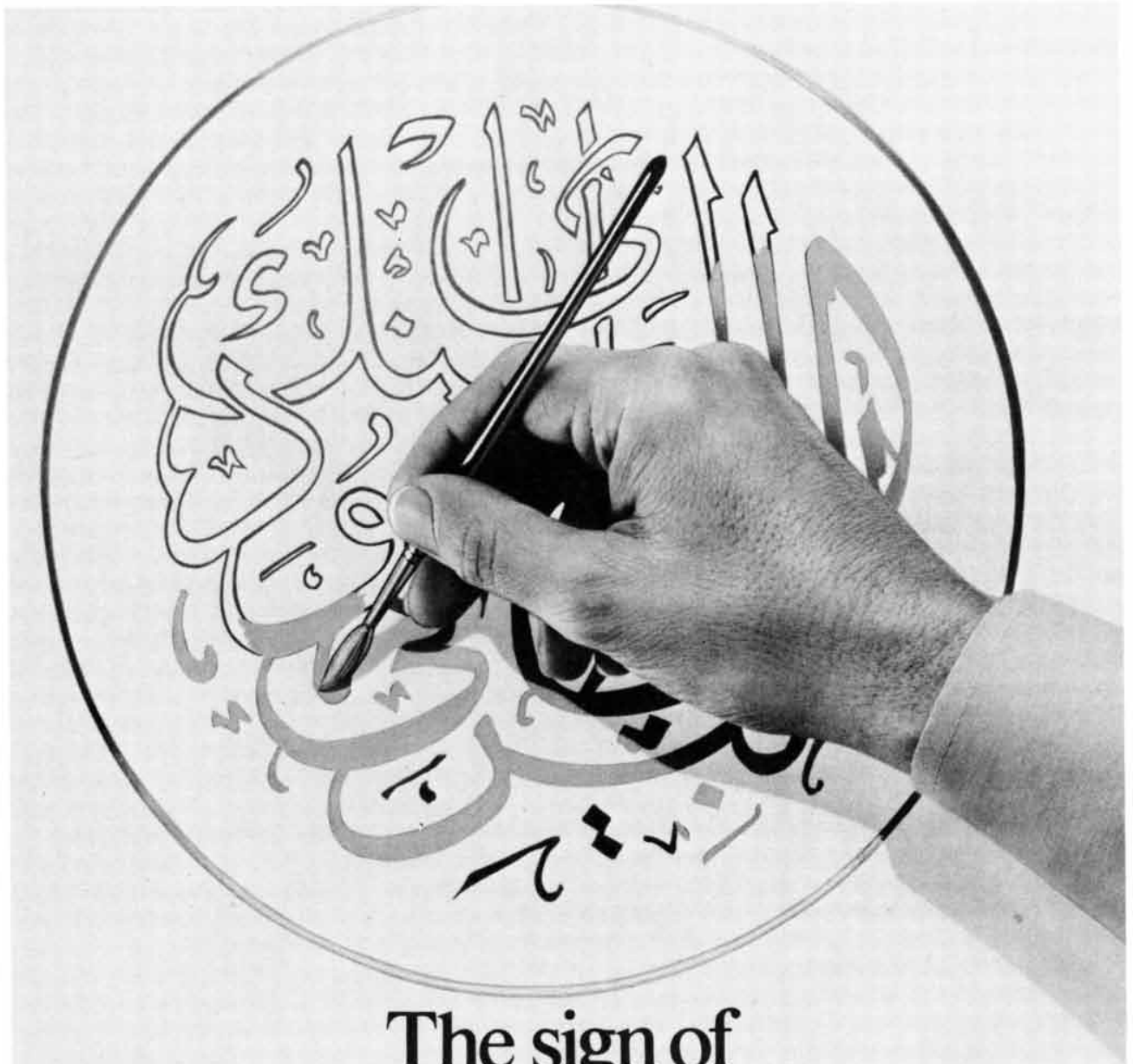
The sign of understanding.

There are branches and offices of The Saudi British Bank throughout the Kingdom.

In Al-Khobar, Dammam, Hoffuf, Jeddah, Jubail, Qatif, Riyadh.

And in each case, the sign outside does more than announce our presence. It shows you that we understand the demands and needs of a fast developing economy and nation.

It shows you that we understand and are ready to help with your financial requirements whether at home or abroad, no matter how small or large.

And it shows that we understand and have a commitment to all our customers from the personal account holder through to the largest international corporation.

البنك السعودي البريطاني

The Saudi British Bank
The bank that understands
Alkhobar · Dammam · Hoffuf · Jeddah · Jubail · Qatif · Riyadh

bil moved into a lube oil venture with Petromin and was a close advisor on the east-west crude oil pipeline to Yanbu which was built and is now managed for Petromin by the Mobil Overseas Pipeline Co.

Mobil Saudi Arabia president Lou Noto recalled the skepticism in some parts of the industry when the Saudis announced the program of large refining and petrochemical projects. Mobil guessed rightly that the Saudis meant to carry out these projects, he said.

"We felt it was a place where we could do business. It is a good environment so that when the Saudis look at us they see enthusiasm. That has helped us to forge a good relationship with local companies."

Mobil like every company private or public in Saudi Arabia, also has to face the problem of labor and training. In a population of 8.6 million there is little tradition of business or industry. Manual work tends to be done by ex-patriate laborers on short term contracts while much of the skilled managerial and technical functions are in the hands of foreigners.

Saudiization also applies to the labor market. Training Saudis to take over from ex-patriates is given top priority at every level.

New projects tend to import the latest capital-intensive technology that require the minimum of supervision. SABIC is insisting that joint venture partners provide training for Saudi personnel often overseas.

Priority is being given to vocational training centers but these have to fight hard for potential candidates with the big companies, the military and even the universities.

During the current Five Year Plan the universities expect to double their enrollment figures. The Saudis are also keen to make a bigger impact at the post-graduate level particularly in the field of research and development (R&D).

SANCST: Coordination of Scientific Research

The promotion and coordination of scientific research in the Kingdom has been entrusted to the Saudi Arabian National Center for Science and Technology (SANCST). The chairman, Dr. Rida Obaid, says it has very close relations with the National Science Foundation in the United States.

SANCST represents Saudi Arabia in bilateral agreements such as the Soleras deal with the United States and other agreements with Taiwan and Canada. It is coordinating 75 projects in Saudi universities all of an applied nature—energy, water, agriculture, construction—and in the longer term will ensure there is an expanded scientific infrastructure including national scientific laboratories.

Dr. Obaid emphasizes the practical nature of R&D in Saudi Arabia which, he says, must learn to transfer and adapt technology for the evolution of the country.

The two bilateral projects with the Taiwanese fall into this category. The production of single cell proteins from hydrocarbons is being studied in partnership with the Chinese Petroleum Corporation.

The second part of the agreement with the Chinese involves fresh and salt water aquaculture. A start has been made on developing the controlled environment needed to breed the tilapia, a small freshwater species common in the Nile, in inland ponds near Riyadh.

A fishpond in a greenhouse where the temperature can be kept within the strict limits of 16 to 20 degrees centigrade (outside the summer highs and winter lows in the area) has been constructed under the supervision of the Taiwan Fisheries Research Institute and the College of Marine Studies from Jeddah.

Air conditioning and recirculating filtration have apparently coped with the summertime temperatures of over 40 degrees centigrade and the high salinity of underground water used in the experiment. The first fry bred in Riyadh are already available and Dr. Obaid is hoping to see commercial fish farms using the stock from the experimental unit.

"It's a very tasty fish and popular but people still have to get used to the availability of fresh supplies rather than frozen fish brought in from outside."

So far the only exception to the practical-studies-only rule is the bilateral agreement with the Canadian National Research Council (NRC) to find the site for a National Observatory in the country.

Teams from NRC and Dominion Astrophysical Observatory have established sites at four locations in Saudi Arabia. Monitoring conditions started in August 1981 and the program will last for a total of 15 months.

Dr. Obaid said that once the monitoring period is complete a decision will be taken on the type and size of telescope to be installed. "This project is completely different from all the others. It will be pure research work.

"The Arabs were the fathers of astronomy and people in Saudi Arabia are fond and interested in the stars. Until recently the Bedouin used to find their way across the desert by this method and you only have to look at the sky on a clear night to know why our people are so fascinated."

Nous parlons le langage de ceux qui font le futur.

POUR LA SCIENCE is read by 41,321 * decision-making professionals throughout France.

The fortunes of major corporations increasingly depend on technically sophisticated people. POUR LA SCIENCE attracts a young audience (average age: 38) of well-paid professionals who are in positions of prominence because of their technical expertise. More than half of them hold top-management job titles.

We speak the language of more than 3,000,000 world shakers in 140 countries each month. We reach the people who make the future happen in French, Spanish, Italian, German, Japanese, Chinese and English.

For more information on the French-language edition of SCIENTIFIC AMERICAN contact:

Francoise Teisseyre
POUR LA SCIENCE
59, Avenue Marceau
75116 Paris, France
Telephone (011-331) 720-7251

*OJD audited

SCIENTIFIC AMERICAN
We reach the people who make the future happen.

BOOKS

*The sky, the solar system, volcanoes,
hands and a patrician life in science*

by Philip Morrison

SKY CALENDAR, by Robert Victor and
Jenny Pon. Abrams Planetarium,
Michigan State University, East
Lansing, Mich. 48824 ($3.50 per year,
starting anytime, with separate monthly
pages mailed quarterly). SKY ATLAS EP-
OCH 2000.0, by Wil Tirion. Sky Publish-
ing Corporation (deluxe edition, bound
and in four colors, $27.50; field edition
of unbound pages with white stars and
desk edition of unbound pages with
black stars, each $14.95). THE NEW SO-
LAR SYSTEM, edited by J. Kelly Beatty,
Brian O'Leary and Andrew Chaikin.
Cambridge University Press and Sky
Publishing Corporation ($19.95). In the
small hours (Eastern Standard Time) of
January 2, 1982, we earthborne travel-
ers reach our nearest orbital point to the
sun, only to start outward again; it is the
closest we come to a physical meaning
for our New Year. Enter 1982 with this
customary comment on a few of all the
many works of high quality that these
days offer guidance to the pointers vis-
ible on the celestial dial. The month-
ly *Sky Calendar* sheet from Michigan,
billed as "information for helping teach-
ers and students observe the sky," seems
to be the best buy. Each monthly page is
a spread of daily boxes. Each day of the
month is given over either to part of a
simple text account (with an occasional
question) or to a small sketch represent-
ing a section of the horizon, usually at
dawn or dusk, with a clear indication
of the sights to be seen there: planets
or bright stars nestling close, the moon
lively and new or waxed old and grave,
eclipses, meteor showers, approaches to
the limits of horizon motion or whatev-
er else the authors expect to watch for
when the sky is clear.

The calendar page sets out the hori-
zon astronomy of the ancient northern
Temperate Zone heritage. Overleaf the
entire starry evening hemisphere of the
month is very simply mapped (by D.
David Batch), with only the most con-
spicuous stars shown. Intended for guid-
ing stargazers in grade school, the de-
tailed and reliable data are helpful to
anyone who wants to look with care. An
occasional extra sheet adds more mat-
ter, such as last year's fine account of the
great conjunction of Jupiter and Saturn,
half a year of ecliptic dos-à-dos along a

few degrees of arc in the sky. The two
close-set cat's-eyes burned until they
were caught in the autumn sun. This
year will be no such vintage one for
conjunctions, although the flirtation of
Mars, Jupiter and Saturn will be worth
watching come late July, and we in the
U.S. can hope to see two total lunar
eclipses. These calendar pages are a bar-
gain entry into direct observation of the
changing order visible in the heavens.
Produced in a cost-cutting format, the
pages are graphically no beauties (off-
set typewriter text), but handsome is as
handsome does (as was remarked here
six years back about this publication).

Look far beyond the wanderers of the
ecliptic to the fixed stars. Every couple
of decades some heroic delineator gives
us an elegant new map of the stars. This
year we have the production of Wil Ti-
rion, no team but a one-man show, a
young graphic artist in the Netherlands
with a tireless, devoted and enthusiastic
pen. He has prepared an atlas of all
the stars brighter than about the eighth
magnitude, mapping nearly 43,000 of
them. This is no unaided-eye view but
one seen with a telescope of modest size.
Within the bowl of the Big Dipper three
dozen stars are entered, with more than
a dozen galaxies. To stars selected from
the three-volume position catalogue of
the Smithsonian Astrophysical Obser-
vatory, Tirion has added the brighter
stars omitted from that scrupulous work
because certain data were unavailable.
The main deep-sky objects are carefully
plotted as well, in all about 2,500 gal-
axies, star clusters and nebulas. The co-
ordinates used are those of the epoch
2000.0, closer to the truth than the 1950
values in most of the current atlases.

The entire sky is shown in these flat
charts, each one carefully projected for
good shape compromises. The scale is
workable, that of a large atlas page; Ori-
on spreads out as big as your hand, still
all on one sheet. The atlas is a useful
reference not only for those who engage
the sky with a telescope but also for
those who do not, although not so much
for disciplined unaided-eye observers.
What the reader can find is the sky po-
sition of many objects we often read
about or see photographically repro-
duced, such as the Horsehead Nebula or

40

BIGGER ISN'T NECESSARILY BETTER IN THE BIG BUSINESS OF OIL.

You don't have to be a giant company to become a big success in the oil business. At Union Texas Petroleum, we take advantage of our size. As a hardworking independent company, our efficiency and speed have let us achieve some giant-sized successes.

For example, when it comes to finding oil and gas, our strength comes from our efficiency. In fact, in 1980, our finding and development costs per barrel were almost 50% lower than the industry average.

Unlike most companies, we have higher oil and gas reserves today than 5 years ago. And in the next 5 years, we'll be spending almost $3 billion to keep them high. We're currently exploring in the U.S. and 20 foreign countries, and in 1981 alone we made significant discoveries in Louisiana, Texas and Pakistan.

Along with energetic businesses like fibers and plastics, chemicals, and electrical products, UTP's oil and gas help fuel Allied Corporation.

True, size can be a large factor in the energy field. But UTP shares a giant advantage with every Allied company.

We mean business

ALLIED Corporation
We mean business.

INTRODUCING THE CAR CONVENTIONAL WISDOM DECREED COULD NOT BE BUILT.

There is a common enemy to innovation in automobiles, as in most everything else.

It is conventional wisdom—a failure of vision or nerve that causes car makers to postpone attempts at the optimum. And settle instead for the comfort of consensus and compromise.

By that standard, BMW has created a luxury car engineered to cause considerable discomfort—not to those who buy such cars, but those who build them.

THE BMW 528e: MAXIMUM PERFORMANCE FROM MINIMUM ENERGY.

The 528e was designed to meet an engineering objective that reads like the modern equivalent of alchemy:

Turn minimum energy into maximum performance.

It succeeds—largely because of the new BMW "Eta" engine.

The conventional engine accepts a loss of response as a fair price to pay for fuel efficiency.

The Eta engine doesn't. It actually develops higher torque (or power) at engine speeds where the car is most often driven. The response is exhilarating, even by BMW standards.

Convention says that an engine must run faster to be so responsive.

The Eta engine doesn't. It actually runs slower than the conventional 6-cylinder engine—and is designed to lessen engine wear accordingly.

Convention often turns to diesels to deliver fuel efficiency, and accepts their unpleasant side effects, such as noise and excessive emissions, as necessary evils.

Since the Eta engine runs on gasoline, no such evils are necessary. And to ensure efficiency, it relies upon another BMW innovation called Digital Motor Electronics.

SPEED: 55 MPH.
FUEL CONSUMPTION: 0.

"DME" is not another control mandated by government regulators.

It is an efficiency system mandated by BMW engineers— a means of maintaining constant electronic surveillance over the engine.

DME assures that the optimum

THE GIFT MOST LIKELY TO SUCCEED.
After all, it's the number one imported distilled spirit in America.

Give V.O. in abundance. Enjoy our quality in moderation.
Canadian whisky. A blend of Canada's finest whiskies. 6 years old. 86.8 proof. Seagram Distillers Co., N.Y.C.

Barnard's star, not easy on the smaller and sparser sky charts meant for the unaided-eye observer. The hand-drawn graphics are crisp and inviting, with an up-to-date air, not the fuzzy characters and rigid typography of the computer printout. The bound edition with its glossy paper stock is worth having for its colors, which not only enhance the pages but make it much easier to distinguish the galaxies (red) and the star clusters (yellow), seen only as open circles and ellipses in the less expensive unbound editions. The field version with white stars on a black ground, meant to serve at the telescope, is none the less striking.

The New Solar System is a summary account of what has been learned about the solar system to date, at a level that is "neither a textbook nor a 'coffee-table' volume—it lies somewhere in between." The authors and editors are nearly all people caught up in NASA space-sciences investigations; quite a few are university faculty members as well. This is no report from the mountaintop; the planetary probes that actually went out to have a look and scrape the soil are much different from the observatory, not only in their investigative power but also in the sustained cooperative effort they demand.

One of the most interesting pieces—quite unlike any essay in solar-system astronomy of an earlier epoch—is one by Noel W. Hinners, the reflective former Associate Administrator for Space Science of NASA during half of the 1970's. He puts that golden age of exploration into perspective: its roots were national prestige, a sense of vision, the thirst for knowledge and perhaps its applications. The mismatch between the yearly budget and the 10-year mission cycle, and the struggle over new starts within a steadily declining total, are grittier questions. He argues that the lean years now ahead must be met not in depression or unhappy silence but with energy and determination to exploit anew what might be done even when costs seem high and budgets tight. The Space Shuttle and the upper stage it needs for deep-space missions are delayed, while the costs of the shuttle take over more and more of the shrinking real NASA budget. Yet over the long run the entire instrumental program has found shelter under "the manned program's budgetary umbrella." Here is a broadly informed account, if a diplomatically framed and optimistic one. The Space Station looms ahead. Paramilitary, it will cost plenty, although most of the money will be spent while the station is in low earth orbit.

Only one of the chapters is at so worldly a level; the rest are about theories and results. Comparative accounts dominate: it is too late to treat our system as nine independent planets, one star and some debris. Surfaces, col-

lisions, atmospheres, rings, magnetic fields and solar-wind interactions form an important part of the volume. Mars, the moon, Jupiter and its four Galilean moons, Saturn, Titan, their phenomena and their models account for most of the rest. Comets, asteroids and meteorites begin to be seen as one physical class. Is Chiron out there near Uranus a comet? It has a comet's orbit, but it is as big as an asteroid. The system actually has three classes of objects: the giant gassy planets and the small rocky ones both wheel in circular orbits, whereas a wide variety of smaller bodies fly in orbits of many forms. There is a continuum of compositions, motions and histories.

The illustrations include a fine set of images of reality, graphs, maps, diagrams of models and a number of imaginative paintings (well executed throughout but a little disturbing in the degree to which their plausible verisimilitude tends to dominate reality). An outstanding evidential painting is one by Andrew Chaikin, a geologist-editor of this book, which shows some 30 larger asteroids to scale, with observed color and texture, arranged diagrammatically by orbital properties. It is all worth long study.

The spiral of ice layers in the Martian polar cap, the photomosaic of Jupiter's ring, the plume of a volcano on Io (juxtaposed with a simple ejection model), samples of the Allende meteorite both at the scale of a hand and close up, and a few kilometers of green laser beam on its way to moon reflectors will serve to indicate the scope and level of the work. Why is the planet Mercury so dense? Did Venus have oceans? Where did the moon come from? Is orange Titan loaded with organic matter? These are answers we do not know, as John S. Lewis concludes in a last theoretical chapter.

The volume comes out at a reasonable time; before the longish gap in exploration that lies ahead only *Voyager 2* is a little slighted. Useful and very readable, the book is not unlike a big single-topic issue of this magazine in its cooperative and expert nature.

VOLCANOES OF THE WORLD: A REGIONAL DIRECTORY, GAZETTEER, AND CHRONOLOGY OF VOLCANISM DURING THE LAST 10,000 YEARS, by Tom Simkin, Lee Siebert, Lindsay McClelland, David Bridge, Christopher Newhall and John H. Latter. Hutchinson Ross Publishing Company, distributed by Academic Press ($19.75). "Volcano lists are not new." Indeed, the first list of three pages appeared in 1650; its Amsterdam author, the young geographer Varenius, already knew of most of the volcanic regions in the world. Since 1951 the international community of volcanologists has maintained a continuing catalogue of volcanoes by region, replete with maps, photographs, references and chemical data, which stands now at 22 big volumes. It is that *Catalog*

of Active Volcanoes of the World from Rome that is the primary source for this single summary volume, along with the annual volcanological bulletin and other current matter. Its senior author not only is Curator of Petrology and Volcanology at the National Museum of Natural History in Washington but also has taken on the task of editing the international *Catalog* itself.

The endpapers define the book. The closing papers hold a world map, the volcanoes numbered and marked in red; the front papers summarize the contents and the abbreviations of the three long computer-generated lists that form the bulk of these 200 big pages. The chief list is the directory, which musters by region 1,353 volcanoes active over the past 10,000 years, a line for each well-defined crater giving some physical details in compact form. There is a chronology of 5,564 distinct eruptions among all these volcanoes from 8000 B.C. to 1980 and a gazetteer with names and synonyms alphabetized in all the tongues of mankind, the worldwide roll call from Aak to Zyogoro-yama.

More than 500 volcanoes have had one or more recorded eruptions in the span of human history: firm evidence of activity. The world of 10 millenniums is most unevenly covered, however, by published records. A hundred more volcanoes have eruptions reasonably well dated by carbon 14, tree rings, varve counts, magnetic-field directions, ash stratigraphy and one or two other such techniques. The rest gain less precise place because they still show hot spots, say hot springs and fumaroles, or even subtler geologic evidence for postglacial eruption. "Anthropology" is one dating category; it includes legendary accounts among peoples living near volcanoes and a few instances of recovered buried artifacts. The outcome is a nearly exponential rise in the number of eruptions reported per year from 1500 to the present. A naive extrapolation suggests the entire world will explode in ash within a modest number of millenniums! Of course, this is a defective inference from a sample of only two parts per million of geologic time. The curve is parallel to the curve of human population; what is in fact measured here is central reporting, the worldwide spread not of eruptions but of recorded public witness.

Classical records included only 10 volcanoes, nine in Mare Nostrum and one in the Cameroons. Indonesian eruptions have been recorded only since A.D. 1000. Iceland was settled at about this time, to swell the European list. A study over time of eruption counts with respect to known volcanoes shows two steps of increase, which are not hard to interpret. The first is the time of the printing press and the galleons: the Renaissance of the West. The second is the 19th-century expansion of trade and reportage, and the extension of European

45

Do you want to sell, license or fund your health-care invention?

We represent a group of European and U.S. companies who want to buy or license—and in some cases invest money in—new pharmaceuticals, diagnostics and devices. Also: OTC oral hygiene, skin care and hair care products and technologies. These companies are totally funding our search efforts. *There are no costs or fees charged to invention-owners.* For information, write to Eugene F. Whelan, Chairman or Alan R. Tripp, President, Product Resources International, Inc., Dept. 33, 800 Third Avenue, New York, N.Y. 10022. Please send no disclosures with your initial letter, except of course for issued patents. We will also welcome an inquiry through your attorney.

control to all the world. Twentieth-century data show a slower steady rise, with sharp drops timed to fit the two wars and a drop during the Great Depression. Preoccupied witnesses and shaky publications do not report eruptions fairly. On the other hand, a few recent peaks imply that a press sensitized by a Krakatoa calamity is likely to rumble with eruption news for five to seven years afterward. The data bear out these remarks, although there is no clear control for such plausible conclusions. One peak in the recent curve in about 1950 has no broad historic correlate. It seems to reflect one special Geological Survey program in Alaska and a coincidental Indonesian catalogue volume, followed by civil strife in that country for several years.

It is not Krakatoa west of Java but Tambora east of Java, in the Sundas, that leads the list in activity and in volume of ash. A three-year eruption episode of Tambora ash seems to have caused the Atlantic crop disaster of 1816, when New England, for example, saw no single month of real summer. The older record is clear not from human witness but from geologic study. Fifteen million years ago the scablands of the U.S. Northwest were flooded within a few days by lava flows an order of magnitude greater than any listed here. The Sumatran volcano Toba erupted on the same inordinate scale 75,000 years before Tambora. Deep in the magma chamber below some long-dormant cone the slow process of gas concentration may go on now at a pace that evades our notice.

The steep ash-built mesas lying silent under the wide skies of the upper Rio Grande preserve evidence that gigantic eruptions exceeding any we know historically may terminate millions of years of volcanic dormancy. Hence this careful and comprehensive list may miss volcanoes that will erupt in 1982. Fewer than one in 10 of the volcanoes mapped here are away from plate boundaries, standing above deep old thermal "plumes" in the earth's mantle. Submarine volcanism is incompletely reported, although the hydrophones are tirelessly listening. No attention is paid to diamond pipes, which might be a rare kind of cool volcano. This work is a fascinating catalogue in which to browse, a substantial reference book and an invitation to many specialists far from volcanology to search the past.

OAKES AMES: JOTTINGS OF A HARVARD BOTANIST, 1874–1950, edited and with an introduction by Pauline Ames Plimpton. Foreword by George Plimpton. Botanical Museum of Harvard University, distributed by Harvard University Press ($12.95). The Huxleys and the Darwins and the Wedgwoods and their kin have for a long time graced the borders of the literature of science

47

with the blossoms of reminiscence and biography. Those books make vivid the experiences of a forward-looking scientific aristocracy within a social milieu of travel and growth, training and cultivation, with easy familial expectation of real accomplishment. The ambience is wealth and stability; ambition is strong but internal. It all sounds quite unlike the contemporary pattern, the Honest Jims or those Snowy manipulators around the High Table.

This delightful book is the American counterpart of the journals we have from the "backs" at the University of Cambridge. Centered in the other collegiate Cambridge, it is a selection of the diaries, letters and personal reflections of the Harvard pioneer of economic botany. Oakes Ames is the quiet, firm scholar celebrated in that classic of popular science by Edgar Anderson: *Plants, Man and Life*. Anderson was Ames's student at Harvard, touched forever by the man and his scrupulous devotion to the best. Ames held: "The fundamental implications of taxonomy should be emphasized through their broad application to human affairs."

It is easy to recognize an aristocracy: the artist Blanche Ames married Oakes Ames, a new graduate but an expert botanical amateur. Blanche's orchid drawings adorned their later monographs. (Blanche's brother Adelbert, a lawyer turned psychologist, made those remarkable studies of perspective illusions known ever since by his name.) There is not much botany in this volume, but it offers a personal view of a remarkable man in a fascinating setting. He recalls his trips to Martha's Vineyard when he was young because of the 'yellow day' when the volcanoe [sic] of Krakatoa sent its ashes around the world." He comments before Pearl Harbor that "the unending work on means to destroy aircraft" goes on uninterruptedly around his country house, part of the early work on radar and countermeasures by the Cambridge physicists. In between he sees Gillette play Sherlock on the London stage, takes the Cunarders past the icebergs in autumn crossings, insists on a small wedding for his daughter, under the wisteria in the rock garden, "as he had a tremendous prejudice against the folly of large weddings."

The editor of this vivid and agreeable book is the daughter whose college-age portrait is here. In a long embroidered dress she poses before the family mantelpiece with its Chinese porcelain figure, in an evocative period delight painted by her mother.

HANDS, by John Napier. Pantheon Books ($13.95). "When the hand is at rest, the face is at rest; but a lively hand is the product of a lively mind." The court dwarf of Velázquez looks slyly up at you while his shuffling hands cleverly deal off the bottom of the deck. The tranquillity of Dürer's praying hands or the marble hand of David in powerful repose are not so striking. These and many more hands, human and monkey, flesh and bone, are pictured here in an engaging and all too brief book by a leading English primatologist, an amateur conjurer and a physician "infatuated with the hand." His book is deliberately "not too technical," light and full of wit, intended both for the general reader in science and the scientist fluent in some other dialect of the single language of science.

The human hand is the agent of our most deliberate motor activity (speech apart), the chief organ of the sense of touch, with a convenient remote-sensing facility, and the basis of an old and subtle signal system. It is not structurally highly specialized, since many of its features are plainly ancestral and few are really newly won. Rich connections with the higher centers of the brain are its true specialization: it is a bioelectronic tour de force.

The precision and power grips (and two others), the role of the rotating opposable thumb, the rise of prehensility and the clear differentiation of the grips in the apes are a familiar story well told here, with glimpses of the bones of our forebears three million years ago in the Afar Triangle and the stocky hands of our Neanderthal cold-country cousins. Those Mousterian bones were clearly suited to the precision grip, but how much use they made of it we do not know. The varied stone-blade tool kit of that period held only power-grip instruments; it is not until the Upper Paleolithic, the time of the working of bone, antler and wood, that we become certain the precision grip came into full play. The two grips are described functionally, but in fact they are more or less discrete even anatomically. That division is neurological: the median nerve controls the muscles mainly responsible for the precision grip, the ulnar nerve the bulk of those responsible for the power grip.

The long evolution of the hand is told here in a fresh and plausible way. The predecessors of the primates were insectivorous mammals, navigating by smell, snout to the substrate, mere "noses on legs." We primates live a hand-to-mouth existence. We carry food to the mouth; they put the mouth to the food. Of course, there are nonprimate analogues: bears and otters, for instance, are also hand feeders but show strong behavioral differences from us.

A shift in diet and an increase in size were steps to the evolution of prehensility. A diet of insects pursued along the bark of trees extended to fruits, buds, leaves. The hungry animal began to venture out from the trunk into the peripheral tree crown, where the new foods are most abundant. Out on small branches small animals can move easily, but as the body size grows a point of instability is reached. Without some compensatory trick large animals cannot move deftly among the slender food-bearing branches of the tree; the choice is to stay small on top of the branch or to suspend yourself in a four-, two- or one-point tensile support, at once doubling the feeding range, now below the branch as well as above it. Then turn the forefoot into a foot-hand, and finally return in part to the ground.

Bipedalism is at least as old as the genus *Homo;* the full story of its origin is far from clear. We have the bones of the feet as rarities from long ago but still fewer bones of the ancient generic hand: they were too small to escape scavengers and decay.

Napier as a man of the hand sides with the view that humankind rose as toolmakers. That was no sudden change; it came about slowly, first through the improvisatory use of the tools at hand in nature, past the clever modification of what could be picked up, on to true planned manufacture, surpassing the apes. Here too the last word is not yet heard. The role of carrying is perhaps underestimated, although hands certainly shape that task. A hint of the complexity of the issue arises in a fine chapter on left- and right-handedness. Individual chimpanzees are left- or right-handed, it appears, although not strongly. There is a sign of a right-hand preference in the stone tools of the Lower Paleolithic. The right-handed dominance that is seen in modern human beings may have arisen from the use of shared specialized tools, such as sickles, or perhaps under the bias introduced by the carrying of infants nestled against the heart, a habit as frequent among left-handed mothers as it is among right-handed ones. The heart's comforting rhythm is familiar to the infant long before birth.

Fingerprints, gestures, nerves, bone, hand hair—all are given attention, not always as fully as one would hope, in this interesting survey. There are many morsels of wonder. Apes lack one special small muscle that helps to cup the hand, a "bona fide hallmark of mankind." We are not anatomically naked apes at all; we have about as many hairs as any chimpanzee; ours are fine, short and pale whereas the chimp's are coarse, long and dark. The palm has neither hair nor a fat layer. (Subcutaneous fat is another characteristic of our species.) Human beings, palms and all, are very sweaty animals, unlike nearly all the rest. Both a Balinese dancer and a general making his point on BBC television exhibit the function of gesture, older than language. "If language was given to men to conceal their thoughts," the author ends the book, "then gesture's purpose was to disclose them." Napier has made his own thoughts as plain as the hand before your face.

48

Sapporo. From Mitsubishi, Master Car Builders of Japan.
More standard equipment than Datsun 200SX or Toyota Celica GT.

In a land where craftsmanship is an honored tradition, where technology has reached superior levels, Mitsubishi builds Sapporo...the totally equipped personal luxury car. At $8,036,* Sapporo offers more than any other import in its class. Elegant, adventurous styling. Highly responsive 2.6 liter engine performance. An exceptional 36 est. hwy., [24] EPA est. mpg.† And more than 40 standard equipment features including full instrumentation, 5-speed manual transmission, AM/FM stereo with four speakers. Sapporo. By Mitsubishi. A unique blend of traditional craftsmanship and advanced technology.

*Base sticker price excluding title, taxes, and destination charge. Road wheel package as shown, $365 extra. †Use EPA est. mpg. for comparison. Mileage may vary depending on speed, weather and trip length. Actual highway mileage probably less. California mileage: 34 est. hwy., [22] EPA est. mpg.

BUILT BY MITSUBISHI   IMPORTED ONLY FOR PLYMOUTH

Plymouth

HOW TO BE THE GUEST OF HONOR.

REMY MARTIN
COGNAC

FINE
CHAMPAGNE
V.S.O.P.

Rémy

THE FIRST NAME IN COGNAC SINCE 1724
EXCLUSIVELY FINE CHAMPAGNE COGNAC FROM THE TWO BEST DISTRICTS OF THE COGNAC REGION

The Social Security System

The viability of the pay-as-you-go system is threatened by economic and demographic trends. The retired population and its benefits are growing faster than the active work force and its tax contributions

by Eli Ginzberg

On the third day of each month some 35 million green U.S. Government checks are delivered to homes or deposited directly in banks across the nation. The checks, which averaged just under $300 each in 1979, represent the monthly installment in the country's largest system of transfer payments: the Social Security system. The total benefits paid out in 1979 to 22 million retired or disabled workers and 13 million dependents and survivors came to more than $104 billion. Unlike a funded pension system, in which contributions to the system are invested to pay future beneficiaries, Social Security is founded on a pay-as-you-go approach. This means that the taxes employees and employers pay into the system provide the funds to cover current payments to beneficiaries. These payments amount to more than 6 percent of the total disposable income of individuals in the U.S.

Transfers of income through the Social Security system fulfill an implicit contract between younger people currently in the active work force and people no longer working. This intergenerational compact has been in effect since 1940, when the first person to qualify for Social Security benefits retired. Now the financial soundness of the contract is in question. Indeed, for much of last year the country was given to believe the Social Security system faced imminent bankruptcy. It did not, but in the absence of remedial actions taken by Congress in the fall it would have developed a serious cash-flow problem in the next year or so. Nevertheless, the cash-flow problem was a manifestation of an underlying imbalance between revenues and expenditures. The success of the remedy for the immediate difficulty

should not divert attention from the root causes of the imbalance and the long-term need for a thorough reform of the system.

The sources of the difficulties can readily be identified. Since Social Security is a pay-as-you-go system, the taxes paid by employees and employers in, say, 1950 provided the cash for 1950's disbursements to beneficiaries, and this year's tax receipts will pay this year's benefits. Such a system is at the mercy of economic and demographic forces. For a long time those forces worked to the advantage of the system. As the U.S. economy expanded, employment increased and wages at least kept up with inflation; a high birthrate (which peaked during the baby boom of the 1950's) and a rise in women's employment brought into the labor force many more new workers than were lost through retirement. As a result income from payroll taxes exceeded benefit payments.

Now all of that is changing. Wages (which, together with the level of employment and the tax rate, determine payroll-tax receipts) have been lagging behind inflation while benefits (which are price-indexed) are protected, even overprotected. What is more important in the long run, lower fertility and greater longevity are changing the age structure of the U.S. population, increasing the number of potential beneficiaries faster than the number of workers who will be asked to support them. An 18-year-old who enters covered employment today is not scheduled to retire until the decade of the 2020's; the system may have to provide benefits for him and his survivors until the 2040's. As one analyst put it, 85 percent of those who will receive benefits in the next 75 years are alive today; 96 percent of all

Social Security benefits to be paid during the next 75 years will go to people who are alive today; 99 percent of the payroll taxes to be paid in the next 25 years and 81 percent of the taxes to be paid in the next 50 years will be paid by people alive today.

Congress has no option but to consider how it is going to cover the liabilities it is assuming today. Before the year 2010 the baby-boom cohorts will begin to retire. We therefore have some 30 years in which to restructure the Social Security system.

The U.S. came late to the idea that people should look to the central government to protect an individual against the loss of income as a result of age or disability and to provide for his dependents in the event of his death. Germany had established a government-sponsored pension program in 1889, the U.K. in 1908 and France in 1910. The Social Security Act was passed in 1935, in the heyday of the New Deal. President Roosevelt decided (over the objections of some advisers) to limit coverage at first to workers in industry and commerce and to limit their benefits to cash payments after they stopped working at the age of 65, a retirement age chosen with less than full deliberation. It was generally understood, however, that the system would subsequently be broadened. Even before the first benefit checks were issued Congress amended the law to provide benefits for the aged wife and the children of a retired worker and to the widow and young children of a covered worker on his death.

Two major programs were grafted onto the original system. In 1956 disability insurance was added for covered

51

workers who become unable to work because of sickness or injury. In 1965 Medicare was added. It has two components: hospital insurance and supplementary medical insurance, which help patients to cover physicians' charges. The original nucleus of the Social Security system is formally called Old Age and Survivors' Insurance (OASI). When disability insurance is included, the program is designated OASDI, and with hospital insurance OASDHI. All these programs are financed by payroll taxes. Supplementary medical insurance is funded separately. It is an optional program, although most Medicare beneficiaries elect to participate in it; premiums paid by the beneficiaries now cover 30 percent of the cost and the remainder is paid from the general revenues of the Federal Government.

About 90 percent of all currently employed people are covered by the system. They are subject to a payroll tax (now 6.7 percent of their wages up to $32,400), which is matched by a tax paid by employers. A large proportion of the more than eight million self-employed people are covered; they pay 75 percent of the combined employee-employer rate. The largest category of people not covered by Social Security are government employees, for whom there are 68 Federal retirement plans and thousands of other plans for employees of state and local governments. Workers in some nonprofit institutions are also not covered.

From its inception the Social Security system was designed to meet two conflicting objectives. The basic intention was to establish a compulsory system of mandated contributions that would provide income after retirement. In keeping with a national predilection for "paying one's own way" and to gain broad public support, the benefits were to be related to the amount of "contributions." The more tax money that was paid by a worker, the higher his benefits would be at retirement.

Another intention, however, was to ensure that even people who had earned low wages would have enough income in old age to meet their basic needs. That could not be accomplished if benefits were directly proportional to tax payments; the benefit-to-tax ratio had to be skewed in favor of low-wage earners. Today the average unmarried male worker who retired in 1979 after having earned wages at the maximum taxable level will have received by the time of his death 4.97 times as much in benefits as he paid in taxes; the ratio of total benefits to total taxes for an unmarried low earner is 7.07 for men and 8.82 for women. For a retired person with a dependent spouse the benefit-to-tax ratio rises to 9.19 for a maximum earner and 13.06 for a minimum earner. What has made such high ratios possible, of course, is the pay-as-you-go scheme, combined with the economic and demographic forces I mentioned above. (It is often pointed out, incidentally, that the Social Security tax is regressive: not only is the tax rate the same at all income levels but also the fraction of the income on which the tax is paid is larger for a low earner than it is for a high earner. On the benefit side, however, the system is clearly progressive: it favors low earners.)

A retiring worker's benefit is based on his earnings in jobs subject to the payroll tax, except that for years when the earnings were higher than the maximum amount taxed, only that maximum is included. To calculate the benefit the earnings are first adjusted for the increase in wage levels since the money was earned. After a certain number of lowest-pay years are excluded the wage-indexed annual earnings up to age 62 for retiring workers are divided by 12 and averaged. Then a benefit formula is applied to the average indexed monthly earnings to determine the "primary insurance amount" (PIA). The formula is complex, but in essence what it does is convert successively higher increments of earnings into the PIA at successively lower percentages. Thus the benefit for a low-wage worker is a larger percentage of his average wage than the benefit for a high-income worker is. That overall percentage is called the replacement rate. It ranges from almost 70 percent for the lowest-wage earners to about 33 percent for someone who has earned the maximum taxable wage each year; the average replacement rate is 40 to 50 percent. The PIA is the benefit for a retired 65-year-old worker without dependents. It is adjusted if he retires before or after 65 or if he has a spouse or dependent children.

To the extent that Social Security is considered a social insurance plan, people who contribute to it are entitled to draw benefits. The concept of entitlement has protected Social Security from political attack (as President Roosevelt foresaw it would); it has also discouraged any reduction in benefits once they are granted. (It seems to me, however, that the high ratio of benefits to tax payments argues against the idea that any reduction in benefits would constitute a breach of faith with those who have been paying the tax, particularly if the adjustment is made after due warning.)

The concept of the system as contributory insurance has been reinforced by the decision to rely exclusively on the payroll tax as the source of revenue to pay OASDHI benefits. Most of the planners of the system assumed that in time there would be partial recourse to general revenues. They had in mind an eventual three-way division, characteristic of most European pension systems, in which employees, employers and the central government each contribute a third of the necessary funds. The direct linkage between taxes and benefits has helped Congress resist pressure to raise benefits without consideration of how they would be financed.

Social Security cannot be discussed without attention to other measures intended to provide income maintenance. The same motives that led the Federal Government to pass the Social Security Act led to the establishment of various programs providing cash or other benefits for people whose income is too low to meet their basic needs. Whereas Social Security benefits are paid regardless of need to all qualified workers whose earnings have been taxed, most other public programs are means-tested: benefits go only to those who have a demonstrated need.
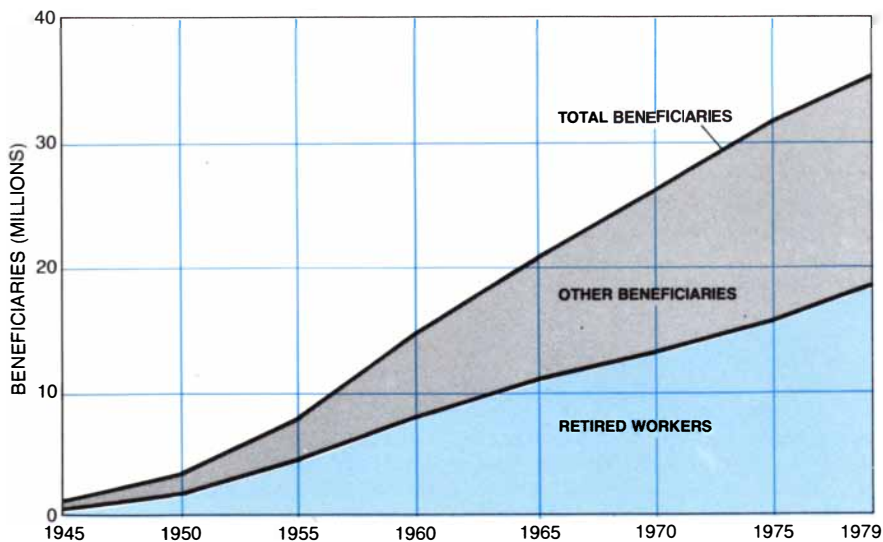
NUMBER OF BENEFICIARIES of Old Age, Survivors' and Disability Insurance (OASDI) has grown steadily since the first covered worker retired in 1940. Most beneficiaries are retired workers. Others include disabled workers, dependent spouses and children, and survivors.

52
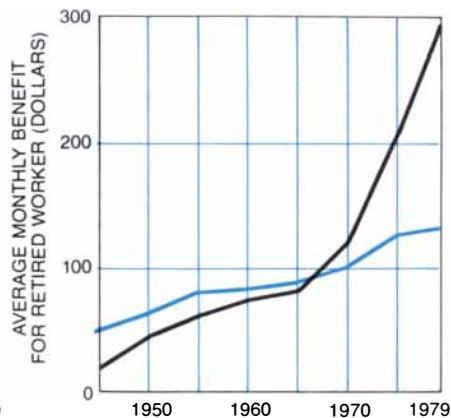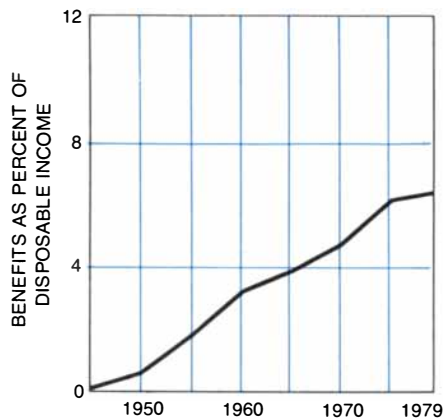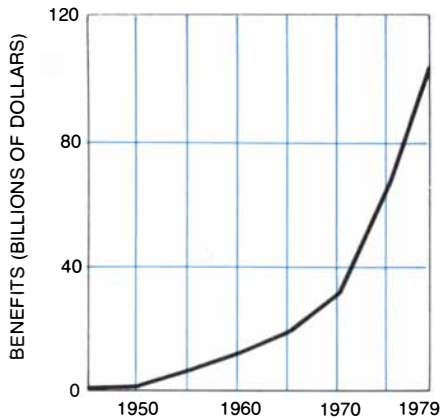
ANNUAL BENEFIT PAYMENTS for OASDI have increased over the years both in their dollar amount and as a percentage of the total U.S. disposable income. The effect of the increases that were legislated in the 1970's can be seen. The average monthly benefit for a retired worker has also risen, but the inflation-adjusted value of the benefit (in 1967 dollars) has risen much more slowly (*color*).

The means-tested programs began modestly in the 1930's when the Federal Government joined with the states to give aid to families with dependent children (AFDC), which still accounts for the bulk of what is referred to as welfare payments. Assistance was also provided for the blind and later for disabled and needy old people excluded from or not adequately supported by Social Security; in the early 1970's the Federal Government assumed responsibility for these groups with the Supplemental Security Income program. In 1965 the joint Federal-state Medicaid program was established for people on welfare and for others defined by each state as medically needy. Other in-kind programs aim to provide adequate nutrition (food stamps and other assistance) and housing (rent subsidies) for low-income people. In 1979 the total Federal outlay for means-tested programs was $35 billion, compared with the total outlay for OASDI of $104 billion.

In spite of the country's commitment to rely on social insurance to cover the basic needs of those unable to work be-

cause of age or disability, more than two million OASDI recipients in 1979 needed supplemental income. When all income from all sources, including Social Security and means-tested programs, cash as well as in-kind, are taken into account, no more than 3 or 4 percent of the elderly fall below the poverty line. About a third of the elderly, however, are not much above that line. Of course, Social Security benefits were never meant to be the sole source of income for older people. Most aged families with a satisfactory standard of living receive less than half of their income from OASDI benefits; they rely also, in different proportions at different income levels, on continued earnings by family members, on savings and on private pensions.

Private pension systems antedated the establishment of Social Security, although they became a significant factor only during and after World War II. Between 1950 and 1975 the number of workers covered by private plans grew from 9.8 to 30.3 million, the number of monthly beneficiaries from 450,000

to more than seven million and annual benefit payments from $370 million to $14.8 billion. In 1979 almost two-thirds of all male workers and half of all female workers were covered by a private plan. Nevertheless, in 1980 such plans accounted for only 14 percent of all retirement, disability and survivors' benefits; 78 percent came from Social Security and other Federal retirement systems and 8 percent came from state and local systems.

The fundamental importance of the Social Security system is clear. It has often been stated (and it is probably true) that the system is the Government's most successful social program. It is all the more imperative, then, that the system be kept financially sound.

The cash-flow crisis that was imminent last fall affected only one part of the OASDHI system. Under current law benefits for each Social Security program can be paid only from the specific trust fund for that program. It is the OASI fund that is in trouble. At the beginning of 1960 the money in that fund
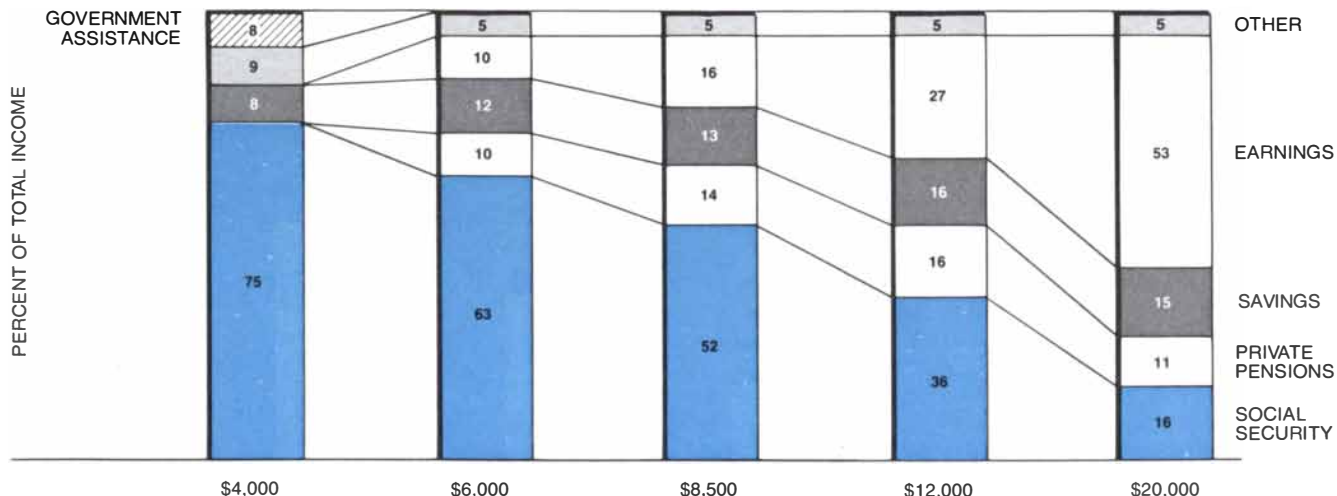
IMPORTANCE OF SOCIAL SECURITY as a source of retirement income declines with increasing total family income, as is shown for five income levels in 1978. For $4,000-a-year families means-tested government-assistance programs provided 8 percent of total income.

53

amounted to 195 percent of the total outlay for 1960; last October 1 (at the beginning of the 1982 fiscal year) the balance was down to 14 percent of the estimated expenditure for 1982. The Congressional Budget Office has a rule of thumb: the starting balance each year must be at least 9 percent of the year's expected outlay in order to cover benefits paid early in the fiscal year, before current payroll-tax money is collected. In the absence of congressional action the OASI balance would have fallen to less than 5 percent by next October.

The primary cause of the decline in the trust-fund balances in the 1970's was a series of three large increases in benefits: 15 percent in 1970, 10 percent in 1971 and 20 percent in 1972. In addition in 1972 the escalation of benefits was

institutionalized: Congress decided that benefits should routinely be raised to match any increase of more than 3 percent per year in the Consumer Price Index. As a result benefits have been raised every year since 1975. The increases were to be financed by raising the taxable wage base as current average wages rose. Prices outstripped wages, however; indexed benefits rose faster than the indexed wage base.

Faced with the combined effects of benefit increases, high unemployment and slow economic growth, Congress (having no inclination to alter reliance on the payroll tax and being loath to reduce benefits) enacted a series of increases in the total OASDHI tax rate in 1977. In conjunction with the rising tax base the rate increases were expected to

ensure the financial stability of the system for the remainder of the century. It was not enough. Less than three years later Congress had to transfer some tax receipts from the disability-insurance fund to OASI. Then last fall it had to take emergency action again, diverting some future tax receipts from the disability and hospital-insurance funds to OASI and also allowing for the possibility of interfund borrowing. According to projections made by the Congressional Budget Office, however, neither interfund borrowing nor even a complete merger of the trust funds will sustain the system for long. By the beginning of fiscal year 1985 the combined balance for all three funds will fall below 9 percent of expected expenditures.

What then? Last February the Con-
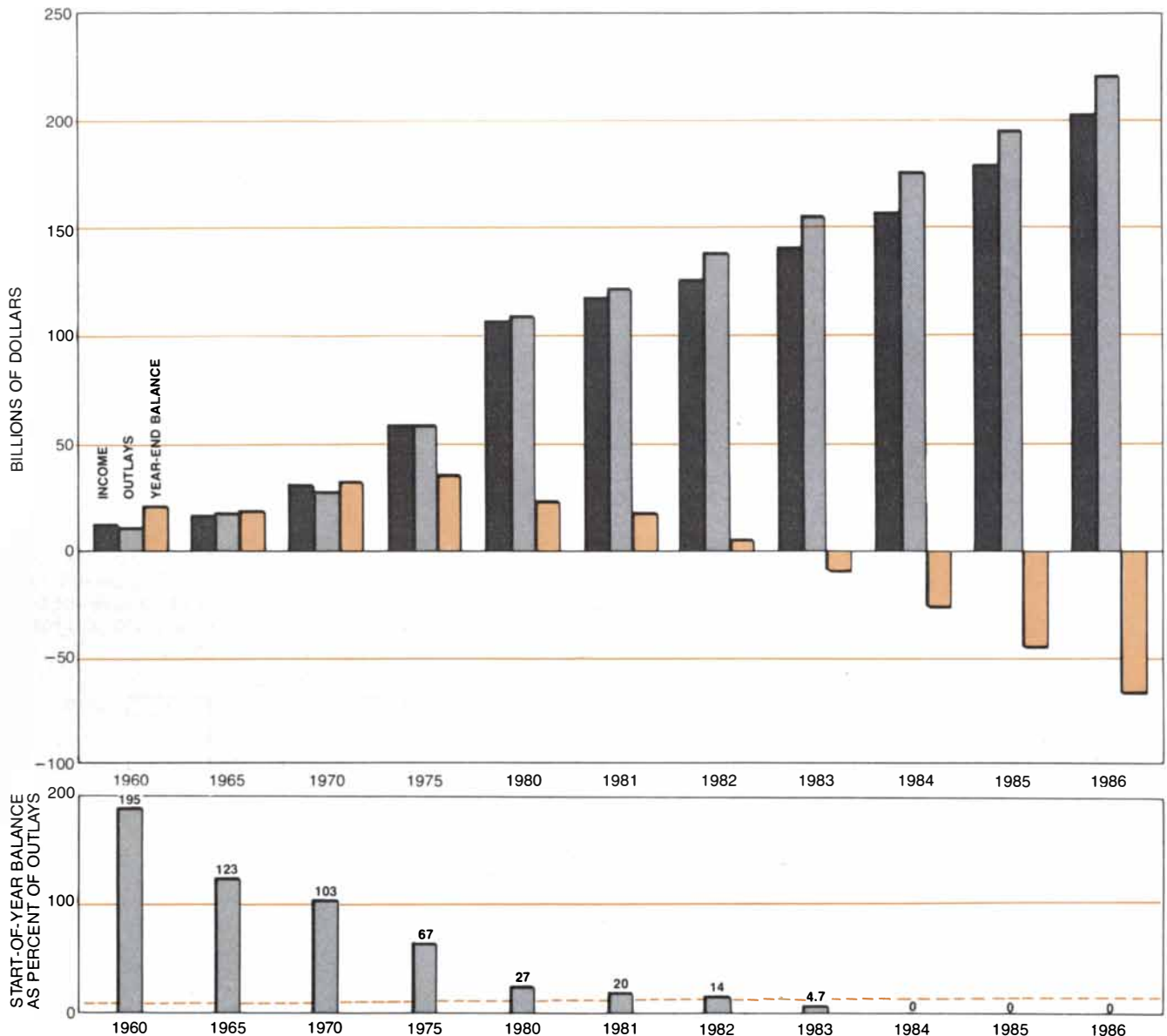
OASI TRUST FUND, from which benefits are paid to retired workers, dependents and survivors, has recently spent more annually than it has received in tax revenue. The upper chart shows how annual outlays have exceeded income, reducing the year-end balance and leading to the expectation of negative balances in the mid-1980's. The lower chart gives the fund's balance at the beginning of each fiscal year as a percentage of that year's outlays. (Up to 1980 the percentages are based on actual expenditures; for 1981 through 1986 they are based on estimates made by the Congressional Budget Office.) A beginning balance of 9 percent is considered essential if the system is to meet its obligations. Next year the balance would have been below that level (*broken line in color*) if Congress had not acted last fall.

54

gressional Budget Office identified a number of alternatives that in various combinations could carry the system through the 1980's. For example, many economists now think cost-of-living increases tend to overcompensate for inflation and that they could be modified. The minimum benefit, which increases the payment that would otherwise be received by certain very low earners, could be eliminated. Congress has already eliminated the "lump sum" death benefit and special benefits for dependent students.

The Budget Office has also suggested a number of ways to increase revenue, such as allowing the system to borrow from the U.S. Treasury, accelerating scheduled tax increases, raising the tax base to include all earnings and raising the self-employment tax. Hospital benefits might be paid in full or in part with money from specifically earmarked income-tax receipts so that the hospital-care payroll tax could be allocated to OASDI. Congress could consider forgoing some future income-tax cuts and instead allocating some portion of any proposed cut to the Social Security trust funds. Others have suggested that half of each benefit payment should be subject to the income tax: the half financed by employers' contributions, on which no tax was ever paid. Although beneficiaries with low incomes would not pay the tax, the yield would still be about $4.5 billion per year.

Given all these options, it seems most unlikely that the system will fail to meet its obligations in the next few years. Combined with the action already taken by Congress, adding only .5 percent to the total payroll tax rate would right the balance between revenues and expenditures unless the economy goes into a long decline.

The real challenge to Social Security lies in the future. The proportion of the U.S. population 65 years old and older has been growing (from 9.2 percent in 1960 to 11.2 percent last year) and will continue to grow (to an estimated 12.2 percent in the year 2000). Retired people and their spouses are living longer: in 1950 the average life expectancy at 65 was another 13 years for men and 15 years for women, whereas in 1980 it reached 15 years for men and 20 years for women. The key factor in determining the financial integrity of the Social Security system is the relation between the number of workers paying taxes and the number of retired workers, dependents and survivors entitled to benefits. In 1945 there were 50 workers paying taxes to support current payments to each beneficiary; now each beneficiary is supported by about three workers, and in 2035 the ratio may be less than two taxpayers per beneficiary.

The long-term outlook has been appraised by the secretaries of Health and
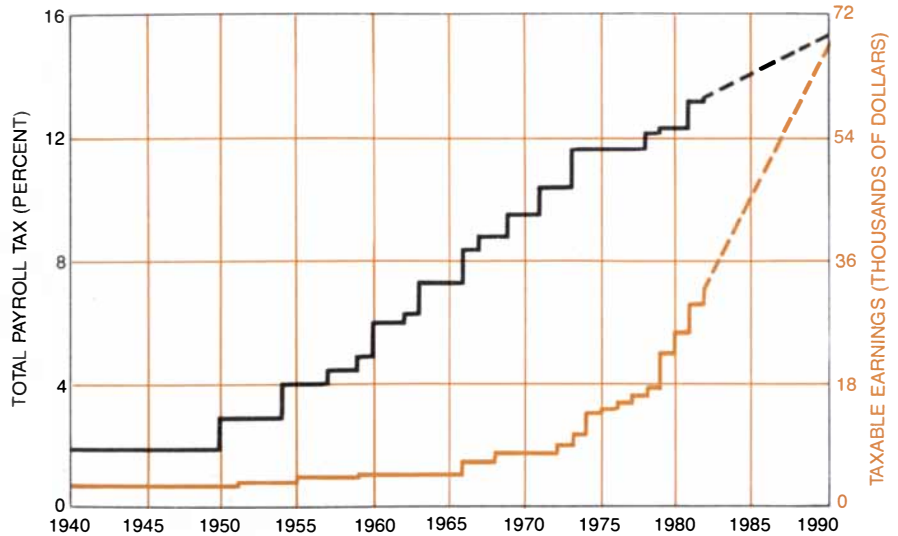
PAYROLL TAX RATE (*black*) and the maximum earnings, or tax base, on which the tax is paid (*color*) have been increasing and will continue to do so. Future increases in the tax rate have been specified by legislation. The maximum tax base, however, has been indexed to the average U.S. wage level since the early 1970's; the future increases given here are estimates.

Human Services, Labor and the Treasury, who serve as the trustees of the Social Security system. In their 1980 report they estimated the average annual balance of revenue and expenditures for the next 75 years, assuming that there will be no change in the currently prescribed tax-base and tax-rate increases. The estimates were made under three sets of assumptions: optimistic, intermediate and pessimistic. Between 1980 and 2004, the trustees reported, the system as a whole should have an annual surplus of revenue over expenditures. In the next 25-year period the intermediate projection puts the annual deficit at about $12 billion. Between 2030 and 2054 the intermediate projection puts the deficit at about $46 billion per year.

The most obvious cure for a deficit is an increase in taxes. Last March the National Commission on Social Security estimated the increase in tax revenues that would be needed to keep the system in balance, assuming that some modest benefit changes recommended by the commission are enacted by Congress. The total tax (the employee's share plus the employer's) is already scheduled to rise from 13.4 percent this year to 15.3 percent in 1990. Additional increases would be necessary, raising the total to 15.7 percent between 2005 and 2009, to 20 percent between 2020 and 2024 and to 23.8 percent between 2035 and 2039. The commission saw two dangers in allowing the payroll tax to rise to these levels: the tax bite might be too big for young workers and other workers earning low wages, and the employers' share of the tax might discourage hiring. The commission therefore recommended that the payroll tax be limited to 18 percent, with the difference between that amount and the recommended tax level being covered from general revenues.

Is a rise of nearly 80 percent in the Social Security tax burden out of the question? Clearly the commission did not think so. The social security systems in some countries of Western Europe have a payroll-tax rate (employee plus employer) of between 18 and 28 percent. In return for higher taxes they supply more comprehensive benefits. Contributions for other public and private social programs can push the total levy in those countries up to 50 percent of a worker's earnings.

Before considering further the financing of the Social Security system, it should be asked whether the benefits now provided by the system are adequate. I would submit that there is room for improvement in the system, and the improvements would add to future costs. Among the conspicuous problems are the inadequacy of the present minimum benefit, inequities for women and the ineffective response of Medicare to the health-care needs of the elderly.

Every retired worker eligible for any benefits receives at least $122 per month. Those who get this minimum benefit are mainly women who worked for very low wages. In 1979 the Advisory Council on Social Security Financing and Benefits held that "single people who have worked full time at the Federal minimum wage do not now receive a benefit sufficient to keep them out of poverty." The Advisory council recommended corrective action. President Reagan went in the opposite direction: he recommended eliminating the minimum benefit on the ground that it is a windfall for people who have retired on other Federal pensions but who also barely qualify for Social Security. Congress went along with the President in the first round of budget cutting but re-

55

stored the minimum payment last fall. It remains inadequate.

When the Social Security system was established, it dealt with women primarily as dependents or survivors. That has changed with the entry of large numbers of women into the work force. By 1978 two in every five female beneficiaries were receiving benefits by reason of their own work, not as a consequence of their marriage; the proportion is likely to increase to 80 percent for women who reach 65 in 2000. Women workers pay the same payroll tax as men, but many women receive higher benefits as dependents than as retired workers because the replacement rate is skewed to their disadvantage. One perverse result is that a 65-year-old couple get 17 percent more in benefits if only one of them was a wage earner than they do if both earned equal amounts. If a woman is divorced before the 10th year of a marriage, she receives no benefits when her former husband retires; if the marriage lasts for 10 years, she receives the full spouse's benefit.

One proposal that might lead to fairer treatment of women is an earnings-sharing plan. Benefits would be based on the combined earnings of the couple, whether or not both had worked and regardless of how much each had earned and therefore paid in Social Security taxes. In effect all earnings would be vested equally in the husband and the wife. The Advisory Council favors moving toward such a plan but suggests that more study is needed to be sure dependent's and survivor's benefits now accruing to many women are not jeopardized. For now the council has recommended two steps. It would allow a surviving spouse to have benefits based on the couple's combined earnings, and it would adopt a limited form of earnings sharing for divorced couples.

Medicare now covers only about 40 percent of the medical costs of the elderly. It largely covers acute hospital care and, for those who elect to buy supplementary medical insurance, physicians' services. Each year a considerable number of older people have very large medical expenses (more than $2,500 or $3,000) that they must pay from their own funds. Much of the care required for chronic conditions is simply not covered. Medicare makes no contribution toward helping a patient remain in his own home with home nursing or attendants when he is chronically sick or feeble, and it makes no significant contribution toward paying the cost of his staying in a nursing home. In failing to support care in the patient's home or in a nursing home Medicare encourages unnecessary hospital stays and thus runs up the country's total health-care bill. In spite of these serious shortcomings, the hospital-insurance tax rate will increase sharply beginning in the 1990's.

The demographic imperatives and the need for improvements in Social Security benefits provoke a series of questions. The first is: Can payroll-tax increases sustain the system into the middle of the next century? A moderate increase in the tax rate is certainly feasible, and the tax could be imposed on all earnings. Because the rate is the same for both low earners and high earners, however, the tax bite would soon become too deep for some employees. As for the employer's share, there is no doubt that too high a payroll tax can inhibit employment, creating serious problems for the economy and the society; this may already have happened in some countries of Western Europe. The tax rate, then, cannot be raised indefinitely without such adverse consequences as depressing workers' wages, enlarging the off-the-record economy and decreasing the total tax revenues collected.

Is there some other way to enlarge Social Security revenues? One good opportunity, which should be pursued aggressively, is to bring into the system as many as possible of the workers in government and in nonprofit institutions who are not now included. Uncovered groups make up about 10 percent of the work force. Perhaps workers currently covered by state and local plans could not be compelled to join the system, but Congress could certainly extend coverage to all Federal employees. Complex adjustments would be needed to protect the interests of such employees; the difficulties are formidable but not insurmountable. Many experts on the Social Security system advocate starting such a program of consolidation soon, arguing that no other system affords workers and their dependents as broad a range of basic benefits. The National Commission on Social Security has calculated that universal coverage would add an average of $5 billion per year to the revenues of the system over the next 75 years. Of course, it would also increase liabilities.

As for expenditures, is there any equitable way the system's future obligations might be significantly reduced? There is. A growing consensus among those who have studied the system in recent years holds that the standard age of retirement should be increased from 65 to 68. The change would be doubly effective in that it would augment the population of taxpayers while reducing the population of beneficiaries. It should be done only after adequate warning and gradually, perhaps at the rate of only a few months per year. The National Commission estimates that the change would save about $11 billion per year from 1980 to 2055.

The system now tends to promote early retirement. A worker can retire at 62 and get only 20 percent less per year than he would have received at 65. As a result about three in four workers retire before they reach 65. President Reagan proposed changing the early-retirement reduction to 45 percent but withdrew the plan when it met vigorous public opposition.

The increase in early retirement is paralleled by increasing unwillingness of most people to work beyond age 65. Whereas one man in three and one woman in 10 worked after the statutory retirement age in 1960, only one man in five and one woman in 12 does so now. This decline in labor-force participation, combined with increases in longevity, generates pressure on the system. In 1940 only 583 men per 1,000 and 687
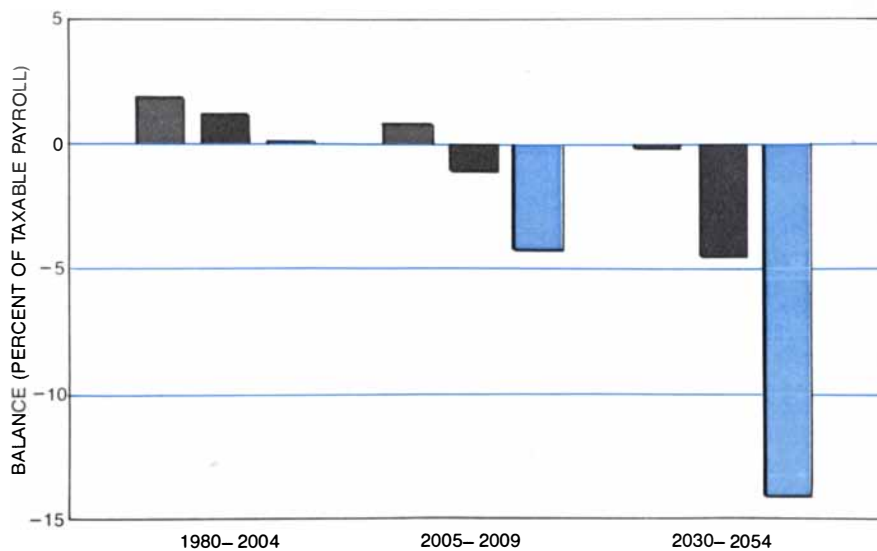
AVERAGE ANNUAL BALANCE for the Social Security system as a whole has been estimated for three 25-year periods under optimistic (*gray*), intermediate (*black*) and pessimistic (*color*) assumptions. The estimates were made (and are shown here) not in dollars but as a percentage of the total taxable payroll, so that they are applicable regardless of inflation. Percentages can be translated into current dollars for the period in question: 1 percent is $10 billion.

women per 1,000 lived to be 65; today the proportions have risen respectively to 711 and 839 per 1,000. Of all people 65 or over fewer than a third were over 75 in 1974; by the end of the century the fraction will be close to a half.

As I indicated above, any increase in the retirement age would have to be phased in gradually. Rita R. Campbell of the Hoover Institution at Stanford University has suggested that allowances might be made for workers who have been employed continuously for as long as 45 years; many private pension plans, she points out, consider the number of years worked as well as age in determining the time of retirement. Furthermore, many people have health problems that make it inadvisable for them to keep working beyond their early sixties, particularly if the job is physically demanding; the definition of disability might be modified to cover such cases.

One alternative to raising the retirement age (or perhaps a companion measure) would be to adopt a recent Swedish approach called phased retirement. Older people can choose a shortened or gradually declining work week instead of going from full-time employment to full retirement in one day.

Might it be possible to limit the increase in the obligations of the Social Security system by shifting some of the burden to alternative sources of income for retired people? Last year the President's Commission on Pension Policy recommended that a compulsory system of private pensions be established. Employers would pay an amount equal to 3 percent of their payroll; changes in tax policy would ease the burden on small businesses. The commission also urged that the tax structure be modified in ways calculated to encourage personal saving.

There is little likelihood that a compulsory private pension plan will be approved by Congress. For one thing, there is continuing mistrust of existing private plans; although Congress attempted in 1974 to regularize such plans and expand their coverage, many workers are still left either without a pension or with a very small one. Private pension plans already cost the U.S. about $15 billion per year in lost tax revenues. Owners of small businesses, many of whom operate close to the margin of profitability, would oppose a compulsory plan bitterly. Effective social control over the reserves of large private pension funds is already a problem; a new program would exacerbate it. Moreover, what is the point of a new compulsory system if the present compulsory system—Social Security—can be made to work?

Many economists contend that the level of personal saving in the U.S. is too low and that this accounts in large part for a deficiency in capital that will have
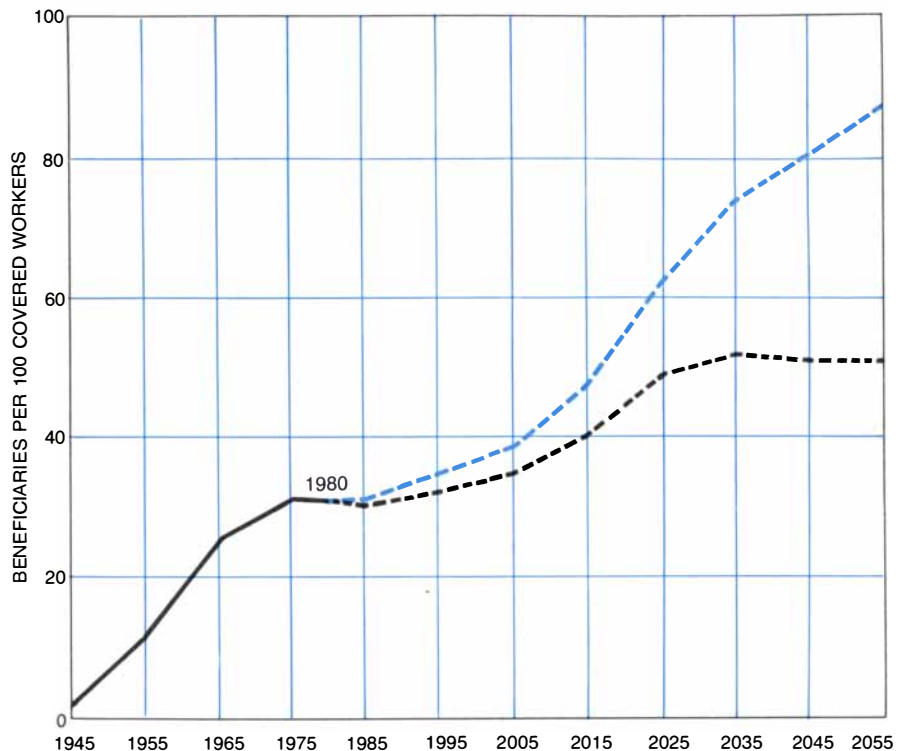
RATIO OF BENEFICIARIES TO WORKERS is the key datum affecting the system's financial soundness. The ratio has increased from two beneficiaries per 100 workers in 1945 to 31 today. Future increases, implying a growing imbalance between benefits and revenue, are projected to 2055 (*broken lines*) on intermediate (*black*) and pessimistic (*color*) assumptions.

to be remedied if the country is to modernize its plant and equipment. (Martin S. Feldstein of Harvard University undertook some years ago to demonstrate that the Social Security system is directly responsible for the low level of savings, but many economists think he is wrong.) Whether or not additional saving is needed to provide adequate capital for industry (and some economists, including Thomas Juster of the University of Michigan, contest that assumption), Congress has already provided tax advantages for individuals who save for their retirement; the Keogh Plan and the individual retirement account (IRA) are two examples.

Additional tax incentives to encourage more people to save for retirement will probably be forthcoming, but only upper-income people can take full advantage of such opportunities. Any serious attempt to get middle- and low-income groups to divert more of their earnings into savings might require Government subvention and protection against inflation for small savers, as is done in Germany. In any case, except for the very poor, private pensions and personal savings already contribute between a fourth and a third of the income of older people. That fraction is unlikely to be raised by more than a few percentage points. There is no alternative, then, to ensuring the financial integrity of the Social Security system so that it can provide basic income for older people.

To accomplish that, it seems to me, the following actions are needed. The Social Security system should be expanded as quickly as possible in order to provide universal coverage, which would add significantly to revenues and also broaden the range of benefits for people now outside the system. An early decision should be made to raise the retirement age gradually, probably to 68. The half of benefits that was based on income and that was never taxed should be made subject to the income tax, but only for those older people whose total income is high enough to make them liable to the tax. It must be determined what changes in Medicare would enable it to provide an acceptable level of medical care for the elderly and to do so more efficiently. The unconscionable inequities imposed on women as workers and as wives should start to be redressed. The benefits for long-service, low-wage workers, which have been adjusted in the recent past, need further adjustment to bring them above the poverty line.

The Social Security system can provide basic support for all older Americans, but only if it is not overloaded. That is why the key reform is to encourage people to work longer, thereby ensuring a comfortable income when they stop working. That is a sensible national goal, and not only for economic reasons. As the Talmud puts it: When you stop working, you are dead.

57

The Mass Extinctions of the Late Mesozoic

Many species of plants and animals, notably the dinosaurs, suddenly died out about 63 million years ago. A thin stratum of iridium-rich material suggests the cause may have been the fall of an asteroid

by Dale A. Russell

One of the most striking events in the record of life on our planet is the simultaneous disappearance at the end of the Mesozoic era, some 63 million years ago, of many kinds of reptiles, certain kinds of marine invertebrates and certain kinds of primitive plants. For generations scholars have sought unsuccessfully to explain this event. New evidence, however, has now led to a novel hypothesis: The disappearances were the result of a catastrophic disruption of the biosphere by an extraterrestrial agency.

Catastrophism is not a new doctrine in efforts to account for episodes in the history of the earth, but it has not been a particularly popular one. Early in the 19th century, when geology was in its infancy, the French anatomist Georges Cuvier suggested that the past had been marked by a series of environmental "revolutions," or catastrophes. In his view such disruptions would account for three animal disappearances: that of the mammoths at the end of the ice age, that of the many primitive mammals fossilized in rocks lying deeper than the ice-age gravels and that of the giant reptiles fossilized in chalk beds lying deeper still. In the decades that followed, however, the work of such pioneer geologists as Charles Lyell made it apparent that the processes of change in earth history were of far greater duration than Cuvier had believed. Catastrophism fell from favor, to be replaced by the doctrine of gradualism. For more than a century now paleontologists have generally agreed that whatever may have caused the disappearances at the end of the Mesozoic era, it could not have been a worldwide catastrophe.

The principal casualties among the reptiles were the dinosaurs. As an example, late in the Cretaceous period, the closing chapter of the Mesozoic, at least 15 separate families of dinosaurs, possibly representing between 50 and 70 distinct species, inhabited North America. In the rocks that were formed immedi-

ately after the Cretaceous not one dinosaur skeleton has been found. That is why the end of the Mesozoic is generally characterized as the time when the dinosaurs became extinct. The dinosaurs were not, however, the only organisms to disappear. Among the 33 other families of reptiles that inhabited North America late in the Cretaceous there were the following losses. All four of the families of marine turtles died out (although three of the four survived elsewhere). One of the three families of crocodilians, the Goniopholidae, also died out. So did two pterosaur (flying reptile) families, the Ornithocheiridae and a family still unnamed, two ichthyosaur (marine reptile) families, the Platypterygiidae and another family still unnamed, and all three of the plesiosaur (also marine reptile) families, the Elasmosauridae, the Polycotylidae and the Cimoliasauridae. And so did two of the eight families of lizards, the Polyglyphanodontidae (primitive skinklike land forms) and the Mosasauridae (large marine forms).

What happened? Was there a gradual or a catastrophic extinction? My own interest is primarily in the larger reptiles of the Mesozoic in North America, and so the examination of these questions I shall undertake here will focus mainly on the disappearance of those animals. Among the many hypotheses put forward to account for their disappearance are disruptions of the food chain, both at sea and on land, a general alteration

of the environment as the sea level began to drop at the end of the Mesozoic, a sharp rise in temperature, a fall in temperature caused by volcanic dust in the atmosphere, and so on. None of these phenomena, however, would seem by itself to be a convincing cause of the reptilian extinctions.

In 1979 paleontologists interested in the problem were presented with a new possibility. A group of workers at the University of California at Berkeley—the geologist Walter Alvarez, his father, the physicist Luis W. Alvarez, and two physical chemists, Frank Asaro and Helen V. Michel—announced the discovery of abnormally large traces of the heavy element iridium in a marine formation near Gubbio in the Apennine mountains of Italy. The iridium was concentrated in a layer of clay, one centimeter to two centimeters thick, that separates marine limestone of late Cretaceous age from an overlying marine limestone of early Paleocene age. The limestone below the clay contains fossil marine organisms typical of the latest part of the Cretaceous. No organisms are preserved in the clay. In the limestone above the clay the Cretaceous organisms are absent; they have been replaced by other organisms typical of the Paleocene.

Iridium is one of several elements geologists call siderophiles, "iron lovers." It is rarely present in the rocks of the earth's crust but is comparatively abun-

LAYER OF CLAY in the photograph on the opposite page is two centimeters thick. It separates two beds of marine limestone exposed near Gubbio in the Apennine mountains of Italy. The white limestone below the clay is late Mesozoic in age and the grayish pink limestone above it is early Cenozoic. Analysis of the clay showed it to be 30 times richer in the heavy element iridium, deposited on the earth's surface in meteoritic material, than the clays from adjacent marine strata. This had led the geologist Walter Alvarez, his father, the physicist Luis W. Alvarez, and two chemists, Frank Asaro and Helen V. Michel, all of the University of California at Berkeley, to the hypothesis that the surplus iridium came from an extraterrestrial object, perhaps an asteroid-size meteorite, that crashed into the earth at the end of the Mesozoic. The Alvarez group further hypothesizes that the collision was the cause of the many extinctions of marine and terrestrial organisms at that time. The coin is the size of a 25-cent piece.

ERODED SEDIMENTS of the late Cretaceous near Fort Peck Reservoir in Montana are 100 meters thick. They were deposited on a coastal plain near the shore of the great inland sea that existed there in Mesozoic times. To the left of the man in the foreground, partially exposed by weathering, is a skull of the dinosaur *Triceratops*. The last articulated skeletal remains of dinosaurs are found at this level, high in the formation. A figure in the background stands in a vertical trench; it cuts down into a narrow, dark stratum. A second trench, to the left, extends up into the same stratum. At this level, 5.5 meters higher than that of the *Triceratops* skull, fossil spores and pollen provide evidence that plants had suddenly become scarce. Thus the extinctions of plants and reptiles were nearly but not quite simultaneous.

THIRD EXPOSURE of sediments at the boundary between the late Mesozoic and the early Cenozoic is near Aix-en-Provence in France. The boundary is near the level in the middle distance where the author is excavating a trench. These sediments are of continental origin.

dant in meteorites. The steady rain of micrometeorites on the surface of the earth (more than 70 percent of which fall into the oceans) results in modest concentrations of iridium and other siderophilic elements in the sediments that accumulate in the ocean basins.

In 1977 Walter Alvarez was working with an international group of scholars, including the paleontologist Isabella Premoli Silva of the University of Milan, who were examining the marine strata near Gubbio that include the layer of clay. Because the infall of micrometeoritic material is thought to be more or less constant, Luis Alvarez suggested that by measuring the amount of iridium in the clay it would be possible to calculate how much time had passed during the deposition of the layer. When Asaro and Michel did so, they discovered to their surprise that the iridium in the clay layer was 30 times more abundant than it was in clays from adjacent limestone strata.

If this excess of iridium had somehow been derived from terrestrial sources, the clay should have shown comparable enhancements in the other elements normally associated with the minerals that form clay. The Berkeley group's analysis disclosed a different pattern of enhancement, closer to that of the relative abundances of the elements found in meteorites. Could the surplus iridium have come from the oceanic reservoir of elements derived from micrometeorites, suddenly precipitated by some chemical event? Evidently not; neither above nor below the clay stratum was there any evidence that the normal rate of siderophile accumulation had fallen off as it should have if a precipitation had occurred. In this connection Charles J. Orth of the Los Alamos Scientific Laboratory and his collaborators have found a similar surplus of iridium at the top of Cretaceous sedimentary strata that are continental in origin; precipitation from an oceanic reservoir cannot, of course, have caused this.

By the time its report was published in 1980 the Berkeley group could add to the Gubbio datum the discovery of strata containing surplus iridium in late Cretaceous marine rocks in Denmark, Spain and New Zealand. (They were later able to add to the list iridium anomalies in deep-sea cores from both the Atlantic and the Pacific.) The report concluded with the suggestion that at the time of the mass extinction of certain marine microorganisms (and by inference at the time many reptiles disappeared) some 500 billion tons of extraterrestrial material had been abruptly deposited on the surface of the earth.

Where might the material have come from? The question is pertinent to further elaboration of the Berkeley group's hypothesis. If, on the one hand, the influx came from within the solar system, the mechanism of its arrival on the earth can be sought in the large body of data on this region of space. If, on the other hand, it came from outside the solar system, where there are comparatively few data on an enormous range of environments, the search for a mechanism would call for much speculation.

Consider the possibility that the extraterrestrial material was produced outside the solar system by a gigantic stellar explosion: a supernova. In support of such a hypothesis Malvin A. Ruderman of Columbia University and James W. Truran, Jr., of the University of Illinois at Urbana-Champaign have proposed that a gigantic burst of gamma rays from such an explosion could have blown micrometeoritic material off the surface of the moon and the earth could have swept it up. They subsequently noted, however, that gamma-ray flashes from supernovas have yet to be observed and furthermore that the transfer of iridium to the earth by such a mechanism would probably fall short of the required amount.

Another potential source of support for the supernova hypothesis is a continuing study being done by Paolo Maffei of the Astrophysical Observatory at Catania in Italy. Maffei is evaluating astronomical evidence for a gigantic explosive event some 1,000 light-years distant from the solar system at the end of the Mesozoic. The hypothesis is considered unlikely, however, by Wallace H. Tucker of the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory. In his opinion the interstellar material swept up by even a very large supernova would not accumulate in sufficiently dense concentrations to account for the amount of iridium in the zone of siderophile enhancement that caps late Mesozoic strata.

The sequence of events in a supernova explosion begins with an initial implosion. In the course of this collapse the nuclei of heavy elements at the core of the star rapidly capture neutrons. Among the new nuclear species formed is plutonium 244. The subsequent explosion distributes this radioactive isotope throughout an enormous volume of space. The chemists in the Berkeley group searched for Pu-244 in the iridium-rich clay, reasoning that if the iridium had been produced by a supernova, the plutonium should be present in detectable amounts. None was found.

They further reasoned that the two isotopes of iridium, Ir-191 and Ir-193, would be produced in different proportions by different supernova explosions because of variations in neutron fluxes and reaction times. When they analyzed the iridium-rich clay, they found that the two isotopes were present not in any exotic proportion but in the proportion that is typical of iridium in the solar system. The same is true of the proportion of the two isotopes of osmium, as has been discovered by workers in two separate laboratories: J. Hertogen of the University of Louvain in Belgium and Ramachandran Ganapathy of the J. T. Baker Chemical Company in Phillipsburg, N.J. Therefore it seems improb-
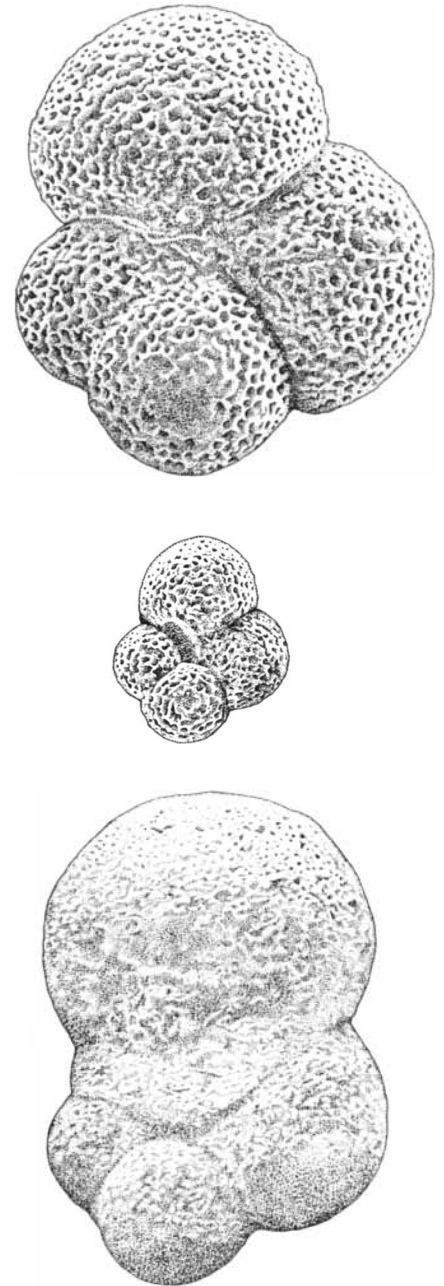
FORAMINIFERA of the late Mesozoic and the very early Cenozoic are seen in these drawings based on scanning electron micrographs enlarged to the same scale (200 diameters). Members of the Globigerinacea, they are, from the top, *Eoglobigerina*, an early Paleocene specimen from Denmark, an *Eoglobigerina* specimen of earliest Paleocene age from a North Pacific deep-sea core and an eroded *Rugoglobigerina* of late Cretaceous age from an oil-well core in Libya. The abnormally small size of the second specimen is taken to reflect a stage of gradual recovery from the environmental stresses that caused most late Mesozoic foraminifera to die out.

61

TURTLES

Archelon

BAENIDAE
DERMATEMYDIDAE
CHELYDRIDAE
EMYDIDAE
TRIONYCHIDAE
PELOMEDUSIDAE
TOXOCHELYIDAE
PROTOSTEGIDAE
CHELONIIDAE

LIZARDS

Clidastes

POLYGLYPHANODONTIDAE
XANTUSIIDAE
TEIIDAE
SCINCIDAE
ANGUIDAE
XENOSAURIDAE
VARANIDAE
PARASANIWIDAE
AMPHISBAENIDAE
MOSASAURIDAE

SNAKES

Dinilysia

ANILIIDAE
BOIDAE

DINOSAURS

(SAURISCHIA)

Tyrannosaurus

OVIRAPTORIDAE
DRYPTOSAURIDAE
TYRANNOSAURIDAE
TITANOSAURIDAE
DROMAEOSAURIDAE
ORNITHOMIMIDAE
SAURORNITHOIDIDAE

(ORNITHISCHIA)

Triceratops

HYPSILOPHODONTIDAE
ANKYLOSAURIDAE
THESCELOSAURIDAE
PACHYCEPHALOSAURIDAE
PROTOCERATOPSIDAE
HADROSAURIDAE
CERATOPSIDAE
NODOSAURIDAE

PTEROSAURS

Pteranodon

ORNITHOCHEIRIDAE
UNNAMED FAMILY

ICHTHYOSAURS

Ophthalmosaurus

PLATYPTERYGIIDAE
UNNAMED FAMILY

PLEISIOSAURS

Muraenosaurus

CIMOLIASAURIDAE
POLYCOTYLIDAE
ELASMOSAURIDAE

EOSUCHIANS

Champsosaurus

CHAMPSOSAURIDAE

CROCODILIANS

Goniopholis

GONIOPHOLIDAE
DRYOSAURIDAE
CROCODYLIDAE
ALLIGATORIDAE

able that the extraterrestrial material present in the clay stratum was the product of a supernova explosion (other than the one that may have been responsible for the formation of the solar system).

If the material came from within the solar system, how did it reach the surface of the earth? Two possibilities immediately suggest themselves: an encounter with a meteorite of asteroid size or an encounter with a comet. Concerning the first possibility the Berkeley group has estimated that the amount of exotic material in the iridium-rich stratum worldwide could have been contained in an asteroid some 10 kilometers in diameter. One problem with the asteroid hypothesis is that material of terrestrial origin in the iridium-rich stratum is not present in the large amounts that should have been gouged out of the earth's crust by the impact of a body of that size. Richard A. F. Grieve of the Canadian Department of Energy, Mines and Resources has suggested a way around this problem. The iridium could have been deposited after the impact in the form of a fallout of relatively pure meteoritic material reinjected into the stratosphere by the force of the collision. Both the Berkeley group and Jan Smit of the University of Amsterdam, a geologist who has studied late Mesozoic limestone strata in Spain, see still another way around the problem. If the asteroid fell into the sea, which is statistically likely, only a small amount of crustal material would have been excavated by the impact.

As for the possibility of an encounter with a comet, comets are low-density bodies composed largely of water ice. It is therefore estimated that a comet containing siderophiles in quantities large enough to account for the observed enrichment would have to be twice as massive as the hypothetical asteroid and therefore very much larger. This, of course, raises the excavation problem again. To counter it Frank Kyte of the University of California at Los Angeles and his colleagues Zhiming Zhou and John Wasson have proposed that as the hypothetical comet approached the earth it was disrupted by gravitational forces. The earth would then have been showered with cometary debris that would not have excavated any major crater or craters.

Either of these hypothetical events might cause extreme short-term stresses within the biosphere. For example, S. V. M. Clube and William M. Napier of the Royal Observatory in Edinburgh estimate that the shock wave generated by the impact of such a big asteroid on land would not only destroy all the earth's forests but also kill all the larger land-dwelling animals. And if the impact were at sea, it would generate tidal waves eight kilometers high.

Extreme stresses such as these, in the opinion of Walter Alvarez and the other members of his group, might not have been enough to have caused the late Mesozoic extinctions. They suggest that the impact would also have injected an enormous quantity of dust-size particles into the stratosphere. These, the Alvarez hypothesis proposes, would render the atmosphere much less transparent and strike at the very foundations of the biosphere by diminishing photosynthesis.

What light does the fossil record throw on this new hypothesis? A considerable amount, but it must be remembered that the fossil record contains a limited quantity of information and is not easy to interpret. For example, although paleontologists have been collecting the remains of Mesozoic animals for more than a century, the total number of known fragments of dinosaur skeletons is only about 5,000. This is largely owing to simple economics: the high cost of collecting dinosaur bones. As a comparison, much useful environmental information can be gleaned from the study of Mesozoic plant pollens. It costs about $500 (including a week's working time) to extract 20,000 pollen grains and mount them for microscopic examination. To collect and prepare the same number of dinosaur bones would cost about $400 million (including a million weeks' working time).

The point is relevant to an important factor in the study of biological extinctions: sample size. For example, one of the most recent Mesozoic dinosaur assemblages, the remains of animals that roamed the interior plains of the U.S. and Canada some 63 million years ago, is characterized by large horned herbivores, such as the genus *Triceratops,* and giant carnivores, such as the genus *Tyrannosaurus*. Strata that are 12 million years older, at Dinosaur Provincial Park in Alberta, have yielded a far greater variety of dinosaurs than the more recent rocks.

Does this mean, as some have suggested, that the dinosaurs were declining in variety as the end of the Mesozoic approached? Were they already heading for extinction millions of years before the hypothetical catastrophe? Not at all.

The Alberta fossils show a greater variety because many specimens have been collected there: more than 300, compared with fewer than 75 and even as few as six or seven at the more recent locales.

The fact is that the diversity of dinosaurs in Europe as the end of the Mesozoic drew near remained about the same; in Mongolia the diversity actually increased. In other areas of the world the samples are too small to reveal trends. In summary, evidence of a long-term decline in the diversity of dinosaurs before the time of extinction is simply not available.

Nowhere else in the world has the fossil record of land organisms over the final 1.75 million years of the Cretaceous been as fully sampled as it has in the 100-meter-thick exposures of grayish brown sediments around the southern edge of the Fort Peck Reservoir in northeastern Montana. The lower, earlier half of the formation is dominated by river-deposited sands, the upper, later half by alluvial silts and clays. The change in depositional patterns presumably reflects environmental changes that would have caused a change in the distribution of the animals, dinosaurs included, that inhabited this former coastal plain. In the lower levels the large predatory genus *Tyrannosaurus* and the duck-billed herbivorous genus *Amatosaurus* predominate. In the upper levels the herbivorous genus *Triceratops* and small browsing dinosaurs of the genus *Thescelosaurus* are more abundant. Similar changes in the plant community are documented by studies of fossil spores and pollen by Robert Tschudy of the U.S. Geological Survey.

An extraordinary pattern of change seems to occur at the top of the 100-meter formation. Here, along one horizon, the relative abundances lower in the sequence are reversed and the remains of large dinosaurs locally outnumber those of small ones. Large animals normally have lower birth and death rates than small animals. Hence an imbalance of this kind, with more large dinosaurs preserved as fossils than small ones, suggests some kind of mass death. (My interpretation is tentative, but further work at Fort Peck Reservoir will surely clarify the situation.)

Above this horizon near the top of the formation the sediments begin to take on a more laminated appearance. The only dinosaur bones are a few fragments that appear to have been eroded from older strata and deposited secondarily in stream beds. At first there is no comparable change in the plant community. Then, at a level about five meters above the horizon that holds the last unredeposited dinosaur skeletal material, the fossil pollen and spores become poorly preserved. David M. Jarzen of the National Museum of Natural Sciences in

REPTILE EXTINCTION late in the Cretaceous period is shown in the chart on the opposite page. Of the 48 families shown, 24 apparently died out at the end of the last stage of the Cretaceous, the Maestrichtian, some 63 million years ago. A colored extension of a black bar in the Cretaceous indicates a record of the family elsewhere than in North America north of the Rio Grande. A colored bar in the Paleocene column indicates the survival of the family outside North America. Except for the snake, all the skulls shown are those of representative extinct species. The dinosaurs were exterminated, and scarcely any reptilian order went unscathed.

63

EASTERN MONTANA | **WESTERN NORTH DAKOTA** | **SOUTHERN SPAIN**

(Vertical axis: MILLIONS OF YEARS BEFORE THE PRESENT — 61, 62, 63, 64, 65; AGE OF MAMMALS above, AGE OF REPTILES below)

EASTERN MONTANA:
- COASTAL PLAIN SEDIMENTS WITH LOW-GRADE COALS
- SECONDARY EXTINCTION EVENT(?)
- COASTAL PLAIN SEDIMENTS WITH TRICERATOPS ASSEMBLAGE
- BEACH SANDS
- SHALLOW MARINE SILTS

WESTERN NORTH DAKOTA:
- COASTAL PLAIN SEDIMENTS
- SHALLOW MARINE SILTS WITH GLOBIGERINA PSEUDOBULLOIDES ASSEMBLAGE
- LOW-GRADE COALS
- COASTAL PLAIN SEDIMENTS WITH TRICERATOPS ASSEMBLAGE
- BEACH SANDS
- SHALLOW MARINE SILTS

SOUTHERN SPAIN:
- OPEN MARINE LIMESTONE
- OPEN MARINE LIMESTONE WITH GLOBIGERINA PSEUDOBULLOIDES ASSEMBLAGE
- OPEN MARINE LIMESTONE WITHOUT FORAMINIFERA
- SECONDARY EXTINCTION EVENT(?) / CLAY LAYER
- OPEN MARINE LIMESTONE

PRIMARY EXTINCTION EVENT(?)

SEQUENCE OF EVENTS at two land sites in North America and a marine site in Spain at the close of the Mesozoic and the opening of the Cenozoic is presented in this table. The presence of the same replacement foraminiferan, *Globigerina pseudobulloides,* in North Dakota and Spain in the Cenozoic suggests a coincidence of marine and terrestrial extinctions.

Thereafter life began to proliferate once more. The deposition of sediments resumed, and the ocean floor was once again plowed by bottom-dwelling invertebrates. A new assemblage of foraminifera arose and was soon followed by another, characterized by the presence of, among others, the species *Globigerina pseudobulloides*. The resurgent protozoans inhabited the ancient Spanish sea for the next two million years.

Half a world away, in what is now North Dakota, a great interior sea spread westward at this same time, flooding a delta area where the remains of *Triceratops* had become fossilized and had been covered by coal-bearing strata. The marine siltstones that were laid down on top of the coal hold the shells of foraminifera species belonging to the same *G. pseudobulloides* assemblage that appeared in the Spanish sea after the great foraminiferan extinction. Given the uncertainties in estimating elapsed time from the thickness of sedimentary deposits, it seems possible that the story told by the sediments here, at Fort Peck Reservoir and in southern Spain is the same. If that is the case, the extinction of the dinosaurs on land and of the foraminifera at sea would have coincided.

The foraminifera were not the only marine organisms to die out at the end of the Mesozoic. As I have noted above, so did a number of marine reptiles. So did various mollusks: the coiled-shell cephalopods known as ammonites, the squidlike cephalopods known as belemnites and the peculiar coral-like bivalves known as rudists. Most of the major families of marine animals survived, but they lost many genera and species.

The fossil record at this crucial boundary is not as well understood with respect to larger marine animals as it is with respect to the microfauna such as the foraminifera. The reason is that the larger animals are numerous and diverse and the number of paleontologists is finite. As one example, even in such relatively well-studied formations as the chalks of Denmark the survival rate among such important animal groups as sponges, lampshells, marine snails and crustaceans remains uncalculated. As another example, the record of animal life in the tropical regions of the globe at this time is still poorly known. In view of the paucity of data it is no wonder that the issue of gradualism v. catastrophism is so vigorously debated.

A crude tabulation is helpful in suggesting the magnitude of the extinctions. Compare the number of animal genera in the fossil record some 10 million years before the end of the Cretaceous with the number of genera in the record in a comparable period after the crisis. It is obvious how insecure these numerical values are. Nevertheless, the numbers appear to reflect a 50 percent decline in

Ottawa has studied the plant evidence. He and I estimate that the five-meter interval represents a depositional period several tens of thousands of years long.

Above the level of poor plant preservation, beds of low-grade coal alternate with laminated siltstones. This separation of the highest occurrence of dinosaurs from the lowest occurrence of coal was first observed some years ago by William A. Clemens of the University of California at Berkeley. Tschudy, who has also studied the spores and pollen in the coal-bearing strata, reports that these evidences of plant life are only a third as abundant as they were in the dinosaur-bone strata below. The age of mammals had begun.

The limestones Smit and his colleagues have been studying in Spain record a remarkable series of events. They were deposited on the floor of an open tropical sea that intruded into southern Spain in late Mesozoic times. The formations are composed almost entirely of the calcium carbonate shells and platelets of tiny foraminifera: free-floating protozoan members of the sea's zooplankton. Here for more than 10 million years planktonic productivity remained high and there was no significant change in the character of the organic debris deposited on the sea floor.

Then within a layer of rock no more than five millimeters thick (representing less than 200 years of deposition) nearly 90 percent of the foraminifera species found lower in the formation simply vanish. Those protozoans that survived attained only a tenth the size of their predecessors. As the rain of shells and platelets nearly ceased so did the burrowing activity of bottom-dwelling invertebrate animals. A blanket of laminated red and green clays accumulated on the sea floor, reaching a thickness of about 10 centimeters. Conditions apparently remained stable for perhaps 20,000 years; by then all but one species of the surviving foraminifera had dwindled to extinction.

generic diversity worldwide. When one repeats this numbers game, counting the number of species recorded for certain plant and animal genera before and after the crisis, the result is similar. In a sample that includes mammals as representative land animals, chitinous marine algae as representative plants, and sand dollars, starfishes and oysters as representative marine animals, the decline in species during the extinction interval is from about three species per genus to 1.5. Therefore it seems reasonable to estimate that the biological crisis associated with the extinction of the dinosaurs also caused 75 percent of the previously existing plant and animal species to disappear. Indeed, this estimate is probably somewhat conservative.

The record of extinctions shows certain anomalies. For example, no land animal weighing more than about 25 kilograms (55 pounds) survived, and many of those that disappeared were considerably smaller. Again, the terrestrial plants of the northern regions of the Temperate Zone suffered more losses than those farther south. Yet the plants and animals of freshwater communities were scarcely affected. Much the same was probably true of deep-water marine mollusks in the opinion of Arthur H. Clarke of Ecosearch, Inc., of Mattapoisett, Mass. Shallow-water marine life, however, particularly the fauna of tropical reefs, was much more profoundly altered.

Even animals that shared the same environment were not identically affected. As Eric Buffetaut of the University of Paris has pointed out, crocodiles that occupied shallow marine waters survived the extinctions but the mosasaurs that occupied the same habitat did not. Whatever the agents of biological stress were, disturbances in food chains included, the ability of the biosphere to resist them was evidently varied.

What is the significance of the apparently dual nature of animal and plant extinctions at the end of the Mesozoic? Were the extinctions truly separate events, with the land animals dying out first and the plants second? If they were, was the second extinction the result of stresses as severe as those that caused the first, or was it simply a quasi-successional phenomenon of biology? Whatever the answer to these questions is, mankind may have been the long-term beneficiary of the evident catastrophe. As the Mesozoic drew to a close certain small carnivorous dinosaurs had achieved the ratio of brain weight to body weight that is characteristic of early mammals. If these presumably more intelligent reptiles had survived, their descendants might conceivably have continued to suppress the rise of the mammals, thereby preempting our own position as the brainiest animals on the planet.
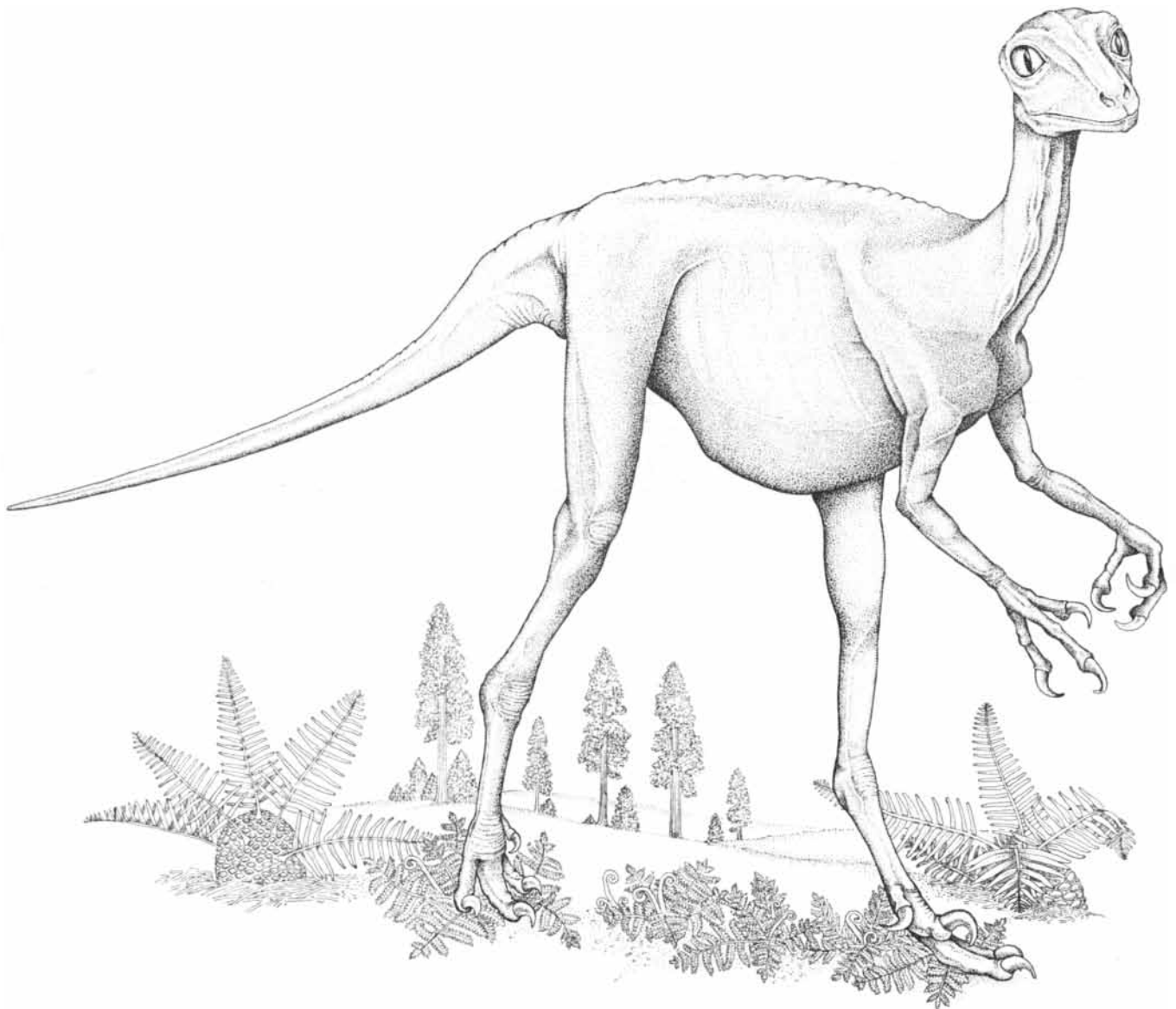
BIG-BRAINED DINOSAUR of the late Cretaceous, *Stenonychosaurus inequalus*, measured some three meters from the tip of its snout to the tip of its tail. This drawing is based on a restoration at the National Museum of Natural Sciences in Ottawa. Its ratio of brain weight to body weight equaled that of early mammals. Had such predators lived on they might have halted the rise of the mammals.

65

The Stabilization of Atomic Hydrogen

The gas combines explosively in familiar environments to form molecular hydrogen. A new technique that inhibits the reaction makes it possible to study certain properties of a quantum gas

by Isaac F. Silvera and Jook Walraven

In the domain of the very small the rule of the quantum theory is absolute. Nevertheless, quantum-mechanical effects are rarely conspicuous on a macroscopic scale, and a substance that exhibits them has properties quite unlike those of ordinary matter. Up to now one of the most striking quantum substances has been helium, which becomes a quantum liquid at low temperature. In the quantum-liquid phase helium flows without viscosity and conducts heat many times better than the best metallic heat conductors; moreover, no matter how low the temperature is, helium does not freeze at ambient pressure. Now another substance with macroscopic quantum properties has been prepared: it is a gas of atomic hydrogen. Indeed, the properties of atomic hydrogen at low temperature may ultimately prove to be stranger than those of helium. For example, atomic hydrogen resembles helium in that it does not solidify, but it also does not liquefy. Because the quantum theory predicts that it remains gaseous at the temperature of absolute zero it is called a quantum gas.

Hydrogen is of course a familiar element, but not in the atomic form; a gas made up of isolated atoms of hydrogen is not stable under the conditions that prevail on the surface of the earth. Instead hydrogen is found strongly bonded to other elements to form chemical compounds such as water, carbohydrates and other organic substances. Two hydrogen atoms can also combine to form a stable diatomic molecule (H_2). Although a gas of diatomic hydrogen does have quantum properties, it will not fail to form a liquid and then a solid if it is cooled sufficiently. It has long been possible to dissociate diatomic hydrogen molecules in the laboratory, but the resulting hydrogen atoms recombine in less than a thousandth of a second to form new hydrogen molecules.

In the past few years we and our colleagues at the University of Amsterdam have developed a way to suppress the recombination of atomic hydrogen even at relatively high densities. We have been able to maintain a gas of the atoms for several hours under conditions of extreme low temperature and high magnetic field. In such a state the hydrogen atoms recombine relatively slowly, and we regard the gas as being stabilized. Our experimental results show that atomic hydrogen remains a gas at the lowest temperature we have yet investigated, .08 degree Kelvin. This finding represents at least preliminary confirmation that atomic hydrogen remains a gas at absolute zero, as is predicted by the quantum theory.

Stable atomic hydrogen gas is a new substance in the laboratory, but it is by no means scarce in the universe as a whole. Interstellar space is filled with an extremely rarefied gas of atomic hydrogen; although the density may be as low as one atom per cubic meter (compared with the 3×10^{25} molecules per cubic meter in air at sea level), the interstellar gas amounts to a major share of all the matter in the universe. Interstellar hydrogen gas is stable because of its extremely low density. Recombination depends on collisions of the atoms, which are so rare that the lifetime of the gas is greater than the expected lifetime of the universe.

In the laboratory, however, one can produce samples of atomic hydrogen at temperatures much lower than those in space and at a density of about 10^{22} or 10^{23} atoms per cubic meter, many orders of magnitude greater than the density of the interstellar gas. According to theoretical calculations, such a density is only one or two orders of magnitude short of the density needed to precipitate exotic quantum phenomena that have been predicted for a quantum gas but have not yet been observed.

The most sought-after quantum phenomenon is a sudden condensation of a large proportion of the atoms in the gas into a state of minimum energy. The condensation is expected to take place at a low temperature that depends only on the density of the gas. For example, at a density of 10^{24} atoms per cubic meter the critical temperature for condensation is .016 degree K., whereas at the density of interstellar hydrogen the critical temperature is 10^{-18} degree K. The critical temperature for the condensation is proportional to the density raised to the 2/3 power.

The idea that a large number of particles in an atomic gas might all sink simultaneously to their lowest energy level at a nonzero temperature was introduced by Albert Einstein in 1923. In the classical, or pre-quantum-mechanical, thermodynamics of the 19th century it is only at absolute zero that a sizable proportion of the particles of a gas (in fact all of them) can attain their lowest energy levels. Above absolute zero the particles of such a classical gas will assume a spectrum of many energy levels, no one of which includes a large fraction of the particles. The distribution of the energy states for a given temperature is described by Maxwell-Boltzmann statistics, after James Clerk Maxwell and Ludwig Boltzmann. The statistics give the probability that an atom will have a certain energy.

In the correct, quantum-mechanical description, however, the statistical distribution of the energy states for an aggregate of identical particles can be quite different. The statistical theory that describes atoms of hydrogen was first studied by the Indian physicist S. N. Bose and is called Bose statistics. The phenomenon predicted by Einstein is a mathematical consequence of Bose statistics, but it was so contrary to the intuition of physicists in the 1920's that it was then regarded as a mathematical oddity that would never be found in a real system. It is now thought, however, that the phenomenon is observable in the labo-

ratory. It is called Bose-Einstein condensation.

A gas of atomic hydrogen that had become Bose-Einstein condensed would be composed of two populations of atoms: the condensate atoms in the ground state, or state of lowest energy, and the atoms that remain in higher energy states. Such a distribution of energy states may not seem unusual, since in any large aggregate of atoms it is statistically likely that a few of them occupy the ground state. Even in Bose statistics the atoms are distributed among numerous energy states at ordinary temperatures, and no one state has more than a small fraction of the total. In a Bose-Einstein-condensed gas, however, a large fraction of the atoms would occupy the ground state at an experimentally accessible temperature, and nearly 100 percent of the atoms would become condensate atoms at temperatures well above absolute zero.

In classical thermodynamics all atomic motion was thought to stop at zero temperature; the atoms would come to rest with zero kinetic energy. In quantum mechanics the motionless state is not possible because it would violate the uncertainty principle formulated concisely by Werner Heisenberg. The principle states that the position and the momentum of an atom cannot both be determined simultaneously with unlimited precision. If an atom were to stop moving entirely, its position could be measured exactly. The momentum of the motionless atom would also be known: since momentum is the product of mass and velocity, the momentum of a stationary atom must be zero.

The uncertainty principle implies that even if the temperature of a substance could be reduced to absolute zero, the atoms would retain some small quantity of kinetic energy (the zero-point energy) and would therefore remain in motion (the zero-point motion). Nevertheless, the motion would be qualitatively different from the random motion characteristic of higher temperatures: all the atoms would have the same energy and therefore the same state of motion. It is the coherent motion of the condensate atoms of a Bose-Einstein-condensed gas that is expected to give rise to extraordinary macroscopic properties at temperatures well above absolute zero.

The atoms that make up about 50 percent of the elements are subject to Bose statistics. Hence it is worthwhile pausing to ask why Bose-Einstein condensation is not a common phenomenon, observed in many substances at low temperature. The reason is that the condensation takes its simplest form (and the form in which Einstein described it) only in an ideal gas, one made up of atoms that do not interact with one another. All real atoms have electromagnetic inter-

MAGNET LEADS

PUMP AND EVAPORATION LINES

LIQUID NITROGEN (77° K.)

LIQUID HELIUM (4.2° K.)

REFRIGERATOR (.005° TO .6° K.)

SUPERCONDUCTING MAGNET

HYDROGEN STABILIZATION CELL

BOLOMETER

FILM OF SUPERFLUID HELIUM

HELIUM VAPOR COMPRESSOR

ELECTRIC DISCHARGE TUBE

MOLECULAR HYDROGEN GAS

HOT ATOMIC HYDROGEN GAS

4.2° K.

77° K

300° K.

ACCOMMODATOR

TEFLON VALVE
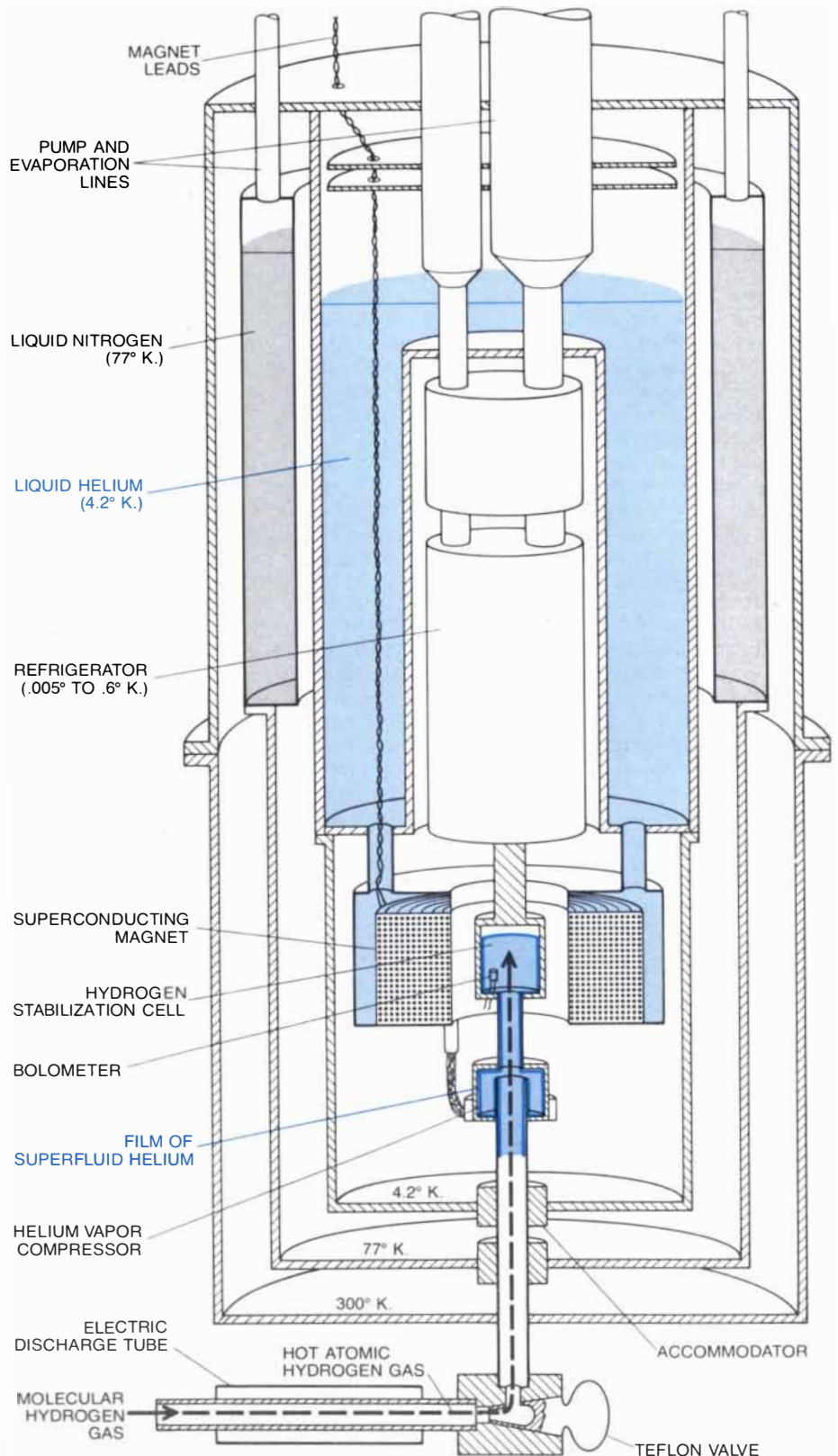
EXPERIMENTAL APPARATUS developed by the authors at the University of Amsterdam breaks up hydrogen molecules to yield a gas of free hydrogen atoms, which is then stabilized at a low temperature in a strong magnetic field. The chemical bond that holds together the two atoms of a hydrogen molecule can readily be broken by bombarding the molecule with energetic electrons, but without refrigeration and a strong magnetic field the atoms would recombine in less than a thousandth of a second. The apparatus incorporates several nested refrigeration units maintained at successively lower temperatures. Hot atomic hydrogen gas is emitted by an electric discharge tube and piped past a series of cold surfaces. The chilled hydrogen then enters the magnetic field, where it can be held for several hours. At a temperature of .3 degree Kelvin densities as high as 10^{23} atoms per cubic meter have been attained for a few minutes.

67

actions that tend to bind them together; indeed, most substances freeze into a solid matrix as a result of such interactions long before the temperature of possible Bose-Einstein condensation is reached.

If the condensation is to be observed at all, the zero-point motions must be vigorous enough to overcome the electromagnetic attraction of the atoms. The magnitude of the zero-point motion increases as the atomic mass or the strength of electromagnetic interactions is reduced. Therefore the chemically inert atoms of the lightest elements are the only plausible candidates. Helium has interesting quantum properties because it is both light and inert, so that the interatomic forces are feeble. It is for these reasons that helium does not solidify.

Hydrogen is of course lighter still and has the most vigorous zero-point motions of any atom; under ordinary circumstances, however, hydrogen atoms also exhibit strong interatomic forces and form diatomic molecules. By subjecting the atoms to extraordinary conditions it is possible to make the interatomic forces weaker than those governing any other atomic system. Under such conditions the kinetic energy of the atoms even in their lowest energy state is sufficient to prevent hydrogen from either solidifying or liquefying. The resulting substance may represent the closest approach possible to the ideal quantum gas imagined by Einstein, and it may offer the best opportunity for observing Bose-Einstein condensation.

The best clues to what can be expected of stable atomic hydrogen come from 70 years of investigating the properties of helium at low temperature, particularly the properties of the common isotope helium 4. (The nucleus of a helium-4 atom is made up of four particles: two protons and two neutrons. Other helium isotopes are differentiated by the number of neutrons in their atomic nuclei.) Helium 4 becomes an ordinary liquid at 4.2 degrees K. If the vapor pressure above the helium is reduced with a vacuum pump, the liquid boils violently and is thereby cooled further. The formation of bubbles in a boiling liquid is caused by relatively poor thermal conductivity.

When the temperature of helium is reduced to 2.18 degrees, however, the liquid suddenly becomes such a good conductor of heat that all boiling stops. Evaporation and cooling continue at the surface, but convection currents alone redistribute the remaining thermal energy throughout the volume. Such currents can provide adequate heat transport only in a fluid that moves without viscosity, or resistance to flow.

Liquid helium 4 at or below 2.18 degrees is therefore called a superfluid. If it is set flowing in a tube closed on itself, the liquid continues to flow without friction, never coming to a stop as a normal fluid would. It flows into the smallest passages of its containing vessel, and it has the remarkable ability to flow through a densely packed powder as if

the barrier were not present. A vessel with microscopic holes that would be impenetrable to a normal fluid can be as leaky as a sieve to a superfluid. Such a vessel is said to have a superleak.

Even without a superleak an open vessel filled with a superfluid will not stay filled for long. A thin film of the superfluid about 200 angstrom units thick will begin to creep up the inside walls of the vessel, over the edge and down the outside surface. The creeping is another consequence of the lack of viscosity. All fluids that can wet a surface tend to creep along it because of the adhesive forces between the atoms of the fluid and the atoms of the surface. In principle any liquid should creep up the walls of its container and minimize its total energy by siphoning to a lower level. In ordinary fluids, however, the flow of the film is inhibited by viscosity. Hence a fluid without viscosity such as superfluid helium forms a continuous link between the inside and the lower external walls of a container, and so the container is emptied.

Other puzzling properties of a superfluid cannot be explained merely by the lack of viscosity. One such property is a tendency to move toward a source of heat; in a thin film the flow can be as fast as 50 centimeters per second. If heat is applied to the upper end of a thin tube whose lower end is submerged in superfluid helium 4, the helium film flows up the tube and evaporates at the heater. If the lower end of the tube is packed with fine powder (which a normal fluid could not penetrate even under pressure), the superfluid flows up the tube so rapidly that it sprays out of the top of the tube like a fountain and showers back into the vessel.

It is highly plausible but not yet definitively established by experiment that superfluid helium 4 is Bose-Einstein condensed. To confirm the hypothesis it would be necessary to measure the fraction of the atoms in the ground state, and so far it has not been possible to do that unambiguously. Fritz London of Duke University first proposed the hypothesis in 1938 when he showed that some of the observed properties of superfluid helium could be accounted for by assuming it is Bose-Einstein condensed. Helium-4 atoms are subject to Bose statistics. Hence London proposed that a superfluid is made up of two components: a condensate, or superfluid component, that can pass through small constrictions and a normal component that cannot.

For example, in the fountain effect only the superfluid condensate can penetrate the powder and reach the upper part of the tube. There heating excites the atoms to higher energy states and so converts the superfluid component into a normal fluid. In order to maintain a
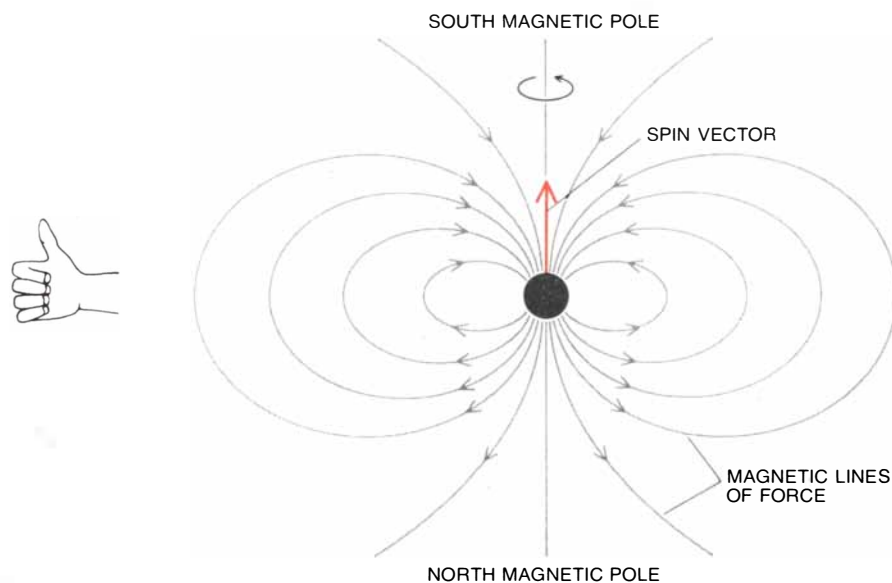
SPIN ANGULAR MOMENTUM is the property of the particles composing a hydrogen atom that can be manipulated to keep the atoms from recombining. The spin is a vector quantity, that is, one having both a magnitude and a direction. The magnitude of the vector can be interpreted as the angular momentum of the spin, whereas the direction of the vector gives the sense of the rotation according to the right-hand rule. (If the fingers of the right hand curl in the same sense as the spin, the right thumb points in the direction of the vector.) A particle with both mass and spin generates a magnetic field, and so the orientation of the particle can be influenced by an externally applied magnetic field. For a negatively charged particle such as the electron the magnetic north pole points in the direction opposite to that of the spin vector.

68

uniform distribution of the superfluid component throughout the volume of the fluid more of the component is drawn up the tube.

In spite of these explanatory successes there are serious problems with London's model that have been the subject of theoretical analyses for decades. Whereas Einstein's prediction was based on a gas of particles that do not interact with one another, the effects of the interatomic forces in a liquid are important and cannot be neglected. The ground-state atoms in such a gas have zero momentum, and so the uncertainty about their position is maximized. Interactions modify the ground state of the atoms in a liquid, however, so that few of the ground-state atoms are in a state of zero momentum. Thus if a liquid is to undergo Bose-Einstein condensation, the condensate atoms must occupy a more complex ground state than the ground state of the atoms of a gas. Moreover, the complex liquid ground state is hard to identify experimentally.

In a superfluid gas of atomic hydrogen such problems can largely be avoided. Whereas a liquid has an essentially fixed density, the density of a gas can be varied by controlling the pressure in the containing vessel. In a low-density gas the effects of atomic interactions are much less important than they are in a liquid. Theoretical calculations suggest that almost 100 percent of the condensate atoms in a gas will be in a state of zero momentum. Moreover, magnetic properties and other properties of the hydrogen atom that it does not share with the helium atom promise to enrich the variety of macroscopic phenomena that can be observed in a superfluid.

Atomic hydrogen can easily be made, but to stabilize it several major experimental difficulties had to be overcome in an area where little prior knowledge existed. Although the goal of our experiments is to examine certain predictions of the quantum theory, the design of our experimental apparatus would not have been possible without guidance from other predictions of the theory.

One of the most important quantum properties of an atom or a subatomic particle is its spin, or intrinsic angular momentum. The spin of a particle is similar to the spin of a ball. It is a vector quantity, that is, one having both a magnitude and a direction. The magnitude of the spin vector can be interpreted as a measure of the angular velocity of rotation; the direction of the vector specifies the sense of the rotation according to the convention called the right-hand rule. If the fingers of the right hand curl in the same sense as the spin, the right thumb points in the direction of the spin vector.

Unlike the spin of a ball, the spin of an atom or a subatomic particle is quan-

STABILIZATION OF ATOMIC HYDROGEN requires that the spin vectors of the electrons in the atoms be polarized, or oriented in the same direction. Two atoms whose electron-spin vectors point in opposite directions tend to share electrons and bind strongly to form a diatomic molecule. In the absence of strong polarizing forces the electron-spin vectors are randomly oriented and virtually all hydrogen atoms soon form pairs. If the electron-spin vectors are parallel, however, the electrons cannot be shared, the interatomic forces become weak and the hydrogen atoms remain in the monatomic state. Only the relative orientations of the two electron-spin vectors (whether they are parallel or antiparallel) need be considered to understand the formation of diatomic hydrogen; the actual directions of the vectors and the spins of the protons are of secondary importance. Thus only two of the 16 possible pairs of atoms are shown.

tized. The magnitude of the spin can assume only integer or half-integer values (when the angular momentum is measured in fundamental units) and the vector can have only a finite number of directions. The proton, the neutron and the electron, for example, have a spin magnitude of $1/2$, and the spin vector can point in one of two directions, either "up" or "down." A particle whose spin has half-integer values, such as $1/2$, $3/2$, $5/2$ and so forth, is called a fermion, after Enrico Fermi.

A particle whose spin magnitude is an integer in the series $0, 1, 2, 3 \ldots$ is called a boson, after Bose. The photon, or quantum of electromagnetic radiation, is a boson, and so is the pion, which is the major carrier of the binding force in the atomic nucleus. A boson with a spin of 0 is of course a particle with no spin at all, so that a spin vector is not required to describe its properties. For a boson with a spin of 1 the vector can have three possible directions.

The spin of a composite entity such as

a hydrogen atom depends on the intensity of the magnetic field in which the atom is embedded. In the case that is simplest to describe the magnetic field is infinite, and for a strong magnetic field an infinite field is a good approximation. In an infinite field the spin of the hydrogen atom is the vector sum of the spin vectors of its constituent particles. In general the sum depends on the orientation of the constituent spins. A hydrogen atom, being composed of one proton and one electron, can exist in any one of four spin states. Both the proton- and the electron-spin vectors can point up, so that the spin of the atom is 1 in the up direction, or both constituent spin vectors can point down, so that the atomic spin is 1 in the down direction. In the remaining configurations the constituent spin vectors point in opposite directions, proton up and electron down or proton down and electron up, so that the total spin of the atom is 0. In all possible orientations it is apparent that the hydrogen atom is a boson.

The distinction between fermions and

69

bosons provided the key to the technique that has made it possible to stabilize atomic hydrogen. The distinguishing property of fermions is that they obey the quantum-mechanical rule called the exclusion principle, formulated by Wolfgang Pauli. If one fermion occupies a certain state, all other identical fermions are excluded from that state. In quantum mechanics the state of a particle is a complete specification of its identity and circumstances, including its energy, its position or momentum and its spin. For example, if two electrons occupy overlapping regions of space and have identical energy and momentum, they must differ in spin. Bosons, on the other hand, do not obey the exclusion principle; there is no limit to the number of bosons that can occupy a single state. Indeed, in Bose-Einstein condensation all the bosons in a macroscopic sample of matter can in principle converge on the same state, namely the ground state.

Although an isolated hydrogen atom can be treated as a composite boson, two atoms that approach each other closely cannot. In the formation of the diatomic hydrogen molecule, for example, the properties of the two electrons and the two protons that make up the molecule must be taken into account. As the atoms are brought together their peripheral electron clouds begin to overlap. Two electrons whose spin vectors point in the same direction cannot occupy the same region of space simultaneously. If the electron-spin vectors point in opposite directions, however, no quantum principle keeps them from being in the same place at the same time. (Electrostatic repulsion tends to keep two electrons apart, but it is a secondary consideration here.)

Two nearby hydrogen atoms tend to assume the configuration of lowest possible energy. In minimizing the energy the constituent electrons and protons of the atoms tend to arrange themselves so that the attractive and the repulsive electrostatic forces acting between them are exactly balanced. For a system of two hydrogen atoms the lowest-energy configuration depends almost entirely on the states of the electrons. For example, if the spins of the two electrons are antiparallel, the negative charges of the electrons can be interposed between the two positively charged nuclei and will partially overcome the nuclear repulsion. With this configuration of the electrons the total energy of the system is much lower when the atoms are close together than it is when they are far apart. The electrons constitute a strong cohesive bond. Except at high temperature the atoms have too little thermal kinetic energy to rupture the bond; the atoms have formed a stable, diatomic molecule of hydrogen.

If the electron spins are parallel, on the other hand, the electronic charge cannot build up between the nuclei without violating the exclusion principle. The resulting forces between the two atoms are repulsive, except that at a nuclear separation of about four angstrom units there is a small attractive force. The attractive force is so weak, however, that the zero-point energy of the atoms is sufficient to overcome the force and prevent the formation of a molecule. The fundamental reason a gas of atomic hydrogen is not stable under the conditions prevailing on the surface of the earth is that the electron spins have an almost random distribution. It is virtually always possible for a hydrogen atom to pair with another atom whose electron-spin vector is pointed in the opposite direction.

Our basic strategy for creating stable atomic hydrogen is to polarize the electron-spin vectors so that they all become oriented in the same direction. If the polarization is complete, the atoms cannot recombine. The substance that results is called spin-polarized atomic hydrogen.

How can one obtain spin-polarized hydrogen from ordinary atomic hydrogen? A particle with both mass and spin has a magnetic-moment vector associated with its spin. The spin state of an electron can therefore be maintained by an external magnetic field just as the direction of a magnetic compass needle can be maintained by the magnetic field of the earth. The degree to which a collection of atomic electrons is polarized can be represented by the ratio of the number of atoms that have spin-up electrons (spin-up atoms) to the number of atoms that have spin-down electrons (spin-down atoms). The ultimate degree of polarization attained depends strongly on the magnetic field intensity and the temperature. In our experiments the spin ratio is about 1 to 10^{20}. It is achieved with a magnetic field of some 100,000 gauss (200,000 times stronger than the earth's magnetic field) and a temperature of .3 degree K.

Unlike a compass needle, whose orientation can be changed as well as maintained by a magnetic field, the spin state of an electron cannot be changed by a uniform magnetic field, no matter how strong. To polarize the electrons we instead exploited the effect of a strong magnetic field gradient: a field that varies rapidly from place to place. The high-gradient field is no more effec-
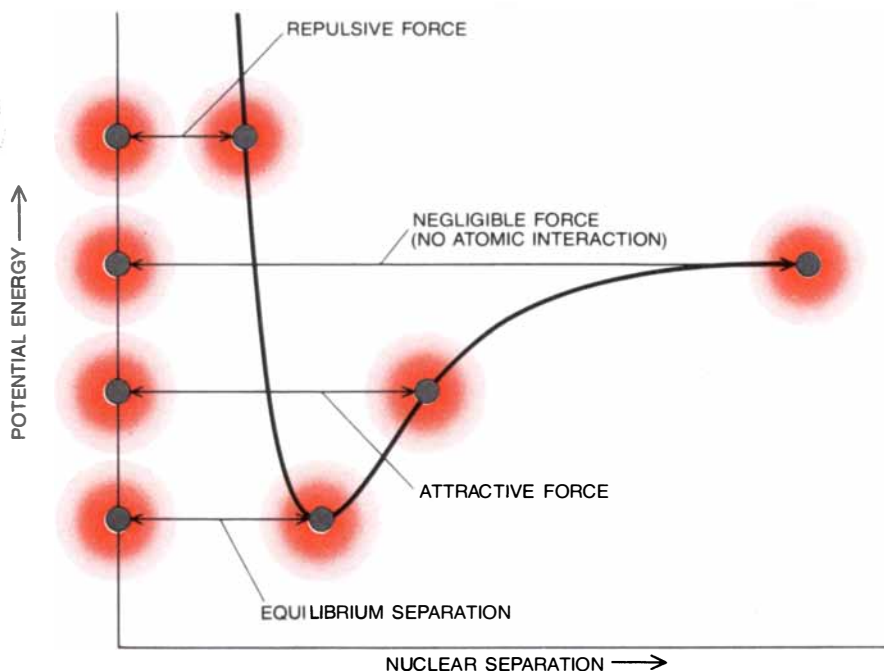
POTENTIAL ENERGY of two hydrogen atoms depends on the distance between them. At a very large separation the potential energy is arbitrarily set equal to zero. At every point on the curve except the minimum there is a net force on the atoms that tends to drive them to a lower energy. The point of minimum potential energy represents the equilibrium separation or rest position of the second atom with respect to the first. If the second atom is moved closer to the first, the potential energy of the atoms increases and there is a net repulsive force between them that tends to restore them to the equilibrium position. Similarly, if the second atom is moved away from the first, the potential energy again increases and an attractive force acts to restore the atoms to the equilibrium position. If the atoms are moved far enough apart, a small change in their separation does not much alter the potential energy of the system: even a small quantity of kinetic energy enables them to move about freely without significant interaction. In this hypothetical potential-energy function the constraints imposed by quantum mechanics on the energy of the atoms and on their interactions have not been taken into account.

REPULSIVE FORCE

NEGLIGIBLE FORCE (NO ATOMIC INTERACTION)

POTENTIAL ENERGY

ATTRACTIVE FORCE

EQUILIBRIUM SEPARATION

NUCLEAR SEPARATION ⟶

tive than the uniform one in flipping the spins, but the gradient can easily sort spin-up atoms from spin-down ones.

Spin-down atoms are forced into regions that have relatively strong magnetic fields, where they are maintained in their polarized state for several hours. (In our apparatus the strongest magnetic fields coincide with the regions of lowest temperature.) The spin-up atoms are repelled to regions that have weaker magnetic fields, and they collide with the confining walls of the apparatus. There they either flip to the spin-down state and proceed to the strong-field regions or recombine to form diatomic hydrogen. The diatomic hydrogen sticks permanently to the cold walls and is of no further consequence.

The spin-down atoms enter a cold stabilization cell at the end of a tube, where the magnetic field gradient and the low temperature serve to confine them. As the polarized atoms travel up the magnetic field gradient they are accelerated. If the atoms did not lose kinetic energy through collisions with the walls of the stabilization cell, the net increase in their kinetic energy would be sufficient to enable them to travel against the magnetic field gradient and so escape from the cell. When the atoms encounter the cold walls of the cell, however, they exchange energy with the walls and become trapped by a magnetic barrier.

Once we understood how to sort the hydrogen atoms and trap them in a spin-polarized state, most of the efforts of the experimental design had to be devoted to preventing the atoms from sticking to the walls of the apparatus and recombining there. In spin-polarized hydrogen a collision between two atoms can sometimes lead to the reversal of the spin of one of the atoms and subsequent molecular recombination. The collision must be mediated by a third particle so that the total energy and momentum of the system can be conserved. When two atoms collide in isolation, without a third particle to carry off the energy liberated in the collision, the spin reversal is forbidden by the conservation laws. Hence in a gas at least three particles must collide simultaneously for recombination to take place, and at low density the probability of such an event is small.

On a surface, however, any collision between two atoms can lead to recombination because the surface atoms can play the catalytic role of the third body. The problem is complicated considerably by the requirement that the walls of the apparatus be maintained at low temperatures in order to cool the hydrogen gas. Unless the walls are coated with special materials, the electromagnetic forces between the hydrogen atoms and the walls become strong enough at low temperatures to bind the atoms to the
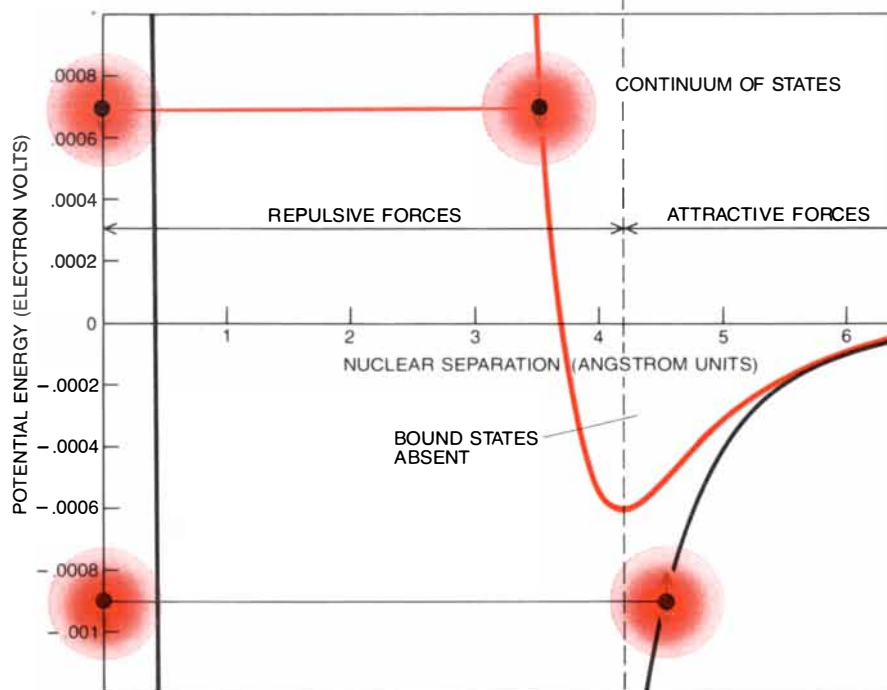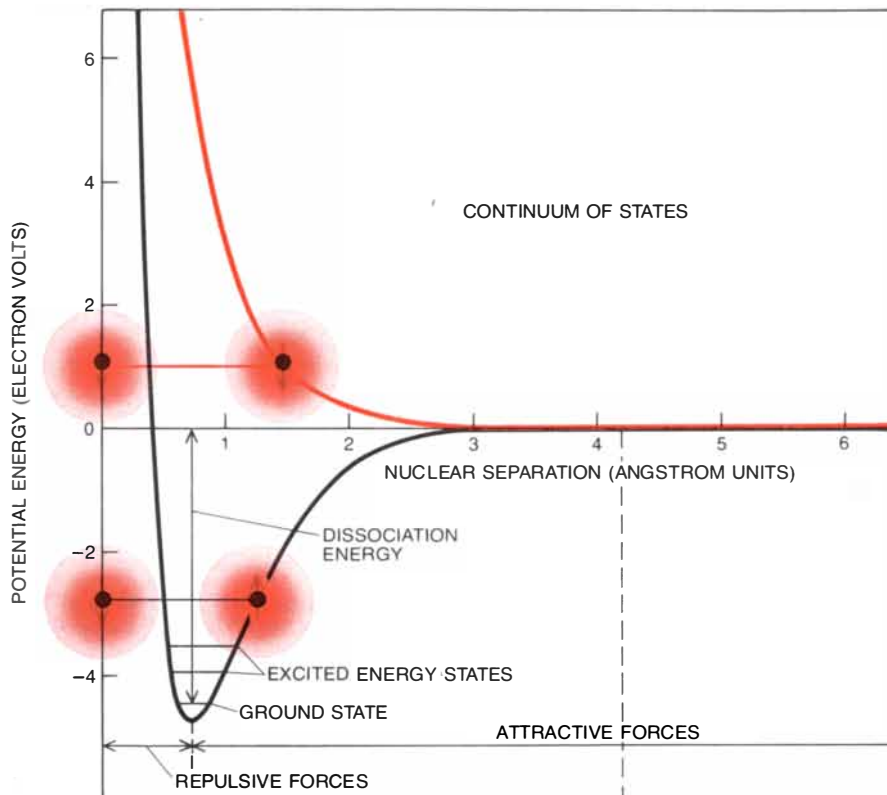
QUANTUM-MECHANICAL POTENTIAL-ENERGY FUNCTION for a pair of hydrogen atoms with antiparallel electron spins (*black*) is quite different from the function for a pair of parallel-spin atoms (*color*). The atoms tend to adopt a configuration that minimizes their potential energy, but they cannot lie at rest at the equilibrium separation because the position of a quantum-mechanical particle cannot be precisely fixed. Instead, if the atoms are bound, they must occupy one of a series of discrete energy states; the lowest state, or ground state, keeps them closest to the equilibrium separation. If the atoms are not bound, they can assume any positive energy state along a continuum of values. When the two electron spins are antiparallel, the ground state lies near the bottom of a deep potential well and the energy required to dissociate the atoms is large. Hence the atoms minimize their total energy by forming a molecular bond. The bond is strong enough to withstand the thermal motions of the atoms up to high temperatures. When the spins are parallel, however, the attractive forces are so weak that even the feeble atomic motions quantum mechanics predicts would persist at absolute zero would keep the atoms from forming a bond. Hence no ground state exists. In the lower diagram the vertical scale is enlarged by a factor of about 5,000 with respect to the upper diagram.

71
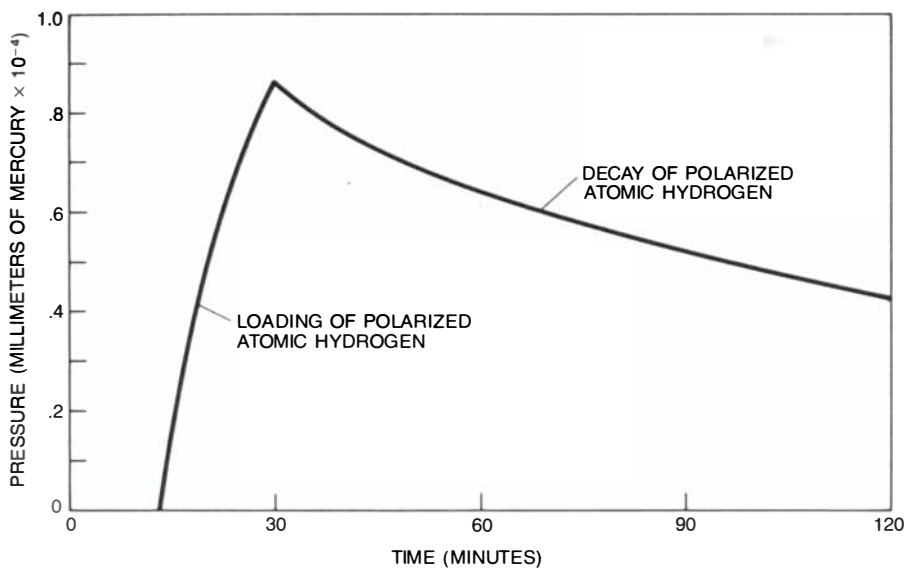
PRESSURE BUILDUP of the stabilized, spin-down hydrogen gas during the loading of the stabilization cell is followed by a gradual decay after the loading stops. The decay is caused by the slow recombination of the hydrogen atoms to form diatomic molecules. According to the quantum theory, in a magnetic field of finite intensity a fraction of the polarized hydrogen atoms are in a "mixed" state where there is a small probability that the spin of a given electron is up as well as down. The presence of the mixed state opens a channel to recombination.

walls. As the density of the atoms adhering to the walls increases, the recombination rate increases; it varies as the square of the surface density. Surface recombination can be so efficient that even spin-polarized atomic hydrogen gas would totally recombine in less than a thousandth of a second on ordinary surfaces.

To prevent surface recombination a surface had to be found that has weak attractive forces for atomic hydrogen. The best possible surface for the purpose is spin-polarized atomic hydrogen itself, but if one could maintain such a surface, surface adhesion would not be a problem in the first place. We found that a surface that has a weak affinity for hydrogen and can prevent its buildup is the surface of liquid helium. The helium
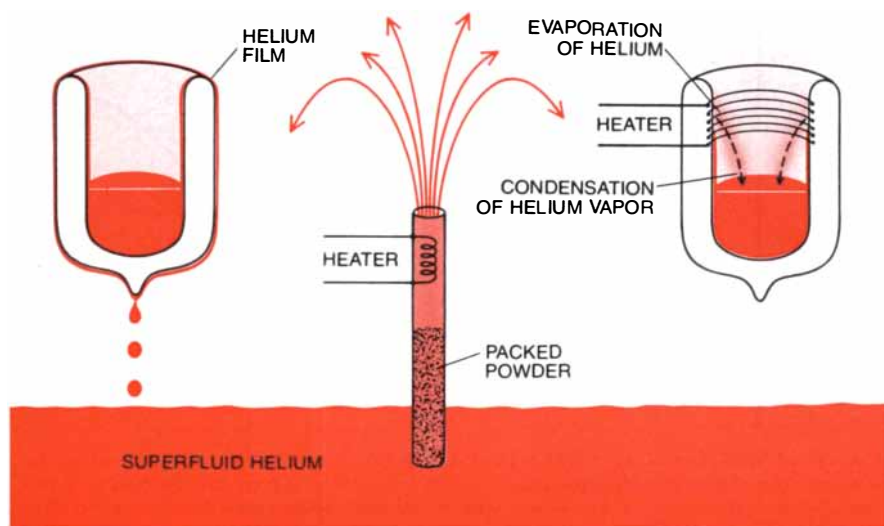
SUPERFLUID HELIUM takes part in several strange phenomena characteristic of a quantum liquid; some of the phenomena may have analogues in atomic hydrogen gas, and some of them were employed in the preparation of the gas. If the superfluid is put in an open container, a film will creep up the sides, over the edge and down the outside. It will pass through a densely packed powder that would be impermeable to a normal fluid, and it will move through the narrow pores of the powder at 50 centimeters per second toward a source of heat. If the upper end of a vertical tube packed with powder is heated, the superfluid helium will rise so rapidly in the tube that it will form a fountain. Heat will also cause a film of helium to creep up the sides of a tube, where it will evaporate, return to the colder region just above the liquid and condense. The circulation of helium vapor was employed by the authors to help compress atomic hydrogen.

surface is a thin film that covers all the surfaces of the stabilization cell, and at the low temperatures necessary for the experiment it is a superfluid. Special cooling stages are necessary to employ a superfluid film and still maintain the low temperature of the cell.

Atomic hydrogen is first produced from molecular hydrogen gas in an electric discharge tube. In the tube free electrons are accelerated by an electric field and collide with the diatomic hydrogen molecules, causing dissociation. Because of the energetic bombardment of the electrons the atomic gas emerges from the tube at a temperature of several hundred degrees K.

The hot gas moves rapidly toward the cold interior of the stabilization cell through a filling tube lined on the inside with the fluorocarbon polymer Teflon. Curiously, the same material that forms a nonstick coating for frying pans also serves to keep atomic hydrogen from sticking to the wall of the tube in the first stages of cooling. In moving through the tube the hydrogen atoms collide repeatedly with the refrigerated Teflon wall, each time giving up some of their kinetic energy; in this way the gas is chilled. Several bounces per atom are usually sufficient for the gas to reach thermal equilibrium with the wall. As long as the temperature of the Teflon wall is above 20 to 30 degrees K. and the passage of the gas is brief there is little recombination of the hydrogen atoms.

The transition from 20 degrees to one degree K. is a critical one, because the Teflon lining is no longer adequate to prevent recombination. Therefore in the next stage of cooling another kind of surface is needed, one with a lower adsorption potential, or propensity to capture hydrogen atoms. Helium, which covers the colder walls of the stabilization cell, is not suitable in the transition stage. At such temperatures it will not form a surface film unless the helium vapor pressure is rather high. The high vapor pressure is unacceptable because the helium atoms will catalyze the rapid recombination of the hydrogen gas.

It turns out that the best surface lining after helium is molecular hydrogen, and it is an acceptable substitute for helium in the transition region. It was shown by Stuart J. B. Crampton and his co-workers at Williams College that hydrogen atoms do not strongly adsorb onto solid surfaces of molecular hydrogen at temperatures above about four degrees K. In our apparatus a solid layer of molecular hydrogen builds up automatically as the atomic hydrogen recombines onto the surface of a metal tube called the accommodator, which is maintained at 4.2 degrees K. After the initial introduction of atomic hydrogen the buildup slows down as the molecular hydrogen

surface inhibits further recombination.

The buildup is greatest in the regions maintained between 10 degrees and one degree K. and depends on the amount of time the atomic hydrogen is allowed to remain in the region. Hence our strategy was to adjust the length of this region so that the atoms undergo just enough wall collisions to cool but not enough to recombine. The atomic hydrogen that does not recombine in this region finally reaches the safety of the region whose walls are covered with helium.

It is in this last stage in the processing of the chilled gas that the atoms are sorted by a magnetic field gradient according to their electron-spin direction. The spin-down atoms enter the hydrogen-stabilization cell, where the final density of the atomic hydrogen is measured. The stabilization cell is maintained at temperatures between .5 and .1 degree K. and is permeated by the high-gradient magnetic field that traps the atomic gas.

Before the hydrogen is injected helium gas is introduced into the cooled stabilization cell, where it condenses onto the surface as superfluid helium 4. The remaining helium vapor pressure is negligible at such low temperatures, so that the cell is effectively evacuated. Once a surface coating of helium 4 has been established, a superfluid effect is exploited. The superfluid helium flows along the surface and out of the cell to the warmer region of the accommodator. There it is vaporized, and the helium vapor tends to return to the cell because of the lower pressure in the colder regions. As the vapor is chilled it again condenses onto the surface and becomes a superfluid.

The condensation of the helium vapor creates a serious problem of refrigera-

CORE of experimental apparatus conducts cool, unpolarized atomic hydrogen toward a stabilization cell permeated by a strong magnetic field. When the hydrogen atoms approach the field, the change in magnetic intensity repels the spin-up atoms (*arrows pointing up*) and attracts the spin-down atoms (*arrows pointing down*). One of the major goals of the design is to keep the hydrogen atoms from sticking to the walls of the apparatus, where they are more likely to recombine. In the lower and warmer part of the apparatus the atoms are cooled to about four degrees K. as they bounce off the walls covered with molecular hydrogen (*red*). In the colder and upper part of the apparatus the surfaces are coated with a thin film of superfluid helium (*blue*), which has a much lower propensity to adsorb atomic hydrogen than solid molecular hydrogen does. The density of the atomic hydrogen gas can be determined accurately by a capacitance pressure detector or by a bolometer, an instrument that registers changes in temperature as changes in electrical resistance.

73
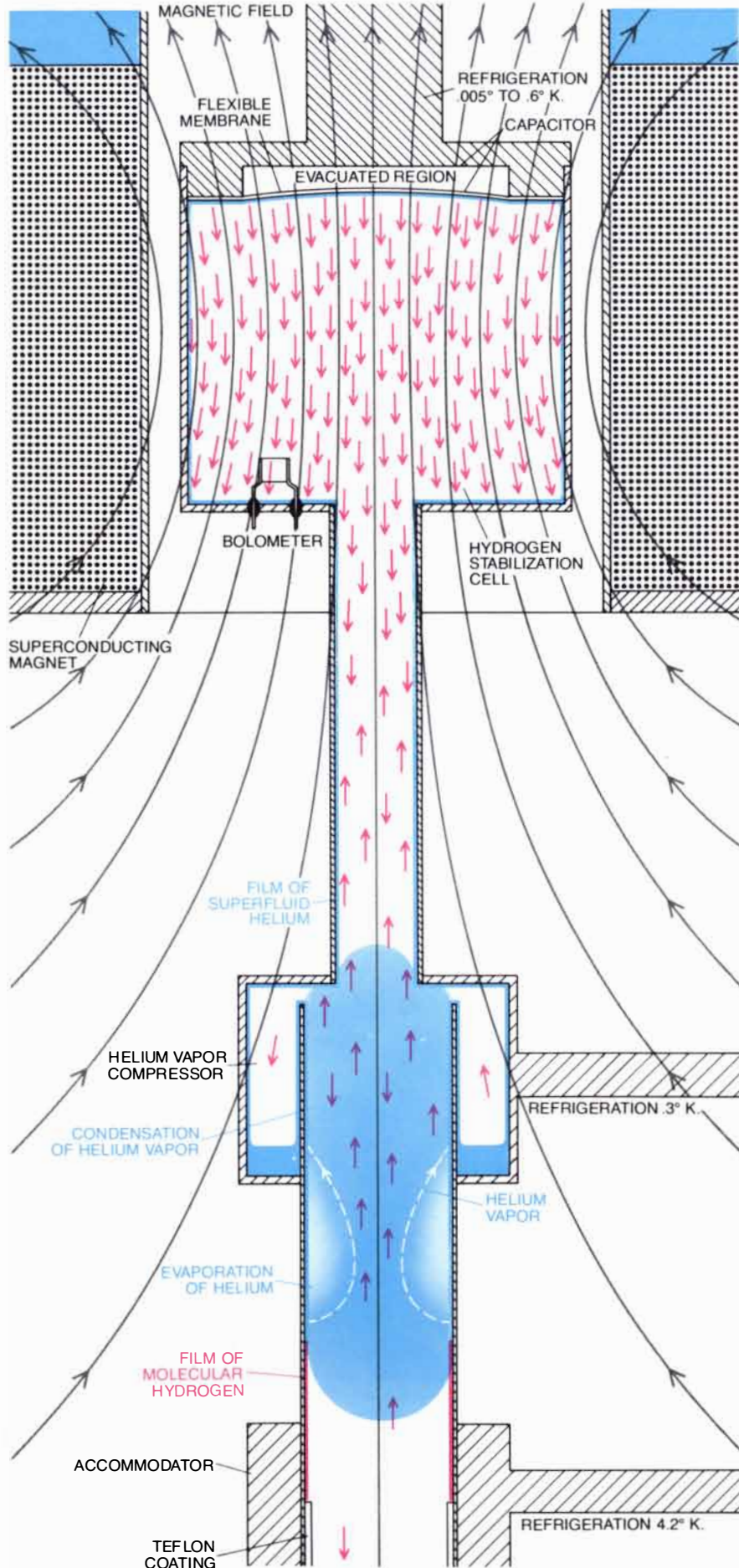
tion technology. If the helium is allowed to condense in the hydrogen-stabilization cell, the heat of vaporization acquired in the warmer region will be released to the walls of the cell. The heat energy can be so great that it prevents the low temperatures required for stabilization from being attained. We resolved the problem by installing an additional stage of cooling between the accommodator and the hydrogen-stabilization cell. The device returns the returning helium gas to recondense before it reaches the cell and thus absorbs the heat of condensation. The returning helium gas also helps to confine the hydrogen gas in the cell. Hydrogen atoms whose trajectory might allow them to escape from the cell collide with the more massive, upward-moving helium atoms and are returned to the cell. In this way the cooling stage acts as a vapor-driven compressor for the polarized hydrogen gas.

To detect stabilized atomic hydrogen we developed a simple and surprisingly sensitive technique. Whenever two atoms of hydrogen recombine, a large quantity of heat energy is released. By causing the gas in the stabilization cell to recombine suddenly, one can measure the temperature increase in the cell. Thus one can determine the density of the atomic hydrogen before recombination from the known recombination en-

ergy and the known volume of the cell.

To induce the sudden recombination and carry out the measurement we installed a square chip of carbon, about a millimeter on a side, in the stabilization cell. The chip was cut from a radio resistor and was suspended in the cell by fine wires. Because the electrical resistance of carbon changes with temperature the chip functions as a bolometer, and the changes in temperature can be calibrated from the changes in resistance. While the stabilization cell is filled with atomic hydrogen gas the fine wires and the carbon chip are coated with a film of superfluid helium.

When the density of the gas is to be measured, an electric current is passed through the bolometer, causing it to heat up. The superfluid helium evaporates from the chip faster than it can be replenished by the superfluid flow along the wires, and so the bare carbon surface becomes exposed to the atomic hydrogen gas. Within about 20 milliseconds the hydrogen atoms in the cell strike the carbon surface and recombine, causing a tiny explosion in the cell and a sudden, measurable increase in its temperature.

Although the bolometer is an effective and inexpensive detector, it does destroy the gas sample. A capacitance pressure gauge does not suffer from this disadvantage. The electric charge that can be held on two parallel plates of a capacitor depends on the distance between the

plates. If one of the plates of a capacitor is a flexible membrane, distortions of the membrane caused by changes in pressure can be registered as an electrical signal.

Given the combination of temperature and density attained so far with atomic hydrogen, what are the prospects for achieving Bose-Einstein condensation? In attempting to reach lower temperatures and higher densities the first problem that has been encountered is that the coating of helium 4 on the wall of the hydrogen-stabilization cell begins to adsorb hydrogen atoms. The attainable density is therefore limited by the increasing rate of recombination. Ideally a surface having a weaker adsorption potential than helium 4 has must be found. One surface coating is a thin film of the isotope helium 3 that floats above a layer of helium 4. The coating has an adsorption potential about a third that of helium 4. Although a film of helium 3 allows the hydrogen density to be increased, the increase is not enough to allow Bose-Einstein condensation.

Current efforts to achieve Bose-Einstein condensation are taking another approach. A detailed quantum-theoretical analysis shows that the spin-polarized hydrogen gas is made up of atoms in two states. In the B state both the electron- and the proton-spin vectors point down. The A state is not as simple. In the A state the configuration of the spins depends on the intensity of the magnetic field, and for all finite intensities the state is a "mixture" of two configurations. Although the electron-spin vector for an atom in the A state is most likely to point down, there is a small but nonzero probability that it points up; the proton-spin vector points in the direction opposite to that of the electron-spin vector.

Because of the mixture of spin vectors for atoms in the A state there is some probability of recombination even among atoms that have been spin-polarized. If atoms in the B state can be segregated from those in the A state, however, no recombination will take place among the atoms in the B state. Richard W. Cline, Thomas J. Greytak and Daniel Kleppner of the Massachusetts Institute of Technology have recently succeeded in preparing an extremely pure sample of spin-polarized hydrogen in the B state. With a gas made up of such atoms it should be possible to maintain a much higher concentration of atomic hydrogen on a surface before the rate at which recombination takes place becomes unacceptable. It may then be possible to compress the gas to the densities required for Bose-Einstein condensation. If that can be done, the scientific community will be presented with a state of matter that may exhibit new and exotic quantum properties.
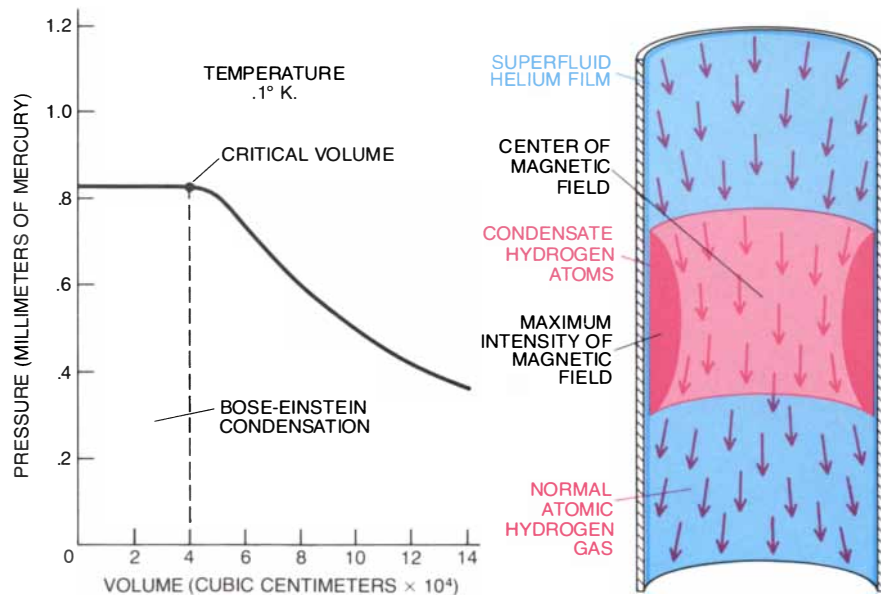
BOSE-EINSTEIN CONDENSATION should be detectable by studying the relation between the pressure and the volume of a gas at constant temperature. For an ideal gas without atomic interactions pressure varies inversely with volume, as long as the volume is significantly greater than the critical volume for the condensation. At the critical volume the atoms begin to enter the ground state and cease contributing to the pressure of the gas. Hence the pressure does not increase as the gas is further condensed. Hydrogen atoms in the gas have a magnetic moment, so that they can be localized in space by the gradient of a magnetic field. Atoms in the lowest energy states tend to occupy the regions of the magnetic field with the highest intensity, so that the Bose-Einstein-condensed atoms should tend to cluster where the field is strongest. The strongest region of the field generated by a solenoid magnet lies at the sides of the central plane of the magnet. Thus if the hydrogen gas is confined to a cylindrical tube in the bore of the magnet, the condensate atoms are expected to form a dense ring in the central plane.

74
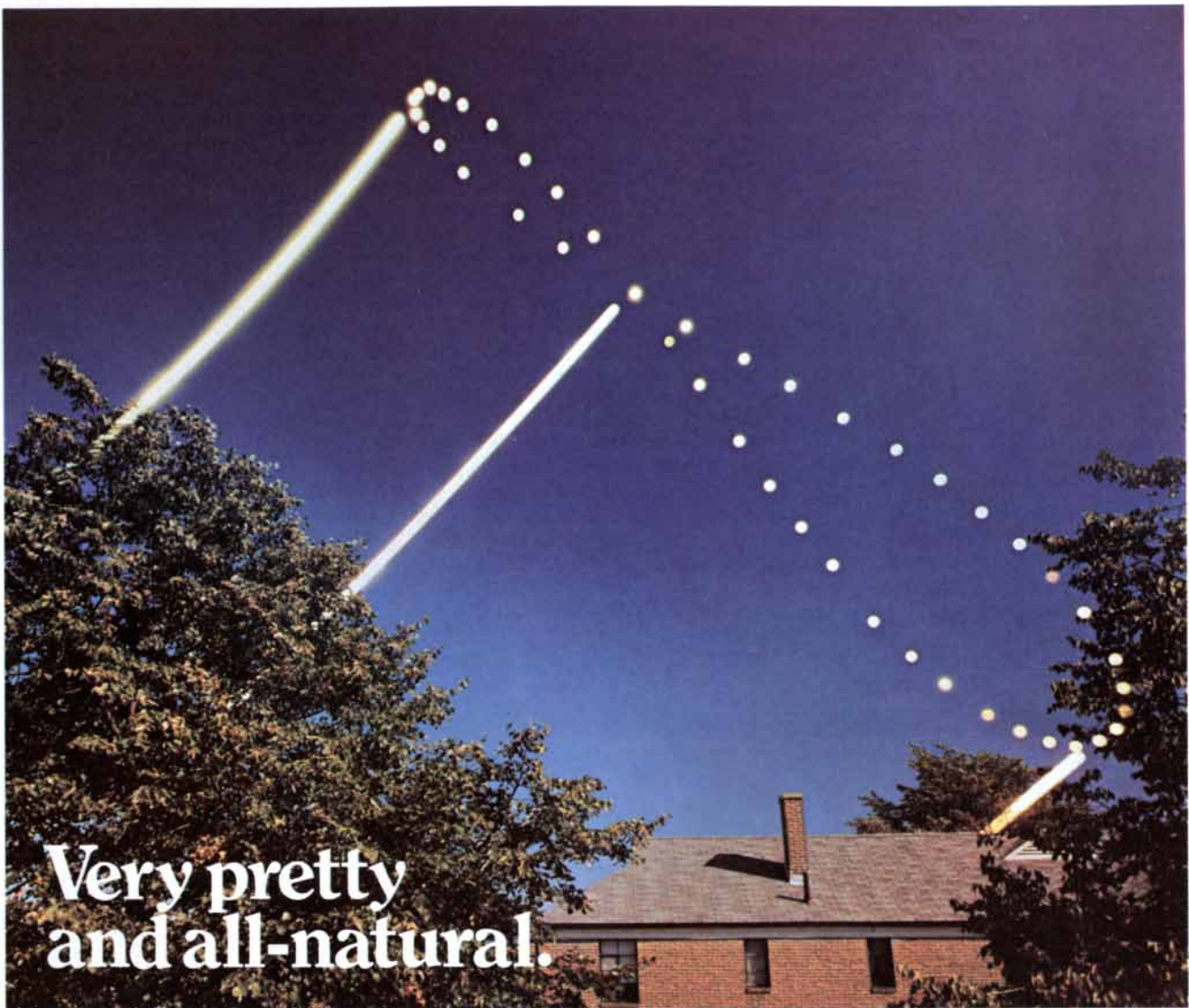
Very pretty and all-natural.

But what does it prove?

It proves how clever were certain parties who, long before the availability of Kodak Vericolor II professional film, drew an 8-shaped analemma figure on globes and sundials to show the daily declination of the sun over the course of a year.

That, though, hardly needs proving. Nor is proof needed that a decent result is obtainable from a sheet of that film exposed 48 times over a full year.

Why, then, do the directions for use say "process promptly"?

ANSWER: This is one of the "professional" films we make for portrait and commercial photographers. These users refrigerate their film and process promptly anyway, out of business necessity. Their raw film inventory justifies dedicating a refrigerator to it. We take this into account in balancing the relative sensitivities of the three color layers for best color rendition of subjects like fashions and flowers by which such users earn their bread. We take enormous pride in being that fussy about nuances of color which most people can scarcely appreciate.

Not everyone blessed with such sensitivity to color depends on it for a living. Fortunately for us, they too buy film and expect the same high level of color rendition though their shooting schedules are not so tight, they may not process so promptly, and their refrigerators are more likely to be stocked with watermelon and hamburger than color film. This we must assume in balancing the relative layer sensitivities of "amateur" films.

All this crafty scheming for fine tuning of photographic properties to differences between use by professionals and non-professionals is just an example of the upgrading in product performance that goes on over the years.

The "process promptly, store at 55° F (13° C) or lower" advises which kind you are using. Whoever reads it as an indication that "professional" film is less stable than "amateur" film is much mistaken.

In *Sky and Telescope* for June, 1979, that magazine's Dennis di Cicco told all about how he took this one-year-long picture. Analemmas at once became a hot subject. The magazine's August '79 issue found it necessary to advise readers how to purchase a 16″ x 20″ print. The deal is still on. If you want to know all about analemmas, see page 20 of *Sky and Telescope* for July, 1972. If you want a copy of Kodak Pamphlet E-36, "Nine KODAK Color Films for Process C-41," write Department 412-L, Kodak, Rochester, NY 14650. If you need detailed information on astronomical photography, make that address "Scientific and Technical Photography" instead of "Dept. 412-L."

© Eastman Kodak Company, 1980

I'll be home for Chivas.

SCIENCE AND THE CITIZEN

Moonshine

The Federal gasohol program is intended to reduce gasoline consumption in the U.S. and thereby lower the amount of crude oil imported. Tax exemptions and other incentives encourage private industry to manufacture gasohol (a mixture of alcohol and gasoline) for use as a motor fuel. The economic consequences of the program have recently been analyzed by Fred H. Sanderson of Johns Hopkins University; writing in *Resources,* a publication of Resources for the Future, he points out that even if the goals of the program are met, the quantity of petroleum saved will be considerably smaller than was initially expected. Furthermore, the diversion of large amounts of corn to the production of alcohol for gasohol could have the unintended effect of sharply raising the price of corn and other agricultural products in the U.S. Such a price increase would reduce exports of some American crops and cause a large rise in retail food prices.

Most gasohol sold in the U.S. is one part ethanol (ethyl alcohol) and nine parts gasoline. (In other countries other formulas have been adopted; Brazil, for example, plans to replace gasoline with pure ethanol by the 1990's.) Automobile engines run reasonably well on the nine-to-one gasohol. When ethanol is added in this proportion to a given grade of gasoline, the number of miles obtained per gallon of fuel is reduced slightly but the octane rating is raised a little. The ethanol will probably be made by fermentation of corn of the grade suitable for animal feed.

The central legislation in the gasohol program is the Energy Security Act of 1980. The act declares that gasohol should be the main motor fuel in the U.S. by 1990. It is estimated that by then U.S. drivers will consume 100 billion gallons of fuel annually. If the fuel is gasohol, 10 billion gallons of ethanol will be needed to meet the demand. Making the ethanol will consume 100 million tons of corn, an amount larger than the current volume of U.S. corn exports.

A variety of incentives to gasohol production were created by the Energy Security Act and other measures. The most substantial incentive was to make gasohol exempt from the Federal tax on highway fuels, which is now four cents per gallon. Although the value of the exemption is small for each gallon of gasohol, because of the nine-to-one ratio of ethanol to gasoline the exemption is worth 40 cents per gallon of ethanol. Other incentives include Federal tax credits and Federal loan programs for investors in industrial plants designed to make ethanol. In addition 25 states have exempted gasohol from state gasoline taxes.

Proponents of the gasohol program have argued that each gallon of ethanol in gasohol reduces the consumption of gasoline by one gallon. Sanderson's analysis suggests that the actual savings are lower. He classifies gasoline, diesel fuel and natural gas as premium fuels whose use should be balanced against that of ethanol. When factors such as the lower fuel mileage characteristic of gasohol and the premium fuels needed to make ethanol are taken into account, the net savings is about .65 gallon of premium fuels per gallon of ethanol. This ratio would lead to the saving of about 170 million barrels of petroleum in 1990, worth about $6 billion in 1979 prices.

Not only will the intended effect of the gasohol program be less dramatic than its advocates expect but also its unintended consequences may be actively disadvantageous. The price of corn and other grains depends in part on the demand for resources such as agricultural land, labor, equipment and seed. An additional 31 million acres of farmland would be required to make 10 billion gallons of ethanol. An increase of this magnitude in the acreage in corn could double the price of corn and other feed grains by the end of the decade. There would be a smaller but still substantial increase in the price of cotton, tobacco, soybeans and some other crops.

Corn and other feed grains grown in the U.S. are now quite competitive in world markets. Doubling their price might reduce demand to some extent, but there would probably still be an increase in the monetary value of exports of these products by the end of the decade. Other agricultural exports, however, including cotton, tobacco and livestock products, are priced less advantageously. Price increases could result in a net loss of as much as $5 billion per year of exports by 1990.

The increase in the price of agricultural exports might have disadvantages beyond the immediate debits in the balance of payments. Because these products are now inexpensive there is little incentive for growers in other countries to increase their crops. Rising prices might induce other countries to attempt to achieve greater self-sufficiency, ultimately reducing the foreign market for American agricultural products.

The domestic cost of the gasohol program will also be high, Sanderson argues. Grain for animal feed makes up a substantial fraction of the cost of production of many foods. According to Sanderson, a doubling of grain prices and a smaller rise in the price of other agricultural goods would raise retail food prices by 12 percent. In 1979 prices this amounts to about $30 billion. The exemption of gasohol from the Federal gasoline tax will reduce Government revenues by $4 billion per year by 1990 if the goal of manufacturing 10 billion gallons of ethanol is reached in that year. The overall loss to the Federal Government will have reached $27 billion by 1992, the year the exemption is scheduled to expire.

Sanderson concludes that the conservation of motor fuels and increased reliance on nuclear power, coal and synthetic fuels derived from coal are cheaper and more effective means of reducing petroleum imports than the gasohol program is. He proposes the early elimination of the tax incentives for the manufacture of gasohol. The longer the incentives are maintained, he states, "the more serious the adjustment problem will become as high grain prices sustained over a long period tend to be locked into land values. The resulting adjustment costs are likely to burden American consumers and taxpayers for many years after the gasohol program has run its course."

Squeeze Play

In the manufacture of microelectronic devices the area on a silicon chip required for a given function has been shrinking by a factor of two every 18 months. The cost per bit of microelectronic memory has been falling along a roughly parallel curve. In the immediate future these trends are certain to continue, but what of the longer term? Current methods of fabrication will reach fundamental limits after only a few more doublings of the density of microelectronic circuit elements. Moreover, the high density tends to raise the proportion of defective chips per batch, and packaging and inspection costs per chip also increase. Gordon E. Moore of the Intel Corporation has predicted that unless there are breakthroughs in manufacturing techniques the cost per bit of memory will cease to decline in about 1985. The question, then, is whether breakthroughs can be expected. As it happens, new methods of fabrication now under investigation could lead to a further increase in density by a factor of 3,000.

Microelectronic devices are now made by photolithography. A silicon wafer a few inches in diameter is coated with a photoresisting agent and exposed to light through a mask on which is traced a pattern for the circuit elements of 100 or more chips. The illuminated areas of the photoresist are then dissolved by a developer solution and the areas of the silicon thus uncovered are

modified chemically. In this way the pattern on the mask can be reproduced in the electrical properties of the silicon. Several masks are needed to define the various layers of semiconductor, insulator and conductor that constitute a single level of microelectronic circuitry.

Photolithography is limited by the wavelength of the radiation employed to expose the resist. With visible light the narrowest features that can be created are about 500 nanometers wide. The most advanced chips in commercial production today have a minimum line width of about 1,000 nanometers. Exposing the resist with ultraviolet radiation or X rays can bring some improvement, but such energetic radiation causes another problem. Electrons knocked out of the material ionize additional molecules of the resist and so blur the etched lines.

One approach to higher line resolution that has long been considered is to etch the resist with a focused beam of electrons. The electrons vaporize the resist and create a channel that is quite deep compared with its width. Electrons are knocked out of the sides of the channel but few of them escape to the surface of the resist, where they would cause excess development and blur the image. Another advantage of electron-beam etching is that there is no need for a lithographic mask; instead a computer can steer the beam and directly control the pattern inscribed on the chip. The creation of the pattern directly on the chip is also the major drawback of electron-beam fabrication. Whereas electromagnetic radiation passing through a mask exposes the appropriate areas of all the chips on a wafer simultaneously, the electron beam must scan each area of each chip sequentially. The process remains too slow to be competitive.

If electron-beam etching does become economically feasible, it has the potential for scaling down the size of circuit elements several thousandfold. Michael Isaacson of Cornell University has employed an electron beam to inscribe lines two nanometers wide on the surface of a salt crystal. Although salt is not a suitable material for microelectronic devices, Isaacson proposes to fill the etched lines with metal to make extremely fine conduction paths. At this scale a square circuit element only 60 nanometers on a side could be made.

A quite different strategy for increasing the density of circuit elements is under investigation by Colin Wood of Cornell. The aim is to build three-dimensional microelectronic devices; all chips being made now have an essentially planar array of circuit elements. In this way a much higher density per unit volume can be achieved, even if the size of the individual elements remains unchanged.

Wood points out that the process of exposing a wafer to radiation and then etching it in a developer can be carried out repeatedly until several hundred layers have been deposited. The focus of Wood's investigations has been to develop a means whereby layers can be put down on a substrate quickly and accurately.

In the method adopted lately the pattern for a given layer is etched in a resist applied on top of all the earlier layers. The wafer is then exposed to a collimated beam of evaporated atoms or molecules, which are deposited in the etched regions to form the new layer. The intensity of the beam can be controlled so precisely that a coating one atom or one molecule thick can be laid down. In this way layers of semiconductor, insulator or conductor can be built up to thicknesses specified to within a monolayer: the depth of a single layer of atoms or molecules. Moreover, the chemical composition of the materials can be varied with each monolayer. Wood and his colleagues deposit alternate layers of gallium arsenide and gallium aluminum arsenide on the wafer; variations in the conductivity of the two substances can be exploited to fabricate circuit elements that enable currents to flow in all three dimensions.

There have been a number of proposals to incorporate biological materials and the technology of genetic engineering into the manufacture of microelectronic devices. Current work takes advantage of the adhesive properties of certain synthetic proteins to affix metallic conductors to a wafer. In principle the proteins might be grown cheaply in genetically modified bacteria.

James H. McAlear of EMV Associates in Rockville, Md., reports that his company has formed conduction pathways by depositing the synthetic protein polylysine on glass. A photoresist is placed over the polylysine layer and a pattern is etched in the resist by photolithography. The polylysine provides a sticky surface on which metal can then be deposited. Ultimately the organic molecules themselves might be made to conduct current, and other molecules might be employed to switch or regulate the current. It may be possible, therefore, to assemble electrical components from proteins and other molecules grown in bacteria.

Toolmaking Tools

The oldest piece of wood known to have been shaped by man, found in 1911 at Clacton-on-Sea in England by an amateur fossil hunter, is a tapered 15¼-inch length of yew. It is thought to be 250,000 years old. Now evidence has been adduced suggesting that human beings have worked wood for at least 1.5 million years. The evidence consists of the characteristic polish on the worn working edges of stone tools from five sites in the Koobi Fora region of Kenya, east of Lake Turkana.

Writing in *Nature*, Lawrence H. Keeley of the University of Illinois at Chicago Circle and Nicholas Toth of the University of California at Berkeley report their analysis of 54 stone tools from Koobi Fora strata reliably dated at 1.5 million years before the present. Sites of similar age in the area have yielded stone tools by the hundreds, but most of them are basalt, a material not suited for the microscopic analysis of wear patterns. The 54 tools are chert, a siliceous stone that can be so analyzed. Nine of them showed clear traces of wear. The appearance of their working edges was compared with that of tools newly made out of the same kind of stone and experimentally subjected to wear by cutting Koobi Fora grass and wood and other local plant and animal tissues.

Four of the ancient tools showed the "greasy" polish acquired by experimental tools used to cut meat and other soft animal tissues; the direction of striations on the edges of the tools indicated that they had been used with a slicing motion. Two other tools showed the highly reflective polish acquired by cutting the siliceous stems of grasses or reeds. The other three showed the less reflective polish acquired by working wood; two of them had been used as scrapers and the third as a kind of saw.

The meat-cutting polish, Keeley and Toth note, independently corroborates the evidence for meat eating at Koobi Fora in the form of animal bones with cutting marks on them. Two of the meat-cutting tools had been found within a meter of the humerus of a large member of the cattle family with cutting marks on it. The tools with the silica polish indicate they had been used to cut grass or reeds but yield no hint of the reason for this activity. The tools with the woodworking polish, however, suggest that the manufacture of such implements as spears or digging sticks was a part of the repertory of human skills more than a million years earlier than anyone has so far had any reason to believe.

Thin Red Line

Red blood cells distribute oxygen to the tissues; white blood cells participate in the immune response to foreign organisms or substances. This division of the blood cells into two functional groups separated by an arbitrary color line may be an investigational artifact, according to a hypothesis presented in *The Lancet* by Israel Siegel, Tian Lin Liu and Norbert Gleicher of the Mount Sinai Hospital Medical Center of Chicago. Certain activities of the red cell do not make sense, they argue, if the cell's only role is a respiratory one; on the contrary, it appears the red cell may be an essential component of the immune system.

Siegel, Liu and Gleicher call attention

78

in particular to the well-established fact that red cells adhere to immune complexes, the aggregates formed when antibody molecules combine with the specific antigen to which they are sensitive. The formation of an immune complex labels the antigen as foreign, but the labeling in itself cannot destroy the antigen or eliminate it from the body. Further action is mediated by complement, the collective term for a group of nine proteins in the blood plasma; complement is activated by certain immune complexes, and the complement in turn precipitates inflammation and other defensive reactions. Eventually most immune complexes are broken up and cleared from the body. It is this final stage of the immune response in which the red cell may take part.

Receptors on the surface of red cells are known to recognize and bind to a component of complement, with the result that the cells adhere to antigen-antibody-complement complexes and so to one another, forming clumps. For a long time this red-cell immune adherence was noted in primates; it was considered an interesting phenomenon but was not generally given any functional significance. Siegel and his colleagues discovered red-cell immune adherence in nonprimates and suggest it is common to all mammals. Studying the effect primarily in rabbits, they found that the extent of adherence varies widely and in ways that imply it is influenced by two factors: an inhibiting agent in the blood serum and an innate receptor-activity level that is characteristic of cells from a particular individual at a particular time. Receptor activity is higher and the level of inhibition is lower in newborn rabbits than in adults; receptor activity is depressed in human cancer patients and enhanced in patients with an autoimmune disease.

Siegel, Liu and Gleicher note that the modulation of red-cell immune adherence suggests it is a functional response to varying physiological requirements. The obvious function is the removal of immune complexes from the bloodstream. Removal has been assumed to be the job of various white blood cells, which also have surface receptors for complement. Calculating the total number of receptors on white cells and on the much more numerous red cells, Siegel and his colleagues conclude that the red cells must be the main agents for clearing immune complexes from the circulation.

In some immunological disorders immune complexes are not promptly eliminated but tend to be deposited in the tissues, where they can cause damage. In a paper published in *The Lancet* a week before that of Siegel and his colleagues a group of investigators at the University of Tokyo reported that red-cell adherence to complement is defective in patients with systemic lupus erythemato-

sus, a disease in which tissue damage has been attributed to the deposition of immune complexes.

Another possible immune function for red cells is suggested by the observation that an individual's red cells adhere to his own *T* lymphocytes. The *T* lymphocytes are white cells that bind to an antigen and thereupon participate in various aspects of the immune response. Experiments done at Mount Sinai showed that red cells, by adhering both to immune complexes and to *T* cells, tend to bring the *T* lymphocytes into close proximity to antigens. The red cells may therefore serve as intermediaries that promote the interaction of a *T* cell and its specific antigen, thereby enhancing *T*-cell activity.

What Leeuwenhoek Saw

The pioneer microscopist Anton van Leeuwenhoek was renowned in his own time (1632–1723) and is remembered today for the many dramatic discoveries and observations he made with his single-lens microscopes. He was the first to see bacteria, spermatozoa and red blood cells; he also examined the structure of the lens of the eye, the striations of striated muscle, the mouthparts of insects, the fine structure of plants and many other things about which little or nothing had been known. His methods, however, have been obscure; in some 375 communications to the Royal Society of London on what he had seen he said little about how he had done it. Early last year Brian J. Ford of the Science Unit in Cardiff found in the archives of the Royal Society several of the very specimens van Leeuwenhoek had examined. The discovery was widely noted at the time. Now Ford has published in *Notes and Records of the Royal Society of London* an account of what he found and what his examination of the specimens has revealed.

Ford began his search "in the hope of uncovering evidence of [van Leeuwenhoek's] work on inanimate specimens and the techniques he might have used in preparing them for examination." In van Leeuwenhoek's fourth communication to the society, sent on June 1, 1674, Ford found a small envelope pasted to the back of the blank sixth and last page.
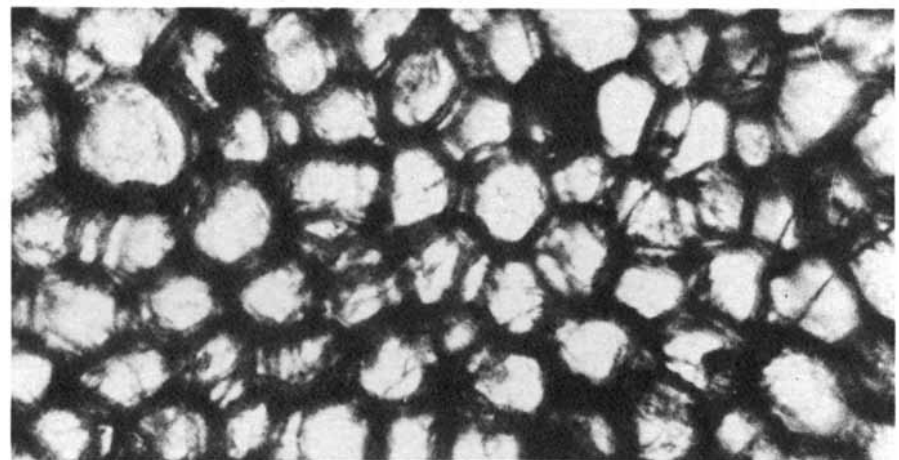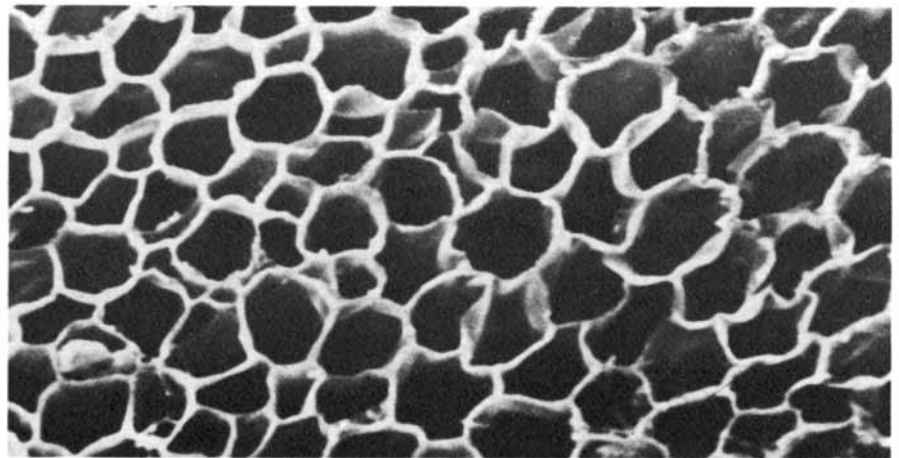
LEEUWENHOEK SPECIMEN is a section of cork cut by the pioneer microscopist and sent by him to the Royal Society of London in 1674, along with several other specimens. They were discovered in the archives of the Royal Society last year by Brian J. Ford of the Science Unit in Cardiff, who made these micrographs. At the top the section is seen in a scanning electron micrograph. At the bottom it is seen through the most powerful of Anton van Leeuwenhoek's original microscopes surviving. The single-lens instrument has a magnification of 266 diameters.

79

"Cautiously I opened it, anticipating that it contained some artifact; inside were four smaller paper packets, and three of these proved to contain finely cut sections that van Leeuwenhoek had prepared with his own hand for the interest of the Royal Society.... Specimen No. 1 was of cork from *Quercus suber,* prepared as finely translucent sections ranging in length from approximately 1.5 millimeters to some seven millimeters. Even a cursory examination showed that they were of histological thinness and compared favorably with the best hand-cut sections that might be obtained today. Specimen No. 2 was clearly elder-pith from *Sambucus nigra,* this time in the form of rounded sections cut from young stems perhaps 15 millimeters in external diameter. Packet No. 3 was empty, but the labeling suggests it contained material from a writing-quill. Specimen No. 4 was not immediately identifiable: it had the appearance of a sectioned aquatic stem with a thickened outer wall and a spongy medullary cavity. However, microscopy proved that this initial impression was erroneous. The sections were of optic nerve, the 'lacunae' being derived from the loss of material through degeneration over the centuries."

In the letter that accompanied the specimens van Leeuwenhoek said he had cut them with a sharp razor and suggested that "when any of these particles is applied to the pin of a Microscope, the instrument may be held within doors and in the shade." Ford points out that the passage gives "an indication of the method van Leeuwenhoek used to cut his sections, an outline of the way they were held in position on the stage bracket of his microscope, and advice on the use of a restricted source of illumination which shows that he was aware of the benefits of limiting the diameter of the source in an attempt to control the aperture of the illuminant cone."

Ford found five more packets of specimens in two later letters from van Leeuwenhoek but concluded that they "are not as interesting in many ways: four of them contain portions of entire specimens, so they do not reveal any degree of microtechnique, and the dissected contents of the remaining specimen show a growth of fungus material that reduces its interest." He also satisfied himself, by experiments with one of van Leeuwenhoek's microscopes and similar ones, that van Leeuwenhoek could resolve bacteria in some of his instruments and almost surely did.

The Year 10^{116}

Cosmologists have extrapolated the laws of physics backward in time to within a fraction of a second of the big bang. They have been less successful in extrapolating them into the deep future. Now Don N. Page and M. Randall McKee of Pennsylvania State University have described the fate of the universe if the current expansion continues forever. The evolutionary scenario they develop is cold, diffuse and chaotic, and the chaos increases without limit. They present their work in a recent issue of *Physical Review.*

It will be at least 10^{116} years before the Page-McKee universe comes to pass. It is a distant future that has only recently become accessible to informed speculation through the development of theories about the decay of elementary particles. How remarkably distant it is can be gauged by the fact that the age of the universe is now thought to be between eight and 18 billion years, or about 10^{10} years.

Page and McKee assume that the large-scale density of the mass in the universe is the same in every direction and just low enough for the expansion to continue forever. After 10^{29} years the universe will expand to 10 trillion times its present size. Virtually all stars and galaxies will have yielded to internal gravitational forces and collapsed to form black dwarfs, neutron stars or black holes. The collapse may engulf masses as large as galaxy superclusters and create huge black holes. In the spirit of their model of a uniform overall mass density, however, Page and McKee assume that black holes larger than superclusters do not form. The gravitationally bound clumps of mass existing at this stage may be huge, but they will be no larger than 100 trillion times the mass of the sun.

At the beginning of this strange epoch the matter in the universe that is not trapped in black holes is expected to decay. Theoretical arguments that are now being tested by experiment suggest that the lower limit on the lifetime of the proton is 10^{29} to 10^{30} years. The neutron and other heavy particles classified as baryons are also expected to decay, so that all matter both within the cold remnants of stars and scattered about as interstellar gas will become a gruel of electrons, positrons, neutrinos, photons, gravitons and perhaps other stable particles as well.

The electron and the positron are antiparticles, so that when they encounter each other, they are annihilated. In the relatively dense environment of a black dwarf or a neutron star the frequency of such encounters will be high. The energetic photons released by the annihilations will tend to redistribute their energy over a wide thermal spectrum, corresponding to a temperature of about one degree Kelvin at the surface of a black dwarf and about 100 degrees K. at the surface of a neutron star. The thermal photon emission, cold as it is, will still dominate the background of microwave radiation that is thought to be left over from the big bang, just as it does today.

Today the background radiation has a temperature of three degrees K.; in the year 10^{30} it will have a temperature of 10^{-13} degree K.

In the extremely rarefied interstellar gas the electrons and positrons that remain after baryon decay are much less likely to encounter one another. The average separation between two such particles will be on the order of one astronomical unit: the distance between the earth and the sun. Nevertheless, since the electron and the positron are oppositely charged, they can form a bound state if their total kinetic energy becomes less than the binding energy. The bound state resembles a hydrogen atom in which the proton has been replaced by a positron; such an atom is called positronium.

As the universe continues to expand, the attractive forces between neighboring particles will decrease, but the kinetic energy of the particles will decrease slightly more rapidly. After 10^{71} years the attractive forces will finally overcome the kinetic energy, and interstellar positronium will begin to form. In this epoch the universe will be 10^{40} times its present size, so that the atoms of positronium will be almost unimaginably rarefied states of matter that will fill regions of space as large as the present observable universe.

Once the electrons and positrons begin to form positronium it is only a matter of time before they spiral into each other and are annihilated. The time required, however, is 10^{45} times as long as the age of the universe when the positronium forms. Thus the annihilation of a large fraction of the electrons and positrons in the interstellar gas is expected to take at least 10^{116} years. The annihilations give rise to high-energy photons that may never encounter one another often enough to become thermally distributed over many energy states.

Meanwhile the black holes formed by the collapse of stars and galaxies may coalesce, but quantum-mechanical effects allow them to evaporate as well. A black hole of one solar mass will evaporate in 10^{66} years, whereas a black hole of supercluster size (about 10^{14} solar masses) will evaporate in 10^{108} years. It is not known which process, coalescence or evaporation, will prevail in the end, but in either case evaporation late in the history of the universe also gives rise to high-energy photons that are not likely to be redistributed thermally.

Therefore although the background radiation of the universe will be cold in the year 10^{116} (about 10^{-70} degree K.), the decay of black holes and positronium atoms will never allow the total energy in the universe to reach a random, thermal distribution. According to Page and McKee, the heat death of the universe that was contemplated by physicists in the 19th century will never come to pass.

80

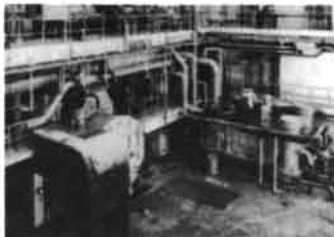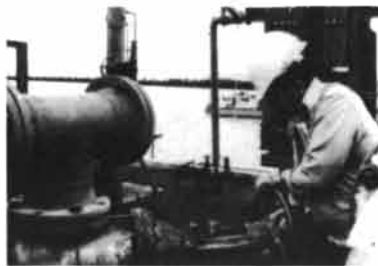# PHILIPPINES 2000: INDUSTRIAL AND TECHNOLOGICAL FUTURES

PHILIPPINES 2000: INDUSTRIAL AND TECHNOLOGICAL FUTURES

by Cecilia A. Sanchez, economic and business consultant and Minerva R. Fajardo, a technical and industrial writer

The Philippines, a developing country in what is predicted to be the world's leading growth area for the rest of this century, faces the eighties with a clear vision of where it has been and where it wants to be. The economy's growth momentum over the last decade has surpassed all its previous records. Its goal of a balanced agro-industrial economy now appears closer and more attainable. In this scenario, the value of science and technology as a tool for development has come into focus.

This special report examines the Philippines' unique approach to the utilization of science and technology (S&T) for development. It describes the country's science institutions and defines specific S&T issues. It indicates the level of technology now present in the country—as reflected in its industries, in the products it manufactures for the domestic and export markets, and the people who man its productive enterprises. It shows how simple indigenous technologies have been improved and disseminated, for people to use in various ways that upgrade productivity and the general quality of life, and how imported technologies have been adapted to suit unique Philippine raw materials and local conditions. Quite obviously, technology in the Philippines lacks the sophistication which characterizes those of modern highly industrialized societies.

Nevertheless, it has served an important role in the task of national development. Finally, the report presents the plans for Philippine development to the year 2000 and the role S&T is expected to play in their realization.

Science and technology (S&T) in the Philippines reflects a growth pattern that parallels the progressive stages of its industrial development.

In the immediate postwar period, the Philippines adopted a policy of import substitution, characterized by an artificially low exchange rate and a protec-

tionist trade policy. This led to the establishment of an inward looking manufacturing sector which limited itself to the assembly, finishing and packaging of products mainly for domestic consumption. Technology utilized by these manufacturing industries tended to be highly capital intensive, having been mostly imported from developed countries.

From the early fifties to almost the end of the sixties, therefore, technological know-how was practically confined to the extraction of mineral and forest resources which were sold to the world commodity markets, and to the processing and finishing of end products to replace imports of consumer goods.

The seventies saw the Philippines embarking on an industrialization program intended to expand and re-structure its industrial sector so that it could better contribute to socio-economic goals. The program addressed itself to the creation of employment through labor-intensive activities, dispersal of industries throughout the country, the promotion of small and medium scale projects, the acceleration of exports of finished manufactured goods, and the further processing of raw materials into semi-finished goods for exports.

Foreign and domestic investors were encouraged to go into projects listed in

the investments and export priorities plans as preferred areas. New and pioneering projects were especially encouraged.

As a result, new projects involving foreign investors were registered with the Board of Investments to produce an increasing range of products for domestic use, among these: nylon—6 filament, polyester fibers and filaments, rayon staple fibers, furfural, stearic acid and industrial explosives, refractory bricks and ceramic dinnerware, high-tension porcelain insulators, as well as distribution and power transformers. Exports of finished and semi-finished products also increased, such as a range of garment and textile products, footwear, leatherwear and sportswear as well as electronic components and parts.

In its efforts to close the apparent technological gaps between the extraction and finishing ends, the nation has experienced and still continues to experience, a number of constraints which are perhaps typical to most developing countries.

One is the relatively low level of public expenditure allocated to enterprises engaged in technology development. This situation is quite understandable since the government has limited financial resources and many other more immedi-

P3

THE PHILIPPINES

THE PHILIPPINES

LUZON SEA

LUZON

PHILIPPINE SEA

MANILA

V

LEGAZPI CITY

III

IV

SIBUYAN SEA

VIII

V I S A Y A S

PALAWAN

PUERTO PRINCESA

CEBU CITY

VI

VII

BOHOL SEA

IX

X

SULU SEA

MINDANAO

XII

DAVAO CITY

ZAMBOANGA CITY

XI

MINDANAO SEA

REGION I	ILOCOS
REGION II	CAGAYAN VALLEY
REGION III	CENTRAL LUZON
REGION IV	SOUTHERN TAGALOG
REGION V	BICOL
REGION VI	WESTERN VISAYAS
REGION VII	CENTRAL VISAYAS
REGION VIII	EASTERN VISAYAS
REGION IX	WESTERN MINDANAO
REGION X	NORTHERN MINDANAO
REGION XI	SOUTHERN MINDANAO
REGION XII	SOUTHWESTERN MINDANAO

0 50 100   200   400 KM

ate development priorities than research and development (R&D). As a result, the resources and infrastructure for the pursuit of new scientific and technological frontiers in the country have been extremely limited. This has severely curtailed the professional opportunities of a number of the country's technical elite and created a "brain drain" in the country's scientific manpower base.

In addition, there have been problems of access to the pool of scientific knowledge.

Viewed negatively, these factors have restrained the country's scientific and technological options. However, the handicaps have enhanced the utilization of inherent native technological capabilities and sharpened the Filipino's natural gift for adaptive innovations.

National science policy has taken a pragmatic turn in recent years. Having survived the economic turbulence of the 70's and with eyes set on an accelerated pace of industrialization while sustaining and improving agricultural productivity gains, research and development are now consciously directed towards searching for appropriate alternative technologies befitting the unique Philippine development setting. Out of specifically identified needs and priorities have sprung S&T's current strategic thrusts: harnessing available indigenous raw materials for food, energy, clothing, housing, medicines, tools and hardware for farm and factory.

Efforts to draw upon available outside sources continue, but in a more selective and discerning manner. The country reaches out for new technologies to enhance its productive capabilities but it is conscious of the fact that even in S&T the path of self-determination is the only one that pays real dividends. As it moves towards 2000, the Philippines appears to be on the final phase of evolving a clear-cut S&T policy that recognizes its needs and opportunities.

R&D INITIATIVES
OF THE PUBLIC SECTOR
AND S&T ISSUES

The National Science Development Board (NSDB) is the major public body responsible for initiating and managing scientific and technological programs in the Philippines. Under its umbrella are several agencies with specific research responsibilities, including the Forest Products Research and Industry Development Commission (FORPRIDE-COM), the Food and Nutrition Research Institute (FNRI), the Philippine Inventors Commission, the Commission on Volcanology, the Philippine Council for Agriculture and Resources Research (PCARR), and the National Institute of Science and Technology (NIST).

Two institutions attached to the Ministry of Trade and Industry are engaged in specialized research: the Philippine Textile Research Institute (PTRI) and the Metals Industry Research and Development Center. Two more technology related agencies have also been recently created: the Technology Transfer Board and the Technology Resource Center.

Organized in 1958, NSDB has been attempting to put shape and system into the country's formerly scattered research efforts. A science policy has been slowly evolving, emphasizing the pivotal role S&T would play in overcoming the country's development problems. NSDB's elevation to a full-pledged ministry in 1978 signaled the government's renewed support and commitment to the country's scientific body. That year also saw the organization of the National Academy of Science and Technology, initially composed of ten leading scientists who make up "a reservoir of competent scientific and technological manpower for the country."

Metal Industries Research and Development Center (MIRDC), serves the metals industry as a non-profit research and technological institution providing technical expertise in the training of engineers and technicians, skills accreditation, quality control and testing of metal products, information exchange, and research and advisory services. A staff of UN consultants lend their expertise to the Center's multifarious efforts.

The Philippine Textile Research Institute (PTRI), provides physical and chemical testing services, standards formulation services and extension services.

The Technology Transfer Board (TTB), created in 1978, ensures that imports of foreign technologies are relevant to the country's industrialization goals and that the terms of their transfer are reasonable. "Our role is to see to it that a genuine transfer of know-how occurs rather than just a package transfer involving payment," says Lilia R. Bautista, executive director of the TTB and a governor of the Board of Investments. "We also make sure that the technology recipient has access to the continuing improvements made on the technology he buys," she continues.

The Technology Resource Center (TRC), organized in 1977, concerns itself with the selection, management and dissemination of appropriate technologies to various types of users, particularly small and medium scale industries.

"We prepare the user to receive the technology and help him to apply it effectively," explains Arthur Alvendia, director-general of the TRC.

Over the past decade, public research has been directed to developing alternative energy sources, increasing agricultural productivity and resources utilization, improving public health, developing low cost housing technology, finding new avenues to increase job opportunities and strengthening scientific manpower and institutional capabilities. These efforts have resulted in some gains relative to the nation's goal of self-reliance but much remains to be done for S&T to generate the kind of impact which an informed public expects.

Questions about the country's science and research priorities are now being raised. One group claims that public monies for research unduly favor applied and technological fields at the expense of more basic, fundamental research. They warn that this could well leave the Philippines out in the cold in its bid for a higher technological level in the future. Another group questions the apparent imbalance in the heavy R&D support for agriculture as against industry. One industry sector maintains that there is little significant relevance between NSDB projects and the requirements of industry. They claim that a gap exists between the production of the new technology and its adoption. The same group suggests that there should be an organized plan to disseminate research results in some useful form.

Dr. Antonio Arizabal, a metals specialist and executive director of the Metal Industries Research and Development Center (MIRDC) would attribute this to the fact that technology is a complex system, not easy to create and difficult to understand. In the absence of the ability to select, adapt or improve upon the technology brought in, the country suffers the short end of the bargain.

Prime Minister Cesar Virata, concurrently Finance Minister, points out that certain Filipino firms have improved upon technologies licensed to them, arising from the need to adapt imported processes to the unique qualities of local raw materials. In due course, technical capabilities have been upgraded.

One case cited is the nickel project of Marinduque Mining and Industrial Corporation which is registered with the Philippine Board of Investments as a pioneering venture. Marinduque's nickel refinery covers the mining of laterite ore and the production of pure nickel metal. As Jones Castro, its senior vice president explains, the commercial processes it employs in the production of nickel from ore into metallic form are not new. In Marinduque's plant, however, the two ends have been joined and it can be said that the Marinduque's Nonoc nickel refinery is the first integrated installation in the world where the nickel laterites are processed into metallic powder and briquettes. Practically 100 percent of Marinduque's production is exported.

Philippine Wallboard, another Filipino firm, ventured into hardboard manufacturing in the mid-fifties, utilizing woodwastes as raw materials. A 50-ton capacity plant from Sweden using the wet process of manufacture was acquired. At every stage from plant start-up and testing until fully commercial operation, technical assistance was availed of. However, due mainly to distinctive differences between Philippine mahogany and the Swedish pine for which the equipment was really suited, adaptations had to be made in the manufacturing process. Now, Philippine "lawanit" has the same qualities as hardboard made of pinewood, except that "lawanit" is darker in color.

Maria Cristina Chemicals Inc. (MCCI), was formed in 1954 by Filipino entrepreneurs, who saw in the establishment of the Iligan hydropower complex in Mindanao an opportunity to establish a business that would take advantage of low power rates and low cost raw materials for making calcium carbide out of limestone charcoal. A Norwegian-made furnace was brought in with the equipment suppliers providing technical assistance during the plant's installation and start-up period.

The chemical properties of coconut shell have made it possible to produce calcium carbide with a very high gas yield in the Maria Cristina plant.

Ricardo Guevara, MCCI president, feels some amount of pride that from calcium carbide production, its company has developed a downstream product—acetylene black for dry-cell batteries—which it is proposing for the ASEAN industrial joint venture program.

Ramie Textiles, Inc. (RAMITEX) started exporting raw decorticated fiber to Japan in the early 50's and importing the finished fabrics under a processing arrangement. After some years, the group of businessman Vicente Araneta decided that foreign technology should be imported so that the ramie fibers could be completely processed locally into finished fabrics.

Ramitex investigated various known technologies for ramie processing: the waste silk system of Japan, the modified worsted system of the Swiss combined with the Fischer Degumming system and the German process (Emindingen) utilizing a high pressure degumming system. After thorough studies, the Fischer Degumming/modified worsted system was adopted.

The Fischer Degumming process involved, among other processes, bacteriological retting of the decorticated fiber under controlled conditions for a period of one week. Because the fiber suffered considerable strength loss from degumming, A. T. Ycasiano, Filipino technician trained in the Fischer Mill in Switzerland, developed in 1964 together with

P5

"THE PHILIPPINES: STRATEGIES FOR SCIENCE AND TECHNOLOGY IN THE EIGHTIES"

Emil Q. Javier, Chairman,

National Science Development Board.

The Philippines is moderately blessed with natural resources. Our farms, forests, marine and inland waters, minerals, geothermal and hydro resources and, hopefully, abundant petroleum give us reasons to be optimistic about our future.

But these natural resources merely represent potentials. They have to be translated into reality by the ultimate resource—people.

SECTORAL RESEARCH COUNCILS

In order to close the gap between the developing world and that of the economically developed countries, it is not absolutely necessary to have a basic research and development establishment as elaborate as those in the West. The experience of some of our Asian neighbors has proven this. Our science and technology (S&T) strategy, therefore, is to develop a minimum core of high level scientists, engineers and academicians so that we can borrow and adopt foreign technology intelligently and most advantageously. This minimum corps of high level expertise should enable us to determine which, from whom and how to acquire technology and to adapt and improve upon such technologies to suit our climate, our raw materials and own culture.

In the agricutural and forestry sciences where the government has lavished ample resources, particularly in the last fifteen years, we have now at the University of the Philippines at Los Banos the capability of turning out 300 masters and 70 doctorates in agriculture and forestry each year.

A new campus of the University of the Philippines is being constructed on the island of Panay to provide for the fisheries and marine resources sector. To aid the fields of natural sciences and mathematics we are implementing an eight-year plan to train 300 Ph.D.'s at the University of the Philippines at Diliman in consortium with Ateneo University, de La Salle University and the Agricultural Sciences Center at Los Banos.

Other plans are underway for stepping up graduate engineering training capability at the University of the Philippines.

All of these high level manpower training programs are directed to satisfy the needs of the universities themselves, the research and development (R&D) organizations under the National Science Development Board (NSDB) and other line ministries as well as the private sector.

Our strategy calls for problem-oriented, applied research in the NSDB system and R&D organizations under the line ministries. Support for the basic sciences will be provided largely to the universities in conjunction with graduate education.

LINKING RESEARCH WITH TECHNOLOGY APPLICATION

Our S&T system in the past has been faulted for failing to decisively link research with technology utilization and application. In the developed economies, this function is more than adequately performed by the private sector. At our stage of industrial development, a more active government hand in this function is deemed necessary.

Thus in addition to aligning our S&T structure with those of industry by way of sectoral S&T councils, we envision even bigger roles for the Technology Transfer Board, the Technology Research Center and the Philippine Inventors' Commission.

RESEARCH THRUSTS

Our research thrusts towards the year 2000 include sustaining our momentum in food and agriculture and moving on beyond primary production into processing and utilization. This will call for more irrigation, cropping intensification to as much as three to four crops per year, accelerated introduction of agricultural machinery, heavier use of fertilizers with increasing reliance on organic forms and biological nitrogen fixation, and increased emphasis on pollution abatement and environmental conservation for sustained productivity.

As our natural forest cover declines, industrial tree plantations are expected to provide more and more of the raw material requirements of our wood-based industries. Modern silvicultural techniques, forest harvesting, pulp and paper chemistry, particle board and other wood technologies will become very important research objectives.

Lower labor costs, availability of raw materials and increasing technological and financial sophistication will determine which of our industries will be competitive in the world markets.

However, certain technologies and industries stand as natural choices for vigorous development. One such broad area is in the chemical processing field where potential exists for the conversion of biomass into food, pharmaceuticals and fine chemicals. Our favorable climate and environment guarantee a bountiful and renewable flow of agricultural, forestry and fishery products. For example, at present, our efforts to produce sugar and coconut are limited only by the volume we could consume locally and market abroad profitably. As the industrial tree plantations and the dendrothermal plantations envisaged under the KKK program go on stream, they will add considerably to the national annual harvest of biomass.

The emerging field of biotechnology and genetic engineering promises new, exciting and efficient ways of making organisms convert biomass into food, pharmaceuticals, fine chemicals and other useful products.

In 1975, an institute of plant breeding was established; later a national crop protection center and an institute of biotechnology and applied microbiology were added to our agricultural sciences complex at Los Banos. A food technology and an animal health and production center are due to join this complex.

Now with relatively well-developed conventional research groups in these institutions, we are ready to move on to molecular biology and fermentation engineering to complete our capability in this area of emerging technology. In time we intend to organize three molecular biology groups, one each for agricultural, industrial and health applications.

Other opportunities such as the exploration of geothermal resources await development. To translate this and the other possibilities mentioned here into real productivity, the role of scientific and technological education will be paramount in the remaining decades before the twenty-first century.

other Filipino technicians, a new system of controlled aerobic retting which reduced the retting time from one week to 48 hours, resulting in a 50 percent increase of tensile strength of the degummed fiber.

In late 1959, initial experiments were conducted by Ramitex on ramie/polyester blends with the assistance of technicians from Imperial Chemicals in the U.K. In 1961–62, further development work and piloting was done at the mill site in collaboration with Teijin technicians. Soon afterwards, Ramitex began producing ramie/polyester blends commercially.

Applications of new processes to ramie are also being tried out, among them, the open-end spinning of combing noils and short staple ramie; cover-spun spinning to attain fine counts such as Nm 73 with chief weight ramie (83 percent ramie) and liquid ammonia treatment of ramie fabrics to attain better crease resistance of even pure ramie fabrics without weakening the fiber.

In order to be able to spin ramie cuts to cotton staple lengths, application of the cotton spinning system to stapled ramie fiber has been carried out by Ramitex researchers at the Philippine Textile Research Institute. Studies to develop a ramie variety which will yield more and finer fibers have been started by the San Miguel Foundation.

PRIVATE SECTOR
R&D EFFORTS

To all appearances, there is very little R&D undertaken by the country's private business sector; there are in fact no official statistics that would indicate the level, amount and value of the private sector's research activities to date. This perhaps can be traced to the tendency of private companies to treat their work in strict confidence mainly to safeguard their competitive positions.

However, the continuing appearance of new products, as well as improvements in the quality levels of existing products which find their way into the domestic and export markets, attest to the presence of active R&D programs in the more dynamic sectors of industry in the Philippines.

Paper Industries Corporation of the Philippines (PICOP) operates the world's first large-scale paper mill using 100 percent mixed tropical hardwoods, and is the only producer of newsprint in the Southeast Asian region.

Although paper-making originated in China, the technology for its commercial production was developed in the West using temperate wood species which are long-fibered. In collaboration with International Paper Company of the U.S., PICOP devoted seven years to experiment-
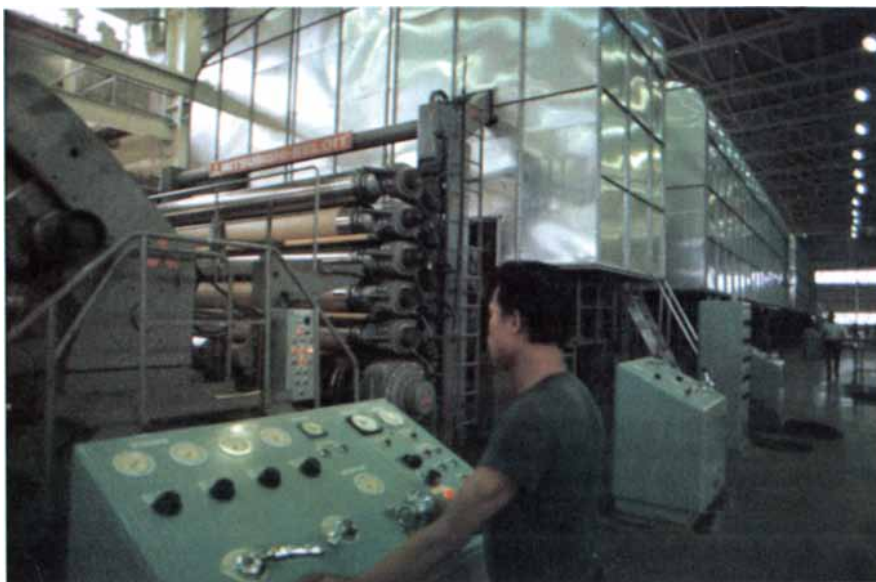
Paper Industries Corporation of the Philippines has pioneered the use of data processing for the control of raw materials in the manufacture of plywood.

CREDIT: PICOP

ing on the use of short fiber mixed tropical hardwoods.

Its efforts resulted in a technological breakthrough in 1971 with its successful first commercial manufacture of paper from short-fiber mixed tropical hardwoods. Today, the industry which started in China during the first century has returned to the region, utilizing hardwoods which are indigenous to Asia.

Continued research has improved the strength properties of PICOP's paper. The company has attained an opacity of 98 percent, ahead of the world standard of 90–95 percent.

Technological improvements have also made possible the operation of PICOP's plywood plant at 105 percent design capacity and a 41 percent increase in newsprint output.

Much of PICOP's pioneering research work is now in plantation forestry, involving establishment methods, stand maintenance and genetic improvement. To sustain the company's wood requirements from its forest concessions, PICOP makes use of a blend of two recognized silvicultural systems under the concept of sustained yield and ecosystem enhancement: selective logging and the industrial tree plantation or the clearcut and plant method. Forest research and protection also play an important role in guiding the management of its forest resources. Experiments are undertaken on different species to identify those which are adaptable to climate and soil conditions. For this purpose, the company maintains a forestry staff whose

experience and qualifications, according to World Wood, the international forest product magazine, "forest agencies in the developing world would find hard to match." PICOP has gained international recognition.

On the health front, United Laboratories, Inc., a Filipino firm and the largest drug and pharmaceutical manufacturing company in the country, undertakes an extensive R&D program to keep abreast of the dramatic changes taking place in the fields of medical science and health care.

Its analytical techniques to test product stability, potency and efficiency include the latest technologies such as high performance liquid chromatography, thermal analysis, computerized scanning spectroscopy and spectro densitometry. Technologies such as radioisotope tracers, fractional and differential centrifugation and all types of chromatographic separation techniques are employed.

Bioavailability testing on Filipino subjects is an important program being carried out by United's research group. Bioavailability testing has been done on many foreign drugs from outside the Philippines.

Consultations are being maintained with international authorities on pharmacokinetics, pharmacodynamics, pharmacology and related fields. The results are medicines which are particularly suitable for the Filipino and Asian patient.

United's research also directs itself to the task of utilizing indigenous technical

P7

know-how, expertise and materials in the synthesis and/or derivation of chemical entities with therapeutic potential. As early as 1958, United already had the capability to develop original formulations. Most of their work was in the area of vitamins and anti-TB preparations. Among the early original formulations were Medicol (an analgesic-antipyretic), Afebrin (a liquid pain-reliever) and Metussan (an antitussive), Trisovit is an anti-TB medication whose dosage was adopted to local needs and to which vitamin B complex was added as a nutritional supplement.

Chemfields Inc., the first semi-synthetic antibiotics plant in the country, started full scale commercial operation in July 1981. A 100 percent Filipino enterprise registered with the Board of Investments as a pioneer project, it currently produces Ampicillin trihydrate and Amoxycillin trihydrate, both semi-synthetic antibiotic raw materials. The plant has also the capability to produce semi-synthetic antibiotics like Caphalexin, Flucloxacillin, Cloxacillin and Dicloxacillin.

Philippine Appliance Corporation (PHILACOR) the local manufacturing licensee of Westinghouse products, was originally offered a total technology package which was adapted to the large production volumes in the U.S. But Dante Santos, Philacor Managing Director, had other ideas. He knew the limits of the Philippine market precluded large volume production, and he wanted to maximize the local content of his product, employing skilled labor instead of complete automation in production. The solution came in tooling and design adaptations, which affected only the physical features and finish of the product but not the performance of its basic functions. Today, Philacor products account for approximately 70 percent of the white line market in the country.

Meanwhile, Filipino inventors, working alone or in teams, continue to come up with viable inventions. The Philippine Inventors Commission reports that the last eight years reaped a particularly rich harvest of Filipino inventions—164 patents were awarded, 143 of them local and 34 foreign.

These inventions have served to improve the performance of crude time-consuming tasks. Thus, a coconut grater created and commercialized by Benjamin Almeda reduced coconut grating time from 15 laborious minutes to a number of seconds. Maggie Villaruz designed and developed a prototype power tiller that could stay afloat on deep muddy rice fields where heavier imported hand tillers sank. This is now in wide use.

A "miracle" cooker was Octaviano Ludeñas' answer to a foreign friend's observation about the wastage in cooking rice. The cooker allows simultaneous cooking of three food varieties in one setting, thus saving on time and fuel.

One recent invention has particularly attracted much attention not only in the Philippines but overseas as well. ZETON POWERFLO is the registered trade name of the fuel vaporizer assembly invented by Jose Ma. Concepcion who has already earned the Presidential Gold Medal award and the UN's Intellectual Property Organization award for 1980 for his invention. The device is also due for publication in the 1981 Rolex Award for Enterprise.

Zeton Powerflo employs a dual flow system where two forms of fuel are simultaneously fed into any form of conventional engine whereby the non conventional fuel enters the engine already in vapor form. It is different from the method which is now under experiment in Brazil where two forms of fuel are first pre-mixed in one fuel tank at a determinate ratio. The Zeton Powerflo's completely independent fuel tank and fuel line system eliminates the common problems associated with the pre-mixed method such as starting difficulties, stalling, corrosion, quick vaporization or pre-mature depreciation.

ENERGY

When in 1973 the OPEC countries turned off the flow of cheap oil, the Philippines was obtaining 95 percent of its commercial energy requirements from petroleum. The balance was supplied by hydroelectric power and a small amount of coal. The dependence picture was hardly any better in 1978. Out of the 86.5 million barrels of oil equivalent (MMBOE) consumed that year, oil contributed 81 MMB or 94 percent.

The country's positive response to the energy crisis was a comprehensive energy development program designed to hold down the economy's oil dependence. The Ten Year Energy Development Program 1979–1988 was drawn up and subsequently compressed to a Five Year Program addressed to providing adequate and secure energy supplies to support production targets in the Philippine Development Plan. A shift in usage to favor indigenous and renewable energy forms is a strong feature of the program.
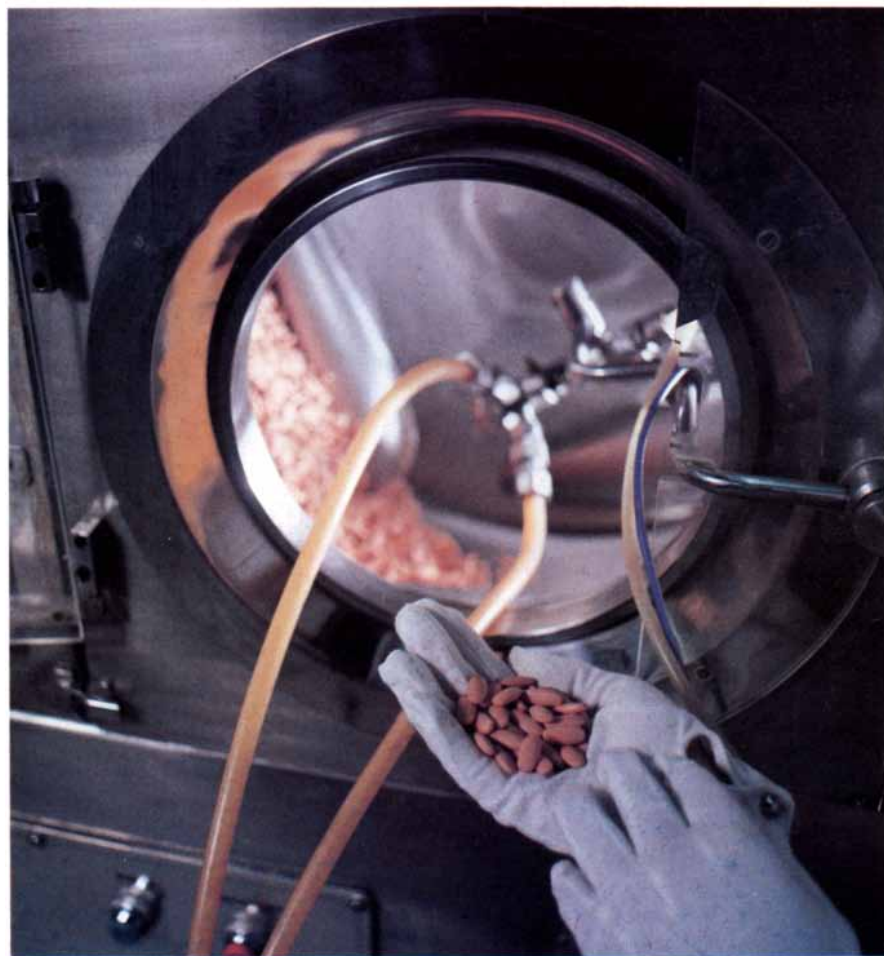
Automated coating of tablets at United Laboratories, the nation's foremost medical research and development company.

This year, based on Ministry of Energy studies, 48 million people will each consume the equivalent of 2 barrels. With an employment base of about 16 million to carry the consumption level at current oil prices, this translates to a US$204 burden per income earner in a country where the annual income is only little over US$700. By 1990, the consumption per capita will have grown by another 50 percent, or 3 barrels per head on a larger population base of 70 million. From the price end, the data is equally staggering. Up to 1973, oil imports remained within 13 percent of the total import bill. Eight years of successive price increases more than doubled the fraction to 30.4 percent in 1980. Thus in terms of 1980 export revenues, 39 cents of every export dollar has had to be set aside for oil imports.

To diversify energy sources and increase domestic supply, the development of indigenous resources such as coal, hydroelectric and geothermal power, as well as non conventional resources, e.g. biogas, marsh gas, biomass, wind and solar power—are being vigorously pursued. A planning estimate of the country's potential energy storehouse has been put together to serve as the basis for directing the energy search.

Coal availability in various parts of the country has been confirmed. Reserve estimates run to 4,000 MMBOE (1.4 billion metric tons), indicating the potential for replacing oil as a major source for industrial and power applications. Hydropower potential has been placed at 50 MMBOE yearly, of which only 6.6 MMBOE is harnessed in the form of electricity. Geothermal potential is estimated to be quite large, theoretically up to 2.3×10^3 million barrels per year, as surface manifestations indicate. This is possible since the Philippines lies somewhere along the Circum-Pacific firebelt where geothermal capabilities have been known to abound. A number of the non conventional sources are also being studied. With at least 2,000 hours of sunshine each year, solar energy could make a viable alternative source of energy.

OIL AND GAS

A strategy to encourage the entry of foreign venture capital has been adopted by the Government to overcome the shortage of both development funds and the technology to explore and produce gas and oil. The service contract scheme allows technically and financially capable foreign investors to participate in the exploration and development of oil and gas resources. The arrangement has attracted a total of 34 petroleum and 16 geophysical contractors to pursue various projects that have made possible the identification of 11 oil-bearing sites now

One of several plants which tap a relatively unexplored resource–geothermal energy

CREDIT: Ministry of Energy

being developed for commercial operation. In 1980 production from four wells of the Nido oilfield reached 3.6 million barrels, providing 5.2 percent of the country's total crude oil consumption and resulting in gross foreign exchange savings of some US$125 million. Last September, the Cadlao structure just 50 kilometers north of the Nido oilfield was inaugurated and its two production wells can be expected to add 9,000 barrels daily to the domestic volume. The current plan calls for the drilling of 148 exploratory wells, 66 percent of them offshore and 34 percent onshore. Possible commercial operation of two new oil-bearing structures could yield an additional 5 million barrels by 1986.

COAL

There were 44 coal service contractors involved in exploration development and production of coal as of year-end 1980. During the same period, 186 million tons of coal reserves had been proven through detailed exploration. This marked a 24 percent increase over the 1979 reserve figure. The bulk of reserves are located in Semirara, Cagayan Valley and Zamboanga del Sur.

Coal production in 1980 amounted to 328,000 metric tons, an increase of 10 percent over 1979's 300,000 metric tons. The coal region of Cebu accounted for about 61 percent of total production but trends indicate that larger volumes will be supplied by other areas in the future.

As supplies have improved, more industrial users have switched to coal for

their energy source. Among them are the National Power Corporation whose coal-fired power plant and cement companies are converting from bunker fuel to fired coal and are projected to be the largest consumers of coal.

From a 1981 demand level of 367,000 metric tons, coal requirements of the cement, mining and other energy intensive industries are estimated to reach 690,000 metric tons by 1986. By then, coal production will have increased from some 373,000 metric tons in 1981 to 374,000 metric tons.

With this projected shortfall in mind, the government has taken a direct hand in assuring adequate supply of coal to industries which are switching to coal for their energy source. PNOC Coal Corporation and the Malangas Coal Corporation are actively developing coal areas. The National Coal Authority is now concerned with installing a Coal Logistics System consisting of loading ports, terminals, blending plants and transport barges to facilitate the delivery of coal supplies to users.
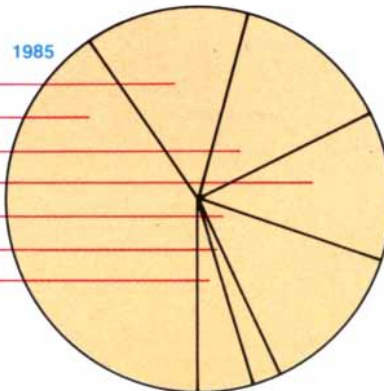
GEOTHERMAL

The first large scale harvest of geothermal power in the country was achieved in 1979 when the two pioneer geothermal fields in Tiwi, Albay and Makiling-Banahaw in Laguna were activated. Four 55 megawatt (MW) plants added 220 MW to the generating capacity of the Luzon grid, displacing 1.1 million of oil equivalent in 1979. Current geothermal production capacity of 446 MW with an

NATIONAL ENERGY SOURCE MIX
(IN MILLION BARRELS/MILLION BARRELS-OF-OIL EQUIVALENT, MMB/MMBOE)

1981	VOLUME	%
DOMESTIC OIL	7.30	7.5
IMPORTED OIL	75.36	77.0
COAL	2.32	2.4
GEOTHERMAL	5.34	5.4
HYDROELECTRIC	7.26	7.4
NONCONVENTIONAL	9.26	0.3
TOTAL	97.84	100 0
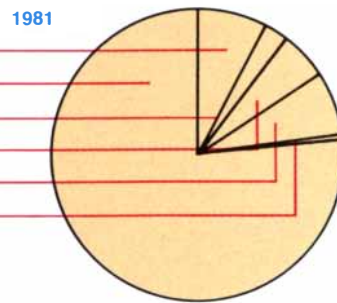
1985	VOLUME	%
DOMESTIC OIL	18.25	13.7
IMPORTED OIL	55.21	41.3
COAL	17.93	13.4
GOETHERMAL	16.34	12.2
HYDROELECTRIC	17.08	12.8
URANIUM	2.81	2.1
NONCONVENTIONAL	6.08	4.5
TOTAL	133.70	100.0

oil displacement value of four million barrels of oil equivalent makes the Philippines the second largest geothermal producer in the world, second only to the U.S.

There are 22 other pinpointed geothermal areas under development. Among them is the Tonggonan field in Leyte which is being developed by the PNOC Geothermal Technology Corporation, in a joint venture with Kingston Reynolds Thom and Allardice (KRTA) of New Zealand. The arrangement is seen by both Philippine President Ferdinand E. Marcos and New Zealand Prime Minister Robert Muldoon as a significant approach toward the more efficient transfer of technology. Tonggonan can log a power steam capacity of some 450 MW, enough to support the power-intensive industrial plants proposed for location in the area, including a copper smelting plant with a production capacity of 318,000 tons which is now under construction and expected to begin commercial operation by 1983.

URANIUM

With no commercial quantity reserves yet proven, uranium exploration was intensified in 1980 through efforts of private companies—the Benguet-Corporation-Getty Group, and Ultrana Minerals.

Benguet-Getty conducted what could be the most comprehensive exploration for uranium in the country to date, starting with the first systematic airborne radiometric and magnetometer survey within its 250,000 square kilometer area and culminating in detailed ground follow-up work. Ultrana on the other hand conducted ground surveys. Uranium would feed the US$1.9 billion 620 MW nuclear plant now being built to supply the Luzon power grid.

In other non conventional energy fields, work on solar heat, wind energy, biogas, alcogas and small hydroelectric resources are being undertaken.

Alcogas is now produced by Victorias Milling Company in Bacolod City from bagasse and other sugar cane wastes. Output is utilized by cars in Bacolod, the first alcogas program area of the Philippine National Oil Corporation. Alcogas will probably lead to the establishment of 70,000 hectares of new sugar cane areas and 25 distilleries to produce alcohol that is hoped to displace about 1.1 billion barrels of premium and regular gasoline consumption by 1988.

A coco-diesel program is planned to augment the supply of diesel oil, thus cutting down imports and providing another use for the versatile coconut. Potential contribution of the program to the energy supply mix is set at about 1.3

MMBOE by 1986 from some 500 thousand barrels in 1981.

SOLAR HEATING

Some 2,000 hours of sunshine is lavished on the country throughout the year but technology has yet to find the means to make its use cost efficient on a large scale. Some small applications however, have been successful.

At the San Miguel Magnolia Dairy Plant, the first industrial scale solar water heating system in the country is now installed, supplying 1,250 gallons of hot water for the plant's sugar hydration process. At the Magellan Hotel in Cebu City, a 134 square meter collector system provides hot water for all the guest rooms and the kitchen. A pilot solar water installation has been installed at the Vitarich plant which is engaged in producing poultry feeds.

Other small but important breakthroughs in the non conventional energy program have also been made and are especially significant in the light of the country's rural electrification drive which stresses self-reliance at the village level, among them:

—The use of agricultural wastes to generate producer gas and to supplement diesel oil in dual-fueled stationary engines successfully demonstrated in two field installations: a 30 kilowatt (kW) electrification system for a Laguna barrio, about 50 kilometers from Manila, providing electricity to 1,000 homes and a 45 HP diesel engine to power a 50-hectare irrigation system also in Laguna;

—The country's first pyrolitic converter which turns agricultural residues to high energy fuels developed and operated at the Cabanatuan National Grains Authority Grains Center in Cabanatuan City in cooperation with the Georgia Institute of Technology and the United Nations Industrial Development Organization. A low cost continuous flow rice-hull fired grain dryer was developed.

—An integrated energy system for small island settlements to answer the needs for potable water and energy supply now in operation on Verde Island off Batangas in Luzon. Constructed on a 693 square meter site, a concentrated pond and salt-making bed was installed to accommodate the salt-rich sea water which is flushed out of the stills. The salt bed produces approximately 2,000 kilograms of salt within a 15-day period. A multi-bladed windmill with a 1-HP capacity is utilized to pump sea water to the solar desalination site. A community management committee maintains and manages the facilities and it is now being monitored closely for possible

replication in other island villages.

—Taking off from basic NSDB-FORE-PRIDECOM studies in 1977, the concept of energy plantations has been pushed. The name dendrothermal was coined and is now becoming part of international technical usage. Physically, two major pilot projects are in place—a 500 kW plant in Mindoro and a 2.5 MW plant in Antique in the Visayas which is in the design phase.

—Biogas: Maya Farms, an integrated swine-raising and feed processing farm, draws 40 percent of its energy needs from a biogas plant which feeds on the waste inputs from a farm stock of 17,000 hogs.

AGRICULTURE AND FISHERIES

Agricultural production has been the brightest spot in the Philippine development scene during the past few years. Significant gains have been achieved, particularly in food crops.

In this year's report of the Food and Agricultural Organization (FAO), the Philippines was ranked as the world's 15th largest food producer. The report further noted that the country registered a high food production index especially during the period from 1969 to 1980.

The most notable achievement in the food production drive during the period was the nation's attainment of self-sufficiency in rice. A traditional rice importer, the Philippines joined the league of rice exporters in 1976. By 1980, rice shipments were being made to Indonesia, Brazil, Malaysia and several other countries to the value of US$76.4 million.

Coconut remains a top dollar earner despite its weakening position in the world market. Sugar, another traditional major export item, posted a 200 percent increase in value of earnings with 1.72 million tons of exports valued at US$629 million in 1980. Fishery production and exports also improved considerably during the period. Poultry and livestock, however, have remained insufficient to meet domestic demand and the country has had to import meat and meat preparations which, during the first seven months of 1981 amounted to US$7.5 million.

Progress in agricultural productivity had been, to a large extent, the result of previous substantial investments in agricultural education and research. Assistance by Cornell University, the Ford and Rockefeller Foundations, the USAID and its predecessors and most recently the World Bank transformed the small college of agriculture at Los Banos into a full agricultural university

with very strong undergraduate and graduate programs, and growing research capability. Much of the manpower and technology in agriculture radiated from this academic institution.

Outstanding varieties of rice, corn, tomatoes, mungbeans, peanuts, cowpeas, cassava and sweet potato developed by the University of the Philippines at Los Banos (UPLB) paved the way for remarkable increases in yields of these crops. Philippine-bred downy mildew resistant corn varieties, high yielding mungbeans and heat tolerant tomatoes are now grown extensively in the southeast Asian region.

The pace of agricultural research stepped up considerably with the establishment of the International Rice Research Institute (IRRI) in the UPLB campus. With the development of IR8 popularly known as miracle rice, by IRRI in the mid-sixties, a quantum leap in tropical rice productivity was achieved. This variety and the ones which were released later spread rapidly in the countryside and together with the government's supervised credit program called "Masagana 99" brought the green revolution to the Philippines.

Little known is the fact that even earlier and simultaneously with the efforts of IRRI, The Bureau of Plant Industry and UPLB were developing outstanding, very high eating quality rices like BPI 76 and C4:63 which together with IR8 brought about this miracle in farm productivity.

Private enterprise has contributed its share in food production. A total of 99 projects have been registered with the Board of Investments for such areas as fruit growing and grain processing, cassava and corn starch production, palm oil growing and processing of tropical fruits, meat and tuna. Most of these companies take advantage of technical assistance from foreign partners in improving their production technology. For instance, the National Development Corporation is getting valuable know-how from Guthrie Malaysia in developing their joint 8,000-hectare palm oil plantation and 50 MT/day capacity oil mill.

The food sector has continued to develop and improve the quality of new processed food products. Food items like dried mangoes, mango nectar and frozen mango puree have been developed locally and have found their way into local and overseas markets. Carnation Philippines now processes avocado oil which is used as a perfume base by its foreign buyer. Another local company, San Miguel Corporation, produces papain, an enzyme derived from the latex of unripe papaya. Papain is used primarily as a stabilizer in beer brewing and as a meat tenderizer. The refined Philippine papain has a potency level five

times that of other refined papain.

A number of local food processing firms take advantage of technology provided by international food processors such as Swift's, Hormel, Gerber, Libby's and others with whom local firms have licensing agreements.

Know-how has filtered down to the smaller firms. It is known, for example, that when the big companies went into poultry processing, standard sizing of chickens developed even among the smaller producers, a number of whom became suppliers of the bigger companies. The smaller firms who remained in the market had to upgrade their standards and quality.

The food processing sector has always looked to an urban market for its production except for canned and powdered milk. There are indications, however, that they are now also aware of a broader market to be found in the rural population—the target of the government's nutrition and food self-sufficiency campaign.

Research in yellow corn has been strengthened by the endeavors of private entrepreneurs such as San Miguel Corporation, the country's largest food and beverage group. San Miguel maintains a 114-hectare experimental center near Manila, where foreign and local varieties of corn are subjected to different agroclimatic environments. High-yielding varieties are crossed to identify heterotic relations and to evaluate their resistance to downy mildew infection and corn borer infestation.

This year, San Miguel is assessing five years of its R&D efforts in corn development. "Results are very encouraging," says an official, "and there are plans to release hybrids this year on a limited scale." The target for commercial operation is set for 1982.

Pioneer Hi-brid Philippines, a subsidiary of Pioneer Hi-brid International, the market leader in the U.S. corn market is also engaged in hybrid corn seed development. Through research stations in General Santos in Davao, Pioneer has developed the high-yield corn seed 6181 which is resistant to downy mildew.

Other agricultural crops which have been the subject of research and experimentation include mungbeans, sweet and white potatoes and cassava varieties. Research endeavors in minor crops with economic potentials led to the breeding of crops like sunflower and castor oil which have found their way into the export market.

Cotton and sericulture have been the subject of extensive R&D. Cotton was introduced by the Spaniards a century ago but with the entry of cheap imports, the local cotton industry died out. The high price of cotton in the world market renewed interest in cotton production.

The NSDB commissioned Central Luzon State University and the Philippine Textile Research Institute to establish a cotton R&D program. Since then the responsibility of commercial promotion and production of cotton has been assumed by the Philippine Cotton Corporation, a government-industry partnership. From nil, cotton growing has expanded to 17,000 hectares in a period of five years. Likewise, the original not-so-well adapted Southeast U.S. varieties are now being replaced by higher yielding and better adapted Philippine-bred cotton.

In fisheries, attention is devoted to technologies which will allow small scale fishermen to farm oysters and mussels. Techniques of pond fertilization to induce "lab-lab" bottom organisms and plankton as food for milkfish have improved yields.

In working with prawns (Panaeus menodon), researchers have induced maturation of ovaries by applying an eyestalk ablation technique. Experimental cage farming of prawns in fresh water areas have enabled them to attain marketable sizes in five months. Similar innovative fish farming technologies are being developed at the Southeast Asian Fishery Development Center, Aquaculture Department in Iloilo.

San Miguel Corporation maintains a nursery and hatchery farm culture for the production of prawn fry (black tiger species). A package of technology is being developed to increase survival rates and ensure the proper growth of the shrimps.

To improve productive use of limited land areas, a trend is being developed for integrating fish culture with livestock or crop farming. A rice-fish culture technology has been developed which has proven effective. Special varieties of fish, raised in rice paddies have produced increased yields. This effort is now the subject of nationwide field tests.

Research in marine fisheries, according to PCARR, will involve assessment of major species of tuna and other migratory fish and invertebrae such as crustaceans, mollusks and echinoderms which have commercial value. Oceanographic studies are also in progress as well as research on fishing gear technology to develop more efficient methods and vessels. An unusual area is the development of "skyponds." In the country's mountainous areas, terrace farming is practiced at altitudes of thousands of feet. Especially in rice growing areas, rain water and the use of natural spring water for short term fish culture can bring fresh fish, a source of protein to the inhabitants of the hinterlands. This practice could eventually lead to commercial undertakings.

Processing and handling studies are also onstream in pilot areas. Simple village type equipment made out of locally available materials have been fabricated; e.g. solar driers and salinometers for measuring brine concentation in smoking and drying fish. A mechanical press for extracting excess moisture and fats from fish and for producing protein concentrates has also been fabricated.

The National Institute of Science and Technology another NSDB agency has developed expertise in fish, vegetable and fruit processing and preservation. Some processed foods for which it has matured technologies ready for commercial spin-off are canned coconut whole milk, vegetable protein and drink from coconut skim milk, vinegar and nata from coco water, coconut chips, low-fat jams and honey, coconut syrup, vegetable loaf, coconut flour and breakfast foods out of coconut.

MACHINERY AND METAL ENGINEERING

A well-developed machinery and metal engineering sector is considered an important component in the country's overall industrial thrust. Thus, there has been a determined effort on the part of the government to support this sector.

With the provision of incentives and assistance, new machinery and fabricated items such as power tillers, steam boilers, air compressors, grain dryers and grinding balls, among others, have been manufactured for the first time in the country. Altogether, 42 projects in machinery and metal fabrication had been registered with the Board of Investments by the end of 1980. Most of these were granted pioneer status in that the products were manufactured commercially for the first time in the country.

The local content manufacturing programs initiated by the Philippine Board of Investments have also been a major factor in inducing increased activity in the metal fabrication and engineering sector. Among the major ones are: the progressive car manufacturing program (PCMP); the progressive truck manufacturing program (PTMP); the progressive motorcycle manufacturing program (PMMP); and the progressive automotive diesel engine manufacturing program. By requiring a yearly increase in the percentage of local content of the finished product, it was expected that wider and faster technology transfers from assembler to suppliers would be effected. The program also would encourage domestic manufacture of components, increase activities in the small and medium size firms and, in the process, upgrade the engineering skills and provide technical know-how to the domestic manufacturing industries.

The result is a considerable expansion of the component parts industry. From about 50 suppliers and manufacturers in 1972, it reached a level of some 280 after five years. Through technical assistance from participants of the different content programs, jobbers and suppliers were able to produce parts applying various levels of technology for engine transmissions, engine assemblies, machining engine parts and brake systems.

A recent demonstration of the capabilities of the local metal engineering and

A fishery project sponsored by the Development Bank of the Philippines.

fabrication sector is the awarding by Polysius Ltd. of England to Pacific Engineering Co., Inc., a local company, of the contract to fabricate and install the local components and all foreign-supplied components of the coal-conversion projects of Floro Cement and Iligan Cement. The components to be fabricated locally would include the grinding mill housing, dosing vessels, kiln burner and mill building structures. Pacific Engineering Company is a metal processing firm specializing in high-technology fabrication of heat transfer components, pressure vessels, materials handling and bulk storage systems, food processing equipment, piping systems and stainless steel products.

The Metals Industry Research and Development Center (MIRDC) has been instrumental in developing the engineering sector. It assists private companies in the theoretical and practical applications of the latest technological advances in foundry, machining, forging and castings. According to an MIRDC report, the low productivity and generally inferior product quality of the local metals and engineering sector have been traced to two major deficiencies—lack of capital and low technological proficiency. The MIRDC, working with industry, is now trying to correct this.

Another agency which has contributed to the development of intermediate technology for low cost agricultural machinery is the International Rice Research Institute. Through its technical assistance and guidance, a number of agricultural machines have been brought within the reach of poor farmers. For instance, a manual rice seedling transplanter costing about US$170 and a root-zone liquid chemical injector costing US$145 have contributed to a reduction of 40 percent and 65 percent respectively, in the amount of fertilizer and insecticides used, without any decrease in crop yields. "We've done well to some extent but we certainly can do better," says Ceferino Follosco, chairman of the ASEAN Agricultural Machinery Club and president of Alpha Machinery Engineering Corporation.

Follosco believes that the nature of incentives has tended to favor end products and retarded the sector's growth. He feels that Philippine industrialization gains in the future will depend very much on the engineering sector. "We should set up the basic technology infrastructure by setting up foundry, machining, forging and tooling facilities," he recommends.

THE PHILIPPINES IN THE YEAR 2000

According to the Long Term Philippine Development Plan, the last quarter of the century will witness a dramatic rise in Gross National Product (GNP). Real GNP is expected to double within the next decade and to exhibit an almost seven-fold increase by the close of the twentieth century. GNP will be growing at an annual average rate of approximately 8 percent to reach close to US$63 billion by the year 2000. Real per capita income, which up until 1975 had averaged 2.7 percent annual growth, should increase by an annual average of 5 percent. That means a per capita income of more than three-fold the 1976 level for the 70–80 million Filipinos living by then.

The Plan predicts a dramatic structural transformation within the economy. The share of agriculture, forestry and fishery in net domestic product would drop 14.2 percent during the period, from 30.9 in 1976 to 16.7 in the year 2000. However, agricultural production would still expand in value by 5.6 percent, supplying raw materials, generating demand for industrial goods and remaining a major source of jobs. Industry's share on the other hand, would increase from 28.3 percent in 1976 to 52.1 percent by 2000, reflecting an annual compounded growth of 11.1 percent, with growth coming from every industry subsector. The share of services will grow in absolute value but go down in percentage terms from 40.8 percent in 1976 to 31.1 percent in 2000.

NATURAL RESOURCES

Pressures from an expandng population and a desire for improved income levels make the efficient management of natural resources for maximum yield an imperative as never before. The strategy calls for an optimum balance "between exploitation and replenishment, between protection and utilization." Technologically, increasingly more sophisticated methods will be applied for surveying and monitoring resource activities.

According to Dr. Celso Roque, a physicist and head of the Land Resources Management Center of the Ministry of Natural Resources, the Philippines has progressed from aerial photography to the use of the American "earth resources satellite." He explains that through computer analysis of images taken during overflights every nine days, good progress has been made with complementary input from conventional methods.

Measuring mineral potential is something else. Geological mapping has been done for only six percent of the total area because of prohibitive costs. Potentially rich mineral areas like the Sierra Madre Mountain and much of the Mindanao remain unexplored. Satellite images would be useful, says Roque. They would allow detection of geological formations and surface plants with soil preferences that indicate the presence of minerals.

The Plan relies on modern technology to accelerate exploration and development of mineral resources. Remote sensing, geophysical surveys, marine, seine and sea-bottom sampling and geochemical surveys should lead to discoveries of new deposits and to a better understanding of ore formation processes. Resources not traditionally tapped such as zinc, sulfur and manganese will be evaluated and mined to generate employment. In the capital intensive areas like copper, computerization will be used to improve extraction methods and enable efficient response to local and foreign demand.

The Plan views the natural resources scenario in 2000 with optimism, dwin-

Rice husk briquettes, one of the appropriate technologies developed by the Technology Resource Center, are an economical and efficient substitute for firewood or charcoal.

P13

Whereas large, expensive technological transfers and development projects figure prominently in the industrial future of the Philippines, smaller scale adaptation of simpler, cheaper tools and machines can play an important role in improving the productivity levels and livelihood of much of the population. In the following article, Arthur Alvendia, Director-General of the Technology Resource Center, elaborates on this view.

TECHNOLOGY FOR THE SMALL ENTERPRISE: A PHILIPPINE APPROACH

Arthur Alvendia

The most basic and permeating changes of tomorrow will be generated by the most fundamental issues of today—the survival issues.

The industrial-technological landscapes of all countries, developed or developing, must undergo dramatic transformation if their people are to survive economically.

In the Philippines today we see the emergence of new trajectories of change that may allow us to leap-frog ahead in industrial development, find new routes to progress, and avoid the worst mistakes of industrialization.

The important areas of agri-business, energy and regional development are discussed elsewhere in this special section, but I would like to make particular mention of the productivity problem of the small producer, a central survival issue because the Philippines is a nation of small producers.

Small capital means backward production facilities or technologies that are incapable of volume production and standardized quality—and therefore of growth. The entrepreneur lacks the credit and the technical and managerial know-how to move from a backyard operation. Essential capital intensive production facilities cannot be financially justified because of limited production volume. Because he is small, he stays small.

A World Bank study of Philippine industry in 1977 showed that about 93 percent of manufacturing establishments were enterprises with fewer than 100 employees. In terms of production share, however, establishments with 100 or more employees accounted for about 50 percent of value added in the manufacturing sector. This is thus both an economic and a social problem. Moreover, the limits of growth in a small domestic market require that the Philippines' enterprises gear up for international competition.

An industry modernization program is now being implemented by the Philippine government and aimed at upgrading production efficiencies. Meanwhile, new possibilities for the small producer are seen in the trend towards decentralization of large integrated production units into smaller specialized units. This pattern, which is creating small component producers to supply for assembly or finishing for the larger end-product firm, was once experienced in Japan, South Korea and Taiwan, and is evident now in the Philippines' garments and electronics sectors. Rising costs of capital and labor, not to mention labor relations problems, have improved the economics of decentralization.

Because the move to integrate Philippine industry with the international economy will lead to reduced tariff barriers, local component producers will have to improve production technologies to compete with foreign suppliers. Modernization of these and other small and medium-scale enterprises will require more cooperative action within industry groups, a need that is leading in the Philippines to the development of specialized industrial estates where related or complementary enterprises can be grouped together.

This will make it feasible for small entrepreneurs to set up large capacity, capital-intensive common service facilities, develop specialization and complementation in product lines. The first such estate is being established in Cavite Province near Manila for the woodworking industry by the Technology Resource Center and the industry associations for wood manufacturers.

The traditional industrialization strategy which holds that the big, basic industries lead growth is being re-examined by Philippine economic planners. The big enterprises are viable only if their supportive base of small enterprises has attained the necessary efficiency levels. During the next two decades in the Philippines, the trend may well be industrialization led by the small producer.

Smallness is a problem also shared by millions of Filipinos in coastal communities whose survival is imperiled by the deteriorating economics of fishing. Returns have declined as fish resources in nearby coastal waters are depleted by over-catching, dynamite fishing and pollution. To go further out to sea increases capital requirements—larger boats, heavier gear and fuel. Filipino fishing, once largely an individual affair with reasonable returns on small investment in an out-rigger and rudimentary gear must now turn to two possible solutions—organized cooperatives now being formed by the government and micro-scale fish farming.

Although fish farming is not new in the Philippines, it has been a capital intensive technology. Now marine research institutions have developed a micro-scale fish farming technology, utilizing a fish cage farming option well within the investment reach of many small fishermen. A cage is constructed from fish nets and anchored in open water tied to bamboo poles. Because cages are submerged, the risk of typhoon damage is reduced and production—hatching and growing frys—has trebled to 100 to 150 tons per hectare per annum.

Another technology breakthrough is the "payao," an anchored barge from which coconut trees and debris hang underwater, eventually attracting coral life and plankton. Large schools of fish, particularly tuna, take shelter in such artificial havens, improving fishing returns. Other technologies with quantum potentials are the spawning of prawns, crabs and milkfish, as well as seaweed farming. Research of the nation's 7,000 islands could give the Philippines a leading role in the "blue revolution."

Many of the necessary technologies appropriate to the Philippines require relatively simple modern machines or tools that can be imported and used straightaway with minor adaptations. Some need to be transferred through licensing agreements under governmental guidance. Still others must be created. But these attempts should be focused in such appropriate areas as agro-industry, microbiology and aqua-culture.

Emphasis is placed on the utilization and propagation of the existing stock of technology to supplement newly created technology. There is a need for new types of technology institution where orientation is technology utilization rather than creation.

The Technology Resource Center has been established as an institution that is primarily oriented to the problems of the technology user and the difficulties of applying technology. The commercial application of any technology requires the coordination of such resources as financing, technical services, information and management support, a difficult undertaking for a small entrepreneur. The TRC has filled this gap by functioning as a centralized body to organize and provide the user with the necessary resources.

The effort to improve and develop the economy of the Philippines cannot be based alone on large-scale implantations of expensive technology. Appropriate technologies which directly serve the needs of the small producer must be introduced and encouraged as a fundamental solution to the problems of low productivity and under employment.

dling forests and other obvious ecological problems notwithstanding. As the frontiers of expansion are reached, an overall strategy of "judicious management" will replace "intensive exploitation." The objectives of resource management are to recapture the resource bases of overexploited areas, utilize modern technology and maintain the ecological balance.

The Plan projects that by 2000, forest lands will make up 46 percent of the country's total land area, though variations in the ideal 40:60 forest to non-forest land balance ratio may occur from region to region. Timber production in the 1990's will be mainly from residual forests but proper silvicultural management techniques will have to be enforced, otherwise yields would decrease with each cutting cycle.

By 2000, reforestation of eroded and critical watersheds will have been completed and a forest road network to facilitate management and communication among forest communities will be in place. Occupancy of forest areas will be controlled and shifting cultivation (the "kaingin" practice) will have been eliminated to protect the forest base. To reinforce the reforestation targets, Executive Order 725 mandating the establishment of industrial tree plantations within forest concessions has recently been issued. It provides that 30 percent of the current 8 million hectares forest concession area or 2.4 million hectares should be developed into industrial plantations. Assuming a yield of 200 cubic meters per hectare, this would mean an additional 480 million cubic meters in timber resources in about 10–15 years from now, before the year 2000.

"This has far reaching technological and marketing implications," says Emilio Jaranilla, assistant director at the Board of Investments and a forest products specialist.

The Plan estimates that fisheries resources will expand tremendously by 2000 with the adoption of the 200 mile limit. Development efforts will stress intensification of yield rather than expansion of hectarage. The target: 1.5 metric tons per hectare by the year 2000. Fish culture from shallow sea ponds will augment inland production. Commercial fishing will expand as more efficient vessels gear for operations in deeper waters and traverse non-traditional fishing grounds.

Demonstration stations and hatcheries will be set up throughout the country to popularize marine culture, and sea farming will be promoted by the National Sea Farming Program. Breeding and culture techniques for "bangus," a quality fish, and prawns will be perfected by 2000, the Plan projects. It will contribute to exports as well as to domestic food supply. But the ecological system will be protected and this will to some degree limit the extent of fisheries development in the country.

REGIONAL DEVELOPMENT

Spreading the benefits of development equitably in a nation of 7,000 islands is a task of some magnitude. The cost of developing basic infrastructure—roads, inter island ports, airports and communications to link the islands—is enormous.

Sixty percent of industry is concentrated in the Manila area. Several factors account for this. Being the seat of government and center of trade and commerce, Manila is the logical site for new industries which require access to financing, ports and government services, according to Ernesto Pernia of the University of the Philippines School of Economics. The franchise system which allows the independent operation of electric power companies in different regions is also a contributory factor, adds Washington Sycip of Sycip, Gorres & Velayo, the country's largest accounting and auditing firm. Until the oil crisis of 1973, Manila with its concentrated distribution system, had among the lowest power rates in the world, he explains. The other regions where distribution was less efficient charged higher rates. There was therefore a tendency for new manufacturing ventures to stay in Manila.

With the increase in fuel costs, it is significant that new power-intensive industries have located in the Northern Mindanao and Eastern Visayas areas where relatively cheap power from geo- and hydrothermal sources is available.

The regional development plans address themselves to correcting regional disparities which stem from uneven dis-

The Popular Technology Program disseminates information, such as the booklets pictured above, on the availability and benefits of simple, income-generating technologies.

P15

tribution of resources among the regions and the concentration of public and private investments in only a few regions. This calls for employment generation, greater access to public services and a more balanced allocation of resources. A major challenge is to divert an increasing urban population, estimated to reach 32 million or about 46 percent of total population by the year 2000, away from the major growth centers towards a number of secondary cities. This will require accelerated public sector investments in infrastructure, notably power, transportation, irrigation, schools and housing in the different regions.

When plans to shift resources and investments to less developed regions are fully implemented, an annual average growth rate in output among the 3 major islands of the Philippines may be expected: 9.4 percent for Mindanao, 7.6 percent for Visayas and 7.7 percent for Luzon during the period 1975–2000. At the same time, GDP will grow at an average of 8 percent. This means that by then, Luzon's share in output would be 59.2 percent instead of the present 64 percent. The share of the Visayas region would also decrease from 20 percent to 18 percent by year 2000. Mindanao's share on the other hand, would increase from the present 17 percent to 23 percent.

However, Manila will continue to be among the fastest growing regions. Southern Tagalog will benefit from the southward sprawl of industries from Manila, says the Center for Research and Communication, while the geographically distant Ilocos, Cagayan Valley and Bicol face, in the short run, unimpressive growth prospects. Outside Luzon, Central Visayas is expected to grow the fastest mainly because of increasing industrial activities in Cebu, the second largest urban center in the country.

Mahar Mangahas, of the Development Academy of the Philippines, points out that attention should be on reducing economic disparities between income groups within the regions, rather than between geographical areas. He bases his remarks on the Gini concentration ratio showing that "within region" inequalities account for .46 and "between region" inequalities only .03 of the overall national income inequality picture.

The recently initiated KKK or Livelihood Program is the umbrella for the total national effort aimed at reaching out to the countryside and initiating economic development at the grassroots level. Some US$380 million in fiscal year 1982 will be available at the local level for employment-generating small and medium scale enterprises.

To support these local industries, the Philippine Inventors Commission under the NSDB has recommended for packaging and development a number of Filipino inventions making use of available local materials and expertise. Among these inventions are mini-ice plants and green charcoal.

The mini-ice plant helps solve the problem of fish spoilage for fishermen in isolated areas in the country. The invention does not use brine for ice-making and could be carried along in any fishing boat. This product is an invention of Luis Camus and Jose Enriquez.

Green charcoal can be made anywhere in the country utilizing grasses, leaves, twigs, giant kelp and other seaweeds, cellulosic garbage and agricultural wastes. It is easy to ignite and burns slowly with a steady low flame. Its inventor, Gonzalo Catan, explains that the product is ideal for household use and for fueling industrial boilers and furnaces. It can also be used for manufacturing alcohol and charcoal for commercial purposes and activated charcoal for purifying air and water.

EDUCATION AND MANPOWER

A premium is placed on education as a way to increase productivity in this country which considers people the most valuable of its national resources.

Access to elementary schooling is now practically universal—free public schools reach out even to the barrios. Thus in 1979, 7.7 million Filipino schoolchildren or 85 percent of those of school age, were enrolled.

By 2000, the Plan sees that the goal of a "complete, adequate and integrated system relevant to national goals" shall be very nearly achieved.

By then, every Filipino 15 years and over, will have completed at least an elementary or equivalent education.

A scientific manpower training plan, initiated by 1982, will have started to produce a broad base of scientists and technologists for the country's expanding economic ventures, adds Dr. Emil Javier, Chairman of the NSDB.

HEALTH, NUTRITION AND FAMILY PLANNING

Philippine health planners are convinced that there is a correlation between health and development. In the words of Jesus Azurin, Minister of Health, "health promotes development and development promotes health."

On this premise, Azurin has initiated a plan for delivering health services to all Filipinos based on "practical, scientifically sound and socially acceptable methods and technology." Azurin explains that the nation's health problems are varied and enormous, and that it is not possible to attend to everything all at once. The current health thrust then limits itself to four major programs which are expected to create an impact on the broad health front:

—a primary health care system involving a multi-sectoral agency approach in which communities, families and individuals participate fully;

—an oral rehydration program designed to prevent deaths from diarrheal diseases, the second most frequent cause of death among Filipino children below 5 years. It consists of providing each household in all barrios throughout the country with a simple inexpensive packet made up of salt, potassium and glucose that effectively make up for rapid body losses of these nutrients;

—the standardization and upgrading of hospitals in the public system;

—the review of selected programs, i.e. malaria/schistosomiasis control, to make them more effective and far reaching.

On the research side, one project is particularly interesting for its potential insofar as making inexpensive medicine available to rural areas is concerned. This is the integrated research project to investigate the curative qualities of common Philippine plants which village "herbolaryos," untrained herbalists, administer to the sick. "There are thousands of such plants being used in rural areas," says Dr. Conrado Dayrit of United Laboratories who cooperate on the project.

For the project's first stage, the College of Medicine of the University of the Philippines would undertake trials on actual patients using the "herbolaryos" ways. Based on these trials, they will recommend what herbs or plants may be propagated and used for medication because they have been "proven."

The second stage would involve the development of crude tablets out of the plants which have been "proven." This means extracting the active ingredients and undertaking chemical and pharmaceutical analysis. "But that would require a lot of research, time and money," says Dayrit. The Philippines has the capability to determine structural form in chemistry but at a later stage, it would have to "farm out" the more difficult and expensive part of the work outside the country, he adds.

TRANSPORT AND COMMUNICATIONS

An efficient transport and communications system is the key to improving the productivity and self-reliance of the various regions.

This strategy calls for the concentration of public investments in areas with intensive production activities, as well as in the planned growth centers. Thus, in transport development, priority will be given to Manila, Cebu and Davao, and to cities with populations of at least half a million.

By 2000, highways will still be the dominant mode for passenger movements in urban areas. To optimize the utilization of infrastructure capacities, high-occupancy transport systems will be preferred. This year, double-decker buses were introduced by the Metro Manila Transit routes within Manila. The first light rail transit line (LRT), the construction of which started this year, will move daily an estimated 28,000 city commuters travelling on short and long-hauls by 1984.

These measures are expected to relieve traffic congestion in urban Manila and to minimize energy consumption as well. Highway transport will be complemented by commuter train services which will be expanded based on the industrial dispersal scheme and projected population growth.

In terms of local transport networks in the rural areas, the Plan estimates that by the year 2000, the density of feeder and farm-to-market roads should more than double at 1.21 kilometers (km) per square km of arable land, from only 0.42 km in the late seventies. This will improve mobility within and between rural communities and urban centers in the transport of goods. Other significant results would be greater accessibility to educational opportunities and health services.

The existing ratio of 3.6 km of road per 1,000 population will have been improved to 9.0 km by then. All told, this involves construction of some 210,000 km of roads traversing areas identified for priority production of food, commercial crops and livestock during the period.

The inter-island/inter-urban network will be served mainly by highways connecting the growth centers. In Luzon, the backbone network will consist principally of the Manila North and South Roads and the Cagayan Valley Road. The flow of people and goods in Mindanao will be served by a highway network connecting growth centers. In the Visayas, the principal coastal and cross country roads linking towns and cities will also serve as access points to port hinterlands. Aside from conventional shipping, the islands will be served by an extensive network of roll-on/roll-off and ferry facilities. At least one principal seaport will be developed in each region or major island to coincide with commodity flow patterns from production location to consumption centers.

The versatile coconut continues to earn considerable revenues, both as a food commodity, and more recently as an important raw material for new applications.

The overseas transport system particularly via seaports, provides a major supportive role to overall export and import activities. By the year 2000, the Port of Manila will continue to be the major transhipment center but goods will also be moved through Cebu, Cagayan de Oro and Davao. Secondary seaports of entry will be strategically located within the regions.

Passenger movements will similarly be improved through the development of regional international airports. The Manila International Airport (MIA) inaugurated in 1981, is one of the most modern terminals in the Asian region. It will accommodate 10 747-jumbo jets, 10 DC-8's, 10 DC-10's and 2 Concordes at any given time.

To bring the transport system from its current state to the year 2000 level, an estimated outlay of $326 million yearly or a total $7,180 million will be required. External finance is expected to meet part of these requirements while domestic sources will account for the balance.

There are now 0.88 telephones per 100 Filipinos; by 1990 this should improve to 3.25 per 100, and 6.97 per 100 by year 2000. The main backbone for communication will also carry other media such as facsimile, radio, TV, telex and data communications.

In setting up the communication facilities, priority is being given to Ilocos and Cagayan Valley regions. The Western Visayas and Central Visayas regions will follow.

At present, the privately-owned Philippine Long Distance Telephone Company (PLDT) controls the major tele-

phone system. There are 10 other companies operating fixed telegraph stations and 12 ship-to-shore lines spread over 4,000 stations and serving over 2,000 points. Eight companies provide domestic telex services; the same firms also offer record services. Facilities for international communications are available to all parts of the world through communications satellites, submarine cables or tropospheric scatter systems. The Philippine Communications Satellite Corporation is a member of the INTELSAT Consortium which has the ground station linked to both the Pacific and the Indian Ocean Communications Satellites.

Jose Lavares, Jr. deputy minister of the Ministry of Transport and Communications explains that there is a low-level of technological capability in telecommunications existing in the country and that to date, multinational companies are a major source of technology. For this reason, the entry of international communications companies has been encouraged. General Telephone (GTE) has joined Filipino partners in the manufacture of telephone sets and parts. Siemens of West Germany is very much involved in the current telephone expansion program of the PLDT.

HOUSING

The annual housing construction target in urban areas is set at 185,000 units to take care of new requirements as well as replacements for substandard units. Apart from urban housing needs, another 200,000 units need to be built

P17

An interior view of General Motors Pilipinas' Pacos Assembly Plant, recently rated second in quality control against many overseas plants, some of which are located in highly industrialized countries.

to cope with rural area needs.

In the past, housing was interpreted to mean just a roof over one's head. Housing projects were initiated anywhere, with careless architecture, with little regard for scientific planning, and with no consideration of other basic needs such as transport, water, food and jobs.

The new housing concept that has emerged takes a total approach, proceeding from the basic cultural premise that housing should not dislocate people or disorient them, that instead it should be the center of a total environment and the focus of economic activity. This implies that housing is not just shelter per se but that with it goes the other basic needs of food, clothing, utilities, employment, mobility, medical care and education, and recreation. Briefly, funds are raised from selling housing bonds with lottery features (Biglang-Bahay) which are then invested in constructing low and medium-rise structures to form the physical core of a housing community. Tenants are awarded units on liberal amortization terms. These communities feature self-administering livelihood and production centers.

The Dagat-Dagatan project in Manila's Tondo foreshore area has been given worldwide exposure by the World Bank as a model in urban squatter upliftment. The area, originally planned to form part of a modern port complex but overtaken by a big squatter population, was converted into a housing community. Similar projects like Kapitbahayan and Lungsod Silangan within Manila illustrate the humanistic approach in the building of housing communities.

The BLISS project engages in research aimed at allaying high materials cost through the development of low-cost construction components.

NSDB-FORPRIDECOM research efforts in this field have concentrated on finding cheaper substitutes for conventional construction materials. Among them are:

—Research breakthroughs in coconut trunk utilization. The coconut trunk has been found feasible for use as lumber for house-framing, millworks such as tongue and groove for flooring and stone-cuts for sidings. Its hard outer-layer is especially useful for sidings, floorings, rafters and trusses; the soft core for ceilings and window jalousies.

—Utilization of agri-wood wastes such as sawdust, bagasse, rice hull and wood chips in the production of hollow blocks and particle board.

By the year 2000, the level of housing investments will reach at least three percent of the gross national product. A significant increase in housing investments by the government will be attained, with the private sector also being expected to accelerate its own investments.

INDUSTRY

By the turn of the century, the industrial sector, composed of the manufacturing, mining, construction and utilities sector will account for approximately 52 percent of the total economic output of the economy. This is foreseen within the context of a broad industrial base characterized by a complementary mix of cottage, small, medium and large scale industries. By then, the country's imports will consist mainly of high technology finished and intermediate products, e.g. high alloy metals, electronic parts and other sophisticated equipment.

The achievement of these goals is not at all far-fetched, according to Hermenegildo Zayco, governor of the Board of Investments and for several years a top officer of the Philippine Refining Company, Unilever's subsidiary in the Philippines.

He anchors his optimism on the fact that the country has a relatively low-wage skilled labor force and that the average Filipino's knowledge of English facilitates technology transfer. He cites the relative success of the Philippine electronics industry which in 1980 exported some US$600 million, from virtually nothing in 1971. Requirements of the basic industries in terms of capital and technology are high, he concedes, but they still must be pursued as they are expected to supply the raw material needs of downstream manufacturers who produce products for domestic and export markets.

He especially believes that Filipinos abroad who are now employed in high technology companies, constitute an untapped technological pool. By 2000, when the country's science and technology infrastructure has considerably improved, he foresees the return of these highly trained Filipinos.

To provide financing for development, Sixto Roxas, vice chairman of the Union Bank of the Philippines and president of Bancom Philippines Holdings, Inc., foresees the emergence of a new breed of bankers who will "commit themselves to the developing community rather than to a product or a service." These financiers would therefore be "sensitive to the totality of the community's needs, and will package an appropriate financial design for the entire community."

Meanwhile, Roberto Ongpin, Minister of Trade and Industry, aptly describes the Philippines of the future when he says that with the development of a basic infrastructure firmly set and self-reliance in food obtained, more improvements in technology would be brought about by native innovation as well as technology imports and adaptations to push the momentum towards full industrialization.

The report was coordinated by Leverage International (Consultants) Inc., Makati, Manila.

© 1981 SCIENTIFIC AMERICAN, INC

The Moons of Saturn

The 17 icy bodies that orbit the planet display a surprising range of geological evolution. Many of them show craters more than four billion years old, but one of them has terrain so new that no craters are seen

by Laurence A. Soderblom and Torrence V. Johnson

Before the spacecraft *Voyager 1* neared Jupiter in March, 1979, only five bodies in the solar system other than the earth had been observed in sufficient detail for their history to be surmised. In essence all of them (Mercury, the moon and Mars and its moons Phobos and Deimos) consist of rocky material. The encounter with Jupiter and its moons doubled the list. Moreover, it marked the first appearance on the list of planet-size moons composed mostly of ices.

The encounters with the Saturn system doubled the list again. In November, 1980, *Voyager 1* flew past Titan, the largest moon of Saturn, at a distance of only 7,000 kilometers. It passed the smaller moons Mimas, Dione and Rhea at greater distances but nonetheless transmitted high-resolution images of each back to the earth. The trajectory of *Voyager 2* had already been devised to bring the spacecraft closer than *Voyager 1* had come to the moons Iapetus, Hyperion and Phoebe; in addition it would come very close to Tethys and Enceladus. In the months before the arrival of *Voyager 2* near Saturn last August the sequence of observations planned for the spacecraft was altered to provide for observations of several newly discovered moons, three of which were found by *Voyager 1*. In spite of a temporary jam in the mechanism that points the

cameras on *Voyager 2,* the mission was successful. In a few short months, therefore, the moons of Saturn have been transformed. Before November, 1980, the ones that were known were no more than dots of light in a telescope. Now they form an array amounting to 17 new worlds.

General Properties

Several generalizations can be made about the moons of Saturn. In the first place only one of them has any appreciable atmosphere. It is Titan, whose atmosphere is opaque to visible light. Since no one has seen the surface of Titan, we have little direct information about its geological evolution. Second, it can be calculated that all but the three outermost moons of Saturn should certainly be in synchronous rotation: they should keep the same face turned toward the planet, just as the moon keeps the same face turned toward the earth. In each such case a planet's gravitation raises a tidal bulge on a moon. Then the gravitational attraction between the bulge and the planet acts as a torque that slows the moon's rotation until the rotation is synchronous. The Voyager images suggest that all but one or possibly two of the moons of Saturn really do rotate synchronously. The definite exception is Phoebe, the outermost

moon, which is too small and too far from the planet to lose its spin to tidal forces. It rotates once every nine hours, whereas its orbital period is 13,211 hours, or 1.5 years. The possible exception is Hyperion, the third-outermost moon. The images made of Hyperion by *Voyager 2* cover only a short arc of the orbit of the moon and leave it uncertain whether Hyperion rotates synchronously or not.

Finally, all but two of the moons of Saturn form a regular system of satellites. That is, their orbits are nearly circular and lie in the equatorial plane of the planet. The two exceptions are Iapetus, the second-outermost moon, whose orbit is inclined 14.7 degrees with respect to the equatorial plane, and Phoebe, the outermost, whose orbit is inclined 150 degrees. (In addition the orbital motion of Phoebe around Saturn is in a direction opposite to that of all the other moons.) Three regular systems of satellites are known in the solar system; they consist of the inner moons of Jupiter, the inner moons of Saturn and the five known moons of Uranus. It is probably no coincidence that those planets all have rings. Rings and a regular system of satellites may form naturally as a by-product of the accretion of a giant planet. In any case each regular system of satellites is thought to have formed from the gas, ice and dust around each incipient giant planet, much as (on a larger scale) the planets formed around the sun.

Observations made with telescopes before the Voyager spacecraft arrived showed that among the moons of Saturn only Titan is as large as the four moons of Jupiter discovered by Galileo; the others are smaller than the earth's moon. The Voyager images show that the moons span a range of sizes from the size of asteroids to that of Mercury. The masses of Saturn's moons remain difficult to specify. An analysis of the mutual gravitational effects the moons have on one another has yielded values for some of the masses, and the tracking of gravitational perturbations in the trajectory of the Voyager spacecraft as they
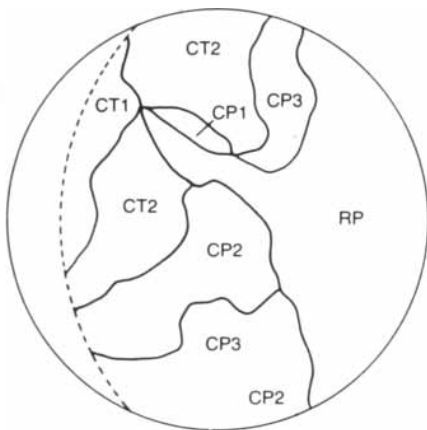
SATURNIAN MOON ENCELADUS was photographed last August 25 by the spacecraft *Voyager 2*. It is some 500 kilometers in diameter and is seen here at a range of 119,000 kilometers in an image that has been processed by computer so that the surface topography stands out. Six different terrains are distinguished. The heavily cratered terrains called *CT*1 and *CT*2 are the oldest; many of their craters are thought to represent the bombardment of Enceladus and other Saturnian moons by debris in orbit around Saturn. The debris was left over from the accretion of the planet, its moons and its rings. The cratered plains *CP*1, *CP*2 and *CP*3 are intermediate in age. The ridged plain *RP* is the youngest; on one hypothesis it lacks visible craters because it consists of upwelling fresh material.

101

passed among the moons yields new values and refinements of some of the earlier ones. It seems clear from the various measurements that the moons of Saturn all have densities of less than two grams per cubic centimeter. In fact, several have densities of less than 1.5 grams per cubic centimeter. Such values suggest that the moons are composed mostly of ice. For most of Saturn's moons a composition of 30 to 40 percent rock and 60 to 70 percent ice by weight would match the calculated density. Only Titan is large enough for its gravitational self-compression to affect its density appreciably. When this self-compression is taken into account, the density estimated for Titan—1.9 grams per cubic centimeter—becomes compatible with a mixture of half rock and half ice.

In one sense the calculated densities are curious. Among the planets of the solar system one finds a trend toward greater density with decreasing distance from the sun, and among the moons of Jupiter one finds a trend toward greater density with decreasing distance from the planet. Such trends are attributed to the influence of the heat from the central body on the temperature of the gas and dust that surrounded it at the time its satellites formed. In the case of Jupiter it appears that ice was unstable in any large quantities at radial distances from the planet less than the distance of the orbit now followed by Ganymede.

In spite of the uncertainties in the measurements no similar trend of density is evident among the moons of Saturn. Instead the current determinations of the values of the densities are consistent with a composition of rock and ice that is similar in all Saturn's moons except for more or less random variations in the exact proportions. On the other hand, the densities of Saturn's moons in general are less than those of Jupiter's. This suggests a greater proportion of ice. The relative lack of rocky material is explained by models of Saturn's history developed by James B. Pollack and his colleagues at the Ames Research Center of the National Aeronautics and Space Administration. The models suggest that rocky material near Saturn was swept into the incipient planet as it contracted before its moons began to form some 4.5 billion years ago.

In any event the surfaces of the moons of Saturn suggest the presence of ice.
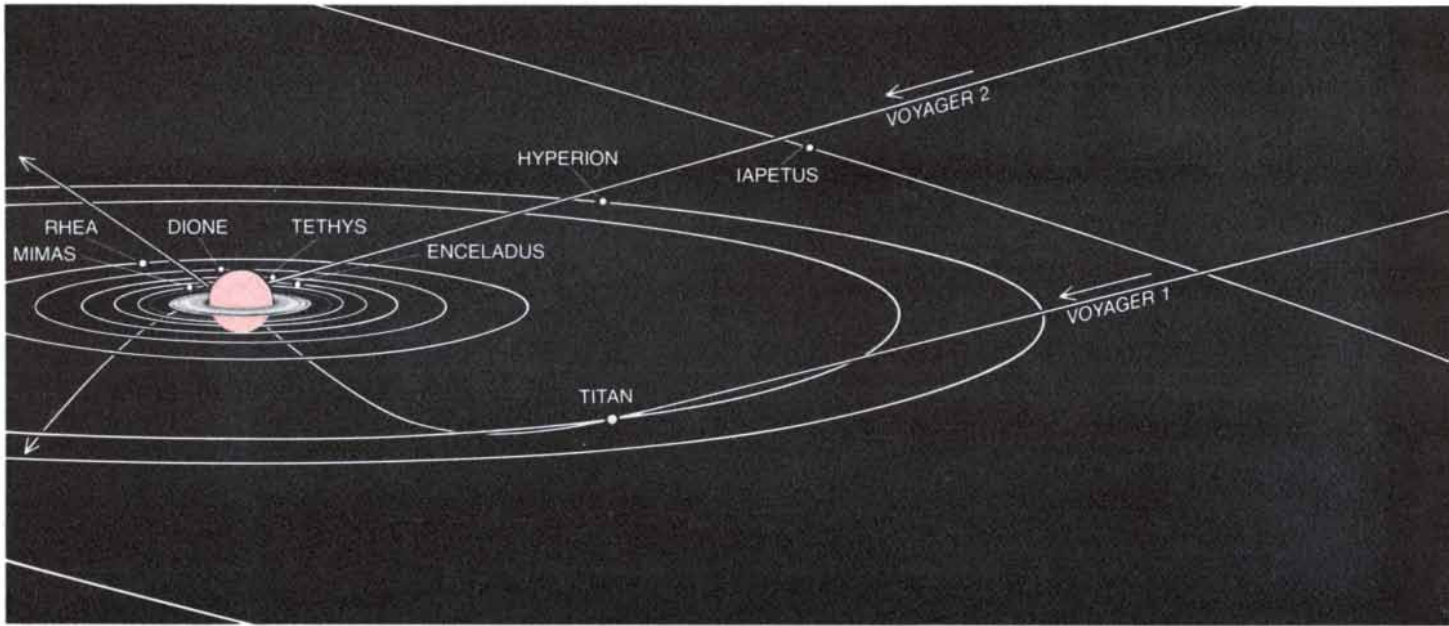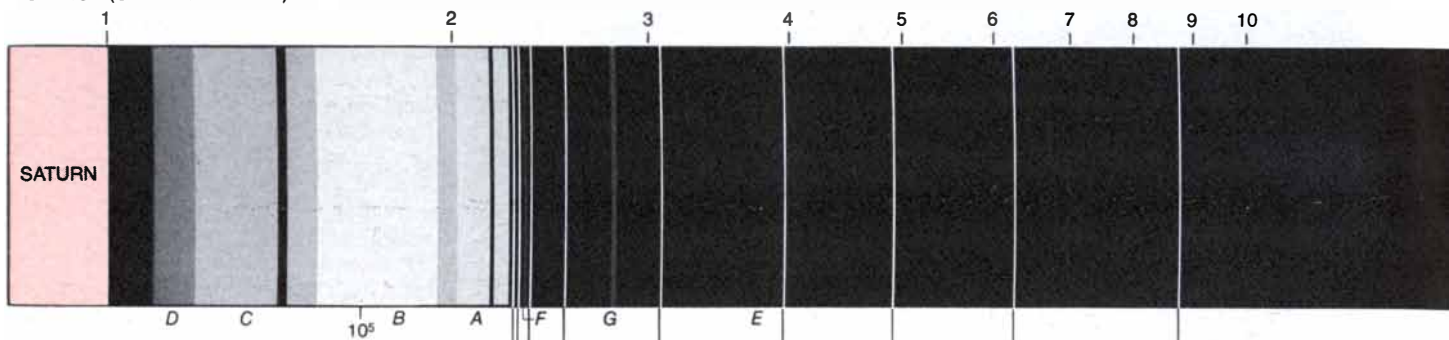
KNOWN MOONS OF SATURN are 17 in number. The top panel shows their orbits. All but two of them lie in the equatorial plane of the planet, which is also the plane of the planet's rings. The exceptions are Iapetus (with an orbit inclined by 14.7 degrees) and Phoebe (with one inclined by 150 degrees). Phoebe's rotation about Saturn is clockwise; all the other moons go counterclockwise. The top panel also shows the trajectory followed by *Voyager 1* in November, 1980, and the one followed by *Voyager 2* last August. The middle panel shows the orbits of the moons on a logarithmic scale; the numbers on the scale are distances from the center of Saturn. Each of the

102

Spectra of the solar radiation reflected by the moons in the near infrared show its absorption at wavelengths characteristic of water frost. Moreover, measurements made by the Voyager spacecraft show that most of the moons reflect between 60 and 90 percent of the radiation that hits them. With an albedo, or reflectance, of nearly 100 percent, Enceladus is the most reflective body in the solar system. If it were at the same distance from the sun as the earth's moon, it would be about five times as bright. Water ice is highly reflective.

The ubiquity of ice in the outer solar system was not entirely unexpected. For one thing, the vapor pressure of water ice (that is, the tendency of the ice to sublimate and lose vapor to space) depends strongly on temperature. Hence at distances from the sun less than the distance of the asteroid belt between the orbits of Mars and Jupiter an unprotected mass of ice will evaporate in a time quite short compared with the age of the solar system. At greater distances a mass of ice will be stable for billions of years. In addition most models of how the solar system formed predict that water should be a major constituent of a body that accreted at low temperatures.

Specifically, if a gas whose composition is much like that of the sun is cooled under the conditions of temperature and pressure thought to have prevailed in the early solar system, some of the oxygen in the gas will combine with silicon to form silicate rock at relatively high temperatures. When the elemental silicon is exhausted, however, a substantial amount of oxygen will remain. As the temperature continues to decrease it will combine with hydrogen, the element most abundant in the gas. Thus water will form. It emerges from such models that a moon condensing at low temperatures should have much the same proportions of rock and ice as those inferred for Jupiter's moons and Saturn's moons from the estimates of their densities.

One might suppose an icy moon would be no more interesting geologically than a cratered ice cube. Ice, however, has a melting point far lower than that of rock, so that relatively little is needed to melt the interior of an icy moon in the outer solar system. On this basis it was suspected even before the Voyager missions that such moons

ORBIT OF PHOEBE

| | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 | 200 | 300 |

HYPERION IAPETUS PHOEBE TITAN

10^6 10^7
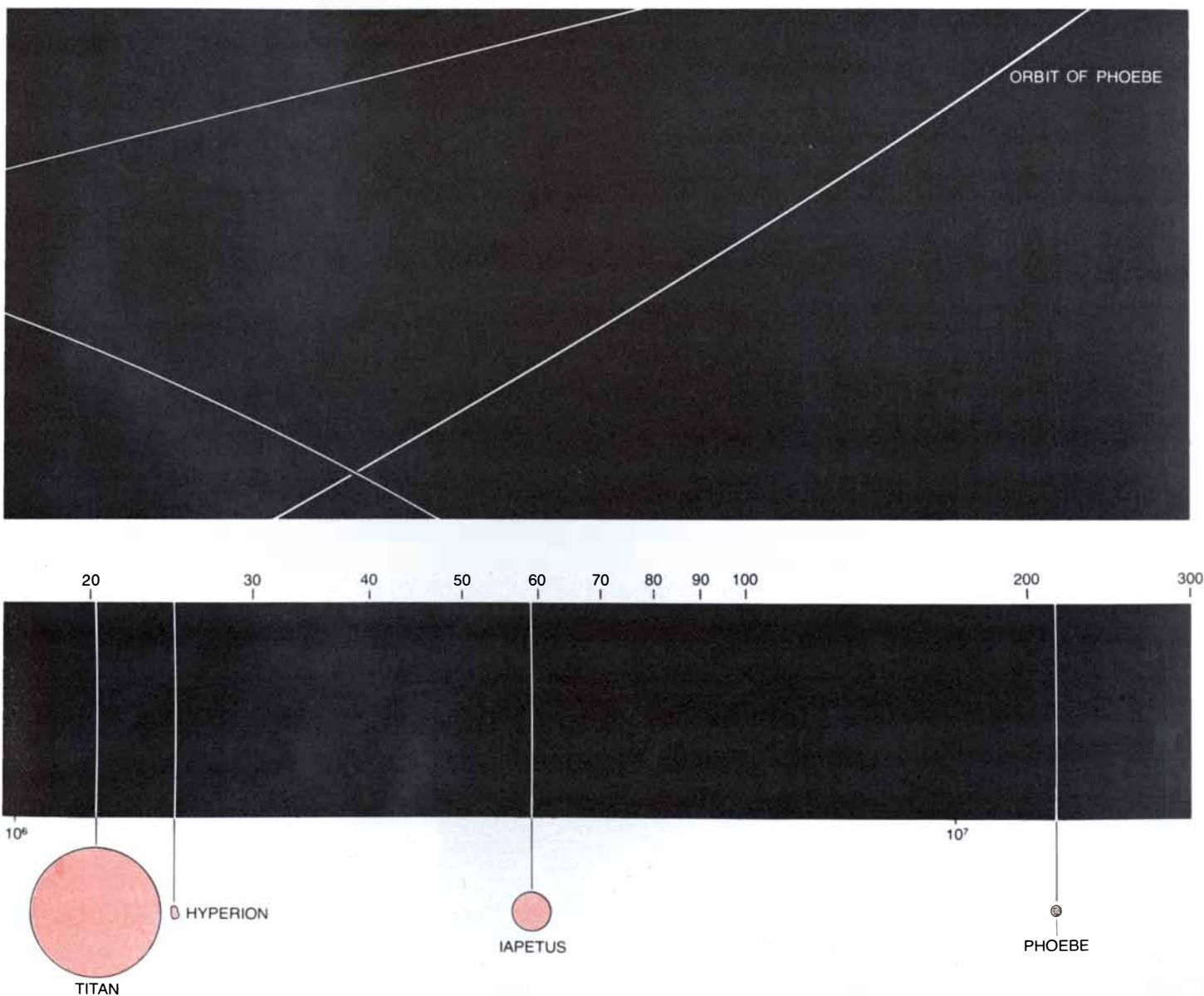
moons discovered recently turns out to have a dynamical vagary. The moon 1980S28 lies near the outer edge of the *A* ring; 1980S27 and 1980S26 bracket the *F* ring, and 1980S1 and 1980S3 have orbits that differ by less than the sum of the diameters of the moons (hence they are "co-orbital"). Finally, 1980S13 leads Tethys by 60 degrees, 1980S25 trails Tethys by 60 degrees and 1980S6 leads Dione by 60 degrees. Each of the three occupies a position of stability studied by Joseph Louis Lagrange in the 18th century and today called a Lagrange point. They are the first known Lagrangian moons. At the bottom of the illustration all the moons of Saturn are shown to scale.

103

might show signs of geological activity. It was also proposed that the moons of the outer solar system might incorporate substances such as ammonia hydrates and the compounds of methane and water known as clathrates. The interior of a moon incorporating such material would melt even more readily than a moon consisting of rock and water ice alone. The results of the Voyager missions surpass the speculations. Hyperion, Mimas and Enceladus, for example, are much the same size, but they display a range of geological evolution far broader than one would have thought was likely.

The New Satellites

The moons of Saturn discovered in the past decade have dynamics that are unusual in one way or another. Consider 1980S28, the innermost moon in the set and also the innermost known moon of Saturn. It was discovered by *Voyager 1* just beyond the outer edge of Saturn's *A* ring. 1980S28 is an elongated body whose diameter is about 40 kilometers. One hypothesis suggests that its gravitational field does much to sculpture the sharp outer edge of the ring. Somewhat farther out from Saturn are the pair of small moons 1980S27 and 1980S26. They too were discovered by *Voyager 1,* although it can now be recognized that at least one of them may have affected the counts of charged particles made by the *Pioneer 11* spacecraft near Saturn in September, 1979. Between their orbits lies the multistrand *F* ring. The gravitational fields of the moons may well confine the ring. For that reason they are called the shepherd moons.

About 10,000 kilometers beyond the *F* ring, or roughly halfway between the *F* ring and the *G* ring, are the "co-orbital" moons 1980S1 and 1980S3. The French astronomer Audouin Dollfus photographed one of them at the Pic du Midi Observatory in 1966. Today it is difficult to say which one it was. In 1978 John W. Fountain and Stephen M. Larson of the University of Arizona determined that there were two moons. Then in 1979 *Pioneer 11* made an image of one of them. The Voyager spacecraft made images of both. The mean orbital radii of 1980S1 and 1980S3 differ by less than the sum of their diameters. Hence their orbital velocities are similar but not quite identical: the inner moon slowly overtakes the outer one. As they approach each other the gravitational attraction between them alters their angular momentum. The inner moon gains momentum; it moves into a larger orbit, where its orbital speed is reduced. The outer moon loses momentum; it moves into a smaller orbit, where its orbital speed is increased. In short, the two moons change places; the inner moon becomes the outer one and begins to fall behind. About once every four years the celestial dance is repeated and the two change places again.

Three other new moons of Saturn were found by earth-based telescopes in 1979 and 1980, a time when Saturn's rings were edge on as they are seen from the earth. This orientation greatly reduces the effect of scattered light from the rings when Saturn is viewed in a telescope and allows the detection of faint bodies close to the planet. Two of the three new moons share their orbit with Tethys. One of them maintains a position about 60 degrees ahead of Tethys; it was discovered by a group led by Bradford A. Smith of the University of Arizona. The other stays about 60 degrees behind Tethys; it was discovered with the aid of a prototype of a planetary camera system designed for the telescope that the U.S. plans to put in orbit around the earth. The third moon stays about 60 degrees ahead of Dione; it was discovered by two French astronomers: P. Lacques of the Pic du Midi Observatory and J. Lecacheux of the Paris Observatory.

Each of the positions occupied by these moons is a point of dynamical stability of the kind first studied by the French mathematician Joseph Louis Lagrange and today called a Lagrange point. In 1772 Lagrange noted that in a system consisting of one body in orbit around another (say the moon around the earth) there are five positions at which a third body can lie undisturbed. Three of the positions are unstable: a body at any one of them is readily driven from it by the influence of gravitational forces other than those exerted by the two bodies that set up the system. The three positions lie (1) inside the orbit of the satellite, (2) outside the orbit and (3) at the point in the orbit opposite the satellite itself. The remaining two positions—the ones on the orbit 60 degrees ahead of the satellite and 60 degrees behind—are quite stable: a body that occupies either one of them will merely drift back and forth along its orbit under the influence of perturbing forces. It has long been known that the groups of asteroids called the Trojans occupy the two Lagrange points 60 degrees from Jupiter along the orbit of the planet around the sun. The three bodies at Lagrange points near Saturn are the first known Lagrangian moons.

In contrast to their dynamical vagaries the newly discovered moons have much the same appearance. Each is rather small, and almost all of them are quite irregular in shape. The irregularity tells much. It suggests that each moon arose from the fragmentation of a larger body. And since a small, cold, icy moon is strong enough to keep itself from being pulled into a sphere by its gravitational self-attraction, the irregular shapes suggest that these moons have not been heated significantly since the time they took their current form.

Iapetus

Iapetus and Rhea are the second-largest of Saturn's moons; they have a nearly identical diameter of about 1,500 kilometers. They move in quite different orbits. The orbit of Iapetus lies about 60 Saturnian radii (R_s) from the center of Saturn, that of Rhea about 9 R_s. They bracket the orbits of two other moons of Saturn, namely Titan and Hyperion. As we have noted, the orbit of Iapetus is inclined.

Iapetus has a density of about 1.1. This means its density is almost as low as that of pure water ice. Since it keeps the same face turned toward Saturn, its trailing hemisphere (the side facing backward with respect to its orbital motion) is always the same part of the surface. The trailing hemisphere is bright. In the visible region of the electromagnetic spectrum its albedo is almost 50 percent. In contrast the leading hemi-
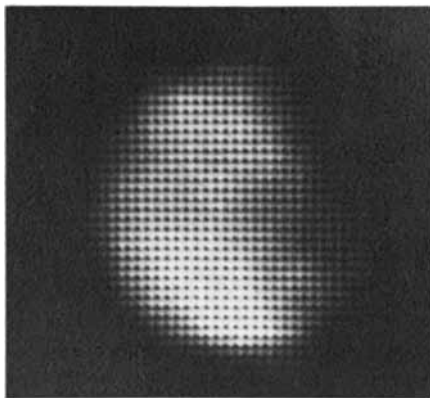
PHOEBE, the outermost moon of Saturn, was photographed by *Voyager 2* at a range of 2.2 million kilometers. Its curious orbit and its low reflectivity (about 5 percent) suggest that it formed elsewhere in the solar system and then was captured by Saturn's gravity. Phoebe is about 200 kilometers in diameter.
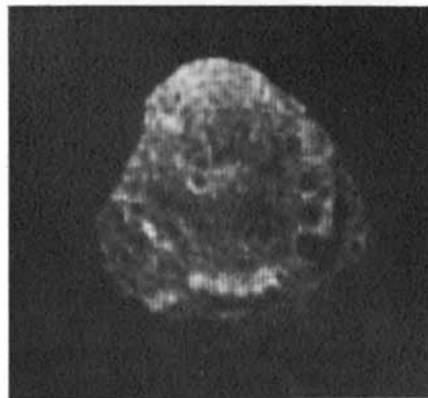

HYPERION, the third-outermost moon, was photographed by *Voyager 2* at a range of 500,000 kilometers. It is an irregular body about 400 kilometers long and 220 kilometers wide. It is four to six times brighter than Phoebe. The concavities along the margin of this image of Hyperion are probably craters.

sphere is extremely dark: its albedo is only 3 to 5 percent. This difference was noted in the 17th century by Jean Dominique Cassini, the discoverer of Iapetus, who found he could see the body on one side of Saturn but not the other. Among the few materials whose albedo is only a few percent are lampblack and the primitive meteorites called carbonaceous chondrites.

The peculiar pattern of darkness and brightness on Iapetus has suggested that dark material falling from space coats what amounts to the prow of the moon. In 1974 it was further suggested by Steven Soter of Cornell University that the source of the material is Phoebe. Phoebe had long been taken to be a moon of low albedo; observations made with earth-based telescopes show that its color is similar to that of many dark asteroids. The idea, then, is that dark matter is kicked off Phoebe by the impact of micrometeoroids. Then the particles of the matter come under the influence of what is called the Poynting-Robertson effect. Specifically, electromagnetic radiation leaves each particle because the particle reflects some of the radiation incident on it and because the particle absorbs some radiation and reemits it later. In either case the radiation leaving the particle in the direction of the particle's orbital motion around Saturn undergoes a Doppler shift that gives it a higher frequency (and thus more energy and momentum) than that of the radiation leaving in the opposite direction. The net result is that the particle loses orbital angular momentum and slowly spirals inward. As it falls (the argument concludes) it is swept up by Iapetus.

The images of Phoebe made by *Voyager 2* show some detail of Phoebe's sur-

IAPETUS was photographed by *Voyager 1* at a range of 3.2 million kilometers and nine months later by *Voyager 2* at a range of 1.1 million kilometers. The image made by *Voyager 1* (*top*) shows the side of the moon that always faces Saturn. (Iapetus and most other Saturnian moons keep the same face turned toward Saturn, just as the earth's moon keeps the same face turned toward the earth.) Iapetus is moving toward the left; the leading hemisphere of Iapetus (the hemisphere facing in the direction of the orbital motion of the moon) is dark. A ring of dark material extends, however, into the bright trailing hemisphere. The image made by *Voyager 2* (*bottom*) shows the side of the moon that always faces away from Saturn. The north pole of the moon coincides roughly with the large crater astride the border between day and night on Iapetus (the terminator), which crosses the top of the image. The dark leading hemisphere is at the right. An equatorial band of darkness extends into the trailing hemisphere. The sharp and complex boundary between bright and dark regions on Iapetus militates against the proposal that dark matter fell from space onto the leading part of the moon. It suggests instead that the dark matter is an extrusion from within Iapetus.

105
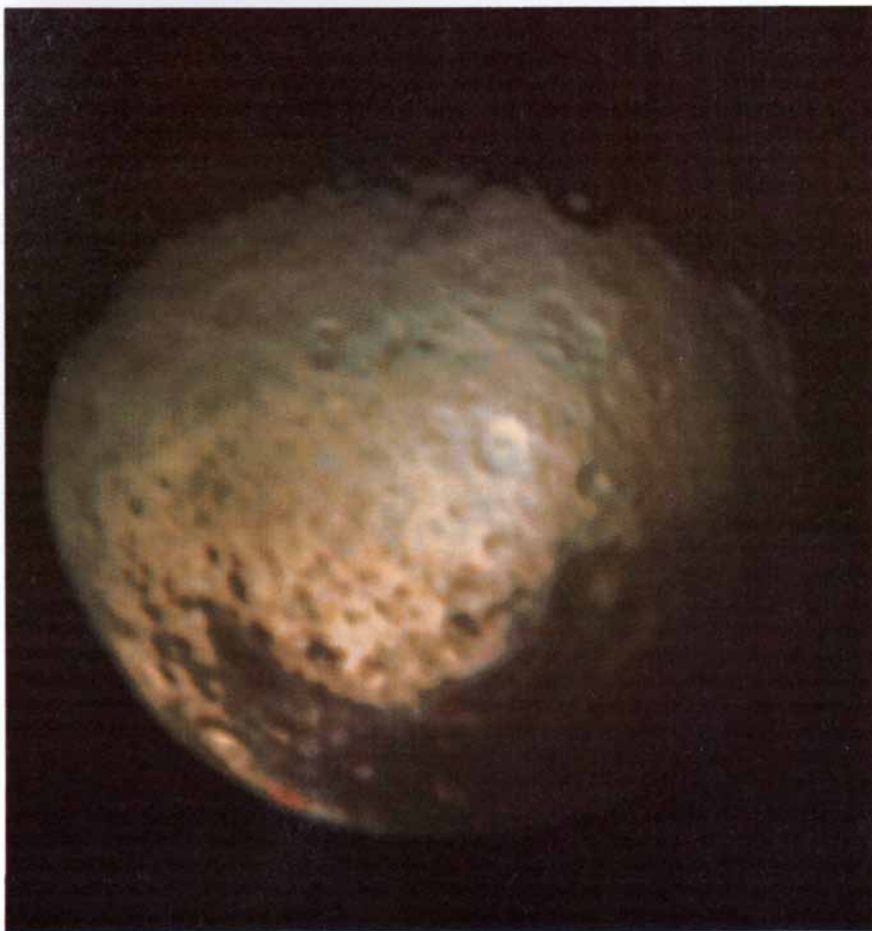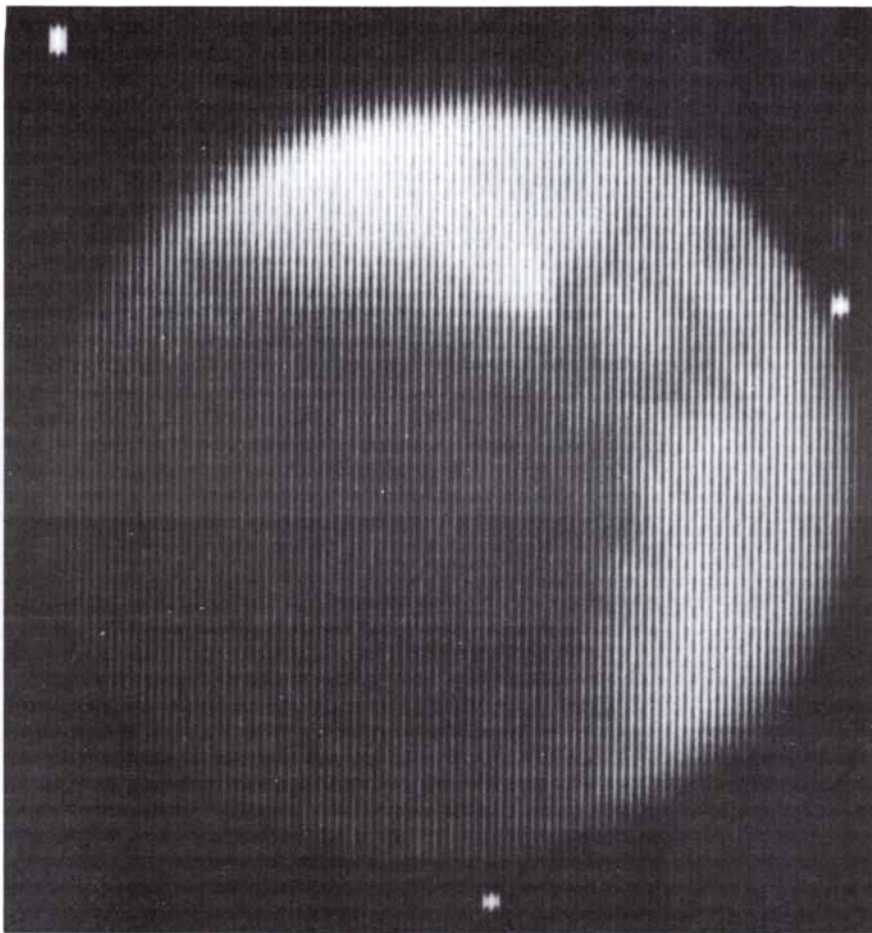
face. Phoebe emerges as being roughly spherical and having an albedo of only about 5 percent. In both of these attributes (and in its inclined, retrograde orbit) it proves to be quite different from the co-orbitals, the ring shepherds and the Lagrangian moons of Saturn. Phoebe may indeed be a captured dark asteroid, unchanged since it accreted early in the history of the solar system. Perhaps it is a primordial body that was ejected from the inner solar system by the gravitational field of the growing planet Jupiter and took up an orbit around Saturn.

In themselves the Voyager images of Phoebe cannot make it plain whether Phoebe dusted the dark leading hemisphere of Iapetus; the idea is still being debated. One test of the idea involves Hyperion, the next moon inward from Iapetus. The test rests on the assumption that Hyperion should have been dusted by material kicked off Phoebe that Iapetus failed to sweep up. The images made by *Voyager 2* show Hyperion to be a remarkable little moon. It is nearly the size of Mimas, a spherical moon with a radius of about 200 kilometers, and yet it is markedly elongated: its short axes are only about three-fifths the length of its longest axis. Hyperion is one of the largest bodies in the solar system that has an irregular shape. (The asteroid Hector is one of the very few other such objects.) An analysis of the Voyager images indicates that Hyperion is somewhat darker than most of the moons of Saturn: its albedo is 20 to 30 percent. This finding is consistent with a dusting by matter from Phoebe. Hyperion appears, however, to lack a dark leading hemisphere. The problem with searching for one is that it is not yet established whether Hyperion rotates synchronously.

An answer to the question of whether Phoebe dusted Iapetus therefore awaits in part the determination of the rotation rate of Hyperion from observations made with ground-based telescopes. Meanwhile the findings of Dale P. Cruikshank and his associates at the University of Hawaii at Manoa may have got the idea into trouble on other grounds. Cruikshank's group has found from telescopic observations that the color of Phoebe and the color of the dark leading hemisphere of Iapetus are different. The dark regions on Iapetus are much redder than Phoebe throughout the visible and the near-infrared region of the spectrum. This makes it difficult to favor a scheme in which material from Phoebe simply coats the dark part of Iapetus without having undergone some kind of change.

The images of Iapetus made by the Voyager spacecraft also tend to subvert the hypothesis that it was dusted by matter from Phoebe. *Voyager 1*'s best image of Iapetus was made at a range of 2.5 million kilometers. The image showed only those features on Iapetus that are more than 50 kilometers across. Even so, a clue to the history of the body emerged. A ring of dark material about 100 kilometers in diameter was found to straddle the border between the hemispheres. It strongly resembles the rings in craters along the margin of large volcanic flood plains on the moon and Mars. Such rings formed there when dark volcanic melts flowed into impact craters and filled each one around its central peak. Perhaps the dark ring on Iapetus formed by a similar process when fluid extruded from the interior of the body darkened half of it. Clearly it is unlikely that a feature with such a peculiar geometry as that of a dark ring could have been formed by matter falling from space.

Voyager 2 made a series of images of Iapetus at a resolution three times greater than that of the images made by *Voyager 1*. The single best image is of the north polar region, mostly in the bright trailing hemisphere. It shows that the trailing hemisphere is heavily cratered. Many of the craters have a dark floor. Such floors resemble the dark floors of craters on the highlands of the earth's moon, which are thought to have formed when volcanic material flowed out over them. Taken together, the dark-floored craters in the trailing hemisphere and the sharp definition and complexity of the boundary between dark and bright terrain imply a history of eruptions from the interior of Iapetus. That is not to say the erupting material resembles any ordinary lava. One can speculate that it is a fluidized slurry of a mixture including ammonia, ice and something dark. The low albedo of the material suggests that whatever its origin, the dark stuff is rich in organic substances such as are found in primitive meteorites.

Rhea

The distant views of Rhea transmitted to the earth by *Voyager 1* as the spacecraft approached the moon showed a bright, bland leading hemisphere marked only by what appears to be a large and relatively recent impact crater. The trailing hemisphere is different. It shows a complex pattern of bright swaths on a background darker than the leading hemisphere. The swaths are thought to result from internal activity. Perhaps the bright material was extruded along lines of fracture at the surface. In any event the bright swaths fail to follow the kind of pattern that cratering would lay down on the surface. The zone containing the swaths is confined within a small circle at the center of the trailing hemisphere. The boundary be-

RHEA was photographed by *Voyager 1* at a range of 1.7 million kilometers. Here the image of the moon is shown in false color. The trailing hemisphere (seen here) is marked by bright wisps that do not seem to follow a pattern laid down by craters. The leading hemisphere is bland.
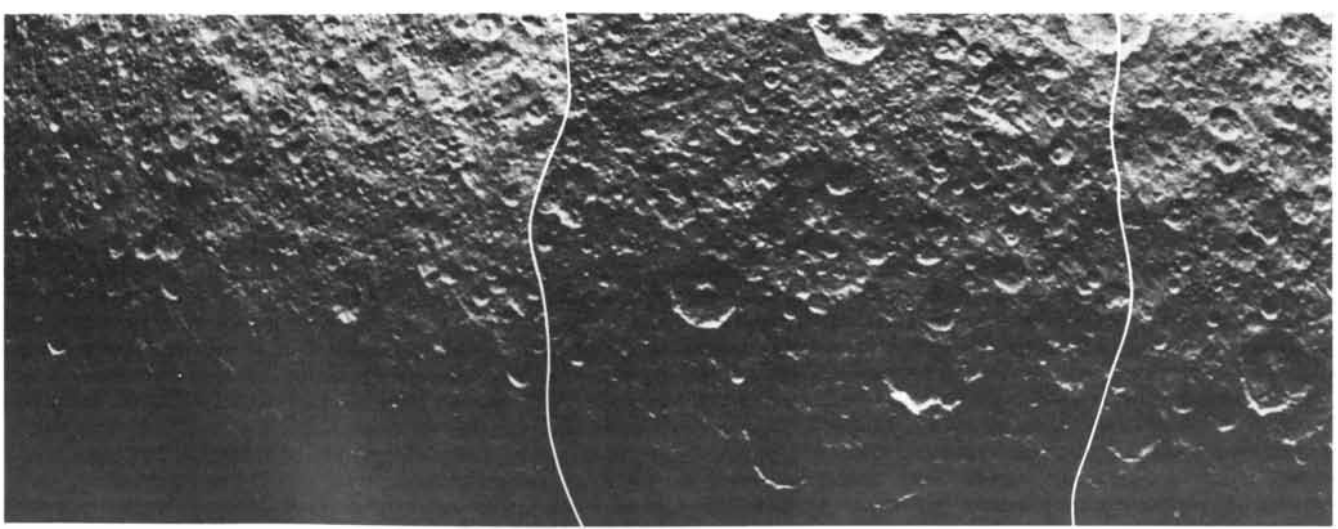
RHEA'S NORTH POLE was photographed by *Voyager 1* at a range of 80,000 kilometers. The pole itself is at the middle of the arc the terminator describes across this mosaic. The terrain to the west of the pole (toward the right in this image) is marked by large craters; in the terrain to the east such craters are missing. Presumably the large craters were made early in the history of the moon; the small craters both east and west were made later, after geological activity had given the terrain to the east of the pole a fresh, uncratered surface.

tween this zone and the bland leading hemisphere is diffuse.

The high-resolution views of Rhea made by *Voyager 1* show the equatorial region of the leading hemisphere somewhat to the east of the putative large, fresh impact crater. The best views show the north pole of the body, which *Voyager 1* flew over at a distance of only 59,000 kilometers. In each view Rhea is found to have a densely cratered surface much like the cratered highlands on the moon and Mercury. The principal difference is that the large, fresh craters on the moon and Mercury are surrounded by blankets of ejecta; the craters on Rhea are not. Presumably Rhea's weaker gravity is responsible. If Copernicus, a large crater on the earth's moon, had formed on Rhea, its ejecta would have been spread over most of the body's surface. On the earth's moon most of the ejecta lie no farther away than a few times the radius of the crater.

The nature and the origin of the bodies that cratered Rhea can be surmised from a mosaic of images showing the body's north polar region. At first the entire region appears to be uniformly pitted with craters. A closer examination shows that the western two-thirds of the polar mosaic is marked by a collection of craters with diameters ranging from 30 to 100 kilometers and in addition a dense population of smaller craters. The smallest ones visible are a few kilometers across. If other craters are smaller still, they are below the limit of resolution. The eastern third of the mosaic is also marked by small craters, but the larger craters are missing. A few subtle depressions in the surface are visible. They suggest that the larger craters in the east were filled or buried.

The difference between the two terrains implies a period of cratering in which projectiles with a wide range of energy formed craters with a wide range of sizes. At some point the remaining projectiles whose impacts could form large craters (craters greater than 50 kilometers in diameter) were swept up. The bombardment by the objects that formed the smaller craters continued. Meanwhile part of Rhea's surface was regenerated, perhaps by fluid extrusion from the interior or perhaps by the flow of a slurry of material driven to the surface by the pressure of gases accumulating within the moon. The resurfacing included the eastern third of the polar terrain. Estimates of the rate at which Rhea is being cratered by comets and asteroids today indicate that few of the visible craters (both large and small) in the images are recent. Hence most of the cratering must have come early in the history of the solar system, and it must have been intense. The projectiles could have been remnants of the accretion of gas and dust in orbit around the sun. They could have been remnants of the accretion of gas and dust in orbit around Saturn. They could also have come in pulses as the debris from bodies that collided in the early Saturnian system.

Dione

In sum, the appearance of Rhea suggests that at least two populations of projectiles marked the surface. The appearance of Dione, the next moon inward from Rhea, allows more detailed deductions. Dione travels in orbit around Saturn at a distance from the planet's center of 6.2 R_s. Its diameter is 1,100 kilometers. With a density of 1.5 grams per cubic centimeter it is the densest of Saturn's moons except Titan. The distant views of Dione made by *Voyager 1* showed a striking asymmetry between the leading and the trailing hemispheres. In the trailing hemisphere a network of intersecting bright streaks on a dark background was visible. It looked much like the trailing hemisphere of Rhea. The leading hemisphere was uniformly bright. Again the region of complex variations in albedo was confined within a small circle at the center of the trailing hemisphere, and again the boundary of this region with the bland leading hemisphere was diffuse.

Closer views of the trailing hemisphere made by *Voyager 1* showed that craters between 50 and 100 kilometers in diameter were crossed by the streaks. Therefore the streaks must have formed well after the torrential bombardment the craters imply. All things considered, the large craters on Dione and on parts of Rhea (including the western two-thirds of the polar mosaic) resemble the highland craters found on the earth's moon, Mars and Mercury. We shall refer to the projectiles that made such craters in the Saturnian system as Population I. The most likely hypothesis is that they resulted from the agglomeration of matter left in orbit around the sun after the solar system formed.

The closer views of Dione also showed plains where the density of cratering is far lower than it is on the "highlands." On the plains, however, the proportion of small craters to large ones is far higher. The proportion is that of the eastern third of the polar mosaic of Rhea. On both the plains of Dione and the terrain east of the north pole on Rhea the distribution of sizes among the craters resembles that of the craters surrounding the principal craters on the moon. It also resembles that of the craters surrounding the places on the earth where nuclear devices have been tested. In general the prevalence of small craters suggests the pocking of a landscape by the ejecta from the impacts of larger projectiles or by the fragments pro-

107
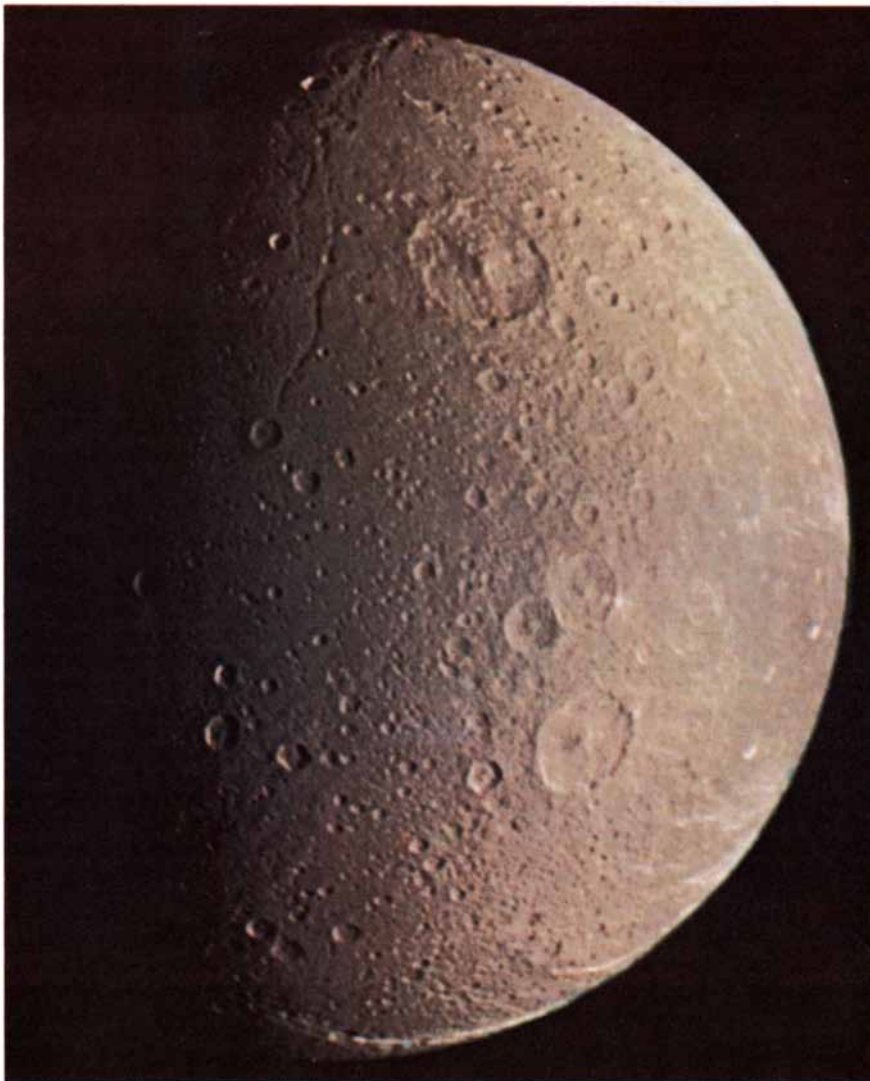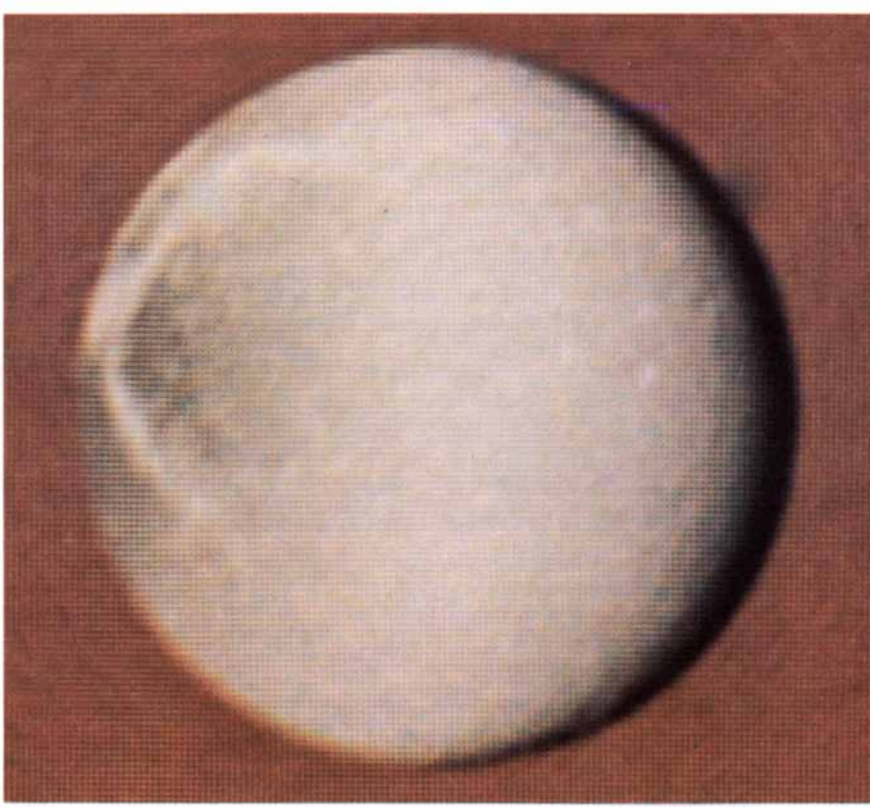
duced by collisions between large bodies. With Dione and Rhea it seems likely the small craters were made by the debris from bodies that collided within the Saturnian system. Presumably, therefore, certain regions of Dione and Rhea bear a record of primordial cratering; other regions were resurfaced by material from the interior that buried the oldest craters. The resurfaced regions were then pocked by a second bombardment. We shall refer to the bodies in the second bombardment as Population II.

A problem remains. The "highlands" bearing evidence of cratering by Population I and the plains bearing evidence of cratering by Population II show no correlation with the global pattern of albedo recorded by *Voyager 1* at a distance from Dione and Rhea. On Dione, for example, the leading hemisphere, which was uniformly bright and bland in the distant imagery, turns out at high resolution to have both highlands and plains. What, then, caused the global albedo pattern? Some calculations made by Eugene M. Shoemaker of the U.S. Geological Survey lead to estimates of the contributions that projectiles make today to the cratering of the Saturnian moons. For virtually any source of projectiles outside the Saturnian system the flux of impacts turns out to vary dramatically from the leading hemisphere to the trailing hemisphere of any given moon. For Dione, Shoemaker calculates a variation of 10 to one. For Rhea the variation is six to one. It makes no difference whether the projectiles are comets or asteroids or whether they have periodic orbits around the sun because the gravitational acceleration imparted to an arriving projectile by Saturn overwhelms the projectile's original trajectory.

Shoemaker further calculates that from the apex of Dione to the side of the body (with respect to the orbital motion of Dione around Saturn) the flux of arriving projectiles changes by a factor of

DIONE was photographed by *Voyager 1*. An image made on November 11, 1980 (*top*), shows the side of the moon that faces away from Saturn. The orange background is the top of Saturn's clouds, 377,000 kilometers from the moon. Dione is moving toward the right; the trailing hemisphere is at the left. The center of the trailing hemisphere shows bright wisps on a dark field. According to one hypothesis, a pattern of wisps once covered the entire surface of Dione. It was erased by the continual impact of small meteoroids over the history of the solar system, except at the center of the trailing hemisphere, where relatively few such projectiles arrive. An image made on November 12, 1980, at a range of 162,000 kilometers (*bottom*) shows what amounts to the prow of Dione. The apex of the leading hemisphere is at about the middle of the terminator along the left of the image. At the right is the side facing Saturn. The beginnings of several bright wisps in the trailing hemisphere are at the right limb of the moon.

LARGE IMPACT CRATER on Tethys is notably different from one on Mimas. The crater on Tethys (*left*) is more than 400 kilometers in diameter. It was photographed by *Voyager 2* at a range of 826,000 kilometers. When the crater was made, it must have been deep, and its rim and central peak must have been high. Today, however, the floor of the crater has rebounded to match the contour of the moon, and both the rim and the central peak have collapsed. Ap- parently Tethys once was warm enough to allow such rearrangement. In this image the apex of the leading hemisphere is near the center of the disk of the moon. The crater on Mimas (*right*) is 130 kilometers in diameter. It was photographed by *Voyager 1* at a range of 425,000 kilometers. It is the only large crater on Mimas. Its rim and its central peak are prominent. Apparently Mimas has long been cold. The crater's central peak lies at the apex of the leading hemisphere.

only two. From the side to the trailing end it changes by a factor of five. As a result the part of Dione least affected by the continuing impact of projectiles is only a small region at the trailing end of the moon.

With the aid of Shoemaker's calculations a history of Dione can be proposed. In this history the early life of Dione was dominated by the impact of large bodies most likely left over from the accretion of the solar system. Then parts of Dione were resurfaced. Meanwhile collisions near Saturn between the bodies left over from the accretion of the solar system yielded smaller bodies. Some of them took up orbits around Saturn much like Dione's. Their impacts with Dione cratered the newly formed plains moderately. At about this time fractures formed in the surface of Dione. The fractures were filled by bright extrusions from the interior.

It is likely that the entire surface of Dione then had the pattern now seen only at the center of the trailing hemisphere of the body. Over the past four billion years, however, Dione has rotated synchronously and its surface has been "gardened" by the impact of small meteoroids from outside the Saturnian system. The craters made by these bodies may be too small and too scattered to be identified in Voyager images. Still, the craters would rework the surface. In this way they erased the pattern of the surface in a region extending from the apex of the leading hemisphere well into the trailing hemisphere. This history accounts for the global pattern of albedo.

The meteoroids arriving well into the history make up Population III.

Tethys

Tethys is the next moon inward from Dione; it travels in orbit around Saturn at a distance of 5 R_s. Its diameter is almost identical with that of Dione, but its density—1.2 grams per cubic centimeter—is lower. Moreover, its appearance is quite different. *Voyager 1* viewed Tethys only at low resolution. The images showed diffuse patches of small variation in albedo on a heavily cratered surface. The pattern did not resemble the more pronounced global pattern seen on Rhea and Dione. One of the best images showed a branching canyon spanning the distance between the north and south polar regions on the side of the moon facing Saturn. It was estimated the canyon was at least 1,000 kilometers long, 100 kilometers wide and several kilometers deep.

Nine months later *Voyager 2* approached Tethys. The images it transmitted to the earth revealed an enormous impact scar in the leading hemisphere. The rim of the scar has a diameter more than two-fifths the diameter of Tethys itself. As the spacecraft continued its approach the scar was photographed progressively closer to the visible edge of the moon. Soon an image showed it in profile. Here it could be seen that the floor of what must once have been a crater now matches the spherical shape of the body. Only a low rim and a subdued central peak remain.

Evidently the interior of Tethys was sufficiently warm early in the history of the moon to allow the collapse of the raised topography. The same could be said of large craters on the icy Jovian moons Ganymede and Callisto.

Finally the north pole of Tethys and the side facing away from Saturn came into view. From this vantage it was apparent that the canyon found by *Voyager 1* extends over the north pole and down to the equator on the outward-facing hemisphere; thus the canyon traverses three-fourths of a great circle around the body. It could also be seen that parts of Tethys (like parts of Rhea and Dione) had been resurfaced. Smooth plains had developed in a small part of the leading hemisphere, and the resurfacing had buried some of the large craters already there.

A tentative explanation for the current appearance of Tethys begins with the surmise that at the time the large impact scar was made the interior of the body was much warmer and more mobile than it is today. Perhaps it was liquid. If Tethys had been cold and brittle when the scar formed, the impact that formed it might well have fragmented the moon. Moreover, the topography raised by the impact has clearly collapsed. The rim and the central peak of the impact scar persist. Hence it seems likely that much of the crust of Tethys in place at the time of the impact remains at the surface today. One can imagine, then, a simple history in which Tethys freezes from the crust down. If Tethys had first been liquid, the freezing would

110

PREHISTORIC SEA CREATURE?
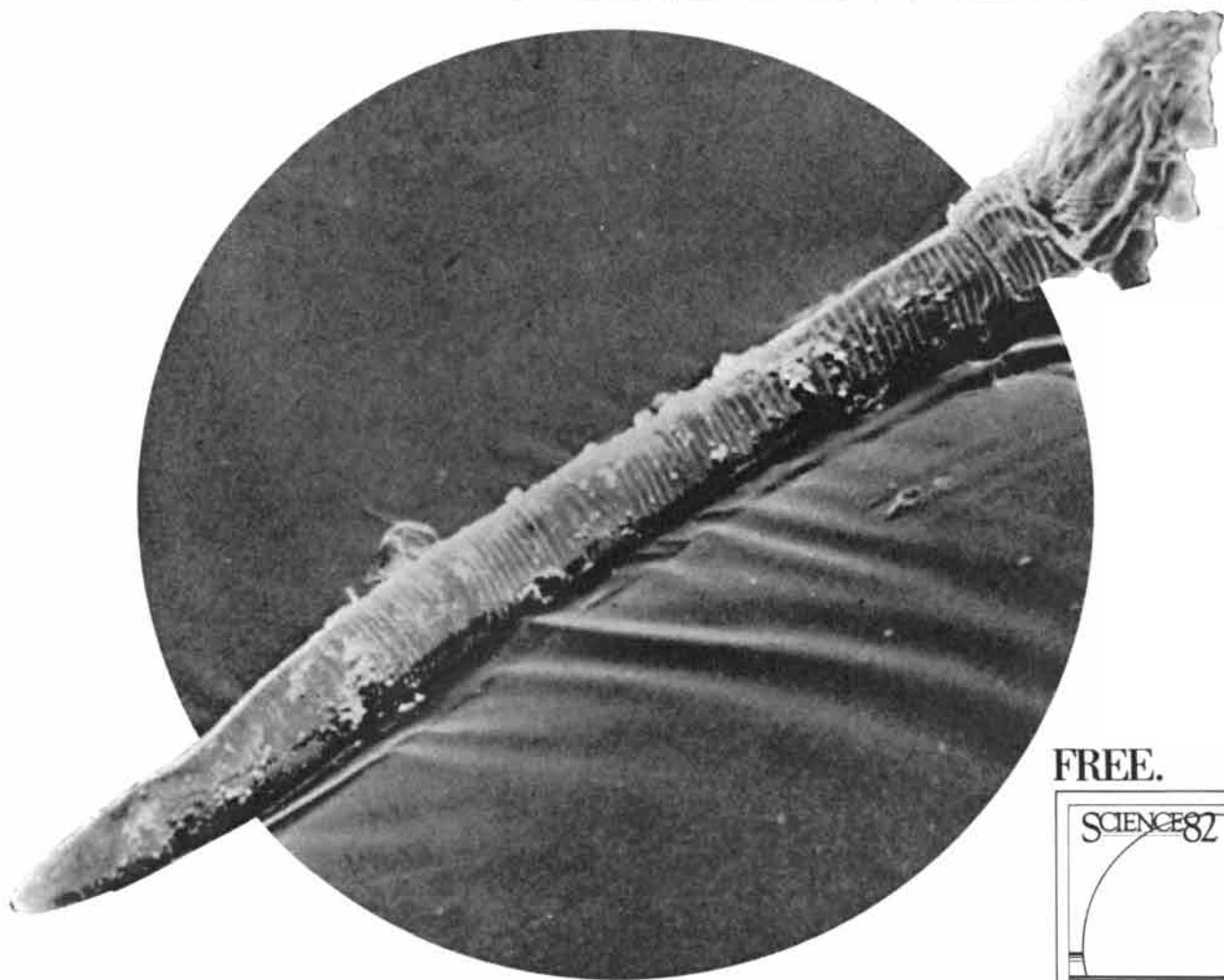
FREE.

No, it's very likely alive and well and living in your eyelashes!

It goes by the fancy name of demodex folliculorum. It lives in the follicle of the human eyelash — and nobody knows where it comes from...or even why it's there!

Fact is, there are scientific mysteries that surround and confound us everyday. And every month, there is one magazine that unveils them in a way that makes the whole world of science come alive. Science 82

Today, that means seeing through the eyes of a satellite the universe that lies beyond our solar system as much as probing through a microscope the myriad worlds that live within us. It's the discovery of a cancer created to fight cancer and the revelations of the

4th dimension . . . constructed by a computer.

Science 82 You can read it for the drama of breakthroughs in genetics, in solar power, in physics and medicine...you can read it for the beauty of its photography and illustration. Best of all, you and your family can enjoy it all—because it's all written to be understood.

Science 82 *Now you can try the next issue free.* Find out why more than 2 million readers every month let Science 82 open up their lives to the wonder of it all. Join them today at no obligation. Use the attached card (or the coupon) to send for your free issue.

111

have increased the volume of the object by about 10 percent and the surface area by about 7 percent. The estimated extent of the canyon on Tethys suggests that it forms 5 to 10 percent of the surface. The canyon may represent the stretching of Tethys' crust over the expanded, frozen interior.

Mimas and Enceladus

The smallest of the nine moons whose presence around Saturn was known before the 20th century are Mimas and Enceladus. Each of them has only a thousandth the mass of a Galilean moon of Jupiter, a thousandth the mass of the earth's moon and a hundred-thousandth the mass of the earth. Their tininess is significant. Before the Voyager spacecraft began to explore the outer solar system the main source of heat in a planet or a moon was taken to be the decay of the radioactive atomic nuclei within it; hence it seemed that the history of a planet or a moon would be determined by the ratio of its volume to its surface area. The volume governs the quantity of radioactive nuclei and thus the generation of heat; the surface area governs the loss of heat and conversely the retention of it. The ratio of volume to surface area is determined by the size of the object: the ratio increases with the radius. One therefore supposes larger bodies are more likely to melt than small ones. Certainly the attribution of internal heat to radioactivity and the resulting correlation of heating with size explains the observations that volcanically the earth's moon is long dead, Mars is moderately active and the earth is quite active. It also suggests, however, that tiny moons such as Mimas and Enceladus should not have evolved substantially since the time they accreted.

The image of Mimas made by *Voyager 1* showed a surface that conforms to this prediction. The surface is uniform in albedo and is saturated with impact craters. It is a surface that has not been reworked by volcanism since the time of the Population I bombardment. In the leading hemisphere of Mimas there is an impact crater 130 kilometers in diameter, about a third the diameter of the little moon. One suspects other craters of that size or even larger should have been made in the course of the bombardment. Why are they not seen? If a cold, brittle Mimas had been blown apart by one such impact, the relative velocities at which the debris would fly apart would probably have been on the order of the velocity required for escape from Mimas' gravity. Such velocities are small compared with the orbital speed of Mimas, which is 30 kilometers per second. Hence the debris would have remained in a narrow band surrounding what had been the orbit of the moon. Gradually the debris would have reaccreted. The result would have been a cold mass of rubble consisting mostly of ice, but with a few percent of rock. (The density of Mimas is 1.2 grams per cubic centimeter.)

The statistics of the craters seen on Mimas suggest this may have happened. On Mimas today the largest crater is the one 130 kilometers in diameter. Then comes a gap in the sizes. The next-largest craters have diameters of only a few tens of kilometers. After that the abundance of craters increases exponentially as the diameter of the craters decreases. It is as if a large-scale impact had destroyed a parent body and the fragments had recombined. The present pattern of craters results, then, from impacts by the last of the fragments. The single largest crater represents a collision not energetic enough to have destroyed the body again; the other large craters made by Population I objects disappeared with the disruption of the parent body.

Voyager 1 provided only distant views of Enceladus, yet if the craters on Enceladus resembled those on Mimas, the im-
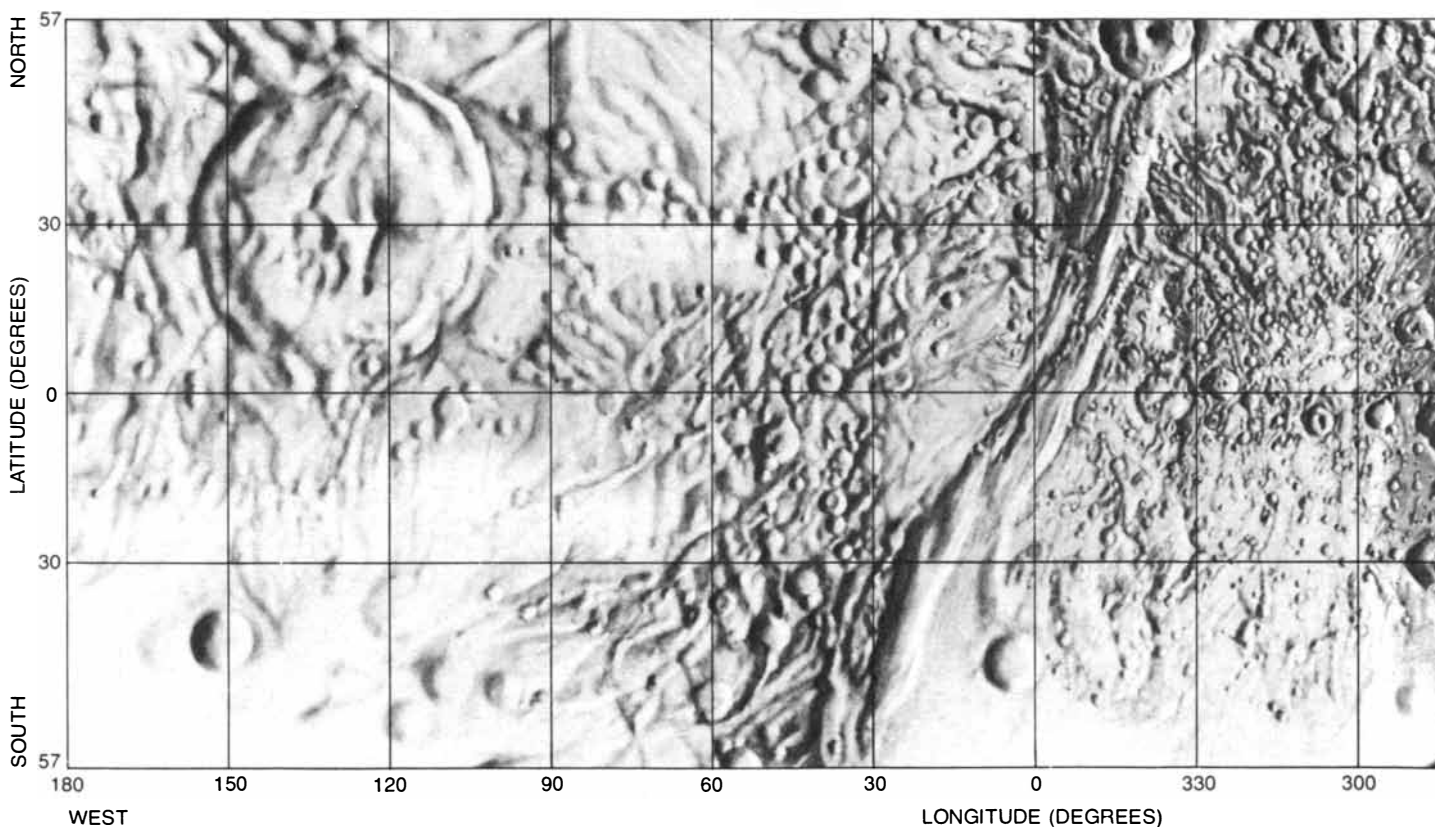
MAP OF TETHYS was prepared by the U.S. Geological Survey from images made by *Voyager 1* and *Voyager 2;* it is a Mercator projection, in which all latitudes and longitudes are made rectilinear. Zero degrees of longitude marks the center of the side of the moon that faces Saturn, 90 degrees marks the center of the leading hemisphere and 270 degrees marks the center of the trailing hemisphere. Two features dominate the topography. One of them is the collapsed impact crater shown in the photograph at the left on page 110. The central peak of the crater lies near 120 degrees longitude and 30 degrees north latitude. The other feature is a great canyon running from south to north between 30 and 330 degrees longitude. The canyon extends over the north pole of Tethys and then back toward the

112

ages should have shown them. A crater as large as the one in the leading hemisphere of Mimas should certainly have been visible. Instead the images showed a surface that seemed to be smooth. More important, the surface of Enceladus was brighter than that of the neighboring moons Mimas and Tethys. This suggested that much of the surface of Enceladus may have been regenerated and is covered with very fresh ice.

Two chains of reasoning had already implied that Enceladus was unusual. In the first place observations made from the earth had shown that the diffuse ring of Saturn designated the E ring has its maximum brightness along the orbit of Enceladus. This suggested that Enceladus might be the source of the ring. Indeed, it suggested that Enceladus might continually replenish the ring, because otherwise the gravitational field of the moon would tend to clear matter out of its orbit. (It is proposed that such gravitational shepherding is responsible for at least some of the banding discovered by the Voyager spacecraft in the rings of Saturn.)

In the second place, an argument developed by Stanton J. Peale of the University of California at Santa Barbara and his colleagues had led to the conclusion that Io, the innermost major moon

270   240   210   180

EAST
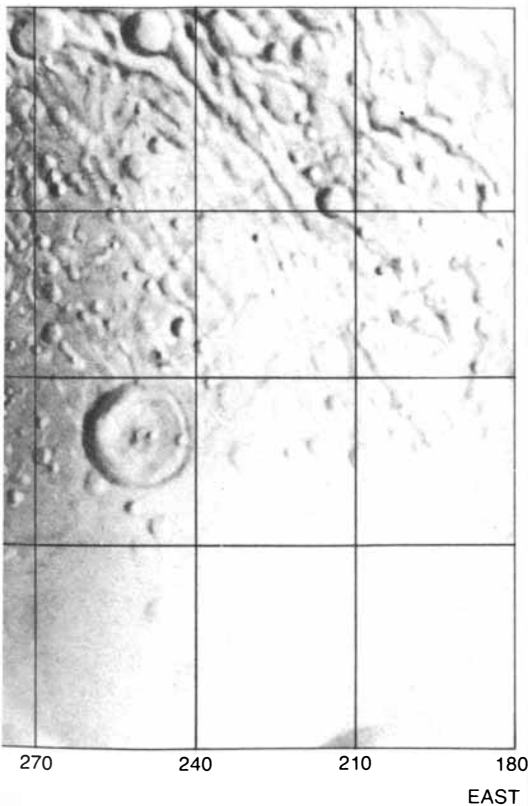
equator; its end is apparent at the upper right of the map. The canyon may represent the stretching of the crust of Tethys over an interior that expanded as it cooled and solidified into ice. The terrain on the map is shown as though it were all illuminated from the west.

of Jupiter, is heated substantially by tidal forces. Then Charles F. Yoder of the Jet Propulsion Laboratory of the California Institute of Technology proposed that a similar mechanism might act on Enceladus. Specifically, the orbital period of Enceladus is half that of Dione.

This means that Enceladus and Dione align themselves with Saturn at fixed points in their orbits. The gravitational attraction of Dione for Enceladus acts repeatedly at these points, so that although the attraction is small, it makes the orbit of Enceladus stay slightly el-

liptical. In such an orbit Enceladus is forced to oscillate radially in Saturn's gravitational field, much as Io oscillates radially in Jupiter's. The internal friction that results from this oscillation might keep the interior of Enceladus warm and mobile. Gases and fluids es-
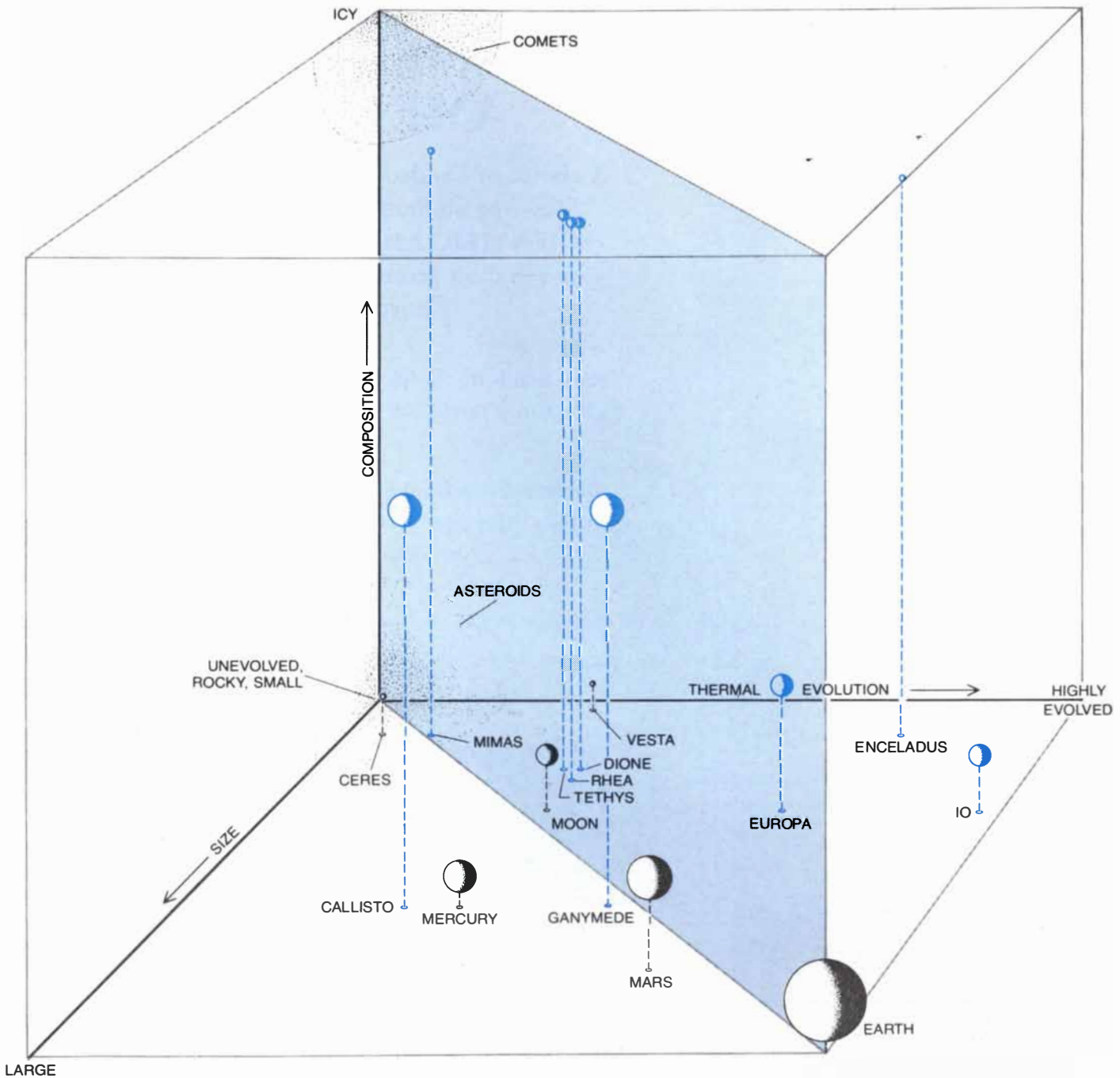
WIDE RANGE OF HISTORIES among planets and moons in the solar system is suggested when planets and moons with a visible solid surface are plotted in a cube whose axes represent their composition (from rocky bodies at the bottom of the cube to icy bodies at the top), their thermal evolution (from cold, unevolved bodies at the left to highly evolved, volcanic bodies at the right) and their size (from small bodies at the back to large bodies at the front). The black spheres are bodies whose position could be plotted before the Voyager spacecraft arrived in the outer solar system. In general these bodies (including Mercury, the earth, the earth's moon and Mars) are rocky. Moreover, the larger they are, the more they show signs of evolution and recent volcanic activity. In the illustration the rocky composition of the bodies means they lie near the bottom of the cube; the correlation

between their size and their thermal evolution means they lie near a plane that passes through the cube diagonally from left to right. The correlation between size and evolution implies they derive their internal heat primarily from the decay of radioactive atomic nuclei. The colored spheres represent bodies that can be placed in the cube as a result of the study of their solid surface in Voyager images. Such bodies occupy a large part of the volume of the cube. They include Io and Enceladus, two small but highly evolved bodies that are thought to derive their heat from internal friction as each one moves in a slightly elliptical orbit around its parent planet. Io is a rocky moon, Enceladus is icy. Io, Europa, Ganymede and Callisto are the four major moons of Jupiter discovered by Galileo; Ceres and Vesta are bodies in the asteroid belt between the orbits of Mars and Jupiter.

114

caping from the interior might then keep the surface bright and remove the traces of craters. Allan F. Cook II of the Harvard College Observatory and the Smithsonian Astrophysical Observatory and Richard J. Terrile of the Jet Propulsion Laboratory have proposed that if Enceladus is geologically active, gases escaping from the interior of Enceladus and ice particles escaping from the surface might be the source of the E ring's material.

As *Voyager 2* flew close to Enceladus last August it transmitted back to the earth images that showed some craters after all. The craters, however, are sparser than the ones on Mimas, and some of them are in a state of collapse similar to that of old craters on Jupiter's moon Ganymede. The images of Enceladus made at the highest resolution show plains marked by a network of braided ridges but no craters whatever down to the limit of resolution. How old could such a surface be? Shoemaker's calculations suggest that the comets currently traversing the outer solar system should make craters on the plains that are visible at the resolution of the Voyager images in a few hundred million years. That amounts to about 10 percent of the age of the solar system. If other unseen bodies also hit Enceladus, the plains must be even younger.

It seems unlikely that Enceladus has been geologically active for 90 percent of its age but has now become inert; it is therefore highly likely that Enceladus is active today. The E ring and the extremely high albedo of the surface may result from this activity. It should be added that estimates of the effectiveness of tidal heating in Enceladus suggest that the amount of energy supplied by tidal forces is too small and too easily radiated into space for the interior to be a liquid if it consists of water. Perhaps, however, the interior includes methane clathrates or hydrates of ammonia. Such substances have a melting point about 100 degrees Celsius below that of pure water ice.

At least six different types of terrain can be distinguished on the surface of Enceladus. The oldest two are cratered terrains designated $CT1$ and $CT2$. They show a population of craters 10 to 30 kilometers in diameter that almost certainly date back to a period of Population II bombardment in which Enceladus was hit by debris in orbit around Saturn. On the $CT1$ terrains most of the craters have collapsed: their floors have rebounded, their rims have sunk and their central peaks are better described as gentle domes. On the $CT2$ terrains the craters are deep and bowl-shaped. The difference suggests that these different parts of Enceladus have had different thermal histories. The $CP1$ terrains are plains that show a relatively sparse distribution of craters between two and five kilometers in diameter. The craters lie on a rectilinear pattern of faults. Shoemaker's calculations suggest a range of ages. If the flux of recent projectiles has been minimal, the $CP1$ terrains may date back to nearly the time of the Population II bombardment more than four billion years ago. If the flux has been substantially greater, they may be a billion or even two billion years younger. In any event three still more recent terrains ($CP2$, $CP3$ and RP) also formed on Enceladus. They are revealed by the presence of corridors that cut through the older terrains so that parts of craters are obliterated.

The RP terrain includes the plains that show no craters. Its margin is marked by braided ridges. Two tentative hypotheses have been offered for the origin of the ridges. One explanation is that a system of faults was invaded by liquid water that froze. The expansion of the ice is what raised the topography. The other explanation is that a zone of solid-state convection slowly makes matter rise in the center of the plain and sink elsewhere, producing a roughly concentric pattern of wrinkles. Clearly Enceladus joins Io in showing that radioactivity does not always dominate the heating of bodies in the solar system. Enceladus has a hundred-thousandth the mass of the earth. It may nonetheless be just as active geologically.

Different Evolutions

The exploration of the Jovian and the Saturnian systems of moons by the Voyager spacecraft has yielded new insights into the evolution of the small bodies in the solar system. Imagine a three-dimensional chart whose axes are determined by measures of a body's size, composition and thermal evolution. Before the Voyager missions the only bodies that could have appeared in such a chart would have been the planets, moons and asteroids of the inner solar system. All of them fall into essentially the same class of objects of rocky composition, and their degree of thermal evolution is closely related to their size—a correlation suggesting that radioactive heating has dominated their history. In the wake of the Voyager missions a large part of the chart gains occupants, and the simple trend of evolution with size is destroyed.

Much remains to be discovered about the evolution of the moons in the solar system, but already it is apparent that nonradioactive sources of energy (for example tidal heating) are important for some of them. Moreover, it is apparent that an icy composition can allow vigorous geological activity even in very small moons. Further studies of moons will surely tell us much about both the conditions and processes in the incipient solar system and the ways planets evolve under a wide variety of circumstances.

116

The world's leading scientists get it every week and so should you... at ½ price.

Scientists who are leaders in their field and who get to reap the rewards of their achievements are among *Nature's* most eager readers.

And they are renewing their subscriptions to *Nature* in record numbers.

If, like them, you keep searching for ways to speed up your research, or if you're looking for advancement where you work, or are thinking of moving on to something more stimulating, how can you afford to be without your own personal subscription to *Nature?*

It's the most certain and consistent way to get the research results, news, opinion, and analysis you need, while they're still fresh and can be applied most usefully to your own work.

With your own *Nature*, you'll be well ahead of the monthly journals whose reports often lag a year or more behind ours. And there isn't a science weekly anywhere that covers research developments in as many parts of the world, as fully and authoritatively, as *Nature* does.

The good news is that we've cut *Nature's* personal subscription rate in half and, to top it off, if you subscribe for a full year, we'll send you, free, a copy of *Nature's* new, annual *Directory of Biologicals.* Planned for fall 1981, the *Directory* will be the first truly worldwide buyers' guide to bio-

Now: Nature at ½ price – plus a bonus book, free

logical products. It will be available only to yearly subscribers (or, for $45, to buyers of the hardcover version, in bookstores).

A personal subscription to *Nature* has never been more attractive. We urge you to clip and mail the coupon today. For credit-card orders only, **call toll-free now, 24 hrs/day, any day: (800) 824-7888 Ask for Operator 130.**

In California (800) 852-7777.
Ask for operator 130

© 1981 SCIENTIFIC AMERICAN, INC

Supercomputers

The Cray-1 and the CYBER 205 can perform 100 million arithmetic operations per second. This kind of "number crunching" is needed for the solution of complex problems such as those in fluid dynamics

by Ronald D. Levine

The home computers now reaching the market in volume can perform a few hundred arithmetic operations per second. At the other end of the scale are the powerful machines known as supercomputers, of which there are only about three dozen in the world, whose peak computing speeds now exceed 100 million operations per second. The first commercial electronic computer, UNIVAC I, delivered to the Bureau of the Census in 1951, was about three times faster than today's home computers and thousands of times more massive. Since UNIVAC the speed of large-scale scientific computers has doubled on the average every two years. These increases of speed have entailed large increases in memory capacity, necessary for storing data and results. The latest supercomputers are the Cray-1, built by Cray Research, Inc., and the CYBER 205, built by the Control Data Corporation.

Although the current performance levels of such machines owe much to the rapid advance of microelectronics, new concepts in computer architecture have been equally important. The term architecture refers to the logical organization of the computer as it is seen by the programmer. The architectural innovations of greatest significance are those that enable the machine to carry out many similar operations in parallel, or concurrently. Whereas early computers obliged the programmer to break his computational problem down into a sequence of elementary steps, which would then be executed one at a time, the latest supercomputers allow him to specify that many different elementary steps be executed simultaneously.

What kinds of problems demand the prodigious computing power made available by successive generations of high-speed machines and how do these problems lend themselves to parallel processing? A large class of the computationally most intensive problems, a class that embraces such disciplines as aerodynamics, seismology, meteorology and atomic, nuclear and plasma physics, is the numerical simulation of continuous fields. A continuous field can be represented by a collection of physical quantities attached to each point in some region of three-dimensional space. The numerical value of such quantities can vary from point to point and typically change with time in response to some driving force. The quantities may represent, for example, the density, velocity and temperature of a moving gas, the mechanical stress in a solid or a component of an electromagnetic force.

The basic physical laws of continuous mediums are expressed mathematically by systems of partial differential equations that relate (in the language of the infinitesimal calculus) the values and the rates of change of the field variables at each point to their values at nearby points. Subject to the specification of the field values at an initial instant or at a boundary of the region, the mathematician can prove that a unique solution for the system of equations exists. The solution constitutes a complete description of the behavior of the physical system being examined. In all but the simplest physical situations, however, an explicit expression of the exact solution is practically impossible. The mathematician must therefore resort to numerical methods, which, although they are approximate, can always be made as accurate as it is desired by increasing the volume of arithmetic computation.

Since every region of space contains infinitely many points, the complete description of a physical field would call for an infinite amount of data. Hence the first step in devising an approximate numerical method of computation is to retreat to a discrete description of the field by introducing a grid of finitely many points distributed throughout the region. Field values are attached only to the grid points. Mathematically speaking, a system of partial differential equations is replaced by a large system of ordinary algebraic equations. The number of grid points needed depends in part on the number of spatial dimensions in the model, the complexity of the geometry of the region and the amount of spatial detail sought in the result.

Three-dimensional grids with as many as a million points have been needed in order to simulate turbulent aerodynamic flow in fairly simple geometries, and tens of millions of points will be needed in order to obtain solutions sufficiently detailed for engineering purposes in complex geometries such as the region surrounding an entire aircraft. For each grid point several numerical quantities must be stored in the computer memory and periodically updated in the course of the computation: one quantity for each physical field variable, perhaps several quantities to describe the geometry and often several intermediate computed results. The number of stored quantities per grid point in current supercomputer programs for three-dimensional simulations of aerodynamic flow ranges from five to 30.

The field computation proceeds through many iterations of a basic step in which new values for the field variables at every grid point are computed from the old values at that point and at neighboring points. In a typical steady-state problem the iterated steps yield successively better approximations of the exact solution, the computation usually being launched by making some reasonable guess about the initial field conditions at each grid point. The computer then applies the governing differential equations at each step to introduce the corrections necessary to bring the current field description closer and closer to a true solution.

The number of iterations needed for convergence to a sufficiently accurate approximation of the true solution varies strongly with the numerical method and can range from several hundred to several thousand. In simulations where conditions in the field evolve over time, the iterated steps are small time steps. Each step uses the differential equations to project from the current field values their approximate values a short time in the future. The time-step intervals selected for the computation must be shorter than the interval required for the physical field itself to exhibit a change,

118
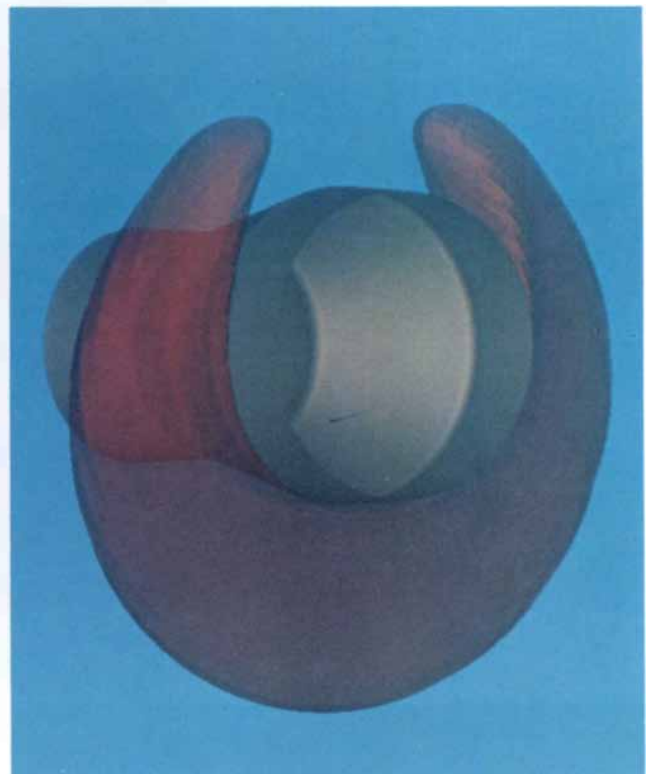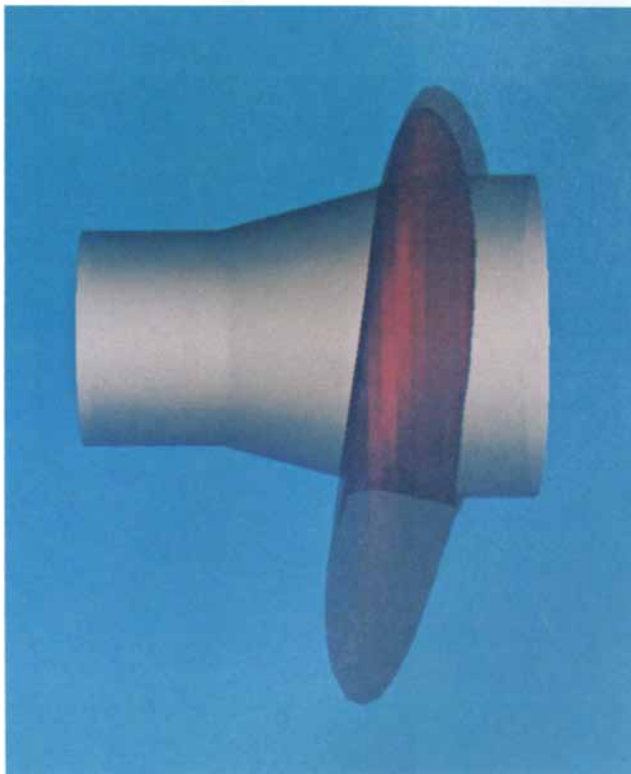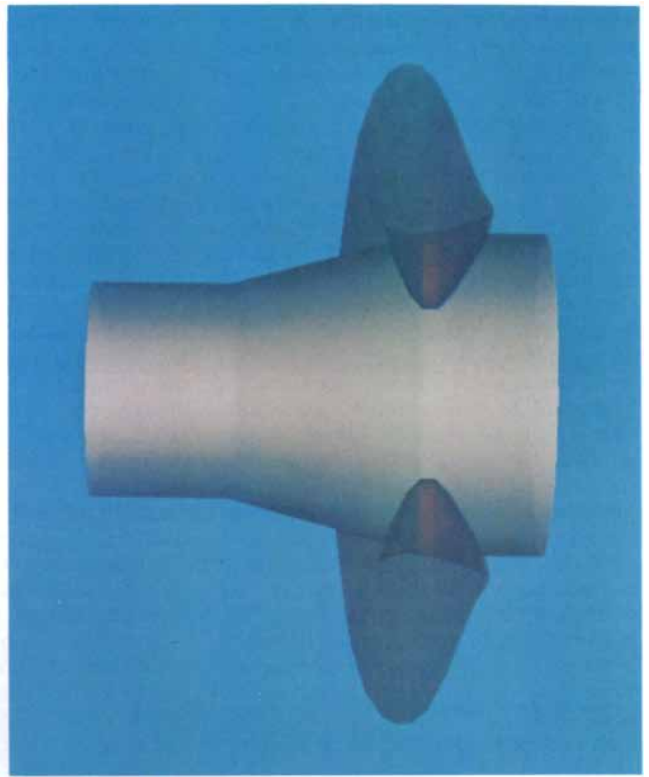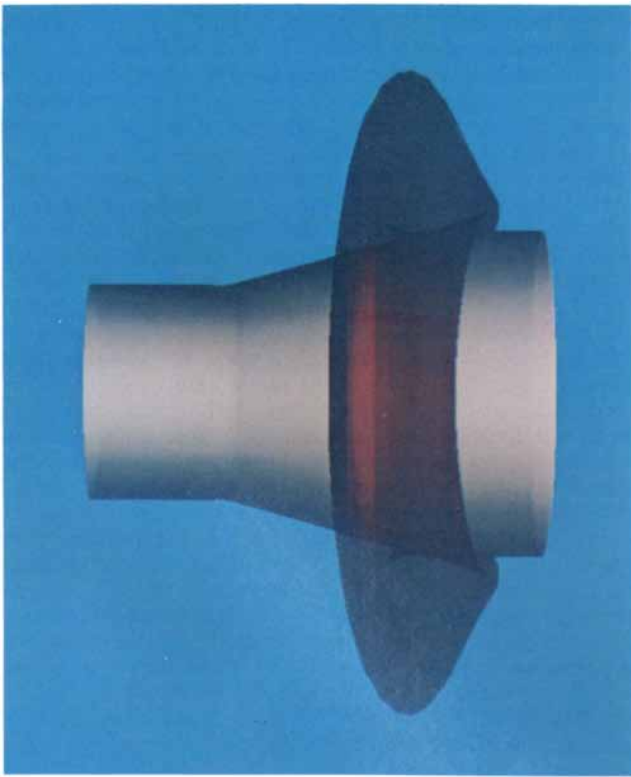
COMPUTER SIMULATION OF FLOW PATTERNS shows the behavior of the air adjacent to the surface of a rocket "afterbody," the region where the main structure of the rocket is tapered to match the smaller diameter of the rocket's propulsion unit. The simulated rocket is "flying" just below the speed of sound. The simulation of aerodynamic behavior in the critical region of the afterbody was one of the last computational tasks of ILLIAC IV before the retirement of this supercomputer last fall from the National Aeronautics and Space Administration's Ames Research Center at Moffett Field, Calif. The computation, which traced the airflow at a network of a quarter of a million points surrounding the afterbody, called for 18 hours of computing time during which ILLIAC IV performed some 10^{11} arithmetic operations. The results of this computation closely duplicate those that can be obtained by placing a model of the rocket in a wind tunnel. The four computer displays depict the "sonic surface," a horseshoe-shaped region where the airflow reaches the speed of sound. The afterbody is shown from below (*top left*), from above (*top right*), from the side (*bottom left*) and turned 45 degrees (*bottom right*). The computation was done under the direction of George Deiwert of the Ames Research Center. The program for the displays was devised by James H. Clark of Stanford University and was implemented by Philip Mast of Silicon Graphics Systems Incorporated and Kelly Luetkemeyer of the Ames Center. Other displays derived from the same computations and similar ones appear on page 132.

119

and preferably they should be much shorter in order to provide suitable numerical accuracy. As the grid points are spaced closer together time intervals must be shortened to reduce the sensitivity to errors that inevitably accumulate from rounding off the number of digits that can be accommodated at each step of the calculation. The overall effect is to increase the number of iterations needed to arrive at a useful answer.

At each iteration the computer may be required to perform anywhere from 10 to 500 operations per data item in order to arrive at a new value for each grid point. In some current applications in aerodynamics and nuclear physics as many as 10^{13} arithmetic operations, absorbing more than two days of computing time, are needed for a single problem. Still bigger problems await more powerful computers. The aeronautical engineer would like to obtain the results of his flow simulations in less than a day, preferably in an hour or so. Other computer users, such as weather forecasters, work under even more imperative time constraints.

In computer simulations of continu-ous fields, and in many other large-scale scientific problems as well, the computational process can be divided into a number of independent parts that can go on concurrently. There are, of course, limits to which concurrency can be pushed. The nth iteration obviously cannot proceed at the same time as iteration $n+1$, since the output results of the first supply the input data for the second. With continuous-field problems, however, the opportunity for concurrent operations follows from the local nature of the differential equations, which endow the grid points with computational homogeneity. In other words, if a particular sequence of operations will describe the change in field quantities that takes place at a given grid point between time t and time $t + 1$, exactly the same sequence of operations can be applied simultaneously to every other grid point, if not throughout the entire grid, then perhaps at all points in a plane cross section. Since the corresponding operations at different points are independent, they can proceed in parallel. Much effort is currently being given to the reexamination of algorithms for large computational problems, aiming to maximize the number of operations that can be carried out concurrently in the new generation of supercomputers and with the help of new programming languages.

Discussions of supercomputer speed often focus on the "megaflop": one million floating-point operations per second. "Floating point" refers to the binary version of the well-known method for representing numbers called scientific notation. In ordinary scientific notation a number is represented as a product in which one factor (the mantissa) has a magnitude between .1 and 1, and the other factor (the characteristic) is an exact power of 10. Thus 6,600 is represented as $.66 \times 10^4$ and 66 is represented as $.66 \times 10^2$. Such a scheme is needed in scientific computation because often the range of magnitudes entering into a problem is very large. The computer relies on a binary version of floating-point notation in which the characteristic is an exact power of 2, and the mantissa and the exponent of the characteristic are expressed in binary notation, consisting of a string of 0's and 1's.

CRAY-1 COMPUTER (*left*), built by Cray Research, Inc., is one of two machines commercially available today capable of a peak computational rate of at least 100 million arithmetic operations per second. The Cray-1 shown here was recently installed at the Ames Research Center, where it has replaced ILLIAC IV. The larger of the two structures in the foreground is the Cray's "main frame," which holds the machine's central processing unit and central memory. The smaller structure is the input/output (I/O) subsystem, made up of three small fast computers and extended memory. The benches at the base of the main frame hold the power supply. The photograph at the right shows part of the maze of wires in the rear of the Cray-1 that interconnect more than 1,600 circuit boards in the machine's central processing unit. The wires, about 300,000 in all, are adjusted in length so that the signals between any two points deviate from the desired travel time by less than one nanosecond (10^{-9} second). The machine goes through one operating cycle every 12.5 nanoseconds.

"The data from this experiment could affect national defense. That's why we use Scotch Brand Disk Cartridges."

IF OTHER MOTORCYCLES WERE BUILT LIKE THIS THEY'D BE EXPENSIVE TOO.

For every motorcycle BMW carefully builds, Suzuki cranks out 45. Yamaha turns out 68. And Honda, a startling 86.

The reason for BMW's strikingly limited production? A passion for extraordinary engineering and hand craftsmanship that high-speed assembly-line techniques simply cannot satisfy.

Each and every BMW frame, for example, is welded entirely by hand instead of machine. It is aligned and scrutinized by a single craftsman and then stamped with his personal seal.

Onto this steady foundation are mounted some of the most precisely tooled components ever to grace a production motorcycle:

A driveshaft, for example, so perfectly formed as to not require a single counterbalancing weight. Cylinders and crankcases that are machined to such infinitesimal tolerances that gaskets are wholly unnecessary.

To insure its durability every critical component on a BMW is actually X-rayed to determine its stress value.

To increase maneuverability these parts are, wherever possible, made of expensive lightweight alloys instead of steel or iron. Culminating in a motorcycle which, model for model, is the lightest in its class.

And not surprisingly, the most expensive in its class. Ranging between $3,600–$7,300, excluding local shipping charges and state taxes.

The rewards? According to Cycle magazine "…high-mileage, years-old BMW's retain their taut, solid feel when lesser motorcycles have long since gotten loose and rattly."

And according to the NADA's motorcycle handbook, BMW's also retain a higher percentage of their original purchase price than any other motorcycle.

Which provides the BMW owner with the singular satisfaction of sitting not only on a superior machine, but on a superior investment as well.

© 1981 BMW of North America, Inc.
The BMW trademark and logo are registered trademarks of Bayerische Motoren Werke, A.G.

THE LEGENDARY MOTORCYCLES OF GERMANY.

BMW

A single floating-point operation is the addition, subtraction, multiplication or division of two floating-point operands to get a floating-point result. Such operations call for somewhat more computational work than the corresponding fixed-point, or integer, operations. In modern computer designs the first three operations take about the same time; division is somewhat slower, but programs can usually be written so that division is fairly rare. In calculating the computing speed in megaflops the time needed to fetch the operands from memory and to return the results to memory should be charged to the operations themselves.

A megaflop rating must be qualified by stating the precision to which the operations are carried out, that is, the maximum number of significant digits, as determined by the length of the machine's data word. Supercomputers are designed to process long data words, usually 64 bits (binary digits). In a 64-bit word 49 bits are typically assigned to the mantissa and its sign (plus or minus) and the remaining 15 bits to the exponent and its sign. This makes it possible to represent numbers with magnitudes ranging from $10^{-2,466}$ to $10^{2,466}$ with a precision corresponding to 15 decimal digits.

One may question the need for such great precision, since scientific measurements are almost always less precise by many orders of magnitude. In many algorithms, however, the inevitable small errors entering into each operation as a result of rounding off can accumulate after hundreds of thousands of iterations to an unacceptably large error in the final result unless very high precision is maintained in the individual operations. On the other hand, many algorithms are stable against rounding-off errors and can make do with less precision. For this reason a desirable feature in a general-purpose supercomputer is the option of a half-precision mode, so that some problems can be run in that mode at twice the megaflop rate and with twice the available working storage.

One cannot assign a speed or a megaflop rating to a particular computer without taking account of the program it is asked to execute. Speed therefore depends on the amount of concurrency inherent in the problem, on how well the inherent concurrency fits the parallel computing capacity designed into the machine and particularly on the skill of the programmer. Performance levels above 100 megaflops quoted by manufacturers of supercomputers are always peak values, attainable only in short bursts and under conditions that arise only occasionally in actual practice. The average megaflop rate that can be sustained over the course of execution of a typical program invariably falls below


CYBER 205, roughly competitive with the Cray-1, is built by the Control Data Corporation. The unit shown here is undergoing final testing at the company's plant in Arden Hills, Minn. The CYBER 205 and Cray-1 are each priced between $10 million and $15 million, depending on the amount of fast memory desired and various other options. The first CYBER 205 was delivered last spring to the Meteorological Service of the U.K. Numerical weather forecasting has provided one of the principal incentives for developing supercomputers, beginning with MANIAC I (Mathematical Analyzer, Numerical Integrator and Computer), designed in the late 1940's by John von Neumann and his colleagues. The CYBER 205, which evolved from Control Data's STAR 100, has a cycle time of 20 nanoseconds. Because the Cray-1 is highly compact, signals can be carried from point to point in it at the velocity attainable with ordinary copper wire: about three-tenths the speed of light. In order to minimize the signal travel time in the considerably larger CYBER 205 its designers have linked its several thousand circuit boards with coaxial cables in which signals travel at about nine-tenths the speed of light.

the quoted peak values and is highly dependent on the program.

For a machine to qualify as a supercomputer in 1982, according to one fairly widely accepted standard, it should be able to sustain average rates of 20 megaflops for a range of typical iterative problems that have a data base of a million or more words. The most successful large-scale scientific computer of the 1970's, the Control Data CDC 7600, can achieve at most five megaflops and that only on much smaller problems. The first computer capable of 20 megaflops or better was the one-of-a-kind ILLIAC IV, which was designed by a group under D. L. Slotnick at the University of Illinois, financed by the Advanced Research Projects Agency (ARPA) of the Department of Defense and built in the late 1960's and early 1970's by the Burroughs Corporation. ILLIAC IV was in-

stalled in 1972 at the Ames Research Center of the National Aeronautics and Space Administration, where it successfully handled some of the largest aerodynamic problems ever programmed until it was finally retired in September.

With the dismantling of ILLIAC IV only two families of supercomputers are fully operational and offered for sale. One is the Cray-1, built by Cray Research; the other is the CYBER 205, built by Control Data. The first Cray-1 was installed at the Los Alamos Scientific Laboratory in 1976. Thirty-four Cray-1's are now in operation at various centers. The first full-fledged CYBER 205 was delivered to the Meteorological Service of the U.K. last spring. Several more have been bought by government and university laboratories and by two oil companies. The CYBER 205 represents an evolution of Control Data's STAR 100

123
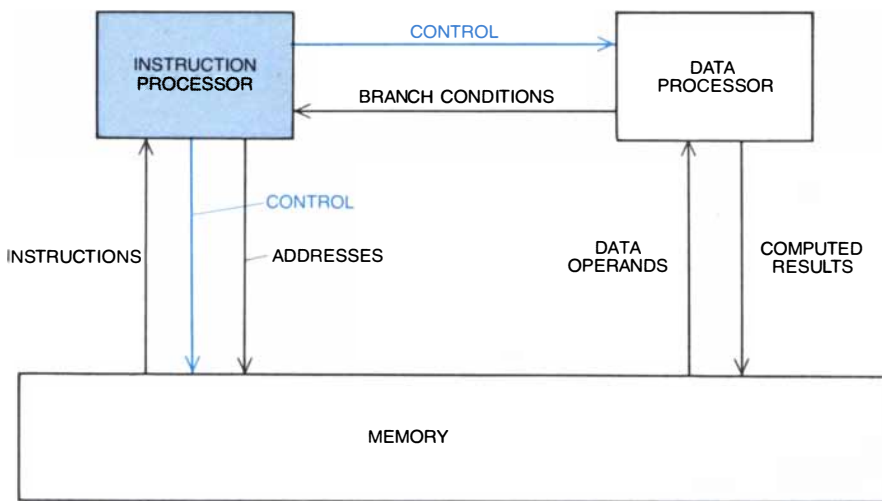
BASIC ELEMENTS OF A COMPUTER are two processors that actively carry out arithmetic and logic operations, and a memory unit that passively stores instructions and the results of continuing computations. At the outset the memory is loaded with the detailed instructions constituting the program to be executed. The instruction processor fetches the instructions in programmed sequence and passes them along to the data processor. Results are stored in memory and recalled as they are needed for further manipulation until the program is executed.

computer, of which four were built in the early 1970's. Three units of a transitional model, the CYBER 203, were delivered last year before that model was superseded by the 205. The 203's will probably be upgraded to 205's, and so the differences between them need not concern us.

Neither the Cray-1 nor the CYBER 205 is indisputably faster than ILLIAC IV. Each of the three machines is (or was) superior for a particular class of scientific problems. Since ILLIAC IV was a one-of-a-kind effort, the cost of its development and construction was high, about $40 million, making it the costliest of the three machines by far. Maintenance and operating costs were also high, about $2 million per year. The purchase prices of the Cray and Control Data supercomputers are between $10 million and $15 million, depending on the memory, peripheral equipment and other options. Cray and Control Data are competing for a world supercomputer market estimated at 100 to 200 machines over the next five years. Moreover, both companies plan to introduce new machines with enhanced capabilities within the same period. By 1985, if not before, several government and company laboratories in the U.S. and Japan may also demonstrate supercomputers capable of outperforming the current generation of Cray and Control Data machines.

The doubling and redoubling of computer speeds in recent years has been made possible in large part by the steady decrease in size of microelectronic circuits. The number of transistors that could be fabricated on a chip of silicon measuring a fraction of an inch on a side increased from a dozen or so in the

early 1960's to several thousand in the early 1970's and has now reached several hundred thousand. These strides in semiconductor technology made it possible for the first time to build large, fast memories at tolerable cost. ILLIAC IV was the first large-scale computer in which the fast central memory was stored on chips rather than on ferrite cores strung on a square matrix of fine wires. ILLIAC's memory chips were at a level of density now called medium-scale integration (MSI): 256 bits per chip. The total capacity of the machine's semiconductor memory was small by today's standards: 131,072 words, each word consisting of 64 bits. The central memories of the Cray-1 and the CYBER 205 exploit large-scale integration (LSI), which provides 4,096 bits per chip. Both machines are available with up to 4,194,304 64-bit words of fast memory. The memories are implemented in the fast bipolar silicon technology known as emitter-coupled logic, at a cost of about $1 per word.

A second benefit of increased circuit density has been a reduction in the time needed for each cycle of logic operations. Digital computers are discrete-time devices: one or more basic clocks emit pulses at the fixed interval called the cycle time or clock period. The state of the machine, defined as the instantaneous content of all its registers, or primary memory cells, is unambiguously determined at each tick of the clock. The change of state between ticks depends on the inputs to the machine. Digitally speaking, nothing happens in less than one clock period, and the computer designer tries to make the clock period as short as possible.

One limit on the clock rate is the response time of the gates, or elementary

switches, that make up a computer. For many years the major limiting factor on computer cycle times was imposed by gate delays. In the primordial vacuum-tube technology cycle times were about 10,000 nanoseconds (10^{-5} second). With the advent of transistors the cycle times were reduced to about 100 nanoseconds. Integrated-circuit technologies have further cut the cycle times to about one nanosecond (10^{-9} second). Times much shorter than one nanosecond can be achieved with technologies now under development, but they remain to be applied in full-scale computers.

When the gate delays are reduced to one nanosecond, the attainable clock rate has a new limit: the speed with which signals can be propagated from one part of the machine to another. No signal can travel faster than the speed of light, which is close to a foot per nanosecond. In practice the speed of pulses through the wiring of supercomputers ranges from .3 to .9 foot per nanosecond. In order to reduce the cycle time of a computer to one nanosecond it would therefore be necessary to limit the distance separating synchronous parts to appreciably less than a foot. Although the increasing density of integrated circuits has led to some amazingly small devices, such as microprocessors (virtually complete computers on a single chip), the miniaturization of supercomputers is inhibited by the vast amount of memory and logic circuitry they must have. The cycle time of ILLIAC IV was 80 nanoseconds; the cycle times of the CYBER 205 and the Cray-1 are respectively 20 and 12.5 nanoseconds. The cycle times of the three machines are roughly proportional to their linear dimensions, evidence that signal-propagation speed is the limiting factor.

Concentrating a large amount of circuitry into a small volume to minimize the length of wires creates a serious problem: the removal of the waste heat generated by electrical energy conversions. The power consumption of fast bipolar logic elements is appreciable, about five watts per chip, and the Cray-1 packs about 300,000 chips into a volume of less than 100 cubic feet. Heat removal is critical because semiconductor failure rates increase rapidly with temperature. ILLIAC IV was cooled by large volumes of refrigerated air blown through its chassis. In the Cray-1 and the CYBER 205 heat is removed by compressed Freon refrigerant carried through cooling channels. In the Cray the individual circuit boards, each six inches by nine inches and bearing dozens of chip packages, are bonded to solid sheets of copper whose edges on two sides slide tightly into massive vertical slabs of aluminum through which the coolant circulates. In the CYBER 205 the coolant tubes are embedded directly in the printed circuit boards, and each chip

package is held firmly in contact with a coolant tube by means of a spring clip.

Digital circuits are subject to errors caused by the failure of components and by random electrical noise. Even though chips have become increasingly reliable, error rates grow as supercomputers evolve because the error rates of a complex system are at least proportional to the number of its parts. In order to protect the user from flawed results supercomputers incorporate mathematical codes capable of detecting and correcting errors. This is done by adding redundant bits of information to the minimum word length required. Mathematical techniques originally devised by Richard W. Hamming of Bell Laboratories in 1950 showed how the redundant bits could be added efficiently. For example, by adding eight bits to a 64-bit word it is possible to detect all two-bit errors and to detect and correct all one-bit errors. Circuitry for generating and checking the redundancy bits must be included at several error-control points in the machine. All supercomputers of the current generation incorporate such error-control schemes. Their absence in ILLIAC IV, together with its more error-prone hardware, was a continuing nuisance. To be sure, error control adds appreciably to machine cost and does not eliminate the possibility of undetected errors (errors of three bits or more), although it does reduce their rate to acceptable levels.

To set the stage for a discussion of computer architecture let me briefly review the basic functions of a computer. A computer has essentially only three parts: a memory, an instruction processor and a data processor. (For the sake of simplicity I omit the input/output system that provides communication with the outside world.) Memory holds both data and instructions; memory locations are referenced by unique addresses. A datum is anything that can be an operand, that is, anything that can be manipulated mathematically, and it includes previously computed results. Instructions are coded patterns of bits that specify the operations to be done and carry addressing information for gaining access to memory. Instructions are organized into programs, often called codes.

The instruction set, or set of defined hardware operations, is the essence of the architecture. The instruction processor fetches instructions from memory, decodes them and issues the correct sequences of control signals and addresses to the data processor and to memory, causing the encoded operations to be carried out. Most often the next instruction to be processed is the one following the current instruction in memory. Few computer programs, however, are simply linear sequences of instructions. Rather, they have loops (sequences of instructions to be repeated many times) and conditional branch points, where the next instruction to be executed is to be selected from two or more possibilities, depending on a computed result the programmer cannot always predict.

In conventional sequential computers each instruction can refer only to a small number of operands. Indeed, the commonest architectures allow only one memory address per instruction. Fetching an instruction from memory and decoding it in the instruction processor takes several machine cycles. The execution phase of an instruction takes several cycles for the memory reference and several more cycles for the arithmetic operation. In early computers the instruction processor was idle during the execution phase of the instruction. It was soon noted that some cycles could be saved by putting the instruction processor to work on fetching the next instruction while the data processor worked on the current instruction, provided the two processors did not come in conflict over access to memory. Such overlapping of fetching and execution was a forerunner of the elaborate "pipeline" concepts that now dominate the design of what are called vector computers.

To the computer programmer a vector is just an ordered list of data items whose elements must be stored in memory in some regular way. The number of elements on the list is called the vector length (a definition that differs from the more familiar one of vector length in mathematics and physics). For any operation that can be applied to a single operand (such as the extraction of the square root) there is a corresponding vector operation that consists of applying the same operation to every element of a vector. By the same token, for any operation that can be applied to a pair of operands there is a corresponding vector operation that applies the same operation to pairs of elements taken in order from two vectors of equal length. To distinguish them from vector operations ordinary arithmetic operations applied to single numbers or pairs of numbers are sometimes called scalar.

Vector computers, including all the supercomputers, rely on several strategies to speed up the execution of vector operations. One strategy is the inclusion of vector instructions in the instruction set. The issuance of a single vector instruction results in the execution of all the pairwise operations that make up the total vector operation. Thus in addition to the operation to be performed a vector instruction specifies the starting addresses of the two operand vectors and the result vector and their common length. More signif-
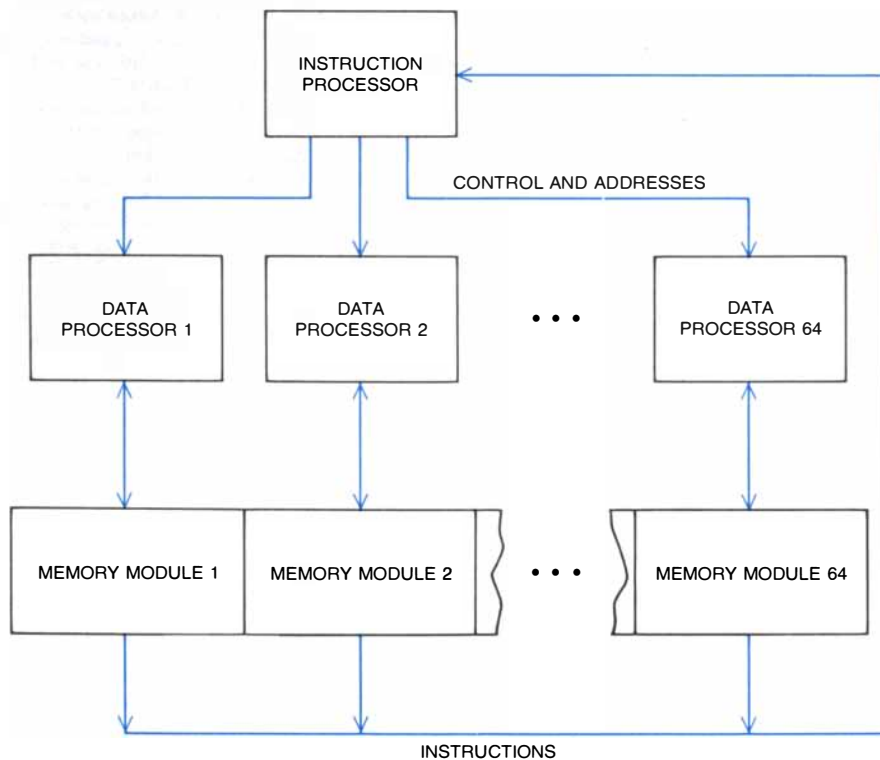
ORGANIZATION OF ILLIAC IV, a one-of-a-kind supercomputer designed at the Center for Advanced Computation at the University of Illinois in the late 1960's, placed 64 independent data processors, each with its own memory module, under the control of a single instruction processor. The machine could therefore perform as many as 64 related computations simultaneously. There is a large class of problems in meteorology, aerodynamics and other disciplines that can be solved in such an all-at-once fashion. For the largest problems ILLIAC IV is unsurpassed. It was the first large computer to exploit semiconductor memory circuits.

125
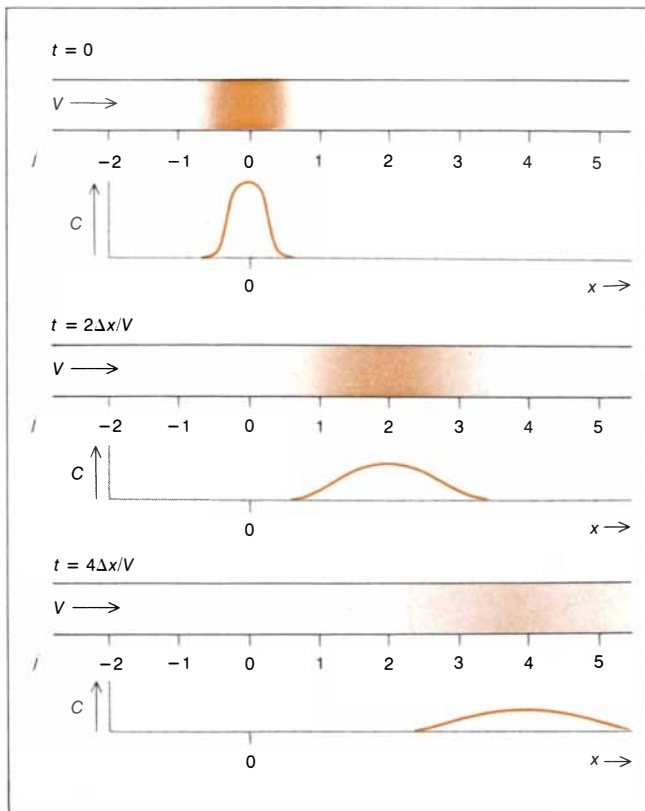
PROBLEM AMENABLE TO PARALLEL PROCESSING in a supercomputer is illustrated by a fluid-flow problem (*left*): in a long, narrow channel or pipe a fluid flowing to the right with a uniform velocity V contains a localized blob of contaminating fluid (*color*). The contaminant is carried along by the ambient flow (convection) and also spreads out because of random molecular motion (diffusion). Below each drawing of the channel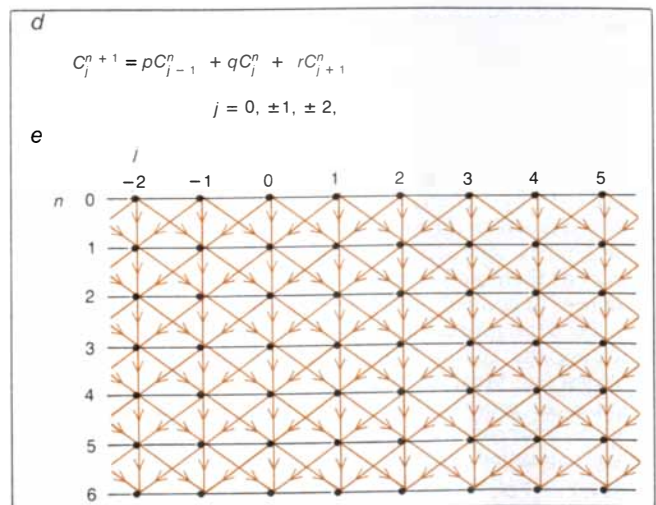 a graph shows the concentration C of the contaminant as a function of distance x along the channel. The mathematical problem is to write equations that will give the value of C at every position x along the channel for every time t. The grid points, labeled j and separated by a distance Δx, serve for the discrete description of the problem for a computer solution. The value Δx must be small enough to yield the desired spatial resolution. The dynamics of the fluid-flow problem can be described by a partial differential equation known as the transport equation (a, *at right*). C, the concentration of the contaminant, varies from point to point and varies with time at any fixed point. The first term in the equation gives the rate of change with respect to time. The second term gives the effect of convection; V is the velocity of the stream and the partial derivative is the gradient of the concentration, that is, its rate of change with respect to its position along the pipe. The third term gives the effect of diffusion. D is a constant that reflects the diffusive properties of the two fluids; the second partial derivative is the gradient of the gradient of the concentration. In b the continuous problem is replaced with a discrete approximation. A small time-step interval Δt is chosen and the concentration is considered only at times $t = 0$, Δt, $2\Delta t,\ldots$, and only at the points labeled j. The term C_j^n denotes the concentration at grid point j at time step n. In order to make the problem finite for the computer left and right limits to j must be chosen. Here it suffices to carry the computation into the region where the concentration effectively vanishes. The number of computations that are needed to follow the system over the time period of interest depends on the value of Δt. Accuracy requires that Δt be small at the cost of lengthening the computation. In c each partial derivative is replaced by its "best estimate" as it will be calculated for specific points of the space-time grid. In each case the best estimate is a ratio whose denominator is the space or time grid-step interval and whose numerator is a difference of the values that have been taken at adjacent grid points. In particular, for the second derivative the numerator is a difference between terms that are themselves such ratios.

COMPUTER'S TASK is to solve a large system of linear equations, d, obtained by substituting the discrete approximation of c, given in the illustration above, in the transport equation. After algebraic simplification one obtains an equation for each value of j that yields the concentration at point j and time $n+1$ in terms of the concentration at the preceding time step n at point j and at its two neighbors $j - 1$ and $j + 1$. The letters p, q and r are constants that depend on V, D, Δx and Δt but that are independent of j and n. In order to advance the solution one time step the computer must carry out five floating-point operations per grid point: three multiplications and two additions. ("Floating point" refers to the scientific notation in which numbers are expressed in powers of 10.) A diagram of the space-time grid (e) shows the data dependencies of the grid points. Before the data at a point can be computed input is required on each colored arrow directed toward that point. (The initial concentration at $n = 0$ for each value of j is assumed to be given.) All values of j for each value of n can in principle be computed in parallel, or concurrently, because there are no dependent linkages between points at the same time level. Values for each level can be computed as soon as values for the preceding time step have been obtained. Moreover, since the form of the required computation is the same for all points j, the parallel computation can be programmed for vector processing.

icant, the hardware provides for the simultaneous execution of a large number of the separate pairwise operations that make up the vector operation. These concurrency provisions generally fall into two classes: multiprocessing and pipelining.

ILLIAC IV embodied a primitive form of multiprocessing. The machine had 64 identical data processors that operated in lockstep under the control of a single instruction processor. Memory was partitioned into 64 modules, each module dedicated to a single data processor. The elements of a vector occupied corresponding locations in the different memory modules. Accordingly the optimal vector length for ILLIAC IV was 64, for which vector operations ran about 64 times faster than ordinary scalar operations. Longer vector operations were divided into segments of length 64, which were executed in a program loop as they are in a sequential machine. For shorter vectors (or the remaining piece of a long vector whose length was not divisible by 64) some of the data processors could be turned off. In both cases the megaflop rate was reduced. ILLIAC IV could function as 128 processors when the word length could be cut to 32 bits without loss of satisfactory precision.

Pipelining is analogous to an industrial assembly line where the manufactured product moves through a series of stations. Each station executes one step in the manufacturing process and all stations work simultaneously on different units in different stages of completion. Floating-point addition, for example, calls for several steps that must be carried out in sequence. A pipelined floating-point adder is divided into segments each of which does a portion of the work that takes one clock period. At the end of the clock period each segment passes the result of its work to the next segment and accepts the partial results from the preceding segment. Hence after a start-up time equal in clock periods to the number of segments in the pipeline, the adder can produce one floating-point sum per clock period as long as a new pair of operands is supplied to the first segment every clock period. The architectures of both the Cray-1 and the CYBER 205 are highly pipelined and have admixtures of multiprocessing.

The organization of computer memory has received much attention over the years. There are two general ways of partitioning memory, which can be called vertical and horizontal. The incentive for structuring memory in a vertical, or hierarchical, manner is that fast memories cost more per bit than slower ones. Moreover, the larger the memory is, the longer it takes to access items that have been stored randomly. The processing units in most large computers communicate directly with a small, very fast memory of perhaps several hundred

words, which are accessible in one machine cycle. This memory is frequently called the register set and is often considered part of the processing unit.

At the next level of the hierarchy is the large, fast memory generally designated the central or main memory. Here resides the bulk of the program and the data base during the execution of large iterative problems. Some computer designs interpose a "cache" memory, holding perhaps several thousand words, between the register set and the main memory.

At the lowest level of the hierarchy is a relatively slow but capacious secondary storage, usually in the form of an array of rotating magnetic disks. The Cray-1 and the CYBER 205 both have the same high-performance disk unit, built by Control Data, which holds about 77 million words. Both machines can be coupled to a battery of such units with a total capacity exceeding a billion 64-bit
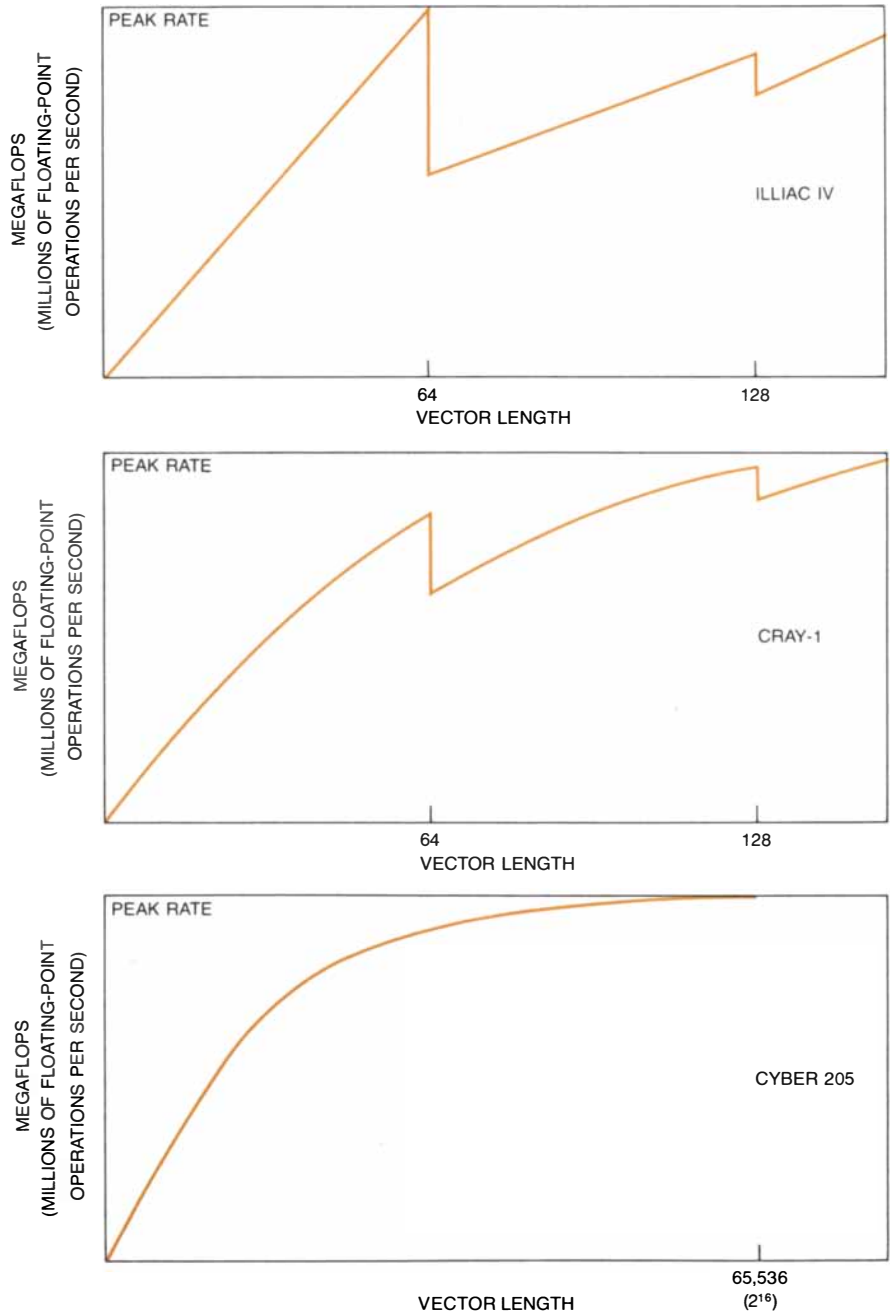
VECTOR PROCESSING is a simple form of computational parallelism available in modern supercomputers. A vector is simply an ordered list of data items that are manipulated pairwise with the items in another list of the same length. The computer hardware provides for the simultaneous execution of long strings of such paired operations. Because of differences in architecture the optimum vector length differs among the three supercomputers (ILLIAC IV, Cray-1 and CYBER 205) represented by these three curves. A "megaflop" is one million floating-point arithmetic operations per second. The peak rate for each machine depends on the particular vector operation involved. It is as high as 400 megaflops for the CYBER 205.

words. Data can be transferred to or from one of these disk units at a maximum rate of half a million words per second, and in practice it is possible to maintain data flow between central memory and several disk units simultaneously.

The maximum rate of transfer of information to or from a memory device is known as the bandwidth. In order for the average computing speed not to be dominated by the smaller bandwidth of the lower memory levels, programs must be arranged so that as much computation as possible is done with instructions and data at the higher levels before the need arises to reload the higher level from one below. As we shall see, this is an important consideration in programming vector operations for the Cray-1, whose central-memory bandwidth is small in relation to the megaflop rate that can be sustained for data held in the register set.

The disk-based secondary storage systems of both the Cray-1 and the CYBER 205 are too slow to allow any feasible solution of continuous-field simulations whose iterative data base is too large to fit in the machines' central memory. IL-

LIAC IV had a unique disk-memory system in which every disk track had its own fixed read/write head. All the disks were synchronized, enabling a good programmer to sustain disk transfer rates of more than seven million 64-bit words per second. As a result ILLIAC IV, in spite of its primitive architecture and small central memory, was the fastest computer yet built for solving the largest problems. The machine routinely solved aerodynamic simulations that had as many as eight million 64-bit words or 16 million 32-bit words in the iterative data base. Both Control Data and Cray Research, and some other manufacturers as well, have recently offered a new level of memory between central memory and disk storage. This level would be implemented with metal-oxide-semiconductor (MOS) technology, and it could hold up to 32 million words at a cost of about 10 cents per word. The first Cray-1 or CYBER 205 so equipped would take from ILLIAC IV its last claim to being "the world's fastest computer."

I have described the horizontal structuring of memory in ILLIAC IV, whose central memory was divided into 64 lockstep modules operating in parallel, each module dedicated to one processor. Several multiprocessing supercomputers currently under development incorporate a refinement of this scheme: a number of independent parallel memory modules are linked to an equal number of independent processors through a high-speed program-controlled switch so that all the memories are equally accessible to all the processors.

For pipeline processors still another kind of horizontal partitioning of central memory has been devised: the memory is divided into a number of "phased" memory banks, so described because they operate with their access cycles out of phase with one another. The rationale for the scheme is that random-access central memories are relatively slow, requiring the passage of a certain minimum number of clock periods between successive memory references. In order to keep vector operands streaming at a rate of one word per clock period to feed a pipeline, vectors are stored with consecutive operands in different banks. The phase shift that "opens" successively referenced banks is equal to one processor clock period. The CYBER 205 has 16 phased banks; the Cray-1 has eight or 16, depending on the memory size. In both machines the bank cycle time is equal to four processor cycles. Thus the Cray-1 central-memory bandwidth is one word per cycle of 12.5 nanoseconds, or 80 million words per second. In the CYBER 205 the phased-bank partitioning is combined with cross-partitioning to form a number of lockstep parallel modules called stacks: eight stacks for a central memory of one million words and 16 stacks for memo-
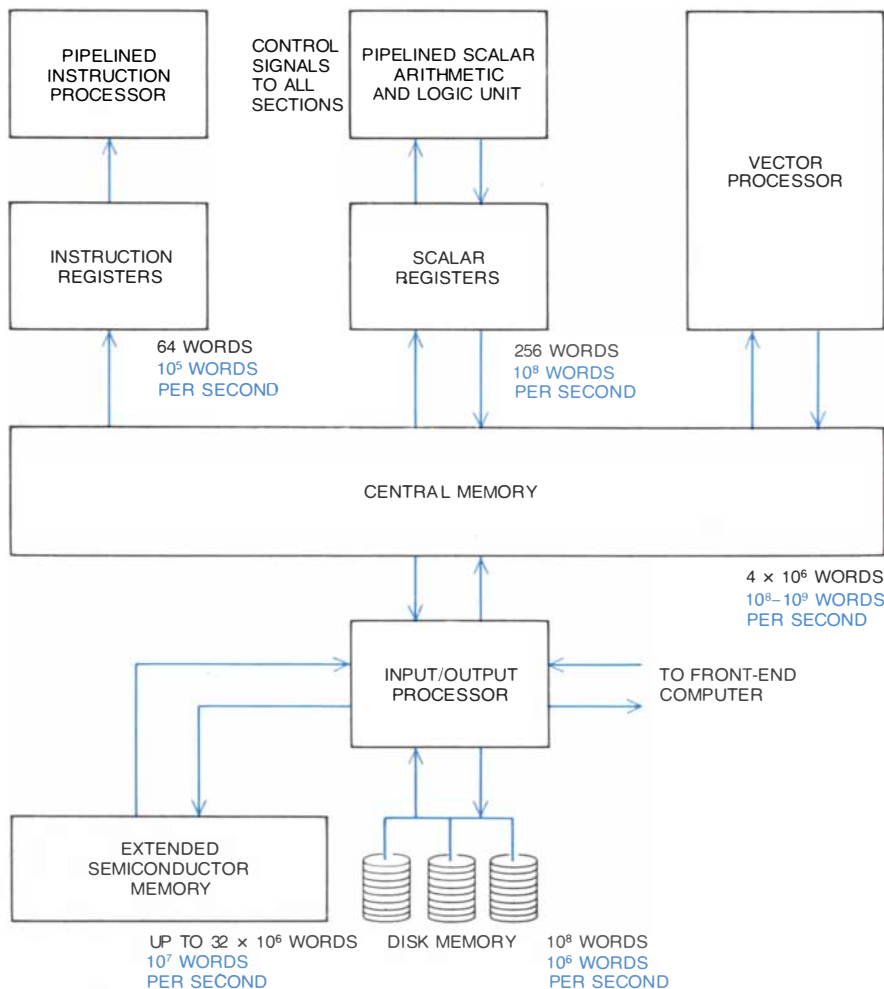
GENERALIZED BLOCK DIAGRAM of a modern vector supercomputer can be regarded as a simplified model of both the Cray-1 and the CYBER 205. Memory is organized hierarchically. The two register memories are the smallest, followed in capacity by central memory, extended semiconductor memory and disk memory. Approximate capacities are shown in black; bandwidth, or transfer rate, is in color. The extended semiconductor memory has just begun to appear in supercomputer installations because rotating-disk technology has not kept pace with the increasing speed of processors. (The largest such memory, with a capacity of eight million words, is in a Cray-1 recently installed at the Exxon Research Laboratories in Houston, Tex.) Here all the functional units are "pipelined," meaning that tasks are broken into elements that can be executed at peak speed and reassembled in a continuous flow, thereby achieving one floating-point operation per clock period: 20 nanoseconds in the CYBER 205, 12.5 nanoseconds in the Cray-1. All the functional units can run concurrently but not all can run at top speed concurrently because they share common resources, such as data paths or memory access cycles. Moreover, conditional branches in the program interrupt the smooth flow of instructions through the instruction processor. Before that processor issues an instruction it must wait until it is clear that all the resources needed for the execution of the instruction will be available when they are needed. In the Cray-1 the register memories incorporate further hierarchical structure, and the vector processor holds additional register memory. The vector processors of the Cray-1 and the CYBER 205 also differ significantly in other respects.

128

"Because great photographic moments only strike once. Olympus OM-2."

Roy Morsch
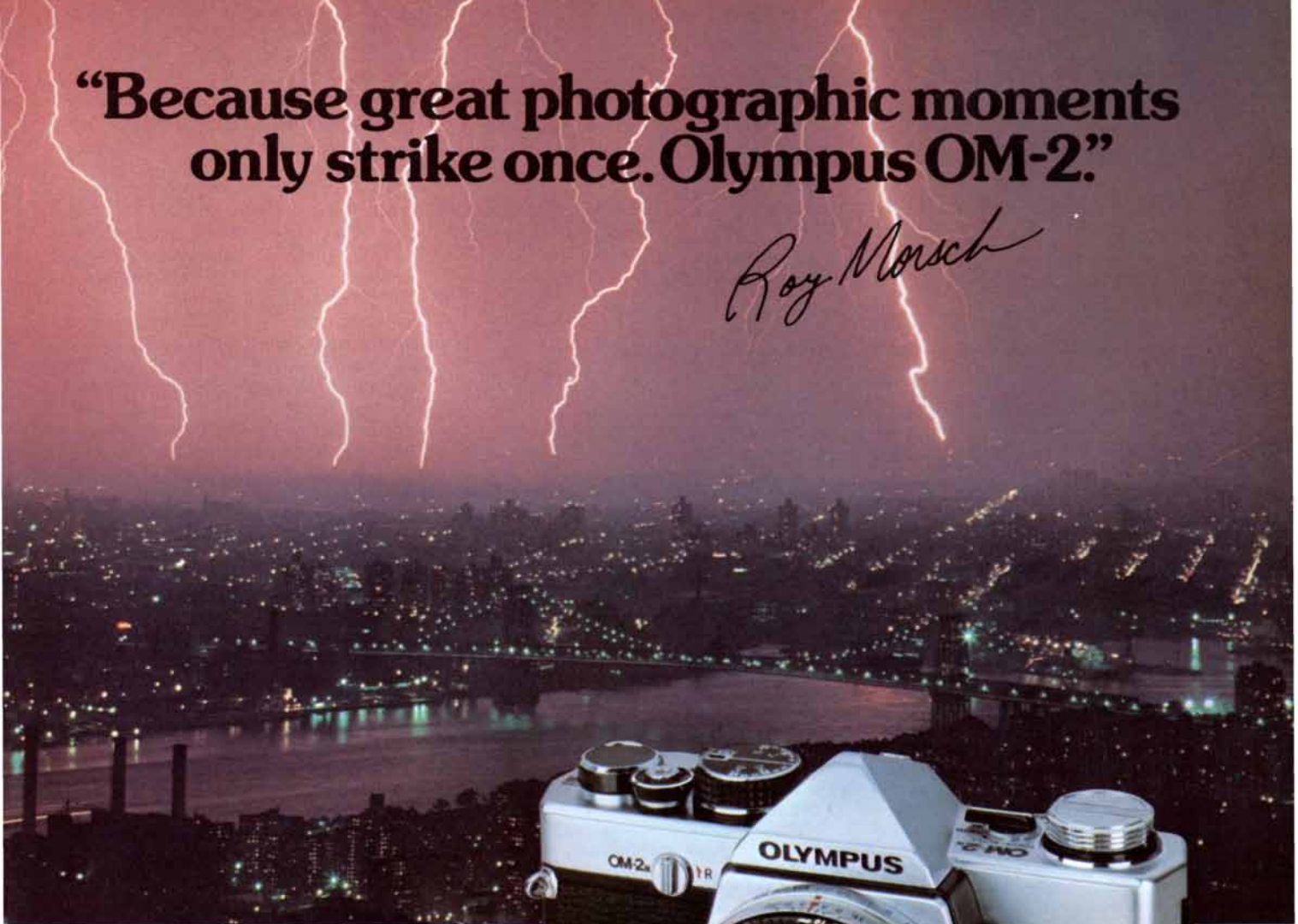
"My OM-2 worked perfectly where most automatic SLR's couldn't work at all," says award-winning commercial and *NY Daily News* professional photographer Roy Morsch. "It automatically made a perfect picture from 5 successive lightning bolts, without guesswork. My OM-2 never misses a shot!"

What makes OM-2 perfect for professionals like Roy Morsch makes it unbeatable for you. Its revolutionary OTF™system sets exposures *as* you shoot, not before. Off-the-film. Automatically fine-tuning changing light in micro seconds. As no ordinary SLR can do.

OM-2 even makes pro-quality flash shots automatically easy. With its OTF™-controlled Olympus thyristor flash units. So you command more creative options than ever before. The OM-2's LED viewfinder ready light even confirms exposure accuracy. Automatically.

And OM-2 is more than an incomparable camera. It's the heart of the incomparable OM system. With 300 superb lenses and accessories to let you do almost anything, anytime. See OM-2 now at your Olympus dealer. You'll see why it's taking the pros by storm…

For information, write Olympus, Woodbury, NY 11797. In Canada: W. Carsen Co. Ltd., Toronto.

OLYMPUS® OM-2
The camera with the press card
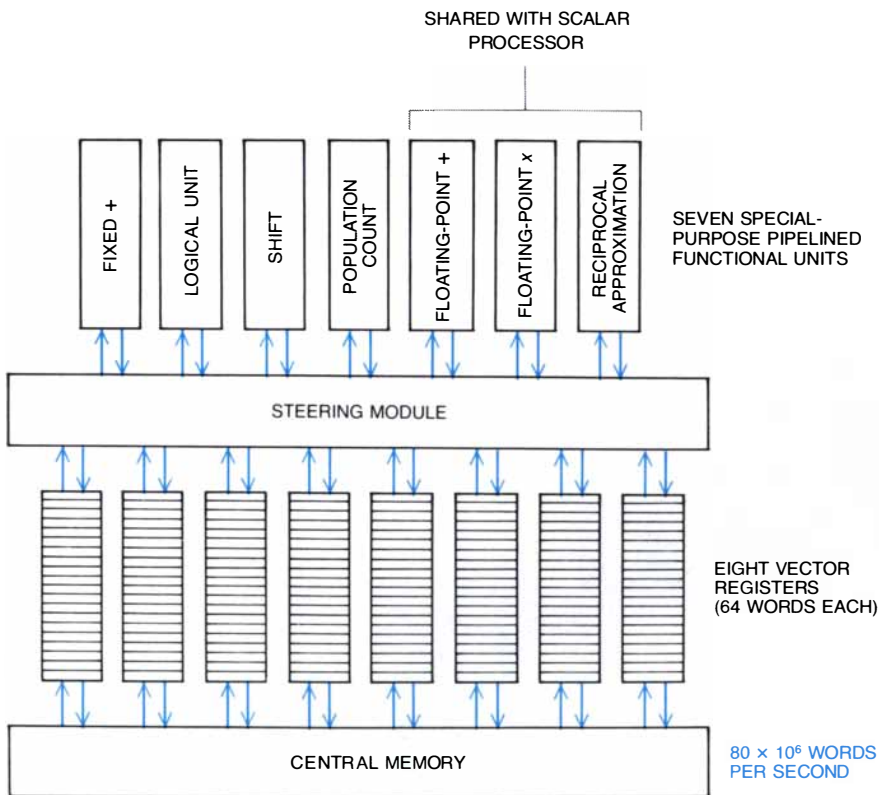
VECTOR PROCESSOR OF CRAY-1 includes seven special-purpose pipelined units for executing specific functions. Three are shared with the Cray's "scalar" processor, which handles nonvectorized operations. (The population-count register counts binary 1's, which is useful in some problems.) Several of the units can work concurrently on different vector operations. Vector data stream from the eight vector registers, through the functional units and back to registers. The steering module switches operands from the registers to the functional units and back again to the registers. While some registers are serving as sources or destinations of vector operations others can be transferring data to or from central memory. As a consequence of register-to-register streaming of vectors pipelines are short and start-up overhead is small.
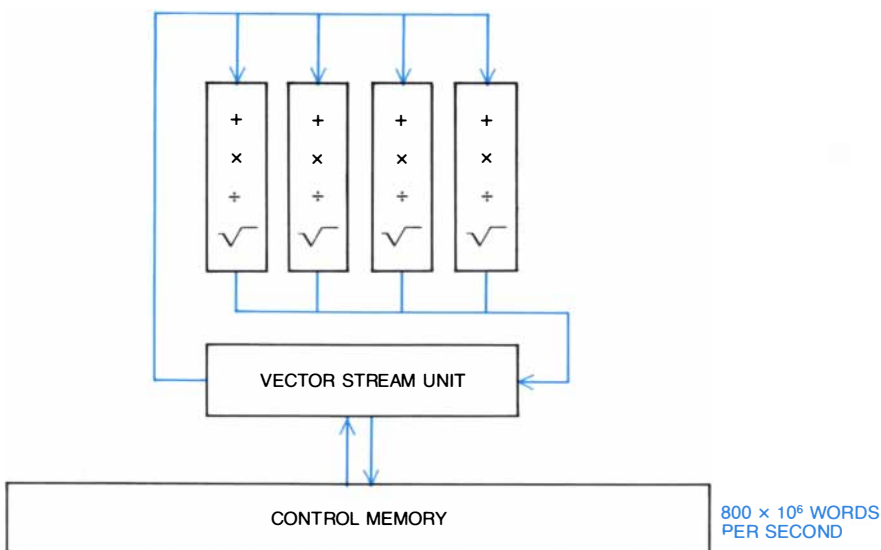

VECTOR PROCESSOR OF CYBER 205 has up to four identical general-purpose floating-point arithmetic pipelines. Each of them can add, multiply, divide and take square roots, but at any given time all of them are performing the same operation on different pairs of elements in one vector operation. The vector-stream unit manages the traffic between the floating-point "pipes" and central memory. Since central memory serves as both the source of vector operands and the destination of the results, the total vector pipeline is long and start-up overhead is large. But since the transfer rate to central memory is much higher than the Cray-1's (800 million words per second), CYBER 205 outpaces Cray-1 when vector length exceeds a certain value.

ries of two or four million words. The maximum bandwidth of a 16-stack CYBER 205 memory is 16 words per clock period of 20 nanoseconds, or 800 million words per second.

The CYBER 205 processing unit has either two or four floating-point arithmetic pipelines (pipes for short). A four-pipe model must have at least two million words of central memory to provide enough bandwidth to keep the pipes full. All the pipes can add, subtract, multiply, divide and extract square roots. For a given vector operation, say addition, all the pipes work simultaneously, with each pipe producing one result every 20 nanoseconds. Vector start-up time is fairly long: about 50 cycles, or one microsecond, between the issuance of an instruction and the appearance of the first result. Therefore peak vector performance can be approached only with long vectors, so that the start-up "overhead" can be amortized over many floating-point operations. In order to realize 50 percent of the peak megaflop rate the vector length must be at least 200. The CYBER 205 offers programmers a useful half-precision feature: each pipe can perform 32-bit operations at twice the 64-bit rate.

The processing unit of the Cray-1 incorporates 13 special-purpose pipelined functional units, all of which can operate in parallel. It also includes a large register set with its own hierarchical structure. Three of the functional units are devoted to floating-point arithmetic: one for addition and subtraction, one for multiplication and one for computing reciprocals. There is an important difference between the Cray-1 and the CYBER 205 in the way vectors are streamed through the functional pipelines. In contrast to the memory-to-memory operation of the CYBER 205's vector pipes the Cray-1's floating-point pipes communicate with a vector register set consisting of eight components, each holding 64 words of 64 bits. One consequence of the register-to-register streaming of vectors is that the curve of efficiency v. vector length shows peaks at vector lengths that are multiples of 64, somewhat analogous to the peaks exhibited by ILLIAC IV. Since the fraction of time lost in pipeline start-up decreases steadily as vectors become longer, however, the Cray-1's efficiency generally increases with vector length.

Another consequence of this scheme is that the pipes of the Cray-1 can be filled and operating in about 50 nanoseconds (four clock periods), or only a twentieth of the time needed by the CYBER 205. The Cray-1's performance for short vectors is therefore superior to the CYBER 205's. The major bottleneck for vector operations in the Cray-1 is the limited transfer rate between the central processor and the main memory. A useful compensating feature is that the
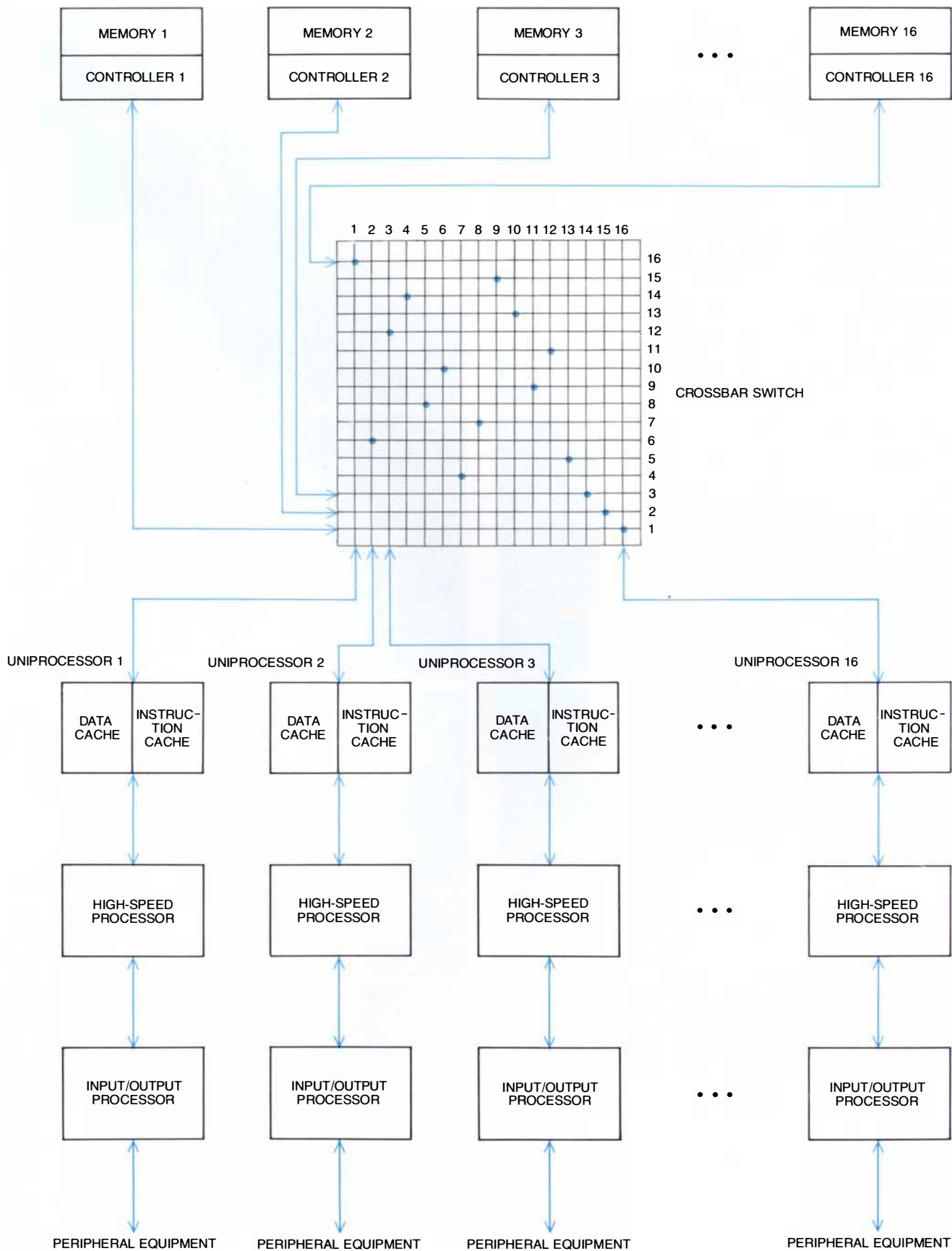
S-1 MULTIPROCESSOR, now under construction for the Navy at the Lawrence Livermore National Laboratory, will coordinate the operations of up to 16 uniprocessors, each one of them a general-purpose computer designed to be as powerful as the Cray-1. A fast-acting crossbar switch will give each processor of the S-1 access to any of 16 large memory banks. Each memory bank is connected to one row of the switch, and each uniprocessor is connected to one column. At any time any row can be connected to any one of the 16 columns. The switching of the connections will be under control of the program. The first of the uniprocessors, Mark IIA, is nearing completion.
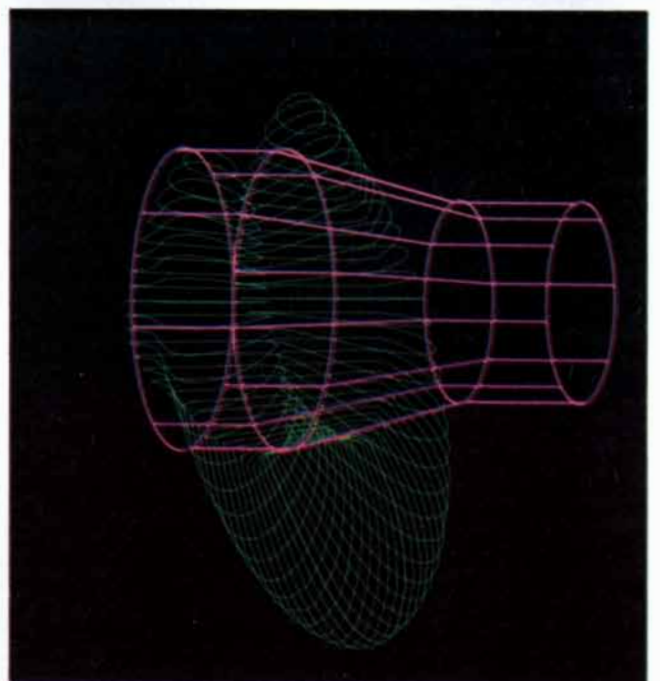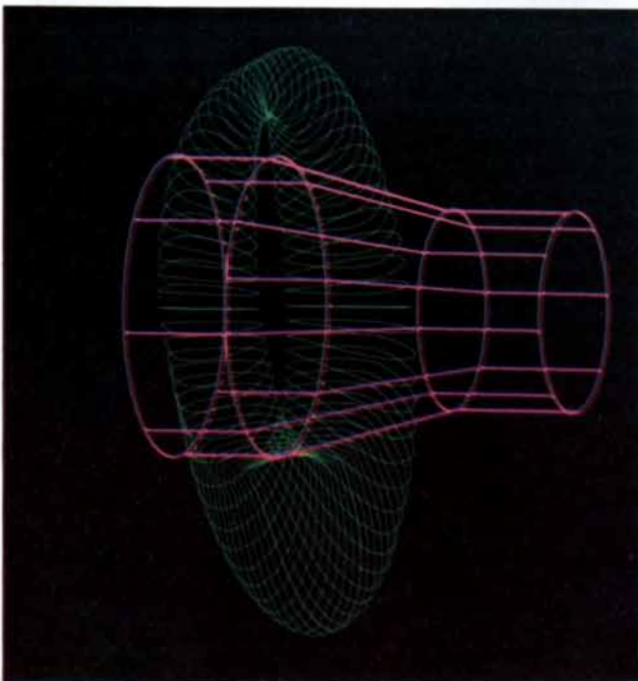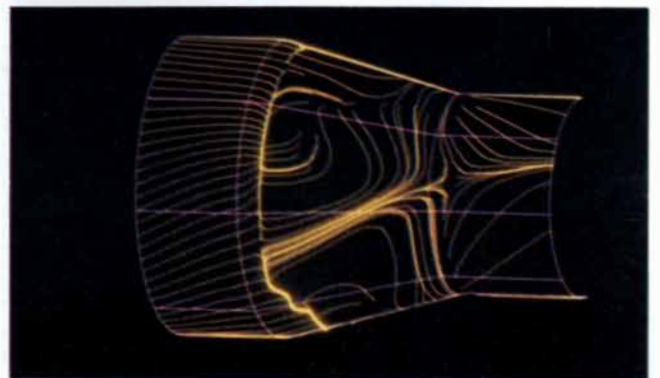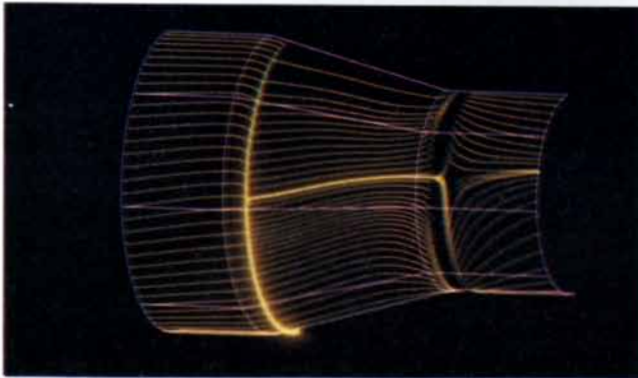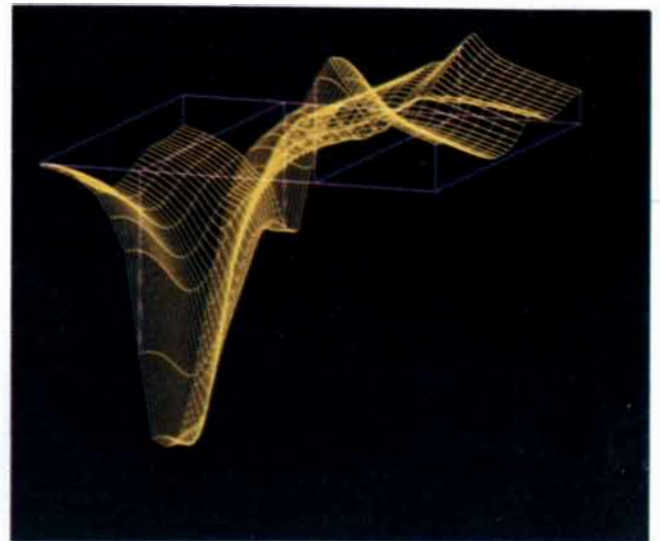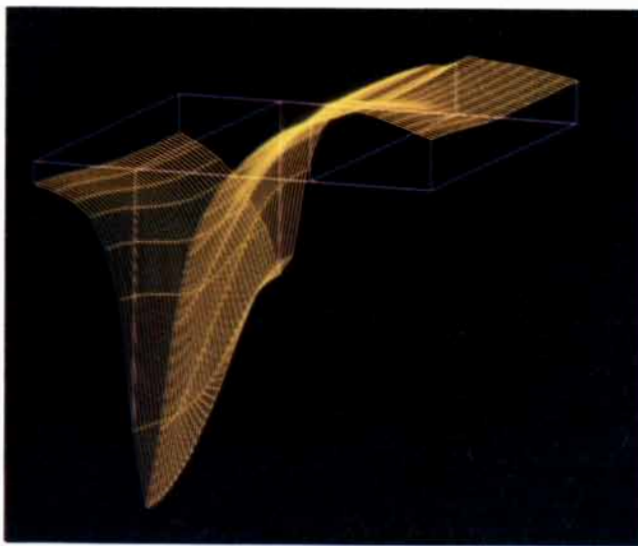
VARIOUS GRAPHIC DISPLAYS can be derived from the computer's solution of a complex problem. The problem in this case is the same one that is represented in the illustration on page 119: the pressure pattern of the airflow around the afterbody of a rocket traveling at a speed just below that of sound. ILLIAC IV computed the flow pattern for several angles of attack, that is, for different angles of inclination between the axis of the rocket body and its direction of flight. The solution for each angle of attack took about 18 hours of computing time. The displays in the two columns, programmed by Adrian Borja, illustrate three ways of presenting a small portion of the computed results for two angles of attack: four degrees (*left*) and 12 degrees (*right*). The veil-like graphs in the top row show the changes in air pressure over the top half of the afterbody. The streak patterns in the middle row depict the surface shear: the flow of air immediately adjacent to the surface of the afterbody. At an angle of attack of 12 degrees the flow actually reverses. The patterns in the bottom row delineate the contours of the horseshoe-shaped sonic surface, where the airflow reaches the speed of sound. The pattern that develops at an angle of attack of 12 degrees is the same one depicted as a continuous transparent surface in the illustration on page 119.

132

Cray-1 can execute a chain of several operations simultaneously on the vector elements while they are in the register set. The optimum utilization of the Cray-1 depends on programs that maximize such chaining. The CYBER 205 provides for a limited form of chaining for triadic operations that involve two vector operands and one scalar operand.

To sum up, the Cray-1, because of its short cycle time, is the fastest computer for problems dominated by scalar and short-vector operations. The CYBER 205, because of its large central-processor-to-memory bandwidth, is the fastest for problems that can be programmed to include long vectors. ILLIAC IV, because of its large-bandwidth link to secondary storage, was faster than any model of the Cray or CYBER yet delivered for problems with the largest data base.

Notwithstanding my emphasis on the value of vector operations in simulating continuous fields, many important large-scale problems cannot be organized into vector form efficiently. Problems that call for much searching and sorting or that are dominated by conditional branching do not vectorize well. Nevertheless, some such problems have other forms of inherent concurrency that can be exploited by parallel modes of computation. Vector computers can execute either scalar or vector instructions, but only in a single sequential stream. They are called single-instruction-stream-multiple-data-stream, or SIMD, computers. (Conventional sequential computers, which lack provisions for concurrent computation, are now sometimes called single-instruction-stream-single-data-stream, or SISD, machines.)

Machines that could handle multiple instruction streams as well as multiple data streams—MIMD machines—have been discussed for many years. Such a machine can be visualized as a collection of general-purpose processors, each processor capable of executing a stream of instructions independently of the others while communicating with the others at some level so that all can work cooperatively on a single problem. Small experimental MIMD machines have been built at several universities. The first superclass MIMD computers may be fully operational later this year.

One obvious way to build an MIMD machine is to link a number of SIMD uniprocessors and a large fast memory by means of a high-speed data-switching network. Alternatively one could implement MIMD processing in a uniprocessor with pipeline architecture by multiplexing, or time-sharing, the hardware resources among instruction streams. An interesting example of such a machine is the Heterogeneous Element Processor (HEP), a multiprocessing system developed by Denelcor, Inc., of Denver, Colo., for the Army Ballistic Research Laboratories. This system's

basic uniprocessor, the processing-element module (PEM), is enabled to carry out multiple processes by time-sharing the hardware assigned to the control and execution of instructions. Each process is periodically given a chance to execute an instruction, and the logic is pipelined so that several processes are in different phases of execution at any one moment. A new instruction can begin execution every clock period, with most instructions being completed in eight clock periods. Although each process can be serviced at most once in eight cycles, the several concurrent processes can be different subtasks of a single program, so that the maximum rate of execution is actually one instruction per cycle.

Synchronization among cooperating processes is aided by hardware "semaphores" attached to memory locations that indicate whether a location is waiting to accept data supplied by a process or whether it contains data waiting to be consumed by a process. Up to 128 processes can be active simultaneously in the PEM; at least eight are needed to utilize the hardware fully and realize the maximum instruction-execution rate. The HEP's circuit chips are fabricated with only medium-scale integration, and so the machine's clock period is a fairly slow 100 nanoseconds, yielding a peak execution speed of 10 million instructions per second. The machine does not provide for vector instructions, so that the sustained megaflop rate would be somewhat smaller than 10.

A HEP system of maximum size would connect as many as 16 PEM's and up to 128 million 64-bit words of memory through a data switch of highly advanced design. The processes making up a single program can be dynamically distributed among the several PEM's, so that the HEP has a desirable fail-safe property: whenever one or more PEM's are down for repair or maintenance, the rest of the system can continue processing the normal work load at a proportionately lower speed. A HEP with 16 PEM's could execute at most 160 million instructions per second, or twice the peak rating of the Cray-1. Because it would cost much more than the Cray-1, however, it would not be cost-competitive for large problems that could be vectorized to suit the Cray's architecture. There nonetheless seems to be no obstacle in principle to implementing the HEP architecture in faster and denser chip technologies to achieve perhaps an eightfold increase in speed and greater cost-effectiveness.

New supercomputers under construction or still on the drawing board show a trend toward increased multiprocessing. Whereas there are fundamental limits to how finely a digital process can be segmented to provide concurrency through pipelining, the multiplication of the computational work force by

the replication of parallel processors is limited only by cost. Most designers agree that a multiprocessor should incorporate a small number of uniprocessors, each very powerful in itself. Although the cost of a multiprocessor system increases linearly with the number of uniprocessors, the speedup in performance increases more slowly because of communications overhead and inevitable inefficiencies in the utilization of the uniprocessors.

Successful competition in supercomputer development will depend heavily on reducing the time needed to design and build a machine in order to capitalize quickly on rapidly advancing technologies. It is widely recognized that computers can be exploited to design better computers. As a first step toward designing an advanced multiprocessing supercomputer for the Navy the group working in the S-1 Project at the Lawrence Livermore National Laboratory developed the system dubbed SCALD (structured computer-aided logic design). SCALD consists of about 60,000 lines of code in the high-level computer language PASCAL. The system enables an engineer to express his design concepts for any level of digital logic in the same abstract graphical language he uses in thinking and in talking with other engineers. The tedious rendering and checking of details is left to the machine.

With the SCALD system a small group of Livermore engineers has created in less than a year a detailed design for a powerful SIMD scientific computer called the Mark IIA. Conceived as a uniprocessor for an eventual multiprocessor, the Mark IIA exhibits a rich and flexible architecture. The goal of the Livermore group is to build a machine that will compare favorably with the Cray-1. It would be difficult, however, to predict the relative performances of the two machines because their memory hierarchies are structured quite differently in terms of their capacity and their bandwith at each memory level. The cost of the Mark IIA will be substantially less than that of a Cray-1 because the Mark IIA will have newer circuitry that is denser and cheaper. The main memory, for example, will be in metal-oxide-semiconductor technology with 65,536 bits per chip, compared with 4,096 bits per chip in the Cray-1. An S-1 multiprocessor would include up to 16 Mark IIA uniprocessors and hundreds of millions of words of shared memory.

For MIMD multiprocessors such as the S-1 and the HEP the speedup attainable from a multiplicity of processors will ultimately depend on the nature of the problem the machine is required to solve. For highly iterative continuous-field problems new algorithms must be developed in order to exploit such machines efficiently. On the other hand, some applications, such as real-time im-

age processing, are naturally amenable to multiprocessing because they can be pipelined in a straightforward manner: while processor No. 1 performs the first stage of data reduction on data frame n, processor No. 2 performs the second stage of computation on frame $n - 1$, processor No. 3 works on the third stage of frame $n - 2$ and so on.

Communications between a computer and the outside world, referred to as input/output (I/O) operations, are very slow in relation to the machine's internal processing speeds because peripheral devices have mechanical components and human response time is slow. To relieve the supercomputer of managing large volumes of input/output traffic and to save its valuable time for "number crunching," most supercomputer installations are equipped with I/O processors and with one or more satellite computers called front-end processors. The latest models of the Cray-1 are available with an array of I/O processors, which are fast computers with short words and a limited instruction set, responsible for managing high-speed data traffic between central memory and the secondary memories, other peripherals and the front-end computers. The front-end computers act as appointment secretaries to the mighty supercomputer. The front-end machines can be fairly powerful scientific computers in their own right, capable of carrying some of the data-processing load, particularly to reduce large volumes of output data to humanly digestible form. For example, the raw output of an aerodynamic-flow simulation can consist of a million numerical quantities, which would overwhelm the engineer if it were not presented in graphic form, perhaps as a motion picture.

Some major centers of scientific computation have developed elaborate data-communications networks linking several supercomputers, front-end processors, other general-purpose computers, mass-storage data banks and graphic-display stations so that each user has rapid access to the most suitable resources for his problem. Although supercomputers are still few in number, they are becoming widely available with the development of national and international high-speed data networks that link the scientific centers.

Supercomputer users demand ever increasing power for attacking the computational problems at the frontiers of their various disciplines. Computer architects have differing ideas for satisfying user demands. It is likely that the coming years will see accelerating development and evaluation of new solutions, spurred by rapid changes in technology, by advanced techniques of computer-aided design and by wide availability of the supermachines for field testing by a critical community of users.

The Development of a Simple Nervous System

Tracing the pedigree of nerve cells in the embryonic growth of dwarf and giant leeches yields preliminary clues to the functioning of the adult nervous system

by Gunther S. Stent and David A. Weisblat

The nervous system presents two of the most challenging questions of contemporary biology. How do networks of neurons, or nerve cells, generate animal behavior? And how do the neurons and their specific connections arise during the development of an animal from a fertilized egg? The second question cannot be considered independently of the first, since the anatomy and the functioning of the adult nervous system represent the end point of neural development. Detailed anatomical and functional knowledge is needed even to ask well-focused questions in developmental neurobiology; to gain such knowledge animals with a simple nervous system are particularly suitable. One such animal is the leech, the blood-sucking relative of the earthworm.

For many people the mere mention of leeches evokes the revulsion expressed by Charlie Allnutt (played by Humphrey Bogart) in *The African Queen:* emerging from a swamp and finding the parasites clinging to his skin, Allnutt shouts, "I *hate* leeches!" Nevertheless, leeches have their good points; unknown to Allnutt, while drawing his blood the leeches would have injected anticoagulant substances into his bloodstream that might have reduced his risk of certain cardiovascular disorders.

More to the point, the leech's simple body plan makes it an attractive animal to the experimental biologist. The tubular body of the leech is built of 32 similar segments and offers the possibility of understanding the entire animal by studying just one of its segments. Of the 32 segments the frontmost four make up the specialized structures of the head, including a pair of eyes on the dorsal, or upper, surface and a front sucker on the ventral, or lower, surface. The rearmost seven segments make up the specialized structures of the tail, including the anus and a large rear sucker. The anatomy of the intervening 21 mid-body segments is highly stereotyped. Each segment has a

complete set of visceral organs, including circulatory vessels, kidneys and gut. The skin of each segment is subdivided into a fixed number of annuli, or rings; the middle annulus bears an array of sensory organs distributed around the circumference of the body tube. The body wall of each segment is girded by circular muscles that can constrict the body tube. Deeper in the wall lie longitudinal muscles; their contraction shortens the body tube.

The nervous system of the leech reflects the segmental body plan. It consists of 32 ganglia, interconnected to form the ventral nerve cord. Each ganglion consists of some 200 bilaterally symmetrical pairs of neurons and a few unpaired neurons. The ganglion is linked to the body wall and the internal organs by two bilateral pairs of segmental nerves and to the neighboring ganglia by bundles of connective nerves. Exhaustive studies of the nervous system of the leech, carried out mainly by John G. Nicholls and his students at Stanford University, have shown that the anatomy of the segmental ganglia is sufficiently stereotyped from segment to segment and sufficiently invariant from specimen to specimen for a large fraction of the neurons to be reliably identified [see "The Nervous System of the Leech," by John G. Nicholls and David Van Essen; SCIENTIFIC AMERICAN, January, 1974].

A neuron of the leech's segmental ganglia can be penetrated with microelectrodes to record the cell's electrical activity. Similarly, a stain or a fluorescent dye can be injected through a micropipette to reveal the anatomical details of an individual neuron. By means of these techniques it has been possible to establish the pattern of neuronal connections and thereby to account for some simple acts of reflexive behavior, such as the shortening of the body tube in response to tactile stimulation, and even for some moderately

complex integrated movements, such as the heartbeat.

The most complex behavior that has been described in terms of identified neurons and their connections is swimming. Work done in our laboratory at the University of California at Berkeley from 1971 to 1977 by William B. Kristan, Jr., Carol Ort, Otto Friesen, Margaret Poon and Ronald Calabrese showed that the contractile rhythm of the longitudinal muscles responsible for the swimming movement of the leech is generated by a set of 12 bilateral pairs of rhythmically active motor neurons in each segmental ganglion. The rhythm of the motor neurons is imposed on them by four bilateral pairs of interneurons (intermediate nerve cells), which form the central swim oscillator.

Since the functional elements of the nervous system of the leech are known in some detail, the system provides the kind of clearly defined conceptual end point needed for asking specific questions about developmental processes. For instance, is the identified swim circuitry organized from the outside in, so that the longitudinal muscles are first connected to the motor neurons, the motor neurons are then connected to the interneurons and the interneurons are finally joined to form the central oscillatory network? Or does development proceed in the reverse, inside-out order? Or are all three levels of connections established concurrently? Indeed, is it possible that the neural circuitry does not develop by establishing specific connections but rather by selecting from an initially overconnected neuronal network those connections that are functionally appropriate?

Seven years ago, in the hope of answering some of these questions, we embarked on a long-term study of the development of the leech's nervous system. Our first task was finding a suitable species that could be cultivated in the

laboratory to provide a constant supply of embryos. To that end our colleague Roy T. Sawyer examined many varieties of leech and found two candidates for our project. One is *Helobdella triserialis,* a native of California that feeds on the body fluids of snails. *Helobdella* grows to an adult length of about two centimeters (less than an inch) and propagates with an egg-to-egg generation time of about eight weeks. The other species is *Haementeria ghilianii,* a native of French Guiana that feeds on the blood of mam-

mals. *Haementeria* reaches a length of about 50 centimeters (more than a foot and a half) and has a generation time of about 10 months.

The short generation time, ease of cultivation and robust embryo of *Helobdella* make it a favorable material for developmental studies, but because of its small size it is less suitable for neurophysiology. In contrast, the enormous size of *Haementeria* makes its nervous system and neuronal precursor cells highly accessible to intracellular electri-

cal recording and other techniques that require the penetration of single cells, but its long generation time, more demanding breeding conditions and less hardy embryo are drawbacks. As it happens, the advantages of both organisms can sometimes be exploited. The two species belong to the same family, and in spite of the disparities in their size and their way of life they are similar in their adult body plan and embryonic development. For many embryological purposes the species can be consid-
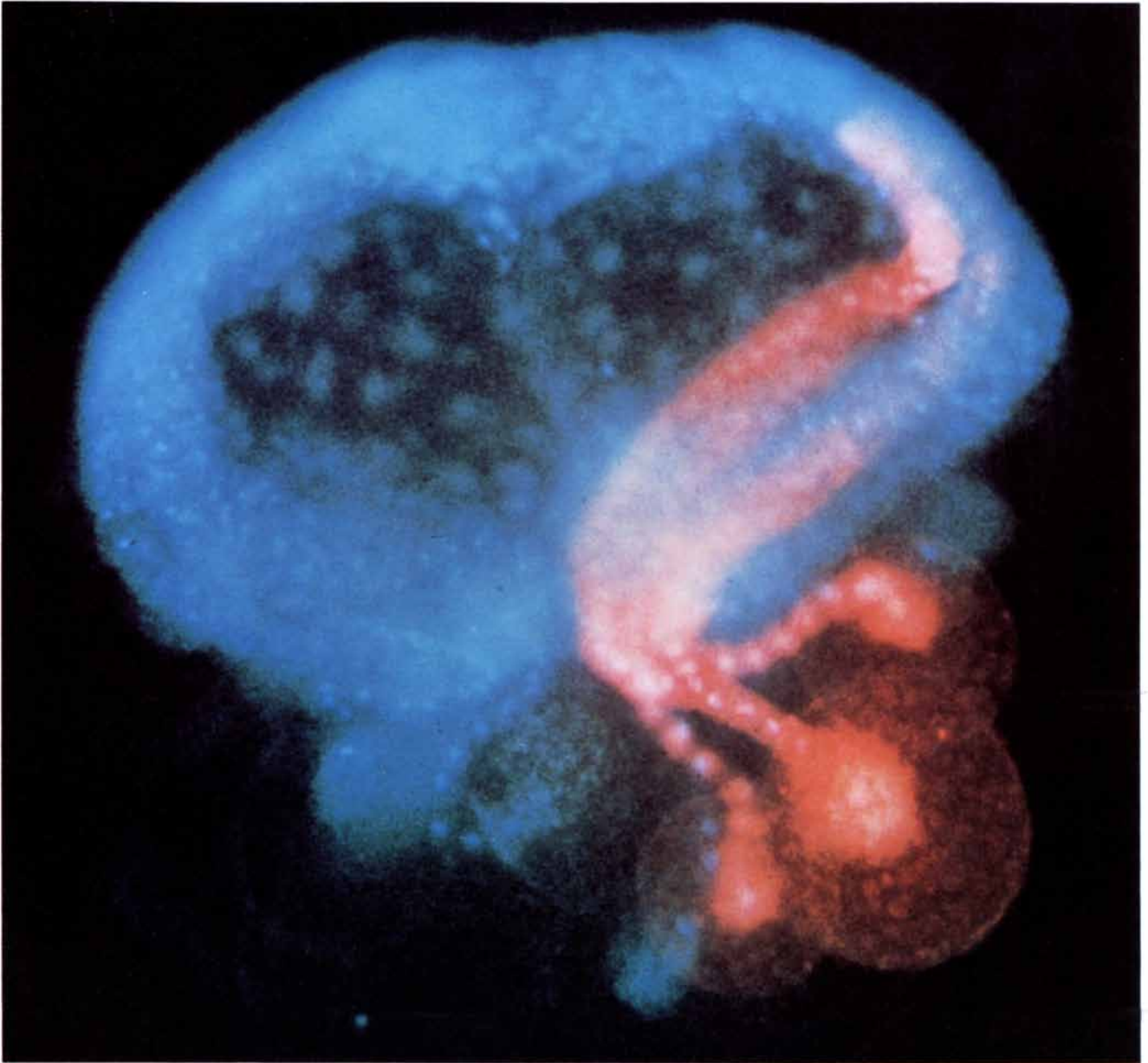
EARLY STAGE in the embryonic development of the dwarf leech *Helobdella triserialis* **is visualized with the aid of a new tracer technique devised by the authors and their colleagues at the University of California at Berkeley. A red-fluorescing dye coupled to a tailor-made polypeptide carrier was injected through a micropipette into a single cell of the leech embryo at an earlier stage of development. The embryo was later fixed, cleared of yolk and treated with a blue-fluorescing stain specific for DNA. It was then photographed through a microscope. The clustered blue dots in the resulting photomicrograph** **are the nuclei of cells that have begun to coalesce to form the two crescent-shaped embryonic structures known as the germinal bands; these cells are particularly concentrated at the anterior, or head, end of the embryo (***top***). The three large, red-stained cells at the bottom right are the immediate descendants of the cell that was injected originally with the red-fluorescing dye. From each of these cells a bandlet of daughter cells, called stem cells, projects into the right germinal band. The nervous system of the adult leech is derived from the stem-cell bandlets. The magnification is approximately 300 diameters.**

137

STAGE 1 — TELOPLASM

STAGE 2 — AB, CD, EGG MEMBRANE

STAGE 3 — A, B, C, D

STAGE 4 — MICROMERES, A, B, DM, C, DNOPQ

STAGE 5 — A, B, M, C, NOPQ, M, NOPQ

STAGE 6 — B, A, C, M, N, N, O, O, P, Q, M, P, Q

STAGE 7 (EARLY) — M, N, N, O, O, P, Q, M, P, Q, STEM-CELL BANDLET

STAGE 7 (MIDDLE) — GERMINAL BANDS, MICROMERE CAP

STAGE 8 (EARLY)

STAGE 8 (MIDDLE) — GERMINAL PLATE, NERVE CORD

STAGE 8 (LATE) — NERVE CORD, LATERAL EDGE OF GERMINAL PLATE

STAGE 9 — EGG MEMBRANE

STAGE 10 — FUSED TAIL GANGLION, ANNULI, FUSED HEAD GANGLION

FIRST 10 STAGES in the embryonic development of *Helobdella* are depicted in this sequence of drawings, beginning with the uncleaved egg (*Stage 1*) and ending just as the body tube is about to close (*Stage 10*). Two key stages (*7, 8*) are broken down into substages. All the drawings in the top two rows are views of the future dorsal, or upper, surface of the embryo. The drawings in the bottom row are lateral views. Early in Stage 4 the cells designated *A, B, C* and *D* sprout four small cells, called micromeres, at the dorsal pole. The subsequent proliferation of these cells leads to the expansion of the micromere cap (*blue area*), which eventually envelops the early embryo. In stages 9 and 10 the cells of the germinal plate (*red area*) proliferate, causing it to expand back over the surface of the embryo. The left and right lateral edges of the germinal plate finally meet at the dorsal midline, closing the body tube. Cells *M, N, O, P* and *Q* are called teloblasts.
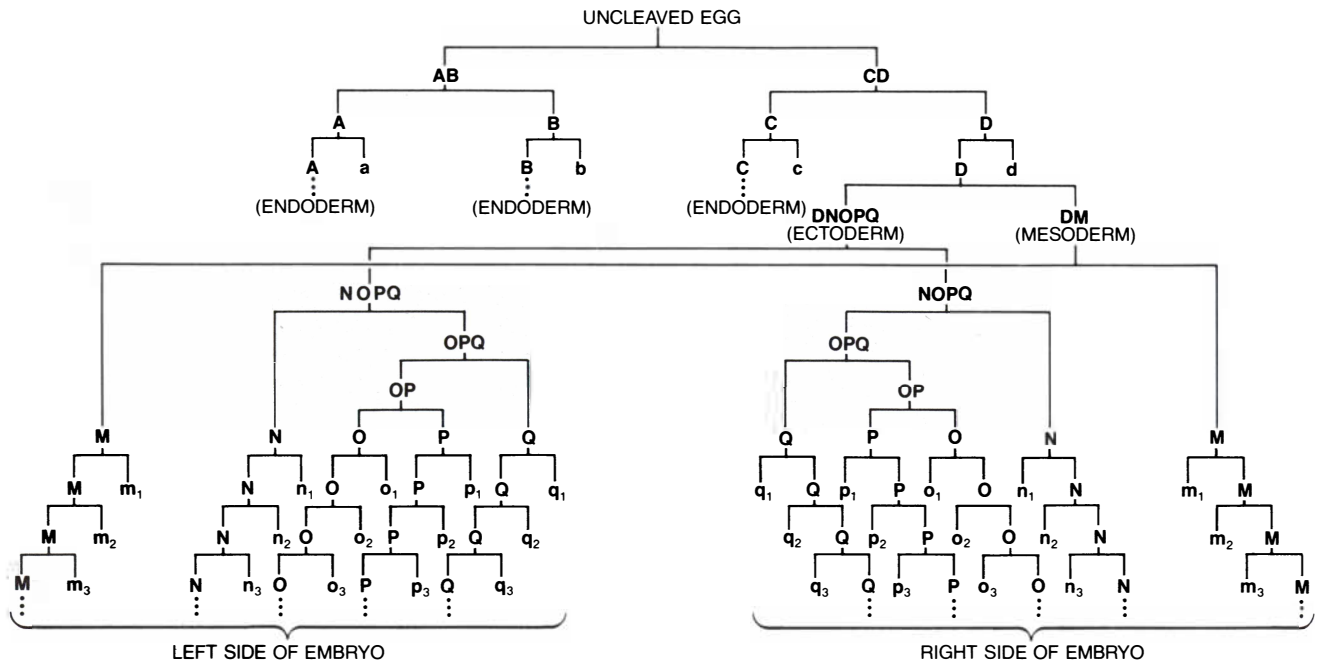
PEDIGREE OF CELLS in the developing *Helobdella* embryo is delineated in this "family tree" diagram. Stem cells are denoted by lowercase letters corresponding to the uppercase letters of the teloblasts from which they are descended. The separation of the embryo into three layers of germinal tissue is indicated. The progeny of cells *A, B* and *C* give rise to the endoderm (the innermost layer), the progeny of *DNOPQ* to the ectoderm (the outermost layer) and the progeny of *DM* to the mesoderm (the middle layer). The lines of cell division responsible for the growth of the micromere cap are omitted. Bilateral symmetry of embryo is established when cells *DM* and *DNOPQ* divide.

138

© 1981 SCIENTIFIC AMERICAN, INC

ered interchangeable, thereby providing greater scope for experimentation than would be offered by either species alone.

Helobdella and *Haementeria* lay yolk-rich eggs measuring respectively about .5 millimeter and 2.5 millimeters in diameter. The eggs are laid in clutches and each clutch is enclosed in a transparent cocoon that remains attached to the ventral body wall of the parent leech. Embryonic development begins as soon as the eggs are laid and lasts for two weeks in *Helobdella* and for a month in *Haementeria*. The egg yolk provides the nutrients needed for embryonic development. Soon after the yolk is exhausted, the juvenile leech, which is similar to the adult in form but smaller, takes its first meal from a host animal. In the course of the subsequent, postembryonic growth and maturation of the juvenile leech its cells increase in size and to a lesser degree in number.

Leeches were a major focus of interest for the 19th-century pioneers of modern experimental embryology. In the 1880's Charles O. Whitman, the first director of the Marine Biological Laboratory in Woods Hole, Mass., described the exact succession of cell divisions by which the leech embryo arises from the fertilized egg; in doing so he presented the first analysis of a developmental cell lineage. On the basis of his observations Whitman put forward the idea that each identified cell of an early embryo, together with the clone of cells descended from it, plays a specific, predestined role in later development. The hypothesis was then quite novel and is still a central issue in discussions of embryonic development.

The interest of embryologists in the leech declined at the turn of the century, after Whitman redirected his attention to animal behavior and laid the groundwork for the discipline of ethology. Meanwhile his student Thomas Hunt Morgan turned to the genetics of the fruit fly *Drosophila melanogaster* and revolutionized the science of heredity. The recent revival of work on the embryology of the leech was initiated by Sawyer and by Juan Fernandez of the University of Chile.

As the fertilized leech egg approaches its first cleavage, or cell division, there becomes visible at each of its two poles a region of colorless cytoplasm, called the teloplasm, that is distinct from the colored yolk filling most of the egg. One of the poles marks the future dorsal surface of the embryo; the other pole marks the future ventral surface. The first cleavage divides the egg into two large cells, designated *AB* and *CD,* with most or all of the teloplasm passing into *CD.* The second cleavage gives rise to four cells, *A, B, C* and *D,* of which cell *D* receives all the teloplasm. Cell *D* then cleaves. One of its daughter cells, designated *DNOPQ,* lies more dorsally and
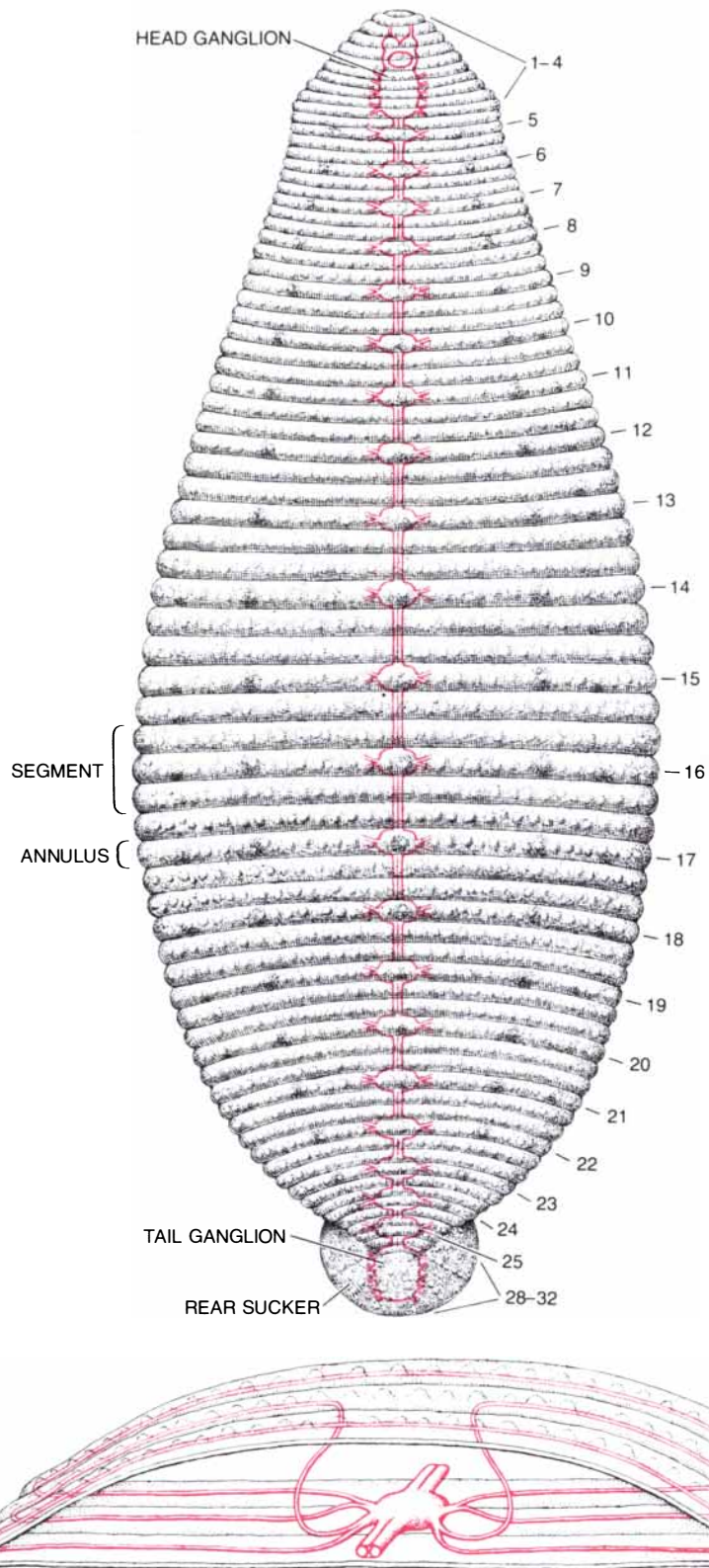
SEGMENTAL BODY PLAN of the leech is evident in the drawing at the top, which presents a dorsal view of *Haementeria ghilianii,* a giant bloodsucking species native to French Guiana. In contrast to *Helobdella,* which reaches an adult length of only about two centimeters, *Haementeria* grows as long as 50 centimeters. Nevertheless, the two species resemble each other in many respects, and they are considered interchangeable for many experimental purposes. The dorsal skin of each mid-body segment in both species is subdivided into three annuli, or rings. One such segment is depicted in the cross-sectional perspective view at the bottom. The main components of the leech's nervous system are shown in color. The ventral nerve cord consists of 32 segmental ganglia. The rearmost seven segmental ganglia are fused to form the large tail ganglion; the frontmost four are fused to form the lower part of the large head ganglion. (The head ganglion also includes an upper part whose development is entirely separate from that of the segmental ganglia.) Segmental nerves are shown in greater detail in the perspective view.

139

receives the teloplasm from the dorsal pole; the other cell, designated *DM*, lies more ventrally and receives the teloplasm from the ventral pole.

At this stage the separation of the embryo into the three layers of germinal tissue has been accomplished: the progeny cells of *A, B* and *C* will give rise to the endoderm (the innermost layer), the progeny of *DNOPQ* to the ectoderm (the outermost layer) and the progeny of *DM* to the mesoderm (the middle layer). The next two cleavages establish the bilateral symmetry of the embryo: cell *DM* divides to yield the left and right *M* cells, and *DNOPQ* divides to yield the pair of *NOPQ* cells lying on each side of the future midline. Three further cleavages of the paired *NOPQ* cells produce four bilateral cell pairs, designated *N, O, P* and *Q*.

The paired *M, N, O, P* and *Q* cells, among which the teloplasm of the egg has now been partitioned, are referred to as teloblasts. As soon as each teloblast has been formed it begins a series of highly unequal divisions. Through these divisions each teloblast gives rise to a bandlet of small stem cells, to which most of the teloplasm but little of the yolk of the teloblast is eventually distributed. The stem-cell bandlets produced by the five teloblasts on each side of the midline grow and merge to form a prominent pair of ridges: the left and right germinal bands. The stem-cell bandlet descended from cell *M* lies under the other four bandlets.

As more stem cells are produced the left and right germinal bands advance toward the front of the embryo along a crescent-shaped path over the future dorsal surface; they converge at the site where the head will form. At the same time the middle regions of the still-lengthening germinal bands migrate circumferentially to the future ventral surface. Eventually the left and right germinal bands meet on the ventral midline, where they coalesce. Coalescence begins at the future head and continues like a zipper to the tail.

When coalescence is complete, the germinal bands have given rise to the structure called the germinal plate, which lies on the ventral midline. In the course of subsequent development the cells of the germinal plate divide to yield the precursor cells of the adult tissues. The proliferation of cells results in a gradual thickening and circumferential expansion of the germinal plate over the surface of the embryo and back into dorsal territory. Eventually the left and right edges of the expanding germinal plate meet and coalesce on the future dorsal midline, thereby closing the body tube of the leech.

Early in this process of expansion the germinal plate becomes partitioned into a series of tissue blocks, separated by septa (transverse walls). Each block corresponds to a future body segment. Segmentation starts at the front, progresses to the rear and is complete on the formation of the 32nd segment, by which time the expanding germinal plate covers about a third of the ventral surface.

Meanwhile the formation of the embryonic gut, derived from the endodermal precursor cells *A, B* and *C*, is under way. The gut first appears as a long cylinder filled with yolk from cells *A, B* and *C*. Later the cylinder is divided into segments by a series of annular constrictions in register with the septa and possibly generated by their growth. As a result of this process a pair of gut expansions, called caeca, develops in each abdominal body segment. The segmentation of the gut is completed concurrently with the closure of the body tube; the embryo now has the general shape of the juvenile leech.

The progressive segmentation of the germinal plate and the general front-to-rear sequence of development are manifest also in the formation of the nervous system. The first indication of the presence of nervous tissue is the appearance of paired hemispheric cell masses on each side of the midline of the germinal plate. Each pair of cell masses is the "primordium" of a segmental ganglion. The paired primordia coalesce sequentially on the ventral midline to form globular ganglia, which already have approximately the adult number of neurons. When the expanding germinal plate covers about half of the ventral surface, all 32 ganglia are present. Initially each ganglion lies in direct contact with the ganglia of the neighboring segments. As the embryo elongates, the ganglia move apart and a short connective nerve grows between them. The chain of connected ganglia constitutes the ventral nerve cord.

Our colleague Andrew P. Kramer has been able to dissect the developing nervous system from *Haementeria* embryos and to penetrate its neurons with microelectrodes and micropipettes. In this way he has made intracellular electrical recordings from embryonic neurons and injected fluorescent dyes into them. Kramer has found that when segmental ganglia first appear in their globular form, their neurons have not yet grown axons and dendrites, the extensions characteristic of adult neurons. The immature neurons are coupled by junctions that allow injected dye molecules to pass from cell to cell. Later the neurons begin to grow axons, and their earlier couplings largely disappear. Even then, however, the embryonic neurons do not yet show the electrical activity of adult neurons.

Electrical activity of this kind appears at a still later stage, when growing axons have begun to enter both the connective and the segmental nerves. By the time of body closure the embryonic nervous system has taken on the overall properties of the adult system. The neurons now show transient changes in electric potential resembling those shown by adult neurons. The electrical activity suggests that the neurons are linked
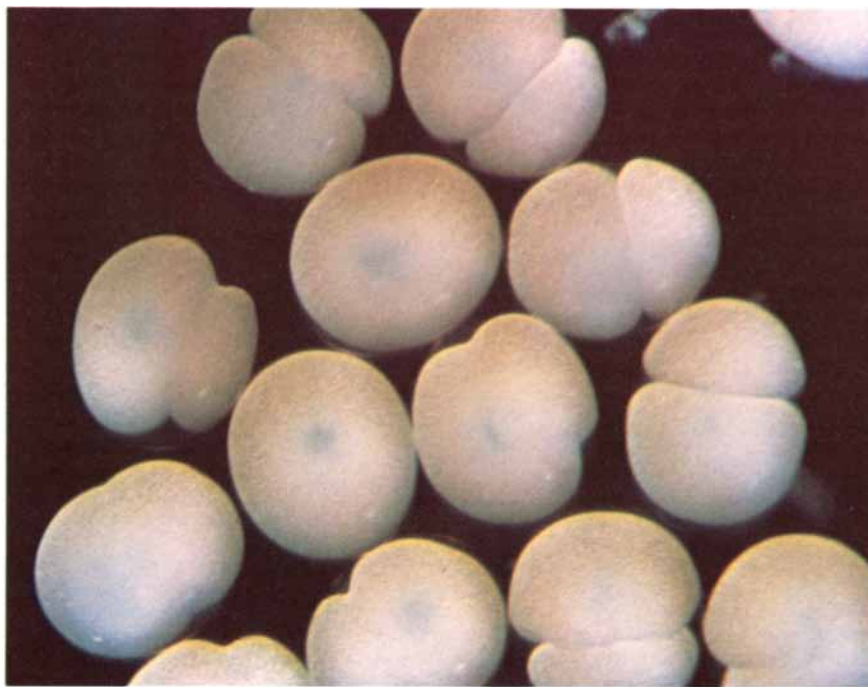
EGGS of the dwarf leech *Helobdella* are enlarged about 50 diameters in this photomicrograph. Some of the eggs have not yet begun to divide; others are in the process of dividing for the first time, and still others have just completed their first division. In two of the uncleaved eggs the teloplasm is visible as a light blue spot at the pole of the egg. In two of the cleaved eggs the teloplasm has been passed to the larger of the two daughter cells, the one referred to as *CD*.

140

by functioning synapses, the junctions through which nerve impulses are transmitted from one cell to another.

Because the functioning of a synapse depends on the release of the substances called neurotransmitters it follows that by the time of body closure the embryonic neurons must be sufficiently mature to synthesize these substances. Studies by our colleague Duncan K. Stuart have shown that indeed they are. To ascertain how the capacity for synthesizing the neurotransmitter acetylcholine develops Stuart dissected the nerve cords from leech embryos of various ages and placed them in a solution containing choline (the metabolic precursor of acetylcholine) that had been labeled with tritium (the radioactive isotope of hydrogen). The rate of formation of tritium-labeled acetylcholine provided a quantitative index of the concentration of enzymes that catalyze the uptake of choline and its conversion into acetylcholine.

Stuart found that the capacity for acetylcholine synthesis is still very low when the ganglia are already morphologically complete and their neurons have begun to grow axons. The synthetic capacity begins to rise rapidly, however, once the neurons have sent axons into the connective and segmental nerves and have begun to exhibit electric action potentials. By the time of body closure the capacity of the nerve cord for synthesizing acetylcholine has increased more than 25 times over its initial value.

Two other neurotransmitters studied by Stuart in identified leech neurons are dopamine and serotonin. When adult ganglia are treated with the fluorescent dye glyoxylic acid, the dye is taken up mainly by dopamine- and serotonin-containing neurons, which can then be identified under the microscope by their green fluorescence. By exposing the nervous system of embryos of various ages to glyoxylic acid Stuart found that the dopamine- and serotonin-containing neurons start accumulating their neurotransmitters at about the same time that the other neurons gain the capacity to synthesize acetylcholine.

The morphological, biochemical and electrical maturation of the leech's nervous system is paralleled by the behavioral evolution of the embryo, which begins at an early stage and proceeds through a stereotyped sequence of motor acts. Analyses by Kramer of videotape recordings of *Haementeria* embryos have shown that the behavior of the embryo progresses from simple, irregular twitching to complex patterns of movement. The simple movements of early embryos are incorporated into the more elaborate movements observed at later stages. The more complex movements are in turn components of the locomotor routines of the juvenile and
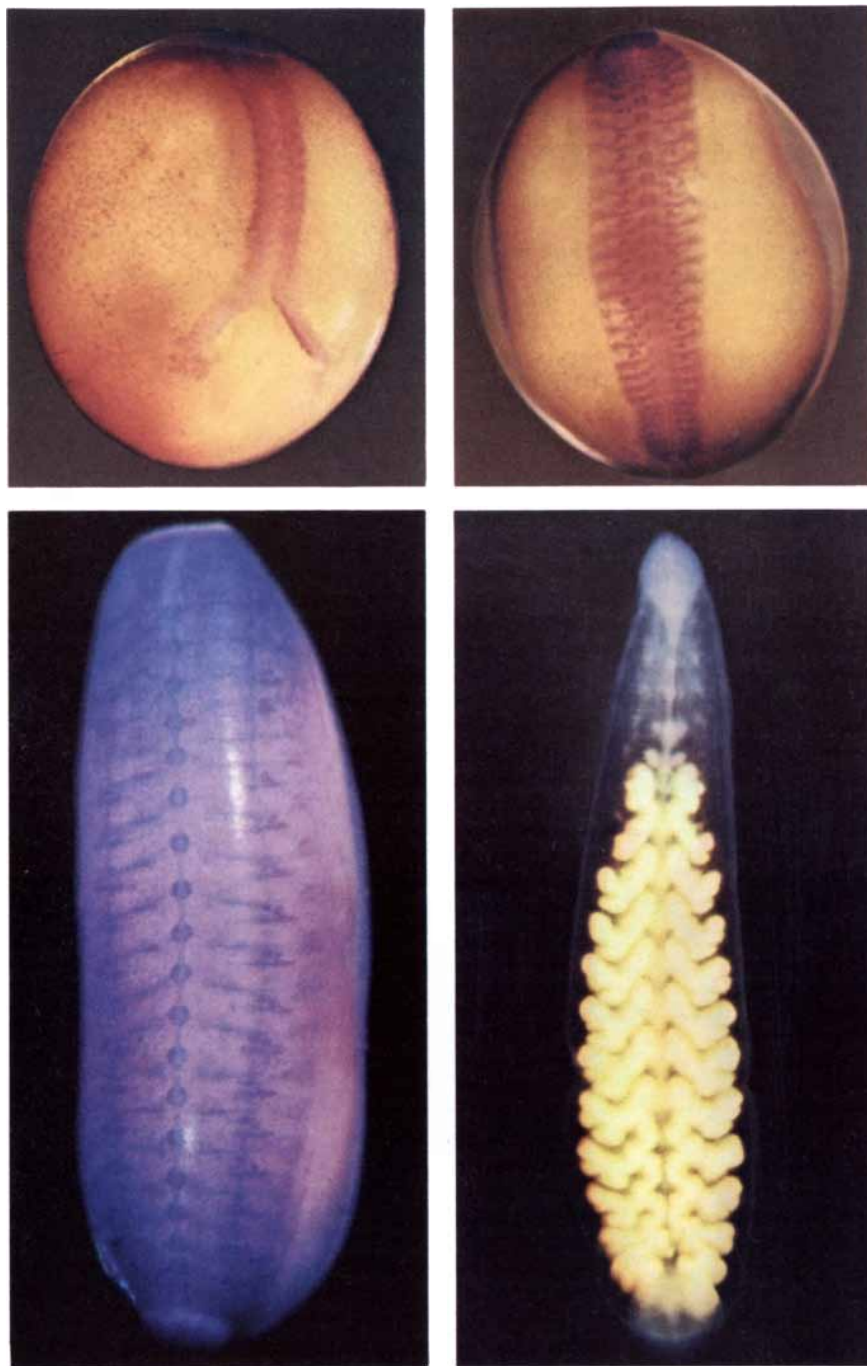
EMBRYOS of the giant leech *Haementeria* were photographed at four successive stages of development by Roy T. Sawyer of the authors' group. In all the photomicrographs the future head is at the top. In general the expansion of the germinal plate is accompanied by a progressive lengthening of the longitudinal axis with respect to both the horizontal and the vertical axes, gradually transforming the ovoid early embryo into the elongated, flattened juvenile leech. At the upper left is a nine-day-old embryo (corresponding to late Stage 8). The formation of the germinal plate on the ventral midline is almost complete, except for the rearmost part, where the germinal bands have not yet coalesced. The front part of the germinal plate is already divided into segmental tissue blocks. The cells of the germinal plate have been stained purple to distinguish them from the yellow of the yolk. At the upper right is an 11-day-old embryo (Stage 9). The expanding germinal plate has covered about a fourth of the ventral surface, and the segments are separated by septa, or transverse partitions. In the front third of the germinal plate each segment has a complete ganglion on the ventral midline; in the middle third the segments have primordial ganglia at various stages of coalescence, and in the rear third no such structures are apparent yet. At the lower left is a 14-day-old embryo (Stage 10). The germinal plate has by now covered all the ventral surface, and the segmental ganglia of the ventral nerve cord are all present, linked by connective nerves. In addition the rear sucker has begun to form. At the lower right is a 19-day-old embryo (Stage 11). The left and right leading edges of the expanding germinal plate have met and coalesced all along the dorsal midline, thus closing the body. The gut, still containing the yolk, is subdivided into the segmental compartments called caeca; the rear sucker is visible as a faint disk. Embryos are shown at different magnifications.

141

adult animals. Some of the embryonic movements may fulfill a physiological function at the stage where they appear; for example, they may help in circulating nutrients or in the hatching of the embryo from the egg membrane. Other movements, however, appear to be no more than preludes to adult behavior, incidental to the formation of functional connections between the components of the developing nervous system.

Overt movements that seem to prefigure adult behavior begin with a peristaltic constriction wave first observed as the germinal plate completes its coalescence. Since there is no nervous system at this stage of development, the peristaltic wave is likely to represent a myogenic rhythm, that is, one originating in the musculature. Peristalsis eventually results in the hatching of the embryo from the egg membrane. After hatching peristalsis tapers off and the embryo begins intermittent lateral bending movements, effected by the contraction of the longitudinal muscles on one side and the distention of those on the other side. By this time the ventral nerve cord of the embryo is mature enough to support the execution of neurogenic movements, or movements originating in the nervous system.

After body closure the lateral bending movements give way to a more complicated cycle of motions, which is evidently the prelude to the inchwormlike crawling of the juvenile leech [see illustration on opposite page]. Later, after the front and rear suckers have developed, the precrawling cycle of the embryo ripens into actual crawling. On reaching the status of a juvenile the miniature leech can also coordinate the contraction of its longitudinal muscles in the rhythm that underlies the swimming movement. Hence the juvenile Haementeria is able to seek its first blood meal by either crawling or swimming toward a prospective host.

We have inquired into the embryonic origins of the component cells of the leech's nervous system. Whitman and his students had established by direct microscopic examination of early embryos the lines of descent from the uncleaved egg of cells A, B and C and of the teloblasts and their stem-cell bandlets. With the continued increase in the number of cells at later developmental stages direct observation becomes too cumbersome a method for following the fate of individual cells. In order to compile more detailed cell pedigrees in the nervous system of the developing leech we devised a new tracer technique.
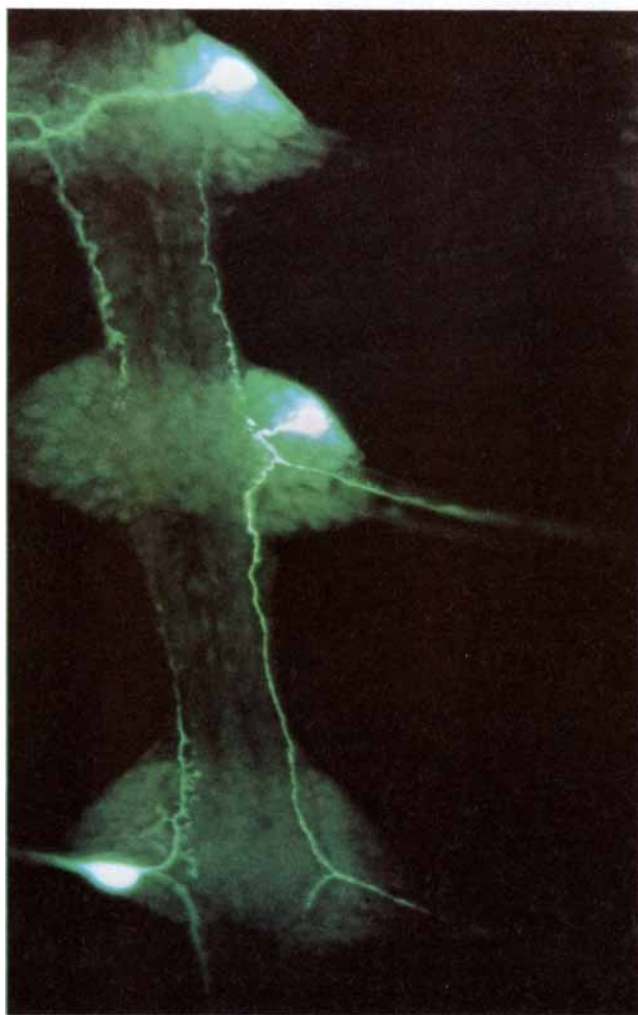
SEGMENTAL GANGLIA of the nerve cord of *Haementeria* embryos are viewed at the body-closure stage. In each of the three ganglia at the left a neuron has been injected with the fluorescent dye lucifer yellow. The injected neuron of the anterior ganglion (*top*) sends its axon into the segmental nerve on the opposite side of the ganglion and a branch into the posterior connective nerve, which links the ganglion to the middle ganglion. The injected neurons in the middle and rear ganglia send their axons into the segmental nerve on the same side and branches into both the anterior and the posterior connective nerves. The injected neuron of the middle ganglion also sends a secondary branch into the segmental nerve of the next posterior ganglion. The ganglion in the photomicrograph at the right has been treated with glyoxylic acid, a dye that induces fluorescence specifically in neurons containing the neurotransmitter substances dopamine and serotonin. The yellow-green fluorescence identifies seven serotonin-containing cells: three cell pairs and one unpaired cell (*top center*). The more diffuse blue-green fluorescence arises from the axons of two pairs of dopamine-containing neurons, whose cell bodies lie outside the ganglion. The photomicrograph at the left was made by Andrew P. Kramer; the one at the right was made by Duncan K. Stuart.

142

In the new technique a tracer molecule is injected through a micropipette into an identified cell of the early embryo. Embryonic development is then allowed to progress to a later stage, whereupon the distribution of the tracer in the tissues is examined. The tracer selected for such an experiment must satisfy three conditions: embryonic development must continue normally after the injection; the injected tracer must remain intact and not be diluted too much, and the tracer must not pass through the junctions that link embryonic cells, so that it will be passed on exclusively to direct descendants of the injected cell.

One of the tracers we chose is the red-fluorescing dye rhodamine, which we coupled to a tailor-made polypeptide carrier molecule with a known sequence of amino acids. In collaboration with Janis D. Young we synthesized the polypeptide from amino acids whose configuration is different from the natural one; in this way the polypeptide was made resistant to degradation by digestive enzymes. The rhodamine-polypeptide complex has proved particularly useful for visualizing the development of the germinal plate. For this purpose a teloblast or a precursor cell of a teloblast is injected with the rhodamine-polypeptide complex, and the embryo is later fixed, cleared of yolk and treated with a blue-fluorescing stain specific for DNA. The result is blue fluorescence of all the cell nuclei and red fluorescence of the cytoplasm of those cells descended from the injected teloblast.

In one such experiment, done by our colleague Saul L. Zackson, the right *OPQ* cell (the precursor of the right *O*, *P* and *Q* teloblasts) was injected with the rhodamine-polypeptide complex at Stage 6 and the embryo was examined at Stage 8 [*see illustration on page 137*]. The three large, red-stained cells visible in the photomicrograph are the right *O*, *P* and *Q* teloblasts, from each of which a bandlet of daughter stem cells can be seen projecting into the right germinal band. The nuclei of the cells belonging to each bandlet can be recognized because they lie within the red cells. Close inspection of the bandlet nuclei shows that some of them are in the process of division.

In another experiment an embryo whose left *M* teloblast had been injected with the rhodamine-polypeptide complex at Stage 5 was similarly examined at Stage 8 [*see illustration on next page*]. Here the large red *M* teloblast lies deep (as might be expected of the cell that gives rise to the mesoderm) and is out of focus. From it a bandlet of red stem cells projects upward; on reaching the surface the bandlet makes a sharp left turn and joins the crescent-shaped left germinal band. At its front end the mesodermal bandlet is subdivided into clusters of labeled cells. The subdivision is the first sign of the segmentation of the mesodermal tissue, which begins even before the left and right germinal bands have coalesced.

The accompanying photographic enlargement of the front end of the germinal band of this embryo shows that the adjacent clusters of cells are similar in appearance. The similarity, technically referred to as an isomorphism, suggests that the early development of the mesoderm of each body segment proceeds by a sequence of stereotyped cell divisions. More recent experiments by Zackson have shown that each of the cell clusters arises from a single stem cell, which is derived in turn from an *M* teloblast. Each cluster is probably the precursor of all the mesodermal tissue on one side of the midline of a segment. Therefore each of the 32 mesodermal body segments may originate as a bilateral pair of stem cells derived from an *M* teloblast.

The other tracer we have employed to establish embryonic cell lineages is the enzyme horseradish peroxidase. In the mid-1970's Kenneth J. Muller and Uel J. McMahan II of the Harvard Medical School pioneered the experimental procedure of injecting horseradish peroxidase into neurons of the adult leech's nervous system. When the tissue is subsequently exposed to hydrogen peroxide and benzidine, the entire injected neuron, down to its finest axonal and dendritic extensions, turns a dark color.

We injected horseradish peroxidase into embryos at Stage 6, when the germinal plate has expanded to cover nearly all the ventral surface and when all 32 segmental ganglia of the nerve cord are present [*see illustration on page 145*]. In one embryo the enzyme was introduced into the left *N* teloblast; the stain appeared in segmentally repeated structures whose size, shape and position indicate that they are the half ganglia on the left side of the ventral nerve cord. It follows that the *N*-teloblast pair is the progenitor of a substantial part of the leech's nervous system. The ganglia were not stained uniformly, howev-

PRECRAWLING MOVEMENT of a *Haementeria* embryo at the body-closure stage of development was recorded on videotape by Kramer. The elapsed time of the sequence is 25 seconds. At a later stage of development the movement seen here will ripen into actual crawling. The mature leech begins its inch-wormlike crawling cycle by attaching its rear sucker to a solid substrate and elongating its body. It then attaches the front sucker and releases the rear sucker. The body is next shortened and bent downward and inward at the ends, after which the rear sucker is reattached at a more advanced position and the front sucker is released. The cycle is completed when the body unbends again and elongates.

143
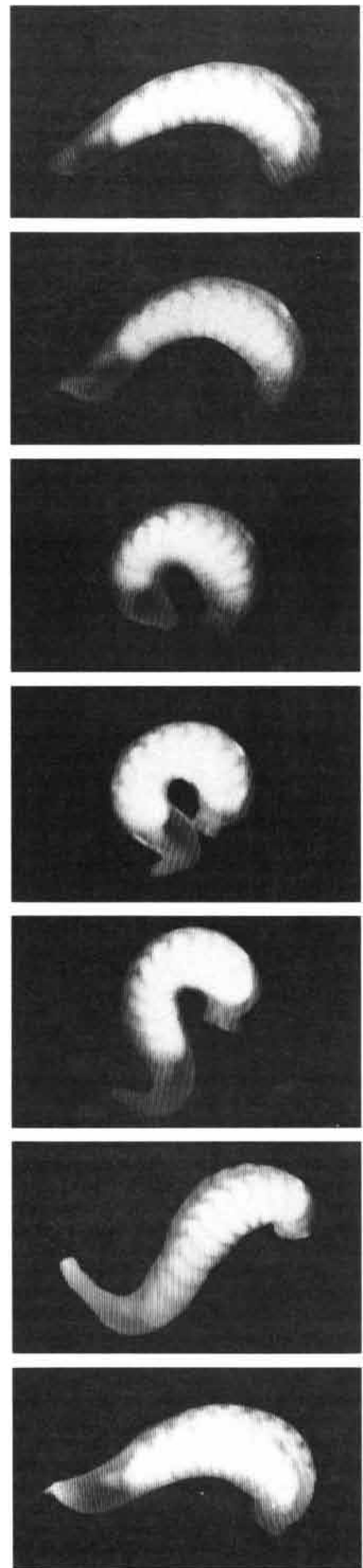
er; within each half ganglion the stain formed a thin longitudinal strip next to the ventral midline and two transverse strips extending laterally from it. The presence of unstained tissue between the two transverse strips in each ganglion indicated that some neurons of the seg-mental nervous system have an origin other than the *N* teloblast.

A second embryo in which the left *OPQ* cell was stained with horseradish peroxidase showed that the remaining neurons originate with the *O, P* and *Q* teloblasts. Although most of the stain was broadly distributed over the ecto-derm of the left germinal plate, there were also some segmentally repeated patches of stain whose position near the ventral midline indicated that they mark neurons within the segmental ganglia. The staining patterns in the two embry-os were found to be complementary: the patches of nervous tissue stained after the injection of horseradish peroxidase into cell *OPQ* are just those left un-stained by the injection of the enzyme into the *N* teloblast.

The cellular distribution of the horse-radish peroxidase stain was ascertained in much finer detail by examining em-bryos sliced into serial sections. Sections through the nerve cord of embryos whose *N* teloblast has been injected indi-cate that the thin longitudinal strip of stain corresponds to a cell cluster in the part of the ganglion nearest the midline; the two transverse strips correspond to an anterior and a posterior cluster of neurons in the part of the ganglion far-thest from the midline. Sections cut through the nerve cord of embryos whose *OPQ* cell has been injected show that here the stained cell clusters corre-spond to the clusters that remain un-stained when the *N* teloblast is labeled.

Further experiments in which *O, P* and *Q* teloblasts (rather than their com-mon precursor) were injected individu-ally with horseradish peroxidase have shown that the descendants of each tel-oblast form characteristic patterns of neuronal clusters in each segmental gan-glion. The size and position of each clus-ter are quite invariant from segment to segment and from specimen to speci-men. This finding suggests that each of the four teloblast pairs is the founder of a distinct neuronal family. Moreover, in view of the known positional invariance of identified neurons in the segmental ganglia it would appear that each identi-fied neuron is the lineal descendant of a given teloblast.

Cell-lineage analyses of this type, which are sometimes called fate maps, show that in the leech neurogen-esis is highly determinate, in the sense that a particular teloblast regularly gives rise to a particular part of the ner-vous system. Another aspect of this de-terminacy is that when a teloblast dies or malfunctions, its developmental role is not taken over fully by any other cell. Killing a teloblast of the early embryo therefore leads to characteristic devel-opmental aberrations. In collaboration with Seth S. Blair we examined the neu-roanatomical consequences of killing identified teloblasts or precursor cells of teloblasts by injecting them with toxic enzymes.

In accord with our cell-lineage analy-ses we found that killing an *N* teloblast at Stage 6 results in the formation of an
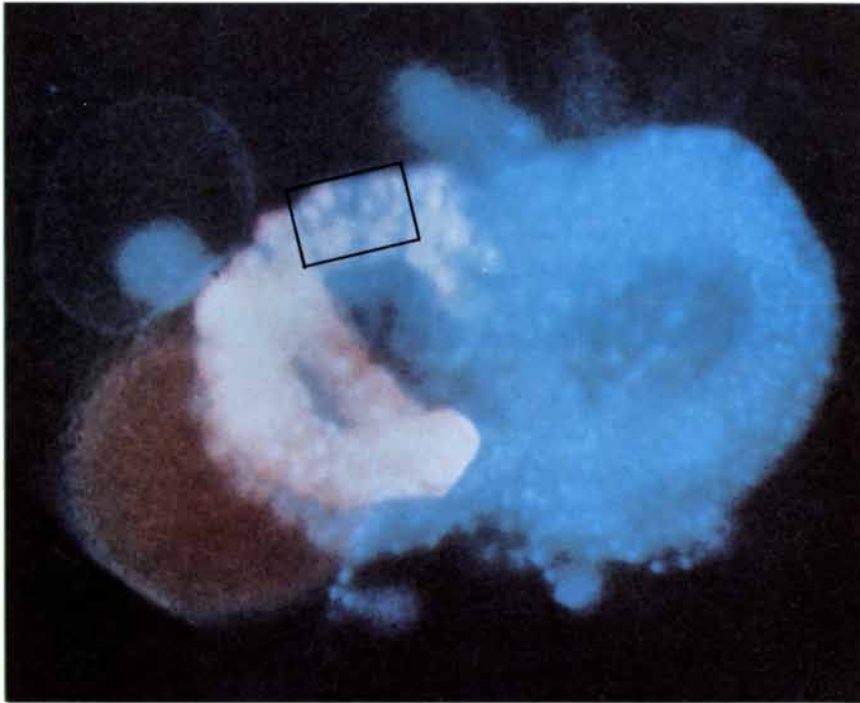
FORMATION OF THE MESODERM, the middle layer of germinal tissue in the leech em-bryo, was elucidated by the authors' new tracer technique. The *M* teloblast of the embryo in the photomicrograph at the top had been injected with the red-fluorescing rhodamine-poly-peptide complex at Stage 5; it was examined at Stage 8 after having been fixed, cleared of yolk and treated with a blue-fluorescing dye. A bandlet of red stem cells can be seen project-ing upward from the large red *M* teloblast and making a sharp left turn at the surface of the embryo before joining the crescent-shaped left germinal band. The black rectangle marks two adja-cent clusters of mesodermal stem cells, which are reproduced at a much higher magnification at the bottom. The clusters are isomorphic (morphologically and topologically similar), sug-gesting that the early development of the mesoderm proceeds by a sequence of stereotyped cell divisions. Both photomicrographs were made by Saul L. Zackson, as was the one on page 137.

144
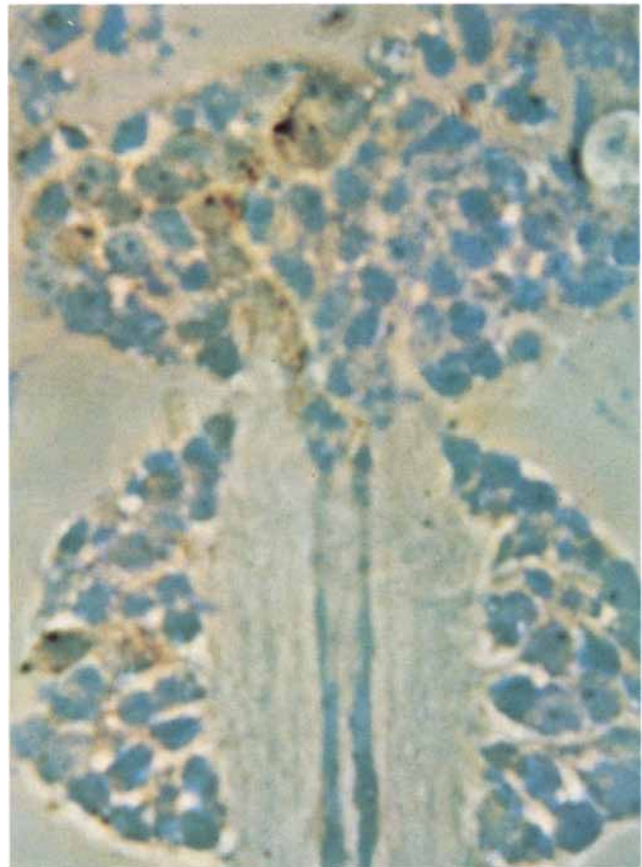
ALTERNATIVE TRACER TECHNIQUE was employed by the authors to establish cell lineages in leech embryos. The two whole embryos seen from the ventral side in the photomicrograph at the left were both injected with the enzyme horseradish peroxidase at Stage 6; at Stage 10 they were exposed to chemicals that turned the injected cells and the cells descended from them dark brown. In the embryo on the left side of the photomicrograph the enzyme was injected into the left *OPQ* teloblast; in the embryo on the right side the enzyme was injected into the left *N* teloblast. The higher-magnification photomicrograph at the right shows a slightly oblique horizontal section cut through two segmental ganglia of the nerve cord of an *OPQ*-injected embryo similar to the one on the left side in the lower-magnification photomicrograph. The nerve cells containing the horseradish peroxidase appear brown against the blue-stained background. The section was cut partway through the central nerve tract, revealing the axons there, which run in the longitudinal direction.

embryonic nervous system whose segmental ganglia show anatomical deficits on the altered side. The pattern of the deficits is not as regular as one might have expected, however: the half ganglia on the altered side vary greatly in size within the same specimen, ranging from an almost normal number of cells to a total absence of segmental neurons.

The half ganglia that are almost normal in size include the cells normally provided by the *O, P* and *Q* teloblasts on the same side. In addition they include some cells derived from the intact *N* teloblast on the other side; the latter cells have made an abnormal crossing of the midline of the germinal plate during the formation of the ganglionic primordia. That some half ganglia are entirely absent in these specimens suggests that the abnormal midline crossing by *N*-derived cells is subject to considerable fluctuation. It also appears that the *N*-derived cells play some organizing role for the *O*-, *P*- and *Q*-derived cells in founding the ganglionic primordia.

As might have been expected, killing an *M* teloblast results in an embryo that lacks all mesodermal structures on the altered side. In addition the altered side has no recognizable nervous system. The normal *N, O, P* and *Q* teloblasts still produce their bandlets of stem cells, but in the absence of the mesodermal tissue blocks the precursor cells of the nervous system are not organized into ganglionic primordia. The deficits caused by the killing of teloblasts show the importance of interactions among cells of different lines of descent in shaping the normal development of the leech's nervous system.

So how do the neurons of the leech's nervous system and their specific connections arise during development? As is evident from the foregoing account, the answers to this grand question we have so far been able to provide are more descriptive than explanatory. Nevertheless, a number of more narrowly focused questions about neurodevelopmental mechanisms can now at least be formulated. For instance, is the precise succession of alternating cell-cleavage planes, first during early embryogenesis and later during the formation of the mesodermal segments, actually predetermined? If it is, the determinacy probably reflects the replication and positioning of the elements of the cytoskeleton of the embryonic cells, because it is the orientation of the cytoskeleton that governs the cleavage planes in cell division.

A related question concerns the developmental role of the teloplasm that is passed on only to cell *D*. As was explained above, the ventral part of the teloplasm is given to cell *DM* and thence to the mesodermal precursor cells, whereas the dorsal part goes to cell *DNOPQ* and the ectodermal precursor cells. Most of the teloplasm is eventually parceled out by the teloblasts to their stem-cell descendants. What kind of counting mechanism ensures that exactly 32—never 31 or 33—stem cells descended from the *M* teloblast found a segmental block of mesodermal tissue on each side of the embryo? How do

145
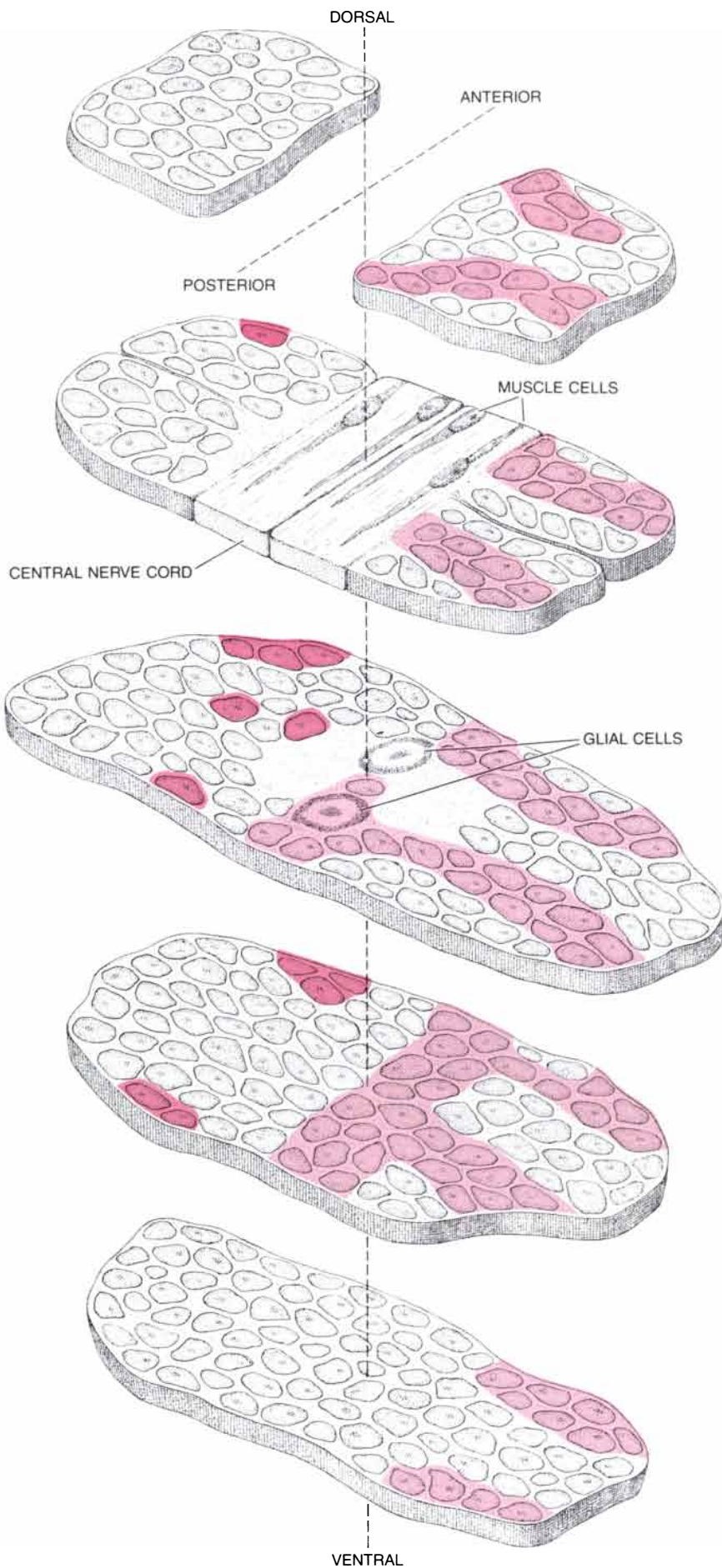
the cells descended from the ectodermal precursor teloblasts *N, O, P* and *Q* lying within such a tissue block assemble to form the neurons of the ganglionic primordia? Is the morphological, electrical and chemical maturation of the neurons dependent on the interaction of their growing axons with target cells or tissues? Does the development of the motor neurons and their connections that underlies the gradual perfection of the crawling and swimming movements require practice, or does locomotor behavior arise autonomously during embryogenesis without the need for any functional feedback?

Finally, how does the line of cellular descent govern the ultimate character of a neuron? Does each neuron receive a particular set of intracellular determinants that were distributed in some regular way among the daughter cells in successive embryonic cell divisions? Or is the nature of the cell determined by its position in the embryo and hence by the regular cleavage pattern through which it arose? We hope that the experimental techniques already in hand will enable us to answer some of these questions before long.
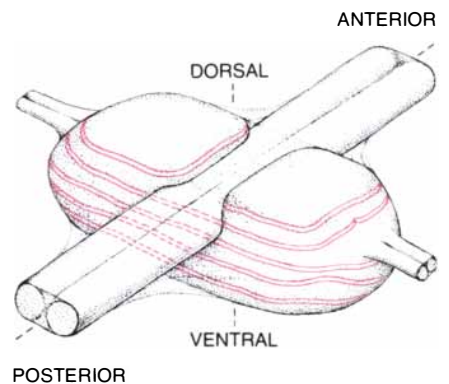
"FATE MAPS" identify the early embryonic origins of the cells in a segmental ganglion of a late leech embryo. The drawing at the left is an exploded view of five horizontal sections sliced through a mid-body segmental ganglion of a *Helobdella* embryo at Stage 10. (An overall view of the intact ganglion is shown above.) The two pairs of dark, elongated shapes at the center of the second section from the top represent identifiable muscle cells in the longitudinal nerve tract. They are descended from the *M*-teloblast pair. The two dark circular contours in the center of the middle section represent two identifiable glial cells, each of which is a descendant of an *N* teloblast. The smaller globular shapes seen in cross section at each level do not correspond to actual cells but are meant to indicate the approximate size, number and disposition of the neurons in each section. In the left half ganglion the domains shown in color contain descendants of the left *Q* teloblast; in the right half ganglion the domains shown in color contain descendants of the right *N* teloblast. Domains not colored contain descendants of the *O* and *P* teloblasts.

An Ancient Greek City
in Central Asia

A legacy of Alexander's conquest of Persia was a Greek kingdom in what is now Afghanistan. A French archaeological group has spent the past 15 years unearthing one of its cities, Ai Khanum

by Paul Bernard

Alexander the Great (356–323 B.C.) is remembered as the liberator of the Greek cities of Asia Minor, the captor of Egypt and the victor over the last Persian emperor, Darius III. It is less well known that his conquest of the Persian realm carried him as far east as central Asia, with the result that a Hellenic state rose in the valley of the Oxus (the modern Amu Darya). In the third and second century B.C. this outpost of Greek civilization controlled territories that lie today in the Soviet republics of Turkmenistan, Tadzhikistan and Uzbekistan, and in Afghanistan and Pakistan. Yet apart from an abundant coinage that testifies to a succession of Greek monarchs nothing of the civilization of these Greek colonists appeared to have survived. Some scholars had even come to refer to it as the "Greco-Bactrian mirage."

Excavations at an ancient city site on the Afghan side of the border between Afghanistan and the U.S.S.R. have now put an end to skepticism about Hellenic Bactriana. Fifteen years of work at the site, Ai Khanum, by a French archaeological group show that for two centuries a Greek state did indeed flourish on the Oxus, more than 5,000 kilometers distant from the cradle of Greek civilization.

Among the handful of facts learned from historical references to this central Asian kingdom is that before the Greek conquest it had been a Persian frontier province. Twenty years after Alexander's death a comrade-in-arms, Seleucus, controlled that part of Alexander's former empire extending from Syria to the Hindu Kush, including the frontier province of Bactriana. Of the "1,000 cities" for which Bactriana was famous Ai Khanum is the only one that has come to light. It was founded either by Alexander himself or some years later by Seleucus, who by 305 B.C. had assumed the title of king.

Under Seleucid rule over the next half century Bactriana remained in direct contact with the Mediterranean world. An influx of settlers both increased the Greek population of the province and reinforced the Hellenic components of its culture. In about 250 B.C., at a time when the Seleucids were preoccupied with rivalries to the west, Bactriana quietly seceded and became an independent kingdom. Limited at first to its Oxus territories, the Greco-Bactrian state later expanded toward the south until its boundaries embraced rich lands in the upper valley of the Indus.

Although this enlarged Hellenic domain was prosperous, it lived under a constant threat from the nomadic peoples who roamed its northern border. Around 145 B.C. the Greeks of Ai Khanum were driven from their city by nomadic invaders, and before 100 B.C. all Hellenic central Asia was in the hands of the nomads. The Greek principalities of the upper Indus were able to hold back the invaders for a few more decades but eventually were forced to capitulate. Thus ended the easternmost extension of Hellenism.

The ruins of Ai Khanum stand on the left bank of the Oxus at the eastern tip of the river's middle valley. Here the river leaves the mountains to debouch onto its first sizable plain and is joined by a tributary, the Kokcha. The site is a long way from the most heavily traveled crossings of the Oxus and from the great caravan road to India, but it is well suited for a military outpost. A force stationed here could control the eastern marches of Bactriana and block a potential invasion route southward along a left-bank tributary of the Oxus upstream from the city.

The plain that extends along the left bank of the Oxus above Ai Khanum is some 10 kilometers wide and 35 kilometers long. Even before the founding of the city this rich loess terrain was already being cultivated with the aid of a large network of irrigation canals. In addition to these agricultural resources the proximity of the site to the mountains upstream for the summer pasturage of flocks must have made the area an excellent one for animal husbandry, as it still is today. The mountains are also rich in minerals. For example, mines in the upper valley of the Kokcha were a rich source of the gemstone lapis lazuli.

The triangular area at the confluence of the Oxus and the Kokcha must have seemed predestined to satisfy the requirements of the Greek city planners. A natural acropolis, some 60 meters higher than the rest of the site, guards the approach from the east, and a sheer drop to the two rivers protects the southern and western flanks of a zone that measures some two kilometers from north to south and 1.5 kilometers from east to west.

The settlers reinforced these natural defenses by completely enclosing the city within a rampart that ran along the banks of the two rivers and up the slope of the acropolis. Along the northern edge of the lower town, where there were no natural defenses, the city wall was built particularly strong. The solid mud-brick masonry was more than 10 meters high and six to eight meters thick; its massive rectangular towers extended another 10 meters in front of the wall. Outside this section of the rampart a steep ditch guarded against the approach of siege machinery. Finally, at the southeast corner of the acropolis, there was a small but virtually impregnable redoubt, measuring 150 by 100 meters. It was protected on two sides by an 80-meter drop to the Kokcha and on the third side by a moat.

The city's residential quarters and most of its public buildings were in the lower part of the site, which was less exposed to the wind than the acropolis and was easily supplied with water from the canals of the plain. The lower town was basically organized around the palace, a complex of monumental buildings that housed the city's administrative apparatus. So that these buildings

148

RUINED WALLS in the palace enclosure at Ai Khanum (*above*) bear traces of the whitewashed clay that once covered the bricks. These are administrative buildings at one corner of the 20-acre area.

MAIN TEMPLE (*below*) stands on a three-step platform off the main street. Deep niches decorate its thick walls. The time is soon after dawn; the mist in the background is rising from the Oxus River.
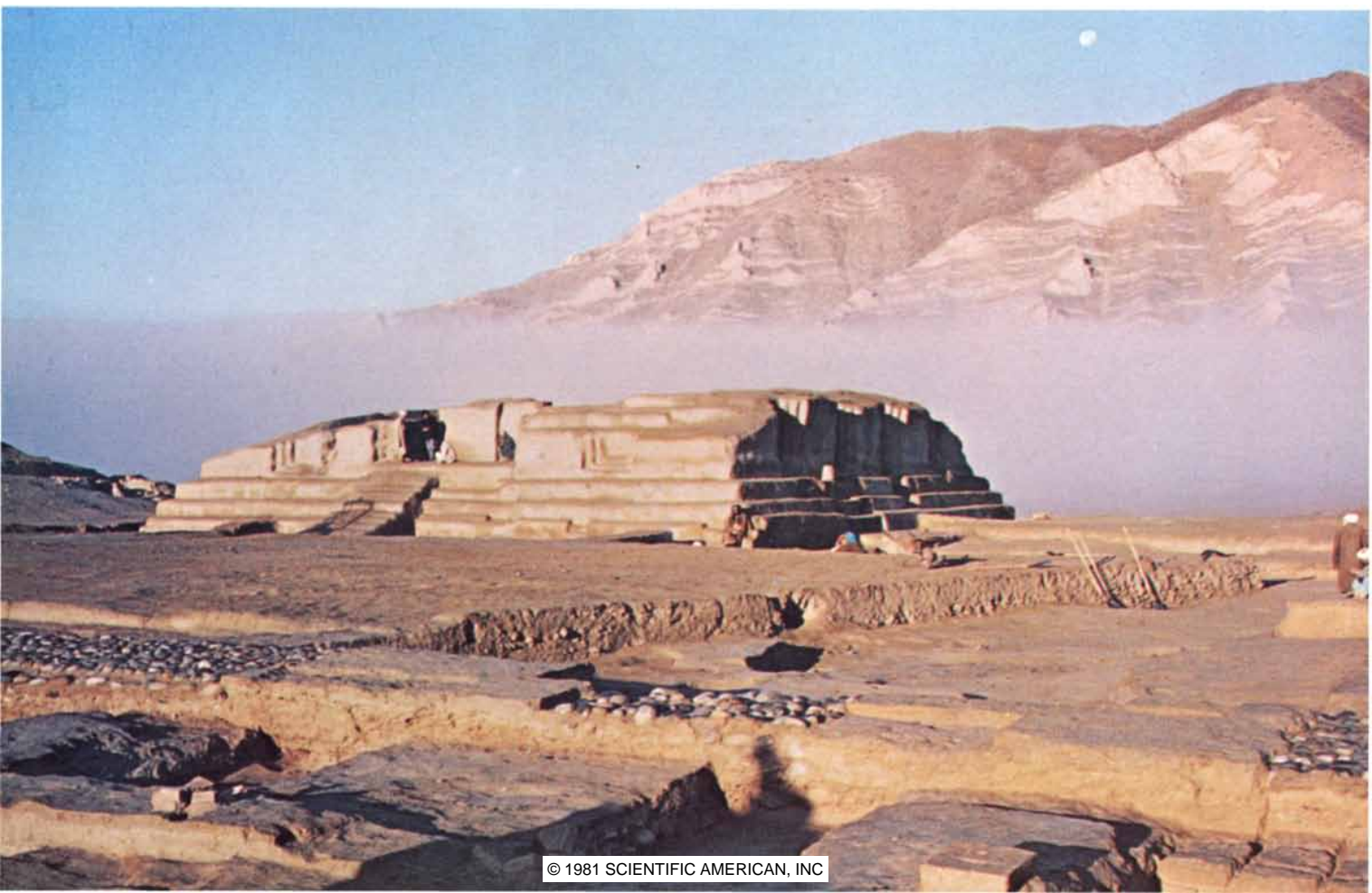
HOUSE WALLS of an elite residence in the lower town are recessed at the center to allow for a pillared porch. In such houses, which were 65 meters wide and 35 meters deep, the Greek tradition of a central courtyard gave way to a front court and a central living room.

MOSAIC FLOOR of a palace bathroom was made by setting a field of dark red pebbles in a stucco floor and outlining decorative figures with white pebbles. As befitted a bathing area, many of these figures represented aquatic animals and mythical ocean creatures.

could extend across the full width of the available space the planners built the city's main street on a terrace at the foot of the acropolis. The street ran from a gate in the city's north wall southward to the bluff above the Kokcha. This arrangement allowed the palace buildings to cover an area in the southern half of the town measuring 250 by 350 meters. A curving road ran from the palace area to the main street. At the junction stood a monumental gatehouse with two colonnaded facades.

Because the concept of a palace was alien to the Greek architects they turned for their models to the Persian royal edifices they had seen in the course of their conquest. Like Darius' palace at Susa, the palace at Ai Khanum simultaneously served three functions: it was a state structure, a residence and a treasury. A great formal courtyard, enclosed by four 118-column porticoes with Corinthian capitals, covered an area measuring 137 by 108 meters. Beyond the portico on the south side of the courtyard a vestibule with three rows of columns led to the heart of the palace. In the southeast corner of the palace grounds stood the administrative buildings, the most noteworthy of which was a square structure measuring 52 meters on a side. Its interior was divided into two symmetrical groups of rooms by corridors that crossed at right angles. On the east side of the building were two reception halls, their walls decorated with stone pilasters; on the west side were what appear to have been two chancery offices. A small courtyard with 60-column porticoes, built along the west border of the palace grounds, would have offered a more intimate official setting.

The private apartments, three in all, occupied the southwest corner of the palace enclosure. They are easily distinguished from other palace structures because each includes, in addition to domestic quarters, bathrooms and a front courtyard. A series of storerooms surrounding a central courtyard, situated on the west side of the palace enclosure, served as a treasury. The plan of this building is of non-Hellenic inspiration. The storage vessels found here and the debris associated with them—pearls and both rough and polished bits of agate, onyx, carnelian, garnet, lapis lazuli, beryl and crystal—are indicative of the building's function as a storehouse for valuables.

This ensemble of structures, covering nearly 90,000 square meters (more than 20 acres), is far grander than one might expect to find on a remote frontier. Its size reinforces the impression that the palace is a royal one. Ai Khanum must have been a capital in its own right, serving the same function in the eastern part of Bactriana that the famous city of Bactra served in the west. One indica-

tion of this is the presence of bronze blanks for coins not yet struck. They show that Ai Khanum had its own mint, a royal prerogative.

Who ruled here? There is reason to suppose the last master of the palace was the noted King Eucratides (170–145

B.C.), who waged a series of victorious campaigns in northwestern India. If that is so, the treasury at Ai Khanum should once have held a wealth of Indian coins and other valuables collected in his forays. Evidence that such was the case has survived in the form of several vases in

SILVER COIN of Bactriana bears the head of King Eucratides (170–145 B.C.) and on the reverse the twin heroes Castor and Pollux. The Greek colony fell at the end of Eucratides' reign.

SANDALED FOOT is one of the few surviving fragments of the cult statue in the main temple. Not enough of the statue remains to indicate whether the deity was male or female.

151

SOME OF THE SMARTEST BUSINESS EXECUTIVES IN AMERICA MIGHT FAIL THIS SIMPLE TEST.

Japan   USA

1. Which country has the world's most productive work force?

2. Which country spends the most for research and development?

3. Which country exported more goods last year than any other country in the world?

We wouldn't be surprised if you answered "Japan" to all three questions. After all, Japan's gains in these areas have been impressive, as has its press.

But in fact, despite recent trends, according to the Department of Labor's Bureau of Labor Statistics, real gross domestic product per employed person—the national measure of productivity—shows Japan behind the United States by 31.6%.

As for research and development, based on the latest National Science Foundation figures available, the United States spent three times more than Japan.

And in exports, the most recent International Trade Statistics Yearbook shows that even in manufactured goods alone U.S. exports had a dollar value about 39% greater than exports from Japan.

So Japan may be gaining, but it hasn't beaten America. In many areas it hasn't even caught up. Which is not to say the challenge from Japan is a hollow one. We know it is real.

We know individual companies in Japan, with which many of us compete, achieve excellent productivity levels.

We know U.S. spending, of itself, will not generate innovation. It takes commitment to leadership.

We know U.S. trade balances with Japan in certain businesses have shifted to Japan's advantage.

But we are a strong country with outstanding resources and a formidable overall lead. As we take notice of things we need to do, and get on with them, we can build on that strength and maintain our ability to compete successfully anywhere in the world.

America is a winner. A winner has confidence. In fact, one can't win without confidence. Yet, the way things have been written and spoken of lately, you'd think we'd lost our winning ways. Not true. We have great strengths. Let's build on them. We have great ability to recommit— to overcome challenge.

As for Motorola, we believe we are already doing much better than the average American company you would compare us to, and most Japanese companies as well. We have plans and programs in place that are working to improve constantly our quality and productivity, and to keep sharp the cutting edge of our technology.

We are confident we can win against competitive challenges. We are committing ourselves publicly to do so.

We know other companies feel the same.

It is only a matter of putting ourselves to the test. And having the right answers.

Ⓜ **MOTOROLA** *A World Leader in Electronics.*

Quality and productivity through employee participation in management.

the treasury storerooms. They are empty now, but they bear Greek inscriptions in ink that indicate their former contents: deposits of Indian and Greek coins stored away after having been counted for the public record. Hence it seems possible that Ai Khanum may be the place Eucratides renamed Eucratidia in his own honor; the city name is preserved in the works of ancient writers.

In any event the palace and its grandiose embellishments display the main characteristics of Greco-Bactrian architecture. They include walls built of mud bricks (with the first few courses occasionally consisting of fired bricks) that were smoothly plastered with a layer of clay and then whitewashed. The flat roofs of the buildings were characteristically Eastern, but an ornamental edging of terra-cotta tiles gave them a Greek look. The decor of cylindrical stone columns and rectangular stone pilasters had a similar Hellenizing effect; the scarcity of good building stone in central Asia made it a luxury rarely indulged there. At Ai Khanum the masons mass-produced column sections with large lathes and chiseled the column capitals in all three classical styles: the simple Doric, the involuted Ionic and the leafy Corinthian. Taken together, the colossal proportions of the palace, the strict geometric character of its plan and the architects' predilection for symmetry contribute to an effect of unmistakable but rather cold grandeur.

The importance of Ai Khanum is further indicated by the city's other public buildings. Three structures may be singled out. The first is the gymnasium, situated on the bank of the Oxus. An inscription states that, as was customary in Greece, this institution of learning and physical training was under the joint protection of Hermes and Heracles. The main building was square, 100 meters on a side, and essentially Greek in plan: a series of rooms and porticoes surrounded a central courtyard. At the same time both in size and in the relentless symmetry of its four sides, each featuring a central columned porch flanked by two long rooms, it exhibits genuinely Greco-Bactrian innovations.

The second great structure, also essentially Greek in plan, is the theater that extends up the inner slope of the acropolis. Its 35 tiers fan out over a semicircular area with a radius of 42 meters, and its capacity—estimated at 6,000—is larger than that of the theater in Babylon, the only other theater of the Hellenistic period known in the ancient Middle East. A touch of Oriental influence is evident in this most Greek of institutions. Halfway up the slope stood three spacious loggias: ostentatious seats of honor that are practically unknown in the theaters of Greece.

The third major building, unlike the other two, shows no debt to Hellenism. It is the arsenal, a rectangular structure on the city's main street. It was 140 meters long and some 100 meters wide, and featured a series of long storerooms that opened on a great central courtyard.

The city's finest residential section was just above the junction of the two rivers. The houses were of such luxurious proportions that no more than 50 could have been fitted into the substantial area available. They were built at regular intervals along parallel streets that run from east to west. Each house faced north, each had a columned front porch and a front courtyard and each covered an area measuring 65 by 35 meters. Where the traditional Greek house was planned around a central courtyard, these elite dwellings were dominated by the living room. It was set off from the other rooms, particularly from the domestic quarters, by a horseshoe-shaped gallery. The courtyard, no longer central, gave access to the living room but not to the work area.

The elite of Ai Khanum evidently prized their bathrooms. Each bath was divided into two or, more often, three units, walled with red-painted stucco and paved with stone or with cement-and-pebble mosaic. (Other rooms had a floor of earth.) The first of the three units was for disrobing. The second was the bathroom proper, where the bather,
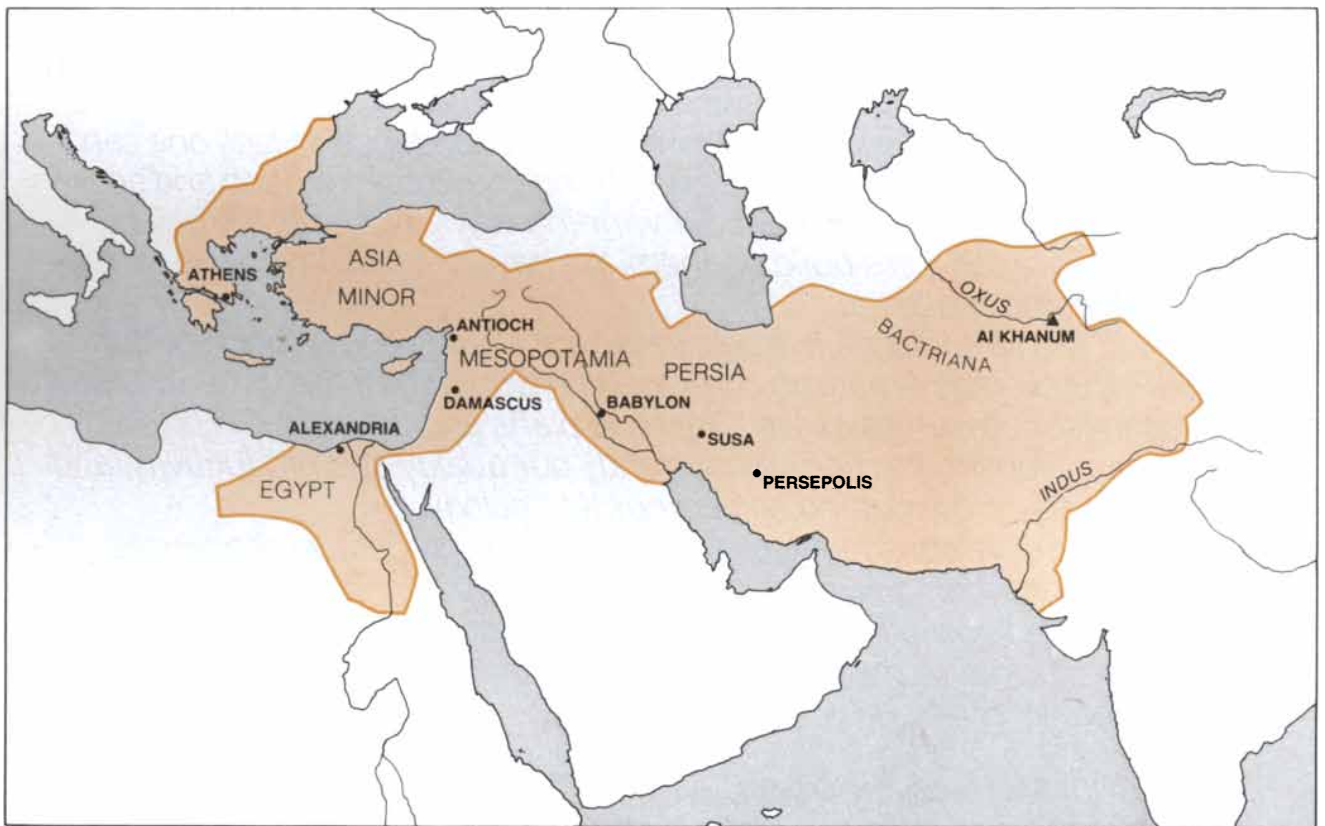
LEGACY OF ALEXANDER was a Hellenized empire that included Asia Minor and Egypt and extended eastward to the Indus River. One Greek frontier colony in the east was the city now called Ai Khanum, built on the Oxus River after Alexander's conquest of a former Persian province, Bactriana. At first a part of the realm ruled by Seleucus, Bactriana became an independent kingdom in 250 B.C.

SCIENCE/SCOPE

A new photochemical vapor deposition process forms oxide layers on selected sub-
strates at low temperatures. The Hughes PHOTOX™ process deposits silicon diox-
ide and other oxide dielectrics on semiconductor devices, and coats temperature-
senstive electro-optical components. The oxide forms when chosen gas phase
reactants absorb selected wavelengths of light. One major benefit of this
process is that it's free from charged species that can damage the substrate.
Because it is done at low temperatures (50° to 300°C), it is useful for making
certain solid-state devices. The process has been used to form pinhole-free
dielectrics for temperature-sensitive elemental and compound semiconductor
materials. The PHOTOX process is available for non-exclusive licenses.

A new all-optical logic device could make many electronics systems immune to
effects of natural or man-made "noise," including lightning strikes and radio
interference. Hughes scientists have fabricated a high-speed optical device
that uses no electronic signals. It is made of discrete components, including
four reflecting surfaces and a slab of non-linear material (gallium arsenide).
The device has shown optical bistability (flip-flop behavior) with switching
times of 3 nanoseconds and switch energies under 100 microjoules. Although
propagation delays have kept the device's speed under the theoretical limit up
to 10 gigahertz, the speed will be increased by further miniaturization.

Eastern Airlines has eliminated lengthy and costly delays by using an automatic
tester to check electrical systems on its L-1011 aircraft. Eastern's main-
tenance center in Miami has equipped a mobile van with a Hughes FACT® (Flexible
Automatic Circuit Tester) system. Now technicians can perform thousands of
high-voltage tests per minute in hangars or on taxi strips. Once faults are
pinpointed, repairs can be made quickly and the aircraft restored to service.

Hughes Industrial Electronics Group offers the advantages and opportunities of a
small company backed by the resources of a $2-billion company. Our facilities
are in the Southern California communities of Carlsbad, Irvine, Newport Beach,
Torrance, and Sylmar. Our programs incorporate 34 different technologies. They
include silicon and GaAs semiconductor technologies, fiber optics, microwave and
millimeter-wave communications, microprocessors, lasers, and solar cells. Send
resume to B.E. Price, Hughes Industrial Electronics Group, Professional Employ-
ment, Dept. SSA, P.O. Box 2999, Torrance, CA 90509. Equal opportunity employer.

Japan's newest weather satellite, now keeping watch over the vast western
Pacific region, represents strong international teamwork. GMS-2 (Geostationary
Meteorological Satellite) was designed, developed, and built by Hughes under
contract to Nippon Electric Company. Nippon Electric assembled many of the
satellite's electronic units and built a scientific instrument called a Space
Environment Monitor, which measures energetic protons in space. Sharp Corpora-
tion of Japan provided a portion of the solar cells. GMS-2 was procured and
launched by Japan's National Space Development Agency. It is operated by the
Japanese Meteorological Agency as part of the World Weather Watch.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY
CULVER CITY, CALIFORNIA 90230

(213) 670-1515 EXTENSION 5964
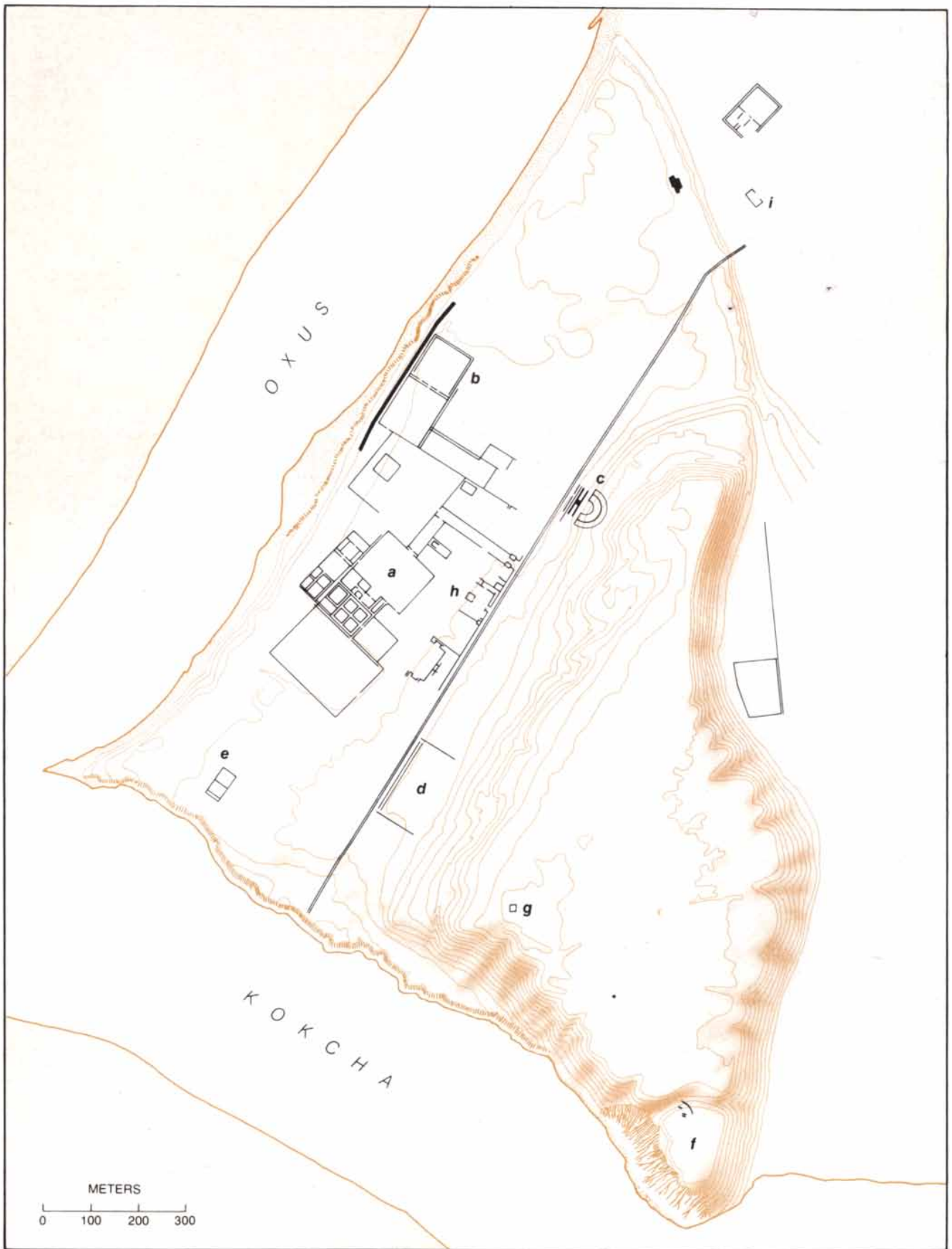
CITY SITE, roughly triangular in outline, lay upstream from the confluence of the Oxus and Kokcha rivers, where a natural acropolis stands 60 meters higher than the rest of the plain. The main street ran from the north gate along the foot of the rise leading to the acropolis. The great palace enclosure and most of the other elite structures occupied the level area between the street and the left bank of the Oxus. Outlined are the palace enclosure (*a*), the gymnasium (*b*), the theater (*c*), the arsenal (*d*), one of the elite residences (*e*), the citadel on the acropolis (*f*), the open-air temple platform there (*g*), the temple on the main street (*h*) and the temple outside the north gate (*i*).

without benefit of a tub, washed with water poured out of a large vessel. The bathwater, both cold and hot, came from the third room; a stove in the adjoining kitchen served for heating the water.

According to Greek custom, most of the city's dead were buried outside the walls, the elite in family mausoleums with multiple vaults lining both sides of a central passageway. The colonists were also familiar, however, with another Greek custom that allowed a city's benefactors to be buried within its walls. This privilege had been granted to two citizens whose tombs, replicas of small temples, stood near the entrance to the palace. The later and more monumental of the tombs was lined on all sides with rows of stone columns. The earlier and simpler tomb, its facade once decorated by two modest wood columns, held a certain Kineas, one of the city's founding fathers. A stone stele stood in front of its porch. The upright itself has disappeared, but around its base are engraved the last five maxims in a series of 150 that appear in Apollo's sanctuary at Delphi. They define the cardinal virtues of the ideal Greek. "In childhood," the maxims read, "learn good manners; in youth learn to control your passions; in middle age learn to be just; in old age learn to be of wise counsel; die without regret." Presumably the missing stele displayed all the other maxims. Their public exhibition in this place of honor offers striking testimony to the colonists' desire to remain loyal to their distant heritage.

Other evidence of that desire is present in the ruins of Ai Khanum. For example, although inscriptions are not plentiful, those that have been found show that the settlers continued to speak and write Greek correctly during their 200 years of residence in central Asia. Even the kinds of lettering that were used, both ornamental and cursive, show the same sequence of development that appears on monuments in Greece.

Another example of this cultural continuity comes from one of the treasury storerooms that evidently served as a library. Several papyrus manuscripts apparently escaped the attention of the conquering nomads. Although the manuscripts themselves had long ago crumbled to dust, the ink of their words had left its imprint in places, making it possible to decipher fragments of text. The fragments include part of a philosophical treatise of the Aristotelian school and parts of a Greek poem. Thus we see that the occupants of the palace, in the best tradition of the Hellenic elite, were readers of philosophical and literary works.

The find gives some credence to Plutarch's assertion, made centuries later, that after Alexander's Asian expedition the people of the eastern territories—at least the ruling aristocracy—read Homer and their children sang the tragedies of Sophocles and Euripides. At least it may be deduced from a fountain spout carved to represent the mask worn by the cook-slave character in Greek comedies that the plays performed in the theater at Ai Khanum belonged to the Greek repertory.

This is not to say that all who lived in the city were Greek settlers or their descendants. Many of the names found in the inscriptions are Greek, but some of them, such as Lysanias, are typical of Macedonia and some are typical of other areas of northern Greece. Triballos is one of them, as is Kineas, whose tomb stood near the palace entrance. Among the accountants who worked in the palace were men with names of local ori-
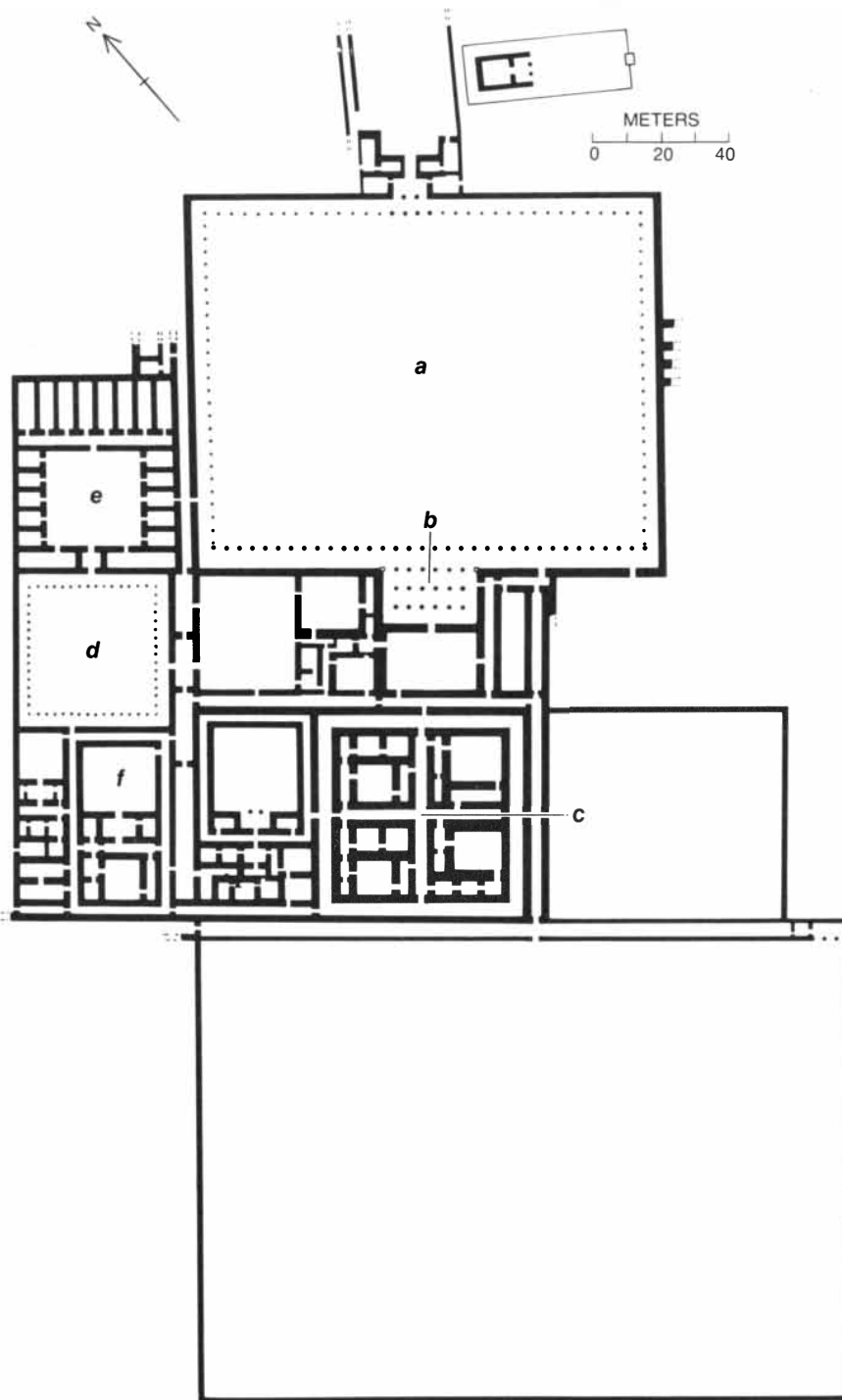
PALACE ENCLOSURE contained a great formal courtyard (*a*) with a vestibule (*b*) that gave access to the administrative buildings. The most notable of these (*c*), a square structure with its interior symmetrically divided by two corridors that cross at a right angle, evidently held two reception areas, at the right, and two suites of offices, at the left. A smaller courtyard (*d*) offered a more informal setting; beyond it in one direction stood the treasury with many storerooms around a central court (*e*) and in the other direction one of three private apartments (*f*).

157

IMPRINT OF INK from a long-vanished Greek manuscript was discovered on the dirt floor of a storeroom in the palace treasury that may have served as a library. A reading of this fragment and a similarly preserved one shows that the Greek elite read poetry and philosophy.

THEATER MASK inspired a fountain spout; this representation shows the traditional mask of the cook-slave character in Greek comedies that may have been performed at Ai Khanum.

gin. For example, two of them—Oxeboakes and Oxubazes—were named after the local river deity, Oxus. Evidently the new masters of Bactriana found it expedient to retain the native bureaucrats trained by the former rulers of the province. It seems safe to assume that the same reasoning was applied to local workers of lower status. A number of modest one-room houses on the acropolis could have been those of local people who were not allowed to settle in the lower town.

In the arts the taste of the colonists remained traditionally Greek, even to the point of perpetuating an outdated Classical style. For example, the mosaics were made up of pebbles rather than the square-cut stones of the later style. The pebbles were widely spaced, in many instances with red stones forming the ground and white stones the figures, and in some instances with black stones inserted for accents. The repertory of motifs was conventional: floral displays, dolphins, sea horses and sea monsters. Moreover, their treatment was flat and lacking in any effort to create an impression of volume or space.

The same may be said of the smaller sculptures that have been found: a woman leaning on a pillar, a man with a wreath of leaves on his head and a bas-relief of a young man with a mantle thrown back over his shoulders. The workmanship is excellent but the treatment is conventional. The Greco-Bactrian sculptors must nevertheless be given credit for one innovation. When they made monumental statues or mural bas-reliefs, they exploited a technique rarely seen in Greece. They first built an armature of lead rods or wood and then modeled the figure in clay or stucco over it.

Cult figures were more susceptible to Eastern influences. For example, figurines modeled in terra-cotta or carved out of ivory represent a female deity. She is shown either fully clothed and laden with jewels or completely nude and quite voluptuous. A gold-covered silver plate bears a religious relief that also shows local influences. Cybele, the Greek goddess of nature, is seen riding in a chariot drawn by lions and driven by another traditional female Greek figure, a winged goddess of victory. Two priests are in the scene, and in the sky shine the sun (anthropomorphically personified as Helios), the moon and a star. The iconographic elements are mostly Greek, but the artist's execution is in accord with Oriental canons. The composition is flat, with no effort at perspective, and the figures are presented either head on or in profile.

Although the official deities of Bactriana, for example those represented on the coinage, belong to the Greek pantheon, the three temples discovered at Ai

Khanum are not Greek at all. The principal one, on the main street near the palace, is a massive square structure that stands on a high three-step podium. The outside wall is decorated with niches, and a wide vestibule opens into a smaller room, where a statue of the deity once stood; the smaller room is flanked by two narrow chambers. Architecturally the building reflects nothing of Greek tradition. The podium is borrowed from Persian and central Asian models; the plan of the building and its exterior decoration are derived from the same two sources and from Mesopotamia. Libation vessels buried at the base of the podium are evidence that the rites performed here were not Hellenic. Although a few fragments of the cult statue have survived, there are not enough of them to indicate whether the deity was male or female. A broken foot wears a Greek sandal, and so the god or goddess may have been Greek in appearance, but its nature was surely in part Oriental.

The second temple stands outside the city wall near the north gate. It strongly resembles the first temple except that three cult rooms instead of one open off the vestibule and the vestibule was left unroofed. The third temple is equally non-Hellenic. It is a large stepped platform at the southwest corner of the acropolis. Greek sources state that the Persians were in the habit of worshiping in the open air at sites chosen for their high elevation; this sacred structure on the acropolis fits such a Middle Eastern model.

Our expedition's concept of what might be found at Ai Khanum when the investigation began in 1965 has changed profoundly in the 15 years of excavation. No one expected to find what had obviously been a powerful and wealthy capital city in this remote corner of Central Asia. Nor had anyone anticipated, on the one hand, the Greek settlers' stubborn loyalty to Hellenic culture in most respects and, on the other, their wholehearted acceptance of Eastern practices in such matters as architecture and religious worship.

Substantial evidence bearing on the culture of pre-Hellenistic Central Asia, for example the discoveries at Altyn in the U.S.S.R., is now coming to light. Such findings reinforce a major conclusion we have drawn from our work at Ai Khanum. This conclusion is that when the Greeks arrived in this part of the world, it was already the seat of a highly developed urban civilization with an ancient tradition of monumental architecture. As the Greek colonists on the banks of the Oxus strove to preserve their cultural identity they were also prepared to assimilate the lessons that could be learned from the local civilization.
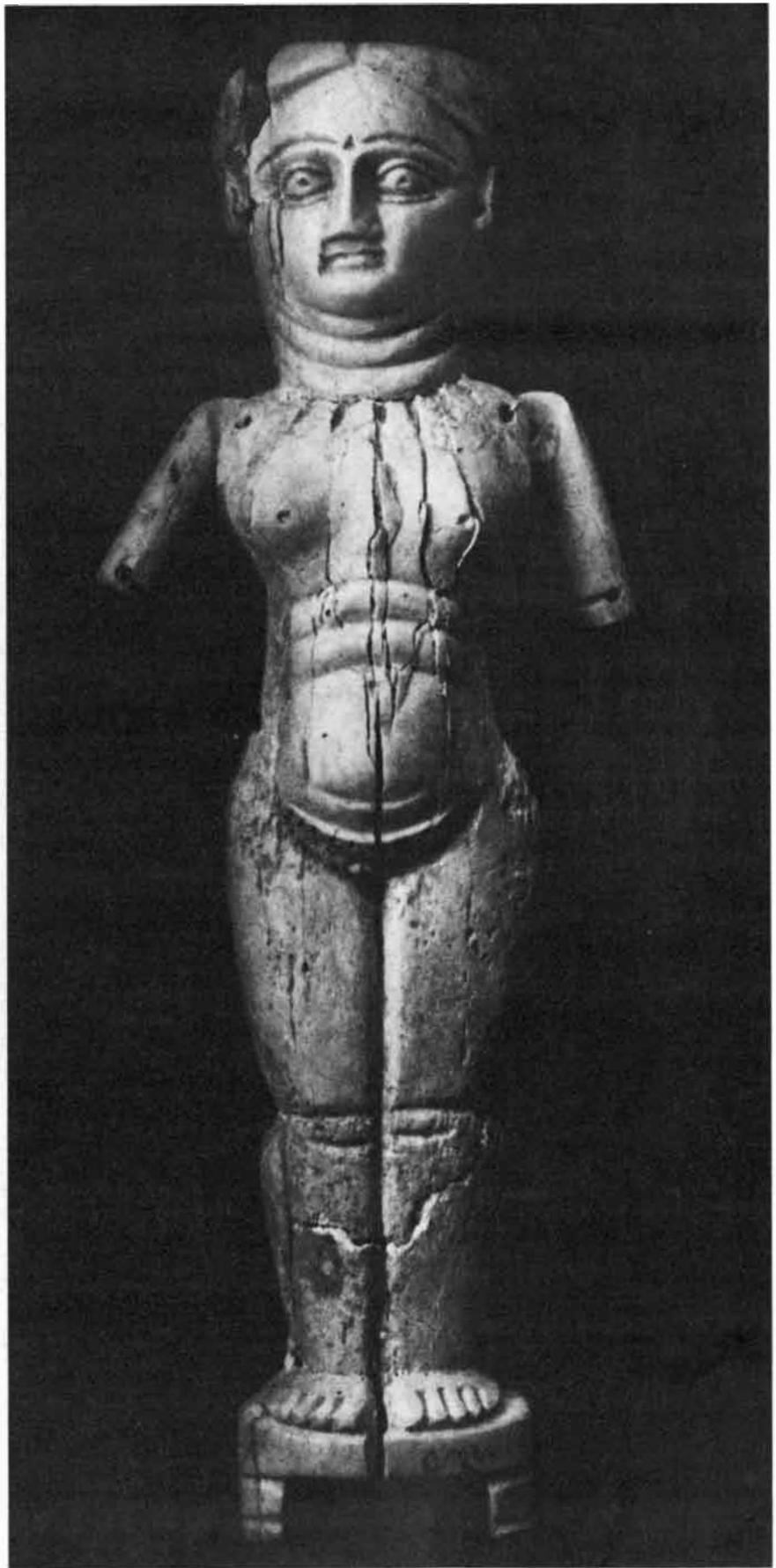
BONE FIGURINE, evidently a representation of a local goddess, shows no debt to Greek artistic traditions. Local influences also left their mark on the architecture of Ai Khanum.

159

The Psychology of Preferences

*When people make risky choices, they often do not do so objectively.
Experimental surveys indicate that such departures from objectivity
tend to follow regular patterns that can be described mathematically*

by Daniel Kahneman and Amos Tversky

Imagine yourself on your way to a Broadway play with a pair of tickets for which you have paid $40. On entering the theater you discover that you have lost the tickets. Would you pay $40 for another pair of tickets? Now imagine you are on your way to the same play without having bought tickets. On entering the theater you realize that you have lost $40 in cash. Would you now buy tickets to the play?

In objective terms the two situations are identical: in both cases you are poorer by $40 than you were earlier, and you face the decision of whether or not to pay $40 to see the play. Nevertheless, most people presented with such a situation say they would be more likely to buy new tickets if they had lost money than if they had lost tickets. One interpretation of the difference between the two situations is that the same loss can be assigned to different "mental accounts." The loss of $40 in cash is entered in an account distinct from that of the play. The loss therefore has little effect on the decision to buy new tickets. In contrast, the cost of the lost tickets is posted to the account of the play. The unexpected doubling of the cost of the play is difficult to accept.

Recent investigations of the psychology of preferences have demonstrated several intriguing discrepancies between subjective and objective conceptions of decisions. For example, the threat of a loss has a greater impact on a decision than the possibility of an equivalent gain. Most people are also very sensitive to the difference between certainty and high probability and relatively insensitive to intermediate gradations of probability. The regret associated with a loss that was incurred by an action tends to be more intense than the regret associated with inaction or a missed opportunity. These observations and others of a similar character contribute to the understanding of how people make decisions and to the elucidation of some puzzles of rational choice.

The origin of the psychology of preferences can be traced to an essay published in 1738 by the mathematician Daniel Bernoulli. In the essay Bernoulli discussed a widespread characteristic of human preferences: risk aversion. To understand risk aversion, imagine that you are given a choice between two options. The first is a sure gain of $80. The second is a risky prospect that offers an 85 percent chance of winning $100 and a 15 percent chance of winning nothing.

Most people who are presented with this choice prefer the certain gain to the gamble, in spite of the fact that the gamble has a higher "monetary expectation" than the certain outcome. The monetary expectation of a gamble is the sum of its outcomes weighted by their probabilities. The monetary expectation of the gamble offered in this case is $100 multiplied by its probability (.85), added to $0 multiplied by its probability (.15), or a total of $85. The monetary expectation reflects the average monetary value of the gamble; if one were to play this gamble many times, the average gain would be about $85 per play. The monetary expectation of the certain gain is $80 multiplied by 1 (certainty), or $80. A choice is risk-averse if a certain outcome is preferred to a gamble with an equal or greater monetary expectation. A choice is risk-seeking, on the other hand, if a certain outcome is rejected in favor of a gamble with an equal or lower monetary expectation.

The hypothesis that people generally make risk-averse choices has been widely accepted by economists who normally assume that a consumer or an entrepreneur will choose a risky venture over a sure thing only when the monetary expectation of the venture is sufficiently high to compensate the decision maker for taking the risk. Psychological studies, however, indicate that risk-seeking preferences are common when people must choose between a sure loss and a substantial probability of a larger loss.

To get a sense of risk seeking imagine you are forced to choose between a sure loss of $80 and a risk that involves an 85 percent chance of losing $100 and a 15 percent chance of losing nothing. Faced with this choice a large majority prefer the gamble to the loss. The monetary expectation of the gamble (− $85), however, is worse than that of the certain loss (− $80). The majority preference is thus an instance of risk seeking.

The contrast between the majority choices observed in these two problems suggests that preferences between gains are risk-averse and that preferences between losses are risk-seeking. This pattern has been observed both in answers to hypothetical questions and in real decision problems where people were paid according to their choices. The same pattern has been confirmed for hypothetical problems that involve outcomes other than money, such as duration of pain and the number of lives that may be lost in an epidemic or saved by medical intervention.

When people make risk-averse or risk-seeking choices they forgo the option that offers the highest monetary expectation. In order to explain such choices Bernoulli replaced the objective criterion of monetary expectation with the subjective criterion of expected utility. According to expected utility theory, each outcome gives rise to a particular degree of pleasure or utility. The utility of a risky prospect is the weighted sum of the utilities of its outcomes, each multiplied by its probability. The central idea of the theory is that utility is not a linear function of money: a gain of $2,000 contributes less than twice as much to utility as a gain of $1,000. As a consequence the prospect with the highest monetary expectation does not necessarily have the highest expected utility. The decision maker is assumed to select the option with the highest utility, whether or not that option also has the greatest monetary expectation.

How do people identify the outcomes of a decision? It has generally been assumed, following Bernoulli, that utilities are assigned to states of wealth. We depart from this tradition and analyze choices in terms of changes of wealth rather than states of wealth. The classical analysis implies that preferences reflect a comprehensive view of the options. In contrast, we propose that people

**Here's why
the aluminum can
is America's favorite
recyclable beverage
container.**

Lightweight, stackable, quick-chilling and unbreakable, the all-aluminum can preserves the flavor and protects the taste of America's best-loved beverages.

And when it comes to saving energy, remember that 15 billion cans were recycled last year. And every time we recycle, we save 95 percent of the energy needed to make new metal from ore.

The next time you shop, look for the all-aluminum, all-American can. Its recyclability ensures many happy returns. For more information, write Aluminum Company of America, 303-A Alcoa Building Pittsburgh, PA 15219.

We can't wait for tomorrow.
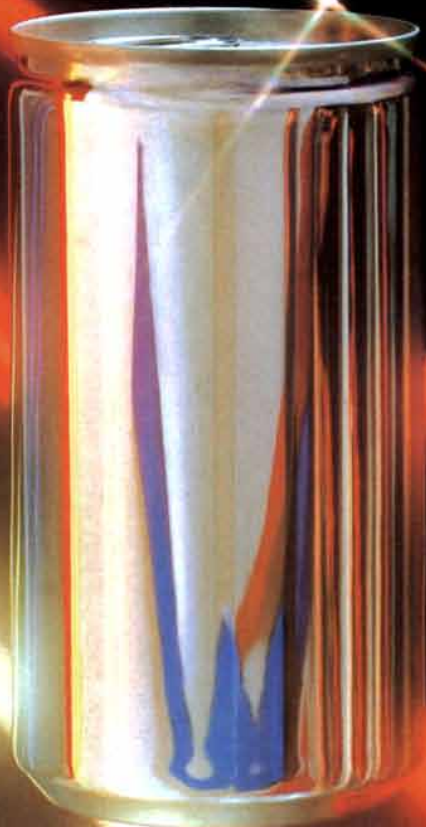
◨ ALCOA

commonly adopt a limited view of the outcomes of decisions: they identify consequences as gains or losses relative to a neutral point. This form of mental accounting can lead to inconsistent choices, because the same objective consequence can be evaluated in more than one way.

To explain choices we introduce a function that associates a subjective value to any amount that may be gained or lost. We call it a value function. As in classical utility theory, risk seeking and risk aversion are explained by the curvature of the function that relates subjective values to objective outcomes. Consider first the value function for gains. We assume that gains have positive value and that a zero gain, which merely retains the status quo, has a subjective value of zero. We also propose that the value function for gains is concave downward, so that each extra dollar gained adds less to value than the preceding one. The concavity of the value function is compatible with the common impression that the difference between a gain of $100 and one of $200 is more significant than the difference between a gain of $1,100 and one of $1,200. This shape of the value function favors risk aversion.

Consider the choice between a sure gain and a gamble that offers some probability of a larger gain and some probability of no gain at all. The advantage of the large gain over the sure thing is evaluated in a shallow region of the value function, where increments of money produce relatively small increments of value. The advantage of the sure outcome over no gain is evaluated in the steepest region of the function, where each dollar makes a greater difference. Because value is nonlinear the sure outcome is closer to the large gain in terms of value than in terms of money. The shape of the value function thus favors a risk-averse preference for sure outcomes over risky prospects [*see illustration below*].

To explain the common observation of risk-seeking preferences in choices that involve losses, we assume that the value function for losses is convex, so that each extra dollar lost causes a smaller change in value than the preceding one. This proposal is compatible with the common impression that the difference between a loss of $100 and one of $200 appears more significant than the difference between a loss of $1,100 and one of $1,200. The convexity of the value function for losses favors a risk-seeking preference for a gamble over a sure loss. The advantage of the best outcome of the gamble (no loss) over the sure loss is evaluated in the steepest region of the value function. On the other hand, the advantage of the sure loss over the worst outcome of the gamble is evaluated in a flatter region. As a result the sure loss is relatively closer to the worst outcome on the value scale than it is on the money scale [*see illustration on page 164*].

The properties of the value function can be inferred from observed choices. To understand how this procedure works imagine you have been offered a choice between winning a cash prize and having a 50 percent chance of winning $100. What amount would make the sure prize just as attractive as the bet? A choice of $35 as the matching prize is common. If you find $35 as attractive as the gamble, your value for this amount must equal your value for the gamble. On the assumptions that we introduced above the value of the gamble is 1/2 the value of $100, since the probability of winning $100 is 1/2 and the value of the other outcome (winning nothing) was assumed to be zero. Consequently the choice of $35 to match the gamble indicates that the value of $35 is half the value of $100.

Now find the prize amounts that match a 50 percent chance of winning $200, $500, $1,000 and $2,000. If you try this thought experiment, you will naturally find that the amount of the prize increases with the size of the bet. You may also observe that the prize is roughly proportional to the stake. A person who matches a prize of $35 to an even chance of winning $100 is likely to match a prize of $300 or $350 to an even chance of winning $1,000. Thus as the amount of the bet has increased by a factor of 10 the prize has grown by almost the same factor.

It can be proved mathematically that the value function of an individual who sets prizes exactly proportional to stakes can be described by a power function. In such a function the value of an amount of money equals that amount, raised to some power. The exponent of the value function can be estimated from choices. If the prize that matches a 50 percent chance of winning an amount of money is always 35 percent
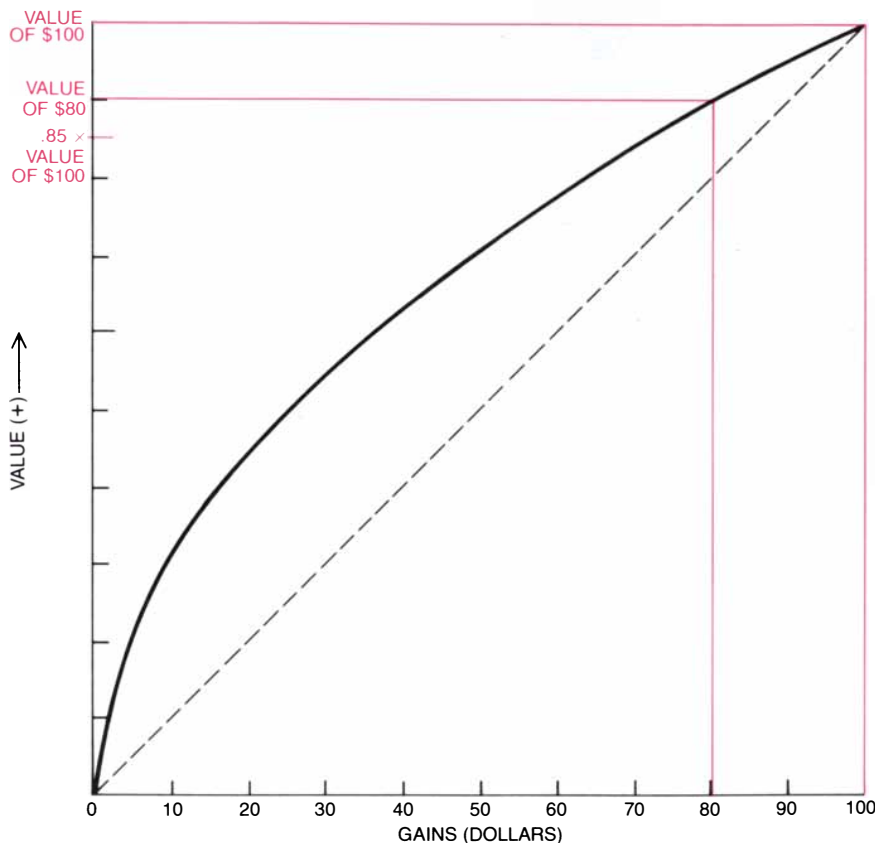
CONCAVE VALUE FUNCTION helps to explain preferences where gains are concerned. A value represents the subjective attractiveness of a gain (or loss). The function assigns a value to each possible gain. Gains are indicated on the horizontal axis, values on the vertical axis. The status quo is assigned the value of zero. (The units of value are arbitrary.) The concave function becomes progressively flatter as the amount of the gain is increased. Risk-averse preferences are common when people make choices among gains. In a risk-averse choice a sure gain is preferred to a gamble with a higher monetary expectation. The monetary expectation of a gamble is the sum of its outcomes multiplied by their probabilities. Hence an 85 percent chance of winning $100 has a monetary expectation of .85 multiplied by $100, or $85. This is greater than the monetary expectation of a sure gain of $80 ($80 multiplied by certainty, or $80). The value of the sure gain, however, is greater than that of the gamble, which is .85 multiplied by the value of $100. The value function thus predicts that the sure gain will be preferred to the gamble.

162

Come to where ideas soar.

The rush. The whistling wind of excitement that comes with reaching a frontier that's never been reached. Having an idea that's never been had before. Touching your own limits.

That's your opportunity at Phillips Petroleum. Where our people pioneered the huge Norwegian North Sea oil fields when many doubted there was any oil there to find. Where we've led our industry in number of patents issued 12 of the last 13 years. Where we allow computer science to become an art, and we hold patents on some of the most advanced hardware and software in the world. Where recognition can take many forms – from crystal trophies to cash awards and almost boundless chances for more achievement.

And where now we're looking for people in engineering, law, geoscience, computer science, technical sales, data processing, business, tax, accounting, and chemistry to join some of the best people at their jobs in the business. Maybe any business.

The opportunity is here. Achievement is the standard. Talk to us, for a chance to spread your wings to the fullest.

PHILLIPS 66

Send resumes in confidence to W. M. Hutchison – Recruitment, Box 9AH, Phillips Petroleum Company, Bartlesville, Oklahoma 74004
Phillips is an equal opportunity employer

of that amount, the exponent of the value function for gains is about 2/3. Actually the proportionality of prizes to stakes is only approximate. The prize generally increases more slowly than the stake. Because the deviation from proportionality is small, however, a power function provides a convenient description of the values of gains.

A similar analysis can be applied to losses. An even chance of losing $100 or losing nothing is often matched by a sure loss of about $40, and an even chance of losing $1,000 or nothing by a sure loss of about $400. If the sure loss that matches a 50 percent chance to lose an amount is always 40 percent of that amount, the exponent of the value function for losses is about 3/4.

In order to piece together the value functions for both gains and losses it is necessary to study risky prospects that involve both positive and negative outcomes. Consider a gamble in which you face a 50 percent chance of losing $100 and a 50 percent chance of winning a cash prize. What is the smallest prize that would make this bet acceptable to you? It is a commonplace that the pleasure of winning a sum of money is much less intense than the pain of losing the same sum. Accordingly people accept a gamble at even odds only when the possible gain from the gamble is substantially larger than the possible loss. For example, you may find that a 50 percent chance to lose $100 is unacceptable unless it is combined with an equal chance to win $200 or more.

By observing the choices people make among gains and losses we have specified some features of the value function. The approximate proportionality of prizes and bets is reflected in the power relation between gains and their values. Risk aversion and risk seeking are described by the concave and the convex segments of the value function. The asymmetry in the response to gains and to losses is expressed in the greater steepness of the value function for losses. It is important to emphasize, however, that the value function is merely a convenient summary of a common pattern of choices and not a universal law.

Naturally individuals differ in their attitudes toward risk and toward money, and no single value function can describe the preferences of all individuals. Choices also can vary substantially with the method used to elicit preferences. For example, the acceptability of a risky prospect may depend on whether the respondent matches a gain to a set loss, matches a loss to a set gain or adjusts the odds when the amounts at stake are fixed. In addition the near proportionality of prizes to stakes breaks down beyond the range of moderate gains and losses. For extremely large gains the value function becomes practically flat as individuals can no longer distinguish one huge gain from another. In the negative domain, however, the function can be very steep in the vicinity of monetary losses that would force a substantial change in a person's way of life.

The discussion to this point has assumed that the contribution of each outcome to the value of a risky prospect is weighted by its probability. There is evidence that this assumption is often violated and that the "decision weights" that multiply the values of outcomes do not coincide with the probabilities.

Some properties of decision weights can be explored by the following informal experiment. Imagine that you can improve your chances to win a very desirable prize. Would you pay as much to raise your chance of winning from 30 to 40 percent as you would to raise your chance from 90 percent to certainty? It is generally agreed that the former offer is less valuable than the latter. It is also agreed that an increase from impossibility to a probability of 10 percent is more significant than an increase from 30 to 40 percent. Thus the difference between certainty and possibility and the difference between possibility and impossibility loom larger than comparable differences in the intermediate range of probability [see illustration on page 140].

Investigations of risky choices tend to confirm these hypotheses. Low probabilities are commonly overweighted but intermediate and high probabilities are usually underweighted relative to certainty. The impossible event is naturally assigned a weight of zero and the certain event is assigned a weight of one. The overweighting of small probabilities can give rise to risk seeking in the positive domain and risk aversion in the negative domain, in contrast to the prevalent pattern of preferences described above.

The inflated effect of small probabilities contributes to the appeal of lottery tickets and accident insurance, which are concerned with events that are highly significant and relatively improbable. The underweighting of intermediate and high probabilities, on the other hand, reduces the attractiveness of possible gains relative to sure ones. It also reduc-

LOSSES (DOLLARS)

−100 −90 −80 −70 −60 −50 −40 −30 −20 −10 −0

VALUE (−)

.85 × VALUE OF −$100

VALUE OF −$80

VALUE OF −$100
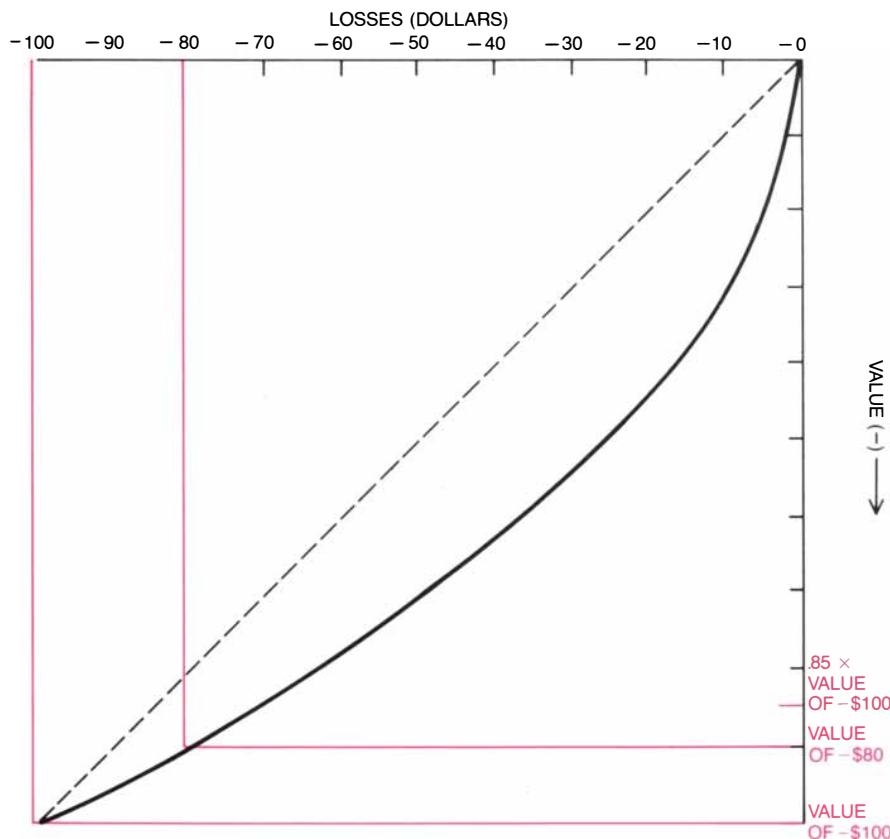
CONVEX VALUE FUNCTION explains preferences where losses are concerned. The function assigns a unique negative value to each monetary loss. Losses are indicated on the horizontal axis to the left of the origin; values are shown on the vertical axis downward. The status quo is assigned the value of zero. The convex function becomes progressively flatter for larger losses. Risk-seeking preferences are common when people make choices among possible losses. In a risk-seeking choice a gamble is preferred to a sure loss that has a greater monetary expectation. The value function predicts that a gamble offering an 85 percent chance of losing $100 and a 15 percent chance of losing nothing will be preferred to a sure loss of $80. The monetary expectation of the gamble (−$85) is less than that of the sure loss (−$80). The value of the gamble, however (which is .85 multiplied by the value of −$100), is greater (has a smaller negative magnitude) than that of the sure loss. Because the gamble has a greater value it will be preferred. The convex value function therefore helps to explain risk-seeking preferences.

164

The lens that sets all standards in photography is standard on only one video camera.

Toshiba has been responsible for so many firsts in home video, it's really no surprise that we've created the first video camera equipped with a Nikon lens.

It's the Toshiba IK-1900 color video camera, with a Nikkor 12.5 mm~100 mm, f1.6 macro zoom lens. And it's going to change recorded history. Because rather than just revamp the video camera, Toshiba reinvented it.

The electronic viewing system has been eliminated (which, incidentally, decreases weight while increasing battery life by 25%) and replaced with our new True Image system.

You look directly through the lens at virtually 100% of the image, something even network studio cameras can't promise. Since you focus on precisely what you see, and not on a black and white electronic image, the subject literally shows its true colors.

Thanks to the True Image system, even the way you adjust the focus is superior. There is no split image to rob you of light.

Instead, a fresnel field system sharpens the image to crystal clarity with only the slightest rotation of the focusing barrel.

Working with Toshiba, Nikon also perfected macro/zoom capability all the way to 0.0 cm. focusing, a feature found on no other video camera. It means, although we don't suggest you try it, that you can actually touch the lens to the subject while keeping it in crisp focus.

Toshiba also gives you the highest resolution of any compact home video camera. And even a unique diopter adjustment feature, which allows people who wear eyeglasses to remove them while using the camera.

The Toshiba IK-1900 color video camera with the amazing Nikkor TV 8:1 macro zoom lens. Anyone interested in the very best home video should consider it standard equipment.

Underexposure   Auto white balance

U   W

V   B

Recording   Battery condition

TOSHIBA
Again, the first.

Toshiba America, Inc. 82 Totowa Road, Wayne, N.J. 07470

Simulated viewfinder image

es the threat of possible losses relative to sure ones. The contribution of decision weights was neglected in the construction of the value function we described above. In that function all observed risk aversion and risk seeking were assumed to be the result of curvature of the value function. If the probability of .5 that was applied in the construction of the function were replaced by a slightly smaller decision weight, the value function would become less curved, although it would retain its general *S* shape.

We have so far been concerned with rules that govern the evaluation of risky options. Another important aspect of the psychology of preferences is how people define the consequences of their choices. The same decision can be framed in several different ways; different frames can lead to different decisions. For example, consider the following problem:

Imagine that in addition to whatever else you have, you have been given $200. You are now asked to choose be-

S-SHAPED VALUE FUNCTION combines a concave segment representing the values of gains and a convex segment representing the values of losses. The concave segment to the right of the origin reflects risk aversion in choices between gains; the convex segment to the left reflects risk seeking in choices between losses. The segment for gains represents preferences in which a 50 percent chance to win a given amount is just as acceptable as a sure gain of 35 percent of that amount. For example, a 50 percent chance of winning $1,000 is as acceptable as a sure gain of $350. The value of $350 is therefore half the value of $1,000. Such a pattern of preferences can be expressed by a power function with an exponent of 2/3. (In a power function the value of a gain is equal to the magnitude of the gain raised to a fixed exponent.) The segment for losses represents preferences in which a 50 percent chance of losing a given amount is just as acceptable as a sure loss of 40 percent of that amount: a 50 percent chance to lose $1,000 is as acceptable as a sure loss of about $400. The negative value of −$400 is therefore half the negative value of −$1,000. Preferences where losses are concerned can be expressed by a power function with an exponent of 3/4. The S-shaped function reflects the common observation that a loss has a greater subjective effect than an equivalent gain. Because the slope of the function for losses is steeper than that for gains, the negative value of a loss of $100 is equal to the positive value of a gain of $200. The function therefore represents the preference of a person who finds an even chance of winning $200 or of losing $100 just acceptable; this is a typical pattern. The function shown summarizes some common features of preferences observed in tests of many people; individual value functions can vary greatly.

tween (*A*) a sure gain of $50 and (*B*) a 25 percent chance of winning $200 and a 75 percent chance of winning nothing.

Most people given these options make a risk-averse choice, preferring the sure gain (*A*) over the gamble (*B*), which has the same expected value. Now consider the following problem:

Imagine that in addition to whatever you have, you have been given a cash gift of $400. You are now asked to choose between (*C*) a sure loss of $150 and (*D*) a 75 percent chance of losing $200 and a 25 percent chance of losing nothing.

Most people have a risk-seeking preference for the gamble (*D*) over the sure loss (*C*). The options presented in the two problems are identical, however, in objective terms. There is no valid reason to prefer the gamble in one version and the sure outcome in the other. Choosing the sure gain in the first problem yields a total gain of $200 plus $50, or $250. Choosing the sure loss in the second version yields the same result through the deduction of $150 from $400. The choice of the gamble in either problem yields a 75 percent chance of winning $200 and a 25 percent chance of winning $400.

If our respondents took a comprehensive view of the consequences, as is assumed by theories of rational decision, they would combine the bonus with the available options and evaluate the composite outcome. Instead they ignore the bonus and evaluate the first problem as a choice between gains and the second as a choice between losses. The reversal of preferences is induced by altering the description of outcomes. We call such reversals framing effects.

Framing effects arise when the same objective alternatives are evaluated in relation to different points of reference. We asked a large number of physicians to consider the following problem:

Imagine that the U.S. is preparing for the outbreak of a rare Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows: If Program *A* is adopted, 200 people will be saved. If Program *B* is adopted, there is a 1/3 probability that 600 people will be saved and a 2/3 probability that no people will be saved. Which of the two programs would you favor?

The majority response to this problem is a risk-averse preference for Program *A* over Program *B*.

Other respondents were presented with the same problem but a different formulation of the programs: If Program *C* is adopted, 400 people will die. If Program *D* is adopted, there is a 1/3 probability that nobody will die and a

166

Qantas perfects Business Class.

Quick check-in at the First Class counter. Priority baggage handling.

Whenever possible, a drink on us at the Captain's Club before boarding.

A special section aboard your Qantas 747. Wide chairs in pairs so you're never more than one seat from an aisle.

Free bar service and in-flight entertainment.

A special menu with a choice of entrées and fine wines, followed by a selection of liqueurs.

Enjoy these and other first-class touches on Qantas Business Class. Twelve flights a week to Australia. Three flights a week to Tahiti.

Ask your Travel Agent about Qantas Business Class. Only $100 or so more than the Economy Class Fare to Australia.

And that's perfect.

"Nobody's perfect, Qantas. You made the bootees too big."
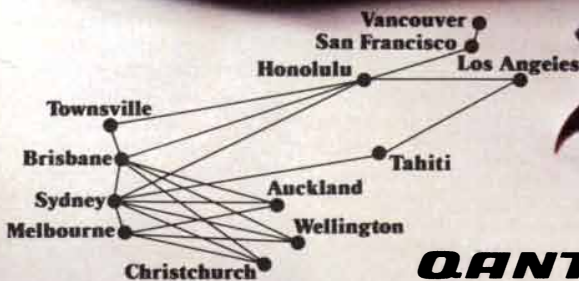
QANTAS The Australian Airline.

2/3 probability that 600 people will die.

The majority choice in this problem is risk-seeking: the certain death of 400 people is less acceptable than a 2/3 chance that 600 people will die.

It is easy to see that the two versions of the problem describe identical outcomes. The only difference is that in the first version the death of 600 people is the normal reference point and the outcomes are evaluated as gains (lives saved), whereas in the second version no deaths is the normal reference point and the programs are evaluated in terms of lives lost. Because of the *S*-shaped value function and the overweighting of certainty the two frames elicit different preferences.

The decisions we have discussed so far involve a single dimension. Many decisions, however, concern transactions, in which the possible outcomes include connected changes in several dimensions of value. The basic example of a transaction is the purchase of goods after which one has more goods and less money than before. A transaction must be evaluated according to the balance of costs and benefits in a mental account.

The framing of a transaction can alter its attractiveness by controlling the costs and benefits that are assigned to its account, as in our example of the theater tickets. Going to the theater is normally framed as a transaction, in which the cost of the ticket is exchanged for the experience of seeing the play. If the ticket has been lost, buying a new one effectively doubles the cost of the play. In contrast the loss of the cash is not considered a debit to the account of the play. The cash loss affects the decision to buy a new ticket only to the extent that a minor reduction of wealth reduces the tendency to make optional purchases. When we asked a sample of students a version of this question involving a single lost ticket, the majority of those who were asked to imagine they had lost the money said they would buy a ticket, but the majority of those who were asked to imagine they had lost the ticket said they would not.

Mental accounting is dominated by the tendency to group the costs and benefits associated with an object, as the following example illustrates:

Imagine you are about to buy a jacket for $125 and a calculator for $15. The calculator salesman tells you that the calculator you want to buy is on sale for $10 at the other branch of the store, a 20-minute drive away. Would you make the trip to the other store?

The majority of the respondents who answered this question said they would drive to the other store. Another group answered a similar question in which the cost of the jacket was changed to $15 and the cost of the calculator was changed to $125 in the original store and $120 in the other branch. Of the respondents presented with this version, the majority said they would not make the trip. The total purchase and the consequences are the same in both versions: one has to decide whether to drive 20 minutes to save $5. The contrasting choices in the two versions suggest that the respondents evaluated the saving of $5 in relation to the price of the calculator. In relative terms a reduction from $15 to $10 is more impressive than a reduction from $125 to $120.

Framing effects in consumer behavior may be particularly pronounced in situations that have a single dimension of cost (usually money) and several dimensions of benefit. An elaborate tape deck is a distinctive asset in the purchase of a new car. Its cost, however, is naturally treated as a small increment over the price of the car. The purchase is made easier by judging the value of the tape deck. independently and its cost as an increment. Many buyers of homes have similar experiences. Furniture is often bought with little distress at the same time as a house. Purchases that are postponed, perhaps because the desired items were not available, often appear extravagant when contemplated separately: their cost looms larger on its own. The attractiveness of a course of action may thus change if its cost or benefit is placed in a larger account.

If a decision is influenced by the reference point with which possible outcomes are compared, what determines the reference point? The dependence of impressions, judgments and responses on a point of reference is a ubiquitous psychological phenomenon. The same tub of tepid water may be felt as hot to one hand and cold to the other if the hands have been exposed to water of different temperatures. A given income may be considered lavish or inadequate depending on whether one's earnings have recently increased or decreased. In these cases the reference point is the state to which one has become adapted. In many cases, however, the reference
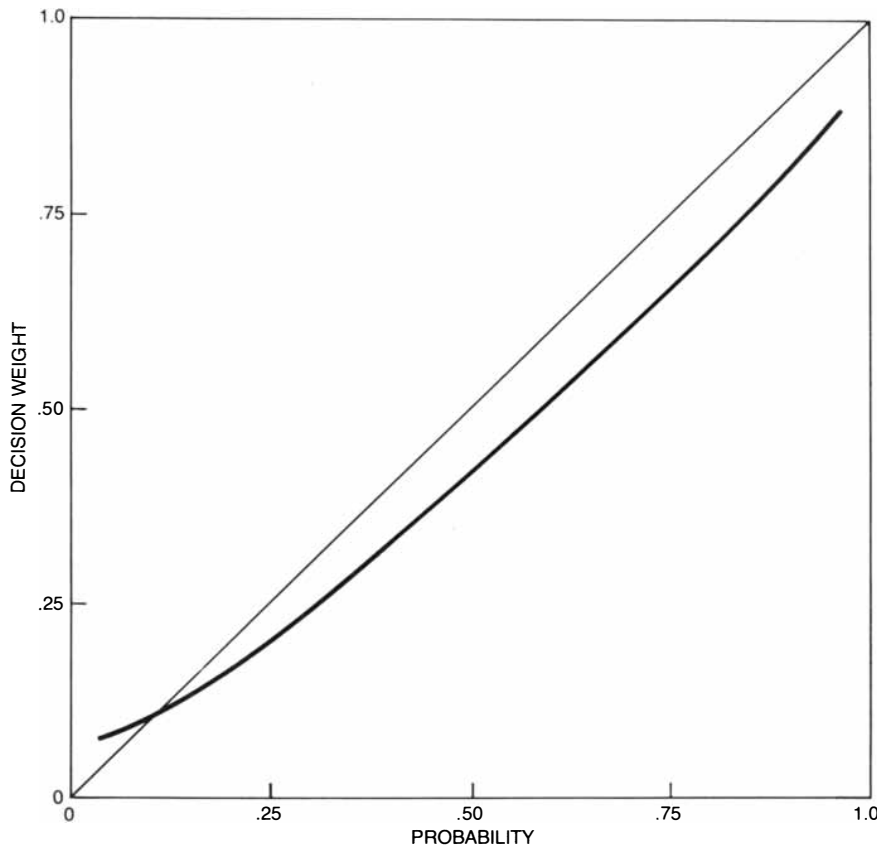
DECISION WEIGHTS express the subjective evaluation of probabilities. The classical analysis of decisions assumes that decision weights (*heavy line*) coincide with probabilities (*light line*). Recent investigations, however, suggest that decision weights do not coincide with probabilities. Because probabilities are used to derive the value of gambles, the subjective response to probabilities complicates the task of formulating a realistic value function. In particular, the difference between certainty and possibility and the difference between possibility and impossibility are given more weight than comparable differences in the intermediate range of probability. The main properties of decision weights are illustrated by the curve shown. Impossibility is assigned a weight of zero; certainty is assigned a weight of one. Small probabilities are overweighted in relation to impossibility; moderate and large probabilities are underweighted in relation to certainty. The overweighting of small probabilities contributes to the attractiveness of both lottery tickets and insurance policies by enhancing the impact of unlikely events.

168

point is determined by events that are only imagined. Consider the following incident:

Mr. Crane and Mr. Thomas were scheduled to leave the airport on different flights at the same time. They traveled from town in the same limousine, were caught in a traffic jam and arrived at the airport 30 minutes after the scheduled departure of their flights. Mr. Crane is told that his flight left on time. Mr. Thomas is told that his flight was delayed and just left five minutes ago. Who is the more upset?

Almost everyone presented with this incident agrees that Mr. Thomas is more upset than ⁻Mr. Crane, although their objective conditions are identical: both have missed their flight. Furthermore, both had expected to miss their flight, so that Mr. Thomas has no reason to be more surprised or disappointed than Mr. Crane. If Mr. Thomas is the more upset, it is presumably because in the act of imagination Mr. Thomas comes closer than Mr. Crane to catching his flight. The frustration experienced in an unsatisfactory situation increases when it is easy to imagine a more desirable alternative. For another example of the same notion, consider the following:

The winning number in a lottery was 865304. Three individuals compare the ticket they hold to the winning number. John holds 361204; Mary holds 965304; Peter holds 865305. How upset are they respectively?

There is general agreement that the experience is devastating for Peter, quite severe for Mary and very mild for John. Here again the ranking corresponds to the degree to which the individuals can be described as having "come close" to winning the prize.

An individual's experience of pleasure or frustration may therefore depend on an act of imagination that determines the reference level to which reality is compared. It is notable that the act of imagination by which one creates alternative realities reflects many realistic constraints. If there were no constraints, Mr. Crane would find it as easy as Mr. Thomas to imagine himself catching his flight and John would find it as easy as Peter to imagine himself with the winning ticket. Imagination appears to be governed by rules, and the rules of imagination affect our experience of reality by controlling the alternatives to which it is compared.

Regret is a special form of frustration in which the event one would change is an action one has either taken or failed to take. A natural extension of the hypothesis we have developed for the analysis of frustration is that regret is felt if one can readily imagine having taken an action that would have led to a more desirable outcome. This interpretation explains the close link between the experience of regret and the availability of
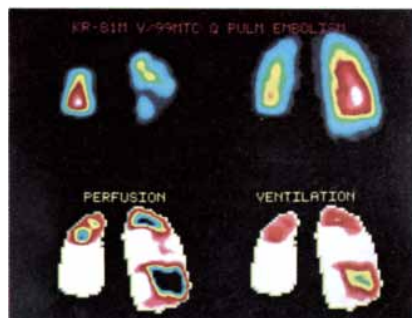
Medicine is always changing, now there's a text that changes right along with it.
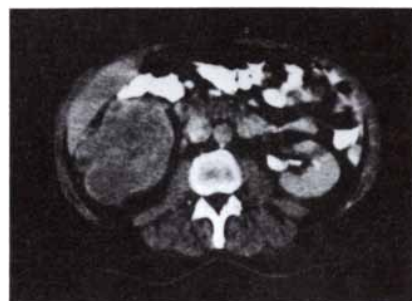
SCIENTIFIC AMERICAN *Medicine.*

The answers that you get from SCIENTIFIC AMERICAN *Medicine* are current, because each month new chapters replace the old. You will know where to look for the answers, because this advanced text follows the familiar structure of medical knowledge, changing as medicine changes.

Comprehensive. Leading authorities from Harvard and Stanford cover the full range of medicine for you the seasoned clinician.

Complimentary CME. At no extra cost, we offer a CME program through which you can earn 32 top credits* per year.

Trial Offer. We invite you to try SCIENTIFIC AMERICAN *Medicine* for two months with our compliments. Send us the coupon and you will receive the two-volume text and two monthly updates. You may also take a CME test for credit. At the end of 60 days, if you decide to continue the subscription, you will owe us $220 for the first 12 months (current renewal is $170); otherwise, return the volumes; there will be no charge.

Please mail the coupon today so that we can put medical progress to work for you.

*As an organization accredited for continuing medical education, the Stanford University School of Medicine designates this continuing medical education activity as meeting the criteria for 32 credit hours in Category 1 for Educational Materials for the Physician's Recognition Award of the American Medical Association, provided it has been completed according to instructions.

This program is approved by the American Academy of Family Physicians for 32 elective hours of CME credit.

Echocardiograms from patients with aortic regurgitation.

Computerized scintigraphy reveals pulmonary thromboembolism.
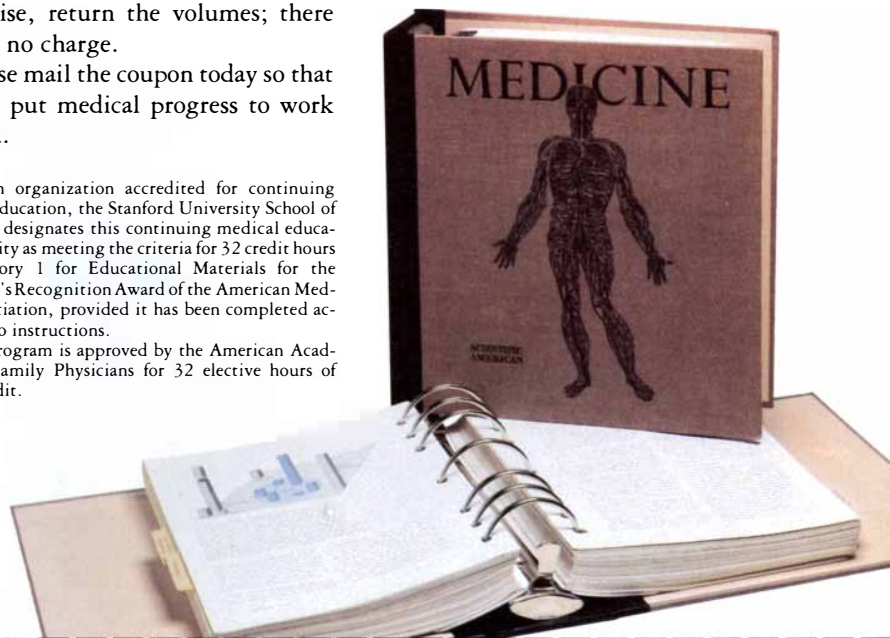

Abdominal computed tomogram reveals large renal carcinoma replacing part of right kidney.

SCIENTIFIC AMERICAN MEDICINE

415 Madison Avenue, New York, New York 10017

Please enroll me as a subscriber to SCIENTIFIC AMERICAN *Medicine*. On receipt of this coupon you will send me the advanced two-volume text described in your announcement and update it regularly by sending me new monthly subsections. I understand that the annual subscription of $220 for SCIENTIFIC AMERICAN *Medicine* is tax deductible. If I am not entirely satisfied, I may cancel at any time during the first 60 days, returning all materials for a *complete refund.*

☐ Please enter my subscription for SCIENTIFIC AMERICAN *Medicine*
☐ I shall also enroll in the CME Program
☐ I enclose a check made out to SCIENTIFIC AMERICAN *Medicine* for $220*
☐ Please bill me.

*Please add sales tax for California, Illinois, Massachusetts, Michigan, Ohio and New York

Name_____MD Specialty_____
Address_____
City_____State_____Zip_____
Signature_____

Subscribers outside the U.S. and possessions will be charged extra for shipping and handling; updates will be sent by surface routes unless airmail delivery is requested. Please allow 6-8 weeks for delivery.   D1

Phil and Steve Mahre, World Cup Race at Aspen

You have what it takes to win.
Help support the U.S. Ski Team.

The U.S. Ski Team is on a winning streak. They need your help to keep it going.

World Cup victories by Phil Mahre and Tamara McKinney highlighted the 1981 season. The Team's wins provide high hopes for the 1982 World Championships.

Our Alpine and Nordic Teams are not subsidized by the government. These fine athletes devote years of their lives training to win. They depend on you, the American public, to support their efforts. Please send your tax-deductible donations to the U.S. Ski Educational Foundation, Box 100M, Park City, Utah 84060.

Please help.

choice: actions taken under duress generate little regret. The reluctance to violate standard procedures and to act innovatively can also be an effective defense against subsequent regret because it is easy to imagine doing the conventional thing and more difficult to imagine doing the unconventional one.

A closely related hypothesis is that it is often easier to mentally delete an event from a chain of occurrences than it is to imagine the insertion of an event into the chain. Such a difference in imaginability could help to explain the observation that the regret associated with failures to act is often less intense than the regret associated with the failure of an action. Consider the following:

Paul owns shares in Company *A*. During the past year he considered switching to stock in Company *B*, but he decided against it. He now finds that he would have been better off by $1,200 if he had switched to the stock of Company *B*. George owned shares in Company *B*. During the past year he switched to stock in Company *A*. He now finds that he would have been better off by $1,200 if he had kept his stock in Company *B*. Who feels more regret?

Here again it is generally agreed that George is more upset than Paul, although their objective situations are now identical (both own the stock of Company *A*) and each reached his situation by deliberate decision.

Apparently it is easier for George to imagine not taking an action (and thereby retaining the more advantageous stock) than it would be for Paul to imagine taking the action. Furthermore, one would expect both men to anticipate the possibility of regret and to act accordingly. In general the anticipation of regret is likely to favor inaction over action and routine behavior over innovative behavior. Our analysis traces these biases of decision to the rules of the cognitive operations by which alternative realities are constructed.

We have considered a variety of examples in which a decision, a preference or an emotional reaction was controlled by factors that may appear irrelevant or inconsequential. Some of these examples illustrate impediments to the achievement of rational decision that were discussed by Herbert A. Simon under the heading of "bounded rationality." The difficulty people have in maintaining a comprehensive view of consequences and their susceptibility to the vagaries of framing are examples of such impediments. The descriptive study of preferences also presents challenges to the theory of rational choice because it is often far from clear whether the effects of decision weights, reference points, framing and regret should be treated as errors or biases or whether they should be accepted as valid elements of human experience.

THE AMATEUR SCIENTIST

Why do particles of sand and mud stick together when they are wet?

by Jearl Walker

Sand castles and mud pies are made for fun but present intriguing questions. Although sand and mud consist of particles of much the same composition, they have very different properties. What forces hold sand castles and mud pies together? Why is it that sand castles disintegrate when they dry and mud pies do not? Why do only the mud pies shrink when they dry? How can water, which normally functions as a lubricant, increase the cohesiveness of sand and clay? The answers to these questions are found in the electric interactions between the water and the particles of sand or clay.

Sand consists of fairly large, roughly spherical grains that remain closely packed even when they are wet. Their close packing is indicated by the fact that sand does not shrink when it dries. A wet grain of sand has positive and negative ions at its surface, usually grouped in pairs with one member a little farther from the surface than the other. Which ion is the outer one varies over the surface. Both orientations contribute an electric field to the space surrounding the grain. When the outer ion is positive, the electric-field vector points away from the surface. When the outer ion is negative, the vector points toward the surface.

Although the electric field surrounding the sand grain should average out to zero, it does not because of a polarization of charge in the negative ions on the surface. On one side of a negative ion there is a positive ion a little farther from the surface; on the other side there is no charge. Since the negative ion has more negative charge than positive charge, its negatively charged electron cloud shifts slightly away from its positively charged nucleus toward the neighboring positive ion. This slight separation of positive and negative charges, called an electric dipole, has a relatively weak electric field away from the surface of the sand grain. A positive ion on the surface has a deficiency of negative charge, undergoes no separation of its charges and has the electric field of a

point source of charge instead of the weaker dipole field. Hence the electric field around a sand grain is primarily due to the positive ions on the grain's surface.

Wet sand is firm because of an electric interaction between these positive ions and the water between the grains. According to a slightly dated model of the interaction, the water provides negative ions (OH^-) that adhere to the grains and electrically shield the sites of positive charge. Positive ions (H_3O^+) remain in the water that is farther from the surface. The attraction between the positive ions in the water and the negative ions adhering to the surface of the grains makes it harder for one grain to slide over another. Although this model does explain the firmness of wet sand, it has a flaw: the water is unlikely to be ionized so that it can provide the required OH^- and H_3O^+ ions.

No current model of wet sand is complete, but one modern interpretation of the electric interaction does not require that the water be ionized. The electric field generated by positively charged sites on the surface of the sand grain itself partially depletes the positive charge of the water adjacent to the surface. This depletion is not ionization, because the charges within a water molecule are not totally removed from each other. The electric field simply shifts the average positions of the hydrogen nuclei (protons) in the molecules that are farther from the surface. The layer next to the surface is left relatively negative; the water farther from the surface is relatively positive. The transition from relatively negative to relatively positive is gradual; this state of affairs is described as a diffuse double layer of charge.

In the older ion model the strength of the wet sand is attributed to the attraction of the oppositely charged ions in the water. In the diffuse-double-layer model the strength arises from the fact that the water tends to be immobilized. Since the fields generated by positively charged sites at the surface of a sand grain redistribute the average positions of the hy-

drogen protons in the water, the water's ability to flow is reduced. The sand is firm because the increased viscosity of the water hinders the sliding of one grain with respect to another.

The immobilization of the water diminishes with distance from the surface of a grain. When wet sand flows to the degree that it cannot be used for building a sand castle, it does so because of the large amount of water between the grains. Although the relatively negative water surrounding a grain still attracts the relatively positive water farther from the grain, the grains are too much separated by water. Some of the water between the grains is not close enough to any one grain to participate strongly in the electric interaction. This less viscous water flows, allowing the adjacent grains (and their halos of relatively charged water) to slide over each other.

It is a common notion that sand castles are held together by the force exerted by the surface tension of the water in the sand. Water molecules are said to be polar because they have a permanent electric dipole. These dipoles attract each other. At a surface between air and water the mutual attraction of neighboring water molecules puts the surface in a virtual state of tension. When the surface is curved, the mutual attraction provides a collective force on it.

On the surface of a sand castle the collective force does help to firm up the sand. The sand is just as firm, however, within the sand castle, where no air spaces remain. Although the water molecules there are loosely attracted to each other by their electric-dipole fields, the attraction is too weak to account for the sand's firmness and resistance to flow. Water molecules in a glass of water are attracted to each other but the water is not firm.

When a sand castle dries, it disintegrates because the amount of relatively charged water in it has decreased. During the transition the charges on the surface of the sand grains rearrange themselves. These ions form a double layer of charge, called the Helmholtz double layer, with a positive ion slightly inside the surface below a negative ion outside. The electric field generated by this positioning of charges is weak except directly at the surface. Wherever two dry sand grains touch they repel each other because both surfaces have an outer layer of negative ions. Hence dry sand grains do not adhere.

The cohesion of sand is destroyed if some of the water is replaced with glycerol. The molecules of glycerol electrically shield the sites of positive charge on the sand grains, reducing the electric field in the water. With weaker fields there is less ordering of the water layers, and with no immobilization of the water the sand does not hold together.

You can demonstrate the effect of
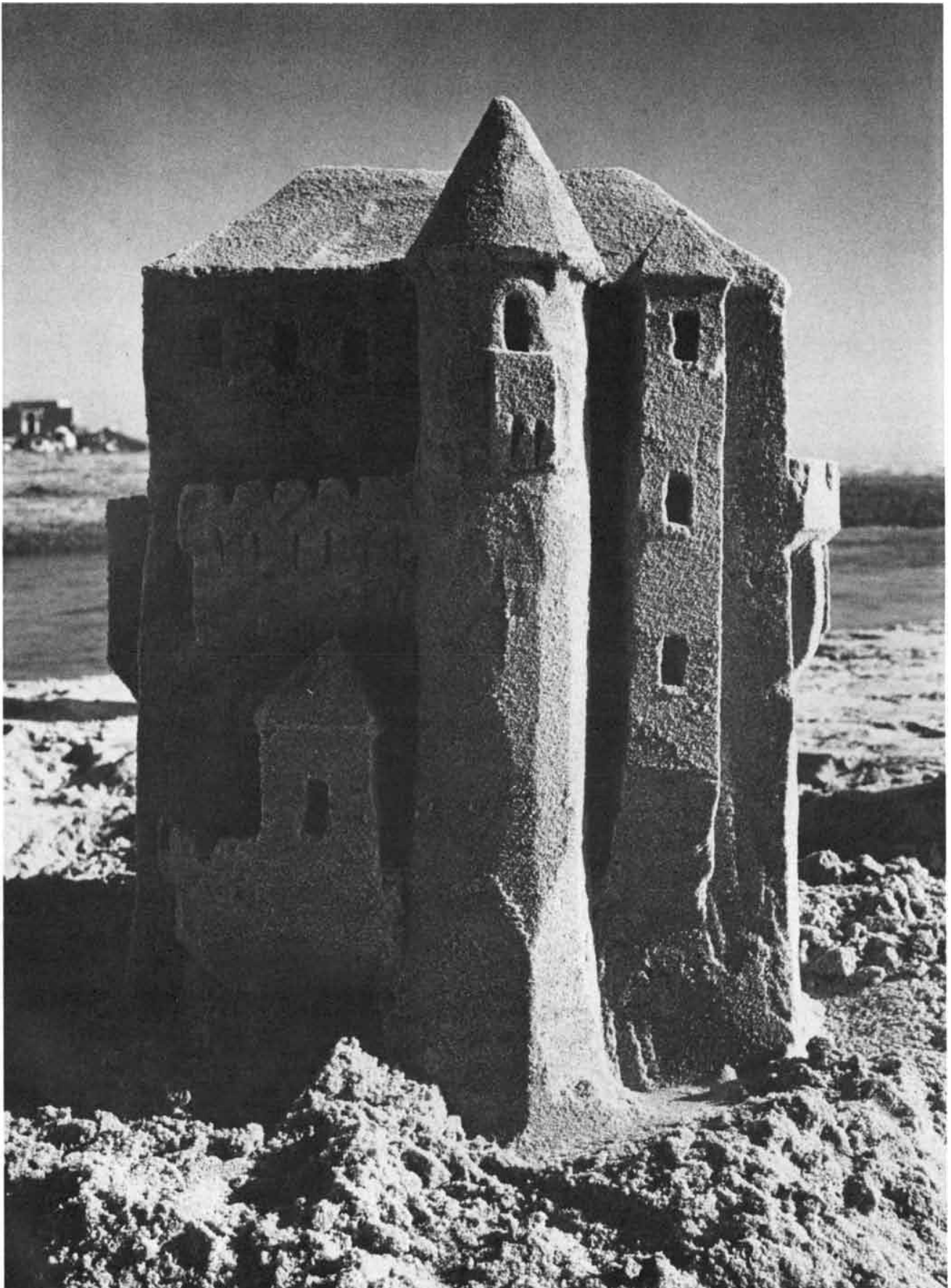
175

glycerol easily. First pack a small container with wet sand so that the sand is reasonably firm. Over the container place a flat plate. Quickly invert the plate and the container without allowing the sand to spill out. With the plate on the floor pull slowly upward on the container so that a mound of sand remains on the plate. Test the strength of the mound by shaking or tapping the plate.

Now repeat the experiment with glycerol added to the water mixed with the sand. When you have the assembly inverted, again pull upward on the container. The chances are that the mound left behind is noticeably weaker than the mound without the glycerol. A shake or a tap will probably cause it to collapse. The more glycerol you add to the sand, the less stable the mound is. The result is somewhat surprising. A viscous fluid (glycerol) is poured onto a very viscous mixture (wet sand), and the result is a decrease in viscosity.

Wet clay differs from wet sand in several subtle ways. Its particles are con-siderably smaller than sand grains and are platelike, with one dimension much smaller than the other two. In wet clay the particles are loosely packed and well separated by water. One indication of the looseness is the fact that clay shrinks as it dries.

On the surface of a clay particle positively charged sites are exposed, just as they are on a sand grain. According to the older ion model of wet clay, OH^- ions adhere to these sites, leaving H_3O^+ ions in the water between the clay particles. The strength of the wet clay is due to the attraction between the two sets of ions. In the more modern model of wet clay the water is said to be immobilized by the electric fields set up by the positively charged sites on the clay particles. As with the model of wet sand, the immobilized water has a high viscosity. The wet clay is firm because the clay particles cannot flow over each other. If the clay is very wet, it is not firm. The water away from the particles is less viscous and can flow.

Drying clay differs from drying sand in an important respect. Whereas the charges on a sand grain can shift and give rise to a Helmholtz double layer on the surface of the grain, the charges on a clay particle cannot. The clay particle is too thin in one dimension for double layers to form on the opposite sides of that dimension. The innermost layers of charge (which would be of the same sign) would repel each other too strong-ly to be stable.

When wet clay dries, the fact that a Helmholtz double layer cannot form requires that the particles retain their diffuse double layer and that the resulting viscous-water remain between them. The clay structure continues to be firm. In fact, its rigidity increases as the particles move closer to one another, further increasing the immobilization of the water. As the structure shrinks and more contact points develop between the particles, the particles bond together by sharing ions. They also resist movement because of the remaining viscous water. Other forces, all electric, also help to bond them together.

Sand grains are closely packed. Clay particles are loosely packed. Their packing differs because of the difference in their shapes and sizes. Suppose a sand grain falls and comes in contact with a stationary grain. Does the fallen grain remain attached to the other grain or does it roll or slide over the other grain and continue its downward motion? At the region where the grains come in contact the attractive forces are strong enough to hold them together. The gravitational pull on the fallen grain, however, puts a torque on it that tends to rotate the grain about its point of contact with the other grain. Some of the attractive forces create a torque opposing the rotation. Because the grain is relatively heavy the torque due to gravity overwhelms the torque due to the attractive forces, and the grain rotates. Continued motion of this kind finally brings the grain downward into the other grains until it is closely packed.

When a clay particle falls onto a collection of other clay particles, it is subject to attractive forces on encountering a stationary particle. Gravity continues to pull it downward, creating a torque that tends to rotate it about the point of contact. The particle is so light, however, that the torque created by gravity cannot accomplish the rotation. Hence the particle remains in the orientation it had when it first collided with the stationary particle. When more particles fall, they too go no farther once they land. As a result the packing of the particles is loose.

When wet or dry sand is compressed, it takes a different form only if the sand grains slide relative to one another. Suppose a pressure is applied to wet sand but not enough to cause relative motion

Ion model of water

Immobilization of water

176
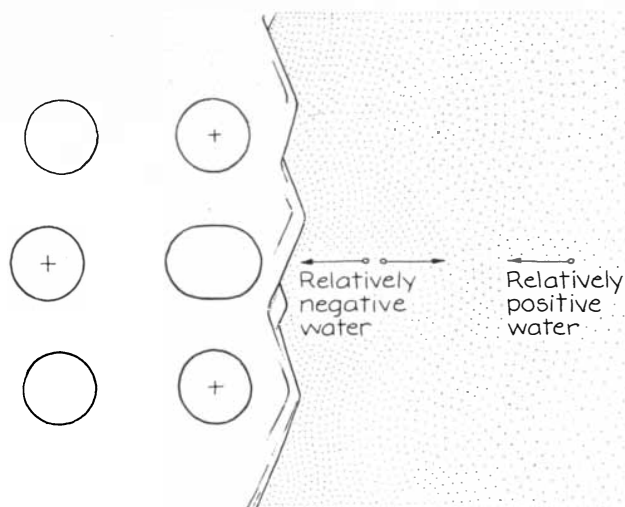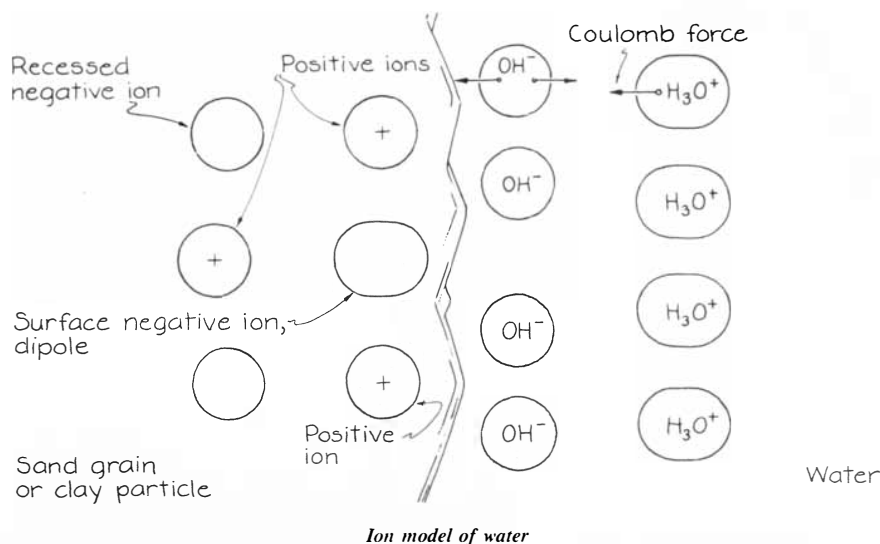
of the grains. The contact area of two grains that are pushed against each other increases, and so does the extent of the electric forces holding them together. At the center of the contact area the additional pressure on the grains irreversibly deforms their surfaces. The energy expended in the deformation cannot be reclaimed when the pressure on the grains is removed. Most of the energy of the compression of the grains, however, goes into deforming the region surrounding the center of the contact area, which is not irreversible. That energy is stored as elastic potential energy, much as energy is stored in a compressed spring.

When the pressure on the grains is removed, the elasticity of the grains pushes them apart. The additional electric forces that are developed during the time of increased contact between the grains are relatively weak and are easily broken by the rebound of the grains. Hence after the pressure is removed the grains return to their earlier form.

Something different happens when wet clay is compressed. Consider the loose collection of clay particles shown in the bottom illustration on this page. The applied force bends one of the platelike particles so that its previously free end is now touching another particle. At the point of contact attractive forces bond the two particles together. When the applied force is removed, the particles remain bonded because the bonding forces are strong enough to hold the particles in their bent configuration. When you push with your fingertips against wet clay, you compress the structure. It does not rebound when you stop pushing. Such material is said to be plastic. Wet clay is plastic and wet sand is not.

The best type of sand for building a sand castle is sand with a little clay mixed into it. Choose sand near the water so that it is already wet enough. Sand right at the edge of the water is probably too wet to be firm. Besides, any sand castle built there would soon be destroyed by waves or the tide.

You can build sand castles in several ways. Wet sand can be dumped carefully from a container so that the shape of the container's interior is preserved. An interesting container leads to interesting designs. The sand can also be shaped by packing it between two lengths of board. Long castle walls are easily constructed in this way. A more difficult technique is sculpture. Heap up a large mound of sand and then, starting at the top, whittle away. Another technique mimics the growth of stalagmites in caves. From a clenched fist filled with wet sand allow a small stream to dribble out. The stream creates a thin spire that resembles the steeple of a futuristic cathedral.

If your castle is to last throughout a hot day, you should periodically spray it
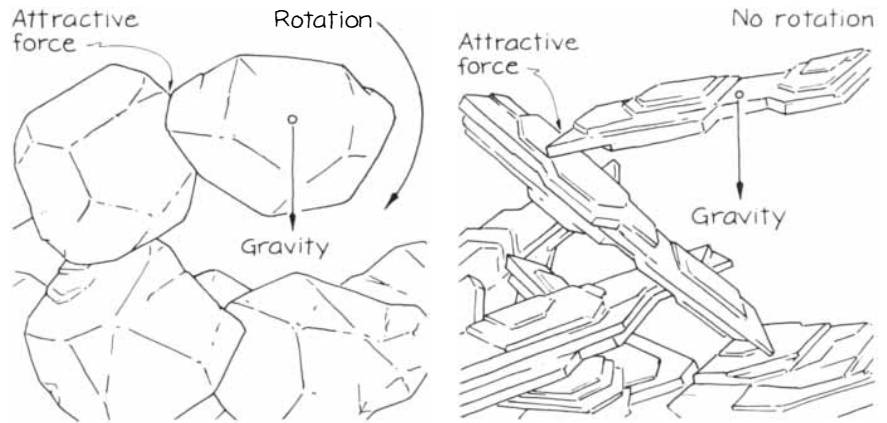
The packing of sand (left) *and clay* (right)

in order to preserve its firmness. The added water seeps between the grains. There the electric fields from the grains shift the average positions of the hydrogen protons in the water and so immobilize the water. The increased viscosity of the water ensures the structure's stability until the next spraying is needed.

Although sand does not change in volume as it dries, it does expand when it is suddenly stressed, a kind of behavior termed dilatancy. You may have noticed such expansion when you walked across a stretch of beach just wetted by a wave. The sand looks wet before you step on it but dries just as you do so. Only after a while does the footprint look wet again.

Wet clay does not dry under a footstep because its particles are loosely packed. Any pressure on the clay moves the particles closer together, decreasing the clay's volume. Dilatancy is characteristic of closely packed particles. When a sudden pressure is applied to the particles, forcing them to move, they can do so only if their average distance

from one another increases. With wet sand pressed down by the sole of a foot this expansion amounts to the surface of the sand's rising. The water is left behind and takes a few minutes to seep to the surface.

The dilatancy of sand also figures in an avalanche of dry sand. A simple demonstration exhibits the important features of an avalanche. Gradually tilt a trough of sand until it flows. The avalanche begins when the tilt of the surface exceeds a certain angle called the dynamic friction angle. The sliding sand decreases the tilt to an angle, called the friction angle, at which the surface is stable. If you again tilt the trough, the surface again exceeds the dynamic friction angle and a new avalanche begins.

At any angle less than the friction angle the sand grains on the slope are stable because they are closely packed. Only if the tilt of the slope reaches the dynamic friction angle can the grains slide. To do so they first must rise, so that they can slide or roll over the grains under them. The entire overburden of
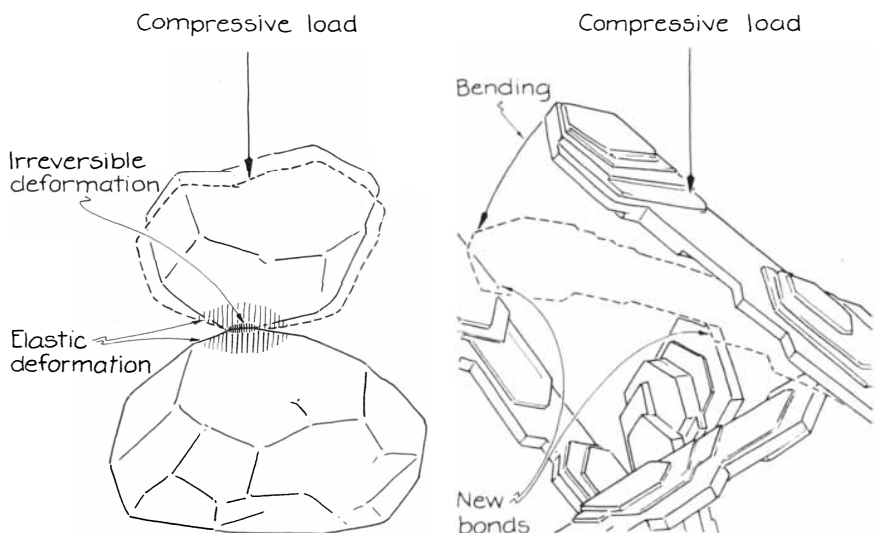
Response of sand and clay to compression
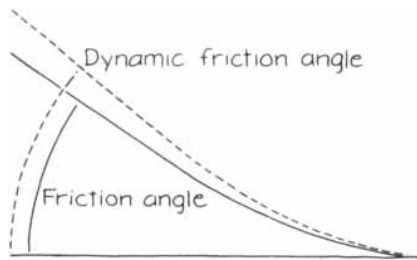
Significant angles in avalanches

the sand enters the state of dilatancy and slides down the slope.

Many granular substances slide in much the same way. Snow is a well-known example. In addition to sand I investigated sugar (both granulated and powdered), salt, cornstarch and cocoa powder. I also studied the avalanching of dry pinto beans, oranges and apples. Each had its characteristic friction angle and dynamic friction angle.

With sand the difference between the two angles is about five degrees. When I spray the sand with water, the grains no longer flow down the tilted surface. Instead the entire bulk of sand in the trough slides across the bottom of the trough once the trough is sufficiently tilted. With pinto beans the difference between the two angles is about four degrees. Cornstarch and cocoa powder are so cohesive that with them I have trouble making the measurement.

Many powders are cohesive: their particles stick to each other and resist flowing. Some, such as face powders, are meant to be cohesive so that they stay put. Others, such as salt and cocoa powder, are a nuisance when they are cohesive. The cohesion in a powder is due to a variety of electric forces.

One such force, the electrostatic, is due to charges on the surface of adjacent particles. The strongest of the electrostatic forces is the Coulomb force, which attracts particles that have the same sign of charge and repels particles that have the opposite sign. Weaker is the attractive force that develops between a charged particle and a neutral (uncharged) atom or molecule. The electric field of the charged particle induces a separation of charge inside the atom or molecule. For example, if the particle is negative, the positive constituent of the neutral atom or molecule moves slightly toward the particle and the negative constituent moves slightly away. Although the atom or molecule is still neutral, it now has the field of an electric dipole and is attracted to the charged particle.

Some molecules are permanent electric dipoles and thus attract each other when they are neighbors; the negative end of the one pulls toward the positive end of the other. The van der Waals force is a similar type of electric force, but it is due to the quantum nature of atoms. Although an atom might be electrically neutral and so not be a permanent dipole, it is nonetheless a dipole at any one instant. In the simple Bohr model of an atom the electron is considered to be in orbit around the nucleus. Therefore at any given instant it is on one side of the nucleus and is hence separated from the positive charge. This instantaneous dipole can induce a dipole in a neighboring neutral atom. Once the two dipole fields are established they pull the atoms together.

The van der Waals force attracts almost any two surfaces that are in contact. Everything in the world is not stuck to everything else only because Coulomb repulsion usually keeps surfaces from adhering. Many common powders are cohesive because of the van der Waals force between the particles.

If the particles in such powders touch, the frictional forces between them oppose the sliding of the particles relative to each other. If there is water between the particles, their cohesion can be due to its immobilization by the charged sites on the particles. Surface tension can also play a role if there is air between the particles or if the outside of a cake of the powder is held together by water.

The cohesion of a powder cannot be predicted from a detailed knowledge of the structure of its particles except perhaps in a very general way. For example, there is probably more friction between irregularly shaped particles, but only if they happen to touch each other. The cohesiveness of a powder is often reduced when the particles are ground into smoother and more regular shapes. The cohesiveness can also be altered if the powder is dried so as to diminish the role of the water between particles. The compression of a powder usually increases its cohesion by increasing the strength of the van der Waals force or the attractive electrostatic forces or by adding to the number of particles touching and interlocking. In general the smaller the grains of a powder are, the more cohesive the powder is. Suppose you prepare a powder whose particles are nearly uniform in size. The powder is placed in a hopper whose bottom has a circular opening to allow the powder to flow out. To measure the rate of flow of the powder you measure the time it takes to empty the hopper by a particular amount.

When the particles are very small, their rate of flow from the hopper is either very low or zero. They encounter large cohesive forces that restrict their motion relative to one another. If progressively larger particles are substituted, the rate of flow increases until it peaks and then decreases. When the particle size is roughly a fifth the size of the escape hole, the flow stops. At this point the particles have locked onto one another, building across the opening something like a dry-stone wall.

Some powders form crumbs that are held together by water between particles. The form of the water layers between the grains falls into one of three categories. If the particles of the powder were originally quite dry, the water lies between the regions of contact of the particles in what is termed a pendular formation. Other regions between the particles are filled with air. When more water is present, the water-filled regions are larger but are still separated by air gaps. This formation is called funicular. If the space between the particles is filled entirely with water, the formation is said to be capillary. The detailed nature of the forces holding together a crumb of particles is not well understood. The adherence is often attributed to the surface tension of the water, and that model may suffice for the pendular and funicular formations. In any event the immobilization of water is important in all three formations.

Powder can cake even if most of it is dry. You may have noticed that when water is added to a powder for hot chocolate, the powder is often so caked that it resists dissolving. Such a cake is dry on the inside but wet on the outside. When a large mound of hot-chocolate powder floats on the surface of the water, the powder along the edge of the water gradually collapses the perimeter of the mound, sending part of the powder avalanching into the water.

Powder cakes when the water between the particles on the outside of the cake is so viscous (owing to immobilization) that additional water cannot easily seep into the interior. Attractive forces between the particles may also play a role. Eventually enough water seeps between the particles so that their electric fields can no longer immobilize all the water or strongly attract each other. Then the cake falls apart. Still, some particles remain clumped long after the main body of the powder has dissolved and disappeared.

Wet cornstarch is similar to many other wet powders except that its viscosity depends on the stress put on the mixture. It is an example of a non-Newtonian fluid, a class of fluids whose properties I discussed in this department for November, 1978. Cornstarch is thixotropic, which is to say that its viscosity increases when the stress on the fluid is increased.

I dissolved a small amount of cornstarch, rolled it into a ball and then examined it periodically as it dried. At first the firmness of the ball depended on the stress placed on the mixture. When I pressed the ball, it was very firm and looked dry. When I released the pressure, the ball immediately flowed and looked wet. As water evapo-

178

rated from the mixture the ball became progressively firmer until its surface was quite rigid to the touch. The structure was weak to any twisting force but strong to any pressure on its surface.

Because cornstarch is composed of very small particles its behavior might be modeled on the electric interactions in wet clay. I am not certain why the mixture is thixotropic. The easiest explanation is that the increase in viscosity is linked to the immobilization of the water between the particles. When I press the ball of cornstarch, I move particles of starch closer to one another. The water between them becomes less mobile because it is now closer to one or more particles. That region of the ball thus becomes much more viscous than it was before I applied the pressure.

Most of the cohesive powders are also adhesive and cling to my finger or a kitchen utensil. Which of the many types of electric force are important in adhesion may be difficult to determine. The van der Waals force usually plays a role, but some of the electrostatic forces probably participate also. If the particles in the powder are wet and fairly large, the surface tension of water may adequately explain the adhesion.

When a fine powder is even slightly damp from the humidity in the room, part of its attraction to your finger is probably due to the immobilization of the water between the finger and the individual particles of powder. On the surface of a particle there is a layer of relatively negative water. Somewhat farther from the particle is relatively positive water. When this system is brought close to a surface such as a finger, the electric field from the relatively positive water forces the surface of the finger to be negative. The particle clings to the finger because of the Coulomb attraction between adjacent layers of charge. The surface of the particle is positive. The water adjacent to the particle is relatively negative. The water farther from the particle is relatively positive. The surface of the finger is negative.

This kind of adhesion is strong, however, only if the particles are in the size range of a few microns. Particles much larger are probably too heavy to cling to a finger. Particles much smaller do not cling either, but that is not because of their weight. Smaller particles are able to generate thicker halos of immobilized water because they are not as good at generating a Helmholtz double layer of charge on their surface. Since they cannot be as close to a finger as the somewhat larger particles with their thinner halos, they fall off. If the adhesion to your finger were not so selective, life would be miserable, particularly on humid days. Everything you touched would stick to your fingers because the water between them and the particles was immobilized.

If you'd like a booklet explaining how we smooth out Jack Daniel's with charcoal, just drop us a line.

ASK THE ROGERS TWINS how Jack Daniel's Whiskey gains smoothness, and the reply will be identical.

Either Ray or Clay, who guide visitors here, will take you to a room where just-made whiskey is seeped through vats of tightly packed charcoal. And they'll point out how we've insisted on this whiskey-smoothing method since 1866. Of course, you may have trouble telling Ray from Clay. But if you compare Jack Daniel's to any other whiskey, you'll spot the difference in just a sip.

CHARCOAL MELLOWED

DROP

BY DROP

Tennessee Whiskey • 90 Proof • Distilled and Bottled by Jack Daniel Distillery Lem Motlow, Prop., Inc., Route 1, Lynchburg (Pop. 361), Tennessee 37352

Placed in the National Register of Historic Places by the United States Government

179

© 1981 SCIENTIFIC AMERICAN, INC

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful

METAMAGICAL THEMAS

METAMAGICAL THEMAS. Douglas R. Hofstadter in *Scientific American*, Vol. 244, No. 1, pages 22–32; January, 1981.

THE SOCIAL SECURITY SYSTEM

SOCIAL SECURITY: TODAY AND TOMORROW. Robert M. Ball. Columbia University Press, 1978.

COMING OF AGE: TOWARD A NATIONAL RETIREMENT INCOME POLICY. President's Commission on Pension Policy. U.S. Government Printing Office, 1981.

SOCIAL SECURITY FINANCING AND BENEFITS: REPORT OF THE 1979 ADVISORY COUNCIL. Social Security Administration, Department of Health and Human Services. U.S. Government Printing Office, 1981.

SOCIAL SECURITY IN AMERICA'S FUTURE: FINAL REPORT OF THE NATIONAL COMMISSION ON SOCIAL SECURITY, MARCH, 1981. U.S. Government Printing Office, 1981.

THE MASS EXTINCTIONS OF THE LATE MESOZOIC

CRETACEOUS-TERTIARY EXTINCTIONS AND POSSIBLE TERRESTRIAL AND EXTRATERRESTRIAL CAUSES: PROCEEDINGS OF WORKSHOP, NOVEMBER 16 AND 17, 1976. K-TEC Group. National Museum of Natural Sciences, 1976.

AN EXTRATERRESTRIAL EVENT AT THE CRETACEOUS-TERTIARY BOUNDARY. J. Smit and J. Hertogen in *Nature*, Vol. 285, No. 5762, pages 198–200; May 22, 1980.

EXTRATERRESTRIAL CAUSE FOR THE CRETACEOUS-TERTIARY EXTINCTION. Luis W. Alvarez, Walter Alvarez, Frank Asaro and Helen V. Michel in *Science*, Vol. 208, No. 4448, pages 1095–1108; June 6, 1980.

THE STABILIZATION OF ATOMIC HYDROGEN

STABILIZATION OF ATOMIC HYDROGEN AT LOW TEMPERATURE. Isaac F. Silvera and J. T. M. Walraven in *Physical Review Letters*, Vol. 44, No. 3, pages 164–168; January 21, 1980.

RECENT DEVELOPMENTS IN CONDENSED MATTER PHYSICS. Edited by J. T. DeVreese. Plenum Publishing Co., 1981.

SPIN-POLARIZED ATOMIC HYDROGEN: A WEAKLY INTERACTING, MAGNETIC BOSE GAS. A. J Berlinsky in *Journal of Applied Physis*, Vol. 52, pages 2309–2315; 1981.

THE MOONS OF SATURN

THE CHEMISTRY OF THE SOLAR SYSTEM. John S. Lewis in *Scientific American*, Vol. 230, No. 3, pages 50–65, March, 1974.

THE NEW SOLAR SYSTEM. Edited by J. Kelly Beatty, Brian O'Leary and Andrew Chaikin. Cambridge University Press & Sky Publishing Corporation, 1981.

VOYAGER 1 ENCOUNTER WITH SATURN *Science*, Vol. 212, No. 4491; April 10, 1981.

A NEW LOOK AT THE SOLAR SYSTEM: THE *VOYAGER 2* IMAGES. B. A. Smith, L. Soderblom, R. Batson, P. Bridges, J. Inge, M. Mazursky, E. Shoemaker, R. Beebe, J. Boyce, G. Briggs, A. Bunker, S. A. Collins, C. J. Kansen, T. Johnson, J. Mitchell, R. Terrile, A. F. Cook II, J. Cuzzi, J. Pollack, G. E. Danielson, A. Ingersoll, M. E. Davis, G. E. Hunt, D. Morrison, T. Owen, C. Sagan, J. Veverka, R. Strom and V. Suomi in *Science*, in press.

SUPERCOMPUTERS

PIPELINE ARCHITECTURE. C. V. Ramamoorthy and H. F. Li in *Computing Surveys*, Vol. 9, No. 1, pages 61–102; March, 1977.

A SURVEY OF PARALLEL MACHINE ORGANIZATION AND PROGRAMMING. D. J. Kuck in *Computing Surveys*, Vol. 9, No. 1, pages 29–60; March, 1977.

VECTOR PROCESSING ON THE CYBER 200. M. J. Kascic. Control Data Corporation, 1979.

PERSPECTIVES ON LARGE-SCALE SCIENTIFIC COMPUTATION. Garry Rodrigue, E. Dick Giroux and Michael Pratt in *Computer*, Vol. 13, No. 10, pages 65–80; October, 1980.

SECOND GENERATION OF VECTOR SUPERCOMPUTERS. Edward W. Kozdrowicki and Douglas J. Theis in *Computer*, Vol. 13, No. 11, pages 71–83; November, 1980.

THE DEVELOPMENT OF A SIMPLE NERVOUS SYSTEM

CELL LINEAGE ANALYSIS BY INTRACELLULAR INJECTION OF A TRACER ENZYME. David A. Weisblat, Roy T. Sawyer and Gunther S. Stent in *Science*, Vol. 202, No. 4374, pages 1295–1298; December 22, 1978.

EMBRYONIC CELL LINEAGES IN THE NERVOUS SYSTEM OF THE GLOSSIPHONIID LEECH *HELOBDELLA TRISERIALIS*. David A. Weisblat, Georgia Harper, Gunther S. Stent and Roy T Sawyer in *Developmental Biology* Vol 76, No 1, pages 58–78, April, 1980

NEUROBIOLOGY OF THE LEECH Edited by Kenneth J Miller, John G Nicholls and Gunther S. Stent Cold Spring Harbor Laboratory, 1981

AN ANCIENT GREEK CITY IN CENTRAL ASIA

LA VIEILLE ROUTE DE L'INDE DE BACTRES À TAXILA. A. Foucher in *Mémoires de la Délégation archéologique française en Afghanistan*. Les éditions d'art et d'histoire, 1942–47

THE GREEKS IN BACTRIA AND INDIA William W Tarn. Cambridge University Press, 1951

ANCIENT ART FROM AFGHANISTAN TREASURES OF THE KABUL MUSEUM. Benjamin Rowland. Asia Society, distributed by H. N. Abrams, 1966.

THE DYNASTIC ARTS OF THE KUSHANS. Jan M. Rosenfield. University of California Press, 1967

L'ORIENT HELLÉNISÉ. D. Schumberger A. Michel, 1969.

THE PSYCHOLOGY OF PREFERENCES

DECISION ANALYSIS: INTRODUCTORY LECTURES ON CHOICES UNDER UNCERTAINTY. Howard Raiffa. Addison-Wesley Publishing Co., 1968.

UTILITY THEORY: A BOOK OF READINGS. Compiled by Alfred N. Page. John Wiley & Sons, Inc., 1968.

PROSPECT THEORY: AN ANALYSIS OF DECISION UNDER RISK. Daniel Kahneman and Amos Tversky in *Econometrica*, Vol. 47, pages 263–291, March, 1979.

TOWARD A POSITIVE THEORY OF CONSUMER CHOICE. Richard Thaler in *Journal of Economic Behavior and Organization*, Vol. 1, No. 1, pages 39–60; March, 1980.

THE FRAMING OF DECISIONS AND THE PSYCHOLOGY OF CHOICE. Amos Tversky and Daniel Kahneman in *Science*, Vol. 211, No. 4481, pages 453–458; January 30, 1981.

THE AMATEUR SCIENTIST

ATOMISTIC APPROACH TO THE RHEOLOGY OF SAND-WATER AND OF CLAY-WATER MIXTURES. W. A. Weyl and W C Ormsby in *Rheology Theory and Applications, Vol. 3,* edited by Frederick R. Eirich. Academic Press, 1960

THE COHESIVENESS OF POWDERS. N Pilpel in *Endeavour*, Vol. 28, No. 104, pages 73–76; May, 1969.

THE PHYSICS AND MECHANICS OF SOIL. Ronald F. Scott in *Contemporary Physics*, Vol. 10, No. 5, pages 449–472; September, 1969.

CRUMB FORMATION. N. Pilpel in *Endeavour*, Vol. 30, No. 110, pages 77–81, May, 1971.

POWER STYLING
DATSUN 200-SX

Now that gas mileage is the hot button, few cars boast about their styling. The new Datsun 200-SX, Hardtop SL is a brilliant exception. Look at the integrated, wrap-around bumpers. The new louvers on the hood. The classic blacked-out grille. Inside: real comfort and the ultimate in sophistication. This car actually "talks" to you. A voice calls your attention to six vital functions like... "Right door is open." On the road a more powerful 2.2 liter fuel-injected engine propels you along without sacrificing economy: EPA estimated 26 MPG, 38 estimated highway. Use MPG for comparison only. Your mileage may differ depending on speed, trip length and weather. Highway mileage will probably be less. The thrilling 200-SX: proof positive that no one marries performance and styling like Datsun.

DATSUN
WE ARE DRIVEN

Product of NISSAN

NISSAN

© 1981 NMC–U.S.A.

Only one tape deck combines the incredible realism of dbx® with the precise sound of direct drive. Technics RS-M270X.

Dynamic range has long been the quest of audio purists because it represents a major difference between live and reproduced sound. And perhaps nothing says dynamic range better than dbx.

Rotational stability is something else audio purists have longed for in a tape transport system, and virtually nothing says that better than Technics direct drive. After all, the majority of the top radio stations that use turntables rely on Technics direct drive.

Listen to the RS-M270X. You'll hear the expansive distinction between loud and soft tones. In fact, a

recording made on the RS-M270X will sound 50 percent more dynamic than the same recording made on a conventional deck.

Of course, dbx also doubles as a noise reduction system. Yet, unlike conventional systems, dbx reduces noise at all frequencies, not just the high ones. And with the RS-M270X, you can even decode dbx Encoded Discs.

The RS-M270X also features solenoid controls, SX sendust heads and fluorescent VU meters.

Listen to Technics RS-M270X. You'll agree you've never heard so much dynamic range, so precisely. ®dbx is a registered trademark of dbx Incorporated.

Technics
The science of sound