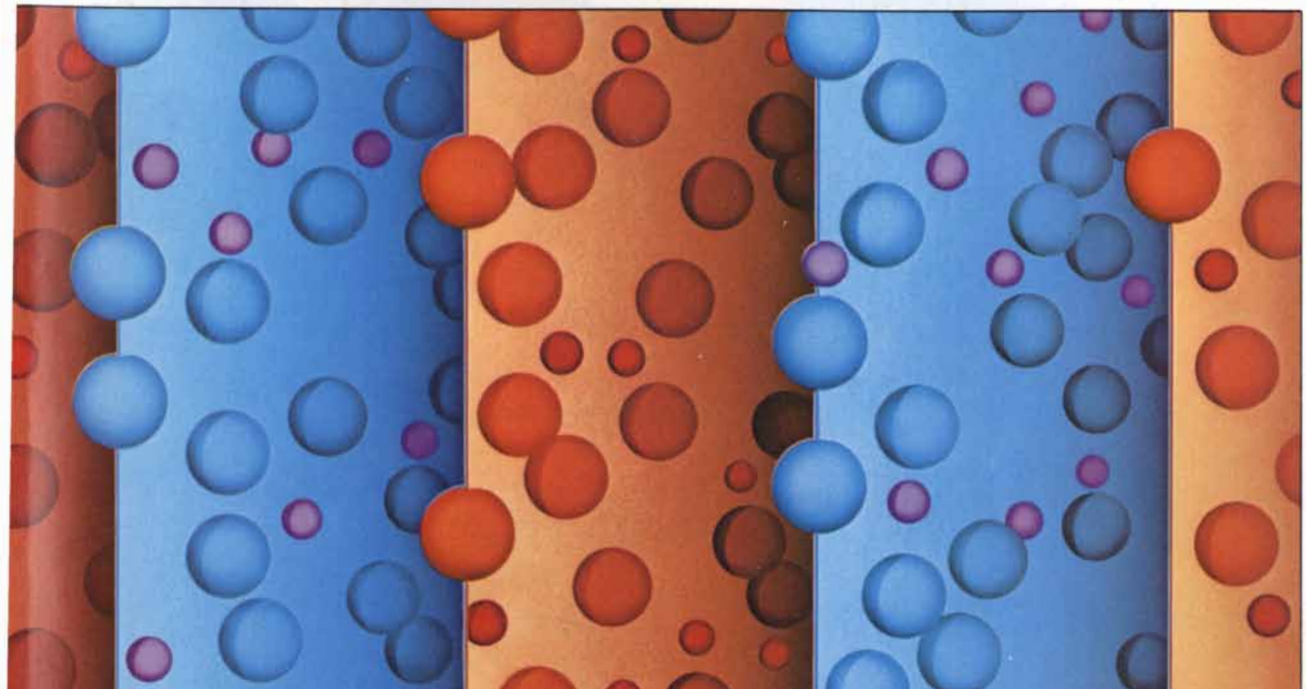
SCIENTIFIC AMERICAN



SEMICONDUCTOR SUPERLATTICES

$2.50

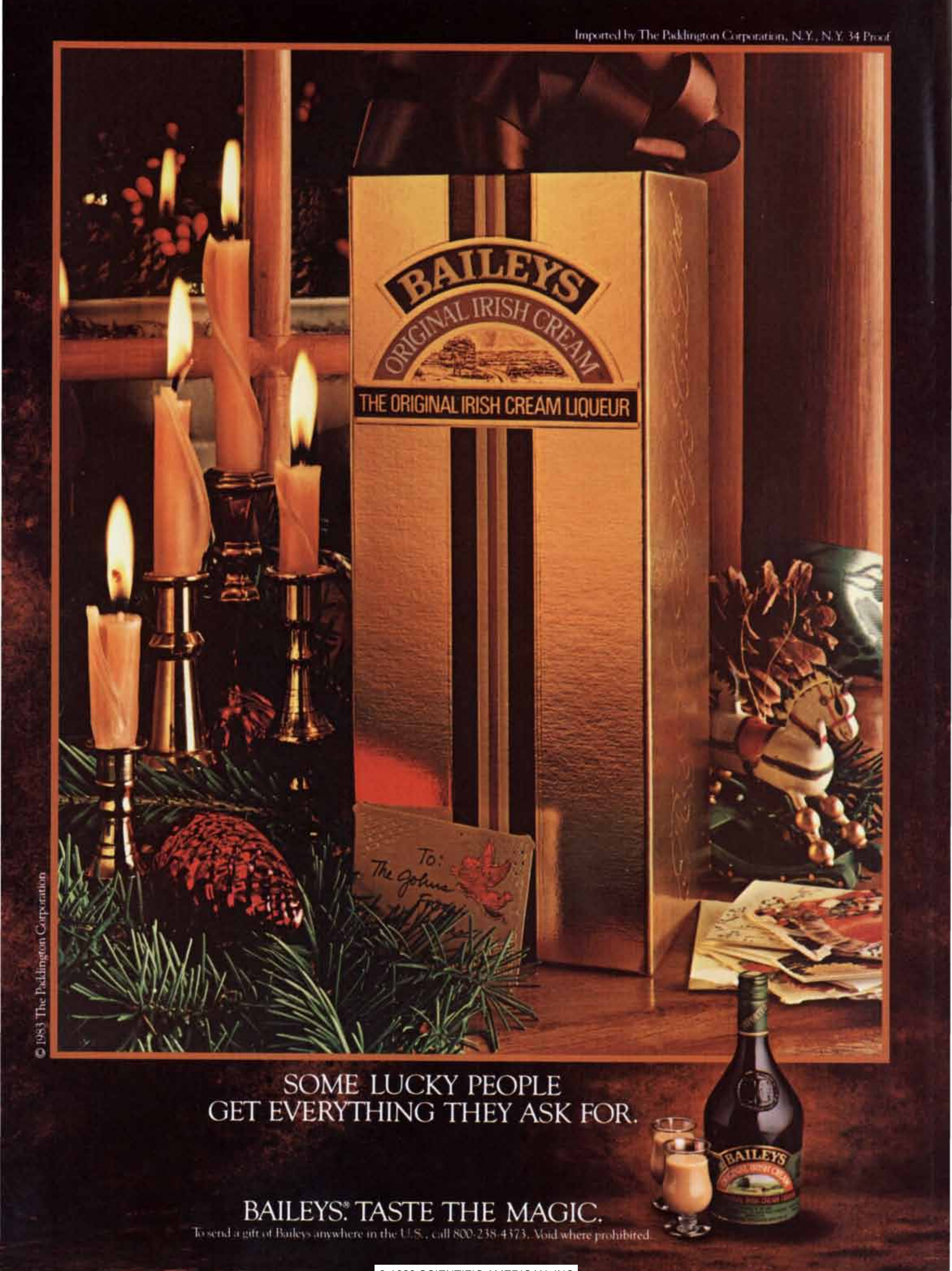
November 1983


ENCORE
ENCORE

Fully independent suspension with twin coaxial torsion bars, front wheel drive, power front disc brakes, and electronic fuel injection makes the performance expressive, and the handling impressive. Impressive fuel economy, too: **52** EST HWY, **38** EPA EST MPG.* The new Renault Encore. *Built in America* from an inspired European design—creatively expressed. Impressive for one other enormous achievement: a price of **$5755**.** Now that you've let go of your preconceptions, take hold of Renault Encore.

RENAULT

THE ONE TO WATCH


Imported by The Paddington Corporation, N.Y., N.Y. 34 Proof
BAILEYS
ORIGINAL IRISH CREAM
THE ORIGINAL IRISH CREAM LIQUEUR
To:
The Johns
From
© 1983 The Paddington Corporation
SOME LUCKY PEOPLE
GET EVERYTHING THEY ASK FOR.
BAILEYS® TASTE THE MAGIC.
To send a gift of Baileys anywhere in the U.S., call 800-238-4373. Void where prohibited.

Established 1845 **SCIENTIFIC AMERICAN** November 1983 Volume 249 Number 5

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $24 PER YEAR IN THE U.S. AND ITS POSSESSIONS; $33 PER YEAR IN ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017

SCIENTIFIC AMERICAN

CORRESPONDENCE

Offprints of more than 1,000 selected articles from earlier issues of this magazine, listed in an annual catalogue, are available at 75 cents each. Correspondence, orders and requests for the catalogue should be addressed to W. H. Freeman and Company, 4419 West 1980 South, Salt Lake City, UT 84121. Offprints adopted for classroom use may be ordered direct or through a college bookstore. Sets of 10 or more Offprints are collated by the publisher and are delivered as sets to bookstores.

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN for the flat fee of $2 per copy of each article or any part thereof. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970. Specify CCC Reference Number ISSN 0036–8733/83. $2.00 + 0.00.

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

Advertising correspondence should be addressed to C. John Kirby, Advertising Director, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, NY 10017.

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 5969, New York, NY 10017. The date of the last issue on your subscription is shown in the upper right-hand corner of each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, on a mailing label of a recent issue) as well as the new one.

Name ______________________

New Address

Street ______________________

City ______________________

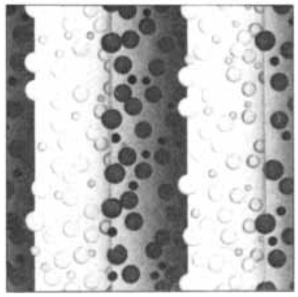
State and ZIP ______________________

Old Address

Street ______________________

City ______________________

State and ZIP ______________________



THE COVER

The painting on the cover represents a superlattice: a crystal whose remarkable electronic and optical properties arise from ultrathin layers of semiconductor (see "Solid-State Superlattices," by Gottfried H. Döhler, page 144). It is a doping superlattice, which consists of a single semiconductor in layers "doped" with different impurities. The orange planes are *n* layers: they are doped with donor atoms (*orange spheres*), which give up electrons and thus become positive ions. The blue planes are *p* layers: they are doped with acceptor atoms (*blue spheres*), which bind electrons and thus become negative ions. A voltage applied to the crystal sends free electrons (*small red spheres*) into the *n* layers, and "holes," or the absence of electrons (*small purple spheres*), into the *p* layers. In a typical semiconductor these charge carriers recombine in billionths of a second, and the energy of each electron-hole pair is dissipated by a photon (a quantum of light). In a superlattice the carriers can survive for several hours.

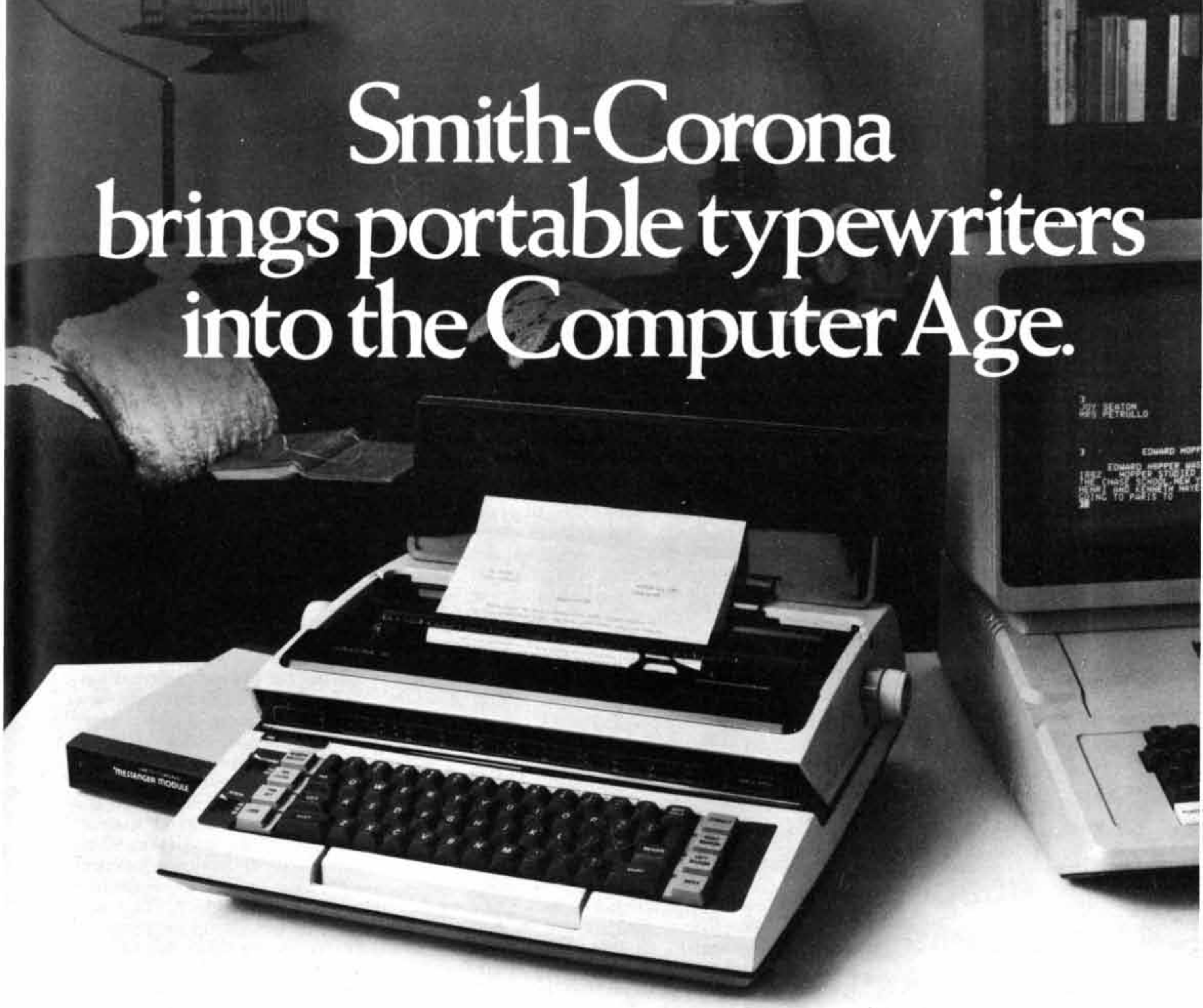
THE ILLUSTRATIONS

Cover painting by George V. Kelvin

Page	Source
18–21	Edward Bell
24D	Ilil Arbel (*top*), Edward Bell (*bottom*)
25–28	Edward Bell
39	U.S. Air Force
40–46	George V. Kelvin
47	Ilil Arbel
49	Edward J. Wasp
50–55	Walken Graphics
57	Jack O. Burns, Eric D. Feigelson and Ethan J. Schreier
58	Jack O. Burns and R. Marcus Price, University of New Mexico
60	Ian Worpole
61	Jack O. Burns, Eric D. Feigelson and Ethan J. Schreier
62–63	Ian Worpole
64	Jack O. Burns, Eric D. Feigelson and Ethan J. Schreier
65	Ian Worpole
127	Erika A. Hartwieg and Jonathan A. King, Massachusetts Institute of Technology

Page	Source
127A–139	Bunji Tagawa
140	Ravi Dhar
142	Bunji Tagawa
145	Pierre M. Petroff and Arthur C. Gossard, Bell Laboratories
146–151	Allen Beechel
153	Colin Renfrew, University of Cambridge
154	Andrew Tomko
157	West Air Photography
158–160	Andrew Tomko
161	Patricia J. Wynne after Alec Daykin
162	Colin Renfrew, University of Cambridge
165	James E. Mattler
166–171	Patricia J. Wynne
173	*Album of Krakatau*
174–183	Todd Pink
186	Maurice Krafft
189	Jearl Walker
190–196	Michael Goodman

Smith-Corona brings portable typewriters into the Computer Age.



Now Smith-Corona® takes computer technology and puts it into a new electronic *portable* typewriter. Typing has never been so simple!

The Smith-Corona electronic portable typewriter has features never even possible without today's advanced technology.

Simple One-Touch Memory Correction.

The Smith-Corona electronic portable has a memory, so it remembers what you type. Then, when you press the correct key, a letter, a word, even a line is lifted right off the page.

Simple Automatic Operations.

No more time-consuming operations like counting and backspacing. Just press the appropriate code key and type. Automatic centering, underlining, decimal tabulation, relocation, carrier return, flush right and reverse tabulation! The Smith-Corona electronic portable produces perfect results every time.

Simple Computer Compatibility.

The Smith-Corona electronic portable doubles as a computer printer. Simply connect the typewriter to the computer with optional Messenger Module.

Impressive Print Quality.

Add to these features a print quality as impressive as on big, expensive office machines. And the Smith-Corona electronic portable can be adjusted to type 10, 12 or 15 spaces per inch. Inexpensive ribbons, and printwheels which offer a choice of type styles, couldn't be easier to change.

Impressive Name.

Over the years, Smith-Corona has always been the leader in portable typewriter technology. No wonder our name has set the standards of the industry!

Impressive Price.

Best of all, Smith-Corona electronic portables are truly affordable. So you can have advanced office typing technology at home.

Try a new Smith-Corona electronic portable at your dealer's. And see for yourself how marvelous Computer Age typing can be.

Some electronic features not available on all models.

Smith-Corona

LETTERS

Sirs:

Philip Morrison writes in his review of Kenneth J. Hsü's book "The Mediterranean Was a Desert: A Voyage of the *Glomar Challenger*" [SCIENTIFIC AMERICAN, September] that the realization that the bed of the Mediterranean was once a desert was a "bright flash of inference ... in the lofty tradition of field geology [going] back to Louis Agassiz and James Hutton." Actually Hutton imagined this possibility almost two centuries before it was confirmed. I quote from Hutton's *Theory of the Earth,* Volume 1, page 75, although he said the same thing in a paper of 1788:

"The formation of salt at the bottom of the sea, without the assistance of subterranean fire, is not a thing unsupposable, as at first sight it might appear. Let us but suppose a rock placed across the gut of Gibraltar (a case nowise unnatural), and the bottom of the Mediterranean would be certainly filled with salt, because the evaporation from the surface of that sea exceeds the measure of its supply."

KEITH J. TINKLER

Brock University
St. Catharines, Ont.

Sirs:

"The Stave Churches of Norway," by Petter Aune, Ronald L. Sack and Arne Selberg [SCIENTIFIC AMERICAN, August], commented on the assimilation of Norse elements in the architecture and decoration of these 10th-century Christian structures. The diagrams and photographs enhanced the text in a beautifully woven manner. Sack's photograph of an exterior carving was not, however, properly described. The designs are an intermingling not of Norse and Christian elements but of diverse cultural influences. Besides the Christian inspiration in the plant tendrils, the dragon was introduced by Roman legions and became the guardian of the grave; the griffin was originally found in Hittite symbolism. The dragon and griffin are not intrinsic to Norse mythology but are cultural accretions of a later period.

WILLIAM ROBA

Scott Community College
Bettendorf, Iowa

Sirs:

"Chapman's rule" for judging a fly ball ["Science and the Citizen," SCIENTIFIC AMERICAN, April] would limit successful outfielders to college graduates with a good grade in mathematics. Who else could judge whether the tangent of the ball's angle of elevation rose at a constant rate?

May I, as an expert on the subject ever since I was a foreign student and was given a free ticket to a game between the Pittsburgh Pirates and the New York Mets, offer a simpler interpretation? The correct position of the outfielder is to stand where the ball appears to rise above the head of the batter at a constant speed. Hoping to have served the interest of the great game, play ball!

GUSTAV WINKLER

Flensburg, Federal Republic
of Germany

Sirs:

Michel Bur has done an admirable job in assessing the historical significance of the motte ["The Social Influence of the Motte-and-Bailey Castle," by Michel Bur; SCIENTIFIC AMERICAN, May]. Bur reports that the network of the first French mottes had no connection with existing population centers. I suspect that engineering factors may have played a significant role in 10th-century motte siting. Consider a motte 100 meters in diameter at its base and 20 meters high. Such an earthwork contains 410,000 cubic meters of earth. This is a substantial amount of material to excavate, transport, dump and form. Even with modern earth-moving equipment such a construction job, although it is simple, is not easy or quick. For example, if one assumes that the excavation is done in earth with a Euclid loader having a capacity of 790 cubic meters per hour and that the earth is hauled over an average distance of one kilometer with no more than a 3 percent grade by seven tractor-pulled bottom dump wagons with a capacity of 18 cubic meters working 10 hours per day, 112 days would be needed to complete the job.

The 10th-century engineers would therefore have faced a significant challenge. If only manually pushed wheelbarrows were used, I estimate that 56 million loads of 140 kilograms each would have been required. Larger loads would of course have been possible with domestic animals.

It seems reasonable that the determination of motte location included the consideration of factors such as the availability of earth, favorable topography and climate, suitable indigenous labor and the ability to protect the construction site for the many months, and possibly years, needed to complete the fortifications.

ROBERT E. SCHAFRIK, PH.D. P.E.

Dayton, Ohio

Test Drive the Leica R4.

We invite you to experience the perfection of the Leica® R4. We invite you to hold, inspect, and then test the Leica that has been acclaimed "...a miracle of modern technology." A limited production SLR of astonishing versatility, precision and beauty. With flawless Leitz® optics, two exposure measurement systems, five operation modes, and a series of advances unequalled in the history of photography.

The Leica R4 is Leica's proudest achievement. A photographic instrument designed for perfection, not price. Like every Leica since the beginning, it is lavished with attention to detail. Hand finishing. Hand assembly. Hand testing. No shortcuts or compromises. Because the Leica R4 is designed and built never to fail. And to have no limit to its useful life.

We invite you to experience the Leica R4. And through it enter Leica's world of superior photography. To see and appreciate for yourself the perfection of the photographic image that is uniquely Leica's.

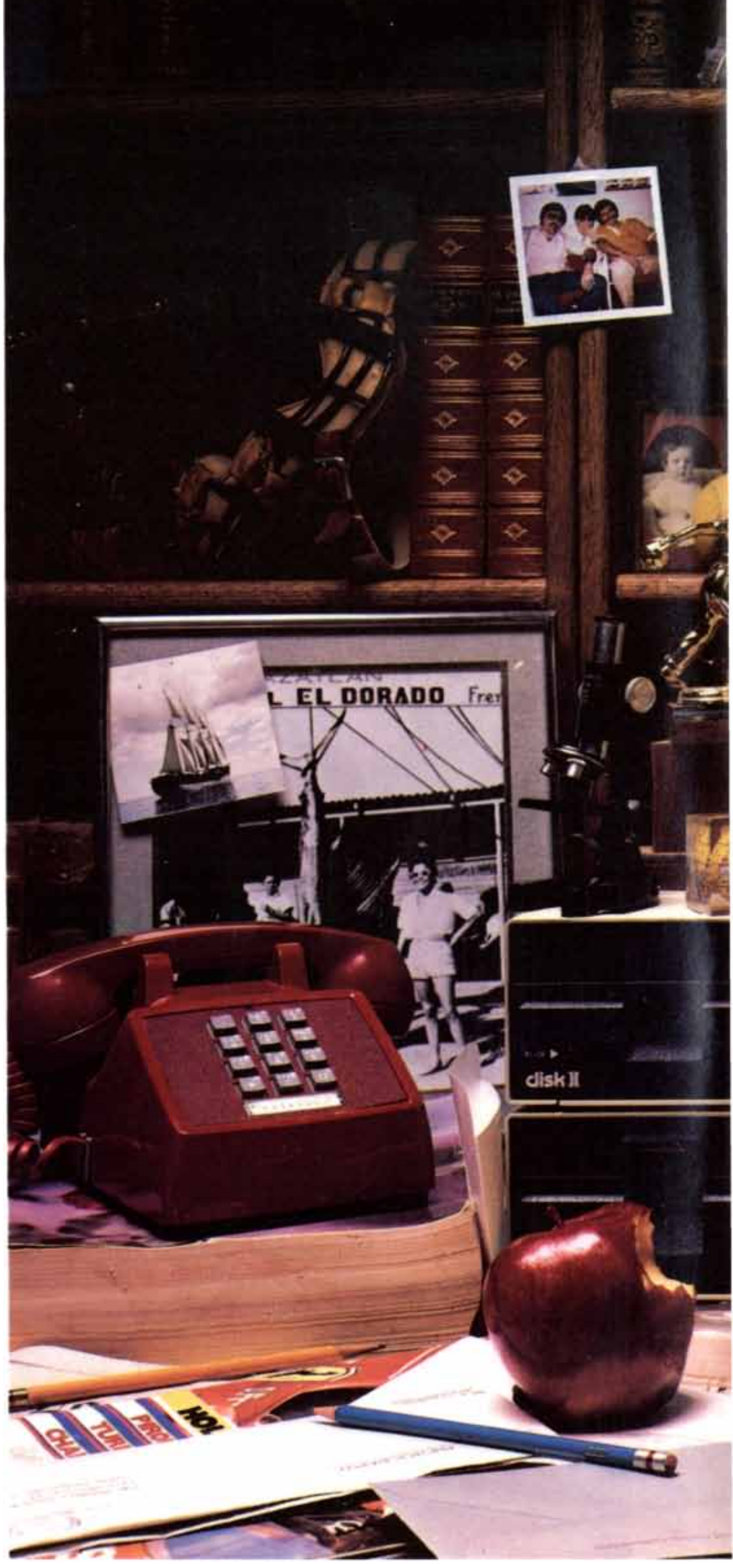
You can test drive the Leica R4 now, at your local Leica dealer. And to show his appreciation for your visit and your interest, your Leica dealer will enter your name in a special drawing for the ultimate sports coupe—the BMW 633CSi.* We think you'll find this high performance German import, the perfect accessory to your new Leica.

For the name of your nearest Leica dealer write E. Leitz, Inc., 24 Link Drive, Rockleigh, N.J. 07647 or call (800) 223-0507, ext. 801. In New Jersey call (800) 223-0509, ext. 801.

And you could win a BMW 633 CSi to drive it home in.

*No purchase necessary. Entries must be made through your Leica dealer before January 31, 1984. The winner will be notified by March 30, 1984. Void where prohibited by law.

The renaissance man is alive and well.



This once endangered species has made an astounding comeback.

Take Tom McWilliams.

He grew up with lots of skis, bicycles and microscopes.

And an Apple® II Personal Computer.

Using the latter to design a few video games, 16-year-old Tom was able to supplement his allowance.

By about $50,000. Without mowing a single lawn.

Of course, as we all know, some of the best renaissance men are women.

Like Margaret Waldmann. Who spent the first 50 years of her life becoming a world class pianist. And the last ten building one of the finest music schools in New York City—with some recent help from *her* Apple II.

After four decades of ever-narrowing specialization, more and more people are broadening their talents in more and more ways.

And more of them are using Apple IIs. Or our newest version, the Apple IIe.

It's even easier to use.

And thanks to the most sophisticated circuitry available, even more reliable.

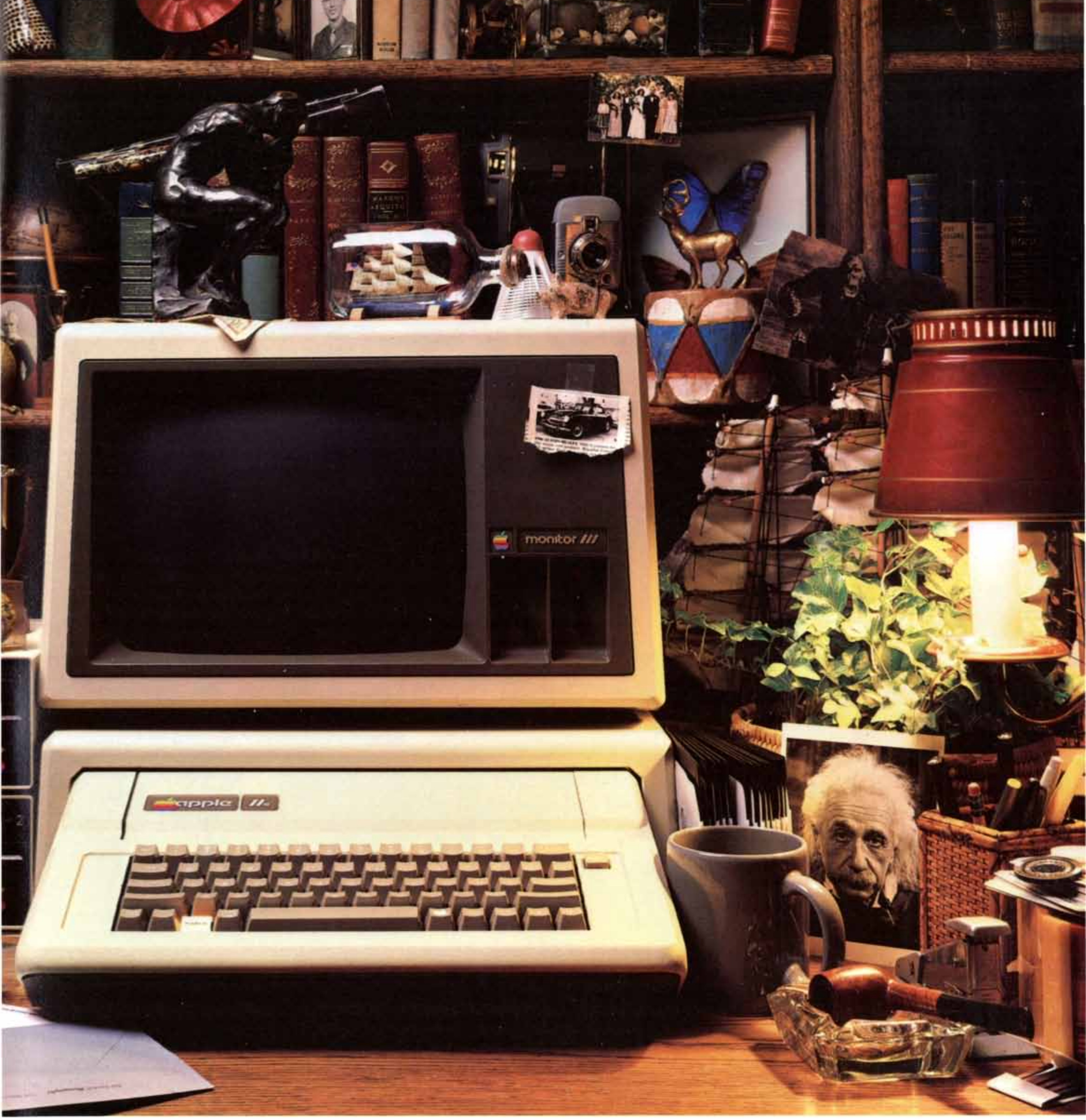
But improvements aside, it's basically the same remarkable design that's already earned access to more software programs, accessories and peripherals than any other personal computer.

With more on the way.

So year after year, Apple IIe's will become even more valuable.

Just like the people who use them.

Call **(800) 538-9696** for the location of the authorized Apple dealer nearest you, or for information regarding our National Account Program. In Canada, call **(800) 268-7796** or **(800) 268-7637.** Or write Apple Computer Inc., Advertising and Promotion Dept., 20525 Mariani Ave., Cupertino, CA 95014. © 1983 Apple Computer Inc.


monitor ///
apple

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

NOVEMBER, 1933: "Lord Rutherford spoke sensibly when he told the members of the British Association for the Advancement of Science at their recent meeting: 'Anyone who says that, with the means at present at our disposal and with our present knowledge, we can utilize atomic energy is talking moonshine.' Among physicists the belief in the eventual recovery of atomic energy has swung back and forth several times over the past two decades. Recently it has been seen that the recovery of even a minute amount of this energy involves putting in far more energy than is taken out. This is not to say, however, that new discoveries may not alter the situation, but one slant that is often given the public—that physicists are working toward the solution of the problem—is a misleading one. They are interested in the purely scientific problem of breaking up, transforming and otherwise manipulating the atom. Any power discoveries that may come from these experiments will be in the nature of a by-product."

"In July, Wiley Post startled the world by flying a 15,400-mile ring around it in seven days, 18 hours, 49½ minutes. And he did it solo to boot. There was a radio device of paramount interest on the *Winnie Mae.* It was mounted on the globe-circling ship by engineers of the U.S. Army Air Corps at Wright Field. Broadly speaking, it is a direction finder. To make use of it the *Winnie Mae* had an aerial from the rudder post to the fuselage, although she carried no radio in the ordinary sense. Stations along the world route had been asked to broadcast to her in flight on specified frequencies, and this they did with great success. It is obvious that if a pilot can receive signals from two stations of known location while he is in flight, he can orient himself and fix his position. But the direction finder of the *Winnie Mae* did more than this. The signals from a single station actuated a pointer that told Post with great exactitude whether he was on course. It may be said without too much stretching of the truth that he rode a radio wave around the world."

"The lowest temperature ever produced and measured by man, .085 degree on the absolute scale, has been achieved in the Kamerlingh Onnes Laboratory at the University of Leiden. This is extraordinarily close to the absolute zero point, at which all atomic motion would cease, electricity would flow without hindrance and other strange things would happen. Prof. W. J. de Haas of Leiden and Prof. H. A. Kramers of Utrecht, who did the experiments, used the method known as adiabatic demagnetization of paramagnetic salts."

"The farmer who through failure or inability to utilize labor-saving equipment is compelled to put too much work into each unit of his production naturally gets a lower rate of pay for his services than does the one who increases his efficiency through the use of machinery. This is true regardless of what prices may be. Studies made at Ohio State University reveal that a farmer who uses two-horse equipment in growing corn and picks it by hand puts 27.63 hours of man labor into growing and harvesting an acre of corn yielding 55 bushels, whereas the one who uses a tractor and tractor-drawn equipment for all operations except planting, and who uses a two-row picker, puts in only 7.48 hours. The one gets two bushels for each hour of his labor whereas the other gets 7½ bushels."

SCIENTIFIC AMERICAN

NOVEMBER, 1883: "Builders of machinery and machine tools are rapidly substituting low-carbon steel for refined iron in the part of machines subjected to strain and yet requiring stiffness. For piston and valve rods, for small finished shafts, rod connections and many other uses heretofore filled by iron, steel is now generally preferred. When well made and rolled or hammered into rods and small bars, the toughness of this kind of steel is remarkable. It contains no single 'sand bars,' or spiculae of hard iron, that take the edge off the turning tool or the planer cutter. This steel is admirably adapted for the feed screws of lathes, particularly screw-cutting lathes. The durability of steel as compared with iron is so much better that the value of the rolling and sliding parts is greatly enhanced, and fits can be made with much closer accuracy, whereas the increased first cost of material is nearly, if not quite, made up in the greater facility of working."

"Baron Nordenskjöld has sent a summary of his explorations and results in Greenland. An inland ice party started on July 4 from Auleitswik Fjord. When it was 140 kilometers east of the glacier border and 5,000 feet above sea level, it was prevented by soft snow from proceeding with sledges. Laplanders were sent farther on snow shoes. They advanced 230 kilometers eastward over a continuous snow desert to a height of 7,000 feet. The conditions for a snow-free interior consequently do not exist here. This expedition, during which men have reached for the first time the interior of Greenland, has given important results as to the nature of the continent. Over the whole inland there is ice."

"Mr. St. George has devised a means of recording a telephone conversation by the aid of photography. A circular plate of glass is coated with collodion and made sensitive as a photographic plate. It is placed in a dark chamber having a small slit, through which a pencil of light can fall upon the sensitive surface of the glass. The vibrating telephone plate actuates a shutter that varies the thickness of the luminous pencil corresponding to the vibrations after a plan introduced by Professor Graham Bell. The pencil falling on the photographic plate prints a dark line on it whose thickness is proportional to the vibrations of the telephone plate. The plate is revolved by clockwork like the barrel of a phonograph, and the record is afterward chemically fixed."

"In our time great popular fêtes can no longer dispense with the wonderful luminous effects that can be obtained by means of the electric light; and so, when last summer the Czar of Russia finally decided to have himself crowned with solemn ceremony, the city of Moscow prepared for the eve of the event a series of luminous decorations that left a deep impression upon the gigantic crowd that had come from all parts of the empire. The promenade of the court at the Kremlin, which was brilliantly illuminated at night by the electric light, took place in St. George's Hall. Never until then had there been lighted within the same space a like number of lamps. Trimmings formed of Edison lamps were mounted on the high tower of Ivan Velikoi, which rises in the center of the Kremlin, and formed silhouettes of fire of the most fantastic forms; and at the same time the belfries of the strange cathedral Wassili-Blajenoi were illuminated with many-colored facets, while a scintillation of electric points, running along the pinnacles of the ancient citadel, and reflected by the thousands in the bed of the Moskowa, seemed to give the waters of the usually dark river the appearance of a vast sheet of molten metal. Upon the 18 towers of the Kremlin there had been arranged in addition enormous regulators, which shot their luminous fascicles to a far distance and illuminated the cupolas of the Terem, of the Temple of the Saviour and of the cathedrals of the sacred precinct as if with a continuous lightning flash."

THE AUTHORS

MATTHEW BUNN and KOSTA TSIPIS ("The Uncertainties of a Preemptive Nuclear Attack") share an interest in the effects of current weapons systems and the means of controlling them. Bunn is an undergraduate at the Massachusetts Institute of Technology, majoring in political science. He also serves as an undergraduate research worker in the M.I.T. Program in Science and Technology for International Security. Tsipis is codirector of the program, which he founded with Bernard T. Feld in 1977. A native of Greece, Tsipis came to the U.S. in 1954 to study electrical engineering and physics. He got a B.S. and an M.S. in physics at Rutgers University before going on to obtain a Ph.D. in the same subject from Columbia University. He joined the department of physics at M.I.T. in 1966 and has remained there since.

EDWARD J. WASP ("Slurry Pipelines") is a chemical engineer who has devoted most of his career to developing methods for transporting solids long distances through pipelines. He went to Cooper Union College as an undergraduate before going on to earn his M.S. in mathematics at New York University. In 1951 he joined the Consolidation Coal Company as manager of process engineering and design. At Consolidation Coal he was responsible for the design and development work on the Eastlake coal-slurry pipeline. The Eastlake line, the first long-distance slurry pipeline in the U.S., ran 108 miles from southern Ohio to Cleveland; it began operation in 1957. In 1963 Wasp moved to Bechtel Petroleum, Inc., and soon afterward he was made manager of slurry systems, a job he still holds. In 1973 he became executive vice-president of the ETSI Pipeline Project, which was formed to build a coal-slurry line between mines in Wyoming and electrical generating stations as far east as Louisiana.

JACK O. BURNS and R. MARCUS PRICE ("Centaurus A: the Nearest Active Galaxy") are respectively assistant professor of astronomy at the University of New Mexico and professor of physics and astronomy and chairman of the department at the same institution. Burns was graduated from the University of Massachusetts with a B.S. in 1974; he went on to get his Ph.D. in astronomy from Indiana University. After two years of postgraduate work at the National Radio Astronomy Observatory he moved to New Mexico in 1980. In addition to the subject of the current article he is interested in galactic superclusters. Price got a B.S. from Colorado State University before going to Australia to continue his education; his Ph.D. in astronomy was awarded by the Australian National University in 1966. He returned to the U.S. to join the faculty at the Massachusetts Institute of Technology. After eight years he left M.I.T. to become the first radio-spectrum manager at the National Science Foundation. At the NSF he also served as head of the astronomy research section. He went to New Mexico in 1979. He was one of the discoverers of the Faraday effect, whereby the plane of polarization of radio waves from sources outside our galaxy is rotated as the waves pass through magnetic fields in interstellar space.

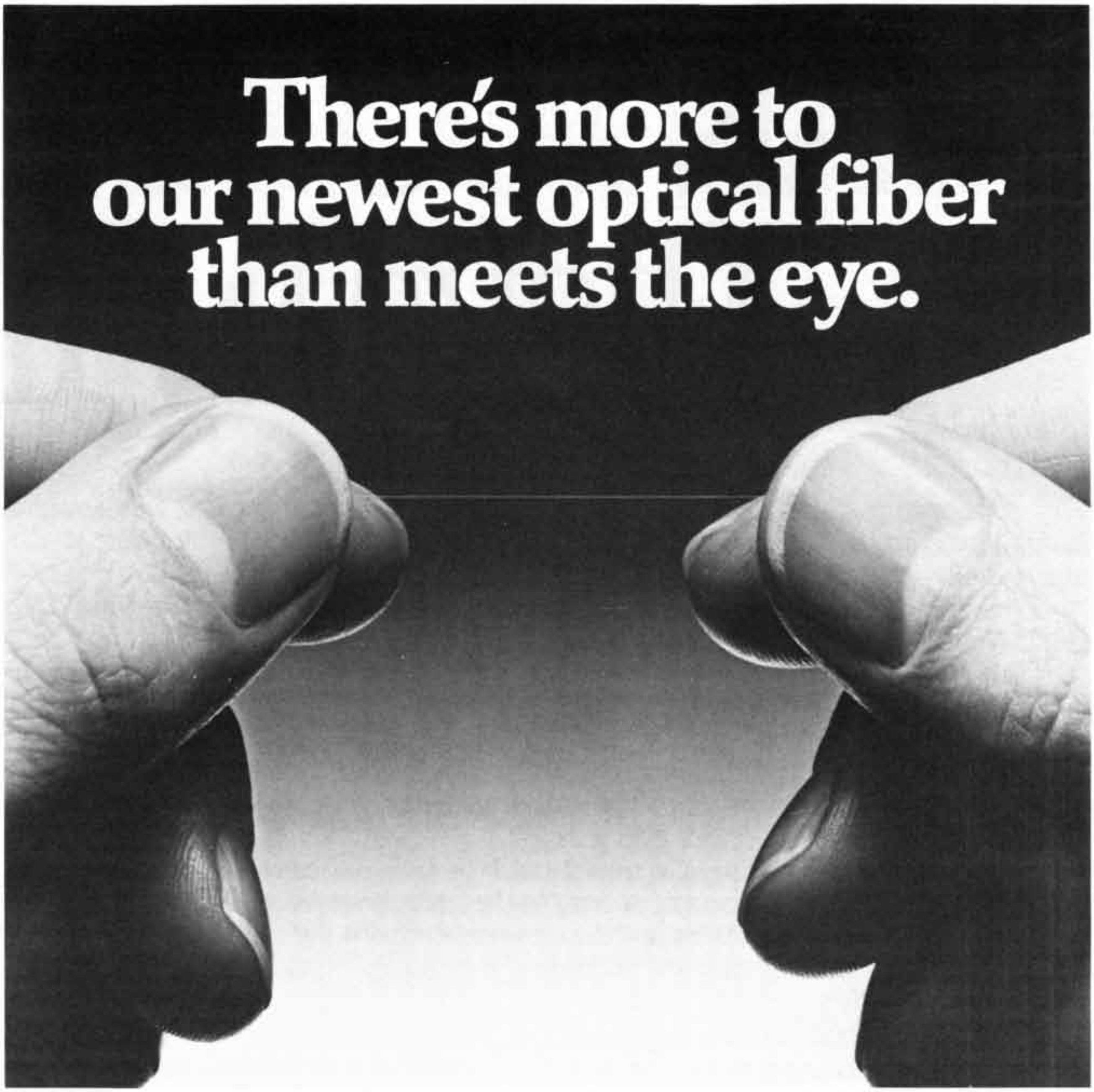
ROBERT A. WEINBERG ("A Molecular Basis of Cancer") is professor of biology at the Massachusetts Institute of Technology. His association with M.I.T. goes back to his undergraduate days. All three of his degrees are from that institution: his B.A., his M.A. and his Ph.D. in biology, which was given in 1969. After leaving to do postdoctoral work at the Weizmann Institute of Science in Israel and the Salk Institute for Biological Studies he returned to M.I.T. in 1972. He became professor in 1982; his appointment is in the Center for Cancer Research as well as in the department of biology. Weinberg is also a member of the Whitehead Institute for Biomedical Research.

GOTTFRIED H. DÖHLER ("Solid-State Superlattices") is a member of the staff of the Max Planck Institute for Solid-State Physics in Stuttgart. He writes: "I became a theoretical physicist as a graduate student, provided the term is taken to mean one who is a member of a theory group. Actually each time I started a theoretical investigation of a topic that had first interested me because of its fundamental physical aspects I soon discovered intriguing implications for experiments or practical application. Thus my work became largely motivated by the experimental and applied aspects and I avoided getting lost in studying higher-order modifications of the original theory. This was true of my work on high-electric-field effects in semiconductors and my later studies of the transport theory of amorphous semiconductors. With artificial semiconductor superlattices I have found a field of study where theory, experiment and application are very closely connected."

COLIN RENFREW ("The Social Archaeology of Megalithic Monuments") is Disney Professor of Archaeology at the University of Cambridge. A native of Britain, he received his undergraduate and graduate education at Cambridge. His B.A. was awarded by St. John's College in 1962 and his Ph.D. in archaeology by the university in 1965. After a tour of duty as a flying officer in the signals section of the Royal Air Force he joined the faculty of the University of Sheffield to teach prehistory and archaeology. After leaving Sheffield he went to the University of Southampton, where he served as professor of archaeology and head of the department. In 1981 he returned to Cambridge.

GERALD A. ROSENTHAL ("A Seed-eating Beetle's Adaptations to a Poisonous Seed") is professor of biological sciences and toxicology at the University of Kentucky. He got a B.S. at the College of Environmental Science and Forestry of the State University of New York at Syracuse. He continued his education at Duke University, obtaining a master's degree in forest pathology before earning his doctorate in plant physiology and biochemistry. He served for several years as a postdoctoral fellow in biochemistry at the National Institutes of Health before going to the department of biology at Case Western Reserve University. He left Case in 1972 to join the faculty at Kentucky. In 1979 he served as Lady Davis Professor of Agricultural Entomology at the Hebrew University of Jerusalem.

PETER FRANCIS and STEPHEN SELF ("The Eruption of Krakatau") are geologists from Commonwealth countries who attended the Imperial College of Science and Technology in London, who share an interest in the effects of volcanic eruptions on climate and who currently work in Texas. Francis was born in Zambia, which was then called Northern Rhodesia. He moved to England to get his education in geology. He attended Imperial College as both an undergraduate and a graduate student, earning his Ph.D. in geology in 1969. Since 1971 he has been a member of the department of earth sciences at the Open University in England. Since 1981 Francis has served as senior visiting scientist at the Lunar and Planetary Institute in Houston. Self is a native of England who attended the University of Leeds as an undergraduate. After getting his B.S. he went to Imperial College, where he obtained his Ph.D. in geology in 1974. After earning his doctorate he went to the Victoria University of Wellington in New Zealand to collect data on eruptions at the Taupo volcanic center. He came to the U.S. to do work on the effect of volcanic eruptions on climate; the project was funded by the National Aeronautics and Space Administration. In 1979 Self joined the faculty at Arizona State University, leaving three years later to take up a job in the department of geology at the University of Texas at Arlington.



In this picture, it's what you can't see that counts.

These hands are holding an optical fiber—an incredibly pure strand of glass that transmits both voice and data as pulses of light.

But as extraordinary as any optical fiber is, this one is even more so.

This is a new kind of optical fiber, with a core only 1/6th as thick as a human hair (or any conventional optical fiber).

You see, the thicker a fiber's core is, the more dispersed—or "fuzzy"—the light signal gets as it passes through.

In our new optical fiber (technically, a "single mode") the core is so fine that it virtually eliminates this "fuzziness."

This allows billions of pulses per second to be sent through the fiber over longer distances. A lot more information can be transmitted in a given time.

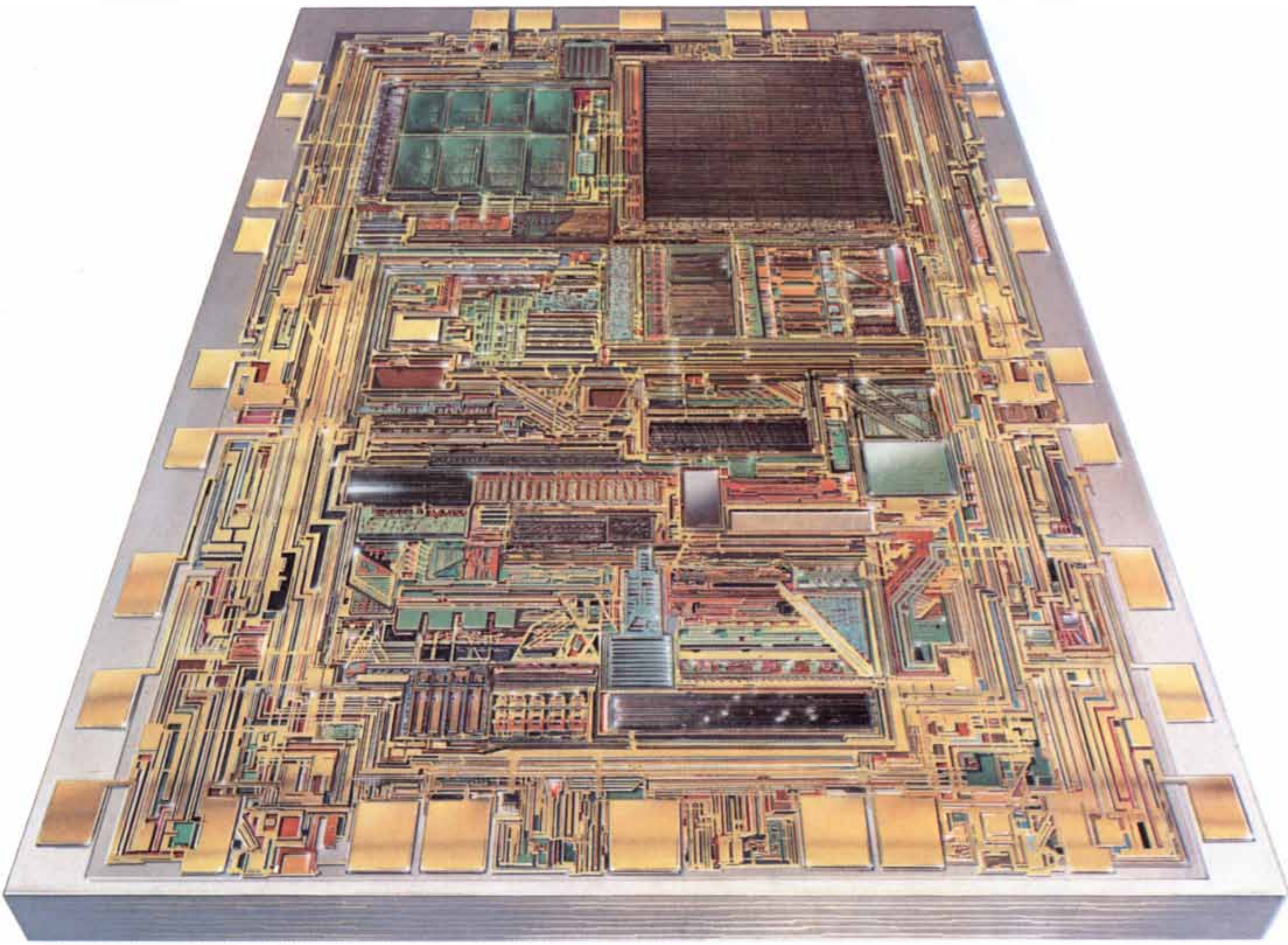
So, a telephone line made of "single mode" fibers can carry more information than any telephone line could before.

Considering the continuing explosion of voice and data traffic, this new kind of optical fiber will be vitally important.

And that's something that's easy to see.

The best ideas are the ideas that help people. **ITT**

THE HIGH TECH IS PICTURE?



What's on the best-seller list in IBM Personal Computer software?

COMPUTER RECREATIONS

A progress report on the fine art of turning literature into drivel

by Brian Hayes

Almost any computer program can be made to yield meaningless results if it is given sufficiently muddled information to work with; this is the sense of the tired adage "garbage in, garbage out." The principle is now so well established that no one would take much notice of another demonstration. With a little thought and effort, however, it is possible to create a program that accepts as its input great masterworks of literature and nonetheless produces as its output utter nonsense. In goes the last act of *Macbeth;* out comes a tale told by an idiot, full of sound and fury, signifying nothing. Now *that* is data processing. (The inverse transformation, alas, seems to be much harder.)

The conversion of literature into gibberish is done in two stages. First a text is "read" by the program, and certain statistical properties are extracted and recorded. The statistics define the probability that any given letter of the alphabet follows another letter, or another sequence of letters, in the source text. In the second stage a new text is generated by choosing letters at random in accordance with the recorded probabilities. The result is a stream of characters that reproduce the statistical properties of the original text but whose only meaning, if any, is a matter of accident.

I cannot imagine a cruder method of imitation. Nowhere in the program is there even a representation of the concept of a word, much less any hint of what words might mean. There is no representation of any linguistic structure more elaborate than a sequence of letters. The text created is the clumsiest kind of pastiche, which preserves only the most superficial qualities of the original. What is remarkable is that the product of this simple exercise sometimes has a haunting familiarity. It is nonsense, but not undifferentiated nonsense; rather it is Chaucerian or Shakespearian or Jamesian nonsense. Indeed, with all semantic content eliminated, stylistic mannerisms become the more conspicuous. It makes one wonder: Just how close to the surface are the qualities that define an author's style?

The process of generating random prose has been investigated in detail by William Ralph Bennett, Jr., of Yale University. He has made the statistics of language a major theme of a course on the applications of computers, and the topic also figures prominently in his introductory textbook on programming, *Scientific and Engineering Problem-solving with the Computer*. (The book is a good deal livelier than the title might suggest. The problems taken up include the aerodynamics of the 1950 Princeton-Dartmouth football game, which was played in a hurricane; the diffusion of syphilis through a population of sailors and prostitutes, and a spectral analysis of the krummhorn, oboe and "mode-locked garden hose.")

Bennett notes that the earliest known reference to the random generation of language is in the *Maxims and Discourses* of John Tillotson, archbishop of Canterbury in the 1690's. In making a case for divine creation Tillotson wrote: "How often might a Man, after he had jumbled a Set of Letters in a Bag, fling them out upon the Ground before they would fall into an exact Poem, yea or so much as make a good Discourse in Prose? And may not a little Book be as easily made by Chance, as this great Volume of the World?"

For most modern considerations of random language the point of departure is Sir Arthur Eddington's statement of 1927: "If an army of monkeys were strumming on typewriters, they *might* write all the books in the British Museum." Eddington too meant to emphasize the improbability of such an outcome; he cited it as an example of an event that could happen in principle but in practice never does. All the same, since Eddington's time the possibility of finding genius in the random peckings of monkeys has taken on a literary life of its own. Bennett mentions works by Russell Maloney and Kurt Vonnegut, Jr., and a nightclub act by Bob Newhart.

The process Eddington envisioned can be simulated by a program I shall call an order-zero text generator. First an alphabet, or character set, is decided on, which determines what keys are to be installed on the monkeys' typewriters. In some higher-order simulations it becomes important to keep the number of symbols to a minimum, and for consistency it seems best to adopt the same character set in the order-zero program. I have therefore followed Bennett's recommendation in choosing a set of 28 symbols: the 26 uppercase letters, the word space (which the computer treats as a character like any other) and the apostrophe (which is commoner in much written English than the three or four least-common letters are).

The ideal, unbiased monkey would at any moment have an equal probability of striking any key. This behavior can be simulated by a simple strategy. Each symbol in the character set is assigned a number from zero to 27. For each character to be generated a random integer is chosen in the same range and the corresponding character is printed. A small specimen of text created by this procedure is shown in the illustration on this page. It bears no resemblance to written English or to any other human language. "Words" tend to be extraordinarily long (on the average 27 characters) and thick with consonants. The reason, of course, is that letter frequencies in real English text are far from uniform. The word space alone generally accounts for roughly a fifth of the characters, whereas J, Q, X and Z together make up less than 1 percent. In an order-zero simulation all the characters have the same frequency, namely 1/28.

The comic routine by Bob Newhart concerns the plight of the inspectors who must read the monkeys' output. After many hours of poring over unintelligible prattle they come upon the phrase, "To be or not to be, that is the gesorenplatz. . . ." In fact, getting even that far is wildly improbable; the first nine words of Hamlet's soliloquy can be expected

PWGMMLTHIDVGRHPEDFCXFEKFNOPYPQSXZRUXG'YS'AEEU PEDEGLQYFUWPO'IKI
QTONIXJKZEUKDXWKKJREHYHPKWUJHLEJNBPLQ AIEOQXUBJYYVIFFDPQGIGZNTI
RQXPDJ NQESPQMCRSNGMKQEZICZV'GSWALK ZZEYIBBOTDCRSMK'VI MRCZXUBI
SNEQ'VQQHFQUCBJXZRVVNIBHFJEFTCFJPWFOIYHOMPNFSFWKNCMVLOJJBX
QV KIZTLNRWGGTZFPZPQQCGVJCPAYRDQJRMYSWCGABRXLERCYYRHQCHTOQ'UT
FMRITFTIZUIWTSTXWQGOCAFXJOZYKSTV'BYOBEUFIRQWQ VOUVQJPRKJWBKPLQZCB

Order-zero random text, drawing on an alphabet of 28 symbols

SELECTAVISION

VIDEO MONITOR

to turn up once out of every 2×10^{46} characters. In a run of 50,000 characters I was able to find one instance of TO and another of NOT; they were many lines apart. (I did not read the 50,000 characters but instead made the search with a pattern-matching program.)

A first step toward improving the monkeys' literary skills is to adjust the probability of selecting a given letter so that it reflects the letter's actual frequency in written English. In effect, the plan is to build a typewriter with, say, 2,500 space keys, 850 E keys, 700 T keys and so on. The letter frequencies might be averages calculated from a large sample of English prose, but it is both more convenient and more interesting to base them on a particular source text. A program that chooses characters with such a frequency distribution is a first-order text generator.

The letter-frequency values can be represented in a one-dimensional array with 28 elements. The array is a block of storage locations in the computer's memory, organized so that any one element can be specified by an index, or subscript, between zero and 27. In order to fill up the array one could count the instances of each letter in the text and enter the values by hand. It is better, however, to let the program do the counting, even when that means the text itself must be prepared in a machine-readable form. The counting program initially sets all the elements of the array to zero. The text is then examined one character at a time, and for each occurrence of a character the corresponding array element is incremented by 1.

First-order random text is generated by making the probability of selecting a character proportional to the character's array element. One method works as follows. A random number is generated in the interval between zero and an upper bound equal to the sum of the array elements (which is also the total number of characters in the source text). The first array element, which might record the occurrences of the letter A, is then subtracted from the random number. If the result is zero or less, an A is printed; otherwise the next element (representing B) is subtracted from the value remaining after the first comparison. The successive subtractions continue until one of them gives a zero or a negative result, and the corresponding character is selected. Note that the procedure cannot fail to make a selection, since the random number cannot exceed the sum of the array elements.

A sample of first-order random text is shown in the upper illustration at the right. It is based on a frequency array compiled from a passage in the last chapter of James Joyce's *Ulysses*, the chapter known as "Ithaca," or Molly Bloom's soliloquy. I had a reason for choosing it: the absence of punctuation in the random text is of little consequence because the source text too is unpunctuated.

The information on letter frequencies embodied in a first-order random text brings an improvement, but one would hardly call the text readable. Although the average word length (4.7 letters) is near the expected value (4.5 letters), the variance, or deviation from the average, is much too great. Words in normal English, it seems, are not only short but also have a narrow range of lengths; in the random text the distribution is much broader. Apart from the question of word length there is the mat-

FIRST ORDER

HUD T ALONIT NTA SN TVIOET ELERFOAD PE TRLTWTL N CABEG TYLUEMU TIGT
BH OFDRRIC O STU HOOOTO YATNDL UYA HWAE SS NLSDB OTRORT DEERARFT
D LBFF HHARE MW OSPE OFOIT SEOUN GTUMG H N GHKOY T EAOS A SD E TNNE
PEHAGIADIHNATO AATSAGI ED INNE ABRA TAAM GT E TWNO HEWIIGUTNCM GA SFHHY
HREBH RARE OOSY LFE OC EGGTA WIFRTYE EUS DA ETO WF EIT ERNETEBTSTTELO
NTAAN O YEETWNSONRNHN TYHVN NLUESETTHLGEAKPNNMTIA TSM REEANTVONC POE
RUTP EOIT L IEETGTWHSW H KHHER W OLIOEWOEPT D AEYBSTNHGDNPT C TNLINHH
KHHE E RTVIOB EI K EOAFPUTSTTAS NA LAN SRDF D NMTHESKO UGEEDICRAWDT OBD
TUIML WSORGNETE

SECOND ORDER

BEGASPOINT IGHIANS JO HYOUD WOUMINN BONUTHENIG SPPRING SBER W IDESE WHE D
OOFOMOUT O CHEDA AFOOIAUDO IS WNY UT DRSASER LD OT POINE ETHAT FOEVEL BE
ORRI IVER BY HE T AS I HET W BE T WAU GIM UTHENTOTETHAVE THIKEWOITOCOUTORE
TATHASTHEE AT D Y WAN TOND SE TEDING US AKIN WING W TE T BO TOTSTHINGATONO
EN T LLY WID OUCOUSIND HEF THIMES AG T BENG LORYE ALLATHOMOFTHER TOUDIMS YS
S ORYRY THERNG S HE M G M ANG S CITOOFO HEN G BEST ONDLOL ANE DO HE
ICISEKERIT ME NKITHADIMUPL WHES HT BATHE T LOR WITULOWAYE WATHEG M
LEROMAUN OUGS POUPO O HASING LIN ON ASHAN AWFAS HET ND MEDE

THIRD ORDER

MAY THOT TO THER YOURS CHIM JOSE EY EILLY JUSED AND HID YEL THE MARK WASK
TROOFTEN HEREY LING SH THAVERED HER INCED I MEA BUT DAY WOM THE EAKIN WIPS
AS SUGH THE WAY LIARADE TH MY HE ALMASEETIR ANICIOUT JOSIDNTO GRATEVE NO
VER BIGH WER ACCOW WAS I GEORE HENDSO EGGET PUT TO SQUAD TRADE OFF GIN
GO ME HER SPING HE CONE WELL FEWHEY THEYES AND AND QUICE YOULDNT HER
ORL SO MAKING RINGS SOMET DREAVE HISETTO COMAD THAT ME WE MIG TOLD THE
THERFUMBECK OT OFF FEELP HE WAST ITS LETHOTTEN ITHEE ROWN YOURS FEL FOR
SOME IF WIS HE STAKED UPPOIS SHENS NO TILL HIM I WAY SO WHATEALWAS WER TWE
NER DING O THIS IT IN ANIGH ACK REAN THAT DO GETHE BITER

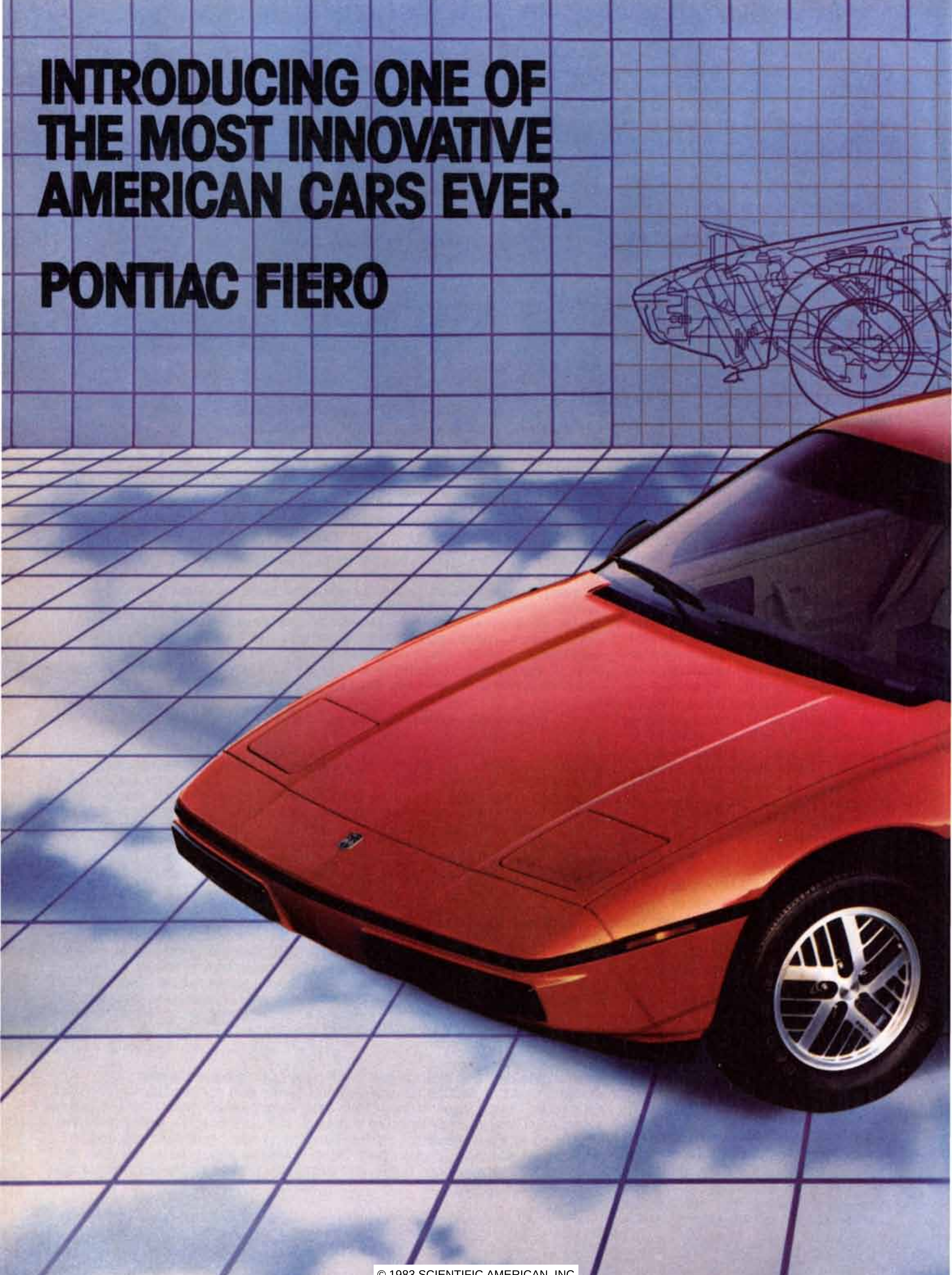
First-, second- and third-order random text based on "Molly Bloom's soliloquy"

AD CON LUM VIN INUS EDIRA INUNUBICIRCUM OMPRO VERIAE TE IUNTINTEMENEIS
MENSAE ALTORUM PRONS FATQUE ANUM ROPET PARED LA TUSAQUE CEA ERDITEREM IN
GLOCEREC IOVELLUM ET VEC IRA AE DOMNIENTERSUO QUE DA VIT INC PARBEM ETUS
TU MEDE DERIQUORUMIMO PEREPORIDEN HICESSE COSTRATQUIN FATU DORAEQUI POS
PRIENS NOCTA CIENT HUCCEDITAM PET AUDIISEDENDITA QUE GERBILIBATIA VOLAEQUE
ORECURICIT FES ADSUE ARCUMQUE LULIGITO PIMOES PERUM NOSUS HERENS EA
CREPERESEM ETURIBUS AVIS POS AT IS NOMINE FATULCHENTURASPARIS AUDEDET PARES
EXAMENDENT DUM REMPET HA REC ALEVIREM ORBO PIERIS ATAE PARE OCERE RAS

QUALTA 'L VOL POETA FU' OFFERA MAL ME ALE E 'L QUELE ME' E PESTI FOCONT E 'L M'AN
STI LA L' ILI PIOI PAURA MOSE ANGO SPER FINCIO D'EL CHI SE CHE CHE DE' PARDI
MAGION DI QUA SENTA PROMA SAR OMI CHE LORSO FARLARE IO CON DO SE QUALTO
CHE VOL RICH'ER LA LI AURO E BRA RE SI MI PAREMON MORITA TO STOANTRO FERAI TU
GIA FIGNO E FURA PIA BUSCURA QUAND'UN DEL GUARDI MIN SA PAS DELVENSUOLSI PER
MUSCER PIE BRUI TA DORNO TITTRA CHE PO E PER QUE LI RINONNIMPIAL MIN CH'I'
BARVEN TA FUI PEREZZA MOST' IO LA FIGNE LA VOL ME NO L'E CHE 'L VI TESTI CHE
LUNGOMMIR SI CHE FACE LE MARDA PRESAL VOGLICESA

PONT JOURE DIGNIENC DESTION MIS TROID PUYAIT LAILLE DOUS FEMPRIS ETIN
COMBRUIT MAIT LE SERRES AVAI AULE VOIR ILLA PARD OUR SOUSES LES NIRAPPENT LA
LA S'ATTAIS COMBER DANT IT EXISA VOIR SENT REVAIT AFFRUT RESILLESTRAIS TES FLE
LA FRESSE LES A POURMIT LE ELLES PLOIN DAN TE FOLUS BAIER LA COUSSEMBREVRE
DE FOISSOUR SOUVREPIACCULE LE SACTUDE DE POU TOUT HEVEMMAIT M'ELQU'ILES
SAIT CHILLES SANTAIT JOU CON NOSED DE RE COMMEME AVAIL ELLE JE TER LEON DET
IL CED VENT J'ARLAMIL SOUT BLA PHYSIS LUS LE SE US VEC DES PEUSES PAU HAS BEAU
TE EMANT ELLE PLANQ HEUR COIRACOUVRE BIENE ET LUI

Third-order Latin (Virgil), Italian (Dante) and French (Flaubert)


INTRODUCING ONE OF THE MOST INNOVATIVE AMERICAN CARS EVER.
PONTIAC FIERO


Fiero
Fiero can be described as an exciting car. It can also be considered a practical and durable car. And, by its technical specifications alone, Fiero can be legitimately called a sports car. The beauty of Fiero is that it's *all* of these cars. Best of all, Fiero is one of Pontiac's best examples of what driving is all about: pure unabridged fun!
Fiero is America's *first* two-seat, mid-engined production car. Fiero features fully-independent suspension, rack and pinion steering and power four-wheel disc brakes, all *standard*. Fiero is powered by a responsive 2.5 liter 4-cylinder engine with electronic fuel injection.
Fiero is designed to cope with the realities of driving in the real world: Its Enduraflex™ panels will *never* rust, the front fenders "give" on minor impacts, and all panels are "ding" resistant.
The mid-engined Pontiac Fiero! Fiero pr ces start at just $7,999.† Even at $8,499† (Fiero Sport Coupe shown), Fiero takes its rightful place among the great cars of the *world*!
†Manufacturer's Suggested Retail Price including dealer prep. Taxes, license, destination charges and optional equipment additional. Fiero Sport Coupe (shown), with WS6 perf. pkg.—$463 additional.
Pontiac Fiero Sport Coupe offers an EPA EST MPG of (26) and a highway estimate of 40. Use estimated MPG for comparisons. Your mileage may differ depending on speed, distance, weather. Actual highway mileage lower. Some Pontiacs are equipped with engines produced by other GM divisions, subsidiaries or affiliated companies worldwide. See your Pontiac dealer for details.
GM
PONTIAC WE BUILD EXCITEMENT

ter of word content. Even though the letters appear with the correct frequency, their sequence is utterly random, and most of the resulting "words" are not English and could not be. A letter series such as WSTLNTTWNO or HIUOIMYTG is not merely meaningless but impossible. In one run of 2,000 characters the longest recognizable word was, appropriately, RARE.

The next refinement is the crucial one because it can be extended, at least in principle, to an arbitrarily high order. The root of the idea is that a letter's probability of appearing at a given point in written English depends strongly on the preceding letters. After a V, for example, an E is most likely; after a Q, a U is all but certain. The procedure, then, is to set up a separate frequency table for each symbol in the character set. The frequencies are recorded in a two-dimensional array with 28 rows and 28 columns, for a total of 784 elements. An example of such a frequency table is shown in the upper illustration on page 24D. (The array is "normalized" by rows, meaning that comparisons are valid only within a row.)

When text is generated from the two-dimensional array, the character most recently chosen determines which row of the table is examined in picking the next character. For example, if the preceding letter is a B, only the elements of the second row are taken into consideration. The largest element of the second row is E, and so it is the likeliest letter; A, I, L, O, R, S and U also have a chance of being selected. Impossible combinations such as BF and BQ have zero frequency, and so they can never appear in the program's output.

Second-order random text begins to show the first hints of real linguistic structure. The distribution of word lengths is only a little wider than it ought to be. Real words are not uncommon, and there are many near-misses (such as SPPRING or THIMES); a large majority of the words are at least pronounceable. Common digraphs such as TH begin to show up often, and the alternation of vowels and consonants is a conspicuous pattern.

The next step should be obvious. A third-order algorithm chooses each letter in the random text according to probabilities determined by the two preceding letters. This calls for a three-dimensional array with 28 planes, each plane being made up of 28 rows of 28 columns. Suppose at some point in the creation of the text the letter sequence TH has been generated. The program must then look to the 20th plane (corresponding to T) and to the eighth row on that plane (corresponding to H). In that row E is the likeliest choice, although A, I, O and the space symbol also have nonzero probabilities. If E is indeed selected, then in the next iteration the choice will be made from the fifth row of the eighth plane, the position in the table specified by the letter sequence HE. Here the leading candidate is the word-space character, followed by R.

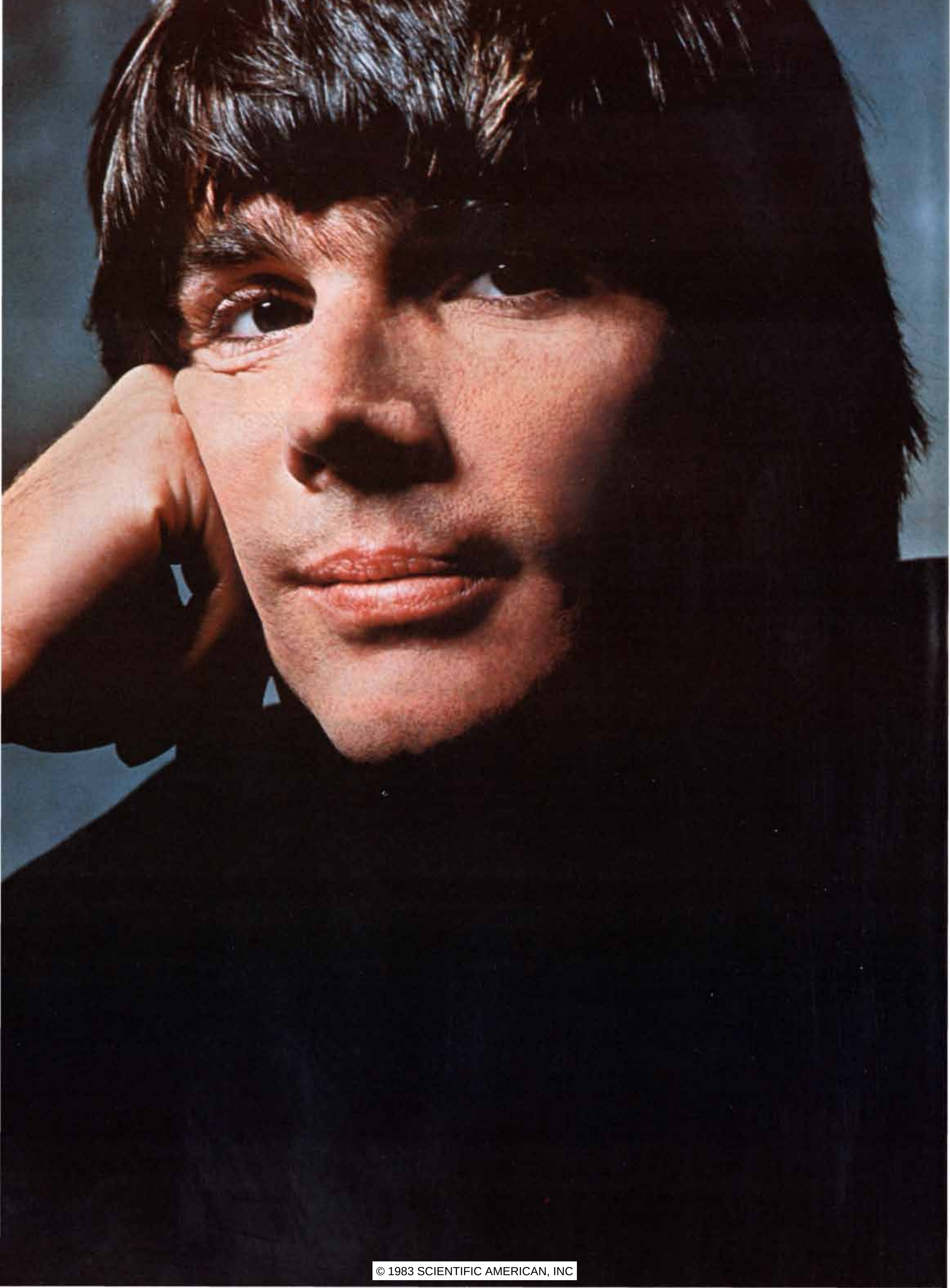
In third-order text no three-character sequence can appear unless it is also present somewhere in the source. Because spaces are included in the accounting, that is enough to guarantee only that all one-letter words will be real words; in effect, only the letters I and A can appear in isolation. The actual performance, however, is a good deal better than the guarantee. Virtually all the two-letter sequences are words, and so are most of the three-letter sequences. Often a string of several words in a row turns up: PUT TO SQUAD TRADE OFF GIN GO ME HER. Even quite long nonwords have a certain phonetic plausibility. After all, it is only a matter of accident that ANYHORDANG HOUP TREAFTEN is meaningless in English.

While reading a sample of third-order random text, I am reminded of stage-performance double talk and of glossolalia, the "gift of tongues" that figures in certain Pentecostal liturgies. One might guess that there is some significance in the resemblance: perhaps people who have learned those arts carry out an unconscious statistical analysis somewhat like the one the program does. I think another explanation is likelier. Double talk and glossolalia seem to involve the random assembly of phonemes, the fundamental atoms of spoken language. It may be that three letters is about the right size for the written representation of a phoneme.

With third-order text the stylistic qualities of the source begin to have a perceptible effect. Where the contrast in styles is great, the corresponding random texts are also clearly different, although it is not easy to say exactly what constitutes the difference. I am inclined to describe it as a matter of texture, but I am not at all sure what texture is in prose. Is it whatever remains when all the meaning is removed?

Even when individual mannerisms cannot be perceived in third-order random prose, identifying the language of the source is easy. Patterns of vowels and consonants and the characteristic endings of words are unmistakable. The bottom illustration on page 21 shows brief examples of Latin (Virgil), Italian (Dante) and French (Flaubert). Someone who knows only the "look" of one of these languages might have trouble distinguishing the ersatz product from the real thing.

Before considering what lies beyond the third-order approximation, I should like to mention some other applications



Bill Budge wants to write a program so human that turning it off would be an act of murder.

ARE YOU SURE YOU WANT TO CALL THIS GUY AN ARTIST?

IN A BEDROOM in a frame house in Berkeley, California, a guy who looks like he might have stepped out of a TV family series is playing with some ideas that could change your life. They are ideas that are amusing, even charming. And they are ideas that are, quite frankly, a little scary.

His name is Bill Budge and he talks about things like how programming for a microcomputer is like writing a poem using a 600-word vocabulary. He talks about how the elements are so limited and how you have to make them mean so much. And he talks about how, if you do it right, you can make those elements suggest something more than aliens—something that begins to make you believe it has a life of its own, something he calls "a software friend."

A software friend. It sinks in slowly.

To create a personality in the computer, you have to come to some decisions about what personality is in the first place. We often think of it in nearly the same way we think of "habit" or "character traits"—a way of describing continuity in our behavior from one moment to the next. (F. Scott Fitzgerald called it "a series of consistent gestures.")

According to Budge, however, the essence of a software friend is quite the opposite. "Creating the illusion of personality," he says, "means creating an intelligence that's always changing. It reacts differently to different situations."

The idea is probably ten years away from actuality. But when it comes to working such a mojo on our home computers, well, Budge stands about as good a chance as anyone of pulling it off.

After all, look at PINBALL CONSTRUCTION SET. Everyone always knew Budge was good, but when he cranked out PINBALL, well, the switchboard lit up like a Christmas tree.

It was a program that changed the way people thought about personal computers. Instead of reacting to the machine, you were suddenly inside it, trafficking information this way and that, making things. It was like programming, but with familiar items—you'd grab this bumper, move those flippers, change the colors, then shoot a ball through it all and wonder. Maybe for the first time in a popular program, you could feel the power of the computer.

Steve Wozniak called it "the best program ever written for an 8-bit machine." And suddenly, what-Budge-would-do-next was something you heard people talking about. To Budge himself, however, things weren't quite that simple.

"Sometimes I worry," he says. "I worry about the ability of software to absorb you, focus on you, steal you away from your family and friends. Because in its short-term excitement, it seems to be more interesting. Of course, it's not."

He leans on his hand. "Not yet."

Bill Budge's classic PINBALL CONSTRUCTION SET is just one of more than a dozen remarkable publications by a company called Electronic Arts. We're an association of software artists, united by a common goal: we want to realize the potential of the home computer. To do this, we're creating software worthy not only of the capabilities of these machines, but also of the minds that use them. If you'd like to know more about our company and its products, call (415) 571-7171, or write us at 2755 Campus Dr., San Mateo, California 94403.

ELECTRONIC ARTS™

PINBALL CONSTRUCTION SET is now available for Apple, Atari, and Commodore computers. To see it, and to receive a free "We See Farther" poster, stop by the authorized Electronic Arts dealer nearest you.

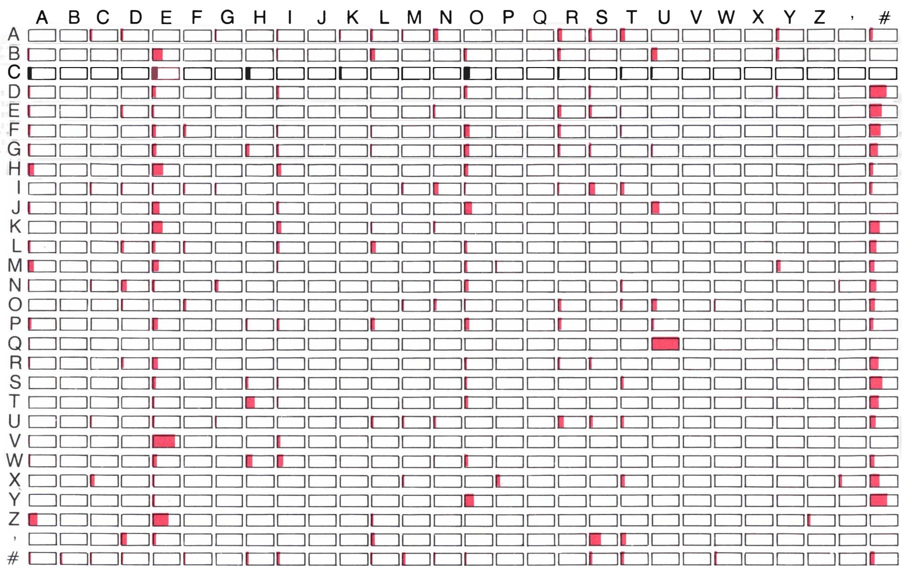
of letter-frequency tables. Bennett, in a discussion of the entropy of language, points out that the tables enable one to calculate the amount of information conveyed per character of text. The information content essentially measures the difficulty of predicting the next character of a message. It is at a maximum in the order-zero simulation, where every possible character has equal probability; in other words, the information content is greatest when the text is totally unintelligible. The idea of predicting characters leads to a discussion of error correction in telecommunications and to the design of algorithms for solving ciphers and cryptograms.

Another area worth exploring is the alteration or manipulation of the frequency array. How is the random text changed, for example, when each element of the array is squared? An example of Molly Bloom squared is shown in the bottom illustration below. Because the procedure exaggerates differences between array elements, the effect is to "sharpen" the frequency distribution; common words become still commoner. Many other transformations are possible. Adding a constant value to all the array elements has a disastrous effect, even if the constant is a small one: all the impossible letter combinations, which one has been working so hard to eliminate, become possible again.

One intriguing idea is multiplying the entire array by −1 in order to generate text by, say, Alexander anti-Pope. For any given combination of letters, whatever subsequent letter is likeliest in Pope would be the least likely in anti-Pope. Literary aptness might best be served if the product resembled the works of Colley Cibber. Actually, it is an almost patternless jumble.

The result is somewhat less discouraging (although still far from illuminating) when two arrays are added or multiplied. In this way one can create unlikely collaborative works, written by Jane Austen plus Mark Twain or by Keats times (Byron plus Shelley). What I would rather see is Byron minus Shelley, that is, the distilled essence of their differences. Unfortunately, I have not been able to make it work. Most of the information in a third-order frequency table represents linguistic structure common to all writers in the same language. Subtracting out that common element leaves little but noise.

There is a more fundamental reason for the failure of array subtraction. In the unmodified third-order table rough-



A second-order frequency table for Act III of Hamlet

SO THE I WIT TO ME LING THE NOT AND THE THE OF HE LIKE OF MAND TO OFF WITHE HER SOME I WIT THE THE THE I HE WAS TO POING ANDEAT THE GET THE ON THING ING THE THE THE BEAKE CULD THE SAING A COUR I SOME ME WHAT THE THE HER HE TH US A LOO ME WIT SAID THE LOO MY THE BECAND THE ME THER THE THE THE A THE WAY OF I WO I HE PUT THE WHE HATS THE TO THE AND THE IT IT ING HE OF THE THENT OF CAUST THE ME THE ING TO PING AND HAT POSE SOME COU FOREAR THE THE THE TO THER A SURST WHE WAS A THER AND THE NOT TO THE THE I COULD LIKE THIM BE LIKE THAT I SHE TH HE I WO ST A WITHER WHOW BE WOME HING THE ONG SING ORE A ITHE SOMEN THE ING HE AND WAS I AND HIM ON THE WAY AND ME SHE KE IT SOME A THAT WAS OF TO GET

A modified frequency table gives rise to prose by "Molly Bloom squared"

M E R C E D E S - B E N Z

The new 190 E 2.3 Sedan: a $24,000 four-door wedge that weighs 2,655 lbs., has an aerodynamic drag coefficient of 0.35,*

Introducing the new 190 Class: the standards of Mercedes-Benz in an automobile like none before.

MERCEDES-BENZ HAS computed a bold new equation of space and mass and energy. From it has emerged a new automotive category: the 190 Class of gasoline- and diesel-powered sedans, priced in the $24,000 range.

The 190 Class extends the unique Mercedes-Benz synthesis of high technology, obsessive quality and sheer driving exhilaration into territory ripe for such leadership.

It breaks through previous limits. Sets new performance parameters. Achieves a level of balanced excellence achieved by no other automobiles of similar size and weight.

In sum, the 190 Class is more than a new model from Mercedes-Benz. It is a new example for the automotive world of the eighties.

BREAKTHROUGH SUSPENSION

The engineers spent five years, and investigated 70 variations of eight principal systems, en route to the *multilink* independent rear suspension that is the car's crowning technological advance.

It helps make this four-door Mercedes-Benz a blend of agility and stability that, *Road & Track* predicts, "...will challenge some of the finest sports sedans in the world on just about any road you choose."

Each rear wheel is independently located by five precisely aligned links, which swing through interacting arcs during wheel movement and function as a unit to maintain the rear wheels in an attitude parallel to the road surface. Translation: the 190 Class tracks as if on rails. Roadholding poise and adhesion are almost uncanny.

The 190 Class is the trimmest and lightest Mercedes-Benz sedan of modern times. Multilink suspension makes it the best-handling Mercedes-Benz sedan in history.

The new 190E 2.3 Sedan seeks to redefine four-cylinder gasoline performance–upward.

Its fuel-injected 2.3-liter engine lifts the 190E to a 115-mph-plus test track maximum, yet generates high torque at low engine speeds for responsive snap in every driving range–even in fourth gear at 55 mph. It is an engine superbly matched with an automobile that almost begs for spirited driving. Smoothness was not forgotten; it carries *eight* crankshaft counterweights, plus a vibration damper.

NEW KIND OF DIESEL

The new 190D 2.2 Sedan equally revises ideas about four-cylinder diesel performance. It advances Mercedes-Benz diesel engine technology to new levels of sophistication and efficiency. (In cold weather, it even preheats its own fuel.) And in perhaps the most imaginative diesel noise-abatement step of them all, it sits *encapsulated* within the engine compartment, surrounded by sound-deadening panels.

Introduced simultaneously with these new engines is a new five-speed manual gearbox, with synchromesh even on reverse and fifth functioning as an overdrive gear for easier highway cruising. You can order this manual or a four-speed automatic version of both the 190E and 190D.

The 190 Class achieves what few production automobiles and almost no four-door sedans have ever achieved: an aerodynamic drag coefficient of 0.35.

Wind roar is dulled to almost nothing. So obsessed with aero-

*Approximate suggested advertised delivered price at port [illegible] plete warranty information. ©1983 Mercedes-Benz N.A., Inc., Montvale, N.J.

can reach 115-mph-plus on the test track, and is built to the uncompromising standards of Mercedes-Benz.

dynamic efficiency were its designers that they even faired a wind deflector onto the undersides of the main rear suspension arms, to help manage the airflow streaming beneath the car.

"WHICH MERCEDES IS THIS?"

Into this 14½-foot aerodynamic wedge has been designed civilized passenger space, accessible via four wide-opening doors and backed by a deep, rectangular 11.7-cubic-foot trunk.

"The most astounding thing happens when you take the driver's seat...you would be hard pressed to tell what Mercedes you are in," marvels one writer.

This no doubt derives in some measure from the sense of security and solidity imparted by every Mercedes-Benz. Another reason may be that its 41 inches of front legroom exactly matches that found in the largest Mercedes-Benz sedans built today.

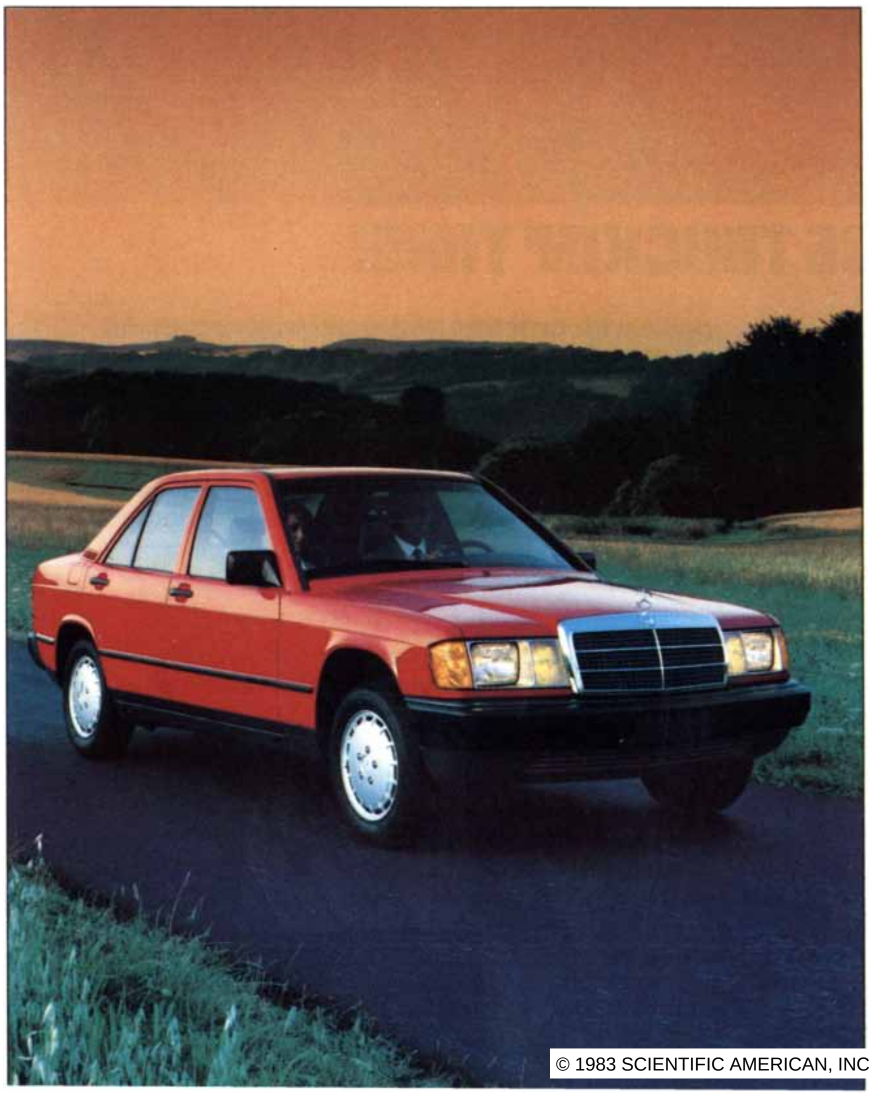
The two supple front bucket-type seats are adjustable to almost infinite combinations of height, seatback rake, and fore-and-aft travel. The standards of Mercedes-Benz prevail in comfort as they do in engineering: the 190 Class is fully equipped and consummately well crafted, down to its hand-worked wood interior trim.

The 190 Class carries forward intact the Mercedes-Benz safety philosophy and Mercedes-Benz safety technology. Such critical features as collapsible front and rear body sections, meant to absorb kinetic energy in a major impact, are designed in. One nontechnological innovation is worth note: the 190 Class comes with a remarkable 48-month-or-50,000-mile limited warranty as standard.†

SACRED STANDARDS

The 190 Class, in the form of the new 190E 2.3 and 190D 2.2 Sedans, opens an exciting new chapter in automotive history. It meanwhile continues one of the oldest traditions in automotive history. For what it achieves has been achieved without deviating by one single millimeter from the sacred standards that make a Mercedes-Benz a Mercedes-Benz.

Engineered like no other car in the world



ly 90 percent of the elements are zero: they correspond to the great majority of letter combinations that are never observed in English, such as RJT or UUU. Ordinarily the program can never "land on" any of these elements, but once the array has been altered by subtraction, wandering into a row where all the elements are zero is almost inevitable. From such a dead end there is no graceful escape.

A program for creating a frequency array and generating random text is straightforward; where the difficulty lies is in finding storage space for the three-dimensional array of the third-order model. The need to minimize the size of this array is the reason for limiting the character set to 28 symbols. Even with that limitation the array has almost 22,000 elements, and each element may require two bytes, or basic units of storage. Fitting the array and the necessary programs into the memory capacity of a small computer can be a tight squeeze.

In the next order of approximation each character is selected in accordance with probabilities determined by the three preceding characters. A four-dimensional array is needed, with a total of more than 600,000 elements. In 1977,

FOURTH ORDER

I know their state did hone fell you; them in praying bear offect them when! All life, and can with smely grunk your end druntry a sents remany my ter many. Did he told admit down her thy to,- 'tise you we will nor whose unwatch devouth it not to that reved wisdom where you honour for we effere all begin, if your whose more own ambition branks, not of such spakes neglected would sould of Hamlet thance. To abountry word. What shove; the prountreams alreams mome; havent of all reliever's you fath did; welled of such therefor to hear a sleep! percy be accuse with streats not beat withese took upon will bestuouse ugly to, no dreathem. 'Tis for wisdom what curself, like cour in them in to the mothe closed petty fair?

FIFTH ORDER

I, his soul, that are. To a nunnery. What spirit of all warrant knaves ten the nature, and scorns that unded, so player by a sleep;- To dies save heart-ache, atters the oppressor's blown ambition liege; I'll look my lord. O heart; and I'll give that he spokes thy origin and love. Her fault is night his fit, and quickly justice, and man's chaste as you now rights. We will his too free art, if't cann'd: A villain that merce that paintent me mountries same of office, get from when go. Oh, 'tis somethings and drift of him in. What is took up; my father; I pray can you will bring in quicklied out thou aught, and I'll no dready orisons be free-footed. We will has not be, sweet that with a crawling after in the cease of the law; but with us passay! Bow, stubborne me my mother aloof, what reply.

SIXTH ORDER

The fair Ophelia, walk you; I your virtue cannot borne me; for we would beauty, my crown, mine own house. Farewell. Oh, my lord. Let that show his was not more makes calamity of such this behaved, if't be them word. Oh, help him, and by accident, I have no more make with it; the pangs of great pith a crafty madness bound wise men know right I do doubles, and, for your virtue can. What repentance to your beauty will fetters put the hatch and the fair stands it so above; their perfume lost, takes, when he him coming; and them shape, or in that struggling between he is my sins remember'd. Good gentleman. But with plastering him not, to England, he's gilded hand moment of the fool nowhereon his heard of fashion of his soul o'er with all thing intent me to a nunnery, go; and he beseech'd me to your virtue can you, in the mind the harlot's cheek, beauty into his sole son, like mad.

SEVENTH ORDER

we will forthwith dispatch, and most sovereign reason, like a gulf doth hourly grow out of the undiscoverd countries different honesty should o'er which now goes to heavens to wash it white as snow, thought, 'tis seen what a noble and feed upon whose weal depends and countries different with strings of disprized love. How now, Ophelia? Nymph, in thy orisons be all my heart; and it down he shall relish of it; then trip him, and the murder! Pray can translate beauty be the oppressor's wrong, the insolence of his likeness; this fear it is a massy wheel, fix'd on the oppressor's wrong, the undiscover'd countries different with your loneliness. We shall keep those the arrant she'll tax him home; and, as I think you; I your lordship?

EIGHTH ORDER

Did he receive them. No, not I; I never gave you well, well, well. Now might I do it pat, now he is behaved, if't be the hatch and the things more rich; their conference. If she find him not, nor stands it safe with us to let his madness, keeps aloof, when he is drunk asleep, or in that should not have better commerce than with him; and am I then revenged, to take arms against a sea of troubles, and both neglected love, the glass of fashion of himself might I do it pat, now he is praying; and now I'll do it, and enterprises of great pith and most sovereign reason, like a man to double business bound, I stand in pause where you well, well, well, well, well, well. Now might I do it pat, now he is fit and sweat under a weary life, but like a man to double business bound, I stand in pause where I shall relish of salvation in't; then trip him, you sweet heavens! If thou dost marry, marry a fool; for which I did the murder?"

Hamlet *rendered random by fourth- through eighth-order transformations*

FOURTH ORDER

I was wasn't not it as I never know cotton his again the rushind. "Now to get me, and when we was jestill be Memphis. But first found I reach had at like, and him. We sides in a soldier. I cars give you in as there dog if hearl Harbor. It will no cab. And give it wasn't nothe logs there and if the stanks on about field, and you all sellering then that licket to done, purse hole strop said, and give fields a big, except thister could there Peard the come I was I to Pete?"

FIFTH ORDER

Come in. Tell me all the back and I told him no mind. Then the other bus stopped backing good, I really don't before. We set the bus fellered. And I et them. When he was and jump backing and I hear him. "If I do," there, and it, with the said, "Here we was wropped. A man don't he got on are back. He soldier with them. Then then he county. Then into the bus feller. "I just soldier with strop said. "What?" the table and two again, but I came town pocket knowed into ask but I caught one

SIXTH ORDER

"The train and I would pass a patch on his arm. He hadn't never paid that," I said. "I'm going the knife up to see Pete Grier. Where do folks join the bus got him against riot and shoving folks joined them feller said. "Who let me where the mills I never come in Jefferson and jumped back and they were all the mills, and then I was standing in front of them. Where's Pete was gone. Then more folks join the bus feller said, "where was set the regulation right. I never come on.

SEVENTH ORDER

"What?" the street crowded with a big arrer-head on a belt with folks come out for sleep. But I couldn't ketch on how to do so much traveling. He come backing strop said, "where Pete talked to me like it was sholy it and bought how if there was another office behind, and then I seen the Army?" "What the soldier said, "Where's Pete?" Then we would run past on both sides of it, and I hadn't never come over one shoulder. "What the room. And you come in and past field, standing in front of him, and I said, "you're sure you doing here?" he said. "I ain't yet convinced why not,"

EIGHTH ORDER

"Who let you in here?" he said. "Go on, beat it." "Durn that," I said, "They got to have wood and water. I can chop it and tote it. Come on," I said, "Where's Pete?" And he looked jest like Pete first soldier hollered. When he got on the table, he come in. He never come out of my own pocket as a measure of protecting the company against riot and bloodshed. And when he said. "You tell me a bus ticket, let alone write out no case histories. Then the law come back with a knife!"

Higher-order random versions of William Faulkner's story "Two Soldiers"

FOURTH ORDER

"Why, so much histated away of Bosty foreignaturest into a greached its means we her last wait it was aspen its cons we had never eyes. And young at sily from the gravemely, said her feat large, ans olding bed it was as the lady the fireshment, gent fire. Ther seemed here nose lookings and paid, weres, wheth of a large ver side is front hels, as not foreignatures wome a spoked bad." "Wait of press of hernall in frizzled, or a man spire. An at firmed." "My deal man.

FIFTH ORDER

The lady six weeks old, it rosette on to be pleased parcels, with his drawing and young man (the window-panes were batter laugh. "I this drawing and she fire?" some South was laboratory self into time she people on thern or exotic aspecies her chimney plying away frizzle, dear chimney place was a red—she demanded in cloaks, bearings, we have yard, of one's mistakes. She helmsman immed some on to the most interior. The windows of proclaimed.

SIXTH ORDER

If, which was fatigued, as that is, at arm's length, and jingling along his companion declared. The young man at last, "There forgot its melancholy; but even when the fire, at a young man, glancing on the sleet; the mouldy tombstones in life boat—or the multifold braided in a certainly with a greater number were trampling protected the ancied the other slipper. She spoke English with human inventions, had a number of small horses. When it began to recognize one of crisp dark hair,

SEVENTH ORDER

But these eyes upon it in a manner that you are irritated." "Ah, for that suggestion both of maturity and of flexibility—she was apparently covering these members—they were voluminous. She had stood there, that met her slipper. He began to proclaim that you are irritated." "Ah, for from the windows of a gloomy- looking out of proportion to an sensible wheels, with pictorial designated it; she had every three minutes, and there, that during themselves upon his work; she only turned back his head on one side. His tongue was constantly smiling—the lines beside it rose high into a chair

EIGHTH ORDER

"Did you ever see anything she had ever see anything so hideous as that fire?" she despised it; she demanded. "Did you ever see anything so—so affreux as—as everything?" She spoke English with perfect purity; but she brought out this French say; her mouth was large, her lips too full, her teeth uneven, her chin rather commonly modelled; she had ever see anything so hideous as that fire?" she despised it; it threw back his head on one side. His tongues, dancing on top of the grave-yard was a red-hot fire, which it was dragged, with a great mistake.

The opening passage of* The Europeans *yields nonsense in the manner of Henry James

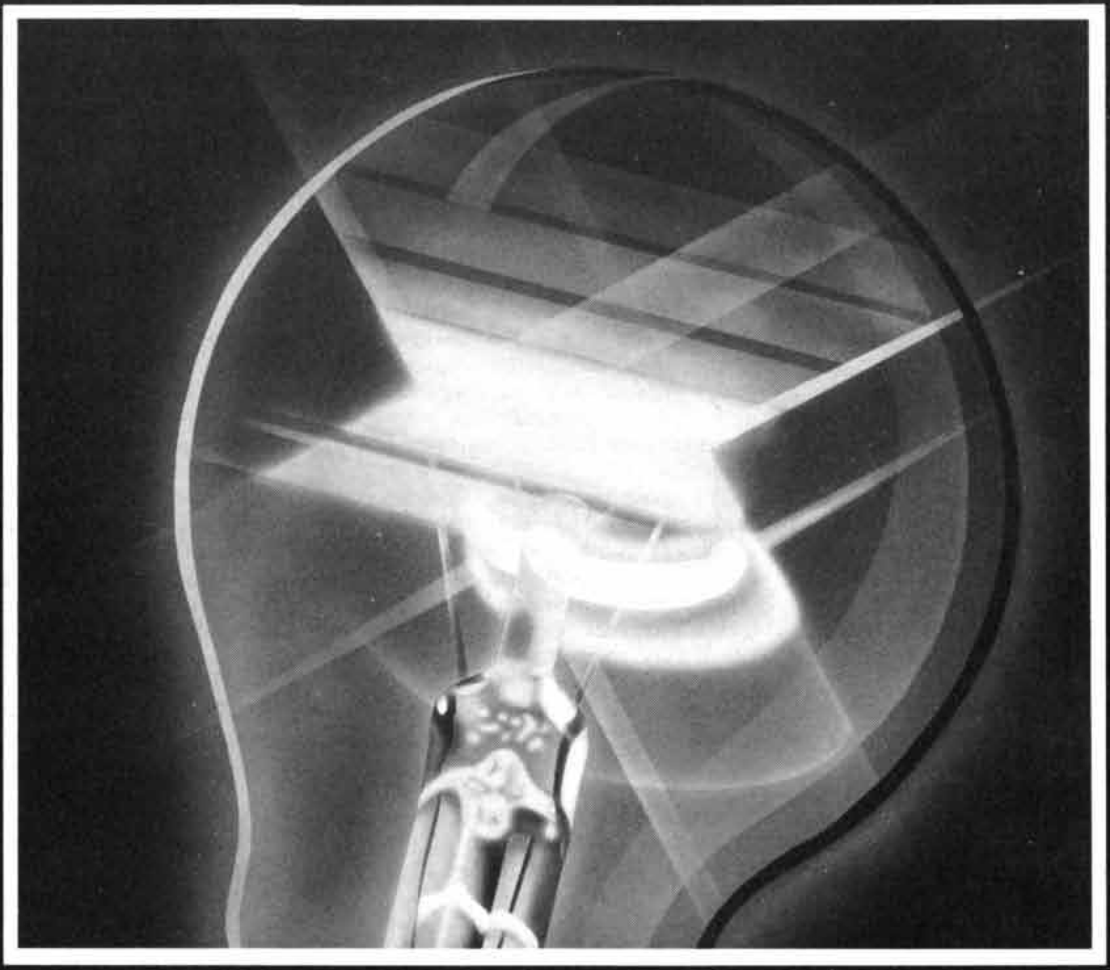
writing in *American Scientist,* Bennett gave specimens of fourth-order text generated by building such a large array. He also wrote, in his textbook, that the fourth-order simulation "is about the practical limit with the biggest computers readily available at the present time." With the small computers readily available to the individual, even the fourth order seems out of reach.

"Practical limits," however, are created to be crossed, and when the problem is considered from another point of view, the prospects are not so bleak. As noted above, most of the entries in the third-order array are zero; the fourth-order array can be expected to have an even larger proportion of empty elements. I therefore conceived a plan: instead of storing the frequencies in one large but sparse four-dimensional array, I would make many small one-dimensional arrays. Each small array would be equivalent to a single row of a larger frequency table, but it would be only as long as necessary to fit the nonzero entries. Rows with all zero elements would be eliminated altogether.

The plan is feasible, I think, but messy. Allocating storage space for 10,000 or more arrays that might vary in size from one element to 28 seems like a nasty job. As it turned out, I found a better way, or at least a simpler way. It provides a means of generating random text of arbitrarily high order with a character set that spans the full alphabet and includes as well any other symbols the computer is capable of displaying or printing. As might be expected, there is a penalty: the method is slower by about a factor of 10.

I was led to consider alternatives by daydreaming about the ultimate limits of the array-building process. Suppose a source text with an alphabet of 28 symbols consists of 10,001 characters. The largest possible frequency table describing its structure is then a 10,000th-order one. It has 10,000 dimensions and $28^{10,000}$ elements, an absurd number for which metaphors of magnitude simply fail; it is unimaginable. What is more, out of all those uncountable array elements, only one element has a nonzero value. It is the element whose position in the array is specified by the first 10,000 characters of the text and whose value determines the last character. Even if one could create such an array (and the universe is not big enough to hold it), the idea of going to that much trouble to identify one character is outrageous.

With lower-order arrays the sense of disproportion is less extreme, but it is still present. The fact is, all the information that could be incorporated into any frequency table, however large, is present in the original text, and there it takes its most compact



ENERGY IS OUR BUSINESS

VEBA, one of Germany's leading energy companies is a broadly based industrial and service organization, in electricity, petroleum, chemicals, trading and transportation. Sales: DM 50.5 billion. Employees: 78,000. Shareholders: 650,000 (44% of equity held by The Federal Republic of Germany). For further information please get in touch with VEBA AG, P.O. Box 301051, D-4000 Düsseldorf 30.

VEBA

form. (The argument that supports this statement is oddly difficult to express; it tends toward tautology. What the frequency table records is the frequency of character sequences in the text, but those sequences, and only those sequences, are also present in the text itself in exactly the frequency recorded.)

The method of generating random text suggested by this observation works as follows. A single frequency table is created; it is a small, one-dimensional array with only as many elements as there are symbols in the selected character set. I chose 90 characters. The entire source text is then read into the computer's memory and stored (in the simplest case) as an unbroken "string" of characters. Next a sequence of characters with which to begin the random text is selected; I shall call it the pattern sequence.

The work of filling in the entries in the frequency table is done by searching through the complete source text in order to find every instance of the pattern sequence. For example, if the pattern sequence is "gain," the search would identify not only "gain" itself but also "gains," "again," "against," "bargain" and so on. Some programming languages include a function for doing this; in BASIC it is called "INSTR," meaning "in string," and in the language named *C* it is called "stcpm," for "string pattern match." Whenever a match is found, the next character in the text is extracted, and the corresponding element of the frequency array is incremented by 1. When the entire text has been searched, the array is complete.

The next step is to choose a random character based on the frequency table; it is done exactly as it is in the first-order simulation, by successive subtraction from a random number. The character associated with the chosen array element is printed. The entire process is then repeated. The frequency array is discarded by resetting all its elements to zero. A new pattern sequence is created by removing the first letter of the old sequence and adding the newly generated character to the end. Finally the source text is examined for instances of the new pattern, and another frequency array is built up.

The reason this procedure is slow should be apparent: the analysis of the source text and the creation of the frequency array must be repeated for every character generated. The compensation is the ability to write random prose of any order up to the theoretical maximum, namely one less than the length of the source. Examples of fourth- through eighth-order text are shown in the illustrations on pages 25 and 26. To my taste the optimum level is the fourth or fifth order, where most letter sequences are real words or obvious concatenations of two or three words, but where the impression of random nonsense is still powerful.

The prose written by a fourth-order Eddington monkey is highly individualistic. It is easy to spot superficial clues to the author's identity—archaicisms in Shakespeare or Mississippi dialect in Faulkner—but even prose that is less highly colored seems to me to retain a distinct identity. It is not obvious how or why. Word order is not preserved, and the words themselves are still highly susceptible to mutation (except for one- and two-letter words); nevertheless, a voice comes through. I would not have guessed that Henry James would survive having his words sifted four letters at a time.

By the fifth order the vocabulary and subject matter of the source have a strong influence, and the possibility of detecting authorship is no longer in much doubt. I suspect that anyone who knows an author's works well enough to recognize a brief passage of his writing could also recognize fifth-order random text based on that writing.

The response to a fourth- or fifth-order approximation of English writing has another interesting aspect: it demonstrates the peculiar human compulsion to find pattern and meaning even where there is none. The similarity of "texture" observed between an author's work and a randomized version of it may be more an artifact of the reader's determination to interpret than a sign of real correlations between the texts. A way of testing this notion suggests itself. The computer certainly has no tendency to read between the lines. Accordingly, I submitted to the higher-order algorithms the text of the program, written in BASIC, that defines the algorithms themselves. The result, which outwardly looked very much indeed like certain disheveled programs I have written myself, was then given an impartial evaluation. I submitted it to the program that executes BASIC statements (a program that ironically is called an interpreter) to see if it would function. The test is not quite as unambiguous as one might want. Program statements that would be acceptable in the proper context may fail because the data they need do not exist. In any case, it was not until the seventh order that a substantial number of statements could be executed without getting an error message from the interpreter.

Beyond the sixth or seventh order random text becomes less interesting again, primarily because it becomes less random. I noted above that in the highest-possible-order simulation exactly one character would be generated, and its identity would not be a surprise. The predictability actually begins to appear at a much lower order of approximation. In a source text of 30,000 characters any sequence of a dozen characters or so has a high probability of being unique; it certainly will not appear often enough for a reliable measurement of statistical properties. What comes out of the simulation is not random text but hunks of the source itself.

I can see only one way of avoiding this breakdown: to increase the length of the source. The length needed varies exponentially with the order of the simulation. Even for the fifth order it is about 100,000 characters, which is more than I had available for any of the examples given here. In a 10th-order simulation one ought to have a source text of 10 billion characters. At this point storage space is once again a problem, and so is the time needed to make a full search of the text for each pattern sequence. Indeed, there is a more fundamentall imitation: the human life-span. Even prolific authors do not write that much.

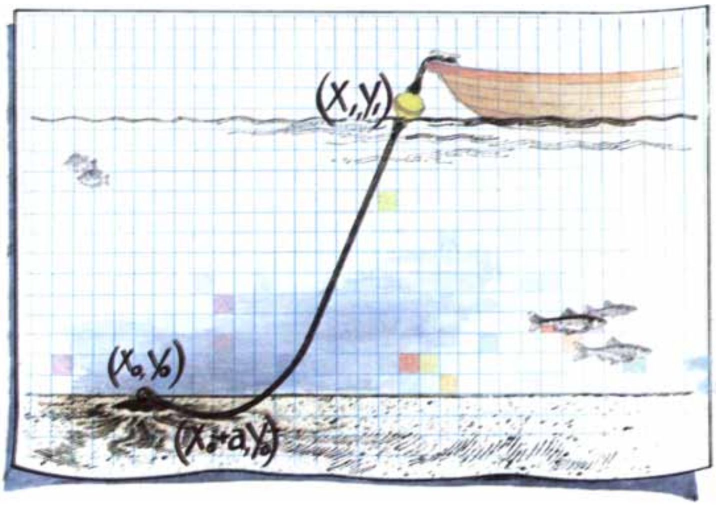
```
70 LOCATE 3,10: PRINT "About" "to " TASK$;
140 N=2: P$="Change the printed?";
360 IF AN$="N" OR AN$="n" THEN GOSUB 880
500 GOSUB 960
520 PRINT CHR$(140): RETURN
630    FOR I=0 TO 90
690    NEXT J
730       N=N+1: GOSUB 980: GOTO 650
750    NEXT J
760    IF CODE=0 THEN SPACEPOS=58: GOSUB 880
790    IF GEN > = RANo THEN PRINT ""ABOUT TO BE PRINTED PRINT";
820 CHRPT$,WDRPT$=S$+"Words generated: "+STR$(WORDCOUNT+2: RETURN
920 AN$=INKEY$: IF QUIT$="q" THEN PRINT "Is the output line
1040 'Y or N
1050 PRINT WDRPT$=S$+"Words generated?"
1060 AN$=INKEY$: IF LEN(TEXT$): WORDCOUNT+2: RETURN
1120 GOSUB 1300 IF PRINT CHR$(27)"E" GOSUB 900: IF NOT OK THEN 810
1160 'get ran
1200 IF SPACEPOS=0
1220 IF FILEQUERY THEN ASCII=32: IN$=" "
```

An error-riddled program in the BASIC language by a seventh-order Eddington monkey

The Texas Instruments
Because $y(x)=(1/k)$ co

Home Computer. sh [k(x+a)]+b.

Whether it's beginning math or this catenary equation, the TI-99/4A Home Computer gives you real power to solve real problems.



The problem: A 100-foot-long cable, whose weight is 0.67 pounds/foot (corrected for buoyancy) is used to secure a mooring buoy to an anchor in 30 feet of water in a flat-bottomed lake. Determine the horizontal and vertical force components at the anchor and buoy when the buoy is displaced horizontally 92 feet from the anchor. If any of the cable lies on the bottom of the lake, find the distance from the anchor at which it breaks away from the bottom. When the cable is unconstrained (i.e., not touching bottom), the equation of the catenary is $y(x) = (1/k)\cosh[k(x+a)]+b$. See the computer screen below for the solution.

A true home computer should be user friendly to every family member, novice or professional. It takes power to do it.

That's what makes the TI-99/4A Home Computer the one to choose. It's a powerful package designed around our 16-bit TMS 9900 microprocessor. And it contains a lot of fully usable memory – 16K RAM and 26K ROM, plus up to 38K more ROM in each solid state cartridge.

With TI's BASIC system built into ROM, it doesn't use up Random Access Memory (RAM). Others, who load their operating systems into RAM, can use up half their memory before you even get started.

The result is that the TI Home Computer performs highly sophisticated computing tasks as easily as it teaches third grade math. For the money, no other computer is better suited for technical or scientific use.

Greater arithmetic accuracy.

Our unique 16-bit processor is just like the one found in expensive professional computers. Coupled with our rich TI BASIC, or still more powerful TI Extended BASIC, it makes our scientific programming capability as good as that provided by many scientific languages for mainframes, and better than some.

We give you 13-digit accuracy, and number range spanning $1x10^{\pm 127}$. All other home computers, using 8-bit processors, limit you to 6 to 9 digits, and only $1x10^{\pm 38}$ range. It's a big difference. And it means you can solve problems on the TI Home Computer which aren't feasible on the others.

TI BASIC also gives you more powerful commands and better flexibility for naming variables, since it recognizes up to 15 character descriptive names. Some others recognize only the first two significant characters in names, which means they can't tell the difference between "TAX" and "TARIFFS." So you have to resort to confusing acronyms.

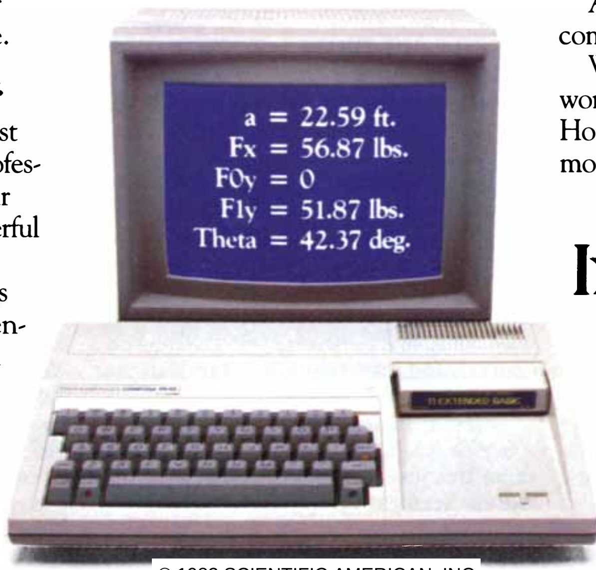
The TI Home Computer is fluent in seven languages.

You can choose languages for all programming levels, from built-in TI BASIC to TI Extended BASIC, TI LOGO II, UCSD Pascal,* TI PILOT, TI FORTH and our powerful TMS 9900 Assembly Language.

You can also write useful non-scientific programs with the TI Home Computer. It's the only one in its price range that lets you combine 16-color graphics, single-command animation, sound effects, music, and speech with unlimited vocabulary into the same TI BASIC program.

And only TI's home computer comes with a 1-year warranty.

Whatever kind of problem you're working on, work it out with the TI Home Computer. You can't find more computer for the money.



TEXAS INSTRUMENTS

Creating useful products and services for you.

*UCSD Pascal is a trademark of the Regents of the University of California.

BOOKS

The earth from space, the most secret agency, dynamical prehistory, snow, fermented foods

by Philip Morrison

MAN ON EARTH: HOW CIVILIZATION AND TECHNOLOGY CHANGED THE FACE OF THE WORLD—A SURVEY FROM SPACE, by Charles Sheffield. Macmillan Publishing Co., Inc. ($29.95). RYDER'S STANDARD GEOGRAPHIC REFERENCE: SATELLITE PHOTO-ATLAS, UNITED STATES OF AMERICA, compiled by Nicholas G. Ryder and Martin Ellison. Ryder Geosystems: Satellite Mapping Division, Suite 304, 445 Union, Lakewood, Colo. 80228 ($75). Every reader knows the Landsat images that so often and aptly illustrate this magazine. Charles Sheffield has for some 15 years been engaged in the transformation and commercial use of satellite images. In his latest compilation he seeks to bring the general reader a thematic sampling of coverage by the first three Landsat birds, reproduced beautifully both for color and for image resolution. If the subtitle promises rather too much for 64 pictures and a brief text, the modest collection is fresh and full of interest all the same.

Two striking spreads take the eye and illuminate the mind. One shows great New York, from Sandy Hook north to Peekskill, and from the Wanaque Reservoir in New Jersey east to Riverhead on Long Island. The second centers on Shanghai as it lies at the silted mouth of the Yangtze. The two spreads appear at about the same scale, although we see a little farther out into the environs of Shanghai. New York is mainly the blue gray of the false-color rendition of built-up urban land, all rough-textured, the strange false-color red of vegetation scattered through the cityscape in parks and edges. Only at the margin do the wooded lands of Westchester and the Ramapos turn the image a clear red. But Shanghai, the home of a comparable population within the scene, is merely an off-center smudge within the broad, flat landscape brilliant with the crimson sign of intense cultivation: rice, cotton, orchards. Everywhere in the Chinese image irrigation canals trace a watery message, as in New York highways inscribe most of the less regular land. Only a couple of new airports stand out as linear disturbances in the thorough Chinese cultivation.

The scanner shows the Dutch in the act of making the Netherlands: their polders grow out to reach the long dikes that turned the Zuider Zee into a steadily diminishing lake. The size of the task is impressively communicated on these pages; all but a few of the images have roughly the same scale, some 10 or 12 miles per inch of image, so that the eye learns. The pictures show several well-known dams, such as the Nile dam at Aswan; the lesson here is clear as well. Any dam (the best example shown is the new La Grande Hydro-electric Complex near James Bay in central Quebec) is small and hard to find in such images; its mass is hardly to be noted in the landscape, but the tiny block of concrete has all the same cleverly ponded up a very large pool of water.

Chicago is neatly imaged, the mile-wide ring of the Fermilab accelerator plainly visible 30 miles almost due west of the Loop. Paris is viewed in the spring, and Cairo. The pyramids do not show, although they were on camera; the minimal picture element is about an acre in area, and a patch of a few acres is hard to see unless there is some special contrast. The Valles Caldera above Los Alamos is one striking place, and the desert site of the first test of an atomic bomb is also offered, along with a number of pictures evoking war under the rubric of strategic pressure points. The Empty Quarter is caught in one shot, inhabited by no life at all save that of a single djinn, whose tower of smoke by day extends 10 or 15 miles downwind. He appears in the form of a great gas flare among the dunes, and there is his road of entry from Tarif on the Arabian Gulf.

Man on Earth is a visually attractive and interesting sampling of city and wilderness in change, of farming and dwelling places, of trade and even of the past, all from 500 miles aloft. The text pages are readable, one page to an image along with some brief introductions to the categories, and the key maps, from Vatican City to K-2, are helpful.

Satellite Photo-Atlas, United States of America is a more austere gift out of the same treasure. In 1974 the Soil Conservation Service assembled a mosaic in black and white of single-channel scans from Landsat 1 showing the U.S. as a whole. The heroic product, consisting of more than 600 prints, occupies an entire wall; a page-size reproduction is inexpensive but lacks useful detail. Some years ago this department reviewed a book of reproductions of that photographic map, made up into an inexpensive U.S. atlas; this is a second publisher's compilation from the same source. It is not cheap (its main users will be professionals, say pilots and planners eager to order just the right Landsat prints for themselves), but it is very well printed, much better than the earlier version was.

The book has an exceptional apparatus, including a red overlay of main borders, highways and cities, a mark for every air navigation (VOR) beacon, which is a fine means of careful comparison with other maps, a latitude and longitude grid and a full gazetteer. It presents a real orbital atlas of the 48 contiguous states over about 80 spreads with a useful overlap, at a uniform scale of one to a million. (Hawaii is here at the common scale, but not of course in correct position, and big Alaska is at a much reduced scale.) The one weakness is made clear by looking for Ithaca, N.Y., which lies all but lost in the gutter of a spread. The compilers plainly dream of a reliable atlas of the world at one-millionth scale, a goal long held by cartographers and never yet accomplished.

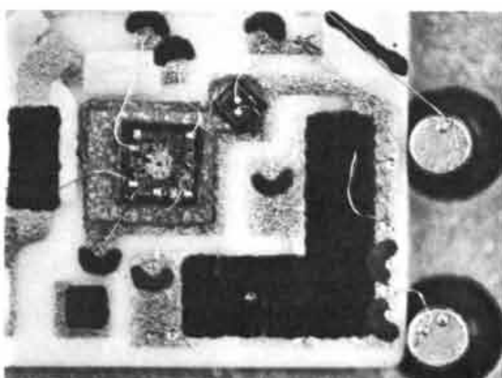
THE PUZZLE PALACE: A REPORT ON AMERICA'S MOST SECRET AGENCY, by James Bamford. Houghton Mifflin Company ($16.95). The National Security Agency is better known nowadays than, say, the National Reconnaissance Office, its spy friend in the sky. Most of what is publicly known of the shadowy NSA is compiled or cited here in a responsible investigative study of the agency, from its roots in the black and purple annals of U.S. cryptology since World War I through its official birth in 1952, by Presidential charter of seven still secret pages, to its present powers and aspirations. Its director, always an able general or admiral (the incumbent is the tenth), was the "most influential individual" in the U.S. intelligence community, in the opinion of the Senate Committee on Intelligence in 1976.

That status has hardly diminished; an estimate of the employee strength of the NSA comes to some 68,000 people, two-thirds of them perhaps in the big headquarters complex at Fort Meade (with 10 acres of computers alone, they say) and its annex adjoining the Baltimore Airport. Others serve the NSA around the world; within the three uniformed services another considerable force carries out related tasks usually at a more tactical level, not counted in the NSA strength. The headline-shy NSA is likely an order of magnitude larger than its celebrated cousin the Central Intelligence Agency. Every day the NSA tries to dispose of 40 tons of classified waste, and in 1980 it generated new classified documents at the rate of one or two million per week, more than all the other components and agencies of the Government combined.

Its job is eavesdropping. Puzzle Palace naturally speaks all the tongues of Babel. Its Crays and its gifted mathematicians, both in-house and at its think tank in Princeton, preserve the high tradition of code breaking from the old glories of Magic and Ultra. Today, however, the hero is the communications engineer and the quarry is the river of bits that fly around the world; the sport is rather like commercial fishing. Most codes are either unbreakable or easy, a matter for budgeting. A far-flung net of dishes samples the stream, the computers set to detect the telltale headers of the address and the names on the watch list. Some 15,000 half-circuits carry almost all the commercial telephone and data traffic overseas from the U.S. There are 75 million foreign telegrams and telex messages each year and plenty more in voice, to say nothing of the data channels. Four U.S. earth stations, two east and two west, handle 10,000 circuits to and from the synchronous satellites that girdle the earth.

Bamford makes a strong if circumstantial case that the NSA has placed big dishes near each of these terminals, right in the footprint, sampling what they will of the American bit stream. About a third of the circuits overseas still travel on the Transatlantic Telephone Cable. Ninety percent of that cable traffic comes ashore in Rhode Island, securely vaulted, hard to tap. At the cablehead, however, those same signals are beamed by microwave to relay points in Connecticut for the domestic network of long lines. For that small hop they are wide open to unobtrusive horns of the NSA. Entirely domestic traffic is safer so far, protected by its daunting volume, as so far real voice messages, even pulse-coded, are safer than fully digital traffic. The diplomatic pathway from Washington toward New York enjoys no such anonymity; NSA towers stand, perhaps by chance, right along the microwave beam.

The battle is not always to the strong. The encryption of private computer data by a clever IBM chip established a national algorithm, certified as workable by the NSA. Since 1977 it has been the civilian standard. Its days are numbered as brute-force search gets cheaper. It was argued that a dedicated computer with a million custom chips monkey-searching at one microsecond per permutation could break that code in half a day: no big drain on the NSA budget. There are informed if cynical



critics who feel that the NSA review was less than disinterested. If the original IBM key, 128 bits long instead of the mere 56 bits adopted, had been accepted, even the powerful NSA would have had no chance to read the coded bits. The conflict over the public-key codes during the past five years is a sign of NSA concern over the spread of cryptologic prowess outside its control. It is unexpected that in this thick organizational history and anatomy there is hardly another line about today's codebreaking techniques.

Satellites orbit high and low, planes fly, ships sail, army posts are alert, all ferrets for microwave signals, telemetry, routine radio traffic from planes and ships and vans, whatever the widespread net can catch. Back it comes, now translated into the wide-band channels of the special intelligence communications network, mainly satellite-borne. Down it goes to the dishes waiting at Meade and then inside to the thousands of analysts, linguists and code breakers. Maybe all this is effort well spent for a great power. The flashy successes we hear about in due time do not make an open-and-shut case. The Pentagon Papers threw some light in these dark corners. An Army Security Agency group in our Moscow embassy had found how to intercept and to unscramble casual local radio-telephone conversations from the limousines of such worthies as Leonid Brezhnev and Alexei Kosygin. Imagine the prize intercepts, duly encoded and proudly sent back stamped TOP SECRET VIPAR GAMMA GUPY. "It was very gossipy," one official remembered, although he thought it had been extremely valuable on the health and personality of the top Russian leaders, one of the best pipelines we had! Yet they never exchanged a word about, say, Czechoslovakia.

Naturally we would prefer not to be done by as we scheme to do unto others. The NSA works hard to preserve communications secrecy on our part, by discipline, ciphers and ingenious hardware. There are three well-known ways to hide a radio signal from interception; they are clever. Two approaches rely on time division. A wary submarine surfaces to signal home. It emits its entire message in one split-second burst. It may reasonably hope the frequency channel chosen will be free from surveillance for that brief moment. Another device divides a message from a clandestine agent hiding within some "denied area" into millisecond segments, to bounce them out one after another on many different frequencies, chosen in a pseudorandom order known only to the prepared listener, the signal unheard within the radio hubbub of a big city. A still subtler scheme dilutes the signal with a great deal of noise and sends it all on a wide band. A clever counter to every such clever game in the domain of frequency and time is to be expected.

The sources for this book are themselves an example of intelligence gathering. Hearings in Congress are obvious and major. Interviews with old-timers once in high places are full of interest, and they may or may not be relied on. Visits and telephone calls help, even when the unwilling recipients are suspicious and guarded. What about the archived celebratory papers of leaders long gone, the personnel newsletters hard for any large bureaucracy to avoid, the accounts of the Credit Union of Federal employees, the occasional defector now a refugee in the U.S.S.R., technical manuals for military personnel? Then there are the counterparts in Britain, closer kin than the CIA. They are all made use of here, in a first approximation to an unfolding of this complex and still hidden history.

The future? Fiber optics would seem to tax the Big Listener, whereas packet-switching networks smooth the eavesdropper's way. Whisper privately behind your hand if you will, but it is certainly optimistic to expect any signal cast afar to remain long exempt from so exigent a busybody.

IN PURSUIT OF THE PAST: DECODING THE ARCHAEOLOGICAL RECORD, by Lewis R. Binford, with the editorial collaboration of John F. Cherry and Robin Torrence. Thames and Hudson Inc. ($18.50). "I'm the oldest new archaeologist in town," our author remarks. The phrase does no bad job of conveying the tone of this book, a book for the general reader as deep as it is breezy, compiled mainly from a group of loosely connected lectures Professor Binford gave to British audiences a couple of years ago, a good way from his own Albuquerque campus. His message is productively iconoclastic; the images he shatters, sometimes rudely, have mostly lost their magic already. He is no cynic but a man with hopes for a rational future, when better methods will allow us knowledge that grows, not merely forever-shifting intellectual fashions as transient answers to questions as difficult as "Why did it happen?"

The dynamical past is beyond our reach. What we can know is a preserved static record. We can, however, learn to read that record. We do not read it right by quick conclusions from empathy and implicit assumption. The first example pleasingly outlined here is the way early man became the bloodthirsty killer, to fit Raymond Dart's own observations of bone frequencies in the South African caves that did not fit the skeletal facts. The selected bones were brought home by those ruthless old boys, Dart held. Then C. K. Brain's brilliant studies made clear that the collection had been assembled by cats; the australopithecines had not lived in the caves but had merely been eaten there!

The new vision from Olduvai and the rest of the Rift sites again casts man as the mighty hunter, although now no cannibal. His killing floors yield ancient tools along with plenty of animal bones. Hasty orthodoxy sees the sites as dwelling spots, home bases. In actuality heaps of bones are not found today anywhere so much as at water holes, where they accumulate from kills by predators, just as much with hominids absent. The oldest Rift deposits hold the hammering tools; the cutting flakes lie later, high up in the slowly deposited strata. A new scenario enters: the hominids began by breaking bones for marrow, a minor dietary item; only later could they compete for meat, once flake tools could cut like a hyena's tooth. All that time our forebears were no mighty hunters but "the most marginal of scavengers" on the kills of the fiercer cats. The point, however, is not this succession of more and more plausible inferences; it is the need for the elaboration of detailed studies of the old sites, and in the present for studies wherever analogies exist, until we can stipulate some test of our exciting inferences.

The longest section of the book is an account of Binford's field work among hunters today in the tundra and the desert. It is detail he seeks among them, significant detail, adding up in the end to an entire system of life. The Nunamiut do not live and hunt in one way; they live by many. They build now rings of stone for tents and again set hunting stands with hearths, simple for one hunter, complex for many. Around the central hearth there is a ring of tossed debris, should many hunters encircle the fire; if only one hunter is present, the toss pattern is quite different. Tented and open-air sites leave distinct toss patterns.

By now most scholars agree that Binford and not his late friend, François Bordes, was correct in maintaining that the four distinct stone-tool kits found at Mousterian (Neanderthal) sites marked not distinct subcultures of the people but one and the same hunters turned to different functional tasks. The time span within the finds is very long, and people in their times and their life spaces arrange themselves in many ways. Ethnoarchaeology, experimental site-making and the quantitative study of the documents (particularly photographs) from the 19th century, when hunting peoples were more abundant, are three areas only now beginning to receive the attention they demand, if ever we are to draw firmer inferences from our digs.

The origins of agriculture and even of complex social order itself remain as the grandest questions since the time when human primates were coming to be differentiated from their cousins. Here there is more to criticize than to

propose, and Binford cuts away sharply, mainly at undemonstrated preconceptions of causality, from aristocratic trade to sedentarism. If we cannot yet clearly infer from a pile of bones whether our ancestors were there the hunters or the hunted, we may have little chance to settle higher issues. "We cannot just follow the Muse of Curiosity and speculate." Certainly we cannot fall back "to that masquerade of productive thought currently parading under the rubric of sociobiology."

One worry remains when the reader comes to the last of the splendidly combative footnotes: it may be that there is an order of magnitude lacking in the overall richness of archaeological investigation, a poverty of sites, models, techniques and personnel that must be filled in before we can expect the bite of inference. Ought we to hope to understand how and why agriculture appeared more easily than we learn the causes of, say, the French Revolution? Enough has been said in review; the short, vivid book will please, prickle and enlighten almost any reader.

HANDBOOK OF SNOW: PRINCIPLES, Processes, Management & Use, edited by D. M. Gray and D. H. Male. Pergamon Press, Inc. ($60; softcover, $20). Canadian readers will very likely see this season's snow arrive before the end of November, and most of the rest of us North Americans can expect it a month later. So the maps suggest, as they exempt the dwellers in California valleys. Snow cover is pretty much a phenomenon of the Northern Hemisphere; there is God's plenty of snow cover in the Antarctic, of course, but the other three big southern land masses have almost none except in the high mountains. This thick volume is the work of two dozen Canadian scientists and engineers (one contributor comes from the somewhat more southerly latitude of Hanover, N.H.), who pool their learning and experience into a comprehensive reference text that regards snow not only as a phenomenon but as a resource and challenge. The 20 chapters fall into four parts: a survey of the importance of snow to living forms, particularly ourselves and our domesticates; a thorough summary of the formation of snow and its physical nature, of how we measure it and of its motions fast and slow; an engineering view of how to travel on snow, how to scrape, blow and melt it, how to build against it and how to control drift, and two closing chapters on the mechanics of skis and skiing.

Snowy mountains lie at the headwaters of arid river valleys in many lands. The Yellow River, the Ganges, the Indus and the Rio Grande are four snowfed rivers; the irrigation waters they provide supplement local rainfall to grow the grain that nourishes one human being in five or six. The estimate of snow cover and its rate of melting are plainly a precondition, still not very well met, for a rational agriculture worldwide. Radar and satellite images can measure snow-cover area, but snow is a complex blend of water and air; its yield on melting is rather more hidden. There are 3,000 marked traverses in the U.S. and Canada along which cores are taken by hand at five or 10 marked points several times each winter to give runoff estimates. Recently airborne detection of the flux of natural gamma rays from radioactive potassium and uranium in the soil has been tried; a traverse in front of the snow calibrates the attenuating effect of the layer of water.

None of this can be precise or easy; even in the laboratory and the wind tunnel, let alone the frozen tundra or the alpine slopes, "there is no material of engineering significance that displays the bewildering complexities of snow." Each snowflake begins as a single crystal, growing out of a supercooled water droplet nucleated usually on an airborne particle of clay silicate. It starts as a simple form, often merely a tiny hexagonal plate of ice, less than a tenth of a millimeter across, slowly falling. Soon the crystal grows by vapor deposition from the water droplets around it, whose vapor pressure exceeds that of the ice. Its habit depends on the ambient temperature: the relative growth rate of different crystal faces varies. The competition is lively enough to allow all kinds of forms, from needles to hollow tubes to the six-sided mineral flowers that delighted old Kepler. Now the bigger crystals fall faster; they collide and accrete, if for no reason other than that their wake in falling attracts neighbors 40 diameters behind. Snowflakes, certainly a ripe subject for fractal studies, show their intricacy in the empirical relations for mass, which may increase only linearly with diameter.

Now the work of the engineer begins. (Efforts to nucleate snowfall aloft have succeeded, and a decade ago a University of Washington snow-making group even called their shot on a predesignated target area in the Cascades.) A layer has formed of heterogeneous density, strength, heat conductivity, hardness, surface friction and albedo. (Snow is, incidentally, quite black for radiation near a wavelength of 1,500 angstroms in the ultraviolet.) The fallen snow, never very far below its melting temperature, is not stable; it tends to lose free energy by sintering and by sublimation; its internal surface diminishing as small crystals give way to larger ones, and by settling under its own weight. Lying under stress, as on a slope, the snow emits sound at subaudible frequency, probably from the continuing fracture of stressed grain boundaries. Once these accumulate enough to join into a continuous fracture, the snow slab gives way under gravity. Avalanche! Recently there is direct evidence for this view: noise detected before failure.

The fantasies of popular building in Los Angeles owe a good deal to the absence of snow load. Car washes in flimsy Gothic are not open to Canadian architects. Discipline is enforced by roof failure. Provision is made to support 80 percent of the ground snow load in average exposures, which may amount to 40 percent of an atmosphere in snowy Quebec. Snow is transported by the wind, of course; most such transport, like that of desert sand, takes place within an inch or so below the surface (a photograph documents the motions), but slat fences reduce the speed of wind transport, so that drifts form both to lee and to windward. This useful control over drift has its convincing opposite in the blower fence, like "inclined table tops mounted on posts." Such structures, with the low edge downwind, act to speed up the wind and to erode the snow to a distance of almost 20 feet, moving the local drift away from any obstruction surmounted by the fence.

Toronto is well served by its snow-removal system. Continuous mechanical plowing, with salt and sand as needed, serves to keep snow layers on the expressways below a depth of an inch for 98 percent of the storms, even those with a snowfall rate of up to nearly an inch an hour. The formula for the number of blade-equipped trucks and sanders per kilometer of highway is offered for a wide range of service quality. The cost is not small. This winter is the 100th anniversary of the big rotary snowblowers by road and rail, the winter glory of the old newsreels. Their design theory is given here, although it is not clear that Orange Jull of Orangeville, Ont., the mechanically proficient miller who designed the first one, needed it.

The two million snowmobiles of this continent, the noisy and effective vehicles that carry their drivers out into once-silent snowy lands mainly for sport, trace their ancestry to the first Ski-Doos, produced in the winter of 1959–60 by Armand Bombardier of Valcourt, Quebec. He began work on an over-snow vehicle in 1922 at the age of 15, with a Model T and parts of a sleigh. The standard snowmobile has a single wide rear track and a steering ski or two in front. They can race at a mile a minute and climb 45-degree slopes. The Eskimo hunters use them, but the fuel-hungry vehicle cannot provide the warmth of sled dogs for survival for hours under the most trying conditions of Arctic travel. Heavy burdens can be carried across snow by a wonderful variety of buses and trucks; the largest designs call for huge, ribbed pneumatic tires at very low pressure, some made easily adjustable to the needs of the terrain. A man walks with a bearing pressure of about a third of an atmosphere; the big soft-tired trucks can follow him pretty easily across a four-foot snowbank with a load of 30 tons. Snowshoes and skis, however, reduce the bearing-pressure requirement by a factor of 10.

The volume is full of interest, a bargain in softcover for any technical reader inclined to the "enhancement of life in a snow environment." It is a model effort to cope as rationally as can be with the genuine complexities of a real world, from the layer of disordered protons that lies at the frictional surface of snow to the tax burden of busy, snowy Ontario.

HANDBOOK OF INDIGENOUS FERMENTED FOODS, edited by Keith H. Steinkraus, editor in chief; Roger E. Cullen, Carl S. Pederson and Lois F. Nellis, associate editors, and Ben K. Gavitt, assistant editor. Marcel Dekker, Inc. ($79.50). The sweet life may benignly turn a little acid, as for one who enjoys a slice of Cheddar on a fragment of the sourdough loaf dear to that luxurious city along the Bay of San Francisco. This particular bread, prized for a moist, chewy texture and the faint sharpness coming from fermentation to lactic acid, is made to rise by a leaven of wheat flour and water kept wet and warm until it begins to ferment. A novel yeast, not at all the usual bakers' yeast, and a number of specific lactose-fermenting bacilli form the essential flora of the sourdough. The organisms nourish each other in a balanced way: the bacteria metabolize maltose to yield glucose, which the yeast requires, while the yeast secretes the amylases that catalyze the change of wheat starch into maltose. The trick for maintenance is frequent subculturing into fresh dough, part of the time-honored process. San Francisco, although a fine place for the flora, is not unique. The very same microecology has been demonstrated in sourdough leavens both in Glasgow and in Poland. The substrate selects for the well-fitted organisms, airborne far and wide in the temperate world.

Here is a thick and authoritative volume in which the sourdough phenomenon makes only one rather provincial example in a catalogue raisonné of fermented foods for the widely cosmopolitan palate. There are about 250 substrates discussed, from almonds to sea cucumbers, from the durian fruit to the winged bean. Even wastepaper nourishes something good to eat: the oyster mushroom. For most of the hundreds of products there is an account of importance and use, a recipe in the form of a flow chart, an introduction to the microbiology and the biochemistry of the process and its results, and an estimate of nutritional changes and values. The book fed on papers delivered late in 1977 by a couple of hundred experts from around the world, assembled in Bangkok to outline the indigenous fermented foods they knew. That rich broth was refined, compressed by a factor of about four, and is presented with something of a flourish by the corps of expert editors, who work in the food sciences at Cornell University.

Few will claim to have appraised the relative qualities of ambali and bongkrek, ginger beer and jackfruit wine, munkoyo and ragi-tapai, trahanas and urwaga. Our world is large, even for the jettiest set. All of those are important items of food or drink somewhere, usually prized by the poor for their good value and favored by the discriminating for variety and pleasure. It should be pointed out that most cheeses made from milk and the hops-flavored malted-barley beers of Europe, now enjoyed the world around, are not systematically included.

Main divisions treat the dishes in their imposing variety in four distinct classes. There is hardly space to summarize even one example of each class to characterize the generous global buffet. In the class of acid fermentations used to preserve and to enhance the qualities of fresh foods there is the section on acid-leavened breads and pancakes, about 60 pages of matter from a dozen countries. Sourdough bread is a simple yeoman of this family. Perhaps the subtlest is idli and its group of kindred foods. These are staples in southern India and in Sri Lanka. Two slurries are carefully made, one of soaked and coarse-ground rice, the other of dehulled and fine-ground black gram, the commonest bean of the region. They are mixed into a thick batter, salted lightly and kept overnight in a warm place. No inoculum is needed; the key organisms, two bacterial forms, are present and outgrow all competitors to produce both carbon dioxide as leaven and a pleasant acid flavor.

The idli are made up into little white cakes and steamed in a special pan. The same batter, a little thinner, makes dosa, a lunch or dinner dish of large crisply fried pancakes. There are many other variants, and a range of traditional hot condiments and accompaniments of virtue. Early work appeared to show that the limiting amino acid of the bean protein was increased in amount, much improving the nutritive quality. This has not been confirmed; it is clear, however, that the vitamin content is enriched. Like fresh beer in the medieval European winter, idli are good for you.

Take another class, the foods developed as protein-rich vegetarian meat substitutes, with a fibrous meatlike texture. A large family of Indonesian preparations, generically known as tempe, fills this part of the volume. Two fermentations are involved. The first is a bacterial acidification; it arises naturally when dehulled soybeans are soaked in water. It is not well understood; in the laboratory adding a little acid during soaking is a replacement. The second stage is grander: a particular mold species grows on the partly cooked and long-soaked beans, but it must be supplied as an inoculum. In Indonesia the mold is sold grown and air-dried on the leaf of a hibiscus; in Tennessee, say, a modern small factory industry supplies the growing American market among vegetarians for tempe, providing spores and soybeans in less romantic packages.

The mold can be any of a number of species of a couple of genera. The mold spores are mixed with the drained beans and incubated for a day or so at blood heat. All these conditions are flexible; the process has been studied in detail for many decades. The mold overgrows the beans completely, knitting them together with the tender fibers of its mycelium. The cake, once overgrown, can be sliced and fried in a few minutes; it takes hours to cook untreated soybeans, and then they lack the nutlike flavor and meaty texture of tempe, which is also a little peppery. Best of all, the process can be used for many kinds of substrates besides soybean: peanut presscake, coconut-milk residues, other beans, peas, even wheat and barley.

A start has been made on the engineering of such solid-state fermentations, with pumped air and plenty of controls. The authors tout the process as one way of making substitutes for animal protein in the world of 10 or 12 billion diners to come: inexpensive, easily digested and good to eat. (This is unconfirmed here; the reviewer has yet to try his first tempe-burger.) Plainly, eating a mold colony entire is ecologically more efficient than dining on a soymeal-fed chicken encumbered with feet, feathers, bones and beak, let alone a horned beast of the field.

The volume is completed by soy sauce and sake, fully industrialized in Japan; sauerkraut and miso; a wide variety of tasty-sounding alcoholic brews from homes around the world, based on honey, sorghum, sugarcane, palm-tree sap, bananas, agave or cornstalk juice; together with a set of complicated pastes and sauces compounded of salted fish and shrimp (including a few flow charts that deeply challenge the cultural conventions of any American). Several interesting essays, pointing out relations between indigenous technology and its adaptation to both large scale and new places, come at the end. There are plenty of problems: sauerkraut is sometimes a surprising pink (a pigmented yeast), some false-banana sourdoughs can induce abortion and there are well-known fatal mycotoxins to be shunned. The book is no ordinary compilation but a powerful eye-opener to any self-centered view of everyday life.

Established 1845 SCIENTIFIC AMERICAN November 1983 Volume 249 Number 5

The Uncertainties of a Preemptive Nuclear Attack

It is said that the U.S. needs new land-based missiles because its present ones are vulnerable to attack. Analysis of uncertainties in such an attack suggests that the vulnerability is exaggerated

by Matthew Bunn and Kosta Tsipis

The deterrence of nuclear war requires that nuclear forces not be vulnerable to a preemptive attack. For decades strategic planners in both the U.S. and the U.S.S.R. have been concerned about the possibility of a disarming nuclear first strike, which would leave the victim of the strike unable to retaliate in kind. Each nation has expended large resources on efforts to maintain the survivability of its strategic nuclear forces.

Concerns over the vulnerability of land-based nuclear forces were heightened by the development of multiple independently targetable reentry vehicles (MIRV's) in the late 1960's. This technology, tested initially by the U.S., makes it possible for one missile to carry several warheads, each warhead capable of striking a separate target. Thus each MIRVed missile might be able to strike several of the opponent's missiles, each of which might itself contain several warheads, giving the attacker a double advantage.

As a result the evolution of Russian MIRV technology generated fears that once the U.S.S.R. had developed MIRV's with the appropriate combination of accuracy, yield and reliability it would have the ability to destroy most of the U.S. land-based missile force in a first strike, using a fraction of its available weapons. Although land-based missiles constitute only a fourth of the U.S. strategic nuclear forces, with the rest represented by submarine-launched ballistic missiles (SLBM's) and bombers, concern has arisen that the vulnerability of the ICBM component of the strategic "triad" alone would present the U.S. with a grave problem of national security.

In the fall of 1977 the worst fears of U.S. strategic planners were realized. The U.S.S.R. began a series of tests of a new guidance system with greatly improved accuracy. Simplified calculations indicated that once the U.S.S.R. had deployed an adequate number of these more accurate MIRV's it would indeed have the ability to destroy the bulk of the U.S. land-based missile force in a first strike. Thus was born what has come to be called "the window of vulnerability." This concept has dominated American strategic thinking for some years, providing the primary justification for the development of a new generation of U.S. strategic weapons, including the MX and Trident II missiles.

Simplified calculations do not, however, do justice to the substantial uncertainties inherent in any assessment of the results of a countersilo attack. Intelligence information is rarely absolute, and when strategic planners are confronted with uncertainty about the actual value of such parameters as the accuracy of the other side's ICBM's, they must make assumptions that are conservative from the defender's point of view. Unfortunately the process often obscures the fact that any attack would also involve substantial uncertainty from the attacker's point of view.

Because of the immense destructive power of modern nuclear arsenals, any nuclear first strike would represent a gamble on a scale absolutely unprecedented in human history; the future of entire civilizations would hang in the balance. As a result of the magnitude of the stakes any uncertainty about the outcome of such an attack will act as a powerful deterrent; such gambles are not taken without extremely high confidence in the outcome. The question for anyone considering how to plan the attack will always be not only "What is the expected outcome?" but also "What is the worst plausible outcome?" Therefore in assessing the possibility of an attack it is crucial to gauge quantitatively the uncertainties involved, in order to develop an assessment that is "attack-conservative." In U.S. assessments of the strategic balance this is rarely done. What is generally presented to Congress and the public are the results of an idealized, nearly flawless attack; the uncertainties inherent in the attack are often ignored. In this article we shall be offering a corrective.

The silos in which modern ICBM's are housed are underground concrete structures, "hardened" to withstand the effects of nuclear blasts. There is a wide range of nuclear effects that might damage an ICBM within such a silo, but spokesmen for the U.S. Air Force have indicated that U.S. ICBM silos are generally most vulnerable to the shock wave of the nuclear blast. Thus the hardness of a given silo is usually expressed in terms of the shock-wave overpressure required to destroy it, measured in pounds per square inch (p.s.i.).

The silos currently housing the 1,000 U.S. Minuteman ICBM's are generally estimated to be capable of withstanding overpressures of up to 2,000 p.s.i. The detonation of a half-megaton weapon, such as those carried by the most accurate Russian MIRV's, would create such overpressures at ranges of roughly 300 meters; therefore to destroy a Minuteman ICBM a half-megaton weapon would have to be detonated within 300 meters of the silo. Thus the accuracy with which the weapon is delivered, although it is not particularly important in an attack on a city, would be of decisive importance in any attack on hardened targets such as ICBM silos.

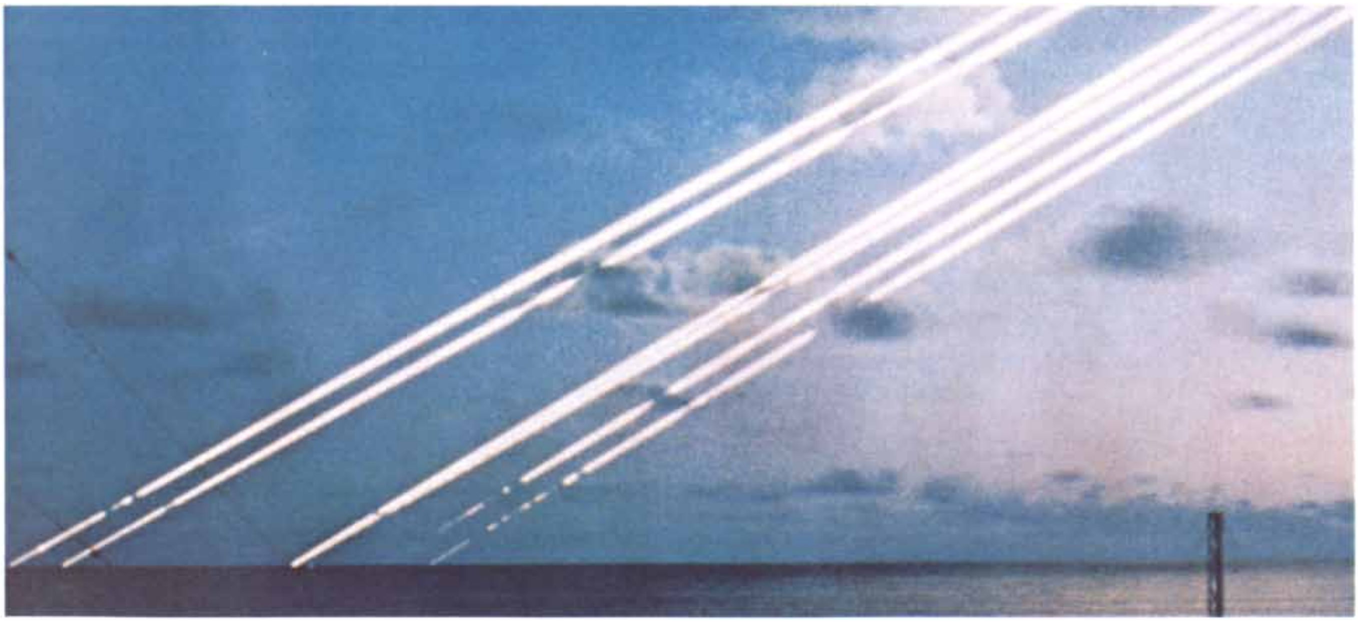
The ICBM's with which such weapons are delivered have three major parts: a rocket, which may have several separate stages; the payload, which consists of one or more reentry vehicles (RV's) armed with thermonuclear warheads, and the guidance system, which directs the rocket thrust in order to place each RV on the appropriate trajectory for it to reach its intended target. When such a missile is launched, the main rocket fires for only the first three to five minutes of the flight. Over the next few minutes a smaller rocket known as the postboost vehicle provides the final trajectory adjustments necessary to set each RV on the path to its intended target. The RV's are then released and fall freely in the earth's gravity field, unpowered and unguided. At the end of a half-hour flight the RV's reenter the atmosphere and detonate over their targets. The missile's flight can therefore be divided into three phases: the boost phase, free flight and reentry.

To deliver the RV's within several hundred meters of their targets over ranges of 10,000 kilometers calls for a highly sophisticated guidance system. Current strategic weapons rely on the technique known as inertial guidance, in which gyroscopes and accelerometers are used to measure the specific forces acting on the missile. These inertial instruments cannot, however, measure the force of gravity, because of the equivalence of gravitation and acceleration described by Albert Einstein: an accelerometer in a free-falling elevator would register zero, even though the elevator would be accelerating toward the ground under the influence of gravity. Therefore in order to account for the effects of the earth's gravitational pull on the motion of the missile a mathematical model of the gravity field as a function of position must be programmed into the missile's guidance system prior to launching. By combining the gravity model with the measurements of the specific forces made by the accelerometers, the guidance computer can use Newton's laws of inertia to calculate the motion of the missile in three dimensions and put it on the appropriate trajectory to its target.

The accuracy of such a weapon can be affected by errors from many sources. The two main sources are errors in the inertial-guidance system and errors associated with reentry into the atmosphere; in some current ICBM's the deviation from the target resulting from either of these sources is 100 meters or more. The errors in the guidance system originate with imperfections in the gyroscopes or the accelerometers; they can be either constant, accumulating with time, or random, arising from vibration, shock or changes in acceleration. The errors associated with reentry originate with unpredictable atmospheric variations over the target, such as winds and variations in atmospheric density, and with uncertainties in the ablation of the RV's nose cone. Because the RV from an ICBM reenters the atmosphere at a speed of some 7,000 meters per second, the nose cone will begin to burn away as the RV passes through the atmosphere; this ablation tends to occur in a rather unpredictable and asymmetric manner, giving rise to aerodynamic instabilities that degrade the accuracy of the RV. Moreover, severe weather effects, such as heavy rain or snow, can cause drastic increases in the rate of ablation, further reducing the accuracy of the RV.

Lesser sources of error include errors in the gravitational model relied on by the inertial-guidance system, errors in warhead fusing and errors in the determination of the position of the target. Each type of error can contribute several tens of meters to the system's overall "error budget." The determination of the initial position and velocity of the launcher is extremely important in systems fired from mobile platforms, such as SLBM's, but is negligible for silo-based ICBM's; indeed, this type of error is one of the reasons current SLBM's are considered to be too inaccurate to strike at hardened missile silos.

The bulk of the error sources of a ballistic missile will be random from one missile to the next. As a result if a large number of such weapons were fired at a single target, they would tend to fall in a random scatter around the target. The spread of the scatter is measured by the parameter known as the cir-



COUNTERFORCE TARGETING PATTERN is suggested by this photograph of six unarmed Mark 12 reentry vehicles streaking toward their targets on Kwajalein atoll in the western Pacific. The reentry vehicles were originally mounted on two Minuteman III intercontinental ballistic missiles (ICBM's), launched in an operational test from Vandenberg Air Force Base in California. The reentry pattern is typical of a countersilo attack in which two thermonuclear warheads from different missiles are directed toward each hardened target.

cular error probable (CEP): the radius of a circle centered on the average point of impact within which 50 percent of the RV's would fall. Thus the CEP is a measure of the precision with which a given missile delivers its payload.

With estimates of the CEP, the explosive yield and the reliability of a given weapon it is possible to calculate the probability of the weapon's destroying a target of a given hardness. As an example, the Russian missile designated SS-19 Mod (for modification) 3 is estimated to have a CEP of 250 meters, and each of its six warheads has a yield of approximately 550 kilotons. The SS-18 Mod 4 has a similar accuracy and warheads with a similar yield, but it carries 10 independently targetable warheads. Assuming perfect reliability, warheads of this type would have a 63 percent probability of destroying a Minuteman silo hardened to 2,000 p.s.i. Assuming statistical independence, two such weapons would have an 86 percent probability of destroying the same silo. This calculation would indicate that with 2,000 warheads the U.S.S.R. could destroy nearly 90 percent of the 1,000 U.S. Minuteman missiles in their silos. This is the kind of alarming theoretical result that has been given wide circulation.

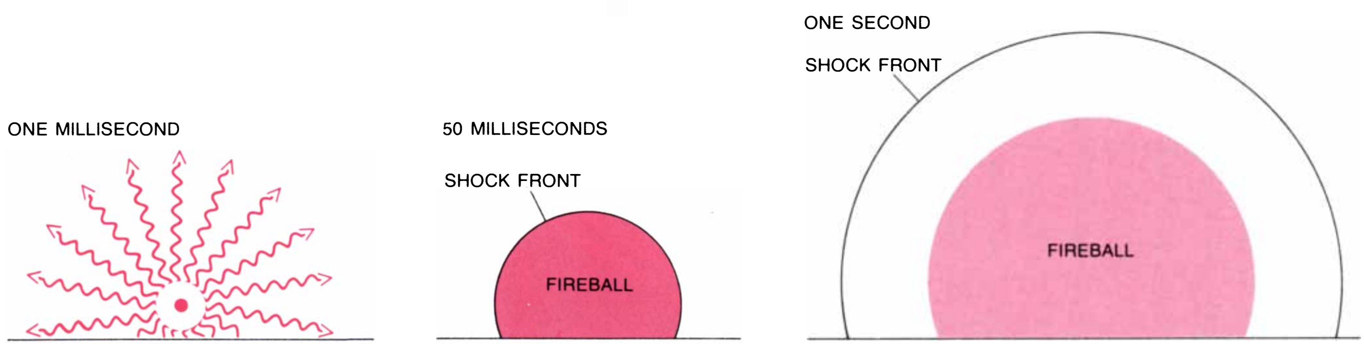
It is unreasonable to expect, however, that a system as complex as a modern ICBM will have a reliability close to 100 percent; a more plausible estimate for the reliability of current Russian ICBM's is 75 percent. Such a reliability would reduce the two-warhead kill probability to 72 percent, which agrees well with the 70-to-75 percent figure currently cited by the U.S. Joint Chiefs of Staff.

Even this figure does not take into account the broad range of uncertainties that must be considered in calculating the possible outcomes of any countersilo attack. The planner of any such attack will always face further uncertainty about the actual accuracy, reliability and yield of his weapons under operational conditions, and about the hardness of his opponent's silos as well. Several less well known factors, such as interference among the weapons used in the attack, known as "fratricide," will also have significant effects on the outcome. As we shall show, these uncertainties would make it impossible for Russian leaders to have reasonable confidence in destroying significantly more than half of the U.S. land-based missile force, given the current capabilities of Russian ICBM's.

Let us first return to the leading source of uncertainty: the accuracy of the attacking weapons. As we mentioned in passing, the bulk of the errors in an ICBM system will be statistically uncorrelated from one missile to another, contributing to the random distribution described by the CEP. As with most other complex electromechanical systems, however, ICBM's show not only random errors but also systematic ones. Therefore whereas the standard calculations of ICBM vulnerability assume that the center of the impact distribution will be directly on the target, the actual distribution will often be offset somewhat. This distance between the target and the average point of impact is referred to as the bias. For example, if the estimated position of a given target is 20 meters south of its actual position, it is likely that the position of other targets in the same general area, such as other silos in the same missile field, will be similarly misestimated; thus on the average the RV's attacking that field will land 20 meters south of their intended targets. Similarly, the gravitational errors for ICBM's launched from silos close together in the attacker's territory will be strongly correlated, as will the errors due to prevailing winds and atmospheric density variations encountered by RV's reentering the atmosphere over the same area in the defender's territory. On the other hand, it should be possible to remove the most significant biases in the guidance system itself through peacetime testing and analysis.

Any bias smaller than roughly 100 meters has essentially no effect on the success of an attack against silos hardened to 2,000 p.s.i. The reason is that any weapon falling within three times that distance of the target will destroy it. Because of the nature of the bias, however, it will be difficult to place limits on the magnitudes of the bias that might be encountered in an actual attack, owing to the statistical properties of large numbers. If an error is simply random from one warhead to the next, then an attack involving 2,000 warheads would provide 2,000 independent trials for that variable; as with 2,000 rolls of a die, the probability of any significant variation from the average outcome is quite small. On the other hand, if an error is completely correlated over some fraction of the force, then for that fraction the attack provides only a single trial. Just as a roll of two dice might well turn up a two or a 12, rather than the more likely value of seven, so biases significantly larger than 100 meters in a single missile attack cannot realistically be discounted.

There is also some uncertainty about the estimate of the CEP for any given ICBM. A variety of sources of information are available on the accuracy of a given weapon. For example, nondestructive tests on the ground can provide



EARLY EFFECTS of the low-level atmospheric explosion of a 550-kiloton nuclear weapon are represented in the sequence of drawings on these two pages. Both the size of the explosive device and the height of its detonation are assumed to be appropriate for an ICBM attack on U.S. Minuteman silos. In the first millisecond after the warhead is detonated the temperature of the nascent fireball is approximately 400,000 degrees Celsius and the overpressure (the increment above ambient air pressure) is on the order of 100,000 pounds per square inch. Radiation (primarily neutrons and gamma rays) that can destroy a nuclear warhead extends outward to a distance of 800 meters from the point of detonation. After 50 milliseconds the radius of the rapidly expanding fireball has grown to about 500 meters and the temperature inside the fireball has fallen to approximately 75,000 degrees C. The overpressure at the shock front (which at this stage is coincident with the surface of the fireball) is 600 p.s.i. and the wind at the perimeter is blowing outward at several thousand kilometers per hour. After one second the fireball has a radius of 900 meters, an internal temperature of 10,000 degrees C. and a surface temperature of 6,000 degrees. The shock front is now expanding faster than the fireball and has reached a radius of 1,400 meters; at that distance detailed performance specifications for every component of the guidance system, which can be combined to yield rough estimates of the overall accuracy of the system. The fact remains that many significant sources of error, such as errors attributable to reentry, cannot be realistically tested on the ground. In addition the interaction of the various components of the guidance system in the demanding vibration, shock and acceleration environment of a rocket boost is extremely complex. Hence only a statistically significant number of realistic full-system flight tests can give accurate estimates of the CEP of a ballistic missile.

Such tests are typically divided into two major categories: (1) research-and-

the overpressure is 40 p.s.i. and the wind is roughly 1,200 kilometers per hour. After 10 seconds the fireball has attained its maximum radius of about one kilometer and has begun to rise. The surface temperature of the fireball is approximately 2,000 degrees C. and the radius of the shock front is about five kilometers. (It is visible only at the top of this frame.) The overpressure at the shock front is about 5 p.s.i. Vertical winds at a speed of about 600 kilometers per hour are beginning to suck dust and other debris from the ground into the stem of the ascending cloud. After one minute the characteristic mushroom-shaped cloud has grown to a radius of 2.5 kilometers and its center has reached an altitude of 6.5 kilometers. The fireball has ceased to radiate at visible wavelengths. Vertical winds at several hundred kilometers per hour continue to hold large particles aloft in the cloud and cloud stem. A second-wave reentry vehicle entering the immediate vicinity at any of these early stages stands a good chance of being destroyed by nuclear or thermal radiation, winds or collisions with the larger particles raised by the detonation of a first-wave warhead. Even at later stages smaller particles raised by the first-wave warheads may affect the second-wave warheads (*see illustration on next two pages*). These destructive effects have been termed "fratricide."

development tests, which serve to suggest design changes and to provide initial information about the accuracy the missile can attain, and (2) operational tests, which provide estimates of the accuracy and reliability of the deployed force. Because a single ballistic missile often costs in excess of $10 million, flight-test programs are generally limited by budget constraints. As a result both the U.S. and the U.S.S.R. have tended to perform a comparatively small number of flight tests of each missile system, with intensive engineering analysis of each test. For example, in the initial stages of deployment a U.S. ICBM is typically subjected to between 25 and 30 flight tests, which yield the primary estimate of the system's operational accuracy and reliability. These are followed by from five to 10 operational tests in each year of the system's life cycle, to monitor any changes that might result from prolonged storage.

The operational flight tests of the U.S. are designed to ensure that the tests are as realistic as possible. The ICBM's to be tested are chosen randomly from the operational force. The chosen missile is then brought to alert in its silo, ready for immediate firing; this is intended to provide a test of the crew and the launch-silo electronics. If the missile fails to come to alert, it is listed as a failure and is not tested further. If it passes the test, the missile is removed from its operational silo and is sent to the test range at Vandenberg Air Force Base in California. The RV's are sent to a special facility where the thermonuclear warheads are removed and replaced with telemetry equipment. The missile with the modified RV's is then fired from the test silo at Vandenberg by a crew randomly selected from the operational missile crews, and the RV's reenter the atmosphere over Kwajalein atoll in the Pacific.

Telemetry equipment on board the missile broadcasts detailed information about such parameters as the flow of fuel, the thrust of the rocket, the performance of the guidance components and the vibration and shocks to which the missile is subjected. This information is picked up by U.S. stations monitoring the tests and by Russian intelligence ships stationed in the Pacific for that purpose. In addition the flight is carefully monitored by radars and optical telescopes based at Vandenberg and on Kwajalein. The information is then intensively analyzed over a period of several months. As a result far more information is generated by a flight test than simply whether or not the missile worked and how far from the intended target it landed.

Significant uncertainties nonetheless persist. First, the number of full-system flight tests is statistically quite small. Second, in spite of efforts to achieve the greatest possible verisimilitude such peacetime testing is still appreciably different from wartime operations. For example, American test vehicles reenter over Kwajalein lagoon, an area where atmospheric conditions are among the placidest in the world. In addition, since the gravity field of the earth varies from place to place, tests over one trajectory cannot in themselves provide estimates of possible gravitational errors over other trajectories.

Perhaps more significant, the range over which most Russian missiles are tested is 6,500 kilometers long, whereas many wartime trajectories would be nearly 10,000 kilometers long. Although the Russians do perform several full-range flight tests in assessing each of their ICBM systems, most of their tests are over the shorter distance. This difference in range has a marked effect on virtually every source of error in the system. Although adjustments to the resulting accuracy estimates can be made, based on a detailed mathematical model of the performance of the system, additional uncertainty will inevitably result. Given these factors, it cannot by any means be ruled out that the CEP in a large-scale countersilo strike could be 10 percent larger (that is, poorer) than the CEP estimated from shots over test ranges; indeed, we believe this is a conservative estimate. For the warheads we have been considering, an unfavorable variation of 10 percent in the CEP alone would reduce the two-shot kill probability against a Minuteman silo from 72 to 66 percent.

Uncertainties in the reliability of an ICBM are separate from uncertainties in its accuracy. Simulations on the ground are considerably more useful for estimating rates of failure than for assessing the accuracy of the weapon system. For example, U.S. Minuteman guidance systems are regularly given simulated flight tests, in which the guidance system is subjected to vibrations and shocks similar to those of a missile flight. As with any sensitive and complex technical system, however, high confidence in estimates of overall reliability can be achieved only through full-system tests, and the number of such tests is quite limited. In addition estimates of the overall operational reliability of ICBM's must take into consideration a broad range of human factors that would be involved in an attack: any large-scale countersilo strike would call for the timely cooperation of several hundred people, whose behavior under such circumstances is unpredictable. Thus 10 percent is probably a conservative estimate of the uncertainty in estimates of overall reliability. Again, an unfavorable variation of 10 percent in the reliability we have assumed for an SS-19 Mod 3 would reduce the number of Minuteman silos destroyed in a hypothetical two-on-one attack from 72 to 67 percent.

Estimates of the explosive power of the thermonuclear warheads have un-

MUSHROOM CLOUDS arising from the simultaneous low-level explosion of several 550-kiloton nuclear warheads are shown at two stages: one minute after detonation (*left*) and 10 minutes after (*right*). The explosions were spaced eight kilometers apart, which is roughly the average spacing between missile silos in a Minuteman field. After 10 minutes the individual clouds have merged into one large cloud approximately eight kilometers thick; the top of the cloud has stabilized at a height of about 18 kilometers. Several second-wave reen-

certainties of their own. This uncertainty is of two interrelated kinds: first, uncertainty about the precise effects of warheads of given yield, and second, uncertainty in the estimates of the yield of a given class of warheads. The peacetime testing of nuclear explosives is limited by considerations of cost, safety, environmental impact, instrumentation and politics. Not least, it is also limited by treaty, including the Limited Test Ban Treaty of 1963 and the still unratified Threshold Test Ban Treaty.

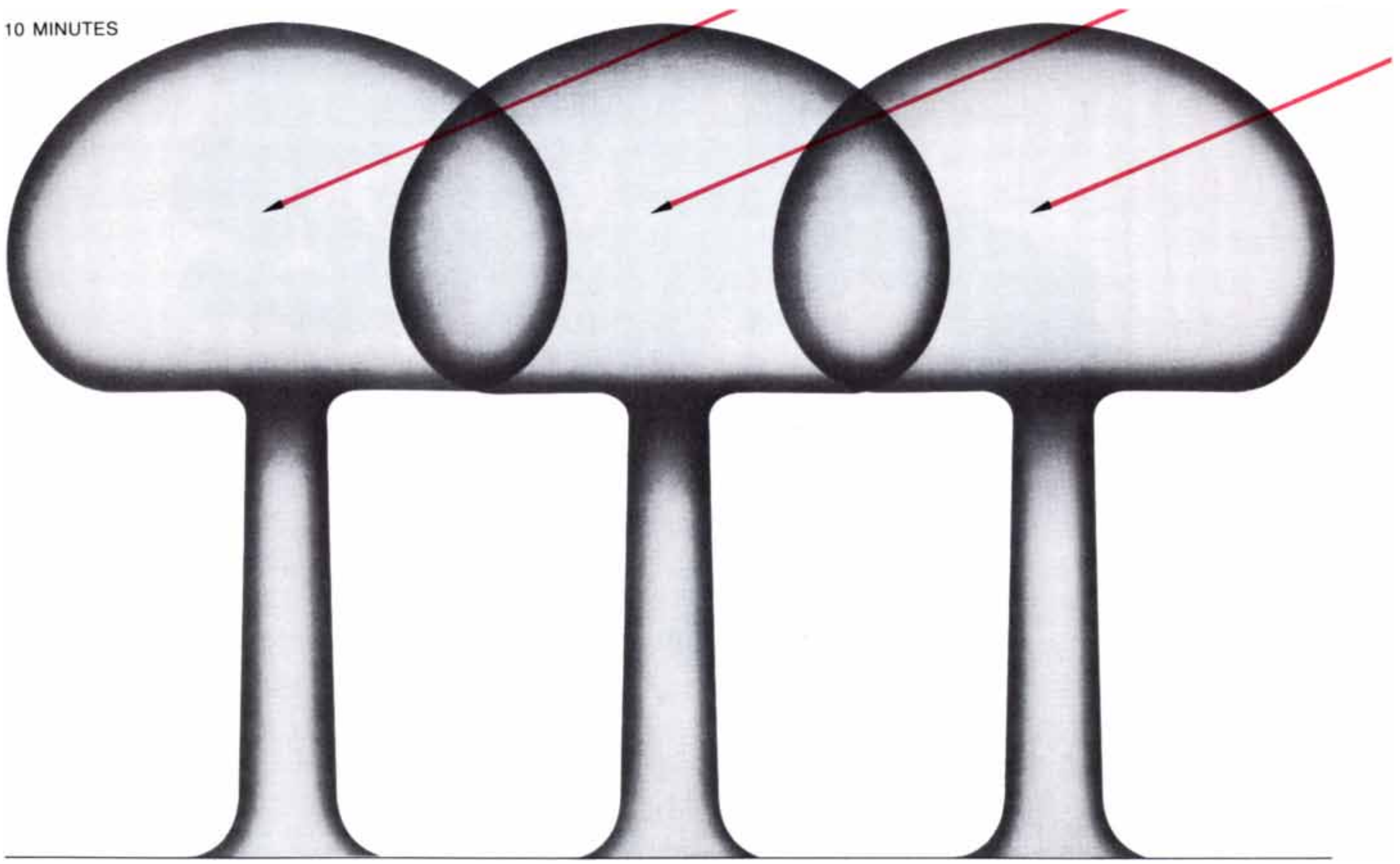
The measurement of overpressures in the extreme range necessary to destroy a modern ICBM silo has been particularly limited, because of instrumental problems and the lack of pressing need. By the early 1960's, when the last American and Russian nuclear tests in the atmosphere were conducted, the hardest targets of interest were roughly an order of magnitude "softer" than current missile silos. As a result there is little in the way of nuclear-test data for overpressures higher than 200 p.s.i., and even less for overpressures higher than 500 p.s.i. The nuclear-test data that do exist for these high overpressures show an enormous degree of scatter, and they often do not agree well with theoretical predictions. Although additional data are available from tests with chemical explosives, large uncertainties are involved in extrapolating such data to predictions of the effects of megaton-range nuclear weapons. The U.S. Defense Intelligence Agency has stated that U.S. estimates of the overpressures to be expected from nuclear explosions in given ranges are uncertain to plus or minus 20 percent.

The second uncertainty in estimating the explosive power of thermonuclear warheads, the uncertainty in estimating the average yield of a given class of warheads, arises from similar sources: the number of tests of any given nuclear weapon is small, and there are substantial instrumental difficulties in measuring the energy yielded by an underground nuclear-test explosion.

The yield to be expected from a given warhead design can be estimated theoretically, but an example of the pitfalls of this approach is provided by the warhead developed for the Mark 12A reentry vehicle recently deployed on the U.S. Minuteman III missiles. The first three tests of the weapon revealed that its yield was considerably less than had been predicted, and the original design had to be modified until, in a fourth test, the weapon achieved its full yield. Although this may be an extreme case, the uncertainty in the average yield of a given class of warheads will be at least 10 percent.

For the sake of simplicity we shall combine the two types of uncertainty in the explosive power of a warhead, describing them both as variations in effective yield. Since the peak overpressure is roughly proportional to the explosive yield, a 20 percent variation in overpressure would result from a 20 percent variation in explosive yield; a conservative estimate of the total uncertainty in the effective yield of a given weapon might then be on the order of 25 percent. An unfavorable variation of 25 percent in the yield of the weapons involved would reduce the effectiveness of the hypothetical attack we have described from 72 to 66 percent.

Although the accuracy, reliability and yield of ICBM's are uncertain, they are at least subject to peacetime testing by a nation that might be considering an attack. This does not hold for the hardness of the silos to be attacked. The overpressure at which a silo will fail depends primarily on the technical characteristics of the reinforced-concrete door at the top of the silo, and it is extremely difficult to obtain reliable intelligence on these characteristics. Indeed, from an attacker's point of view this may well be the largest uncertainty of all. Even U.S.



try vehicles aimed at more-distant targets are shown traversing the cloud blanket. High-speed collisions with comparatively small particles in the clouds could have catastrophic fratricidal effects on the second-wave reentry vehicles, severely degrading their accuracy and perhaps even destroying them. ICBM's that survived the first-wave attack could be safely launched through the cloud cover before the second wave of incoming warheads could safely enter the target area. Hence the attacker might succeed only in destroying empty silos.

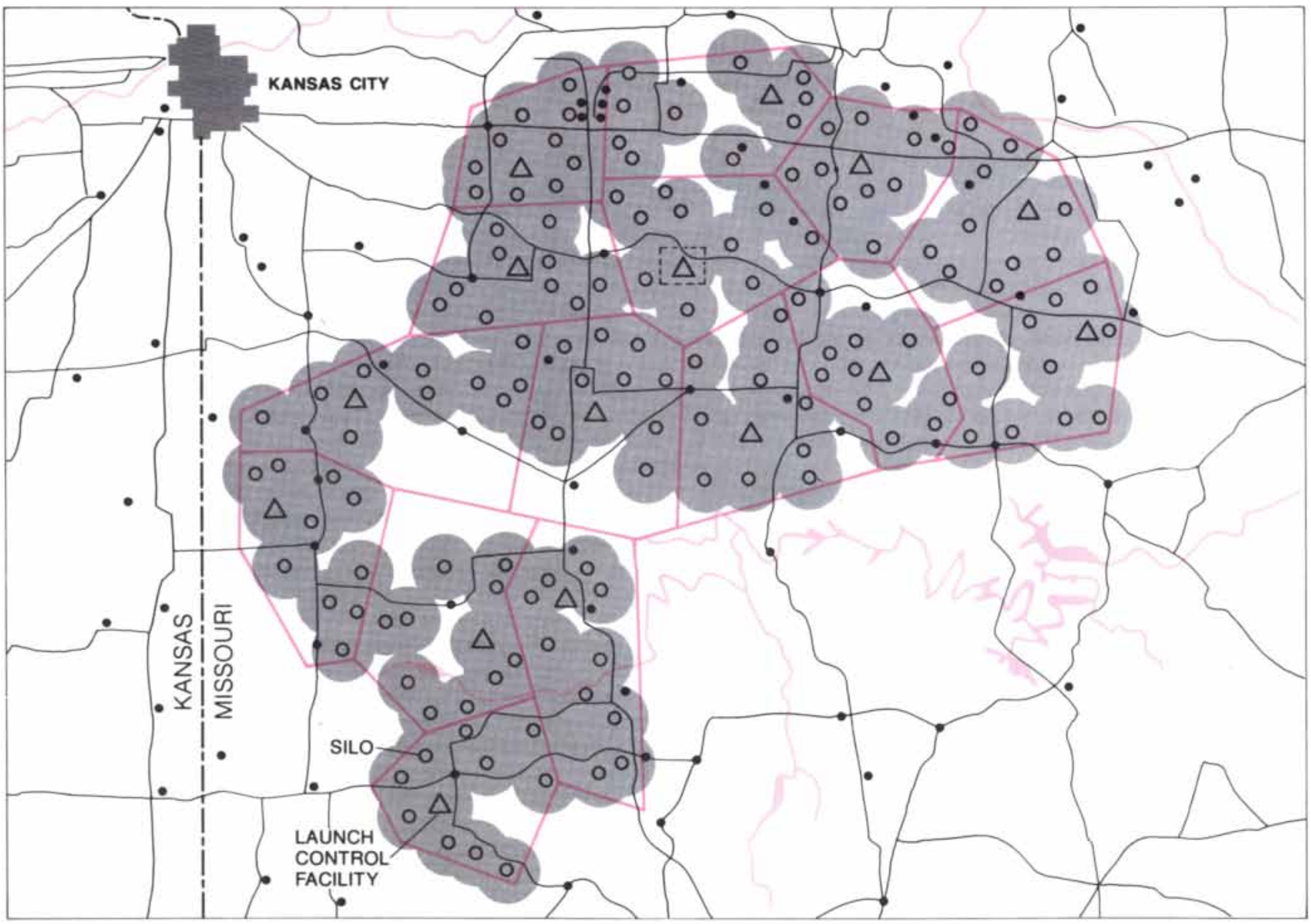
estimates of the hardness of U.S. silos include significant uncertainties; the fact is that no silo has ever been exposed to a nuclear weapon in any test. Estimates of silo hardness are based entirely on theoretical structural considerations and tests of scale models with chemical explosives. Thus from an attacker's point of view the uncertainty in the hardness of the silos to be attacked is likely to be at least 20 percent, if not considerably more. If the silos under attack were capable of withstanding overpressures 20 percent higher than expected, the effectiveness of the hypothetical attack we have been describing would be reduced from 72 to 68 percent.

Another significant uncertainty arises from the interference among the hundreds of warheads that would be involved in an attack. Up to this point we have been assuming that the warheads involved in a countersilo attack would be statistically independent of one another, in other words, that the detonation of one warhead would have no effect on other warheads participating in the attack. This is not at all the case: thermonuclear explosions can have extremely destructive effects on other reentry vehicles, constituting the phenomenon ironically termed fratricide.

In its first milliseconds a thermonuclear explosion gives off an intense burst of radiation, including neutrons, X rays and gamma rays, which in turn generate a powerful electromagnetic pulse. This short-lived burst of radiation is followed by the rapid expansion of a fireball of hot, compressed gases. The fireball expands faster than the speed of sound, reaching a radius of several hundred meters in less than a second. Since the superheated gases in its interior are orders of magnitude less dense than the surrounding air, the fireball begins to rise quite rapidly, much like a hot-air balloon. This creates a vertical wind of several hundred kilometers per hour; indeed, the drag of the wind created by a half-megaton explosion is sufficient to hold aloft a two-ton boulder. In the case of a weapons burst at the optimum height for an attack on a silo hardened to 2,000 p.s.i. the fireball will be in contact with the ground for several seconds; as a result the powerful updraft winds will suck up thousands of tons of dust and other debris, creating the characteristic mushroom cloud.

The cloud rises at high speed; within a minute it reaches an altitude of several kilometers. It then slows down, reaching its maximum height some 10 minutes after the detonation. In the case of a half-megaton weapon the top of the cloud will stabilize at an altitude of some 18 kilometers, with the bottom of the cloud roughly eight kilometers below. By 10 minutes after the detonation the cloud will have covered an area of some 100 kilometers; indeed, the cloud from such an explosion will be so large that in an attack on the ICBM silos in a U.S. Minuteman field the clouds from explosions over individual silos will merge into a single blanket of dust over the entire field.

The significance of the dust and other debris raised by the explosion arises



EXTENT OF CLOUD COVER that would be created by the explosion of the first wave of nuclear warheads in a hypothetical counterforce attack on the Minuteman field in the vicinity of Whiteman Air Force Base in Missouri is indicated on this map. By 10 minutes after the detonation of a 550-kiloton nuclear warhead over each of the 150 Minuteman silos and the 15 launch-control facilities in the target area the merged mushroom clouds would have essentially covered the entire field. If the first-wave warheads were detonated at ground level, the resulting cloud cover would contain millions of tons of dust and larger debris. Even at the greatest height possible for an effective attack on such hardened targets with 550-kiloton warheads, the interfering blanket of particles would still contain hundreds of thousands of tons of material sucked up by the rising fireballs. Population centers, main roads and major rivers are also represented.

from the fact that reentry vehicles traverse the atmosphere at extremely high speed. When an RV reentering the atmosphere in the area of an earlier explosion encounters the cloud, it will be traveling some six kilometers per second. Therefore if the RV collided with a particle weighing several grams, it would probably be destroyed outright; such a collision would take place at several times the speed of a rifle bullet. The smaller particles and dust in the cloud could have catastrophically abrasive effects on an RV: the effect of such high-speed passage through the dust cloud would be equivalent to exposure to an extraordinarily powerful sandblaster. The resulting ablation of the RV's nose cone would severely degrade its accuracy; in extreme cases it could cause the RV to burn up. Moreover, the earlier explosion (or explosions) would have completely changed the density and wind profiles throughout the area in ways that are essentially unpredictable and would have drastic effects on the accuracy of reentering RV's.

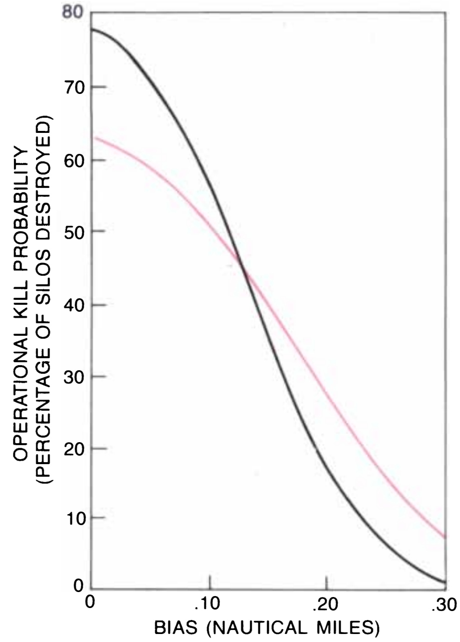
Obviously the problem of fratricide would have to be carefully considered in planning the timing of any attack calling for more than one RV targeted on each silo. If two RV's were targeted on each silo simultaneously, and neither RV failed, then the radiation from the first warhead to explode would destroy the second. If the attack were organized in two waves separated by several seconds, the RV's of the second wave would encounter the fireballs, powerful winds and lethal particles raised by the explosions of the first wave. Although the high temperatures and the outward winds generated by the explosion would subside within the first few seconds, particles large enough to be lethal to an RV would take up to 20 minutes to fall back to the ground, and the clouds of smaller particles and dust would remain until they were dispersed by atmospheric winds. Therefore unless the attacker allowed at least several minutes between the first wave and the second, so that the largest particles would have fallen out of the cloud before the second wave arrived, it seems unlikely that any very large proportion of the second wave would get to its target and do so with the accuracy necessary to destroy it. Even if the attacker allowed several tens of minutes between the two waves, the second wave would still face the dust blanket and the severe atmospheric disturbances created by the explosion or explosions of the first wave.

Since fratricide has never been tested, its effect on incoming RV's cannot be assessed precisely; only the roughest order-of-magnitude estimates are possible. Suppose, in the case of a second wave reentering 10 minutes after the first wave, that passage through the dust cloud and encounters with large particles destroyed only 5 percent of the RV's, and that on the average the effect of the dust and atmospheric disturbances increased the fraction of the CEP attributable to reentry by a factor of two. This, in our view, is a quite conservative estimate of the fratricide such a wave would suffer, given the extreme conditions we have described. These effects alone would reduce the kill probability of a two-wave attack by SS-19 Mod 3 warheads on U.S. Minuteman silos from 72 to 65 percent.

Fratricide introduces an uncertainty that may be even more significant: if the attacker must allow several minutes to elapse between the first and the second wave of his attack, it is quite possible that the ICBM's surviving the first wave will have left their silos before the second wave arrives. Since an ICBM rising out of its silo travels much slower than an RV entering the atmosphere, the particles and dust in the cloud will have comparatively little effect on it. As a result the surviving ICBM's could be safely launched before the second wave of RV's could safely reenter the atmosphere. Although a nearby nuclear explosion that failed to destroy the silo might still keep an ICBM from getting off immediately, in those cases where the first-wave warhead failed the ICBM in its hardened silo would remain undamaged, still available for launching. In the attack scenario we have been considering, involving weapons with a reliability of 75 percent, this would mean that at a minimum 25 percent of the ICBM's could be launched between the first and the second wave of the attack. Combined with the fratricide effects postulated above, this would reduce the fraction of the ICBM silos destroyed in the hypothetical attack we have been considering to 56 percent. Since the vulnerability of an ICBM increases sharply when it leaves its protective silo, however, it is possible that the attacker could keep these weapons from being launched by detonating additional warheads over the silo field at regular intervals so that any ICBM launched from its protective silo would be destroyed; this tactic is called pindown.

Up to this point we have been considering the uncertainties of a countersilo attack on an individual basis. In any real countersilo attack, however, all these uncertainties would be present, making the final outcome of the attack even harder to predict. For example, if all the variables we have been describing were to turn out unfavorably for an attacker, even in the absence of significant bias only 45 percent of the U.S. Minuteman silos would be destroyed. Although it may be unlikely that an attack would be subject to large unfavorable variations in all the uncertain parameters simultaneously, it should be noted that an unfavorable variation in any two of the parameters, combined with fratricide, would reduce the effectiveness of the attack to less than 55 percent, even ignoring the possibilities of bias error and of ICBM's escaping between the two waves. Moreover, we believe the degrees of uncertainty we have postulated are conservative. Thus it would be difficult for a planner to have reasonable confidence that a two-wave attack by SS-19 Mod 3 and SS-18 Mod 4 warheads would destroy significantly more than half of the U.S. Minuteman force.

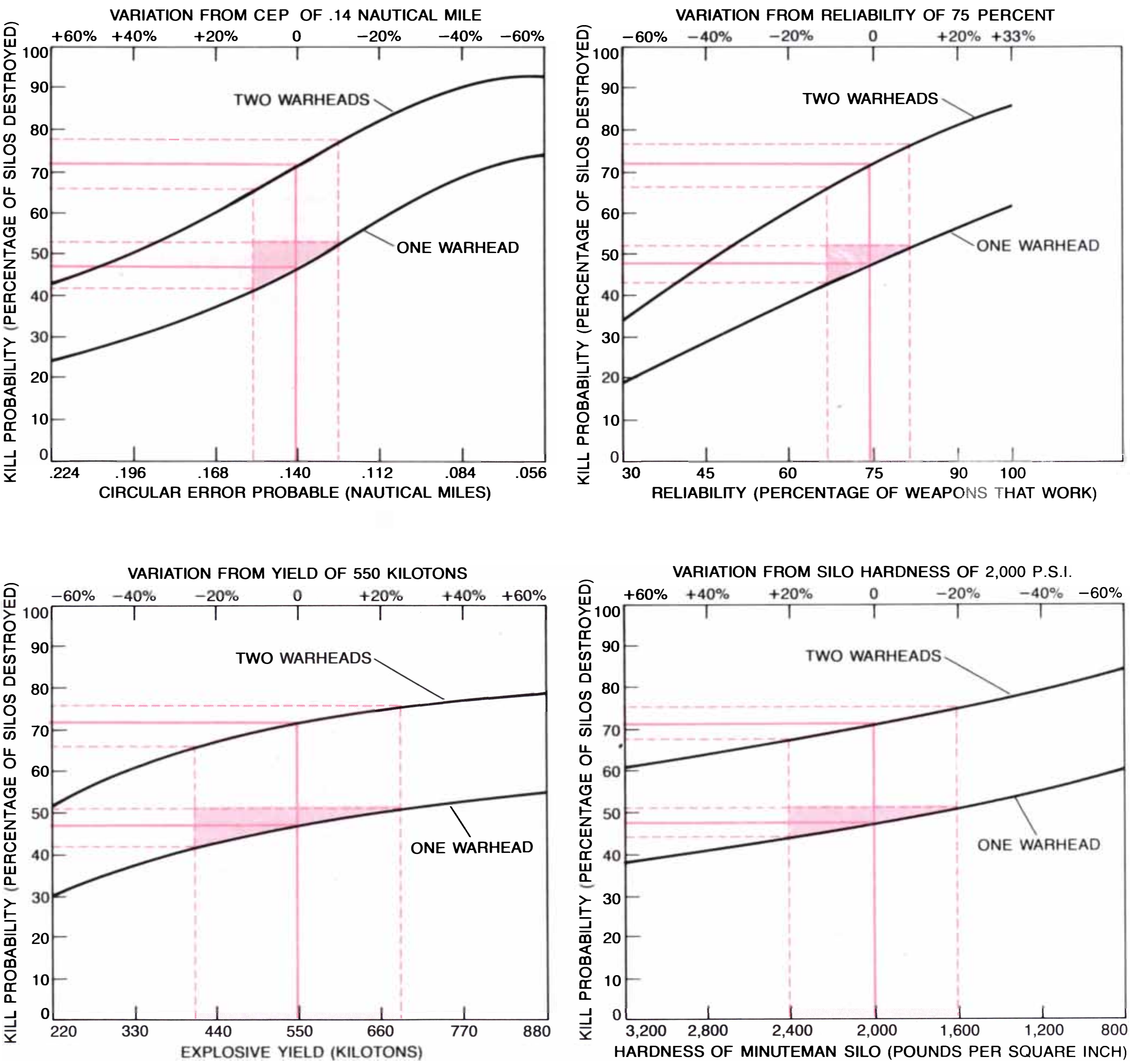
It is possible that a planner would choose to launch more than two warheads against each silo, or to launch larger warheads. In the case of an attack



EFFECT OF BIAS on the operational kill probability of an incoming warhead in a hypothetical countersilo attack is shown in this graph for two ICBM reentry vehicles: a Russian SS-19 Mod 3 warhead (*colored curve*) and a U.S. Minuteman III Mark 12A warhead (*black curve*). (Bias is defined here as the distance from the target to the average point of impact of a random sample of warheads aimed at the target.) In both cases the warheads are assumed to have been fired at targets hardened to survive an overpressure of 2,000 p.s.i. In addition it is assumed that each system has a reliability of 100 percent, that the explosive yield of each warhead is 550 kilotons for the SS-19 and 335 kilotons for the Minuteman III and that the circular error probable is .14 nautical mile for the SS-19 and .1 nautical mile for the Minuteman III. (Circular error probable, or CEP, is defined as the radius of the circle within which half of the warheads aimed at the target will fall.) The important thing to note is that for both missiles a bias of less than about .05 nautical mile (roughly 100 meters) would have very little effect on the kill probability, but that a much larger bias could be quite significant.

on the U.S. Minuteman force, however, neither of these options would be very attractive. For example, a third wave would encounter fratricide effects from both of the first two waves, and it would allow more time after the second wave during which the surviving ICBM's might be launched. Even if the surviving ICBM's did not escape, a third wave would only raise the overall percentage of the silos destroyed from 45 to 57 percent. Since the 1,000 U.S. Minuteman silos contain 2,100 warheads, this would mean an expenditure of 1,000 additional Russian warheads in order to destroy 250 American warheads; such an attack would disarm the U.S.S.R. faster than the U.S. In addition to its ICBM's armed with MIRV's the U.S.S.R. has 100 SS-19's and SS-18's armed with single large warheads; these, however, are clearly much too few for an attack on the 1,000 U.S. ICBM silos.

For a smaller number of more important targets neither of the two preceding arguments would apply. For example, if the U.S. were to deploy 100 MX ICBM's in Minuteman silos, as has been proposed by the Reagan Administration, the 100 Russian large-warhead missiles could be used to attack them. If these weapons were equipped with the accurate guidance systems now deployed on the MIRVed versions of the same missiles, an attack involving one such warhead on each MX silo, followed by a half-megaton warhead, would destroy 77 percent of the MX force, even allowing for unfavorable variations and fratricide. Since each MX carries 10 warheads, this would mean that even under less than ideal conditions 200 Russian warheads could destroy 770 MX warheads. The expenditure of additional Russian warheads would provide even higher operational kill probabilities.



EFFECT OF INDIVIDUAL VARIATIONS in each of four parameters—CEP, reliability, explosive yield and silo hardness—on the outcome of a hypothetical countersilo attack on the U.S. Minuteman force is projected in this set of graphs. In each case it is assumed that the attack is carried out with one or two waves of SS-19 Mod 3 or SS-18 Mod 4 warheads having a nominal CEP of .14 nautical mile, a nominal reliability of 75 percent and a nominal yield of 550 kilotons. The Minuteman silos are assigned a nominal hardness of 2,000 p.s.i. (Nominal values are indicated by solid-color lines.) The light-color bands show the effects on operational kill probability of a conservatively estimated uncertainty of 10 percent in both CEP and reliability, of 25 percent in explosive yield and of 20 percent in silo hardness.

Thus although the current Russian ICBM force could present a severe threat to a small number of particularly valuable targets, it could not provide a planner with reasonable confidence of destroying significantly more than half of the current U.S. ICBM force. The common practice of citing the probable result of such an attack to two significant figures, with no mention of the attending uncertainty, is grotesquely misleading; all that can realistically be said is that such an attack would *probably* result in the destruction of between 50 and 90 percent of the U.S. ICBM force. We conclude that the magnitude of the threat presented by the current generation of Russian ICBM's has been greatly exaggerated.

This comparatively comforting conclusion may not, however, remain valid indefinitely. Strategic-weapons technology is almost never in stasis, and foreseeable improvements in weapons-delivery systems could drastically alter the situation we have been describing. In the past the accuracy of American and Russian ICBM's has generally improved by roughly a factor of two every seven years; the improvement of strategic-weapons technology may slow somewhat as the room for improvement narrows, but there is little reason to expect that in coming years the pattern will be fundamentally different.

Hence sometime in the late 1980's or in the 1990's the U.S.S.R. may have deployed a force of ICBM's twice as accurate as its current weapons. In addition reentry vehicles specifically designed to penetrate dust clouds are currently under development in the U.S., and possibly in the U.S.S.R. as well. These technological changes could significantly reduce the effect of the uncertainties we have described. If the CEP of the SS-19 Mod 3 were reduced by a factor of two, for example, a two-on-one attack would result in the destruction of more than 80 percent of the U.S. Minuteman force, even allowing for a certain amount of fratricide and unfavorable variations in the attack parameters. In essence, an uncertainty of 10 or 20 percent no longer has a significant effect if the weapon is twice as accurate as it needs to be to destroy its target.

The purely technical uncertainties we have been describing, however, are only the tip of the iceberg. The planner of a real countersilo attack would have to consider a host of other uncertainties, mainly centering on the reaction of the nation under attack. For example, a nation with ICBM's has the option of adopting the policy of "launch on warning" or "launch under attack": launching missiles immediately on receiving intelligence that another nation has launched missiles against it. The U.S. has renounced such a policy, because of the possibility of catastrophic error, but it would be difficult for a planner to have any confidence that this policy would hold in the event of an attack. Nearly 30 minutes would pass between the first detection of a massive launch and the detonation of warheads over U.S. silos; the actions of U.S. policy makers in this period and afterward would be impossible to predict.

If in addition to the U.S. land-based ICBM's the rest of the U.S. strategic nuclear forces are taken into account, a serious problem would arise for the putative planner of an attack. First, the thousands of U.S. nuclear weapons in submarines at sea will be invulnerable to attack for as far into the future as it is possible to predict. Second, it would not be possible to destroy both the U.S. ICBM force and the U.S. strategic bomber force. Part of the bomber force is on a constant 15-minute alert, meaning that the only missiles that could destroy them before they got off the ground would be submarine-launched ones with short flight times. Current Russian SLBM's, however, are much too inaccurate to be effective against hardened silos, meaning that the silos would have to be attacked by ICBM's with longer flight times. If this were done, nuclear weapons would begin detonating over U.S. bomber fields more than 15 minutes before the RV's targeted on the ICBM silos could reach their targets, making it very likely that the ICBM's would be launched before the attacking RV's arrived.

Even if these difficulties could be surmounted, an attack involving more than 2,000 near-ground explosions of megaton-range nuclear weapons would cause between 20 and 40 million civilian casualties. Such an attack cannot realistically be described as "a surgical nuclear strike." It is impossible to believe the U.S. would not respond, relying on some fraction of the thousands of bomber and submarine-based warheads that would surely have survived. In all probability the conflict would quickly escalate to an all-out strategic exchange, which would destroy the attacker as completely as it would destroy the victim of the initial attack.

In the more distant future even these inherent uncertainties might be somewhat reduced by advances in technology. The development of accurate SLBM's, already under way in the U.S., will erode the relation between the alert bomber force and the ICBM force. If extremely accurate maneuvering reentry vehicles are deployed, hardened silos could be destroyed with weapons of much smaller yields, sharply reducing the number of civilian casualties such an attack would inflict. This might at first glance seem desirable, but the possibility of a genuinely "surgical" strike, combined with increased confidence in predictions of its outcome, would increase the temptation to launch such a strike, lowering the nuclear threshold and increasing the probability of nuclear war.

Thus although the current situation is stabler than is commonly believed, the progress of weapons technology bodes ill for the future. It is still possible, however, that stringent limitations on the testing and deployment of ballistic missiles could forestall many of these undesirable technological advances. Far from locking the U.S. into a position of vulnerability, such limitations, if they were effective, could prevent the rapid erosion in U.S. security that will otherwise occur in the years to come. Limitations on the testing and deployment of ballistic missiles could be an important component of arms-control efforts, and they deserve more careful study than they have had to date.

ASSUMPTIONS	BIAS (NAUTICAL MILES)			
	0	.05	.10	.15
100 PERCENT RELIABILITY	86%	84%	76%	63%
75 PERCENT RELIABILITY	72%	70%	62%	50%
LIGHT FRATRICIDE	65%	62%	56%	45%
UNFAVORABLE VARIATIONS	45%	43%	38%	31%

EFFECT OF COMBINED UNCERTAINTIES on the outcome of a countersilo attack is presented in this table. The columns correspond to four different projections of bias. The rows indicate the effects of different assumptions about the technical capabilities of the weapons employed in the attack (and, in one instance, of the hardness of the silos under attack). The first row shows the outcome in terms of operational kill probability of an ideal attack with perfectly reliable weapons on silos with no variation in hardness from the nominal value. The second row shows the results of a somewhat more realistic attack employing weapons with a nominal reliability of 75 percent. The third row shows the effect of a minimal amount of fratricide. The fourth row shows the outcome of an attack in which light fratricide is combined with unfavorable variations (for the attacker) in all the other parameters. In this case the accuracy and reliability turn out to be 10 percent less than the nominal values, the yield of the attacking weapons is 10 percent less and the silos under attack are 25 percent harder than predicted. Even in the absence of significant bias only 45 percent of the Minuteman silos would be destroyed.

Slurry Pipelines

They currently serve to transport metal ores and coal suspended in water. Their greatest potential is for transporting coal from Western states to power stations in other parts of the country

by Edward J. Wasp

Most of the coal in the U.S. (an estimated 72 percent of the minable reserves) is in seven Western states. Although the country is not currently suffering from a shortage of fuels other than coal, experience has shown that the fuel situation can change rapidly, and in any case a cheaper fuel, other things being equal, is always preferable to a costlier one. The reason Western coal is not cheaper than it is currently is the cost of transporting it: most of the electric-power generating stations, which would be the consumers of the coal, are outside those seven states. To build roads and railroads to the mines would call for an enormous capital investment. An alternative that would call for a smaller investment, although still a large one, is slurry transportation: pipelines through which coal is pumped suspended in water.

The longest coal-slurry pipeline currently in operation is the Black Mesa line, which transports coal 273 miles from a mine in northeastern Arizona to a power plant in the southern tip of Nevada. At the mine end the coal is crushed, formed into a slurry and set in motion by three giant reciprocating pumps. At the power-plant end the slurry enters special agitated storage tanks and is fed into continuously operating centrifuges that remove water from it. Out of this process comes a material that looks like damp black beach sand. It is coal for burning in the power plant.

The Black Mesa line has been in steady operation since 1970. It is regarded as the prototype for the much larger systems that will probably be built in the West before the end of the decade. One such system is the one planned by Energy Transportation Systems, Inc. (ETSI), which will move coal from mines in Wyoming south as far as 1,400 miles to Oklahoma, Texas and Louisiana.

Slurry transport is a phenomenon that has been building river deltas and scouring canyons throughout geologic history. It works on the principle that a fluid sufficiently turbulent and fast-moving can transport a particulate solid an indefinite distance. One of the great natural slurry transportation systems is the Mississippi River in the spring: soil is fed into the muddy brown flood from headwaters and tributaries and is carried to the Mississippi delta more than 1,000 miles away. There in an average year the delta is extended six feet by 495 million tons of silt.

These giant natural riverine slurries and manmade pipeline slurries have several things in common. Both consist of finely divided solids suspended in a liquid. In the natural system the solids are ground by erosion; in the manmade one they are ground by machine. Both must be kept moving at or above a critical velocity or they will settle out of suspension. Both eventually do separate out of the transporting fluid. In the natural system they form sandbars or deltas; in the manmade one they are removed mechanically or by evaporating the transport water. Both end up transporting solids from one place to another.

The physics of slurry transport encompasses the principles of fluid dynamics and hydrostatics, whose study goes back at least to Archimedes. For slurry transport, however, the most important scientific findings were those of the 19th-century British physicist Osborne Reynolds. His 1883 paper "An Experimental Investigation of the Circumstances Which Determine whether the Motion of Water in Parallel Channels Shall Be Direct (Laminar) or Sinuous (Turbulent), and the Law of Resistance of Parallel Channels" cleared the way to the understanding, which came later, that turbulence plays a fundamental role in the transport of solids by water. By observation Reynolds developed a formula that generates an index, now called the Reynolds number, indicating whether or not the flow in a pipe is turbulent.

A simple interpretation of the physical significance of the Reynolds number is that the numerator represents the forces tending to cause disorder (higher velocity increases momentum and the degree of randomness in the flow) and the denominator represents the force opposing disorder (greater viscosity damps randomness in the flow). The higher the index, the greater the turbulence, and when in a round pipe the ratio reaches 2,300 (that is, 2,300/1), the flow becomes turbulent. The turbulence in turn increases the capacity of the pipe to carry solids. This last point is easily recognized in the flow of a river: at its head high in the mountains the river flows fastest and carries the coarsest solid material (including rocks); at its mouth it flows slowest and carries only the finest silt.

Modern commercial slurry systems were not developed until after World War II. Since then several commodities have been successfully transported by slurry systems. Iron-ore concentrate, copper-ore concentrate, phosphate-rock concentrate, limestone and the asphalt-like mineral gilsonite are transported by pipeline at several locations around the world. Moscow uses a slurry pipeline to transport garbage from the city to a disposal plant. Moreover, a number of liquids other than water have been examined for slurry-transport purposes: crude oil, methanol (methyl alcohol) and liquid carbon dioxide.

Coal suspended in water, however, was the first slurry system and remains the one with the greatest potential. This is in spite of the fact that it presents special problems. Whereas the metallic ores that have been transported by pipeline are slurried as part of the refining process, coal must be first slurried and then dewatered, steps not normally part of the coal-handling process. Dewatering brings the coal as much as possible back to its original (dry) state for burning in the furnace of a power plant.

The first major experimental work on a coal-slurry system with commercial potential was done in New York City by Wallace C. Andrews, president of the New York Steam Company, which began distributing steam to city residents in 1882. Andrews developed his concept to move coal cheaply and efficiently from mines to his boiler plant. In the early 1890's he received one patent for his coal-slurry pipeline concept, a second for "An Apparatus for Preparing Coal for Transportation" and a third for "Settling or Storage Pond or Basin for Pulverized Coal." He also built a model coal-slurry pipeline system at his boiler plant at 58th Street and Madison Avenue, complete with pipe running over scaffolding and around corners to simulate cross-country pumping conditions.

It was not until 1914, however, that the first coal-slurry pipeline went into operation. It was in London. Designed by Gilbert G. Bell, it had a pipe eight inches in diameter and a mere 1,980 feet long. It nonetheless did yeoman work, moving 50 tons of coal per hour from barges at a dock on the Thames to a nearby powerhouse.

A major advance then came not from coal and electric-power interests but from the petroleum industry. As the industry expanded, well-drilling technology quickly evolved. The drilling called for methods of preparing and handling drilling muds: slurries pumped down the drill pipe to lubricate and cool the drill bit and to remove rock chips from the well hole. Powerful pumps are needed to force the drilling mud down the drill pipe at pressures approaching 5,000 pounds per square inch. The need for equipment that will operate reliably at such pressures and under the severest conditions has contributed greatly to the development of the massive positive-displacement reciprocating pumps that power slurry-pipeline systems today.

The first modern study of coal-pipeline transportation technology was undertaken in 1951 by the Consolidation Coal Company in Ohio. The company supplied coal from its Cadiz mine in the southern part of the state to the Cleveland Electric Illuminating Company's Eastlake Generating Station 108 miles to the north. At that time coal was shipped a few cars at a time coupled into ordinary freight trains, and the coal company thought it might do better with a slurry line.

Consolidation Coal had to start from scratch, as there were no operating coal pipelines from which to draw technical experience. The coal-slurry research team, of which I was the leader, first built a demonstration plant at Cadiz consisting of two loops of 12-inch pipe. One loop was 4,000 feet long and laid on level ground; the other was 1,200 feet long and laid on an incline. The plant included its own coal-preparation plant, capable of crushing up to 20 tons per hour.

The purpose of the test loop was to examine the distribution and behavior of slurry particles inside the line and to determine the optimum pipeline velocity. If the velocity was too low, the slurry particles would start settling out of suspension, scrape along the bottom of the pipe and erode it. If the velocity was too high, energy would be wasted propelling the slurry and again there would be excessive erosion. The research team also needed to find out if the slurry particle size would be reduced as the slurry traveled down the line. Perhaps most important, the degree to which pressure would drop along the pipeline had to be determined, so that when a full-scale line was built, pump stations would be properly placed and would include sufficiently powerful pumps.

An additional major concern was determining the normal rate of internal corrosion and erosion inside the slurry pipe. The research team needed to find conditions that would allow the pipeline to last for the life of the overall slurry system as well as transport a stable slurry a long distance. To collect the corrosion and erosion data the team designed a short section in the 4,000-foot test loop that included 10 horizontal sections resembling barrel staves. These sections could be removed, weighed and examined after various slurries were run through the loop.

Tests were begun with a slurry in which the coal was in small lumps averaging three-eighths of an inch in diameter. The loss of metal along the bottom of the pipe was enormous compared with that at the top. The plain carbon-steel pipe of the test section was replaced with a stainless-steel one; although stainless steel is slightly less re-

IRON-ORE-SLURRY PIPELINE crosses a gorge of the Savage River in northwestern Tasmania. It carries ore concentrate from the Savage River mine 53 miles over rugged terrain to processing facilities at Port Latta on the coast. Diameter of the pipe is $9^5/_8$ inches.

sistant to abrasion than carbon steel, it is much more resistant to chemical corrosion. The metal losses were still high at the bottom of the pipe but were not at the top. The high metal losses at the bottom could now be definitely identified as being caused by coal sliding along it. As increasingly fine coal sizes were tried, the coal began to stop dragging along the bottom, and erosion measurements in the test section showed progressively less wear. Furthermore, probes inserted into the pipe near the top, near the bottom and in the middle showed that as smaller particles were slurried the slurry became increasingly homogeneous.

A basic conviction arose that the key to the design of slurry systems that would operate reliably lay not in the selection of exotic pipe materials or the design of special equipment but in the understanding and control of the slurry environment. More specifically, it was thought that if the flow was homogeneous, that is, if the solids were uniformly suspended along the vertical axis of the pipe, the pipeline would be stable with a pressure drop that would be constant with the distance traveled. Two other results would follow naturally from a properly controlled slurry environment. First, if the corrosive environment was controlled and made benign, the wear in the pipe would be uniform and of such a minimal magnitude that a pipe life of more than half a century could be expected. Second, in a homogeneous environment the coal itself would not undergo any attrition, that is, degradation in particle size, during its travel, even over very long distances.

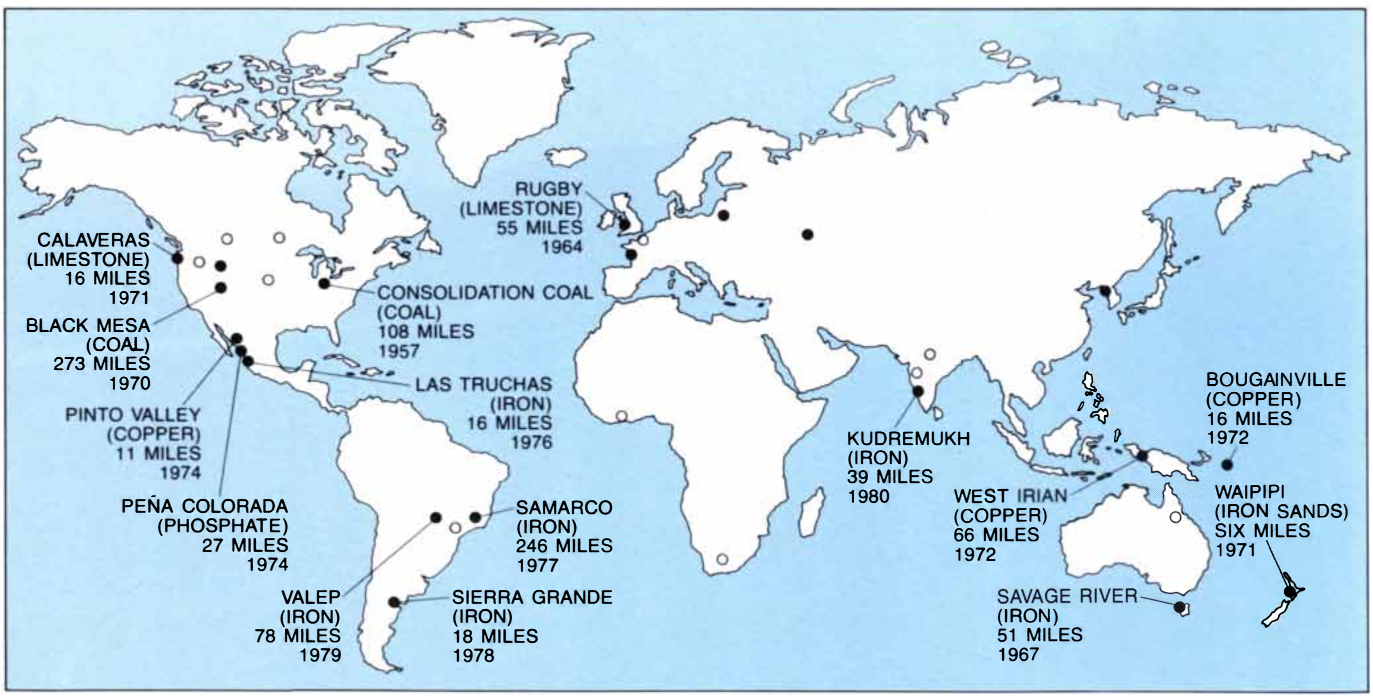
If this basic conviction were proved correct, conventional materials and well-established construction methods that had been developed for pipelines by the oil and gas industry over the first half of the 20th century could serve for slurry-pipeline systems. That could well lead to a rapid, inexpensive startup for a viable coal-pipeline industry.

Several problems remained. Coal particles in the test loop continued to degrade into progressively smaller fines, and the pressure drop in the loop decreased with the time the coal particles spent in it. The research team had to determine whether the level of fines in the full-scale pipeline would increase because of attrition and whether this would affect the pressure drop in the line. Until Consolidation Coal's Eastlake line was completed in 1957 a long-distance, once-through slurry system was not available for experimentation; it was therefore necessary to extrapolate the pressure-drop information obtained from the 4,000-foot test loop to a 108-mile continuous system. Furthermore, since the increase in the level of fines and the decrease in the pressure drop were caused in part by attrition, it was not possible before building the full-scale line to separate the relative contributions to the attrition made by the pipe and by the pump in the system.

To play it safe, the Eastlake coal-slurry pipeline system was designed to accommodate the maximum pressure drop recorded on the test loop and to deal with the fine particles that resulted from the attrition of the coal in the test loop. When the Eastlake system began operating, there was absolutely no measurable attrition in particle size over the 108-mile distance traveled. All the attrition observed in the test loop was caused by a centrifugal pump through which the slurry had to pass more than 100 times as the loop simulated the full-scale system. The pipeline was installed with only three pump stations equipped not with centrifugal pumps but with reciprocating ones.

There was heated debate about what would happen to the turbulence in the pipeline as the slurry traveled down the line away from one of the pump stations. Would it diminish, as some contended, so that the coal particles would lose their support and settle out? Since there was no way to test for the problem until the line was built, the slurry research team had to rely on a "reverse experiment." I reasoned that if the pipeline was shut down and started up again, and if the turbulent forces in the line were strong enough to redistribute the solids uniformly in it before the turbulence generated by the pump reached the sampling point, then the forces in a continuously operating line would be strong enough to maintain the distribution of the solids.

To test this hypothesis the 4,000-foot loop was shut down and then started up. Probes were inserted in the loop at frequent intervals at the top, middle and bottom of the pipe. When the flow was restarted, the course of the slurry resuspension was observed with hundreds of measurements of particle concentration taken at the top, middle and bottom. It was determined that within a minute after the pump started, the slurry had re-



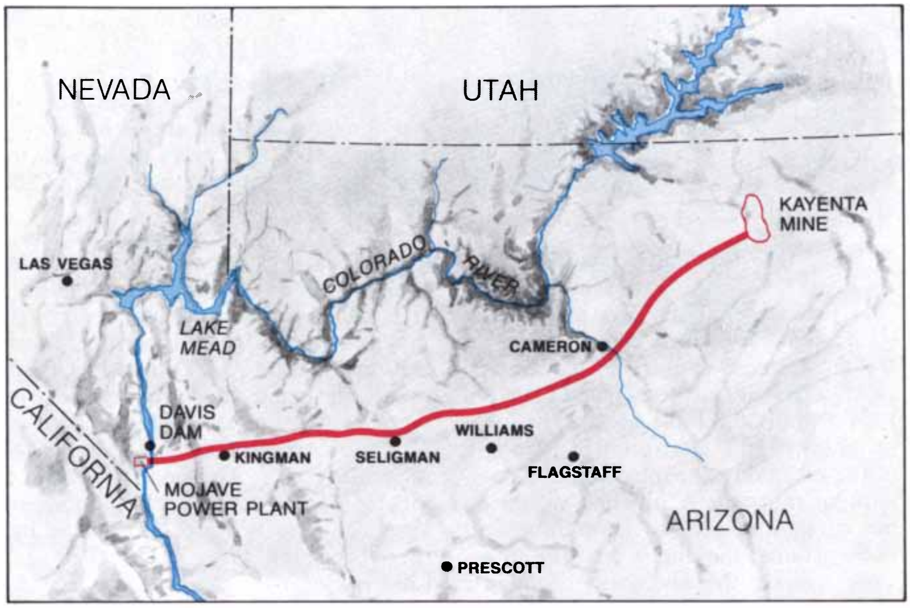
PRINCIPAL PIPELINE PROJECTS are given with their length, the material transported and the date the pipeline went into operation. The Consolidation Coal pipeline closed down after operating successfully for six years. Circles are where pipelines are being considered.

suspended all along the length of the loop. After this experiment worry about the effect of distance on turbulence vanished.

The Eastlake pipeline operated successfully for six years until June, 1963, when it was mothballed after having transported some seven million tons of coal. The competition it had given coal-hauling railroads had forced them to operate more efficiently, to develop the concept of the unit train (a train hauling only coal) and to drastically cut their rates for four times the volume of coal that was being moved by the pipeline. In these particular circumstances the railroad had become the cheaper means of transportation.

The next major U.S. coal-slurry pipeline project was sponsored by a railroad: the Southern Pacific Transportation Company. That was the Black Mesa line. Before Black Mesa went into operation, however, several other slurry pipelines were built at widely dispersed locations around the world. The decision to rely on slurry lines at these locations was made easy because material had to be moved across mountainous or otherwise difficult terrain in areas where the construction of roads or railroads would have been prohibitively expensive. Modern pipeline technology made it possible to build the lines on steep grades over routes much more direct than those needed for alternative transportation methods.

At the same time that the Eastlake coal-slurry line was going into operation in 1957, the American Gilsonite Company began operating a gilsonite refinery in Grand Junction, Colo., to process the mineral into coke, gasoline and fuel gas. The gilsonite is mined underground in the Uintah basin of northeastern Utah. High-pressure water jets pulverize the mineral in its seam, and the resulting slurry is pumped to the surface. Once the slurry is aboveground it is moved 72 miles over rugged mountains (including an 8,500-foot pass) through a six-inch pipe to the Grand Junction refinery. If the slurry line had not been built, the gilsonite would have had to be hauled 182 miles from the mine to the refinery by truck or narrow-gauge railroad.



LONGEST SLURRY PIPELINE currently in operation is the Black Mesa line of Southern Pacific Pipelines, Inc. It moves coal 273 miles from the Peabody Coal Company's Kayenta mine in Arizona to the Southern California Edison Company's Mojave power station in Nevada. For most of its length the pipeline is 18 inches in diameter. The Black Mesa line carries 5.5 million tons of coal per year with 99 percent reliability. It went into operation in 1970.

The next major long-distance slurry line was built on behalf of Pickands Mather & Co. in northwestern Tasmania. The company-was interested in developing and shipping iron concentrate from the Savage River mine, situated in rugged, semimountainous terrain with few roads and no railroads. The relatively low capital investment needed to construct a slurry pipeline and the low operating costs of this bulk-transportation method convinced the mine's developers that they should rely on a slurry system to ship the ore 53 miles from the mine to processing facilities at Port Latta on the Tasmanian coast.

Major construction barriers had to be surmounted, including two river gorges. The Savage River gorge was more than 1,200 feet wide, and the pipeline crossing was 450 feet above the river. A suspension bridge was built that carried the $9\frac{5}{8}$-inch pipe across in a graceful arc. The experience gained in building the Eastlake coal line, including the mathematical formulas and models developed as part of Consolidation Coal's research program, proved to be sound groundwork for the iron-ore line. When the line was first filled with an iron-concentrate slurry, the observed discharge pressure was within 2 percent of that projected by laboratory studies and by mathematical methods based on those applied for the Eastlake line. The Savage River line has been operated with 99 percent reliability for more than 15 years. It moves some 2.5 million tons of iron concentrate per year.

Iron ore is also transported in the Waipipi slurry line in New Zealand. Placed in operation in 1971, the Waipipi Iron Sands Project consists of a relatively short land pipeline and a ship-loading system. A four-mile land slurry line feeds ore to a shore installation. There a 1.8-mile undersea line carries the slurry to an offshore mooring buoy, where special Marconaflo tankers load the slurry and then dewater it on board. After the transport water is removed from the cargo and piped ashore, the tanker sails to its destination with the iron ore. On arrival the ore is slurried again for unloading.

The Savage River iron-slurry line is far exceeded in length by the Samarco iron-slurry line in Brazil, which hauls ore concentrate from mines in the interior through more than 400 miles of jungle terrain to ship-loading facilities on the coast. The 20-inch Samarco line was completed in 1976 and transports 12 million tons of iron-ore concentrate per year.

Two copper-ore-slurry pipelines went into operation in 1972: a six-inch, 17-mile line on the island of Bougainville northeast of Australia and a four-inch, 69-mile line in the West Irian section of New Guinea. The West Irian line receives copper-ore concentrate at an elevation of 8,500 feet and transports it down a mountain to a processing and loading facility on the Arafura Sea.

A limestone slurry is transported by pipeline from the Calaveras limestone mine in California 17 miles through the Sierra Nevada foothills to a cement plant. This major slurry project was completed in 1964 and since then has had an operating availability of more than 98 percent.

Unfriendly terrain and the availability of pipeline technology were important considerations when Southern Pacific Pipelines, Inc., decided to build the Black Mesa pipeline. The line crosses a high desert plateau and four mountain ranges with grades of up to 16 percent along its 273-mile route from the Peabody Coal Company's Kayenta mine to the Mojave power plant of the Southern California Edison Company. Between Pump Station No. 2 at Gray Mountain and No. 3 near Williams, Ariz., the line climbs 3,000 feet in 25 miles. The last 12 miles of the route drops 3,500 feet down the side of a mountain range into the valley of the Colorado River. For most of its length the Black Mesa pipeline is 18 inches in diameter. On the steep final 12-mile downgrade, however, the pipe's diameter is reduced to 12 inches to diminish the excess pressure caused by the rapid descent.

The Black Mesa line is capable of carrying 5.5 million tons of coal per year, and its reliability now approaches 99 percent. The route crosses some of the most scenic country in the American Southwest near the Grand Canyon, Lake Powell and Lake Mead. The local geography and environment strongly militated against the intrusion of a coal-hauling railroad. A pipeline would be out of sight, buried at least three feet underground, making it possible to restore quickly the landscape along the route to its natural condition.

At the Kayenta end of the pipeline the Peabody Coal Company transports two-inch lumps of coal by conveyor belt to the Black Mesa preparation plant. There cage mills crush the dry coal to 3/8-inch lumps. The smaller lumps are then fed to a battery of rod mills that mix the coal with water and wet-grind it until it can pass a No. 14 mesh (a mesh with interstices of 1.19 millimeters). The resulting slurry, which varies slightly in concentration from 46 to 48 percent coal by weight, is first stored in three agitated tanks, each holding a two-hour supply of slurry for the system, that feed Pump Station No. 1.

The positive-displacement pumps at the first station feed the slurry into the main line at 600 pounds per square inch. Pump Station No. 2 at Gray Mountain has four pumps discharging at 1,600 p.s.i. in order to generate the head needed to get the slurry over the Gray Mountain grade. The remaining two stations, No. 3 near Williams and No. 4 near Seligman, are identical with station No. 1. The rates of flow in the pipe are held between five and 5.7 feet per second.

At the Mojave power plant the slurry discharges into one of three agitated tanks capable of holding it for 24 hours. It is then pumped into the centrifuges that dewater it. In addition the Mojave station has eight 75,000-ton Marconaflo ponds designed to store and recover large volumes of slurry coal. There are 20 centrifuges for each 750-megawatt generating unit. The coal cake that is yielded by the centrifuges still contains about 25 percent water. It must be further dried and pulverized before it is blown into the boiler fireboxes with heated air.

At about the same time that the Black Mesa system was being designed and built, an even more ambitious project was envisioned: a massive coal-slurry system linking the newly developed coal mines in Wyoming's Powder River Basin with power stations in states to the south as far east as Texas and Louisiana. Several other coal-slurry pipelines have been proposed, including the Powder River Pipeline, which would transport Wyoming coal to Wisconsin and Illinois; the Coalstream system, which would transport coal from both West Virginia and southern Illinois mines to power stations in Florida; the San Marco system, which would transport Colorado coal to the Houston area; the Alton system, which would transport coal from southern Utah mines to Utah and Nevada power stations, and the Pacific Bulk Commodity Transportation System, which would transport Utah coal to a ship-loading facility on the California coast for export to Japan.

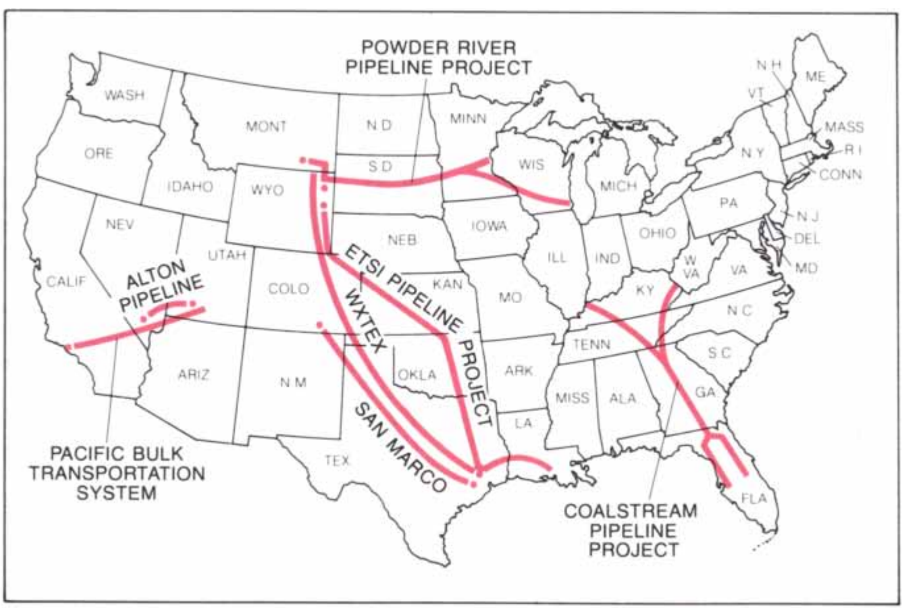
Bechtel Petroleum, Inc., is currently negotiating the construction of two new coal-slurry pipelines in Asia. It is conducting a feasibility study jointly with the China National Coal Development Corporation on a 450-mile pipeline to run from Inner Mongolia to the port of Quinhuangdao and to four power plants. Together with Armand Hammer of the Occidental Petroleum Corporation, Bechtel is also discussing with Russian officials a coal-slurry pipeline from Siberia to Moscow. Proposals for two new coal-slurry pipelines, both about 1,200 miles long, are being studied in India.

To make the pipeline from the Powder River basin to the south a reality Bechtel joined with the investment-banking firm of Lehman Brothers Kuhn Loeb, Inc., to form Energy Transportation Systems, Inc., and began the ETSI Pipeline Project in 1973. Today the joint venture is owned by the Texas Eastern Transmission Corporation, Kansas-Nebraska Natural Gas Co., Inc., InterNorth, Inc., and Bechtel. Lehman Brothers left the joint venture at the beginning of this year.

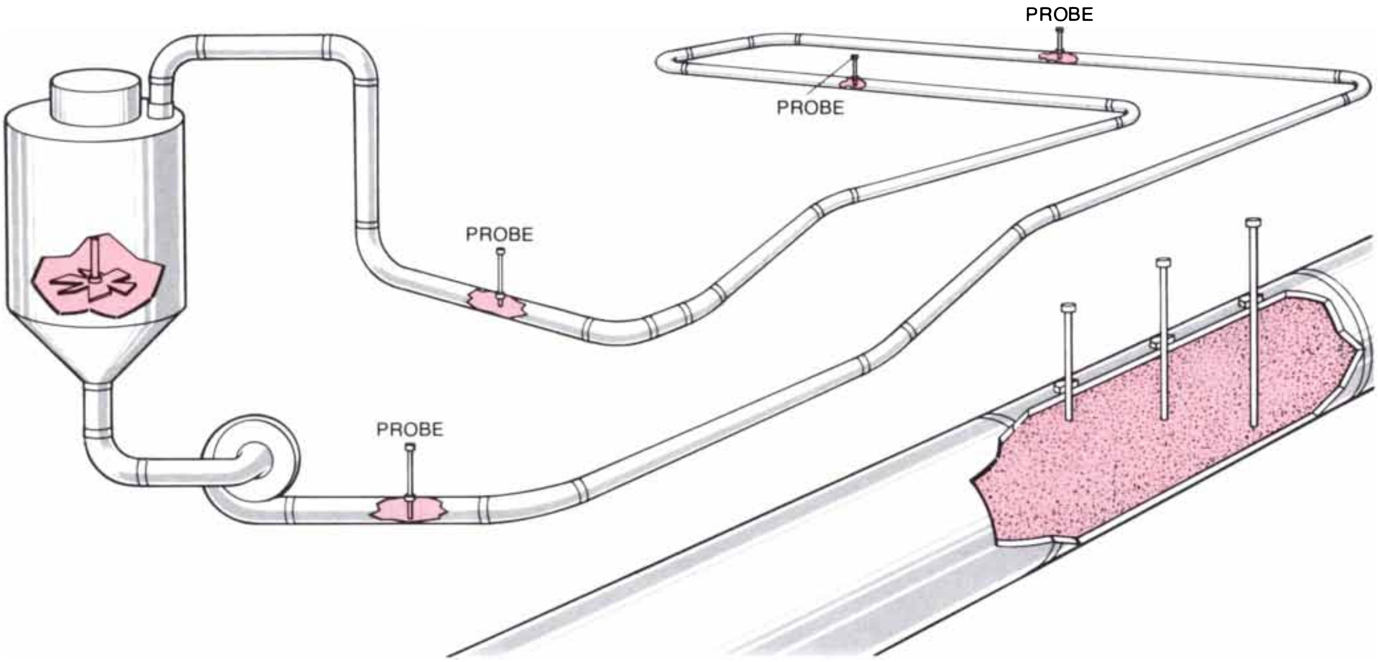
Of several large-scale slurry-pipeline systems being planned today the ETSI project is the one closest to construction. During its lifetime on the drawing boards the proposed ETSI system has been subject to change, reflecting changing market conditions. As the system was originally conceived, it would include three coal-preparation plants that would receive and slurry coal from several different mines near Gillette, Wyo. Two of the outlying preparation plants would be linked by gathering lines to the third: the Thunder Basin plant at the head end of the main slurry line.

Water would be transported to the preparation plants from the Lake Oahe reservoir on the Missouri River near Pierre, S.D., through a buried 34-inch steel pipeline. The 20,000 acre-feet of water that will be used in the slurry pipeline to transport 30 million tons of coal per year is less than .1 percent of the annual flow of the Missouri at Pierre. The 30 million tons of coal per year represents about 4 percent of all U.S. coal production.

At the coal-preparation plants the same type of cage mills and rod mills installed at Black Mesa will grind and slurry the coal received from the mines by conveyor belt. The slurry will consist of approximately 50 percent coal and 50 percent water by weight. The main line will be a carbon-steel pipe 40 inches in diameter between its Wyoming end and the first delivery points in Oklahoma. Electrically powered 1,500-horsepow-



LONGER COAL-SLURRY PIPELINES are scheduled to go into operation by the end of the decade. The longest pipeline is the ETSI Pipeline Project of Energy Transportation Systems, Inc. It will carry coal from mines in Wyoming as far as 1,400 miles into Texas and Louisiana.



TEST LOOP was needed before construction could start on the first coal-slurry pipeline: one from the Consolidation Coal Company's mine in Cadiz, Ohio, 108 miles north to a generating station of the Cleveland Electric Illuminating Company. Actually two loops were built, one 1,200 feet long and the other 4,000 feet long. The slurry was kept agitated in a tank. Probes along the loop monitored velocity, pressure and particle distribution. Sections of the pipe could be removed to determine whether the flow caused wear. In this way optimum particle size, particle concentration, velocity and pressure could be established. Full-scale pipeline went into operation in 1957.

er positive-displacement reciprocating pumps in stations from 80 to 100 miles apart will be used to propel the slurry through the line at approximately the walking speed of a man, which is required to keep the slurry in turbulent flow. The slurry will take about two weeks to travel from one end of the line to the other. The line will carry the slurry almost due south across Wyoming, then turn southeast across the northeast corner of Colorado before crossing Kansas in a southeasterly direction into Oklahoma.

When the line reaches Oklahoma, it will split. One leg will head almost due south into Texas. The other will serve power stations in northern and northeastern Oklahoma. Alternative configurations to serve additional power stations in Texas and Louisiana are being studied and may be included in the final system, depending on the outcome of ETSI's marketing efforts.

Coal from specific mines will be transported in batches to specific generating stations to meet the ash, sulfur and B.t.u. requirements of their boilers. Each public utility in the ETSI system would have a contract to buy coal from individual Powder River basin mines. To ensure that the right coal reaches the right destination, batches of coal will have a recorded history prepared when they arrive from the mines at a coal-preparation plant. The history will include an analysis of the batch's chemical properties and heat content. The information will be compared with that for batches arriving at each terminal facility. If the analysis of the arriving batch matches that of the batch expected for delivery, the arriving batch will be delivered to the dewatering plant for processing.

To aid in tracking individual batches through the main slurry line electrolytic markers consisting of a salt solution will be placed in the line ahead of and behind each batch. When the markers pass a pump station, a dewatering terminal or some other way station on the line, they will cause the electric potential across the line (measured by resistance meters) to increase sharply, signaling the beginning or the end of a batch. A group of agitated tanks at each terminal facility will enable the terminal operator to choose between processing the slurry immediately or storing it for later dewatering or later shipment to a downstream terminal in a predetermined sequence.

The entire cycle of coal preparation, pipeline transportation and dewatering has been thoroughly and successfully evaluated with more than 1,000 hours of operation at ETSI's Coal Evaluation Plant in Arkansas. The Coal Evaluation Plant was constructed for full-scale equipment test, so that the behavior of various coals could be studied through the grinding, slurrying, shipping and dewatering processes. The evaluation program demonstrated two things. The first is that coal slurry meeting pipeline specifications can be produced with commercially available cage mills and rod mills, each mill consuming about the same amount of energy as the mills at the Black Mesa coal-preparation plant. The second, at the other end of the circuit, is that the Coal Evaluation Plant consistently produced dewatered slurry coal having between 7 and 11 percent surface water with the reliability needed for commercial operation.

Through the use of a continuous-belt filter press in the dewatering cycle, the dewatering process recovered 100 percent of the coal from the transport water; no more than 30 parts per million of suspended solids remained behind in the water after the coal was removed. The dust-collection equipment on the vibrating-bed dryers worked more efficiently than had been expected. This fact, together with the complete recovery of the solids from the water, assures that both Federal and state environmental standards for industrial air emissions and wastewater discharges at the dewatering plants of the system will easily be met.

Western coal has one disadvantage in addition to its being far away from most existing power stations. That disadvantage, like the one of remoteness, is negated by slurry-pipeline technology. It consists in the fact that Western subbituminous coal has a relatively low heating value and a high inherent moisture content. A hydrothermal coal-treating process to upgrade B.t.u. content and reduce moisture had been explored on a small scale in the laboratory. This treatment was found to permanently raise the B.t.u. content and reduce the inherent moisture content of coal by heating it in slurry form to above 600 degrees Fahrenehit. To find out if this heat treatment could be applied in a large-scale continuous process, Texaco, Inc., joined with Bechtel in running a 20-ton-per-day pilot plant at Montebello, Calif.

The coal to be processed was first ground and then slurried in water at concentrations of between 45 and 50 percent. The slurry was heated to between 500 and 650 degrees F. in heat exchangers pressurized to keep the slurry liquid. The resulting dewatered coal had a higher heating value than the coal fed into the process because much of the volatile matter had been driven out of it. The B.t.u. content was raised by up to 25 percent; the final test run increased the heat content of a pound of Western coal from 8,500 B.t.u. per pound to 10,500. (High-volatile Kentucky bituminous coal has a heat content of 14,800 B.t.u. per pound.) The process could easily be adapted for minehead coal-preparation plants or dewatering terminals on slurry systems.

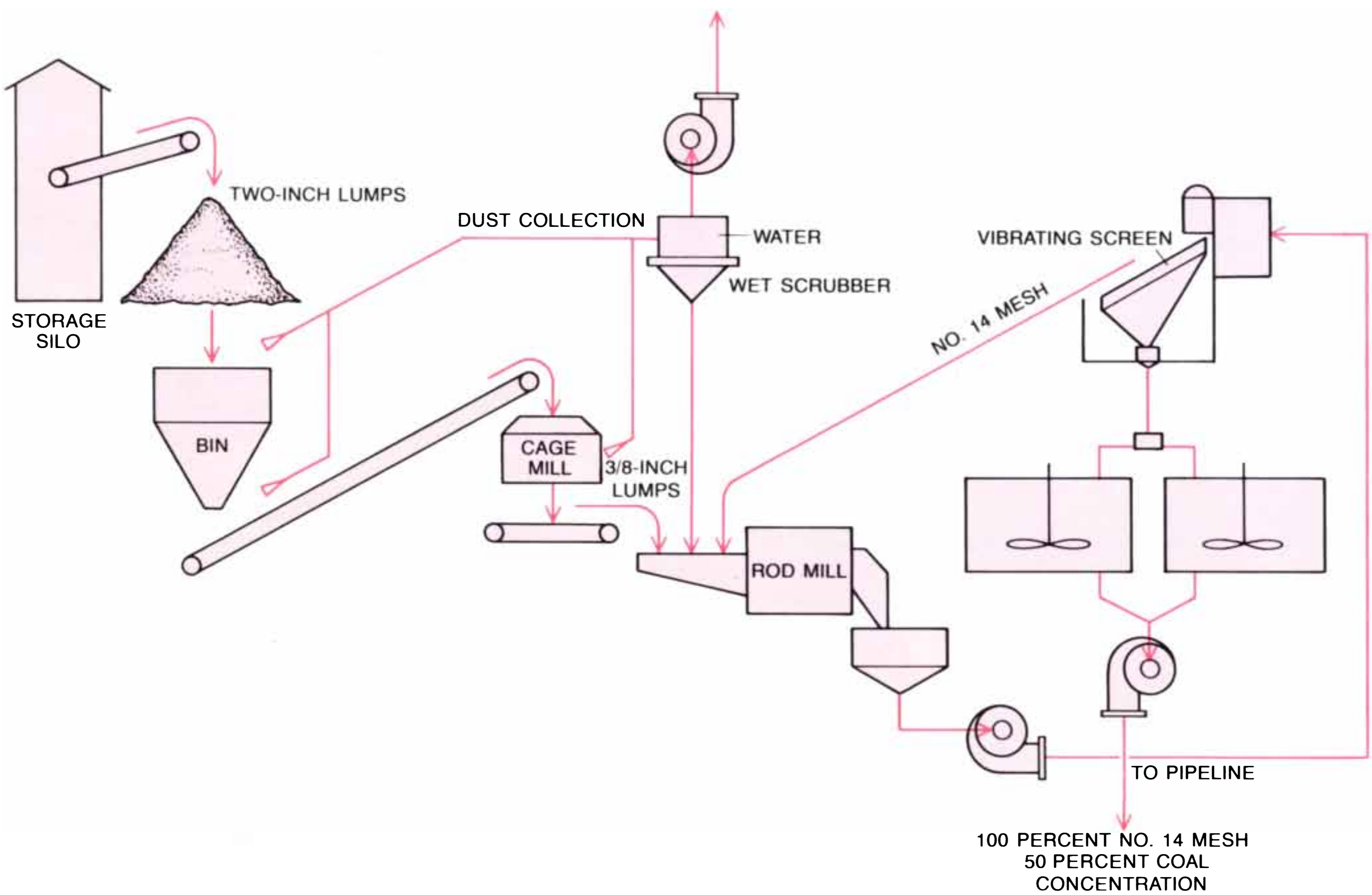
The distance penalty for the use of Western coal is outweighed by a single great advantage: it has the lowest sulfur content of any coal found in the U.S., making it ideal fuel for today's environmentally clean generating stations. Sulfur dioxide is the chief pollutant in coal smoke and may be the cause of "acid rain." Modern coal-fired generating stations rely on scrubbers to remove sulfur dioxide from their stack gases. The less sulfur there is in the coal, the less difficult it is for the scrubbers to clean up the sulfur dioxide.

Two of the major successes of the ETSI project are the experience with the coal-evaluation plant and the results of engineering design. Others include the acquisition of about 90 percent of the rights of way needed for the main slurry system, the acquisition of 70 percent of the easements needed for the water-supply line, the securing of a stable supply of water and the granting of a favorable environmental-impact statement by the Department of the Interior.

In order to acquire the rights of way opposition from railroads had to be overcome. The railroads' major tactic has been to refuse ETSI crossing rights under their tracks. At first this appeared to be an obstacle impossible to overcome. Legal research into the original granting of Western railroad rights of way more than 100 years ago revealed, however, that in many places the railroads had only a surface easement for their tracks; the subsurface rights were retained by the landowners. Title searches made in counties through which the ETSI pipelines would pass located these "windows," and ETSI negotiated easement agreements with the landowners. Then, through a series of court cases filed to validate the agreements, ETSI overcame the railroad-crossing obstacle. Of the 65 window cases filed in the lower courts, all were decided in ETSI's favor. Four subsequent railroad appeals to higher courts were dismissed, and ETSI is now free to build its pipeline under railroad tracks without getting any further approvals of window crossings.

If there had been Federal eminent-domain laws for coal-slurry pipelines, ETSI's window program would have been unnecessary. Such laws are vitally needed for the construction of the other planned coal-slurry systems in this country, particularly those crossing Eastern states where railroads hold full title to their rights of way.

The first coal-pipeline legislation was proposed in 1962 by President Kennedy. His bill, which would have inaugurated the building of a coal-slurry pipeline from West Virginia to New York and New Jersey, was opposed by Eastern railroad companies and was defeated. In 1974, at the height of the Arab oil embargo, a second coal-pipcline bill was introduced in Congress. This time, still opposed by railroad companies, it passed the Senate by voice vote but was defeated in the House Committee on In-



SLURRY IS PREPARED at the mine end of a coal-slurry pipeline. Typically the coal is first crushed into two-inch lumps (*top left*). It is reduced to smaller lumps in a cage mill, water is added and it is further crushed in a rod mill. It is then passed through a mesh to ensure that no large lumps remain in it. The coal slurry is kept suspended in agitated tanks (*bottom right*) until it is pumped into the pipeline.

terior and Insular Affairs. A third bill was introduced in 1975, only to be tabled in committee the following year. The coal-pipeline bill of 1977 reached the House floor and was defeated. Last year legislation cleared House and Senate committees but did not come to a vote before Congress adjourned. In September of this year similar legislation was defeated on the floor of the House. This, however, does not affect the eminent-domain arrangements that will enable ETSI to go forward.

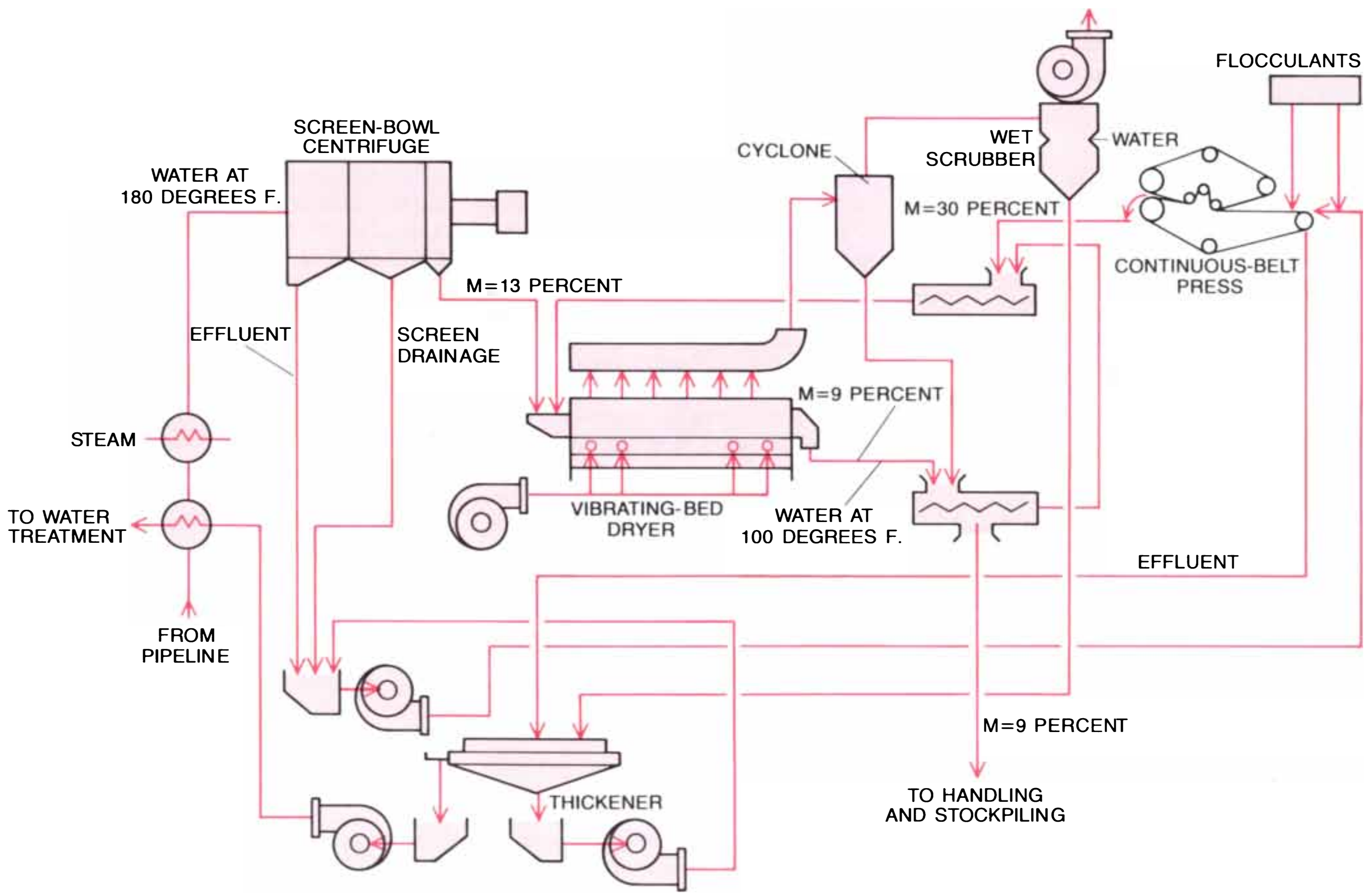
By the end of March of this year ETSI had acquired a total of 1,161 miles of right of way for its main slurry line and 174 miles of right of way for the South Dakota water line, the results of the largest right-of-way acquisition program ever undertaken at one time in the contiguous 48 states. Of the total rights of way acquired only 58.7 miles were across Federal lands and only 40 miles were across state lands.

The Federal lands it is necessary to cross are chiefly the national grasslands in Wyoming and Colorado and other lands in the same states administered by the Federal Bureau of Land Management. It was for this reason that ETSI needed to have an environmental-impact statement. The statement was issued by the Department of the Interior after two years and $6 million in studies. It found the proposed ETSI system was both environmentally acceptable and a "cost-competitive option for transporting coal." It noted that the buried pipeline would have a minimal effect on the environment and that vegetation would restore itself over the right of way within a few years.

In December, 1981, ETSI signed an agreement with the state of South Dakota to buy up to 50,000 acre-feet per year from Lake Oahe. The lake is one of several "main stem" reservoirs built and managed by the U.S. Army Corps of Engineers under the Pick-Sloan plan to end flooding along the Missouri River. At present no use is being made of the water stored behind the Oahe Dam, although Federal and state studies indicate that three million acre-feet per year is available for industrial purposes. As part of its agreement with the state ETSI has agreed to transport 4,300 acre-feet of water per year free of charge to communities and rural water districts along its aqueduct route in western South Dakota.

When construction on the ETSI system begins, the work will proceed simultaneously on the Oahe water line, the main slurry line, the preparation plants and their gathering-line systems, and the dewatering terminals at ETSI customers' generating stations. ETSI expects that as many as 5,500 workers will be employed building the system. The entire effort will create more than 12,000 jobs on the system during construction and some 24,000 additional jobs in related industries. The value of the materials and equipment needed to build the system is expected to total more than $1.5 billion out of a total cost of about $3.5 billion, a cost estimated to be 9.5 percent of the $36.1 billion in commercial and industrial construction contracts awarded in the U.S. in 1982.

Black Mesa showed how to build and run a reliable long-distance slurry pipeline. The ETSI system would dramatically scale up the Black Mesa concept into a pipeline system carrying slurry across a third of the country. One would have to go back more than 100 years to the construction of the transcontinental railroad to find a transportation project equal in size and potential impact on the U.S. economy. There is every reason to believe that because coal-slurry pipelines will contribute to the economy and secure energy supplies for the future the necessary legislation will finally be written into law and the infant industry will be able to grow to adulthood.



SLURRY IS DEWATERED at the power-plant end of the pipeline. The coal slurry coming in from the pipeline is first steam-heated to 180 degrees Fahrenheit. It is run through continuously operating centrifuges that remove all but 13 percent of its surface water (*M*). Other components of the system remove an additional 4 percent of the water. The water removed meets all Federal environmental standards.

Centaurus A: the Nearest Active Galaxy

Active galaxies are galaxies that radiate as much as a million times more energy than typical galaxies. The question of what powers the radiation is studied by observations of Centaurus A

by Jack O. Burns and R. Marcus Price

Among the largest and most intriguing objects that have yet been observed in the universe are the objects known as active galaxies. Active galaxies, which make up only a few percent of all known galaxies, can emit a million times as much energy in the form of electromagnetic radiation as an ordinary galaxy. The emissions from an active galaxy extend over many different frequencies, including gamma rays, X rays, visible radiation, infrared radiation and radio waves. Much of the recent observational work has been done at radio frequencies, where the emissions from an active galaxy can be 100,000 times as strong as those from an ordinary galaxy.

The active galaxy nearest our own galaxy is Centaurus A, which is about 15 million light-years away. The visible outline of Centaurus A is a few tens of thousands of light-years across. Extending at an angle from the central plane of the visible structure, however, is an elongated radio envelope some three million light-years from tip to tip. If the radio-emitting region of Centaurus A were visible to an observer on the earth, it would appear to be 20 times as wide as the moon.

How is the enormous radio structure of an active galaxy such as Centaurus A created and maintained? The combination of observations made at X-ray, radio and optical frequencies is beginning to tell astronomers a great deal about how Centaurus A is organized. The material of the radio structure is expelled like a stream of water from a hose by an engine at the center of the galaxy. From the engine comes a thin beam of radio plasma: a highly ionized gas whose main components are electrons (moving with nearly the speed of light), magnetic fields and possibly protons. As the electrons travel outward from the center of the galaxy they emit radiation with a wide range of frequencies. In doing so they lose energy, and their motion would soon stop if they were not continuously supplied with energy. As the beam moves outward from the engine it passes into three radio lobes arranged in order of increasing size. The middle lobe coincides with a series of optical filaments, thin regions that give off radiation in the visible range. The filaments have recently been shown to be regions where new stars are being born.

Although much has been learned about active galaxies in the past decade, some of their most fundamental features are not understood. The nature of the engine that drives the plasma beam is not known with certainty and the details of the mechanism can only be guessed at. Furthermore, there are several possible explanations of how the electrons in the beam are accelerated as they move outward from the galactic core. We cannot choose the correct explanation on the basis of current knowledge. We also do not know how the radio and optical components of the galaxy interact in the regions where the young stars form. Recent observations of Centaurus A at radio and X-ray frequencies have greatly increased our knowledge of what is going on in this active galaxy. In one sense the new information has made Centaurus A more familiar. In another sense the observations have revealed physical processes so large and powerful that they make the active galaxy seem an even stranger and more wonderful phenomenon.

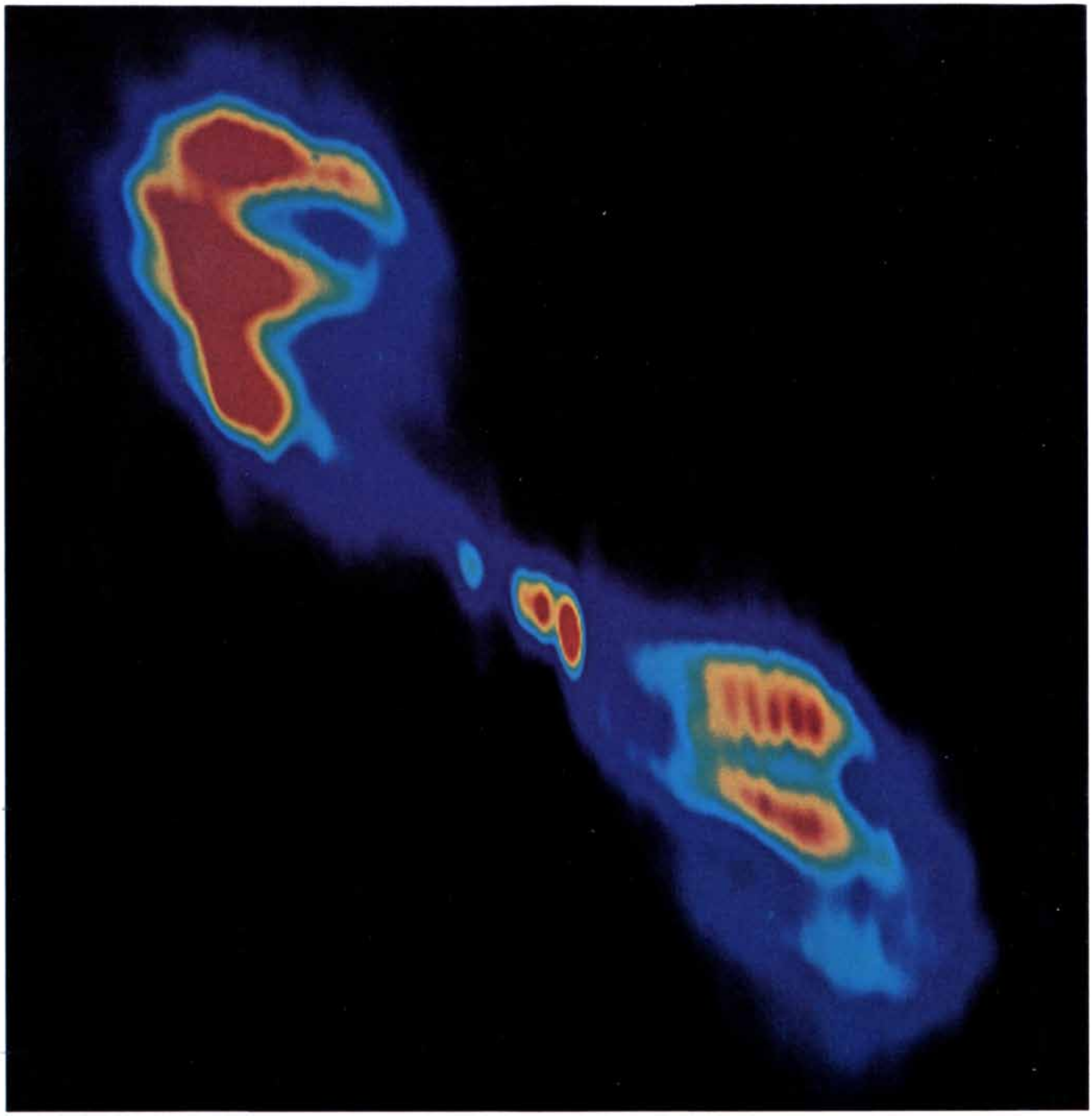
Centaurus A, which in observations at visible wavelengths is designated NGC 5128, was given its label because it is the most luminous source of radio waves in the constellation Centaurus. It occupies a position far south in the sky seen from the Northern Hemisphere. Seen from near Australia in the Southern Hemisphere, the constellation is overhead. Seen from New Mexico, where many of the recent radio observations have been made with the Very Large Array radio telescope (VLA) of the National Radio Astronomy Observatory, the galaxy is above the horizon only a few hours a day; its maximum elevation is about 14 degrees.

Centaurus A is one of the strongest sources of radio waves in the entire sky. The strength of the emission and the proximity of the object to our galaxy make it possible to observe processes and resolve structures in it that may be present in all active galaxies but that cannot be detected in more distant ones. As a result Centaurus A has had much attention from astronomers in the past few years. It has been studied with many of the most sensitive detectors of astronomical radiation, including in addition to the VLA the four-meter optical telescopes at the Cerro Tololo Inter-American Observatory in Chile and the Anglo-Australian Telescope at Siding Spring in Australia and the Einstein X-ray satellite observatory.

Much of the radiation detected in the work on Centaurus A originates in one of three physical processes. In all three processes radiation is emitted when an electron loses energy. The means by which the electron's energy changes, however, is different in each process, and the characteristics of the emitted radiation also differ. In the first process clouds of interstellar gas are energized by incident radiation. In such clouds atoms of gas absorb radiant energy and are thereby ionized (meaning that they gain or lose electrons). When an electron in the cloud comes near such an ion, the electron is deflected and accelerated; most of the ions are protons, hydrogen nuclei stripped of their single electron. As the electron's path shifts it emits a photon, or quantum of electromagnetic energy. Radiation generated in this way is referred to as thermal radiation because its spectrum depends strongly on the temperature of the gas.

Electromagnetic radiation can be described either as a stream of photons or as a train of waves. In the photon description the radiation can be characterized by the energy of the photons; the corresponding property in the wave description is the frequency, which in turn is directly related to the wavelength. The greater the energy of the photon, the higher the frequency of the waves and the shorter the wavelength. In thermal radiation the energies of the emitted photons or the frequencies of the emitted waves are distributed continuously, that is, they form a smooth curve when they are plotted on a graph. The radiation tends to be most intense, however, in a fairly small range of frequencies. The waves are randomly polarized, meaning that the plane of the electromagnetic oscillations can have any orientation.

In the second process an electron is captured and accelerated in a helical path around a line of force in a magnetic field. As the electron is accelerated it

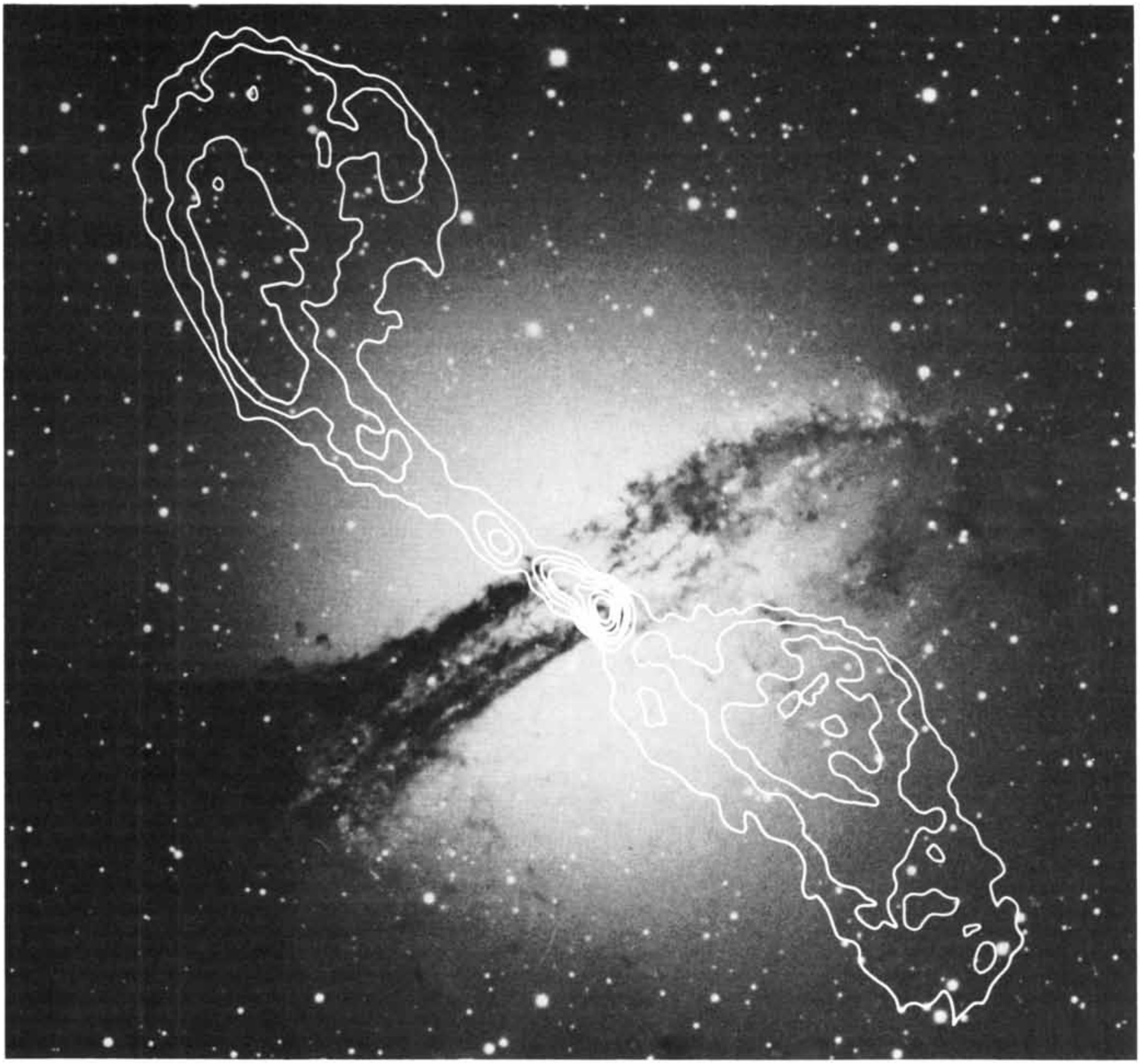


INNER RADIO LOBES of Centaurus A are shown on a false-color map of the radio emissions from the central region of the galaxy, which at visible wavelengths is designated NGC 5128. In the arbitrary color scheme red stands for the most intense radio emissions and blue for the least intense. Here and in other images of the galaxy on the following pages north is at the top and east is at the left. The inner lobes are the symmetrical rounded structures aligned on a northeast-southwest axis. The distance from tip to tip is about 60,000 light-years. The inner lobes are the smallest and innermost of a series of radio structures that extend from the center of the galaxy on each side. The narrow channel between the center of the galaxy and the northern inner lobe is called a jet. It is thought to be a beam of plasma, a highly ionized gas. The plasma is ejected from the galactic core, which is marked by the larger of the red ovals between the inner lobes. As in other active galaxies, the emissions from the plasma have a broad range of frequencies. The radio emissions were observed by one of the authors (Burns), Eric D. Feigelson and Ethan J. Schreier with the Very Large Array radio telescope (VLA) in New Mexico. The map was made by means of a computer image-processing system developed by Dennis Ghiglia of Sandia National Laboratories.

emits electromagnetic radiation tangent to the helical path. The emission constitutes what is called synchrotron radiation. Synchrotron radiation, like thermal radiation, has a continuous frequency distribution, but it is spread out over a broader range of frequencies than thermal radiation. Moreover, in synchrotron radiation the waves are polarized: the electric-field components of all the waves are aligned in the plane perpendicular to the magnetic line of force around which the electron spins.

In the third process a photon is emitted when an electron moves from one of the allowed energy levels in an atom to a lower level. When an atom in a cloud of gas absorbs energy from its environment, an electron in the atom can move to a higher energy level. Having ascended, the electron tends to quickly return to the original energy level. In the downward transition a photon with an energy equal to that of the photon absorbed in the upward transition is released. The photons emitted in this way have a set of discrete frequencies, each frequency corresponding to a particular transition within the atom. On a continuum representing the total electromagnetic spectrum the emitted radiation appears as a group of the bright lines known as atomic emission lines. The radiation in the emission lines is randomly polarized.

As we shall see, synchrotron radiation has turned out to be of great significance in understanding the physics of Centaurus A. The earliest observations of the galaxy, however, were made by means of radiation in the visible range, which is not surprising since NGC 5128 is the seventh most luminous object in the sky outside our own galaxy. From a wealth of optical observations beginning early in the 19th century it has become clear that the visible part of Centaurus A has a very unusual structure, with features of both elliptical and spiral galaxies.



OPTICAL IMAGE AND RADIO CONTOURS of Centaurus A are aligned perpendicular to each other. The illustration combines the radio contours of the inner lobes with a picture of the galaxy made with radiation in the visible range. The visible area of NGC 5128 has an unusual form with features of both elliptical and spiral galaxies. The large luminous circle is much like the body of an elliptical galaxy. The dark stripe that bisects the circle is a disk-shaped dust lane composed of gas, dust and associations of stars; the dust lane resembles a spiral galaxy. The disk is warped in opposite directions at its ends. Although the ellipse is rotating slowly if at all, the disk of the dust lane is rotating steadily. The radio structure emerges from the center of the dust lane more or less along its axis of rotation.

The main body of NGC 5128 is a bright ellipse about seven minutes of arc in angular extent. At the estimated distance of 15 million light-years each minute of arc is approximately 5,000 light-years, and hence the ellipse is about 35,000 light-years across at its widest. The mass of the galaxy has been estimated at about 300 billion times the mass of the sun. This mass and size are typical of elliptical systems.

What is unusual is the presence of a broad absorption lane, or lane of dust, in the middle of the ellipse. In optical images the dust lane appears as a dark stripe aligned with the short axis of the ellipse. The lane has the shape of a disk with its edges warped in opposite directions, like the brim of a fedora.

The dust lane is populated by dust, associations of stars and regions filled with ionized atomic hydrogen. The neutral, or un-ionized, hydrogen atom is designated H I and the ionized atom is designated H II. Therefore the regions of ionized hydrogen observed in the dust lane of NGC 5128 are referred to as H II regions. In the H II regions the hydrogen gas has a temperature of some 10,000 degrees Kelvin (degrees Celsius above absolute zero). The gas is heated by radiation from hot young blue stars that have recently formed in the gas cloud. The H II regions give off a set of characteristic atomic emission lines.

The main body of Centaurus A, like that of many elliptical systems, is rotating very slowly if at all. The dust lane, however, is rotating steadily and in a pattern typical of spiral or disk galaxies. The dust lane and the ellipse are centered at the same point, and the gravitational potential of the large ellipse dominates the system composed of the two components. Such a hybrid configuration is extremely rare and could well be connected with the violent activity underlying the radio structure.

Before turning to the radio and X-ray observations that are the center of current interest in Centaurus A, one other unusual optical feature is worth noting. Long-exposure photographic plates made by J. A. Graham of Cerro Tololo show there is a remarkable elongated luminous region in the northeast quadrant of the galaxy. The plates were made with the light of the atomic emission line corresponding to the recombination of an electron with a proton in the H II regions, which is referred to as the H-alpha transition. The plates show that at a distance of 50,000 light-years from the center of the galaxy there is a linear filament including three large H II regions aligned radially outward from the center of the galaxy. Farther out, at distances of up to 130,000 light-years, are regions yielding emission lines and including both filaments and diffuse gas. Spectroscopic studies show that the excitation of the gas decreases with distance from the galactic core, suggesting the source of the excited gas is in the center of the galaxy.

Near some of the elongated H II regions are what appear to be chains of bright blue young stars; the chains were first noted by Victor M. Blanco and his colleagues at Cerro Tololo. If these are indeed stars less than 10 million years old, they must have formed near where they are now observed rather than forming at the center of the galaxy and then being ejected. The stars are about 65,000 light-years from the galactic core, and even at the unusually high velocity of 1,000 kilometers per second (which is many times the velocity needed to escape the gravitational field of the main body of the galaxy) the stars could not have existed long enough to travel from the center of the galaxy to where they now are.

The first detailed radio observations of Centaurus A were begun in 1961 when the 210-foot radio telescope at the Australian National Radio Astronomy Observatory came into service. One of us (Price) was fortunate enough to work as a graduate student with J. G. Bolton and B. F. C. Cooper of that observatory in utilizing the new telescope, which was then the premier instrument in radio astronomy, to observe Centaurus A. It was quickly found that the radio emission from the galaxy is plane-polarized, suggesting it comes from the synchrotron process.

Further work showed that the plane of polarization of the emission is rotated as the radio waves pass through clouds of cosmic-ray electrons and the weak magnetic fields in the interstellar space of our galaxy. Such rotation, which is called the Faraday effect, is proportional to the square of the frequency of the radiation. Although the Faraday effect had been predicted theoretically, the observations of Centaurus A with the Australian radio telescope marked the first time the effect had been clearly detected observationally.

Thus the early radio investigation of Centaurus A had two significant results. First, the existence of the interstellar magnetic fields in our own galaxy was confirmed. Second, the confirmation of the hypothesis that the radio emissions from NGC 5128 come from a synchrotron process was a significant step in understanding the active galaxy.

Furthermore, the Australian radio observations provided the first reliable map of the overall radio structure of Centaurus A. It was found that the linear extent of the radio region is 2.7 million light-years, covering about 10 degrees in the sky. When Centaurus A was first mapped in the early 1960's, it was the largest discrete astronomical object that had ever been observed. In the past 10 years larger radio sources have been found, but Centaurus A is still among the largest astronomical objects known.

The extended radio-emitting regions on each side of the galactic center change their direction with the distance from the nucleus. Near the nucleus the alignment is northeast-southwest but the outer regions are aligned almost north and south. The shift in direction of about 65 degrees from the center of the galaxy to the tip of the radio envelope is caused by a continuous curvature along the length of the radio region. As a result the overall radio structure has the shape of a huge *S*. The curvature indicates that the flow of material into the radio envelope is being continuously perturbed. The perturbation could be due to a precession (a wobble like that of a top) of the nuclear engine or to the effect of gas in the galaxy outside the nucleus pressing on the plasma beam.

Between the galactic nucleus and the tip of the radio envelope is a series of complex radio structures. The size of these substructures increases from the nucleus outward, and the scale of the largest ones is about a million times that of the smallest. The details of the radio components have been resolved only recently in work with the VLA done by us, Eric D. Feigelson of Pennsylvania State University, Ethan J. Schreier of the Space Telescope Science Institute and George W. Clark of the Massachusetts Institute of Technology.

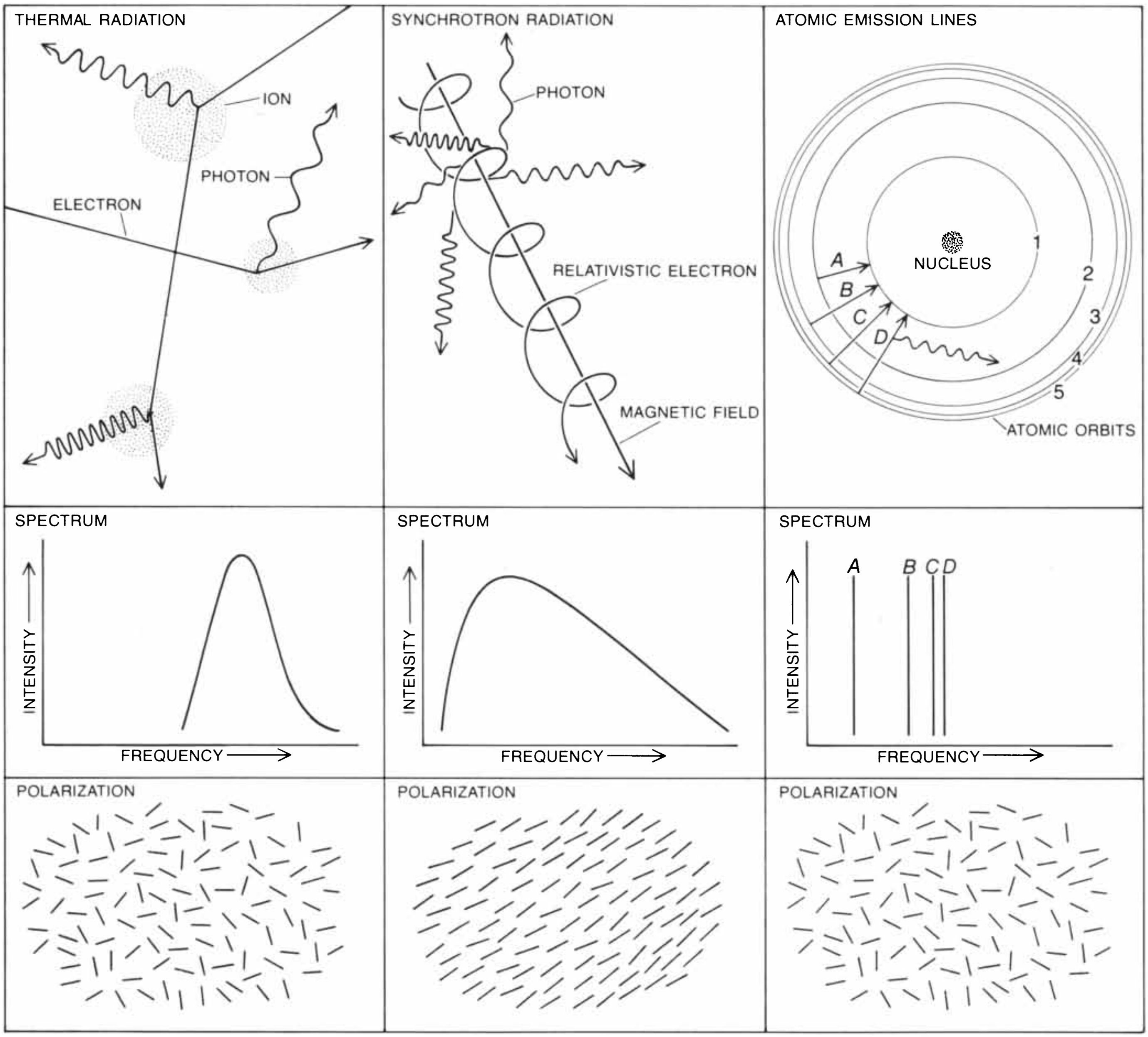
In a radio interferometer such as the VLA radio signals from a single source are received simultaneously by an array of several antennas. In the VLA there are 27 antennas arranged along three intersecting arms. Comparing the phase relations of the waves arriving at the receivers can give a rich picture of the radio source. With the VLA we have been able to achieve a resolution of one second of arc. At the distance of Centaurus A such a resolution is equal to about 80 light-years and is comparable to that in photographic plates made with visible radiation gathered by large telescopes.

Moving inward toward the core from the ends of the radio structure an asymmetry becomes clear. To the northeast, between 70,000 and 130,000 light-years from the nucleus, is a large, roughly circular middle radio lobe. No comparable middle lobe is seen at the corresponding position in the southwest. The middle lobe is a region of intense emission that is distinct from the diffuse outer radio envelope. The position of the middle lobe coincides with that of the optical filaments where young stars are seen. As we shall see, it also overlaps a newly discovered area of X-ray emission.

Closer to the nucleus is a pair of inner radio lobes, one lobe on each side of the core. The inner lobes, which are also distinct from the outer envelope, are symmetrically positioned 30,000 light-years from the center of the galaxy. With a diameter of 3,500 light-years they are considerably smaller than the middle lobe. No visible radiation or X-ray emission has yet been observed in association with the inner lobes.

Connecting the northern inner lobe and the nucleus is a narrowly collimated channel of radio and X-ray emission called a jet. Such jets, which are observed in many radio galaxies, could be channels that carry freshly energized electrons and magnetic fields from the central engine to the inner lobe. If this is so, the jet is a leaky and inefficient channel, since some of its energy is radiated outward as synchrotron emission.

Perhaps the most striking asymmetry in Centaurus A is the fact that only the northern part of the radio envelope has a jet; no comparable structure is observed in the southern part. Many radio galaxies have a jet extending from the galactic core, and in some of them the jet is one-sided. The reason for the asymmetry is not known, but several hypotheses have been proposed. Lawrence Rudnick of the University of Minnesota has suggested that periodically the jet switches sides. As a result over any long period the lobes would be supplied with roughly the same amount of energy. The oscillation could be due to the interaction of the jet with the dense gas in the region of the nucleus. Other workers have proposed that jets actually extend from both sides of the core but that because of relativistic effects the emission from the jet coming toward us is greatly enhanced and the emission from the jet going away from us is greatly reduced



RADIATION EMITTED BY CENTAURUS A could originate in one of three physical processes. In all three an electron loses energy, and the mechanism by which the energy is dissipated determines the frequency and intensity of the radiation. In thermal radiation (*left*) electrons are deflected and accelerated as they pass near ions (mostly hydrogen nuclei, or protons) in a cloud of hot gas. When an electron is accelerated, it gives off a photon, or quantum of radiation, which can be said to have an energy, a frequency and a wavelength. The frequencies of thermal radiation are continuously distributed, but the radiation is most intense in a fairly narrow range. The polarization of thermal radiation is random. In synchrotron radiation (*middle*) photons are given off by relativistic electrons: electrons with a velocity approaching that of light. The relativistic electrons are accelerated in a helical path around lines of force in a magnetic field and photons are emitted tangent to the path. Synchrotron radiation has a broad range of frequencies and is polarized. An electron can also give off radiation when it makes a downward transition from one allowed energy level in an atom to a lower level (*right*). The spectrum of radiation that is emitted in this way consists of a series of discrete lines. The radiation in these atomic emission lines is randomly polarized.

[see "Cosmic Jets," by Roger D. Blandford, Mitchell C. Begelman and Martin J. Rees; SCIENTIFIC AMERICAN, May, 1982]. This explanation, however, requires that the bulk speed of the plasma be close to that of light, which appears to be unlikely in Centaurus A.

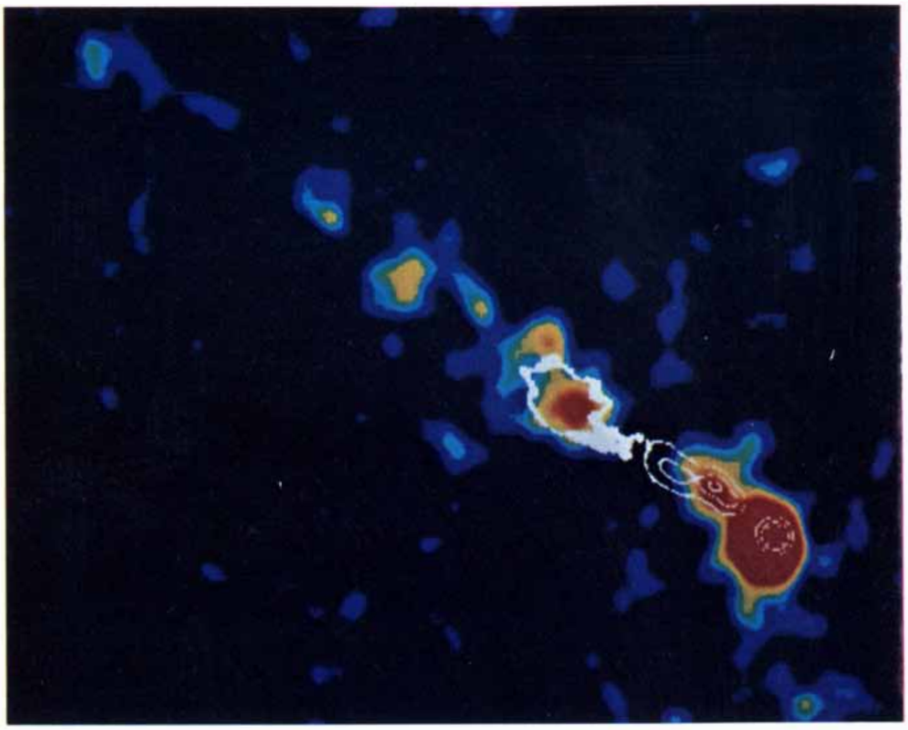
In radio interferometry the resolution of the observations increases when the base line, or the distance between antennas, is increased. The resolution that can be achieved with a radio interferometer has recently been improved by putting antennas on different continents. The data from each instrument are recorded on a magnetic tape and are later correlated and compared by computer. Robert A. Preston and his colleagues at the Jet Propulsion Laboratory of the California Institute of Technology employed radio telescopes in Australia and South Africa to observe Centaurus A and were able to resolve features as small as three milliseconds of arc in the nucleus of the galaxy.

Deep in the core of Centaurus A, Preston and his colleagues found a jet about 50 milliseconds of arc long. At the distance of Centaurus A, 50 milliseconds of arc is equal to four light-years. The small jet emerges at the same angle as the larger radio jet found with the VLA. This remarkable collimation of the beam from four light-years out to 20,000 light-years means that the engine in the core maintains a consistent spatial orientation over long periods. The small inner jet is very close to the engine and is probably the initial point where the plasma beam becomes collimated.

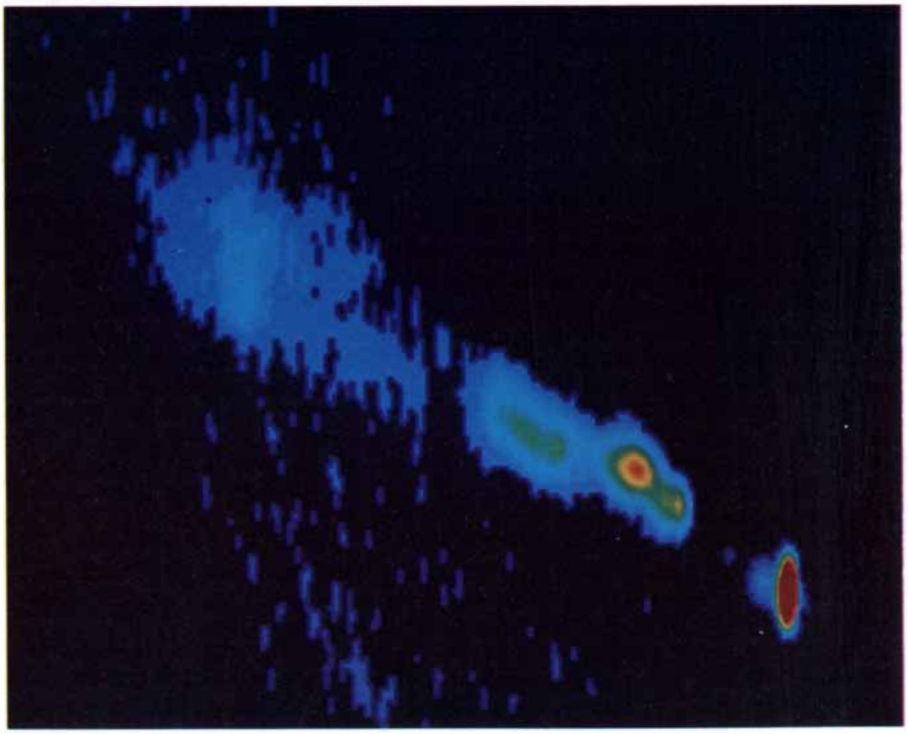
Completing the radio map of the galaxy is the core itself. In radio observations the core appears as a strong source of emissions at the very center of the galaxy; it cannot be resolved into its components with current instruments. While the detailed structure of Centaurus A was being examined by means of radio interferometers another significant kind of information about the galaxy was being accumulated.

The X-ray telescope in the Einstein observatory satellite, launched in November, 1978, was the first instrument capable of detecting and mapping soft X rays from sources other than the sun. Within the X-ray part of the electromagnetic spectrum soft X rays are those with a relatively low energy: from about 1,000 to 10,000 electron volts. Hard X rays correspond to photons with a high energy: from about 10,000 to 50,000 electron volts. The two main detectors mounted on the Einstein satellite were able to resolve X-ray features respectively 1.5 minutes of arc and five seconds of arc across, for the first time giving X-ray astronomers a resolving power comparable to that of radio telescopes and optical telescopes.

Soon after the Einstein satellite was



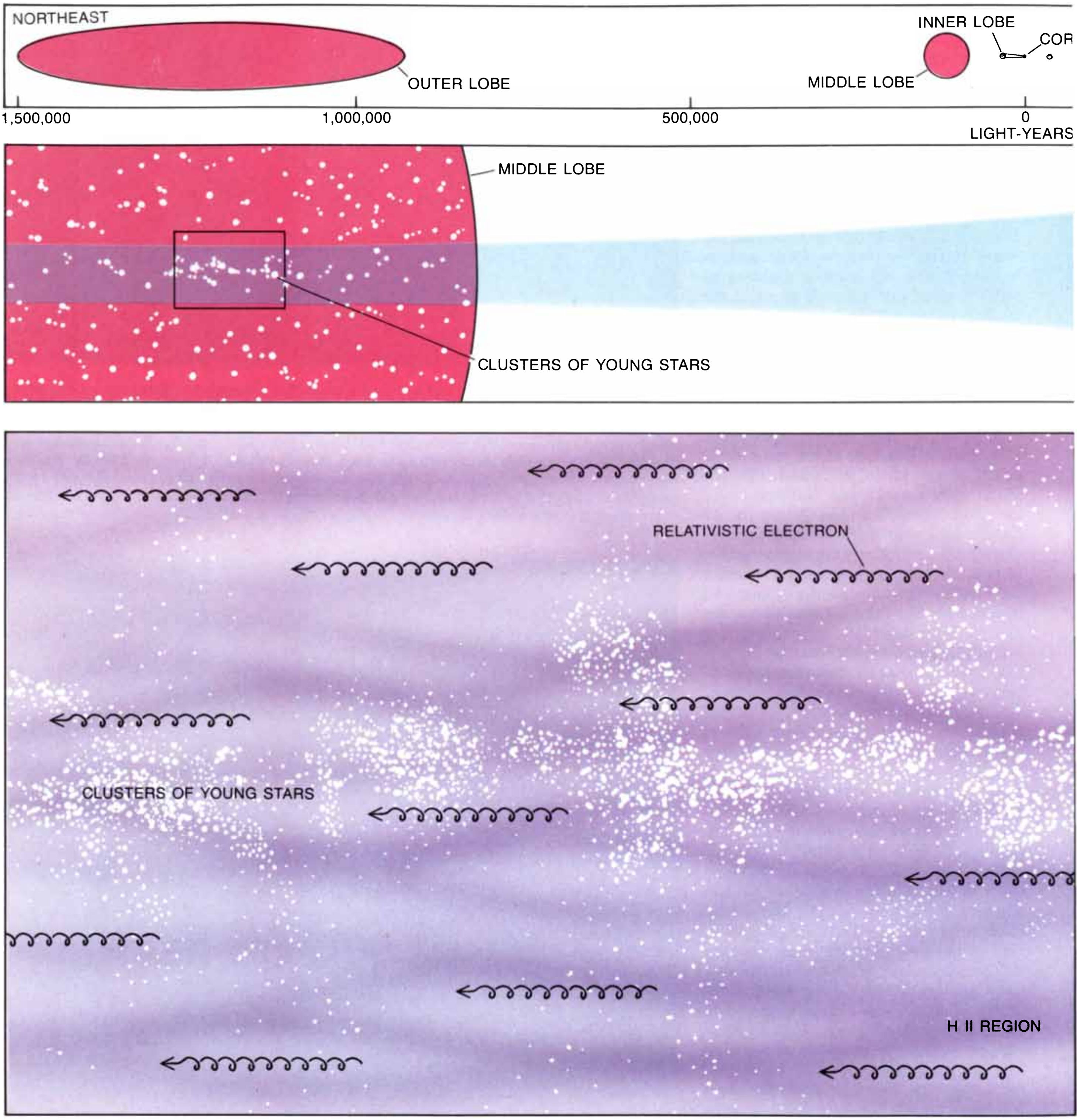
INNER JET is a highly collimated beam of plasma that channels freshly energized electrons from the core of Centaurus A to the northern inner lobe. The galactic nucleus is the compact red region at the lower right with white contours superposed on it; the jet extends to the northeast from the core. The jet is about 3,000 light-years long. The computer image combines radio data with X-ray data gathered by Feigelson and his colleagues using the Einstein satellite observatory. The colored areas correspond to X-ray emission, the white contours to radio emission. The conjunction of X-ray and radio emissions suggests that the radiation from the plasma comes from a synchrotron process. No corresponding jet is seen to the southwest of the core.



HIGH-RESOLUTION IMAGE of the inner jet is a detail of the illustration at the top of the page: it expands a small area near the center of the galaxy. The galactic core is the red oval at the lower right. The colored areas to the northeast of the core are "knots," regions of particularly intense radio emission. Since the relativistic electrons in the plasma can retain their high energy for only about 50 years, they must be reaccelerated in order to reach the inner lobe, which is 3,000 light-years from the core. The knots could be turbulent regions in the jet where the electrons are reenergized. The image was made utilizing the VLA in a configuration that makes it possible to resolve features as small as 80 light-years across at the distance of Centaurus A.

launched Feigelson and Schreier, together with Riccardo Giacconi of the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory, directed the satellite's instruments at Centaurus A and found a wealth of detail unprecedented in extragalactic X-ray astronomy. At about the same distance and angle from the nucleus as the northern middle radio lobe is a region of X-ray emission 24,000 light-years long. The radiation is more intense than would be expected from synchrotron emission. Therefore it is probably thermal radiation; its source could be a cloud of gas with a temperature of 10 million degrees K. The gas might be heated when it is compressed by turbulence in the radio plasma of the middle lobe.

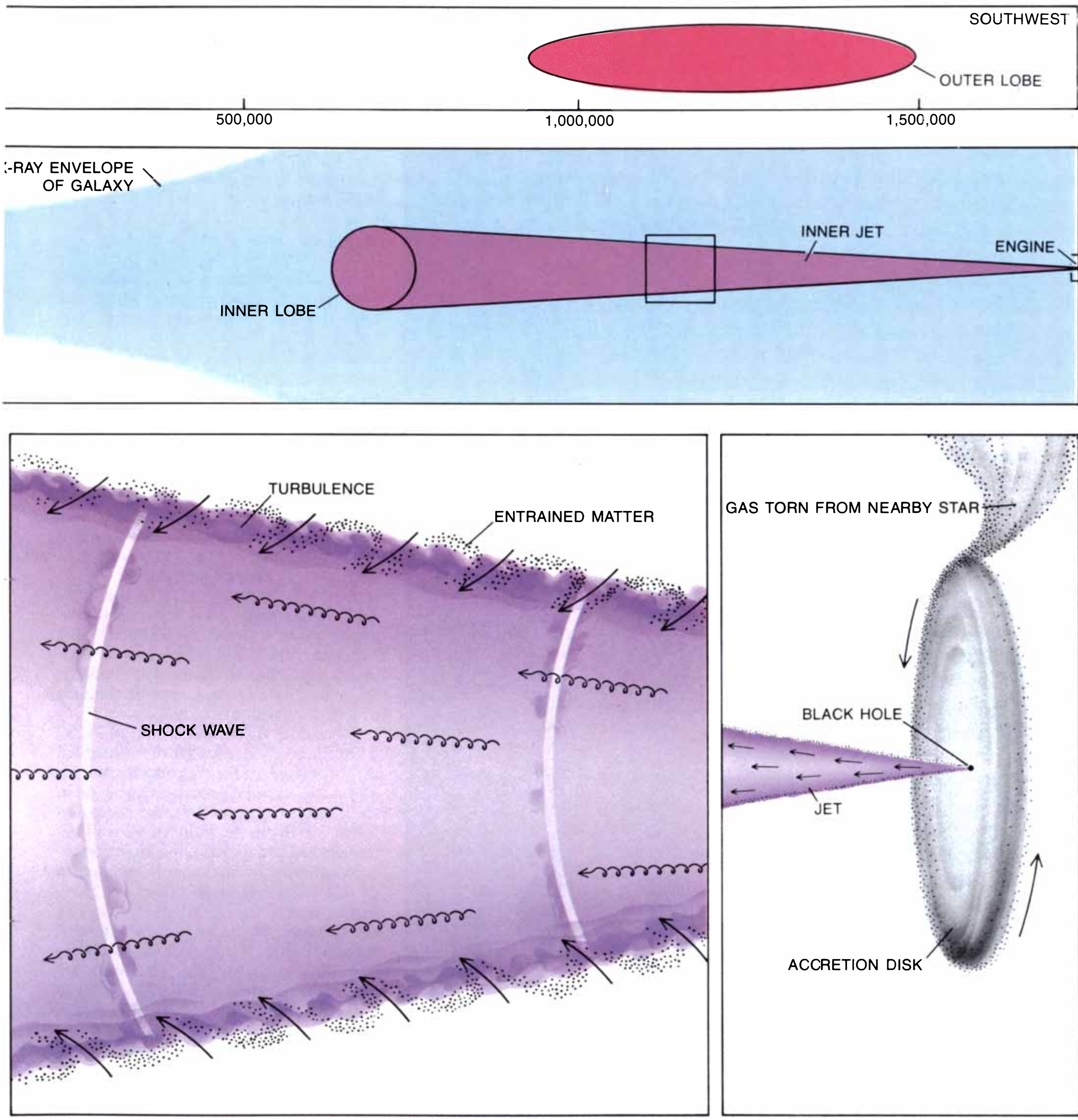
An envelope of less intense X-ray



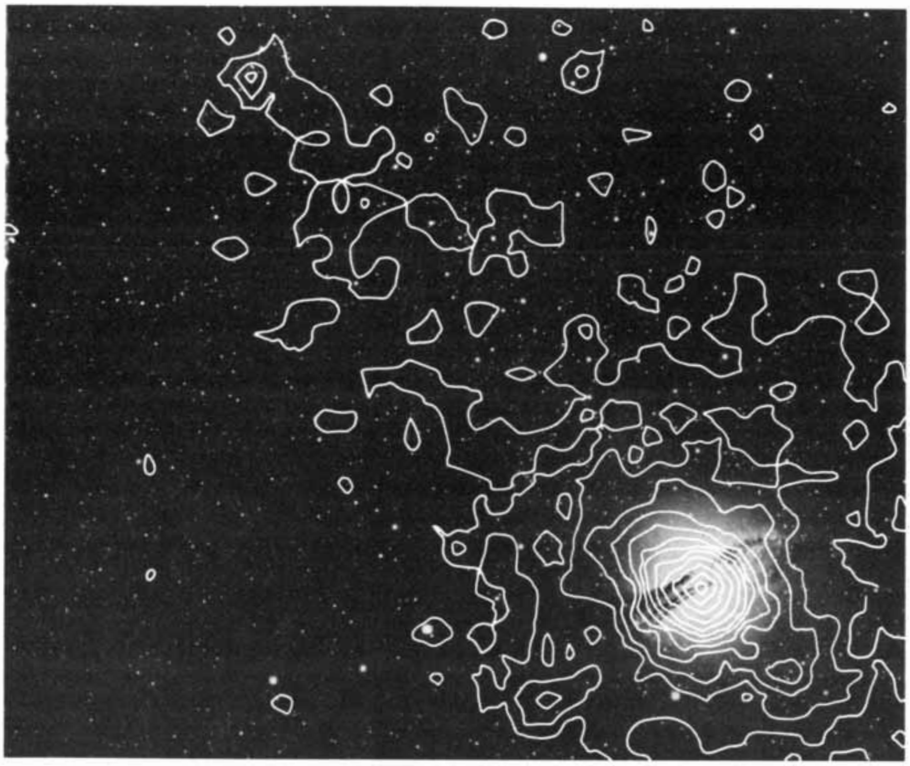
DYNAMICS OF THE EMITTING REGION of Centaurus A are shown schematically in panels with a progressively smaller scale. The full extent of the emitting region is three million light-years (*top panel*). Radio emissions are shown in red and X-ray emissions in blue. The pair of inner lobes on each side of the center of the galaxy are areas of relatively intense radio emission within a diffuse outer emitting envelope. The inner jet that connects the core to the northern inner lobe gives off much emission at both radio and X-ray wavelengths (*middle panel*). It lies within a larger region of diffuse X-ray emission that comes from the interstellar medium. The region of diffuse X-ray emission extends to the northern middle lobe. One edge of the middle lobe is also the site of a group of extended filaments and bright blue point sources of visible radiation. One hypothesis that accounts for many recent radio and X-ray observations of Centaurus

emission surrounds the galaxy, extending out to a distance of about 10,000 light-years from the core. The diffuse emission is probably generated by a cloud of interstellar gas with a temperature of 20 million degrees K. and a total mass 200 million times that of the sun. The detection of the hot gas by the instruments on board the Einstein satellite is one of the few times an interstellar medium has been found within the visible body of an elliptical galaxy.

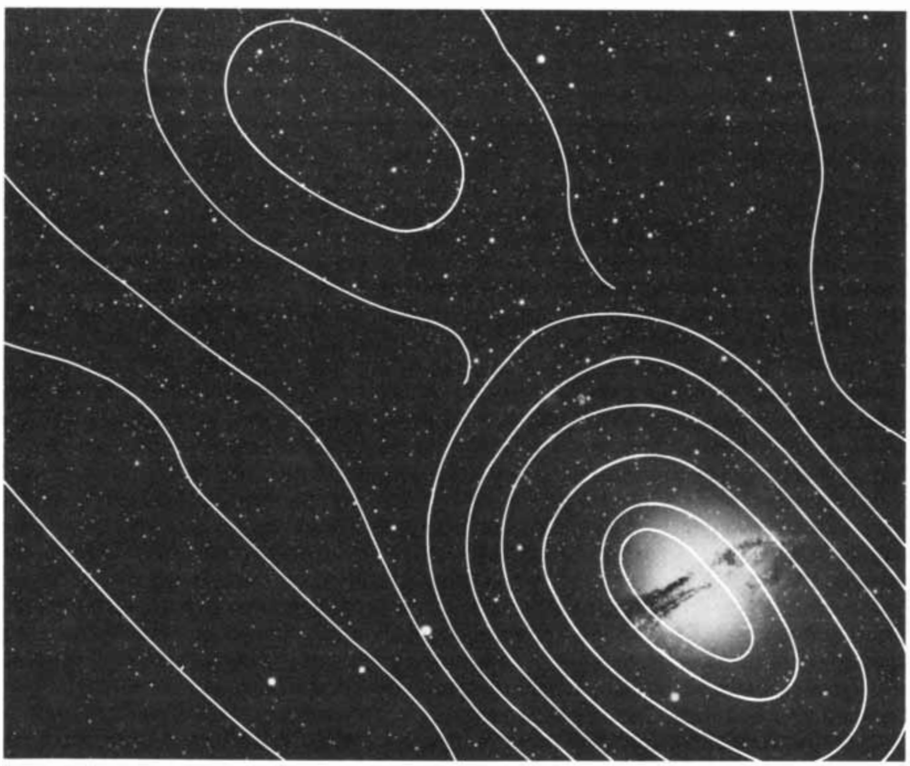
With a density of about one particle per 300 cubic centimeters the interstellar medium discovered in the Einstein satellite observations is fairly dense by astronomical standards. The dense medium could confine the plasma in the radio jet and the inner radio lobe and thereby contribute to the maintaining of their shape. Such confinement could be



A is the following: The engine that generates the radio plasma has at its center a black hole with the mass of a billion suns (*bottom right*). The black hole is surrounded by a toroidal accretion disk made up of gas and dust. Fuel for the engine comes from infalling gas, which could be tidally torn from the atmosphere of stars near the disk. The interaction of the gas with the black hole and the accretion disk yields a narrowly collimated beam of electrons and magnetic fields: the inner jet. As the jet moves outward from the core, turbulence at the edge of the jet causes gas to be picked up and carried along with the plasma (*bottom middle*). Within the jet shock waves reenergize the relativistic electrons. When the plasma reaches the region of the middle lobe, some of the entrained gas cools and contracts and forms stars (*bottom left*). At the end of their life they explode as supernovas, putting energy back into the radio plasma and energizing the electrons.



X-RAY CONTOURS mapped with the Einstein satellite observatory are superposed on an image of NGC 5128 made by means of visible radiation. The main body of the galaxy is at the lower right. The optical filaments and young stars near the northern middle lobe are at the upper left; they are 130,000 light-years from the core. The chief X-ray emissions outside the main body of the galaxy coincide with the optical filaments. The optical image is made by the emission line resulting from the recombination of an electron and a proton to form a hydrogen atom.



RADIO CONTOURS made with the 210-foot radio telescope at the Australian Radio Astronomy Observatory are superposed on the same optical image of NGC 5128 in the illustration at the top of the page. As in the case of the X-ray emissions, the region of the most intense radio emission outside the main body of the galaxy (the northern middle radio lobe) coincides with the optical filaments and young stars. The overlapping of the sources of radiation at X-ray, visible and radio frequencies suggests that the same process could be responsible for all three.

critical. Without it the radio jet might lose its collimation and the lobes would expand greatly, diluting their energy density. If the hypothesis is correct, the structure of NGC 5128 confirms the thermal confinement of radio structures, a phenomenon that had been predicted but had never been observed.

The most exciting finding made with the Einstein satellite, however, was the discovery of a jet of X-ray emission coming from the core in much the same position as the VLA radio jet. This was the first time such an intense elongated structure had been found in X-ray data. Indeed, the X-ray jet was discovered before the VLA jet, and the X-ray findings motivated much of the recent work at radio frequencies.

Not only is the X-ray jet the same size and shape as the radio jet but also at the same places in the two jets there are "knots": small regions where the emissions are particularly intense. The structural similarities, along with the continuous spectrum of the emissions and their polarization properties, led us to conclude that the X-ray emission from Centaurus A is also synchrotron radiation.

When the data from more than two years of operating the Einstein satellite were analyzed, only one other galactic X-ray jet like the one in NGC 5128 was found. The second jet is in the next-closest active galaxy: M87. This galaxy is about three times as far from us as Centaurus A is. The resolution of the satellite's instruments, however, is not sufficient to show structural details in the M87 jet. The X-ray jet in Centaurus A has details that can be observed individually, and it therefore currently remains a unique astronomical object.

What does the detailed map of NGC 5128 that has been patiently assembled at X-ray and radio frequencies tell us about the physical processes going on in the galaxy? Consider the inner radio and X-ray jet. From the point of view of astronomical observation the most significant component of the jet is the stream of relativistic electrons (electrons moving at nearly the speed of light). The electrons are the source of the synchrotron radiation. It does not follow, however, that because the electrons are moving at relativistic velocities the jet as a whole has such a velocity. The beam includes a considerable quantity of gas in addition to the electrons, and the gas has a much lower speed than the electrons do. The best current estimate is that the bulk-flow velocity of the jet is about 5,000 kilometers per second.

The conclusion that the X-ray photons detected by the Einstein satellite come from a synchrotron process puts some severe constraints on the model of what goes on in the jet. The X-ray photons have an energy of about 2,000 electron volts. In the synchrotron process such photons are emitted by relativistic electrons with an energy of 3×10^{13} electron volts. Electrons with that much energy cannot stay in a magnetic field for long before they emit photons.

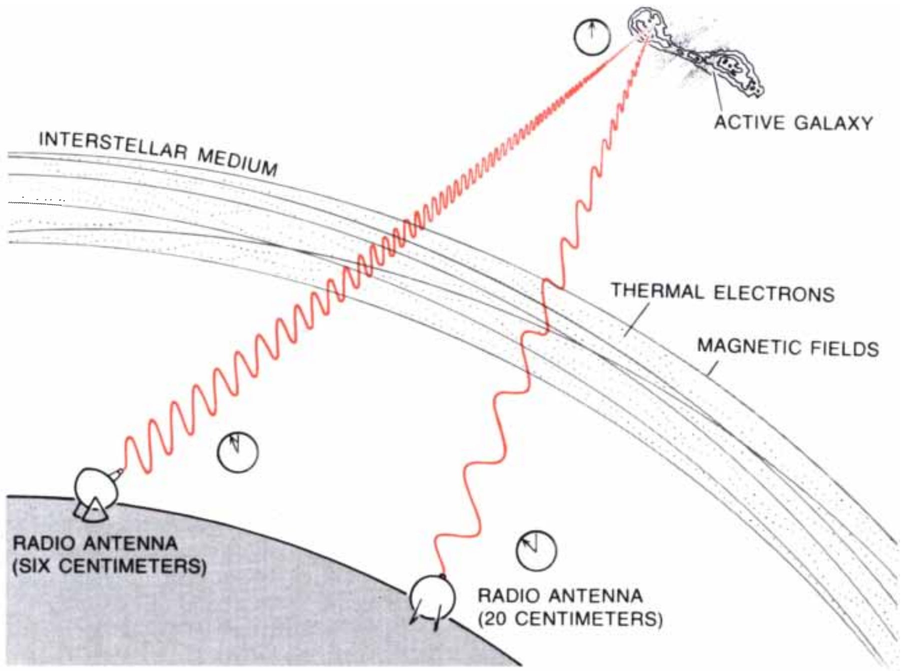
If the emission continues, eventually the electron will be depleted of energy and will not be able to emit any more photons. Indeed, given the energies of the electrons and the X-ray photons the depletion will occur in 50 years or less. The radio and X-ray jet in Centaurus A, however, reaches the inner lobe, which is 20,000 light-years from the nucleus. Therefore if the electrons are traveling at about the speed of light, they will have to be reenergized many times in order to reach the inner lobe.

The relativistic electrons in the beam could be accelerated by any one of three mechanisms. The first mechanism entails the existence of shock waves in the jet much like the ones formed in the earth's atmosphere when an airplane is flying at a velocity greater than that of sound. Such shocks could result from the collision of the jet with interstellar clouds. They could also form internally if the beam is unsteady or unstable. The knots in the beam would then correspond to regions where the plasma is subjected to strong shocks.

In the second mechanism the energy of turbulence in the beam provides the energy to accelerate the electrons. Instabilities and turbulence can form on the boundary layer between a beam of plasma and the medium that surrounds it. The turbulence could propagate into the beam in the form of waves whose motion would be damped by interaction with the electrons. In interacting with the waves the electrons would be accelerated. The net result of the process would be to transfer the bulk-flow energy of the plasma into the kinetic energy of the waves and then into the acceleration of the relativistic electrons.

The third possibility is that the electrons could be reenergized by collisions with the protons in the beam. The protons in the plasma cannot be detected directly because they do not emit a detectable quantity of radiation. For the same reason, however, they could retain their high kinetic energy much longer than the electrons do. Collisions between protons and electrons would eventually deplete the reservoir of proton energy, but the process would be slow and the reservoir might be large enough to get the plasma all the way to the inner radio lobe.

The process by which the beam of electrons, protons and magnetic fields gets from the inner lobe to the middle lobe is not yet well understood. However it operates, when the plasma reaches the middle lobe, it could have an important role in the formation of the young stars and the emission-line regions observed at distances of about 130,000 light-years from the core.



FARADAY EFFECT rotates the plane of polarization of radiation traveling through our galaxy. Radiation from an active galaxy such as Centaurus A is often plane-polarized because it originates in the synchrotron process. As the radiation passes through the interstellar medium in our galaxy it interacts with thermal electrons and magnetic fields. As a result the plane of polarization of the waves is rotated. The degree of rotation is proportional to the wavelength: the longer the wavelength, the greater the rotation. Radio observations of Centaurus A that were made in the early 1960's provided the first direct confirmation of the Faraday effect in the Milky Way galaxy. The observations also helped to strengthen the conclusion that the broad-spectrum radiation emitted by Centaurus A has its origin in a synchrotron process.

David S. De Young of the Kitt Peak National Observatory proposes the following scenario for the interaction of the radio structures and the optical structures. Eddies in the turbulent layer at the edge of the jet can entrain nearby material, which is mainly gas and dust. As the material is entrained it is heated by shocks in the plasma beam to about 10 million degrees K.

If the elements in the gas surrounding the jet have the same abundances as they do in the sun, which is a reasonable assumption for an evolved galaxy, the gas will cool to 10,000 degrees K. in from 10 million to 100 million years. At a temperature of 10,000 degrees the gas would show optical emission lines like those seen in the northeast quadrant of NGC 5128. Moreover, in the time it took the gas to cool it would have traveled between 33,000 and 330,000 light-years, a range that could put it near the position of the optical emission seen in NGC 5128 and in several other radio galaxies where optical emission lines coincide with radio lobes.

A fraction of the entrained gas would continue to cool and would eventually reach temperatures much lower than 10,000 degrees. The gas would then begin to coalesce by gravitational attraction and stars would form in it. De Young holds that the clusters of optical emission observed in the northeast quadrant of NGC 5128 support the conclusion that stars are forming in the region of the middle lobe.

The plasma contributes to the formation of stars by entraining ambient matter, but the stars could also help to keep the plasma beam energized as it moves outward from the center of the galaxy. Some of the stars formed in the northeast quadrant would probably have masses greater than 10 times the mass of the sun. Such massive stars have a quite short lifetime: about 10 million years. At the end of their relatively short life the massive stars would explode as supernovas.

The effect of the explosion would be to reenergize the electrons in the beam. Therefore the flow in the northeast quadrant could be self-perpetuating: the plasma beam gives rise to the formation of giant stars, which in turn explode and return their energy to the beam.

This account of how the plasma beam could be energized and how the optical filaments form is based on substantial observational data. Two of the most fundamental questions about Centaurus A and other active galaxies, however, must be answered in large part on the basis of conjecture. The questions are: What is the basis of the galactic activity? How does the central engine work?

It is possible that the activity observed in NGC 5128 is directly related to the galaxy's unusual structure, which combines an ellipse with a disk. Some 30 years ago Walter Baade and Rudolph Minkowski of the Mount Wilson and Palomar Observatories hypothesized that the form of NGC 5128 is the result of a merger between a spiral galaxy and an elliptical one. Recent theoretical work by Allan D. Tubbs of the National Radio Astronomy Observatory has shown that gas and dust tidally torn from a nearby galaxy can fall into an elliptical galaxy and settle in a shape much like that of the dust lane in NGC 5128.

Such a merging process has been called galactic cannibalism. Cannibalism has been invoked to explain some powerful radio sources other than Centaurus A. Certain clusters of galaxies that include a giant galaxy at their center are five times as likely to be a strong radio source as the average cluster of galaxies. It has been proposed that the giant central galaxy is formed by repeated gravitational encounters deep in the cluster. The encounters generate a frictional force that causes the larger galaxies to spiral slowly inward toward the center of the cluster. Over the lifetime of the cluster the most massive galaxies continue to coalesce, ultimately forming a single massive central unit. The central giant galaxy then cannibalizes the smaller galaxies nearby.

The cause of the radio activity of such giant galaxies is almost certainly the matter accumulated in the process of cannibalism. The accreting gas, dust and stars provide the fuel for the radio engine at the core of the galaxy. A more limited cannibalism in NGC 5128 could also have triggered violent activity in the galactic nucleus and the corresponding radio activity. One problem with the cannibalism hypothesis is that Centaurus A is quite isolated: its only companions within a million light-years are a few dwarf galaxies. Perhaps the galaxy has cannibalized all its neighbors and the nonthermal radio and X-ray emissions are signs of cosmic indigestion in the aftermath of a hearty meal.

How might the central engine utilize the fuel the cannibalism provides? The fuel could probably be consumed most efficiently in the form of a cloud of gas. Observing with the VLA, Jan M. van der Hulst of the Westerbork Observatory in the Netherlands recently found clouds of neutral hydrogen gas within about 500 light-years of the core in Centaurus A. The pattern of the clouds' motion suggests that they are falling toward the nucleus. Such clouds could each year provide the engine with fuel equivalent to at least a tenth of the mass of the sun. If we assume that the engine has an efficiency of 10 percent in converting mass into energy, a tenth of a solar mass per year would be more than enough to account for the output of the engine, namely the extended radio region.

Fuel for the engine might also come from the atmosphere of stars in the central region of the galaxy. Jack G. Hills of the Los Alamos National Laboratory and others have suggested that stars near the core are subject to large tidal forces, implying that the strength of the gravitational field varies considerably from one side of the star to the other. As a result the star could be broken up or at least stripped of its atmosphere.

Hence matter is available as fuel for the engine at the core of the galaxy. The next problem is to find out how the clouds of gas, which would tend to circle the nucleus, are drawn into it. One way would be for two clouds of gas to collide in the vicinity of the nucleus. In such a collision one cloud might lose angular momentum and fall into the nucleus while the other gained momentum and moved outward.

Having accounted for the fuel and suggested a feeder mechanism whereby the fuel could be delivered to the engine, it is of great interest to consider what is going on in the engine itself. To do so it is useful to have an estimate of the size of the machine. The size of a source of radiation can be estimated by measuring the period over which the intensity of the emission varies. The reasoning underlying such an estimate is as follows. For two widely separated regions of the source to turn on or off simultaneously they must have some means of communication: a physical signal must pass between them. Since the fastest such a signal can propagate is the speed of light, it follows that if a significant fraction of the emission varies in, say, two hours, the source itself cannot be more than two light-hours across.

Although this reasoning is not incontrovertible, it is probably valid for a source with a continuous emitting surface, and we have employed it to estimate the size of the central engine of NGC 5128. Observations of the variability of the X-ray and radio fluxes from the core of Centaurus A show that the emissions vary significantly over a period of months. Faster variations, however, are also observed, with the fastest taking less than 24 hours. Therefore we conclude that the core has components ranging in size from light-months to light-days.

Although any account of what actually goes on in the substructures of the galactic engine is highly speculative, one point seems clear. Whatever the engine's mechanism is, it is not nucleosynthesis, the process of nuclear fusion that supplies the power for most common stars. The assumption of an efficiency of 10 percent for the central engine rules out nucleosynthesis, which has an efficiency of less than 1 percent and would thus require more fuel and more time to generate the radio structure of Centaurus A than is suggested by recent observations. Moreover, nucleosynthesis yields thermal radiation rather than the nonthermal synchrotron emissions observed in NGC 5128.

A more efficient way for energy to be extracted from matter is for the matter to fall into a strong gravitational field. When that happens, the matter gains considerable kinetic energy. If the energized fuel then collides with structures in the core, energy can be released in the form of high-energy electromagnetic waves and even high-speed particles.

For this to be the mechanism of the central engine of Centaurus A there would have to be a source of a very strong gravitational field at the center of the galaxy. Such a field could be provided by what is referred to as a collapsed object: a black hole with a mass of about a billion solar masses. If the collapsed object exists at the center of the galaxy, it is undoubtedly spinning, since it is difficult to imagine a process that would lead to the creation of a black hole without at the same time giving it considerable angular momentum.

The spinning collapsed object would have three main effects. First, it would provide the well of gravitational potential into which the fuel could fall. Second, its spin axis would orient the entire engine. The spin would cause nearby matter to precess around the engine and form the disk of gas known as an accretion disk. Third, the black hole could have a magnetic field associated with it, and the lines of force in the field could accelerate charged particles to a high energy, extracting energy from the black hole as they did so.

Imagine a cloud of relatively cold gas falling onto the accretion disk. The disk might have the shape of a torus with the black hole in the center. Some of the infalling gas would be collimated within the narrow confines of the black hole and driven back out along the rotation axis of the hole by the radiation pressure of emission from the inner surface of the accretion disk. In this way the disk could both collimate and accelerate the particle beam that forms the inner radio jet.

For the moment this is as far as even the boldest speculation can take us in understanding active galaxies such as Centaurus A. The galaxy remains the focus of intense scientific activity. Because of its proximity and its intrinsic interest, it will undoubtedly continue to be so over the next few years. The current work has amassed much information on the structure and physical processes of Centaurus A. The most intriguing aspect of an active galaxy, however, is the mechanism that underlies its huge emitting region. We believe the study of Centaurus A will help to solve this fundamental problem in astrophysics.


Ford
Created for the State-of-the Artist.
POWER
101.9 FM
VOL
BASS
TREB
SPKRS
SEEK
AM FM
MEMORY
SCAN
AUTO-REVERSE
REV
EJECT
DOLBY NR
CrO2
Ford Electronic
Beethoven was the classic performer of his time. Today, it's the Ford Electronic Premium Sound System. Designed, engineered and quality tested by Ford, specifically for Ford and Lincoln-Mercury automobiles. This highly sophisticated stereo system will faithfully reproduce the musical performance of the most demanding artist. From Beethoven to Bach. From Bach to Rock. And everything in between. The Ford Electronic Premium Sound System. The classic performer in a full line of high performance sound systems by Ford.
Available at your Ford and Lincoln-Mercury dealers today.
Beethoven never heard it so good.
Ford Electrical and Electronics Division

PHOTO CREDITS Kurt Vonnegut Jr. by Jill Krementz

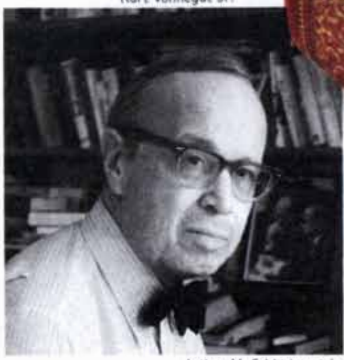
Walker Percy

Gore Vidal

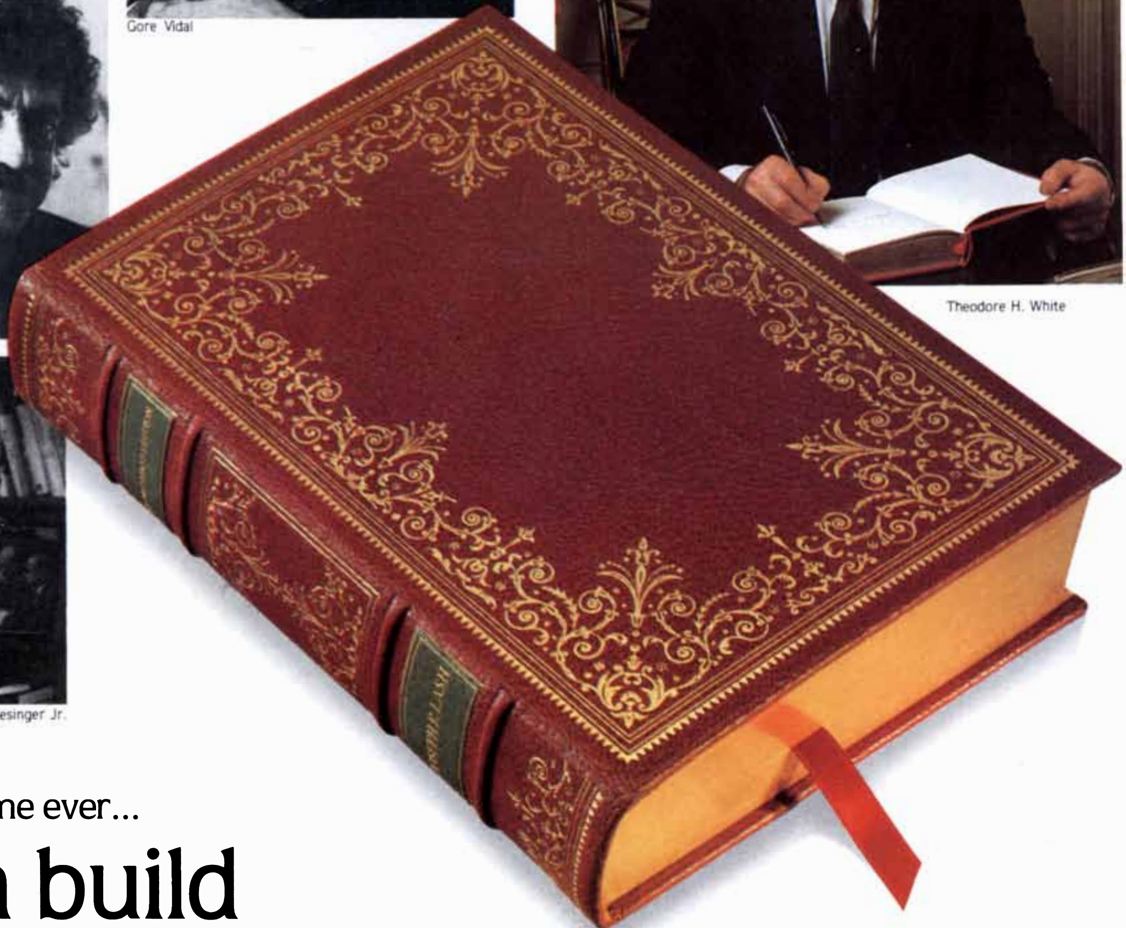
Theodore H. White

William Styron

Kurt Vonnegut Jr.



Arthur M. Schlesinger Jr.



Now, for the first time ever...

You can build a private library of author-signed First Editions.

Important new books by today's great writers, in First Editions that may well become tomorrow's collecting treasures.

- **Published in limited edition, solely for members of The Signed First Edition Society.**
- **Charter Members accepted for a limited time only.**

Applications must be postmarked by December 31, 1983.

You are invited to enjoy the exclusive privilege of building a private library unlike any that has existed before. A unique collection of leather-bound First Editions of today's major new books. Each book signed by its author. And each volume a limited First Edition that may well become a book collector's "find" of tomorrow.

By arrangement with America's leading publishing houses, each of these Signed First Editions will mark the very first publication of an outstanding new work. To document this fact for future generations, every volume will contain a statement of authenticity attesting to its First Edition status and certifying the author's signature.

Furthermore, these volumes will be privately printed, the spines richly embellished with 22 karat gold, and published solely for members of The Signed First Edition Society.

By enrolling now—as a charter member—you can build this extraordinary library from the beginning. And, for as long as you wish to continue in the Society, you will be

© 1983 SFES

among a select number of people guaranteed the opportunity to own every one of these Signed First Editions.

Major works by great authors

The Society has already arranged to obtain and review new manuscripts by such prominent authors as Allen Drury (Advise and Consent), John Hersey (The Wall and A Bell for Adano), William Styron (The Confessions of Nat Turner and Sophie's Choice), Wallace Stegner (The Spectator Bird), Leon Uris (Exodus, Armageddon, and Trinity).

Indeed, the authors to be published by the Society in the coming months have accounted for eight Pulitzers and six National Book Awards. There will be a searing new novel by Philip Roth. A fascinating historical study by Louis Auchincloss. A biographical novel by Irving Stone. And eagerly awaited new books by such best-selling authors as Irwin Shaw, Joyce Carol Oates, Walker Percy, Gore Vidal, Arthur Schlesinger Jr., Robert Penn Warren and Kurt Vonnegut Jr.

The First Editions of all of these new books will be reserved exclusively for members of the Society.

The value of owning author-signed First Editions

Among knowledgeable book collectors, First Editions have traditionally been held in highest esteem. Especially prized are First Editions that have been signed by their authors. For then the book—like a signed work of art—serves as a direct link between the owner and the writer.

Accordingly, as each new book is prepared for publication, every aspect of the volume—from the marbled endleaves of the book to the unique design of its leather binding—will be carried out to reflect the spirit and intent of the work. And the author's hand-written signature in the book will be a final mark of approval.

In addition, each author will write a special foreword that will appear only in the Signed First Edition issued by the Society—adding to the importance of owning these superb volumes.

As a charter member, therefore, you will possess perhaps the most desirable books of all: important new First Editions, bound in leather, and signed for you by the greatest authors of today. A private library to be treasured throughout a lifetime—and by generations to come.

Leather accented with 22kt gold

Every one of these Signed First Editions will be created to enhance the collector's library and the reader's enjoyment.

Each volume will be fully bound in carefully selected hand-cut leather, deeply stamped with an original cover design. The leather of the spine will be "hubbed" with raised horizontal ridges, and richly decorated in 22 karat gold. And throughout the collection, the books will vary in height, size, cover design, and in the color and grain of their leather bindings.

Leading artists and book illustrators will be commissioned to create original works of art exclusively for these volumes—they will never appear in any other editions.

Further, only quality bookpapers will be used—custom-milled to be free of the acids that cause ordinary papers to yellow. Pages will be thread-sewn into separate "signatures" and the edges gilded with a tarnish-free finish to seal the volume from dust and moisture. Finally, a ribbon page-marker will be bound into each book—an elegant touch and added convenience.

These will be volumes of such beauty and craftsmanship that their authors will be proud to sign them. And you will be equally proud to own them.

Long-term price guarantee—at no obligation

The total number of each new First Edition will be limited to just one volume per member. No additional copies will ever be made available.

By enrolling now, you will receive each new Signed First Edition—approximately one book each month—at the original issue price of just $35. This price will be guaranteed to you for three full years. And you will have the right to cancel your membership at any time, simply by giving thirty days' written notice.

Application deadline: December 31st

To enroll as a charter member, simply complete and return the accompanying application by December 31, 1983. After that date, the charter membership rolls will be permanently closed.

In future years, the Society may accept non-charter members, who will be able to acquire only those volumes made available during the periods of their membership—at prices then prevailing.

To be sure of acceptance now, mail your Charter Membership Application by the deadline of December 31, 1983. No payment is required now, but it is important to respond by that date.

THE FRANKLIN LIBRARY
A FRANKLIN MINT COMPANY

CHARTER MEMBERSHIP APPLICATION

The Signed First Edition Society

Must be postmarked by December 31, 1983.

The Franklin Library
Franklin Center, Pennsylvania 19091

Please enroll me as a Charter Member of The Signed First Edition Society, and send me the First Editions to be published by the Society during the period of my three-year membership. Each volume will be fully bound in leather, the spine accented with 22 karat gold, and personally signed by the author. I understand there will be approximately 12 books published each year, and that I can cancel my membership at any time upon 30 days' written notice.

I need send no money now. I will be billed $35.* for each leather-bound volume in advance of shipment. This price will be guaranteed to me as a Charter Member for three full years.

*Plus my state sales tax and $2.50 for shipping and handling.

Signature ______________________
ALL APPLICATIONS ARE SUBJECT TO ACCEPTANCE

Mr.
Mrs.
Miss ______________________
PLEASE PRINT CLEARLY

Address ______________________

City ______________________

State, Zip ______________________

Limit: One Membership per person.

Canadian residents will be billed for each book in advance of shipment at $47. plus $2.50 (Cdn. $).

60

Lost Ground

Although the U.S. is in principle an egalitarian society, it is well known that some social groups have faced greater barriers to social and economic achievement than others. The group that has been confronted by the most substantial obstacles consists of the black citizens of the U.S. Many efforts have been made to remove such obstacles. Have they succeeded in bringing blacks into any kind of parity with whites? According to a recent report by the Bureau of the Census, the answer is clearly no. Indeed, the report suggests that by several significant measures of social welfare, including average income and the integrity of the family, between 1970 and 1982 blacks lost ground to whites.

The report was compiled from several sources, among them the censuses of 1970 and 1980, monthly surveys of statistically representative samples of the U.S. population done between censuses by the bureau and reports from the Bureau of Labor Statistics, the National Center for Health Statistics and the Department of Defense. The accumulated data show that blacks living in an intact family are doing better economically in relation to whites than they were a decade ago. Families without a husband present, however, are doing worse. Moreover, the proportion of households headed by a woman is increasing, and hence since 1970 the standard of living of blacks as a group with respect to that of whites has declined.

In 1980 blacks made up about 12 percent of the U.S. population. From 1971 through 1980 the median income for all black families, corrected for the effects of inflation, decreased by 8.3 percent. The median income for all white families also decreased over the same period, but only by 2.7 percent. Thus the long-standing economic disparity between blacks and whites in the U.S. increased. In 1971 the average income of all black families was 60 percent of that of white families, but by 1981 the figure had dropped to 56 percent.

The overall economic loss was not distributed equally in the black population. The real income of black families that included both a wife and a husband rose 7 percent in real terms over the 10-year period. The average income of such families in 1981 was about $20,000. The average income of all black families headed by a woman without a husband present, however, decreased by 8 percent. In 1981 the average income of such families was $7,500.

Much of the economic setback suffered by the black community as a whole in the 1970's was due to the fact that the proportion of families headed by women increased. In 1972, 32 percent of black families did not include a husband; by 1982 the fraction had increased to 41 percent. In 1982 only 12 percent of white families were headed by a woman with no husband in the home.

Several factors contributed to the rise in the proportion of black families without a husband. Black women who have children are less likely to have been married than white women with children. In 1982, 32 percent of all black women who maintained families had never been married, compared with 11 percent of white women.

In addition blacks are much more likely than whites to be divorced, although in the past decade the divorce rate in both groups has risen substantially. The divorce ratio, or the number of divorced people for each 1,000 married people living with a member of the opposite sex, was 265 per 1,000 for black women in 1982, a sharp increase from the ratio of 104 in 1970. Among white women the divorce ratio rose from 56 to 128 in the same period. As a result of the divorce rate and the tendency for women to remain unmarried even after they have had children, 34 percent of all the children who lived with one parent in 1982 were black, although only 15 percent of all children in the U.S. that year were black.

The familial and economic difficulties of blacks intensify each other, with the result that a substantially greater number live in poverty than was the case a decade ago. In 1970 eight million blacks had incomes below the poverty level as it was then defined by the Federal Government. In 1981 nine million blacks lived in poverty. In both years the figures corresponded to 34 percent of the total black population. In 1981, 11 percent of whites had incomes below the poverty level.

What is the cause of such disproportionate hardship? The answer is that the removal of legal barriers has not eliminated the economic obstacles that keep blacks confined to a limited sector of the economy and cause them to be laid off first in times of economic recession. In 1982, when the unemployment rate was the highest since World War II, the rate of unemployment among blacks was 19 percent, or about double the rate of 9 percent that prevailed among whites. The ratio of two to one between black unemployment and white unemployment was roughly the same as it was in 1972, when the overall levels of unemployment in the U.S. were much lower. Thus in generally worsening economic conditions the historical disadvantages of blacks persisted.

At the end of the 1970's the blacks who had jobs were still concentrated at the bottom of the occupational hierarchy. According to the results of the 1980 census, blacks in the experienced civilian labor force worked mainly in three broad job categories: (1) operators, fabricators and laborers, (2) technical, sales and administrative-support occupations and (3) service occupations. The three groups accounted for 74 percent of the black workers.

In some of the more detailed occupational classifications utilized in the census blacks are heavily overrepresented. For example, although blacks make up only about 10 percent of the civilian labor force, more than half of all private-household cleaners and servants were black, as were one-third of all garbage collectors and one-fourth of all postal clerks, nurse's aides, hospital orderlies and hospital attendants. Conversely, only 3 percent of all experienced lawyers, physicians and engineers were black, and only 4 percent of all managers and administrators.

Reynolds W. Farley of the University of Michigan Population Studies Center and Suzanne M. Bianchi of the Bureau of the Census have concluded on the basis of the information in the current report and other surveys that the black community in the U.S. is being split into two groups: an upwardly mobile group with middle-class aspirations and a permanent, poverty-ridden underclass. They argue that the social and financial distance between the best-off and worst-off blacks has increased considerably in recent decades: "In the 1940's almost all employed blacks worked as farm laborers, unskilled factory workers or domestic help. As blacks became better educated and moved into higher-status professional areas, the prestige gap between the top and bottom quarters of the occupational ladder widened."

Farley and Bianchi observe that for the worst-off blacks things are grimmer than the data on occupation and employment alone would suggest. One group of the most oppressed are not included in figures on unemployment: those who have become so discouraged about finding a job that they have given up looking for one. According to Farley and Bianchi, the fraction of discouraged job seekers "has increased among white men from 3 to 5 percent, but it has more than doubled among black men, from 5 to 12 percent, between 1970 and 1980."

They conclude that "two signs point to trouble for the future of black Americans and for society as a whole. One is the growing number of working-age


THE GIFT OF LOVE
SARONNO
AMARETTO
di
SARONNO
ORIGINALE
A LIQUEUR PRODUCED BY
ILLVA SARONNO ITALY
750 ml
56 PROOF
AMARETTO DI SARONNO
ORIGINALE·1525
TO SEND ANYWHERE IN THE U.S. CALL (800) 528-6148. VOID WHERE PROHIBITED.
LIQUEUR 56 PROOF. SOLE U.S. DISTRIBUTOR, FOREIGN VINTAGES, INC., NEW YORK, N.Y. © 1983. A UNIQUE PRODUCT FROM ILLVA SARONNO, ITALY

men who do not hold or seek regular employment; the other is the growing economic gap between husband-wife families and families headed by women. In 1960 one-quarter of all blacks in poverty lived in female-headed households; today the figure is 60 percent. The children who grow up in these families may be leading candidates for the next generation of poor and chronically unemployed, signaling an even greater division among black Americans in the future."

Cruising for Trouble

The 464 ground-launched cruise missiles (GLCM's) scheduled to be deployed by the U.S. in Europe beginning next month are representative of a rapidly proliferating class of weapons that threaten to make the arms race even less controllable than it is today. In addition to the intermediate-range, nuclear-armed GLCM's intended for deployment in Britain, Italy, West Germany, Belgium and the Netherlands, the U.S. is currently installing air-launched cruise missiles (ALCM's), also armed with nuclear warheads, on B-52 bombers stationed in the U.S. and is vigorously pursuing the development of an assortment of sea-launched cruise missiles (SLCM's), both nuclear and non-nuclear, for service on submarines and surface warships. Within a few years, it is projected, the U.S. will have augmented its already formidable strategic nuclear forces with thousands of highly effective cruise missiles mounted on a variety of mobile launching platforms around the periphery of the U.S.S.R.

Estimates of the status of cruise-missile developments in the U.S.S.R. vary, but it is generally agreed that the Russians are substantially behind the U.S. in this particular area of advanced military technology. As in the case of multiple independently targetable reentry vehicles (MIRV's), however, there seems to be little reason to doubt that the Russians will eventually be able to match the American achievements and hence be in a position to present a comparably enhanced threat to targets in the U.S. Thus the net result of the advent of the cruise-missile era is likely to be a further reduction in the national security of both sides.

From the standpoint of arms control the main drawback of cruise missiles arises from precisely those qualities that make such weapons militarily attractive in the first place: they are smaller, cheaper, easier to hide and more versatile than long-range ballistic missiles, and hence they lend themselves to more flexible and deceptive basing modes. (Although cruise missiles are much slower than ballistic missiles, this disadvantage is thought to be more than offset by the superior accuracy of their continuous-guidance systems and by their ability to fly low over the terrain, evading radar tracking.) Because cruise missiles are hard to identify reliably by "national technical means" of surveillance (chiefly satellite photography), they are expected to present unprecedented difficulties to any nation attempting to verify another nation's compliance with an agreement limiting their number. Restrictions on cruise missiles, based on novel "counting rules" and other special verification aids, were incorporated in the SALT II Treaty and Protocol, but they remain in abeyance, since the treaty has not been ratified by the U.S. Senate. Given the extraordinarily strict standards of verifiability espoused by representatives of the Reagan Administration, the outlook does not appear to be bright for the inclusion of significant controls on cruise missiles in any agreement that might emerge from the current START negotiations in Geneva.

Meanwhile the U.S. is already preparing for the arrival of the next generation of cruise missiles. Early this year the Air Force announced that it would halt the current ALCM-B program in 1984, cutting off production at approximately 1,700 missiles rather than the 4,348 previously planned. The decision is expected to facilitate the timely procurement of the "follow on" weapon, the ACM, or advanced cruise missile, which will incorporate several improvements over its predecessors: (1) a more efficient engine, giving the ACM a much greater range than the ALCM-B; (2) a combination of microwave-absorbing materials and smoother design features (referred to collectively as "stealth" technology) that will enhance the missile's ability to elude Russian defense radar, and (3) a new passive guidance system, called Autonomous Terminal Homing (ATH), that will make the new missile not only more accurate but also less vulnerable to detection. According to one report, the advanced cruise missiles of the late 1980's may be capable of operating at supersonic speeds over intercontinental distances, giving them the potential of serving as strategic "counterforce" weapons against "hard" targets such as missile silos and command centers.

Appropriate Medicine

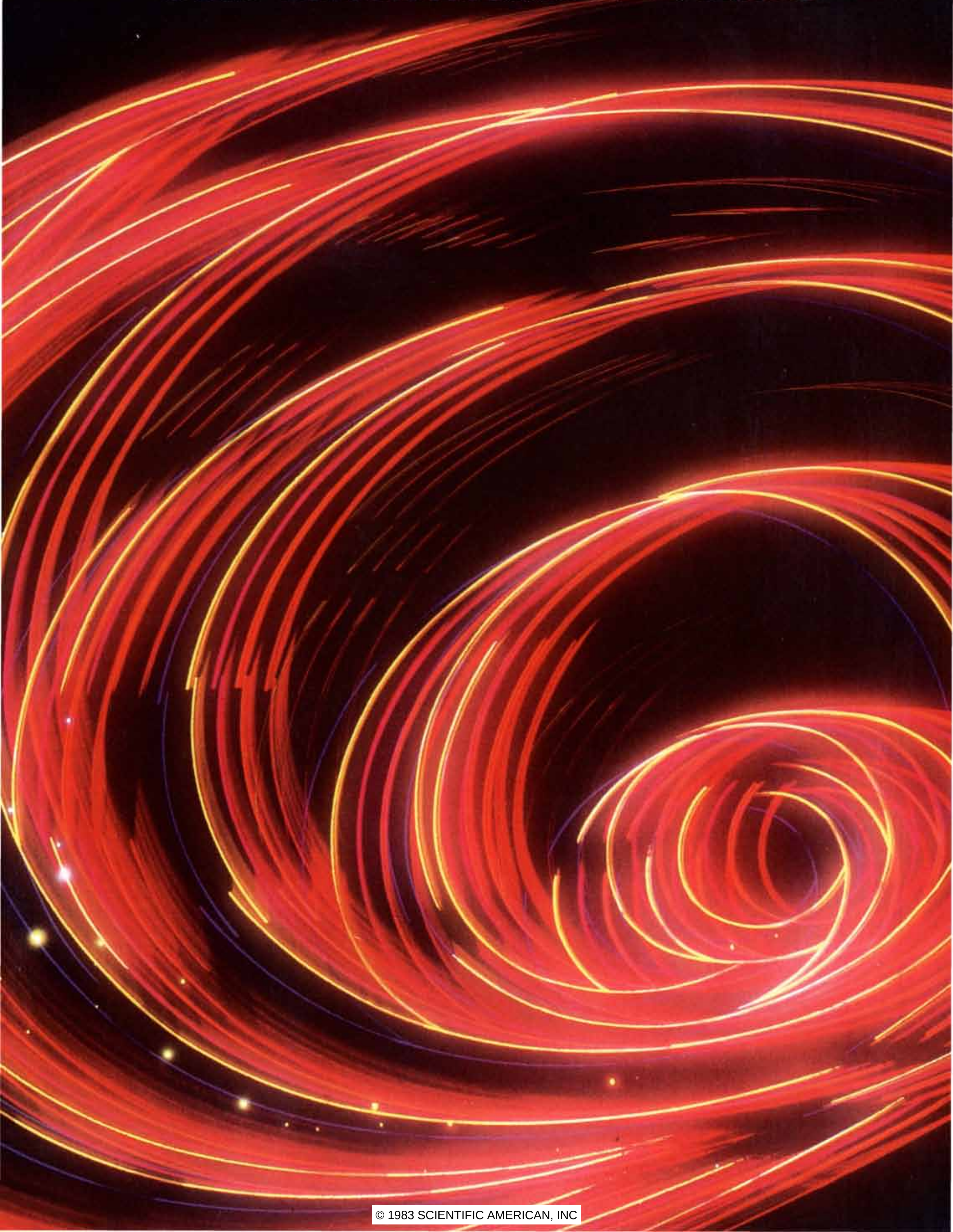
Diarrhea kills at least five million children a year, mostly in underdeveloped parts of the world, and contributes to the malnutrition of many millions more. It kills by dehydration. The treatment is obvious: replace the lost water and salts. This can be done by intravenous injection, but such therapy calls for sterile fluids, special equipment and skilled personnel, and it is expensive. It is generally unavailable precisely where childhood diarrhea is most prevalent. There is an effective, inexpensive alternative: oral rehydration therapy. A simple mixture of glucose and salts is mixed with local water and the solution is given by mouth; the patient's thirst regulates the intake to match what is lost in the watery stools. The potential of oral rehydration therapy has been clear for two decades, and recent data indicate that its routine application can virtually eliminate mortality from diarrheal dehydration. The task is to make the treatment available where it is needed.

The development and efficacy of the oral treatment are reviewed in an annotated bibliography published by the Pan American Health Organization and the World Health Organization. Acute diarrhea can be caused by a variety of agents, most of which are spread in contaminated water. Whatever the agent, the essential outcome is usually a reversal of the normal functioning of the small intestine. Sodium, and with it water, fails to be absorbed as usual from the gut into the bloodstream. Instead water and essential salts are secreted into the gut and are excreted in the stools. Even after it became clear that replacement of water and salts can save even most victims of cholera, the severest diarrheal disease, it seemed that only intravenous injection could circumvent the inability of the intestinal wall to absorb any orally administered fluid.

The key to oral therapy came more than 30 years ago. As early as 1949 it was found that glucose and some other simple sugars dramatically increase the absorption of sodium and water by the intestinal wall. A solution of glucose, sodium chloride (ordinary salt), potassium chloride and sodium bicarbonate in water could be given by mouth as maintenance therapy after initial rehydration by the intravenous route. In time it became clear that the oral treatment alone would suffice in most cases. The WHO developed a packet containing the correct amounts of glucose and the salts to be dissolved in a liter of water. The packets can be distributed through health-care systems and can be made available to families for administration in the home. If the packets are not available, it has been shown that a simple salt and sugar solution (a pinch of salt and a small handful of sucrose, or ordinary sugar) is often enough.

Oral rehydration therapy is a good example of effective "appropriate technology," in which laboratory investigation and clinical trials have developed a scientifically sound method suitable for widespread and inexpensive application under even primitive conditions. Commenting on an international conference on oral rehydration held in June in Washington, *The Lancet* recently pointed out that the problem is now one of inadequate implementation. Physicians, including those serving poor populations in developed countries, need to be

The Illuminated Vortex



The Illuminated Vortex

Understanding how the in-cylinder flow of the fuel-air mixture is influenced by chamber geometry provides a key to improving engine performance. By applying a laser measurement technique, a researcher at the General Motors Research Laboratories has gained new insight into the behavior of the flow.

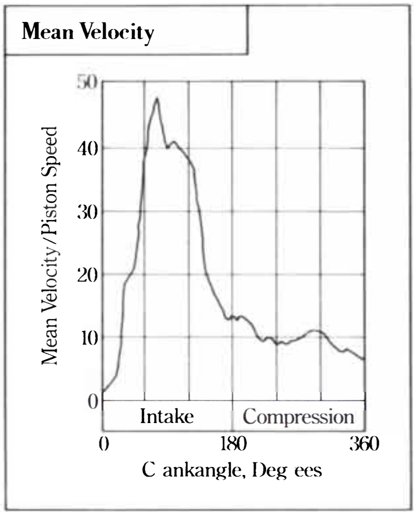
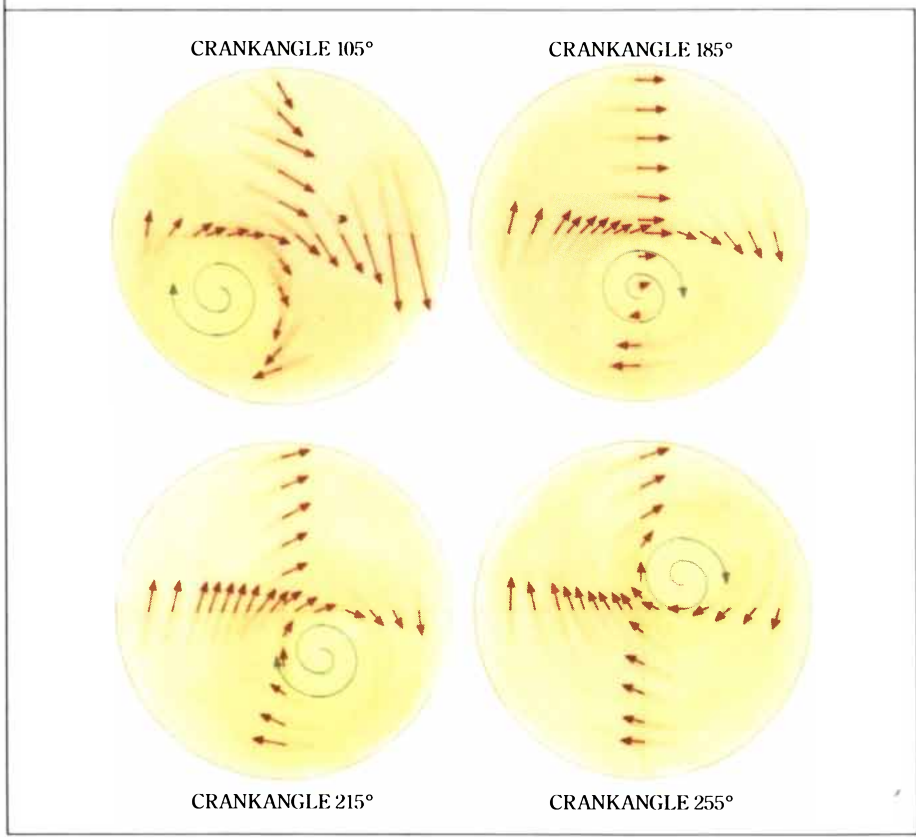


Figure 1: History of mean velocity at a single engine location.

Figure 2: Panoramic view of engine flow patterns. With changing crankangle, the center of rotation precesses from the cylinder's lower left quadrant to its upper right quadrant.



THE FLUID motions inside engine cylinders have considerable influence over the progress of combustion. Mixing of air and fuel, combustion rate, and heat losses from the cylinder are all important transport processes strongly dependent on fluid motions. The motion inside the cylinder has two components. Mean velocity influences the transport of momentum, energy, and species on a cylinder-wide scale, while the turbulence component influences the same phenomena on a local basis. The in-cylinder flow field depends primarily on the geometry of the cylinder and inlet port. Hence, decisions made in the engine design stage exert a controlling influence over the flow. But before questions about how different geometrical features affect the flow field can be answered, the problem of how to measure the flow must be solved. By applying Laser Doppler Anemometry (LDA), Dr. Rodney Rask, a researcher at the General Motors Research Laboratories, has obtained detailed measurements of the flow field.

LDA is a technique in which two focused laser beams pass into the cylinder through a quartz window. In the minute measuring region where the laser beams cross, a regular pattern of interference fringes is created. As the 1-micron particles, which have been added to the engine inlet flow, cross the measurement region, they scatter light in the bright fringes. In Dr. Rask's LDA system, the scattered light is collected by the same lenses used to focus the laser beam, and measured by a photomultiplier tube. The resulting signal is processed electronically to determine the time it takes a particle to traverse a fixed number of fringes. Since the fringe spacing is a known function of the laser beam crossing angle, this transit time provides a direct measure of velocity.

During operation of the LDA, measurements of velocity as a function of engine rotation (crankangle) are made at a number of locations within the cylinder. The instantaneous velocity at each point must then be separated into mean and turbulence components. The simplest technique is to declare that the mean velocities for all cycles are identical and ensemble average the data. However, this approach ignores the cyclic variation in the mean velocity. Another technique looks at individual cycles and uses a variety of methods, including sophisticated filtering, to split the instantaneous velocity into its components. This approach is consistent with the LDA measurements, which clearly show that the mean velocity does not repeat exactly from one engine cycle to the next.

Differences in the flow field from one cycle to the next can seriously compromise engine efficiency. Near the end of the compression stroke, it is important to maintain a consistent velocity at key cylinder locations (e.g., at a spark plug). Dr. Rask's LDA measurements have identified design features that control cyclic variability.

FIGURE 1 shows mean velocity measured at a single location during an engine cycle. High velocity exists during the intake stroke when the inlet flow is rushing through the narrow valve opening. This jet-like flow into the cylinder causes large velocity differences between adjacent cylinder locations and produces strong turbulence. As the end of the intake stroke is approached (180 degrees in Figure 1), the levels of both mean velocity and turbulence drop rapidly. This decrease is a result of the changing boundary conditions for the cylinder—from strong inflow to no inflow. During the compression stroke the flow field evolves, but it undergoes no drastic changes. However, in a high-squish chamber, where the flow is forced into a small bowl in the piston or cylinder head, considerable turbulence is generated near the end of the compression stroke.

Measurements from many cylinder locations are necessary to make the flow field understandable. Figure 2 shows four flow patterns covering a period from near the end of intake into the compression stroke. Note the strong vertical flow, with the center of the vortex away from the cylinder center and precessing with changing crankangle.

By experimenting with geometrical variables, Dr. Rask has gained new understanding of phenomena observed in operating engines. The resulting knowledge has guided the design and development of new engines with a minimum of trial-and-error testing. The LDA findings are also being used to validate and calibrate engine flow computer models under development.

"From our measurements," Dr. Rask states, "we have been able to deduce how changes in the geometry of the port and combustion chamber modify the velocity field. These flow field effects are now being used to help designers tailor engine combustion for optimum performance."

General Motors

THE MAN BEHIND THE WORK

Dr. Rodney Rask is a Senior Staff Research Engineer in the Fluid Mechanics Department at the General Motors Research Laboratories.

Dr. Rask received his undergraduate and graduate degrees in mechanical engineering from the University of Minnesota. His Ph.D. thesis concerned the Coanda effect.

Prior to joining General Motors in 1973, Dr. Rask worked on the design of nuclear reactors at the Knoll's Atomic Power Laboratories. In addition to further refinements in LDA measurement techniques, his current research interests include computer simulation of engine systems, with special emphasis on the intake manifold.

You can count on 3M diskettes. Day after day.

Just like the sun, you can rely on 3M diskettes every day. At 3M, reliability is built into every diskette. We've been in the computer media business for over 30 years. And we've never settled in. We're constantly improving and perfecting our product line, from computer tape and data cartridges to floppy disks.

3M diskettes are made at 3M. That way, we have complete control over the entire manufacturing process. And you can have complete confidence in the reliability of every 3M diskette you buy.

Look in the Yellow Pages under Computer Supplies and Parts for the 3M distributor nearest you. In Canada, write 3M Canada, Inc., London, Ontario. If it's worth remembering, it's worth 3M diskettes.

3M hears you...

3M

made more aware of the serious impact of childhood diarrhea and of the efficacy of oral treatment. Educational materials are needed to show mothers how to administer the treatment and to discourage reliance on traditional antidiarrheal medicines, most of which do more harm than good. The salient need is the development of effective networks for the delivery of primary care to provide the necessary education and materials.

Quarked Spark

What makes the sun shine? The short answer is that thermonuclear reactions deep in the sun's interior release energy in the course of fusing hydrogen into helium. A more complete account, however, must specify the pathways of the reactions, and the only direct evidence bearing on the pathways is the flux of neutrino by-products from the intermediate reaction steps. In principle such a flux can be detected on the earth because neutrinos interact with matter so rarely that they pass through the outer layers of the sun virtually unimpeded. Since 1967 Raymond Davis, Jr., of the Brookhaven National Laboratory has measured the neutrino flux at the bottom of the Homestake Gold Mine in Lead, S.D., and his results have presented a profound puzzle: the measured flux is lower than the predicted one by a factor of four. Now four physicists at Ohio State University have come forward with a new pathway for the nucleosynthesis of helium that may resolve the discrepancy. Richard N. Boyd, Ronald E. Turner, Michael C. T. Wiescher and Lawrence J. Rybarcyk invoke the presence of exotic atomic nuclei near the center of the sun whose electric charge, instead of being some integral multiple of the charge of the electron, is fractional. They describe their work in *Physical Review Letters.*

Fractionally charged particles must be extremely rare in nature if they exist at all; the idea that all electric charge is integral has been accepted as one of the basic properties of matter ever since Robert A. Millikan measured tiny electric charges early in this century. According to the theory of quantum chromodynamics (QCD), however, all nuclear matter is made up of quarks, which bear a charge of 1/3 or 2/3 times the charge of the electron. In most interpretations of QCD theory isolated quarks cannot exist; the quarks bind together to form particles with integral charge. The proton, for example, is made up of two "up" quarks, each with charge 2/3, and one "down" quark, with charge $-1/3$; the net charge is 1. The existence of bound states of quarks whose net charge is nonintegral, however, is not ruled out by QCD; indeed, such particles may have been found. In 1981 William M. Fairbank, George S. LaRue and James D. Phillips of Stanford University reported the detection of small concentrations of nonintegrally charged particles in the metal niobium. Although the Stanford findings have not been replicated, the sensitivity of the experiment is so high that the results cannot easily be dismissed.

The Ohio State theorists suggest that nonintegrally charged nuclei in the sun could catalyze a cycle of helium nucleosynthesis whose existence has not previously been recognized. Until now two major reaction pathways have been held responsible for the production of solar energy. The dominant pathway begins with the fusion of two protons to form a deuterium nucleus, and so the reaction is called the proton-proton reaction. It results in the emission of a neutrino, but the average energy of such neutrinos is too low to be detected by Davis' apparatus. Subsequent steps in the reaction, however, release neutrinos of much higher energy; it is primarily these neutrinos that can be observed. According to the standard theoretical model, reaction pathways beginning with the proton-proton reaction account for about 98 percent of the helium nucleosynthesis in the sun and for 98 percent of the sun's energy.

The second major reaction pathway is the carbon-nitrogen-oxygen (CNO) cycle. It was first proposed in 1939 by Hans A. Bethe of Cornell University and independently by C. F. von Weizsäcker in Germany. Nuclei of carbon 12 in the sun's interior combine with three protons, one at a time, to form increasingly heavy nuclei. The accretion of protons continues until it has built up a nucleus of nitrogen 15, consisting of seven protons and eight neutrons. (Two of the accreted protons decay into neutrons; each decay is accompanied by the emission of a barely detectable neutrino.) With the addition of a fourth proton the growing nucleus splits into a helium nucleus and a nucleus of carbon 12, whereupon the cycle begins again. In the standard model of solar burning the CNO cycle accounts for the remaining 2 percent of the sun's energy output.

The flux of high-energy neutrinos that can be measured on the earth varies widely for relatively small changes in the temperature at which the solar reactions take place. The Ohio State workers therefore realized that one way to reconcile the observed neutrino flux with the predicted one is to assume a lower temperature for the region of the sun where the neutrinos are emitted. If the assumed temperature is reduced by a million degrees Kelvin, from 14.4 to 13.4 million degrees, the production of neutrinos whose energies are high enough to be detected falls off by a factor of four. Without some additional reaction pathway for producing helium, however, such a temperature would also lower the

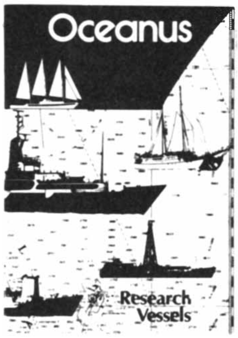
Oceanus®

The International Magazine of Marine Science and Policy

With offices in one of the world's leading marine research centers and published by the

Woods Hole
Oceanographic Institution,

Oceanus magazine is in a uniquely favorable position to monitor significant ocean research and related policy issues.



A case in point is the recent discoveries of deep ocean hydrothermal vents at various locations in the Pacific Ocean. These discoveries are causing a revolution in Oceanography.

Many long-standing biological and geological theories are being revised as a result of this exciting research. Exploration of the vents by the submersible Alvin is turning up new life forms (derived chemosynthetically rather than photosynthetically) and possible new sources of metals, such as zinc, iron, and copper. Scientists are getting a first-hand look at how the earth breathes while garnering supporting evidence for the theories of sea-floor spreading and plate tectonics.

Economically, it is a time of great challenge, too. The cost of running submersibles and support ships, for example, is high and rising. And federal funding for marine research all the while is shrinking. Many critical decisions must be made in nearly every area of oceanic concern.

Oceanus is meeting this challenge by publishing articles that help our readers grasp the significance of present research and that expose them to the substance of important public policy questions.

Recent thematic issues have dealt with:

- Deep Sea Mining
- Research Vessels
- Sharks
- Coastal Problems
- Pollution
- Oceanography from Space

Upcoming issues will look at:

- Marine Birds
- Oil & Gas Activities
- Industry and the Sea
- Polar Research

and much, much more.

Handsomely illustrated with many sharp photographs and line art, **Oceanus** is edited for all those with a serious interest in the three-quarters of our planet that is covered by water. The general readership includes many students, educators, representatives from government and industry, environmentalists, conservationists, as well as scientists in the disciplines of marine biology, geology and geophysics, chemistry, ocean engineering, and physical oceanography.

In addition to the thematic editorial material, **Oceanus** also features:

PROFILES of selected oceanographers that give a glimpse of the human side of science.

A CONCERNS SECTION that exposes the reader to controversial subjects and opinions.

A BOOK REVIEW SECTION that gives readers a convenient way to choose new and timely reading material.

LETTERS TO THE EDITOR or a forum for debate on views and thoughts expressed in the magazine.

Won't you join the Editors of **Oceanus** — published quarterly — in the exploration of the last great wilderness on earth, our oceans. Subscribe today!

Oceanus
SUBSCRIPTION ORDER FORM

Please enter my subscription for:

- ☐ one year at $20.00
- ☐ two years at $35.00
- ☐ payment enclosed ☐ bill me

(prepayment preferred)

Please send MY Subscription to:

Name (please print)

Street Address

City State Zip

Please make checks payable to Woods Hole Oceanographic Institution. Checks accompanying foreign orders must be made payable in U.S. currency and drawn on a U.S. bank. (Outside U.S., add $3 per year to domestic rates). Mail to: OCEANUS, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543.

INSURE OPTIMUM RESULTS

The people behind Hughes Support Systems are talented . . . determined . . . proud. They lead the challenge to new frontiers of technology, from a newly created, ultra-modern, working environment.

At Support Systems, Hughes people show the world how to use the systems of tomorrow . . . providing a full compliment of essential support services and equipment. Including training simulators, portable and depot test systems. Field and logistics support. Technical training for customers and development of training manuals. Covering programs ranging from electro-optics, lasers and missile systems, to airborne radar.

At Hughes, there is a stimulating relationship between the people and their work. A relationship that provides opportunity for substantial individual contribution. That's what Hughes Support Systems is all about. People.

People like you.

With Electro-Optics test experience (12 years + preferred):

Sr. Scientists — Responsible for application and development of EO test and measurement devices. Systems expertise in FLIRs, detectors, lasers, imaging devices, fiber optics, MS or PhD in EE or Physics required.

Project Managers — Management responsibility for technical performance, cost and scheduling in development of EO test sets. Plus direct feasibility and development-mode efforts. Project management background and BS in Physics or EE required. MBA desired.

Send your resume to:
Hughes Support Systems,
A1/1B404, Dept. 017,
P.O. Box 9399, Long Beach,
CA 90810-0463.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY

Proof of U.S. Citizenship Required.
Equal Opportunity Employer.

SUPPORT SYSTEMS

energy output of the sun by 22 percent. The proposed pathway restores the emitted solar energy to its observed value without releasing neutrinos whose energy is high enough to be detected.

The nonintegrally charged nuclei play the same role in the new reaction as carbon-12 nuclei play in the CNO reaction: they catalyze the accretion of free protons. The reaction begins with what is called a quarked helium nucleus, made up of two protons, two neutrons and an "up" quark; the nuclear charge is $2\frac{2}{3}$. Three protons are added to this nucleus one at a time until a particle made up of three protons, four neutrons and an "up" quark has been built up. Such a particle is essentially a lithium-7 nucleus combined with an "up" quark. With the addition of the fourth proton the quarked lithium-7 nucleus splits into a helium nucleus and a quarked helium nucleus, and the cycle begins again. Just as in the other two reactions, the essential change is the fusion of four protons into a helium nucleus, with the release of energy and neutrinos. According to the Ohio State group, the reaction can proceed in such a way, however, that the energy of the neutrinos is too low for them to be detected on the earth.

The newly proposed reaction pathway can generate the missing 22 percent of the sun's energy if nonintegrally charged nuclei make up only one part in 10^{15} of the sun's interior. The proportion is consistent with the error limits given so far by the Stanford experimenters. Nevertheless, the model must be viewed with caution. The Stanford result is not fully accepted, and quarked nucleosynthesis can account for the missing neutrinos only if physical quantities such as the internal temperature of the sun fall within extremely narrow limits. According to William A. Fowler of the California Institute of Technology, the proof of any theory of solar-energy production must await the construction of a more sensitive neutrino detector. One such detector that has been proposed would essentially consist of 50 tons of the rare metal gallium. The main obstacle to its realization is that this amount of the metal is more than the world's annual production and would cost $25 million.

Catalytic Cup

The shape of many proteins is central to their function. Enzymes, for example, offer binding sites with particular shapes so that particular molecules are brought into positions favoring a particular chemical reaction. The reaction would otherwise be improbable. The question arises: Can a chemist design a molecule that has a useful shape, so that it facilitates a reaction as an enzyme does? In an article in *Accounts of Chemical Research* C. David Gutsche of Washington University in St. Louis describes his effort to synthesize molecules that resemble the bell-shaped Greek vase known as the calix crater. He calls them calixarenes.

By a quirk of chemistry the methods Gutsche employs are reminiscent of a synthesis developed by Leo Baekeland at the beginning of this century. Baekeland heated formaldehyde (HCHO) and phenol (C_6H_5OH), a six-carbon ring in which five of the carbon atoms are linked to hydrogen atoms and the sixth is linked to a hydroxyl group (OH). The result was a hard, resinous material: the first commercial plastic, Bakelite. Gutsche and his colleagues start with a phenol that has a second substitution: the hydrogen atom across the ring from the hydroxyl group is replaced by a tertiary butyl group (C_4H_9), which consists of three methyl groups (CH_3) bound to a central carbon atom. When this compound is treated with formaldehyde, it forms calixarenes, a mixture of cyclic substances: mostly four, five, six or eight phenol rings forming larger, cup-shaped molecules.

Each phenol ring is one wall of such a cup. The butyl groups protrude above its rim; the hydroxyl groups jut inward at its base, closing off the cup at its bottom. In essence, then, each calixarene circumscribes a central cavity. The cavity acts as a trap; thus the calixarene has a "propensity to form molecular complexes with smaller molecules." The trapping depends on the size of the cavity. The four-sided calixarene *p-tert*-butylcalix[4]arene forms complexes with chloroform, toluene and pyridine, the five-sided calixarene *p-tert*-butylcalix[5]arene forms complexes with isopropyl alcohol and acetone, the six-sided calixarene *p-tert*-butylcalix[6]arene forms complexes with chloroform and methyl alcohol. X-ray crystallography confirms that the small molecule is indeed sequestered inside the central cavity. The calixarenes therefore simulate enzymes that bind a small molecule inside a pocket-shaped site.

How well might a calixarene mimic an enzyme? Gutsche and his colleagues are considering the enzyme aldolase. In animals it serves an early stage in glycolysis, the metabolism of glucose. For the most part it catalyzes the splitting of the six-carbon molecule fructose 1,6-diphosphate into two three-carbon molecules, dihydroxyacetone phosphate and glyceraldehyde 3-phosphate. In green plants it does the reverse, thus serving a late stage in photosynthesis. Its action is known in detail. First an amino group (NH_2) protruding outward from the amino acid lysine, a part of the enzyme, binds the middle carbon of dihydroxyacetone phosphate. The binding is assisted by the electrostatic attraction between the negative charge of the phosphate group of the three-carbon molecule and a positive charge on the enzyme. Next the three-carbon molecule is stripped of a proton (a hydrogen nucleus). Its resulting negative charge facilitates a linkage to glyceraldehyde 3-phosphate, which aldolase also has ways to capture.

Gutsche and his colleagues hope to graft three features of aldolase onto a four-sided calixarene some of whose sides have been induced to flip downward so that the calix has what they call a partial-cone conformation. An amino group is to be stationed partway up the exposed inner surface of the cup; it is intended to mimic the ability of aldolase to bind the middle carbon of dihydroxyacetone phosphate. An ionized metal atom is to be stationed at the bottom of the cup; it is intended to attract the phosphate group of dihydroxyacetone phosphate. A basic group (CO_2^-) is to be stationed at the rim of the cup; it is intended to strip a proton from dihydroxyacetone phosphate. The simplest ways to synthesize calixarenes cannot produce such features, and so Gutsche and his colleagues are devising more complex synthetic techniques.

The Wonders of Sian

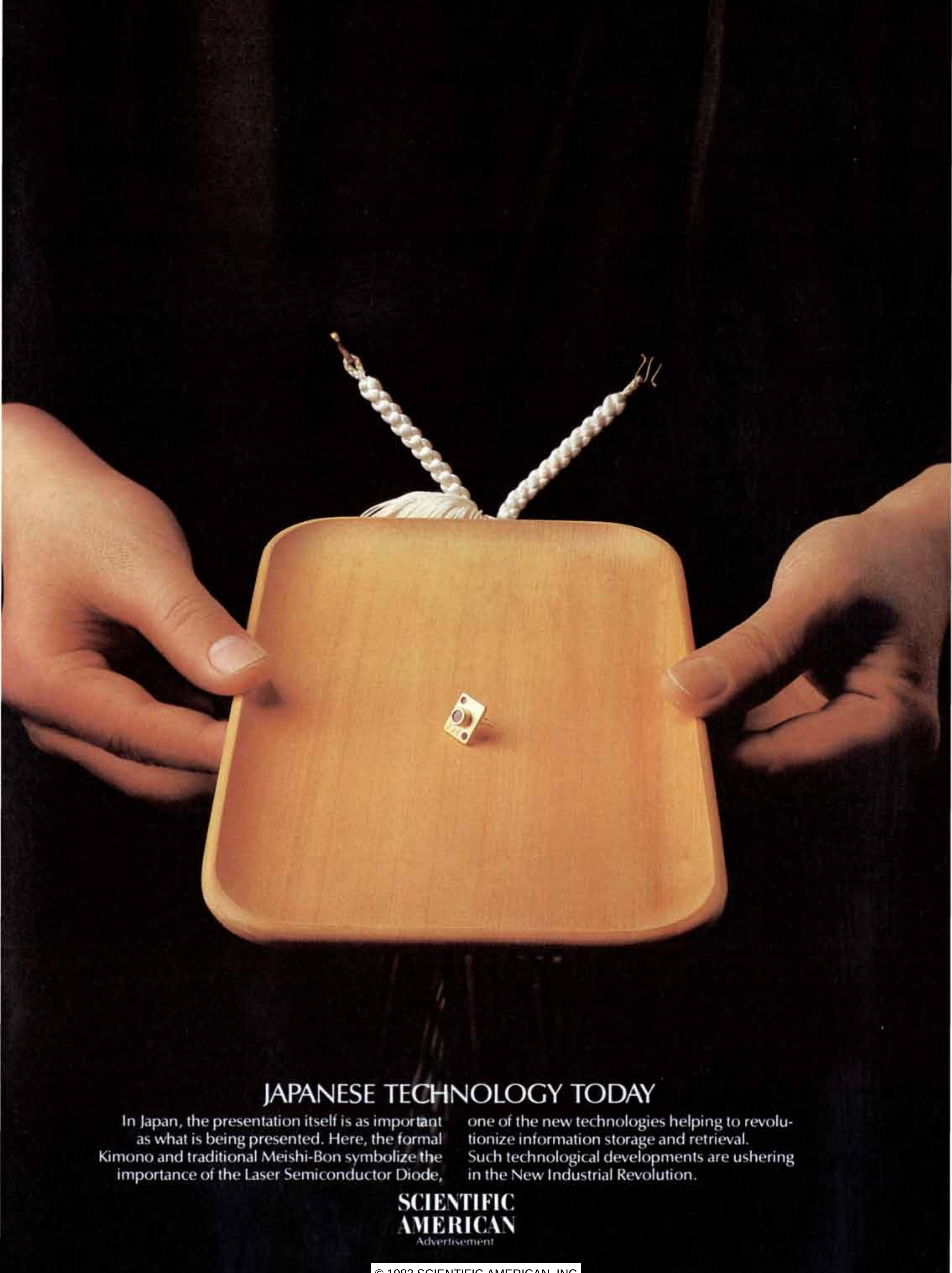
Ever since a well digger exposed the first of some 6,000 life-size terra-cotta soldiers buried two kilometers from a great earth tumulus in southeastern Shensi in the summer of 1975, it has been apparent that the Chinese had struck an incomparable archaeological bonanza. The tumulus was traditionally held to be the tomb of China's first emperor: Ch'in Shih Huang Ti. His burial place, prepared years before his death in 210 B.C., had been the subject of speculation, much of it the stuff of legend, ever after. Until 1975, however, digging in the enormous mound by would-be thieves had revealed nothing, and no official excavation had been attempted.

Since that time archaeologists from the Chinese Academy of Social Sciences, which is responsible for historical research, and from the Shensi provincial government have been excavating elsewhere near the great tumulus, adding to the list of recovered grave offerings hundreds of life-size terra-cotta warriors and horses, together with bronze weapons and other artifacts. One of the most recent finds consists of the first bronze effigies found so far: two teams of four chariot horses, with their harness virtually complete, each team drawing a two-wheeled bronze chariot complete with a driver, arms outstretched to hold the reins, clad in a long-sleeved winter cloak and wearing a cap covering his ears and neck and decorated with a cockade. The effigies are half life-size; one of the two has recently been placed on display at the museum erected on the site not long after the first discoveries.

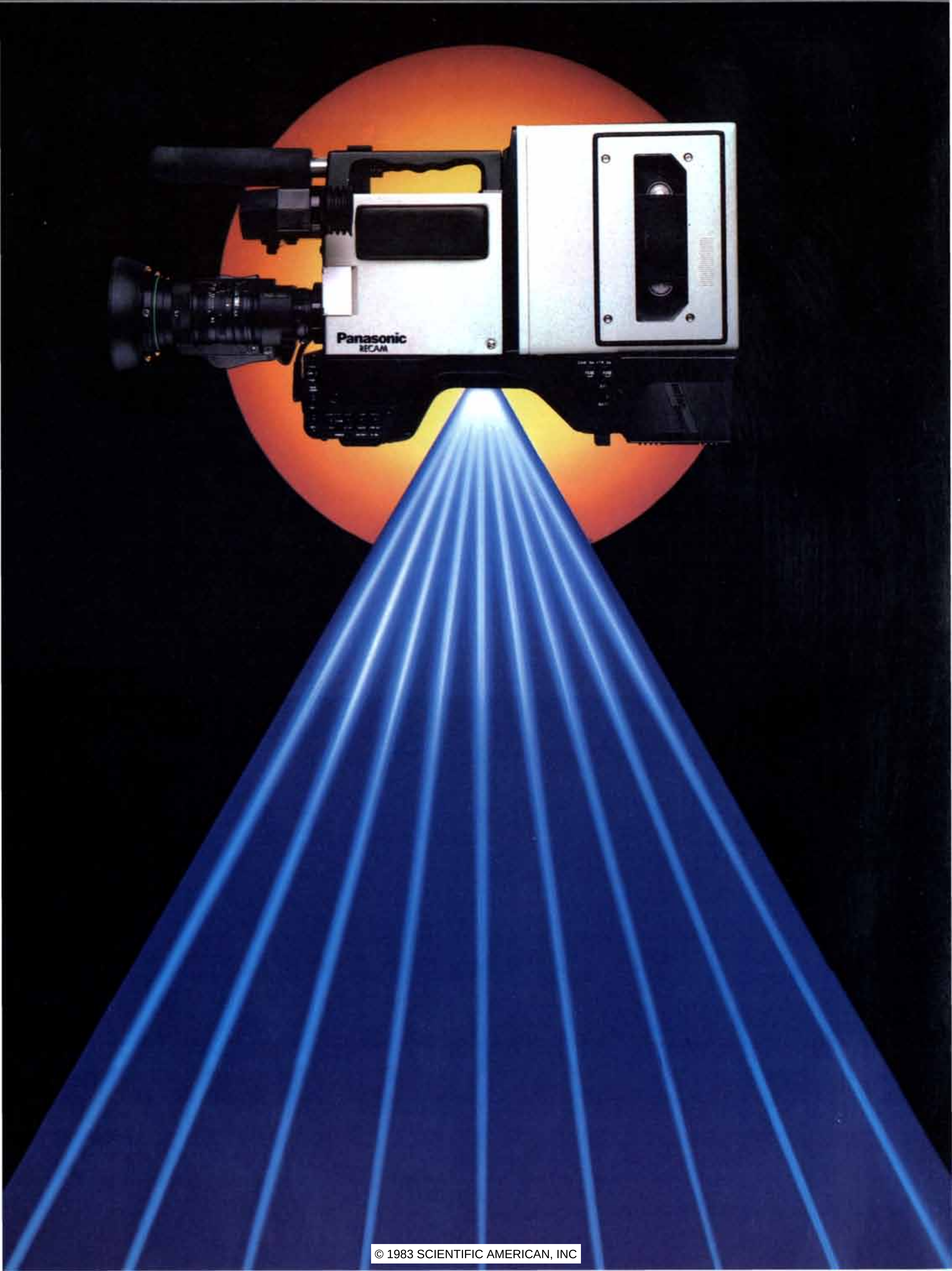
Ch'in Shih Huang Ti holds a unique place in Chinese history. According to tradition, his royal house was instructed in about 770 B.C. to police the western Chou capital city of Anyang after the Chou ruler, then the most powerful of China's feudal kings, had moved his capital to the east. Over the next 150 years the Ch'in state grew to be the most powerful in western China, and by 338 B.C. it was also the best-governed in all China. Shih Huang Ti succeeded to the Ch'in throne in 246 B.C., some 15 years after his predecessor had effectively destroyed the military power of the paramount feudal state, the Chao. (Chou kings, although they were still given some lip service, were by then powerless.) Young Shih Huang Ti also went campaigning, and between 246 and 221 B.C. he brought every other feudal state in China to its knees. With his power consolidated, he declared himself emperor of all China and went on to effectively dismantle the feudal systems of his former rivals and establish his own central authority. He partitioned the nation into 40 provinces, which he ruled through his own civil service: inspectors who oversaw the activities of local magistrates.

Fearful of death, he prepared heroically for his afterlife. Legend has it that within the mountainous tomb he had raised a few score kilometers from his capital (modern Sian) was a great relief map of his empire, with mercury endlessly flowing along its river channels. When the emperor died at the age of 49 in 210 B.C., legend further relates, his burial chamber was filled with molten copper so that his coffin was sealed within a single huge metal ingot.

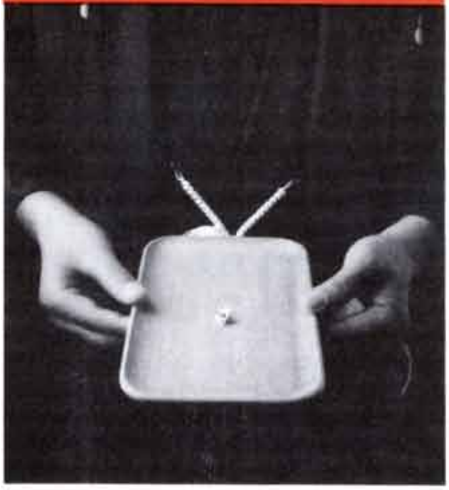
Chinese officials state that the bronze chariot-team effigies, found 20 meters outside the western edge of the great tumulus, are the most important find in the vicinity of the tomb since the initial discovery of the terra-cotta army in 1975. Considering that some 12 square kilometers, including the tumulus itself, remains to be explored, the Shensi site represents one of the great archaeological enterprises of all time. The excavation of Herculaneum and Pompeii, the towns buried by an eruption of Vesuvius in A.D. 79, has been in progress since 1748; the clearing of Shih Huang Ti's immense funerary monument may well take longer.


JAPANESE TECHNOLOGY TODAY
In Japan, the presentation itself is as important as what is being presented. Here, the formal Kimono and traditional Meishi-Bon symbolize the importance of the Laser Semiconductor Diode, one of the new technologies helping to revolutionize information storage and retrieval. Such technological developments are ushering in the New Industrial Revolution.
SCIENTIFIC AMERICAN
Advertisement

MATSUSHITA IN VIDEO TECHNOLOGY:



JAPANESE TECHNOLOGY TODAY



THE AUTHORS

The 1983 edition of "Japanese Technology Today" was written by Dr. James C. Abegglen and Mr. Akio Etori.

Dr. Abegglen is Vice-President and Director of the Boston Consulting Group, Inc., and concurrently Professor of International Management at Sophia University, Tokyo. He has written numerous articles and books including *The Japanese Factory,* a pioneering study of Japanese corporate organization.

Mr. Akio Etori is an award-winning and respected writer on science and technology, having covered these areas for over 20 years. He is managing editor of *Saiensu,* the Japanese language edition of *Scientific American* and Japan's leading and most prestigious publication in this field.

THE COVER

Photographer Arthur Beck captures the importance of one of the latest developments in component technology from Matsushita Electric. Presented on the traditional Meishi-Bon, the new Terraced-Substrate Laser Semiconductor has made possible such information processing devices as optical memory disk systems, digital audio disk systems, and laser printers in computer terminals.

Art direction and supervision of text and illustration were provided by Ted Bates Advertising/New York.

The dynamism of the Japanese economy over the past 30 years, and the impact of that dynamism on the world, must be considered a pivotal phenomenon of the 20th century. For the first time since the industrial revolution, a non-Western nation has moved to the highest levels of output and income. The world now has a new, and unexpected, source of capital and technology that bids to overshadow the traditional Western sources. An English observer has suggested that just as the historical period 1775–1875 might be termed the British Century, and 1875–1975 the American Century, so the next hundred years may come to be seen as the Pacific Century. If so, the driving force will have been Japan.

The magnitude of Japan's change is very great indeed, and in a short period of time. As recently as 1960, only 20-odd years ago, the Japanese economy made up less than 3 percent of the economy of the world. A 3 percent economy may be treated with condescension by a 33 percent economy, as indeed Japan was treated by the United States in those days. *(See Table 1.)* A 3 percent economy may provide occasional nuisances in trade relations, but is hardly likely to be the object of systematic protectionism. Nor is such an economy likely to be the subject of admiring studies of its management systems.

In 20 years, the Japanese economy became a 10 percent economy. (Some license is needed here to accommodate current aberrations in exchange rates, which now somewhat understate Japan's position.) A 10 percent economy attracts very serious attention. This is more the case when its principal trading partner has over the same period experienced a very substantial reduction in relative economic position. While Japan moved from insignificance to major importance economically, the United States' share of world GNP moved from about one-third to about one-fifth.

This kind of change in both absolute and relative position is very hard to accommodate. The issues raised in terms of economic relations, especially trade, are clear enough. More subtle issues of changes of attitudes on the part of the Japanese, and toward the Japanese, are, nonetheless, real problems. The difficulties of adjustment are seen in trade tensions, in problems of redefining strategic relations, in efforts to analyze and better understand the Japanese accomplishment and its sources, and in Japan's own problems of internal social and political adjustment to the consequences of drastic economic change.

Even in the midst of these tensions and readjustments, there is still in Japan a considerable propensity to look ahead, to gauge probable directions and problems, as well as to attempt to measure outcomes. The current view is that Japan will continue to outperform the other developed economies to the year 2000. *(See Table 2.)* That is, the relative economic position of the U.S. is seen as continuing to decline, but at a slowing rate, with Japan continuing to grow at a relatively rapid, but again slowing, rate. In this process per capita output in Japan will come to be greater than that in the U.S. by a significant margin.

These continuing changes in the size of the total economy have implicit in them very great changes in the structure of the economy. Such structural changes have been taking place, and at a rapid rate, as indicated in *Table 3.* There has been a massive shift of the Japanese labor force out of the primary sector of agriculture, forestry and fisheries. The shift was initially into the construction and manufacturing sector through the 1960's and early 1970's. It is already clear, however, that a further shift is in process in recent years, with employment in manufacturing declining relative to employment in the tertiary sector of distribution, utilities, services and the like. That is, the structure of the Japanese economy, in terms of employment, is already moving in the direction of services that has been conspicuously the case with the U.S. in recent years.

As Japanese analysts look forward, they see this shift in the structure of the economy accelerating. As is indicated in *Table 4,* the level of employment in the primary sector is expected to continue to decline rapidly. By the year 2000 less than 5 percent of the Japanese labor force will be in farming, forestry and fishing. (With this, there is very likely to be a steady reduction in political resistance in Japan to greater imports competitive with these primary industries.)

Of even greater interest in terms of the pattern of expected growth is the view that there will be very little increase in employment in the manufacturing indus-

TABLE 1

SHARE OF WORLD GNP
1960–1980

	1960	1970	1980
US	33.7%	30.2%	21.5%
EEC	17.5	19.3	22.4
Japan	2.9	6.0	9.0
USSR	15.2	15.9	11.6
PRC	4.4	4.9	4.7
Other	26.3	23.7	30.8
Total	100.0	100.0	100.0

Source: 2000 Nen No Nihon (Japan in the Year 2000), Keizai Kikaku Cho Sogo Keikaku Kyoku Hen, Nihon Keizai Shimbunsha, Tokyo, 1982, p. 31.

“ ”

“The Honda Accord is far and away the best compact in the land. Its engine hums contentedly. Its ride is smooth and quiet. Its interior is a model of good taste and expensive-looking accoutrements.” (*Car and Driver,* May 1982)

“In short, the Honda Accord gets my vote for best all-around car under $10,000.” (*Washingtonian,* January 1983)

“From the standpoint of product value, though, Honda has given the new model what it takes to capture the attention of thoughtful buyers.” (*Road & Track,* January 1982)

“It is to Honda's everlasting credit that the cars are such all-around sweethearts to drive.” (*Car and Driver,* January 1982)

“They are beautifully put together, very comfortable, and rewarding to spend time with.” (*Car and Driver,* January 1982)

“The new 4-speed automatic overdrive transmission merits high marks.” (*Motor Trend,* June 1983)

“Hard cornering is something the Accord does quite well for a family sedan. After all, why should driving cease to be pleasureful just because you have a family?” (*Motor Trend,* January 1982)

“The Accord is one of the best sedans of any size that money can buy.” (*Car and Driver,* May 1982)

“An acknowledged benchmark.” (*Motor Trend,* June 1983)

“How do you follow a benchmark?” (*Car and Driver,* January 1982)

HONDA

The Accord 4-Door Sedan

JAPANESE TECHNOLOGY TODAY

tries in Japan between now and the end of the century. Indeed, employment in most sectors of manufacturing is expected to decline, with the machinery sector alone showing substantial increase. This sector, of course, takes in the whole range of electronics and other high-technologies.

The dramatic change in employment patterns over the period to the end of the century is expected to be in services. In fact, this sector is expected to account for essentially all of the increase in employment over the period. The employed population is expected to increase by about eight million and employment in services is expected to increase by a like amount. Services is a broad category, of course. It includes auto repairs and beauty parlors, architects and advertising agencies, computer advisers and tax consultants. The significance of this projection lies in the view that Japan's economy will be shifting over the next two decades very rapidly from an emphasis on hardware to a focus on software, to the highest value added sector of the economy.

The underlying causes of this expected shift in employment patterns can be seen in *Table 5*, showing expected changes in the composition of the economy. Such sectors as chemicals and metals will make up only a small part of the Japanese economy at the turn of the century. Machinery manufacturing will largely take the place of these declining sectors. At the same time, services will move from about one-third of GNP to nearly one-half of GNP.

It is against this broad economic background that the present situation and future prospects of Japanese technology must be seen. Technology is both cause and result of these structural changes in the economy. The structural changes themselves are the cause of rapid growth. Technology makes the changes possible. In turn, the changes demand new technologies. In this cycle of change, increases in productivity are possible, and the shift to higher value added is accomplished.

The successive oil shocks and their consequences have imposed on all economies an urgent requirement for structural change. The transitions are difficult for all economies. Japan has managed the transitions with more speed and thoroughness than most. Further, Japanese analysts, as we have seen, expect a series of continuing structural changes in the economy—indeed, welcome them despite the problems they present—since it is with these changes that income levels are raised, and standards of living improved.

Japan has by definition in this post World War II period of unprecedented economic growth managed a number of restructurings of industry. The first was in coal mining, which in its peak years in the early 1950's employed more than 450,000 workers in more than 900 mines and has since the early 1970's employed fewer than 50,000 workers in less than 50 mines. In this same period Japanese coal imports went from virtually nil to some 80 percent of total consumption.

Following mining, the next very large industry that came under pressure as the Japanese economy advanced was the textile industry. Here the effect has been more that of increased productivity than decreased output as employment in fibers and fabrics dropped from nearly 1.5 million in the mid-1960's to less than 900,000 by the end of the 1970's. Imports of fibers, fabrics and apparel have increased sharply, the more so since Japan now has the lowest tariffs on textiles in the world, and maintains no Multi-Fiber Agreement import restrictions.

TABLE 2

WORLD GNP, YEAR 2000 (Percent)

	1980		2000		Annual Growth
Advanced Economies	63%		58%		2.8%
Japan		10%		12%	4.0
US		22		20 }	2.5
EEC and Other OECD		31		26 }	
Developing Economies	15		20		4.6
NICs		4		7	6.0
LDCs		11		13	4.0
Socialist Economies	22		22		3.2
USSR		13		12 }	3.0
Eastern Europe		5		5 }	
PRC		5		5	4.0
World Total	100%		100%		3.2

Source: Japan in the Year 2000

TABLE 3

THE STRUCTURE OF JAPAN'S LABOR FORCE (Percent)

	Agriculture, Forestry, and Fisheries	Mining, Manufacturing, Construction	Utilities, Transport, Services
1955	37.6%	24.4%	38.1%
1960	30.2	28.0	41.8
1965	23.5	31.9	44.6
1970	17.4	35.2	47.4
1975	12.7	35.2	52.0
1980	10.4	34.8	54.8

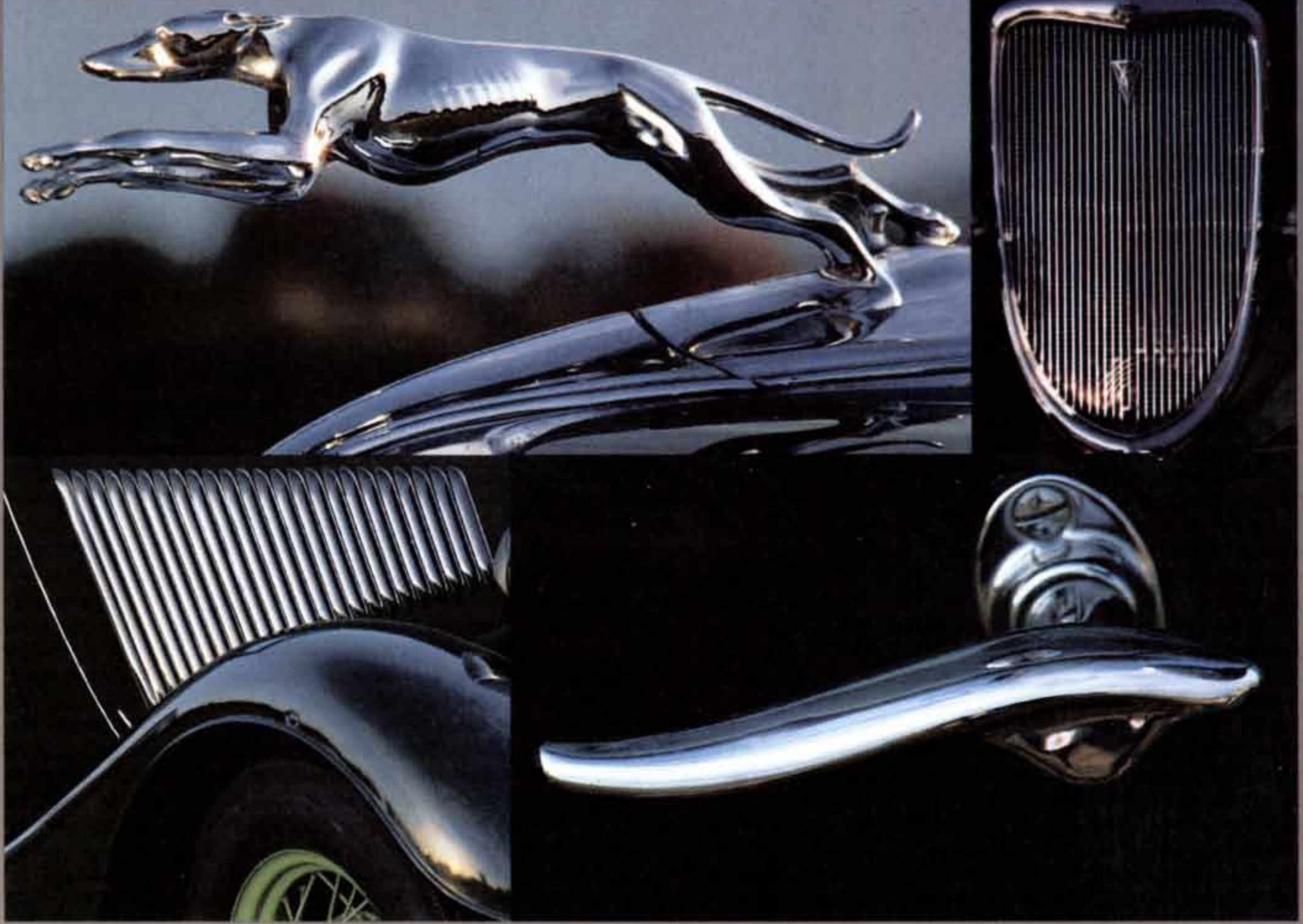
Source: Japanese Ministry of Finance

A more recent and dramatic case of industry restructuring in Japan was that of shipbuilding, where with the collapse of world demand from 1974, the industry in Japan, with government support, moved to cut capacity. Japan had 50 percent of world shipbuilding capacity, at the peak of demand. In two years in the late 1970's, 40 percent of that capacity—or 20 percent of world capacity—was closed down.

For all of these past successes the task facing Japanese industry and government in the 1980's to achieve the restructuring necessary to meet national aspirations is a formidable one. The context needs to be sketched briefly. By the late 1960's it began to become apparent to many in Japan that the industries on which Japan had built its initial successes as rapid growth began in the mid-1950's were in the process of becoming inappropriate as Japan sought higher levels of productivity and value added. These industries included many that had earlier been singled out for rapid development, such as petrochemicals and non-ferrous metals. They had formed the core of Japan's shift from labor intensive sectors like textiles into more sophisticated capital-intensive industries.

Still, in the emerging context of the late 1960's and early 1970's these raw-material-intensive, pollution-intensive, land-extensive industries were of less efficiency and relevance. This was the more true as the raw material supplying countries, Japan's trading partners, sought to put more value added to the materials before exporting them—smelting bauxite to alumina or to ingot before exporting, for example. It was in Japan's interest to shift these industries offshore, both in terms of better using Japan's own resources of capital and labor, and in terms of raising the level of trade with supplying countries.

The abrupt and catastrophic increase in



See beyond the ordinary.

The Canon A-1 is no ordinary camera. It is a creative tool. Conceived as the ultimate in automatic SLR's, the A-1 is unsurpassed in providing exposure control options. There are six, to be precise, allowing you to select the one best suited to your subject. Choose a shutter speed to control and interpret action. Select a lens opening and blur away a background.

In the programmed mode, the A-1 makes both of these decisions for you so you can really concentrate on your subject. You just focus, compose and shoot.

And create. A bright digital display in the viewfinder shows the speed and aperture in use, whether the choice was yours or the camera's.

The A-1 provides the versatility to match your imagination. Add any of over fifty Canon FD lenses. A Canon Speedlite for automatic flash. You can shoot at up to five frames-per-second with the optional Motor Drive MA. But most important, the A-1 does everything automatically. Freeing you to shoot a special subject in your own special way, and make a picture that nobody else saw.

The Canon A-1. It's half of what you need to turn photography into fine art.

The Official 35mm Camera of the 1984 Olympic Games

Canon A-1

Canon USA, Inc., One Canon Plaza, Lake Success, New York 11042
140 Industrial Drive, Elmhurst, Illinois 60126 / 6380 Peachtree Industrial Blvd., Norcross, Georgia 30071
123 Paularino Avenue East, Costa Mesa, California 92626 / 2035 Royal Lane, Suite 290, Dallas, Texas 75229
Bldg. B-2, 1050 Ala Moana Blvd., Honolulu, Hawaii 96814 / Canon Canada, Inc., Ontario

TABLE 4

STRUCTURE OF EMPLOYMENT
(Thousand Persons)

	1980		2000		Annual Change
Primary Sector	5,770		3,080		-3.1%
Secondary Sector	19,250		21,110		0.5
Manufacturing		13,770		14,200	0.2
Chemicals		1,750		1,450	-0.9
Primary Metals		670		540	-1.1
Machinery		5,380		8,930	2.6
Other Manufacturing		5,970		3,280	-3.0
Construction		5,480		6,900	1.2
Tertiary Sector	30,190		39,120		1.3
Utilities		300		330	0.4
Finance, Insurance, Real Estate		1,910		2,410	1.2
Transport and Communications		3,500		3,550	0.1
Services, etc.		24,480		32,830	1.5
Total	55,360		63,290		

Note: Population aged 15 to 64 years: 1980, 67,390; 2000, 67,230.
Source: Keizai Shingikai

energy prices in the 1970's turned what had been seen as a long-term process into one of great urgency. For among their other characteristics these industries are energy-intensive and notably unsuitable for siting in Japan, with the world's highest energy costs and a notably unreliable energy supply. Thus in the early 1980's Japan confronts a massive problem of industry restructuring of the greatest urgency. Something of the nature of the problem can be appreciated from the data in *Table 6*. A series of important industries, employing upward of 250,000 persons directly, with several hundred corporations, are in deep and abiding difficulty.

There is general agreement in Japan that the scale of these industries must be reduced, and resources that might be devoted to maintaining these industries will be better deployed elsewhere. Indeed, it might be argued that the Japanese are especially persuaded followers of the Schumpeterian concept of "creative destruction." Perhaps because of the dynamism of their economy, where economic cause and effect can be observed directly and in a short period, there seems to be in Japan an unusual appreciation of the fact that movement to higher standards of living does in fact require that some industries are losers in the process while others become the next generation of winners.

As an aside, it might be noted here that these cases are useful in providing some perspective on and perhaps a partial antidote to the currently popular literature in the U.S. that uncritically extols Japanese management methods. All of the companies in all of these industries are cases of "Japanese-style management." The often-cited characteristics of Japanese management—consensual decision-making, concern with quality, special attention to employee needs and the like—do not help these industries survive. Or put more precisely, whatever the considerable merits of many aspects of Japanese practices, these management methods are not able to override fundamental economics of a company's or industry's position. Further, when discussing the Japanese economy and its corporations, these cases should serve as a helpful reminder that Japan too has winners and losers, that no amount of management skill, government competence, or labor quality can prevent, or should be allowed to prevent, uneconomic companies and industries from confronting basic economics.

It is this process of continual restructuring that poses Japan's technology challenge. Just as the movement out of the "cheap" trinkets and textiles of Japan's earlier, poorer era required the technologies that established petrochemicals and metals industries, so now newer technologies must be put in place to continue the move to yet higher levels of value of output. As we have seen, the generally held view of Japan's future assumes that these transitions will be accomplished and that the technology is in place already or will be forthcoming. It is, however, a formidable challenge.

The challenge is the greater since the process is a continuing one. It is now beginning to impact an industry in which Japan has for some time held uncontested technological and economic leadership—the steel industry. The quite exceptionally high level of Japanese steel-making technology can be seen from the data in *Table 7*. Japanese steel companies are now suppliers of technology to nearly every U.S. steel-making concern. The U.S. government's import controls—the trigger price mechanism—have assumed that Japan

TABLE 5

COMPOSITION OF GROSS DOMESTIC PRODUCT
(Percent)

	1980		2000	
Primary Sector	3.7%		4.2%	
Secondary Sector	38.2		31.6	
Manufacturing		29.3		21.6
Chemicals		5.4		1.5
Primary Metals		3.6		0.8
Machinery		11.9		15.7
Other		8.4		3.6
Construction		8.9		10.0
Tertiary Sector	58.1		64.2	
Utilities		3.0		1.5
Finance, Insurance, Real Estate		15.5		8.5
Transport and Communications		6.6		5.6
Services, etc.		33.0		48.6
Total	100.0		100.0	

Source: Keizai Shingikai; Choki Tembo Iinkai

MAXELL ADVANCES STATE OF THE ART TECHNOLOGY TO A HIGHER STATE.

WITH THREE CLEARLY SUPERIOR PRODUCTS.

Over the years, Maxell has been the leader in the audio/video/computer media business. And the advanced tape technology used to create our three new products will keep us on top for years to come.

For the new XL-S recording tape, we developed a superfine Epitaxial formulation. This greatly increases its dynamic range, making the XL-S the ideal tape for all high performance recording needs and perfect for digital audio disc recording.

We also developed a totally new Epitaxial formulation for HGX Gold video tape. Its superfine particles deliver brighter chroma levels and higher audio sensitivity. And the binding process, using molecular fusion, gives our new VHS and Beta videocassettes unmatched durability.

Gold Standard floppy disks are widely accepted as the choice for critical new disk drive testing. But the new 5¼″ high density disk, equal to an 8″ double sided, double density disk, with a capacity in excess of 1.6 megabytes, took our involvement one step further. Developed in conjunction with a new drive design, it helped redefine 5¼″ capacity.

Behind the technical excellence of Maxell audio, video and computer media is the human element—Maxell people. We're dedicated to creating the kind of superior product performance that brings out the best potentials of high end equipment. In fact, we insist on it. Shouldn't you?

maxell®
IT'S WORTH IT.

was the world's most efficient steel-producing country.

Yet Japanese steel exports peaked several years ago. There have been no new capacity additions, and none are planned. According to current press reports, the Japanese steel industry is considering dumping and countervailing duty suits against foreign steel suppliers. Imports began a few years ago, and in 1982 increased 41 percent over 1981 to the considerable total of 19 million metric tons. In the first seven months of 1983, imports were up another 25 percent over the same period in the preceding year.

Whatever the validity of the industry's complaints about dumping, and they leave curious echoes in Western ears, the fact is that steel too will have its day in Japan, and will then yield pride of place to yet higher-growth, higher-technology industries. This is not to suggest that the industry in Japan will die; rather, it will be forced to focus on the more sophisticated steel products as South Korea, Brazil, Taiwan and other newly industrializing countries become in turn the lower-cost producers of standard grades of steel. For all of its notable success in management, the steel industry of Japan cannot escape the consequences of Japan's continued economic success.

No doubt in their turn too such current Japanese successes such as the automobile industry will encounter the problem of "creative destruction" that now looms over Japan's very successful steel industry.

The technology that has powered this continuing industrial restructuring has had its phases, just as the industry's changes have had. For industry, the shift broadly has been from labor-intensive to capital- and energy-intensive to technology-intensive sectors, as we will see below with Japan's electronics industry. In technology the transition has been from massive introduction of imported technologies, to adaptation and further development of those technologies, to the present situation where Japan must in greatly increased measure initiate its own technologies.

Japan's long dependence on the importation of technologies is shown clearly in its continuing deficit on its technology trade balance. Japan still pays out more than $3 for technology for every $1 it receives in payment for the export of technologies. *(See Table 8.)* This deficit is steadily diminishing over recent years as Japan has moved into surplus on new technology trade. That is to say, Japan's large current deficit in technology trade is the result largely of major contracts entered into some time ago. Its current account is favorable, one measure of progress in Japan's independent development of technology.

It is curious too that while the U.S., France and the U.K. have run surpluses on technology trade over the past decade, both West Germany and Japan have been in consistent and substantial deficit. No doubt some part of this is a function of catching up after wartime lags and destruction. Some part too, it has been suggested by the Japanese, may be the result of greater flexibility in these two countries and a greater willingness, and ability, to absorb technologies from abroad.

Japan's earlier dependence on imported technologies, and current need to move forward in independent development is very widely recognized in Japan both in public and in private circles. In the most recent *Science and Technology White Paper,* the issue was stated as follows:

"Our country has achieved very rapid progress since World War II, and in the advance of scientific knowledge as a technological revolution has contributed greatly to the progress of our economy and society. Further, in such sectors as electronics and industrial robots, a world-leading position has been attained. However, technology in our country as a whole depends largely on imported technology.

TABLE 6

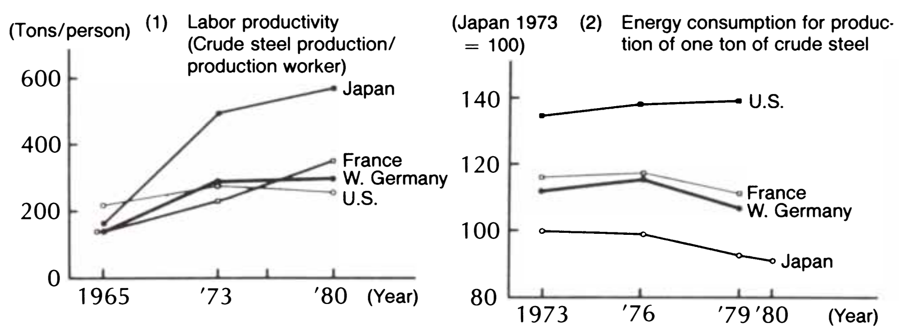
CONDITIONS IN PRINCIPAL DEPRESSED INDUSTRIES, 1981

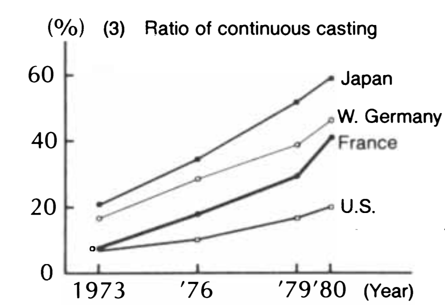
Industry (Number of Companies)	Employees (Number)	Capacity Utilization (Percent)	Operating Profits (¥Billion)
Petrochemical (12)	137,013	59.0%	–57.7
Chemical Fertilizer (34)	1,200	62.0	–30.5
PVC (17)	1,887	54.8	–47.5
Electric Furnace Steel (58)	31,300	80.0	–11.5
Steel Alloys (31)	6,978	51.0	–1.6
Aluminum Refining (5)	5,844	59.0	–56.3
Paper and Pulp (600)	67,500	79.8	27.9
Synthetic Fibers (18)	29,928	82.7	0.7

Source: Asahi Shimbun 9 December 1982.

TABLE 7

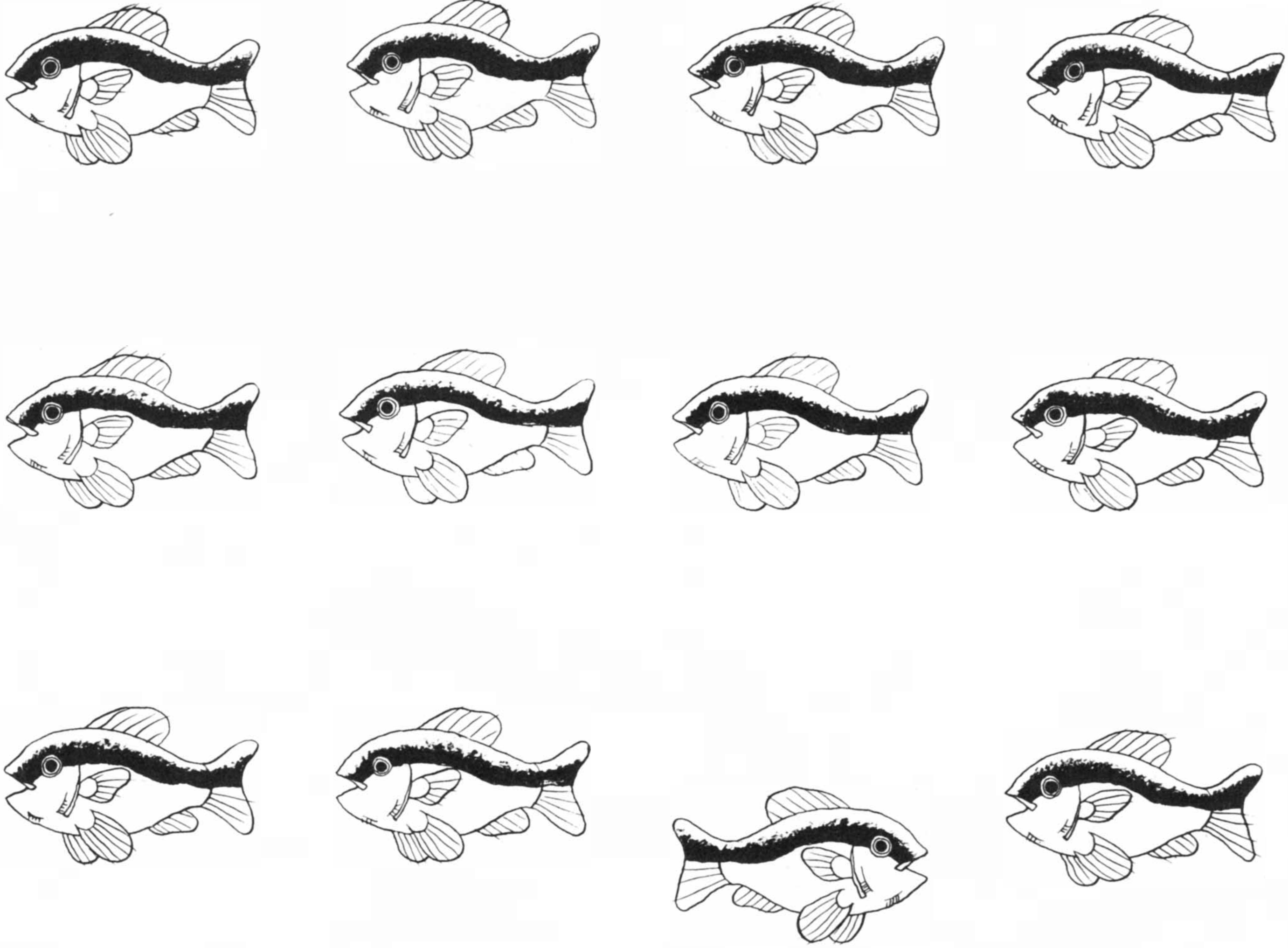
A COMPARISON OF STEEL-MAKING INDUSTRIES OF SELECTED COUNTRIES





Notes:
1. Sources: The Japan Iron and Steel Federation. *Summary Iron and Steel Statistics* and *Japanese and Foreign Iron and Steel Industries in 1981.*
2. Ratio of continuous casting = (production of continuous casting steel ingot) ÷ (production of crude steel) × 100

Source: Economic Survey of Japan 1981/1982. Economic Planning Agency, Japanese Government. The Japan Times, Ltd., Tokyo, p. 310.



A leader should occasionally strike out in a new direction.

Through the years, Sumitomo
has often attempted the unconventional,
a strategy not always successful.
But the number of times that it
has been has helped Sumitomo grow
into the world's 6th largest steelmaker,
a company well respected
in the industry as a technological
innovator and developer of high-grade
specialty steels,
many unavailable elsewhere.
Today, Sumitomo Metals is going
beyond steel production, exploring
challenging new areas like
energy and electronics.
But that shouldn't surprise you.
We've been striking out
in new directions for years.

SUMITOMO METALS

SUMITOMO METAL INDUSTRIES, LTD. Tokyo & Osaka, Japan
Sumitomo Metal America, Inc. Head Office: 420 Lexington Ave., New York, N.Y. 10017, U.S.A. Phone: (212)949-4760 Offices in Los Angeles, Chicago, Houston.
MAIN PRODUCTS: Tubes & pipe, flat rolled products, bars, shapes & wire rods, rolling stock parts, steel castings & forgings, and a wide range of steel products.

Instead of owning a cam
maybe you need a cam

In a world full of cookie-cutter SLRs, Nikon introduces the world's

Over the years millions of people have bought millions of SLRs designed for beginners.

But what happens when you don't want to take pictures like a beginner anymore?

Many stick with their mass-produced SLR and wonder why they can't get better results.

For the few, however, who are determined to improve as photographers, Nikon has created an engineering marvel called the FE2. An automatic SLR so different, there are no other cameras to compare it with.

For instance, take the matter of Selective Focusing. Which is the ability to vary apertures to express your impression of the subject. An absolute must for anyone who wants to improve as a photographer.

The FE2 gives you more options for Selective Focusing than any other 35mm SLR in the world.

To begin with, the FE2 is an aperture-priority automatic. So you have total control over depth of field when you shoot in automatic.

But, as we said, that's just the beginning.

The FE2 is the first 35mm automatic with a 1/4000 shutter speed. An innovation made possible by Nikon's development of an ultra-thin honeycomb titanium shutter. It means photographers like yourself will, for the first time, be able to use a full range of apertures in bright sunlight. And because the FE2 is faster than any other 35mm automatic, it gives you greater control over exposure than other cameras when using the

If you love photography, don't miss Nikon's "The Photographer's Eye" on SPN Cable TV. Check local listings. For further information write Dept. 50, Nikon Inc., 623 Stewart Ave., Garden City, N.Y. 11530. © Nikon Inc. 1983.

era designed for millions era engineered for a few.

fastest 35mm SLR automatic. The FE2.

latest 1000 ASA color print film.

The FE2 is also the only 35mm automatic with a 1/250 flash synch speed. Which means you'll have the ability to select a wider range of lens apertures when taking flash pictures outdoors. An achievement that was once considered impossible with a 35mm SLR.

In addition, the FE2 is one of the few cameras equipped with through-the-lens flash metering (TTL). Which gives you unparalleled accuracy because flash exposures are measured by the camera as they pass through the lens. And not by a sensor on the flash unit itself. TTL allows you to achieve virtually unlimited aperture settings when shooting flash indoors with a dedicated Nikon Speedlight.

Even when you're not taking photographs with the FE2 it's an astonishing camera. Its balance and 'feel' make other SLRs seem bulky and unresponsive. And its copper silumin aluminum body makes it one of the most rugged Nikons ever made.

So while other cameras are designed for millions, the FE2 is engineered to take the kind of pictures only a few strive for.

One-in-a-million.

Nikon

We take the world's greatest pictures.™

JAPANESE TECHNOLOGY TODAY

It might be said that developing on this basis has been a special feature of our economy.

"Our country has established a position as one of the advanced countries even in terms of technology. In order to deal with numerous problems both domestic and foreign, our country must free itself from dependence on imported technology, must open new paths to independent power regarding technology, and must press with greatest efforts creative research and development for finding and nurturing innovative technologies."

Given that Japan's social and economic ambitions require steady technological progress, and given that these increasingly depend on Japan's own efforts in development, a first issue is the scale of current investment in research. Two measures of Japan's investments in research are shown in *Table 9*: research expenditure and research personnel compared with other major economies. The U.S. spends by a great margin more on research than other countries. Japan's expenditure is about equal to that of West Germany, rather more than Britain and France combined.

However, Japan's rate of increase in research expenditure is much the highest of these countries, with West Germany next in rate of expenditure growth. An interesting, if somewhat academic, exercise is to note that if Japan and the U.S. continue to increase research budgets at the rate that they have over the past 15 years, by 1990 Japan's total research expenditure will exceed that of the U.S. Given straight-line projection of historical growth rates, the expenditure in 1990 for Japan would total ¥23,099 billion, and for the U.S. ¥21,325 billion. This would be a per capita expenditure for Japan twice that of the U.S. (The projection is fanciful since it assumes constant exchange, inflation and growth in expenditure rates.)

In terms of research personnel, Japan has about half the research personnel of the U.S. This suggests that on a per capita basis the labor effort in Japan in research is at about the same level as that in the U.S. It appears that Japan has about as many researchers deployed as do West Germany, the U.K. and France combined. By these measures, the Japanese research effort is now considerable, compares favorably with the research investments of the other major economies and is growing.

TABLE 8

TRENDS IN TECHNOLOGY BALANCE OF PAYMENTS
(Ratio: income/payments)

	United States	France	United Kingdom	West Germany	Japan Total	Japan Contracts
1971	10.56	0.85	1.07	0.39	0.12	0.71
1973	8.38	1.13	1.17	0.40	0.12	1.27
1975	9.09	1.27	1.02	0.42	0.23	1.42
1977	10.57	1.43	1.22	0.41	0.23	2.15
1979	8.77	—	1.29	0.46	0.27	1.94
1981	8.03	—	—	—	0.31	2.68

Source: Kagaku Gijutsu Hakusho (Showa 57 Nenhan) (Science and Technology White Paper). Kagaku Gijutsu Cho. Tokyo, 1983.

There is, of course, continuing and often heated discussion of the Japanese capacity to move to discovery and invention, as contrasted with application technology and adaptation. In a recent review of Japan's high technology industries, the Long-Term Credit Bank of Japan offered a balanced assessment:

"Several years ago it was a subject of debate among entrepreneurs, engineers, economists and other specialists as to whether the Japanese would ever take the lead in the newest technologies as they had done with conventional technologies.

"At that time the pessimists maintained a negative outlook for Japanese industry relative to frontier technology, for the following reasons. First there are no suitable grounds for invention and cultivation of original technologies in Japan. The inventive powers of the Japanese educational system and the R&D systems of private enterprises must be radically improved. In reality, however, such radical reforms are extremely hard to achieve. Second, Japan has no large-scale government-sponsored agencies like the U.S.'s NASA that support private industries' R&D efforts.

"Those pessimistic views were challenged by the positivists, however, who had ample evidence to the contrary. They argued that first, the Japanese have already invested in R&D efforts on the high technology frontiers and Japanese industries are energetic enough to meet the technological challenge of the advanced Western nations. In the past Japanese industry did not get enough government support for its basic research efforts, but recent years have witnessed the creation

TABLE 9

RESEARCH EXPENDITURE IN MAJOR COUNTRIES
(¥ billion)

	Japan	United States	USSR	United Kingdom	France	West Germany
1965	426	7,246	2,772 (1966)	911	730	716
1970	1,195	9,339	4,655	929	982	1,454
1975	2,622	10,438	6,885	1,410	1,813	2,293
1980	4,684	13,864	7,319 (1978)	1,463	2,241	4,899
CAGR*	17.3%	4.4%	8.4%	3.2%	7.8%	13.7%

RESEARCH PERSONNEL IN MAJOR COUNTRIES
(Thousand)

	Japan	United States	USSR	United Kingdom	France	West Germany
1965	117.6	494.5	664.6 (1967)	56.6	44.0	61.6
1970	172.0	540.9	927.7	54.7	57.3	78.4
1975	255.2	533.1	1,223.4	78.8	62.0	94.1
1980	302.6	643.5	1,373.3 (1978)	104.4	72.9	122.0
CAGR*	6.5%	1.8%	6.8%	4.2%	3.4%	4.7%

*Compounded Annual Growth Rate
Source: Professor K. Uno, Tsukuba University in *Nihon Keizai Shimbun,* 18 June 1983, p. 10.

of various government programs, such as the Creative Scientific Technology Promotion and Next-Generation Industrial Infrastructure Development. The outcome of these basic research programs will lay the foundation for their active application in the private sector, in the future. . . .

"As noted above, there are two conflicting views on the Japanese high-technology industry. However, judging from the recent spectacular growth of the semiconductor, robot, computer, optoelectronics and other high-technology industries, it appears that the Japanese will eventually gain an international edge in those industries. Still, there are various future technology areas, such as aerospace development, nuclear fusion, Josephson-junction technology, and discovery and invention of new principles where Japan will have difficulty catching up with advanced Western nations."

This assessment seems balanced and realistic. It should be added, however, that it provides an impression of more government support for research in Japan than the expenditure data indicate is warranted. The share of government in total research expenditures in Japan is notably low. The share of government expenditure in research is more nearly comparable, at least between Japan and the U.S., if defense-related research budgets are omitted from the figures. *(See Table 10.)* However, the U.S. defense budget does include substantial support for the semiconductor, computer, special metals and other industries' research. In fact, given U.S. budget pressures, the defense budget is no bad place to put a research subsidy. In contrast, given Japanese political attitudes and pressures, the defense budget is likely to be the last place in which to seek to get approval of research funding.

In any event, the U.S. nongovernment expenditure on research and development is, according to these sources, the equivalent of about ¥7.2 billion, while the nongovernment research expenditure in Japan is about ¥3.7 billion. On a per capita basis, with a population half that of the U.S., Japan's private sector R&D expenditure is now equal to that of the U.S.

These data provide only some measure of the effort that Japan has under way in developing technologies that will allow the process of continual renewal and restructuring of industries to take place. In fact, in the Japanese view progress in new technologies has been so rapid as to justify use of the term "new industrial revolution," in Japanese, *shinsangyo kakumei.* The term is widely used in Japan. It denotes in general new approaches to organization and production systems, first in the factory and now in the office as well. The core technology for this new industrial revolution, as the Japanese see it, is microelectronics and all of its applications through to what is being called the "information industry," the collecting, transmitting, analyzing and presenting of data.

Japan's electronics industry is rapidly on the way to displacing such traditionally critical industries as textiles and steel in the economy. *(See Table 11).* The long decline of textiles is clear and expected. The beginnings of the relative decline of the steel industry can be seen. The auto industry continues to increase its share of Japan's total manufacturing, but the fast growth is that of the electronics industry, now overtaking steel in the position it occupies in total manufacturing.

The industry is about equally divided in terms of sales value into three major categories: consumer electronics, industrial electronic products and electronic components. In a real sense the three sectors are mutually dependent, since developments and cost improvements in one sector can favorably impact other sectors. Further, as new approaches are developed, the interactions among the sectors become more evident. The computerized home will use the television set and the telephone for data transmittal. The components supplied to the consumer sector—integrated circuits and cathode-ray tubes, for example—gain volume and reduced cost for application to industrial products. It is an interactive industry, despite its disparate product applications.

Japan's first successes in electronics were in the consumer sector. It is a sector of particular interest in terms of technology because it not only illustrates the steady restructuring of an industry to higher value added goods but also illustrates especially well the changing position of Japanese technology in terms of world competition.

Table 12 displays the shifts in Japanese consumer-electronics output over some three decades. In the immediate postwar period the industry was largely made up of the production of radios. Radio receiver production has for some years now largely moved out of Japan into the newly industrializing economies of East Asia, with Japan retaining production of only the most sophisticated sets. There are following a series of product life cycles, as monochrome television grew, reached market saturation, gave way to color television and declined. Again monochrome television has for some years now been produced largely in Taiwan and South Korea. Color television production as a percent of the total of consumer electronics in Japan has also had its day, and audio equipment went into a sharp decline in the past few years (to be revived by the new digital player technology?).

The consumer-electronics sector in Japan is now becoming dominated by videotape recorders, now making up nearly 40 percent of total consumer-electronics

TABLE 10

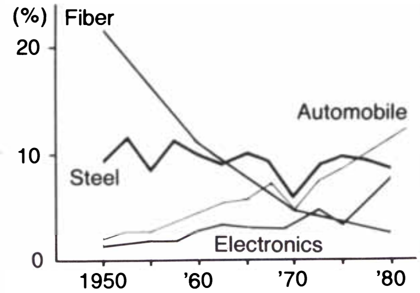
GOVERNMENTS' SHARE OF RESEARCH BUDGETS

	Total Research Budget (¥ trillion)	Government Share: Including Defense Research Expenditure (percent)	Government Share: Excluding Defense Research Expenditure (percent)
US	13.9 (1980)	48%	33%
West Germany	4.3 (1979)	44	41
France	2.2 (1979)	58	47
UK	1.5 (1978)	48	32
Japan	5.2 (1980)	28	28

Source: Journal of Japanese Trade & Industry, No. 4, 1983

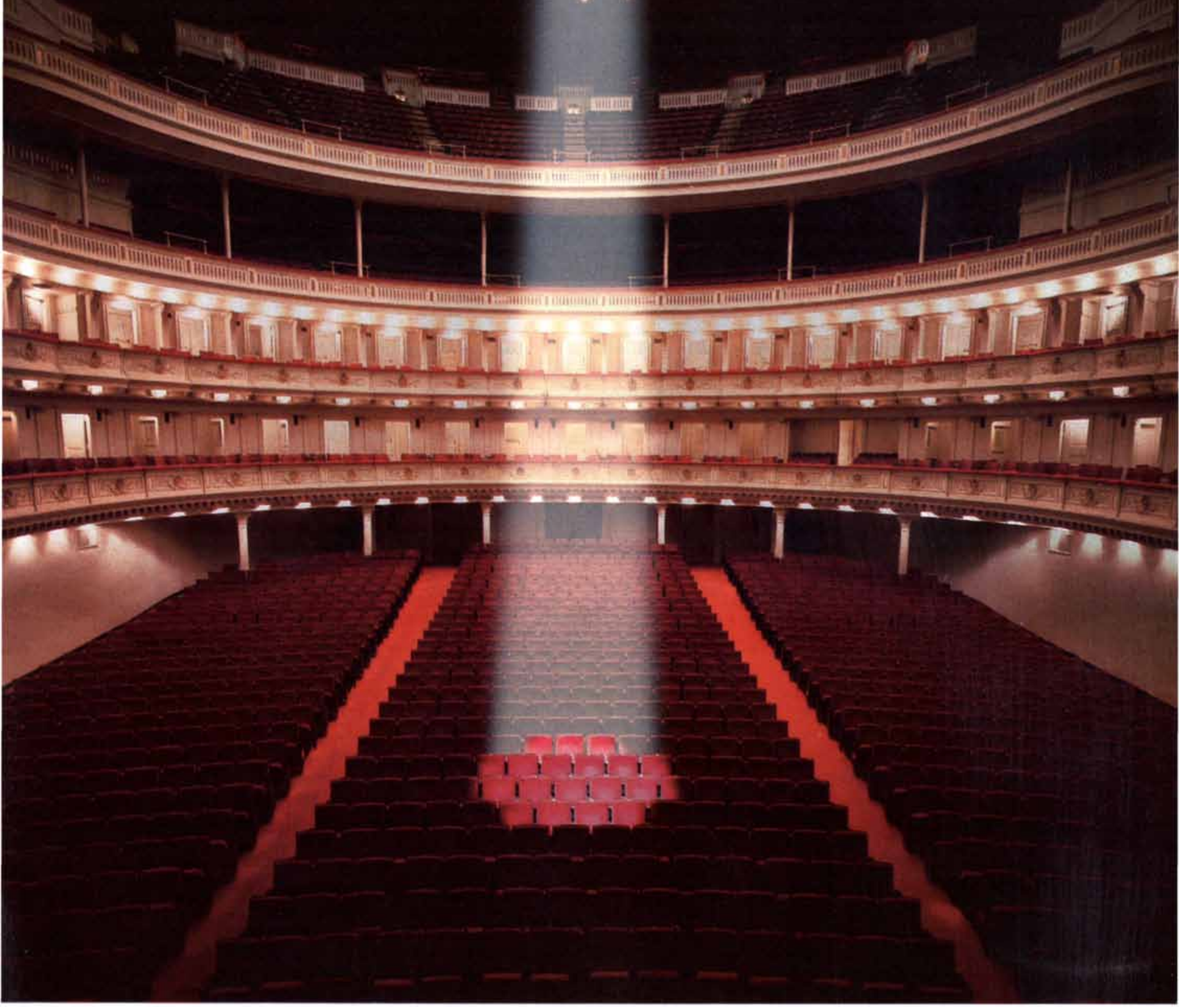
TABLE 11

TRENDS IN COMPOSITION RATIO OF MAJOR MANUFACTURING INDUSTRIES



* Based on the shipment amounts in the Table on Manufacturing Statistics by the Ministry of International Trade and Industry, Japanese Government.
* The figures for the electronics industry exclude those of electronic desk calculators.

Source: Annual Data on Japan's Electronics Industry, 1983 Edition, Electronics Association of Japan, Tokyo, 1983.



output. The product has had an astonishing history. Production began in 1975. The compound annual growth rate of production in seven years since has been 96 percent, barely failing to meet the level of an annual doubling. Exports began in 1976, and have in fact somewhat more than doubled each year since, with a compound annual growth rate of 106 percent. The growth rate was at a more than doubling pace between 1980 and 1981 but slowed in 1982, no doubt due to depression-induced slowing of demand and necessary inventory adjustments. News reports suggest demand is surging again in 1983, both in Japan and abroad.

The product progression is an interesting one in terms of technology. Clearly consumer electronics has moved very rapidly to higher technologies across these several decades. Clearly too Japan was dependent on foreign sources for virtually all of the technology employed even to the stage of color television. RCA licenses made Japanese color-television production possible. Corning Glass supplied glass-tube technology to Japan's Asahi Glass, and lived to regret the fact bitterly and loudly. The early semiconductor production in Japan for the whole range of consumer electronics was available to Japan under license from Western Electric, Fairchild, Texas Instruments and others who held the basic patents. Perhaps more development of Japan's consumer electronics industry is the result of imported technology in the hands of aggressive and highly competitive Japanese firms that adapted, improved, drove costs down and came to dominate world markets.

It is hardly necessary to note as well that Japan's consumer electronics industry has been the subject of especially bitter trade complaints with recurrent charges of dumping, copying, conspiracy, government subsidy and the like. Quotas have been imposed by the U.S. and other countries. Japanese producers have found it necessary to set up production in Western Europe and the U.S. to avoid such trade barriers (which explains in part the life cycle configurations in *Table 12)*.

It is this background that makes the case of the videotape recorder the more interesting. This is essentially Japanese technology. Indeed, Japan very nearly monopolizes world production, with its share of world VTR output generally estimated at more than 90 percent. Only Philips and associated Western European companies are in VTR production, and they have contrived to impose import quotas on Japanese VTR exports to the EEC market. There is no VTR production by U.S. producers. Instead, U.S. consumer electronics firms source VTR from Japanese producers for sale under their own brands in the U.S. market. The VTR thus represents a full turn of the wheel, a Japan-developed product with trade friction minimal because Japan is virtually the sole producer.

Several themes implicit in this case have disturbing implications for U.S. and European trade positions: the unwillingness of U.S. producers to risk investment in this new product, the speed with which the Japanese moved in developing the VTR market and investing accordingly, the dominant position this success threatens technology is extended to cameras and films and other applications. In only 30 years Japanese consumer electronics has moved from marginal production of the cheapest and lowest-technology radio receivers to a dominant position in a massive new product area.

If consumer electronics as an industry segment is especially visible, and therefore especially trade-sensitive, so electronic components may be considered the basis for an overall position in the industry. It has been suggested frequently that the semiconductor industry is the equivalent in the late 20th century of the steel industry in an earlier period of industrialization—the basic industry that must be in place in order to maintain a strong economy. Certainly the explosion in microelectronics is based on semiconductors, and advances in semiconductor technology largely determine the direction and rate of developments across the whole range of electronic-related products.

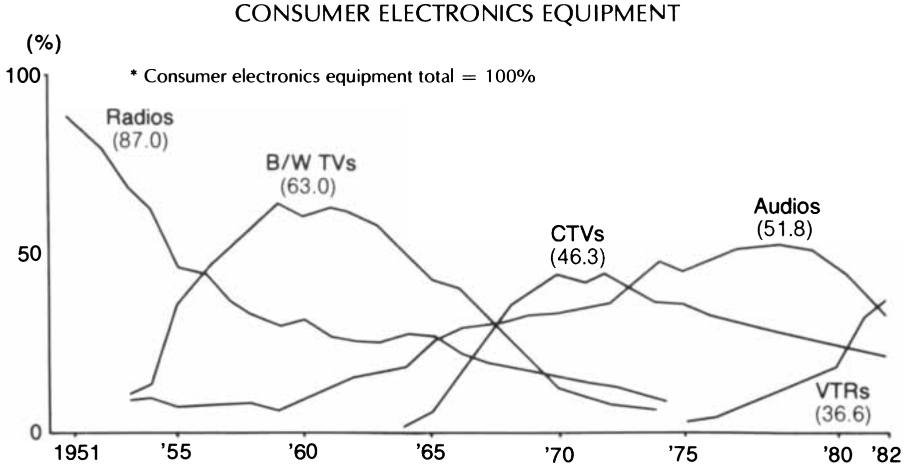
Table 13 reviews the Japanese experience to date in integrated circuits. Production in Japan began in 1970 in small volumes, and in the usual and expected pattern, exports were insignificant while domestic demand was being focused on and while cost and quality were improving to world levels. Japanese integrated-circuitry production has been growing in value terms at about 25 percent per year since volume production began, and is growing at about that rate into 1983. These value data mask a much faster increase in unit output, since prices of integrated circuits fall rapidly as yields improve with expanding volume. Japan continues to import substantial quantities of integrated circuits with import sales increasing about 16 percent per year. The result, however, is that Japanese dependence on integrated-circuit imports has been falling steadily from a 1974 high of about 30 percent of domestic demand, and continues to decline as Japanese producers increase both their share of the domestic market and increase the proportion of production being exported. Japan's trade balance for integrated circuits has swung from negative only four years ago to a strongly positive balance.

It is of special interest to note the destination of Japanese IC exports. In 1982, 38 percent of Japan's total exports of IC's went to Asian markets, while 41 percent went to the U.S. and only 16 percent to Europe. This is yet another example of the benefits to Japan's industry of the industrializing countries of Asia. As production of simpler electronic products moves offshore from Japan to developing Asian economies, those economies in turn become large markets for the more sophisticated products Japan is moving to. The cycle is beneficial to both parties.

The competition in semiconductors is

TABLE 12

TRENDS IN THE PRODUCTION COMPOSITION RATIO OF MAJOR CONSUMER ELECTRONICS EQUIPMENT



Source: Annual Data on Japan's Electronics Industry, 1983 Edition. Electronics Association of Japan, Tokyo, 1983.

慮

CONSIDERATION

Japan is a country where a visitor is indeed a guest. And treated as such. With special considerations you'll discover throughout your stay.

In this tradition of hospitality, Japan Air Lines extends as much consideration to business travelers preflight and postflight as inflight. JAL offers unique Executive Service that not only makes business travel more pleasant; it helps you get the most out of your stay.

BILINGUAL BUSINESS CARDS A necessity for doing business in the Orient. For a small fee, JAL can translate business cards into Japanese or Chinese and have them awaiting your arrival in Japan or Hong Kong.

HOTEL RESERVATIONS AND SPECIAL RATES Use JAL's computer system to obtain reservations, business rates and special services through JAL's Executive Hotel Service. The hotels are among the most convenient and luxurious in the Orient.

LANGUAGE CASSETTES, TRAVEL INFORMATION Enjoyable, informative books on everything from business protocol in the Orient to nightlife. Plus the JAL Jet Age Language Courses, with Chinese and Japanese cassette tapes to help overcome the language barrier.

AND MORE There are other ways to fly to the Orient. But only Japan Air Lines is prepared to offer you so much consideration even before you depart.

TO BE PREPARED IS EVERYTHING.

JAL's unique Executive Service is available to all JAL passengers. For details, call **800-835-2246 Ext. 281.** Or send us this coupon.

JAPAN AIR LINES
Japan Air Lines, P.O. Box 10618,
Long Island City, New York 11101

Dear JAL: Please send me the JAL Executive Service Folder.

Name ______________________________

Address ______________________________

City ______________ State ______________ Zip ______________

TABLE 13

BALANCE OF JAPANESE TRADE IN INTEGRATED CIRCUITS
(¥ billion)

Year	Production	Imports	Exports	Reliance on Imports (percent)	Export Ratio (percent)	Trade Balance
1970	53.2	20.6	—	27.9%	0.0%	—
1972	72.3	16.4	—	18.5	0.0	—
1974	125.5	51.1	Y6.7	30.1	5.3	−44.4
1976	197.1	62.7	22.7	26.5	11.5	−40.0
1978	281.4	61.3	52.2	21.1	18.6	−9.1
1980	570.2	108.4	182.1	21.8	31.9	73.7
1982	835.2	127.4	285.1	18.8	34.1	157.7
CAGR*	25.8%	16.4%	59.8%	—	—	—

*Compounded Annual Growth Rate

Note: First five months 1983, compared with the same period 1982: production +25.5%; imports +7.8%; exports +56.3%; exports to U.S., +68.9%.

Source: Japan Long-Term Credit Bank; Electronics Industry Association of Japan.

essentially between U.S. and Japanese firms. According to a recent report in *Business Week*, 23 May 1983, of the top 10 producers of integrated circuits in the world, four are Japanese firms. Hitachi and NEC are ranked third and fourth in volumes after Texas Instruments and IBM, with Fujitsu and Toshiba also in the top 10. Only one European firm is in this group, Philips, whose position declined from fourth to sixth in production ranking from 1979 to 1983.

As earlier with consumer electronics, and no doubt in part due to having that example in mind, the U.S. industry has for several years been seeking various forms of support and protection from the U.S. government on the grounds that the Japanese industry enjoys unfair advantages from assistance provided it by the Japanese government. In some respects the U.S. industry's concerns seem unwarranted. For some time U.S. companies have held about a two-thirds share of world production, with the Japanese share steady at perhaps one-quarter. Japan now supplies about 5 percent of U.S. demand, while the U.S. producers supply some 10 percent of Japanese demand.

However, when looked at from a U.S. perspective, some of the trade trends are alarming. *Table 14* shows U.S. and Japanese trade balances in semiconductors over the past five years, and the level of exports between the two countries. Within this short period the U.S. has moved from a modest trade surplus in semiconductors to a considerable deficit in 1982. Japan has been in steadily increasing trade surplus through the period. U.S. exports to Japan of semiconductors have not quite doubled over the five-year period, while Japanese exports to the U.S. have increased by more than four times.

There is the possibility currently that the relatively stable overall shares of U.S. and Japanese producers that has prevailed for some time on a world-market basis may be deteriorating to the disadvantage of the U.S. There is first the fact that the Japanese position in semiconductors has tended to improve relative to competitors with each succeeding generation of technology. Thus Japan's position in the 64K RAM has been stronger than in the earlier generations of memory technology. The Japanese producers have been early in announcing sampling, and initial production of the 256K RAM, the follow-on generation of technology.

In addition to this technological progress, and to a considerable degree its cause, has been the high level and growth rate of Japanese investment in this industry. The situation was reviewed recently by the Japan Economic Institute:

"U.S. semiconductor makers have to shoulder a large part of the responsibility for letting their Japanese rivals get a foothold in the American MOS memory market. When semiconductor orders hit the skids during the 1975 American recession, the industry slashed capital spending, cutting it in half according to some estimates. Companies also pared their work forces . . . . Japanese manufacturers, in contrast, continued to add capacity and invest in new equipment over the course of the recession, with an eye to remaining competitive in the computer, telecommunications and other end-equipment markets they supply as vertical producers."

This earlier strategic error was perceived by U.S. producers, who determined not to repeat it in the future, regardless of profit consequences. This resolution underestimated both the depth and duration of the recent recession, and the capacity of Japanese competitors to step up investment levels. The Japan Economic Institute reviews the situation:

"With prices firming and then turning up this year, the 64K RAM finally became a consistently profitable item for suppliers. . . . MOS memory makers in the U.S. as well as Japan will need to wring all the money they can from the chip to underwrite the huge ($50 to $100 million minimum per company) R&D and equipment costs of the 256K RAM generation.

"In fact, overall semiconductor R&D and capital spending in the U.S. has been unusually strong in recent years, even during the recession. Plant and equipment expenditures, for example, rose sharply during the late 1970's and then jumped in 1980. Determined not to repeat the mistake they made during the 1975 slump, U.S. semiconductor makers maintained 1980's high level of capital investment through 1981. They had hoped to duplicate this feat again last year, but weak prices and the recession dashed these plans. Still, capital spending was off only modestly—perhaps 10 percent—from the 1980-81 peak.

TABLE 14

TRADE BALANCE IN SEMICONDUCTORS
($ million US)

	United States	Japan	United States Exports to Japan	Japan Exports to United States
1978	$166.9	$91.6	$105.6	$134.1
1979	191.4	187.5	172.2	242.0
1980	136.9	469.2	151.5	397.2
1981	19.5	545.5	168.9	385.4
1982	−330.3	780.8	190.4	583.6

Source: Japan Economic Institute Report, 12 August 1983, from U.S. Department of Commerce.

There are plenty of sporty cars that will turn heads today. But there aren't many that will turn corners like a Toyota Celica GT. Because Celica has more than just an eye-catching exterior. It has the right stuff underneath, too.

The 1983 Celica GT Liftback has the most powerful engine ever offered in a Celica, a 2.4 liter power plant with a new electronic fuel injection system. From now on, heads better turn quickly, or they won't see Celica at all.

Celicas also come standard with a close ratio 5-speed overdrive transmission. And a 4-speed automatic overdrive is available.

There's even a new Celica model for 1983 — the Celica GT-S (pictured below). It introduces racing-type independent rear suspension to the Celica line. Celica GT-S also comes standard with the biggest tires in the class — extra-wide 225/60 HR 14 steel-belted radials. And they're mounted on gleaming 14" x 7" aluminum alloy wheels. So Celica turns corners with a precision the rest of the sporty car field just can't match.

Other Celica GT-S features include fender flares, for a racy look. And inside, multi-adjustable sports seats that really hug you as you turn those corners.

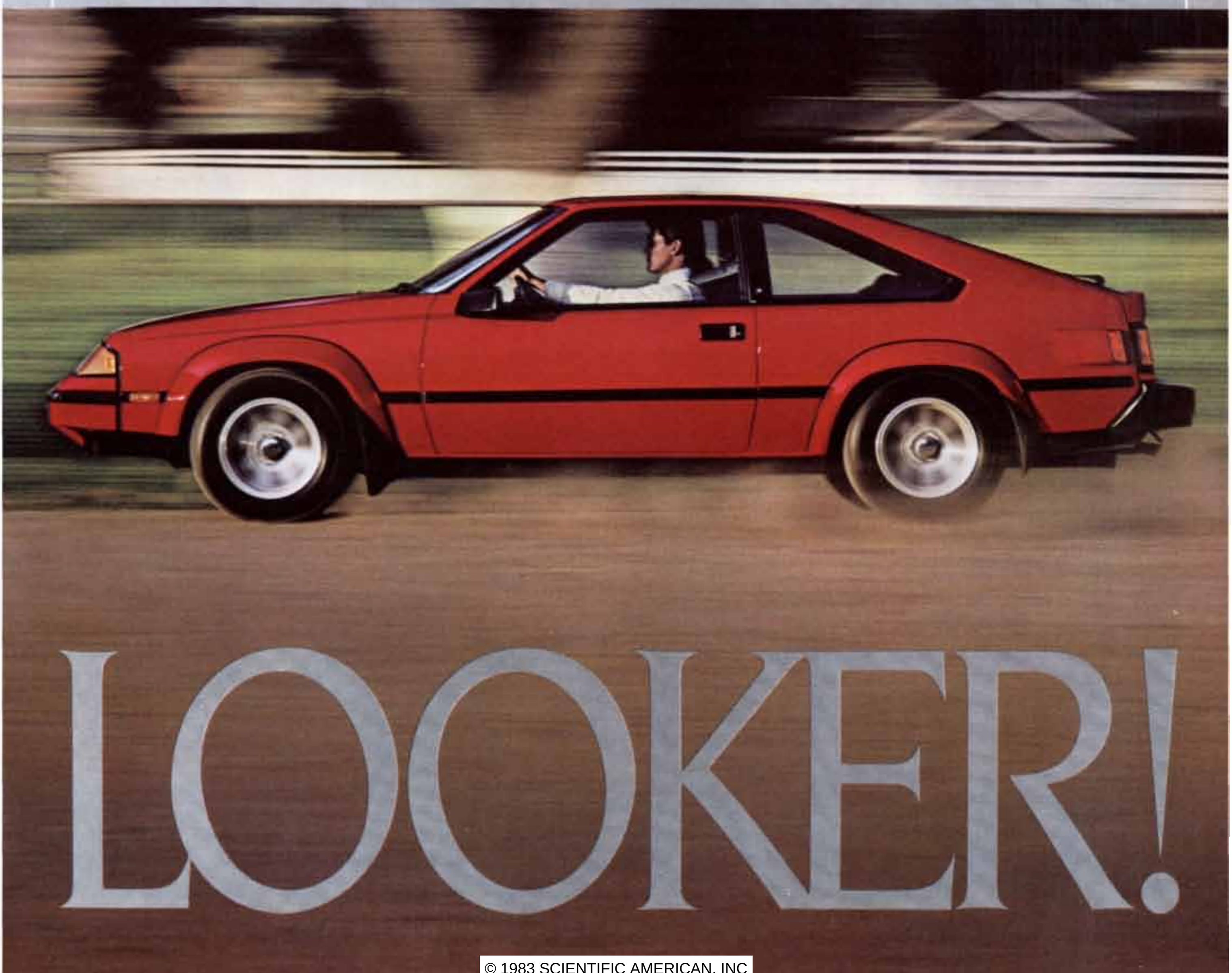
The 1983 Toyota Celica GT. More than just a hot-looking car. It's a turn for the better!

BUCKLE UP—IT'S A GOOD FEELING!

OH WHAT A FEELING!

TOYOTA

CELICA TURNS HEADS AS EASILY AS IT TURNS CORNERS.



future communications

Beyond electronics lie optoelectronics

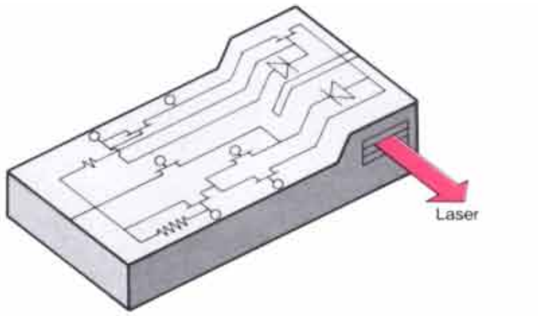
Today's digital data are usually transmitted using electronic circuits capable of frequencies of several hundred megahertz at most. But to attain the quality and reliability required of large-capacity communications in tomorrow's information society, optical transmissions in gigahertz or higher frequency bands have to be achieved. Only an optoelectronic integrated circuit (OEIC) combining a semiconductor laser or photodiode and an electronic integrated circuit (FET) would be able to provide the needed attributes. Yet, with present-day electronic circuit technology, basically a planar IC structure, it is difficult to construct both an electronic circuit and an optical device with a 3-D structure on a single substrate.

Hitachi applies micrometric skills

Using experience and know-how acquired in creating very-large-scale ICs and semiconductor lasers, Hitachi engineers have succeeded in producing a hybrid structure of optical and electronic devices on a single semi-insulating gallium/arsenic substrate. The laser diode which results emits light with an output power of several mW. And because it is completely integrated with FETs, the finished unit is compact, lightweight, and capable of high-speed gigahertz operation with improved reliability—another major breakthrough for Hitachi technology.

Limitless possibilities unveiled

This prototype optoelectronic IC is the very first important step toward realization of large-capacity communications and computers using high-speed optical processing. It can be used as a light transmitter, modulator or repeater in optical communications, or as a low-noise laser source or sensor in optoelectronic equipment. In fact, it represents an entirely new approach to handling data signals and coded information at the speed of light.



Technical excellence is embodied in all Hitachi products

The development of an optoelectronic integrated circuit is just one case demonstrating Hitachi's technological strength. You'll find other examples in all of Hitachi's products, from laser diodes and very-large-scale ICs to a whole host of high-quality electronic appliances. Our comprehensive technical expertise is your guarantee of convenience, easy operation and high reliability in every product that bears the Hitachi brand.

HITACHI

A World Leader in Technology

"Aside from worrying about financing escalating equipment and development requirements out of weak profits, U.S. semiconductor makers were looking nervously over their shoulders at what their Japanese rivals were doing. What they saw was disquieting. Like American firms, Japanese manufacturers significantly expanded capital spending during the late 1970's, and they continued to do so in both 1981 and 1982. By fiscal 1982, as a result, investment in yen terms was almost three times higher than five years before. Equally frightening to U.S. companies, Japanese semiconductor producers spent almost as much on plant and equipment in 1982 as American merchant makers did last year. Moreover, the six firms that compete in the world MOS memory market invested the equivalent of 20.3 percent of their overall sales in plant and equipment in 1982, compared with 13.7 percent in 1978, while capital spending by leading U.S. merchant companies average an estimated 15 percent of sales last year, down two points from 1981."

As the U.S. market strengthens, it is quite possible that there will be a shortage of product in 1984. This will again provide Japanese makers an opportunity to take share from capacity-constrained U.S. makers. Further, these investment rates make clear that Japan's semiconductor industry will rather soon be at least equal in size to that of the U.S. Finally these investment rates help ensure Japanese success in succeeding technologies.

The rate of technological advance in an economy is, of course, in part a function of the research capability and effort. It must be evident, however, that the application of that technology depends basically on the rate of capital investment. Invention will flow to and be applied in that economy that is fastest growing, i.e., investing most heavily in new plant and equipment. Given reasonably effective management decisions, new plant and equipment will by definition incorporate the latest and most efficient technology.

It has often been noted that in the past decades the U.K. has been an important source of significant developments in science, and has often been the initiator in the application of new technologies. However, the U.K. has for the most part during this period not been the successful site for the full application and commercialization of these technologies. No doubt the causes are many, but surely the continuing low rate of capital investment in manufacturing is significant.

The Japanese case is at the other extreme. While Japan has as yet no great distinction in discovery and invention, as the cases of consumer electronics and semiconductors suggest, Japan has been the site of successful application.

Japan's capital investment rate is exceptionally high. From 1962 to 1980, the annual increase in capital investment per employee in Japan has been about 11 percent. At the other extreme, over the same period capital investment per employee in the U.K. has been about 1 percent per employee, and it has been only slightly over 3 percent in the U.S. West Germany is intermediate, with an annual increase in capital stock per employee of about 7 percent. (See *Economic Survey of Japan 1981/1982,* p. 195.)

With the decline of the raw-material-processing industries of Japan, which are capital-intensive and had earlier drawn most of Japan's capital investment, these investment flows are increasingly into the so-called high-technology sectors, and electronics is the center of this investment flow. The capital requirements of semiconductor manufacturers are very great, as we have seen. Investment largely determines level of technology. It determines capacity levels and therefore determines the market-share winners under conditions of high demand.

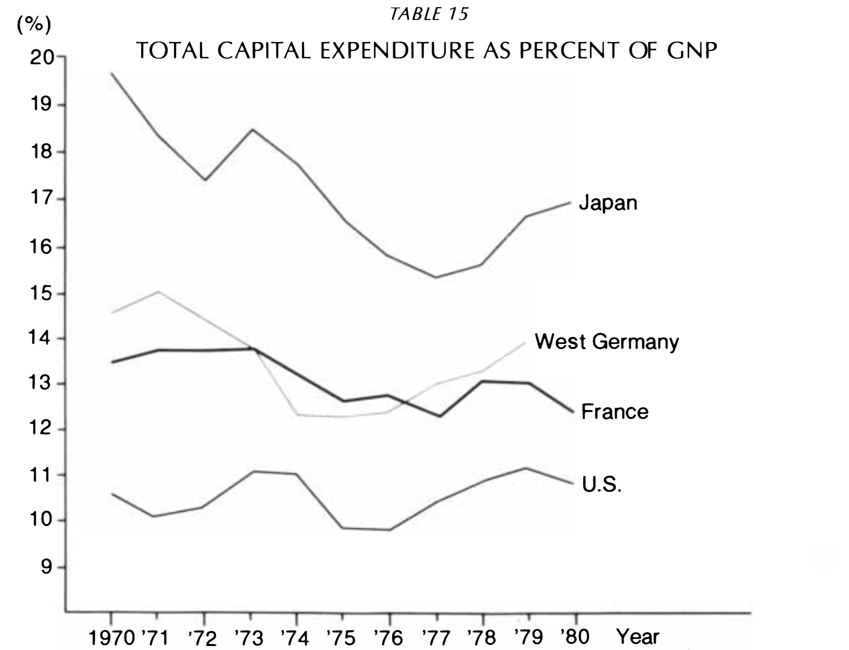
There has been a certain fashion in the U.S. to speak of "supply-side" economics, presumably a focus on savings and investment as the key to sustained economic growth in contrast to a focus on demand stimulation. It is not too much of an exaggeration to suggest that "supply-side" economics is alive, well and living in Japan. Japan's high savings rate provides the base for the kinds of investment levels shown in *Table 15*. The competitive context for Japanese business provides the stimulus to investment of these savings. The propensity for Japanese management and shareholders to take a long view of the company's future—with a focus on growth and market share—helps ensure that investment levels continue to be high, even though near-term returns would appear not to justify the investments.

From all this emerges the dilemma of the U.S. semiconductor industry when faced with Japanese competitors. The U.S. industry complains of Japanese government support and "targeting" in explaining the continuing advance of the Japanese competitors into what had been an American domain. This has given rise to much controversy about which government, the U.S. or Japanese, provides the greatest support to the industry. The argument seems largely to miss the basic issue. The Japanese government has not imposed investment targets on Japanese firms, nor has the U.S. government put constraints on the investment levels of U.S. firms. Investment is driving this process, and the differences in levels of investment between the two countries are caused by quite basic differences in how the two economies function.

This is not to say that the semiconductor race goes by default to Japanese producers. The top U.S. companies in the field are holding position, and are determined to continue to do so. It is to say however that the future of electronics is

TABLE 15

TOTAL CAPITAL EXPENDITURE AS PERCENT OF GNP



Source: MITI (1981)

TABLE 17

INDUSTRIAL ROBOT DEMAND PROJECTION
(¥ billion)

	1980	1985	1990	Annual growth rate 1985/1980	Annual growth rate 1990/1985
Domestic demand	60	200 – 250	370 – 500	30%	14%
Exports	(Negligible)	40 – 50	80 – 100	—	15%
Total	60	240 – 300	450 – 600	35%	14%
Annual production (thousand units)	19.7	43 – 55	75 – 94	—	—

Source: Japan Industrial Robot Association

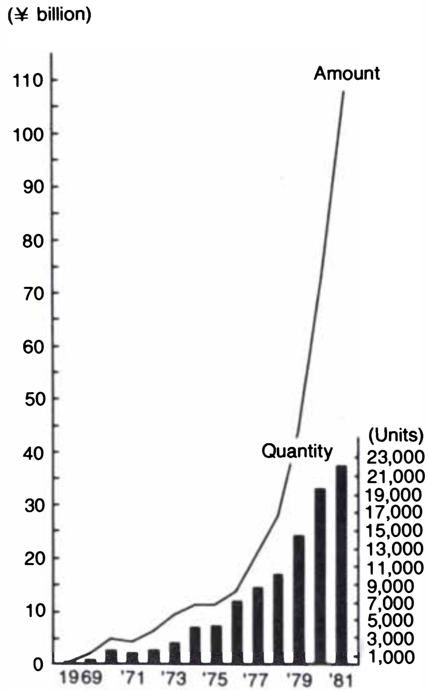
no longer to be determined only in the U.S.; Japan is fully equipped to play an equal role in future developments and will do so, and efforts to obtain trade protection for the U.S. industry will not be helpful either for the U.S. or for Japan.

Semiconductors by themselves are, as steel has been, the base on which industries are to be built. It is in their applications that their value is realized. In the context of Japan today, two broad areas of technology development based on semiconductors are showing rapid development. These are the application of microelectronics to the factory and to the office, the areas of factory automation and office automation. In each of these areas technological developments in Japan are moving very rapidly indeed.

The industrial robot and its development in Japan incorporates nearly all the main themes of Japanese technological advance. Robots in the modern industrial sense were first developed in the U.S., with substantial support from U.S. government agencies interested in their applications in such areas as nuclear plants and space programs. The first robot for commercial use was shown in Japan by a U.S. firm in 1967. The initial response in Japan, as in the U.S., was limited since the machines were still crude and costly. Initial production in Japan was undertaken under license from a U.S. firm, Unimation, a pioneer in commercialization of robots.

TABLE 16

JAPANESE PRODUCTION OF INDUSTRIAL ROBOTS



Source: Japan Industrial Robot Association

The field developed very slowly in Japan, and it is important to note, in terms of the conventional views of Japan, that the Japanese government and its various agencies showed no interest in the product. The application of robots developed very little until the mid-1970's. It was at that time that Japanese firms came under tremendous pressure to reduce costs in every way possible. The focus of management attention in Japan shifted from dealing with fast growth to dealing with the consequences of exploding energy costs, very high inflation and a depression. It became the era of *genryo keiei*, of slimming or dieting management. The application of robots to the factories of Japan took off in that period. *(See Table 16.)*

Robots are not easy to define, in the sense that any sophisticated machine tool has some of the attributes of robots, and the definition of a robot can be quite broad or very narrow and specific. By any definition, however, Japanese industry now has in place on the factory floor a multiple of the number of robots to be found in any other economy. By one recent survey, using a very demanding definition, Japan has in operation about twice as many robots as does the U.S., or four times as many on a per capita basis.

The field of robotics has exploded in Japan in recent years. It is estimated that 190 firms are in robot production. Fundamental studies of robotics are being undertaken in 85 universities and public institutions across the country, the number of participating institutions having doubled in the last six years. The main themes of robot R&D are increasing speed, size reduction, computer control, weight reduction and modular design. The Long-Term Credit Bank concludes that "assembly robots that are capable of performing sophisticated functions will, in all likelihood, come into widespread use by 1985."

Clearly, this development is driven by the same factor that is driving Japanese progress in semiconductors, the high level of capital investment in Japan. It is not surprising to find that Japan's electrical and electronics companies are at the forefront both of development and application of robots. The high rates of capital investment in new plant and equipment provide the opportunity to set up new production facilities with the latest in all forms of automation, including robots.

Robots are expected to become a big business in Japan. *(See Table 17.)* The robot market is expected to reach about $1 billion by 1985, and to double again in the next five years. Unit output is expected to increase by some four times over the decade of the 1980's. These estimates may well be too low, as some manufacturers are talking of having tens of thousands of robots in place in a single company by the end of the decade. In any event, increased volumes will reduce costs and prices, and functions can be expected to be added in parallel. This seems likely to be a product area in which the considerable lead Japan now holds over other economies is likely to be maintained or to increase. Japanese capital investment levels alone virtually guarantee this outcome.

Just as robots are typical in being an improved-on imported technology, and in being driven in development by Japan's investment rates, so the field is representative of the responses of the Japanese government. It was only well after the initial surge of interest and growth of the robot field that the Japanese government became involved. There was no special prescience shown by the government. It is sometimes suggested that the government of Japan has shown unusual skills in "picking winners," products, companies and

NEC is the title sponsor of the Davis Cup and Federation Cup worldwide

C&C
COMPUTER AND COMMUNICATIONS
TECHNOLOGY

How NEC helps curb rising costs

At NEC, we have always believed that automation effectively raises productivity and lowers business costs.

Through "C&C," our unique integrated computer and communications technology, we're proving this to be true.

The NEC digital EPBX system, based on 80 years of know-how in the telecommunications field, illustrates this cost-cutting potential. Utilizing the latest fiber optics transmission techniques, it lets you link voice and data systems into a single cost-effective communications network accessible to all.

Computers incorporating our own advanced semiconductors plug nicely into this network, trimming profit-eating expenditures all along the way. This applies as much to our large general-purpose computers and ASTRA business computers as it does to personal computers like PC6000, 8800, and 16-bit APC.

The NEC teleconference system takes our philosophy yet another step. Combining a variety of data processing systems with communications systems, it can save businesses millions of dollars annually—and business people thousands of hours of needless travel.

These are only a few of over 15,000 different industrial, business and home electronics products manufactured by NEC. They spearhead our drive to boost productivity and curb costs by way of "C&C"—a concern shared by our customers in over 140 countries.

NEC

NEC Corporation

P.O. Box 1 Takanawa Tokyo Japan

industries that will provide for the future of the economy. There is very little evidence of any special skill in this regard on the part of the government of Japan. That government is, however, quite capable of mobilizing effective supporting programs when it becomes apparent that a sector is of potentially high value.

In robots it was not until 1980, well after considerable private-sector activity and developments, that government programs in support of robot development took place. These have, sensibly, focused on supporting the rapid growth of the market in Japan for robots. The Long-Term Credit Bank notes four measures to promote the industrial robot industry: (1) A 13 percent write-off of purchase price as a special depreciation; (2) Smaller companies using robots in dangerous workplaces can borrow low-interest money for their purchase; (3) Smaller companies using robots to modernize production facilities can borrow from the Medium and Small Scale Business Modernization Funds; (4) The Japan Robot Leasing Company, Ltd., was set up to lease robots drawing on capital from the Japan Development Bank and other city banks.

Robots are by no means the only sector in which electronics applications to industrial production can be made. The most pervasive application is the numerical control of machine tools. Again U.S. companies were pioneers in the field, but, as *Table 18* shows quite clearly, it is in Japan that numerically controlled machine tools have found most rapid acceptance. Through the mid-1970's production in the two countries was about equal in volume. Japan now out-produces the U.S. industry by several times. Again we are no doubt seeing the impact of relatively heavier capital investment in manufacturing.

The rates of growth in turn lead to sharply lower costs and prices for this equipment. According to the Japan Machine Tool Builders Association, in 1981 a numerically controlled lathe was priced in Japan at about $55,000, compared with a price in the U.S. of $218,000, a difference of nearly four times. Thus a cycle of increased use leads to lower prices, which in turn lead to increased use. As a result robots and numerically controlled machine tools are at a price level where they can be purchased by the large numbers of small and sub-small businesses and sub- Japan's total production capacity.

The rapid application of electronics to the factory production system has now a parallel in Japan in the automation of the office. It has long been customary to describe the Japanese office as notably inefficient, overstaffed and poorly organized. Since office productivity is generally difficult to measure, it is rather difficult to know the extent to which this conventional view has been correct. In any event, as productivity improvements through automation have been achieved in the factory, it is natural enough that the office becomes a focus of attention. This is the more true as Japanese firms have a propensity to value information highly and as the accumulation, processing and transmitting of information becomes both more important and more difficult from the sheer volumes of information available.

The progress of office automation in Japan has been much influenced by the special requirements of the Japanese language. Effective written communication in Japanese requires the use of two syllabaries of about 50 elements in each, and in addition a minimum of about 2,000 Chinese-derived ideographs. Easy written communication would require more ideographs, to a total of perhaps 3,500. It is obvious that under these conditions an effective mechanical typewriter is simply not possible. Voice communications—meetings and telephony—and handwritten communications are required.

To a Western executive struggling from under a mountain of typewritten memoranda and other communications, the Japanese system might seem to have its merits. Some part of the seeming excess of office workers in Japan can be attributed to the constraints on communication from the complexity of the written language.

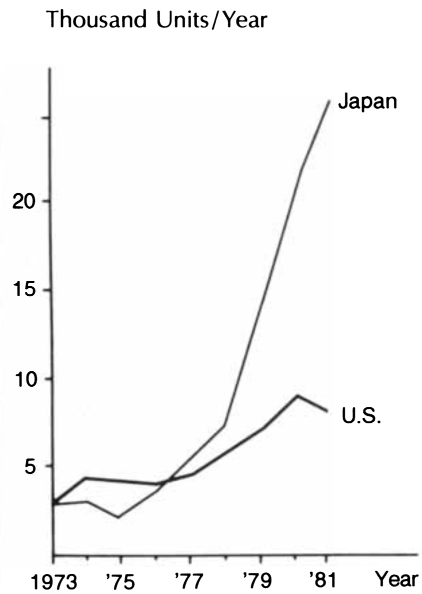
In two product areas related to office automation this disadvantage in written communications was converted to an advantage. With the problems of typing, photocopies and photo-transmittal have a high pay-off. Thus Japanese producers have become the leading suppliers of the lower-cost, lower-speed copying equipment. As Xerox aimed at the higher end of the market, holding high price levels for its products, it left an opening at the lower end of the market that Japanese producers took full advantage of. (The situation has some analogues to small cars, small farm tractors and smaller-screen television sets. As Western producers aimed their designs to attract higher prices for more complex equipment, Japanese producers focused on the lower end of the market—eventually achieving commanding positions.)

Thus Japan now produces annually some $2 billion of photocopier equipment. Moreover, the need to transmit nondigital data such as handwritten memos and the like created a very large The production of facsimile equipment in the world is now entirely dominated by Japanese producers. The value of production is now about $500 million and is expected to double in the next five years, with both production and exports up in 1982 over 1981 by 75 percent. Japanese equipment is now for quality, speed of transmission, reliability and price very nearly unique. Again, however, two elements that might make up part of the office of the future, the automated office, developed in Japan in order to meet the exigencies posed by Japan's written language. It was only with increased memory capacity and sharply lowered prices for memory capacity that it was possible to consider the next element of office automation, the word processor. The first Japanese-language word processor came on the market in 1978-1979. Demand has increased very rapidly. *(See Table 19.)*

Just as in the U.S., the office- and personal-computer market in Japan is in a period of very fast growth, and the word-processor capability is readily available through computer-software packages. The market for these several elements of the automated office is now some $4 billion in Japan and is expected to continue to grow very rapidly indeed. Japan, like the U.S., is fully caught up in the consequences of the microelectronics revolution.

TABLE 18

NUMERICALLY CONTROLLED MACHINE TOOL PRODUCTION



Sources: Nikkei; Japan Machine Tool Builders Association

WHEN PIONEER® LASERDISC™ CHANGED THE NATURE OF STEREO AND TELEVISION, IT BEGAN TO CHANGE THE WORLD.

Less than three years ago, Pioneer introduced LaserDisc™ brand videodisc players. Since then, they have dramatically changed the way people experience music and movies. And they've done even more. They have begun to change both education and business itself.

Very simply, the LaserDisc brand disc is a video record. A record that brings sight and sound together like they have never been experienced before.

A REVOLUTION IN RESOLUTION.

At a time when a multitude of videotape players are crowding the marketplace, the picture on a LaserDisc is astoundingly 40% better than any VCR. And even more remarkable is LaserSound. Simply stated, the stereo LaserDisc delivers is as good as any conventional recording you've ever heard.

And there's a lot to look at and listen to. Today, there are over 500 discs to choose from. Pioneer Artists offers music of every kind. Rock, pop, country and classical.

You can choose from ballet, Broadway theater and opera. And with constant participation from every major motion picture studio, the selection of movies grows bigger every day. For children, there are learning discs and games to play. Yet many people are playing games on LaserDisc brand players outside the home.

THE ARCADE INDUSTRY IS FINDING THE LASERDISC BRAND VIDEODISC PLAYER INCREDIBLY ENTERTAINING.

At the same time LaserDisc was making a great impact on home entertainment, it was changing all sorts of entertainment outside the home. Because of its ability to store and project images, it is being used to create the most exciting video games available today.

*Dragon's Lair,** the first of such, is the only video game done in animation with movie-quality images. This, combined with the sophisticated strategy needed to win the game, is drawing huge crowds into arcades all over the country. And people are paying double the price to play.

And because of the interchangeability of the program disc, arcade owners should soon be able to change games without changing machines.

BIG BUSINESS IS FINDING LASERDISC A BIG ASSET.

Companies like American Bell, Inc., IBM, Apple Computers and Atari are finding LaserDisc brand players extraordinary teaching and selling tools.

American Bell Telephone employees report that training programs which use LaserDisc are more successful than those employing any other teaching medium (including videotape and classroom instruction). Learning centers combining LaserDisc players with computers are providing customers with a unique hands-on learning environment.

The scanning mechanism of LaserDisc allows you to access information anywhere on the disc in seconds. And because one can go in slow motion and even frame-by-frame, forwards and backwards, people can learn at their own speed. In addition, since the disc never wears out, the amount of learning and selling is limitless.

In fact, the next time you go shopping, you might see a LaserDisc unit set up in your department store. Some are being used as electronic catalogues and, when coupled with computers, actually sell products. Others simply help shoppers locate what they are shopping for.

THE FUTURE OF LASERDISC.

The capacity for LaserDisc to prove useful in today's society is endless. Not only have they already changed the way we listen to music and watch movies. But by providing a high caliber of audio and video input never before possible, they are quickly influencing the way we learn.

The Pioneer® LaserDisc™ brand videodisc player many people now have in their living rooms might one day be a required teaching tool in the classroom.

And in the tradition of all new technology, its application promises men and women more free time, higher levels of understanding and an incredibly exciting new way to look at and listen to the world.

PIONEER®

*Available through Cinematronics, Inc.

TABLE 19

PRODUCTION OF OFFICE AUTOMATION EQUIPMENT
(¥ billion)

	1975	1979	1981	1985 (forecast)
Office computer	42.7	146.3	306.7	750.0
Word processor	—	1.2	19.3	109.0
Facsimile	11.8	61.8	107.8	257.9
Copier	112.7	259.6	420.6	582.3
Total	167.2	468.9	854.4	1,699.2

Note: Office computers are small business computers priced at ¥30 million or less.

Source: Long-Term Credit Bank of Japan.

The real constraint on the application of the technologies of automation in Japan, the U.S. and elsewhere is clearly not a technological one. It is, as we have seen, in significant part an investment constraint. If savings and investment rates are high, the speed of application will increase. There is a further financial constraint, however, especially appropriate at the present time, which is that of the cost of capital. This is a subset of the savings and investment issue but is worth noting separately. To the degree that costs of capital are higher in one location than another, then investment in capital equipment will be lower. The high interest rates in the U.S. relative to rates in Japan, means that while this situation prevails, the introduction of automation will be more rapid in Japan than in the U.S.

The other major constraint on introduction of automation is that of labor relations. In the West the introduction of automation is very much slowed by the resistance of the labor force, both in the factory and office, to "labor-saving" methods. The resistance is natural enough, since introduction of automation threatens pay levels and jobs. There has been little opposition in Japan to automation. This results both from the greater job security offered employees by the Japanese company, and from the fact that compensation in Japan, through seniority pay and bonus systems, is closely tied to the success of the company. To the extent that new methods and equipment help secure the future prosperity of the company, to that extent they improve job security and income levels of the work force.

A survey by the Federation of Metal Industry Trade Unions is of interest in this connection. *(See Table 20.)* Trade-union members are reported to be positive toward or indifferent to the introduction of automation, with only a scatter of negative or critical attitudes. This is not to say that Japanese workers are not quite able and willing to be negative or critical if circumstances warrant. However, these introductions are accomplished without layoffs or firings, through retraining and transfer. The technology is not seen as a threat.

Japanese industry is no doubt much aided too in the introduction of new technologies into the workplace by the relatively high level of education of the labor force. There is no immigrant labor in the Japanese factory or office, with the problems that poses for literacy and work discipline. There is no discussion in Japan of what has come to be called functional illiteracy in the U.S.—individuals who have completed a reasonable level of formal education but who lack the ability to do simple sums or to read and write at a functional level. There are complaints about Japanese education, but these are of two kinds for the most part—first, that children are required to study too hard, not a usual complaint in other countries, and second, that Japanese universities are not sufficiently demanding, a problem indeed but not at the factory or office-work level.

Given that Japanese technology is moving forward in both office and factory automation at least as rapidly as is that technology in other countries, and given the high levels of capital investment in Japan, the equipment is becoming available rapidly for what the Japanese like to call the "new industrial revolution." This revolution is unlikely to be impeded by reactionary units of the labor force or from lack of necessary skills in the labor force. The conclusion is inescapable that automation will progress more rapidly in the Japanese economy than in other economies, and that the process will greatly increase the productivity and extend the life cycle of Japan's manufacturing sector, while moving Japan rapidly into the information-related businesses of the future.

The steel industry is both symbol and core of the first industrial revolution. As technology has changed, and as Japan and other competitors have gained cost and technology advantages over Western steel producers, the industry in Europe and the U.S. has come to be a symbol in turn of obsolete facilities, redundant personnel, protectionism and of failure to compete.

In its turn the Japanese steel industry is under pressure now from new competitors from South Korea, Taiwan and Brazil. With this, demand for steel is down. Some U.S. steel companies have responded to market pressures by retreating to diversification through acquisition of non-steel companies. Against this background of Western response to pressure it is of special interest to examine the response of the Japanese industry.

Mr. Hiroshi Kojima, Executive Vice President of Sumitomo Metal Industries, describes the Japanese steel situation:

"The first industrial revolution was led by the era of volume of material of which steel and coal were representative. In contrast, the core of the current revolution is

TABLE 20

REACTIONS OF UNION MEMBERS TO INTRODUCTION OF AUTOMATION
(number of firms reporting)

	Robots	Office Automation
Positive to introduction	51	51
Negative	2	4
Critical	3	0
Indifferent	45	54
Total firms reporting	106	109

Source: National Federation of Metal Industry Trade Unions
From: Economic Survey of Japan 1981/1982.

COME ALIVE, COME AND DRIVE
MAXIMA
MAJOR MOTION
FROM NISSAN
NISSAN

THE PERSONAL PRINTER BUILT LIKE A SHERMAN TANK

PERFORMS LIKE A STRADIVARIUS.

A Tough Act To Follow. Frankly, an Okidata printer is the best printer on the market today. Why? We pack more performance features per dollar into our dot matrix printers than anybody. Bar none. Data processing at speeds from 80 to an incredible 350 characters per second. Exceptional letter quality printing at three times the speed of most daisywheels. Draft and color printing. High resolution all points addressable graphics for charts, graphs, illustrations and photos. Even alternate character sets for self-designed typefaces and symbols. (Print a G-clef, if you like.)

Long-Term Engagement. Our high performers are the most reliable printers available, with rugged steel frames, laser-welded parts, and a print head (the most vital part of all) so strong we guarantee it for up to one full year. No wonder our warranty claim rate leads the industry at less than ½%.

Perfect Harmony With Your Computer. Every Okidata printer works in concert with the major names in personal computers. Better and faster, in fact, than the major computer name printers. (Which, by the way, aren't even made by the major computer companies.)

Best Selection. Right Price. Nobody offers you a better choice. There are eight Okidata printers in all, ranging in price from $449 to $2995, suggested retail. Suggest you call 1-800-OKIDATA (in NJ, 609-235-2600) to find the dealer nearest you. You'll get more for your money from your personal printer. Not to mention your personal computer.

OKIDATA
Mt. Laurel, NJ 08054
A subsidiary of Oki Electric Industry Company Ltd.

function. Thus, if we speak of steel, we must add functional qualities to steel.

"It is not simply a matter of electronics developments. The oil shock had its influence as well. Formerly, it was possible to make money simply through setting up giant facilities, but that era is ended. Still, steel remains the basis of industry. It is inexpensive relative to its strength compared with titanium or aluminum. It remains the basis for construction. However, we cannot expect demand to grow in the future, and we are being pressed by the developing countries. Therefore, for growth, we must aim at upgrading steel. Specifically, because adding to the strength of steel is difficult, we need to make it lighter, to add materials that will strengthen it, and provide surfaces that will not rust.

"We need to develop and are now developing composite material of resins and steel. We purchase the resins, mainly from Sumitomo Chemical, but Sumitomo Metals itself is now hiring chemists, aiming at developing functional materials.

"Looking to the future, we are pursuing two approaches. One is the question of how to change our present manufacturing methods. A stable, high-efficiency system with high quality. We believe that the late 1980's will be a period of reflation for us. To this end we established last year a Facilities Technology Center for the study of new steelmaking methods. Also this past July we established a Design Control Center. This will work to develop sensors and related equipment for the automation of continuous operations. In addition, development work is proceeding on surface control and on composite materials. The shift is from volume to quality, and from volume to functional materials. In this process we must get our costs down.

"The second approach is diversification. One product is titanium alloy, initially for use in aircraft. Another area is that of chemicals from coal as oil-based byproducts become expensive."

Clearly the strategy is not a retreat from steel but an expectation that Japanese producers must focus on developing those steel products that are beyond the technical capability of developing-country producers, while also looking at related technologies and products that will provide further growth.

The theme of a move to higher value added, and of the need to incorporate new technologies into mature products, is a broadly agreed-on concept among Japan's business leaders. The steel industry might be seen as a special case, but even in Japan's still-expanding and highly successful auto industry there is the view that new social and technical developments offer opportunities rather than signaling the end of an industry.

FROM VOLUME TO QUALITY . . . A MACHINE FOR SOCIAL LIVING

Masatoshi Morita, Executive Vice President of Toyota, sees the change as one from volume to quality, from mass production to more specialized, higher value added output. He states:

"In my view the 'new' of the 'new industrial revolution' is a function of the electronics boom. In addition, there has been added the new materials boom, the bio-industries boom, etc. Even with the electronics phenomenon, the real basis is from the properties of materials. It came from that kind of research. From special properties of materials came the 'new-materials' industry, and on the other hand bio-technology rose from the life sciences. This has been a process taking place in the bewilderingly short period of only six years. Thus 'technology industries' arose, seen from one point of view. Again, still differently, looking at it from the auto point of view, there has been a change in the environment for use of the auto, the result of a very complex change in society and in life patterns. User demands, or what might be called user needs, have become very diverse. At the same time domestic competition has intensified and international trade frictions have arisen.

"In all of this the former desire for expansion of volume is gone, and instead there is the problem of seeking higher value added. In the case of autos there is the question of whether this is by materials or by electronic or other means. This is the new industrial revolution, that is to say, technology. The automobile is caught up in all this. In Japan the licensed-driver population is 40 million. Life expectancy is now 74 years for males and 79 for females. Women users are increasing. Drivers are more often older persons. On the other hand, younger persons are seeking in the auto the functions of speed or of play. In these ways too the demands for autos are becoming diverse.

"Growth in volume is largely over. To realize higher value added, this increasing diversity of needs and demands must be met. To do so we must adopt the opportunities of this technical revolution.

"In this context the technical revolution has appeared. Thus in terms of making an auto there is the design issue, as well as production, supply and maintenance. It includes the problems after making the sale. In that sense the concern is with the auto before it is born and after it has died. The technical revolution appears in a number of aspects. The auto, which meets diverse interests and fits diverse objectives, also has an impact on the social economy, on trade, on social structure and as a machine for social living. We must give thought to what is called the new industrial revolution."

If there is general agreement on the broad directions that industry must pursue, toward value rather than volume, and toward full application of new technologies, there is on the part of Japan's top management, a more mixed view of the advantages and potential reach of automation. Nissan Motor's plants are a showplace for visitors to Japan who want to see robots in action. Nissan's original approach to the problem is therefore of special interest, and Executive Vice President Kaichi Kanao of Nissan provides a view of how the process was initiated:

"In the case of Japan our auto industry was late developing and we had somehow to overtake Europe and the U.S. There was a sense of danger that made for joint efforts in the company management, workers and union, and for step-by-step progress. For example, with robots it was not simply a matter of where the best place for them might be. The operators were keeping an eye out every day for where the most difficult work was. Also the Ministry of Labor conducted studies of the degree of work exposure—which tasks were the most exhausting, for example. In this way special work allowances were paid. That is, the work was especially difficult, and so special allowances were attached. Or again, for some jobs it was required that after a certain time at the job a rest period was necessary. It was in these jobs that robots were first introduced.

"From the union point of view this was an improvement of the workplace. Management was making an effort to make the workplace better. However, there was then the question of what happens to the workers involved. These people were trained for better jobs, not simply the heavy jobs of before. These included robot control and robot maintenance. This is a rather positive example of what happened.

"In addition there were men who because of their age could not do these jobs. Since a revolution cannot be made in a single stroke, they continued in their earlier work until retirement. There were also some younger men who could be better placed in quite different kinds of work. When changing jobs, of course, all levels of supervisors are different persons. These men changed jobs, but in the end there was no case of unfair handling. Everyone took notice of all this. These men changed jobs, but in the end were not held back by that fact. Of course, these were not men who played around, and even though young, were good workers. We are a company that endeavors to have a work system and personnel system that does not cause harm to the employees. If you do not have the full efforts of the individuals, and of the company, and of the union—of all three—there cannot be a comfortable company, I believe."

Japan has moved rapidly into robotics, more rapidly by far than any other economy. Nonetheless, the economic constraints on the applications of automation and robots are still considerable.

ROBOTICS IN MASS PRODUCTION . . . RETURN ON INVESTMENT

Senior Managing Director Yoshinori Suzuki of Olympus Optical Co. discusses the issue of automation in his operations:

"We have introduced robotics into our mass production. For example, robotics is well suited to camera production, and there is an adequate return on investment. The problem is, of course, where product variety is great and where volumes are small. Where cameras are concerned, we can invest in robots and automation. However, in plating, for example, in inserting items in the plating tank, the variety of items is very great, almost infinite. Because of the various shapes of the items, the process is not mechanized.

"It is not that it could not be automated, but that the return would not warrant the investment. The process is done by people, inserting items for plating overnight and replacing them during the day. The plating charge and its control are done by computer, however. With volume products, all conveying is done automatically, but as I said, where the volumes are small it is done by hand. Still, as robots' functions are further developed, certainly they will take over the job. It is a matter of investment. The automation could be done now but it is a question of cost.

"An example is the shutter of the single-lens reflex camera. The shutter is flexible and a cord is attached. This cord must be wound on the axle, and the assembly is very complex. The work is extremely difficult, and was done in the past by young girls working most diligently. Now all of the assembly is automatic. I was trained in the old days, and did not believe it could be done by machine. But with sensors and computers, these kinds of operations are entirely possible."

Senior Managing Director of a company that represents both the surge in electronics developments and the most recent advances in factory automation, Mr. Masahide Nanbu of Toshiba notes that the unmanned factory is the concrete representation of the new industrial revolution. He notes, however: "While in general the trend is certainly in the direction of the unmanned factory, a truly unmanned plant is still some way in the future. We are involved both in unmanned-plant developments and in preparing products for sale that have application to unmanned plants. Our Nagoya plant is not unmanned but has gone far in that direction. At one of our television plants, printed-circuit-board production is unmanned.

"We have had no issue of redundancies. Staff has been transferred to other assignments. It has not been a problem. Work shifts from direct to indirect. Even in the era of automation, the need for basic technical skills does not go to zero. The staff who have developed the bases for automation have a function subsequently in building the software aspects. There is need to consider more than work methods. Worry about employment is not necessary even in an era of automation.

"Abroad, concern is voiced about robotics, but it is not warranted. It is people who make the machines work. Their intelligence is needed."

It is not surprising perhaps to find a special preoccupation with very advanced automation on the part of Fujitsu, one of Japan's leading electronics companies. Its executives propound the most elaborate conceptual approach to integrated company-wide computer-controlled operations. Executive Director Shoichi Ninomiya of Fujitsu discusses his company's current activities and plans in automation:

"What is called factory automation in our company is the integrating, in a single divisional location, of production planning, engineering, production engineering and production. Our plan is as follows. Based on office automation, we will bring together design information, production control, laboratory automation and computerized engineering, for total automation of IC design, printed boards and semiconductors. We are thinking of moving to a higher level of Flexible Manufacturing Systems. Further, this will not stop at the divisional location level. We have already completed tying sales information, production control and computers of each factory into a complete circuit. We are aiming at a system that will integrate design information.

"We have introduced robots here and there in the factories. There are two ways of thinking about robots. One thinks of the stand-alone robot, which has caused something of a fever in the world, fanned by robot specialists. The other is robots as equipment in a future automated line, integrating the robot functions. The second is the direction we are taking.

"As products become more diverse, the life cycle shortens. There would be the problem of increasing scrapping of automatic lines, and therefore a basic condition is to make them flexible and capable of producing new products.

MORE WHITE-COLLAR THAN BLUE-COLLAR WORKERS

"In terms of office automation, there is office automation both in the office activities and in engineering. There is a good deal of such equipment in the company. We have over 600 personal computers in use and more than 300 word processors. The numbers are increasing annually. Currently all written material is by word processor.

"Our major planning for office automation is in design. The field might better be called automation of computer aided design. The Kawasaki plant is the main focus. Because engineers are located in the plant itself there are 180 computers there. There is perhaps one terminal for each five persons. These 180 computers at the Kawasaki plant are for hardware design only. Gradually pencils will vanish and only the word processor will be used.

"In our company there are more white-collar than blue-collar workers, so we must plan increased efficiency in the office. The Japanese language was a problem, but with Japanese word processors we can use human power in dealing with computers. The technique of converting *kana* to *kanji* is because we can use the computer. The use by computers of Japanese is becoming customary. The several difficulties of the Japanese language are disappearing."

Even in Japan, however, the enthusiasm for automation is not total. There are industries and executives who look on human skills as irreplaceable. In the view of President Shigetada Fukuoka of Nippon Kogaku, the range of production to which automation technology can be applied is still restricted:

"In our products, as in personal computers, the core is the unseen part, the very large integrated circuit, which carries the opticmechanics, the most important part. The manufacturing technology for the ULSI is not especially difficult. For example, if we produce 10,000 cameras monthly, and the factory hours are 10,000 minutes in a month, a camera is produced each minute. Thus the worker quickly becomes skillful through 10,000 repetitions in a month. However, if we are building


"Near the end of the movie, it looked as if they were both going to escape. Then they nearly got trapped, and zapped. But at the last second they fought their way out. And you know what? Julie and I were right there with them in space, without ever leaving our couch. When we want to go to the movies, we just rent some videocassettes and the movies come to us."
Enjoying the luxury of home screenings is simple with a JVC videocassette recorder. The new HR-D120 and HR-D225 models have controls for all the most popular features—even shuttle search, freeze-frame, and frame advance—designed and arranged so simply, if you can push a button you can go to the movies without ever leaving home.
We take you there.
JVC® JVC JVC JVC JVC JVC JVC
JVC COMPANY OF AMERICA, Consumer Video Division, 41 Slater Drive, Elmwood Park, NJ 07407 JVC CANADA INC., Scarborough, Ont.
WARNING TO PURCHASERS: One Federal Court has held that in-home use of video-tape recorders for off-air recording of copyrighted TV programming is copyright infringement.

equipment for making a camera, the work may take three months, and various technologies must be remembered. However clever the man is who is called for the work, it is necessary to train him in the technology. Of course, if production is increased, and the work becomes a flow, the scope for activity narrows and the amount of technology he needs to remember diminishes.

"This is to say that after the production equipment for the ULSI is produced, the ULSI production is steadily improved, but in building the ULSI machine, automation is difficult. There is much talk around of factory automation, and it would not do to miss the boat. Still, if it is something like steel, or a machine that can be slapped together, perhaps it is all right, but it will not do for precision instruments. Therefore things like robots still cannot be used. There is still the necessity for the skill of the human fingertip."

10 COPYING MACHINES FOR EACH 100 EMPLOYEES

The executive attitude toward factory automation is positive, with some reservations, and this is clearly reflected in the explosion of the market for robotics and flexible manufacturing systems. The situation with regard to office automation is rather different, reflecting in part no doubt the fact that it is only very recently that equipment appropriate to Japanese needs and to the Japanese language has become available. No doubt too many of the reservations about office automation—as well as the strong expression of need for it—would be reflected in conversations of Western executives who similarly feel a need for reduction of office overheads and who similarly do not yet see systems for office automation developed to a satisfactory level.

President Yoshihiro Mita of Mita Industrial Co. directs a leading firm in the copier field and is himself a director of efforts in office automation. Yet he too expresses some frustration over developments:

"Japan is a developed country for copiers. In Japan, for each 100 employees in a company there will be 10 machines in use, a ratio of 10 to one. In Europe and America there might be one machine for 100 employees, and they are not scattered around the offices. In Japan they are dense. This is to say that there is a considerable potential demand in the West. Canon recently has undertaken an ambitious test to develop this base. They are calling it a 'personal copier,' although whether it will be for personal use is another matter. They are already pushing it in Japan for small-scale users. If it succeeds, it will open a big market.

"The entire copy-machine industry is very active. Until recently there was little by way of smaller-size, increased functions. Now there is a sense that if the copy machine does not have these, it is not a copy machine. The market is moving to shortening the life of a machine through replacement of machines that have not been miniaturized or do not have additional functions. One could go so far as to say that a copy machine can be used for 10 or even 15 years. If the attitude can be created that one should be embarrassed to use a two- or three-year-old machine, it could increase the replacement market by several times.

"I do not really understand what is meant by office automation. The use of a copier rather than writing, of a calculator rather than doing one's own calculations, of a word processor rather than a Japanese typewriter—these working tools continue to be upgraded without limit. As a result office help can be reduced, or using the same staffing one can add to the volume of information managed.

"Information maintains the company's competitiveness. If company A manages a great deal of information, then company B does too, and their relationship doesn't change. Because the volume of information increases, people are not freed from managing data, and this causes the introduction of various machines into the office. Demand and production are born. In order to live human beings need food, clothing and shelter of some sort. More than these are non-essentials. How do we produce these non-essentials?

"Because this is the only security for a capitalist society, it is important that we think of these things. Still, in my view, office automation is proceeding extremely slowly. We still have too many people in our company."

The contrast between developments in factory automation and office automation is brought out sharply by Dr. Masaru Yamano, Executive Managing Director of Sanyo Electric:

"We have made considerable changes in our factory and office systems. Because we are a manufacturer, automation in order to get our costs down is our biggest challenge. The robot is one type of automation, as I see it, We have certainly been introducing various types of equipment. Looking at it in terms of types of products, such things as mixers, juicers, toasters and the like are older products with fixed demand levels. These are more or less necessities. Thus with these kinds of products, we can get the advantage of improved quality and lower prices through automation.

"In discussions in our industry the automation of production of goods is considered to be quite advanced. However, concerning the point of numbers of office workers in the cost structure—that is to say, overhead costs—there is a good deal of attention to the issue currently. That is, to the topic of office automation.

THE VOLUME OF PAPER REMAINS LARGE

"There is no question but that the volume of paper is very large. Even with every intention of not having right at hand those papers that are not essential, one soon has a hundred volumes. Thus I believe a paper-filing system becomes important. At the point of introducing a computer for handling these files, the current personal computer and word processor are a considerable help. However, there is still paper. Thus this is a major issue, I believe. However, I also believe it is absolutely impossible to do away with office paper. Since the era of papyrus, all management has been done through the means of paper. In order to manage volumes of important information, it is necessary to go to computer systems. Because we are in an information society, huge amounts of information have to be handled, and I believe this means that we must be able to store, retrieve and edit under any foreseeable circumstances. Still, the notion of a paperless office is doomed to failure."

President Hideo Tashima of Minolta views the office-equipment market, in contrast to the camera market, as an immature one, in which the opportunity for product development is considerable:

"Compared to cameras, I believe there is great potential for further development of office equipment. Office-machinery technology seems not yet matured. Even though various kinds of machines have been offered, there are serious service problems. While the products are maintained, the service conditions are a function of the design of the product. There has been a good deal of progress, but it is still the case that if an amateur handles the product carelessly, there can be a good deal of trouble.

"Certainly too it is the case that the price of many of these products is not high, and after-service provides income. From the supplier's point of view a good product is unreasonable. But that position cannot be defended from the point of view of the user.

"Gradually these products will develop to the point that they can be used immedi-


We've got the economics of copier technology down to a science.
There's never been a copier like the new Mita DC-312RE. A high performance machine with more advanced electronics and microcomputers than other copiers at twice the price. With Mita's full-featured copying system, you get practically every technological advancement available: two reduction modes, enlargement, self-diagnostic maintenance plus a 15 bin sorter and document feed. And, at 30 copies per minute, you'll spend less time copying, too. Nothing beats Mita for reliability and copy quality, consistently rated high by leading consumer reports. And with nearly a half century of experience plus a nationwide service network, Mita remains at the forefront of high technology innovation and cost-efficient performance. See the state-of-tomorrow copiers from Mita, today.
mita
mita DC-312RE
DC-312 RE
Full Service Copiers
mita
MITA COPYSTAR AMERICA, INC.
Heights Plaza, 777 Terrace Avenue
Hasbrouck Heights, NJ 07604
Tel (201) 288-6900

JAPANESE TECHNOLOGY TODAY

ately after purchase without problems. It is in that sense that I speak of office equipment as being still immature. In addition, they provide only black and white.

"Providing color is very difficult. Because this equipment uses ink, it is difficult to change colors. They can be developed gradually by adding first one color and then another, but it is an interesting product area."

FROM IMPORTED TECHNOLOGY TO SELF-SUSTAINED GROWTH

As Japanese companies draw even with the West in technology, the burden of new product development falls directly onto the Japanese companies themselves. The progression in the development of Japanese industry from the exploitation of imported technology to growth self-sustained by the independent generation of new technology is exemplified by the history of NEC, Ltd. Founded at the turn of the century as the joint venture of Japanese financial interests and International Western Electric, a subsidiary of the manufacturing arm of American Telephone and Telegraph Co., NEC had its first growth as supplier of communications equipment to Nippon Denshin Denwa Kosha, the "Bell System" of Japan. The majority shares held by International Western Electric passed to an International Telephone and Telegraph subsidiary under the terms of an antitrust decree. Those shares in time passed to Japanese hands. Under the leadership of Dr. Koji Kobayashi, Chairman, NEC has now emerged as a world enterprise in semiconductors, computers, consumer electronics, robots and in pioneering of the integration of computer and telecommunications technology.

Of his company's latest investment in factory automation, Dr. Kobayashi says: "We have just opened a new plant in Chiba (near Tokyo) that is designed to provide prototype manufacturing methods. As we move to production of a new product, the manufacturing methods will be developed at this plant. It is completely computerized, with a fiber optics network to integrate all functions in the plant's operations. As the production methods are worked out in Chiba, the actual volume production will be established in one of our main manufacturing facilities."

Advances in factory automation undertaken for its own internal "market" has given NEC significant product lines for industrial markets outside. NEC is now pressing parallel advances in office automation. The Tokyo office of Mr. Toshio Ono, the company's General Manager, is directly tied to his office at the new Chiba plant. He is able to open his mail, direct it to the proper person—or throw it out, file it, or answer it—from the computer console in Tokyo tied to and displaying his desk in Chiba. At a single touch to the keyboard of his computer work station, Mr. Ono can summon any category of information a corporate officer can require. He can for example key in to a presentation of the day's news relevant to him and to the company, scan reports prepared by company staff, and make hard copy of items that he might wish to refer to later. Still something of a showplace, the equipment provides a clear indication of where NEC sees office automation and computer/telecommunications integration moving in the future.

Not surprisingly, many of Japan's top business leaders have given a good deal of thought to the question of invention and innovation. Chairman Hideo Sugiura of Honda accepts that the process of development as a follower, drawing on the developments of others, is basically different from the process of innovative development. He sees in the breadth of application of current developments a compensating balance to the fact that true innovations are less frequent today.

QUANTITATIVE CHANGE BECOMES QUALITATIVE CHANGE

"If development is late, it can be broadened very rapidly. Earlier, research work, applied research, development, application and variations could be done in one step. That is no longer possible. For example, there is an enormous quality difference between the 8K and 256K. But from 8 to 16, from 16 to 32, from 64 to 256 was only a difference in quantity. This quantitative change can bring about an enormous qualitative change. Five or six years ago it was said that the age of innovation is over. I have a negative reaction to that. Certainly it is different now from the time that Edison could make a sudden invention. But in various places, various needs and various seeds are located. At some time, on some day, when the fireworks have been placed, a chain reaction will occur and a new thing will appear. Is it a truly qualitatively new thing? Whether it is or not, the point is that from the transistor to the IC was not a qualitative difference. From the IC to the ULSI is a quantitative change that becomes a qualitative change because of the great breadth of application.

"The earlier industrial revolution was a one-way street. Now the contents of development are broad, and from the user's side the applications are broad. From both sides it has become a mutual traffic in all directions. That seems to me to be the great difference between the present and the past."

As Japanese firms gain experience in new-product development and introduction, their sensitivity to the demands of the marketplace in terms of adapting products to users' needs has become acute. Pioneer has had an enormous success especially in the field of audio equipment. It has also had some difficulties with the introduction to the market of the videodisk machine. Dr. Takeo Yamamoto, Managing Director of Pioneer, draws some conclusions from Pioneer's experience:

"There is much discussion of 'new media' as a basis for new industries, especially among engineers who are concerned with hardware. But it is clear that it will be some time before these new industries emerge. In developing new products it is necessary to establish the technology that will provide the base. But on the other hand the question must be asked: What are the real needs to be met? There must be a matching of the product technology with user needs. Again, if needs are identified, there must be software to serve as an intermediary between the hardware and the users' needs.

"However much hardware is developed, the issues are how to use it and what software can be applied to it. Taking a long view, the first requirement technologically is for hard technology. In the case of the videodisk, the first stage was to conceive of the configuration of the player. With that in hand, the need was for disk technology, that is, to go from hardware to software. To actually realize a commercial product, the next step is to pay adequate attention to the software that will actually be used, that is, to move from software to software.

THE MAN-MACHINE INTERFACE

"In this process there must be careful study of the users' needs, and an effort to gradually understand those needs. If there has only been the development of the player—the hardware—and no attention has been paid to software, the product has no meaning. One can have a product, but it is a product that meets no need. The interaction between needs and products is a seesaw game."

Despite the challenges that product innovation presents, the confidence of Japanese executives in their companies' capacity to accomplish major technological innovations is high. Dr. Fusao Mori, former Managing Director of Mitsubishi Elec-


SEIKO
SEIKO
PLAY

JAPANESE TECHNOLOGY TODAY

tric, reviewing his and others' efforts in computer developments, is confident Japan will succeed in its program to develop the "fifth-generation computer."

"The biggest problem when I was in the research area concerned what is called pattern recognition, the technology of the interface with the machine. This is not the problem of output from the machine, but of input—the effort to bring together voice, letters and patterns. That is difficult. I think the technology of entering the computer from humans—the man-machine interface—became important from that time. There was also the concern of not being able to meet the competition from IBM. Our goal was to compete, because however we tried, it was not enough. The man-machine interface was the most important and difficult issue.

"The 10-year pattern recognition project was the first research effort. There is no solution with the present computers, using von Neumann methods. There has been a second- and third-stage effort, but the outlook presents many difficulties.

"We may finally be seeing a path ahead. Concerning the 'fifth-generation computer' one cannot say yet whether it will succeed or not, but I believe the value of the personnel assigned is high enough. There is a plan for a prototype in 10 years. In terms of research methods, if there is an understanding that there must be a breakthrough of the wall presented by the von Neumann methods, there should be a sufficient effort to achieve the breakthrough. There are many approaches to research, and it is important to maintain an optimistic view of what can be done. I believe that sufficient effort is being made to allow the Japanese to make the breakthrough."

As Japan has moved to product innovation, its export successes with those products have created substantial trade tensions. Japan Victor is an outstanding innovator in consumer electronics, and leader in the videotape-recorder field. The company's export success is outstanding. It is therefore of special interest to note that Mr. Toshiya Inoue, Senior Managing Director of Victor, feels that the future lies more with trade in software than with trade in hardware. He is further concerned with the present constraints in Japan on software from the still-limited development in Japan of alternative media such as cable television. He is confident this situation will soon change.

"The world is in an era of technological development, and various new products are appearing. From the time these products appear, thinking internationally, that technology will be transferred. A country will select a suitable product, build factories, hire people and make profit in the country. Whether royalties are paid does not matter. In any event we must keep this way of thinking in mind.

"In contrast, it is entirely all right to export information. That will never be a problem. That is, when a given system is in place, one country will perhaps use the software for entertainment. Again, another country will perhaps convert the information to the development of industry. Because information and software will be converted in this way, we can take the example of a computer. Its software will be appropriate to the using country. Or take the videodisk. Because it is simply a container, it will use the software that suits that country. But the hardware itself will be produced in that country.

"In the case of Japan, if the information pipeline were opened entirely and at once, there would be a big change in the country. Until now software has been tightly controlled. At about the time of the development of the VHD, there was a discussion with staff from an American broadcasting company. Broadcasts can in any event only last for 24 hours. They were concerned about videodisks and satellites. America has free competition. In a free system, good things can last, and thus there was a sense of concern and of a need to prepare for new media.

"In Japan they have been able to sit comfortably protected by controls. However, gradually CATV and the like will be coming. The broadcasting stations, the newspapers—they will face a revolution."

FROM MANPOWER, TO MACHINE POWER, TO ELECTRONICS

The advance of electronics is by no means, however, without its critics in Japan. Executive Vice President and Director Shinji Seki of Mitsubishi Motors sees problems in the continuing diffusion of electronics and automation:

"The earlier industrial revolution's origins, for example Watt's development of the steam engine, came I think with no consciousness of an industrial revolution. That is, the term industrial revolution was added by historians. The leap has been made from manpower to machine power and now to electronics. Perhaps the new industrial revolution had already begun with the invention of the diode. I feel we are already immersed in it.

"I feel there was a strong sense of thankfulness in moving from manpower to machine power, but at present? Is there a sense of thankfulness in going to the brain function of a computer in place of the human brain? I have the feeling that rather a lot of people feel frightened.

"This is a period of surplus or excess information. This is probably not a problem for people who are indifferent to information, but I think it is difficult for sensitive people. The functions of machines are surpassing that of humans. The thinking faculties of humans are, of course, clearly superior, but computers and the like are surpassing humans in reaction time and in time-sharing. This sort of thing is likely to bring about nervous problems for humans. In the first industrial revolution the machine power did not exceed the control of human power and there was an opposite effect on sensibilities.

Mr. Tozo Hikichi, General Manager, Information Systems Administration of Hitachi, Ltd., emphasizes the efforts of Hitachi to create a "computer culture." "Of perhaps 20,000 white-collar workers in our company, about half are able to write computer programs. We provide training and place importance on this not because we expect that all of these people will or should write programs in the course of their work. Rather, we think it is important that all of the staff of Hitachi have a full appreciation of what computers can do."

Japan's fast-growing, small small-computer company, Sord, and its dynamic President, Mr. Takayoshi Shiina, are often cited as evidence that entrepreneurial opportunity remains very much open in Japan today. The company got started modestly a decade ago on limited family resources; Mr. Shiina's mother remains its chief financial officer.

It has long been held that a principal obstacle to the growth of new firms in Japan is the complexity of the Japanese employment system. Employment is for the entire career, with an expectation that the employee will not leave the firm, and will not be laid off or fired during his career. In keen competition of giant firms for qualified personnel, a new small firm like Sord might well find the recruiting of such personnel virtually impossible. Asked about this barrier to the further expansion of his company, Mr. Shiina showed no concern. "I have been able to bring first-class men into Sord. You must understand that we have now some 27 subsidiaries. To attract these men, Sord must offer them the kind of opportunity that I have found. That is, they must be given a chance to run a business rather independently and given position and authority to attract them to join us."

As Japanese consider their position relative to the world there is nearly always considerable attention to those differences that may exist between Japan and the West, as sources of differences in performance. Technology is no exception, and as Japanese executives examine relative performances they find the causes in such cultural differences as the educational system, the basic values of different religious systems and differences in the basis of the social contract.

KANJI CULTURE AND ALPHABET CULTURE

Dr. Motokazu Uchida, General Manager, Research and Development Promotion Center of Hitachi, Ltd., is of the view that differences in the education systems of the U.S. and Japan help to account for differences in software development.

"In these discussions of the relative position of Japanese technology, I believe that America is ahead in software in the sense of assembling conventional equipment and creating new functions. Also in applications software the varieties in America are much greater than in Japan. However, it is said by some that it will not be possible to catch up. That is to say, there are the differences between a *kanji* culture and an alphabet culture.

"There is the question of whether or not culture and related matters are suitable to software development. The educational system is important. There are some who argue that the Japanese educational system is still aimed at hardware. Certainly I feel it is. For example, Americans are very good at presenting and selling themselves. They have a perfect understanding of the knowhow of selling oneself. Japanese have not gotten to that point yet."

Senior Managing Director Shigeru Hayakawa of Matsushita Electric has an interesting view of the different bases for the introduction of office automation in Japan and the U.S.

"I believe that there are endemic reasons for it being difficult to introduce office automation in Japan. There also are reasons why it was absolutely necessary to introduce office automation in the U.S. The reason being that basically people are not trusted. That is to say, in a Christian country there is certainly the doctrine of original sin, of the view that human nature is evil. Thus it is a mistake to trust people. Again, it starts with the concept that people will selfishly refrain from work, that trustful relations are suspect. Thus there is the view that with office automation and factory automation, there will be fewer mistakes and selfishness will be minimized. Therefore it is easy to move to office automation.

"In contrast, in the case of Japan the ethical doctrine is that man's inborn nature is good. Thus the question in the case of Japan is: How much will office automation raise the efficiency of people? As a result, because in currently available office-automation systems the potential for improved efficiency is limited, not so many have been introduced. A large number of individual products have been brought in, but they are not yet linked together. The American system of suddenly installing a full on-line system with a whole building automated is different. We first introduce a word processor, or office computer. As usage of these increases, they gradually spread through the office. If one thinks of office automation as a total system, it seems rather easy to introduce these in America.

"In the case of Japan, rather than as total systems, office automation is entering by individual components. This seems to me a rather different way of entry than has been the case in America. For the U.S., while there is a concern with efficiency, other factors come into it. In Japan it is an issue of efficiency only. Because the use of humans in Japan is quite flexible, perhaps total systems will not spread so widely."

DIFFERENCES IN VALUES AFFECT USE OF TECHNOLOGY

Mr. Takao Nawate, Managing Director of Ricoh, believes part of Japan's less strong position in software stems from the contractual nature of Western societies and a different view in Japan of the need to compensate for services:

"I think there are differences in values between Japan and the West that affect technology advances. It concerns intangible things. In the case of Europeans and Americans, because theirs are contractual societies, naturally contracts are seen as extremely important. Even when there is no formal agreement, service is recognized as having a value, and it is well understood that something of value should not go without remuneration. However, in the case of Japan, this sort of thing is very loose. Even with computers, software may go unpaid for. It is a condition that Europeans and Americans would not be able to believe. In any event Japan is a country where being of assistance or providing service has no tradition of being tipped. I believe this difference in values is most basic."

For all of Japan's success and increasing technological independence, the spur to continued effort is still in good part the competitiveness of the Japanese company and executive, and a keen sense of the need to make further progress still. From autos to electronics the West remains the standard by which technological progress is measured. Both Mr. Kenichi Yamamoto, Senior Managing Director of Toyo Kogyo, and Mr. Susumu Aizawa, Senior Managing Director of Epson, feel their industries still learn from Western developments. Mr. Yamamoto states:

"To meet the sometimes conflicting demands for developments in autos, the basis for development will be the application of new materials and electronics to fundamental technologies. It is necessary to realize fuel savings, safety, comfort, pleasant handling and social responsibility. There are limits to the realization of all this. However, with electronic control and new materials it is possible to do the impossible. The contradictory demands must be met. The usual complaints—the ride is not comfortable, fuel consumption is high, efficiency is not good—must be done away with. This is the direction of development of the auto."

DIFFERENT APPROACH TO EACH COUNTRY

Mr. Aizawa has a world view of his company's market, yet sees the U.S. market as a testing area. He says:

"Seventy percent of our sales are overseas. We may soon encounter some troubles with the problem of exchange rates. Not all of our products can be applied to this rule, but I think it is a fundamental mistake to look on the Japanese market and overseas markets separately. In other words, what is called the domestic market is only one unit of the world market. For example, the effort suitable for the French and German markets is precisely the same as the effort required for the Japanese market. There is no reason for special handling of the Japanese market.

"Still, when one thinks of the future of world trade, with the export rate high, there are problems. In particular, when involved with the software of special-purpose computers, the approach to each country must be different. When the overseas market is in a lead position, there must be thought to how to approach the Japanese market. It is necessary to think of a U.S. approach, a British approach, a Japanese approach. There are also of course the differences between developed, newly developing and less developed markets. My point is that in the case of the developed countries, there is sometimes something of a time lag.

"For example, in the case of computers, America is four or five years ahead of Japan. This is true not just of personal computers; it is the case with all computers. The U.S. market is perhaps four times the size of the Japanese market. In these cases, there must be a different approach to dealing with the market.

"Because one can deal with the Japanese market does not mean that one can handle the U.S. market. Ultimately it is a case of learning in the advanced market. Basically it is a matter of applications software. From that consideration, the response must be different depending on the conditions of a given country."

In their overall view of the position and prospects for Japanese industry, the views of these business leaders is generally in accord with the conclusions recently presented by the Ministry of Finance's Committee for the Study of Economic Structural Change and Policy. The Committee found it necessary to employ several neologisms to deal with its topic, and indeed titled its report, "A Dissertation on Softnomics," to convey its emphasis on the information industry in general and software and services in particular.

THE SOFTNOMICS OF INFORMATIONIZATION

In a partial summary of its conclusions the Committee states: "The new era is a continuation of industrialization and might be called 'the era of informationization.'

"The major direction of culture in the future, at least in the advanced economies, is not 'quantitative expansion' but is rather the 'raising of quality.' The technology that will be key is information technology. With the development of the semiconductor there has been an astonishing advance in only 30 years in the technology of information management. It has permeated all industries and even family living standards and brought about a major revolution. This trend can be called 'informationization.'

"In terms of industry this revolution has two especially striking aspects: 'the high development of basic elements' and 'systemization.'"

Whatever one may think of the aesthetics of the term "softnomics," it does convey a sense of where the Japanese see their technology and economy moving. The shift is from sheer quantity toward quality, from hardware toward software, from manufacturing toward services. It is the direction that an increasingly affluent and productive economy must move if it is to continue to progress.

With all of this affirmative view of the future, some cautionary notes may be in order. It is much in fashion currently in Japan to describe the nation's economy as being at the leading edge of technology. Yet on the whole this is not the case. In such truly frontier technologies as biogenetics, aerospace, nuclear energy and ocean-resource development the Japanese position is marginal. The strength of Japanese technology is now in the high level of development that Japanese industry has achieved in what are on the whole intermediate technologies. Japan's exceptional combination of high savings and investment, a well-educated and well-motivated labor force, an intensely competitive domestic economy and a need to seek economic security have combined to bring Japanese industry to the very front edge of conventional technologies. The issue of Japan's success at the frontiers remains to be resolved.

The publishers, authors and Ted Bates Advertising/New York acknowledge the important contributions by the members of the international business community whose messages appear on these pages:

American Honda Motor Co., Inc.
Canon USA, Inc.
Hitachi, Ltd.
Japan Air Lines Co., Ltd.
Matsushita Electric Industrial Co., Ltd.
Maxell Corporation
Mazda Motors of America, Inc.
Minolta Corporation
Mita Industrial Co., Ltd.
NEC Corporation
Nikon Inc.
Nissan Motors Corporation in U.S.A.
Okidata Corporation
Pioneer Electronic Corp.
Seiko Corporation of America
Sony Corporation of America
Star Micronics, Inc.
Sumitomo Metal Industries, Ltd.
Toshiba America, Inc.
Toyota Motor Sales U.S.A., Inc.
US JVC Corp.

In addition, we acknowledge the significant contributions to this section by the following distinguished business leaders who were interviewed:

Mr. S. Aizawa
Epson Corporation

Mr. S. Fukuoka
Nippon Kogaku

Dr. S. Hayakawa
Matsushita Electric

Mr. T. Hikichi
Hitachi

Mr. T. Inoue
Victor Co. of Japan

Mr. K. Kanao
Nissan Motor Co.

Dr. K. Kobayashi
NEC Corporation

Mr. H. Kojima
Sumitomo Metal Industries

Mr. Y. Mita
Mita Industrial Co.

Dr. F. Mori
Mitsubishi Electric

Mr. M. Morita
Toyota Motor Co.

Mr. M. Nanbu
Toshiba Corporation

Mr. T. Nawate
Ricoh Company

Mr. S. Ninomiya
Fujitsu

Mr. T. Ono
NEC Corporation

Mr. S. Seki
Mitsubishi Motors

Mr. T. Shiina
Sord

Mr. H. Sugiura
Honda Motor Co.

Mr. Y. Suzuki
Olympus Optical Co.

Dr. H. Tashima
Minolta Camera Co.

Dr. M. Uchida
Hitachi

Mr. K. Yamamoto
Toyo Kogyo Co.

Dr. T. Yamamoto
Pioneer Electronics

Dr. M. Yamano
Sanyo Electric

Special acknowledgment to the officers and staff of Japan Society for their contributions to "Japanese Technology Today." For more information about Japan Society write to 333 East 47th Street, New York, N.Y. 10017 or call (212) 832-1155.

EXPERIENCE MAZDA 626.

MAZDA 626 SPORT COUPE. UNUSUAL PERFORMANCE, LUXURY, AND VALUE IN ONE GREAT ROAD CAR.

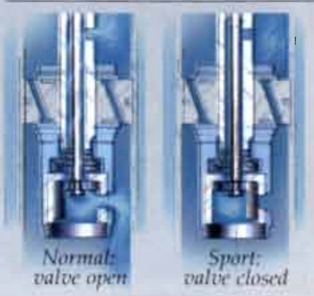
The word got out fast.

Motor Trend magazine named the all-new front-wheel-drive Mazda 626 its 1983 Import Car of the Year. *Car and Driver* said: "The Mazda 626 does everything well, and that makes it the standard of comparison in its class as far as we're concerned."

The public responded by making it one of the most popular road cars ever introduced in America.

And why not? The innovative 626 gives you a lot to like. An advanced 2-litre overhead cam engine that moves you from zero to 50 in 8 seconds flat. A years-ahead suspension system for exceptional handling. An interior spacious enough for five people. A 6-way adjustable driver's seat. A stunning list of standard features.

41 EST. HWY. MPG / **29** EST.* MPG



Electronic Variable Shock Absorbers are controlled by buttons on the dash linked to a solenoid valve atop each shock. NORMAL setting provides a softer ride. In AUTOMATIC mode, the front shocks stiffen above 50 mph for greater stability at cruising speed. SPORT, as the name implies, gives you firmer damping in all four shocks.

In truth, the 626 is one road car that permits you to experience something highly unusual.

Namely, the performance and luxury you look for—at a price you hardly dared hope for.

1984 Mazda 626 Sport Coupe

$8295**

Standard features include 5-speed overdrive transmission (3-speed automatic optional) • Steel-belted radial tires • Rack-and-pinion steering • Power-assisted front disc brakes • Front and rear anti-sway bars • Electric rear window defroster • 60/40 split fold-down rear seatbacks • Full cut-pile carpeting • Quartz digital clock • Tilt steering wheel • Tachometer • Carpeted trunk.

Experienced drivers buckle up.

*EPA estimates for comparison. Your mileage may vary with trip length, speed and weather. Highway mileage will probably be less. **Manufacturer's suggested retail price. Actual price set by dealer. Taxes, license, freight, options (tires/al. wheels shown) and other dealer charges extra. Price may change without notice. Availability of vehicles with specific features may vary.

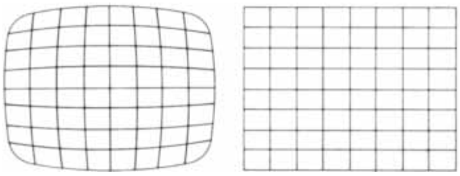
mazda

THE MORE YOU LOOK, THE MORE YOU LIKE.

UNTIL NOW EVERYONE HAD A SLIGHTLY DISTORTED VIEW OF COLOR TV.

If you've ever looked at yourself in a funhouse mirror you know why Toshiba's new picture tube is clearly the most advanced ever created.

Any image, regardless of its original quality, becomes distorted when seen on a curved surface—like your present TV screen.



When it comes to TV picture tubes the world looks better flat.

Toshiba's Flattest Squarest Tube (F.S.T.) is flatter than any other picture tube made. And it's going to change the shape of television for years to come.

A SHARPER, BRIGHTER, BIGGER PICTURE

Unlike other picture tubes, the corners of the F.S.T. BLACKSTRIPE® are squared-off so visibility is increased. In effect, you get a 20″ picture from a 19″ tube. And there's much less distortion—especially at the sides and around the corners. In fact, you can watch it from a 27% greater viewing angle and still be able to see the picture.*

The flatness of the screen also cuts the reflection of ambient light. Conventional convex screens reflect much more daylight and artificial light, so our picture is sharper and brighter. But that's not all Toshiba's advanced technology has accomplished.

As is befitting a television so futuristic in its scope, Toshiba F.S.T. BLACKSTRIPE® TVs are the hearts of entire audio-visual entertainment systems—with capabilities for everything from video movie production and stereo sound to computer applications.

TRUER COLOR THAN ANYONE ELSE.

For years you've listened to television manufacturers waxing poetic about their color. They cite surveys and opinion studies to prove their point. Yet, the truth is, you can't have truly better color unless you create better color phosphors. Toshiba has done just that. Our new phosphors show you reds, greens and blues never before seen on TV. Color that is truer than anybody else's. And that isn't just boast or opinion, but fact. In laboratory tests we reproduce a color picture that's 21% truer than conventional sets and 8% truer than the set acknowledged as the standard of the industry.*

There's also a four-lens focus system instead of the usual two or three. So picture sharpness is greater and there's less convergence error.

From the point of view of accuracy, sharpness, brightness, color clarity, features and design we think the Toshiba F.S.T. BLACKSTRIPE® TVs will revolutionize the way the world looks at TV. But lest you think we're creating a distorted image for these remarkable sets, we encourage you to visit your local retailer and see them side by side with our elliptical competitors.

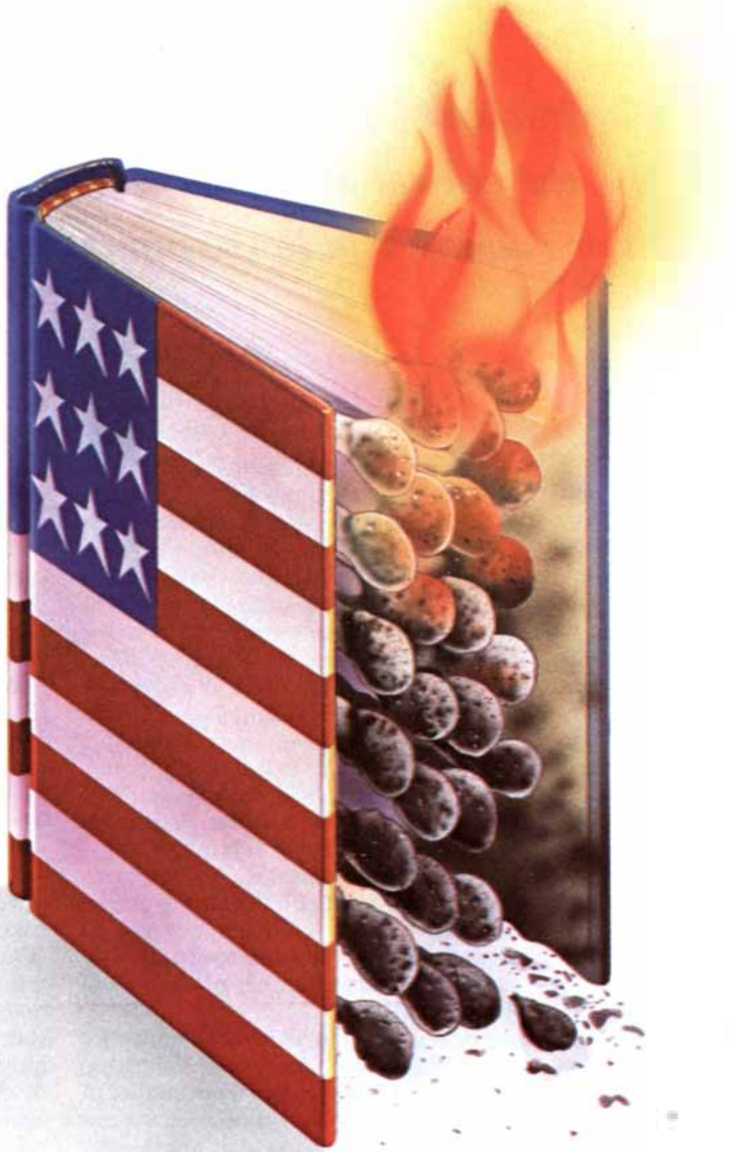
We think you'll agree—to put it flatly, we're the best.

*Toshiba test data.

In Touch with Tomorrow

TOSHIBA

For further information write: Toshiba America, Inc., 82 Totowa Rd., Wayne, New Jersey 07470 or Toshiba Hawaii, Inc., 327 Kamakee St., Honolulu, Hawaii 96814



A Molecular Basis of Cancer

Human cancers are initiated by oncogenes, altered versions of normal genes. In one case the critical alteration is a single point mutation that changes just one amino acid in the protein encoded by the gene

by Robert A. Weinberg

A tumor isolated from a patient is a large aggregation of cancer cells, all of them descended from a single "founder" cell. The single ancestor was once a normal cell, with a normal function in a particular tissue. Somehow it underwent a fundamental change. As a result of that change it began to divide and proliferate in response to some imperative of its own rather than in response to the external stimuli ordinarily required for cellular growth. Eventually this cell spawned the billions of similarly altered cells constituting the tumor mass.

The crucial event in the development of a cancer must be traced back to alterations of the founder cell. What happens to the cell to free it from the normal constraints on cellular growth? In the past few years some answers have begun to emerge. Cancer genes have been discovered in the chromosomes of tumor cells. These genes, often called oncogenes, represent the driving force behind the uncontrolled growth of many cancer cells. It is these genes that become activated in the conversion of the normal founder cell into a cancer cell. Once activated, the genes function continuously to direct cells toward the abnormal behavior that is called the cancerous state.

Pleiotropy

The abnormal behavior of cancer cells is characterized by many distinctive traits. Uncontrolled growth is the most obvious of them. Cancer cells often exhibit a shape that is very different from that of their normal counterparts. They fail to respect the territorial rules that confine normal cells to particular tissues. Many of them import sugar molecules at an unusually high rate. They rely to an unusual extent on anaerobic metabolism: energy-converting processes that do not depend on oxygen. The outer membrane of cancer cells is different from that of a normal cell and displays special tumor antigens giving the cells distinctive immunological properties.

That is only the beginning of a long list. The length of the list and the implied complexity of the cancer phenotype (the total set of structural and functional characteristics that define a cancer cell) raise important questions. Which of these traits are essential to the cancerous state and which are peripheral? If a cancer cell has 100 distinctive traits, is each of them acquired individually as the consequence of a discrete step in carcinogenesis? Does the cell pass through 100 stages in the course of its evolution from the normal state, with each stage signaling a different change? Or is a much simpler, "pleiotropic" mechanism at work? Perhaps a single, centrally acting cellular element is turned on and is then able to elicit simultaneously a large number of changes in the phenotype.

Fortunately the evidence points toward simplicity. An early indication came about 20 years ago from studies of several small DNA viruses that induce cancer in some animals. Marguerite Vogt and Renato Dulbecco, who were then working at the California Institute of Technology, were able for the first time to transform normal cells into tumor cells within the confines of a laboratory tissue-culture dish. They applied polyoma virus to hamster-embryo fibroblasts (connective-tissue cells) growing in Petri dishes and observed that the infected cultures yielded foci, or colonies, of altered cells that piled up on one another instead of forming flat layers only one cell deep as normal fibroblasts do. When the altered cells were inoculated into young rats, they proliferated to form tumors. Tumor cells, in other words, could be created by manipulating normal cells in a Petri dish; the normal cells could be induced to undergo "transformation" into tumor cells. No longer was the induction of cancer a mysterious process taking place only within an animal's inaccessible tissues.

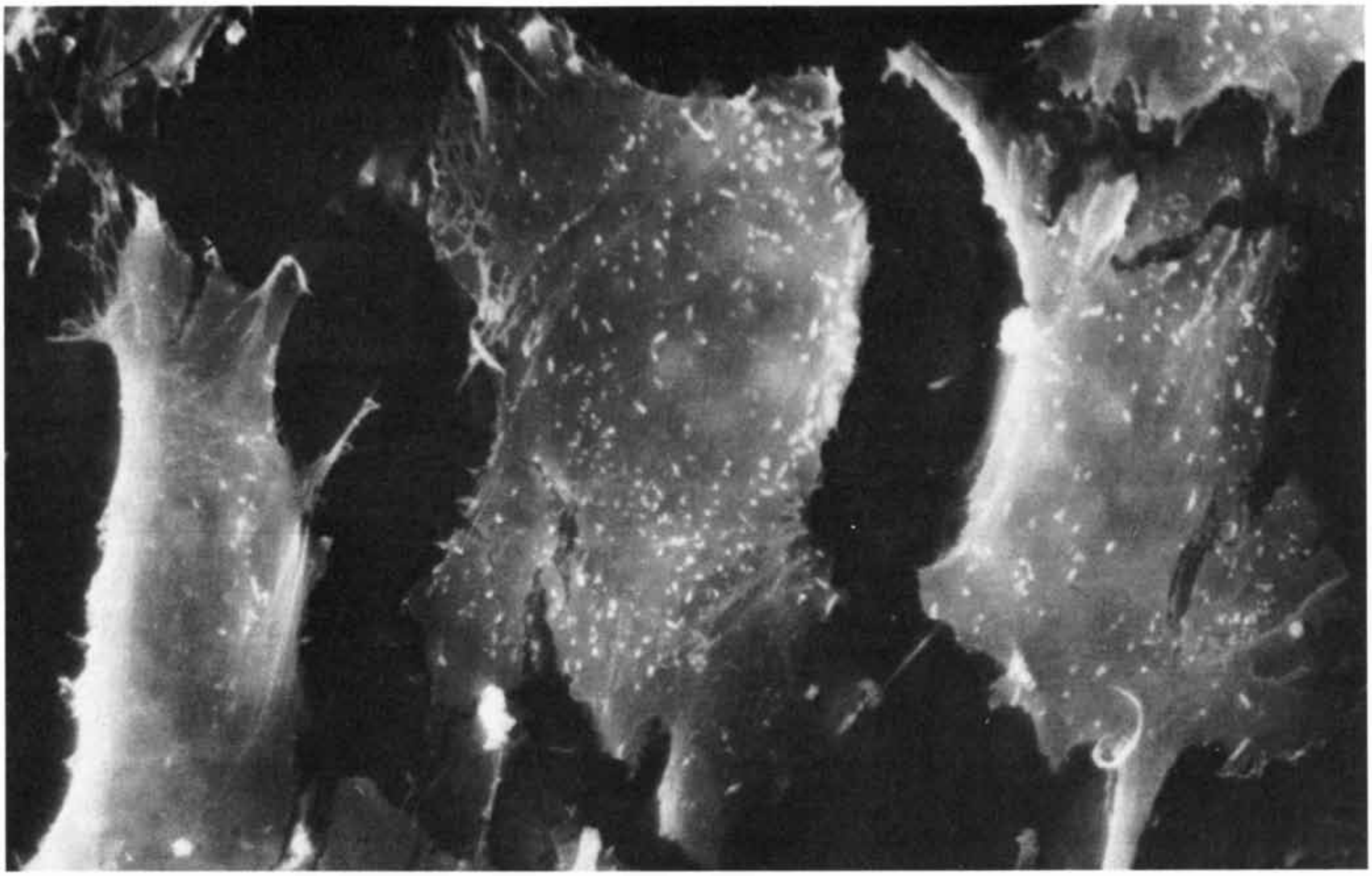
The virally transformed cells exhibited many of the altered traits associated with the cancer phenotype. Dulbecco and others showed that the transformation was somehow induced by viral genes, the segments of genetic information the virus brought into the cell in the process of infection. This information is carried by DNA molecules within the virus particles. An infecting tumor virus brings into an animal cell only a very small amount of DNA, perhaps a millionth of that which is present in the cell's own chromosomes. A very small complement of viral genes had therefore been able to elicit dozens of changes in the cells' structure and behavior. The role of pleiotropy in transformation was clearly established.

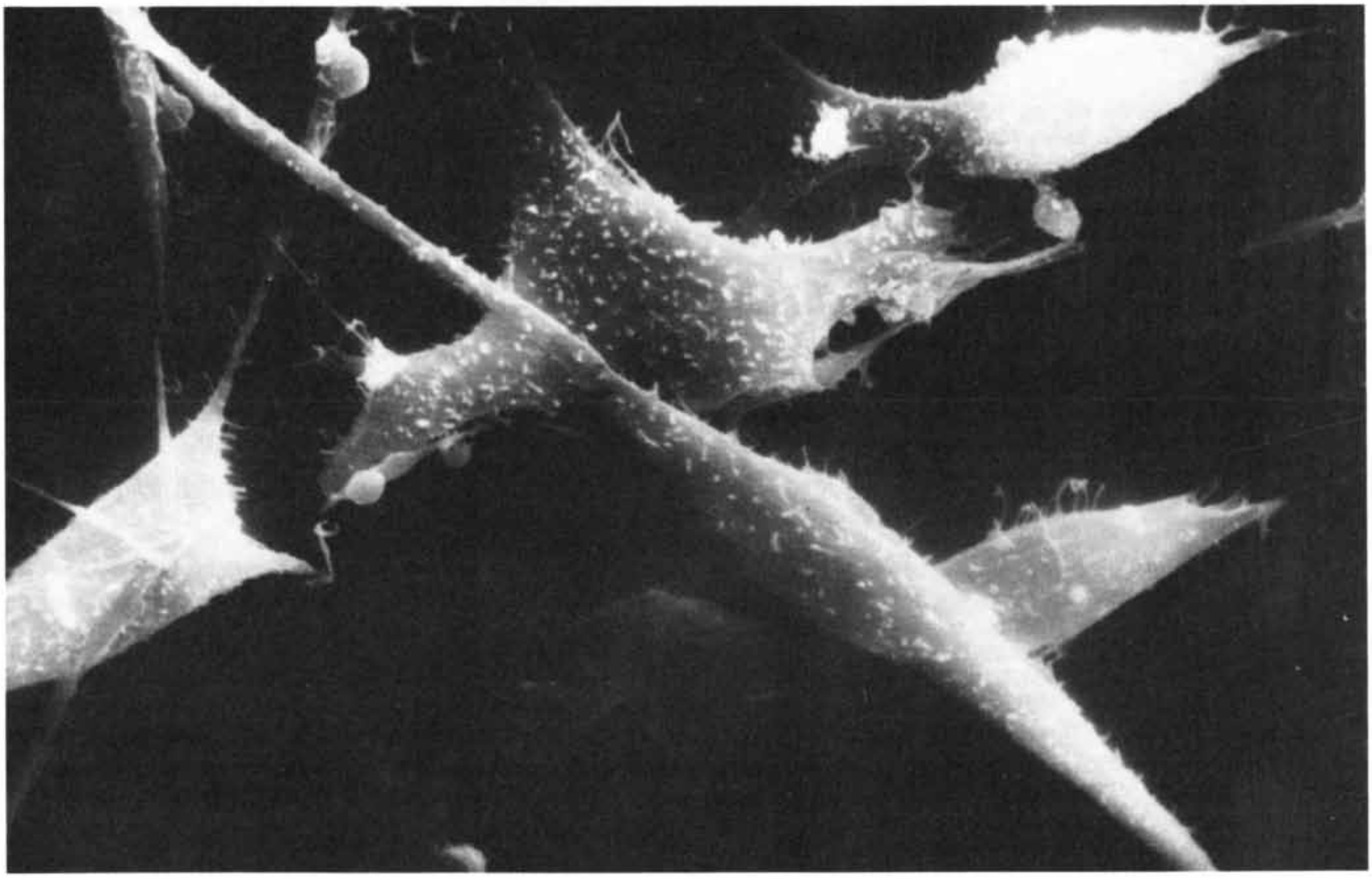
This early work led eventually to another finding bearing on the molecular basis of the transformation. If the viral genes were lost from the proliferating tumor cells or were inactivated experimentally, the cells reverted to a normal state. Not only were the viral genes required to initiate the process of transformation; their continued presence and activity were necessary to maintain the tumor phenotype.

A Genetic Basis

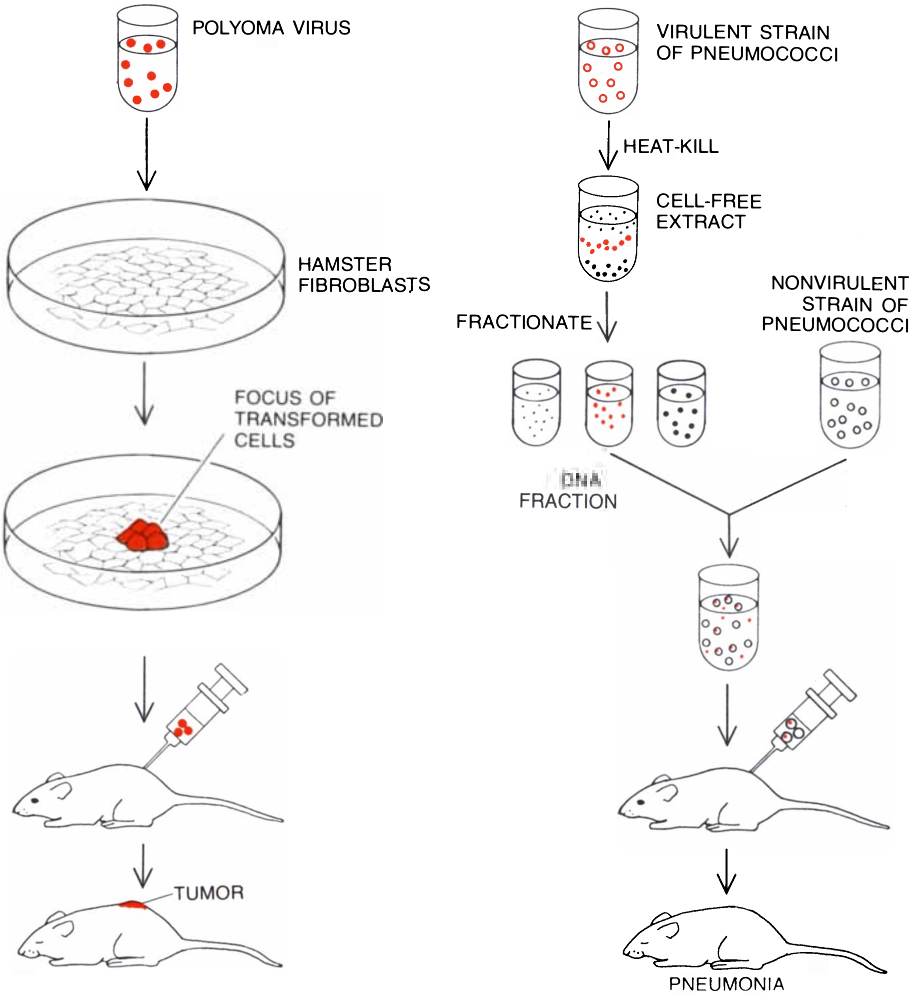
At that stage in the research one might have speculated, in a leap of logic, that analogous mechanisms operate in other cancer cells, transformed by agents other than viruses. Might not all tumor cells carry a small number of genes whose continuing activity is required to orchestrate the complex behavior of a tumor cell? Such thinking implies that the basis of cancer is genetic: that the cancerous state is maintained by genes, and not by other, "epigenetic" regulatory systems affecting the cell.

Twenty years ago this idea was far from being generally accepted, even though there had long been other indications that genes are involved in cancer. Investigators had been accumulating data on nonviral carcinogens, including various forms of radiation and a wide range of chemical agents that can induce cellular transformation. Instead of bringing novel genetic information into the cell, such carcinogens act to alter





TRANSFORMATION OF NORMAL CELLS by an oncogene is demonstrated in scanning electron micrographs made by Erika A. Hartwieg and Jonathan A. King of the Massachusetts Institute of Technology. Normal mouse fibroblasts (*top*) were treated with DNA extracted from the human tumor-cell line designated *EJ,* derived from a bladder carcinoma. The tumor DNA "transfected" mouse cells, giving rise to colonies of transformed cells (*bottom*). Whereas the normal cells lie flat and keep their distance from one another, the transformed cells round up, proliferate and climb onto and over one another. The cells were fixed with glutaraldehyde and osmium, dehydrated, "critical-point" dried and thinly coated with gold-palladium. Cells in both micrographs are enlarged about 2,000 diameters.



TRANSFORMATION OF CELLS BY VIRAL DNA was demonstrated (*left*) by Marguerite Vogt and Renato Dulbecco. They treated hamster fibroblasts with the polyoma virus. Only the small viral genome (the complete set of the organism's genes) enters an animal cell, but it was enough to give rise to foci of transformed cells with many altered traits. Injection of transformed cells into rats gave rise to tumors. A direct demonstration of the role of cellular DNA in transformation was still required. An approach to such a demonstration was suggested by the historic experiment of Oswald T. Avery and his colleagues that in 1944 defined DNA as the genetic material (*right*). A cell-free extract of virulent pneumococci was fractionated, and a fraction consisting of DNA alone was shown to render a previously nonvirulent strain virulent.

existing genetic information: the genes of the target cell. Many carcinogens act by damaging DNA and thus causing gene mutations. Once again a central role for genes in carcinogenesis was suggested.

"Suggested" is all one can say. Direct evidence was lacking. An experiment was required to demonstrate directly the presence in the cancer cell of altered genes responsible for orchestrating the cell's aberrant behavior. Such a demonstration required analysis of the cell's very complex DNA, not the relatively simple DNA of a tumor virus. One way to do this was suggested by the historic experiment that 39 years ago first demonstrated the central role of DNA as a carrier of genetic information. Oswald T. Avery, Colin M. MacLeod and Maclyn McCarty of the Rockefeller Institute for Medical Research studied two strains of pneumococcus bacteria, only one of which caused fatal pneumonia in mice. What was the molecular basis of the difference between the two strains? The investigators analyzed an extract of the virulent bacteria and isolated a purified "transforming principle" that was able to convert the nonvirulent bacteria into actively pathogenic organisms. They established that the transforming principle was DNA. The transfer of DNA molecules could transmit the genetic information for virulence from one bacterial cell to another. DNA was shown to be the carrier of genetic information specifying a discrete trait.

Gene Transfer

Five years ago Chiaho Shih undertook an analogous gene-transfer experiment in my laboratory at the Massachusetts Institute of Technology. He asked: Could the cancer trait be transmitted from one mammalian cell to another by the transfer of DNA molecules? A way to transfer genes between mammalian cells had been developed a few years before by workers in the Netherlands. Following their procedure, one traps DNA molecules in crystals of calcium phosphate. The crystals are then applied to cultured cells and act to facilitate the DNA's entry into the cells. Once inside the recipient cells, the transferred DNA becomes integrated into the chromosomal DNA.

Shih extracted DNA from tumor cells, in this case mouse fibroblasts that had been transformed into a tumor-cell line by exposure to the chemical carcinogen methylcholanthrene. The donor-cell DNA was coprecipitated with calcium phosphate and applied to cultures of untransformed mouse fibroblasts designated *NIH*3*T*3. Two weeks after the donor DNA was applied, foci of transformed cells appeared in the *NIH*3*T*3 cultures. Under the microscope the transformed colonies looked very much like those induced by tumor-virus infection. When the transformed cells were inoculated into mice, tumors grew.

The lesson was clear. The information for being a tumor cell had been transferred from one cell to another by DNA molecules. The expression of at least some part of the cancer phenotype could be traced directly to information carried in the DNA of mammalian cells that had been transformed by a chemical carcinogen. When the DNA of normal cells was similarly transferred, it had no such effect. This showed that DNA sequences in the donor tumor cells differed from analogous sequences present in the DNA of normal cells. We suspected that these differences were due to mutation of the donor tumor-cell DNA by methylcholanthrene.

The generality of those first observations could be questioned because the donor DNA had been taken from chemically transformed mouse fibroblasts and introduced into untransformed cells of the same type. It was possible that the DNA-mediated transformation we observed was characteristic only of donor and recipient cells of this special type, and not of other types of cells. That turned out not to be the case. Workers in our laboratory and in several others repeated the initial experiment with DNA's taken from a wide variety of tumor cells and found the same transforming activity. Among the human tumors that served as DNA donors were carcinomas of the bladder, colon and lung as well as fibrosarcomas, neuroblastomas and even leukemias.

It appeared, then, that some common feature must link the chemically transformed mouse fibroblasts with the cells of all these spontaneous human tumors, since the DNA of all of them has the same effect when it is transferred. Moreover, if the DNA of a human colon carcinoma could transform mouse fibroblasts, it was clear that the transforming sequences of the tumor DNA can function in cells that differ both in their species of origin and in their tissue of origin. Finally, the transforming DNA introduced into cells by gene transfer seemed to act pleiotropically, because the cells transformed by the introduced DNA now exhibited a number of traits characteristic of tumor cells. A limited amount of donor genetic information was shown to elicit several different behavioral responses in recipient cells, just as in the case of the polyoma-virus DNA.

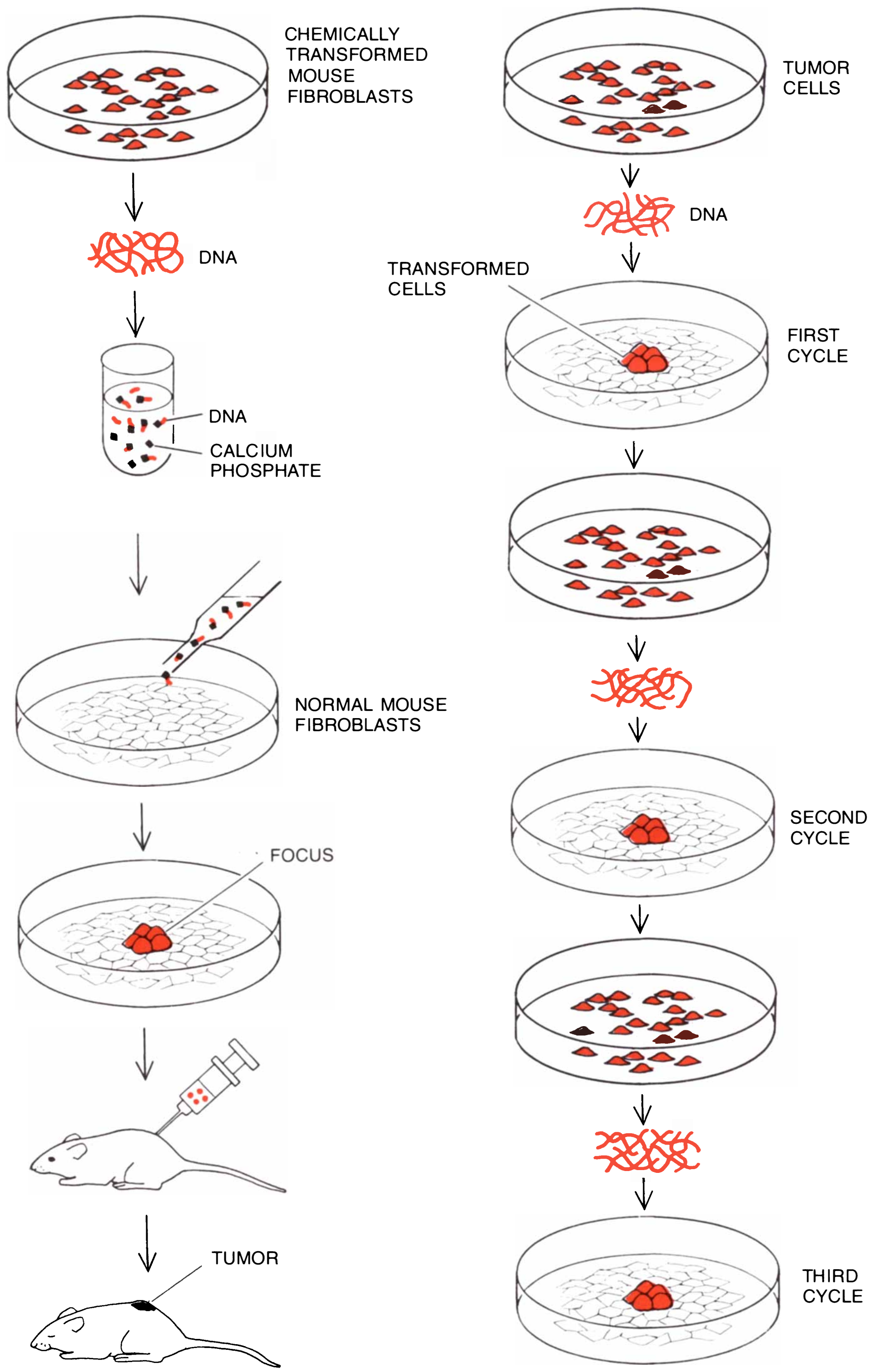
Defining the Transforming Principle

The gene-transfer experiments were informative, but further advance depended on pinning down and characterizing the specific sequence or sequences responsible for transformation. So far not even the most basic property of the tumor-DNA "transforming principle" had been established. Was it assignable to one segment of DNA or to several independent genetic elements whose cooperation was required to transform a cell? If there was indeed a single active segment, that would suggest the presence of a unique gene that functions like other discrete, definable genes in the cell's genome: its complete set of genes.

One experiment had already pointed toward localization of the transforming principle within a single DNA segment. DNA extracted from donor tumor cells and applied to mouse fibroblasts, as described above, induced several foci of transformed fibroblasts. Cells from one of the transformed foci were isolated and grown into a large culture. When DNA was extracted from this culture and applied once again to normal cells, foci of transformed cells appeared a second time. The process could be repeated through a third and even a fourth cycle of gene transfer. This suggested that the transforming principle, whatever its nature, resided in a single DNA segment whose integrity had survived the serial transfers. The successive extractions and related manipulations had the effect of breaking the long strand of cellular DNA into thousands of small segments, only a few of which could be expected to enter any single mouse cell. If the transforming activity depended on the cooperation of multiple, unlinked elements, it would surely have been lost in the course of the serial transfers since the cooperating partners would have been dissociated from one another.

The problem, then, was to find the transforming segment and study it. The design of the experiments was dictated by the details of DNA structure. The DNA molecule is a double helix composed of two strands of paired nucleotides, each characterized by one of four chemical groups called bases: adenine (*A*), guanine (*G*), thymine (*T*) and cytosine (*C*). It is the sequence of the four DNA nucleotides that carries the genetic message. The expression of this message occurs when a single DNA strand is transcribed into the nucleic acid RNA and the RNA is translated into a chain of amino acids constituting a protein. We suspected that these differences were due to mutation of the donor tumor-cell DNA by the carcinogen methylcholanthrene.

The cellular genome is composed of some six billion base pairs of DNA. An average gene is constituted from 5,000 or 10,000 base pairs of DNA. The chal-

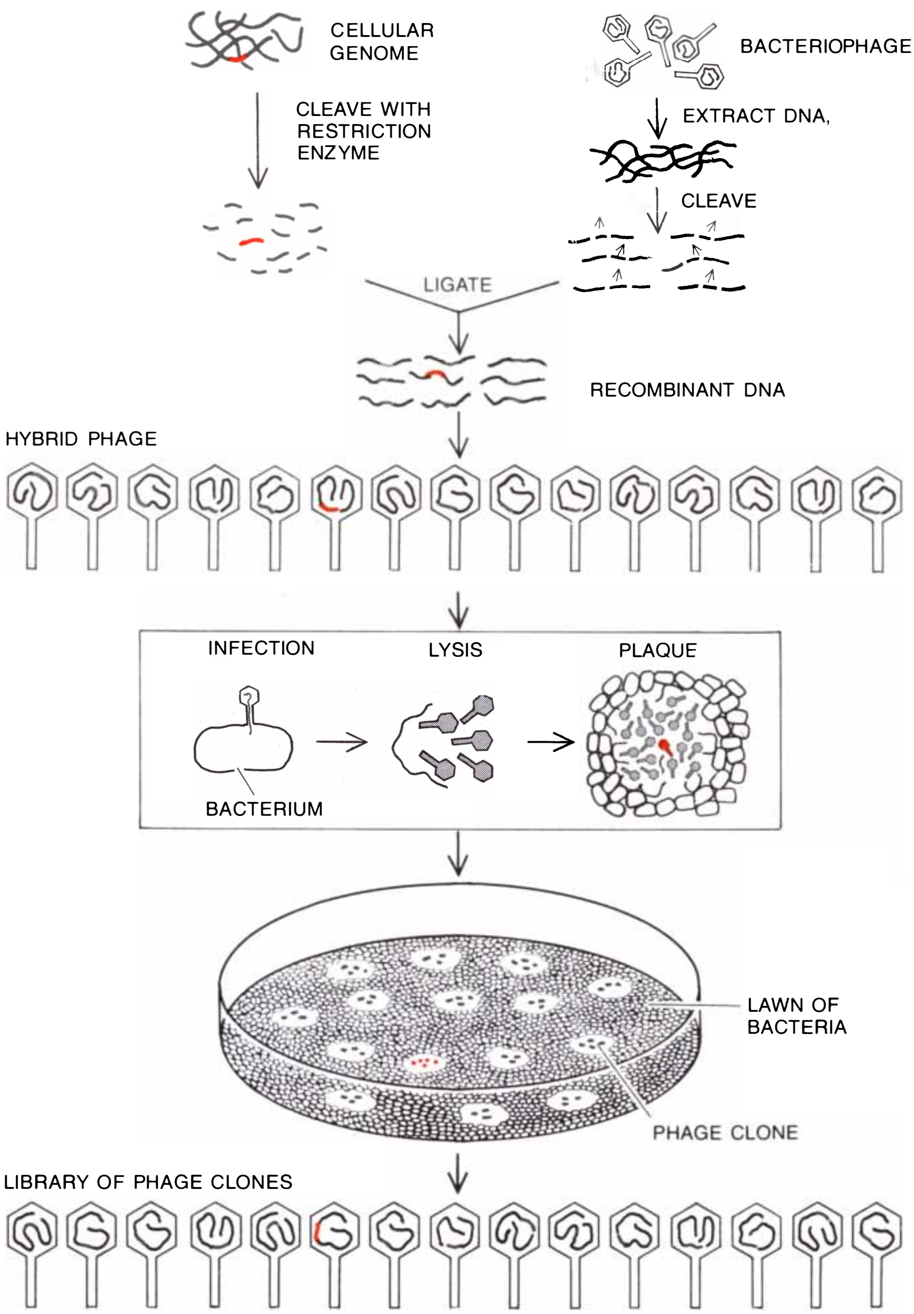


TUMOR-CELL DNA was shown to encode cancer traits by a gene-transfer experiment (*left*). DNA was extracted from mouse cells that had been transformed by a chemical carcinogen. The DNA was coprecipitated with calcium phosphate, which facilitated its entry into normal mouse cells. Transfected normal cells gave rise to foci of transformed cells. Injection of the transformed cells into mice gave rise to a tumor. In a serial-transfer experiment (*right*) DNA extracted from human tumor cells transformed normal mouse cells. The transformed cells were grown into a large culture whose DNA was extracted in turn and again served to transform mouse cells. The process was repeated in a third cycle and even a fourth. Only short segments of DNA could survive the process of repeated extraction and coprecipitation, and so the successive transformations showed that the transforming agent must reside in a single segment.

lenge was to dissect a single genetic segment, carrying the transforming principle, away from a millionfold excess of unrelated genetic material in the cell. The newly developed techniques of gene cloning made it possible. They enable one to retrieve a single genetic segment from the cellular genome and amplify it to make many thousands of identical copies. With this material in hand one could study a pure gene, uncontaminated by the enormous jungle of complexity that is its normal genetic environment.

The various strategies of gene cloning, some involving bacterial plasmids (small circles of extrachromosomal DNA) and some involving bacteriophages (viruses that infect bacteria), have a common logic. The cellular genome is broken into several hundred thousand segments. Each of these segments is inserted into the genome of a vector, often the small DNA genome of bacteriophage lambda, which infects the bacterium *Escherichia coli*. Each resulting hybrid phage carries a single inserted DNA segment along with some of its own genes. The collection of thousands of hybrid phages is often called a genomic library because in aggregate it carries the entire store of genetic information in the cellular genome: every part of the original genome should be present in one or more of the phages constituting the library.

Each of the phages can be amplified independently by growth on *E. coli*. When the phages are applied to a "lawn" of *E. coli* in a culture dish, each phage infects a separate bacterial cell and proceeds to make many copies of itself. The phage progeny kill the cell and go on to infect adjacent cells, yielding thousands of descendant phages within a few hours, all of them localized within a single plaque, or hole, in the lawn of bacteria. The population of phages in a plaque is clonal: it is descended from a single ancestor. From a clone of phages one can isolate thousands of copies of the inserted DNA, and in this way one obtains a clone of the DNA segment.



GENOMIC LIBRARY can be established by cloning genes in the virus known as bacteriophage lambda. The cellular and the phage genome are cleaved with a restriction enzyme. A segment of the phage DNA is discarded and the remainder is joined to a cellular-DNA fragment with the enzyme DNA ligase. The recombinant DNA is packaged in the protein coat of phage lambda. When a hybrid phage infects the bacterium *Escherichia coli*, the phage multiplies, lysing (bursting) the cell; phage progeny go out to infect and kill nearby cells in the bacterial culture, forming a plaque, or hole. When a "lawn" of bacteria in a culture dish is infected by a large number of different hybrid phages, each plaque in the lawn is inhabited by a single clone of phages descended from the original infecting phage. Each clone carries a different fragment of cellular DNA. The problem now is to identify the clone carrying a particular gene of interest (*color*). It is ordinarily done by probing the clones with DNA or RNA known to be related to the desired gene, but no such probe was available to aid in the search for transforming DNA.

Scanning a Phage Library

The central problem in such a cloning experiment is to identify those few phage clones in a library that carry the inserted DNA of interest. The most convenient way to do this is with a probe, a single-strand DNA segment whose nucleotide sequence is closely related to that of the gene to be cloned. Because the two strands of a DNA molecule are complementary (*A* pairs with *T* and *G* pairs with *C*), the probe "hybridizes" specifically with single strands of the sought-after DNA. By applying a probe labeled with a radioactive isotope to the phage plaques, one can trace and precisely localize the plaque carrying the DNA of interest. (Detection is made possible by the procedure termed autoradiography.)

The transforming sequences, however, could not be identified by this standard procedure. No sequence-specific probe was available. The DNA we sought was defined and recognizable only in terms of its biological activity: the ability to induce transformation. Novel techniques were therefore required.

Three laboratories undertook the cloning task, each with a different strategy. One group, led by Geoffrey M. Cooper of the Dana-Farber Cancer Institute, associated with the Harvard Medical School, simply did without a

The all-new Dodge Daytona Series proves it doesn't take a big engine to make a hot car. With a simple 4-cylinder engine—turbocharged and fuel-injected—Daytona Turbo Z can do 0-50 in 5.76 seconds.** Yet it can also earn mileage numbers of 35 est. hwy., [22] EPA est. mpg.***

The basic Daytona gives you the 2.2 block with electronic fuel-injection. A quick, close-ratio 5-speed. Tight, power-assisted, rack-and-pinion steering. Power brakes. Steel-belted radials. Front-wheel drive. Aerodynamic styling that helps achieve 43 est. hwy., [24] EPA est. mpg.*** Go for the Turbo or Turbo Z and you'll get even more power and handling features that really let you play with the big boys.

Daytonas are backed by our standard 5/50 Protection Plan.* And all will be available to buy or lease* at your Dodge dealer this fall.

THE AMERICAN SPORTS CAR HAS JUST GROWN UP.

WE ARE DODGE. Dodge **AN AMERICAN REVOLUTION.**

BUCKLE UP FOR SAFETY.

*5 years/50,000 miles, whichever comes first. Limited warranty on outer body rust-through, engine and powertrain. Excludes leases. Deductible may be required. SEE DEALER FOR DETAILS.
**Based on acceleration tests performed by NHRA on Daytona Turbo Z with standard engine and transmission. ***Use EPA est. mpg to compare. Your mileage may vary depending on speed, weather and distance. Actual hwy. mpg lower. CA estimates lower.

The average executive spends a lot of busy hours sitting behind a desk.

While the average desk sits there doing nothing.

But equip any desk with an IBM computer or terminal—like our new Personal Computer XT—and suddenly a dull piece of furniture becomes a very smart piece of business: The Smart Desk from IBM.

The Smart Desk can give you all sorts of timely information, any time you need it.

It can produce graphs and spreadsheets. Help you compile reports. Keep your calendar, even send and receive electronic mail.

And because no one likes to wait for important facts and figures, our desktop computers are designed to connect you directly with your company's central data files and office systems. Or with the Smart Desks of managers, secretaries and other professionals around the office.

IBM Smart Desk products are easy to use, so you'll start reaping the benefits almost as soon as you sit down. And when you put yourself behind The Smart Desk, we put our famous service network behind you.

To learn more about IBM desktop computing, call your IBM representative or visit your nearest IBM Product Center.

And we'll show you a smart place to spend your career.

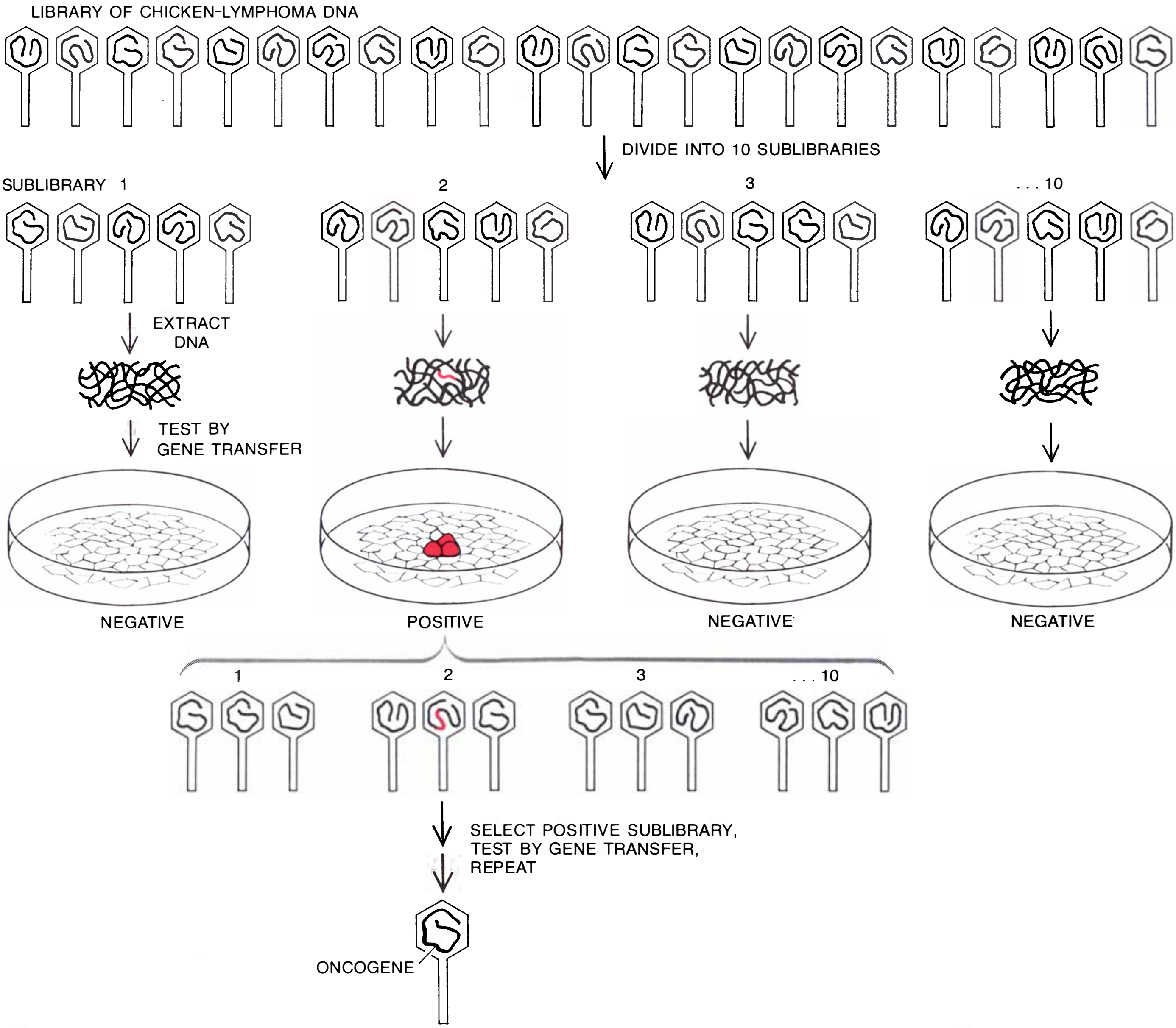
IBM®

The Smart Desk from IBM.

probe. A genomic library was constructed from the DNA of chicken lymphoma cells, which were known from gene-transfer experiments to carry a transforming sequence. The resulting collection of several hundred thousand hybrid phages was divided into 10 sublibraries. DNA was prepared from the phages of each sublibrary and tested by gene transfer for its ability to transform mouse fibroblasts. Any positive sublibrary was further divided into 10 sub-sublibraries, each of which was screened by gene transfer. This exponentially narrowing search routine eventually led to a single phage clone carrying potent transforming activity, and thus to the transforming DNA.

An alternative approach was taken by Shih in my laboratory. He began with the DNA of a human bladder carcinoma. He transformed mouse cells with the human tumor DNA and then used the transformed mouse cells' DNA to transform a second cycle of mouse cells. The library was constructed from the DNA of these secondarily transformed cells. The DNA of the mouse cells now contained only a very small amount of human DNA: that directly associated with the transforming activity. The remaining human DNA segments had been discarded in the course of the serial gene transfer, which selected only for DNA with transforming activity. Shih pursued this small amount of DNA with a probe that would recognize only DNA of human origin: a probe for what are called *Alu* sequences. These sequences are present only in human DNA, where they are scattered randomly through the entire genome. By screening his library with the *Alu*-specific probe, Shih identified and isolated a phage clone carrying human DNA. On testing by gene transfer this human DNA was found, as it was hoped, to contain the sought-after transforming activity.

Michael Wigler's group at the Cold Spring Harbor Laboratory began by linking a bacterial marker gene to each segment of the human bladder-carcinoma DNA. By serial gene transfer they prepared a mouse DNA carrying only the human transforming DNA and a closely linked copy of the marker gene. They constructed their library by inserting the mouse DNA into special phages having defects in the sequences encoding several phage proteins. Only phages



SEARCH ROUTINE was devised by Geoffrey M. Cooper's group to isolate the transforming DNA. They constructed a library of some 200,000 phage clones from the DNA of chicken lymphoma cells. They divided the library into 10 sublibraries and tested the DNA of each sublibrary by gene transfer. If the transfer yielded transformed cells, the sublibrary involved, which necessarily included the DNA of interest, was divided in turn into 10 sub-sublibraries, each of which was tested. Positive subgroups were tested successively until a single phage clone was found that carried transforming activity. From it the chicken-lymphoma transforming gene, or oncogene, was isolated.


A full carat or more. Halfway isn't your style.
Your jeweler can show you many diamond solitaires of a carat or more. This necklace features a 2.65 carat diamond.
A carat or more–one in a million.
Every diamond is rare.
But of all diamonds found, a solitaire of
a carat or more is only one in a million.
And, like love, becomes more precious
with time.
A miracle among miracles.
Born from the earth. Reborn on a woman.
The extraordinary diamond
of a carat or more.
Show the world you couldn't have made
it without her.
A diamond is forever. De Beers

Human Factors Engineering. Interiors designed scientifically to minimize the possible distractions from your driving. A seat has to look great, but it has to fit great too.

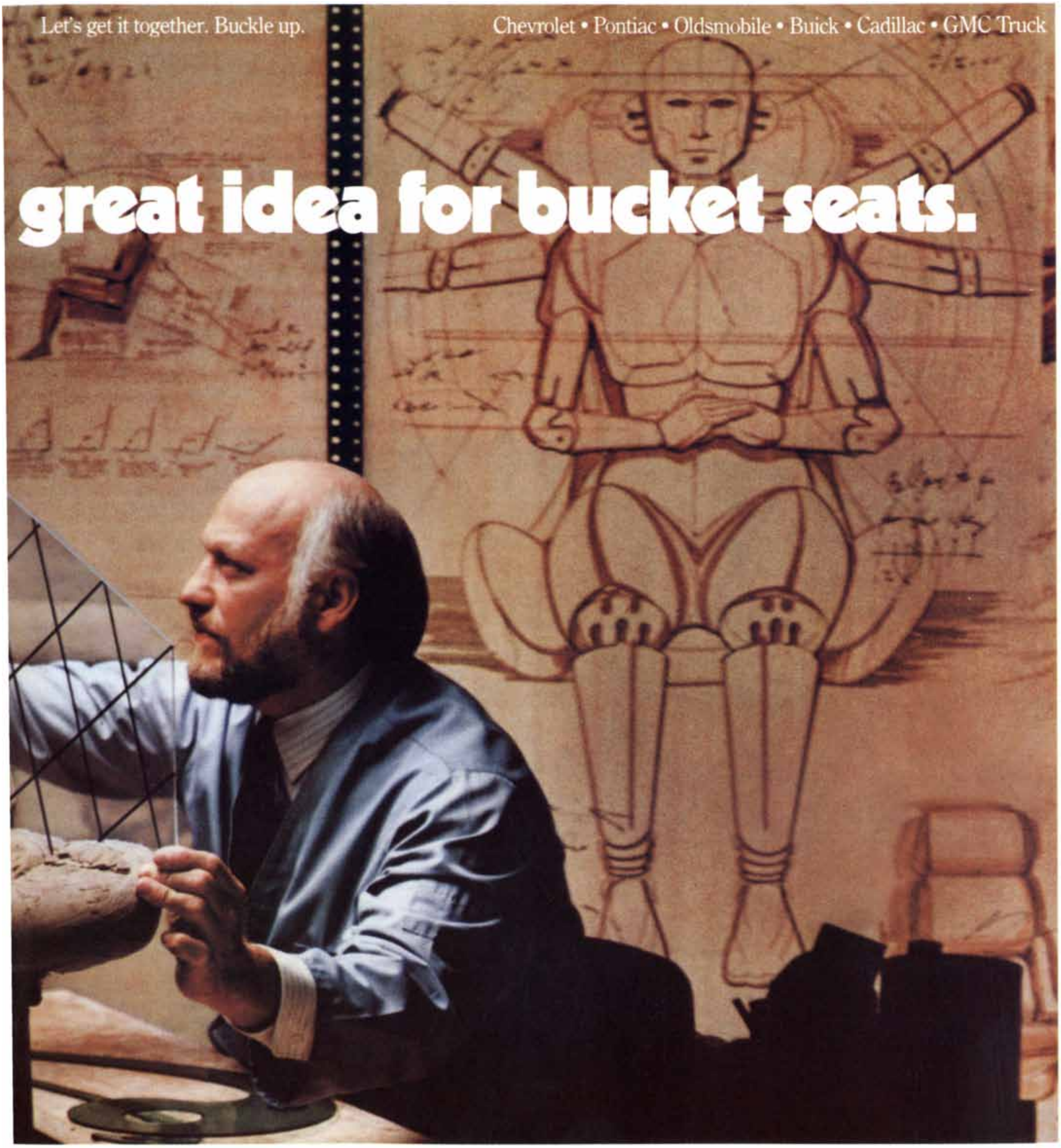
It may take us two years and countless clay models to arrive at a more comfortable, durable seat for new GM cars and trucks. But we think it's worth it.

And we believe old Leonardo would have thought so, too.

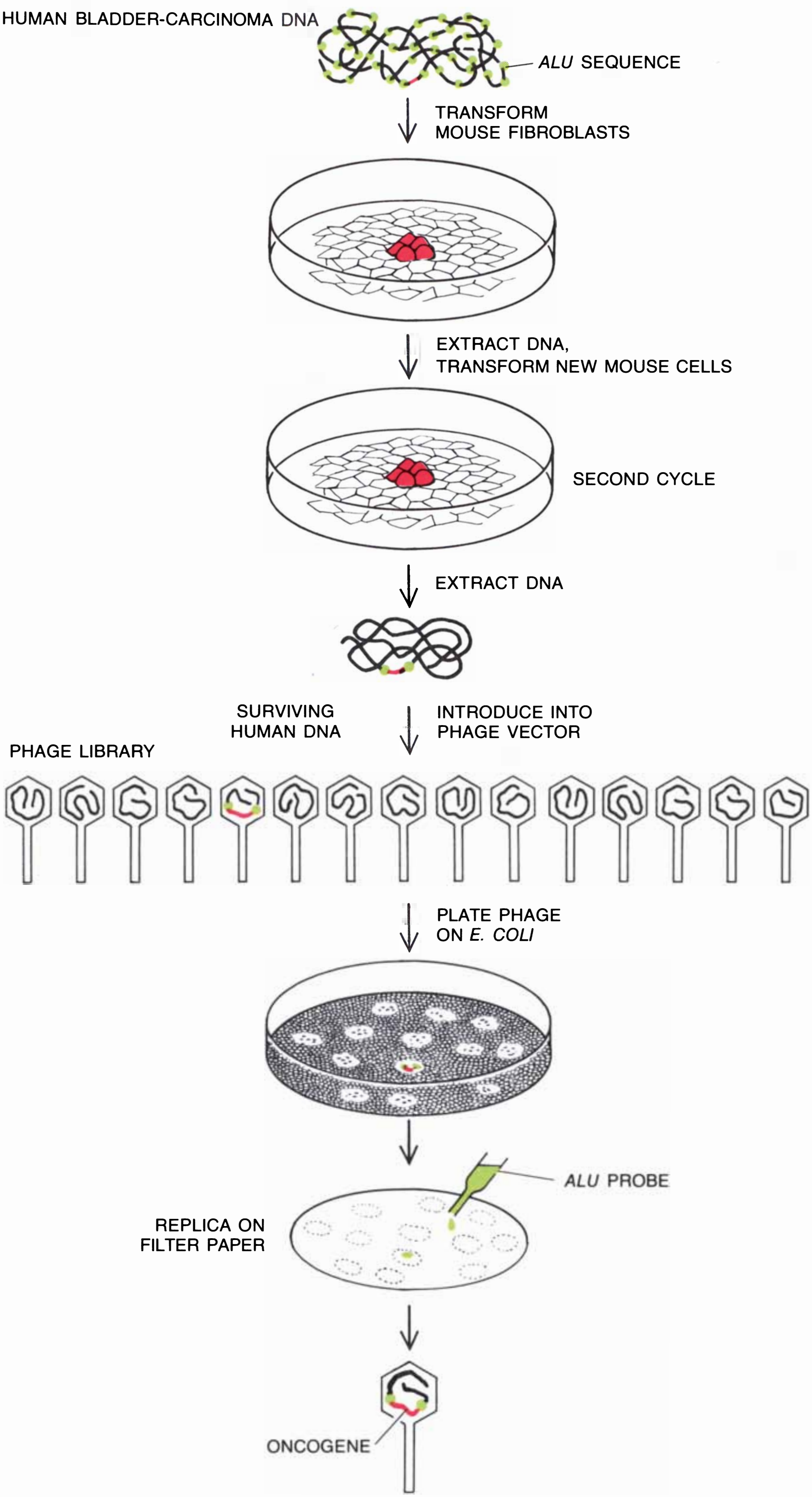
We believe in taking the extra time, giving the extra effort and paying attention to every detail. That's what it takes to provide the quality that leads more people to buy GM cars and trucks than any other kind. And why GM owners are the most loyal on the road.

That's the GM commitment to excellence.

GM



Nobody sweats the details like GM.



PROBE FOR HUMAN DNA enabled Chiaho Shih in the author's laboratory to find the transforming gene of a human bladder cancer. Distinctive short DNA sequences called *Alu* (*green*) are scattered throughout the human genome. Mouse cells were transformed with the tumor DNA; the DNA of the transformed cells was applied to transform new mouse cells in a second cycle. The DNA from the serially transformed cells could include only the small amount of human DNA that was directly associated with transformation (*red*). A phage library constructed from the DNA of the serially transformed cells was plated on *E. coli*. The DNA of the resulting plaques was transferred to a filter-paper replica, where it was probed with *Alu* sequences labeled with a radioactive isotope. Autoradiography revealed the location of the plaque carrying the DNA with wbich the probe hybridized. The phage clone from the plaque was tested for transforming ability and was shown to carry the bladder-cancer oncogene in its DNA.

carrying the bacterial marker gene, which counteracted the genetic defects of the crippled phage, could grow well on a bacterial culture. Among the few viable phages that grew out were several carrying the elusive transforming segment.

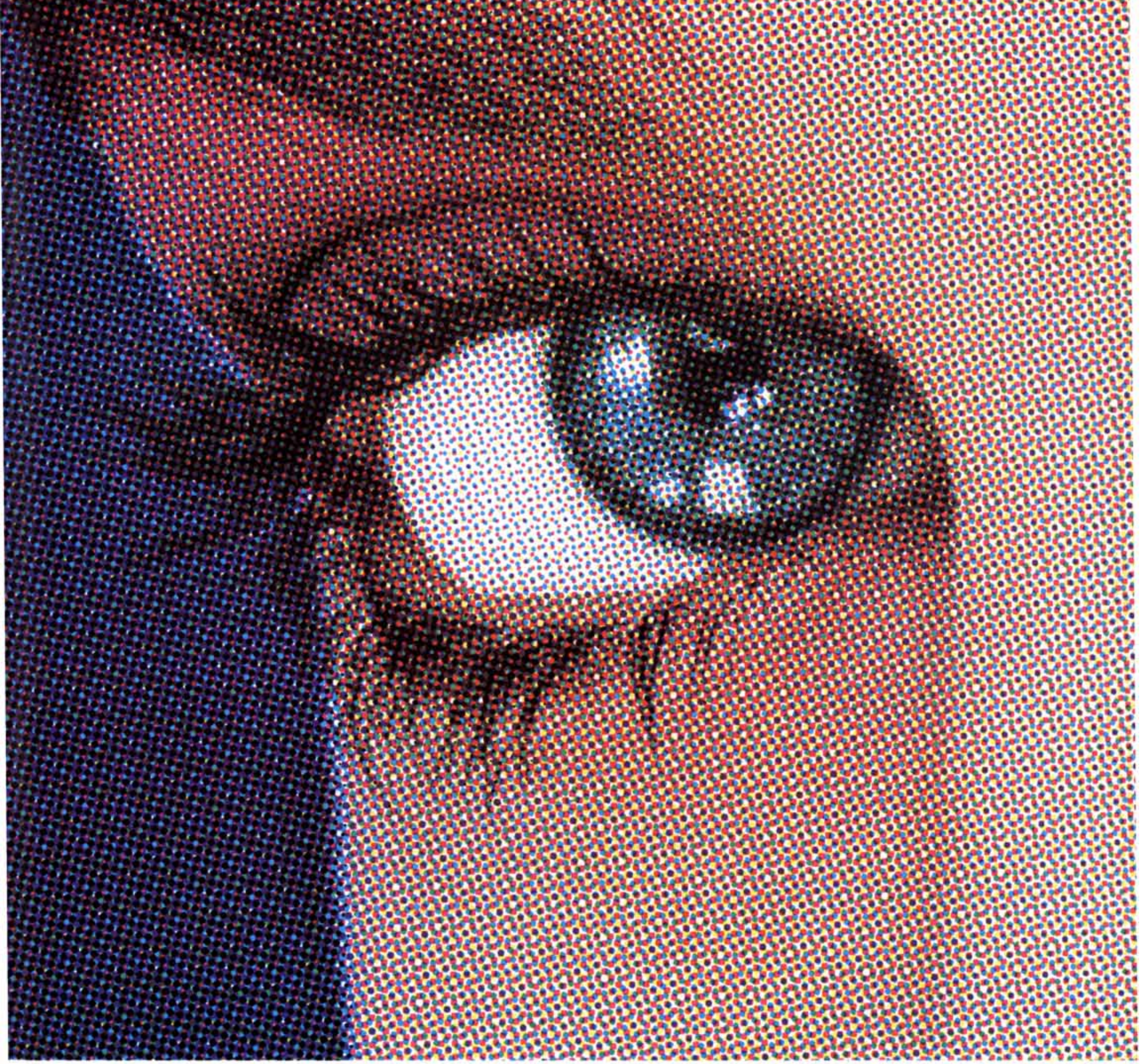
Oncogenes

The successful isolation of transforming DNA in three laboratories by three different methods directly associated transforming activity with discrete segments of DNA. No longer was it necessary to speak vaguely of "transforming principles." Each process of molecular cloning had yielded a single DNA segment carrying a single gene with a definable structure. These cloned genes had potent biological activity. Whereas two micrograms of the original bladder-carcinoma DNA had induced, on the average, one colony of transformed cells, a comparable mass of the cloned transforming gene induced as many as 50,000 foci. The transforming activity previously attributed to the tumor-cell DNA as a whole could now be assigned to a single gene. It was an oncogene: a cancer gene.

What is the origin of a human oncogene? The answer emerged, in Wigler's laboratory and my own and in Mariano Barbacid's laboratory at the National Cancer Institute, from a study of the human bladder-carcinoma oncogene that had been isolated, as described above, from a tumor-cell line variously designated *EJ* or *T*24. When the cloned oncogene was exploited as a probe for exploring the normal human genome, it hybridized readily and strongly with a related DNA sequence in the normal genome. Moreover, the oncogene and the normal gene were found to be the same size. Further characterization of the two genes showed that they were almost identical, but of course we knew they could not be absolutely identical because they functioned so differently: clones of the oncogene transformed cells, whereas clones of the related normal gene had no such effect. They could not be identical twins. Rather, it was clear that the oncogene is a slightly altered version of the normal gene, which is often called a proto-oncogene.

The precursor of the bladder-carcinoma oncogene is not the only proto-oncogene being carried around in human DNA. To date three other active oncogenes of human tumors have been traced to corresponding proto-oncogenes. Why are genes maintained in the human genome that, with slight alteration, become agents able to transform cells and generate tumors? Why should an organism carry the seeds of its own destruction?

All the answers to these important questions are not yet in, but one point is



Sometimes it's important to color the truth.

Did you ever hold a printed picture close enough to see the colors break into dots?

Most color printing is done with dots; tiny dots whose precise arrangement is extremely critical.

The 3M "Matchprint" system tells printers the truth about dot patterns with the most accurate close up available prior to printing. Corrections can be made, true colors reproduced and costly extra proof press runs reduced. Color proofing costs can be cut in half.

We answer the needs of the graphics industry with everything from photographic film to printing plates; even a cost-efficient way to buy advertising in magazines.

Listening to people's needs has helped 3M pioneer over 400 products for the communication arts field. We now make everything from lithographer's tape and overhead projectors to video recorders.

And it all began by listening.

3M hears you...

For a free brochure on all 3M products and capabilities write: Dept. 080711/3M P.O. Box 22002, Robbinsdale, MN 55422.

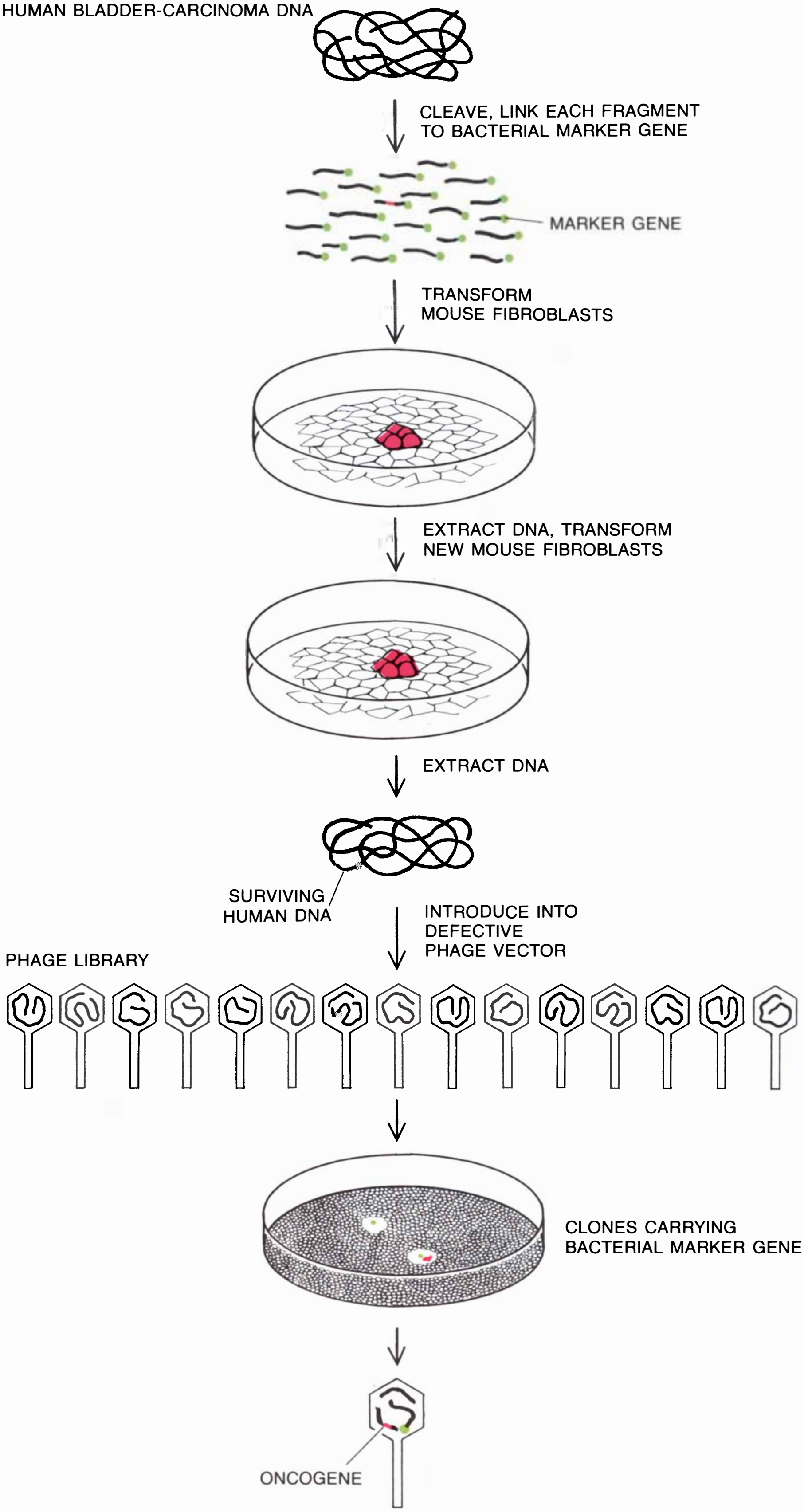
Name____________________

Address____________________

City__________ State_____ Zip______

Or call toll-free: **1-800-33M-HEAR** or 1-800-336-4327.

3M



BACTERIAL MARKER GENE served as a probe for Michael Wigler's group. They cleaved the bladder-cancer DNA and linked each fragment to a copy of the bacterial gene (*green*). By serial gene transfers they prepared mouse DNA carrying no human DNA except for the transforming gene (*red*), which was closely linked to a copy of the bacterial marker gene. A phage library was constructed by introducing the fragmented mouse DNA into a strain of phage that has a defect inhibiting its growth. The bacterial gene had been chosen to overcome that defect. As a result only phage clones carrying the bacterial gene grew when the library was plated on *E. coli*. Some of these clones were found to carry the closely linked bladder-cancer DNA.

clear. The proto-oncogenes would not have been conserved in the genome unless they have some vital role in normal cellular metabolism. And they have indeed been conserved. Relatives of human proto-oncogenes have been found in the DNA of a number of mammals, in chickens and even in the fruit fly *Drosophila*. The ancestors of human proto-oncogenes must therefore already have evolved when a common ancestor of human beings and flies lived, more than 600 million years ago. Proto-oncogenes would not have been kept almost unchanged over such a long time unless they were and continue to be indispensable. The precise role of such genes in normal metabolism is still not known, but there are some indications that it has to do with the control of cellular proliferation. I shall have more to say about that below.

Viral Oncogenes

The gene-transfer experiments revealed one class of oncogenes and proto-oncogenes. Another class had been discovered earlier through work on retroviruses, a group of viruses that cause cancer in some animals and whose genetic material is not DNA but RNA, which is reverse transcribed into DNA when the virus infects a cell. The retroviruses were shown to carry a single gene that is responsible for the induction of cancer: an oncogene. J. Michael Bishop and Harold E. Varmus and their colleagues at the University of California at the San Francisco School of Medicine showed that these viral oncogenes are not truly viral at all but are of cellular origin. They are proto-oncogenes that were picked up and carried (transduced) by retroviruses infecting animal cells. Somehow they are activated to become oncogenes after incorporation into the retrovirus genome and are then exploited to transform cells the viruses happen to infect. To date 17 such cellular proto-oncogenes have been defined, each of them known from its association, as an active oncogene, with a particular retrovirus.

There seemed, then, to be two sets of normal proto-oncogenes and their activated oncogenic derivatives, each set defined by a different method of activation. One group, discovered by gene transfer, comprised oncogenes that arise from the mutation of proto-oncogenes. The other, discovered previously by the virologists, comprised genes that are somehow activated by retroviruses. We now know that the two groups are not mutually exclusive. In the past year or so hybridization experiments have revealed that some oncogenes of human tumors are closely related to the oncogenes carried by several retroviruses that infect rats. The human bladder-carcinoma oncogene referred to above, for example, is a close relative of the oncogene carried by the Harvey sarcoma virus, which acquired it from the rat genome. This means the same proto-oncogene can be activated by two independent mechanisms: by mutation in human beings and by retrovirus acquisition from rats. A second cellular proto-oncogene has been found to follow this pattern. It is activated either by incorporation into the Kirsten rat-sarcoma virus or by mutations that lead to oncogenes present in a number of human carcinomas of the colon, lung, bladder and pancreas, as well as in some human sarcomas.

The two cellular proto-oncogenes, associated respectively with the Harvey and the Kirsten viruses, are themselves related evolutionarily. They are members of a gene family, designated *ras,* that has been studied in detail by Edward M. Scolnick and his colleagues, formerly of the National Cancer Institute and now at the Merck Laboratories in West Point, Pa. A third human oncogene has recently been associated with the *ras* gene family by Wigler's group (but it has not yet been found in a retrovirus). This oncogene, called *N-ras,* is in the DNA of various leukemias, a lymphoma, a neuroblastoma, a colon carcinoma and several sarcomas. The presence of the *ras* oncogenes in a variety of tumor DNA's shows that the activation of a particular cellular proto-oncogene is not confined specifically to a single tissue. The same gene can be activated in any one of several tissues, giving rise in each to a different type of tumor.

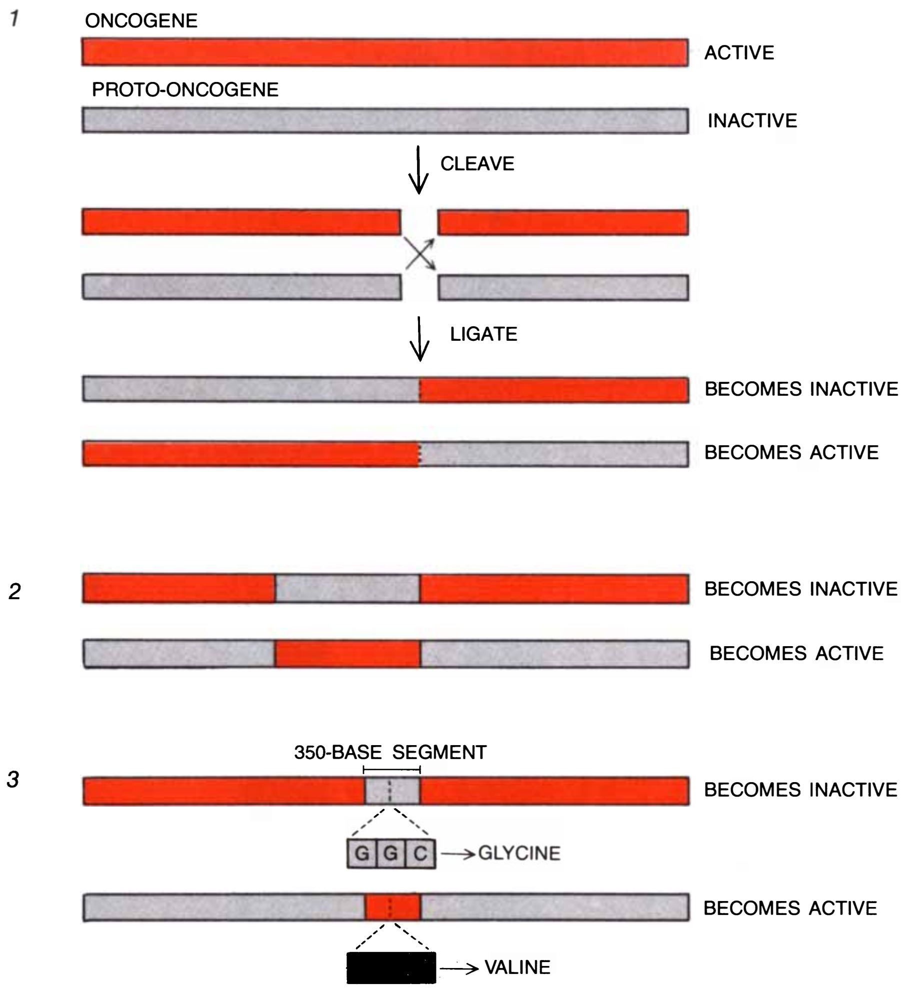
Such a simple insight represents a major advance. From the perspective of a molecular biologist cancer is no longer more than 100 diseases, each of them characterized by a different tumor type. Instead it begins to look as if there are only a small number of molecular mechanisms, which are common to all tumor types. This notion needs to be strengthened by more investigation because so far only a small number of human tumors have been studied in detail.

Defining the Mutation

A mutation was implicated as the mechanism that converts a benign proto-oncogene into an active oncogene, but the nature of the mutation remained obscure. Once cloned versions of both genes became available, the mutation could be localized with precision. In principle one could determine the entire nucleotide sequence of each of the 5,000-base-pair genes and compare the two sequences. In practice it seemed better first to define a region within the genes carrying the critical difference in sequence. This could be done by generating recombinants between parts of the two genes, applying the techniques of genetic engineering to mimic the double crossovers of classical genetics [*see illustration on this page*]. Such experiments were undertaken by three groups of investigators, in my own laboratory, in Barbacid's and in Wigler's.

First the two genes were cleaved into segments at analogous sites with restriction endonucleases, enzymes that cut DNA molecules at points defined by particular nucleotide sequences. Specific segments cut from the two genes were then joined with the enzyme DNA ligase to make two hybrid genes, part of which came from the oncogene and part from the proto-oncogene. The hybrid molecules were tested by transfer into mouse fibroblasts to see which one had lost oncogenic activity and which one had gained it.

In my laboratory Clifford J. Tabin and others narrowed down the region of interest after repeated crossovers of successively smaller segments of the bladder oncogene and its related proto-oncogene. Ultimately they found that introducing an oncogene segment only 350 nucleotides long into the proto-oncogene turned the proto-oncogene into an actively transforming molecule. This meant that the critical lesion distinguishing the two closely related genes must lie within the 350-nucleotide segment. Sequence analysis of the active and inactive versions of the segment by Ravi Dhar of the National Cancer Institute led to a dramatic and unexpected finding. The critical 350-nucleotide sequence in the oncogene and the one in the corresponding segment of the proto-oncogene differ in only one base: a guanine in the proto-oncogene is replaced in the oncogene by a thymine. This meant that a single nucleotide change, a "point mutation," in a 5,000-nucleotide normal human gene could convert it into an oncogene. For the first time a genetic lesion that can lead to the growth of a human tumor had been precisely defined. It may be that a critical event in the initiation of many cancers is the mutation of a



POINT MUTATION that converts a proto-oncogene into the active oncogene of the *EJ* bladder-carcinoma cell line was identified by first locating the segment that, in each of the genes, is the site of the critical difference and then sequencing the two segments. The proto-oncogene and the oncogene (*1*) were cleaved at the same site and the resulting segments were ligated as is shown by the broken lines. The recombined genes were tested by gene transfer to see which had become active and which had lost activity. Successively smaller segments were similarly interchanged (*2*) until a sequence only 350 bases in length was shown to be critical (*3*). There is a single mutation in that segment: a guanine (*G*) in the proto-oncogene is converted into thymine (*T*) in the oncogene, and so one triplet codon specifies the amino acid valine rather than glycine.

proto-oncogene as the result of some environmental insult to its DNA.

Recently other mechanisms of proto-oncogene activation have been described. One mechanism, chromosomal rearrangement, was first discovered by Grace L. C. Shen-Ong and Michael D. Cole of the St. Louis University Medical Center and by Rebecca A. Taub and her colleagues in the laboratory of Philip Leder of the Harvard Medical School. This mechanism relates to the activation of a proto-oncogene called *myc,* whose actively oncogenic version has been found in mouse myelomas and in a human cancer, Burkitt's lymphoma. The cellular proto-oncogene becomes separated from its usual position at the end of one chromosome and is transposed to the end of a second chromosome. There it becomes juxtaposed with genes responsible for the synthesis of immunoglobulins (antibodies). Such genes are transcribed at a high rate in the course of an immune response to an antigen. The association of the proto-oncogene with an immunoglobulin gene apparently deregulates the proto-oncogene and thereby imparts to it a novel biological function.

In other cases a proto-oncogene has been found to be present in large numbers of copies per cell instead of the normal two copies characteristic of most cellular genes. This amplification of the gene, and the resulting high level of gene expression, seems also to have an oncogenic effect. There are, in other words, a number of molecular mechanisms that can force a normal gene into the role of an oncogene. In the case of mammalian tumor genes they include point mutation, chromosomal rearrangement and gene amplification. In the case of retrovirus-associated genes the transduced gene may undergo mutation or may end up adjacent to a viral regulator that increases its level of expression.

Cancer Proteins

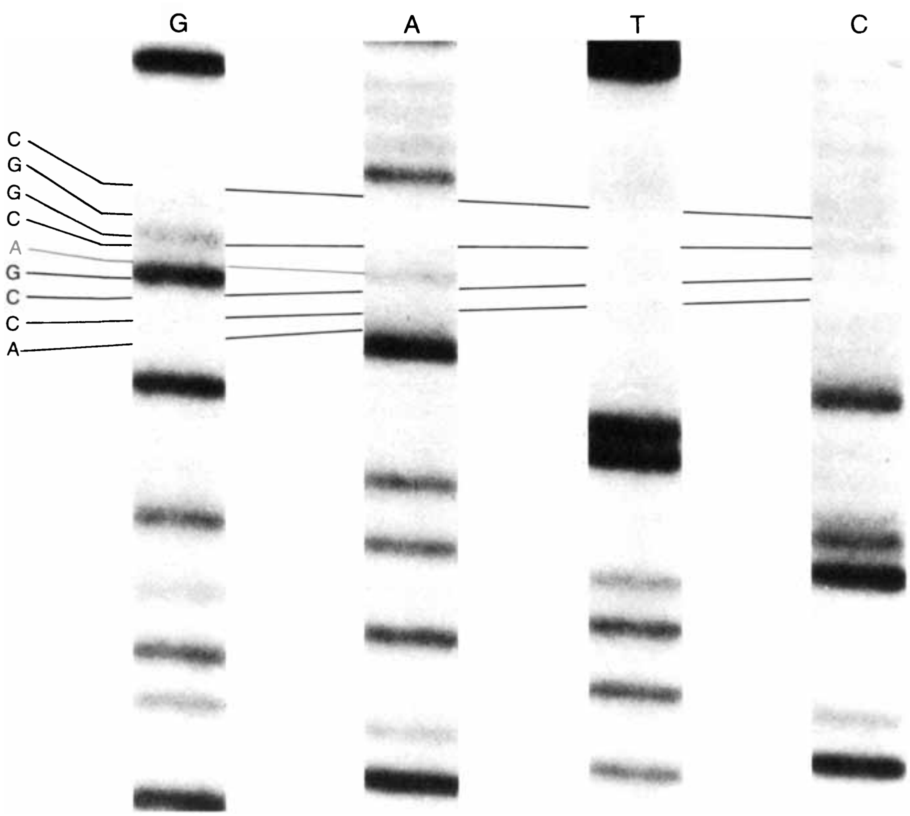
The search for the underlying mechanisms of cancer has progressed a long way, from vaguely described genetic factors to molecules of DNA, and then on to particular genes within the DNA. The search seemed to end with the finding of a precisely defined molecular alteration in one such gene. In fact, the most difficult problems are yet to be resolved.

Having granted that the creation of an oncogene may give rise to a cancer, one still has no insight into how the altered gene contributes to this process. What is the function of an oncogene? Like other genes, it encodes the structure of a protein. Scolnick's group has shown that many of the *ras* genes direct the synthesis of a single protein having a molecular weight of 21,000. The point mutation observed in the bladder-carcinoma oncogene alters the structure of the protein. Each amino acid of a protein is specified by a three-base codon in the gene encoding the protein's structure. The observed point mutation changed a *GGC* codon of the proto-oncogene into *GTC* in the oncogene. This resulted in turn in a change in the structure of the oncogene protein, which carries the amino acid valine at one site instead of the usual glycine. Apparently that single substitution causes the protein to assume novel functions enabling it to profoundly alter cellular metabolism.

Cellular elements such as these oncogene proteins must lie at the heart of the cancer process, but just how such proteins work remains unclear. One assumes that they somehow regulate the growth of cells, with the protein encoded by a proto-oncogene controlling normal growth whereas the version encoded by the oncogene forces cancerous growth.

An understanding of the precise functions of oncogene proteins may eventually make it possible to develop antagonists that inhibit those functions, and so perhaps to develop a therapy striking at one central defect of the cancer cell. There are probably, however, a number of defects. Carcinogenesis appears to be a multistep process. Evidence from a variety of sources indicates that a normal cell must suffer several independent alterations before it becomes a bona fide tumor cell. A point mutation that gives rise to an aberrant version of a single protein represents only a single step affecting a single gene. Presumably the creation of an oncogene fulfills one requirement for making a tumor, but there are other necessary steps; the oncogene may be necessary, but it is hardly sufficient. Clues are beginning to emerge as to the nature of the other steps.

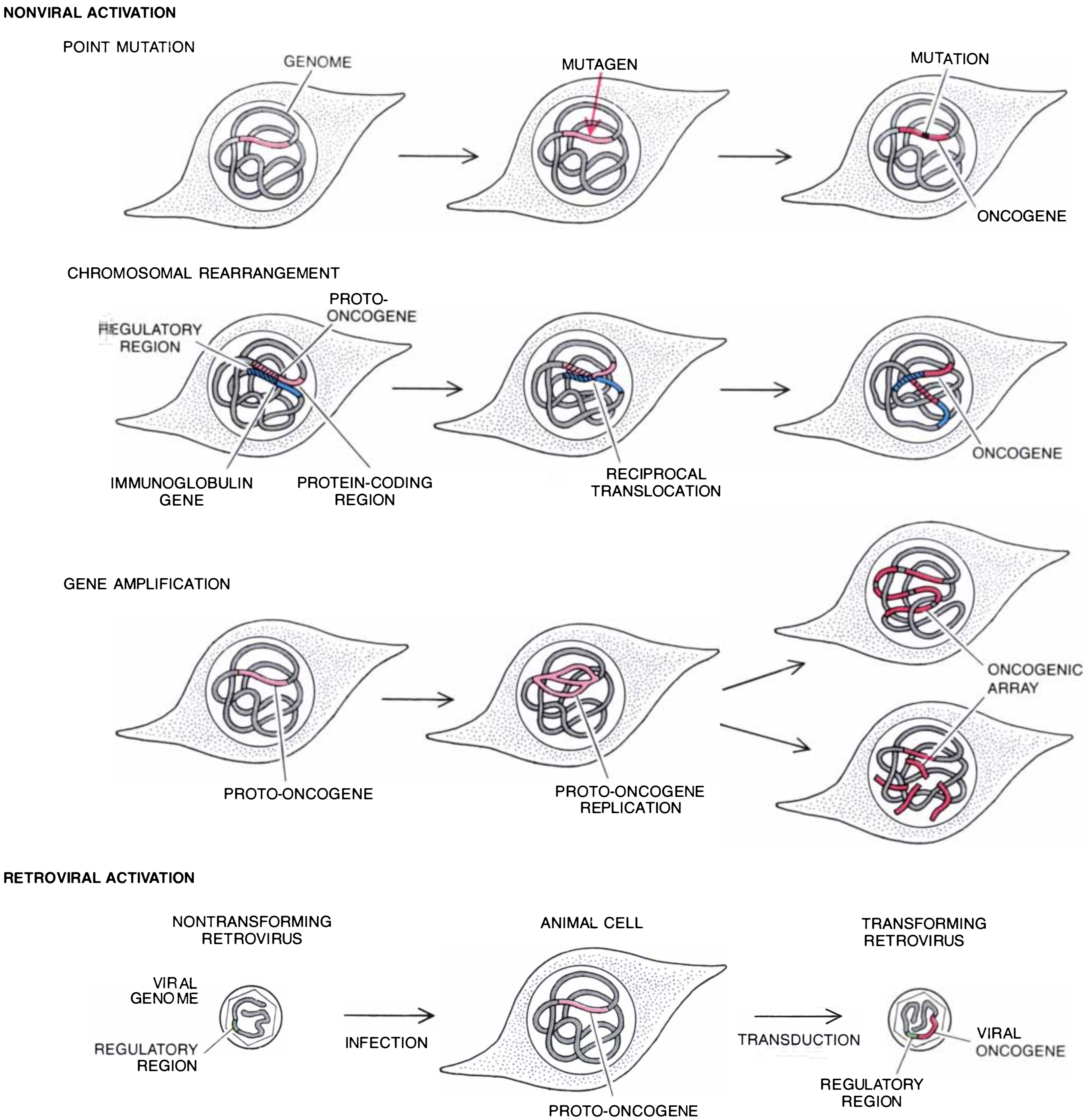
A variety of tumors have been found in which two distinct oncogenes have arisen as a consequence of separate activation events. For example, William S. Hayward of the Rockefeller University and Susan M. Astrin of the Institute for Cancer Research have found that the *myc* proto-oncogene is activated in chicken lymphomas, whereas Cooper's group has found that a second proto-oncogene, designated *B-lym,* is activated in the same tumor cells. In both a Burkitt's lymphoma and a human promyelocytic leukemia two oncogenes have been found in the activated state. Such data suggest that the multistep nature of carcinogenesis may derive in part from a requirement for the activation of several



BASE SEQUENCE of the 350-base segment in the proto-oncogene and in the oncogene was determined by Ravi Dhar. This autoradiograph of an electrophoresis gel shows the critical mutation in the oncogene. The ladderlike array of bands represents successively shorter (from top to bottom) fragments of the segment. Four sets of fragments were prepared. Each set, labeled by the dideoxy method to end at occurrences of one of the four nucleotides, was run in a different channel on the gel. Careful inspection of the position of each band allows one to read off the sequence. Because this gel was prepared from the oncogene's complementary strand rather than the coding strand, the mutation appears (*color*) as adenine (*A*) rather than thymine (*T*).

distinct oncogenes. Each oncogene may elicit a different alteration within the cell, and together they may collaborate to achieve important aspects of the tumor phenotype.

I have raised only two of the many issues that remain to be explored. What is most heartening is that the confluence of evidence from a number of lines of research is beginning to make sense of a disease that only five years ago seemed incomprehensible. The recent findings at the level of the gene are consistent with earlier insights into carcinogenesis based on epidemiological data and on laboratory studies of transformation. Powerful tools are available for making even more rapid advances in the coming years. The molecular mechanisms underlying cancer should be well defined by the end of this decade.



PROTO-ONCOGENE CAN BE ACTIVATED in four ways, three of them involving the alteration of cellular DNA and one involving intervention by an infecting retrovirus. One nonviral way is point mutation. If a proto-oncogene in a normal cell undergoes mutation by radiation or a chemical carcinogen, the mutation may change the information encoding a protein. The altered protein may induce one or more of the changes associated with cancerous growth. Alternatively the protein-encoding region of a proto-oncogene may become recombined, in the course of chromosomal rearrangement, with the regulatory region of an immunoglobulin gene, causing the proto-oncogene's genetic information to be expressed at an inappropriate level. A third method of activation is amplification: a proto-oncogene is somehow replicated until it is present in a cell in multiple copies, either sequentially repeated segments along the chromosome or dissociated extrachromosomal particles; again the result is inappropriate overexpression. Retroviruses, which cause cancer in many animals, acquire their oncogenes by infecting an animal cell. In the course of infection the virus picks up a cellular sequence that includes the proto-oncogene and integrates it into its own genome. The proto-oncogene may then become activated by association with a regulatory region of the viral genome, as is shown here, or by undergoing mutation while it is being carried (transduced) by the retrovirus.

"Microsoft® Multiplan™ allows me to explore more alternatives in less time. I think it leads to better management decisions."

Edward R. Schwinn, Jr.
President, Schwinn Bicycle Company

"Multiplan software helps me make better business decisions. It's as simple as that. With Multiplan on my microcomputer, I can explore a number of alternatives faster. That's really important to me in planning and tracking my totally hand-crafted Schwinn Paramount Bicycle. As with anything handmade, cost control and product planning are major considerations. For instance, Multiplan lets me study the effect of the learning curve on overall productivity. And although we have a mainframe computer, I find that Multiplan is a tool that allows me to personally examine my business options—without spinning my wheels."

A tool for decision makers. Microsoft Multiplan is a microcomputer software program that can help you too. In many ways. You can analyze cash flow. Plan budgets. Forecast income. Manage production. Multiplan allows you to set up an electronic worksheet for whatever your need may be. And, it lets you explore alternatives faster because when you make one change you immediately see the results of that change throughout the worksheet and on any related worksheets. Just change one number and every number that depends on it is adjusted automatically.

In plain English. You won't have to learn a cryptic language to use Multiplan. It takes commands in plain English. And, it will prompt you as you go along by telling you what to do next. There's even a HELP key in case you need help at any point. Press it and it gives you information to get you going again.

See it in action. If you'd like to see Multiplan in action, and receive our latest brochure, call us toll-free:

1-800-426-9400
In Washington State call 206-828-8080

We'll gladly locate the Microsoft dealers nearest you. They'll demonstrate Microsoft Multiplan and explain its compatibility with most popular personal computers. Including Apple® and the IBM® PC. It's compatible, in fact, with any computers that use the CP/M®-80 or MS™-DOS operating systems.

Microsoft Multiplan is a good management tool for building bicycles. It's also a good management tool for building businesses. Like yours.

BETTER TOOLS FOR MICROCOMPUTERS

MICROSOFT™

Microsoft is a registered trademark, and Multiplan, MS, and the Microsoft logo are trademarks of Microsoft Corporation.

Solid-State Superlattices

They are crystals grown by depositing semiconductors in layers whose thickness is measured in atoms. Their properties promise to make them important in solid-state physics and in technology

by Gottfried H. Döhler

"You see things; and you say, 'Why?' But I dream things that never were; and I say, 'Why not?'" These sentences by George Bernard Shaw reflect the spirit in which a few solid-state physicists began a decade ago to speculate about the physical properties, and even the usefulness, of nonexistent semiconductors. The physicists were realistic enough to confine their thoughts to semiconductors that could exist in principle, although nature had never made them. Specifically they envisaged solid-state superlattices: semiconductors in which two materials with different electronic properties are interleaved in thin layers by depositing sheets of two semiconducting materials in alternation or by introducing impurities into layers of a single semiconducting material.

Their hopes have been honored by twofold success. First, it has turned out that superlattices can be "grown" with precision. Second, the properties of the superlattices have confirmed the investigators' predictions. Some of the properties are exotic: a paradoxical decrease in electric current with increasing voltage, for example, or the survival of free electrons in a superlattice for periods of hours rather than billionths of a second. The prediction of such properties before superlattices even existed is indicative of the progress in solid-state physics. It also signals a tendency that will gain increasing importance: the use of the quantum theory of solids to guide the engineering of new and useful materials.

As a prelude to describing the properties of solid-state superlattices it is appropriate to review the fundamental characteristics of semiconductors. Consider a pure and perfect crystal of a semiconducting material such as silicon. At a temperature of absolute zero (zero degrees Kelvin) the material would be an insulator because its electrons (the charge carriers responsible for electrical conduction in metals) would not be free to move. The explanation for this lack of mobility is that the electrons in the semiconductor can have only certain levels of energy and that the levels form energy "bands." A "band gap," an appreciable difference in energy, separates the material's "valence band," which is completely occupied by the electrons forming the chemical bonds that hold the crystal's atoms together, from its "conduction band," which is empty. At zero degrees the band gap would be insurmountable; at greater temperatures an occasional electron acquires the energy needed for it to reach the conduction band. That is, the electron becomes mobile.

What does the "occupation of a band" really mean? Suppose N atoms are bound in a solid. Each atom, if it were isolated, would have a number of energy states available to its electrons. The tenet of quantum physics known as the Pauli exclusion principle dictates, however, that each state could accept only two electrons at most and that the spin (the intrinsic angular momentum) of the two would have to be aligned in opposite directions. In the solid these conditions are changed: each state is split into a set of N substates. In a solid of appreciable size N is at least 10^{23}; thus the substates are so numerous and so closely spaced in energy that they form a quasi-continuum. Still, each state in each atom can accept only two electrons.

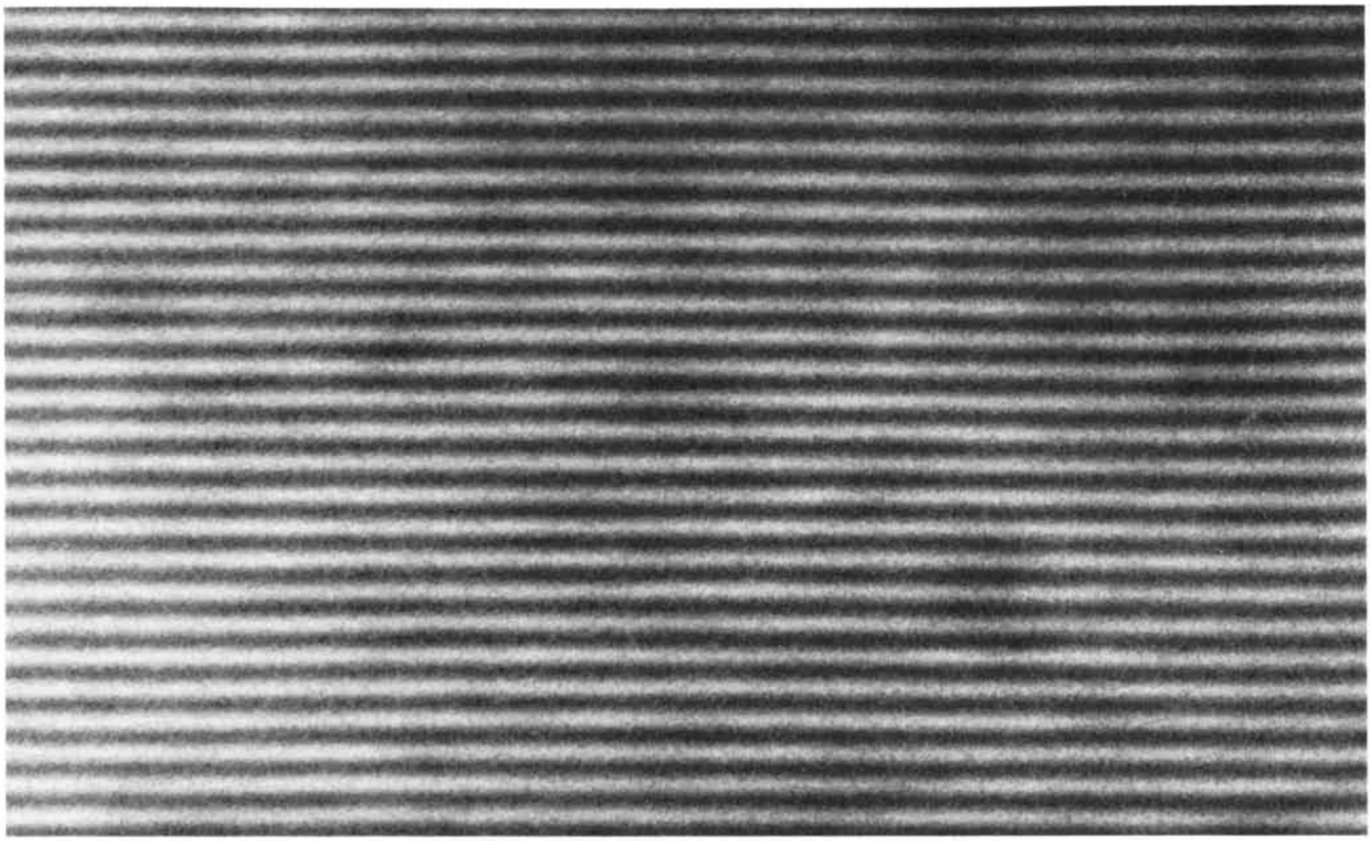
Now suppose each atom has an even number of electrons. In such atoms the states of lowest energy are fully occupied, each state by two electrons; hence the valence band is full. Suppose further that the valence band and the conduction band do not overlap in energy, yet are not very far apart. The solid is then a semiconductor. (If the valence band and the conduction band did overlap, electrons from the uppermost levels of the valence band would "flow" freely into lower-lying conduction-band levels.) Suppose the number of atoms is odd, so that the uppermost state in the valence band is only half filled. Then the solid is a simple type of metal.

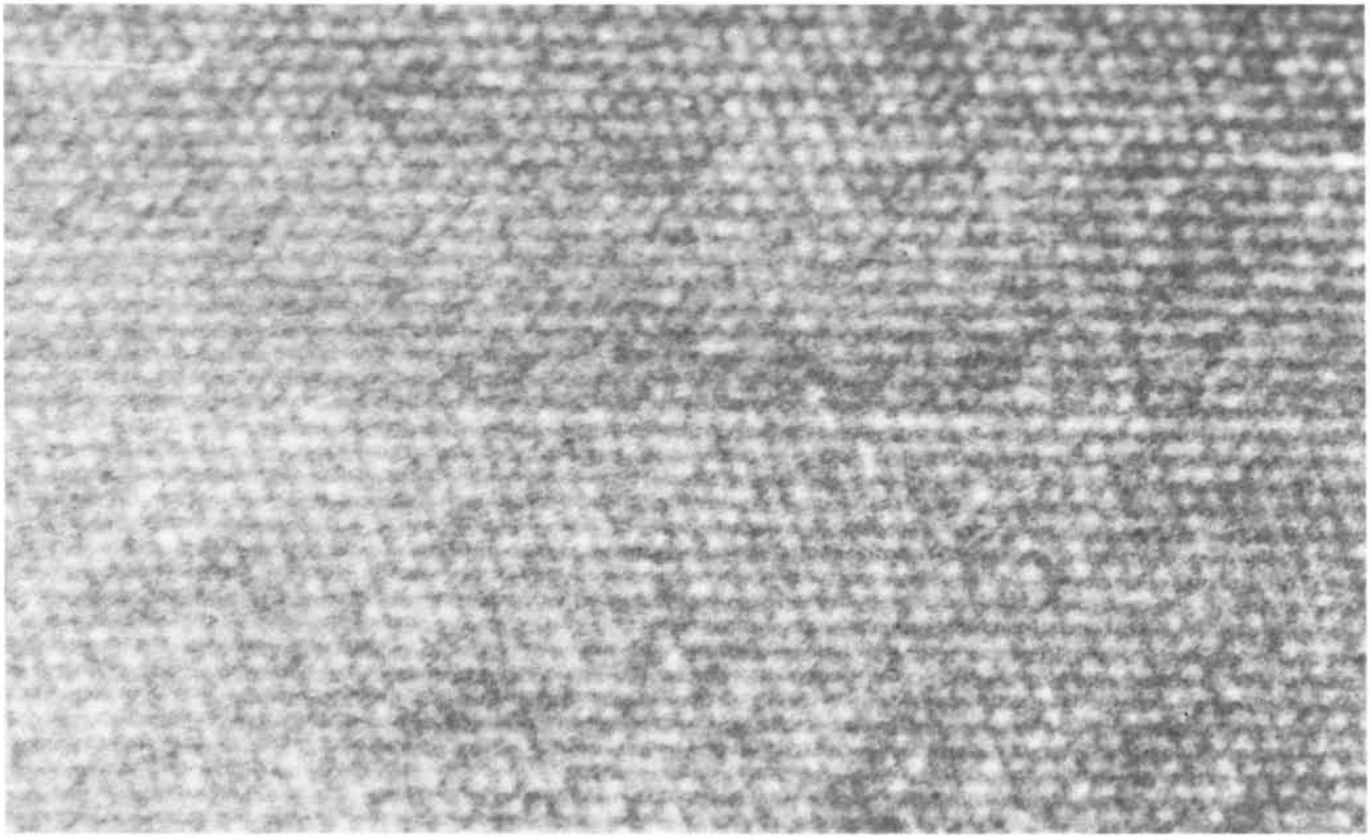
It is at least plausible from this discussion why a simple metal is a conductor. The electrons in the metal are readily accelerated by an electric field because the difference in energy between an occupied substate and unoccupied substates in a quasi-continuum of 10^{23} substates is negligibly small. In contrast the electrons in a semiconductor must be transferred across the band gap and into the conduction band before they can gain energy gradually in an unoccupied quasi-continuum. The transfer can be achieved in several ways, such as when the semiconductor absorbs a photon (a quantum of light) with an energy greater than the width of the band gap.

One further characteristic of semiconductors is also fundamental. Each electron excited into the conduction band of a semiconductor leaves behind a "hole" in the valence band, that is, it leaves behind the absence of an electron in one atom or another. The absence can shift from atom to atom; hence the hole acts in every respect as a carrier of charge, and it contributes to the crystal's conductivity. (In a way, then, the terms conduction band and valence band are misleading.) The charge of an electron is negative; that of the hole is positive.

Suppose a free electron and a hole are generated in a semiconductor by the absorption of a photon whose energy exceeds the band gap. They will lose the excess energy quite fast by warming the crystal. Thus the electron will "tumble downward" to the lower edge of the conduction band. The hole, meanwhile, can be thought of as "ascending" to the upper edge of the valence band. If the two recombine, the semiconductor will emit a photon whose energy is equal to that of the band gap. Clearly the energy value of the lower edge of the conduction band and that of the upper edge of the valence band are important quantities for describing a semiconductor.

The first solid-state superlattice to become a reality was a compositional (or





COMPOSITIONAL SUPERLATTICE is a periodic array consisting of alternating layers of two different semiconductors. In the electron micrograph at the top the semiconductors are gallium arsenide and aluminum arsenide. In particular, each dark band is six atomic planes of gallium arsenide; each bright band is four atomic planes of aluminum arsenide. The distance between successive atomic planes is 2.83×10^{-10} meter. In the electron micrograph at the bottom a single interface between a layer of gallium arsenide (*top half*) and a layer of gallium aluminum arsenide (*bottom half*) is shown. The interface, which runs across the middle of the image, is a single atomic plane thick. The distance between planes is 3.26×10^{-10} meter. Each bright point in the image is a column of atoms viewed head on in the superlattice crystal. The electron micrographs were provided by Pierre M. Petroff and Arthur C. Gossard of Bell Laboratories.

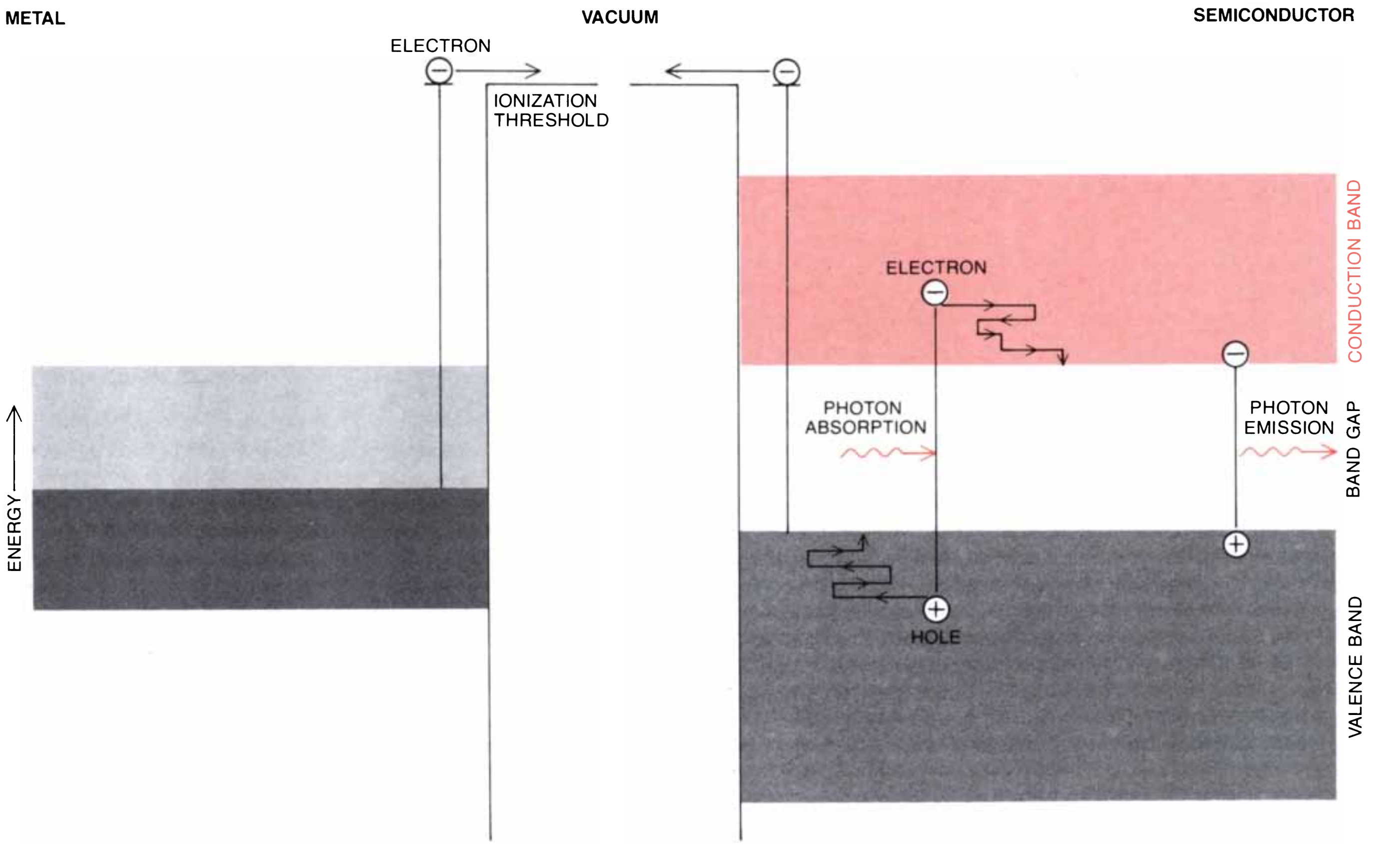
heterostructure superlattice. This novel semiconducting material had been proposed in 1970 by Leo Esaki and Ray Tsu of the Thomas J. Watson Research Center of the International Business Machines Corporation. It was grown a few years later by Esaki and Tsu and their colleagues.

A compositional superlattice is a periodic array of ultrathin layers of two different semiconductors in alternation. Each layer is no more than a few hundred atoms thick. The semiconductors are chosen so that their band gaps are different. Thus the simple banding of energy levels in a bulk semiconductor yields to a more complex pattern resulting from the spatial variation in the material's electronic properties. The most straightforward way to determine the pattern is to treat the superlattice as simply a series of semiconducting layers, with each layer contributing its characteristic properties. More sophisticated calculations show that this strategy gives satisfactory results, even in the case of extremely thin layers. It can be said, therefore, that the periodic alternation of layers gives rise to a periodic alternation in electric potential. Specifically, each layer of the semiconductor with the smaller band gap produces what is called a potential well [*see illustration on opposite page*]. Inside each potential well only certain energy states are available to conduction-band electrons. (Again each state is split into a quasi-continuum, or miniband.) The situation is similar to that of a crystalline solid, where periodicity is established by the presence of atoms at regular intervals. Each atom creates a potential well, and in each atom only certain energy states are available to electrons.

There is an important difference, however. The electronic properties of the atom are predetermined by nature. The electronic properties of the superlattice can be designed. In the first place the values of the energy levels available to electrons can be tailored by the appropriate choice of semiconductors and of the width of their layers. (Basically the same is true of the levels available to holes.) In addition the width of each miniband can be tailored. The width (which turns out to be determined by the strength of the interaction between neighboring potential wells) increases with decreasing width of the layers of the semiconductor whose band gap is greater.

In practice the semiconductor with the smaller band gap can be a material such as gallium arsenide (GaAs) and the one with the larger band gap can be aluminum gallium arsenide (AlGaAs). (Any ratio of aluminum atoms to gallium atoms is possible, but the total number of aluminum and gallium atoms must equal the number of arsenic atoms.) Calculations for a compositional superlattice consisting of layers of such materials show that the minibands are much narrower than the bands in any bulk semiconductor and also that the minibands are separated from each other in the conduction band by relatively large "minigaps." The only requirement is that the thicknesses of the two types of layers be in the range from four to 10 nanometers. (A nanometer is a billionth of a meter.)

The minibands provide the ideal environment for the phenomenon of solid-state physics known as Bloch oscillation, which seems to be undetectable in any solid occurring in nature. It relies on the fact that a high electric field tilts the bands in a semiconductor. The slope of the tilt is *e* (the charge of an electron) times the voltage applied to the crystal, divided by the length of the crystal; thus the slope increases with increasing voltage. In a conventional semiconductor the electrons in the conduction band are driven toward the upper edge of the tilted band [*see upper illustration on page 148*]. Long before they reach it, how-



CONTRASTING ELECTRONIC PROPERTIES of a metal and a semiconductor are displayed in diagrams that show the energy levels available in each material to carriers of electric charge. In the metal (*left*) the carriers are electrons, which move freely in a partly filled band of available energies. In the semiconductor (*right*) the electrons that form bonds between atoms fill up a "valence band." They are excited into a "conduction band," which otherwise is empty, by the absorption of photons (quanta of light) whose energy exceeds the "band gap." The excitation leaves "holes" in the valence band. When electrons and holes recombine, their energy is dissipated by photons.

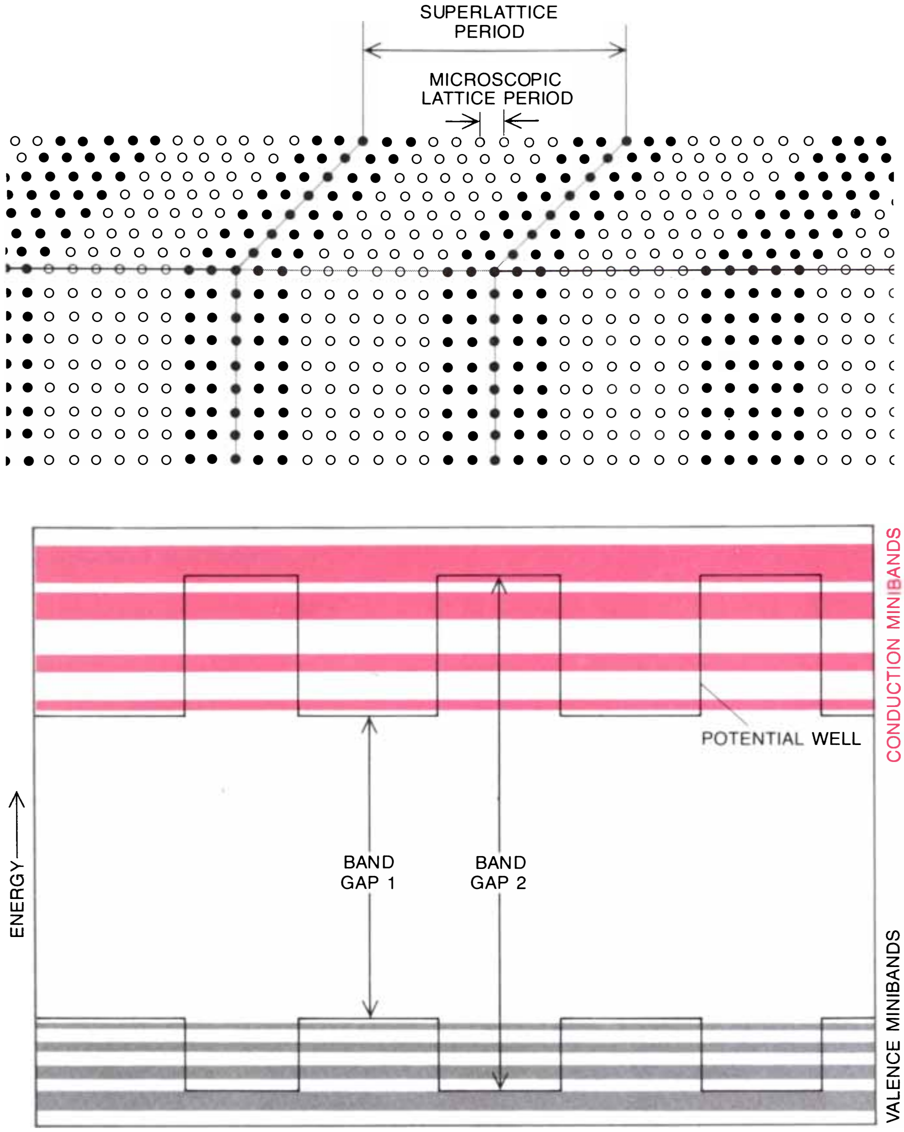
ever, they lose the energy they have gained from the field by emitting "phonons," that is, by exciting thermal vibrations in the crystal.

In a superlattice the situation is different. The minibands are narrow, and so the probability of an electron's reaching the upper edge of a tilted miniband is high. At the upper edge the electrons are reflected, in essence because they cannot pass through a minigap into the next miniband. After such a reflection (the quantum phenomenon known as Bragg reflection) they turn back toward the bottom edge of the miniband. Indeed, they can be reflected repeatedly so that they shuttle back and forth between the upper and the lower edge of the miniband, performing many of these Bloch oscillations before they emit a phonon and "fall" to a lower energy.

The average distance by which the emission of a phonon shifts the average position (the center of mass) of an electron along the conduction band decreases with increasing tilt of the band. Thus an increase in the voltage applied to a superlattice can have the curious effect of making the current flowing through it decrease. In other words, the crystal can exhibit a negative resistance: it can refrain from consuming energy like a resistor and instead can feed energy into an oscillating circuit. A superlattice might therefore serve as the active element of an electromagnetic wave generator. Unlike other negative-resistance devices it would react almost instantaneously to a voltage; hence it could generate microwave radiation with wavelengths of less than one millimeter. Efficient generators for this range of wavelengths are not available today.

A remarkable thing about the Bloch oscillations in a compositional superlattice is that Esaki and Tsu predicted them before superlattices existed. Then came the fabrication of superlattices, by the technique of molecular-beam epitaxy, which Alfred Y. Cho and John R. Arthur, Jr., had developed at Bell Laboratories in the late 1960's. In this technique crystals are grown by directing beams of their constitutent atoms or molecules at a substrate such as a gallium arsenide wafer. The beams deposit matter at a rate of about one atomic layer per second. A few years later Esaki, working with L. L. Chang, reported the first experimental verification of the prediction that had started it all.

This first success was soon followed by another. Raymond Dingle, William Wiegmann, Arthur C. Gossard and Charles H. Henry of Bell Laboratories had been investigating the absorption of light by a compositional superlattice consisting of gallium arsenide and gallium aluminum arsenide. They could predict that photons would be absorbed only if their energy equaled or exceeded



PROPERTIES OF A COMPOSITIONAL SUPERLATTICE result from the periodic alternation of its two different semiconductors. The two are chosen to have band gaps that differ in width, and the difference creates potential wells: abrupt drops in the potential energy inside the crystal (*bottom chart*). In turn the potential wells split the valence band (*gray*) and the conduction band (*color*) into minibands. The crystal can be tailored (that is, the width and the energy of the minibands can be set) by the choice of semiconductors and of the thickness of their layers.

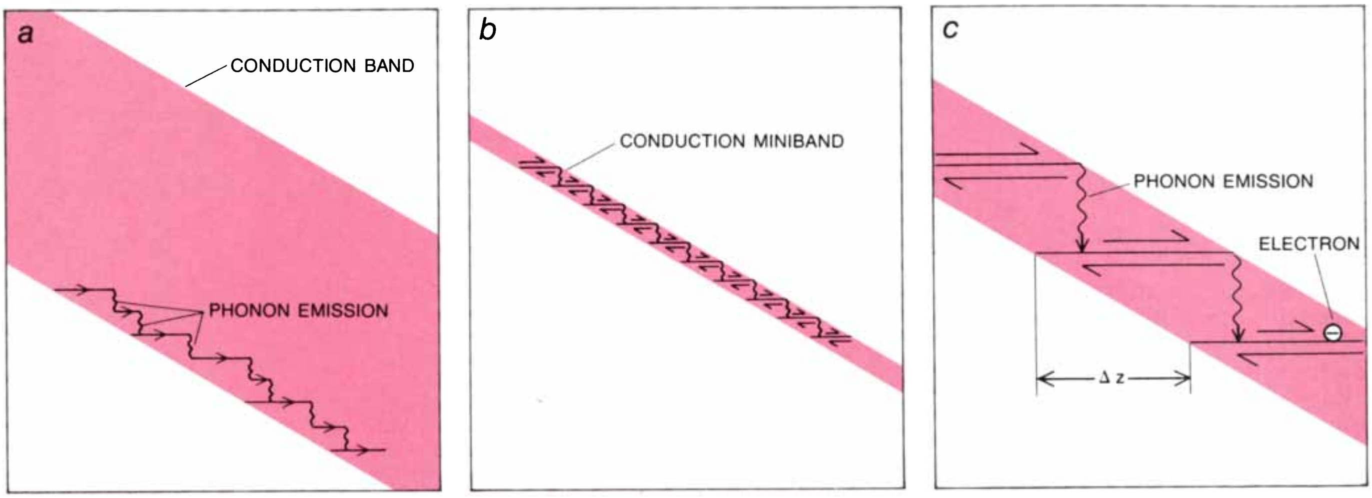
a threshold value determined not by the band gap of either semiconductor alone but by the effective band gap: the difference in energy between the least energetic conduction miniband and the most energetic valence miniband [*see illustration above*]. Further, they could predict that the absorption would sharply increase near photon energies corresponding to the differences in energy between other pairs of minibands.

Experiments confirmed these predictions; thus quantum physics proved itself in a novel manmade material. Conversely, the confirmation of the predictions established that the superlattice had indeed been grown with precision. This raised the possibility that superlattices might serve the study of complex quantum-mechanical phenomena. Specifically, the interactions among the particles in a quantum-mechanical many-body system might be investigated under simple and well-defined conditions if a rectangular potential well engineered with great precision in a superlattice were occupied by many charge carriers.

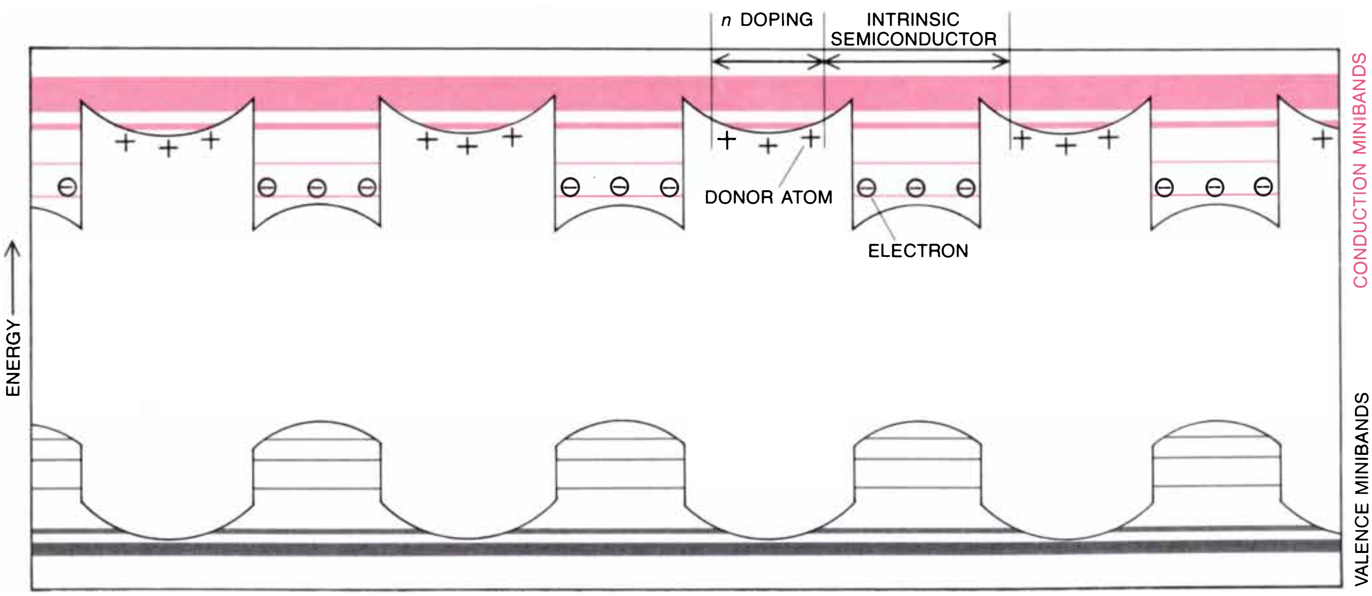
In this respect too superlattices exhibit remarkable properties. So far I have described only semiconducting materials in which charge carriers are generated by the absorption of light. Instead of such carriers, which disappear rapidly owing to electron-hole recombination when the excitation of the crystal has stopped, one can introduce a permanent population of electrons or holes by "doping" the semiconductor with impurity atoms. Atoms whose number of valence electrons (electrons that can enter into chemical bonds) is one greater than that of the atoms for which they substitute in a host crystal are donors: they can easily give away the loosely bound excess electron. Thus a semiconductor doped with donor atoms is no longer an insulator even at a temperature of absolute zero: it is an electron conductor, or *n*-type semiconductor, where *n* stands for the negative charge of the electron. Conversely, atoms whose number of valence electrons is one lower than that of the host-crystal atoms are acceptors: they can easily capture a valence electron from a host-crystal atom, thereby producing a hole. Hence a semiconductor doped with acceptor atoms is a hole conductor, or *p*-type semiconductor, where *p* stands for the effective positive charge of the hole.

Unfortunately the donors and acceptors in a semiconductor do more than provide charge carriers. At the same time they reduce the mobility of the carriers. The reason is that a donor, having given up an electron, or an acceptor, having gained one, becomes an ion and therefore gets an electrostatic field that impedes the motion of charge carriers by scattering them. For the engineer this is disadvantageous: it means the electrons and holes will not move as fast as one would wish. For the investigator of quantum-mechanical many-body systems it is also disadvantageous: it means the interaction among electrons will be obscured by the interaction between electrons and impurity atoms, which is of comparable strength.

In compositional superlattices the disadvantage can be overcome by a simple trick. A doping of donor atoms can be confined to the layers of the semicon-



BLOCH OSCILLATION depends on the fact that a voltage accelerates electrons in the conduction band of a semiconductor but at the same time tilts the band. Electrons are therefore propelled toward the upper edge of the band. In a typical semiconductor (*a*) they never arrive. Instead they emit phonons, that is, they excite thermal vibrations in the crystal lattice and "fall" to a lower energy. In a superlattice, where the conduction band is split into narrow minibands, they do arrive. Specifically, they arrive at the upper edge of the lowest conduction miniband (*b*), and on arrival they are reflected. They may oscillate repeatedly between the miniband's edges before they emit a phonon. Since the miniband is tilted, the emission moves the electron's center of mass a certain distance Δz (*c*). That distance decreases with increasing tilt; thus an increase in voltage can have the curious effect of decreasing the current that flows through the crystal.



DOPING OF A COMPOSITIONAL SUPERLATTICE (the introduction of impurities into the lattice of the crystal) increases the mobility of electrons. The semiconductor with the greater band gap is doped with donor atoms, each of which contributes an electron to the crystal and thereby becomes a positive ion (*plus signs*). The electrons (*minus signs in circles*) then populate the lowest available conduction miniband, which is in the other layers. The resulting spatial separation between the charge carriers and the positive ions reduces their interaction, and so the conductivity of the crystal can be several hundred times greater than it would be in a uniform semiconductor with the same carrier concentration. The bending of the periodic potential wells in the superlattice is caused by the electrons and the ions.

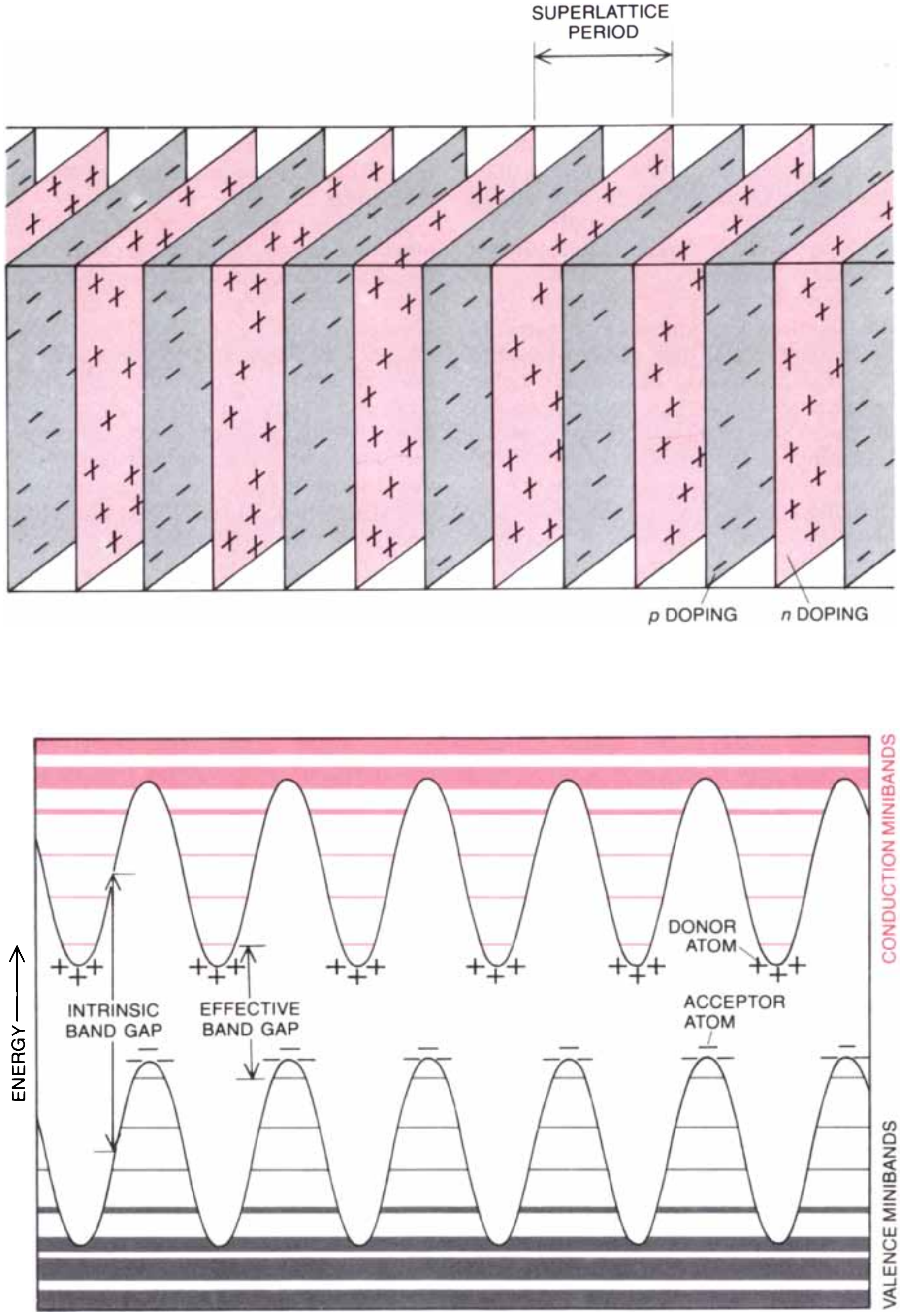
ductor that has the larger band gap [*see lower illustration on opposite page*]. Each donor will contribute a free electron to the crystal, and the electrons will seek the available states that have the lowest energy. These states, however, will be in the layers of the other semiconductor (they will be the lowest conduction minibands there). The electrons will thus gain spatial separation from the donors, and the conductivity of the material may be made several hundred times greater than it would be in a uniform semiconductor with the same concentration of carriers. The trick was first tested experimentally in 1978 by Horst L. Störmer, Dingle, Gossard, Wiegmann and Ralph A. Logan at Bell Laboratories. Once again the experiment confirmed the prediction. Today improved designs and crystal-growing techniques in laboratories in the U.S., Europe and Japan seem to be yielding new records in electron mobility (and new gratification for engineers interested in fast electronic devices) every few weeks.

The more the investigations of compositional superlattices proved successful, the more I grew convinced that another kind of superlattice was also practicable. It would be a doping superlattice, that is, a single bulk semiconductor modulated only by periodic *n* and *p* doping. The problem was to capture the interest of people who might be able to grow such structures. It seemed to me that the technical difficulties facing the growers of a doping superlattice would be no worse than those that had faced the growers of compositional superlattices, and the electronic and optical properties of a doping superlattice, which I had predicted as early as 1972, were certainly no less dramatic than those of a compositional superlattice.

Consider a periodic array of *n* and *p* doped layers, separated, perhaps, by undoped layers of the same semiconductor. (These latter layers are *i* layers, where *i* stands for intrinsic, and so the superlattice is an *n-i-p-i* crystal.) A certain fraction of the donors and acceptors would be ionized, because electrons leaving the donors would recombine with holes leaving the acceptors as long as the process continued to lower the total energy of the material. The resulting charge in the doped layers (positive charge in each *n* layer, negative charge in each *p* layer) would produce a periodic electrostatic potential, and this potential would modulate the conduction and valence bands much as it does in a compositional superlattice.

One great advantage of a doping superlattice is the flexibility with which it can be tailored. For one thing, any semiconductor can be the host material for a doping superlattice, provided only that both *n* and *p* doping are possible. In contrast, the choice of two semiconductors for a compositional superlattice turns out to be strongly constrained by the requirement that the interatomic distances in one of them match at least approximately the distances in the other. Moreover, the effective band gap of a doping superlattice can be given any value from zero to the band gap characteristic of the undoped host material by choosing the appropriate combination of doping concentration and layer thickness. If the doping concentration and layer thickness are sufficiently great, the superlattice becomes a semimetal: a material with a band gap of zero and a residual concentration of free electrons and holes.

Even after a doping superlattice is fabricated it can be tuned. That is, the electronic and optical properties of a doping superlattice can be designed to have certain values and then the values can be modulated within wide limits by a weak excitation of the crystal, say the absorption of a low-intensity light signal



DOPING SUPERLATTICE is a periodic array consisting of layers of the same semiconductor doped in two different ways (*top drawing*). In the *n* layers donor atoms (*plus signs*) contribute electrons; in the *p* layers acceptor atoms (*minus signs*) bind electrons. The resulting distribution of electric charge creates a set of potential wells (*bottom chart*). The value of the effective band gap (the energy difference between the highest valence miniband and the lowest conduction miniband) depends on the choice of dopant concentrations and of layer widths.

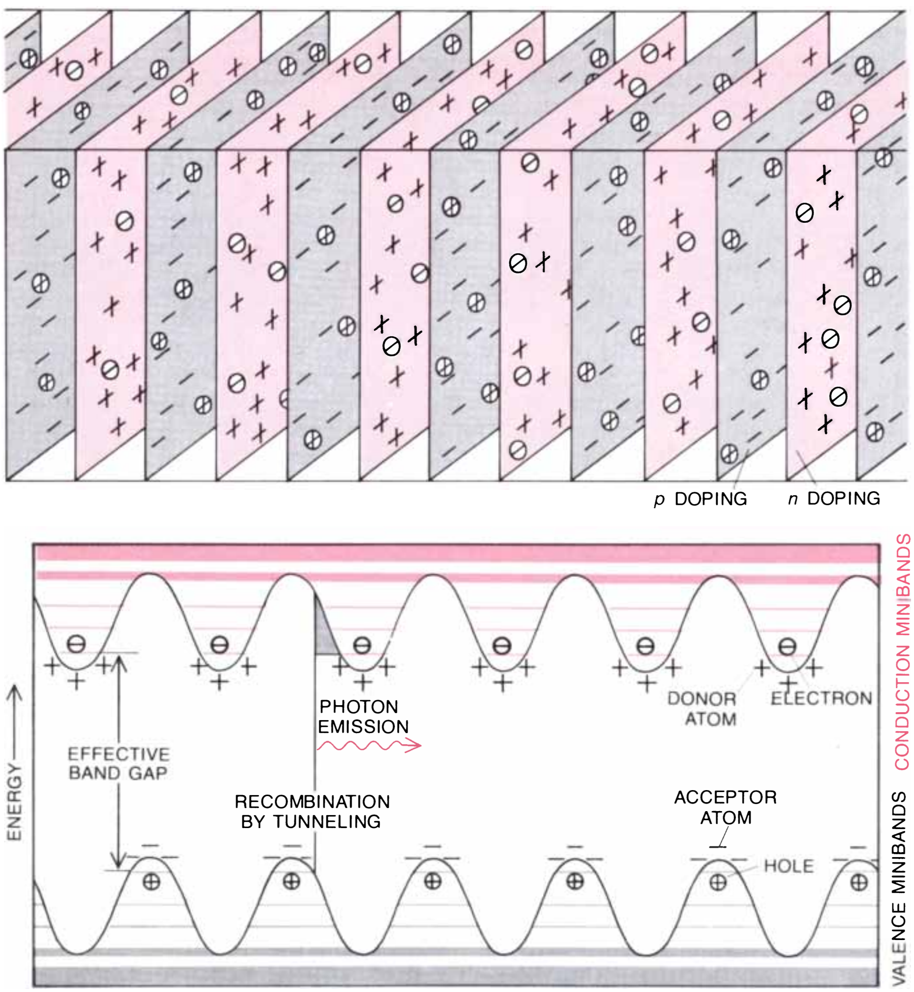
or the application of a small electric current. This tunability, which differentiates doping superlattices from all other semiconductors, arises from the spatial separation between the electrons and the holes that the excitation produces. The electrons are in the n layers; they "tumble downward" into the lowest conduction miniband. The holes are in the p layers; they "ascend" to the highest valence miniband.

The n and p layers are separate; thus the electrons and holes lack partners for recombination. If they obeyed the laws of classical physics, they would never recombine, unless, say, thermal energy in the crystal pushed them to the same location. Recombination by this mechanism is likely only at high temperature. By the laws of quantum physics, however, another means of recombination exists: the particles can "tunnel" through a potential barrier. Still, the probability of recombination by tunneling decreases dramatically with increasing height and width of the barrier. Calculations show, therefore, that doping superlattices can be tailored so that free electrons and holes can have lifetimes ranging from a few nanoseconds (the lifetime of the charge carriers in a uniform semiconductor) up to several hours.

It follows that in a doping superlattice large deviations from the ground state (the state of lowest total energy) can be made to decay very slowly. Indeed, the deviations can be stabilized by simply injecting electrons and holes (hole injection amounts to electron extraction from the valence band) or by causing their generation at a low rate. The concentrations of electrons and holes in a doping superlattice are therefore tunable quantities. The tunability is crucial. In the n layers the electron concentration partly neutralizes the positive charge of the ionized donor atoms. In the p layers the hole concentration partly neutralizes the negative charge of the ionized acceptor atoms. Hence the superlattice potential flattens, and the effective band gap increases.

Perhaps it will not surprise the reader to learn that the electrical conductivity, the optical absorption, the light emission, in fact essentially everything about the electronic and optical properties of a doping superlattice, can be tuned. The electron-hole pairs created by the optical excitation of the crystal become separated with great efficiency. Since recombination is unlikely, their concentrations rise; thus the potential flattens and the effective band gap widens. The flattening increases the likelihood of tunneling; thus the lifetime of the charge carriers decreases. Finally the rate of recombination becomes equal to the rate at which electron-hole pairs arise. The crystal is now in a steady state with its properties tuned to specific values.

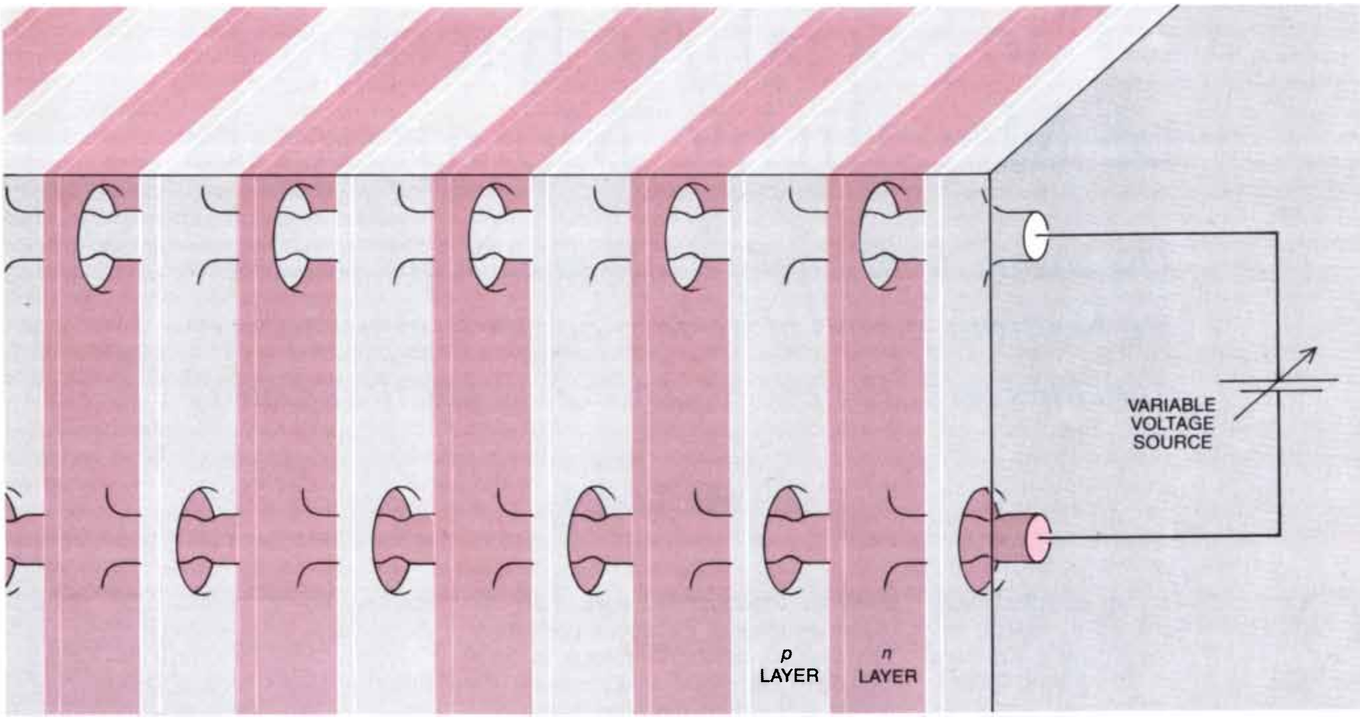
At first the excitation of the crystal by the injection of electrons into only the n layers and holes into only the p layers may seem to be an insoluble problem in microengineering. Actually it can be managed rather easily. A selective electrode to the n layers or the p layers can be a region of strong n or p doping respectively that extends through the superlattice [*see illustration on opposite page*]. At the intersection of such an electrode with a layer of the opposite type of doping the charge carriers one does not want to enter the layer in fact are kept out by electric fields. In the n layers the ionized donors repel the holes that would otherwise enter; in the p layers the ionized acceptors repel the electrons.



EXCITED STATE of a doping superlattice results when electrons (*minus signs in circles*) enter the n layers and holes (*plus signs in circles*) enter the p layers. These charge carriers partially neutralize the charge of the ionized dopant atoms, so that the potential wells flatten and the effective band gap increases. Thus the properties of a doping superlattice can be tuned by current injection or by the optical excitation of electron-hole pairs. The electrons and holes recombine if electrons "tunnel" through a potential barrier (*gray*). The rate of tunneling depends on the width and height of the barrier; hence the lifetime of charge carriers can also be tuned.

The investigator who finally undertook to grow a doping superlattice was Klaus Ploog of the Max Planck Institute for Solid-State Physics in Stuttgart. Thanks to his care and that of his colleagues, the effort, which was made in 1980, succeeded immediately. The host material was gallium arsenide; the dopants were silicon atoms, serving as donors, and beryllium atoms, serving as acceptors. The method of growth was molecular-beam epitaxy.

First the predicted relation between the concentration of charge carriers and the effective band gap was tested by measuring the electron and hole conductivity in the layers of the lattice. H. Künzel, working with Ploog, J. Knecht, A. Fischer and me, found that the conductivity vanished at values of electric potential for which the concentration of carriers should be zero. Next the predicted optical properties were tested. In one experiment, done by Künzel, Ploog and me, the absorption of light by a doping superlattice generated charge carriers with lifetimes of more than 1,000 seconds. This represents a trillionfold improvement over the lifetimes of



SELECTIVE ELECTRODES to the *n* and *p* layers of a doping superlattice are columns of material strongly doped with *n*- and *p*-type impurities respectively. Ionized donors in each *n* layer repel holes in the *p* electrode; ionized acceptors in each *p* layer repel electrons in the *n* electrode. As a result a voltage difference applied across the electrodes sends holes exclusively into *p* layers and electrons exclusively into *n* layers, thereby tuning the crystal. The injection of charge carriers stops when the effective band gap equals the voltage difference.

charge carriers in typical semiconductors. A similar factor of improvement turned out to apply to the superlattice's sensitivity to light. Clearly, doping superlattices are worth examining for their usefulness in light detectors.

The testing of optical properties also enabled us to verify the predicted tunability of doping superlattices. The most impressive proof was provided by experiments with the luminescence of the crystals. Here the investigators I have named, working with H. Jung, D. Olego, W. Rehm and H. J. Stolz, measured the spectrum of the light emitted by a doping superlattice under various intensities of the illumination that served as excitation. At increasing intensities of the incident light, the carrier concentration, the effective band gap and consequently the energy of the photons the superlattice emitted proved to increase by as much as 20 percent. Similar increases have since been observed in doping superlattices excited by the injection of charge carriers. Hence it appears doping superlattices are also worth examining for their usefulness as modulators of light, say in electronic systems that process laser beams.

Of course, the emission of light by a superlattice requires that electrons and holes recombine, which may seem inconsistent with charge-carrier lifetimes on the order of 1,000 seconds. Actually it is not inconsistent. A doping superlattice may indeed exhibit charge-carrier lifetimes in the 1,000-second range when it is near its ground state and the effective band gap is close to zero. (At a band gap of zero recombination can no longer decrease the total energy of the crystal; thus the carrier lifetimes are long.) Under strong illumination, however, the effective band gap approaches the band gap characteristic of the undoped semiconductor. The barrier to tunneling flattens out and the charge-carrier lifetimes shorten.

If one seeks, then, to produce a doping superlattice with high luminescence at photon energies far less than the band gap characteristic of the undoped semiconductor, one will tailor it in the appropriate way. Specifically, one will give it narrow layers and a high degree of doping. The narrow layers entail narrow potential wells, and they will ensure a degree of "transparency" for free electrons in the crystal even if the height of each potential well is great.

The list of the properties of doping superlattices predicted by quantum physics and then verified by experiment must now include the experimental verification of the predicted energy differences between minibands. Investigators at the Max Planck Institute in Stuttgart, including G. Abstreiter and P. Ruden, directed photons into a doping superlattice made from gallium arsenide. They measured the energy of the photons that bounced back out. Some of the scattered photons had excited electrons from populated minibands into higher, empty minibands and so had lost that amount of energy.

Today solid-state superlattices represent one of the fastest-growing fields in solid-state physics. For one thing, the semiconducting materials being made into compositional superlattices now go well beyond gallium arsenide and aluminum gallium arsenide. One of the most intriguing new crystals is surely the compositional superlattice made of layers of indium arsenide (InAs) and gallium antimonide (GaSb), which was proposed and first investigated by George A. Sai-Halasz, Chang, Tsu and Esaki at IBM. Peculiarities in the electronic properties of this superlattice arise from the unusual positions of their bands. It happens that the lower edge of the conduction band in indium arsenide is below the upper edge of the valence band in gallium antimonide. Thus there are energy levels at which electrons and holes coexist, the former in the indium arsenide, the latter in the gallium antimonide. The coexisting carriers mix their characteristics and are becoming known as hoctrons.

Solid-state superlattices are far more than elaborate toys for playful solid-state physicists. In my opinion compositional superlattices, and particularly doping superlattices, with their unique tunability, represent an important new class of semiconductors, one that will have quite an impact not only on solid-state physics but also on future electronic technology.

The Social Archaeology of Megalithic Monuments

The change from simple tombs to elaborate henges in the Neolithic period of western Europe appears to have coincided with the rise of centralized political control

by Colin Renfrew

Over the past two decades many prehistorians have moved beyond the concept of re-creating the past in terms of culture history toward a concept of "process" that seeks to explain past events rather than being content to narrate them. As Kent V. Flannery of the University of Michigan has put it, some New World archaeologists are not "ultimately concerned with 'the Indian behind the artifact' but rather with the system behind both the Indian and the artifact: what other components does the system have, what energy source keeps it going,... and so on?"

The distinction between the two approaches to prehistory is made clear in the Old World by the case of the European megaliths (from the Greek *megas,* "great," and *lithos,* "stone"). These impressive monuments have intrigued students of the past for more than a century. They are found in all the countries of Atlantic Europe, from the Mediterranean seaboard to Sweden, and the human bones many of them contain show that they served as tombs. Some are very simple, although they could not have been easy to build. For example, the monuments called dolmens may consist of three or four large boulders supporting a massive capstone. There are some 50,000 dolmens in Europe, and the labor required to group the boulders, to build up earth ramps and then to move the capstones into position must have been prodigious. Other monuments are far more elaborate than dolmens and must have required man-hours of labor totaling in the hundreds of thousands. Examples include the great passage graves of Newgrange in Ireland and of Maes Howe in the Orkney Islands.

Even the early students of the megalithic monuments of Europe realized that they were pre-Roman and indeed prehistoric. The scale and sophistication of such structures as the passage graves, however, made the same scholars reluctant to believe that the megalithic monuments were the unaided work of the barbarian peoples of European prehistory. The absence of metal objects in them indicated that their makers were the simple farmers of the New Stone Age, but before the development of absolute dating techniques their actual age could only be estimated.

Did the structures represent the influence of more advanced ideas, percolating from areas that supported high cultures, such as Crete and Greece, westward to Spain and then northward along the Atlantic coast? Given such an assumption, the "diffusionist" view that barbarian Europe merely mirrored the civilizations of the Near East was quick to rise. Scholars of the culture-history school even mapped the successive advances of the supposed Mediterranean influences.

The development of carbon-14 dating caused the complete collapse of the diffusionist view. It was soon learned that the megalithic monuments in several areas of Europe were nearly 2,000 years older than any of their supposed Mediterranean predecessors. Some of the earliest of them, in Brittany, could be dated back to 4500 B.C. There could no longer be any doubt that the monuments, the earliest stone structures still standing anywhere in the world, were of local European origin.

The dual problem that then faced the archaeologist was not only to find some explanation for the origins of these monuments but also to give some account of their function that could make the monuments intelligible to us. The various historical reconstructions that were once offered are now seen to have failed. Instead one has to think in terms of process, considering the system behind both the monuments and the societies that built them. Today archaeologists still have only a few clues to why the megalithic monuments rose precisely where they did, and not elsewhere in Europe, but at least one can begin to think about them in social terms.

Although the structures may have had other functions, most of them served as tombs. In some areas, notably in Britain, there are prehistoric stone monuments of other kinds, notably the great stone circles of Stonehenge, Avebury and the Ring of Brogar in the Orkneys. In Brittany great "alignments" of standing stones arrayed in parallel rows three-quarters of a mile long are found at Carnac. Carbon-14 dating and the variety in the form of the monuments in different areas suggest that they may have had five or six quite independent places of origin. One of those places was surely Brittany. Another was in Portugal and perhaps in Spain; others were in Denmark, probably in Ireland and perhaps in southern England. Glyn Daniel of the University of Cambridge has suggested that some of the monuments took their form through the imitation in stone of the wood houses of the local inhabitants and their ancestors. This is in some instances a plausible suggestion, but it does not state precisely why the monuments came into existence.

Faced with such a problem, the archaeologist would do well to ask what the role of these diverse monuments was in the societies of the time. This is in part a question about the actual function of the monuments: How were they used, how did they facilitate the workings of the society? In part it is a further question about the nature of those societies. What kind of societies were they in order for such monuments to have a meaningful place in them? And of course both questions go beyond the issue of "usefulness" in any narrow materialist or "functionalist" sense; it is evident at the outset that these great feats of construction must in many instances have had an enormous symbolic value. They were surely a source of pride to their makers and perhaps of envy to their neighbors. In a social approach to

DITCH SURROUNDING THE RING OF BROGAR, one of two henge monuments on the central Orkney island known as Mainland, was trenched in the course of the author's studies there. It proved to be the result of quarrying in the sandstone bedrock. The labor investment required to complete the late Neolithic ditch, more than 330 meters in circumference, is estimated to have been 80,000 man-hours.

these questions the human qualities of pride and emulation, of solidarity and competition must have a role.

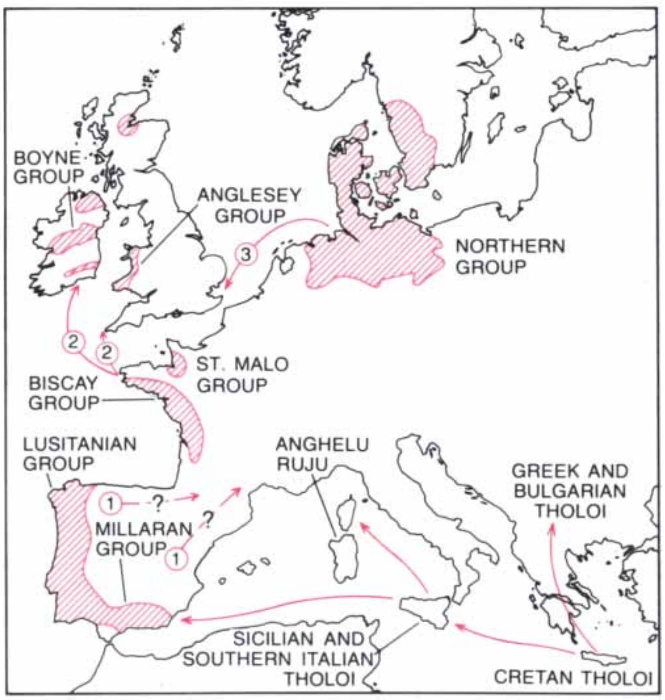
Whatever their general interest, the two questions must in practice be asked first about a specific region and place. The monuments differ from area to area, and so must the specific interpretive approach, whatever regularities may underlie them. One of the first areas to be examined in this way was prehistoric Wessex: those several counties of southern England where Stonehenge, Avebury and many lesser monuments are found. The period is between the first development of farming communities there in about 4000 B.C. and the development of a technology based on the working of bronze in about 2000 B.C., when the building of megalithic monuments had effectively gone out of fashion. Until recently little was known about the domestic settlements of this area. Even today most of our knowledge comes from the considerable profusion of monuments, many of them on the rolling chalklands that were among the first areas to be settled.

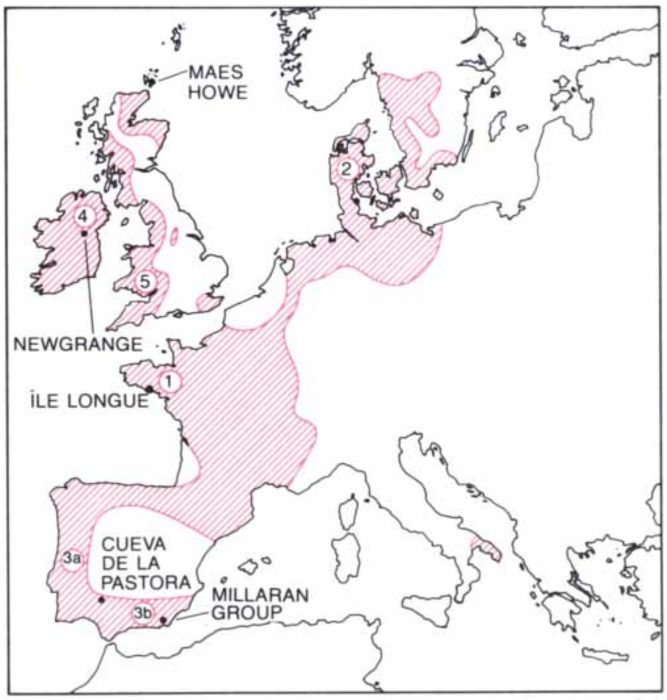
The earliest monuments in Wessex are not in fact megalithic: they have no large stones. They are nonetheless burial monuments. These great oblong mounds, up to 70 meters long, are known as long barrows. Made of chalk, they are in some ways the local equivalent of the stone-built tombs commonly found to the west and, more sparingly, in the Wessex area itself. Because early investigators did not find tomb chambers in them, the structures were at first called unchambered long barrows. Later excavations, however, revealed that most of them did originally contain wood structures that had long since collapsed. Like megalithic chambered tombs, the barrows undoubtedly served as places of burial, but the number of individuals represented by the remains in them is in most instances small. Moreover, the remains are incomplete, suggesting that the dead had been allowed to decompose elsewhere in preparation for secondary burial in the barrow (a custom that is well documented for cultures in many parts of the world).

Grave goods are sparse in the long barrows, as they are in nearly all the European megalithic tombs, and what is there is very simple. The finds are generally restricted to the normal range of Neolithic artifacts: pottery, polished stone axes and chipped-flint tools. Except in the tombs of Spain and Portugal hardly any attempt was made to immure richly decorated objects such as stone carvings or pieces of incised or otherwise decorated stone. It is thus fair to say, both for Wessex and a larger area, that the tombs did not serve for displays of personal wealth, nor do the objects included in them indicate that the dead were of high social status. This contrasts markedly with the finds from the succeeding early Bronze Age in Wessex. In the centuries immediately after 2000 B.C. individual burials were accompanied by rich grave goods, including objects of bronze, amber and gold.

A second class of stoneless Wessex monument is known as a causewayed camp. Most of the camps are roughly circular enclosures about 200 meters in diameter surrounded by concentric rings of ditches. The ditches are not continuous; they are broken by undug areas (the causeways), and it is now clear that the banks thrown up with the chalk excavated from the ditches, rather than the ditches themselves, are what were originally significant. Recent excavations at Hambledon Hill in Dorsetshire by Roger Mercer of the University of Edinburgh have shown the enclosure to have been defended by outlying timber palisades on the approaches to it. Moreover, Mercer found evidence, in the form of human skulls and skeletal fragments in the ditches, that amplified earlier interpretations of these sites as the meeting places of scattered communities from the surrounding areas. Mercer suggests that the enclosure served as a "corpse exposure center, . . . a vast, reeking open cemetery, its silence broken only by the din of crows and ravens."

Both the long barrows and the causewayed camps belong to the earlier part of the British Neolithic, from about 4000 to 3500 B.C. Later, in about 3000 B.C., a different kind of monument made





SUPPOSED DIFFUSION of the megalithic tradition in Atlantic Europe, as mapped in the days before carbon-14 dating, was based on the assumption that the beehive tombs of Crete and Greece, with their corbeled, earth-covered roofs, were older than similar tombs in Italy and Sicily and that the latter in turn had inspired such premegalithic oddities as the rock-cut tombs of Malta and Sardinia. Thereafter diffusion of the megalithic tradition through the Iberian peninsula to western France, northern Germany and parts of Scandinavia and finally to England, Wales, Scotland and Ireland would have followed (*map at left*). With the development of absolute dating, which showed that certain French megalithic monuments had been built in about 4500 B.C. and that even the monuments in the remote Orkney Islands were centuries older than the pyramids of Egypt, the diffusionist view gave way to one emphasizing local origins (*map at right*).

SCIENCE/SCOPE

A Very High Speed Integrated Circuit chip built for the U.S. military uses technology that makes it inherently hardened against radiation. The chip, produced after less than two years of development, draws on complementary metal oxide semiconductor/silicon on sapphire technology. It has circuit dimensions of 1.25 micrometers, or about 50 millionths of an inch. The VHSIC program is being conducted by the Department of Defense to develop chips that will give electronic systems a tenfold increase in signal processing capability. The high-speed, compact VHSIC chips will be more reliable and will require less power than integrated circuits now in use. Hughes Aircraft Company is the only contractor in the tri-service program pursuing CMOS/SOS technology.

Studies have begun to see how an advanced airborne surveillance radar might serve military forces late in this century. The radar would have a large phased-array antenna capable of generating many pencil-shaped beams and would complement the Airborne Warning and Control System (AWACS). One use of the new radar might be to listen in directions other than that of its transmitted beam. If it were to detect another active radar transmitter, the radar could turn its transmitter off (thus foiling an enemy's antiradiation missile) and do its surveillance by using the other radar's transmitted pulse. These concepts are being investigated by Hughes under several study contracts for the U.S. Air Force.

Among many innovations built into the new AMRAAM missile are a special safety mechanism and a high-power coaxial cable. The safety device will prevent the missile from exploding when subjected to fire, yet will not be activated by the high temperatures generated by burning fuel when the missile is launched. The new cable handles much more power than conventional cables and yet costs about one-tenth as much. Hughes designed and developed the Advanced Medium-Range Air-to-Air Missile for the U.S. Air Force and Navy.

The first electro-optical use of a flexible machining system will be for manufacturing large numbers of ultra-precision optical housings. The new Hughes "flex-fab" system is a combination of nine computer-controlled milling machines connected by carts that are pulled on an endless chain towline built into the floor. Each machine has 68 different tools to choose from. Altogether there are 612 tools available, enabling flex-fab to do the work of 25 individual machines. At first, flex-fab will machine aluminum chunks into housings for TOW antitank missile systems with an exactness to one thousandth of an inch. Soon, design engineers will be able to ask flex-fab to build parts, eliminating blueprints.

Hughes Missile Systems Group, located in Canoga Park, California, an attractive suburb of Los Angeles, is seeking engineers and scientists for such developmental and engineering programs as AMRAAM, multimode guidance, Phoenix, and IR Maverick. Openings are in radar and electro-optical systems design, systems software and hardware/software integration, analog and digital circuits design, hybrid process engineering, and systems performance. Qualified applicants are assured prompt replies. Please send resume to Hughes Engineering Employment Manager, Dept. SE, Fallbrook at Roscoe, Canoga Park, CA 91304. Equal opportunity employer.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY

For additional information please write to:
P.O. Box 11205, Marina del Rey, CA 90295

its appearance: the henge, a term derived from the famous Wessex monument Stonehenge. Most henges are not, however, a circle of great stones. They are a circular enclosure surrounded by a single ditch with a well-constructed bank that is usually on the outside of the ditch. There are generally either one or two entrance causeways, and the entire arrangement is much more regular than the earlier causewayed enclosures. Excavations at a number of henge sites (notably those at Durrington Walls undertaken by Geoffrey Wainwright, Inspector of Ancient Monuments for the U.K. Department of the Environment) have shown that some of them included massive circular timber buildings. Henges are sometimes very large, up to 500 meters in diameter.

Finally in this brief review of the Wessex monuments mention must be made of Stonehenge itself, with its circle of stones linked together by lintels, and of the colossal artificial earth mound, Silbury Hill, near a stone circle far larger than Stonehenge: Avebury. Until the time of the Industrial Revolution, Silbury Hill ranked as the largest manmade structure in Europe.

SILBURY HILL, a large artificial earth mound near Avebury in Wiltshire, appears at the center of this aerial photograph. It was the largest manmade structure in Europe until the time of the Industrial Revolution, requiring an estimated investment of 18 million man-hours to build.

In trying to make sense of such a profusion of material two analytical approaches are potentially useful. The first approach takes into account spatial distribution. The second considers absolute scale. Both approaches may give indications of hierarchy, and this in turn may reflect an underlying social ranking within the society itself. For example, when the Wessex monuments are mapped, it becomes clear that the 132 long barrows outnumber the seven causewayed camps by nearly a factor of 20. The map also shows a dispersed distribution of long barrows. They fall into perhaps five groupings, and within each of the groupings is a convenient causewayed camp. (The two camps left over do not fit as well among the long barrows.) Thus it is already possible to speak of a modest spatial hierarchy if one regards each causewayed camp as a local center for the small region defined by each group of long barrows. When one goes on to look at the distribution of the major henges (those more than 200 meters in diameter), a spatial pattern emerges that is similar to the pattern of the causewayed camps. Each causewayed camp, again with two exceptions, is superseded in this later period by a major henge.

The analysis is taken further when we turn from spatial distribution to scale. For example, the long barrows are of rather modest size. Calculation of the amount of labor required for their construction (following estimates based on simple earth-moving methods and the use of very basic tools, such as wood shovels and antler picks) suggests that each would have taken 5,000 to 10,000 man-hours to build. This is an investment of labor with an order of magnitude of from 10^3 to 10^4 man-hours: the amount of work 20 men could do in 50 days. The labor required for the causewayed camps, next in hierarchical rank, is an order of magnitude greater. They would have required between 40,000 and 100,000 man-hours, that is, up to about 10^5 man-hours.

The great henges of the late Neolithic represent a labor investment again an order of magnitude larger, approaching 10^6. For example, Wainwright has estimated a total of 900,000 man-hours for the construction of Durrington Walls. And if the great henges rank as superstars among prehistoric monuments, Wessex provides two megastars: Silbury Hill and Stonehenge. Richard J. C. Atkinson of University College, Cardiff, has estimated the labor requirement for Silbury Hill and Stonehenge (including the labor necessary for the transport of the stones to Stonehenge) at 18 million and 30 million man-hours respectively. This is well in excess of 10^7 man-hours.

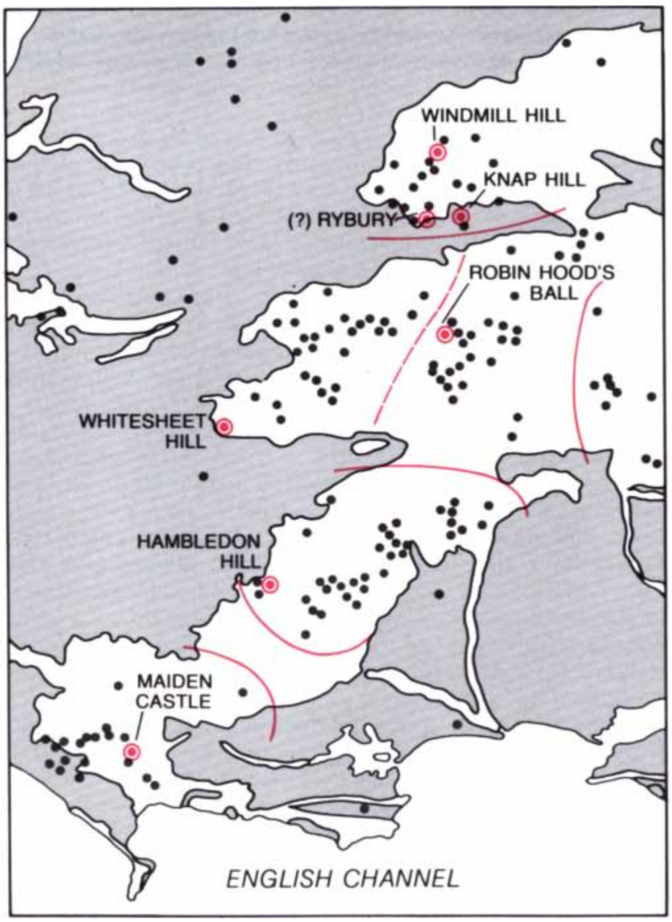
There is thus a man-hour hierarchy to set beside the evident spatial hierarchy. Some years ago I suggested that we should postulate social structures correlating with these space-cum-energetics hierarchies and that, for the late Neolithic, it might be appropriate to talk of "chiefdom" societies centered on the five very large Wessex henges. The approximate territories can be inferred by dividing up the landscape with each of the major henges as the dominant focus of its region. I pointed out that for so great an investment of labor it is necessary to think of some central organizing authority and that it is possible to list a number of relevant features of chiefdom societies as ethnologists recognize them. Fundamental among these features, of course, is the existence of a ranked society (with a chief at its head) sustained by a system based on a redistribution, organized by the chief, of produce, including some of the necessities required for subsistence. Chiefdoms characteristically have a higher population density and more clearly defined territorial boundaries than more egalitarian societies. They have centers that coordinate social and religious activities as well as economic ones, and they show an increase in specialization, as well as the ability to organize and deploy labor for major works.

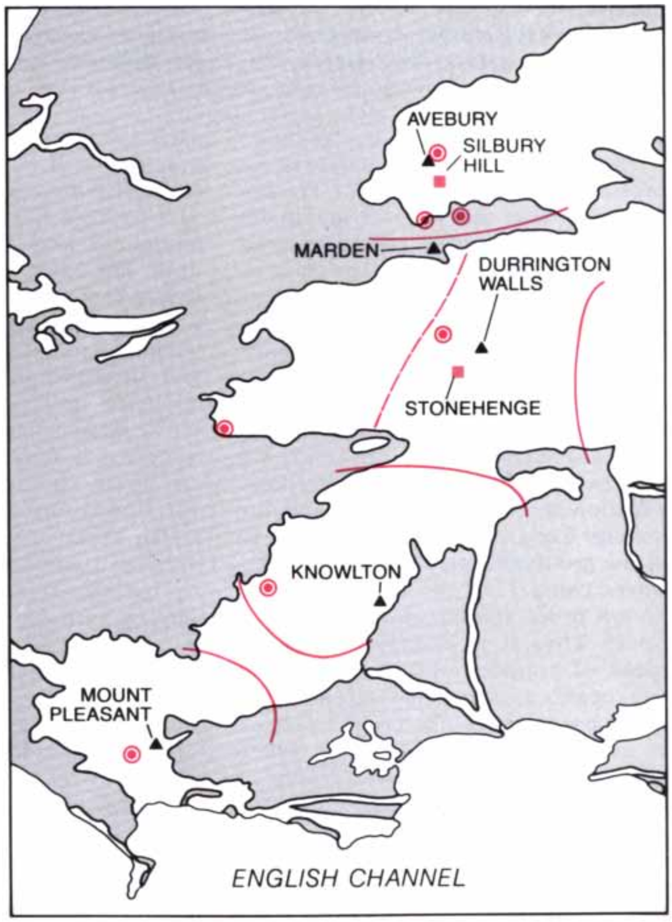
Some of these features might already be inferred for polities with sites like Durrington Walls or Avebury at their center. As the next step upward some further centralization would have arisen: an amalgamation or association of formerly independent chiefdom territories into a federation, perhaps with Stonehenge as its ritual focus and Silbury Hill as its largest monument. On the basis of his excavations at Silbury Hill, Atkinson has suggested that the monument was built within the space of only two years. Even with workers laboring eight hours a day for 300 days a year a full-time team of 3,700 men would have been needed to construct the great tumulus, and each member of the team would have needed to be supplied with food by the organization of his respective chiefdom.

Such a "social archaeology" approach calls for thinking in terms of a well-defined social context for the larger monuments. It situates them within a dynamic development where the simpler societies of the early period developed into the more centralized social units of the later period, which should have had a greater population density and hence a larger population. It is an approach that also allows one to separate the question of the larger monuments (the great henges and Silbury Hill), which we can regard as the product of chiefdom societies, from smaller local monuments such as long barrows. This is important because the large majority of the European megalithic monuments can be more appropriately compared with long barrows. They too are burial monuments, often requiring about 10,000 man-hours for their construction, and many of them are as dispersed as the long barrows. Major monuments are much rarer: henges and stone circles are hardly found on the European mainland (although, at least as far as scale is concerned, the great alignments at Carnac undoubtedly belong in the Stonehenge category).

Concepts of scale and spatial organization are again useful whether one looks more closely at the long barrows of Wessex or at stone monuments of the same antiquity, such as the megalithic tombs on the island of Arran off the west coast of Scotland. In either case if one draws lines midway between the monuments to divide the landscape into separate units (a process geographers call forming Thiessen polygons), one finds that with few exceptions the landscape is divided into roughly equal productive units. For example, on Arran each such polygon contains at least some of the island's limited arable land. Setting aside the instances where the monuments are found in pairs, the pattern is also a dispersed one. If in addition one takes into account the size of the monuments, there is still no sign of a hierarchical structure: at least within one order of magnitude the tombs are approximately the same size. The pattern is not chiefly but egalitarian.

The obvious suggestion arises that these monuments of 5,000 or 6,000 years ago serve as indicators of small, rather scattered groups of farmers living in societies that were not centrally organized. No doubt these groups were linked by marriage and hence by kinship ties (and perhaps also by exchanges of goods), but they may well have been politically independent. Tribal ties between them, initially on a rather modest scale, may be indicated by the next phase: the construction of causewayed enclosures. We may imagine the dead of all the 20 or so long-barrow territories in the Hambledon Hill region, so vividly described by Mercer, being brought





BURIAL MONUMENTS of ancient Wessex (modern Dorsetshire and Wiltshire) in England form clusters that are concentrated largely in the unwooded chalk uplands (*white areas*) that were preferred by the Neolithic farmers who first settled the region. Their burial mounds (*dots*) were long earth barrows and therefore not megalithic monuments at all. When they are grouped into five or six clusters (*colored boundary lines*), however, each of the four lower clusters contains a single causewayed camp (*colored bull's-eyes*), which seems to have been an area where the dead were first exposed before their bones were finally interred in the associated group of long barrows (*map at left*). The grave goods found in the long barrows are few and simple, suggesting that the farmers formed egalitarian social groups. When, about 500 years later, new kinds of circular henge earthworks that may all have included timber structures within their bounds were built in this same part of England, several of the larger ones (*triangles on map at right*) lay within the same bounds of clustered long barrows as the earlier causewayed camps did. The henge monuments, which rarely include megaliths, are substantially more elaborate than the earlier causewayed camps. This suggests that the social organization they served had become more hierarchical. Further evidence to the same effect is the appearance of two "megastar" monuments in the region (*rectangles*): Stonehenge and, to the north, Silbury Hill.

there for exposure and the bones of some of them being later taken back to the local long barrow for interment. Since the long-barrow territories are about 10 square kilometers in area, taking a modest figure of between 10 and 50 hectares of land to support each person in a simple farming economy, we may think in terms of populations of between 20 and 100 people per territory. These small units may be regarded as "segmentary societies," to use the social anthropologist's term, each segment being a group of people, an economically and politically autonomous, self-sustaining perpetual body exercising effective control over its productive resources.

From this perspective we can begin to think more clearly about the far more ambitious megalithic chambered tombs. They were, of course, burial places. Indeed, with the possible exception of the timber-chambered long barrows, they were collective burial places. But a group of perhaps 50 people, even a permanent group occupying the same piece of land for centuries, does not actually need a monumental tomb to dispose of its dead. There would be nothing simpler than to bury them quickly, without any fuss. The monumentality of the chambered tombs, therefore, is not logically implied by their funerary function. It hints at another purpose altogether. In the words of Andrew Fleming of the University of Sheffield, these were "tombs for the living," that is to say, their monumental character is the result of the deliberate activities of the living members of the society and must have served the objectives of that society. Three different explanations can be put forward for the role of the chambered tombs. The three are not contradictory but serve rather to complement one another.

First, with respect to these ambitious monuments, I have suggested a function in terms of what may be called social cohesion. In times of stress, perhaps particularly in circumstances of high population pressure, a community could function more effectively when its social unity was emphasized and felt by all. The building of such an enduring monument was a major symbolic act asserting just this unity. It should be remembered that these people lived in a world where other obvious lasting human creations were few: even the houses were usually made of wood and other perishable materials. Once the monument was built, however, it would be the major manmade landmark of the territory. Sometimes there must even have been competition between neighboring territories to have the largest and most conspicuous center.

I argue, then, that we should view these monuments not merely as tombs but as public centers. Often they might have served as meeting places: perhaps the locus for an entire range of religious rites relating the community as a whole to its ancestors as well as to its more recent dead. The building of a focal monument could have given visible expression to the communal aspirations of the local group, serving to ensure the group's continuing cohesion and hence its survival.

Second is a closely related idea developed by Arthur A. Saxe of Ohio University, who has pointed out the relevance of such funerary activities to a community's maintenance of its right to ancestral lands. As he puts it: "To the degree that corporate group rights to use and/or control crucial but restricted resources are attained and/or legitimized by means of linear descent from the dead (i.e., lineal ties to ancestors), such groups will maintain formal disposal areas for the exclusive disposal of their dead." Title to communal land is thus asserted by the maintenance of the ancestral tomb. Both views may be summarized as saying that the monuments served as the territorial markers of segmentary societies.

Third, Christopher Tilley of the University of Lund has taken a neo-Marxist approach to suggest a further function of the monuments, one that does not contradict the two preceding ideas. Tilley stresses the role of mortuary ritual, in common with other life-cycle rituals, in "legitimizing sectional interests" within the society. Even within the relatively egalitarian framework of a segmentary, lineage-organized society, he suggests, there are class contradictions, notably between the elders of the group and the junior members. He argues that control over ritual, emphasized by such monuments, helped to facilitate the continuing dominance of the elders. I am not sure that such an argument takes us much further than seeing the monuments as symbolizing and facilitating social cohesion, as has already been suggested, but Tilley certainly sets this idea in a different framework.

Each of the three explanations is functionalist. That is to say no one of them effectively explains the origin of megaliths but each shows how the monuments played a useful and meaningful role in the societies that erected them. This point of meaningfulness must be at least one component of any satisfactory explanation for them, and these are certainly explanations in social terms.

Such ideas, of course, are difficult to test directly in the way that one would wish to test a generalization or a hypothesis in the exact sciences. They can nonetheless be examined afresh in the light of new work in the field. It was with this intention that between 1972 and 1974 I undertook field work in the Orkney Islands at the northernmost tip of the British Isles.

The Orkneys are virtually without trees and have been since the last ice age because of the prevailing strong winds. The local rock is an easily fractured laminar sandstone, so that the prehistoric population had readily available an abundant building resource other than timber. All Orkney buildings were therefore built of sandstone, and they have survived particularly well. Until very recently farming practices on the islands were not mechanized, and so the cairns of stone that are today the indicators of buried prehistoric remains have not been demolished.

The prehistoric antiquities of the Orkneys have long been famous, notably two henge monuments with circles of standing stones, the settlement at Skara Brae and the splendid passage grave of Maes Howe. Skara Brae, excavated in the 1930's by V. Gordon Childe, is a magnificently preserved village from the time of the island's early farmers (about 3000 B.C.). The walls of the houses still stand to a height of up to two meters, and some internal furnishings such as beds and cupboards, being also built of stone, have survived. Maes Howe has an impressive stone-built chamber entered by a passageway. The stonework is of such quality, the stones having been carefully fractured, that only with difficulty can one accept that no metal tools were used in its construction. On the old diffusionist theory for the origin of the megaliths, Maes Howe, being the finest monument in the Orkneys, was also considered to be the oldest: the first inspired effort of the supposed new immigrants as they landed with their civilized ideas still unimpaired by decades of living in the remote north.

At the time I began work, no carbon-14 dates at all were available from the Orkneys, and one of my aims was to obtain datable samples of organic material relevant to the chronology of the Orkneys chambered tombs. The main site chosen for excavation was the tomb of Quanterness. From a report written early in the 19th century I knew that Quanterness had been entered at that time and had then been closed. My colleagues and I hoped to find evidence for its date of construction and by conducting further studies to contribute to the understanding of the islands' monuments in general. At about the same time fresh excavations directed by David Clarke of the National Museum of Antiquities of Scotland were beginning at Skara Brae, and Graham Ritchie of the Royal Commissions for Ancient and Historical Monuments of Scotland was starting work at one of the Orkneys henge monuments, the Stones of Stenness.

The roof of Quanterness had collapsed, and we had to enter the chamber from above. As the 19th-century sketch plan indicated, there were six side chambers leading off the rectangular main chamber, with an entrance passageway from the east, now blocked. The early account had not, however, prepared us for the handsome, regular stonework, which makes this one of the most impressive buildings of its period in existence anywhere. Nor had we expected the tangled profusion of human and animal bones we encountered once the debris had been removed from the upper levels. The floor of the main chamber was covered with disarticulated human remains, strewn around in great confusion, together with fragments of pottery and other objects. It was necessary to use a water-sieving technique on all the material from the tomb in order to extract small fragments of artifacts and of bone, including the bones of fishes and birds.

The clearance of these main-chamber finds enabled us to enter the side chambers of the tomb. Four of the six were intact. Once the connecting passages were cleared it was evident that these side chambers were not blocked; it was possible to bend down, crawl in and stand upright in a perfectly complete stone chamber, beautifully roofed, that had existed for more than 4,000 years. Carbon-14 dating later revealed that the tomb had been in service from about 3200 B.C., and so the structure was fully 5,000 years old: older by some centuries than the oldest pyramids of Egypt.

There were few grave goods in the tomb. We found the fragmentary remains of some 34 pots, only three polished stone knives and an occasional piece of chipped flint. The animal bones, clearly the remains of joints of meat that had been brought into the tomb either as offerings to the dead or as food for consumption during funeral rituals, represented parts of a minimum of seven sheep, 18 lambs and five oxen. Also present were the bones of red deer and other wild mammals, of at least 35 birds of various species and of seven species of fishes.

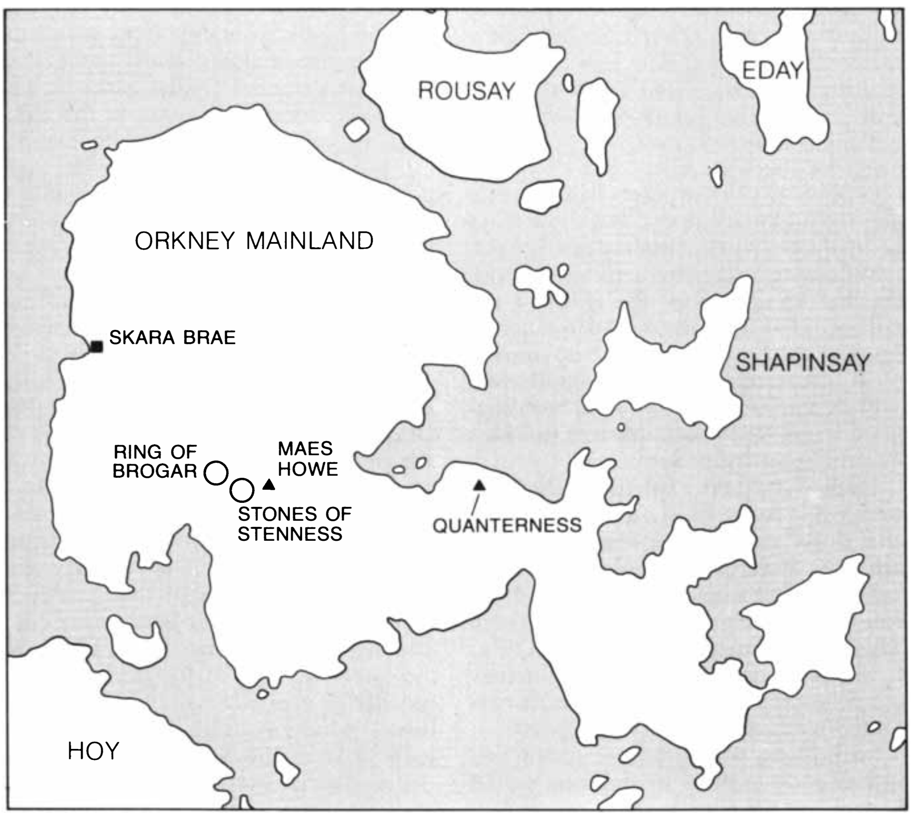
Analysis of the very fragmentary human bones by Judson Chesterman of the University of Sheffield produced surprising results. From the sample of 12,600 pieces we recovered he was able to estimate that the partially excavated main chamber of the tomb and the one side chamber we fully examined held parts of at least 157 individuals. As in the Wessex long barrows, the dead had clearly been exposed before their bones were collected for final burial in the tomb. If this sample figure is increased by a suitable factor to allow for the 20 percent of the main chamber and the other five side chambers that remain unexcavated, Quanterness probably holds the remains of no fewer than 400 individuals.

The ratio of males to females (54 : 46) is sufficiently close to equality to suggest no exclusion from the tomb on the grounds of sex. With the exception of infants all age groups are represented. The age distribution shows that the mean life expectancy was between 20 and 25 years. Although such a life span is short, it is not out of keeping with finds from other Neolithic sites, and it suggests that what we uncovered may have been the dead from an entire living community. From the carbon-14 estimate of how long the tomb was in service, about 550 years, we can place the size of that community at between 13 and 20 individuals.

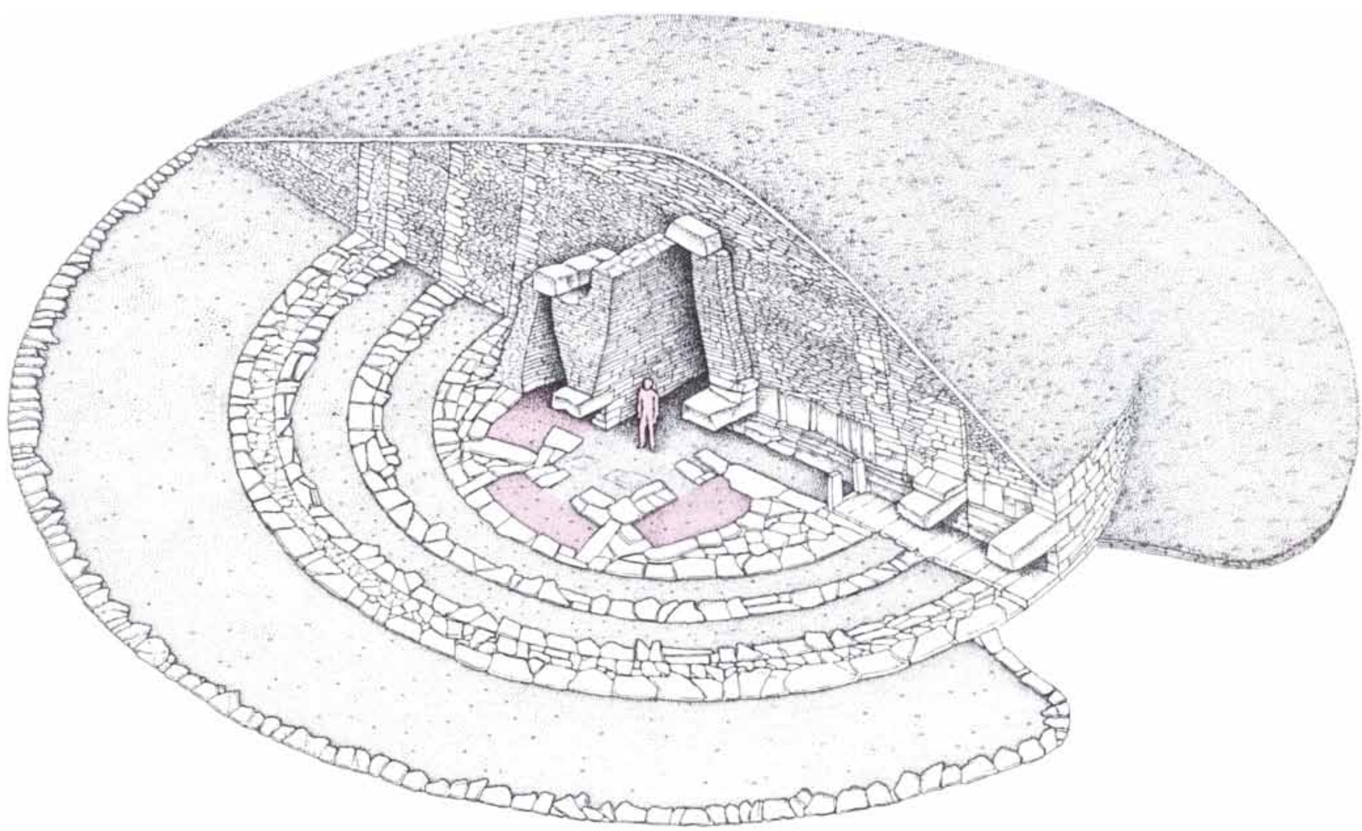
While excavations were continuing at Quanterness we were able to examine some other Orkney sites. Calculation showed that Quanterness represented a labor input of about 7,000 man-hours for the tomb and its enclosing cairn, although this figure takes no account of the skill of the builders. It is clear that most of the other Orkney chambered tombs, which have a notably dispersed distribution, represent labor inputs of the same general order of magnitude.

Our work at the other henge, the Ring of Brogar, showed that the ditch surrounding the henge was a product not of earth-moving but of splendid rock-cutting, representing a labor investment of about 80,000 man-hours (not taking into account the erection of the standing stones of the circle). Unfortunately we were not able to date the monument satisfactorily. The best estimate of its age comes from the carbon-14 date, about 3000 B.C., obtained by Ritchie for its smaller sister monument, the Stones of Stenness. We were able to obtain dates from the ditch surrounding the great tomb, Maes Howe, which suggest that it was built in about 2800 B.C. Because the construction stone had to be brought from a considerable distance, Maes Howe represents a labor input more like that of the Ring of Brogar than that of Quanterness.

These various results allow us to draw conclusions of a social nature about early Orkney that can be compared with the more general social-archeological picture outlined for Wessex. First, at the time of its early use the tomb at Quanterness was not stratified spatially above or below monuments built on a different scale. There simply was no such spatial hierarchy. Second, Quanterness was an "equal access" tomb, containing a balanced representation of both sexes and of all ages except for infants. Third, the



FOUR MEGALITHIC MONUMENTS, two henges and two chambered tombs, are among the eight found on the largest of the Orkney Islands. The author and his colleagues reopened the chambered tomb Quanterness, built in about 3200 B.C., for the first time in nearly two centuries and partially excavated it. They also trenched the perimeter of the Ring of Brogar, a henge monument probably put up in about 3000 B.C., perhaps two centuries before the more elaborate chambered tomb of Maes Howe was built. The simplicity of Quanterness, compared with the other structures, suggests that the farmers who built it formed one small community.



QUANTERNESS TOMB is shown in this cutaway drawing as its northern half probably appeared soon after construction. The human figure at the center (*color*) gives the scale; the outer stone perimeter is 45 meters in circumference and the tomb's height at the center was originally some 3.5 meters. Three of the six side chambers can be seen in the cutaway (*light color*); the author and his colleagues excavated one of the side chambers completely and partially cleared the main chamber. The estimated total number of burials is 400.

grave goods were very simple, indicating no prominent ranking among the occupants of the tomb. This finding, along with the equal-access one, supports the view that the people belonged to a relatively egalitarian society. Fourth, burial practices were elaborate: inhumation of the bones within the cairn was the last stage in the treatment of the deceased, having been preceded by the decomposition of the corpse. Finally, the chamber was in use for at least five centuries.

Let us now consider still other facts that we learned bearing on the nature and sequence of events at the site. Item: The group that used the cairn probably numbered no more than 20, including men, women and children. Item: The labor required for the construction of the monument, less than 10,000 man-hours, could without difficulty have been invested by such a group over the span of a few years, or perhaps over an even shorter time if the builders had the assistance of neighboring groups. Item: Quanterness is just one of a number of similar Orkney burial cairns of comparable scale. Item: The distribution of these cairns is fairly dispersed, suggesting that the group making use of each cairn occupied and farmed the surrounding territory. And lastly, some centuries after the construction of Quanterness a few monuments built on a far larger scale, including the Ring of Brogar and Maes Howe, were raised in a single central area of the same main island of the Orkneys. The new construction can be taken to imply the emergence of some form of more centralized social organization.

The finds from Quanterness and nearby sites thus lead to, from excavational data, conclusions that harmonize well with those already arrived at Wessex. Whereas the time elapsed between egalitarian and centralized societies is not long enough in the Orkneys to demonstrate this sequence of social development conclusively, it does seem clear that the raising of the major monuments there, indicative of a greater degree of centrality, was relatively late in emerging and that most of the cairns belong to an earlier, more egalitarian phase.

The abundant skeletal remains from Quanterness help to confirm the view that these cairns were not built just to house one or two wandering missionaries. Quite simply they were erected by the local inhabitants: a relatively dispersed farming population. The considerable sophistication of their architecture and their skill in construction both point to the great pride taken by the parent community in its monument.

It is not argued here that the monuments of the Orkneys had an origin independent of the other areas where collective burial in monumental tombs was practiced. I suggested above that there may have been five or six primary megalithic centers in western Europe, including one in Ireland and perhaps still another in southern England. The spread of farming techniques to Scotland is likely to have been from one or another of these areas, and the colonists probably brought the custom of tomb-building with them. The first inhabitants of the Orkneys in turn must have crossed to the islands from northern Scotland and were no doubt familiar with the rather simple megalithic tombs constructed in the north at the time. What was special about the Orkneys was the impressive development of a local architecture, based on the excellent local sandstone, to give such masterpieces of prehistoric architecture as Quanterness and Maes Howe. And the particular interest of the Orkneys to the archaeologist is the great abundance of prehistoric remains, which makes the attempt to think coherently about these considerable achievements in social terms less difficult.

The evidence from the Orkneys adds significantly to the picture one gets from Wessex. It is possible to suggest what the function of these monuments may have been, at least in some of the societies that built them, but if one is now to go on to explain megalithic origins, one must also indicate why they came about when they did and where they did.

The "when" is not so difficult. These were the monuments constructed by the first farmers in the various regions of Atlantic Europe soon after they had settled down and established a stable way of life in the lands along the Atlantic coast. No monumental tombs of this scale were built by the earlier and much sparser population of Mesolithic hunters, gatherers and fishermen who exploited some of these same coasts before the development of farming.

The "where" is less easy. If the earliest farmers everywhere in Europe had developed similar monuments, there would be less of a problem. Why did the monuments go up in these particular coastal lands and not in either central or eastern Europe? To answer the question one has to isolate some factor special to the Atlantic areas and to show why that factor is relevant.

One such factor is the existence of the sparse but settled Mesolithic population in areas along the Atlantic coast, notably in Portugal and Brittany. We know that these people were already burying their dead systematically in the middens of discarded shells they left behind them. No one has yet explained, however, exactly why the existence of this earlier population should make it more likely that the early farmers would build such monuments.

I propose a different factor and shall do so in demographic terms. It has been suggested that the spread of the early farming economy caused a substantial rate of population increase in every area as it emanated from Greece and the western Mediterranean, until a ceiling was reached at what was for the time a rather high population density. In each instance the approach to this ceiling would necessitate a decline in the rate of increase. Such a decline could not be brought about without some difficulty. In central Europe there was always the possibility of shedding some of the surplus population by means of emigration toward the west and the north, so that the stress of approaching the population ceiling would have been alleviated. In the lands bordering the Atlantic, however, this safety valve was not available. There was simply nowhere to emigrate to, and the demographic consequences had to be faced as best they might.

It is precisely in these circumstances, I would argue, when the population density had become inconveniently high and a need to curb the rate of population increase was evident, that devices favoring good social cohesion within small farming groups and ways of asserting one's right to farming land would be most useful. Communities that had put up monumental territorial markers would have had an adaptive advantage over less assertive and less cohesive neighboring groups. The seeds of these ideas may already have been present in the burial practices of the earlier, midden-depositing population. In any event the unprecedentedly high rate of population increase, impossible to mitigate by an outflow of emigrants, would make useful any devices favoring both stability and social order.

My proposed explanation is only a sketch. It needs further exploration by means of further excavation. For example, an examination of settlements where the earliest farmers may have come into contact with the last of the midden people would be useful. More work and more thought will be necessary before one can maintain either that

JUMBLED BONES, both human and animal, were found on the floor of the main chamber of the tomb of Quanterness. Grave goods were scarce: a few pieces of chipped flint, three polished stone knives and fragments of some 34 pots. Construction of the tomb is estimated to have required some 7,000 man-hours, or about the same labor investment as that for a long barrow.

the problem has been correctly framed or that the mechanism for its satisfactory solution has been found.

If the picture I have sketched can be accepted as plausible, can it be any more than that? Can it be tested with anything like the rigor that one would wish of a scientific theory? The spatial patterns can certainly be tested in principle with the modern geographer's techniques of locational analysis. The real terrain, however, does not usually approximate the level, "isotropic" plain the geographer often assumes, and the ravages of time have caused the disappearance of many monuments. Moreover, it is often difficult to show that the different sites were in use at the same time, as a strict quantitative analysis would require.

Further, it is clear that there can be more than one motive for monument-building, and even in prehistoric times there are notable concentrations of tombs, not dispersed but clustered in cemeteries. Some of the arguments used here regarding dispersed distributions would not apply to the clustered grave sites. In spite of these limitations the archaeologist need not lose heart. Social archaeology, after all, inevitably labors under the same difficulties as the other social sciences. It is dealing with the activities of human beings, which under the best of circumstances are not easy to quantify and are very difficult to predict. That we can deal at all with social issues in a remote past, datable only with the aid of carbon-14, when half the evidence is forever gone, is remarkable enough.

The French historian Fernand Braudel has differentiated three levels at which the processes and events, the "conjunctures" of history, work themselves out. The personal dramas, the "headlines" of the past, are the short-term crises. Underlying these are the economic and social movements that operate over the medium term. Underlying these again is the secular trend, the *longue durée*, where much larger time factors operate. Among them are the almost constant constraints of the landscape and the basic realities of peasant life. In a curious way the archaeologist, who usually can have little to say about the single individual or the short-term event, can sometimes get much closer to the long-term processes than can the modern historian, overwhelmed by the noisy profusion of his evidence for the recent past. That is why archaeologists working on topics of cross-cultural significance, such as the origins of agriculture, the processes of state formation or (as in the present case) the appropriate way of explaining, in a general social perspective, the building of the world's first monuments, feel that in spite of their practical limitations, they have ideas worth discussing.

A Seed-eating Beetle's Adaptations to a Poisonous Seed

The beetle larvae feed on a seed containing an amino acid that is severely toxic to other insects. The larvae circumvent the toxic effect of the amino acid and also utilize its nitrogen

by Gerald A. Rosenthal

Most insects depend on plants for their food, and so insects are among the most aggressive and destructive adversaries of plants. Terrestrial plants characteristically lack mobility and cannot elude these destructive pests by flight. Yet plants are far from passive participants in the struggle that has gone on between insects and modern plants for more than 50 million years. The primary weapons of plants against the feeding ravages of insects are the plants' own chemical constituents. The thousands of distinct compounds higher plants synthesize in their metabolism are classified either as primary metabolites or as secondary metabolites. Primary metabolites are substances such as nucleic acids, adenosine triphosphate (ATP) and glucose that are common to all life. Secondary metabolites vary among plants, but plants would be unable to wage chemical warfare without them.

Many biologists believe the feeding activity of insects and other herbivores and the ravages of a host of pathogenic organisms have provided the selection pressure for the elaboration and maintenance of these sophisticated chemical barriers to predation and disease. In such a situation natural selection would have capitalized on the inherent variability in secondary plant chemistry to favor the metabolites that can enhance the organism's Darwinian fitness by affording an effective level of protection. Indeed, in considering the relation between certain insects and plants, Paul P. Feeny of Cornell University has said: "We are thus witnessing an evolutionary arms race in which the plants, for survival, must deploy a fraction of their metabolic budgets on defense (physical as well as chemical) and the insects must devote a portion of their assimilated energy and nutrients to various devices for host location and attack."

The point is aptly illustrated by the plant *Dioclea megacarpa,* a vinelike legume. It has only one insect predator: *Caryedes brasiliensis,* a small beetle of the family Bruchidae, seed-eating beetles of worldwide distribution. One reason the plant is so remarkably successful in repelling other insects is that it stores L-canavanine, an insecticidal amino acid that among other things disrupts the production of normal insect proteins. What is even more remarkable is the adaptation of *C. brasiliensis* to the chemical defense of the plant.

Our work at the University of Kentucky has yielded some insight into the interactions between insects and the secondary metabolites of higher plants. Canavanine is one of about 250 amino acids that are synthesized by higher plants but are not utilized as building blocks of proteins. This nonprotein amino acid is synthesized by members of the Lotoideae (Fabaceae), a group of leguminous plants, and is found in such agronomically important crops as alfalfa and numerous trefoils, including clover. It is also synthesized by annual and perennial ornamental and arborescent species such as *Wisteria* and *Robinia.* There is considerable evidence that canavanine functions as an important nitrogen-storing metabolite, particularly in the seed, where it supports the growth of the newly developing plant.

The molecule L-canavanine is a structural analogue of L-arginine, one of the 20 protein-forming amino acids. In canavanine the terminal $-CH_2$ group of arginine is replaced by oxygen. This difference is of little consequence in the metabolism of canavanine, which therefore serves in virtually every enzyme-controlled reaction for which arginine is the preferred substrate. For example, canavanine is activated by arginyl transfer RNA synthetase, an enzyme that activates arginine and then links it to its transfer RNA. Once canavanine has been linked to the transfer RNA, which normally transports arginine to the protein-assembly sites on the ribosomes of the cell, canavanine is inevitably inserted into the growing protein chain in place of arginine.

Canavanine is chemically much less basic than arginine, and under physiological conditions it would be less positively charged than arginine. This difference in charge can affect the interactions that cause the protein chain to fold up into the uniquely correct conformation of a given protein molecule. There is increasing evidence that proteins structurally altered by the incorporation of canavanine instead of arginine do not function properly. Herein lies an important toxic effect of canavanine on most insects.

My colleague Douglas L. Dahlman and I have established experimentally the potent insecticidal properties of canavanine. In our work we employed the tobacco hornworm, *Manduca sexta,* a lepidopteran consumer of plants such as tobacco and the tomato, none of which store canavanine. Hence *M. sexta* has lacked the opportunity to adapt to this toxic amino acid and nullify it. When we injected canavanine labeled with radioactive carbon (C-14) into the hemolymph, or circulatory fluid, of the tobacco hornworm, we found that at least 3.5 percent of the labeled canavanine was incorporated into newly formed insect proteins.

Providing the fat body, an internal organ, of the migratory locust *Locusta migratoria migratorioides* with canavanine results in its incorporation into vitellogenin, a major protein of the fat body that is critically important in egg development. The canavanine-containing vitellogenin of the locust exhibits a higher mobility in an electric field than the natural protein does. This finding reveals a significant change in the physicochemical properties of the protein resulting from the incorporation of canavanine. The canavanine is assimilated exclusively at the expense of the protein's content of arginine.

In 1973 I learned of the work of Daniel H. Janzen, a tropical ecologist who is now at the University of Pennsylvania. Part of Janzen's work in Costa Rica involved studies of the interaction of the bruchid beetle *Caryedes brasiliensis* and the vinelike legume *Dioclea megacarpa*. My interest in this particular insect-plant interaction developed from Janzen's report that the seeds of the plant contain a considerable amount of canavanine. (Later I determined that canavanine can account for as much as 13 percent of the dry weight of the seed, exclusive of the seed coat, and for as much as 55 percent of all the nitrogen in the seed.) The accumulation of so much canavanine provides a highly effective chemical barrier to predation: the only known insect consumers of the seeds are the larvae of *C. brasiliensis.*

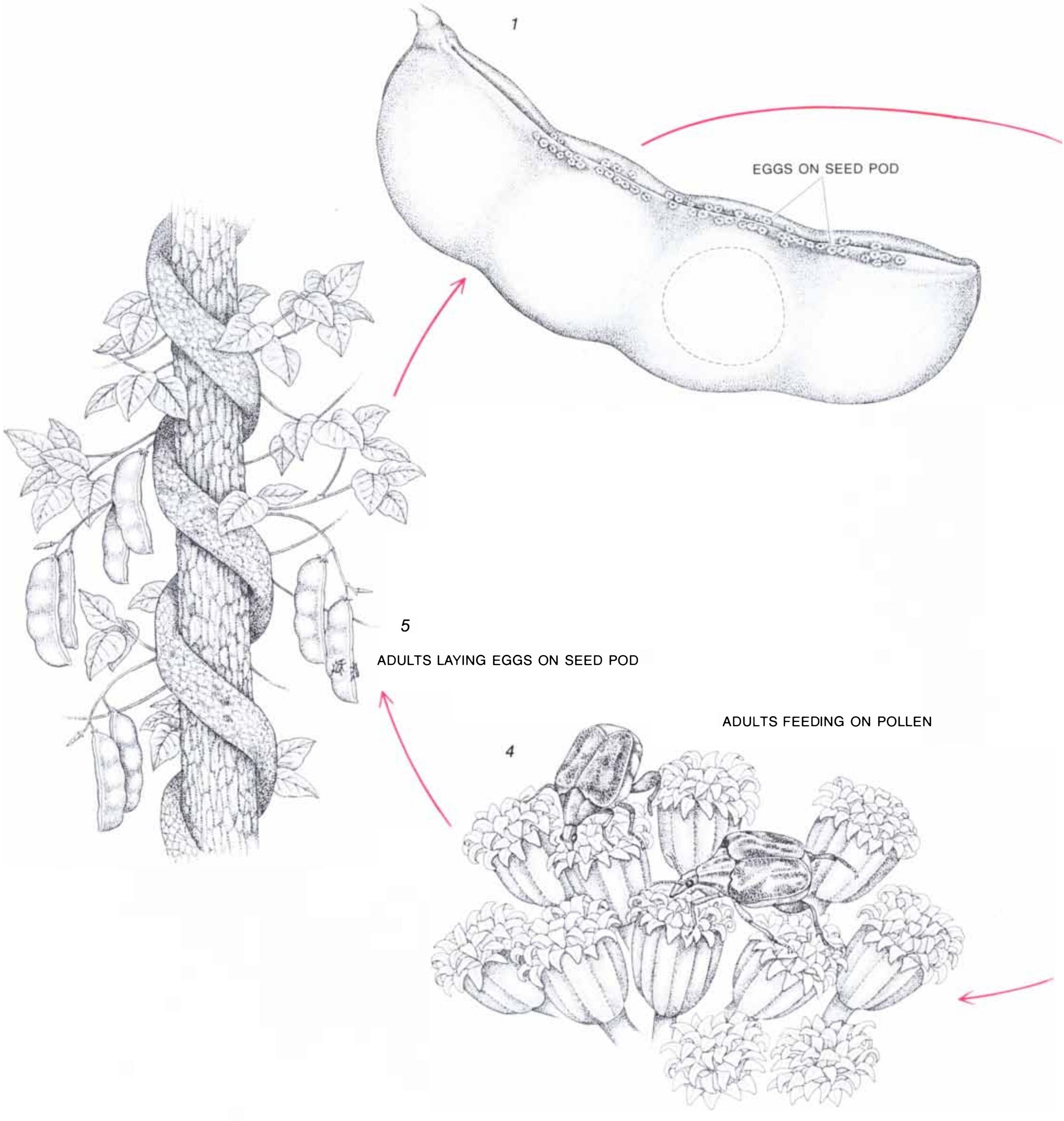
The female beetle lays her eggs on the wall of the newly ripe fruit of the plant, often near the fissure that is characteristic of legumes. The eggs are protected in structures termed oothecae. The time of laying is late fall, at the end of the rainy season and about a month before the pods mature. *D. megacarpa* pods are suitable for the deposition of eggs for only a few months. Bruchid beetles do not lay their eggs on mature or rotting pods, and if the females miss the right period, the beetle population must await the end of the next rainy season before the life cycle is completed with the deposition of eggs on the next crop of pods.

The newly hatched larva bores through the wall of the fruit and the seed coat and makes a small chamber in the storage tissues of the seed. The chamber is the larval home for the several months required for the larva to grow, pupate and emerge as an adult. As many as 50 of the beetles can develop in a single

ADULT BEETLE of the family Bruchidae, the seed beetles, emerges from a seed of the legume *Dioclea megacarpa* after spending several months in it during the larval and pupal stages. The beetle (*Caryedes brasiliensis*) has cut a hole in the coat of the seed in order to get out. The plant makes the amino acid canavanine, which is toxic to other insects. *C. brasiliensis* is its only insect predator. Through a series of biochemical adaptations the larvae (as many as 50 of them in a single seed) avoid the toxic effect and utilize the nitrogen of canavanine.

seed. The young beetle carves an exit portal in the seed coat and escapes ultimately through the suture along the pod or through an opening created by a vertebrate that tears apart the pod to gain access to the seeds. The adult beetle, released from April through June, feeds solely on pollen and does not ingest canavanine.

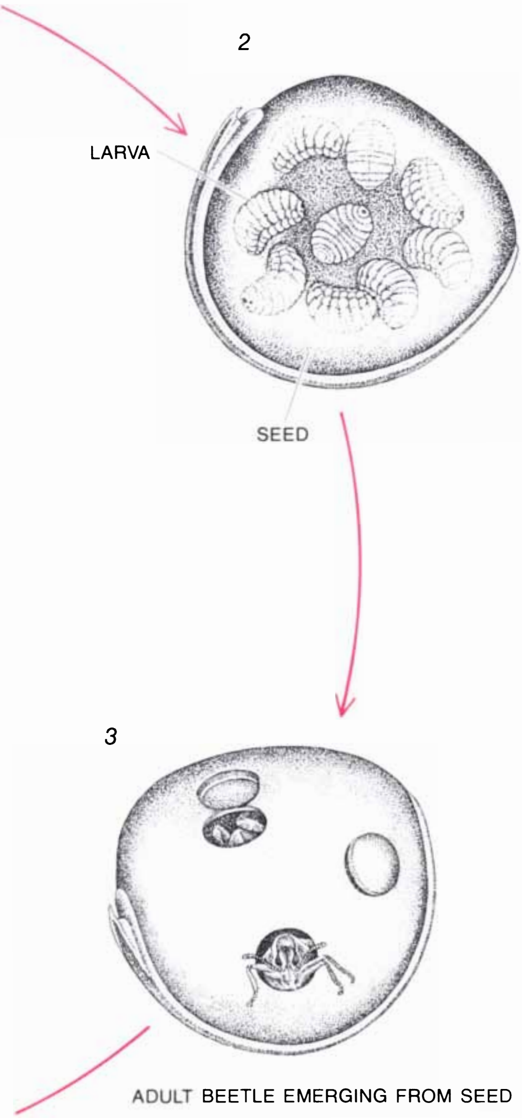
This seed predator is an ideal system for investigating questions of how insects adapt to the toxicants produced by plants. The ability of the beetle to pursue its specialized feeding habit so successfully results from biochemical adaptations that enable it to cope with the canavanine, which would otherwise be highly poisonous. As part of a research program in which I was sponsored by the National Science Foundation I undertook a long-term, systematic investi-



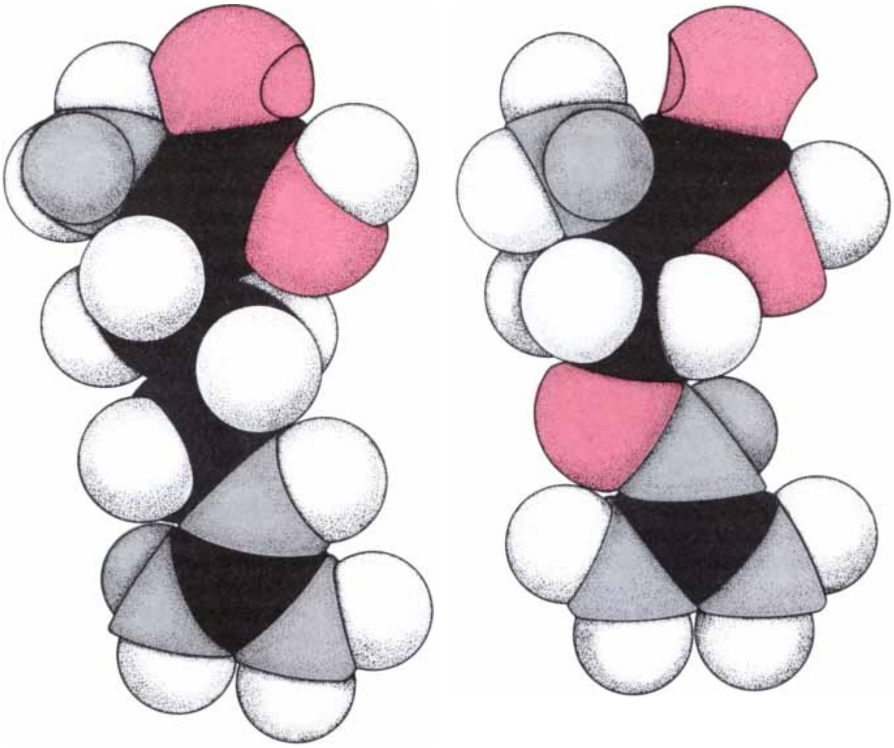
LIFE CYCLE of a new generation of the bruchid beetle *C. brasiliensis* begins when the adult female lays her eggs on a pod of the vinelike legume *D. megacarpa*. The eggs are laid in the late fall, about a month before the pods mature. When the larvae emerge, they bore through the wall of the pod and the coat of a seed, taking up residence in the seed tissues. After several months the larvae have metamorphosed into pupae and then into adults; the young beetles cut exit portals in the seed coat, emerging into the pod and thence through the suture of

gation of these biochemical adaptations.

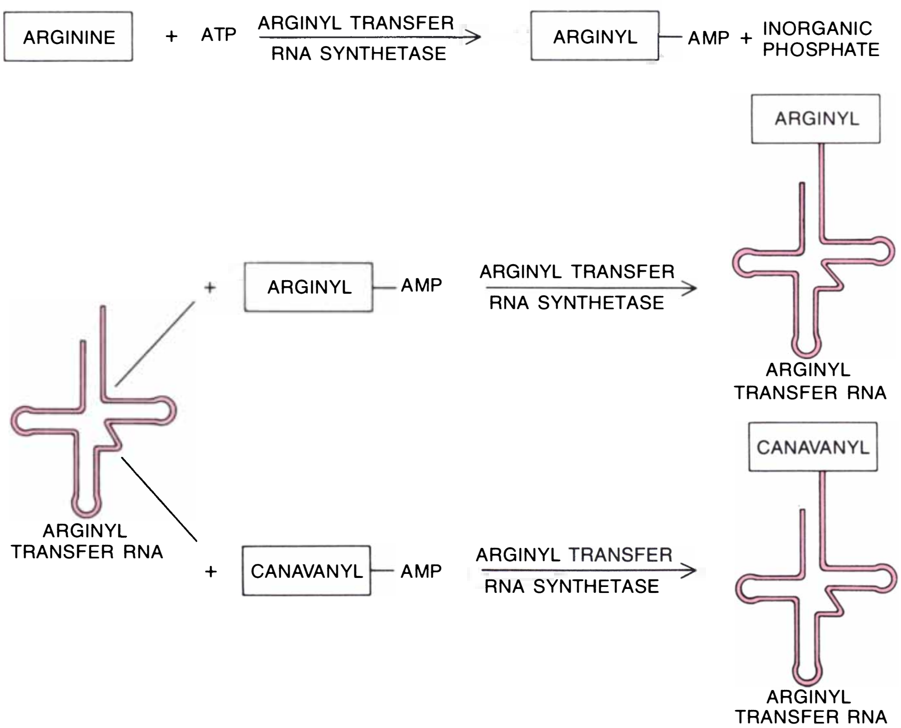
As a beginning my associates and I asked whether or not this bruchid beetle incorporates canavanine into its proteins. If it avoids producing aberrant proteins, how does it manage to do so? To answer these questions we injected



the pod or an opening made by an animal other than an insect. The adult beetles of the species feed on the pollen of various plants, beginning the cycle again by laying eggs in the fall.



MOLECULAR STRUCTURES of L-arginine (*left*) and L-canavanine (*right*) are quite similar. The constituent atoms are carbon (*black*), hydrogen (*white*), nitrogen (*gray*) and oxygen (*color*). Arginine is a protein-forming amino acid; canavanine is one of some 250 amino acids synthesized by higher plants but not utilized to make proteins. The structural difference between the two amino acids is small: the terminal methylene group ($-CH_2$) of arginine is replaced by oxygen in canavanine. In addition the guanidino group of arginine is in the imino form ($=NH$), whereas the corresponding group in the canavanine molecule is in the amino form ($-NH_2$).



ROLE OF TRANSFER RNA in dealing with arginine or mistakenly with canavanine in the production of proteins is shown. Normally (*top*) the enzyme arginyl transfer RNA synthetase reacts with arginine in the presence of adenosine triphosphate (ATP) to make an activated form of the amino acid. The enzyme also mediates the linkage of activated arginine to the correct transfer RNA. Usually (*middle*) arginyl transfer RNA synthetase binds arginine to its transfer RNA before delivering arginine to a ribosome complex, where protein is made. If canavanine is bound instead of arginine (*bottom*), it is transported to the assembly site by the transfer RNA molecule that normally carries arginine. Protein synthesis does not require verification of the amino acid attached to the transfer RNA, and so canavanine ends up in the protein.

larvae, obtained from infected *D. megacarpa* seeds, with radioactively labeled canavanine. Only the terminal carbon atom of the injected canavanine was radioactive. This unique labeling pattern was important because the treatment of radioactive canavanine with the enzymes arginase and urease releases the radioactive carbon atom as carbon dioxide, a gas that can be trapped chemically and then quantified by liquid-scintillation spectroscopy.

The newly synthesized and slightly radioactive proteins obtained from the larvae were isolated and digested with a strong acid to release their constituent amino acid building blocks. The protein digest was then purified by ion-exchange chromatography, which isolated basic amino acids, such as canavanine. Treatment of the isolated amino acids with arginase and urease failed to cause the release of appreciable amounts of radioactive carbon dioxide. This finding established that the larvae were not incorporating canavanine into protein in significant amounts, although some insertion of the radioactive carbon atom of canavanine into protein amino acids had occurred.

The biochemical basis for the finding became apparent when the arginyl transfer RNA synthetase of this insect was compared with that of *Manduca sexta,* the tobacco hornworm. The enzyme from *M. sexta* attached both arginine and canavanine to arginyl transfer RNA, but the arginine-activating enzyme from *C. brasiliensis* attached only arginine. It thus appeared likely that among the biochemical adaptations achieved by the beetle was the development of an arginyl transfer RNA synthetase capable of discriminating between canavanine and arginine. In this way canavanine is not linked to the transfer RNA for arginine and the beetle avoids the production of aberrant, canavanine-containing proteins.

In a recent companion study we gained additional insight into the discriminatory capability of the protein-synthesizing system of the bruchid beetle. The study required the synthesis of a group of radioactively labeled amino acids related structurally to arginine and their testing, along with canavanine and arginine, to determine if any of them were incorporated into the protein of *M. sexta* and *C. brasiliensis.* The tobacco-hornworm larvae fixed each of the compounds into newly synthesized protein; the bruchid-beetle larvae did not. The marked ability of this seed predator to distinguish arginine from molecules structurally akin to it confers a general resistance to the error of incorporating the wrong amino acids into proteins.

Many important questions remained, but the one we directed our attention to was how the bruchid-beetle larva handles canavanine. It may excrete the substance or in some other way avoid its adverse biochemical effects. An exciting alternative possibility was that the beetle had developed some means of utilizing the toxic natural chemical as a food resource. It was reasonable to think the insect might not sacrifice such an abundant source of nitrogen.

Our first effort was to analyze the canavanine content of *D. megacarpa* seeds infested with bruchid-beetle larvae. We applied a statistical technique to estimate the weight of the seed prior to insect attack from the weight of the intact seed coat. The amount of canavanine in the uneaten part of the seed and in the frass (the fecal matter of the larvae) enabled us to calculate how much canavanine the seed had contained originally. These calculations indicated that more than half of the original seed canavanine was consumed in larval feeding.

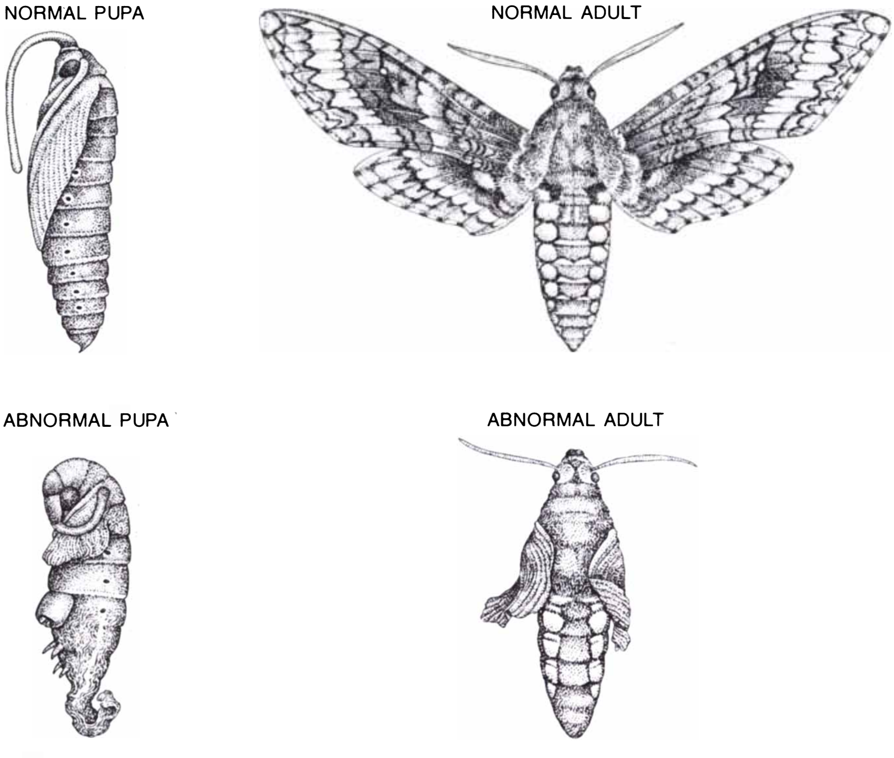
This finding still left the question of how the ingested canavanine was metabolized. Arginase, an enzyme that cleaves L-arginine into L-ornithine and urea, also acts on L-canavanine, cleaving it into L-canaline and urea. The enzyme is distributed widely among insects and has several important functions, one of them being to provide ornithine from arginine for the formation of glutamic acid. The bruchid beetle produces arginase and would be able to convert canavanine into canaline and urea: $H_2N(C{-}NH_2){=}N{-}O{-}CH_2{-}CH_2{-}CH(NH_2)COOH$ [L-canavanine] $\longrightarrow$ $H_2N{-}O{-}CH_2{-}CH_2{-}CH(NH_2)COOH$ [L-canaline] + $H_2N{-}C({=}O)NH_2$ [urea].

Urease, the enzyme that acts on urea, is seldom found in insects. Our analysis of the bruchid-beetle larvae disclosed an extraordinarily high urease activity. The larvae cleave canavanine into canaline and urea and then draw on urease to generate ammonia and urea. In this way half of the nitrogen stored in canavanine becomes available as ammonia for metabolic reactions.

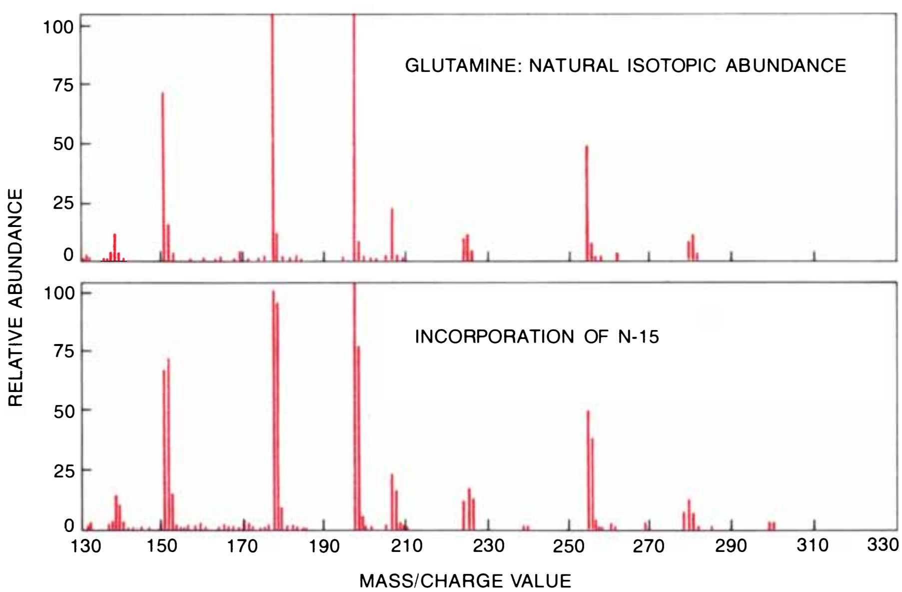
Does the larva utilize the ammonia as a source of nitrogen for its metabolic reactions? Is it not only detoxifying canavanine but also incorporating part of the nitrogen from it into newly made amino acids? These are important questions because with the possible exception of the incorporation of cyanide into L-asparagine, an amino acid of proteins, no other instance is known of an insect utilizing a toxic plant compound for amino acid production.

The commercial availability of urea labeled with the heavy isotope of nitrogen (N-15) and the facile ability of the bruchid-beetle larva to convert canavanine into urea made it possible to test the hypothesis experimentally. Larvae were injected with urea containing the heavy nitrogen isotope, and the newly synthesized free amino acids were isolated and converted chemically into a more volatile form. This conversion made it possible to separate the amino acids by gas chromatography. As each compound emerged from the chromatographic apparatus it entered a mass spectrometer. (These experiments were done at the excellent mass-spectroscopy facilities of the Tobacco and Health Research Institute of the University of Kentucky.)

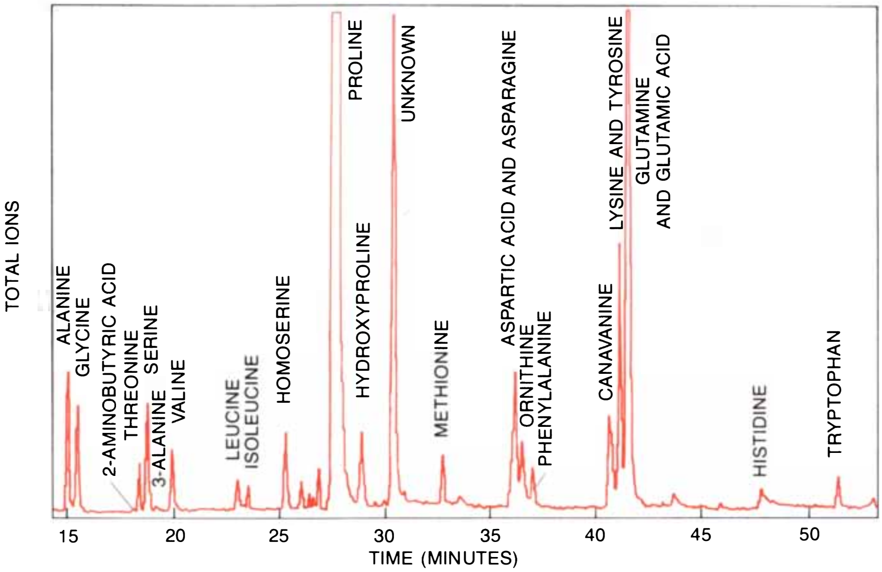
In the mass spectrometer the compound is bombarded by electrons that break the amino acid into a unique pat-



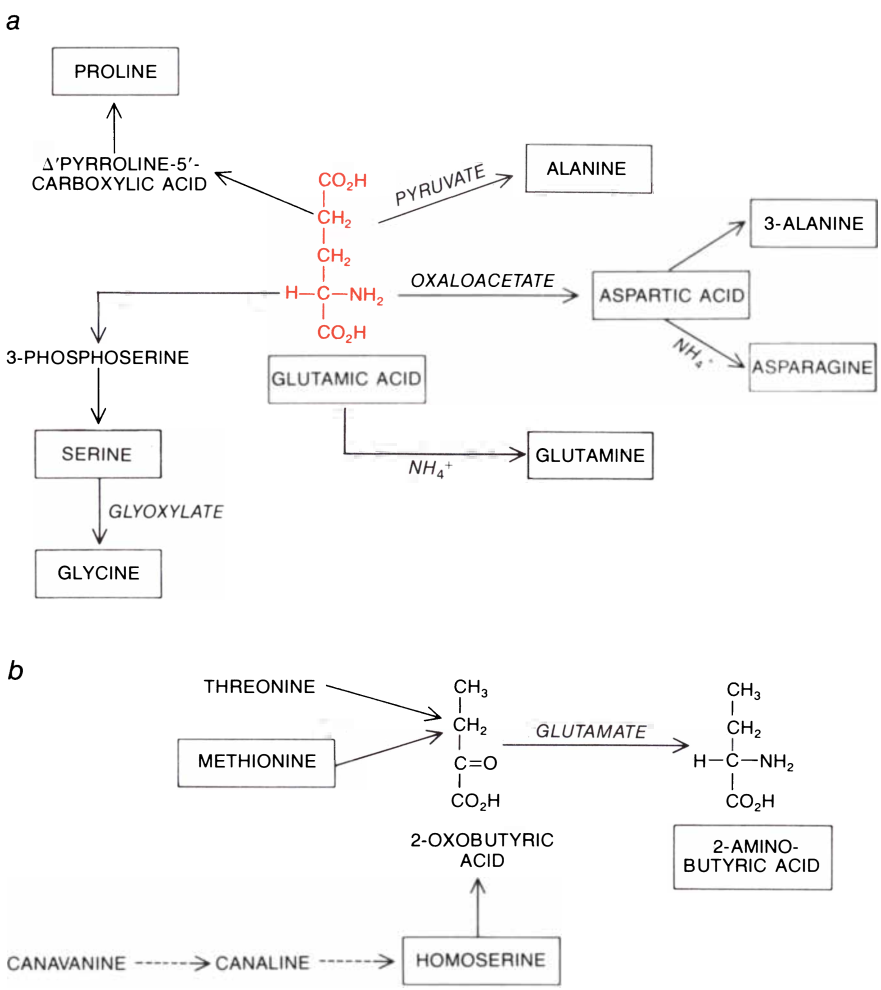
EFFECTS OF CANAVANINE are seen in *Manduca sexta*, the tobacco hornworm. The abnormal pupal and adult forms resulted from the inclusion of canavanine in the diet fed to the larvae in the laboratory many days earlier. Comparably severe developmental abnormalities are elicited by injecting canavanine directly into the hemolymph, the circulatory fluid of the insect.



INCORPORATION OF NITROGEN from canavanine into amino acids was determined by mass spectroscopy. Larvae of *C. brasiliensis* were injected with urea labeled with the heavy isotope of nitrogen (N-15). The amino acids they synthesized were separated by gas chromatography, and each compound then entered the mass spectrometer. The results show that the larvae incorporated N-15 into the protein amino acids that they can synthesize but not into the ones that they must secure in their diet. One of the protein amino acids that incorporated significant amounts of N-15 was L-glutamine; the mass-spectrometric data for it are represented here.



COMPUTER ANALYSIS of the incorporation of heavy nitrogen into the amino acids of *C. brasiliensis* yielded these results. The data show that the nitrogen of canavanine, released by the formation and degradation of urea, is utilized by the bruchid beetle in the synthesis of the amino acids it can manufacture but not in the amino acids the insect can get in its normal diet.



POSSIBLE BIOSYNTHETIC PATHWAYS from glutamic acid (*color*) to the N-15-labeled amino acids produced by the larva of *C. brasiliensis* are charted (*a*). The amino acids that took up heavy nitrogen are indicated by rectangles. The formation of 2-aminobutyric acid from methionine and homoserine, both of which were found to contain N-15, is also shown (*b*). The synthesis of homoserine by *C. brasiliensis,* the only insect known to do so, cannot have been through the degradation of canavanine by way of canaline, because the canavanine was not labeled.

tern of fragments. Computer analyses yield an accurate characterization and compilation of the fragments obtained for each compound. Knowing the relative abundance of the fragments containing the two isotopes of nitrogen makes it possible to accurately reconstruct the proportions of N-15 and N-14 in the original amino acid.

These determinations revealed a significant incorporation of heavy nitrogen into alanine, glycine, serine, proline, methionine, aspartic acid (and/or asparagine) and glutamic acid (and/or glutamine). They are the protein-forming amino acids that insects can synthesize from suitable precursors. On the other hand, no appreciable heavy nitrogen was found in threonine, leucine, isoleucine, histidine, lysine or hydroxyproline. They are the protein-forming amino acids that insects cannot synthesize and must get from their food. Valine is also taken to be essential in the insect diet, but *C. brasiliensis* was able to synthesize very small amounts of it. Tyrosine, phenylalanine, arginine and tryptophan were not evaluated because of technical limitations. Interestingly enough, the bruchid-beetle larvae also incorporated appreciable amounts of N-15 into several nonprotein amino acids: 2-aminobutyric acid, 3-alanine and homoserine.

How might the nitrogen of canavanine that is ultimately released as ammonia support the manufacture of these amino acids? It may be that glutamic acid and glutamine play a central role in the amino acid metabolism of the insect. Enzymes that can catalyze the synthesis of alanine and aspartic acid from glutamic acid are distributed widely among insects. Glutamine and asparagine are readily obtained from glutamic acid and aspartic acid, respectively, by the simple addition of ammonia to those amino acids. Glutamic acid is also linked directly to the synthesis of proline and serine; glycine forms readily from serine. The considerable amount of heavy nitrogen found in 2-aminobutyric acid may seem unexpected, but the compound forms from an organic acid that is readily synthesized from threonine and methionine. Thus virtually all the amino acids known to be synthesized by insects, including those synthesized by *C. brasiliensis,* contain appreciable amounts of nitrogen equivalent to the amount readily supplied by canavanine.

A further problem is that L-canaline, a nonprotein amino acid produced by the cleavage of L-canavanine into urea, is highly toxic in its own right. It is distinctive in being the only nonprotein amino acid with a free terminal aminooxy group: $-ONH_2$. Experiments with larvae of the tobacco hornworm revealed that L-canaline retards growth, causes severe developmental aberrations, increases mortality and interferes with nerve function. Canaline reacts with the aldehyde group of pyridoxal phosphate, a cofactor essential for the function of certain enzymes. The canaline–pyridoxal-phosphate complex is stable, and its formation essentially shuts down the catalytic action of certain enzymes that contain pyridoxal phosphate. On first consideration it appears that the bruchid beetle is merely exchanging one poison for another.

By examining the metabolic fate of canaline in the bruchid-beetle larvae we found that the insect has an enzyme capable of cleaving canaline to yield homoserine: $H_2N–O–CH_2–CH_2–CH(NH_2)COOH$ [L-canaline] $\longrightarrow$ $HO–CH_2–CH_2–CH(NH_2)–COOH$ [L-homoserine] + NH_3. With this finding we confronted the question of how the insect deals with all the ammonia, which is a toxic form of reduced nitrogen. Uric acid is the principal nitrogenous excretory product of insects; ammonia is normally a minor excretory product. Analyzing the frass of the larvae, we found that nearly 90 percent of the nitrogen in it was in the form of ammonia and urea. Only 11 percent of the total excreted nitrogen was in the form of uric acid. Clearly this seed-eating bruchid beetle can excrete some of the extraneous ammonia directly in the frass.

The elimination of ammonia by way of the frass is not the only mechanism available to the insect for avoiding the toxic effects of too much ammonia. The larva of the beetle has a highly active glutamine synthetase, an enzyme that can add ammonia to glutamic acid to form glutamine. Glutamine is a principal free amino acid of the insect's hemolymph. Analysis of the amino acids of the hemolymph also reveals a massive peak for the amino acid proline. Perhaps proline serves as a source of carbon skeleton for the formation of glutamic acid before the formation of glutamine. The reaction sequence from proline to glutamic acid to glutamine may provide an important means of channeling toxic ammonia into a harmless form.

As for homoserine, the investigations of John Giovanelli and his associates at the National Institute of Mental Health have established that plants employ homoserine derivatives as precursors in the manufacture of certain indispensable amino acids containing sulfur. One would expect many types of insect (for example, the aphids that feed on the seed of the pea plant, *Pisum sativum*) to utilize homoserine because this nonprotein amino acid can account for more than 10 percent of the dry weight of certain parts of a plant.

After administering homoserine labeled with carbon 14 to the bruchid-beetle larvae we found a detectable but quantitatively insignificant transfer of the radioactive atoms to the amino acids methionine and cysteine. Automated amino acid analyses of the extract of these homoserine-treated insects revealed a massive peak of radioactivity that was not caused by an amino acid. This was evident because the unknown substance did not form a colored complex with ninhydrin. That reagent reacts with amino acids to form typically a blue purple color; it is the classic colorimetric test for amino acids.

It may be that the bruchid beetle uses homoserine to provide nitrogen for the synthesis of other amino acids through the action of an aminotransferase. The reaction of this enzyme with homoserine would cause the homoserine to lose its normal ninhydrin-positive response. In any event it is clear that the insect is synthesizing significant amounts of homoserine for a purpose other than the manufacture of amino acids containing sulfur.

A fascinating aspect of these biochemical processes is that the bruchid beetle has adapted to a detrimental substance in its food (canavanine) by functioning in much the same way as a plant synthesizing the same substance. This convergent evolution is revealed by four lines of evidence. The first is the formation of an enzyme (arginyl transfer RNA synthetase) capable of discriminating between arginine and canavanine. The ability to distinguish between an amino acid that is normally part of a protein and one that is not is a well-recognized mechanism whereby plants avoid the adverse effects of their own nonprotein amino acids.

The second line of evidence is that the formation of urea from canavanine coupled with the formation of ammonia from urea is the principal means by which plants that make canavanine mobilize the nitrogen stored in it. The third is that the insect detoxifies canaline by converting it into homoserine; in plants synthesizing canavanine the conversion of homoserine into canaline may be part of the synthesis of canavanine.

Finally, we found that the arginase of *C. brasiliensis* has a greater apparent affinity for canavanine and reacts faster with canavanine (compared with arginine) as a substrate than the arginases of the other insects we tested. A similar distinction is found in comparing the arginase of *Canavalia ensiformis,* a legume that utilizes canavanine, with the arginase of the soybean *Glycine max,* which does not synthesize canavanine.

Another conclusion to be drawn from our investigations is that this seed-eating bruchid beetle has achieved many distinct but interrelated biochemical adaptations in its utilization of *D. megacarpa* as a food source. It is reasonable to propose that among the first of the adaptations was the development of the discriminatory ability enabling the insect to avoid the synthesis of dysfunctional proteins. The bruchid beetles that originally invaded the legume may have simply excreted canavanine and canaline or somehow avoided them. Other nitrogen-rich metabolites stored in the seed (such as proteins) could have satisfied the need of the growing larva for nitrogen. Over a period of time, through natural selection, the insects may have become progressively better equipped to cope with canavanine and canaline and ultimately could utilize the nitrogen they contain.

Such adaptive successes would have fostered the dependence of the beetle on *D. megacarpa*. In time the benefits of its association with the plant may have led to the failure to lay its eggs on other seeds in its habitat. In considering the narrow feeding range of *C. brasiliensis* and the danger inherent in depending on a single food and egg-laying resource, it is important to think of the potential advantages. The insect is relatively free from competition for the seeds and does not have to invest in a capacity to process a wide range of plant toxins. The developing larvae are assured a reasonably safe haven, since the toxicity of canavanine should protect the seed from the larvae of other species.

A final point on which these studies bear is the remarkable ability of insects to adapt to a host of detrimental substances, including the insecticides devised by human beings. Is it possible that microbes living symbiotically in the insect gut have a major role in processing toxic compounds such as canavanine? The seed of *D. megacarpa* has an array of microbial symbionts and pathogens that *C. brasiliensis* could have acquired, and they may account for some or even all of the biochemical adaptations required for the utilization of canavanine and canaline. We have isolated microbial symbionts from the beetle larvae and found several that can subsist on canavanine and canaline (or one of them) as the sole source of both carbon and nitrogen. I intend to ascertain how much of the acquired biochemical abilities of this seed predator derives from its microbial symbionts.

Our continuing studies of the relation between the bruchid beetle and its leguminous host have provided a basic understanding of how an insect has adapted at the biochemical level to toxins presented by its host plant. It is regrettable that similar biochemical investigations have not been done, so that the present information is limited to this single association. Nevertheless, what we have found is of practical value in helping to understand the response of insects not only to natural toxins but also to chemical control measures. In addition the studies reveal the kinds of insight that can be gained by the union of ecological and biochemical approaches.

The Eruption of Krakatau

The explosions that obliterated most of the Indonesian island 100 years ago are only now beginning to be understood. The evidence is largely the volcanic deposits and the timing of air and sea waves

by Peter Francis and Stephen Self

One hundred years ago, on the morning of August 27, 1883, a series of intermittent volcanic explosions culminated in the paroxysmal blasts that accompanied the destruction of most of Krakatau, a small island in the Sunda Straits. Tsunamis ("tidal waves") set in motion by the eruption killed more than 30,000 people on the neighboring Indonesian islands of Java and Sumatra. The event attracted worldwide attention, and the ensuing scientific investigations made important contributions to the fledgling field of volcanology. Only now, however, is it becoming possible to explain some of the major events in the eruption sequence in terms of the underlying volcanic processes. Here we address three much-debated questions the eruption raises: What triggered it? Why were there so many violent explosions in the eruption sequence? What was the relation between the devastating tsunamis and the massive explosions?

When Krakatau erupted, the explosions were heard in central Australia, Manila, Sri Lanka and on Rodriguez Island, more than 5,000 kilometers away in the Indian Ocean. Lower-frequency atmospheric waves (at frequencies too low to be audible) were detected worldwide; barometers in Tokyo, 5,863 kilometers away, registered a pressure increase of 1.45 millibars. The sea waves generated by the eruption traveled not only across the Pacific but also across the Atlantic: they were detected by tide gauges in the Bay of Biscay, 17,000 kilometers away. The dust and gases injected into the atmosphere by the eruption produced spectacular sunsets worldwide for months afterward. The mean temperatures recorded in the Northern Hemisphere during the same period were from .5 to .8 degree Celsius lower than normal.

Krakatau's reputation as the classic volcanic eruption, however, is perhaps due as much to the date of its occurrence as to its violence. The eruption, which was one of the first to be the subject of intensive scientific investigation, came at a time in the Victorian era when science was followed by a large and enthusiastic audience. The eruption of the volcano Tambora on the Indonesian island of Sumbawa in 1815 attracted comparatively little attention even though it was much larger; Tambora is estimated to have ejected between 150 and 180 cubic kilometers of pumice and ash whereas Krakatau ejected 20 cubic kilometers. Moreover, the eruption of Tambora caused the death of more than 90,000 people either directly or as a result of tsunamis and an ensuing famine [see "The Year without a Summer," by Henry Stommel and Elizabeth Stommel; SCIENTIFIC AMERICAN, June, 1979]. At the time, however, no one made the connection between the unusually cold weather in Europe and North America in the summer of 1816 and the eruption of Tambora the year before, and the eruption has yet to be studied in detail.

In contrast, both the Royal Society of London and the Dutch government, which was then the colonial administrator of the Indonesian islands, published lengthy reports about the eruption of Krakatau soon after it occurred. The Royal Society's report emphasized the worldwide atmospheric effects of the eruption. Of its 494 pages 312 were devoted to "the unusual optical phenomena of the atmosphere, 1883–6, including twilight effects, coronal appearances, sky haze, coloured suns and moons, etc." The report issued by the Dutch committee, led by Rogier D. M. Verbeek, a mining engineer and geologist, covers the geologic aspects of the eruption in greater detail. The members of the committee visited the scene of the eruption on October 15, 1883, and repeatedly thereafter. They mapped the new islands and the remnants of the old ones and measured the changed contours of the ocean bottom. Verbeek himself collected samples of the volcanic ejecta, which he and his team later examined under the microscope.

The conclusions Verbeek drew from this diverse evidence were for the most part remarkably prescient. For example, on the basis of depth soundings and measurements of the area of ashfall he made an estimate of the amount of material ejected by the volcano that has stood without substantial revision to this day. Finding that the rock specimens he had collected were magmatic rock rather than older rock, he correctly proposed that the old volcanic cone had not been blasted into the air but had foundered into the sea when the magma chamber underlying it was exhausted.

The full story of the eruption based on the accounts of witnesses makes compelling reading. Here we shall concentrate on those events of the eruption sequence that seem to be correlated with the emplacement of volcanic deposits, because it is on this correlation that reconstructions of what happened must be based.

The eruption sequence has been established largely from records kept by Dutch administrators living well above the shoreline in Sumatra and Java and in towns in the interior and from reports made by the officers on watch on ships passing through the Sunda Straits; several of the ships sailed near Krakatau during the most violent part of the eruption. The descriptions of the nature and stratigraphy of the volcanic deposits are based largely on a field study by one of us (Self) and Michael R. Rampino of the Goddard Institute for Space Studies of the National Aeronautics and Space Administration. We studied and sampled the deposits when we visited the island in 1979 in the course of a broader investigation into the atmospheric effects of volcanic eruptions.

Navigational charts of Krakatau and the surrounding areas of the Sunda Straits made before the eruption show that the island consisted of three volcanic cones aligned roughly northwest to southeast. The largest cone, Rakata, 813 meters high, was at the southern end of the chain. A lower cone, Danan, was in

BISECTED VOLCANIC CONE is all that remained of the island of Krakatau after the eruption of August 27, 1883. The island originally consisted of three volcanic cones aligned roughly northwest to southeast. The main vent of the 1883 eruption is thought to have lain between the two northernmost cones. Late in the eruption sequence two-thirds of the island foundered into the sea as the roof of the magma chamber underlying it collapsed. The northern face of the southern cone, Rakata, which was perched on the edge of the new caldera (the submarine depression formed by the collapse), was left virtually unsupported and subsequently slumped into the sea. Clearly visible are the interior structure of the old volcanic cone (including the rock whitened by hydrothermal alteration near the central vent), the feeder dikes (dark columns) leading down to the magma chamber and the alternating lava flows and ash layers that made up the cone. The layer of white pumice deposits laid down on both flanks of the cone is also visible in this chromolithograph from *Album of Krakatau.*

TWO NEW ISLANDS named Steers and Calmeyer in the Sunda Straits north of Krakatau consisted of deposits laid down on the floor of the straits that were exposed above sea level. The islands were created by successive pyroclastic flows: ground-hugging clouds of pumice and ash driven by gravity and fluidized by hot gases. The flows traveled an average of 15 kilometers from the vent, much of it through or over water. Hot material caused the explosive vaporization of seawater; some of the many large explosions late in the eruption sequence may have been such secondary explosions. This chromolithograph from *Album of Krakatau* shows a large secondary-explosion crater on Calmeyer. The crater closely resembles the craters formed by pyroclastic flows from Mount St. Helens that entered Spirit Lake. *Album of Krakatau* was published in 1886; the loosely packed deposits eroded quickly and the islands soon vanished below the surface.

the middle and a much lower one, Perbuwatan, was at the northern end. Lava flows around Perbuwatan showed that it had been active in the geologically recent past, and it is thought to have been the site of a pumice eruption in 1680. Two smaller islands near Krakatau—Sertung and Rakata Kecil (Little Rakata)—and the southern end of Krakatau itself were probably remnants of the rim of a submerged caldera: a large volcanic crater formed by collapse.

When Perbuwatan burst spectacularly into life with a series of deafening explosions on May 20, 1883, after nearly 200 years of inactivity, the eruption came as a surprise. Krakatau was uninhabited and was visited only occasionally by fishermen and woodcutters. Thus any small-scale activity that might have preceded its reawakening went unnoticed. There was a period of markedly increased seismic activity around the straits before the eruption, but at the time no link was made between this activity and Krakatau.

The May eruption of Perbuwatan was accompanied by explosions that could be heard more than 150 kilometers away. At the same distance atmospheric pressure waves of very long wavelength were energetic enough to stop clocks, rattle doors and windows and dislodge hanging lamps. Since the pressure waves were inaudible, their effects were often mistaken for those of earthquakes. Although some seismic activity was recorded during the climactic phase of the eruption in August, at this stage almost all the energy seems to have been transmitted through the air.

Perbuwatan continued to erupt intermittently throughout May, June and July, but the activity was relatively unimpressive. According to Captain Ferzenaar, a Dutch surveyor who visited Krakatau on August 11, trees were still standing, although they had been stripped of foliage by the falling ash. Ferzenaar noted that a layer of ash about 50 centimeters thick covered the island. Today only very limited outcrops of fine-to-medium-grained ash from this early phase of the eruption can be seen.

The relatively minor explosions continued into August and culminated in massive explosions on August 26 and 27. The scale of activity on these two days was so large that it has been difficult to piece together exactly what happened, and much detail is surely missing. No one within close range of the volcano survived. Reconstructions of the eruption sequence in this crucial period are based largely on instrument records and the volcanic deposits. The enormous explosions on August 27 generated air waves so powerful that they registered on a recording pressure gauge at the gasworks at Jakarta, which thereby preserved a record of their timing and relative amplitude. The timing and magnitude of the tsunamis were recorded by tide gauges along the straits.

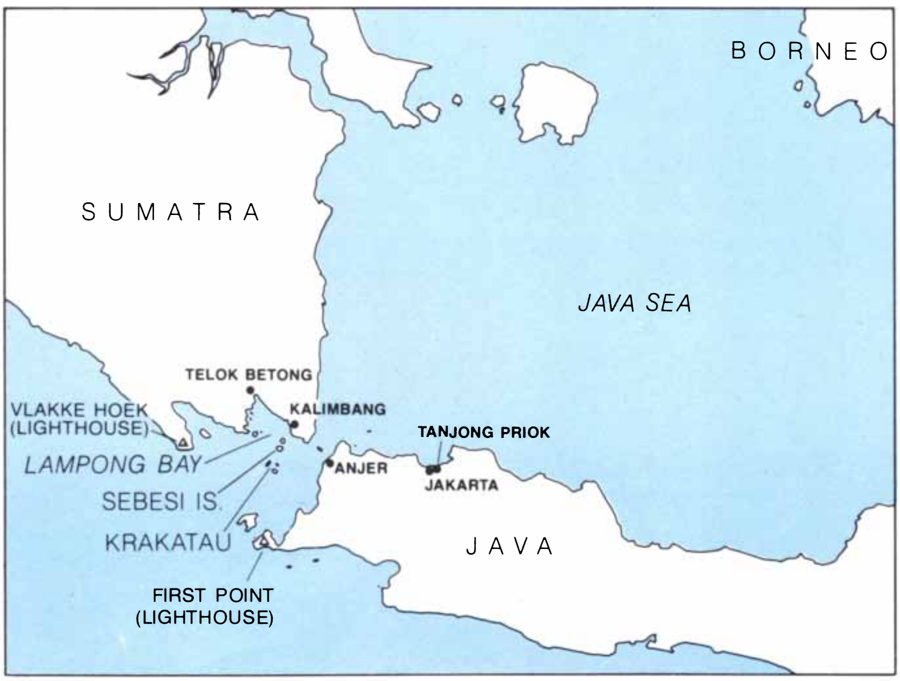
The climactic phase of the eruption can be divided into two stages on the basis of the nature and sequence of the deposited ash layers. Beginning at 1:00 P.M. on August 26 a series of explosions at intervals of about 10 minutes created a more or less sustained eruption column over the island that is reported to have reached a height of about 25 kilometers (82,000 feet). The explosions produced primarily air-fall material: pumice and ash carried upward in the atmosphere by a column of hot convecting gas. (Both pumice and ash are frothy glassy materials created by the chilling of vesiculating magma. The distinction between them is largely one of size; fragments smaller than two millimeters in diameter are generally called ash.)

Although pumice and ash deposits from this phase of the eruption accumulated to thicknesses of up to 20 meters on the islands of Sertung and Rakata Kecil and ships within 20 kilometers of the volcano reported heavy falls of ash accompanied by large pumice clasts (fragments) up to 10 centimeters in diameter, the heavy ashfall was limited in extent. Only a minor fall of ash was reported on Sumatra and on western Java.

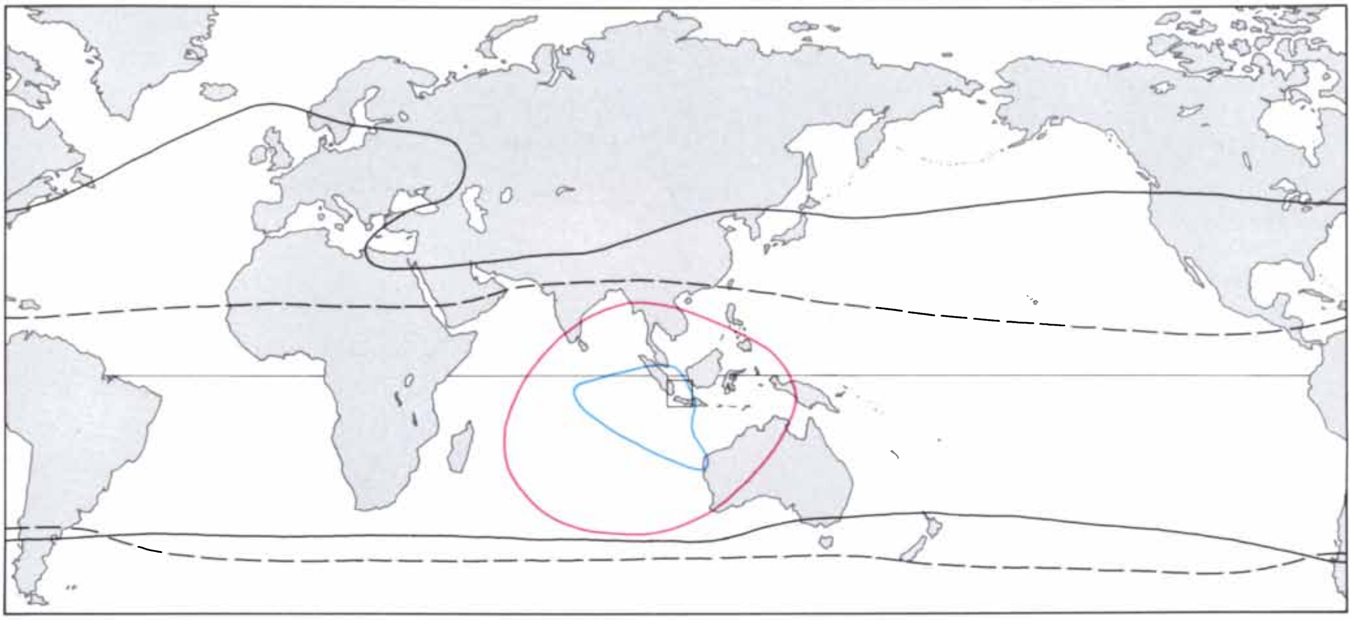
At 5:30 A.M. on August 27 the character of the eruption changed dramatically. In the course of the day there were many enormous explosions. The largest explosion, the one heard as far away as Rodriguez Island, came at 9:58 A.M.; it was associated with the largest tsunami, estimated to have crested at 40 meters, which caused most of the deaths on the neighboring coastlines.

During this period the explosions were characterized by the formation of pyroclastic flows as well as a sustained column of airborne material. The explosions of August 27 seem to have been paroxysmal rather than continuous. Each blast propelled a large amount of pumice and ash high into the air, perhaps as high as five kilometers. The aggregation of material was too dense and heavy to remain airborne for long, however, and most of it immediately fell back to the ground. There it formed incandescent, ground-hugging clouds, driven by gravity and fluidized by hot gases, that moved rapidly off the island and into the sea. A fraction of the ash was lofted much higher by convection currents set up in the atmosphere by local heating. The resulting ash cloud may have reached a height of about 40 kilometers. The ash from the towering cloud fell out over a wide area. The neighboring coastlines were plunged into darkness and ashfall was reported as far away as the Cocos Islands, 1,850 kilometers from Krakatau.

The pyroclastic flows laid down a dis-



KRAKATAU WAS A SMALL UNINHABITED ISLAND about 32 kilometers west of the narrowest part of the Sunda Straits between the Indonesian islands of Sumatra and Java. More than 30,000 people were killed by the eruption that obliterated most of the island. The majority were victims of the tsunamis that swept over the low-lying coastlines of the neighboring islands. Along the east side of Lampong Bay on the southern coast of Sumatra, however, some bodies were found buried in ash. Those people were probably killed by pyroclastic flows that had traveled more than 40 kilometers from Krakatau on the surface of the sea and yet were still lethally hot. The names on the map are locations mentioned in the graph of the eruption sequence on page 152. Modern names have been substituted for their 19th-century equivalents.



GLOBAL EFFECTS of the Krakatau eruption are plotted on this world map. Shown are the range within which ash fell out (*blue line*), the range within which the explosions were heard (*red line*) and the range within which atmospheric effects caused by the volcanic ash and aerosols injected into the upper atmosphere were reported before September 22, 1883 (*broken black line*), and by late in November (*solid black line*). Why the final explosions were so violent has not yet been explained. Ash from the eruption fell out over an area of 700,000 square kilometers, the largest range of ashfall ever produced by a volcanic eruption in historic times. The major explosions generated dense clouds of fine ash over the Sunda Straits that probably attenuated sound quite effectively. One of the consequences may have been that people living on the coastlines of Java and Sumatra, soon to be victims of the tsunamis generated by the eruption, did not hear the explosions that were heard by others as far away as central Australia. A study by the British physicist Rollo Russell of the path and travel time of the high cloud, based on reports of atmospheric effects, provided the first evidence for stratospheric circulation patterns.

tinctive deposit called ignimbrite; these deposits account for the largest fraction of the material erupted by Krakatau. Because the material in the pyroclastic flows was fluidized (given a low density and viscosity) by hot gases, it made extremely efficient use of the kinetic energy it acquired in falling from the height of the eruption column and so traveled long horizontal distances. Deposits of ignimbrite as thick as 40 meters were laid down as far as 15 kilometers from the primary vent, thought to have been between Danan and Perbuwatan.

The pyroclastic flows seem to have spread out preferentially to the north and northeast, covering the islands and the surrounding sea floor with a blanket of ignimbrite. The distribution was probably caused by the high cone of Rakata, which forced material from the collapsing eruption column northward. Reports of burns caused by hot ash from the area around Kalimbang in southern Sumatra provide evidence that some of the flows traveled to the northeast as far as 40 kilometers. It is clear that the injuries were due to horizontal flows rather than vertical ashfall because in one instance the survivors described hot gases and ash blowing up through the floorboards of a house.

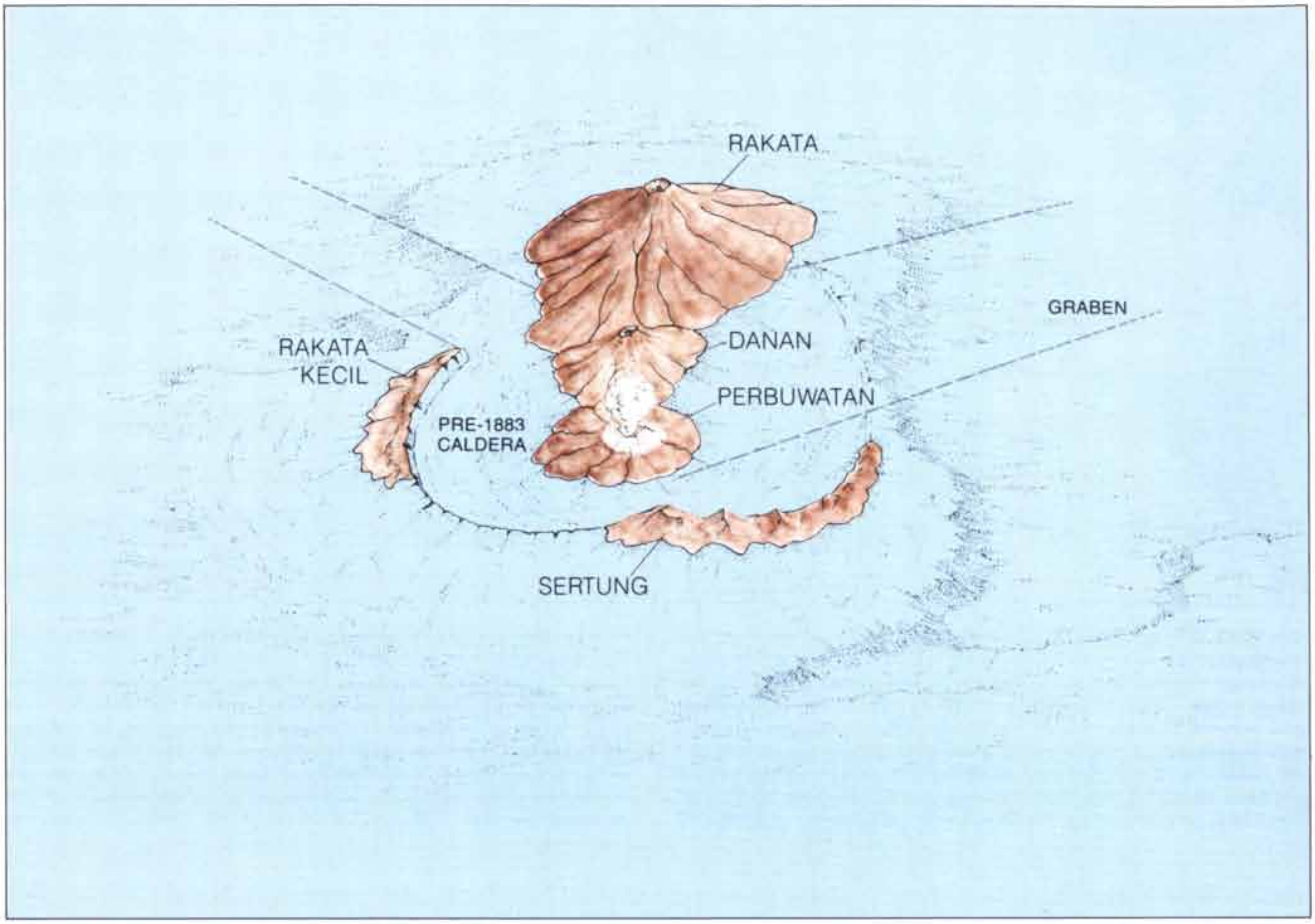
On August 28 the Dutch ship *Gouverneur-Generaal Loudon* attempted to sail from Telok Betong on Sumatra through the Sunda Straits north of Krakatau to Jakarta on Java. Parts of the straits had suddenly become too shallow to be navigable and were also blocked by floating islands of pumice. The ship was forced to deviate widely from its accustomed route, sailing west rather than east and finally passing south of Krakatau. The route the ship took roughly followed the outer edge of the ignimbrite that had been laid down on the floor of the sea.

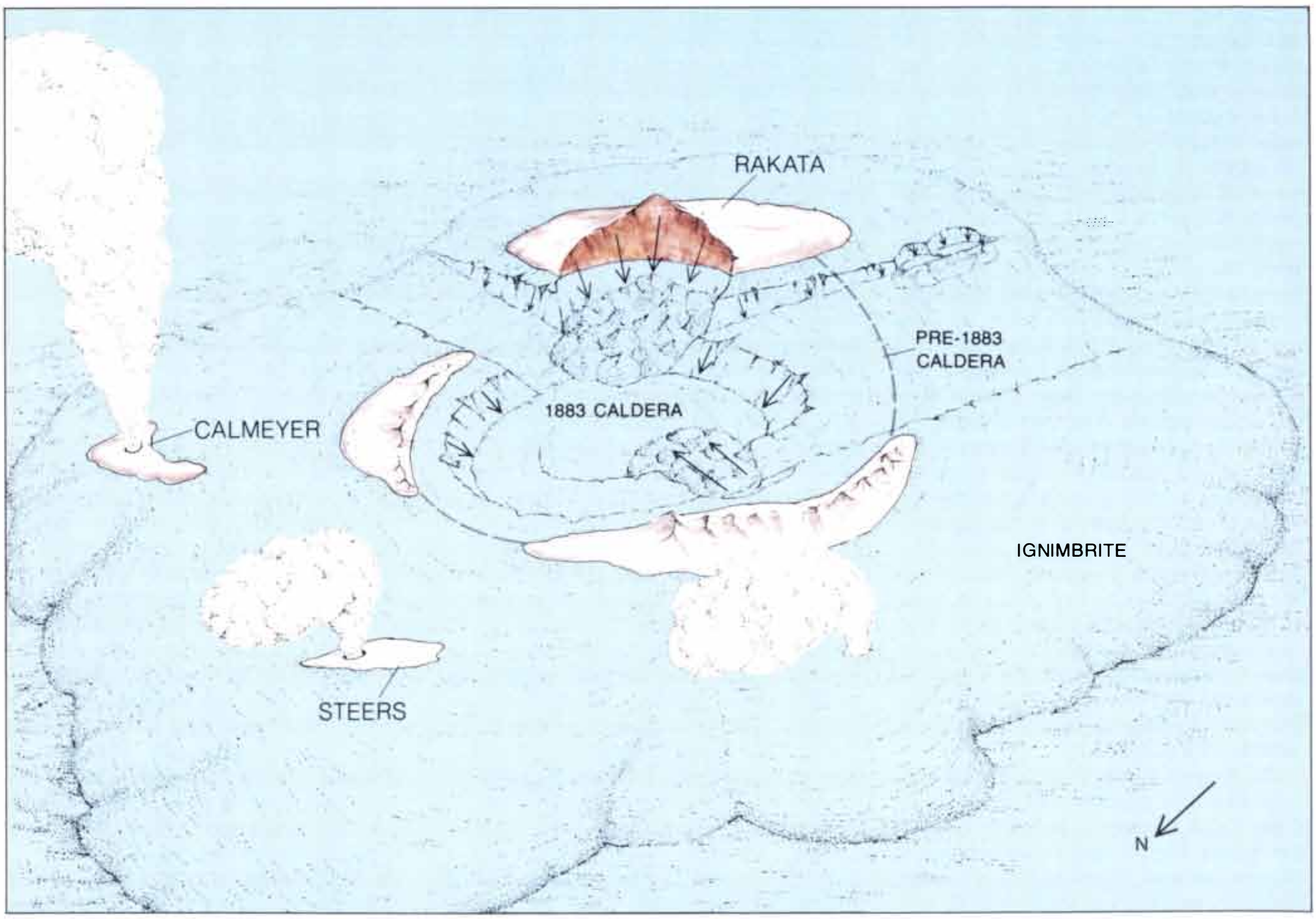
The detailed hydrographic charts Verbeek's team of investigators made in mid-October record the topographic changes caused by the eruption. The northern two-thirds of the island of Krakatau had disappeared. The coastlines of Sertung and Rakata Kecil had been extended by as much as three kilometers by deposits of pumice and ash. Even the southern remnant of Krakatau was girdled by deposits of white ignimbrite. Most of the volcanic material, however, had gone into the sea. Parts of the Sunda Straits originally from 20 to 60 meters deep had been filled by ignimbrite. In the straits to the north of Krakatau ignimbrite exposed above sea level had formed two islands, later named Steers and Calmeyer.

A new caldera, 290 meters deep at its southern end, had formed within the older and now partially filled prehistoric caldera. The distribution of the pyroclastic-flow deposits of the southern and eastern flanks of Rakata suggests they were laid down before the caldera formed, indicating that the magma chamber collapsed relatively late in the eruption sequence. Most of the island foundered into the sea when the roof of the chamber collapsed, but the cone of Rakata was apparently left perched on the southern rim of the caldera, its northern face virtually unsupported. This side of the volcano then slumped into the sea, leaving behind a spectacularly bisected volcanic cone, all that remains of the island of Krakatau.

One reason little progress was made until recently toward understanding what happened at Krakatau is that little attention was given to the physical characteristics of the volcanic deposits, particularly those left by the pyroclastic flows. Verbeek studied the petrology of the deposits and C. E. Stehn (who visited the area in 1927 when the eruption of a new, still-underwater volcano inside the 1883 caldera became perceptible at the surface) studied their stratigraphy. Howel Williams of the University of California at Berkeley made an important contribution in 1941, when he pointed out that much of the exposed pumice had been emplaced by pyroclastic flows rather than by the fallout of airborne material. Only within the past 15 years, however, have volcanologists begun to realize the significance of the physical characteristics of volcanic deposits, such as grain size and internal structure. These characteristics of the Krakatau deposits were studied for the first time when one of us (Self) and Rampino visited the island in 1979.

In many respects it is hardly surprising that these aspects of the deposits had


RAKATA
GRABEN
DANAN
RAKATA KECIL
PERBUWATAN
PRE-1883 CALDERA
SERTUNG


RAKATA
PRE-1883 CALDERA
1883 CALDERA
CALMEYER
IGNIMBRITE
STEERS
N

not been examined before. Not only was most of the ignimbrite deposited in the sea but also the islands of ignimbrite, Steers and Calmeyer, were rapidly eroded and soon disappeared below the sea surface. The deposits that fell on the older islands were also rapidly eroded. Verbeek noted two months after the eruption that steep-sided gullies 40 meters deep had been cut into the deposits. The highly irregular topography was then enveloped in luxuriant and soon impenetrable vegetation. Today the deposits can be sampled only from boats at the bottom of crumbling sea cliffs. Yet it is on the physical characteristics of the deposits that many of the questions about the eruption turn.

Studies of the deposits provide the basis for our answer to the first question: What triggered the eruption? Major volcanic eruptions are caused by the sudden decompression of magma that is saturated or supersaturated with dissolved gases such as carbon dioxide and water vapor. The dissolved volatiles can be released in two ways. The pressure in the magma chamber may gradually increase as more volatiles come out of solution; when the pressure exceeds the strength of the overlying rock, the magma forces an opening to the surface. Alternatively, a tectonic process such as a landslide caused by an earthquake or movement along faults overlying the magma chamber may create an opening that causes instantaneous decompression. There is no evidence that any tectonic process was responsible for the eruption of Krakatau, and so an explanation must be sought in processes within the magma chamber itself.

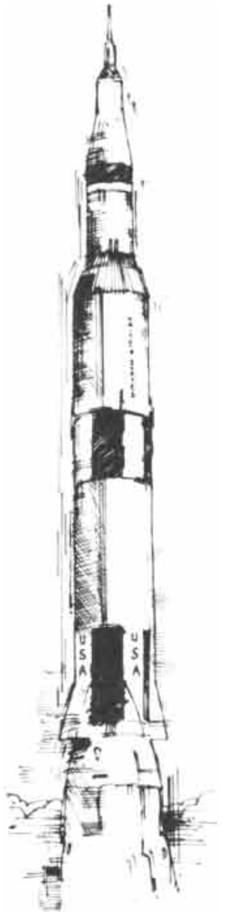
Verbeek suggested that the final explosions began when seawater penetrated the magma chamber and reacted violently with the hot magma. This hypothesis was at one time widely accepted and was generalized to include the earlier explosions. Volcanic eruptions caused by the violent interaction of water and magma, called phreatomagmatic eruptions, do indeed take place either when magma heats and cracks rock close to an underground water table or when seawater somehow gains entry to the magma chamber. The mixture of groundwater or seawater and magma is highly explosive and usually gives rise to a distinctive deposit of very fine-grained and widely dispersed ash. The eruptions at Krakatau left rather different deposits that do not yield unequivocal evidence for this type of eruption.

The intrusion of large amounts of seawater could be expected to cool the magma, so that the ejecta, particularly the finer-grained ash, would cool faster in a phreatomagmatic eruption than it would in other types of eruptions. The field evidence at Krakatau is somewhat equivocal on this point. On the one hand, the early air-fall deposits on Rakata Kecil, only 2.3 kilometers from the postulated vent, were so hot that the glassy pumice fragments were soft enough to weld together in places. The nature of this deposit thus makes it unlikely that the ejecta had been cooled by contact with water.

On the other hand, George P. L. Walker of the University of Hawaii at Manoa has argued that the character of some of the ignimbrite deposits suggests the magma had been cooled by water. The centers of the larger pumice clasts are frothier than the peripheries. In addition the clasts have an exterior skin such as might be formed by sudden cooling. These characteristics suggest the clasts were originally hotter than the surrounding matrix of ash and cooled from the outside in. It is, however, possible that seawater either cooled the erupted material as it left the vent or chilled the ignimbrite as it was deposited. (Most of the deposits visible today are at sea level.)

The geologic evidence therefore does not indicate clearly that there was direct explosive contact between seawater and hot magma in the chamber or conduit of Krakatau. Since the vent was close to sea level, it nonetheless seems probable that there were minor phreatic explosions, which may not have involved direct contact between the water and the magma, late in the sequence, beginning perhaps on the evening of August 26. These explosions may have weakened the roof of the main magma chamber, causing sudden decompression and the release of large volumes of vesiculating magma. We propose, however, that the

DRAWINGS OF KRAKATAU before and after the eruption are reconstructed from contemporary hydrographic charts. The island consisted of three volcanic cones: Rakata, Danan and Perbuwatan. The outlying islands Sertung and Rakata Kecil, and perhaps the southern edge of Krakatau itself, were remnants of the rim of a prehistoric caldera. Two submarine grabens (depressions) near the island indicate that the crust in the area was under extensional stress; this may have thinned the crust enough to accommodate the 1883 magma chamber. Most of the magma was ejected in the form of pyroclastic flows, which deposited on the sea floor a layer of ignimbrite (pumice fragments) as thick as 40 meters. The flows spread preferentially to the north and northeast, probably because the high cone of Rakata to the south acted as a barrier. The two new islands Steers and Calmeyer were areas of the ignimbrite exposed above sea level. Once the eruption had exhausted the supply of magma the chamber collapsed to form a new caldera and most of the island foundered into the sea. The 1883 caldera was probably elongate because the collapse followed the fault lines of the grabens. The drawings are not to scale.

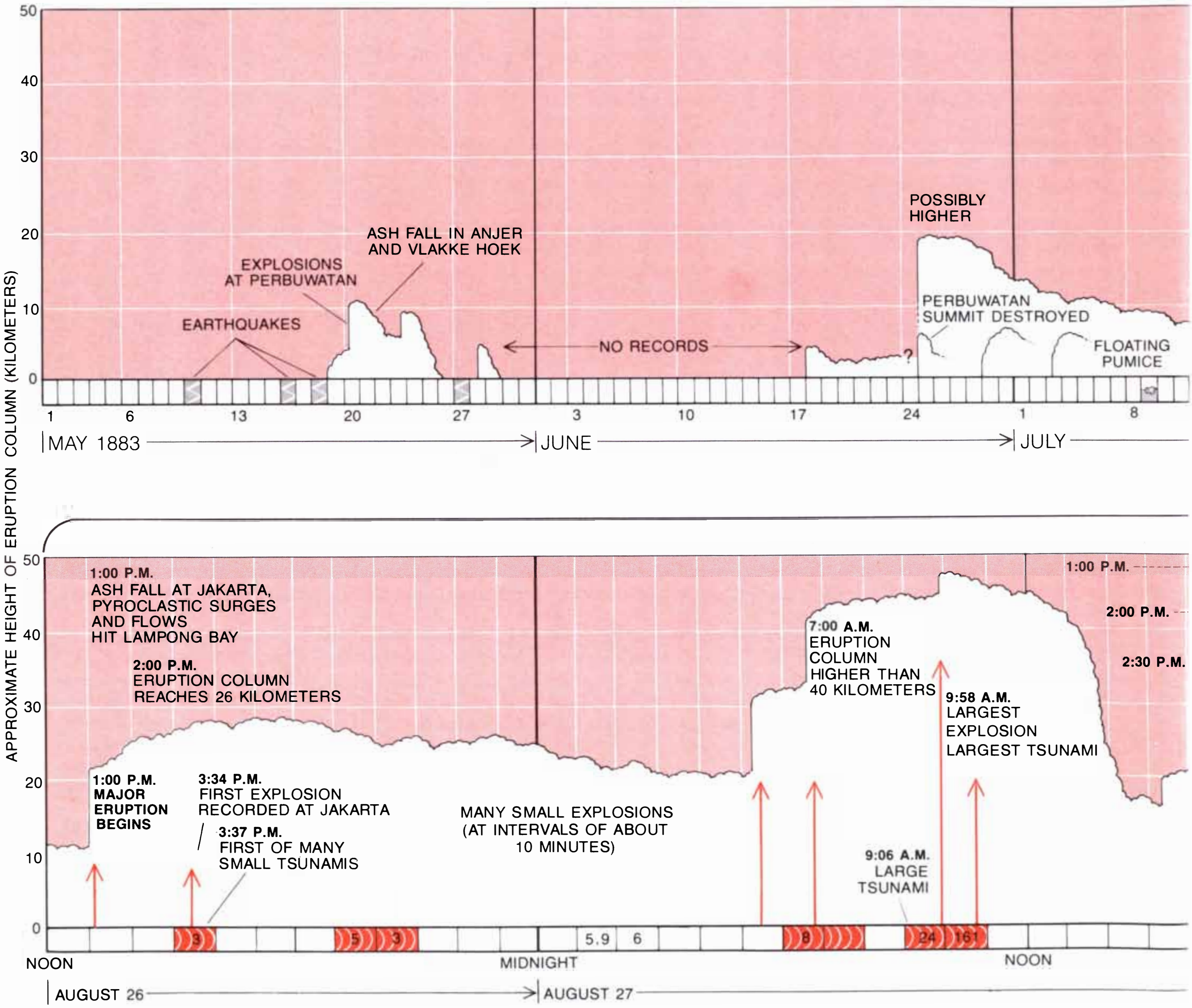


interaction of water and magma was a contributory cause rather than the primary cause of the eruption.

The nature of the pumice fragments themselves points to a more likely triggering mechanism for the eruption. The frothy glass of the pumice includes a small number of mineral crystals that started to form in the magma chamber before the eruption. (When the magma cools relatively fast, glass is formed; slow cooling favors the formation of crystals.) The composition of a pumice and of any crystalline inclusions depends on the composition of the magma from which they formed. Magma ranges from basalt, a dark material relatively poor in silica, through andesite and dacite to rhyolite, a light gray material rich in silica. The lighter-colored magmas are called silicic; they form minerals with a high content of silica, such as quartz and feldspar. The darker magmas are called mafic; they form minerals rich in magnesium and iron, such as pyroxene.

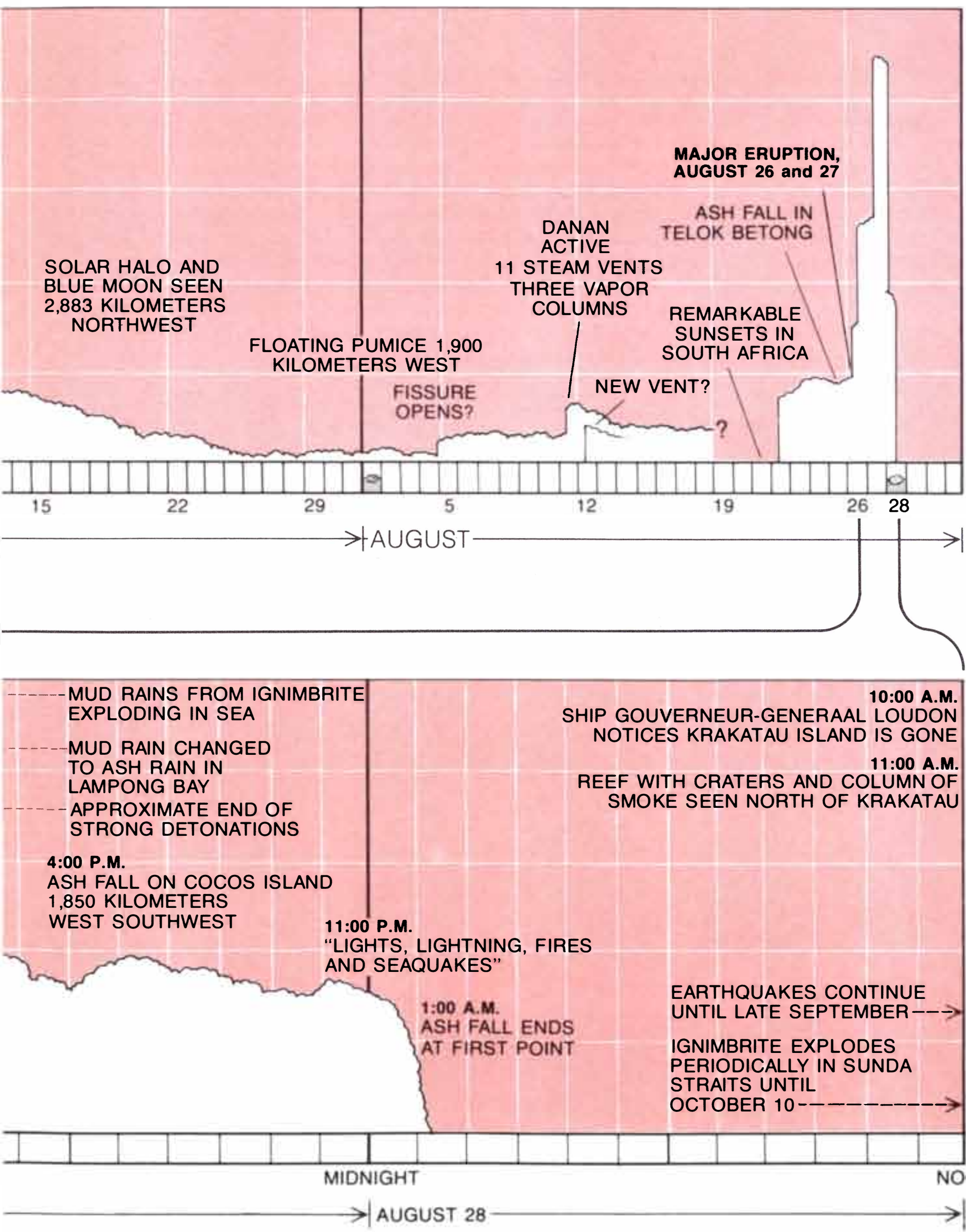
Most of the Krakatau deposits consist of pale dacitic pumice that contains relatively few mineral crystals. Some of the deposits are strikingly different, however; streaks or bands of darker glass are intermixed with the lighter glass, and some clasts are entirely dark. The darker pumice also includes crystals with compositions different from those in the lighter pumice. Such mixed pumices are by no means rare in pyroclastic deposits. It has been suggested that they may be formed when hot basaltic magma intrudes into light rhyolitic or dacitic magma and the two mix.

The intrusion of fresh hot basaltic magma into the base of a chamber of dacitic magma may cause a violent convective overturn within the chamber. Silicic magmas tend to be less dense than more mafic magmas, largely because of the differences in their composition. Most dacitic magma chambers are thought to be stratified according to composition, with the most silicic magma at the top and the somewhat denser, more mafic material lower down. Such a system is stable. Stephen Sparks and Herbert Huppert of the University of Cambridge have suggested, however, that a pulse of hot basaltic magma that intrudes into the chamber may initiate convective overturn by superheating the



ERUPTION SEQUENCE OF KRAKATAU has been pieced together from many sources: diaries and notebooks kept by Dutch administrators and others living on the neighboring islands, logs kept by the officers on watch on ships in the Sunda Straits, events recorded by a pressure gauge at a gasworks in Jakarta and by tide gauges along the Sunda Straits and occasional observations made by people adjacent layer, causing it to become less dense than the overlying silicic magma. A second possibility is that the hot basaltic magma initially rests quietly on the floor of the magma chamber and violent overturn takes place only when it cools. As the basaltic magma cools, volatiles come out of solution and crystals settle out, decreasing the density of the remaining liquid magma until it is lower than that of the overlying layers.

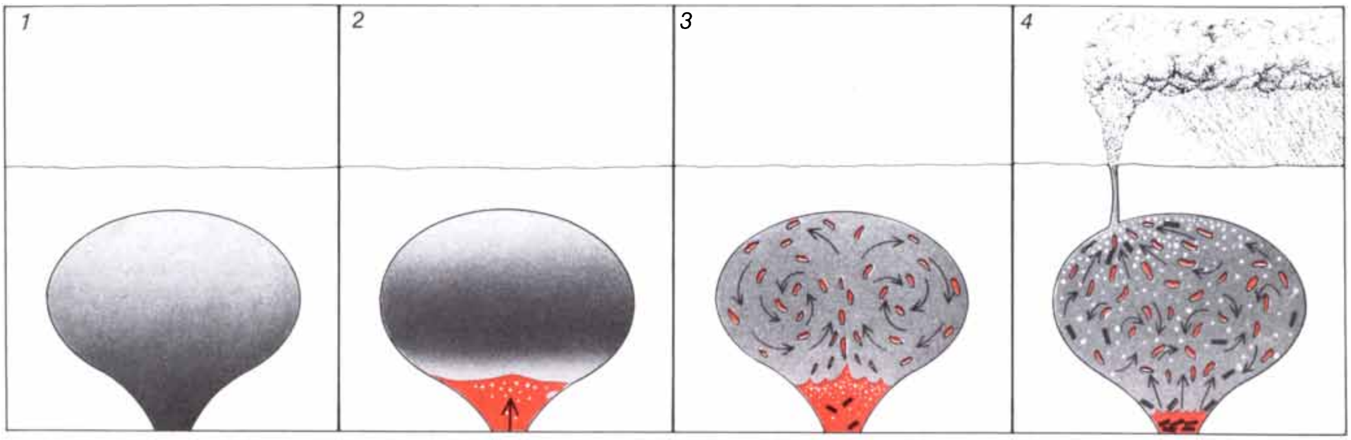
Why should convective overturn initiate an explosion? In general the eruption of basaltic magmas tends to be less explosive than that of more silicic magmas because they are less viscous and the bubbles of gas created by volatiles coming out of solution easily percolate through or effuse out of them, whereas the bubbles formed in more viscous silicic magmas are trapped there and cause explosive fragmentation of the magma when their internal pressure exceeds its strength. In an impressive series of laboratory experiments, however, J. Stewart Turner of the Research School of Earth Sciences of the Australian National University and Sparks and Huppert have demonstrated the importance of the volatile content of the newly introduced magma to the process of magma-chamber mixing. Fresh basaltic magma may be rich in volatiles kept in solution by the pressure at the depth at which the magma formed. Convective overturn carries the volatile-rich material to shallower levels in the chamber, where the ambient pressure is lower. As the magma rises, the volatiles rapidly come out of solution, the pressure with-



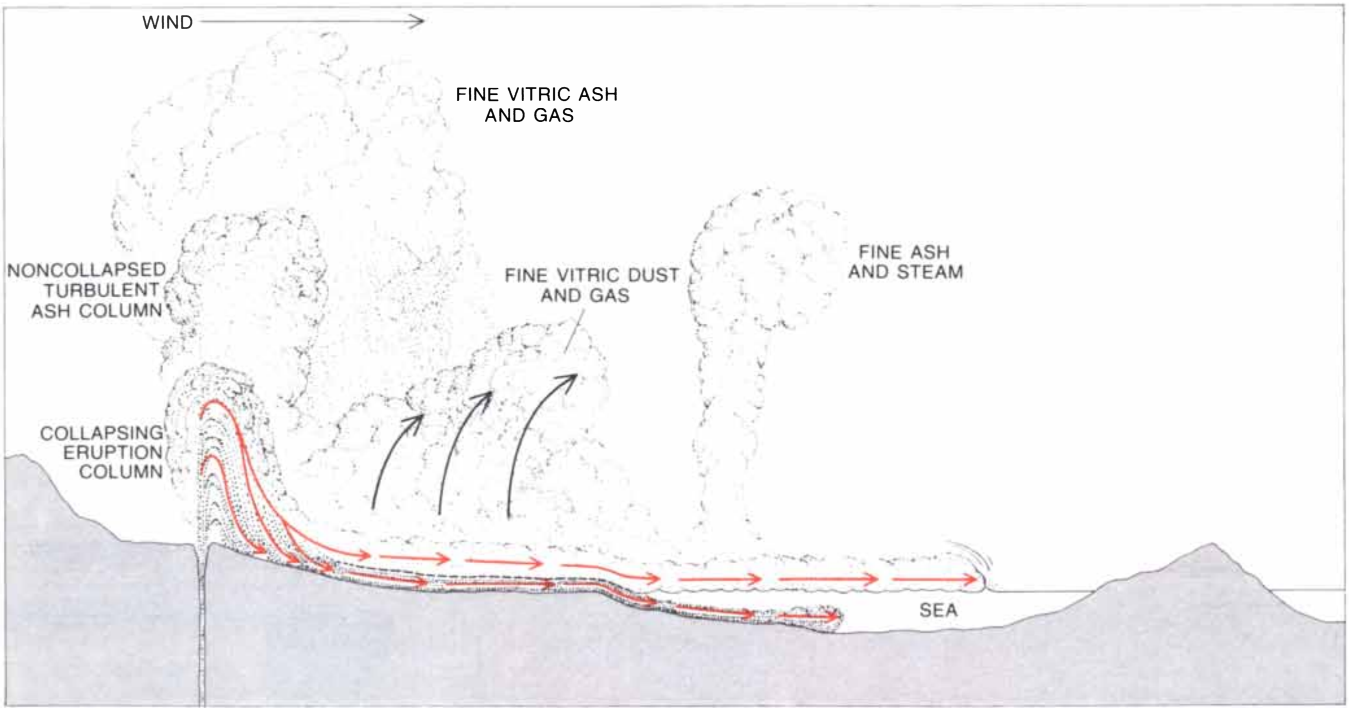
as far away as South Africa. Numbers in colored boxes give the relative magnitude of the tsunamis. The relative magnitude of the explosions is indicated by the height of the arrows. Only a few of the many explosions that occurred in the course of the eruption sequence are shown.

in the magma chamber increases and an explosive eruption is likely to follow.

An important piece of historical evidence supporting the hypothesis that the eruption of Krakatau was caused by magma mixing came to light during the preparation of this article. On May 27, 1883, soon after the first explosions of Perbuwatan in May, a party of no fewer than 86, including the Dutch mining engineer J. Schuurman, visited Krakatau. Schuurman wrote a detailed account of his visit, noting that 60 centimeters of dark "ash" overlay 30 centimeters of lighter-colored "pumice." He also collected samples of the ejecta, which were later analyzed. The samples of pumice are similar in composition to the dacitic pumice ejected in the August explosions. The gray ash, however, is basaltic in composition. It seems likely, therefore,



VIOLENT MIXING OF MAGMAS of differing composition may have triggered the eruption of Krakatau. Magma ranges in composition from silica-poor basalt through andesite and dacite to silica-rich rhyolite. The less silica there is in the magma, the denser it tends to be. The mixture of dacitic magmas in the chamber of Krakatau may originally have been stratified according to density and therefore been stable (*1*). A pulse of fresh basaltic magma, hotter than the material already in the chamber and also rich in volatile substances, may have intruded into the bottom of the chamber, and at first it may have rested quietly on the chamber floor (*2*). The hot magma would superheat the layer of dacitic magma immediately above it, however, making that layer less dense than the overlying layers. At the same time, as the basaltic magma began to cool, crystals would settle out and volatiles would begin to come out of solution, making the basaltic magma itself less dense than the lower layers of dacitic magma. Either or both of these processes would cause violent convective overturning of the material in the chamber (*3*). As the volatile-rich basaltic magma rose to shallower levels of reduced ambient pressure, dissolved volatiles would rapidly come out of solution. The pressure within the chamber might therefore increase to the point where the magma would force its way to the surface and erupt (*4*). Mixed pumices found at Krakatau suggest this mechanism caused the eruption.



PYROCLASTIC FLOWS are masses of incandescent volcanic material fluidized (given a density and a viscosity lower than those of an aggregation of solid material) by hot gases. As the flow travels it segregates into a dense lower zone in which the flow is generally laminar and a light upper zone in which the flow is turbulent. When the flow reaches the sea, the denser material may plunge into the water, but the upper part of the flow may temporarily have a density less than that of seawater and therefore may travel on the surface. As the fluidization dissipates, the remaining material becomes denser and more of it enters the sea. The process continues until only a turbulent cloud of ash and steam is left. The hot material can cause secondary explosions as it sinks into the sea. It has been argued that the grading process that takes place within the flow itself and the winnowing of the ash by the secondary explosions produce a fine, crystal-poor ash that is distributed over a wide area by the secondary explosions. Rogier D. M. Verbeek reported finding deposits of this type near Krakatau.

VISTA

THE IMPORT WAGON WITH FLEXIBILITY AND ROOM FOR SEVEN.

"Now that's a purpose!"

Purpose: to satisfy the ever-changing transportation needs of people on the move.

Car in point: Vista, a new kind of wagon, imported for Dodge and Plymouth, built by Mitsubishi Motors Corp.

Vista is the wagon reinvented for today. Vista seats 7 passengers and those seats flip up or down, to handle just about any combination of people and things.

36 EST. HWY **28** EPA EST. MPG*

* Use EPA EST. MPG for comparison. Your mileage may vary. Hwy. mileage probably lower. Calif. est. lower.

Vista is quick and agile, easy to park. It has front-wheel drive for sure-footed handling in all kinds of weather or road conditions. It also gets excellent gas mileage, and has a very low base sticker price—yet Vista is not a bare-bones purely functional economy wagon at all. Vista's interior boasts cloth and vinyl seats and heat ducts for rear-seat passengers. And Vista comes with power brakes and electric rear window defroster.

But above all, you just have to love a wagon with all this room and flexibility. Vista. See it at your Dodge or Plymouth dealer.

$8,115

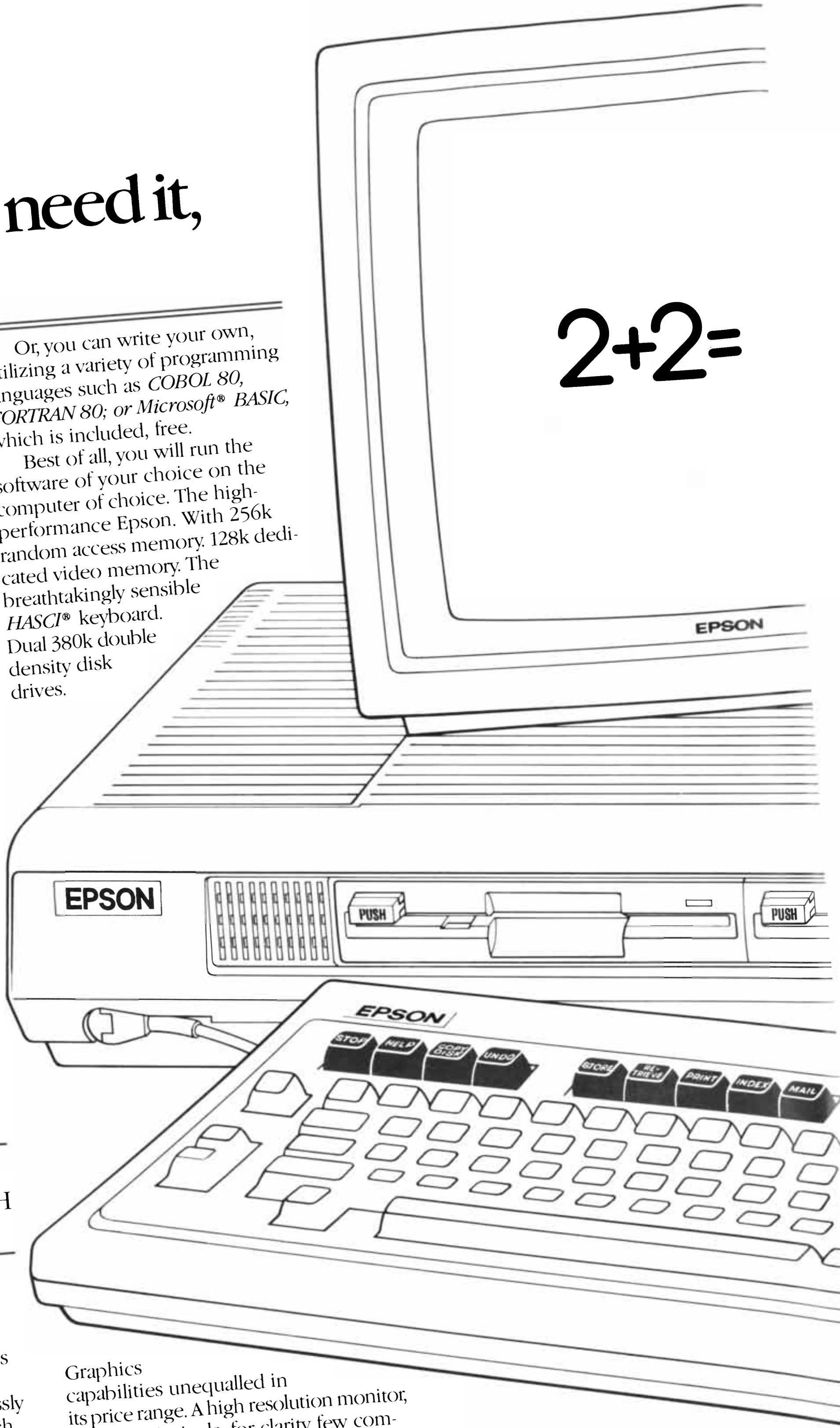
Base Sticker Price excluding title, taxes and destination charges as of October 15. Price subject to change.

IMPORTS WITH A PURPOSE.

Dodge Plymouth

BUCKLE UP FOR SAFETY

Epson. For those who need it, simplicity.



One computer.
Two points of view.
The Epson QX-10 personal computer.

To many, the Epson represents the ultimate in simplicity.

Just press a single key for the function you require: word processing, scheduling, business graphics, address book or file management. One keystroke produces your program. There are no rigamaroles to remember. No disks to change.

The result: you start to work immediately. And you start being productive, immediately. With step-by-step prompts. In plain English, not computerese.

Simplicity itself.

Or is it?

The plain fact is that the ease of operation the Epson offers today is accomplished with a degree of technological sophistication most other computers can only promise for tomorrow – specifically, fully integrated software, operating in an interactive environment.

The few other computers offering such "simplicity" cost $5,000 to $15,000 more. And most other computers can't offer it at any price. Which makes one wonder exactly what they do offer, in terms of either simplicity, or performance.

HOW MUCH CAN YOU DO ON THE EPSON? HOW MUCH ARE YOU READY TO DO?

The Epson's ease of operation may spoil you, but it certainly won't limit you.

Case in point: every Epson comes complete with an integrated software system – *Valdocs®* – to effortlessly provide the basic functions for which most people buy computers. But it is important to note that the Epson also comes with *CP/M®*-80 2.2. So you can choose from hundreds of programs: *dBase II,® Wordstar,® Personal Pearl,® Supercalc,® Microplan,® MBA General Ledger,® Multiplan,® Peachtext,® Spellwriter,® and Dataview,®* just to name a few.

Or, you can write your own, utilizing a variety of programming languages such as *COBOL 80, FORTRAN 80; or Microsoft® BASIC,* which is included, free.

Best of all, you will run the software of your choice on the computer of choice. The high-performance Epson. With 256k random access memory. 128k dedicated video memory. The breathtakingly sensible *HASCI®* keyboard. Dual 380k double density disk drives.

Graphics capabilities unequalled in its price range. A high resolution monitor, 640 by 400 pixels, for clarity few computers in any price range can offer. Plus, an RS-232C interface, a parallel printer interface, and internal space for up to five peripheral cards so you can expand your Epson as your needs require.

One further point: all these features, and quite a few more, are included in the Epson's $2,995 price. Some com-

Epson.
For those who can handle it, performance.

puter companies ask you to pay extra for features like these. Most can not offer them at any price.

That, too, is performance. The kind of performance that can make choosing a personal computer very simple, indeed.

EPSON QUALITY. OR, WHY WONDER WHAT TERRIBLE KLUDGES LURK IN WHICH SLEEK BOXES.

If you know computers, you know Epson. Epson printers set the industry standard for quality, reliability and value. Rest assured, the same can be said for the Epson personal computer.

The satisfying silence of the slim, Epson-designed disk-drives is one way for you to judge or, for an inside-out perspective, here is an excerpt from a review by Jim Hanson in the April, 1983 issue of *Microcomputing*.*

"The Epson QX-10 is soundly designed and executed. I looked hard and found no evidence of kludging or shorting out anything in the name of economy. All the connectors have gold on them and are of quality manufacture. The printed circuit boards are heavy, with soldermarks on both sides of double-sided boards. The circuit boards are completely silk-screened with component labels, and the layout is as professional and clean as you will find anywhere."

Isn't this what you expect? After all, it's an Epson.

A WORD TO THE WISE: GET YOUR HANDS ON THE EPSON.

Is the Epson a simple, easy-to-use computer for beginners? Or a sophisticated high-performance computer for the experienced? The answer is "yes." And when you think about it, aren't those two computers the one you need now.

For technical specifications, and the complete, 3-part *Microcomputing* review, along with the name of your nearby Epson dealer, call toll-free (800) 421-5426. California residents, call (213) 539-9140.

*Excerpt reprinted by permission of Microcomputing Magazine. All rights reserved.
CP/M is a trademark of Digital Research, Inc.
Valdocs and HASCI are trademarks of Rising Star Industries.
dBase II is a trademark of Ashton-Tate.
Wordstar is a trademark of Micropro.
Personal Pearl is a trademark of Pearlsoft.
Spellwriter and Supercalc are trademarks of Sorcim.
Microplan is a trademark of Chang Labs.
MBA General Ledger is a trademark of Micro Business Applications.
COBOL 80, FORTRAN 80, Microsoft, and Multiplan are trademarks of Microsoft.
Peachtext is a trademark of Peachtree Software.

EPSON

STATE-OF-THE-ART...SIMPLICITY.

that a small amount of a basaltic magma was ejected in the early phase of the eruption, soon after an initial ejection of dacitic magma. This indicates that magmas of sharply contrasting composition underlay the volcano in May and lends further support to the hypothesis that the August explosions were caused by magma mixing following the intrusion of fresh pulses of basaltic magma.

Although it is impossible to prove that magma mixing actually caused the eruption of Krakatau, the laboratory experiments of Sparks and Huppert convincingly demonstrate that the mechanism is plausible. It has been proposed to account for other major eruptions, such as the 1875 eruption of Askja in Iceland and the 1902 eruption of Santa María in Guatemala, sites where similar mixed pumices have also been found. The magma-mixing hypothesis for the eruption of Krakatau nonetheless leaves one problem unsolved. Even if it is assumed that magma mixing initiated the eruption, it is still not clear why some of the explosions in the sequence were so extraordinarily violent.

The next question we address is why the eruption of Krakatau was accompanied by many separate explosions. One possibility is that some of the explosions were secondary ones caused by hot pyroclastic flows entering the sea. Little is known about what happens when a large pyroclastic flow comes in contact with seawater, because few examples have been documented. Does the flow displace the water and travel along the sea floor or does it travel over the surface? The answer seems to be a little of both, although exactly what happens probably depends on the characteristics of the flow.

As the flow enters the sea the denser underflow, consisting largely of pumice, fine-grained glassy material and fragments of older, nonvolcanic rock, may plunge into the water. The more diffuse turbulent upper layer, consisting largely of pumice and ash fluidized by hot gases, may temporarily have a bulk density less than that of seawater and may travel over the surface. As turbulence becomes less effective, the larger clasts begin to drop out and eventually the flow is depleted of solid material. Only a turbulent mass of gas and fine particles remains; such flows, still hot, probably hit Kalimbang on Lampong Bay.

A pyroclastic flow deposited 6,000 years ago by the eruption of the volcano Koya on the Japanese island of Kyushu provides evidence that the interaction of hot ignimbrite with seawater may in some circumstances be as powerfully explosive as one might intuitively expect. On the basis of studies of these deposits Walker suggests that the explosions caused by flows entering the sea may lead to ashfall over many thousands of square kilometers. The distinctive feature of these ash deposits is that their thickness and their grain size vary little with distance from the source, a distribution that suggests the ash was generated by explosions of remarkable power.

The Rotoehu deposits erupted from the Okataina volcanic center in New Zealand some 50,000 years ago yield evidence that such ash deposits may also be crystal-poor. Walker argues that the crystals become separated from their parent pumice fragments by a natural grading process. As the flow moves toward the sea the denser crystals sink and the pumice dust is lofted. Further winnowing of the crystals from the dust accompanies the explosions that occur when the flow enters the sea, as powerful turbulent vortexes sifting through the cloud of ash created by the explosions carry away the lighter dust.

At this remove in time it is, of course, difficult to determine exactly what happened when the Krakatau pyroclastic flows entered the sea. Contemporary illustrations, however, clearly show large circular craters on Calmeyer, one of the two islands created by the deposition of ignimbrites. The craters strikingly resemble those made by the secondary explosions that occurred when pyroclastic flows erupted by Mount St. Helens in May, 1980, entered Spirit Lake. Moreover, Verbeek noted that the uppermost layers of ash on the island of Sebesi were extremely fine-grained and consisted almost entirely of glass. This fine crystal-poor ash may well have been the fallout from powerful secondary explosions. For these reasons we suggest that although the gigantic blasts at Krakatau up to 9:58 A.M. on August 27 were probably caused by the expulsion of the pulses of magma that generated the pyroclastic flows, at least some of the many powerful explosions in the eruption sequence were explosions caused by pyroclastic flows entering the sea.

It was the great tsunamis set in motion by the eruption that wreaked the most havoc. Even though the tsunamis have been studied in detail by many investigators, their cause or causes remain a matter of debate. Tsunamis are generally caused by the sudden vertical movement of the sea floor, usually as the result of an earthquake. As late as the 1960's it was argued that the Krakatau tsunamis were caused by the similar mechanism of caldera formation. Discussion of this point has centered on the question of when and how quickly the roof of the magma chamber collapsed and on the question of whether the tsunamis were preceded by a lowering of the sea level, such as would be expected when water rushed in to fill the newly formed caldera. The question of the exact timing of the caldera collapse has not been settled, but it seems to have

What if you chose as a technical

"At Union Carbide, an HP computer network automates production management and has paid for itself in less than two years."

Union Carbide's Linde Division in Tonawanda, N.Y., manufactures custom production equipment. Factory systems manager Doyce Coffman says, "When production demand fluctuates, it complicates our scheduling, product costing, and long range planning.

"To solve these problems we automated production and materials management with a network of HP 1000 and 3000 computers and 43 data capture terminals.

"Because these systems continuously monitor operations and respond in real time, we can identify and eliminate production bottlenecks on line. The payback? Improved efficiency factory-wide and a dollar savings that covered the cost of our entire HP network in less than two years."

What should you expect from a working partnership with Hewlett-Packard?

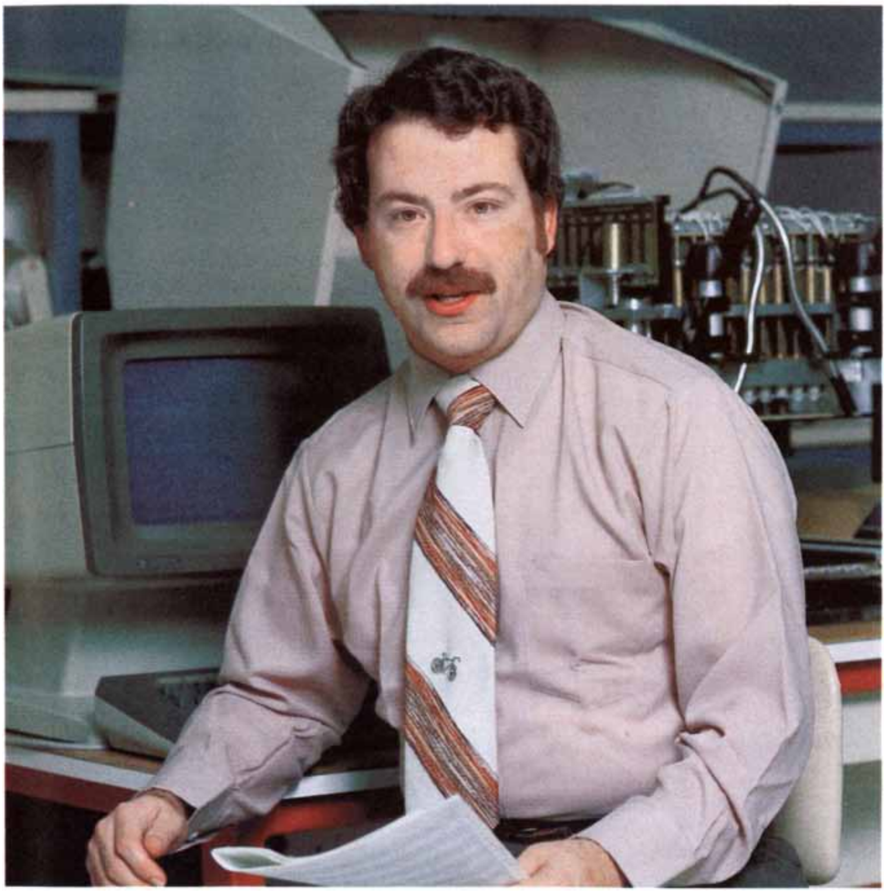
Results with assurance! DataPro Research Corporation's annual survey of computer users has ranked Hewlett-Packard *number one* in Customer Support Satisfaction for three years in a row. Why? Because the same quality and reliability engineered into HP products is built into our customer support programs.

From the outset, a Local Support Team of specialists who understand your needs will help you develop a cost-effective strategy to meet your goals—from selecting the right equipment to training your people.

You can tie your HP system into one of our Technical Centers where HP specialists will monitor your system's performance. Our routine, remote diagnostic tests can help iden-

The HP circle of quality customer support

Hewlett-Packard computer partner?



"At Amdahl, an HP computer network helps control testing. And it's paid for itself in two years."

Amdahl manufactures large-scale general purpose computers. In Sunnyvale, Calif., Bob Henderson, Manager of Test Systems Software, says: "We use a network of HP 1000 computers to control various tests on our Model 580 computers. With the HP systems, we can achieve the advanced testing and stringent quality control our high-speed, high-density products demand.

"Replacing batch systems with an HP computer network saved us nearly $100,000 in equipment costs. And by using HP 1000's we're obtaining test data ten times faster than before.

"Our test network helps streamline manufacturing processes and has paid for itself in just 2 years. In addition, the network saves us up to $34,000 a year by reducing paperwork and test cycle time."

tify and prevent problems before they occur.

The reliability designed into our systems and maintained by our worldwide support network enables us to offer Guaranteed Uptime Service—guaranteeing that covered systems will be up and running at least 99 percent of the time. **Results with assurance:** that's what a working partnership with Hewlett-Packard is all about.

You should consider a working partnership with HP. Now.

For a free copy of our brochure explaining HP's practical, proven approach to meeting your long-term computer and information needs, write to A.P. Oliverio, Senior Vice President, Marketing, Dept. 207, Hewlett-Packard Company, P.O. Box 10301, Palo Alto, CA 94303.

When performance must be measured by results

hp HEWLETT PACKARD

0002343

TK!SO
EXPLA

TK!Solver does for equations what word processing did for words. The first thing you should know about the TK!Solver™ program is that it is not a spreadsheet. Instead, it does something completely unheard of (until now)—it turns your personal computer into a voracious equation processor.

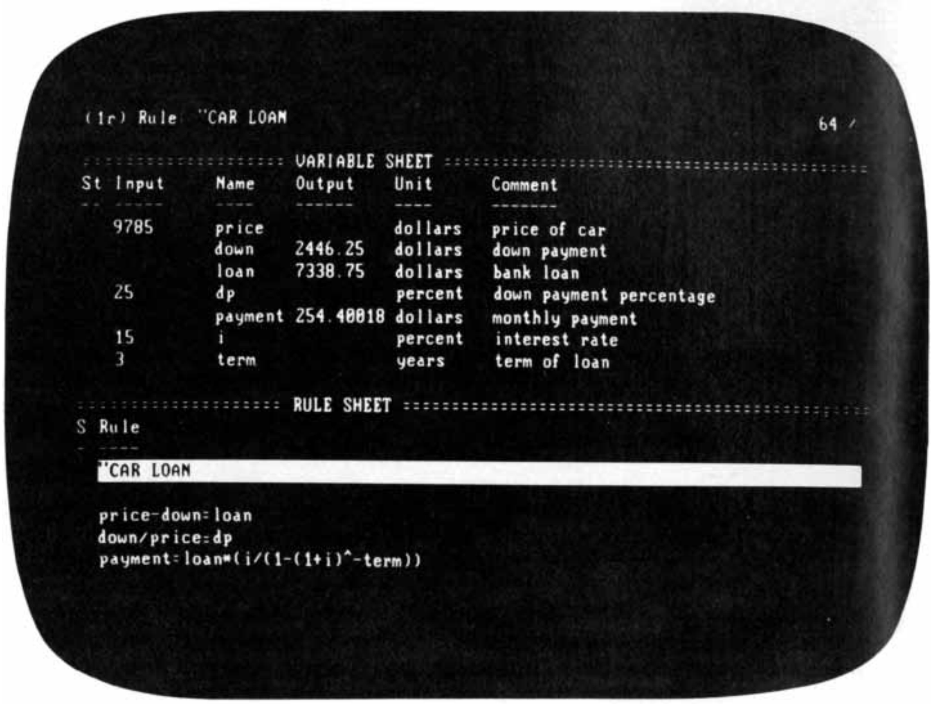
The next thing you should know is that if the TK!Solver program can't make life with your personal computer easier (and pay for itself), even if you use it only 15 minutes a week, you are a very rare person.

And finally, you should know exactly what equation processing is, and how it works. If you keep reading this, you will.

Equation processing with TK!Solver, or problem solving made easy. The best way to understand what the TK!Solver program *is,* is to understand what it *does.* The following simple example is designed to do just that. If you're still a little in the dark after reading it, stop in at your local computer store for a very enlightening hands-on demonstration.

Begin by setting up your problem. The TK!Solver program lets you do it quickly, easily, and naturally. For example, a car costs $9785. What would be the monthly payment on a three-year loan if the down payment is 25% and the interest rate is 15%?

STEP 1. Formulate the necessary equations to solve your problem and enter them on the "Rule Sheet" simply by typing them in (as in the screen photo). For example: "price—down = loan."



STEP 2. Enter your known values the same way on the "Variable Sheet." For example: "9785" for price. You may also enter units and comments, if you want.*

STEP 3. Type the action command ("!" on your keyboard) to solve the problem.

STEP 4. TK!Solver displays the answer: the monthly payment is $254.40.

Backsolving, the heart of TK!Solver. Now that you've defined the problem and solved it, TK!Solver's unique backsolving ability also lets you think "backwards" to solve for any variable, regardless of its position in the equation. For example, if you can only afford a monthly payment of $200, you can re-solve the problem in terms of that constraint. The TK!Solver program will solve the problem, displaying your choice of a higher down payment, a longer loan term, or a lesser interest rate. This unique backsolving capability forms the basis of TK!Solver's remarkably flexible problem-solving ability.

OLVER™

AINED:

Also, as you can see from the example on the screen, TK!Solver deals not only with single variables, but with entire equations and sets of simultaneous equations. It also deals with much more complicated problems than this one. How complicated? That's up to you. What kinds of problems? That's up to you, too, but popular applications include finance, engineering, science, design, and education.

Other extremely useful and interesting things TK!Solver does. Aside from its basic problem-solving abilities, the TK!Solver program performs a number of pretty fancy tricks. Like: *Iterative Solving;* in which TK!Solver performs successive approximations of an answer when confronted with equations that cannot be solved directly,(like exp (x) = 2 – x • y and sin (x • y)= 3 – x – y). Like: *List Solving;* in which TK!Solver attacks complete lists of input values and solves them all, allowing you to examine numerous alternative solutions, and pick the one you like best. Like: *Tables and Graphs;* using the values you produced with the List Solver, the TK!Solver program will automatically produce tables and graphs of your data. You can look at your formatted output on the screen or send it to your printer with a single keystroke. And like: *Automatic Unit Conversion;* in which TK!Solver lets you formulate problems in one unit of measurement, and display answers in another. Very convenient what with all this talk about going metric.

The TK!Solver program also provides a wide variety of specialized business and mathematical functions like trig and log and net present value.

Then, there's TK!Solver's on-screen Help facility that provides information on commands and features any time you want it. Just type "?" and a topic name.

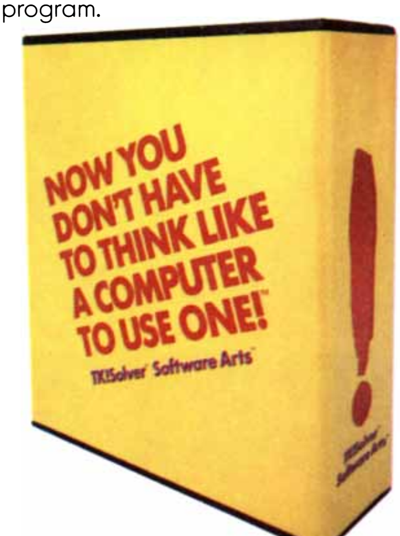
And of course the TK!Solver program combines all these features in one *integrated program.*

TK!SolverPacks make problem-solving a picnic. TK!SolverPack™ application packages are specially developed by experts in specific fields. Each package contains a diskette with about a dozen models that include the necessary equations, values, and tables for solving a particular problem. The models are usable as-is or you can easily modify them.

TK!SolverPack application packages available from Software Arts include Financial Management, Mechanical Engineering, Building Design and Construction, and Introductory Science. Additional TK!SolverPacks are on the way from Software Arts, McGraw-Hill,™ and others.

We know you're out there. No matter who you are, or what you do, if it involves using equations, the TK!Solver program is an indispensable* tool for you.

So, visit your local computer store today, and see TK!Solver in action. You'll be amazed at how much faster and more effectively you'll be able to work when you discover the power of equation processing with the TK!Solver program.



Software Arts™
The inventors of VisiCalc®

27 Mica Lane, Wellesley, Massachusetts 02181
617/237-4000

*You can easily define appropriate unit conversions on the unit sheet.

P/N 100-092 P 8/83

SEE ALL THE LUXURY VAN CONVERSIONS AT YOUR FORD DEALER.

Your Ford Dealer's got it—the widest choice of vans for premium conversions.

Ford provides you with America's best-built van, complete with rugged body-on-frame construction plus exclusive outfront design.

Ford gives you a wide choice of gas or diesel power for great towing capacity. Then you and your Ford Dealer select the level of luxury you've been dreaming of!

Van modifications not by Ford. Ford assumes no responsibility for modification, availability, price, quality, safety, durability.

Get it together—Buckle up.

Quality is Job 1.

Ford's commitment to quality results in the best-built American trucks, based on a survey of owner-reported problems during the first three months of ownership of 1983 trucks. And the commitment continues in 1984.

BUILT FORD TOUGH

Ford

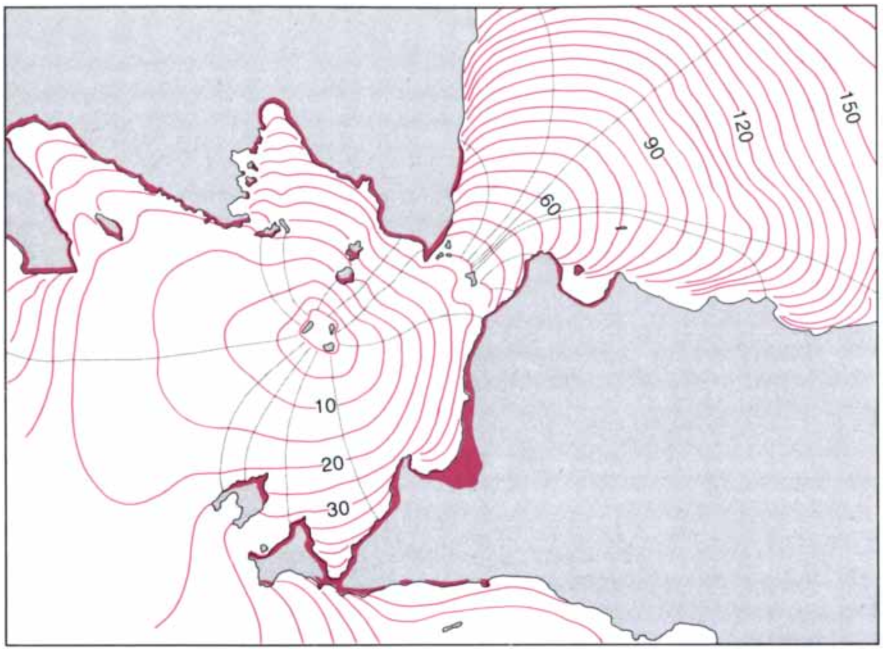
come late in the sequence, perhaps even after many major tsunamis. The evidence on the second question is also somewhat equivocal. At many points along the Sunda Straits the arrival of most of the tsunamis was marked by a rise in the sea level. A few retreats were also apparently recorded, however, and most of the tide gauges were so far away from Krakatau that an initial retreat may have been too small to register on them.

There are three alternative causes for the tsunamis. Some workers have suggested that submarine eruptions, which like a depth charge would create an outward-propagating water dome, may have caused some of the tsunamis. Verbeek suggested in 1884 that the largest tsunami was set in motion by the slumping of the northern half of the volcanic cone Rakata into the newly formed caldera. Verbeek also suggested that the sudden displacement of water by volcanic ejecta "falling" into the sea could have set the waves in motion. Given that many cubic kilometers of material entered the sea in the form of pyroclastic flows, this possibility seems to us the likeliest one.

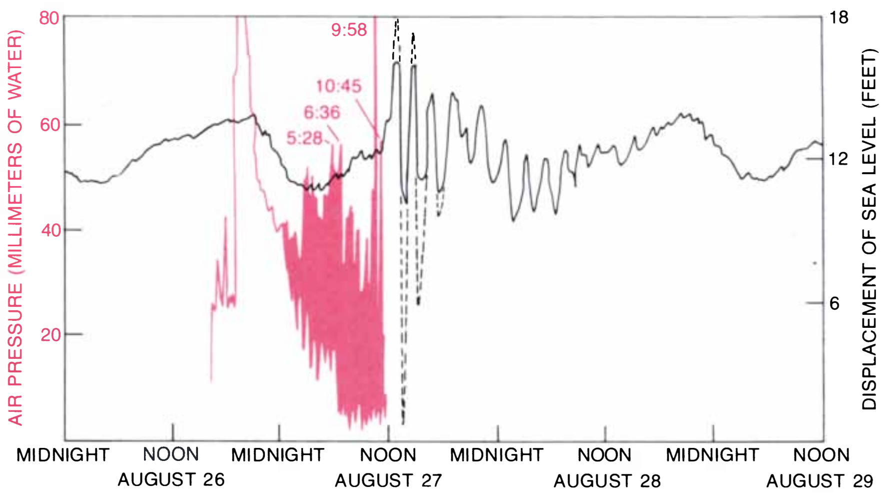
The correlation of the time of arrival of the waves at points along the Sunda Straits with events in the eruption sequence is critical to arguments about the origin of the sea waves. Establishing the chronology has been complicated by two problems. First, not all the tsunamis were true tsunamis, that is, waves that had traveled entirely by water. Second, it has proved difficult to synchronize the eruption sequence with the timing of the atmospheric disturbances recorded by the gasworks pressure gauge at Jakarta and sea waves recorded by the tide gauges along the straits.

Many of the sea waves, particularly those recorded at points far from the volcano, appear to have traveled extremely fast: at the speed of air waves rather than at speeds typical of sea waves. The velocity at which a wave moves through water is proportional to the depth of the water. More precisely, the velocity of the wave is equal to the square root of the acceleration of gravity multiplied by the depth of the water. The waves from Krakatau reached Honolulu in only 11 hours, a travel time that implies an average water depth of 17 kilometers. The average depth of the deep ocean is much less: about four kilometers. Furthermore, sea waves attributed to the eruption were observed in parts of the oceans where they could not reasonably be expected to appear; for example, waves were detected on the far side of island chains, a barrier through which they could not possibly have propagated.

These discrepancies came to light soon after the eruption but for many years were unexplained. In 1955, after microbarographs capable of registering small atmospheric pressure variations were developed, Maurice Ewing of the Lamont-Doherty Geological Observatory and Frank Press of the Massachusetts Institute of Technology were able to demonstrate that seismic waves through the solid earth can be coupled to the atmosphere. In 1967 David G. Harkrider of Brown University and Press showed that the long trains of pressure pulses set up in the atmosphere by nuclear explosions transfer some of



MAP OF TSUNAMI TRAVEL TIME is based on the times of arrival of the tsunamis recorded by tide gauges at various points along the Sunda Straits. The arrival times of the tsunamis were distinguished from those of the trains of secondary waves set up when the tsunamis entered the bays where most of the tide gauges were installed. Then the tsunamis were traced backward in time; their speed at each point in time was determined from the depth of the water through which they were then moving. Contour lines (*colored lines*) were derived from the paths and travel times of a number of tsunamis (*gray lines*), making it possible to find the travel time of a tsunami from Krakatau to any point on the coast. The inundated areas are in color.



PRESSURE AND TIDE GAUGES at Jakarta recorded the major explosions and the ensuing tsunamis on August 27, 1883. A pressure gauge at the gasworks (*color*) fortuitously recorded the arrival of the atmopsheric pressure waves generated by the explosions. The broad peak in the trace on August 26 is unrelated to the eruption; it may have been caused by an increase in the pressure of the gas in the tank the gauge was monitoring. The tide gauge at Tanjong Priok (Jakarta's harbor) (*black*) recorded the arrival of the sea waves set in motion by the same explosions. Some of the waves were so high that they exceeded the range of the gauge. The broken-line peaks are estimates of the displacement in sea level based on various observations, such as the number of steps in a flight of stairs leading down to the harbor that were covered by water. The highest peak corresponds to the tsunami that arrived at Jakarta at 12:16 P.M. The peaks following it include secondary waves set up in the harbor and the straits by the tsunami.

Who on earth will install enough lightguide fiber to circle the earth 10 times?

Western Electric, an AT&T company.

In just five years, Western Electric has already created and installed enough lightguide fiber to circle the earth. That's 25,000 miles of fiber in 60 lightwave communications routes.

Those totals should increase at least tenfold by the year 2000, as Western Electric continues to install lightwave systems to transmit the high-speed voice and data services of the Information Age.

Far more efficient and economical than copper cable, lightwave communications speeds your messages over 90 million light pulses per second through hair-thin glass fibers. A single lightguide cable can carry 240,000 telephone calls at once.

Right now Western Electric is installing the world's largest lightwave system in the busy Northeast Corridor. Another will transmit the entire 1984 Los Angeles Summer Olympics to ABC television studios.

And our research and development arm, Bell Labs, is working on a transatlantic lightwave cable that will triple calling capacity and enhance transmission quality.

These systems are forerunners of a continually expanding lightwave network that one day will literally circle the earth, bringing everyone exciting new information services.

Western Electric, an AT&T company. Applying the technologies of lightwave, microelectronics, and software to make the dream of the Information Age a reality.

AT&T

their energy to the ocean, causing sea waves in regions far from the explosion.

The coupling mechanism is complex; the amount of energy transferred depends strongly on resonance, that is, on the natural vibrational frequencies of the atmosphere and the ocean. Nevertheless, the general principle that governs the interaction can be stated simply: The ocean can be thought of as responding to alternating increases and decreases in atmospheric pressure in such a way as to maintain hydrostatic equilibrium. Increased atmospheric pressure causes the ocean to become shallower and decreased pressure causes it to become deeper, so that the weight of the ocean-atmosphere column above each unit area remains as nearly constant as possible. Harkrider and Press recognized the analogy between a nuclear explosion and a volcanic one and suggested that the same kind of phenomenon might explain the anomalous tsunamis that followed the eruption of Krakatau.

Much patient detective work was still needed to determine what caused the true tsunamis that inundated the coastlines along the Sunda Straits. The arrival of sea waves recorded by tide gauges at various points along the straits had to be distinguished from the arrival of secondary waves set up by the tsunamis within the enclosed bodies of water, such as bays, where most of the tide gauges were installed. Then the tsunamis had to be traced to their point of origin, taking into account refraction, or changes in velocity, caused by the varying depth of water in the straits. In 1981, I. Yokoyama of the Usu Volcano Observatory in Japan analyzed the refraction of the Krakatau waves and published a map of the travel time of the tsunamis. On the basis of the map he concluded that some of the minor tsunamis were probably caused by volcanic ejecta entering the sea but that the largest tsunami, which reached Jakarta at 12:16 P.M. on August 27, must have been caused by an underwater eruption.

John H. Latter of the New Zealand Department of Scientific and Industrial Research came to somewhat different conclusions in a paper published the same year. Latter set out to establish the chronology of events as accurately as possible from the records kept by both the pressure gauge at the Jakarta gasworks and the tide gauge in Jakarta's harbor. He calculated the times of explosions at Krakatau from the times of arrival of atmospheric pressure waves at Jakarta, taking into account the atmospheric travel time of eight minutes and the five-minute difference between Jakarta time and Krakatau time. He then showed on the basis of the timings of the explosions and the map of tsunami travel times worked out by Yokoyama, according to which the travel time for a tsunami traveling from Krakatau to Jakarta was two hours 25 minutes, that the chronometer on the tide gauge at Jakarta was inaccurate by three and a half minutes with respect to the pressure gauge. Having established the chronology of events, Latter was able to show that the arrival times of atmospheric waves and sea waves at Jakarta caused by some of the minor events of the eruption sequence were correlated.

In several instances, however, the sea wave could not be correlated with a matching atmospheric wave and thus with the explosive event that caused both. The largest atmospheric wave, caused by the massive explosion heard as far away as the Indian Ocean, reached Jakarta at 10:08 A.M. If the atmospheric travel time and other factors are taken into account, it must have been caused by an explosion taking place at Krakatau at 9:58 A.M. A large sea wave was recorded on the tide gauge at Jakarta at almost exactly the same time as the atmospheric pressure wave was recorded at the gasworks. Given that the travel time of a sea wave from Krakatau to Jakarta is approximately two hours 25

ANAK KRAKATAU ("Child of Krakatau") is a new volcanic cone that is building up in the caldera left by the 1883 eruption. It is at roughly the same site where Danan once stood. The new cone broke the surface of the sea late in January, 1928. Successive eruptions of Anak Krakatau have since largely filled the caldera left by the 1883 eruption with debris. The photograph was made by Maurice Krafft.

minutes and that Latter could find no evidence for an event taking place at Krakatau at about 7:40 A.M., he concluded that the large sea wave was caused by the coupling of the atmospheric wave to the water. The true tsunami caused by the explosion at 9:58 A.M., which had a much larger amplitude than the false tsunami, did not arrive at Jakarta until 12:16 P.M.

At this point the details of the timing become all-important. Given the sea-wave travel time, the large tsunami that arrived at 12:16 should have originated at Krakatau at about 9:45 A.M., that is, before the major explosion took place. Latter was confident of the accuracy of his meticulously checked timings, and he was certain there must be a causal link between the largest air and sea waves, and so he concluded that the tsunami originated not at Krakatau itself but rather at a point closer to Jakarta's harbor by a distance equivalent to the time discrepancy. He then used Yokoyama's map to show that the point of origin of the tsunami set in motion by the great explosion at 9:58 A.M. was 10 to 15 kilometers from Krakatau. This point corresponds roughly to the outermost edge of the newly formed island Calmeyer. Similar analyses of the timing of paired explosions and sea waves led Latter to conclude that at least three of the Krakatau tsunamis were probably caused by massive pyroclastic flows advancing into the sea, such as the flow that formed Calmeyer.

Although the Krakatau tsunamis are the largest tsunamis known to have been generated by pyroclastic flows, the same mechanism has been shown to be responsible for tsunamis following the eruption of other volcanoes. In 1980 Juergen Kienle and Samuel E. Swanson of the U.S. Geological Survey reported that the nine-meter tsunami that swept through English Bay in Alaska after the eruption of the volcano Augustine in 1976 was set in motion by a pyroclastic flow entering the bay.

Since the eruption of 1883 a new volcano, Anak Krakatau ("Child of Krakatau"), has been building up at a site roughly corresponding to the position of the old volcanic cone Danan. The new volcano was initially under water, but the cone emerged late in January, 1928. Since then successive eruptions have gradually filled the northern sector of the 1883 caldera with lava and pyroclastic flows. The focus of volcanic activity is shifting southward and may well be following the same fissure that determined the alignment of the volcanic cones on the original island. A large island, aligned like the old one northeast to southwest, may eventually develop. It seems that a new Krakatau is rising like the phoenix from the ashes of its own destruction.

THE AMATEUR SCIENTIST

Looking into the ways of water striders, the insects that walk (and run) on water

by Jearl Walker

To me water striders are among the most fascinating of all insects. Members of the order Hemiptera, they live on the surface of the water in a slowly moving stream. Some 75 to 85 species are found in North America, and hundreds more, including many that are exclusively oceanic, are known in other parts of the world. I have been studying some of the long-legged types (of the family Gerridae), and their way of life is a marvel.

Water striders dash, leap and scramble across the water if they are disturbed, moving at a rate of about a meter per second. At other times they may move quite slowly. They do not swim. Instead they glide over the water surface by pushing horizontally against the water, much as a human sprinter pushes against starting blocks. They locate one another by means of the waves they generate as they move. They also seem to locate objects in their environment by means of waves.

Freshwater striders spend the winter in hiding under rocks or vegetation or lying on the bottom of a stream or pond. In the spring they become active, move to the surface and mate. Eggs are left by the female on submerged objects. After two weeks the nymphs hatch and move to the surface. Maturing to the adult stage takes a little more than a month.

The water striders I studied are skittish insects that apparently can see well enough to detect my approach. When I waded through streams, they were obviously also scared off by the waves I made. I learned to move slowly and then stand still for a long time so that they would eventually ignore me.

I caught several striders in a small, fine-mesh butterfly net. The long handle enabled me to avoid tipping off my presence. It was hopeless to try to chase them; they could easily outrun me in the water. When I caught one, I dropped it into a small glass jar. With this last I had to be quick; the insect could jump several centimeters in the air and out of the jar before I got the lid on.

I believe the striders I saw were all of the species *Gerris remigis.* Like all freshwater striders they had some kind of wing growth along their back. (Marine water striders are wingless.) The specimens I caught ranged in length from less than half a centimeter to about 1.5 centimeters. The immature ones were only a few millimeters long.

I rarely saw any striders along the edges of lakes, ponds or streams more than a few meters wide. On smaller streams the striders were absent from turbulent water and from places where algae covered the surface. They were also absent from regions with sparse vegetation. Areas with slowly moving water and some emergent vegetation were heavily populated with the insects. The water they seemed to like best was only a few centimeters deep. Presumably the striders congregate there because the water is too shallow for any fish large enough to eat them.

The striders were easy to detect. As soon as I waded into an appropriate habitat the water surface became a frenzy of small ripples as they scurried off. On my first explorations I found dozens of areas covered with small water striders. Then in a stream I happened on a pool busy with large ones.

For hours I sat on the shore or squatted in the water to watch their movements. Later I caught several of them and released them into a plastic container filled with water. If I took care to move slowly to avoid scaring them, they would stay on the water surface, dashing about, and would not attempt to climb over the edge.

I lay next to the container with a 20-power magnifying lens to examine how the insects manage to stand and to move on the water. Patience was called for. I waited until an insect moved into the focal plane of the lens. Then I could follow it for a short time before it became alarmed and skated off.

To photograph striders on streams I put a telephoto lens on my 35-millimeter camera, a single-lens reflex model. When I had an insect on the water in my plastic container, I hovered patiently and motionlessly with a close-up lens on the camera.

If you intend to photograph these insects in nature, you would do well to eliminate some of the light reflected from the water surface. Otherwise the photographs will be dominated by images of the surrounding trees. You can diminish the reflections by taping a polarizing filter over the lens. Mount the filter so that its axis is vertical.

All water striders have three pairs of legs. The front legs are usually short; they help to support the insect. On the striders I caught the other legs are much longer than the body. The two middle legs are primarily responsible for propulsion. The two rear legs may also aid in propulsion, but more often they serve for steering as the insect glides on the surface.

Each leg consists of several segments. From the body outward the segments are the coxa, the trochanter, the femur, the tibia and the tarsus. The tarsus itself may consist of several segments.

The long-legged water striders are distinguished by the position of a claw on the last tarsal segment, just above the end of the leg. The location of the claw may help the insect when it stands on the water, taking advantage of surface tension.

A stationary water strider rests its weight on all six legs. On the front and middle legs only the tarsus touches the water. On the rear legs the tibia and the tarsus touch. These sections are not submerged but lie in shallow depressions in the water surface. The depressions for the front legs are small. Those for the other legs are elongated because more of the leg is in contact with the water.

When I am close to a water strider with the sunlight at the right angle, the depressions are easily spotted. At other times they are apparent in the shadow the insect casts on the bottom of the stream or the container. There I see a slim shadow for the body itself and more prominent dark ovals at the end of the thin shadow of each leg. The ovals result from the bending of light rays by the curving of the water surface at the depressions.

With a flat water surface the rays of sunlight illuminating a given region would all refract into the water in the same way. Thus they would evenly illuminate the bottom of a stream or a container. When instead the rays pass through the depression surrounding a tarsus of the insect, they are sent off to the sides. The shadow cast on the bottom is therefore considerably larger than the tarsus. These large shadows proved to be useful. When the insect moved a leg, I had better luck following the oval than the leg.

Often the striders glide slowly, presumably searching for food. At other times they dart over the water. The motion is always in a straight line. At the end of a glide the insect stops, repositions one middle leg or both of them to reorient its body and then propels itself forward again.

Although I can follow the slow movements, I cannot see the details of the fast ones. In a slow movement the insect propels itself by moving its middle legs toward the rear. The rear legs are nearly stationary but might move slightly rearward too. The front legs appear to serve only as a support.

The faster movements have been recorded in high-speed motion pictures. Here the legs serve basically the same functions but the rearward motion of the middle legs lasts only about 20 milliseconds. The acceleration is large, about 10 times the acceleration of gravity. The front legs are momentarily lifted from the water surface, as is the upper end of the tibia on each rear leg. After the acceleration both sets of legs are returned to the water surface and the insect glides. The rear legs then serve as stabilizers to hold the glide straight. Eventually the kinetic energy of the movement is dissipated and the insect skids to a stop. The energy is lost to the waves generated by the movement and to the friction between the insect and the water surface.

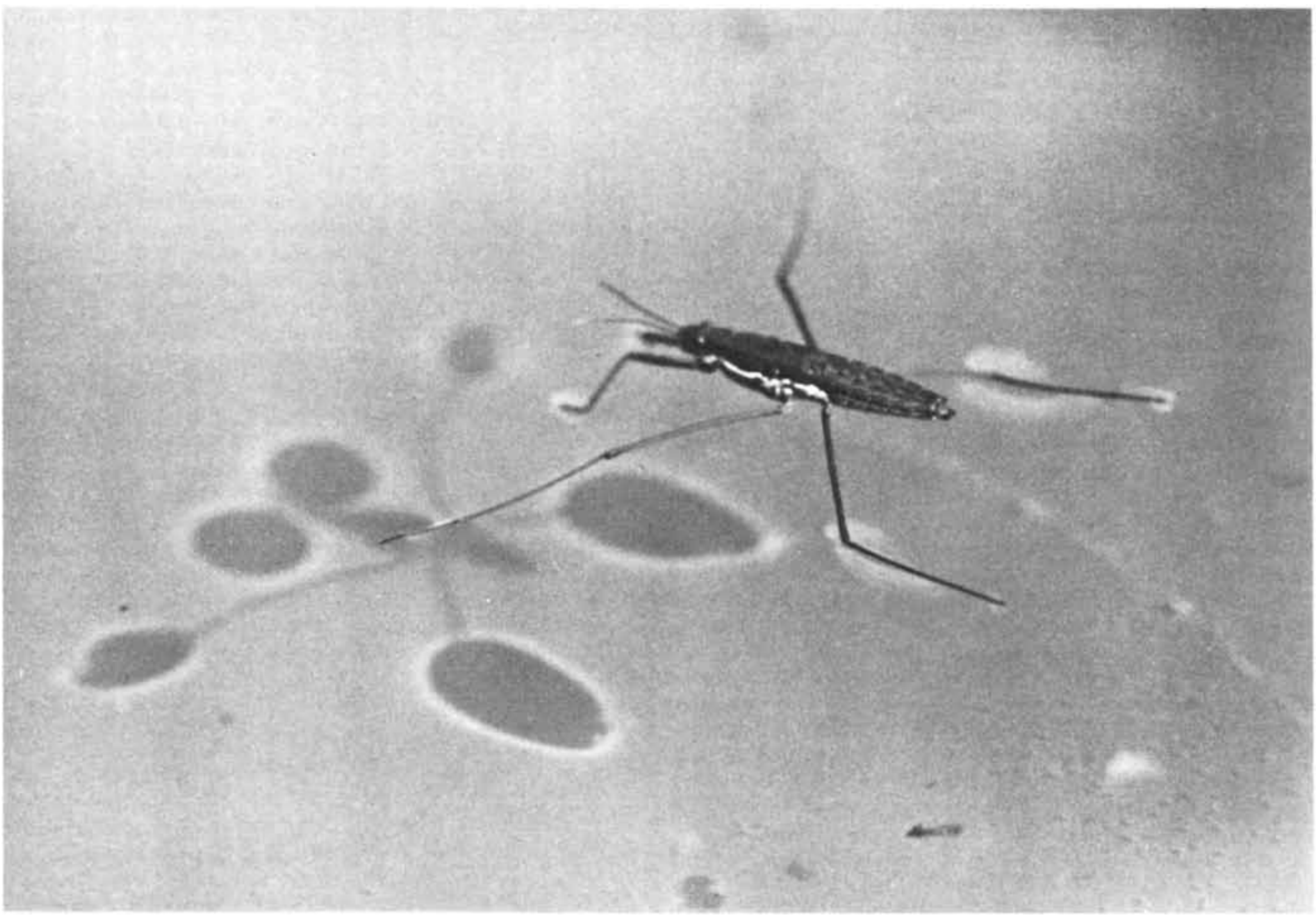
The propulsion from the middle legs involves several rotations. The tarsus and the tibia rotate about the joint between the tibia and the femur faster than the femur rotates about its attachment to the trochanter. These two rotations push the tarsus against the rear wall of the depression in which it initially rests. The resistance of the water to the push supplies the force of propulsion.

In this motion one can see two advantages of the long middle legs of the insect. First, the length provides a long lever arm for pushing against the water surface. A human oarsman likewise finds that a long oar makes for easier rowing.

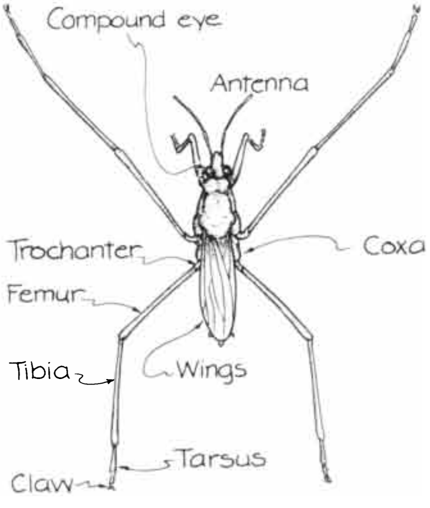
The second advantage is that the long tarsus provides more friction with the water, increasing the effectiveness of a push of the leg backward. The front legs have shorter tarsi so that the friction there is minimized, since friction shortens a glide. Some species of water strider have fans deployed from the tarsi of the middle legs to aid the push against the water. Some have projections from the legs that effectively grab the water surface, as metal studs on a snow tire bite into the snow for traction.

Much of the support for the insect as it rests or glides arises from the surface tension of the water. Surface tension is the name given to the cohesion of the water at the surface. In textbooks this phenomenon is described in terms of the electrical attraction between the water molecules. Consider one molecule that lies on the surface of pure water. Electrical forces pull it horizontally toward its neighbors, but all those forces balance out by symmetry. The molecule is also pulled downward by the neighbors just below it, but it cannot enter the bulk water because of collisions with those underlying molecules, all of which are in motion. Thus the surface is said to be in a state of tension because every molecule on it is subject to a net downward electrical force.

A second way of examining the surface is in terms of energy. Suppose the surface is to be expanded. If more surface is to be gained, molecules must be brought there from the bulk water. Such a molecule initially has no net force acting on it because on the average the forces from its neighbors balance out. At the surface, however, a net downward force develops. The movement



A water strider and its shadow



An adult strider (**Gerris remigis**)

from the bulk to the surface therefore requires energy from whatever is causing the expansion of the surface. In this description the surface is said to be in a state of tension because the water attempts to decrease its energy by contracting its surface.

The surface tension of water can be surprisingly strong. If a sewing needle is laid horizontally and carefully on water, the surface tension can hold the needle in place. The support comes from the depression of the surface produced by the weight of the needle. At first the needle starts to sink, curving the water surface, but then the surface begins to put a net upward force on the needle. Part of the upward force is due to buoyancy (because the needle has displaced some of the water). The rest of the upward force comes from the tension along the curved surface of the depression. When the demonstration is done right, the needle receives enough upward force to balance its weight, and so it floats.

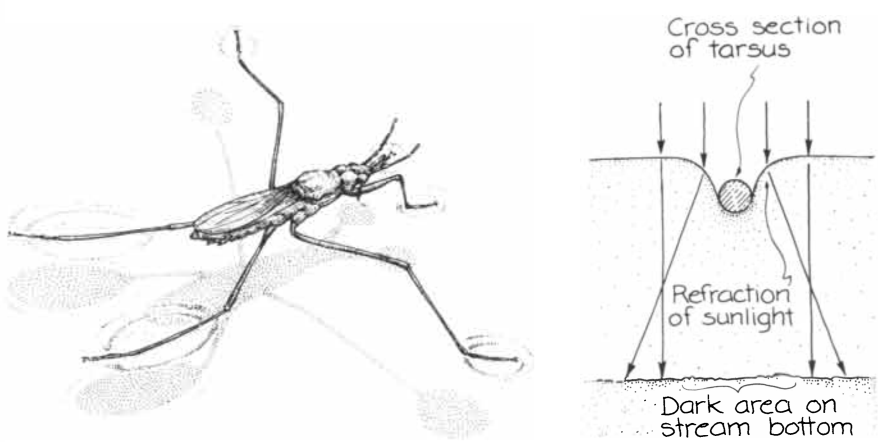
Similar support is given to an insect resting its weight on the tarsi of its legs. In addition the insect has evolved other support mechanisms. Closely spaced fine hairs on the legs and body of many species of water strider trap a layer of air. The hair, legs and body may also be coated with a waxy material to resist wetting.

The layer of trapped air provides additional buoyancy. I demonstrated the presence of air by dunking a water strider. Underwater the air on its surface shimmered in the sunlight. When I released the insect, it quickly regained the surface and scampered off, obviously still dry.

Another source of resistance to wetting has recently been discovered in the oceanic water striders. They need it because they are more exposed to rain and splashing and rarely find any floating object they can climb on to dry off. They differ from the freshwater striders in the structure of the microtrichia, the microscopic fixed hairs on the surface of the body. In the freshwater species the microtrichia are shaped like pegs and jut outward from the surface of the body. In the oceanic species they are shaped like mushrooms, presumably to aid in trapping tiny air bubbles next to the body for additional buoyancy.

Water striders do sometimes become wet, which suggests that the waxy covering or the air trapped in the hairs can fail. I observed several water striders that seemed to sink progressively into the water. The sinking strider would climb onto a rock or a lily pad to dry off.

As I was trying to get close-up photographs of one strider I provoked it into skittering across the water in my plastic container for a long time. As it began to sink it moved to the side of the container, hoisted one leg onto the plastic and hung there for a while. If it did not find the side of the container, it groomed itself dry by rubbing one front leg along the length of the adjacent middle leg, both of which were lifted out of the water. Special hairs on the tibia of the front leg serve for removing water from the other legs in such circumstances.



How the pattern of shadows from a strider is formed

The buoyancy, surface tension and resistance to wetting provide strong support for a strider when its legs are dry. The insect not only can stand on the water with all six legs but also can shift its weight to four legs while grooming. I often saw striders leaping several centimeters into the air. Although the downward push required for this maneuver created waves in the water, the insects did not break through the surface. Sometimes I saw two striders tangle in what appeared to be a fight. It always ended with one insect or both leaping into the air and then running off. Even in these leaps the insects did not break the water surface.

There is more to surface tension than I have said. Perfectly clean water is never found in nature. Even when water is thoroughly cleaned in the laboratory, its surface will be coated with a monolayer of other molecules after a few minutes of exposure to the air. Water found in nature certainly has such a top coating, which reduces the surface tension.

I believe another factor, surface viscosity, aids in the propulsion of a water strider. When the middle legs are brought rapidly to the rear, they push against the back wall of the depression they create. Because of surface viscosity the leg does not simply slide up and out of the depression in the course of the motion. It gets to give a good push. The viscosity is associated with the resistance of the top layer of molecules (the contaminating layer) to sliding over the layer of water slightly lower.

I once watched a water strider that was having difficulty in moving across a stretch of water. The water was almost stagnant and in one area was covered with a thick layer of scum that was obviously highly viscous. When I chased a strider into this area, it could no longer push off from the water and then glide. Its motion was more of a forward leap: it kicked off with its middle legs and flew through the air for several centimeters before it landed, immediately coming to a stop. After I chased it back into cleaner water it was able to glide again.

When a water strider glides, the fluid surface passing under its supporting tarsi must flow quickly into the curved shapes of the depressions. In the scummy region the high viscosity prevented this flow, but it also provided a firm support for the insect's spectacular leap.

A water strider is so agile that a female can move even while mating. One afternoon I watched a pair of the insects mating for almost an hour. They stopped only when I captured them. The male had mounted the female from the rear with his weight resting on the female and on his own rear legs. The male's front legs were around the female's body just forward of her middle legs, and his middle legs were swiveled upward. The arrangement left the female's legs entirely unhindered. As they mated the male was motionless. The female, however, continued to move slowly over the surface. She appeared to be incapable of rapid gliding but could still move in glides of a centimeter or so.

I had supposed the ability of a water strider to stand on the water and glide over it depended critically on the wide separation of the places where the legs touched the water. At each such place the curvature of the surface in the depression around the tarsus provides an upward force on the insect. What would happen if these several depressions were brought close to one another? The surface between nearby tarsi would be flatter and would provide less upward force. Hence I supposed that a water strider would be unlikely to bring its legs close together.

My supposition was wrong. Clearly the propulsion achieved by an insect demands that the middle legs be brought close to the rear legs. For an instant as a water strider is accelerating the front legs are even lifted off the water, requiring that the middle and rear legs fully support the insect. The tarsi of those legs are then quite close to each other. Yet the insect does not break through the surface.

The mating pair I monitored had most of their weight resting on the water directly behind the female. They may have been low in the water, but they were apparently in no danger of breaking through the surface layer even when the female brought her middle legs rearward for propulsion. Furthermore, the water striders of the family Velidae have short legs, and so their support regions are closer. The insects of this family are just as agile as the long-legged striders I studied.

The wave patterns the striders create when they are moving are striking. A simple push and glide builds up a circular group of waves that expand from the site of the push. The waves arise during the acceleration stage, when the insect forces each middle leg against the rear wall of the depression in which it rests.

The full rearward motion of each leg sends out a semicircle of waves to one side. Since the two legs are synchronized, the two sets of waves merge at the front and the back of the insect. The merger is nearly perfect. When I am close to a strider, however, I can see two wakes left in the group of waves expanding to the rear. I believe they arise from the rear legs.

It is difficult to analyze the production and propagation of these waves. The waves are somehow generated when a disturbance, such as the rearward push of the strider's legs, upsets the initially flat surface of a pool of water. Two forces, operating to restore the surface flatness, are responsible for the propagation of the waves away from the site of the disturbance.

When the wavelength of the wave is short (thus when the shape of the water surface is highly curved), surface tension is the important force. The waves are called capillary waves. For long wavelengths gravity is important, and the waves are said to be gravity waves. With intermediate situations one must consider both forces. In no instance is a wave traveling over water a simple sine wave, notwithstanding what is said in many textbooks.

The structure traveling over the surface is considered mathematically to be a superposition of many sine waves of differing wavelength. In the mathematical model they are called phase waves. Since the speed of each wave depends on its length, the waves do not travel in lockstep; hence they end up interfering with one another.

The result is often an impressive pattern. A disturbance such as an insect's kick sends out a short series of phase waves. Together they form what is called a wave group, which spreads out from the insect. On each side of the group the water is flat. Within the group some five crests can be seen. The distance between the crests is approximately the average wavelength of the phase waves making up the wave group.

The speed of the movement of the crests over the water is termed the phase speed. It differs from the speed of the wave group as a whole. When gravity dominates the propagation of the waves, the phase speed is larger than the group speed. Therefore the crests appear at the the rear of the wave group, overtake the middle and disappear at the front of the group. Since their amplitude is greatest at the middle of the group, one can follow the change in the amplitude of a crest as it traverses a wave group.

The waves generated by a water strider are capillary waves, because their wavelengths are short enough for the waves to be dominated by surface tension as the restoring force. With capillary waves the phase speed is less than the group speed. Hence a crest in an expanding wave group from a water strider appears at the front of the group, grows in amplitude as the center of the group overtakes it, decreases in amplitude as the rear of the group reaches it and finally disappears. At any instant the wave group has approximately five crests, separated by a millimeter or so.

This wave group expands roughly as a circle until it runs into some obstacle or its energy is dissipated. After gliding the insect may generate another such pattern of capillary waves. If the insect's speed relative to the water is less than .23 centimeter per second, no waves are generated. That is the minimum speed at

which waves can propagate over the surface of water. Insects moving at less than this speed leave the water undisturbed except for the wave group initiated by their propulsion.

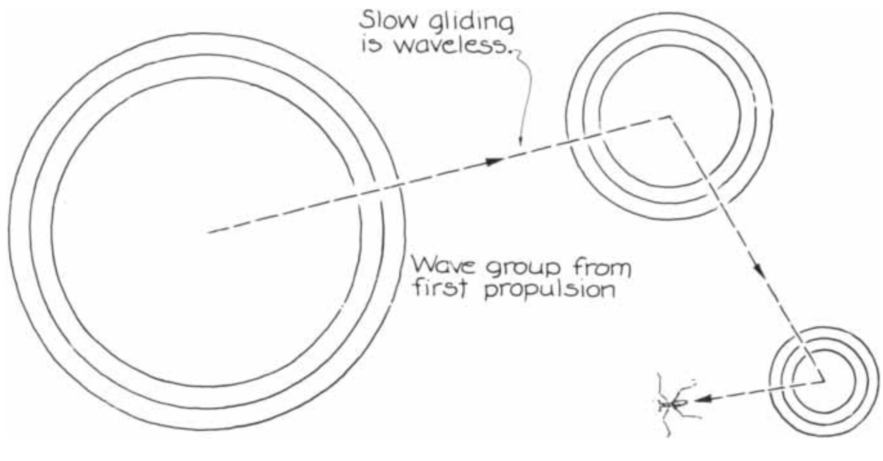
If the insect glides faster than the minimum wave speed, it continuously generates waves. The pattern that results has two curious features. One feature is that the waves in front of the insect are more closely spaced than the ones in the rear. The second feature is that the wave pattern takes the shape of a *V*, with the apex at the insect's head.

In 1883 Lord Rayleigh demonstrated mathematically that a small object moving in water with a relative speed in excess of .23 centimeter per second sends out waves shorter in length toward the front than at the rear. He was not concerned with wave patterns from insects but with the pattern produced by a fishing line in a moving stream. The line served as an obstacle to the flow of water. (Although the line was stationary and the insect moves, the difference is irrelevant. It is the relative speed of the object and the water that matters.) The generation of waves by the obstacle builds up closely spaced wave crests upstream of the obstacle. They are parts of capillary waves. Downstream the perturbation of the water by the obstacle sends out widely spaced wave crests. They are parts of gravity waves.

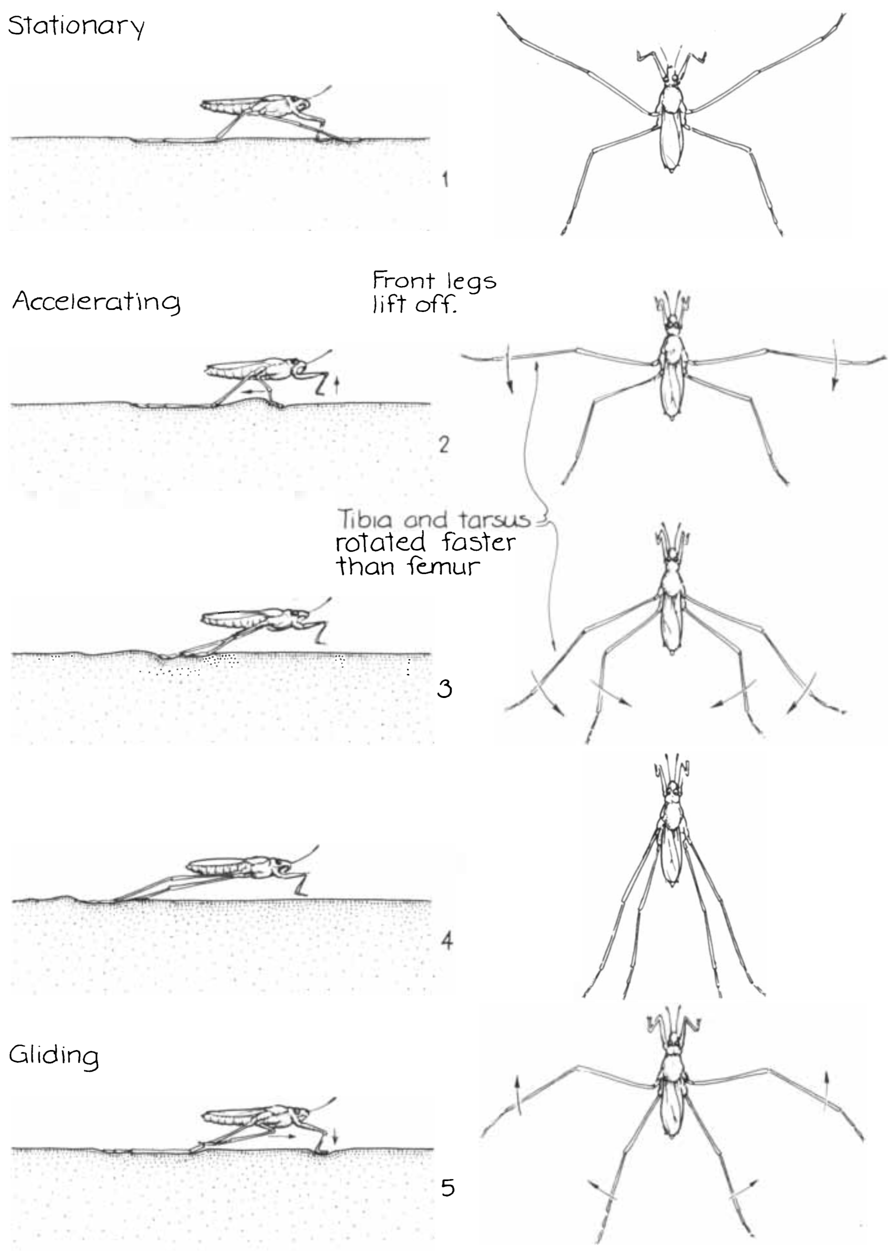
In the *V* wave pattern around a water strider the capillary waves are outside the *V*, the gravity waves inside. The angle of the *V* reflects a mathematical relation of twice the minimum wave speed of .23 centimeter per second divided by the relative speed of the obstacle and the water. A water strider in a faster glide ends up generating a narrower *V* wake.

In 1972 R. Stimson Wilcox of the Australian National University demonstrated that the surface waves generated by water striders play an important role in their mating. He studied the genus *Rhagadotarsus* Breddin by capturing several specimens and installing them in a large tray of water. The male initiated mating by generating wave signals. It gripped a support in the water with its front legs while its middle legs (and possibly the rear ones too) created the waves.

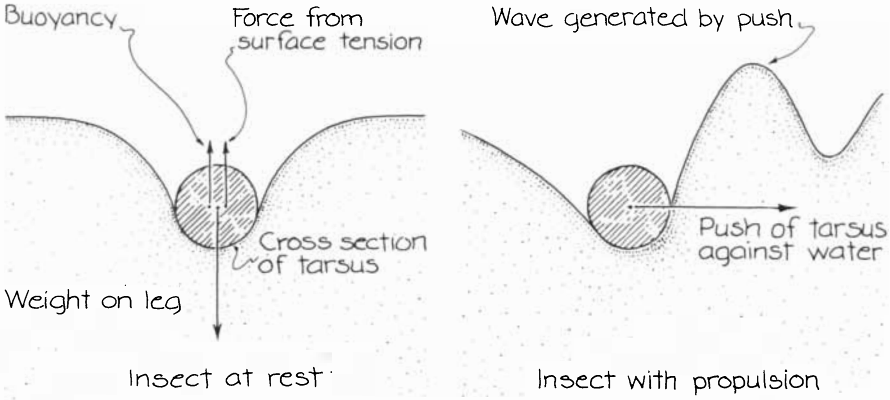
Stimson discovered three main types of signal preceding copulation, all with a frequency of between 17 and 29 waves per second. The male could call for a female by sending out a group of from seven to 15 waves at high amplitude. Alternatively it could transmit a group of only two or three waves in what Stimson interpreted as more a signal for courtship. If a female approached, the signal switched rapidly to a strictly courtship mode of 30 waves at low amplitude. When the female was within a few centimeters of the male, the female



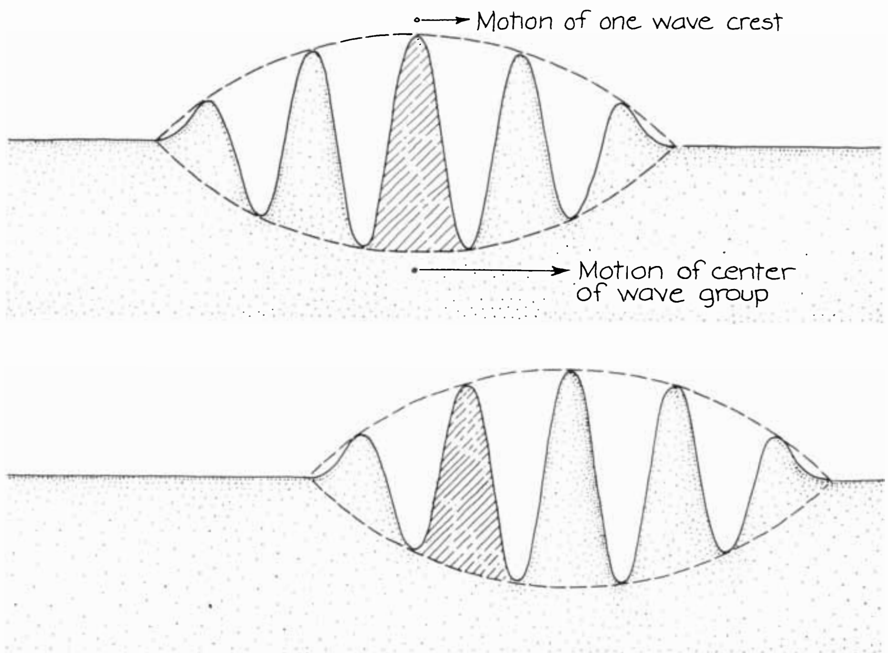
The skitter path of a waterborne insect



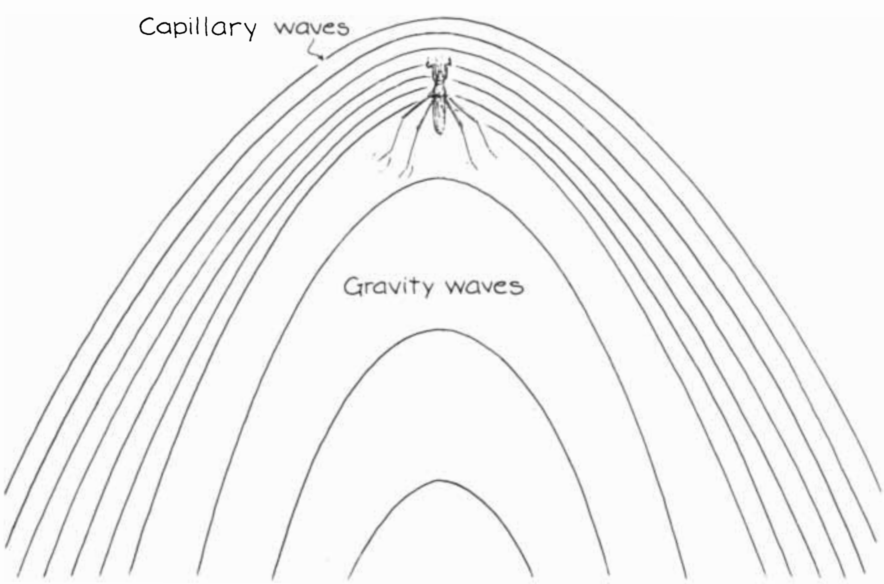
Stages in the movement of a strider



The interaction of a strider's leg with the water surface



The motion of a capillary-wave group



The pattern of waves around a rapidly moving strider

too would send such a signal. Copulation then began.

A male sometimes generated another signal of higher frequency in an apparent warning to other males to avoid the mating area. If two males closed on each other, they would fight, perhaps for some minutes. I saw several such fights as water striders circled and then leaped at each other, after which one would chase the other.

On all my trips up and down streams I found water striders apparently locating one another by means of the waves they generated. When a wave group reached a water strider, the insect paused and then oriented itself perpendicular to the wave crests and toward the source of the crests. Then it dashed a short distance toward the source. After waiting for a fresh set of waves to pass, it would again dash off in the direction of the source. I cannot believe the insect located the source entirely by sight; if it did, the occasional pauses would be unnecessary.

Striders also employ waves to find prey. When a fly fell into the water and flapped violently on the surface, a water strider located it by means of the resulting waves. Sight might have played a role when the strider got near the fly, but not before.

Although the strider's exploitation of waves seemed obvious to me, I wanted a better test. I tried oscillating a twig in a stream, but only one strider came to investigate. It might have arrived there even if I had not been dabbling in the water. I tried a vibrator sold for massage. The oscillating part, which is fully encased in rubber, is powered through a wire running to two small batteries. (A vibrator run on household current would invite electrocution.) A switch on the battery holder varied the frequency of oscillation.

I waded into the center of a pool of water about six meters wide along a stream. Several water striders collected at the edges of the pool after sensing my arrival. I submerged part of the vibrator and set its oscillation at about 20 hertz. The striders immediately paused in their motion, turned toward me and within about three seconds raced to within a few centimeters of the vibrator. Neither the vibrator nor I looked like a water strider, and so the insects surely were attracted by the waves.

Insects of many other types live on or near the water surface in a stream. You might like to search for whirligig beetles. The capillary waves they generate were studied by Vance A. Tucker of Duke University, whose work provided the basis for my discussion of the waves from a water strider. Another interesting aquatic subject is the back swimmer. This insect travels upside down and just below the surface of the water by rowing with its legs.

HOW TO REACH THE PEOPLE WHO MAKE THE FUTURE HAPPEN IN SPAIN

INVESTIGACION Y CIENCIA is the Spanish edition of SCIENTIFIC AMERICAN which attracts a young audience of affluent professionals who have come to the top because of their technical expertise. Almost half of them hold top management job titles. The fortunes of major corporations in Spain increasingly depend on technically sophisticated people. INVESTIGACION Y CIENCIA keeps these people in touch with the advances in science that drive the growth of Spain's industry.

SCIENTIFIC AMERICAN speaks the languages of more than half the world's population. We are the one publication in the world today providing efficient coverage of technology-based management. We reach men and women in industry, whose qualification to make technical decisions places them in key positions in their country's government and industry.

SCIENTIFIC AMERICAN in eight languages: English, Spanish, Italian, French, German, Japanese, Chinese and now Russian has gathered in its audience three million people who make the future happen around the world.

For more information on our Spanish-language edition contact:

Elena Sanchez-Fabres
Prensa Scientifica, s.a.
Calabria 235-239
Barcelona 29, SPAIN
Telephone (011) 343-322-0551

or in New York
John Kirby,
V.P./Advertising Director
SCIENTIFIC AMERICAN
415 Madison Avenue
New York, New York 10017
Telephone 212-754-0262

SAY "CELESTRON"

NOVEMBER 11th AT YOUR CELESTRON DEALER

Free telescope demonstrations, free gifts, free rebates!

If you've ever looked at the night sky with wonder, November 11th is your day and night!

First the entertainment. All day on November 11th, and until closing, participating Celestron Dealers will be featuring exciting demonstrations and displays of a wide variety of Celestron telescopes — including our new inexpensive Cometron refractors, our new quality Celestron refractors and our legendary Cassegrain folded-optics telescopes. Many dealers will feature our new Super C8 — an upgraded version of the world's best selling modern telescope. There will be experts available for questions, fun and festivities. And at night on the 11th many dealers will set up scopes outside their stores — if you've never seen the wonders of the universe, this is your night. And it's all free.

Free gift. Just for walking in and saying "Celestron," participating Dealers will give you a free moon map.

Finally, the rebates. You don't have to wait till Nov. 11th, the money's right here. For the first time ever, a $10 to $20 rebate on Celestron's precision binoculars — including zoom, armoured and waterproof models and the NOVA and GIANT series. Plus a chance at every participating dealership to win a pair of Celestron 8 x 21 binoculars. Mark your calendar. And call our toll-free number.

$$ REBATES! $$

To celebrate, Celestron is offering up to $20 rebates on our full line of quality binoculars. All models are included — zoom, armoured, waterproof, mini's, Nova's and Giant series. Offer ends 12/31/83. At participating dealers.

CELESTRON®
PRECISION OPTICS

For the Names of Participating Dealers near You

CALL FREE

1-800-421-1526

(Mon-Fri; 8-4:00 Pacific Time; Continental USA only)

CELESTRON INTERNATIONAL, P.O. Box 3578 SA11
2835 Columbia St., Torrance, CA 90503 (213) 328-9560

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

COMPUTER RECREATIONS

SCIENTIFIC AND ENGINEERING PROBLEM-SOLVING WITH THE COMPUTER. William Ralph Bennett, Jr. Prentice-Hall, Inc., 1976.

HOW ARTIFICIAL IS INTELLIGENCE? W. R. Bennett, Jr., in *American Scientist,* Vol. 65, No. 6, pages 694–702; November–December, 1977.

THE UNCERTAINTIES OF A PREEMPTIVE NUCLEAR ATTACK

WHY ICBMs CAN SURVIVE A NUCLEAR ATTACK. Lt. Col. Joseph J. McGlinchey and Dr. Jakob W. Seelig in *Air Force Magazine,* Vol. 57, No. 9, pages 82–85; September, 1974.

STRATEGIC VULNERABILITY: THE BALANCE BETWEEN PRUDENCE AND PARANOIA. John D. Steinbruner and Thomas M. Garwin in *International Security,* Vol. 1, No. 1, pages 138–181; Summer, 1976.

BALLISTIC MISSILE GUIDANCE AND TECHNICAL UNCERTAINTIES IN COUNTERSILO ATTACKS. Matthew Bunn and Kosta Tsipis. Program in Science and Technology for International Security, Report No. 9. Massachusetts Institute of Technology, July, 1983.

SLURRY PIPELINES

THE LAMINAR-TURBULENT TRANSITION FOR FLOW IN PIPES, CONCENTRIC ANNULI, AND PARALLEL PLATES. Richard W. Hanks in *A.I.Ch.E. Journal,* Vol. 9, No. 1, pages 45–48; January, 1963.

CROSS COUNTRY COAL PIPE LINE HYDRAULICS. E. J. Wasp, T. J. Regan, J. Withers, P. A. C. Cook and J. T. Clancey in *Pipe Line News,* Vol. 35, No. 7, pages 20–28; July, 1963.

SLURRY PIPING SYSTEMS: TRENDS, DESIGN METHODS, GUIDELINES. Thomas C. Aude, Norman T. Cowper, Terry C. Thompson and Edward J. Wasp in *Chemical Engineering,* Vol. 78, No. 15, pages 74–90; June 28, 1971.

CENTAURUS A: THE NEAREST ACTIVE GALAXY

ACTIVE GALACTIC NUCLEI. Edited by C. Hazard and S. Mitton. Cambridge University Press, 1979.

THE X-RAY STRUCTURE OF CENTAURUS A. E. D. Feigelson, E. J. Schreier, J. P. Delvaille, R. Giacconi, J. E. Grindlay and A. P. Lightman in *The Astrophysical Journal,* Vol. 251, No. 1, Part 1, pages 31–51; December 1, 1981.

EXTRAGALACTIC RADIO SOURCES. Edited by David S. Heeschen and Campbell M. Wade. D. Reidel Publishing Co., 1982.

THE X-RAY JETS OF CENTAURUS A AND M87. Eric D. Feigelson and Ethan J. Schreier in *Sky and Telescope,* Vol. 65, No. 1, pages 6–12; January, 1983.

A MOLECULAR BASIS OF CANCER

PASSAGE OF PHENOTYPES OF CHEMICALLY TRANSFORMED CELLS VIA TRANSFECTION OF DNA AND CHROMATIN. Chiaho Shih, Ben-Zion Shilo, Mitchell P. Goldfarb, Ann Dannenberg and Robert A. Weinberg in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 76, No. 11, pages 5714–5718; November, 1979.

ONCOGENES OF SPONTANEOUS AND CHEMICALLY INDUCED TUMORS. R. A. Weinberg in *Advances in Cancer Research,* Vol. 36, pages 149–163; 1982.

ONCOGENES. J. Michael Bishop in *Scientific American,* Vol. 246, No. 3, pages 80–92; March, 1982.

SOLID-STATE SUPERLATTICES

SUPERLATTICE AND NEGATIVE DIFFERENTIAL CONDUCTIVITY IN SEMICONDUCTORS. L. Esaki and R. Tsu in *IBM Journal of Research and Development,* Vol. 14, No. 1, pages 61–65; January, 1970.

ELECTRON STATES IN CRYSTALS WITH "NIPI-SUPERSTRUCTURE." G. H. Döhler in *Physica Status Solidi (b),* Vol. 52, No. 1, pages 79–92; July 1, 1972.

ELECTRICAL AND OPTICAL PROPERTIES OF CRYSTALS WITH "NIPI-SUPERSTRUCTURE." G. H. Döhler in *Physica Status Solidi (b),* Vol. 52, No. 2, pages 533–545; August 1, 1972.

CONFINED CARRIER QUANTUM STATES IN ULTRATHIN SEMICONDUCTOR HETEROSTRUCTURES. Raymond Dingle in *Festkörperprobleme/Advances in Solid State Physics,* Vol. 15, pages 21–48, 1975.

ELECTRONIC PROPERTIES OF TWO-DIMENSIONAL SYSTEMS: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON THE ELECTRONIC PROPERTIES OF TWO-DIMENSIONAL SYSTEMS. In *Surface Science,* Vol. 73; May, 1978.

INAS-GASB SUPERLATTICES—SYNTHESIZED SEMICONDUCTORS AND SEMIMETALS. L. Esaki in *Journal of Crystal Growth,* Vol. 52, Part 1, pages 227–240; April, 1981.

ELECTRONIC PROPERTIES OF TWO-DIMENSIONAL SYSTEMS: PROCEEDINGS OF THE FOURTH INTERNATIONAL CONFERENCE ON ELECTRONIC PROPERTIES OF TWO-DIMENSIONAL SYSTEMS. In *Surface Science,* Vol. 113, Nos. 1–3; January, 1982.

THE SOCIAL ARCHAEOLOGY OF MEGALITHIC MONUMENTS

BEFORE CIVILISATION: THE RADIOCARBON REVOLUTION AND PREHISTORIC EUROPE. Colin Renfrew. Cambridge University Press, 1979.

INVESTIGATIONS IN ORKNEY. Colin Renfrew. Society of Antiquaries, distributed by Thames and Hudson, 1979.

MEGALITHIC MONUMENTS. Glyn Daniel in *Scientific American,* Vol. 243, No. 1, pages 78–90; July, 1980.

A SEED-EATING BEETLE'S ADAPTATIONS TO A POISONOUS SEED

CHEMICAL ECOLOGY. Ernest Sondheimer and John B. Simeone. Academic Press, 1970.

INTRODUCTION TO ECOLOGICAL BIOCHEMISTRY. J. B. Harborne. Academic Press, 1982.

THE NON-PROTEIN AMINO AND IMINO ACIDS: BIOLOGICAL, BIOCHEMICAL AND TOXOLOGICAL PROPERTIES. Gerald A. Rosenthal. Academic Press, 1982.

THE ERUPTION OF KRAKATAU

THE KRAKATOA ERUPTION. R. D. M. Verbeek in *Nature,* Vol. 30, No. 757, pages 10–15; May 1, 1884.

THE ERUPTION OF KRAKATOA AND SUBSEQUENT PHENOMENA. Edited by G. J. Symons. Trübner & Co., 1888.

THE KRAKATOA AIR-SEA WAVES: AN EXAMPLE OF PULSE PROPAGATION IN COUPLED SYSTEMS. David Harkrider and Frank Press in *The Geophysical Journal of the Royal Astronomical Society,* Vol. 13, Nos. 1–3, pages 149–159; July, 1967.

TSUNAMIS OF VOLCANIC ORIGIN: SUMMARY OF CAUSES, WITH PARTICULAR REFERENCE TO KRAKATOA, 1883. J. H. Latter in *Bulletin Volcanologique,* Vol. 44, No. 3, pages 467–490; 1981.

THE 1883 ERUPTION OF KRAKATAU. Stephen Self and Michael R. Rampino in *Nature,* Vol. 294, No. 5843, pages 699–704; December 24/31, 1981.

THE AMATEUR SCIENTIST

WAVE-MAKING BY WHIRLIGIG BEETLES (GYRINIDAE). Vance A. Tucker in *Science,* Vol. 166, No. 3907, pages 897–899; November 14, 1969.

COMMUNICATION BY SURFACE WAVES: MATING BEHAVIOR OF A WATER STRIDER (GERRIDAE). R. Stimson Wilcox in *Journal of Comparative Physiology,* Vol. 80, pages 255–266; 1972.

Today, you're doing more than ever. So should your phone.

AT&T introduces GENESIS™ Telesystem.

Genesis telesystem is designed to help organize and streamline your life in so many ways.

The Genesis telesystem remembers all your important numbers, including emergency numbers. And it dials with just a touch. It times your calls and even has a built-in speaker.

But the really terrific thing about the Genesis telesystem is that you can customize it to fit your own needs. Add an optional cartridge, and you've got one-touch access to call forwarding, call waiting, and three-way calling. Or add another cartridge, and you have the convenience of automatic re-dial of busy or unanswered numbers. And that's only half of it.

Soon, you'll be able to add a module with an electronic directory that memorizes names and numbers in alphabetical order. Or the reminder cartridge, to remind you of special occasions and appointments.

Genesis telesystem. The more you do, the more it will help you do it.

We set the standards.

AT&T

See it at your AT&T Phone Center, now at thousands of leading retail stores.

Today's Chevrolet

Cavalier. The hottest-selling has a cool, low price.

Chevy Cavalier. The hottest-selling Chevrolet for a lot of good reasons. Hottest-selling because of some of the most advanced design and engineering you'll find on a front-wheel-drive car today. Hottest-selling because of high-level fit and finish, thanks to dedicated workers using computerized robots and lasers. Hottest-selling because Cavalier offers more horsepower and more sedan room than the three top-selling imports. Hottest-selling because of a cool, low price.

Hot-selling performance. Cavalier goes with a high-compression, electronically fuel-injected 2.0-liter engine with an on-board computer that constantly fine-tunes your engine while you drive.

Upbeat style. Cavalier is putting the heat on the competition with new front-end styling which makes its clean wedge shape even cleaner looking. And Cavalier's handsome interior is laid out in a fashion that allows you to drive comfortably and enjoy the challenges of the road.

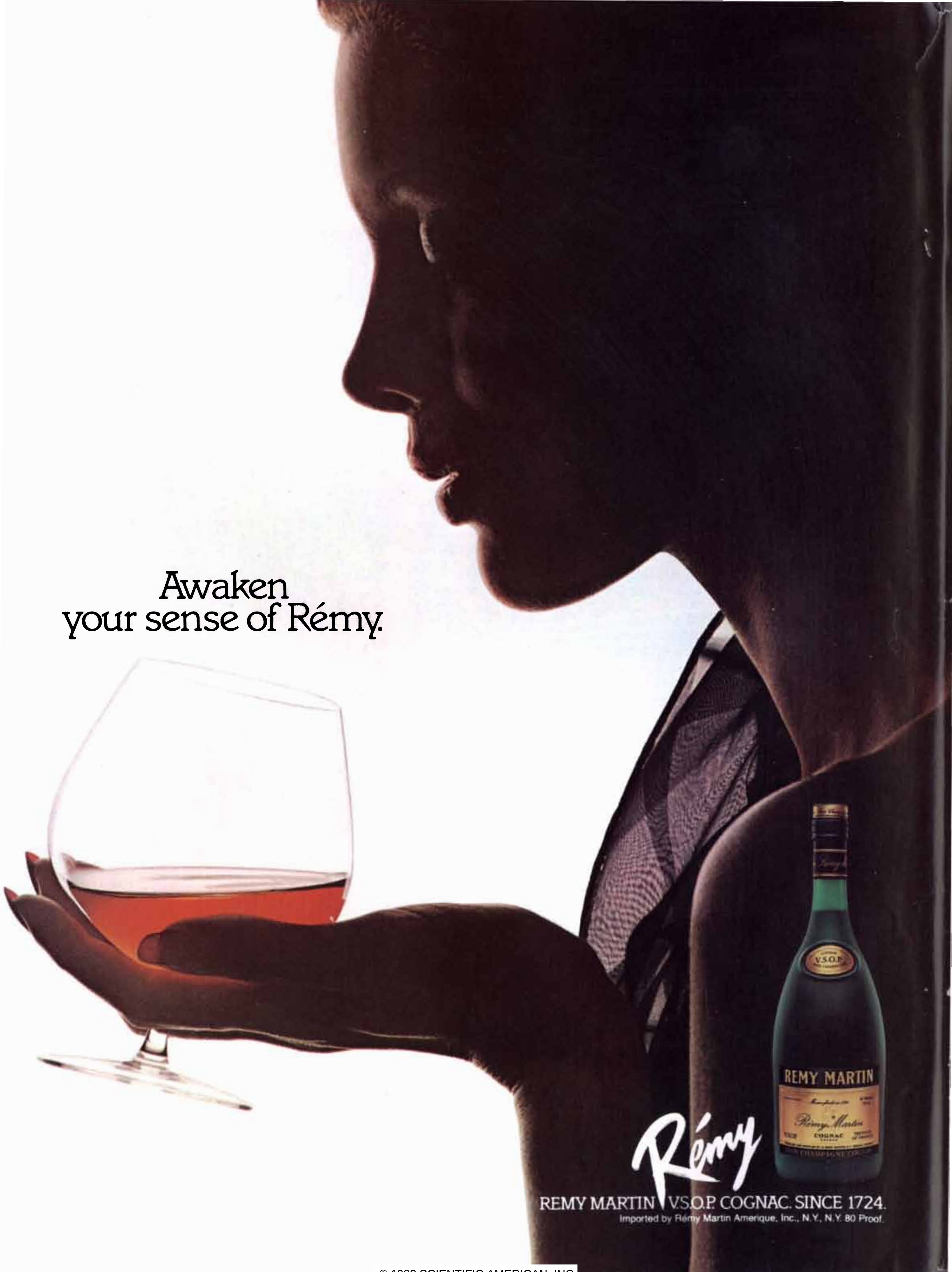
All at a cool, low price. Best of all, Cavalier gives you all this at a low Chevrolet price. So if you're looking for enlightened thinking and a low price in a responsive, fun-to-drive automobile, grab a hot- selling Chevrolet Cavalier. Today's Chevrolet. Bringing you the cars and trucks you want and need. That's what Taking Charge is all about.

Some Chevrolets are equipped with engines produced by other GM divisions, subsidiaries, or affiliated companies worldwide. See your dealer for details

OFFICIAL CARS AND TRUCKS OF THE XIV OLYMPIC WINTER GAMES

TM

Sarajevo '84


Awaken
your sense of Rémy.
REMY MARTIN
V.S.O.P.
Rémy Martin
COGNAC
Rémy
REMY MARTIN V.S.O.P. COGNAC. SINCE 1724.
Imported by Rémy Martin Amerique, Inc., N.Y., N.Y. 80 Proof.