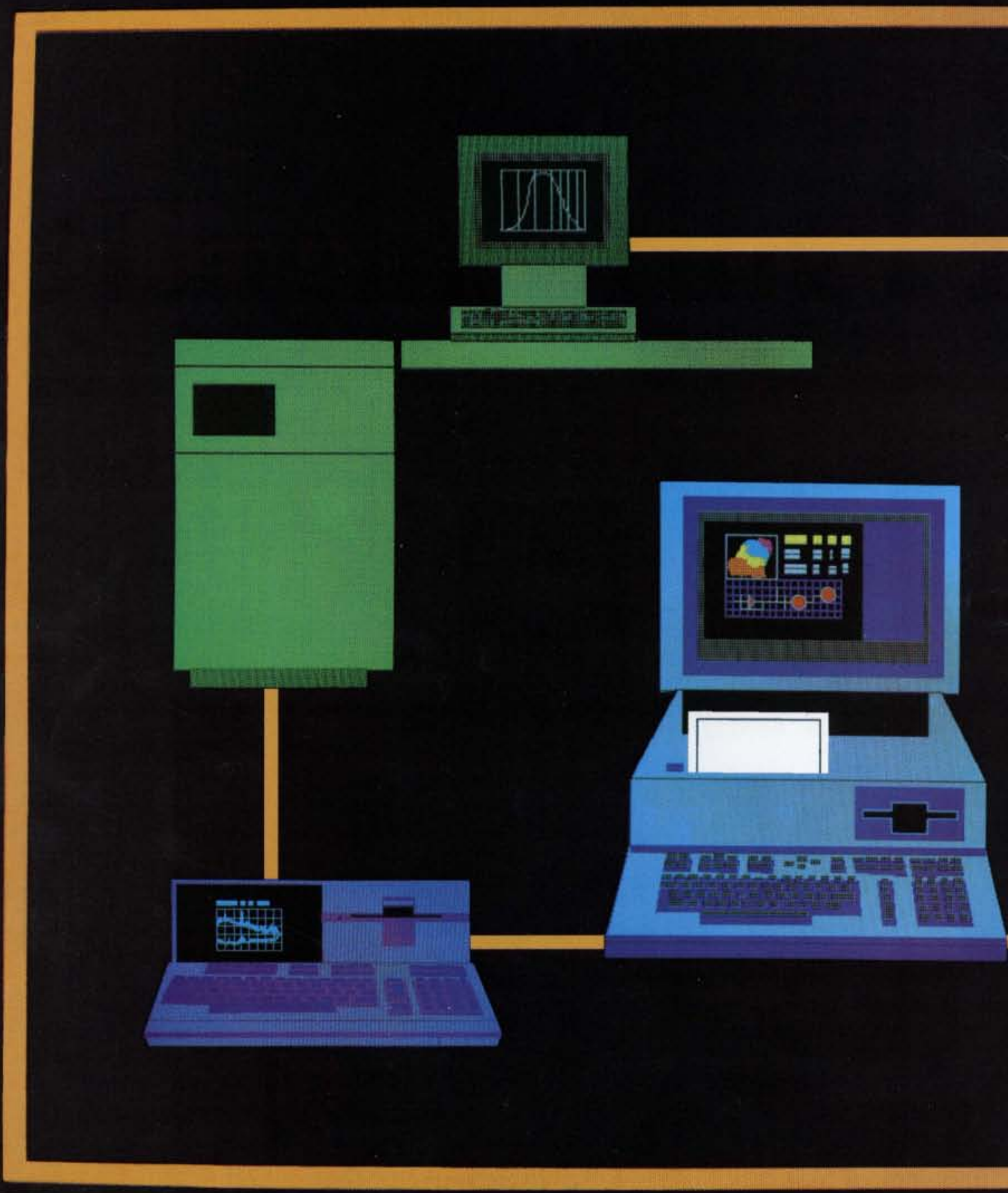
SCIENTIFIC AMERICAN

AZTEC MONUMENT

$2.50

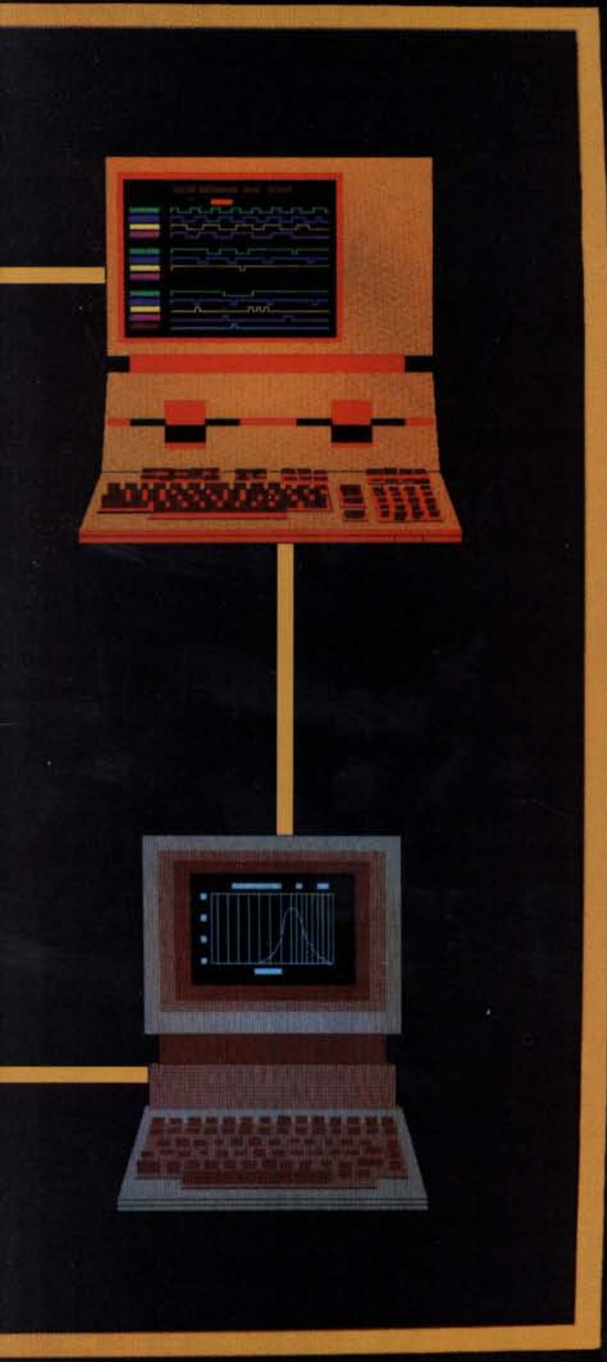
August 1984

The Engineering Pr



With HP's broad range of workstations, you can choose the one that best fits your application and budget. From personal computers for engineers to desktop mainframes, with beginning prices ranging from $3,950 to $28,250.

oductivity Network.



Presenting your broadest choice of engineering workstations anywhere. From a low-cost personal computer to a powerful 32-bit system.

No doubt about it; when you put a computer on every engineer's desk, it helps them be more productive.

In fact, at Hewlett-Packard, the workstation approach has worked so well for our own engineers that we've taken it to an entirely new dimension.

It's called the Engineering Productivity Network. As you can see, it includes the broadest range of workstations available anywhere. From the recently-introduced 32-bit HP 9000 desktop mainframe to our personal computer designed just for engineers.

With such a wide range to choose from, you can match the computing power to the individual task. And give every engineer the tool it takes to do the best possible job.

To help you tailor every workstation to a task even more closely, we have over 200 specific application software packages. And since our workstations use HP-UX—our enhanced version of the UNIX* operating system—we'll be able to bring you a lot more in the future.

Of course, when you tie all these individually-optimized workstations together into a network, you can look for an even higher level of performance. Whether you choose our HP-UX networking, Ethernet local area network, or Shared Resource Manager, your engineers can share data and peripherals, communicate much better, and get more done than ever before.

The HP Engineering Productivity Network. You can start with one workstation, or with dozens. But no matter where you start, you're actually making your whole department more productive. One step at a time.

To find out more, just call the nearest HP office listed in the white pages of your telephone book, and ask for a Technical Computer Representative. Or write to Pat Welch, Dept. 41189, Hewlett-Packard, 19447 Pruneridge Avenue, Cupertino, CA 95014. In Europe, write Henk van Lammeren, Hewlett-Packard, Dept. 41189, P.O. Box 529, 1180 AM, Amstelveen, the Netherlands.

We'll be happy to show you how far you can really go with a good idea.

*UNIX is a trademark of Bell Laboratories.

HEWLETT PACKARD

BD-02316

Established 1845 **SCIENTIFIC AMERICAN** August 1984 Volume 251 Number 2

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 $1.25 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 $1.25 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 0036-8733/84. $1.25 + 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

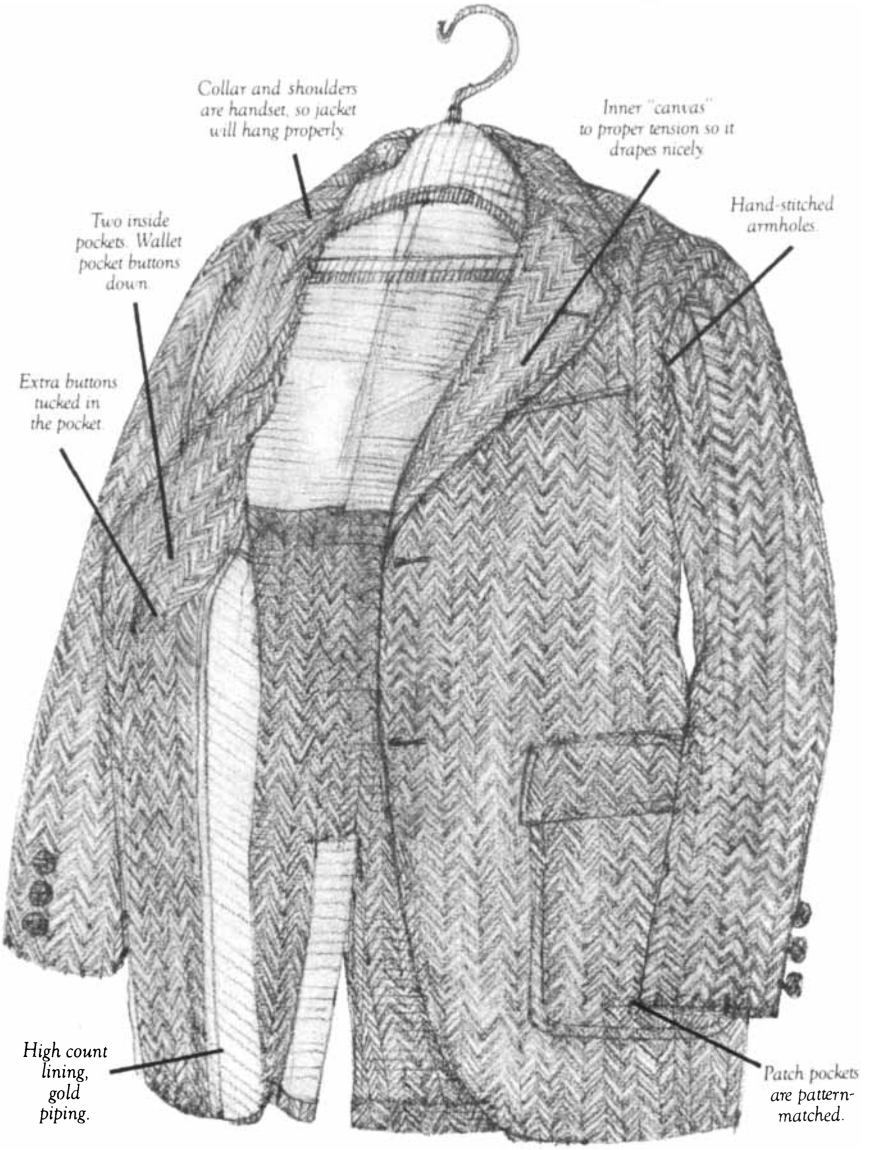
The painting on the cover shows the Stone of Coyolxauhqui, a huge sculptured disk uncovered in the excavation of the Great Temple of Tenochtitlán in Mexico City (see "The Great Temple of Tenochtitlán," by Eduardo Matos Moctezuma, page 80). The Great Temple was the Aztec monument to Huitzilopochtli, the god of war, and Tlaloc, the god of water; it was destroyed by the Spanish conquistadors. According to Aztec myth, Coyolxauhqui was the sister of Huitzilopochtli and was killed by him in a battle on the hill of Coatepec. The disk, which is 3.25 meters in diameter, shows her dismembered body. When the disk was uncovered, it was cracked in two; the crack is eliminated in the painting.

THE ILLUSTRATIONS

Cover painting by Marvin Mattelson

Page	Source
20–23	Alan D. Iselin
34–43	Ian Worpole
44–51	Jerome Kuhl
53	John S. Mathis, Blair D. Savage and Joseph P. Cassinelli, University of Wisconsin at Madison
54	Cerro Tololo Inter-American Observatory
55	John S. Mathis, Blair D. Savage and Joseph P. Cassinelli, University of Wisconsin at Madison
56–59	Allen Beechel
60	John S. Mathis, Blair D. Savage and Joseph P. Cassinelli, University of Wisconsin at Madison
71	G. Steven Martin, University of California at Berkeley
72–75	Ilil Arbel
76	Ilil Arbel (*top*); Tony Hunter, Salk Institute of Biological Studies (*bottom*)
77	Ilil Arbel (*top left*); Tony Hunter, Salk Institute of Biological Studies (*top right*); Ilil Arbel (*bottom*)
78–79	Ilil Arbel
81–82	National Institute of Anthropology and History, Mexico City
83–84	Patricia J. Wynne
85	Emily Umberger
86	National Institute of Anthropology and History, Mexico City (*top*); José A. Naranjo, National Gallery of Art (*bottom*)
87–89	José A. Naranjo, National Gallery of Art
91	Dina F. Mandoli and Winslow R. Briggs
92	Ian Worpole
93–98	Tom Prentiss
101	Hank Iken, Walken Graphics
102–103	Gabor Kiss
104–105	Berthold K. P. Horn and Katsushi Ikeuchi, Massachusetts Institute of Technology
106–111	Gabor Kiss
115	Walter D. Koenig, University of California at Berkeley
116–121	Patricia J. Wynne
122	R. F. Bonifield
124	Lorne A. Whitehead
126–129	Michael Goodman

The Lands' End herringbone jacket. A classic when we introduced it. And, at $115, an even greater value now.



What first led us to think about tailored clothing was your active response to our Oxford buttondowns, our Lands' End knit ties, and other classic items like our Lands' End turtlenecks.

We reasoned you might like suitable jackets to wear over these items, *provided*—and this was a big "if"—we could offer them with the kind of quality features you expect from Lands' End, and at Lands' End prices.

Then we went to work.

We sent our buyers on the prowl for fabrics. We challenged our tailors to add quality features, not eliminate them for cost reasons. And, finally, we encouraged them to practice their Old World tailoring skills; and how they responded when they realized we were serious!

The results are now in.

The herringbone classic described above is typical of what Lands' End now offers in our *Charter Collection*. A small but excellent group of traditional jackets, plus some well-tailored slacks that harmonize with them, and some appropriate shirts and ties.

There's no way we can properly describe the look and feel and fit of this clothing. You need to experience it, yourself. Even then you'll find it hard to believe prices ranging from $105 in a jacket of cotton corduroy and $115 in a herringbone to $195 for a top of the line camel's hair.

Open our catalog.

See the complete *Charter Collection* in our catalog. Shop us for sweaters, ski wear, deck wear and soft luggage, too. And, remember, everything Lands' End offers carries our incredible guarantee.

In two words:
GUARANTEED. PERIOD.

Send the coupon for a free catalog or call for one toll-free at 800-356-4444. One of our operators will answer you in person, 24 hours a day.

LANDS' END
DIRECT MERCHANTS

of fine wool and cotton sweaters, Oxford button-down shirts, traditional dress clothing, snow wear, deck wear, original Lands' End soft luggage and a multitude of other quality goods from around the world.

☐ **Please send free catalog.**

Lands' End Dept. Q-06
Dodgeville, WI 53595

Name____________________

Address____________________

City____________________

State__________ Zip__________

Or call Toll-free:

800-356-4444

LETTERS

Sirs:

The Romans did indeed have the threaded nut and bolt, contrary to the assertion of Frederick E. Graves in his article "Nuts and Bolts" [SCIENTIFIC AMERICAN, June]. The interested reader is referred to F. M. Feldhaus' *Technik der Vorzeit,* W. Engelmann, Leipzig and Berlin, 1914, and particularly to Henry C. Mercer's invaluable work *Ancient Carpenters' Tools,* Horizon Press, 1975 (fifth edition). Mercer reproduces a photograph of a Roman threaded bolt nut of wrought iron, one and five-eighths of an inch square by five-eighths of an inch thick, "excavated between 1893 and 1900 with other Roman relics, of date c. A.D. 180 to 260, at the Roman fortified boundary ('Limes') at Niederbieber, near Neuwied, Nassau, Germany." Although to my knowledge no Roman threaded bolts or wrenches have been recovered, I believe we may safely infer that they existed. The inconspicuousness of the nut, which was on display at the Provincial Museum in Bonn at the time of Mercer's writing (1928), certainly contrasts with its significance. One wonders what unrecovered mechanism the nut and bolt may have enabled the Romans to construct; there is, of course, the melancholy but not unlikely possibility that the nut and bolt were invented for purposes of war.

ROBERT FIENGO

City University of New York
New York, N.Y.

Sirs:

"The Descent of Hominoids and Hominids," by David Pilbeam [SCIENTIFIC AMERICAN, March], points up the great progress that is being made in filling in gaps that have been styled for more than a century as "the missing link." A biologist regularly encounters the usual creationist arguments, of which one they consider among the strongest is the question of missing links, a lack they regard as virtual proof that no species is or ever was derived from another. They assert that the only radical changes since the creation have involved loss of created species, that otherwise we see only minor alterations leading to new varieties of any given species, but no new species. They have a little trouble with a few examples, such as *Archaeopteryx* (which, in spite of many clearly reptilian characteristics, they regard as a sort of delinquent bird). The degree to which the primate picture is being filled in they find quite unconvincing, and they totally ignore the relations among species revealed by molecular biology. But what of the general argument?

In a scientific sense the lack of intermediate forms has obvious explanations. The acquisition of new successful genetic traits is very rare, and any population of individuals with adaptations will typically be small and restricted to some highly localized area. One would assume that even under the most favorable circumstances new traits would at first be of neutral value or possibly disadvantageous except in some particular habitat in which the altered organism could survive. Only when an accumulation of advantageous traits resulted in real survival value would the mutants begin to radiate, compete successfully and increase in numbers exponentially.

But only when exponential population expansion had occurred to a very marked degree would the chances of finding the novel form be finite. Still more stringent and rare, the link would need properties and location favorable to fossilization. Perhaps an analogy will make this point more realistic.

One might hypothesize that the airplane was an evolutionary development that began with the bicycle. (The Wright Brothers as bicycle mechanics used a number of features of the bicycle in their early planes.) In the first 10 years the airplane was hardly more than a curiosity, and the number of these frail machines in existence was minuscule until World War I. Even then we are talking about only hundreds of airplanes basically made of wood and cloth, and hence quite poorly adapted to stand the ravages of time and neglect.

It was only with the development of all-metal planes in the 1930's that we could begin to think of the airplane as an object that, in a discarded form, might exist for an appreciable time. As everyone knows, World War II induced improvement and proliferation of airplanes in an exponential fashion to achieve the state we see today of a world in which nearly every hamlet has an airstrip with at least a few planes and in which the number of airplanes worldwide is in the hundreds of thousands.

The point, of course, is that were it not for museums and the written record, the "fossil evidence" would lead us to the conclusion that the airplane sprang on us in the 1940's in a most elaborate and perfected form having no linkage with any possible forebear such as the bicycle. If we were not thoroughly acquainted with the steps in the development of this now immensely complex, highly diverse and seemingly miraculous machine, we too might be tempted to think of the supernatural.

DEAN FRASER

Indiana University
Bloomington

WHAT A CLOCKMAKER IN 18TH CENTURY ENGLAND TAUGHT US ABOUT MAKING QUALITY PERSONAL PRINTERS.

Each gear finely honed. Each pendulum carefully balanced. Each timepiece a combination of precision and function. He was a craftsman, building a quality product to stand the test of time.

Today, Okidata makes the most technologically advanced dot matrix printers the same way. With quality, performance, and a healthy respect for value.

Function with affordability. All printers print, but an Okidata does it with more performance and versatility than any other. There are seven models to choose from, priced at $349 to $2,595.

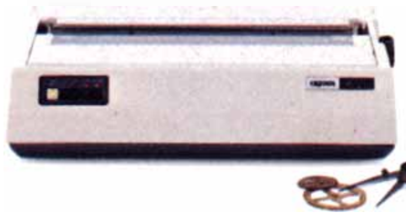
Affordability with flexibility. Okidata offers three print modes, too. The data processing mode lets you print up to 350 characters per second. That's five pages a minute. Another mode lets you print emphasized or enhanced text for more effective presentations, while the correspondence mode prints letter quality at up to 85 characters per second, with clarity that rivals daisywheel. And an Okidata can print graphics and charts, which a daisywheel can't. This allows you to fully use the latest integrated software packages like Lotus 1-2-3™ and Symphony™.

Flexibility with compatibility. Each Okidata printer is fully compatible with all popular software packages and personal computers. Special interfaces are available for IBM and Apple, including the Apple Macintosh.

Compatibility with reliability. Here's where Okidata quality really shows. With a warranty claim rate of less than 1/2 of 1%. With printheads that last well beyond 200,000,000 characters and come with a full year guarantee. With service available nationwide through Xerox Service Centers.

Precision and performance. Quality and value. That old English clockmaker would have been very proud of us.

Call 1-800-OKIDATA (609-235-2600 in New Jersey) for the name of the Authorized Okidata Dealer nearest you.



OKIDATA
an OKI AMERICA company
Technological Craftsmanship.

Lotus 1-2-3 and Symphony are trademarks of Lotus Development Corp.

Mercedes-Benz, which formerly stood virtually alone among automobiles, has a new and worthy competitor. The Continental Mark VII LSC. That's the conclusion reached by automotive experts who, after driving it, made some interesting observations.

"A substantial and gutsy departure...designed to create a direct challenge to some of the great Autobahn cruisers of Europe." —*Motor Trend*

The Mark VII LSC is powered to keep company with some very illustrious touring machines. With an advanced, electronically fuel-injected 5.0 liter V-8 or a new European designed 2.4 liter Turbo Diesel. Another factor in the way this car moves is its shape. With a drag coefficient of just .38, the flow of air actually helps it hold the road more solidly.

"We can't think of another car in this class, domestic or imported, that has its manners and comfort." —*AutoWeek*

Those aforementioned road manners are due, in large measure, to an electronically

Mercedes now has something it hasn't had before.

controlled air suspension system. A technological advance offered by no other automaker in the world.

Coupled with quick-ratio, power-assisted rack-and-pinion steering, front and rear stabilizer bars, and nitrogen pressurized shocks, the result is a unique combination of disciplined handling and plush ride.

As for comfort, there are a number of interior accoutrements including a rear passenger compartment heating vent and, as an option, individually heated contoured front seats.

"Move over Mercedes and BMW, there's a new show in town." —*AutoWeek*

The single guiding philosophy behind the Continental Mark VII LSC is this: That it be rewarding to drive, not just ride in. It's a philosophy that guided the engineering approach to every aspect of this automobile. Its ride, performance, even the way it looks.

The new Continental Mark VII LSC. You could buy a much more expensive European equivalent. But why would you want to?

THE NEW CONTINENTAL MARK VII LSC.

Competition.

Introducing Firestone's S-211 Import Car Radial.

WORLD PROVEN PERFORMANCE. FROM YOUR HOMETOWN FIRESTONE RETAILER.

With over 10 million sold all around the world, Firestone's S-211 import car radial has proved itself in use in over 80 countries.

And it's approved on the cars of 14 leading international manufacturers from Alfa Romeo and Audi to Toyota and Honda to Volkswagen and Volvo.

But you don't have to search all over the world to find it. Because now it's come to America at over 3,500 hometown American Firestone retailers.

In a full range of sizes for your import or small domestic car to give you the performance you want backed by the service you need.

Advanced design for world driving conditions with straight sidewalls, elliptical footprint, dual tread radius, and laterally stable steel belt construction for traction, handling, and long wear.

The Firestone S-211. See it at your Firestone retailer.

World proven performance.
Hometown American service.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

AUGUST, 1934: "It took chemistry hundreds of years to discover in nature all of the 92 chemical elements that Mendeleeff was able to predict in advance through his great generalization, the periodic table. Now the Italian physicist Enrico Fermi believes he has given us Element 93. Does this mean that, with nature's normal 92 elements already found, we are at the beginning of a new series of discoveries of superelements higher in atomic weight than those already known? It must be kept clearly in mind that the new 'discovery' (which is actually a man-made synthesis) of Element 93 had nothing to do with the earlier predictions. It was the logical outcome of research done recently by the French physicists Irène Curie Joliot (daughter of Madame Curie of radium fame) and her husband, Professor Frédéric Joliot. These two announced earlier this year that they had created new forms of nitrogen, silicon and phosphorus by bombarding the nuclei of boron, magnesium and aluminum. The elements of higher weight were thus synthesized from elements of lower weight, but they did not 'stay put'—they proved to be radioactive and transitory."

"After years of litigation in various courts, Dr. Lee De Forest's claim to the invention of the 'feed-back' and oscillating vacuum-tube circuits has been upheld by the Supreme Court of the United States. The Court stated that many experiments were made with a view to exploring the vacuum tube's capacities and developing them. Among those interested and curious was Edwin H. Armstrong. He conceived the idea in about January, 1913, that through a hook-up or coupling of the output and input circuits there would be a feed-back or regeneration of energy whereby the plate in the vacuum tube would become an independent generator of continuous oscillations. De Forest with his assistant Van Etten had been working during the summer of 1912 on the development of the vacuum tube as a generator of alternating currents for any and all uses. On August 6, 1912, a diagram showing a feed-back hook-up of the input and output circuits is recorded in Van Etten's notebook. Armstrong does not deny that all this was done just as stated by De Forest. What he does deny is that anything done or recorded in August, 1912, is in anticipation of his own invention. He says the sustained oscillations generated at that time were audio and not radio frequency. He says there was no perception or thought that the plate could be made to oscillate at radio as well as audible frequencies through a coupling of the circuits."

"Between Admiral Richard Evelyn Byrd and the bleak Antarctic, where the thermometer drops to almost 100 degrees below zero Fahrenheit, there are only the slim walls of his hut, a bare four inches in thickness. Within the hut he has a cook-stove, a small heater and a lighting arrangement—together consuming only four quarts of oil a day. Yet the Admiral reports that his quarters are decidedly livable. Four quarts of oil... four inches of wall... on the hem of the South Pole! It doesn't sound possible. And it wouldn't be, except for the intervention of science in the form of two paper-thin layers of aluminum foil, or metallation. Embedded in the wall, they throw the heat, generated by the cook-stove, the heater and the lights, and even from the Admiral's body, back into the room, much in the manner of mirrors. It has been demonstrated that only about 5 per cent of radiant heat that strikes aluminum foil goes through—95 per cent is reflected back. This white man's igloo represents so remarkable a feat in the difficult art of keeping warm that it is being exhibited in replica at the Century of Progress Exposition in Chicago."

SCIENTIFIC AMERICAN

AUGUST, 1884: "Very few people living in the United States, except residents of the Pacific Coast, had ever felt a decided earthquake shock previous to the afternoon of Sunday, August 10, when one was experienced on the Atlantic seaboard from Maine to Virginia, extending as far inland as West Virginia, and over the greater portion of Pennsylvania and New York. The shock was most severely felt on the south shore of Long Island, near New York City, and in the southern part of the city itself, and on the adjacent coast of New Jersey, between 2:05 and 2:07 P.M. The length of its duration is variously given at from five to 20 seconds, the latter probably including the beginning and ending of the tremor, not so plainly perceptible as the very violent shaking so plainly perceived by everyone for from five to seven seconds. There was no damage done of any consequence anywhere—only some glass and crockery broken, ceilings cracked and loose chimney bricks dislocated—but the most massive buildings in New York were shaken to their foundations, and people who happened to be on third or fourth stories, or higher, thought that only slight additional force would have been needed to bring down many structures and cause a great loss of life."

"In view of the possible invasion of the United States by the cholera during the present summer, the following comments by Florence Nightingale will be read with interest: 'Our whole experience in India, where cholera is never wholly absent, tends to prove that cholera is not communicable from person to person. Cholera is a local disease—an epidemic affecting localities, and there depending on pollution of earth, air and water and buildings. The isolation of the sick cannot stop the disease, nor can quarantine, cordons or the like. The only preventive is to vigorously enforce sanitary measures, *e.g.,* scavenge, scavenge, scavenge; wash, cleanse and limewash; remove all putrid human refuse from privies and cess pits and cesspools and dustbins; look to stables and cow sheds and pig sties; look to common lodging houses and crowded places, dirty houses and yards.'"

"The brilliant discoveries by Pasteur and by Koch are as much due to the perfected microscope as to any one cause. The nature and habits of the tubercular bacillus have only been capable of study since the microscope was so improved that organisms heretofore unrecognizable stand revealed. Disease has been traced to its source, the presence of bacteria and germs, by the use of the finest microscopic appliances; and in fact a thorough course study in the art of intelligently using this instrument is becoming yearly a greater necessity. And when another potent servant of man, electricity, is summoned to aid the microscope, the power of the latter is increased to an astonishing degree. Recently in London such an apparatus threw upon a screen the image of a cholera germ, magnified two million times, and in which these long hidden and minute organisms appeared the size of a human hand."

"It is not generally known that there is an American town in the realms of the Czar, yet such is a fact, it being near Moreton Bay in Kamtschatka. The colony has been formed, gradually, by immigrants attracted by the establishment of important lumbering operations, including saw mills, by an American company, and the town itself has so far been practically ignored. It is not on any known map and does not appear in the Russian real-estate register or on any tax list. The consequence is that the inhabitants pay no kind of tax and, until recently at least, have remained independent of the Russian authorities."

NOW IN PRODUCTION: THE WORLD'S FIRST PORTABLE CELLULAR TELEPHONE.

Here's a phone you can take to work, to the golf course or to lunch. Motorola's hand-held cellular portable is a true portable, reflecting the most innovative cellular technology in production anywhere in the world today.

It is also self-powered, self-contained, weighs only 30 ounces, and can call any other telephone in the world, from any U.S. city with cellular service. In short, it's unmatched for convenience, unsurpassed in cellular electronics, and unavailable from Japan.

In the United States today, cellular telephone service is becoming a reality on a city-by-city basis. As systems are installed, the remarkable mobile telephones you've been hearing about are being introduced. These telephones open new possibilities in personal communication and are being enthusiastically received.

And we are now delivering the most innovative cellular telephone of all, the amazing portable. As this instrument becomes more and more popular, a quiet revolution in personal communication will occur.

Now you can take it with you!

The cellular age is the realization of a plan we've been working on for 14 years. The result will soon be all around us. Already, some new car manufacturers are listing cellular telephones as accessories.

We'll soon see personal telephones in boats, briefcases, and perhaps even pockets. And the smaller and lighter these units are, the more convenient they'll be. The Motorola portable already becomes the smallest mobile cellular phone of all, when used with its vehicle adapter.

This remarkable telephone fairly abounds with technological breakthroughs in miniaturization. More than 20 distinctively different patents are represented in its ultra-compact design.

Motorola is no stranger to mobile communications. Over the years, we've built more mobile telephones, two-way radios and pocket pagers than any other company in the world. Our preeminence in these areas and in cellular technology may surprise people who habitually expect such leadership from the electronics companies of Japan.

But leadership in any field is the result of commitment, talent, timing and investment. The proof of that is in the picture. For more information, call 1-800-752-4752.

MOTOROLA A World Leader in Electronics.

Quality and productivity through employee participation in management.

THE AUTHORS

RICHARD P. TURCO, OWEN B. TOON, THOMAS P. ACKERMAN, JAMES B. POLLACK and CARL SAGAN ("The Climatic Effects of Nuclear War") make up a group that has been collaborating for the past several years in work on the global effects of nuclear war. Turco is research scientist and program manager at R&D Associates, Inc. He got a B.S. at Rutgers University in 1965 before earning his Ph.D. in electrical engineering and physics from the University of Illinois. He is an atmospheric scientist whose work has for the most part been directed at understanding the delayed consequences of perturbing the earth's atmosphere. Toon, Ackerman and Pollack are research scientists at the Ames Research Center of the National Aeronautics and Space Administration. All three are interested in the consequences of injecting particulate matter into planetary atmospheres. Sagan is David Duncan Professor of Astronomy and Space Sciences at Cornell University. He was educated at the University of Chicago, where he received four degrees: a B.A. (1954), a B.S. (1955), an M.S. (1956) and a Ph.D. in astronomy and astrophysics (1960). He taught at Harvard University before moving to Cornell, where he is director of the Laboratory for Planetary Studies.

GUY OURISSON, PIERRE ALBRECHT and MICHEL ROHMER ("The Microbial Origin of Fossil Fuels") are organic chemists who began to work together on the subject of their article when all three were at the Université Louis Pasteur in Strasbourg. Ourisson earned a Ph.D. from Harvard University and a D.Sc. from the University of Paris before joining the faculty at Strasbourg; he has been professor of chemistry since 1958. From 1971 to 1975 he was president of the university. He has worked on plant biochemistry, organic geochemistry, microbial biochemistry, sterol conjugates, cytotoxic sterols and allergenic substances. Albrecht and Rohmer did the work that led to their doctorates under Ourisson's direction. After postdoctoral work at the University of California at Berkeley, Albrecht returned to Strasbourg as head of the organic geochemistry group of the CNRS (the French national research organization). Rohmer spent a year as a postdoctoral worker at Stanford University before becoming professor at the Université de Haute-Alsace.

JOHN S. MATHIS, BLAIR D. SAVAGE and JOSEPH P. CASSINELLI ("A Superluminous Object in the Large Cloud of Magellan") are professors of astronomy at the University of Wisconsin at Madison. Mathis got a B.S. in physics at the Massachusetts Institute of Technology and a Ph.D. in astronomy from the California Institute of Technology. He spent two years on the faculty of Michigan State University before joining the Wisconsin faculty; he was made professor in 1968. Savage received his B.S. at Cornell University. His M.A. and his Ph.D. in astronomy are from Princeton University. After a year as research associate at Princeton he moved to Wisconsin in 1968, becoming professor in 1978. Cassinelli had his undergraduate education at Xavier University before going on to earn his M.S. in physics at the University of Arizona and his Ph.D. in astronomy from the University of Washington. He was research associate at the Joint Institute of Laboratory Astrophysics for two years before taking up a job at Wisconsin in 1972; he was made professor in 1981.

TONY HUNTER ("The Proteins of Oncogenes") writes: "My interest in biology stems from an upbringing in a medical family. I was trained in natural sciences at the University of Cambridge, getting both my B.A. and my Ph.D. there. I was awarded a research fellowship at Christ's College, Cambridge, which I held in the department of biochemistry from 1968 to 1971. At the end of that time I moved to the Salk Institute in San Diego. I spent two years at the institute studying the mechanism of DNA synthesis in animal cells using the polyoma virus as a model. That was my first exposure to the workings of tumor viruses, and with one brief exception I have been working on RNA and DNA tumor viruses ever since. I returned to the Salk Institute in 1975 and was appointed professor in 1982."

EDUARDO MATOS MOCTEZUMA ("The Great Temple of Tenochtitlán") is general director of the Centro de Investigaciones y Estudios Superiores en Antropología Social at Tlalpán in Mexico. He received his master's degree at the Escuela Nacional de Antropología e Historia and then did several years of graduate study at the Universidad Nacional Autónoma de México. Matos has done extensive field work on the pre-Columbian sites of Mexico; the latest excavation done under his supervision is the subject of his article.

DINA F. MANDOLI and WINSLOW R. BRIGGS ("Fiber Optics in Plants") have worked together on the subject of their article. Mandoli is a postdoctoral fellow in the department of biochemistry at the Stanford University School of Medicine. She got her B.A. in 1976 at Wellesley College and earned her Ph.D. in biology from Stanford in 1983 for work on photophysiology and the discovery of the fiber-optic properties of plant tissues. Briggs is director of the Department of Plant Biology of the Carnegie Institution of Washington; the Department of Plant Biology is on the Stanford campus. His Ph.D. in biology is from Harvard University. After getting his doctorate he joined the Stanford faculty, eventually becoming professor of biological sciences. He returned to Harvard as professor of biology in 1967 and went to the Carnegie Institution in 1974.

BERTHOLD K. P. HORN and KATSUSHI IKEUCHI ("The Mechanical Manipulation of Randomly Oriented Parts") did the work that forms the basis of their article when both were at the Massachusetts Institute of Technology. Horn, a native of South Africa, writes: "I received my first degree in electrical engineering at the University of the Witwatersrand. From there I went to M.I.T., where I have been engaged in research on robotics almost since the inception of that subject at the M.I.T. Artificial Intelligence Laboratory in the late 1960's. My work has centered on vision by machines." Ikeuchi, a native of Japan, got a bachelor's degree in 1973 at Kyoto University and went on to earn his master's degree (1975) and his doctorate in information engineering (1978) from the University of Tokyo. He came to the U.S. and spent two years at the A.I. Laboratory at M.I.T. before returning to Japan in 1980.

PETER B. STACEY and WALTER D. KOENIG ("Cooperative Breeding in the Acorn Woodpecker") are zoologists who are independently doing long-term studies of the social system of the acorn woodpecker. Stacey got his B.A. at Middlebury College. His M.A. and Ph.D. are from the University of Colorado at Boulder. After two years of postdoctoral study at the University of Chicago he took up his current job as assistant professor of life sciences at Indiana State University. Koenig writes: "In college I quickly abandoned my early desire to be an astronomer and began pursuing an interest in natural history. This led to my getting a Ph.D. from the University of California at Berkeley. For my dissertation I decided to continue an ongoing project on the acorn woodpecker at the Hastings Natural History Reservation of the Museum of Vertebrate Zoology, which is in the Upper Carmel Valley. After obtaining my doctorate I taught at Occidental College for a year and then returned to Hastings in 1982 as assistant research zoologist and adjunct assistant professor of zoology at the University of California at Berkeley."

"dBASE II® helps keep us on our toes."

Robert Hubert
Marketing Director
The Boston Ballet

"The Boston Ballet was a company in search of a computer when I joined the organization earlier this year. And, after discovering that a large computer system was being considered, I urged a smaller, more sensible first step.

"Since almost no one at The Boston Ballet had any previous computer experience, I strongly recommended the purchase of a microcomputer and dBASE II.

"dBASE II, the relational database management system from Ashton-Tate, would give the people in the Company the time and the opportunity to get used to computing before bringing in a larger, more expensive system later on.

"dBASE II is a command-driven, highly flexible system that can be used for a great variety of applications ranging from very simple to highly complex."

A premier performer from Day One.

"dBASE II manages our extensive season subscriber mailings; keeps track of all our advertising insertions, costs and efficiency; and makes project time management a snap. We are now making plans to use dBASE II in handling the special promotions and manpower analyses critical to our day-to-day management.

"dBASE II made an immediate and sizeable impact on the efficiency of our operation."

The real kicker.

"Recently, we hired a custom systems house to develop a long-range computer program for The Boston Ballet. Without prejudice, they came back and said the new system should be based on dBASE II."

dBASE II can provide you with virtuoso performance, regardless of your application.

For more about dBASE II, contact Ashton-Tate today. 10150 West Jefferson Boulevard, Culver City, CA 90230. (800) 437-4329, ext. 217 . In Colorado (303) 799-4900. In the U.K. (0908) 568866.

ASHTON·TATE

dBASE II is a registered trademark of Ashton-Tate.
Suggested retail price for dBASE II is $700.


BREAKTHROUGH:
ADD LIFT TO LET A HUGE PLANE LAND ON A POSTAGE STAMP.
Since most large planes need a lot of room to touch down and come to a stop, they can't always land where they want to.
But we've developed a way for very big planes to land on very short runways.
The plane's engine exhaust is blown across special titanium wing flaps to create added lift at low speeds. And that allows large airplanes to approach at steeper angles and make more precise touchdowns.
This revolutionary technology, seen on this aerodynamic test model, will let the McDonnell Douglas C-17 deliver large equipment virtually anywhere. It will be able to fly from the United States directly to more than 10,000 runways around the world.
We're creating breakthroughs not only in aviation but also in health care, automated factories and information systems.
We're McDonnell Douglas.


USA
20c

Number One Warranty.

Epson® printers last longer. So we back them longer.

Every Epson printer sold in the U.S. is backed for one full year on parts and labor. That's four times longer than the "average" printer warranty.

That's also why we call it the Number One Warranty.

Taking care of No. 1.

Epson makes the best-selling printers for personal computers in the world. Part of the reason is a reliability rate approaching 100%, and still climbing.

And in the unlikely event that you ever do need service, it is comforting to know that there are more than a thousand Epson Service Centers from coast to coast.

It's also comforting to know that the most you can pay for in the first year are ribbons and paper.

Why are we so generous?

Every single Epson component must meet rigid quality control standards before assembly. And every finished printer is extensively tested before it leaves the factory.

So frankly, we don't anticipate you'll have any trouble.

A logical conclusion.

If you want a printer that works and keeps on working, year after year after year, you want an Epson.

Number one. And built like it.

EPSON

EPSON AMERICA, INC.

2780 Lomita Boulevard, Torrance, CA 90505 • Call (800) 421-5426 for the Epson dealer in your area. In California call (213) 539-9140.

EPSON is a registered trademark of Epson Corporation.

COMPUTER RECREATIONS

A computer trap for the busy beaver, the hardest-working Turing machine

by A. K. Dewdney

With the possible exception of bees, beavers are the busiest animals alive. All day they ply quiet northern waters bringing twigs and branches to their dam. It was undoubtedly this behavior that led Tibor Rado of Ohio State University to name a certain Turing-machine problem the Busy Beaver Game. In the early 1960's Rado wondered how many 1's a Turing machine could be made to print before it halted. Specifically, if a Turing machine with n possible states begins work on a tape filled with 0's, what is the largest number of 1's it can print on the tape before coming to a stop? The answer is known for $n = 1$, $n = 2$, $n = 3$ and $n = 4$ but not for $n = 5$ or for any value of n greater than 5.

Last year a contest was held in Dortmund, West Germany, to see who could discover the busiest beaver with five states. In the year preceding the contest, programs were written to generate candidate Turing machines, and hardware was developed to test the machines. In the course of this work a number of strangely behaved beavers were discovered, and the genus *Castor* had to be expanded to include several species hitherto unknown to zoologists.

The nature of the Turing machine and its place in computer science have recently been discussed in these pages by John E. Hopcroft of Cornell University [see "Turing Machines," SCIENTIFIC AMERICAN, May]. A Turing machine consists of an infinite tape, a head for reading and writing symbols on the tape and a control unit with a finite number of internal states [*see top illustration on next page*]. These components can be thought of as the hardware part of the device, whereas the contents of the control unit are the software—the Turing-machine program. It is the program that distinguishes one Turing machine from another. The program is a table the machine consults to determine what action to take next. For each possible state of the control unit and for each possible symbol at the current position of the tape head an entry in the table tells the machine what symbol to print on the tape, in which direction to move the head and what state to enter next. All the Turing machines discussed here begin in state 1.

The actions of a Turing machine can be traced by writing down the state of the control unit and the symbols marked on the tape (or a region of it) at successive moments; one should also indicate which square of the tape is currently being scanned. The bottom illustration on the next page is a trace of the Turing machine shown above it. Each line in the sequence is an "instantaneous description" of the machine. The format of the description is different from Hopcroft's, but the information is the same. I have also made the tape infinite in both directions, and I have allowed a symbol to be printed in the course of the machine's final transition (as it enters the halted state), contrary to the conventions adopted in Hopcroft's article. These differences do not change what a Turing machine can or cannot do. The format chosen here for the instantaneous description is compatible with the one used in the busy-beaver contest.

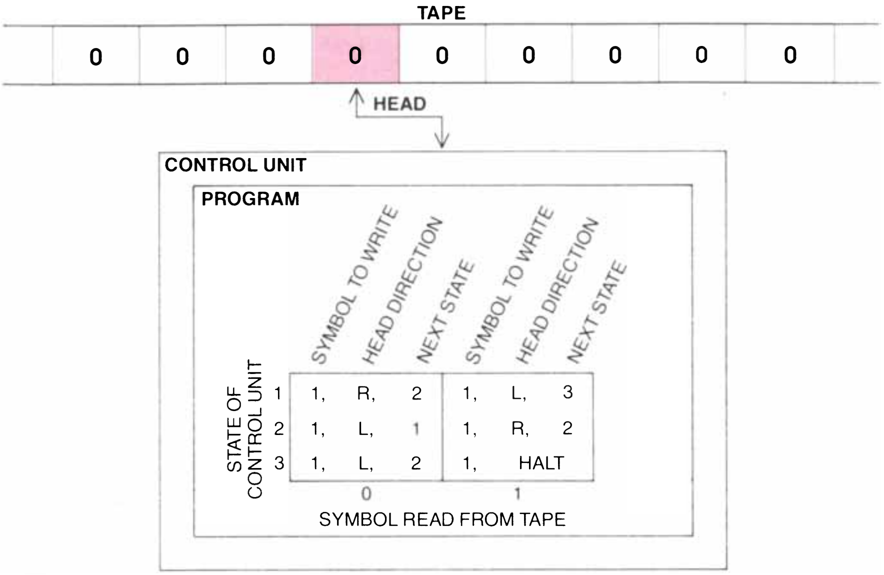
A busy beaver with n states is an n-state Turing machine that meets two conditions. First, when it is started on a tape filled with 0's, it eventually halts; second, it writes at least as many 1's as any other n-state machine that halts. Busy beavers with one and three states are shown in the top illustration on page 21. Each Turing machine is represented by a state-transition diagram, in which a state is a numbered circle and a transition between states is an arrow. The labels on the arrows describe the action of the Turing machine. For example, suppose the three-state busy beaver is in state 1 and it reads a 0 on the tape. The arrow followed under these circumstances is labeled "0,1,*R*" and leads to state 2. Hence the machine, having read a 0, writes a 1 on the tape, moves the head one square to the right and enters state 2.

The maximum number of 1's that can be produced by an n-state Turing machine that halts is denoted $\Sigma(n)$. As is indicated above, the value of $\Sigma(n)$ is known only for the first four values of n. The one-state busy beaver writes a single 1 before it halts; in other words, $\Sigma(1)$ is equal to 1. A two-state busy beaver produces a sequence of four 1's. Can readers devise such a machine? A three-state busy beaver writes six 1's; one three-state beaver is the machine whose program and sequence of instantaneous descriptions are shown in the illustrations on the next page and whose state-transition diagram is given in the top illustration on page 21. The three-state beaver was discovered in 1962 by Rado and by Shen Lin of AT&T Bell Laboratories. In 1973 Bruno Weimann of the University of Bonn found a four-state busy beaver, whose output consists of 13 consecutive 1's. Since then theorists have been searching for a five-state busy beaver.

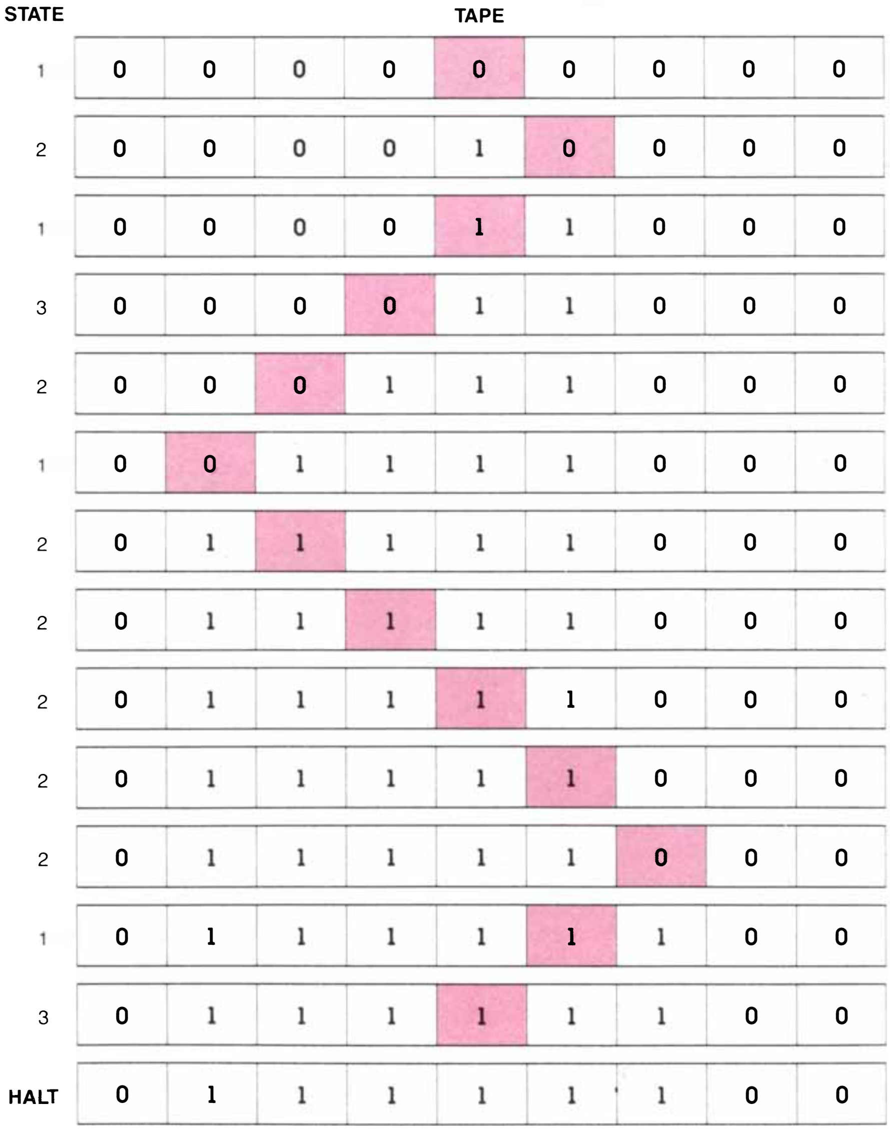
The busy-beaver contest was organized by Frank Wankmuller and held in January, 1983, at the University of Dortmund during a conference on theoretical computer science. Some 133 five-state Turing machines were entered. Uwe Schult of Hamburg won with a machine that produced 501 1's before halting. The state-transition diagram of the winning machine is shown in the bottom illustration on page 21. The runner-up was Jochen Ludewig of the Brown Boveri Research Center in Baden, whose Turing machine printed 240 1's.

Is Schult's Turing machine a busy beaver? Schult, along with Wankmuller and Ludewig, conjectures that it is. In other words, he suspects that no Turing machine with five states can produce more than 501 1's before halting. How could such a claim be proved? The answer lies in exhaustive search by computer, a search of the kind that Schult used to find his champion Turing machine in the first place. Before describing Schult's attempt to trap the five-state busy beaver in his computer, I should like to take a closer look at the function $\Sigma(n)$ to get some insight into why the Busy Beaver Game is so hard to play, even with the aid of a computer.

The function $\Sigma(n)$ has an extraordinary property: it is not computable. It simply grows too fast. From the first four values of $\Sigma(n)$—namely 1, 4, 6 and 13—it might seem that the rate of growth is only moderate. If 501 is indeed the maximum number of 1's for a five-state machine, the increase in $\Sigma(n)$ would still appear to be no faster than that of an exponential function. Schult has found a six-state Turing machine that produces 2,075 1's, which again suggests a quite tractable rate of growth. On the other hand, Schult has also found a 12-state machine that generates so many 1's that the number must be ex-



A Turing machine and its program



"Instantaneous descriptions" trace the operation of the Turing machine

pressed by the following mind-boggling formula:

$$6 \times 4096^{4096^{4096^{\cdot^{\cdot^{\cdot^{4096^{4}}}}}}}$$

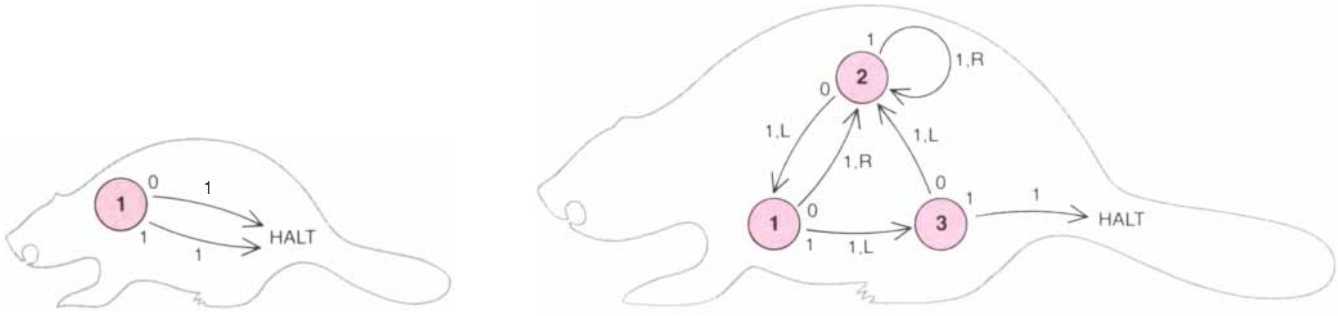
The number 4,096 appears 166 times in the formula, 162 times in the "twilight zone" represented by the three dots. The formula can be evaluated from the top down: first raise 4,096 to the fourth power, then raise 4,096 to the power of the resulting number, then raise 4,096 to the power of *that* number, and so on. When you reach the bottom, multiply by 6.

Anyone whose mind does not boggle when confronted by a string of 1's that long is welcome to construct an even bigger number. Write down any formula you like in which numbers are multiplied or raised to a power; you may even replace any of the numbers with n. No matter what formula you devise, for some value of n that is large enough the n-state busy beaver will produce more 1's than the formula specifies. It follows that $\Sigma(n)$ cannot be calculated for arbitrarily large values of n. The best one can do is to calculate $\Sigma(n)$ for some small, fixed value of n.

It is hardly surprising that the Busy Beaver Game is most often played with the aid of a computer. The essential method is to examine systematically all Turing machines with n states. Each time a new machine is generated its behavior on a tape filled with 0's is simulated. If the machine halts after no more than a specified number of steps, the number of 1's it printed is compared with the score of the "busiest" Turing machine found so far. From time to time a new champion is discovered.

This method of searching for the n-state busy beaver has two major flaws. First, the number of Turing machines to be generated is immense; for example, there are 63,403,380,965,376 five-state machines. Second, it is not known how long one should wait for a machine to halt; the maximum number of transitions an n-state machine can undergo (and still eventually halt), a function denoted $s(n)$, is itself a noncomputable number. Obviously $s(n)$ grows even faster than $\Sigma(n)$, since a Turing machine must make a state transition each time it prints a 1. As Hopcroft pointed out, computing $s(n)$ is equivalent to solving the halting problem for Turing machines, one of the first problems shown by Turing to be undecidable.

In 1982 Schult converted his Apple II personal computer into a busy-beaver trap. He augmented the computer's



Busy beavers with one state and three states

original central processor with a circuit board bearing a Motorola 6809 microprocessor; he wrote his search program in the machine language of the auxiliary processor. To test the vast numbers of Turing machines generated by the program Schult built an actual hardware Turing machine out of standard electronic components mounted on another circuit board that plugs into the Apple II. The device provides a simulated tape of 4,096 squares as well as registers for storing the program and the current state and head position of the Turing machine. Schult estimates that without such specialized hardware his search would have taken 20 months of computer time. Even with the hardware extensions the Apple II took 803 hours to find the winning Turing machine.

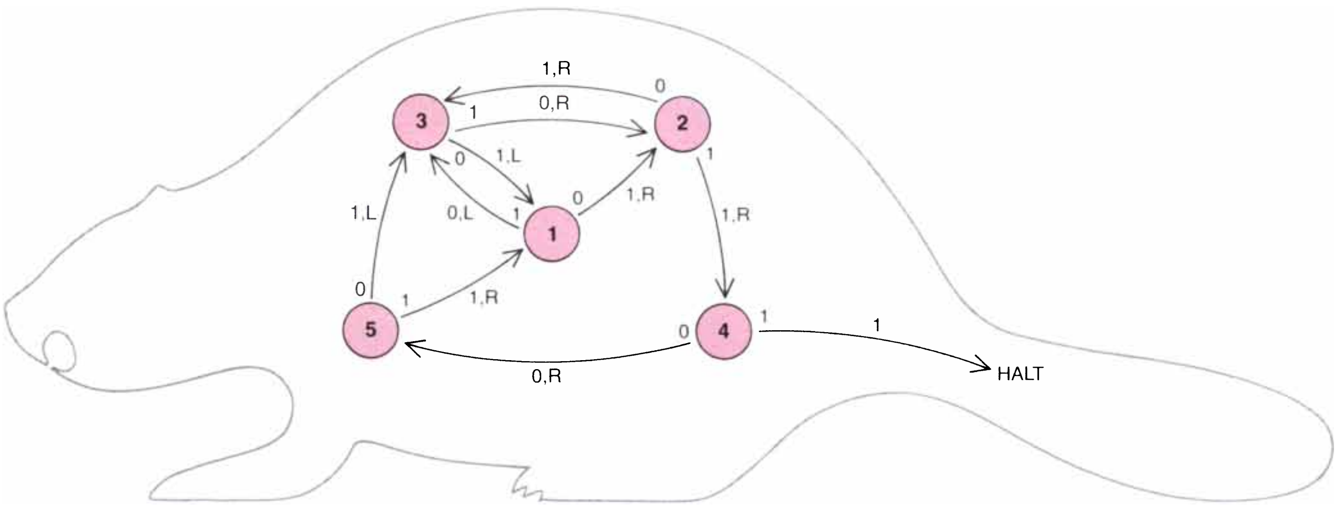
In designing the necessary software Schult also gained by making the search program and the Turing-machine hardware interact closely. The program systematically filled in the transition table for a five-state Turing machine in all possible ways. Even before a table was completed it was submitted to the Turing-machine hardware for testing. In many cases an incomplete table was found to specify a machine that ran out of time or space before any of the undefined entries was reached. Thus the incomplete table and all possible completions of it could be rejected.

Although Schult in large measure overcame the problem of managing multitudes of Turing machines, his approach to the halting problem for five-state busy beavers is not watertight, so to speak. In the absence of exact information about $s(5)$—the maximum number of transitions a five-state halting Turing machine can make—the number must be guessed. Schult set the limit at 500,000 transitions; in other words, he adopted the working hypothesis that if a machine had not stopped after 500,000 transitions, it never would. Of necessity he also imposed space limitations on his candidate busy beavers; since the simulated tape had only 4,096 squares and since his Turing machines always started at the middle of this finite tape, a candidate was considered a "runner" if it moved more than 2,048 squares from its initial position. A runner is a Turing machine that not only fails to halt but also continues indefinitely to visit new tape squares.

Of the 133 Turing machines entered in the Dortmund contest, only four produced more than 100 1's. The operation of each Turing machine was simulated with a Siemens 7.748 computer. More than an hour of processor time was needed to determine the winner.

Ludewig, the runner-up, wrote his busy-beaver search program in the Pascal programming language and ran it on a large minicomputer, the VAX, made by the Digital Equipment Corporation. In spite of a more sophisticated analysis of candidate Turing machines, 1,647 hours of central-processor time were spent in discovering his entry—the Turing machine that produced 240 1's. Schult, not surprisingly, also found Ludewig's machine; of equal interest, he found no machines between Ludewig's and his own. Apparently any halting five-state Turing machine that prints more than 240 1's must print at least 501.

Ludewig, in the course of his investigations, discovered a number of strange Turing machines with beaverlike behavior. Besides printing 1's there are other ways for a beaver to keep busy. For example, without printing many 1's a Turing machine may move a considerable distance from its starting square and then halt. Alternatively, without printing many 1's or even moving very far, it may go through a great many transitions before it halts. Among the machines tested at Dortmund, Schult's won in all three categories. On the other hand, Ludewig discovered three beavers that

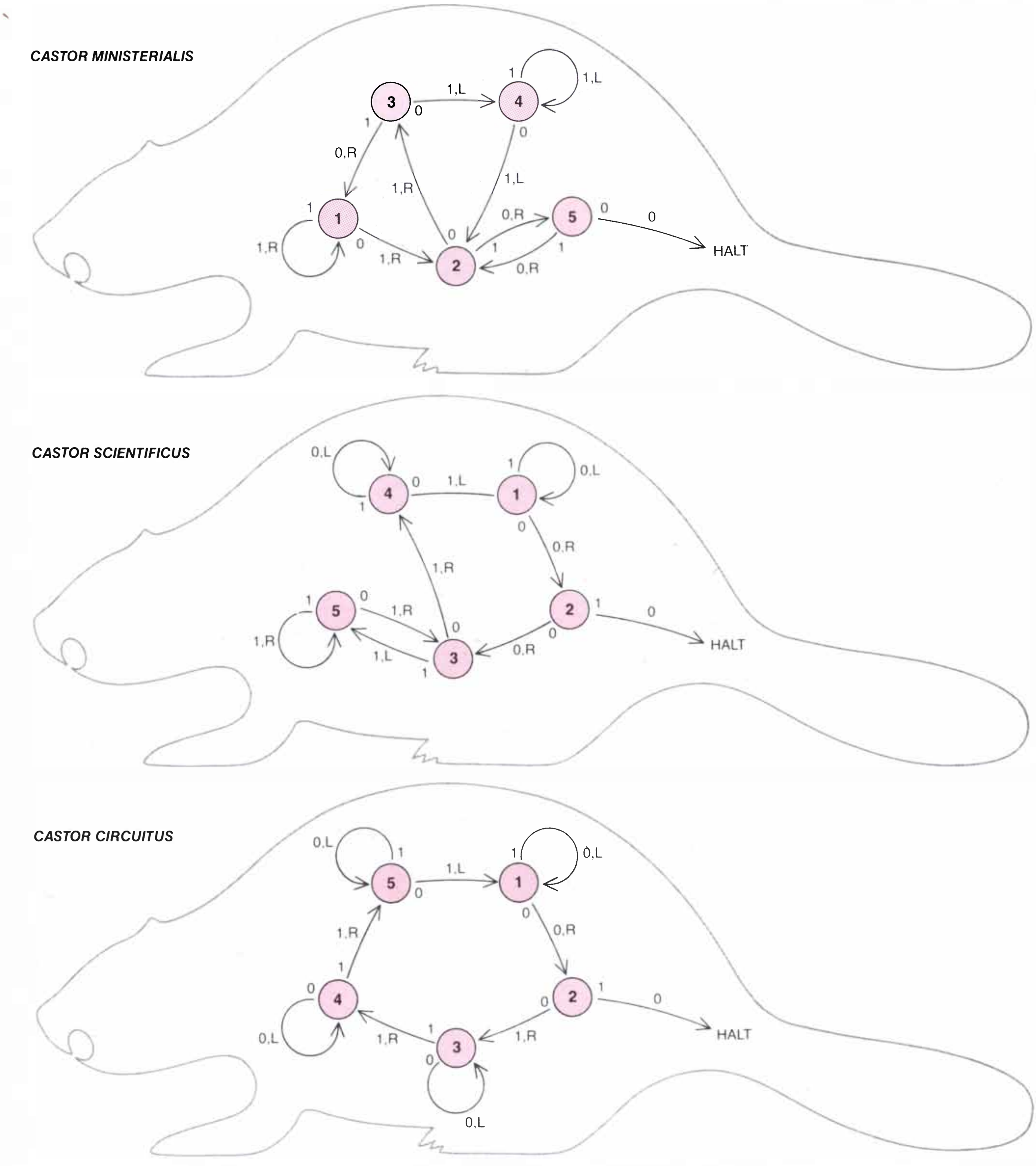


Uwe Schult's candidate for a five-state busy beaver generate no 1's at all but nonetheless either explore a wide territory or waste much time in profitless activity [*see illustration below*]. Accordingly three new species of beaver have been named:

Castor ministerialis (common name, civil-servant beaver). This enterprising creature seeks to advance itself as far as possible without producing anything. The type specimen is a five-state beaver that produces no 1's and moves 11 squares from its starting position.

Castor scientificus (common name, scientist beaver). Again without actually producing anything, this animal seeks to maximize its total activity, perhaps in an effort to attract grants. A five-state member of the species has been observed to make 187 transitions without writing a single 1.

Castor circuitus (common name, dizzy beaver). The dizzy beaver produces nothing and goes nowhere, but in the process it generates a maximum amount of activity. As the state-transition diagram suggests, it tends to spend a lot of time spinning its wheels. The busiest



Three new species of beaver that after much activity leave no 1's on the tape five-state specimen found so far undergoes 67 transitions before it finally halts exactly where it started.

It would be interesting to see some three-state examples of these odd beavers. Any attempt to find them would certainly benefit from the use of a computer (personal or otherwise), even if only to test Turing-machine programs devised in one's mind.

A Turing-machine simulator is easy to write. Use a one-dimensional array to represent the tape; the contents of the array, which consist exclusively of 0's and 1's, can be shown on the computer's display screen. The display is most informative if the position of the head is indicated. For example, the machine's current state might be displayed directly below the symbol being scanned.

A two-dimensional array is needed to represent the Turing-machine program. Each element of the array is a set of instructions for the machine; instructions must be provided for each state of the control unit and for each possible tape symbol. For a three-state Turing machine the array has three rows and two columns; its structure is exactly that of the program shown in the top illustration on page 20. The state of the machine specifies a row in the array, and the symbol under the tape head specifies a column; the instructions found at the intersection of the designated row and column define the Turing machine's next action.

Suppose the machine is in state 1 and the symbol on the tape is a 0. Consulting row 1 and column 0 of the array, the simulator finds the instructions "1,*R*,2." Hence the machine is to write a 1 on the tape, move the head to the right one square and enter state 2. One way of implementing such instructions is to define three variables, say STATE, HEAD and SYMBOL. At the beginning of a cycle the values of STATE and SYMBOL determine where in the table the machine looks for its next instructions. The first component of the instruction found there (in this case a 1) is written on the tape; the second component (*R*) becomes the new value of HEAD, and the third component (2) becomes the value of STATE. The head is then moved (in the direction indicated by the value of HEAD) and the symbol found at the new position is made the value of SYMBOL. The cycle then begins anew.

Various strategies can be adopted to make the programming of such a scheme easier and more efficient. For example, the letters *L* and *R* can be replaced by numbers, which are generally easier to manipulate in the computer. Moreover, the transition that leads to the halted state demands special treatment in the program.

A Turing-machine simulator could be used to test your answers to the following little puzzles, but it is by no means necessary to their solution.

Imagine you have bought a supply of used Turing-machine tapes at your local computer store. Before turning your busy beaver loose on them, the tapes must be cleaned up: any 1's on them must be changed back to 0's. Instead of cleaning the tapes yourself, you decide to devise a simple Turing machine to do the job for you.

One of the tapes has a single 1 on it but is otherwise filled with 0's. You must create a Turing machine that finds the 1, erases it (by changing it to a 0) and then halts. Naturally the fewer states your tape-cleaning machine has, the more elegant it will be. The tape cleaner in the illustration on this page is extremely elegant. Unfortunately it only works half of the time!

The remaining tapes are just like the first one except they have more 1's on them, although in each case the number of 1's is known to be finite. Can you construct a tape cleaner that changes all the 1's back to 0's? Of course, it will never halt.

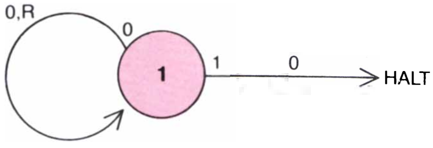
Responses to the May column on Core War ranged from simple requests for the supplementary guidelines on the game to descriptions of complete Core War systems already in operation. In between were numerous anecdotes about Creeper-like programs inhabiting real systems (including worms in Apples), discussions of programs as genes and speculations about defensive and offensive strategy. Only a few important developments can be mentioned here; others will have to wait for a future column on the subject, which I hope will appear before the end of the year.

I have been told by Douglas B. McIlroy of AT&T Bell Laboratories that it was not he but Victor A. Vyssotsky of the same institution who invented the game Darwin. McIlroy did, however, invent an unkillable organism.

What happens when an Imp runs into a Dwarf? One possibility was explained in the May column: Dwarf transfers control to Imp's code and becomes a second Imp endlessly chasing the first one. Another possible outcome has the opposite effect. Suppose Dwarf has just jumped back to its first instruction when Imp copies itself over Dwarf's data location. The situation is then as follows:

```
  Imp →MOV   0     1
Dwarf →ADD  #5    −1
       MOV  #0   @−2
       JMP  −2
```

Since it is Dwarf's turn to execute, it adds 5 to Imp's code, turning it into MOV 0 6. Then Imp executes, copying itself six spaces ahead, well clear of Dwarf, which then bombs its next address (specified by the numerical code



A tape-cleaning machine

corresponding to MOV 0 6). On Imp's next turn something curious happens: it executes the first line of Dwarf's program, so that for a time the game is played by a "double dwarf" pointlessly shooting up the core array while the object of its attack inhabits its own body and does exactly the same thing!

David Menconi of Milpitas, Calif., a game designer at Atari, Inc., has suggested making this very phenomenon a regular feature of Core War by allowing each battle program to execute in two places at once. Thus even if a program loses one "self," a second self might be able to repair the damage. Edsel Worrell of Bethesda, Md., suggests the somewhat more general scheme of *n* selves, all executing the same program at different addresses.

Robert Peraino of George Mason University wrote a Core War system for the Apple II+ computer, compensating for the machine's small word size by using a two-dimensional array of 2,000 by two bytes. Bill Dornfield of AMF, Inc., wrote a complete Core War system in extended BASIC on a Hewlett-Packard 9816/26 desktop computer.

The most impressive system to date was constructed by three graduate students: Gordon J. Goetsch and Michael L. Mauldin of Carnegie-Mellon University and Paul G. Milazzo of Rice University. Mauldin demonstrated the program on a VAX computer in my department at the University of Western Ontario. In an impressive screen display the entire core array is shown, with the position of each contending program marked by a capital letter and the areas affected by the program marked by the corresponding lowercase letter.

Mauldin has invented a battle program called Mortar that operates like Dwarf except that its bombs are directed according to the sequence of Fibonacci numbers (1, 1, 2, 3, 5 and so on, each number being the sum of its two predecessors). Oddly enough, Dwarf beats Mortar 60 percent of the time, but Mortar invariably kills a three-part self-repairing program called Voter. On the other hand, Voter survives attacks by Dwarf and regularly defeats it.

Goetsch, Mauldin and Milazzo have analyzed Mortar and conclude that if a battle program is longer than 10 instructions, it must be self-repairing in order to defeat Mortar. No program longer than 141 instructions, however, can repair itself fast enough to survive an attack by Mortar.

Where the

Today's engineers and scientists keep proving one thing again and again. And that is to do highly technical work you need highly technical help.

Like The Smart Desk from IBM.

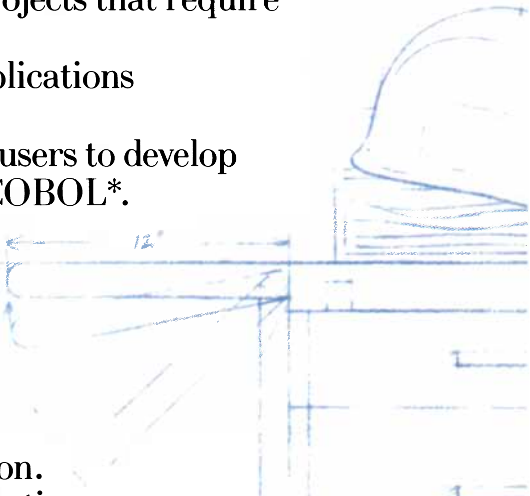
The Smart Desk is any desk with an IBM computer or terminal on it. And with IBM's powerful System 9000, your desk can be a match for many technical tasks.

The 9000 can tackle a wide variety of applications including numerical control, design engineering, process control, laboratory automation and data acquisition.

Designed for multifunction capability in a real-time environment, the 9000 is based on a 32-bit internal architecture. The system can prioritize operations and has programmable function keys to handle projects that require an array of repeatable tasks.

As for languages, the standard operating system runs applications written in FORTRAN 77, Pascal or BASIC.

An optional operating system, XENIX*, allows up to four users to develop multitask applications in C, FORTRAN 77, Pascal and RM/COBOL*. Also, many engineering, scientific, and business applications available on UNIX* can be made to run on XENIX.



The 9000 can have up to 5 megabytes of RAM and up to 40 megabytes of online storage. And its display and optional printer can handle high resolution graphics.

Put yourself behind The Smart Desk. Because these days, what you work at should be just as advanced as what you work on.

IBM 8-84
Dept. 2U3/428
400 Parson's Pond Drive
Franklin Lakes, NJ 07417

I am interested in using the IBM System 9000 in:
☐ laboratory, ☐ science, ☐ engineering, ☐ other________

☐ Please send me free literature on the IBM 9000.
☐ Please have an IBM representative call me.

Name____________________

Company__________ Title__________

Street Address____________________

City________ State________ Zip________

Phone____________________

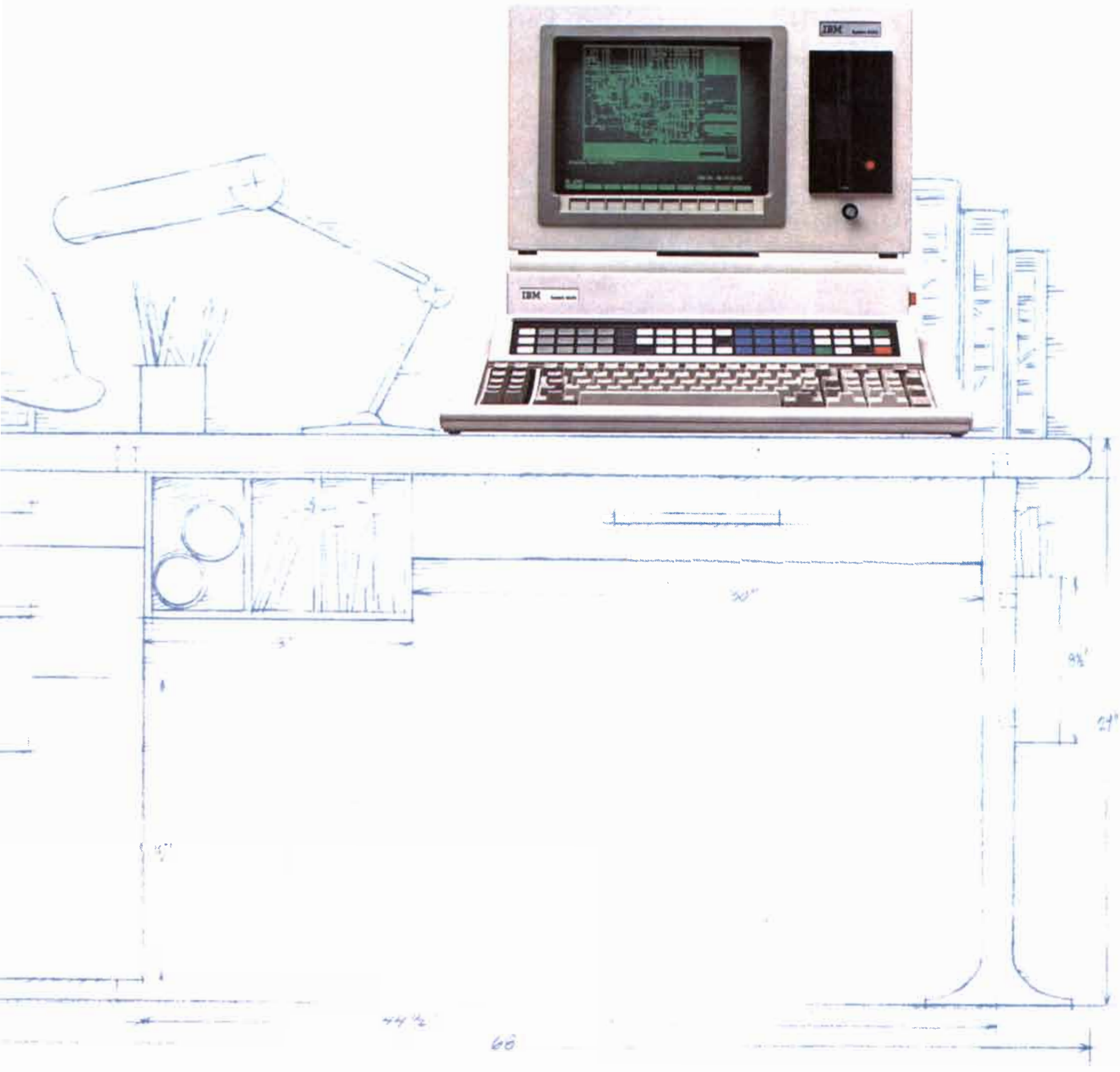
For more information, just fill out the coupon, or call 1 800 IBM-2468. Ext. 428

IBM®

*XENIX is a trademark of Microsoft Corporation and is derived from UNIX under license from AT&T. UNIX is a trademark of Bell Laboratories. RM/COBOL is a trademark of Ryan McFarland Corp.

The Smart Desk from IBM

the talented tackle technical.



BOOKS

The rain forest, Darwin's orchids, textile machinery, Harappans, diverse climates

by Philip Morrison

TROPICAL NATURE, by Adrian Forsyth and Kenneth Miyata. Illustrations by Sarah Landry. Charles Scribner's Sons ($16.95). THE VARIOUS CONTRIVANCES BY WHICH ORCHIDS ARE FERTILISED BY INSECTS, by Charles Darwin. University of Chicago Press ($9.95). Our northern forest floor is carpeted for a footsore hiker: soft deep loams fragrant with humus and springy duff under the trees feel easy underfoot. In the tropical rain forest, where winter never comes, walking is harder. There the litter is thin, typically a couple of inches in typical forests; decay is simply too swift under year-long steady moist incubation. A single good scrape of the boot will disclose the pallid mat of fungal mycelia that dwell in the soil, a lowly root symbiont of the towering trees. The heavy leaching at the rain-drenched surface would carry away all the soluble minerals if that mat did not catch and recycle key ions.

Those mycorrhizal associations live a fast biochemical life. As in an automobile plant in Japan, the steady state of production maintains a frugal inventory: things are used right away. Even the lush green canopy above, where the solar input is made, is not quantitatively as rich a store as it is in the boreal forest by summer. But it is in those leaves on high, and most of all in the great supporting cellulose cylinders of the tree stems, whether erect or new-fallen, that the main stores reside, the essential atoms and the light-derived chemical free energy for all the life of the forest.

"Consider the sloth of the sloth," write the authors of *Tropical Nature,* the first of these two books. That proverbial beast is genuinely lethargic; its metabolism runs at about half the norm for mammals of its size. It is a slow, steady grazer on tree leaves. Herbivores of the forest understory are rare; grass and herbs do not grow in that deep shade. A sloth ranges little; over months one individual may feed within only a few dozen big trees. A sloth prudently avoids the ground, where most carnivores prowl. In its leafy canopy it is hard to see; its fur is camouflaged, greened by algal growth. Its fecal pellets are small, hard and infrequent, as befits an animal that must live high overhead where water is hard to come by. Yet it does not drop the pellets "the easy way" from above. No, the beast laboriously descends each week from its high tree, pokes a hole in the ground at the base of the tree with its stubby tail and carefully buries its scat. The howler monkeys, which range far in the same forests and browse on many of the same leaves, defecate freely from the treetops. The agile monkeys also venture most reluctantly to the ground, even though the watchful eyes of the entire howler band are at the service of any single member. The solitary sloth has no sentinels.

Why? It turns out that a sloth not only dwells within a small set of trees but also spends disproportionate amounts of time grazing in a single tree. One or two sloths in a tree can take as much as a fifth of the annual product of that tree in the course of their deliberate grazing. If they dropped their pellets from 100 feet up, it is easy to believe the stuff would scatter widely as it fell through the vines and branches below. The single sustaining tree could not much benefit from the recycling of the sloth's nutrient. The sloth's feces may return about half of the nutrient to the tree roots. It is a long-term investment by the sloth tribe through a risky and exacting husbandry of the resource, one that is less important to the wide-ranging howler monkey. The scenario is untested, but the authors "like to believe that in the arduous, dangerous descent of the sloth there is a metaphor."

We who saunter in parks and gardens notice easily that plants protect themselves against animals that eat them by visible structures such as spines and thorns. Animals that live by gathering, from our sapient ancestors to Baird's tapir, the largest herbivore in the New World Tropics, learn well that the chief defenses stoutly erected by plants are molecular. Hallucinogens are the innovations not of "clandestine Berkeley chemists" but of fungi and plants seeking cheap ways of discouraging small mammalian herbivores. If 13 seeds of a morning glory can cause wild hallucinations in a man with a body weight of 70 kilograms, consider what a single seed is apt to do to a mouse with the same neurochemistry.

But the powerful alkaloids are expensive for plants, particularly in the nitrogen-short Tropics. A plant that colonizes the neighborhood of tree falls, where for a few years sunlight reaches the ground, may sacrifice chemical defenses for fast growth. It has no time to spare. A balsa tree will grow rapidly toward a gap in the canopy, its leaves pocked with insect-chewed holes, four-tenths of the leaf area gone.

The tapestry of the forest bears such an intricate design that plant-animal interaction can on occasion reverse the chemical conflict. A striking example is the relation between certain colorful bees and fragrant orchids. Male orchid bees visit and pollinate a variety of orchids, but these nonsocial bees have no use for the pollen they carry away and the flowers offer them no nutrient nectar. To summon the bee species the flowers manufacture a mix of fragrances, often rather simple aromatic chemicals. The bees build up a store of varied scents kept in specialized leg pockets. Once the mix is right, the perfumer becomes irresistible. He does not, however, attract the wanted female, as they might have planned it in Paris. No, the sweet-smelling bee gathers a small crowd of males of his species, which buzz and flitter around their beau. Then in some favorable spot of sunlight the bright, dancing aggregation draws from far around those orchid-bee females for which the entire elaborate scheme has coadaptively come to be.

In 17 chapters, each a brief essay on tropical nature observed, these two young field biologists have made a model of contemporary natural history, cheerfully speculative, concerned as much with large pattern as with diversity, chemically informed, thoroughly ecological and Darwinian to the core. Most of the chapters were not jointly written, but their book is nonetheless a whole. They have for a dozen years visited one of "the most complex biological communities that exist," the neotropical rain forest, particularly in Costa Rica and in Amazonian Ecuador. The book is based on what they saw and tasted and heard there, although it is by no means only a journal. They know and call on the relevant specialized literature of today and yesterday, although never technically, but the work is not a textbook either. Its epigraphs are drawn from pop songs as well as from William Blake; the style is of our years, although the authors warmly praise and well employ their great Victorian predecessors.

The larger issues, like the reasons for the tropical diversity they celebrate, are examined, but large answers will remain tentative until that living fabric is more intimately known. Each chapter opens with a handsome drawing of some or-

ganism from the chapter text. The authors supply an appendix for the beginner, with knowing advice on how to make your own trip to these remarkable forests, from the choice of tours to toilet paper (always travel with a roll) and clothing (do not bring much). It is sad to learn that Kenneth Miyata drowned last fall in a rapid of the Big Horn River, the galley proofs of this fine book unseen on his desk at home.

The paperback reissue of Darwin's first published book after his *Origin* (the reprint is actually that of the second edition of 1877) brings the reader to the same kind of detailed example, to repeat the title: the various contrivances by which orchids are fertilized by insects. The argument is founded on evolutionary inference, drawn mainly from homology and observed function, and tested experimentally. We see it here as it looked in the vigorous youth of evolutionary science. The contesting explanations of the day seem frivolous: "to complete the scheme of nature," say. Darwin's work, pushing a flower hinge with a fingertip or some dangling filament with a human hair, or placing small bees within a flower cavity to watch their escape, is entirely different. There are no appeals to final causes or ideal schemes, only, as Michael T. Ghiselin writes in his foreword, "reconstructed history and laws of nature."

As in all historiography, successful reconstructions are tentative but wonderfully fruitful. It was Darwin who first found those remarkable devices of the orchid. Contrivances is an apt term, although it may have had a certain allusive irony for those who first saw the title. One orchid manages to paste a pollen mass onto the right part of the right bee, to be removed by a well-placed sticky cleft when the bee in time visits the right female flower. So different are the male and female blossoms of the orchid genus *Catasetum,* intricately formed to carry out the vital implantation and removal, that they were regarded as distinct genera until they were found growing on a single plant. The question Darwin puts, as do his powerful modern colleagues, is always: What can have happened to bring about the state we uncover by a close look?

CONTEMPORARY TEXTILE ENGINEERING, edited by F. Happey. Academic Press ($74). The historical incubator of the factory system and then of power-driven machinery, particularly the transition from millpond to steam, the textile industry is still a part of daily life in every land, even where no machines whir at all. Each year the worldwide production of fabric weighs about as much as that of automobiles. The technology is little known, however, outside the circles of the industry itself; the casual reader knows about new fibers, of course, but otherwise the ubiquitous output of the industry seems not much different from that of the days of the spinning jenny and the mule.

That impression is not entirely mistaken; in this industry, unlike, say, the chemical or the automobile industry, change is seen less in the product than in the fabrication. A hand-spun, hand-woven, cotton cloth functions as well as (perhaps better than, allowing for the aesthetic dimension) the output of the newest automated mills, albeit at a high cost. The spinning wheel improved on the productivity of the dangling spindle of the skillful spinster by a factor of four, and power spinning with the machines of the early 19th century made an eightfold gain on top of that. Modern techniques have inserted a factor of several hundred more, of course at heavy capital cost for machines that can spin delicate fibers onto 100 spindles whirling 100,000 revolutions per minute.

This book will alert the general reader to an unexpected world of engineering virtuosity, a domain where fundamental limits to mechanism and motion are steadily approached under heavy economic pressure. Topology is at home there along with dynamics, and fine mechanisms abound. The volume offers a dozen chapters or so by as many experts, mostly British, but including authors from the U.S., Switzerland and Australia. Each chapter reviews the state of the art in some important subdivision of the industry and summarizes recent developments in theory and practice. Any technology so mature is bound to have its jargon, and this one can have pride of ancient usage; otherwise the chapters are clear and information-rich. The authors intended to serve engineers, technicians and scientists in or around the textile trade; others will have to be nimble in using context to understand words of art, such as slivers, pirns and doffers.

The book is above all a book of marvelous and daring mechanisms; it does not much enter the world of molecular technology, for producing fiber and adding color. Take it that a large quantity of short lengths of some fiber is given to you; the engineer's task is the swift increase of dimensionality, first spinning those quasi-points into a strong linear thread and then weaving many of those long lines into a two-dimensional web. In this volume is a view of how that can be done on a large scale today, and perhaps tomorrow.

By the middle of the 20th century the spinning jenny and the mule were gone or were found only in specialized habitats. All over the world cotton thread is spun by ring spinning, "the established order," some 150 million spindles at work daily, an enormous capital investment, nearly half of it in Asia. More than half of the cost of turning fiber into yarn is absorbed in spinning alone. Speed is limited by the brute fact that the continuous strand must be taken up (or removed) from some rotating package if the indispensable twist is to be set into the fast-thinning yarn.

In the ring spinner the yarn is twisted as it passes over a smooth, circular ring around and around which the central rotating spindle drags a light traveler along with the yarn. The speed of the fragile yarn is limited by friction and air drag. The radius of the take-up spindle must be kept small to reduce centrifugal tension on the delicate fiber. A speed increase by a factor of a few has been coaxed out of this system by good engineering since its early Victorian beginnings; modern ring spindles, driven by variable-speed electric motors, run at 10,000 or 15,000 r.p.m. Their low-radius spindles can pick up, however, only a few ounces of yarn; elongated spindles raise more difficulties than they cure.

There is an ingenious way out, long suggested, first developed into a practical rotor machine by Czech engineers in about 1970 and now found in many variants. A couple of million rotors go into service each year, every one the productive equivalent of several ring spindles. In the rotor spinner the Gordian thread is cut. The flux of fibers is reduced by a very high rate of flow to the thickness of only a few fibers. Indeed, the probability of zero fibers is quite high. Allowing for such breaks in continuity means that neither the fiber supply nor the take-up package needs to rotate rapidly, even though fiber is twisted at a high rate. Airflow drags fibers a few at a time into the fast-spinning rotor, where many fibers are twisted together and taken up at modest speed and low force.

A high-speed photograph of the yarn as it forms within a transparent rotor shows the process. Matter is conserved; like flowing water, the fibers must move fast when they are few but can slow down as the thread thickens. The fiber count drops by a factor of 10,000 before the fiber enters the rotor, and it increases there by a factor of 100 or so; then the fiber is stored as yarn, condensed by a factor of 100 from the feed of sliver: loosely wound cotton staple. Along the way the thread receives a useful twist of, say, one turn per millimeter.

Weaving too is increasingly done on a new generation of high-speed automatic machines. The swift shuttle carries the weft thread to and fro, interlacing it with the selected set of stationary or slow-moving warp threads, the shed. In the ancient hand loom the weaver might insert 12 meters of new weft per minute; the power looms of 1785 increased that by 25 or 30 times for each machine, and a weaver could tend half a dozen looms, his or her work being mainly the refill of the shuttle and the repair of broken threads. J. H. Northrop's automatic shuttle loom ended 90 years ago the loom tender's heaviest task, reshuttling, so that one person could tend dozens of looms, mainly for patrol and repair.

A transition is under way. The world weaves most of its cloth on about three million power looms. (It has always taken many spindles to feed one loom.) Most of the looms are modern versions of the Northrop machine, now well engineered and fitted with electronic stopping controls. Their replacement is not fast, say 1 or 2 percent per year. Over the past decade, however, nearly all new weaving machines delivered in Europe and the U.S. have been shuttleless. Such a loom no longer parades its entire supply of weft from side to side until the shuttle is empty. That inert load makes all those stops and starts of the heavy loaded shuttle costly in energy terms; emergencies require even quicker stops when the yarn breaks. Yet if the shuttle load is kept light, shuttle refill must be more frequent, and so the fast weaving tempo is lost.

By far the most used loom without a shuttle is the product of the Swiss firm of Sulzer, which introduced its projectile loom three decades ago. Its machine throws a projectile to carry the end of the weft yarn within guides lined up between the warp threads across the width of the cloth. Neither the projectile nor the weft yarn touches the warp in passage. A conveyor brings back the projectile under the cloth after the single yarn width is cut off. The cut ends are tucked by a special unit into the next shed.

A weft projectile flies across a few yards' width at 40 miles per hour. Still newer and faster looms carry the weft across the warp width with a fierce jet of water or air. The nozzle blows a length of weft across the cloth width each time the yarn is measured out and cut to length. Air is free, but high-speed jets able to drive the yarn fast are costly in power; water jets are mainly suited to nonwetting fibers and require controlled viscosity and scrupulous cleanliness, with no algae in the water and no objectionable algicides either. The spray and surplus water have to be sucked away. Yet the author of this chapter believes the jet machines may dominate weaving-machine production by the end of the 1980's. Jet looms are no longer limited by the speed of weft insertion; the other steps of the process will need speeding up if the decline of human effort per kilogram of spun or woven yarn—the two go together neatly—is to continue its steady course.

There is more to fabric production, and to this volume, than spinning and weaving. Other chapters treat many interesting auxiliary processes and specialties, from recycling systems to the statistics of fibers, the optical and mechanical properties of yarn, the application of electronic sensors and microprocessors, knitwear, piled and tufted fabrics, and even the operations analysis of yarn breaks in spinning and weaving. There are plenty of people at work even in the most automated mill; one misses a chapter on the engineering of suitable conditions of work in that noisy, at least micro-dusty and occasionally dangerous environment filled with fast-moving machinery. At one time the unmodified Lancashire climate suited the humidity demands of most cotton fabrication, but these days cotton is worked in quantity from Quebec to Surabaja. If the weaving room is to be air-conditioned to control jet properties, weavers worldwide may yet have a healthier time.

HARAPPAN CIVILIZATION: A CONTEMPORARY PERSPECTIVE, edited by Gregory L. Possehl. Distributed in the U.S. by Humanities Press, Inc., Atlantic Highlands, N.J. 07716 ($45). The first director general of the Archaeological Survey of India wrote of one site in the West Punjab, now in Pakistan, in about 1875: "Perhaps the best idea of the extent of the ruined brick mounds of Harapa [*sic*] may be formed from the fact that they have more than sufficed to furnish brick ballast for about 100 miles of Lahor [*sic*] and Multan Railway." In spite of "this abominable act" that dismantled city, named for a nearby village close to one of the five rivers that flow together to form the braided Indus, gave its name to an entire civilization, more extensive than any other of the Bronze Age. Over the past 60 years 1,000 ruined Harappan settlements have been found scattered across the map of southwest Asia from the Hindu Kush nearly to Bombay, and from west of Karachi almost to Delhi.

This rich, argumentative and fascinating volume published in 1982 presents some 40 papers by archaeologists from America, Europe, India and Pakistan. New material is abundant: a single paper announces and locates no fewer than 370 new sites clustered in the desert of Cholistan, one of which has the same area as the largest Harappan city known: Mohenjo-daro, southward along the Indus in Sind. In its day Mohenjo-daro was home to 40,000 or 50,000 citizens.

The roots of Harappan culture have deepened in time as well; as late as 1977 evidence of agriculture on the Indian subcontinent went back only 5,000 years. Now a Neolithic village has been uncovered at the western fringe of the Indus valley. It was a permanent settlement, beginning before 5000 B.C., with many layers of mud-brick structures and a long record of steady change from flint tools to sickles and from copper beads to the coming of pottery thrown on the wheel. Barley and wheat cereals were early cultivated and ground to flour, although it is not known whether it was fiber or oil that was the aim of those who there planted cotton. The upper levels yield stamp seals and display the mass production of drilled soapstone beads and terra-cotta figurines, merging recognizably into the early Harappan before Mehrgarh was abandoned in about 2500 B.C.

It is not easy to endure in those parts. The hundreds of newly found small sites clustered in the extreme aridity of the Cholistan desert were not built amid the dry mud flats and sand dunes where now they sleep. They were established rather on the fertile floodplain of a mighty perennial river, the Hakra or Ghaggar, as it is now called in Pakistan and in India. For 300 miles the settlements flank the vanished river, whose dry bed averages five miles across. There is much uncertainty over the past climate of the region. It is quite nonuniform even today, and it seems beyond question that at least some watercourse changes were important over the thousands of years.

Whether Mohenjo-daro was brought to sudden ruin by invading Aryan warriors who worshiped Indra, "the fort-destroyer" of the Vedas, or by a hinterland flooded and muddied by the tectonically disturbed and impounded Indus can both be doubted. The earlier evidence is now in question. For example, the massacres once reported in Mohenjo-daro are hard to document; only two of thousands of skeletons found show signs of wound trauma, and neither armor nor weapons are to be found. (Indeed, most of the thoughtful theoretical essays in this book seek to refute the views of the pioneer interpreters of the Indus civilization that it had a late, sudden rise, a static course, a swift, catastrophic decline and no clear heritage.)

It can hardly be doubted that an agricultural civilization whose chief watercourses display "the extreme vagrancy of the rivers in the Indus and the Ghaggar valleys" will not finally be understood without a secure grasp of its environment in the past, both on the local scale and in the large. Two physical scientists, D. P. Agrawal and R. K. Sood, end their account of the inconstant ecological past they have discerned by satellite photography and other geomorphological means with a fresh suggestion. Harappan settlements, even the small ones, seem more like towns than villages; they are more or less dominated by craftsmen and administrators. The people on the land had perhaps to shift locations often as vagrant waters wandered over the floodplains. The urbanism of the Harappan world might have involved some unusual adaptation of that kind, a tendency to a diminution of the traditional institution of the settled village in favor of little towns as working centers for a more scattered and mobile farming population.

Big questions are not ripe yet for

broad and final answers. One marvelous small entry in the history of technology is not to be slighted. "Harappans loved ornaments," begins the paper by three scholars at the University of Baroda. Beads are found in abundance at nearly all Harappan sites (as they are in today's India). Half a dozen well-studied sites yield very regular hard white cylindrical microbeads by the tens of thousands; multiply twisted strands of them would make wonderful decorations "à la Ajanta." The microbeads are only one millimeter to three millimeters long, each with a neat circular perforation only half a millimeter in diameter. How were they made? Drilling and turning would demand enormous labor and keen eyes indeed: 300 beads make up only a gram.

It turns out on chemical analysis, microscopic examination and X-ray-diffraction study that the microbeads were not turned or drilled but were extruded under pressure as a soft wet paste made from fine-ground naturally occurring talcose steatite. The perforation was probably made by a centered mandrel of fine copper wire over which the paste emerged like so much macaroni as it was squeezed through a hole. A horsehair would serve to cut off bead after bead as it came through; the job could be managed by three painstaking people working together. Baking in a kiln at 900 degrees Celsius hardened and finished the beads beautifully. The authors have sketched a simple, plausible device consisting of cloth, a copper disk and wire that would make microbeads, entirely within known Harappan technology.

Where is the legacy of the Harappans? The newest work on the mysterious script of Harappa (reported in this journal by Walter A. Fairservis, Jr., of Vassar College in March, 1983) supports older conjecture. The Harappan language was an early Dravidian tongue; the influence lives within village India today, still strong in the art and fable of the people under a polyglot overlay of 35 centuries. The tiger, like the bull, the elephant and the buffalo, is depicted on Harappan seals; the lion is not. A subtle commentator has seen a dual metaphor in this: the maned lion on the antique emblems of the modern Indian state is Aryan, Vedic. It evokes the sun-bright steppe in a clear sign of overt power and authority. The tiger instead prowls the darkness of the Indian forests, her older fiery powers at once intimate and irresistible.

WORLD-CLIMATES, by Willy Rudloff. Wissenschaftliche Verlagsgesellschaft, Stuttgart. Distributed in the U.S. by Heyden & Son, Inc., 247 South 41 Street, Philadelphia, Pa. 19104 ($96). Planning a trip to Perth in West Australia, where the swans are black and the yachts swift? The seashore climate is inviting enough: it can be described as a subtropical winter rain climate, with hot summer maximums, about 100 Fahrenheit, and mild winter minimums, with no freezing weather. Sunshine averages 64 percent over the year, with the sunniest month November, about as much sun as Los Angeles' brightest month. The tourist can expect heat stress frequently in the southern summer months, and should dress accordingly, with tropical wear in January and February. In winter the rainfall is heavy; it is hard to match the Perth rain pattern in the snowfree parts of the U.S.

Such a game can be played out in this volume among nearly 500 pages of tabulated climatic data, with monthly temperature maximums and minimums and precipitation information given for almost 1,500 locations in 140 countries. There are special remarks about each country and climatic area, and a couple of summary pages for each continent. Quite a few oceanic islands are included (both the tempting and the desolate), a fair sample of Arctic stations (both on the Canadian side and on the Russian) and one chilly page on Antarctic climate (not one monthly mean above freezing for six listed stations).

Exhaustive data of this kind are available to the professional meteorologist in a variety of publications, mainly multivolume official tables, the primary source of this compilation. It is the author's purpose to present the rich worldwide information in a form simple enough for easy use, and to adjoin to the numbers a set of hints on comfort and clothing. These rather subjective-sounding qualitative entries are actually well defined and persuasive, if somewhat aimed at the European traveler, because they include complete scales of gradation that help a reader to judge their meaning well.

Almost 100 pages provide a readable review of the principles of global climate, meant for the general reader, accompanied by maps and tables usually found only in the specialized texts. World maps of thunderstorm activity, snowfall, fog, runoff, tropical storms and atmospheric pressure and its changes are all full of interest. The largest belt of the heaviest rainfall is at sea, just north of the Equator eastward from Java; thunderstorms are most at home in West Africa; the Persian Gulf and the Gulf of Oman demand the most summer cooling for human comfort; summer fogs dominate both polar regions, but away from the poles only the regions of the Grand Banks, the western end of the Strait of Magellan and coastal southwestern Africa grope through more than 80 foggy days per year.

This is a carefully and devotedly produced reference digest, about as good for the browser as for brisk functional consultation; only the steep price stands in the way.

Established 1845 SCIENTIFIC AMERICAN August 1984 Volume 251 Number 2

The Climatic Effects of Nuclear War

New findings tend to support the view that the immense clouds of smoke and dust raised by even a medium-scale nuclear war could bring about a global "nuclear winter"

by Richard P. Turco, Owen B. Toon, Thomas P. Ackerman, James B. Pollack and Carl Sagan

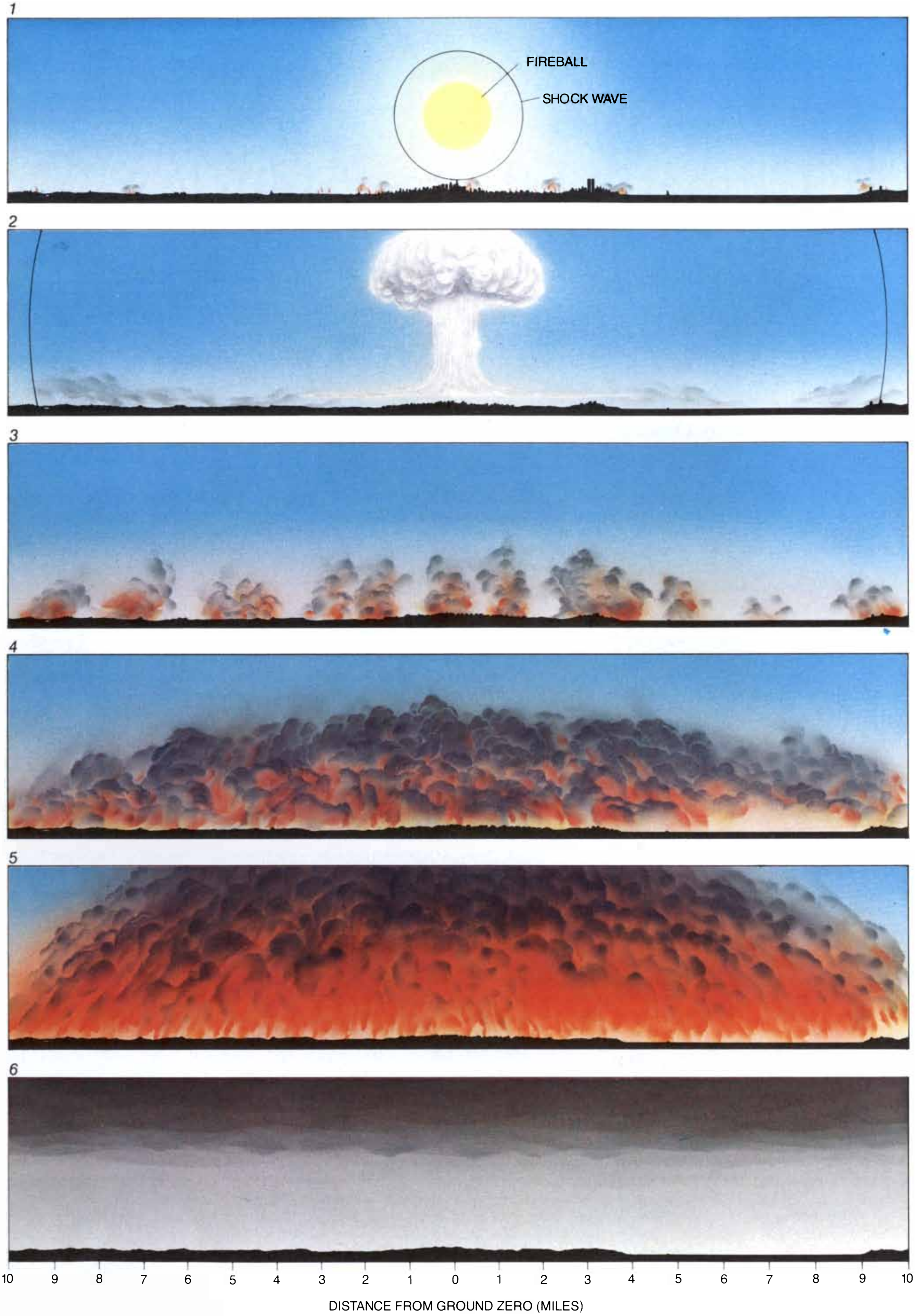
Since the beginning of the nuclear arms race four decades ago it has been generally assumed that the most devastating consequence of a major nuclear war between the U.S. and the U.S.S.R. would be a gigantic number of human casualties in the principal target zones of the Northern Hemisphere. Although in the wake of such a war the social and economic structure of the combatant nations would presumably collapse, it has been argued that most of the noncombatant nations—and hence the majority of the human population—would not be endangered, either directly or indirectly. Over the years questions have been raised about the possible global extent of various indirect, long-term effects of nuclear war, such as delayed radioactive fallout, depletion of the protective ozone layer in the upper atmosphere and adverse changes in the climate. Until recently, however, the few authoritative studies available on these added threats have tended to play down their significance, in some cases emphasizing the uncertainty inherent in any attempt to predict the combined effects of multiple nuclear explosions.

This comparatively optimistic view of the potential global impact of nuclear war may now have to be revised. Recent findings by our group, confirmed by workers in Europe, the U.S. and the U.S.S.R., suggest that the long-term climatic effects of a major nuclear war are likely to be much severer and farther-reaching than had been supposed. In the aftermath of such a war vast areas of the earth could be subjected to prolonged darkness, abnormally low temperatures, violent windstorms, toxic smog and persistent radioactive fallout—in short, the combination of conditions that has come to be known as "nuclear winter." The physical effects of nuclear war would be compounded by the widespread breakdown of transportation systems, power grids, agricultural production, food processing, medical care, sanitation, civil services and central government. Even in regions far from the conflict the survivors would be imperiled by starvation, hypothermia, radiation sickness, weakening of the human immune system, epidemics and other dire consequences. Under some circumstances, a number of biologists and ecologists contend, the extinction of many species of organisms—including the human species—is a real possibility.

Our own involvement in the reassessment of the global effects of nuclear war originated in a confluence of several lines of inquiry. Before joining forces we had separately and collectively been engaged in research on such phenomena as dust storms on Mars and the climatic effects of explosive volcanic eruptions on the earth; more recently we all became interested in the hypothesis that one or more of the mass extinctions of species evident in the geologic record were caused by immense clouds of dust raised by the impact of an asteroid or a comet. In 1982 a committee of the National Academy of Sciences, recognizing the parallels between the dust raised by nuclear explosions and that raised by other cataclysmic events, such as volcanic eruptions and meteorite impacts, asked us to look into the possible climatic effects of the dust likely to result from a nuclear war. We had already been considering the question, and to address it further we had at our disposal sophisticated computer models of both large- and small-scale atmospheric phenomena; the models had been developed over the preceding decade primarily to study the origins, properties and effects of particles in the atmosphere.

At about the same time another important aspect of the question came to our attention. An article in the Swedish environmental journal *Ambio,* coauthored by Paul J. Crutzen of the Max Planck Institute for Chemistry at Mainz in West Germany and John W. Birks of the University of Colorado at Boulder, pointed out that fires ignited by nuclear explosions could generate massive amounts of smoke, severely attenuating the sunlight reaching the ground. Accordingly we added smoke to dust as a likely perturbing influence of nuclear war on the climate.

In brief, our initial results, published in *Science* in December, 1983, showed that "the potential global atmospheric and climatic consequences of nuclear war... are serious. Significant hemispherical attenuation of the solar radiation flux and subfreezing land temperatures may be caused by fine dust raised in high-yield nuclear surface bursts and by smoke from city and forest fires ignited by airbursts of all yields." Moreover, we found that long-term exposure to nuclear radiation from the radioactive fallout of a nuclear war in the Northern Hemisphere could be an order of magnitude greater than previous studies had indicated; the radioactivity, like the other nuclear-winter effects, could even extend deep into the Southern Hemisphere. "When combined with the prompt destruction from nuclear blast, fires and fallout and the later enhancement of solar ultraviolet radiation due to ozone depletion," we concluded, "long-term exposure to cold, dark and radioactivity could pose a serious threat to human survivors and to other species." Subsequent studies, based on more powerful models of the general circulation of the earth's atmosphere, have tended to confirm both the validity


1
FIREBALL
SHOCK WAVE
2
3
4
5
6
10 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 10
DISTANCE FROM GROUND ZERO (MILES)

of our investigative approach and the main thrust of our findings. In what follows we shall review the current state of knowledge on this vital issue.

Before one can understand the climatic effects of nuclear war one must first understand how the earth's radiation budget is normally balanced. The amount of sunlight absorbed by the atmosphere and the surface of the earth, averaged over time, is equal to the amount of thermal radiation emitted back into space. Because the intensity of the thermal radiation varies as the fourth power of the temperature, both the surface temperature and the atmospheric temperature can adjust fairly quickly to maintain the overall energy balance between the solar energy gained and the thermal energy lost.

If the earth were an airless body like the moon, its surface would radiate the absorbed solar energy directly into space. In this case the globally averaged temperature of the earth would be well below the freezing point of water, and life as we know it could not exist on our planet. Fortunately the earth has an atmosphere, which absorbs and traps some of the heat emitted by the surface, thereby raising the average ground-level temperature to well above freezing and providing a favorable environment for forms of life, such as ours, that are based on liquid water.

The thermal insulation of the earth's surface by the atmosphere—the "greenhouse effect"—arises from the fact that sunlight passes through the atmosphere more readily than thermal radiation does. The radiation emitted by the sun is mainly in the visible part of the electromagnetic spectrum, whereas the thermal radiation emitted by the earth's surface is concentrated in the infrared part. The main infrared-absorbing components of the atmosphere are water (in the form of ice crystals, liquid droplets and vapor) and carbon dioxide gas, both of which are essentially transparent to visible light. Hence the atmosphere generally acts as a window for sunlight but as a blanket for heat.

Under normal conditions the temperature of the troposphere, or lower atmosphere, decreases gradually with increasing altitude up to a height of about 12 kilometers, the boundary called the tropopause. Heat from the earth's surface is transferred upward through the atmosphere by several mechanisms: thermal radiation, small-scale turbulence, large-scale convection and the release of latent heat through the condensation of ascending water vapor. In a purely radiative atmosphere (that is, one in which the air does not move vertically and all the energy is transferred by radiation) the lower layers of air, where most of the solar energy is absorbed, would be warmer than the higher layers; in this situation the upward thermal radiation would exceed the downward thermal radiation, allowing the excess heat to escape into space. If the opacity of the atmosphere to infrared radiation were to increase (with no change in the opacity to visible light), the temperature would increase. For example, if carbon dioxide, a good infrared absorber, were added to the atmosphere in sufficient quantities, it would warm the surface.

Conversely, if some component of the atmosphere were to reduce the amount of sunlight reaching the surface without significantly increasing the infrared opacity, the ground temperature would decrease. For example, if all the sunlight were absorbed high in the atmosphere and none reached the ground, and if the surface could radiate energy to space without hindrance, the surface temperature would fall to that of an airless planet. If the absorption of solar energy were to take place above most of the atmosphere, the earth's radiation budget would be balanced without the greenhouse effect. (Accordingly we refer to this condition as the "anti–greenhouse effect.") Below the layer where the sunlight was absorbed the temperature of the atmosphere would not vary with altitude: at each lower level the upward infrared flux would equal the downward infrared flux and the net energy transfer would be negligible.

Particles in the atmosphere can affect the earth's radiation balance in several ways: by absorbing sunlight, by reflecting sunlight back into space and by absorbing or emitting infrared radiation. In general a cloud of fine particles—an aerosol—tends to warm the atmospheric layer it occupies, but it can either warm or cool the underlying layers and the surface, depending on whether the particles absorb infrared radiation more readily than they reflect and/or absorb visible light.

The anti–greenhouse effect of an aerosol is maximized for particles that are highly absorbing at visible wavelengths. Thus much less sunlight reaches the surface when an aerosol consists of dark particles such as soot, which strongly absorb visible light, than when an aerosol consists of bright particles such as soil dust, which mainly scatter the light. Consequently in evaluating the possible climatic effects of a nuclear war particular concern should be focused on the soot particles that are generated by fires, since soot is one of the few common particulate materials that absorb visible light much more strongly than they absorb infrared radiation.

How much an aerosol will cool the surface (by blocking sunlight) or warm the surface (by enhancing the greenhouse effect) depends on the size of the particles. If the average diameter of the particles is less than a typical infrared wavelength (about 10 micrometers), the infrared opacity of the aerosol will be less than its visible opacity. Accordingly an aerosol of very fine particles that even weakly absorb sunlight should have a visible effect greater than its infrared effect, giving rise to a significant cooling of the lower atmospheric layers and the surface. In the case of soot this is true even for somewhat larger particles.

The visible and infrared radiation effects associated with particle layers also depend on the thickness and density of the aerosol. The intensity of the sunlight reaching the ground decreases exponentially with the quantity of fine, absorbing particulate matter in the atmosphere. The infrared radiation reaching the ground, however, depends more on the air temperature than it does on the quantity of aerosol. Hence when a large amount of aerosol is present, the dominant climatic consequence tends to be strong surface cooling.

The "optical depth" of an aerosol (a measure of opacity equal to the negative natural logarithm of the attenuation of an incident light beam by absorption and scattering) serves as a convenient indicator of the aerosol's potential climatic effects. For example, a cloud with an optical depth of much less than 1

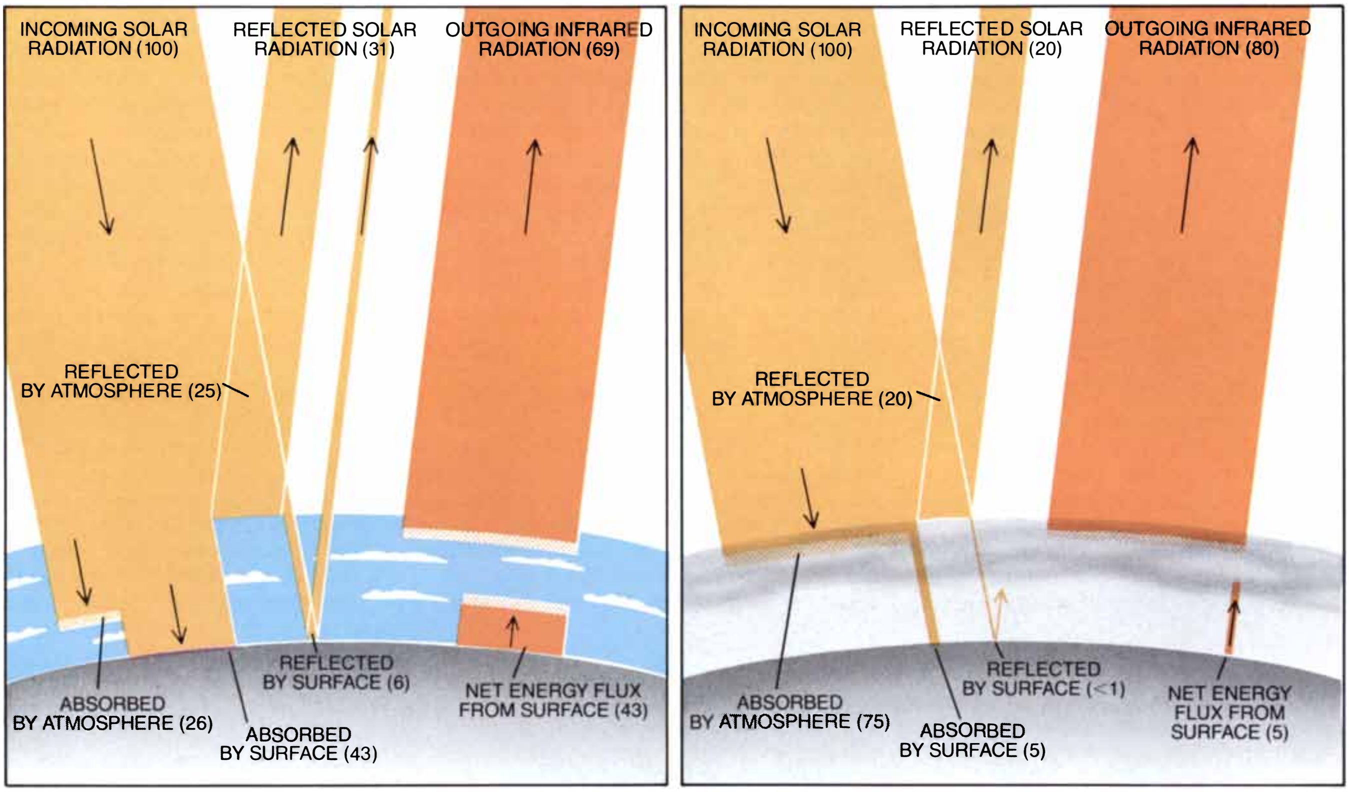
FIRESTORM DEVELOPS in the aftermath of a one-megaton nuclear explosion over the heart of New York in the hypothetical sequence of events depicted on the opposite page. (The skyline of the city, viewed here from the west, is drawn to scale; the detonation point is assumed to be at a height of 6,500 feet directly over the Empire State Building.) In the first few seconds after the detonation the initial flash of thermal radiation from the fireball would spontaneously ignite fires in combustible materials at a considerable distance (*1*). Many of these fires would be promptly snuffed out by the passage of the spherical blast wave (*black arcs*) and the accompanying high winds, but those two effects would also start a large number of secondary fires in the process of destroying most of the city's structures (*2*). Some of the individual primary and secondary fires could then merge into major conflagrations (*3*), which in turn might coalesce into a single massive fire covering most of the city (*4*). If such a fire were intense enough and the meteorological conditions were favorable, a full-scale firestorm could ensue, driven by winds of 100 miles per hour or more in the vicinity of the central updraft (*5*). Eventually the fire would burn itself out, leaving a smoldering residue (*6*). The smoke and dust thrown up by thousands of such explosions could extend over a large region, effectively blocking the sunlight and drastically reducing the surface temperature, regardless of the season. In the authors' "base line" 5,000-megaton nuclear-war scenario little soot is injected into the stratosphere by urban firestorms; if such firestorms were common in a nuclear war, the resulting nuclear winter would be much severer than the one predicted by the computer models.

would cause only minor perturbations, since most of the light would reach the surface, whereas a cloud with an optical depth of 1 or more would cause a major disturbance, since most of the light would be absorbed in the atmosphere and/or scattered away into space. Although volcanic particles happen to have an optimal size for enhancing visible effects over infrared effects, the magnitude of the induced surface cooling is limited by the modest optical depth of volcanic aerosols (less than about .3) and by their very weak intrinsic absorption at visible wavelengths. Nevertheless, the largest volcanic clouds may disturb the earth's radiation balance enough to cause anomalous weather. Much more significant climatic disturbances could result from the huge clouds of dust that would be thrown into the atmosphere by the impact of an asteroid or a comet with a diameter of several kilometers or more. These dust clouds could have a very large optical depth, perhaps initially as high as 1,000.

The radiative effects of an aerosol on the temperature of a planet depend not only on the aerosol's optical depth, its visible absorptivity and the average size of its particles but also on the variation of these properties with time. The longer a significant optical depth can be sustained, the closer the surface temperature and the atmospheric temperature will move toward a new state of equilibrium. Normally it takes the surface of the ocean several years to respond to changes in the global radiation balance, because of the great heat capacity of the mixed uppermost layer of the ocean, which extends to a depth of about 100 meters. In contrast, the air temperature and the continental land temperature approach new equilibrium values in only a few months. In fact, when the atmosphere is strongly cooled, convection above the surface ceases and the ground temperature falls rapidly by radiative cooling, reaching equilibrium in a few days or weeks. This happens naturally every night, although equilibrium is not reached in such a short period.

Particles are removed from the atmosphere by several processes: falling under the influence of gravity, sticking to the ground and other surfaces and scavenging by water clouds, rain and snow. The lifetime of particles against "wet" removal depends on the frequency of cloud formation and precipitation at various altitudes. In the first few kilometers of altitude in the normal atmosphere particles may in some places be washed out in a matter of days. In the upper troposphere (above five kilometers) the average lifetime of the particles increases to several weeks or more. Still higher, in the stratosphere (above 12 kilometers), water clouds rarely form and so the lifetime of small particles is typically a year or more. Stratospheric removal is primarily by gravitational settling and the large-scale convective transport of the particles. The deposition of particles on surfaces is very inefficient for average-size smoke and dust particles, requiring several months for significant depletion.

Clearly the height at which particles are injected into the atmosphere af-

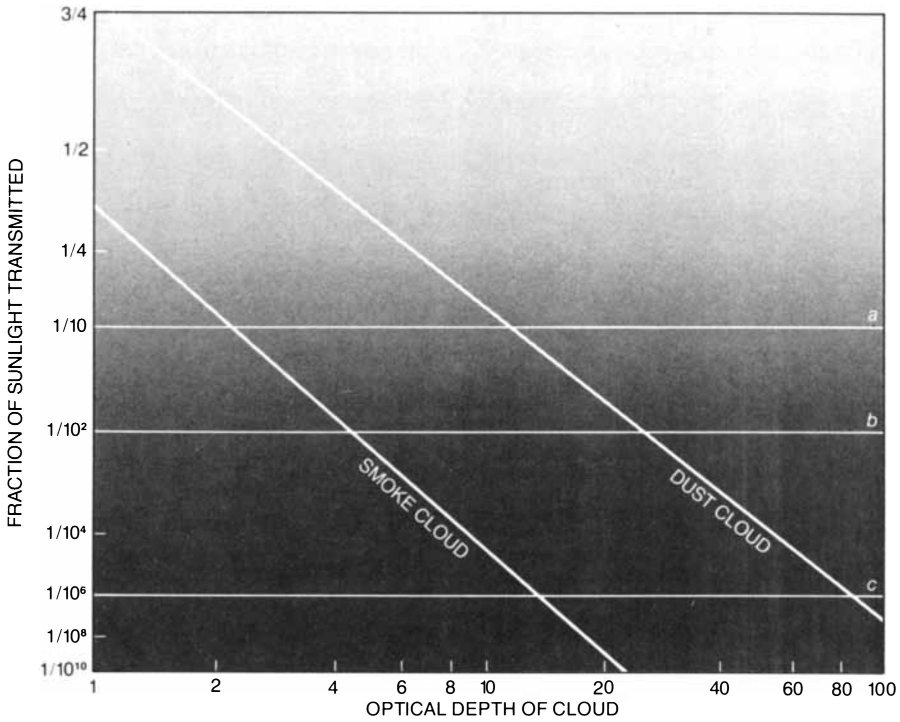


"GREENHOUSE EFFECT," the phenomenon whereby the earth's surface is warmed by the insulating properties of the atmosphere, could be negated over a large area by the cloud of smoke and dust resulting from a nuclear war. Under normal conditions (*left*) the atmosphere is quite transparent to radiation at visible wavelengths; accordingly a large fraction of the incident sunlight passes through the atmosphere and is absorbed by the surface in the form of thermal energy, or heat. Thermal radiation from the surface is emitted predominantly at longer, infrared wavelengths, which are strongly absorbed by the lower atmosphere, raising the temperature near the surface to levels well above the freezing point of water. Under normal conditions the net heat balance at the ground is the difference between a downward infrared flux of 101 units and an upward energy flux of 144 units; the latter value in turn consists of an upward infrared flux of 115 units plus an upward flux of latent and "sensible" heat totaling 29 units. The energy absorbed by the atmosphere and the surface in the form of sunlight, averaged over time, is equal to the energy emitted back into space from the surface and the atmosphere in the form of infrared radiation. A dense cloud of smoke and dust in the upper or middle atmosphere would rebalance the earth's energy budget (*right*). Most of the sunlight would then be absorbed by the cloud, and a large fraction would be reradiated directly back into space at infrared wavelengths without ever reaching the surface. Under these perturbed conditions the net heat balance at the ground would be the difference between a downward infrared flux of 65 units and an upward infrared flux of 70 units; the two fluxes would almost balance, canceling any heating effect in the atmosphere. The surface and the lower atmosphere would cool rapidly as the residual heat dissipated.

fects their residence time. In general, the higher the initial altitude, the longer the residence time in the normal atmosphere. Massive injections of soot and dust, however, may profoundly alter both the structure of the atmosphere and the rate of particle removal.

In our analysis of the climatic effects of nuclear war we have adopted a number of specific scenarios, based on what is publicly known about the effects of individual nuclear explosions, the size and deployment of the world's present nuclear arsenals and the nuclear-war-fighting plans of the U.S. and the U.S.S.R. Among the several dozen cases we have analyzed are a 100-megaton "countervalue" attack directed strictly against cities, a 3,000-megaton "counterforce" attack directed strictly against missile silos and a 10,000-megaton "full-scale exchange" directed against an assortment of targets on both sides. Our "base line" case is a 5,000-megaton nuclear exchange, with about 20 percent of the total explosive yield detonated over urban, suburban and industrial areas. All the postulated attack scenarios are well within the present capabilities of the two superpowers.

A nuclear explosion can readily ignite fires in either an urban or a rural setting. The flash of thermal radiation from the nuclear explosion, which has a spectrum similar to that of sunlight, accounts for about a third of the total energy yield of the explosion. The flash is so intense that a variety of combustible materials are ignited spontaneously at ranges of 10 kilometers or more from a one-megaton air burst detonated at a nominal altitude of a kilometer. The blast wave from the explosion would extinguish many of the initial fires, but it would also start numerous secondary fires by disrupting open flames, rupturing gas lines and fuel storage tanks and causing electrical and mechanical sparks. The destruction resulting from the blast wave would also hamper effective fire fighting and so promote the spread of both the primary and the secondary fires. Based on the known incendiary effects of the nuclear explosions over Hiroshima and Nagasaki in 1945 it can be projected that the fires likely to be caused by just one of the far more powerful strategic nuclear weapons available today would extend over an area of from tens to hundreds of square kilometers.



TRANSMISSION OF SUNLIGHT through smoke clouds and dust clouds is represented in this graph as a function of each cloud's optical depth, a measure of opacity equal to the negative natural logarithm of the attenuation of an incident light beam. At a given optical depth a smoke cloud evidently absorbs much more sunlight than a dust cloud. Three typical levels of transmitted light are given for comparison: they correspond to a very cloudy day (*a*), a light level at which photosynthesis is barely possible (*b*) and a clear night with a full moon (*c*).

Nuclear explosions over forests and grasslands could also ignite large fires, but this situation is more difficult to evaluate. Among the factors that affect fires in wilderness areas are the humidity, the moisture content of the fuel, the amount of the fuel and the velocity of the wind. Roughly a third of the land area in the North Temperate Zone is covered by forest, and an equal area is covered by brush and grassland. Violent wildfires have been known to spread over tens of thousands of square kilometers from a few ignition points; in the absence of a nuclear war such fires occur about once every decade. Although most wildfires generated by nuclear explosions would probably be confined to the immediate area exposed to the intense thermal flash, it is possible that much larger ones would be started by multiple explosions over scattered military targets such as missile silos.

The total amount of smoke likely to be generated by a nuclear war depends on, among other things, the total yield of the nuclear weapons exploded over each type of target, the efficiency of the explosions in igniting fires, the average area ignited per megaton of yield, the average amount of combustible material in the irradiated region, the fraction of the combustible material consumed by the fires, the ratio of the amount of smoke produced to the amount of fuel burned and the fraction of the smoke that is eventually entrained into the global atmospheric circulation after local rainfall has removed its share. By assigning the most likely values to these parameters for a nuclear war involving less than 40 percent of the strategic arsenals of the two superpowers we were able to calculate that the total smoke emission from a full-scale nuclear exchange could easily exceed 100 million metric tons. In many respects this is a conservative estimate. Crutzen and his co-workers Ian Galbally of the Commonwealth Scientific and Industrial Research Organization (CSIRO) in Australia and Christoph Brühl of the Max Planck Institute at Mainz have recently estimated that the total smoke emission from a full-scale nuclear war would be closer to 300 million tons.

One hundred million tons of smoke, if it were distributed as a uniform cloud over the entire globe, could reduce the intensity of sunlight reaching the ground by as much as 95 percent. The initial clouds would not cover the entire globe, however, and so large areas of the Northern Hemisphere, particularly in the target zones, would be even darker; at noon the light level in these areas could be as low as that of a moonlit night. Daytime darkness in this range, if it persisted for weeks or months, would trigger a climatic catastrophe. Indeed, significant disturbances might be caused by much smaller amounts of smoke.

Wildfires normally inject smoke into the lower atmosphere to an altitude of five or six kilometers. In contrast, large urban fires have been known to inject smoke into the upper troposphere, probably as high as 12 kilometers. The unprecedented scale of the fires likely to be ignited by large nuclear explosions and the complex convective activity generated by multiple explosions might cause some of the smoke to rise even higher. Studies of the dynamics of very large fires suggest that individual smoke plumes might reach as high as 20 kilometers, well into the stratosphere.

During the World War II bombing of Hamburg the center of the city was gutted by an intense firestorm, with heat-generated winds of hurricane force sweeping inward from all directions at ground level. Rapid heat release over a large area can create fire vortexes, heat tornadoes and cyclones with towering convective columns. The sheer intensity of such fires might act to reduce the smoke emission considerably through two processes: the oxidation of carbonaceous smoke particles at the extremely high temperatures generated in the fire zone and the washout of smoke particles by precipitation formed in the convective column. Both effects were taken into account in our estimates of the total smoke emission from a nuclear war.

The climatic impact of smoke depends on its optical properties, which in turn are sensitive to the size, shape and composition of the smoke particles. The most effective light-screening smoke consists of particles with a radius of about .1 micrometer and a very sooty composition rich in graphite. The least effective smoke in attenuating sunlight consists of particles larger than .5 micrometer with a predominantly oily composition. The smoke from a forest fire is typically composed of extremely fine oily particles, whereas the smoke from an urban fire consists of larger agglomerations of sooty particles. Smoke from fierce fires usually contains large particles of ash, char, dust and other debris, which is swept up by the heat-generated winds. The largest of these particles fall out of the smoke clouds just downwind of the fire. Although very intense fires produce less smoke, they lift more fine dust and may burn metals such as aluminum and chromium, which efficiently generate fine aerosols.

The release of toxic compounds in urban fires has not been adequately studied. It is well known that many people who have died in accidental fires have been poisoned by toxic gases. In addition to carbon monoxide, which is produced copiously in many fires, hydrogen cyanide and hydrogen chloride are generated when the synthetic compounds in modern building materials and furnishings burn. If large stores of organic chemicals are released and burned in a nuclear conflict, additional airborne toxins would be generated. The possibility that vast areas could be contaminated by such pyrotoxins, adsorbed on the surface of smoke, ash and dust particles and carried great distances by winds, needs further investigation.

Nuclear explosions at or near ground level throw up huge amounts of dust. The principal dust-forming mechanisms include the ejection and disaggregation of soil particles from the crater formed by the explosion; the vaporization and subsequent renucleation of soil and rock, and the lifting of surface dust and smoke. A one-megaton explosion on land can excavate a crater hundreds of meters in diameter, eject several million tons of debris, lift between 100,000 and 600,000 tons of soil to a high altitude and inject between 10,000 and 30,000 tons of submicrometer dust particles into the stratosphere. The height at which the dust is injected depends on the yield of the explosion: the dust clouds produced by explosions with a yield of less than about 100 kilotons will generally not penetrate into the stratosphere, whereas those from explosions with a yield of more than about a megaton will stabilize mainly within the stratosphere. Explosions above the ground can also raise large quantities of dust, which is vacuumed off the surface by the rising fireball. The combined effects of multiple explosions could enhance the total amount of dust raised to high altitudes.

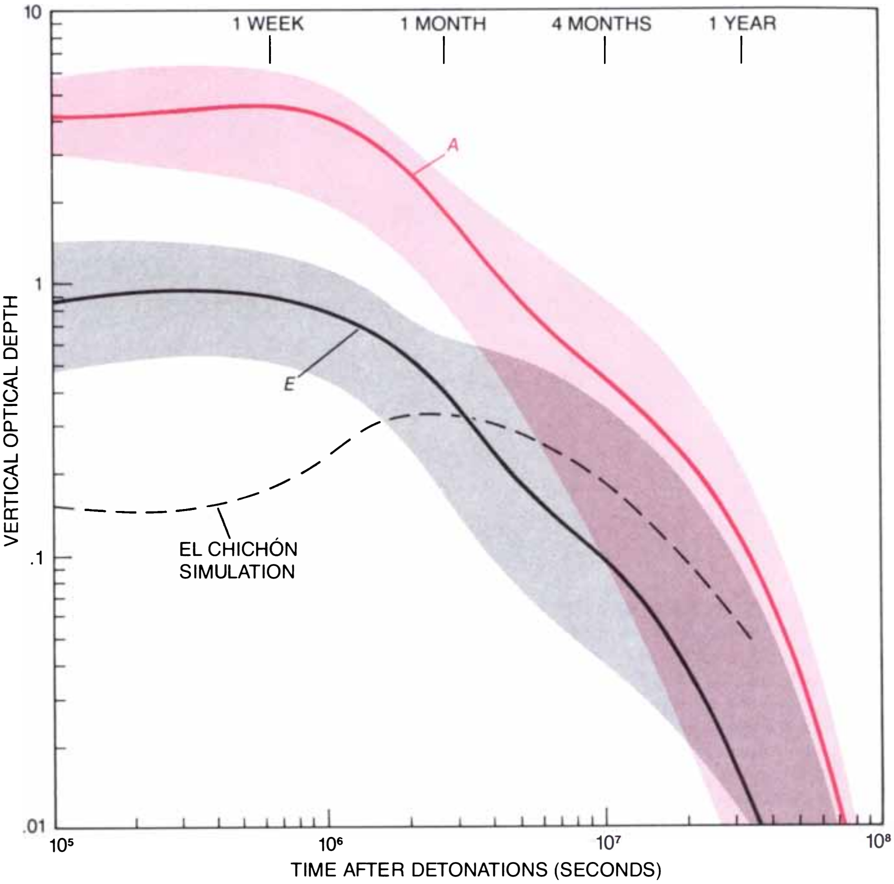
The quantity of dust produced in a nuclear war would depend sensitively on the way the weapons were used. Ground bursts would be directed at hard targets, such as missile silos and underground command posts. Soft targets could be attacked by air bursts as well as ground bursts. There are more than 1,000 missile silos in the continental U.S. alone, and at least two Russian warheads are probably committed to each of them. Some 1,400 missile silos in the U.S.S.R. are similarly targeted by U.S. warheads. Air bases and secondary airfields, submarine pens and command and control facilities are among the many other strategic targets to which ground bursts might be assigned. In short, it seems quite possible that at least 4,000 megatons of high-yield weapons might be detonated at or near ground level even in a war in which cities were not targeted, and that roughly 120 million tons of submicrometer soil particles could be injected into the stratosphere in the North Temperate Zone. This is many times greater than all the submicrometer dust lifted into the stratosphere by the eruption of the volcano El Chichón in Mexico in 1982 and is comparable to the global submicrometer dust injections of much larger volcanic eruptions such as that of Tambora in 1815 and Krakatau in 1883.

	SCENARIO	TOTAL YIELD (MEGATONS)	SURFACE BURSTS (PERCENTAGE OF YIELD)	URBAN OR INDUSTRIAL TARGETS (PERCENTAGE OF YIELD)	YIELD OF WARHEADS (MEGATONS)	TOTAL NUMBER OF EXPLOSIONS	MASS OF SUBMICROMETER SMOKE (MILLIONS OF METRIC TONS)	MASS OF SUBMICROMETER DUST (MILLIONS OF METRIC TONS)	OPTICAL DEPTH OF SMOKE	OPTICAL DEPTH OF DUST
A	BASE-LINE EXCHANGE	5,000	57	20	.1–10	10,400	225	65	4.5	1
B	LOW-YIELD AIRBURSTS	5,000	10	33	.1–1	22,500	300	15	6	.2
C	FULL-SCALE EXCHANGE	10,000	63	15	.1–10	16,160	300	130	6	2
D	MEDIUM-SCALE EXCHANGE	3,000	50	25	.3–5	5,433	175	40	3.5	.6
E	LIMITED EXCHANGE	1,000	50	25	.2–1	2,250	50	10	1	.1
F	GENERAL COUNTERFORCE ATTACK	3,000	70	0	1–10	2,150	0	55	0	.8
G	HARD-TARGET COUNTERFORCE ATTACK	5,000	100	0	5–10	700	0	650	0	10
H	CITY ATTACK	100	0	100	.1	1,000	150	0	3	0
I	FUTURE WAR	25,000	72	10	.1–10	28,300	400	325	8	5

NUCLEAR-WAR SCENARIOS outlined in this table were drawn from a much longer list of possibilities considered in detail by the authors in their original study. The last four columns at the right give the outcome of each scenario in terms of the resulting clouds of smoke and dust. All the scenarios shown here except the last one are thought to be well within the present capabilities of the two superpowers.

Analogies between the atmospheric effects of a major volcanic explosion and a nuclear war are often made for convenience. Nevertheless, there is no straightforward way to scale the effects of a volcanic explosion against those of a series of nuclear detonations. The aerosol particles produced by volcanoes are fundamentally different in composition, size and shape from those produced by nuclear explosions. We have therefore based our calculations on the properties of dust measured directly in nuclear-explosion clouds.

The only proper comparison between a volcanic eruption and a nuclear explosion is the optical depth of the long-term aerosols that are produced. In fact, we utilized data on global "dust veils" generated by volcanic explosions to test and calibrate our climate models. In so doing we have been able to account quantitatively for the hemispheric surface-cooling effect observed after major volcanic eruptions. The present nuclear-dust calculations are entirely consistent with observations of volcanic phenomena. For example, it is now clear that violent eruptions can lead to a significant climatic cooling for a year or more. Even so, in recorded history volcanoes have had only a rather modest climatic role. The fact that volcanoes are localized sources of dust limits their geographic influence; moreover, volcanoes inject comparatively little fine dust (and no soot) into the stratosphere. Nuclear explosions, on the other hand, are a powerful and efficient means of injecting large quantities of fine soot and dust into the atmosphere over large regions.



VARIATION IN OPACITY OVER TIME of the clouds of smoke and dust resulting from a number of nuclear-war scenarios is traced here in terms of the clouds' vertical optical depth averaged over the Northern Hemisphere. The optical depths were calculated for visible light at a wavelength of 550 nanometers. The results tend to fall into two bands. The upper band includes those scenarios that gave results close to the authors' 5,000-megaton base-line case, indicated by the solid-color curve; the total explosive yields of the scenarios in this band range from 100 to 10,000 megatons. The lower band includes those scenarios that gave less drastic results; the scenarios in this band range from the 1,000-megaton limited exchange indicated by the black curve to a 3,000-megaton counterforce attack aimed at military targets only. Significant climatic disruptions are expected whenever the average optical depth over a hemisphere is greater than about 1. The average optical depth over the Northern Hemisphere of the cloud of dust raised by the eruption of the volcano El Chichón in 1982 is given for comparison.

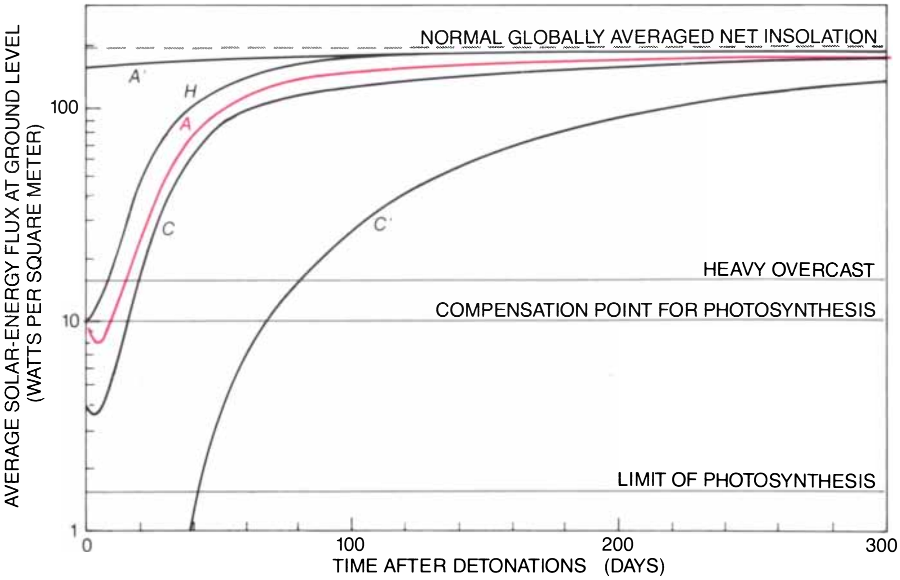
The atoms produced in the fission reactions of a nuclear explosion are often in unstable isotopic states. Radioactive decay from these states releases alpha, beta and gamma radiation. In most nuclear weapons at least half of the energy yield is generated by fission and the rest by fusion. About 300 distinct radioactive isotopes are produced. Most of them condense onto aerosols and dust formed in (or sucked into) the fireball. Accordingly the dust and the radioactivity generated by nuclear explosions are intimately related.

Of particular interest here are the prompt and the intermediate radioactive fallout. The former is associated with short-lived radioactive isotopes that condense onto large soil particles, which in turn fall to the ground within hours after an explosion. Intermediate fallout is associated with longer-lived radioactive isotopes carried by smaller particles that drift in the wind and are removed by settling and precipitation in the interval from days to months. Prompt fallout is generated by ground bursts, and intermediate fallout is generated by ground bursts and air bursts in the yield range from 10 to 500 kilotons, which deposit their radioactivity in the middle and upper troposphere.

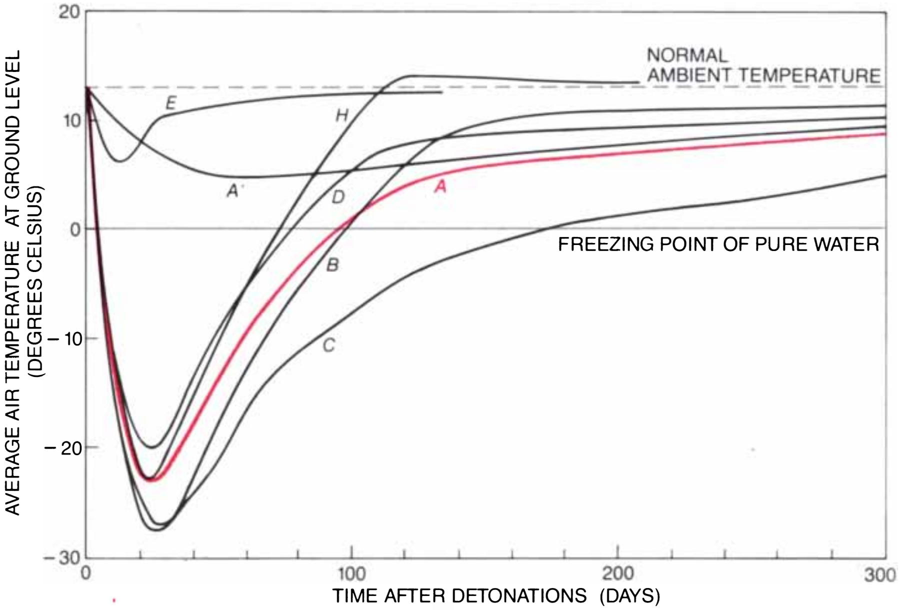
The danger from radioactive fallout is measured in terms of the total dose in rads (a unit of radiation exposure equivalent to 100 ergs of ionizing energy deposited in one gram of tissue), the dose rate in rads per hour and the type of radiation. The most deadly effects are caused by the intense, penetrating gamma radiation from prompt fallout. The widespread intermediate fallout delivers a less potent long-term gamma-ray dose. A whole-body gamma-ray exposure of 450 rads, received over several days, is lethal to half of the healthy adults exposed. Chronic doses of 100 rads or more from intermediate fallout could suppress the immune system even of healthy people and would cause long-term increments in the incidence of cancer, genetic defects and other diseases.

Our most recent studies of the effects of radioactive fallout in our base-line case indicate that the prompt fallout could contaminate millions of square kilometers of land with lethal radioactivity. The intermediate fallout would blanket at least the North Temperate Zone, producing average long-term, whole-body gamma-ray exposures of about 50 rads in unprotected populations. Internal exposures of specific organs to biologically active radioactive isotopes such as strontium 90 and iodine 131, which enter the food chain, could double or triple the total doses. According to Joseph B. Knox of the Lawrence Livermore National Laboratory, if nuclear power plants were targeted directly, the average long-term gamma-ray dose could be increased to several hundred rads or more.

The computer models we have employed to define the potential magnitude of the long-term global aftereffects of nuclear war are one-dimensional: they treat only the vertical structure of the atmosphere. Obviously the atmosphere is a complex three-dimensional system



VARIATION IN THE AMOUNT OF SUNLIGHT reaching the ground through a widely dispersed cloud of smoke and dust created by a nuclear war is shown for several of the nuclear-war scenarios considered by the authors. Each curve gives the average solar-energy flux at the ground in the Northern Hemisphere for a given scenario. The fluxes correspond to horizontally averaged smoke clouds and do not take into account possible patchiness in the clouds. The color curve corresponds to the 5,000-megaton base-line case (*curve A*). Curve *A'* is a variant of the base-line case that takes into account only the effects of dust. Curve *C'* is a variant of the 10,000-megaton full-scale exchange that assigns more extreme values to the smoke-emission parameters. Also indicated are the energy level at which photosynthesis cannot keep pace with the respiration of a typical plant (a level known as the compensation point) and the energy level at which photosynthesis ceases. These two thresholds vary considerably from plant to plant.



VARIATION IN TEMPERATURE at ground level in the Northern Hemisphere with time after a nuclear war is shown for several scenarios. Again the color curve corresponds to the authors' 5,000-megaton base-line case. In general smoke in the troposphere (below 12 kilometers) would have a substantial short-term cooling effect, whereas dust in the stratosphere (above 12 kilometers) would result in a less pronounced but longer cooling trend. The calculated average temperature decrease in each case is probably the largest that would be registered, and then only in the interior regions of the continents. The temperature decrease over the ocean would be only a few degrees Celsius or less. Hence the mixing of continental and marine air masses would lead to a less drastic temperature decrease over land, particularly along coastlines. The same phenomenon, however, would also lead to prolonged severe coastal storms. The temperature changes plotted in this graph are seasonally averaged. If the war were in the summer, the temperature decrease would be larger than that indicated in each case; if the war were in the winter, the temperature decrease would be smaller. A decline in temperature of only a few degrees may have significant adverse effects on agriculture. In a wide range of cases the temperature drops to −20 degrees C. or lower and does not return to the freezing point for months.

whose intricate interactions determine its response to perturbations. At present, however, there are no three-dimensional models with the appropriate features to treat the nuclear-winter problem with high precision, although several such models are under development. The existing models of the climate and the general circulation of the atmosphere incorporate a number of empirical approaches to physical processes that are not well understood. In the nuclear-winter scenario the climate is so seriously perturbed that such treatments are of dubious applicability.

Our approach has therefore been to estimate the first-order effects through detailed microphysical, chemical and optical calculations in a one-dimensional format. Even this simplified approach had not been attempted before our work, and it was not clear then that more sophisticated three-dimensional work was justified. Based on the predicted first-order one-dimensional effects, the principal three-dimensional meteorological interactions that would have to be treated in more refined studies were deduced. The three-dimensional results generally confirm our one-dimensional results.

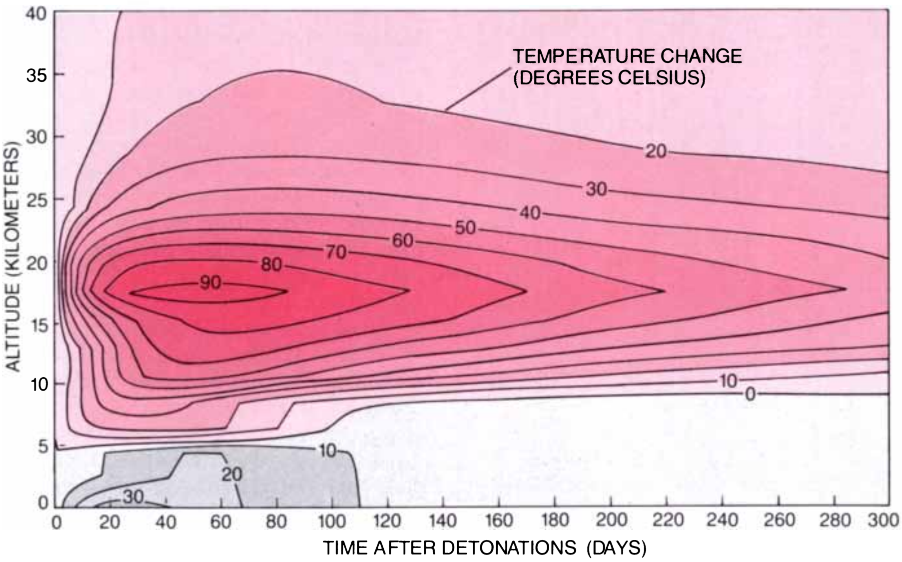
Three basic models were used in our study: a nuclear-scenario model, a particle-microphysics model and a radiative-convective climate model. The nuclear-scenario model determines the quantities of smoke, dust, radioactivity and pyrotoxins generated by a specific nuclear exchange, relying on (among other things) the smoke and dust estimates cited above. The microphysics model simulates the evolution of the quantity and the size of the smoke and dust particles and the deposition of radioactivity by accounting for the physical interactions and vertical transport of particles at all altitudes. The radiative-convective model calculates the optical and infrared properties of the evolving particles, the visible and infrared energy fluxes and air temperatures as a function of time and altitude. Because the predicted air temperatures are sensitive to surface heat capacities, separate calculations were done for land and ocean environments to define possible temperature differences.

One-dimensional models cannot accurately forecast the short-term or local effects of a nuclear war. The applicability of the predictions based on such models depends on the rate and extent of the dispersion of the smoke plumes and dust clouds. Soon after a nuclear exchange thousands of individual smoke and dust clouds would be distributed throughout the North Temperate Zone at altitudes of up to 20 kilometers. The action of horizontal turbulent diffusion, vertical wind shear and continuing smoke emission would almost certainly spread the clouds of nuclear debris over the entire zone and fill in gaps in the clouds within a week. Local effects in this period could vary considerably from the average effects simulated with our one-dimensional models.

Since large uncertainties are inherent in any assessment of the effects of nuclear warfare, a variety of simulations were made to investigate the sensitivity of various outcomes to changes in scenarios and in key physical parameters. Although all these results cannot be discussed in an article of this length, some selected findings are summarized in the illustrations accompanying this article.

In general our results suggest that the optical depth of the cloud of smoke and dust resulting from a massive nuclear exchange would at the very least be comparable to or larger than that of the cloud resulting from a major volcanic eruption, and that the most probable optical depth would be an order of magnitude larger. The likeliest outcome of a nuclear war would therefore be a climatic catastrophe.

The ultimate impact of smoke and dust on climate depends on the fraction of sunlight they screen from the surface. This suggests that a saturation effect will occur. For example, a smoke cloud with an optical depth of 3 can prevent 95 percent or more of the sunlight from reaching the ground. Greater optical depths cause only negligible further decreases in the average surface sunlight. Thus an average optical depth of about 3 is roughly the saturation level for smoke-induced climatic effects. The saturation level for dust occurs at an optical depth of about 20, and its onset is much more gradual with increasing optical depth than the saturation level for smoke. It is doubtful whether enough dust could be raised in a nuclear war to create optical saturation on a global scale.



VARIATION IN TEMPERATURE WITH ALTITUDE over the northern continents is plotted in this contour diagram for the first year after the 5,000-megaton base-line nuclear war. The contours give the average change in temperature at intervals of 10 degrees C., based on the authors' one-dimensional radiative-convective model. The gray area indicates parts of the atmosphere that are cooled below their normal temperature; the color area indicates parts that are heated above their normal temperature. The strong heating effect in the upper troposphere and lower stratosphere is attributable mainly to the absorption of sunlight by smoke and dust.

How a smoke cloud extinguishes light also differs from how a dust cloud does so. A sooty pall of smoke absorbs most of the incident light and scatters only a small fraction back into space or down toward the surface. The absorption rapidly heats the smoke clouds, inducing powerful air motions and winds. Dust clouds, on the other hand, primarily scatter the incident sunlight and absorb only a small fraction. To block light effectively clouds that are purely light-scattering must be very thick, because much of the light is scattered forward toward the earth's surface; for example, ordinary water clouds typically have an optical depth of 10 or more.

We find that for many scenarios a substantial reduction in sunlight may persist for weeks or months after the war. In the first week or two the clouds would also be patchy; hence our calculations probably underestimate the average light intensity at these early stages. Nevertheless, within the target zones it would be too dark to see, even at noon.

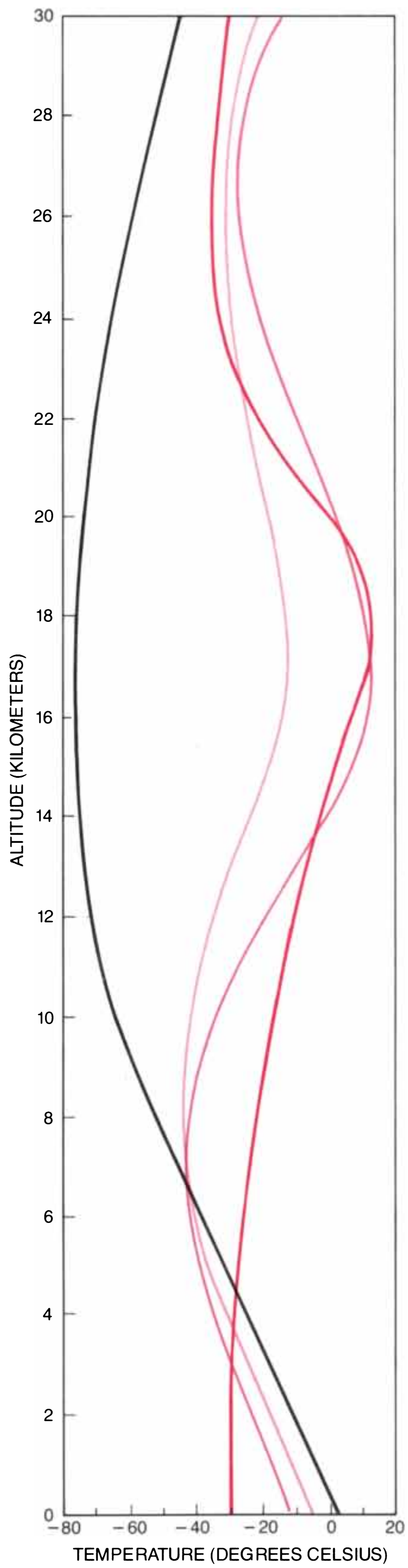
The large amount of smoke generated by a nuclear exchange could lead to dramatic decreases in continental temperatures for a substantial period. In many of the scenarios represented in the illustrations accompanying this article land temperatures remain below freezing for months. Average temperature decreases of only a few degrees Celsius in spring or early summer could destroy crops throughout the North Temperate Zone. Temperature drops of some 40 degrees C. (to an absolute temperature of about −25 degrees C.) are predicted for the base-line case, and still severer cooling effects are possible with the current nuclear arsenals and with those projected for the near future.

The predicted changes in air temperature as a function of height and time for our 5,000-megaton base-line scenario reveal several important features. First, the upper atmosphere is heated by between 30 and 80 degrees C. as the sunlight, which normally warms the ground, is absorbed in the highest smoke layers. At the same time the ground cools in darkness. The hot clouds, like hot-air balloons, would not remain stationary but would rise and expand.

A month after a massive nuclear exchange the entire troposphere over land could be thermally brought to a standstill. Even after three months only the lowest few kilometers would receive enough solar energy to drive weak convection. In effect the stratosphere would descend to the surface, creating an alien atmosphere. In some places warm currents of ocean air would still sweep into the continents at ground level, but this heat source would be able to drive convection only within the lowest few kilometers of the atmosphere. The intense temperature inversion would effectively damp deep convective activity. Elsewhere cold air flowing off the continents might warm over the oceans, rise and recirculate over the continents and finally subside over the land.

One possible consequence of the temperature inversion caused by such a smoke cloud would be an increase in the atmospheric residence time of the smoke and dust. This outcome represents a positive feedback effect, not taken into account in any calculations so far, that would increase both the severity and the duration of the nuclear winter. The temperature inversion reduces the convective penetration of moist air from below, inhibiting the condensation of water in the sooty air and hence greatly limiting precipitation at altitudes higher than a few kilometers. The longer soot and dust remain in the atmosphere, the farther they spread horizontally and the more widespread their climatic impact is. Under these conditions the particles are removed mainly by continuing coagulation and fallout and by transport in global-scale wind systems and turbulence to low altitudes where precipitation scavenging still takes place.

Our calculated temperature changes over extended land masses do not account for the initial patchiness in the clouds or the later dilution of cold continental air by warm marine air. Michael C. MacCracken of Livermore has investigated the combined effects of patchi-



TEMPERATURE PROFILES of the atmosphere before and after a 5,000-megaton nuclear war are compared. The black curve is the normal temperature profile. The other curves show the profile one month after the war (*dark color*), three months after (*medium color*) and six months after (*light color*).

ness in clouds and the transfer of heat from the ocean, working with a general-circulation model to trace large blobs of smoke; he has also worked with a two-dimensional climate model to calculate land temperatures corresponding to the smoke emission in our 5,000-megaton base-line scenario. He finds average temperature decreases on land that are roughly half our continental-interior temperature drops. Even more sophisticated three-dimensional general-circulation-model calculations for conditions similar to our base-line scenario confirm that temperature drops of between 20 and 40 degrees C. are possible over vast continental areas.

The results of our computations indicate that the motions induced in soot clouds by the absorption of sunlight might cause the soot cloud to rise and spread out horizontally. This phenomenon could accelerate both the early dispersal and the global spreading of smoke plumes, a process that is otherwise dominated by wind shear and turbulence. Recently a group at the National Aeronautics and Space Administration's Ames Research Center, consisting of Robert M. Haberle and two of us (Ackerman and Toon), applied an advanced two-dimensional global-circulation model to compute the motion of heated soot clouds in the earth's troposphere. The Ames group considered a uniform soot cloud between 30 and 60 degrees north latitude, encircling the earth at these latitudes and extending from the ground to an altitude of eight kilometers. This smoke simulation shows massive fragments of the cloud rising high into the stratosphere and moving briskly toward the Equator and the Southern Hemisphere.

Although these calculations are preliminary, they support a major hypothesis of our initial study: that self-propelled smoke and dust clouds from the Northern Hemisphere could be rapidly transported to the Southern Hemisphere, causing large climatic anomalies there as well. Such accelerated dispersal could have the most severe consequences in the Tropics of both hemispheres, where the indigenous organisms are extremely sensitive to dark and cold. A nuclear winter extending to the Tropics would represent an ecological disaster unprecedented in history.

Our speculations about major meteorological disturbances and interhemispheric transport following a nuclear conflict have received further support from sophisticated calculations with three-dimensional models of global circulation. These models are not yet designed to move smoke and dust as tracer elements or to make the required detailed radiative-transport calculations. Nevertheless, they are able to define the initial three-dimensional perturbations in winds and temperatures caused by massive smoke injections. Two research groups have made these advanced climate studies: Curt Covey, Stephen H. Schneider and Starley L. Thompson of the National Center for Atmospheric Research (NCAR) in Boulder, Colo., and Vladimir V. Alexandrov and Georgi L. Stenchikov of the Computing Center of the Academy of Sciences of the U.S.S.R.

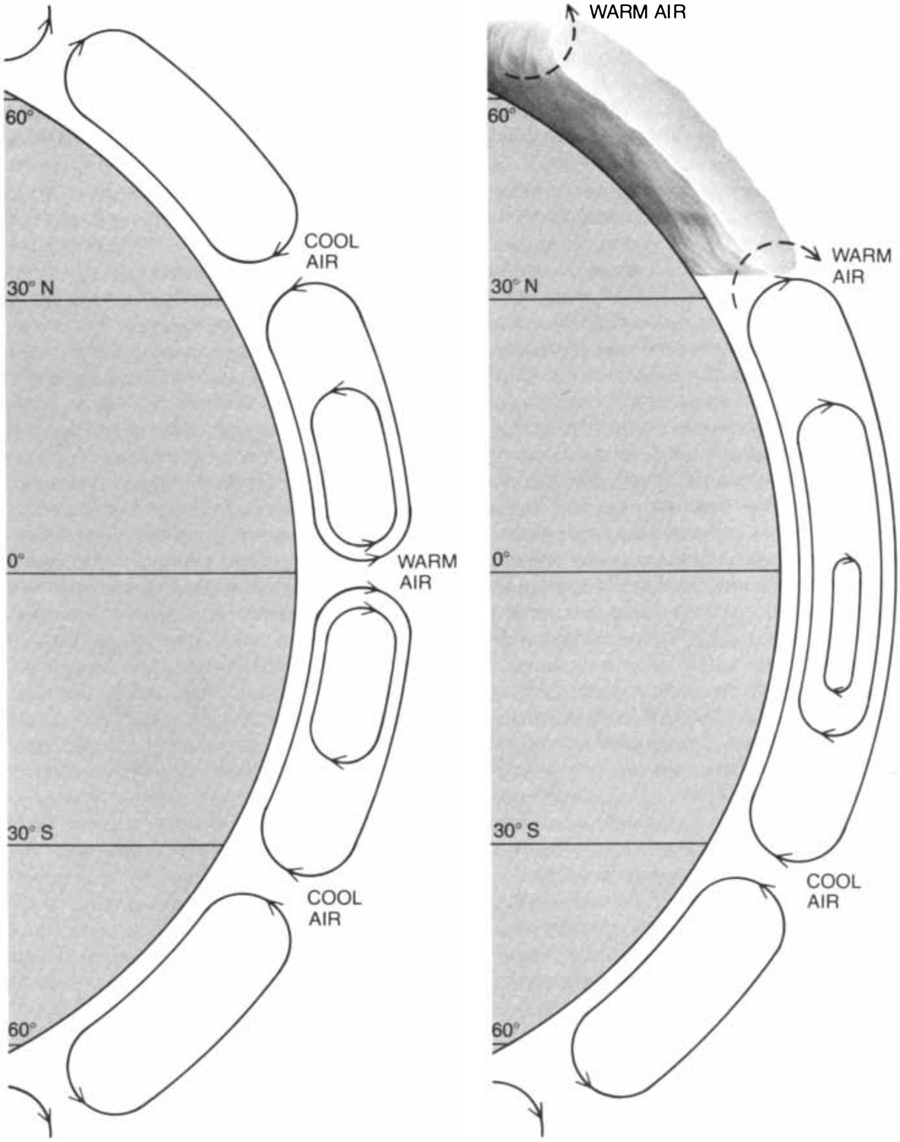
The predictions made by both groups of the normal and perturbed meridional, or north-south, circulation of the atmosphere several weeks after a nuclear exchange in the Northern Hemisphere in the spring or summer lead to the same conclusion: the normally bifurcated "Hadley cell" circulation in the Tropics would be transformed into a single intense cell with strong winds in the upper troposphere flowing directly from the Northern Hemisphere to the Southern Hemisphere. This would represent a profound change in the global wind system.

The average meridional circulation is the residual motion of large-scale planetary-wave oscillations. The global-circulation models predict anomalies in the planetary-wave motions, and here too the results are surprising. The NCAR group finds that continent-size bodies of heated air could penetrate deep into the Southern Hemisphere in a matter of days. Essentially all the habitable land masses of the earth could be subject to rapid blackout by soot. The global-circulation models also forecast subfreezing temperatures over most of the northern continental regions. What is startling is that local freezing could occur within two or three days; the NCAR group refers to it as a "quick freeze." Under such circumstances practically no area of the globe, north or south, would be safe from nuclear winter.

Consideration of the possible weather activity near coastlines during the nuclear winter suggests that even if the incident sunlight were reduced significantly, the oceans would continue to feed heat and moisture into the marine boundary layer near coastlines. In some regions cold offshore winds would interact with the marine environment to produce intense storms and heavy precipitation. In other regions, as prevailing winds swept ocean air onto cold continents, thick stratus clouds and continuous precipitation could ensue. It is not known how far this severe weather might extend inland from the coastlines, but a 100-kilometer margin would probably include most of the activity.

There are many additional questions about nuclear winter that remain to be answered. With rapid surface cooling widespread fogs would develop, and they might affect the radiation balance at the surface. The presence of millions of tons of nuclear debris in the atmosphere could modify the properties of cloud droplets and so the removal rate of the debris. The nuclear clouds and modified natural clouds would also affect the overall infrared-radiation balance of the atmosphere, but the implications for surface temperatures remain uncertain. Daily variations, which have not yet been treated in the climate models applied to the nuclear-war problem, could also influence the dynamics and removal of the nuclear debris. Water injected by nuclear explosions and fires might affect atmospheric chemical and radiative processes. All these effects are important second-order refinements of the basic climatic theory of the nuclear winter. On the basis of the existing scientific evidence, however, none of these effects appears to be capable of significantly altering the major climatic impacts now predicted for a nuclear war.

There is also a lack of understanding of the interactions of the atmosphere with the oceans, which may have a major influence on short-term climatic changes. Through what was perhaps a series of coincidences, the eruption of El Chichón in the spring of 1982 was followed by an unusually intense El Niño warming of the South Pacific in the winter of 1982 and spring of 1983, associated with an unexpected calming of the southerly trade winds. These events were followed by unusual weather in North America and Europe in the winter of 1982 and throughout 1983. Most of North America suffered record-breaking cold that winter, and Europe enjoyed a balmy spring in December. Although proof that these events were related is lacking, the evidence suggests a potentially significant coupling of ocean currents, winds and weather on a comparatively short time scale; such a relation has yet to be defined rigorously.



GLOBAL CIRCULATION PATTERN of the atmosphere could be disrupted by a major nuclear war in the Northern Hemisphere. In the spring and summer (in the Northern Hemisphere) the average global circulation in the meridional, or north-south, direction is dominated by the large convective feature called a Hadley cell, in which the air rises over the hot, humid Tropics, splits into two streams and descends over subtropical and middle latitudes in both hemispheres, establishing secondary circulation cells at higher latitudes (*diagram at left*). If a large, dense cloud of smoke and dust were to be introduced into the troposphere in the North Temperate Zone during these seasons, the tropospheric heating effect at the southern edge of the cloud might be intense enough to reverse the normal mid-latitude subsidence, converting the Hadley circulation into an unusual pattern characterized by a single dominant cell with upper-level winds blowing briskly across the Equator from north to south (*diagram at right*). This novel circulation pattern has been observed in sophisticated computer models of the general circulation of the atmosphere developed by investigators in the U.S. and the U.S.S.R.

Our study also considered a number of secondary climatic effects of nuclear war. Changes in the albedo, or reflectivity, of the earth's surface can be caused by widespread fires, by the deposition of soot on snow and ice and by regional modifications of vegetation. Short-term changes in albedo were evaluated and found to be unimportant compared with the screening of sunlight. If significant semipermanent albedo changes were to occur, long-term climatic shifts could ensue. On the other hand, the vast oceanic heat source would act to force the climate toward contemporary norms following any major disturbance. Accordingly we have tentatively concluded that a nuclear war is not likely to be followed by an ice age.

We have also analyzed the climatic effects caused by changes in the gaseous composition of the atmosphere. The maximum hemispheric temperature perturbation associated with the production of oxides of nitrogen and the accompanying depletion of ozone is a cooling of no more than a few degrees C. The concentrations of greenhouse-effect gases would also be modified by a nuclear war; such gases might produce a surface warming of several degrees after the smoke and dust had cleared. These mutually offsetting temperature perturbations are uncertain, however, because the chemical and physical changes in the atmosphere caused by a nuclear war would be coupled through processes that are not adequately treated in existing models. Further analysis is clearly needed on this point.

Of course, the actual consequences of a nuclear war can never be precisely foreseen. Synergistic interactions among individual physical stresses might compound the problem of survival for many organisms. The long-term destruction of the environment and the disruption of the global ecosystem might in the end prove even more devastating for the human species than the awesome short-term destructive effects of nuclear explosions and their radioactive fallout. The strategic policies of both superpowers and their respective military alliances should be reassessed in this new light.

The Microbial Origin of Fossil Fuels

Chemical analysis of the most varied organic sediments, including coal and petroleum, reveals a surprising commonality: all derive much of their organic matter from once unknown microbial lipids

by Guy Ourisson, Pierre Albrecht and Michel Rohmer

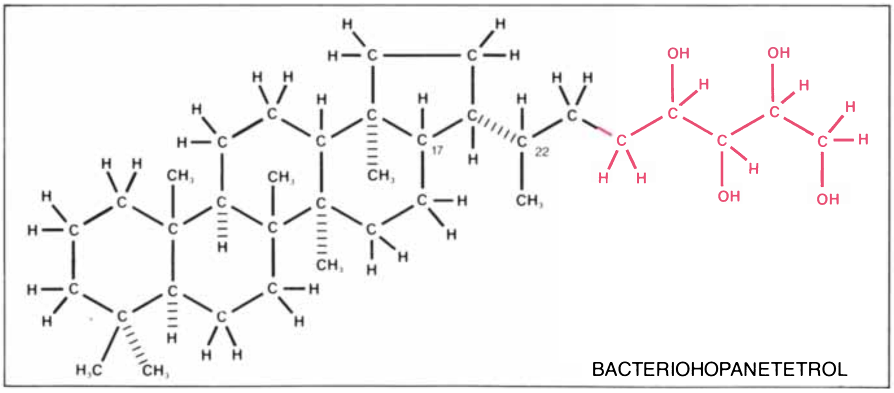
Coal, petroleum and natural gas are fossil fuels, but fossils of what? And what processes transformed the remains of living organisms into anthracite or Arabian crude? At one level geologists and geochemists have had answers to these questions for some time. Coal originated in vast primeval swamps, where the partial decay of dead trees and plants formed thick beds of the concentrated organic matter called peat. Petroleum, on the other hand, began for the most part as dispersed organic material in the sediments of inland seas or coastal marine basins; the sources of this material were long thought to have been plankton (which account for the bulk of all living matter in the ocean) as well as plant debris swept from the land by rivers. Both coal and petroleum form only if the organic matter is buried before it can be completely oxidized to carbon dioxide by microorganisms. As the carbon compounds sink deeper into the earth under accumulating sediment they are subjected to increasing temperatures and undergo chemical reactions. Oxygen and other elements are eliminated, resulting in a mixture composed in the case of petroleum and gas almost entirely of hydrocarbons and in the case of coal mostly of carbon.

Fossil fuels are thus the product of a small but steady leak in the circulation of carbon between the atmosphere and the surface of the earth, a process dominated by the uptake of atmospheric carbon dioxide by photosynthesizing organisms and the release of carbon dioxide by respiring plants, animals and microorganisms. Actually fossil fuels constitute only a small part of the organic matter that has been buried and thereby escaped oxidation. Most of this organic carbon (a term that excludes carbonates such as limestone) is stored in dilute form in sedimentary rocks; it becomes concentrated in usable deposits only under certain geologic conditions. The amount of carbon stored in sediments since photosynthesis began has been estimated at 10 quadrillion (10^{16}) tons, or roughly 10,000 times the quantity in all living organisms.

For the past 20 years we have been studying the genesis of fossil fuels and of all kinds of organic sedimentary deposits at a different level of detail: the level of the molecule. Our goal has been to determine the structure of individual fossil compounds, to deduce the structure of their precursors in living organisms and whenever possible to describe the reactions the substance has undergone in the sediment. Because each type of organism synthesizes its own characteristic compounds, chemical fossils whose carbon skeleton has survived more or less intact serve as biological markers, enabling investigators to estimate the contributions of various forms of life to a fuel or other deposit.

This field of study, which might be called molecular paleontology, is still relatively young, but already the work has yielded interesting results. We have found a striking similarity in hundreds of sediment samples from throughout the world: they all seem to be made up principally of microbial cell debris. The compounds in petroleum are derived from precursors found in the cell membrane of unicellular plankton and of bacteria and other microorganisms that inhabit the sea floor. In the case of coal the conclusion is more tentative. Nevertheless, our analysis suggests that many of the substances that can be extracted from coal are not derived directly from trees and higher plants but instead have been reworked by bacteria and fungi dwelling on the bottom of the swamp.

When we began our work, we were not optimistic about our chances of reaching such a general conclusion. The task before us was daunting. To an organic chemist a crude oil or the organic extract of a sedimentary rock is an appallingly heterogeneous mixture of hundreds or even thousands of compounds. Even when they can be separated in the laboratory, they are not always easily identified. At the time, owing largely to research sponsored by the American Petroleum Institute, more than 350 hydrocarbons (compounds consisting solely of hydrogen and carbon) and about 200 organic compounds containing sulfur had been identified in crude oils, but nearly all of them had relatively simple structures based on fewer than 15 carbon atoms. These molecules are well known and are easily recognized in an extract when they are



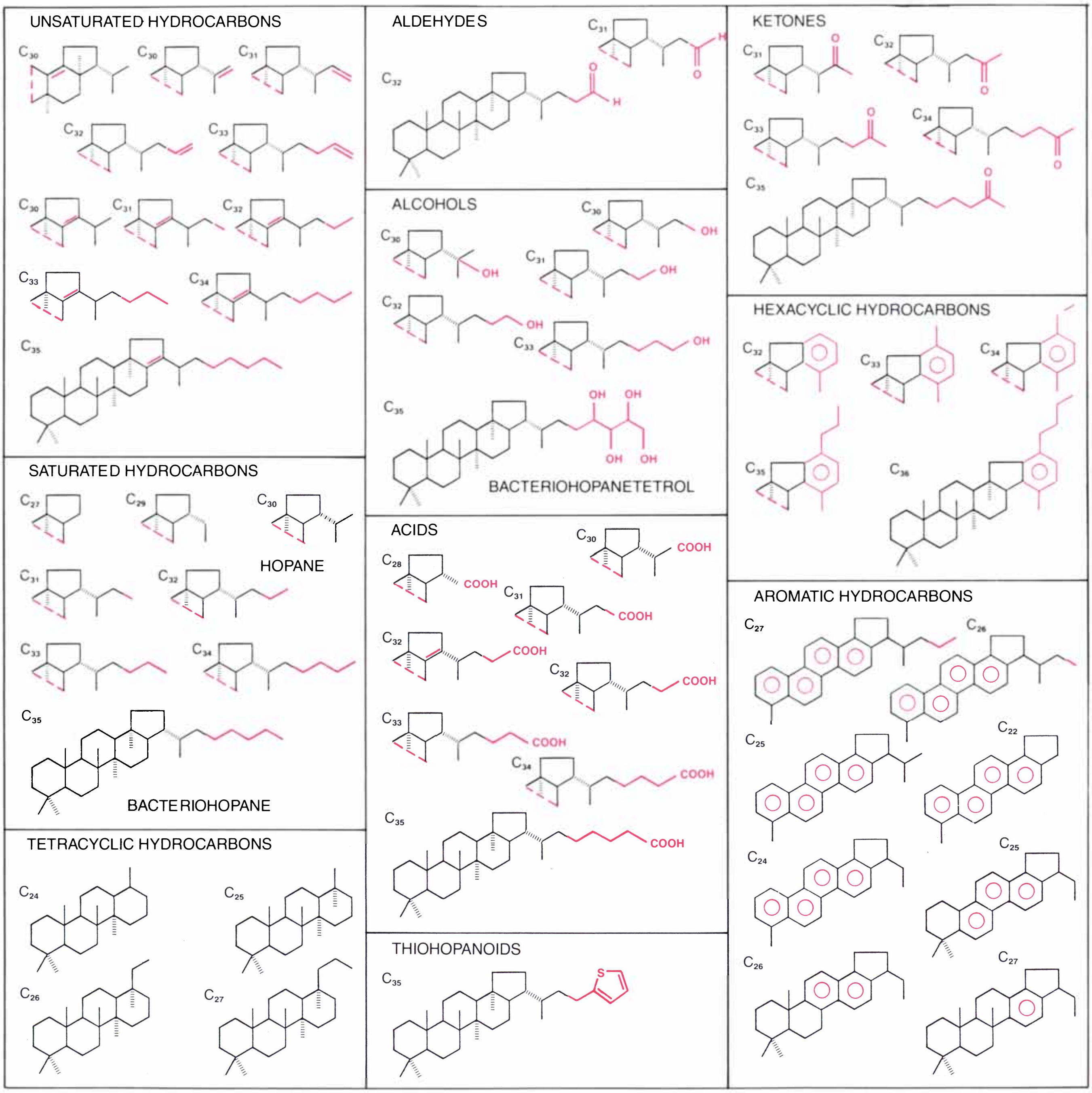
GEOHOPANOIDS, derivatives of the hydrocarbon hopane ($C_{30}H_{52}$), have been detected throughout the world in every sediment or fossil fuel that has been analyzed for them. The structure of more than 150 of these fossil molecules has been determined; some of them are shown here, with departures from the structure of the basic hopane molecule drawn in color. The ubiquity of the fossils is evidence of their microbial origin. Biohopanoids such as bacte-

compared with samples available in the laboratory. Determining the structure of less familiar and more complex molecules is harder, not least because the possible configurations are much more numerous. The simplest hydrocarbon, methane (CH_4), has only one possible structure, whereas C_{30} molecules can exist in hundreds of forms.

Yet it was the complex compounds that we had to analyze, because their structures reveal the most information about their origin. Methane and graphite (pure carbon) are totally uninformative, because any organic compound will decay into these substances when it has been heated enough. The structure of complex fossil molecules has been altered the least and thus points most unambiguously to the structure of the biological precursors of the molecules.

When we started our research, only a few such compounds had been isolated. Phytane ($C_{20}H_{42}$) and pristane ($C_{19}H_{40}$), branched-chain hydrocarbons built out of five-carbon units, had been found in many samples. This was not surprising, because both compounds can easily be derived from chlorophyll, the green pigment present in all photosynthesizing plants. Several other complex substances had been identified, but they had been found in only a few sources. Given the diversity of sedimentary environments and of possible source materials, there was little reason to doubt that heterogeneity would be the rule.

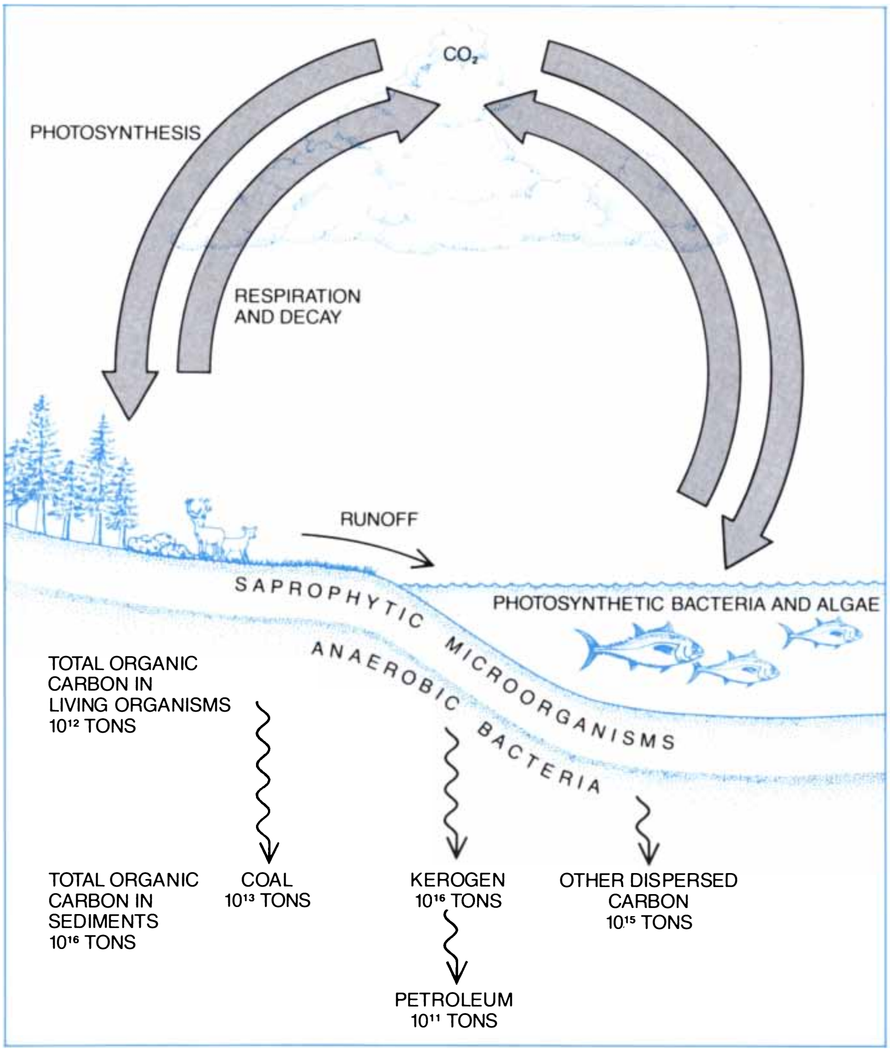
The methods we use to isolate individual compounds are familiar to every or-



riohopanetetrol, which has the formula $C_{35}H_{62}O_4$, are constituents of the cell membrane of many modern species of microorganisms. Some 15 biohopanoids are known; they were recognized in living microorganisms after their molecular fossils, the geohopanoids, were found. In the schematic drawings each angle denotes a carbon atom. Each carbon atom has four bonds, and at angles where fewer than four lines meet hydrogen atoms are implicit. Hatched lines represent bonds extending below the plane of the molecule, double lines stand for individual double bonds and circles for benzene rings. Isomerizations (spatial rearrangements) are common at positions 17 and 22.

ganic chemist. The first step is to extract the organic matter, which makes up, to pick a typical value, about 2 percent of the mass of sedimentary rock. The extraction is done with an organic solvent such as ultrapure chloroform. Great care must be taken at this stage and subsequent ones to avoid contamination of the material. The dissolved material is poured on a silica gel in a glass column, and the silica adsorbs the organic matter. Fractions of increasing polarity are then drawn out of the silica by pouring increasingly polar solvents through the column. A polar molecule is one with regions of opposite electric charge; in water, for example, the hydrogen atoms carry a positive charge and the oxygen atoms a negative charge. A substance is most likely to dissolve another substance if it is of approximately the same polarity. Water dissolves organic salts, alcohols and acids because they contain oxygen and are polar, but it does not dissolve nonpolar hydrocarbons.

This process, called column chromatography, separates the extract into saturated hydrocarbons (compounds without double bonds between the carbon atoms), aromatic hydrocarbons (those with benzene rings or similar structures), alcohols (compounds with hydroxyl, or OH, groups), carboxylic acids (with COOH groups) and other major fractions. Some of the fractions can be further broken down by treating them with a molecular sieve: an inorganic substance such as a zeolite (a type of silicate) with a fine-mesh structure. If saturated hydrocarbons, for example, are treated with a molecular sieve that has a mesh diameter of five angstrom units, molecules with a straight chain, which are about 4.5 angstroms in diameter, enter the sieve and become trapped, whereas the larger branched-chain and ringed molecules are held back.



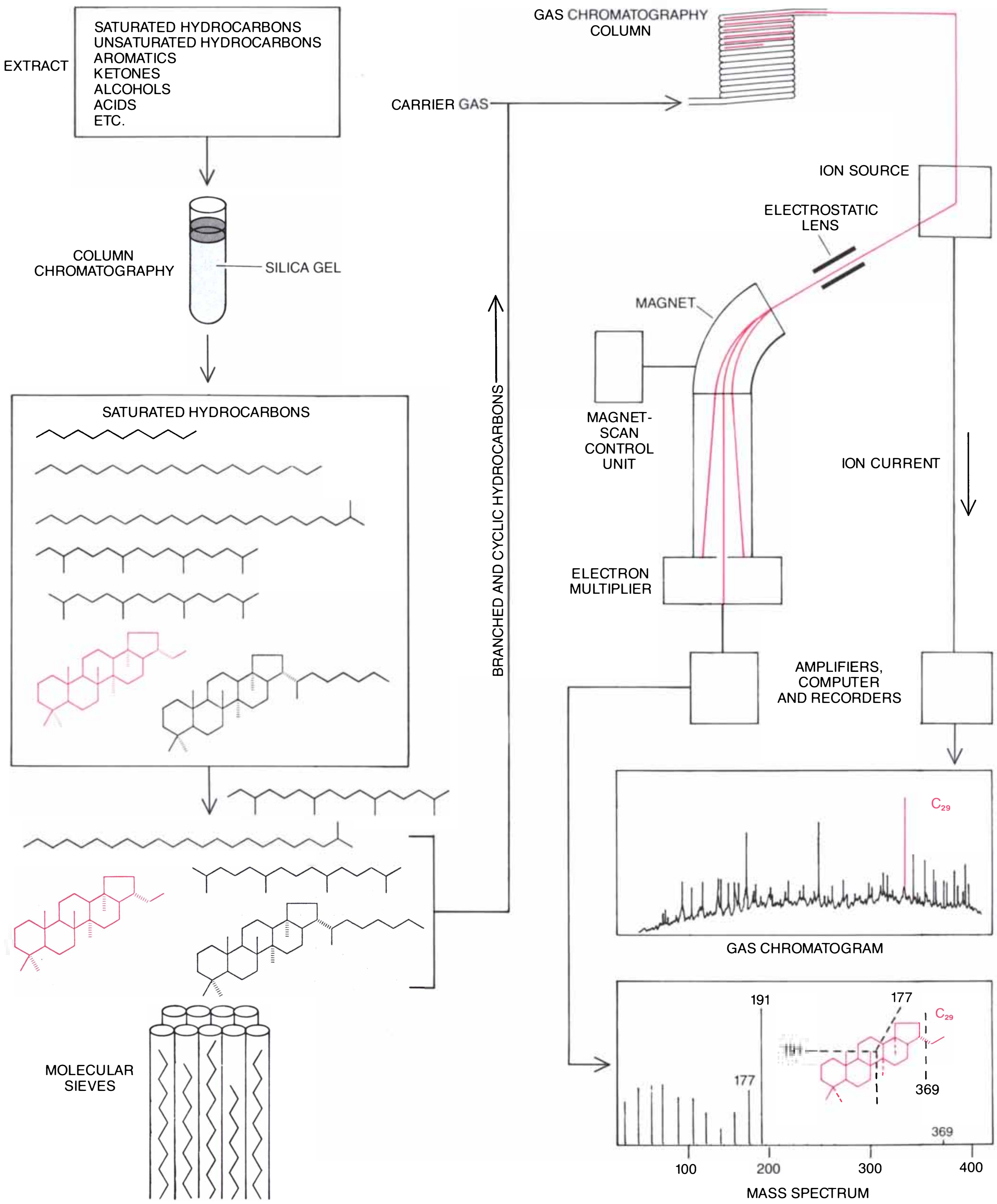
LEAK IN THE ORGANIC CARBON CYCLE has over hundreds of millions of years led to the accumulation of organic matter in sedimentary deposits. The cycle begins with photosynthesizing plants and microorganisms, which convert atmospheric carbon dioxide into organic compounds. The plants themselves and their predators break down some of these compounds, converting them back into carbon dioxide in the process called respiration. All but a tiny fraction of the remaining organic matter is degraded by saprophytic microorganisms and anaerobic bacteria when the larger organisms die. Microorganisms also rework organic material, synthesizing their own compounds. Some microbial matter as well as some of the remains of higher organisms escapes decomposition. Most of this material is in the form of the dispersed mixture called kerogen, but under the right geologic conditions it accumulates and forms fossil fuels.

The next step in the analysis is gas chromatography. Each family of molecules is vaporized and mixed with an inert (nonreactive) carrier gas such as helium. The gas mixture flows through a long, narrow capillary tube whose inner surface is coated with a thin film of nonvolatile liquid. As the organic molecules move through the tube, they are repeatedly retained and released by the liquid, but not to the same degree: those with longer carbon chains or more carbon rings tend to be retained longer on the average, because they are less volatile. (Indeed, the largest molecules are retained in the liquid solution for so long that the temperature in the tube must be raised gradually during the analysis in order to drive them out.) Different substances thus emerge from the capillary column at different times.

There are various ways to measure the quantity of each substance, but in our equipment, which combines a gas chromatograph with a mass spectrometer, the molecules are ionized by an electron beam as they emerge from the chromatography column. The magnitude of the resulting ion current is plotted on a chromatogram, and each peak generally indicates the proportion of molecules with a particular number of carbon atoms. The gas chromatogram does not, however, reveal the structure of a molecule, or even its mass. That is where the mass spectrometer comes in.

The electron beam not only ionizes the molecules but also breaks them into fragments of various sizes. In the mass spectrometer the electrically charged fragments are focused into a concentrated beam by an electrostatic lens and then channeled into a magnetic field. The heavier the fragment is, the less it is deflected by the magnetic field. The intensity of the field is made to vary rapidly, and when a fragment is deflected just enough to pass through a thin slit and strike the detector, the magnetic field strength at that instant is a measure of the mass of that fragment. Each peak emerging from the gas chromatograph yields a particular distribution of fragment masses—a mass spectrum. Ideally each peak and each mass spectrum correspond to a single compound, but sometimes a peak may actually be a superposition of several compounds with similar structures, and this complicates the task of interpretation.

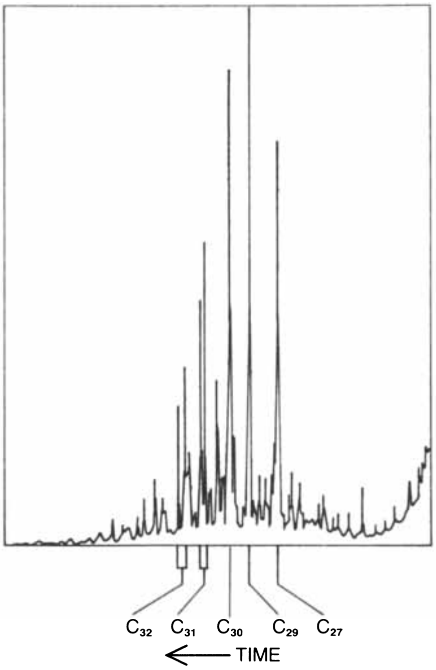
A mass spectrum would be of little help if it were not for the fact that carbon compounds tend to fragment in characteristic ways, breaking at their

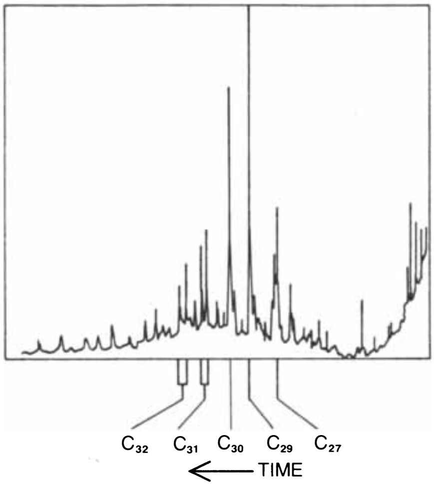


ANALYSIS OF AN ORGANIC MIXTURE by a series of separation procedures elucidates the structure of individual compounds. The mixture is first poured onto a column of adsorbent silica gel. Molecular families of increasing polarity are successively drawn through the silica when solvents of increasing polarity are added to the column. The saturated hydrocarbons are separated further by means of molecular sieves: minerals with a fine-mesh structure that admit the narrow straight-chain hydrocarbons and turn back the branched and cyclic molecules. Each fraction is vaporized, mixed with an inert carrier gas and channeled into the capillary tube of a gas chromatograph. The bore of the tube is coated with a liquid that repeatedly retains and releases the organic molecules. Larger molecules are retained for longer periods and therefore take longer to move through the column. Individual compounds emerge from the column at different times and are broken into ionized fragments by an electron beam; the peaks of the ion current, plotted on the chromatogram, measure the relative proportions of various compounds in the mixture. In the magnetic field of the mass spectrometer the fragments are separated by mass, because heavier fragments are deflected less. The fragmentation pattern is a clue to the structure of a molecule.

weakest links. The spectrometer also measures the mass of the entire molecule, because some molecules pass through the electron beam and the rest of the apparatus without being fragmented. The molecular mass together with the fragmentation pattern provides a firm basis for forming hypotheses about the structure of the molecule. These hypotheses must then be tested by comparing the mass spectrum of the fossil molecule with that of a known specimen. If, as is usually the case, a specimen is not on hand, it must be synthesized, a painstaking process that can require several years of work.

The ultimate proof of a molecule's structure is obtained by means of X-ray crystallography, which enables one to construct a "picture" of the position of each atom in a molecule. To grow the perfect crystal needed for crystallographic analysis, however, one must first isolate a relatively large quantity—a few milligrams, say—of a pure compound. This is very difficult, because the output from a gas chromatograph of an individual compound is on the order of micrograms or even nanograms. As a result we have been able to prove structural hypotheses with X-ray crystallography in only two cases.





GAS CHROMATOGRAMS of the saturated branched and cyclic hydrocarbon fraction of two unrelated samples show surprising similarities. The upper chromatogram was obtained from a Lorraine coal between 250 and 300 million years old; the lower one comes from a heavy crude oil about 150 million years old from the Aquitaine basin. Because of the difference in age and because the organic compounds in coal and petroleum were thought to come from different sources, the correspondence of peaks in the C_{27}-to-C_{32} region was unexpected. Each peak represents a hopanoid with a particular mass spectrum.

Soon after we began analyzing fossil fuels and sediment samples with these methods totally unexpected regularities became apparent. Specifically, the pattern of peaks in the C_{27}-to-C_{35} region of the gas chromatogram turned out to be very similar for all samples. A good instance of this surprising finding was our comparison of two samples completely dissimilar in both origin and age: a coal from Lorraine and a crude oil from Aquitaine [*see illustrations on these two pages*]. When we separated out the fraction of branched and cyclic (or ring-shaped) saturated hydrocarbons from each sample, we found not only that the chromatogram peaks corresponded but also that corresponding peaks had identical mass spectra. In other words, the substances producing the peaks in the coal were the same as those in the petroleum. Furthermore, the mass spectra of different peaks were similar in a way that suggested we were dealing with a homologous family of substances, the members of which were derived from some common ancestor. In particular they were all dominated by a fragment with a molecular mass of 191.

From the mass-spectrometry data we were eventually able to identify the family. Since then we and other investigators have analyzed thousands of sediment samples taken from all over the world, from every type of sedimentary environment and ranging in age from recent topsoil to rocks more than 500 million years old. Every sample has been found to include members of this family, compounds related to one complex molecule: hopane, a C_{30} pentacyclic (five-ring) triterpenoid. (Terpenoids are a large class of biological molecules synthesized from five-carbon units; they include, in addition to the phytane and pristane derivatives of chlorophyll cited above, steroids such as cholesterol and carotenoids such as beta-carotene, an orange pigment common in carrots and many other plants.) Hopane derivatives—we named them hopanoids—are a highly diverse group. So far we have isolated more than 150 different molecules from sediments and fully determined their structures. The ubiquity of sedimentary hopanoids prompted one of us (Ourisson) to offer, only half in jest, a doctoral degree to any graduate student who could find a sediment that does *not* contain hopanoids—and explain why.

So far no one has accepted the offer. As we shall explain below, other complex organic substances have been identified in sediments. Some are quite widespread, but only the hopanoids are ubiquitous. The precursors of these fossils had to be equally universal.

When the importance of fossil hopanoids as sediment constituents first became evident in the early 1970's, their precursors were not immediately obvious, because hopanoids had been detected in only a few species of living organisms. They had been isolated first from several trees and grasses and particularly from ferns. That primitive fernlike plants had contributed to the formation of some coals was and is beyond question, but clearly none of the trees and higher plants could account for the presence of fossil hopanoids in all sediments, particularly those of marine origin. Some simple hopanoids had also been found in a few species of blue-green algae and in one bacterium, and C. W. Bird and W. W. Reid of Queen Elizabeth College in London had suggested these discoveries might explain an early finding, by E. V. Whitehead of the British Petroleum Company, of a few hopanoids in one Nigerian crude.

Not only were these modern hopanoids too rare to explain the omnipresence of fossil molecules in sediments but also they all had either 29 or 30 carbon atoms, whereas many fossil hopanoids were C_{35} or even C_{36} compounds. It is highly unlikely that these compounds acquired additional carbon atoms after deposition, and so the origin of sedimentary hopanoids remained a mystery. Then in 1973 Hans J. Förster and Klaus Biemann of the Massachusetts Institute of Technology and their co-workers, who were studying a specific bacterium, *Acetobacter xylinum,* for unrelated reasons, isolated from it a C_{35} hopanoid. The compound was called bacteriohopanetetrol, the -tetrol suffix indicating that it has four alcohol, or hydroxyl, groups attached to a side chain in place of hydrogen atoms. (An alternative name is tetrahydroxybacteriohopane.)

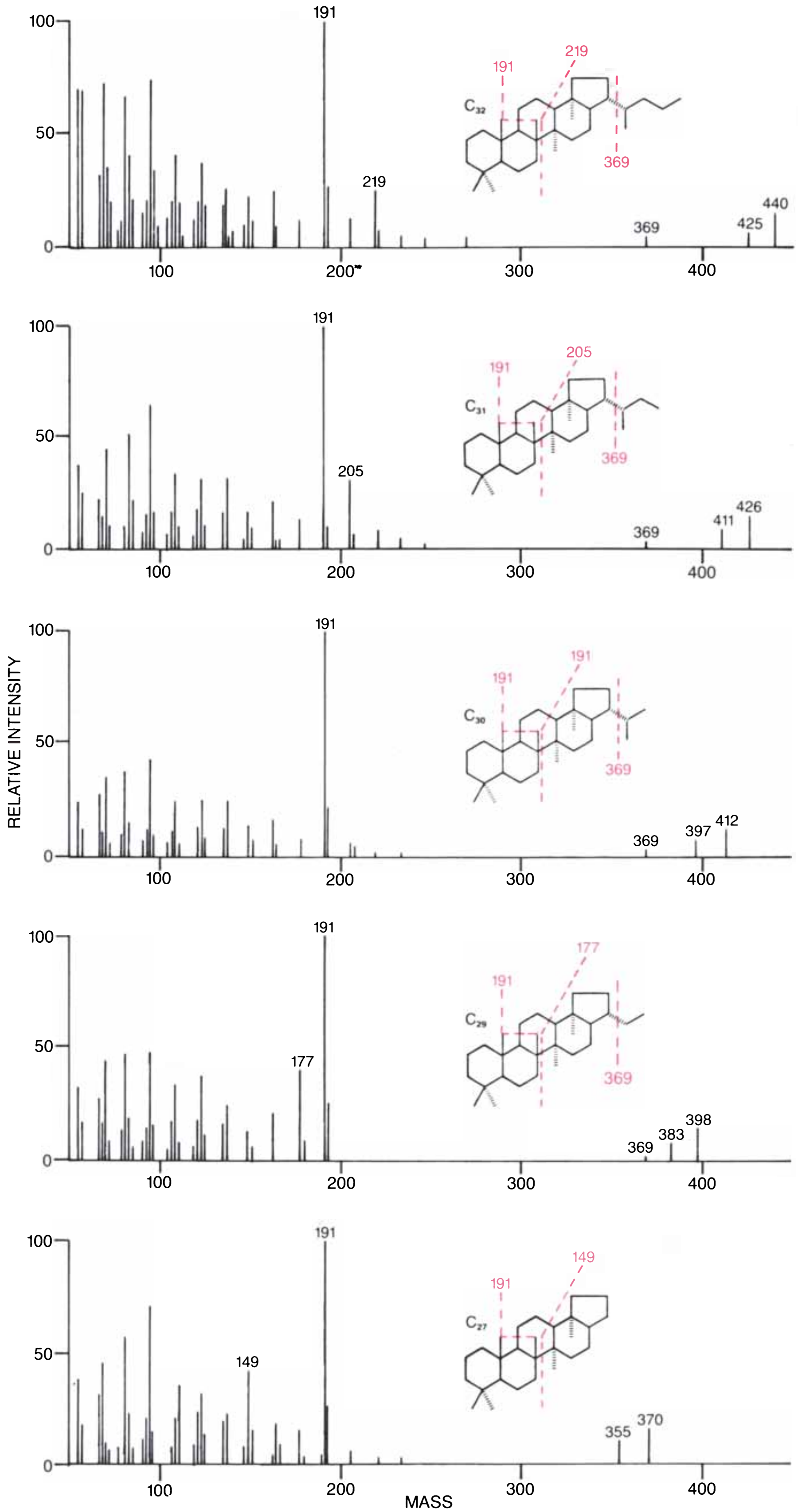
Förster, Biemann and their colleagues suggested two potential configurations for bacteriohopanetetrol. The structure they preferred would not have allowed it to serve as a precursor of the known fossil hopanoids, because that particular carbon skeleton had not been found among the fossils. When we isolated bacteriohopanetetrol in our laboratory, however, we were able to show that the other possible structure, which is compatible with sedimentary hopanoids, is the correct one. Essentially we subjected the compound to a chemical reaction designed to simulate the degradation it might undergo in sediment and showed that the resulting compound was identical with a known sedimentary hopanoid.

We then began to look for bacteriohopanetetrol in other microorganisms. It turns out to be one of a family of highly polar hopanoids, and we were able to isolate C_{35} molecules that were potential precursors of fossil hopanoids from about half of the many species of bacteria and blue-green algae we analyzed. In some bacteria we also found C_{36} hopanoids to account for the C_{36} fossils we had already identified.

The wide distribution of biohopanoids explains the ubiquity of geohopanoids: the fossils are ubiquitous because the ocean floor and the top layer of soil on land are teeming with bacteria. When organisms die, much of their organic matter is broken down into simple molecules by these enzyme-wielding bacteria. Most of the organic matter is thereby converted into gases such as carbon dioxide, but the bacteria also rework some of the simple components into their own complicated molecules, including hopanoids. As the bacteria are buried under accumulating sediment, they themselves die and decay. The hopanoids are buried progressively deeper and are transformed by the action of other microorganisms as well as by thermal reactions catalyzed by clay or other minerals. Bacteriohopanetetrol, for example, may be reduced to a hydrocarbon through the replacement of its hydroxyl groups by hydrogen atoms, or it may be oxidized, its side chain thereby being cut in various places. This is just one of many chemical reactions, most of them still understood only in general terms, that produce the amazing array of fossil hopanoids from a relatively small family of biological molecules.

It was not merely the ubiquity of fossil hopanoids that persuaded us of their microbial origin. The C_{30} fossils, in fact, could have come from higher plants. So far, however, C_{35} biohopanoids have been found only in microorganisms. The one-to-five-carbon tail of many fossil hopanoids is the most convincing evidence of their microbial origin. To say they are derived from microorganisms, however, is to solve only half of the problem. What role do hopanoids play in living microorganisms? Since they are so widespread, their role must be an important one, which raises an interesting subsidiary question: Why did biohopanoids go undetected for so long?

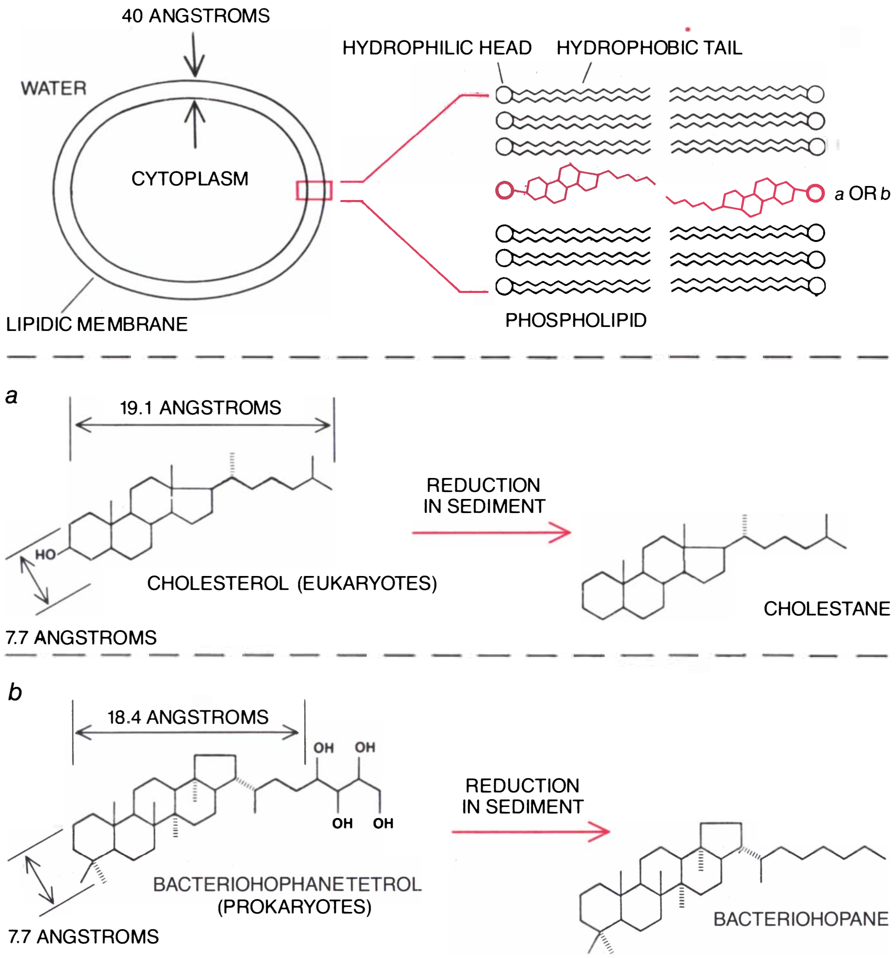
To this last question there is a simple answer: in their biological form hopanoids can be detected only if they are de-



MASS SPECTRA of the five corresponding chromatographic peaks in the illustration on the opposite page are identical, indicating that the peaks are produced by the same compounds in the coal and the petroleum. The presence of fragments with a molecular mass of 191 is characteristic of hopanoids, which tend to fracture at weak carbon bonds in their third ring. The peak farthest to the right in each case represents the mass of intact molecules that have escaped fragmentation by the electron beam. The molecular mass and the fragmentation pattern enable investigators to form a hypothesis about the structure of a molecule; the hypothesis must be confirmed by comparison with the mass spectrum of a known sample synthesized for that purpose.

liberately looked for, and no one had looked for them. Biohopanoids are difficult to isolate because they are amphipathic lipids: one part of the molecule is essentially a hydrocarbon and can be dissolved in a nonpolar organic solvent such as chloroform; the other part, the carbon side chain, is loaded with hydroxyl or nitrogen groups, is therefore polar and dissolves only in polar solvents such as water or alcohol. As a result no single solvent will dissolve biohopanoids. They can be extracted only with a mixture of a nonpolar solvent and a polar solvent, say chloroform and methanol. Most fossil hopanoids, in contrast, have had the polar oxygen or nitrogen groups stripped from their side chain and, since they are pure hydrocarbons, dissolve easily in a nonpolar solvent. This explains why a broad class of bacterial lipids could have been discovered first through their fossils.

The amphipathic structure of biohopanoids is also a clue to their function in microorganisms. Cell membranes are made up of amphipathic lipids. In an aqueous environment these molecules naturally form a closed membrane two molecules thick, with their hydrophobic hydrocarbon ends pointing toward each other in the inside of the membrane and their water-soluble ends pointing toward the outside and the inside of the vesicle. The hydrophobic end of the primary membrane lipids in most cells is simply a pair of straight hydrocarbon chains, but a membrane made up solely of such molecules would be too flexible and unstable. In the membranes of all eukaryotic cells (nucleated cells, including those of all higher organisms) some of the straight-chain lipids are replaced by cholesterol, an amphipathic lipid whose four-ring structure makes it more rigid and thus braces the membrane.



AMPHIPATHIC LIPIDS, the main structural constituents of all cell membranes, have a hydrophilic (polar) head and a hydrophobic (nonpolar) tail. The membrane is formed from a bilayer of such molecules, with the hydrophilic heads pointing toward the aqueous environment both inside the cell and outside it and the hydrophobic ends facing one another in the middle of the membrane. In a typical membrane lipid, a phospholipid, the head is a phosphorylated glycerol and the tail is a pair of highly flexible straight-chain hydrocarbons. In the eukaryotic cells of higher organisms the fluid membrane is reinforced by more rigid, ringed cholesterol molecules. The authors have found that biohopanoids, which have almost the same dimensions as cholesterol and are also amphipathic, have the same function in many bacterial (prokaryotic) membranes. Recognizable fossil steroids and hopanoids are abundant in sediments because membrane lipids are plentiful in living cells and resist decay. The fossil molecules are often reduced, losing oxygen and gaining hydrogen; they may also be oxidized to acids or aromatics.

Hopanoids serve the same purpose in the membranes of prokaryotes (cells without nuclei) such as bacteria and blue-green algae, almost none of which have been found to contain cholesterol. Our first clues to this equivalence of function were the similarities in structure between cholesterol and bacteriohopanetetrol and their almost identical size. The conclusion has since been confirmed in our laboratory and by Karl Poralla of the University of Tübingen.

Furthermore, it appears likely, although it has not been demonstrated, that hopanoids are not only the structural and functional analogues of sterols but also their evolutionary precursors. Both sterols and hopanoids are terpenoids, and the biosynthetic reactions required for their formation are fundamentally similar. There is, however, one significant difference: sterols, like many other essential constituents of eukaryotic cells, can be formed only in the presence of free oxygen. The primitive atmosphere was anoxic, but the appearance more than two billion years ago of photosynthesizing blue-green algae is thought to have led to the rise in oxygen levels that was a precondition for the later emergence of eukaryotes and other aerobic organisms.

Although we have no direct evidence, it is reasonable to believe that early blue-green algae, like their present-day relatives, had membranes reinforced with hopanoids, and that eukaryotic sterols are the evolutionary descendants of these prokaryotic lipids. We have even identified a substance that might have been an intermediate between the two molecular families: isoarborinol, a pentacyclic terpenoid with structural similarities to both hopanoids and sterols. So far isoarborinol has been found only in several sediments as well as in some higher plants, but it may one day be isolated from some uninvestigated bacterium. These observations and others have led us to a general hypothesis, which we are still in the process of testing and elaborating: that the function of reinforcing cell membranes is carried out in all living organisms by polyterpenoids.

Other terpenoid fossils in addition to hopanoids provide evidence of the microbial contribution to the organic carbon content of sediments. Steroids are common in sediments, and some of them may come from higher, sterol-containing organisms. Many steroid fossils, however, have an additional methyl group in the ring bearing the hydroxyl group, and these derivatives are usually present only in trace amounts in eukaryotes. Their common occurrence in sediments must therefore be due to microorganisms. Geoffrey Eglinton of the University of Bristol has recently shown that the most likely source of these molecules is the dinoflagellates, unicellular planktonic eukaryotes found in both the ocean and fresh water. Some of them, however, may also come from methylotrophic bacteria, in which the oxidation of methane is the basis for the synthesis of organic molecules.

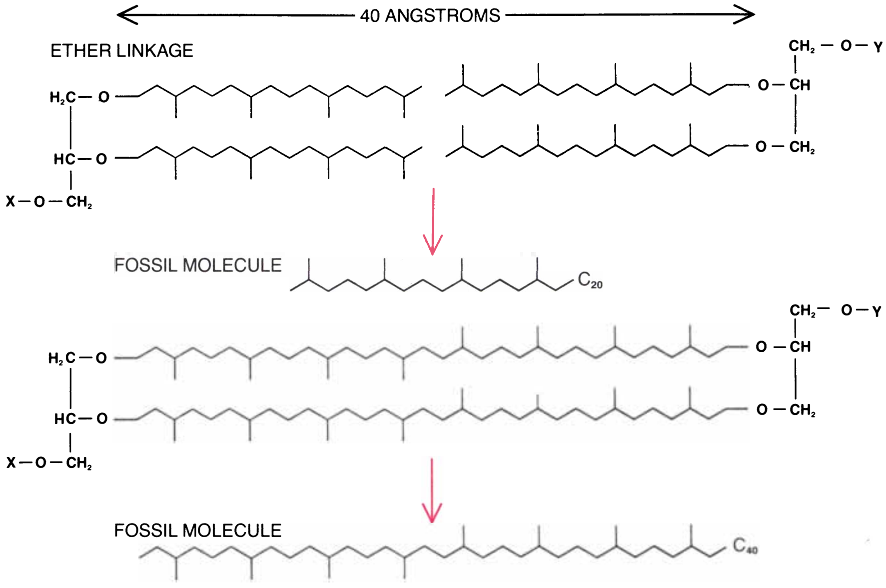
The methane that methylotrophic bacteria thrive on is produced in part by anaerobic bacteria in lower sedimentary strata. These methanogenic bacteria have unusual membrane lipids: the polar groups are connected to diterpene (C_{20}) and tetraterpene (C_{40}) branched-chain hydrocarbons by an ether linkage, in which two carbon atoms are connected by an oxygen atom. The branched ethers yield fossils that are easily recognized, in particular by the coexistence in close proximity of C_{20} and C_{40} molecules. (The latter span the membrane, thereby increasing its stability; methanogens do not synthesize steroids, hopanoids or similar cyclic terpenoids as membrane reinforcers.) Although so far we have not been able to analyze many samples containing such fossils, it is highly probable they are as widespread as hopanoids: the presence of methane in every conceivable organic-rich, anaerobic environment suggests methanogens live in all sediments.

Hopanoids, steroids and branched ethers—the inventory of microbial products we have identified in sediments is a rich one. And we have not mentioned many molecular fossils whose structure, like that of isoarborinol, strongly suggests a microbial origin but for which precursors have yet to be found in living organisms. One important question remains: Do these complex molecules constitute, in quantitative terms, a significant portion of sedimentary organic matter, or are they merely eccentricities we happened to identify?

Consider first an example. A lignite we studied from Yallourn in Australia contained several hundred parts per million of a single C_{32} hopanoid acid. That may not sound like much, but it means that a one-meter cube, weighing about two tons, contains approximately one kilogram of the acid, making it easily the most abundant defined organic substance in the coal. Nor is the Yallourn lignite a trivial deposit: it is an important Australian export to Japan.

Moving to global evaluations, which are always risky, we encounter a special problem. Approximately 90 percent of the 10^{16} tons of organic carbon estimated to be in sediments is in the form of insoluble kerogen, the disseminated organic matter that through thermocatalytic reactions is converted into petroleum as sedimentary rocks sink into the earth. Because kerogen is insoluble, there is no easy way to determine its content of microbial matter. According to our analyses, however, hopanoids account for about 5 to 10 percent of the earth's 10^{15} tons of soluble organic carbon. By this estimate, which can hardly be off by more than an order of magnitude, the global stock of hopanoids alone would be at least 10^{13} or 10^{14} tons, more than the estimated 10^{12} tons of organic carbon in all living organisms. The molecular fossils of methanogens might prove to be of equal importance, but it is difficult to evaluate their input, because the main source so far of these fossils appears to be kerogen, from which they must be painstakingly extracted with reactions specific to ethers.

Many of the simpler constituents of crude oils and sedimentary extracts can come from a number of sources and therefore may also be microbial in origin. The common straight-chain or branched hydrocarbons of the methane series are good examples. Others are phytane and pristane, which clearly can be derived from chlorophyll and therefore from plants or photosynthetic algae but which can also be fossils of methanogenic membrane lipids. A lot of information about the sources of fossil organic matter is lost as a result of two processes: the thermal, clay-catalyzed "maturation" reactions responsible for isomerizations and for the breakdown of complex molecules into simpler petroleum hydrocarbons, and bacterial degradation of organic compounds in shallow deposits.



METHANOGENIC BACTERIA have unusual membrane lipids: the hydrocarbon tails are branched rather than straight and are connected to the polar head by ether linkages (–O–) rather than the ester linkages (–CO–O–) found in eukaryotes and most other bacteria. (*X* and *Y* represent hydrophilic groups.) The C_{20} lipids (*top*) form an ordinary bilayer, whereas the C_{40} molecules (*bottom*) span the membrane and strengthen it by acting like rivets. Although phytane, a C_{20} fossil molecule, can also be derived from chlorophyll, the fossils of methanogenic lipids are often recognizable by the coexistence of the C_{20} and the C_{40} compounds; intact branched ethers have also been found in sediments. Given the ubiquity of methanogens, their membrane lipids are probably an important contributor to organic sedimentary matter.

Although maturation reactions can make the detective work of the molecular paleontologist more difficult, they are also the source of the most important practical application of such research: its use in petroleum prospecting. (This explains why the Société Nationale Elf-Aquitaine, the French oil company, has supported our work.) If the rate of a particular maturation reaction (say an isomerization) is known as a function of temperature, it can serve as an indication of whether a given sedimentary deposit has been heated to the point where it is a potential petroleum source rock. Moreover, when a petroleum reservoir has been found, the extent to which maturation reactions have proceeded in the oil can be used to help locate the source rock, which may be at a considerable distance, and thus to guide further drilling. The study of maturation reactions is still in the early stages, but the application of existing knowledge to petroleum exploration has already started, in particular by a group led by Wolfgang K. Seifert of the Chevron Oil Field Research Company in Richmond, Calif.

The molecular makeup of kerogen is one more important example of the terra incognita stretching before us. The work so far has been rewarding. Where chaos was expected, we have found enough order to prove beyond doubt the microbial origin of many of the important constituents of fossil fuels and of sedimentary organic matter in general. It is also particularly gratifying that our research has had resonances in other fields, leading us and other workers to open new chapters in bacterial biochemistry and biochemical evolution.

A Superluminous Object in the Large Cloud of Magellan

A giant nebula in this small galaxy close to our own holds an object that is 50 million times brighter than the sun. If it is one body, it is far more massive than any known star

by John S. Mathis, Blair D. Savage and Joseph P. Cassinelli

The mass of the sun (2×10^{33} grams) is a standard of measurement for the mass of other celestial objects. Until recently it was the generally accepted view among astronomers that no star much more massive than about 100 times the solar mass could form. Now this view has been challenged by the observations that have been made of a superluminous and possibly supermassive object in a small galaxy close to our own: the Large Cloud of Magellan. The object is designated R136. It is in the Tarantula Nebula, which is also known as the 30 Doradus nebula because it is in the southern constellation Doradus. If R136 is a single star, it could be as much as 1,000 times more massive than the sun.

Glowing gaseous nebulas are among the loveliest and most impressive objects in the universe. The 30 Doradus nebula is the brightest and largest gaseous nebula in the 30 or so galaxies of the local group that includes our own galaxy. It is of irregular shape and extraordinary size. Whereas to the unaided eye the Great Nebula in Orion looks like a fuzzy star, 30 Doradus covers an area of the sky comparable to that occupied by the sun or the moon, in spite of the fact that it is more than 100 times farther away than the Orion nebula. Its diameter is about 1,000 light-years, compared with the Orion nebula's three. Its gas is highly ionized: most of its atoms have lost one electron or more. Indeed, it contains 1,500 times more ionized gas than the Orion nebula. The ionization must be the result of ultraviolet radiation emitted by the massive hot young stars embedded in the nebula.

R136 is the brightest object in 30 Doradus. (The designation comes from a catalogue of the brightest stars in the Large Cloud of Magellan prepared by Michael Feast, A. D. Thackeray and A. J. Wesselink of the Radcliffe Observatory in South Africa.) It is near the center of the nebula and is surrounded by dozens of fainter stars, each of which would ordinarily be considered bright. The central object radiates about a million times more light than the sun at visible wavelengths and is another factor of 50 more luminous when ultraviolet wavelengths are included. Because of its great luminosity, it may account for more ionization than any other object in the nebula.

In 1980 J. V. Feitzinger, Wolfhard Schlosser, Theodor Schmidt-Kaler and Christian Winkler of the University of the Ruhr got excellent photographs of R136 with the 3.6-meter telescope of the European Southern Observatory in Chile. They found that the object actually has at least three distinct components, arranged in a curve resembling a comma. The brightest component was designated R136a. R136b and R136c are fainter, redder and probably cooler. The German workers concluded that R136a might be a star considerably more massive than the commonly accepted upper limit for stellar masses.

The first visible-light spectra of R136, obtained in 1950 by Feast at the Radcliffe Observatory, indicated that the object has a peculiar spectrum characteristic of extremely hot stars. Later work by Nolan R. Walborn of the Space Telescope Science Institute, Peter S. Conti of the University of Colorado at Boulder and Dennis Ebbets of the Space Telescope Institute confirmed that finding.

Stars with a surface temperature higher than 30,000 degrees Kelvin emit most of their radiation at the shorter ultraviolet wavelengths that cannot penetrate the atmosphere of the earth. Therefore ultraviolet observations of R136a were essential to an understanding of the object, but they had to wait for the technical advances brought by the space age. One such advance was the International Ultraviolet Explorer satellite (IUE), a space observatory operated jointly by the National Aeronautics and Space Administration, the European Space Agency and the Science Research Council of the United Kingdom. With this satellite ultraviolet spectra of R136a were secured in 1978.

The satellite revealed a spectrum resembling in many respects that of the hottest normal stars known, type O3. (The *O* designates the hottest spectral class. The lowest associated numeral designates the hottest star in a class. No O1 or O2 stars have been defined.) It was Walborn who first recognized the O3 stars as a distinct type. They have a surface temperature estimated at 50,000 degrees K. and are among the most luminous stars known.

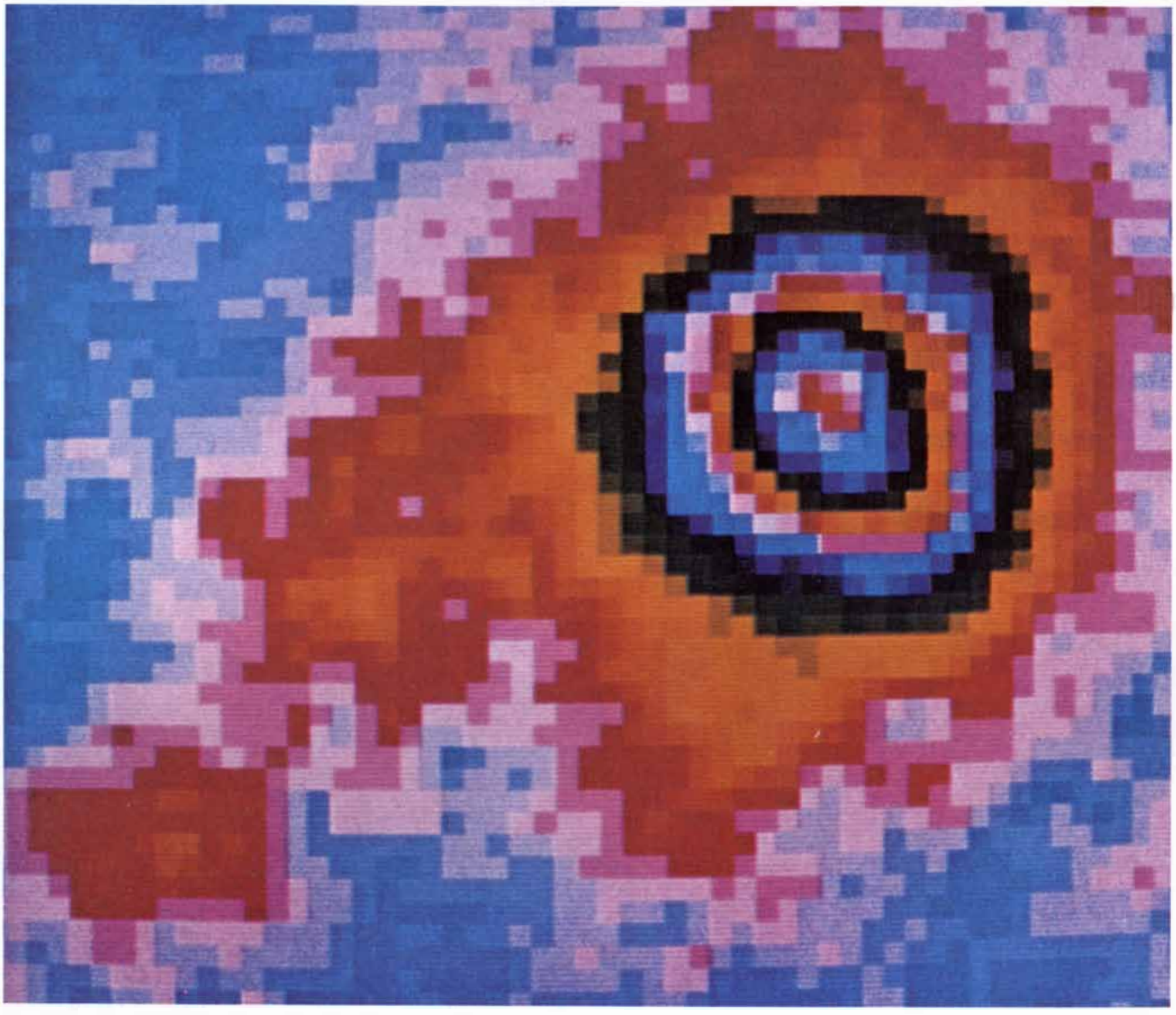
The ultraviolet spectra of R136a show what are called P Cygni lines, after a highly luminous star in the constellation Cygnus. At ultraviolet wavelengths such lines often originate with highly ionized atoms of carbon, oxygen and nitrogen: C IV, O IV, O V, N IV and N V. (The Roman numerals specify the number of electrons removed plus one; for example, C IV is carbon with three electrons removed.) The P Cygni lines have a peculiar profile: they look like an absorption line on the short-wavelength side of the rest wavelength and like an emission line on the long-wavelength side. The rest wavelength is one that would be emitted by an object at rest with respect to the observer, and so the P Cygni lines imply that the object is losing mass in the form of an envelope of gas being expelled from it as a stellar wind. The gas on the far side of the object from the observer is going away from him, so that the wavelengths of its emissions are lengthened; the gas on the observer's near side is coming toward him, so that the wavelengths of absorption are shortened.

The velocity of the outflow of matter from R136a can be estimated from the profile of the P Cygni lines. In particular the wavelengths of the absorption on the short-wavelength side of the C IV line extend 18 angstrom units from the rest wavelength of the line. This indicates that matter is moving away from R136a at a velocity of 3,500 kilometers per second. The stellar wind of R136 is best described as a hurricane; for a normal hot star the average wind speed is about 2,000 kilometers per second. Only type O3 objects exhibit a wind speed as high as that of R136a.

The high surface temperature of R136a is implied by the high state of ionization of the matter in its atmosphere. As the temperature of a star rises its outer layers become more highly ionized because of the increasing number of energetic photons in its radiation and also because collisions between particles become more violent. Thus the presence of some ions and the absence of others as represented in a star's spectrum are diagnostic of temperature. For R136a the presence of lines representing C IV, O V and N V and the absence of lines for Si IV (triply ionized silicon) imply a surface temperature in the range from 45,000 to 80,000 degrees K. In a cooler atmosphere there would be conspicuous lines of Si III and Si IV.

The ultraviolet spectrum of R136a resembles that of O3 stars except for the exceptionally strong and broad emission line of singly ionized helium (He II) near a wavelength of 1,640 angstroms. Although some O3 stars show He II emission, none exhibit a line as strong and broad as R136a's. It is noteworthy, however, that the peculiar objects known as Wolf-Rayet stars show exceptionally strong He II emission because of their massive stellar wind. These hot stars are thought to be in a stage of evolution more advanced than that of normal O stars. The similarity of the spectrum of R136a to the spectra of both O3 and Wolf-Rayet stars suggests it might even be produced by a collection of stars of both types.

The extreme nature of R136a becomes evident when one considers the number of "normal" stars that would be needed to reproduce its spectrum. For example, R122 is the most luminous normal star in the Large Cloud of Magellan, being three times as bright as the next brightest star in that galaxy. It is three million times more luminous than the sun; most of the other supergiant stars are only 500,000 times more luminous. R122 has an O3 spectrum roughly resembling that of R136 but with weaker He II emission. To get the R136a spectrum from a group of normal stars would call for about 13 stars like R122 or about 40 less luminous O3 stars plus



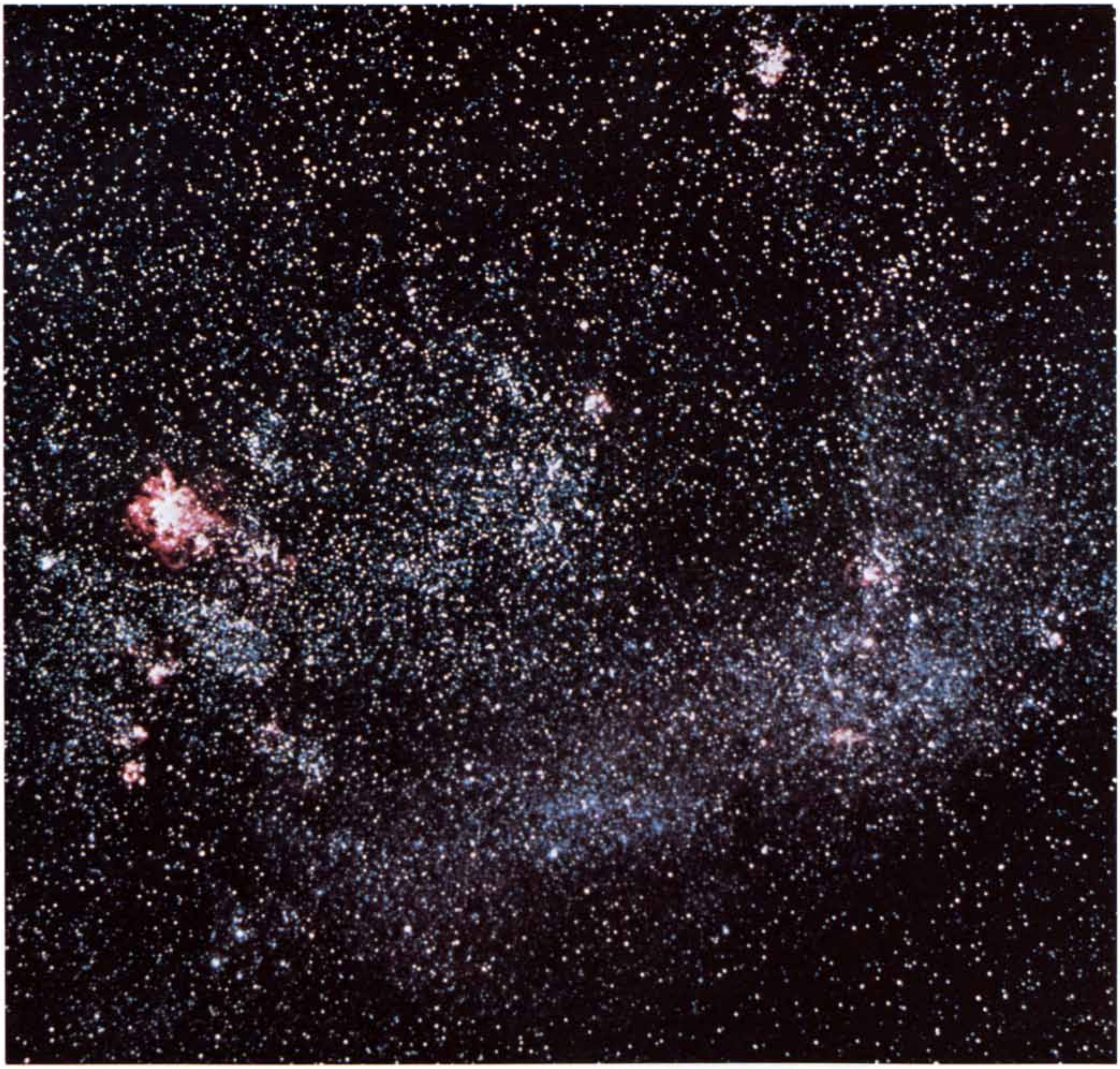
AREA OF THE OBJECT R136 appears in this computer-processed photograph made at visible wavelengths. The area is near the center of the Tarantula Nebula (also known as 30 Doradus because it is in the southern constellation Doradus). The photograph indicates that R136 has at least three components. The brightest one is the irregular, predominantly blue area at the center right; it has been designated R136a. Immediately to its left is R136b, which is about 20 percent as bright as R136a. The reddish area at the lower left is R136c.

20 Wolf-Rayet stars of normal brightness. O3 stars are exceedingly rare; only four are known in the entire Large Cloud of Magellan and only 10 in our own galaxy. All these presumably rare objects would have to be packed into a volume with a diameter of one light-year or less.

In terms of stellar evolution it is not too unreasonable to suppose a mixture of O3 and Wolf-Rayet stars could be identified; indeed, such a mixture exists in the giant nebula in our own galaxy called the Carina Nebula. The four O3 stars in the Carina Nebula, however, extend in space over a distance of 10 light-years.

One can try to determine the nature of R136a by studying the structure of the R136 region on photographs made with telescopes on the ground. The main trouble with this approach is that the earth's atmosphere blurs the image. Nevertheless, at a time of excellent "seeing" last year Y. H. Chu of the University of Wisconsin at Madison made photographs of the inner region of 30 Doradus in order to sort out the types and numbers of stars near R136.

Many hot stars lie within several minutes of arc of R136, and Jorge Melnick of the University of Chile had identified several of them as stars of type O3. Chu has carefully analyzed many images of the R136 region and has concluded that within the R136a component, which has a diameter of only three seconds of arc, one can identify at least four stellar objects. Unfortunately the instruments of the International Ultraviolet Explorer can resolve individual stars only if they



LARGE CLOUD OF MAGELLAN was photographed with the 61-centimeter Schmidt telescope at the Cerro Tololo Inter-American Observatory in Chile. The Large Cloud, which is visible only from the Southern Hemisphere, is the galaxy closest to our own. The 30 Doradus nebula is the pink area at the center left. The nebula harbors many massive stars hot enough to ionize the interstellar gas; they radiate most of their energy at short wavelengths and appear blue in the photograph. The red patches are ionized gas emitting the red Balmer line of hydrogen in the visible region of the spectrum. The 30 Doradus nebula is the largest such object in the entire local group of galaxies. Some 1,000 light-years in diameter, it occupies an area in the sky that is approximately a third as large as the sun.

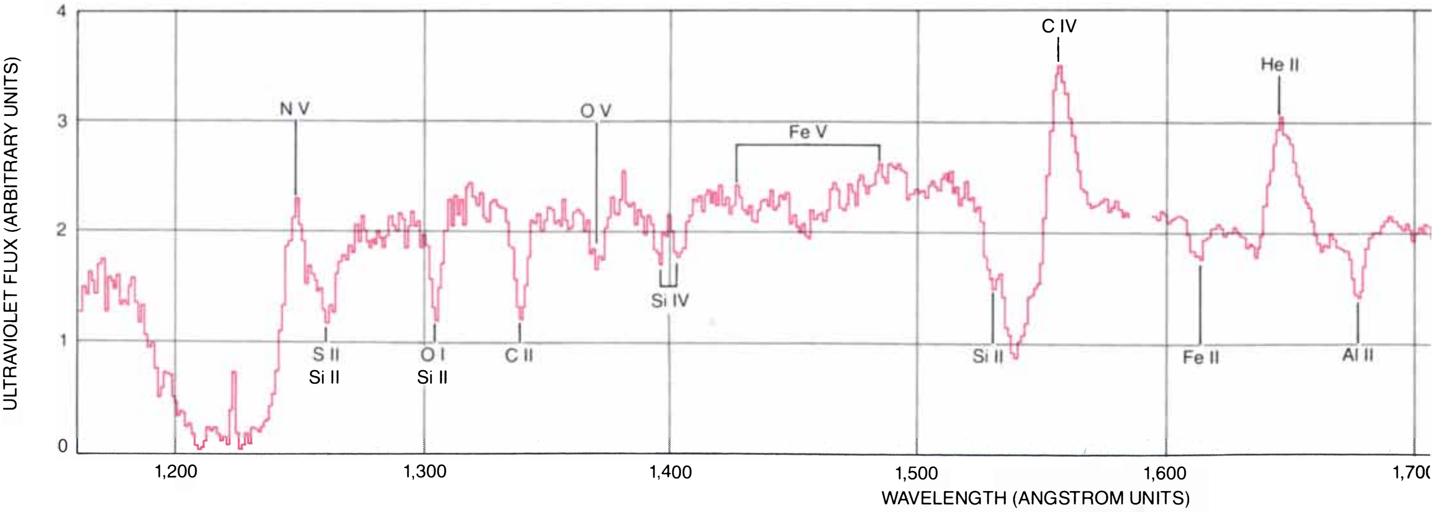
are separated by at least three seconds of arc, and so the spectra from the satellite include contributions from all sources of ultraviolet radiation in the R136a area. Chu has labeled the dominant source R136a1. It is this object that is now the candidate for a superluminous star. Chu has also identified a fainter point source, R136a2, about .5 arc-second from a1.

A technique that can be exploited to determine the details of an image distorted by the earth's atmosphere is speckle interferometry. The technique consists in making hundreds of very short exposures in rapid succession. Each exposure shows the image when the atmosphere is in a definite (but unknown) state. By applying certain mathematical procedures to the individual images the investigator can extract the true structure of the image almost as well as the mirror of the telescope could resolve it if the telescope were above the atmosphere.

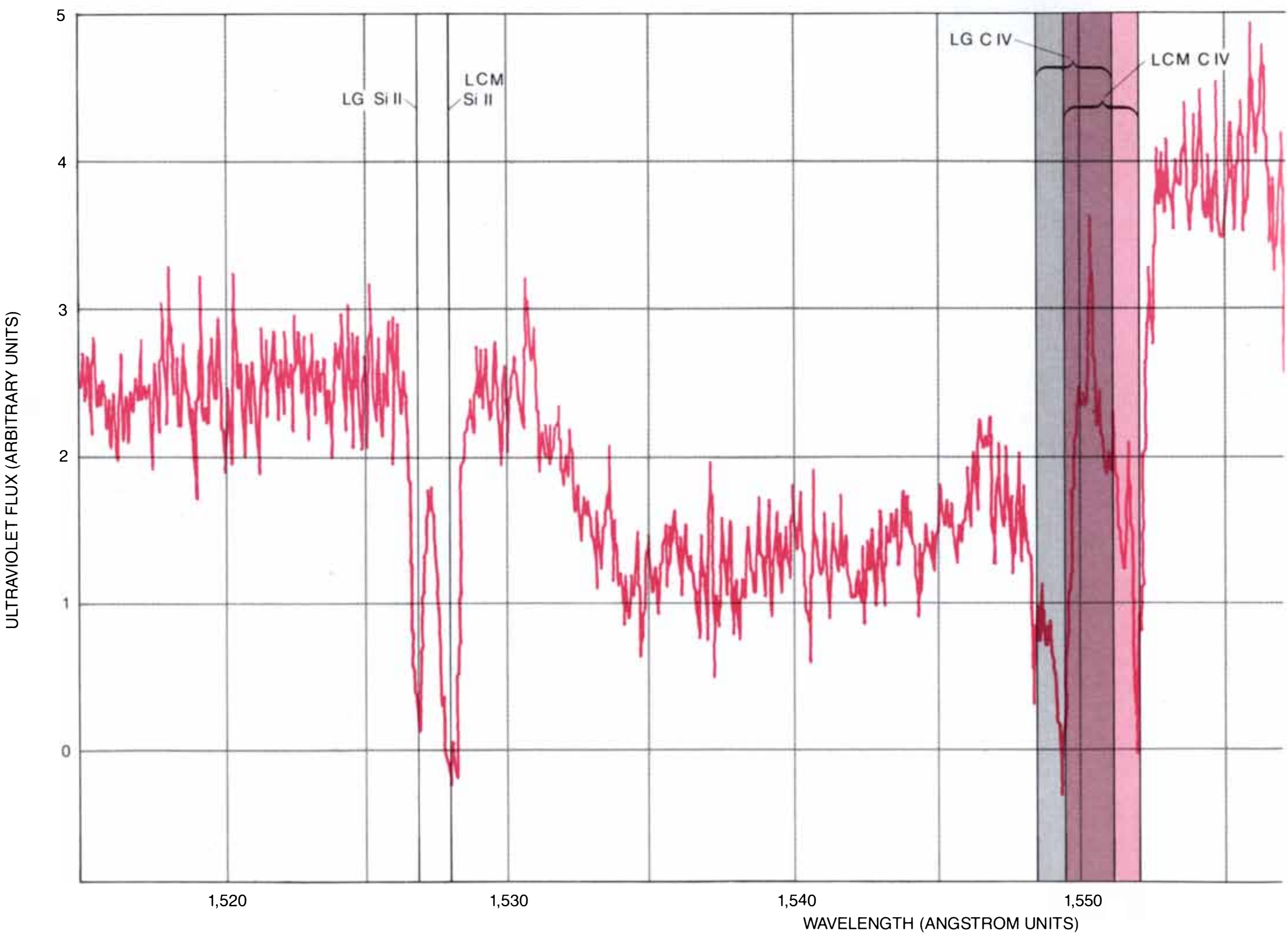
Gerd Weigelt of the University of Erlangen-Nuremberg applied this technique to images of R136a1 made with the European Southern Observatory's 3.6-meter telescope. At about the same time John Meaburn of the University of Manchester and his colleagues got somewhat different results with the 3.9-meter Anglo-Australian telescope in Australia. Presumably the difference arises from the way the two groups analyzed the images.

The speckle-interferometry results agree on the fact that R136a1 consists of a dominant unresolved object with several fainter objects close to it. Meaburn believes nothing else in the area is of

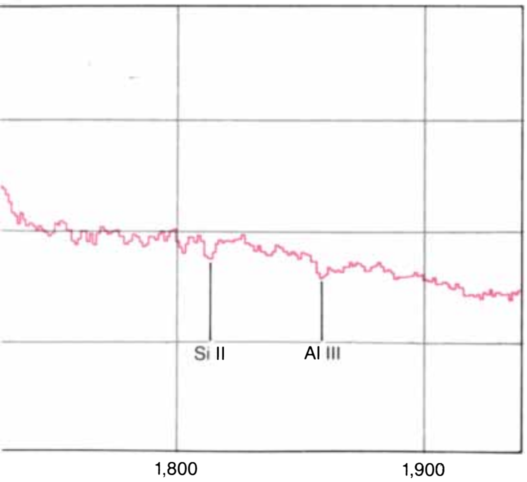
CLOSE VIEW of the inner region of the 30 Doradus nebula was obtained by John Wood with the four-meter telescope at the Cerro Tololo Observatory. R136 is the tight knot of stars at the center of the photograph. Although it is some 50 million times more luminous than the sun, it is dimmed by interstellar dust. Ultraviolet radiation from R136 ionizes much of the gas in the nebula. R136a and other nearby stars emit powerful stellar winds that have presumably pushed the gas into the arclike surrounding structure visible in the photograph. The central object is either the most massive star known or an extremely dense cluster of massive stars. The data obtained so far by observations from the ground and from the International Ultraviolet Explorer satellite do not rule out either of these two hypotheses.



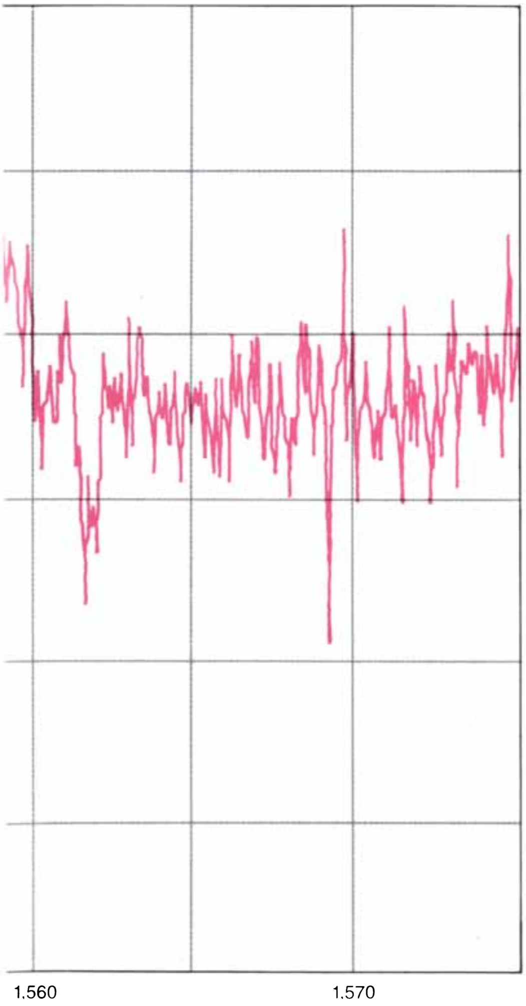
ULTRAVIOLET SPECTRUM of 136a obtained by the International Ultraviolet Explorer satellite reveals the unusual nature of the object. The intensity of the ultraviolet radiation is plotted against wavelength. The notable emission and absorption features associated with R136a are indicated by the symbols above the spectrum; absorption lines resulting from intervening stellar gas are marked below the spectrum. The symbols above the spectrum represent multiply ionized atoms: nitrogen, oxygen, iron, carbon and helium. The most prominent features are C IV, He II and N IV. (The Roman numerals are one higher than the number of electrons removed in ionization.) The spectra of these ions have a profile characteristic of P Cygni, a luminous star in the constellation Cygnus: absorption is shifted to short-



HIGHER RESOLUTION of the spectrum of R136a was obtained with the high-resolution spectrograph of the International Ultraviolet Explorer. The spectrum resolves wavelength to within .1 angstrom unit, a resolution about 70 times higher than that in the spectrum in the upper illustration. Here the focus is on a line in the spectrum of C IV: triply ionized carbon. Various narrow absorption lines that originate in the interstellar gas of our own galaxy (*LG*, for local galaxy) and of the Large Cloud of Magellan (*LCM*) are marked. The broad C IV feature at the right, originating in R136, is an absorption line at shorter wavelengths and an emission line at longer ones. The absorp-



er wavelengths and emission to longer ones. The high degree of ionization implies a high temperature for the object or objects generating the spectrum. The P Cygni profiles imply that R136a is losing mass in a powerful stellar wind. The spectrum centers on R136a.



tion at shorter wavelengths implies, on the basis of the Doppler effect, that matter is flowing outward from R136 at the high velocity of 3,600 kilometers per second. The spectrum is a profile of the characteristic P Cygni type.

comparable brightness, but Weigelt maintains that a second unresolved object is about .5 arc-second away (in the same position where Chu had placed R136a2) and is about a fifth as bright as R136a1 at visible wavelengths. In addition Charles E. Worley of the United States Naval Observatory has confirmed the position of R136a2 by visual observation. We therefore believe the reality of R136a2 has probably been established.

Weigelt's speckle-interferometry results suggest that the morphology of R136 is even more complex. He finds another component, which Chu calls a3, only .1 arc-second from a1—too close for Chu to resolve it. The speckle-interferometry results do not help much in determining relative brightness, but a3 seems to be comparable to a2, that is, it is about a fifth as bright as a1. All observers who have been studying R136 from the ground agree that there is also a background of still fainter stars within its three-arc-second image.

With Weigelt's results Chu was able to estimate the relative brightness of the three components. The light from the dominant component, a1, is equivalent to that from six stars like R122 or perhaps 20 of the more typical O3 stars. According to Weigelt, the angular diameter of a1 is no more than .08 arc-second, or about 24 light-days. By astronomical standards that is a small region. For example, the star nearest the sun is four light-years away. Nevertheless, the radius of the 24-light-day volume is 55 times the distance from the sun to Pluto. In that space there is certainly room for six stars or even 20.

Is R136a1 a single star six times more luminous than any now known or is it a cluster of stars that are individually like, say, R122? The question is of great interest because either hypothesis extends astronomy to objects of larger mass and greater luminosity than anything now recognized.

If R136a1 is a single star, it must have a mass of between 400 and 1,000 solar masses, making it at least twice as massive as any previously known star. The estimate of mass comes rather directly from luminosity. The outer layers of the atmosphere of a star must be held down by gravity against the pressure of the emerging radiation. The minimum mass for a star of a given luminosity was established by the British astronomer A. S. Eddington some 60 years ago. A star cannot violate the Eddington limit and be in even approximate mechanical equilibrium.

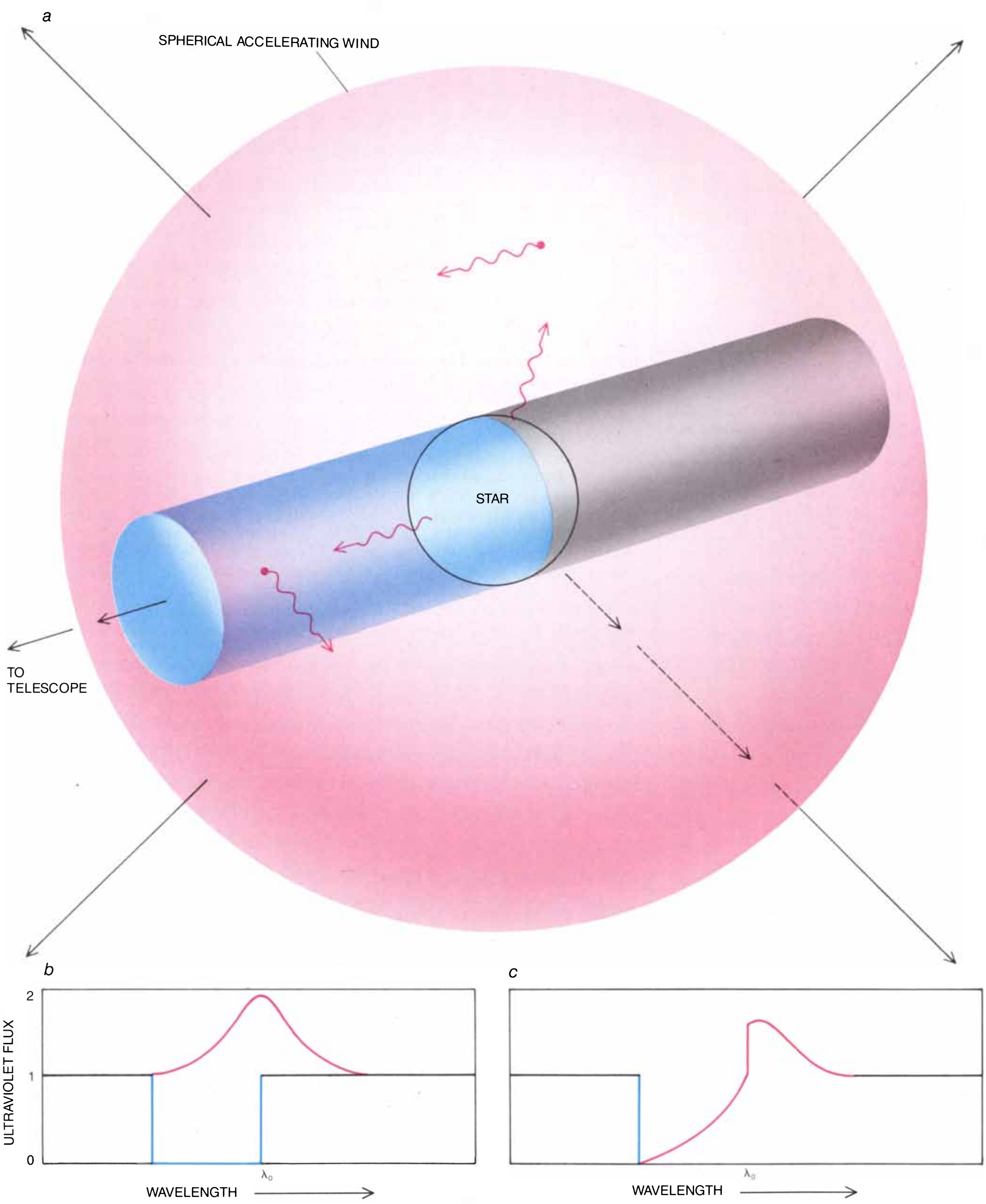
In 1970 Franz D. Kahn of the University of Manchester showed that very massive stars may not be able to form out of the clouds of cold gas and dust in which all stars originate. The problem is that as the cloud collapses, the density in the center increases much faster than the density in the outer regions. The core heats up, its dust grains are sublimated and its gas is ionized. It becomes a luminous, continuously growing object, radiating away approximately half of the gravitational energy converted into heat in its collapse. The result is a protostar surrounded by a dust-free zone surrounded in turn by a dusty shell.

In the dusty shell the light and the ultraviolet radiation from the protostar are converted into infrared radiation characteristic of the lower temperature at which dust can exist (about 1,000 degrees K.). The infrared radiation is absorbed in the outermost parts of the collapsing cloud, where the infalling gas and dust are only loosely bound by gravity. The outward-directed momentum of the absorbed radiation tends to reverse the infall of the dust grains. The now outflowing dust in turn drags along its gas. The growth of the star stops.

Hence there is an upper limit (only 40 solar masses in Kahn's original paper) on the central-star mass that will allow mass to continue falling inward. The Kahn model is sensitive, however, to the properties assumed for the dust, as one might suspect from the central role dust plays in the reversal of flow. Our University of Wisconsin colleague Mark G. Wolfire has shown that if Kahn had assumed a dust-sublimation temperature of 2,000 degrees K. instead of 3,600, the upper limit on the mass of the protostar could have been raised to about 1,000 solar masses, keeping all the other approximations of the model unchanged. The lower sublimation temperature is reasonable; Edward P. Ney and his colleagues at the University of Minnesota have found that in the expanding shell of gas around a nova dust condenses when the gas cools to a temperature of about 2,000 degrees. There are also other processes for the destruction of grains, such as collisions with helium atoms as the dust is pushed through the gas by radiation, that could make the collapse of a larger mass possible.

The original dust content per gram of the protostellar cloud is also important. In our own galaxy the ratio of dust to gas in interstellar matter tends to be large in the central plane and in regions of the galaxy closer to the center than the sun is. Toward the outer regions of the galaxy, however, the abundance of dust decreases.

A supermassive star would be unlikely to form in a region that has a large amount of dust, since radiation pressure pushing against the dust would tend to prevent the necessary collapse of a large dust cloud. The ratio of dust to gas near 30 Doradus has been found to be about



P CYGNI SPECTRAL PROFILE is formed in the expanding atmosphere (the stellar wind) surrounding a star that is losing mass (*a*). The shortward-displaced absorption (the decreased flux of radiation at wavelengths shorter than λ_0, the wavelength of the line if the atmosphere were not expanding, as shown in *b*) is caused by photons scattered out of the line of sight by ions in the blue region on the near side of the star. In a spherically symmetrical outflow every photon scattered out of the line of sight is matched by one scattered into it by ions in the red emission lobe. Photons scattered toward the observer from behind the star (*black region*) are not seen because they are blocked by the star. The wavelength at which scattered light appears in the spectrum depends on the angle between the outward direction of the stellar wind at the point of scattering and the position of the observer. Some of the ions in the emission lobe have a component of velocity away from the observer, so that the light they scatter is shifted toward the longer wavelengths; some have a component of velocity toward the observer and are shifted toward the shorter wavelengths. The P Cygni profile (*c*) results from a combination of three effects: scattering out of the line of sight from the blue region, scattering into the line of sight from the red region and the absence of photons from behind the star. The P Cygni profile of C IV of R136a has a large shortward-displaced absorption, indicating wind of very high velocity.

a third of the ratio in the neighborhood of the sun. It is therefore not unreasonable to suppose a star of 400 or more solar masses could have formed in the 30 Doradus region.

The structure of a star is determined by several balancing processes, such as an equilibrium between the inward force of gravity and the outward force provided by pressure. In a very massive star the rate at which nuclear energy is generated is quite sensitive to the temperature at the center. Slight perturbations have a tendency to grow. In addition the pressure at the center is primarily the pressure not of gas but of radiation. A star supported entirely by radiation can easily be dispersed; in fact, with radiation and gravity in balance it can disperse without any further input of energy. It is only the small fraction of the pressure arising from the thermal motion of the gas that stabilizes the massive star. Calculations have suggested that stars more massive than about 60 suns might be dispersed by internally driven pulsations.

Exceptionally massive stars can be described mathematically in a fairly simple way. In 1962 Fred Hoyle of the University of Cambridge and William A. Fowler of the California Institute of Technology derived the relevant equations. (They were interested in objects that might have a mass of up to a million solar masses, which they had proposed to explain the recently discovered quasi-stellar objects, or quasars.) The interior of a massive star is well mixed by rising and sinking convection currents. In a star smaller than 60 solar masses the interior mixing can lead to a change in chemical composition from the interior layers, where energy is transported by convection, to the outer layers, where energy is transported by radiation.

André Maeder of the Geneva Observatory has shown that stars larger than 60 solar masses should be almost homogeneous because their outer layers are removed rapidly by stellar winds. In such a star evolution proceeds straightforwardly. At the beginning the star's nuclear fuel is hydrogen, which in the course of thermonuclear burning is converted into helium. The composition of the star therefore gradually changes from being about 70 percent hydrogen to being primarily helium. In the process a homogeneous star decreases in radius and increases in temperature. Stars of lower mass, in contrast, evolve from relatively compact hydrogen stars to cooler and larger giant stars. The surface temperature of a massive star increases from about 60,000 degrees K. to about 90,000 degrees as it ages and becomes a helium star.

The absence of the usually strong Si IV lines in the ultraviolet spectrum of R136a1 is evidence for a gas temperature consistent with the hypothesis that the object is a single star. One might expect, however, that a very massive star would be variable in its output of light (and thus unstable) because of the marginal stability of its internal structure. Richard Stothers of the Institute for Space Studies of NASA has investigated several processes that can in principle bring about stability in such a star. They include high-speed rotation and the entanglement of lines of force in interior magnetic fields. Assuming processes of this kind to be at work in R136a1, the ultraviolet spectroscopic data are consistent with the single-star hypothesis. The data do not, however, rule out other possibilities.

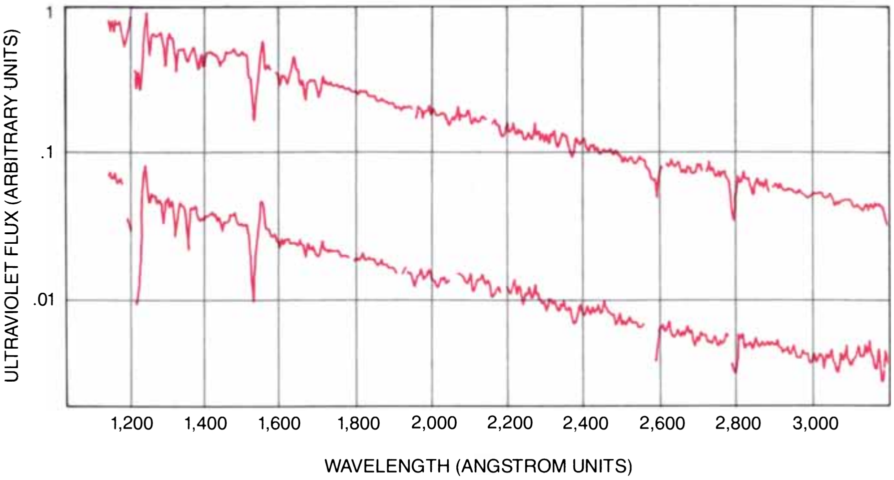
Anthony F. J. Moffat of the University of Montreal and Wilhelm Seggewiss of the Hoher List Observatory in West Germany believe R136a1 is a tight cluster of stars, each of a mass comparable to the mass of stars already known. Probably the most massive individual stars known today are R122 and Eta Carina, an unstable object surrounded by a thick cloud of dust. Stellar theory predicts that each of these stars has about 200 times the mass of the sun. As we have pointed out, a cluster supplying the energy of R136a1 would require at least six R122's within a remarkably small volume.

That hypothesis raises two questions. Would not the first star, as soon as it had formed, prevent the formation of others by heating the collapsing gas? How can so much gas get into such a small volume without having collapsed earlier? To date these questions have no answers, but the same thing can be said of many other puzzles associated with the formation of stars.

Chu's visible-light images indicate that the bright component, R136a1, is near the center of a compact cluster of stars. This observation raises the interesting possibility that R136a1 has formed as the consequence of the dynamical evolution of a star cluster. In regions with a high density of stars gravitational encounters between stars can result in the evaporation, or loss, of low-mass stars, causing the cluster to shrink. Eventually supermassive stars or small groups of supermassive stars might be created by the actual physical coalescence of stars in such encounters. The theory of the dynamical evolution of clusters has received considerable attention in recent years because of its possible application to the theory of the origin of the peculiar luminous objects found in the nucleus of galaxies. It is exciting to think R136a1 might represent an example of stellar coalescence in a galaxy next door to our own.

So far we have discussed R136a1 with the implied assumption that the object burns hydrogen like an ordinary star. Perhaps we should consider certain more exotic possibilities. For example, if black holes exist, the interstellar matter spiraling into one of them would form an accretion disk. The process would generate large quantities of radiation, giving rise to the kind of brightness observed in R136a1. Another exotic possibility is the spinar, a rotating mass of ionized and magnetized gas that has been put forward as an explanation of the large quantities of radiation emitted by quasars.

Our strongest reason for not proposing such hypotheses for R136a1 is that the observational data now available can be explained by a fairly straight-



ULTRAVIOLET FLUX of R136a is charted (*upper curve*) as it is measured from above the earth's atmosphere. The range of wavelengths was observed by the International Ultraviolet Explorer spectrometer, and the reading was corrected for attenuation by dust in our galaxy and the Large Cloud of Magellan. For comparison the flux distribution of R122, a type O3 star at the same distance as R136a, is also shown (*lower curve*). The spectral lines are similar except for the helium-ion line (He II) at 1,640 angstroms. R136a is, however, about 13 times brighter than R122, which is the most luminous "normal" star in the Large Cloud of Magellan.

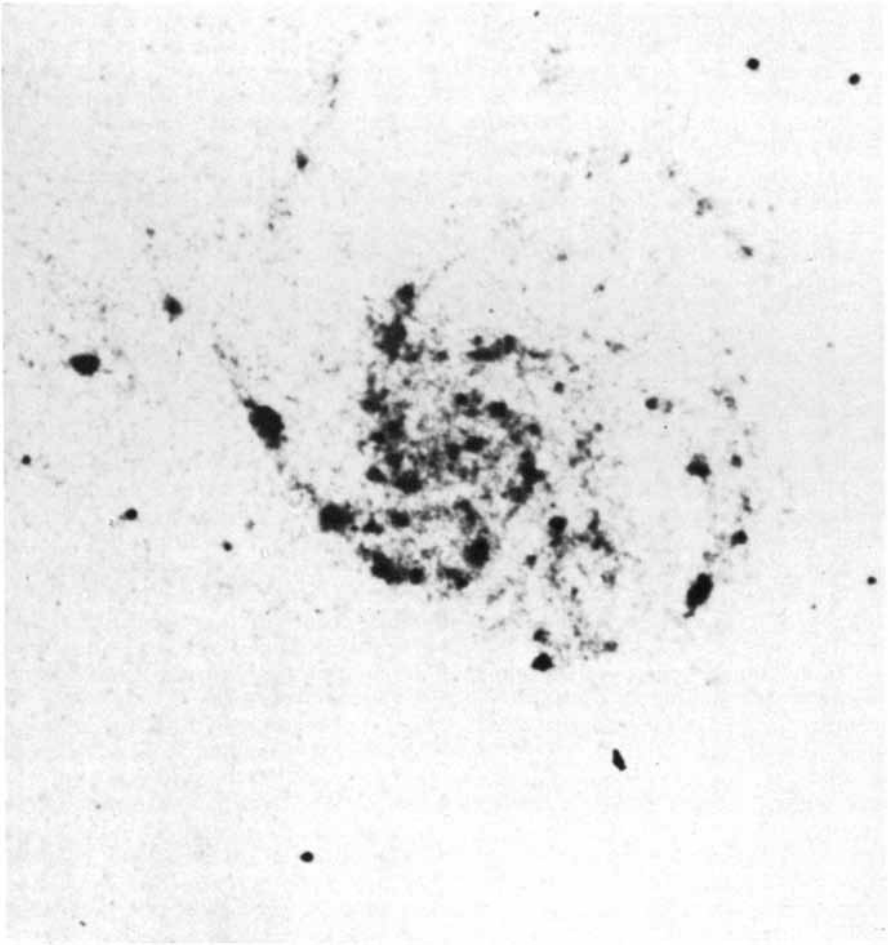
forward extrapolation from massive hydrogen-burning stars on the main sequence of stellar evolution. The distribution of energy output across the spectrum of R136a1 does not perfectly match current theoretical models of other stars, but the differences are no worse than those encountered with the luminous type-O stars and are probably due to the effects of massive stellar winds. The velocity of the R136a1 wind is high, but it is not out of line with that of the wind from ordinary O3 stars. The rate at which R136a1 loses mass is huge (one earth mass per week), but even that is in line with an extrapolation from type-O and Wolf-Rayet stars. Finally, one would expect black holes and other exotic objects to give rise to a large flux of X rays. Knox S. Long, Jr., of Johns Hopkins University, working with the Einstein observatory satellite, has observed X rays from objects in the center of 30 Doradus. Although the X-ray output from R136a and its environment is large (equivalent to about 100 times the total energy output of the sun), it too is not significantly out of line with an extrapolation from O stars.

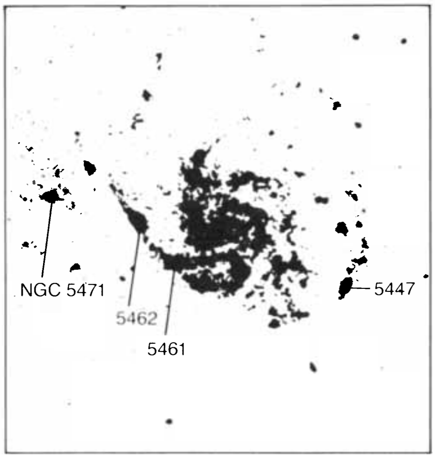
Whatever the true nature of R136a is, it is almost certainly not a unique object. In our galaxy there is a giant ionized region (NGC 3603) with a luminosity about a seventh that of 30 Doradus. Walborn first noted the strong similarity between the core object of NGC 3603 and R136a. Perhaps the former is a scaled-down version of the latter.

One must go to galaxies beyond the Large Cloud of Magellan to find objects that rival the 30 Doradus nebula. Frank Israel of the Leiden Observatory and others have studied the largest nebulas in galaxies out to about 25 million light-years. Within that volume are a few spiral galaxies with notably active regions where hot, massive stars are forming and the associated ionized nebulas can be perceived. Israel lists seven other galaxies, all at least 10 million light-years away, with nebulas more luminous than 30 Doradus. The most spectacular is the galaxy M101, which harbors four nebulas that are more than five times as luminous as 30 Doradus. One of them is NGC 5461, the most luminous nebula known; it is the equivalent of 11 nebulas like 30 Doradus. Little is known about the sources that excite these huge gas complexes; they are about 10 times farther away than 30 Doradus, meaning that the earth receives only about a hundredth as much radiation from their individual stars as it does from the stars of 30 Doradus.

Philip L. Massey and John B. Hutchings of the Dominion Astrophysical Observatory in British Columbia have worked with the International Ultraviolet Explorer satellite to examine the luminous stars in the large ionized regions of the spiral galaxy M33 in the local group of galaxies. Six of the seven objects they studied have ultraviolet spectra quite similar to R136a's. Unfortunately M33 is 10 times farther away than R136, so that it is impossible to tell whether the spectra represent superluminous stars or highly compact groups of stars with normal characteristics.

The Space Telescope that is to be launched by NASA in 1986 should help to clarify the uncertainties about the physical nature of R136a. With an angular resolution 10 times better than can be achieved from the ground and with ultraviolet-spectroscopic capabilities, the Space Telescope will determine the relative ultraviolet brightness of the various components of R136a. The high stability of the telescope for purposes of measurement will make possible a careful search for variability in R136a. The detection of such variability would limit the range of possible explanations for this bizarre object. The telescope will also be employed to study the central objects of other supergiant nebulas beyond 30 Doradus. It may turn out that objects as unusual as R136a lie at the core of most giant ionized regions in other galaxies.





ULTRAVIOLET IMAGE of a distant galaxy, the supergiant system M101, was made from a high-altitude rocket in a project carried out by T. P. Stecher and R. C. Bohlin of the Goddard Space Flight Center of the National Aeronautics and Space Administration. The image emphasizes the hot stars that ionize the gas around them. In the accompanying map the four leaders point to nebulas that are all at least five times more luminous than 30 Doradus. The nature of the sources that excite the radiation of these nebulas is not clear. Moreover, it is difficult to determine because of their distance (about 20 million light-years). The spectra obtained by the International Ultraviolet Explorer nonetheless suggest that the four sources resemble R136a.

SCIENCE AND THE CITIZEN

Son of MX

"Politics makes strange bedfellows," goes the old saw. No issue in contemporary American politics has brought together an odder assortment of backers than the plan to build a new land-based strategic missile, known variously as Midgetman, Singlet or SICBM (for small intercontinental ballistic missile). Politicians of both parties, liberals and conservatives, proponents and opponents of greater emphasis on arms control, have all found their own reasons to approve of the proposed single-warhead missile. The widespread support for the new weapon contrasts sharply with the repeated clashes in Congress and elsewhere over the MX missile, the other major part of the Reagan Administration's program for the "strategic modernization" of the U.S. land-based ballistic-missile force.

In retrospect it seems clear that the seed from which Midgetman sprang was planted more than a decade ago, when the U.S. decided to reequip its existing force of Minuteman ICBM's with highly accurate MIRV's (multiple independently targetable reentry vehicles). As opponents of the decision predicted at the time, the U.S.S.R. promptly followed suit by pushing ahead with its own MIRV program, which in due course came to be perceived as threatening to the U.S. land-based missile force.

The initial response of the U.S. to the "window of vulnerability" said to be opened by the new generation of Russian MIRV's was the MX: the giant, 10-warhead ICBM that has been a focus of controversy throughout most of the Carter and Reagan administrations. The MX was originally intended to be deployed in a mobile "basing mode" in order to increase its chances of surviving a preemptive "counterforce" strike by the U.S.S.R. and thereby lessen the confidence of Russian military planners in the efficacy of such an attack. The inability of both administrations to find a politically acceptable basing mode for the MX finally led President Reagan to appoint a special Commission on Strategic Forces, headed by Brent Scowcroft, a retired Air Force general.

Confronted with the failure to close the window of vulnerability by their initially preferred method (by deploying the MX in a clearly "survivable" basing mode), the members of the Scowcroft Commission in effect disposed of this much-heralded threat by simply acknowledging in their 1983 report what critics of the MX program had been stressing for years: the inherent implausibility of a preemptive strike by the U.S.S.R. on the land-based missile force of the U.S. in the face of the unquestioned deterrent value of the rest of the U.S. strategic arsenal, particularly submarine-launched ballistic missiles. Accordingly the commission members were able to recommend as part of the current program the immediate deployment of 100 MX missiles in existing Minuteman silos, where they would of course be no less vulnerable than the missiles they were designed to replace.

In a seeming contradiction, however, the report of the Scowcroft Commission went on to advocate a continued pursuit of the long-term goal of "ICBM survivability," calling simultaneously for the development of a small, single-warhead ICBM, to be deployed in the 1990's, and for further investigation of the entire question of ICBM basing, with particular emphasis on mobile basing in "hardened vehicles." The Reagan Administration and its backers have enthusiastically embraced the commission's recommendations, insisting that they constitute a compromise package whose parts cannot be treated properly in isolation.

Many congressional opponents of the MX program have chosen to ignore the foregoing stricture and have come out selectively in favor of Midgetman, apparently in the hope that it could prove to be a less dangerous substitute for the MX. According to their reasoning, the proposed deployment of MX missiles in Minuteman silos would have a destabilizing effect on the strategic balance in two ways: by appearing to enhance the threat of a U.S. first strike against the U.S.S.R. and by presenting a particularly attractive target for a potential Russian first strike against the U.S. A force of 1,000 Midgetman missiles dispersed in mobile, hardened vehicles, it is argued, would be less destabilizing on both counts. Furthermore, if the U.S.S.R. could be induced to emulate the U.S. move, replacing its current force of large MIRVed ICBM's with a comparable force of dispersed mobile SICBM's, the negative trend in the security of both sides that has resulted from the deployment decisions of the past decade could be reversed.

There are a number of problems with this position. For one thing the U.S. Air Force has made clear that it views Midgetman basically as a disaggregated MX; in keeping with the recommendation of the Scowcroft Commission it has set out to develop a single-warhead missile with "sufficient accuracy and yield to put Soviet hardened military targets at risk," relying in some cases on components adapted directly from the MX program.

The deployment of Midgetman in a dispersed, mobile basing mode can also be expected to face many of the same obstacles that led ultimately to the rejection of the original MX system, and it would have the added disadvantage of costing considerably more per warhead deployed. (Current estimates run as high as $100 billion for deploying and operating the proposed 1,000-missile system.) In the absence of a strategic-arms agreement limiting the total number of land-based Russian missiles, the U.S.S.R. could always add enough warheads to its attacking force to be confident of destroying the entire U.S. ICBM force, regardless of its basing mode. Under the circumstances the U.S.S.R. is unlikely to be persuaded to give up its existing force of large, MIRVed missiles in favor of small, single-warhead missiles, particularly on the strength of strategic theories invented in the U.S. Given the ultimate dependence of any scheme for ICBM survivability on future arms-control agreements, Midgetman appears to be no more promising in this respect than the MX. The fatal flaw of both programs is that they are aimed toward a purely technological solution of what has all along been essentially a political problem.

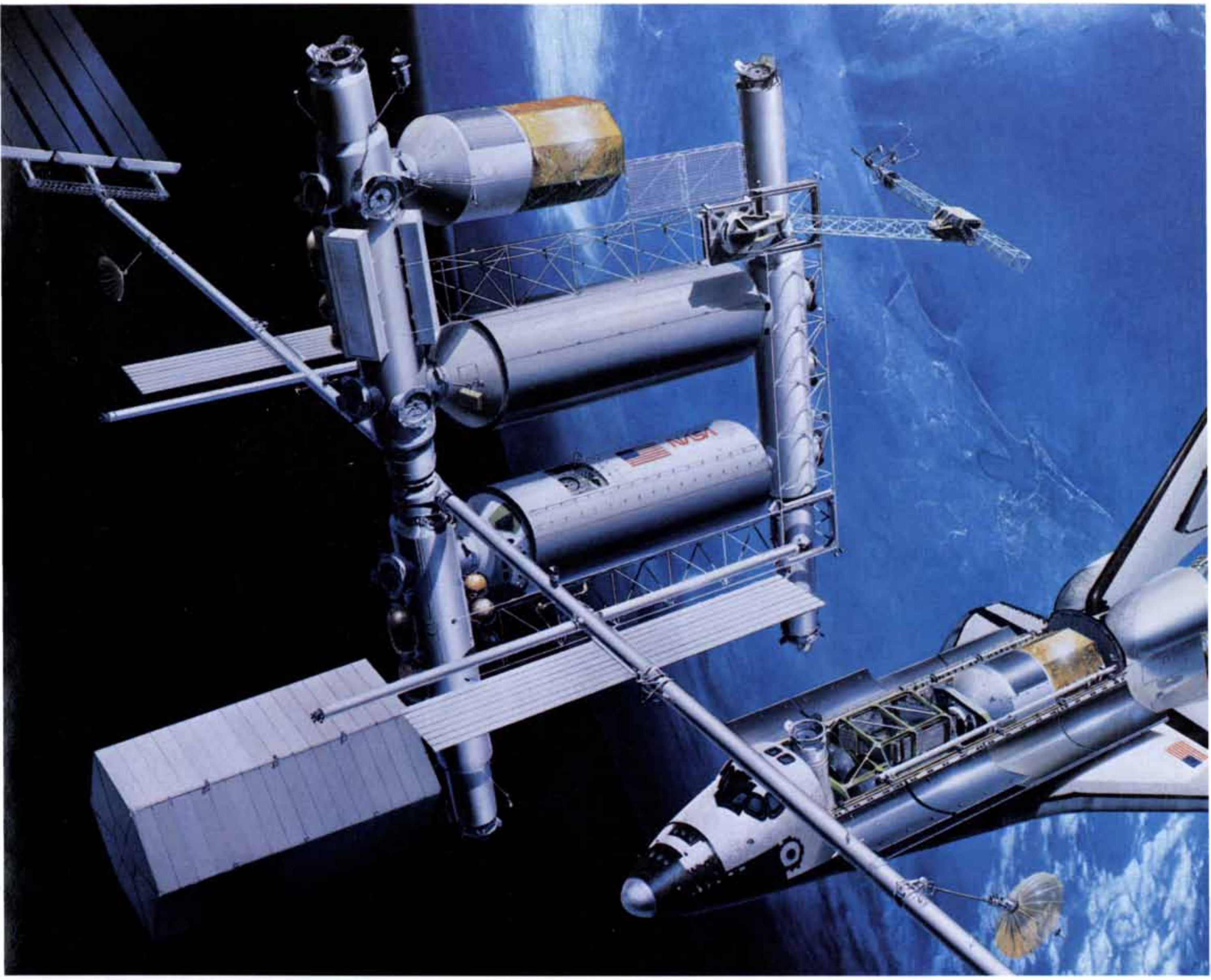
Fast Track

The Superconducting Super Collider (SSC), the proposed high-energy particle accelerator that would make it possible to explore the foundations of the unified theory of weak and electromagnetic interactions, will probably pass one of its first major political milestones this month. A recommendation is expected from the Department of Energy that Congress provide funding for the first of two to three years of formal research and development for the machine. The recommendation, if it comes, will be based on a conceptual design study completed in April under the direction of Maury Tigner of Cornell University. Last month physicists concluded a round of meetings held throughout the past year with a major conference at Snowmass, Colo., aimed at matching the design as closely as possible to future experimental programs.

The conceptual study, which is known as the reference design, presents three major options for the machine. All three are based on the general requirement that two beams of protons be accelerated in opposite directions around a ring and then made to collide. The protons in each beam are to reach an energy as high as 20 TeV (trillion electron volts), and the flux of each beam is to enable the beams to cross each other at a rate of 10^{33} protons per square centimeter per second. At that rate collisions between


CHIVAS BROTHERS INTRODUCES ROYAL PAISLEY SCOTCH.
Scotch drinkers have come to know that when you see the words "Chivas Brothers" on the outside of a bottle, you're sure to find exceptional Scotch on the inside.
And Royal Paisley is no exception.
Made from Scotland's most prized whiskies, Royal Paisley is remarkably rich and incomparably mellow. Yet it's priced at about $11 a bottle.*
Best of all, Royal Paisley offers you something only a few of the world's finest Scotches possess.
The name Chivas Brothers.
ROYAL PAISLEY
ROYAL PAISLEY
FROM CHIVAS BROTHERS
BLENDED
SCOTCH WHISKY
100% SCOTCH WHISKIES
BLENDED & BOTTLED IN SCOTLAND
BLENDED AND BOTTLED BY
CHIVAS BROTHERS, LTD.
CHIVAS BROTHERS
The taste is guaranteed right on the bottle.
*750 ML size. Price varies slightly by locale.
© 1984 GENERAL WINE & SPIRITS CO., N.Y.C.

Out in Space.



orbiting space stations. We're helping NASA design modules that go together like an erector set and house a crew of 4 to 12 for at least three months. It's a giant step from the Lunar Rover to space stations. And Boeing is on top of the program — every step of the way.

d Trust

BOEING


ATTENTIVELY HYATT.
At Hyatt, we have a special welcome waiting every time you arrive. Because people who expect the best from themselves expect a lot from their hotel.
We usher you through check-in without a moment wasted. We'll clean and press your suit and have it back before evening. From our famous room service, we'll send up a meal as thoughtfully prepared as any in town. And with our express check-out service, you never have to wait. Just drop off your key and you're on your way.
When you're our guest, special treatment is something you learn to take for granted. An attentive touch of Hyatt. Don't you WISH YOU WERE HERE℠
ATLANTA
Discover magnificent Peachtree Center, 15 minutes from Atlanta International Airport.
CHICAGO
In Illinois Center overlooking Michigan Avenue and Lake Michigan.
HOUSTON
In the very heart of exciting Houston, walk to business through climate-controlled tunnel system.
NEW YORK
Grand Hyatt on fashionable Park Avenue at Grand Central.
PRINCETON
Two miles from Princeton University, midway between Philadelphia and New York City.
HYATT HOTELS
For reservations, call your travel planner or 800 228 9000. © 1984 Hyatt Hotels Corp.

counterrotating protons would be frequent, and each collision could make up to 4 TeV available for the materialization of new particles at the site of the collision. Particles could thereby be created with a mass 40 times greater than those seen in any present accelerator.

The three options in the reference design differ primarily in the strength of the superconducting magnets needed to bend the beams of protons into circular paths. For a given energy the smaller the ring is, the more powerful the magnets must be. The smallest ring, suggested by the Lawrence Berkeley Laboratory of the University of California and the Brookhaven National Laboratory, would be 90 kilometers in circumference and would call for magnets with a field strength of 6.5 teslas. (The strength of the electromagnet commonly employed for picking up scrap iron is .3 tesla.) The Fermi National Accelerator Laboratory (Fermilab) envisions a ring 113 kilometers around and magnets with a strength of five teslas. Perhaps not surprisingly, a consortium of four Texas universities led the work for the design of the largest version of the ring: it would be 164 kilometers around and the magnets would have a strength of up to three teslas. The trade-off between the magnetic field strength and the real estate needed for the machine makes the estimated costs of the three options roughly equal: from about $2.5 to $3 billion in current dollars. An additional $200 million will be needed for research and development. Construction could take about six years at a site to be chosen within the next three years. The SSC could be ready for experimental work by 1993 or 1994.

The basic design of the SSC is dictated by theoretical predictions that suggest important experimental tests at energies of a few TeV. According to what is now regarded as the standard theory, electromagnetic interactions and weak interactions such as the decay of the neutron into a proton within the atomic nucleus are superficial aspects of the same underlying phenomenon. Many predictions of the standard theory have now been borne out, notably the prediction that any weak interaction is mediated by one of three particles called intermediate vector bosons.

The mass of a particle that mediates a force sets the limit on the range of the force. In the case of the weak force the three vector bosons are extremely massive, about 80 to 90 times the mass of the proton, and the force has a range of only 10^{-16} centimeter. In contrast, the mass of the photon, which mediates the electromagnetic force, is zero, and so the range of the electromagnetic force is infinite. One of the major questions that must be confronted by any theory seeking to unify the two forces is how to account for the difference between the mass of the vector bosons and the mass of the photon.

In the standard theory the differences in mass become manifest only at relatively low energy. When the available energy is equal to or greater than the mass of the intermediate vector bosons (as it was, for example, during the first fraction of a second after the start of the big bang) vector bosons materialize freely and then disappear again into the energetic background. At such high energies the distinction between the vector bosons and the photon is lost, and the electroweak interaction is said to be symmetrical. The symmetry can be likened to that of a collection of water molecules above the critical temperature and pressure of water. When the pressure is greater than 218 atmospheres and the temperature is above 374 degrees Celsius, there is no distinction between the solid, liquid and gas phases of water. Near that critical point, however, the interactions among the water molecules become highly complex and the symmetry of the system is broken as groups of molecules assume one of the three distinct phases. Similarly, when the available energy falls below about 90 times the mass of the proton, the symmetry of the electroweak interaction is broken into the distinct forms it now assumes everywhere in the universe except in the high-energy laboratories of physicists.

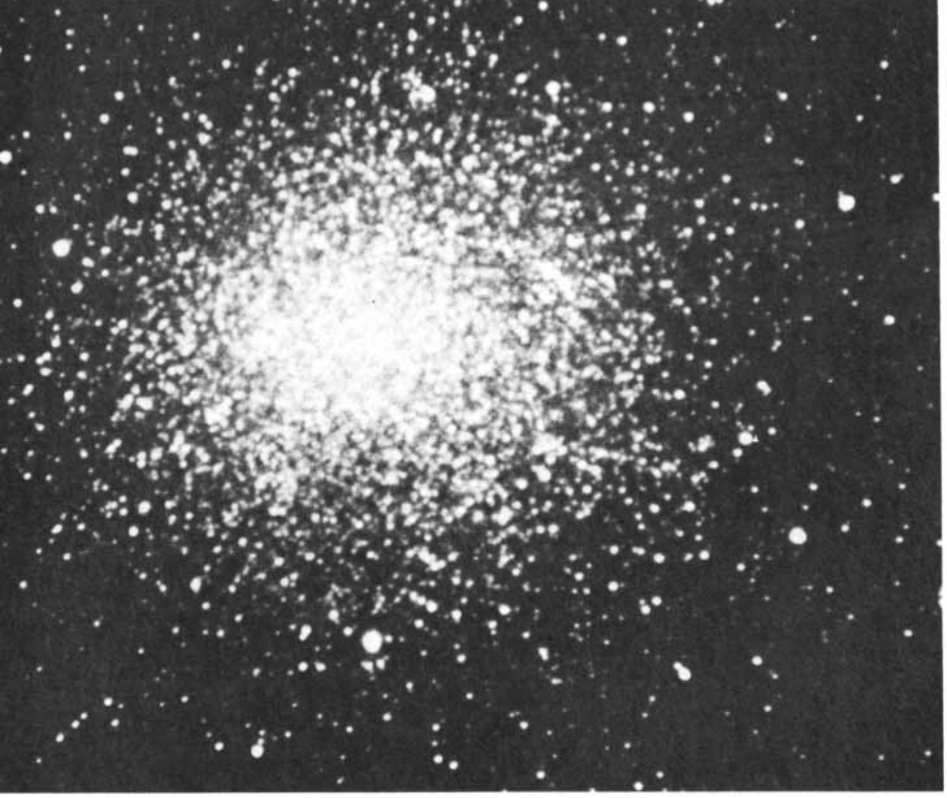
There is a general argument, independent of the standard electroweak theory, that suggests the mechanism for symmetry breaking should become manifest at energies of less than a few TeV. Because the intermediate vector bosons acquire their mass when the electroweak symmetry is broken, a correct account of symmetry breaking should also lead to a partial understanding of the origin of mass. Hence there is a strong theoretical incentive to explore the energy spectrum up to a few TeV by experiment. Although the reference design for the SSC is explicit in making this energy range accessible, particle physicists have continued to hold workshops on various special topics in Texas, at the University of Chicago, at several of the national accelerator laboratories and, most recently, at the Lawrence Berkeley Laboratory in June. The general aim of the workshops has been to seek the most effective and economical ways of exploiting the SSC for detecting phenomena that would test competing theories.

According to the standard theory, symmetry breaking is caused by the particle called the Higgs boson, after Peter W. Higgs of the University of Edinburgh. Up to now the Higgs boson has not been observed, and the theory gives little guidance about where in the energy spectrum such a particle might be found. At the Berkeley workshop theorists held three seminars to discuss various theoretical versions of the Higgs mechanism and their implications for the design of the SSC. In one seminar it was assumed the mass of the Higgs boson is greater than 1 Tev, which could be found only by an accelerator like the SSC. The second seminar considered the possibilities that there are many Higgs bosons and that each of them is a composite made up of particles so far unknown. In the third seminar it was assumed the mass of the Higgs boson is relatively small: roughly equal to the mass of one or two intermediate vector bosons. In that case the Higgs boson could be detected by the accelerators that will be exploring the energy spectrum up to several tenths of 1 TeV in the next two to five years.

There is a growing suspicion among theorists, however, that the Higgs boson as it was initially conceived does not exist. Accordingly several working groups at Berkeley discussed the kinds of experiment that could test alternative theories of symmetry breaking. One group focused on the detection of supersymmetrical particles: massive counterparts of every known elementary particle. If the supersymmetrical particles exist, they too could play a role in symmetry breaking; their discovery would open up a rich and unexplored vein of physics at extremely high energy.

Another working group studied the possibility of detecting the signals that would be emitted if the particles that now seem to be elementary turn out to be composite. There is strong historical precedent for such a development. Where the atom was once thought to be indivisible, it was found to be a composite made up of electrons, neutrons and protons; by the same token the neutron and the proton were found to be made up of quarks. The quarks, the electron and several other particles like the electron that are generically classified as leptons have so far shown no internal structure. If internal structure does exist, the SSC may be able to detect it.

There are many other exotic particles conjectured by theory whose existence could be confirmed or ruled out by the SSC. Given the constraints that will inevitably be imposed by budget, physicists must decide in the next few years how flexible a machine to build and what kind of detectors to install. For example, should the design of the SSC allow one of the proton beams to be extracted from the ring and collided with a fixed target? If fixed-target physics is done, should the neutrinos or muons generated in the collisions of protons with the fixed target be employed as a secondary beam directed against a second fixed target? Should detectors seek to cover the entire space around the site of a collision or should they focus only on certain directions with respect to the beams? Such decisions will indirectly



determine the phenomena most likely to be found; consequently they have great bearing on the course of elementary-particle physics for the 1990's.

Cholesterol Catchers

Epidemiological studies associate a high level of cholesterol in the blood (as well as cigarette smoking and hypertension) with both atherosclerosis and the incidence of heart attacks and strokes; the data also associate high blood cholesterol levels with ingestion of diets high in cholesterol and saturated fats. The question remains: Can the reduction of cholesterol and fat in the diet lessen atherosclerosis? Early this year the National Heart, Lung, and Blood Institute reported the results of a 10-year prospective study indicating that lowering the blood cholesterol level does reduce the risk of heart attack. The test involved only middle-aged men with high initial cholesterol levels, however, and the lowering was attained by drug treatment, not diet. Extrapolation of the results to imply that dietary modification would be effective for the general population has therefore been attacked by many investigators of atherosclerosis, and so the cholesterol-and-diet controversy continues.

Some light is shed on the controversy by the work of Joseph L. Goldstein and Michael S. Brown of the University of Texas Health Science Center at Dallas. They have described in detail the mechanisms whereby the level of circulating cholesterol rises, promoting the formation of the atherosclerotic plaques that can eventually close off an artery.

Cholesterol circulates in the bloodstream packaged in globular particles of low-density lipoprotein (LDL). An atherosclerotic plaque forms when LDL invades the wall of an artery and deposits its cholesterol, and so the most important risk factor for atherosclerosis and heart attack is the blood LDL level. LDL is ordinarily removed from the circulation when it is taken up by cells, which extract from it the cholesterol they need. All cells need some cholesterol to build their plasma membrane; some specialized cells need more of it to make steroid hormones, but most of it is taken up by liver cells that convert it into bile acids. In 1973 Goldstein and Brown discovered LDL receptors: proteins embedded in the plasma membrane that bind circulating LDL and initiate the process of receptor-mediated endocytosis, whereby the LDL is taken into the cell and broken down to yield its cholesterol.

A cell's supply of cholesterol is regulated by a feedback system whose most important element is the synthesis of LDL-receptor molecules. When a cell has all the cholesterol it needs, expression of the gene encoding the receptor is reduced; the cell makes fewer receptors and therefore extracts less LDL from the circulation. A general reduction in the number of receptors (in the liver in particular) leads to a rise in the blood level of LDL—and to atherosclerosis.

The link between LDL receptors and atherosclerosis has been demonstrated most clearly by studies of familial hypercholesterolemia (FH), a genetic disease caused by mutation of the LDL-receptor gene. The cells of heterozygotes, who inherit one mutant gene and one normal gene, synthesize half of the normal number of receptors; the cells of homozygotes, who have two mutant genes, synthesize no normal receptors. The plasma LDL level is elevated at least twofold in heterozygotes, whose risk of a heart attack before the age of 60 is 25 times as great as that of the general population; homozygotes have LDL levels more than six times the normal level and usually have fatal heart attacks before the age of 20.

Goldstein and Brown have treated FH heterozygotes successfully with drugs that create a cholesterol deficiency within cells and thus enhance the expression of the LDL-receptor gene, so that cells synthesize and display more receptors. One drug interrupts the recirculation of bile acids to the liver, thereby increasing the liver's demand for cholesterol. The cells respond by making more receptors, but they also step up their de novo synthesis of cholesterol. To block this second feedback response and so force the cells to rely on circulating LDL for the cholesterol they need, a second drug is administered that inhibits an enzyme in the cholesterol-synthesis pathway. The two-drug treatment forces the FH heterozygote's single receptor gene to make more receptors and reduces the circulating-LDL level by about 50 percent.

How relevant are these findings, in patients with a severe genetic deficiency, to the development of atherosclerosis in people with a normal complement of LDL-receptor genes? Goldstein and Brown think much of the atherosclerosis in the general population is attributable to high levels of circulating LDL resulting from inadequate production of receptors. At birth human beings have LDL levels comparable to those of other animal species, but the level rises with age, presumably because the number of receptors decreases. There are probably several reasons for the decrease, but one reason is probably the ingestion of a diet high in cholesterol and saturated fats. The accumulation of cholesterol in cells acts directly to turn off receptor synthesis; epidemiological data show clearly that saturated fats raise the cholesterol level, but the mechanism remains to be established. Goldstein and Brown think moderate restriction of dietary cholesterol and fats is indicated for most people, severe restriction for people with a strong family history of heart attacks and strokes. The kind of drug treatment developed for FH heterozygotes may possibly prove to be effective and safe for the general population, perhaps eventually making dietary restriction less essential.

Earthscan

Geophysicists have long gleaned information on the earth's inaccessible interior from the behavior of seismic waves triggered by earthquakes. Seismic studies have revealed much of the earth's average radial structure: the planet has a solid inner core, a molten outer core and a solid yet ductile mantle stirred by convective currents that drive the motion of the continental and oceanic plates. This one-dimensional picture, however, leaves unanswered questions of great interest. For example, the nature and the scale of the convective flow in the mantle are not known, nor is the fate of slabs of the earth's crust that are subducted back into the mantle at oceanic trenches. A new analytic technique, made feasible by the vast expansion of seismic data in recent years and the availability of large computers to analyze them, now promises to supply some of the detail needed to sketch a truly three-dimensional picture of the earth. The technique is called seismic tomography, and as its name suggests it is analogous to computer-aided tomography, or CAT scanning.

A ray is affected by the medium it passes through and thus contains information on the structure of the medium. The X rays employed in CAT scanning, for instance, are partially absorbed as they pass through the body; the denser the tissue, the greater the absorption. Seismic waves are not significantly attenuated by passage through the earth, but they do travel faster through cold, dense material and slower through hot, less dense material. The arrival time of a seismic wave at a measuring instrument is therefore a measure of the "velocity structure" of the earth along the wave's path. Single rays yield only average values, but by combining measurements from many rays with crisscrossing paths it is possible to locate inhomogeneities, which in the case of seismic tomography are regions of anomalously fast or slow wave propagation.

From the velocity anomalies investigators can derive, with the help of laboratory data on the materials thought to make up the earth's mantle, the magnitude of associated temperature and density anomalies. It is these variations in temperature and density that generate the slow convective flow in the mantle. In a sense, then, seismic tomography is providing the first images (in the form of color-coded maps at various depths and vertical cross sections through the earth) of the vast machine that drives the motion of the plates.

Earthquakes radiate waves that travel along the surface as well as body waves that plunge through the interior of the earth. The velocity of long-period surface waves is affected by the earth's structure down to a depth of 700 kilometers, and so these waves yield information on the upper mantle. (The boundary between the upper and the lower mantle is usually placed at a depth of about 650 kilometers.) Don L. Anderson and his colleagues at the California Institute of Technology and Adam M. Dziewonski and John H. Woodhouse of Harvard University have analyzed surface waves from 60 large earthquakes that occurred between 1977 and 1982. The arrival times of these waves, which travel along great-circle paths, were recorded at some 20 special digital seismometers distributed around the globe.

The maps prepared by the Cal Tech and Harvard workers suggest a flow in the mantle more complex than that assumed by some theoretical models. As one would expect, midocean ridges and volcanic regions where hot magma wells to the surface are underlain in the shallow mantle by low-velocity anomalies corresponding to regions that are hotter than average. At a depth of 250 kilometers, however, only isolated low-velocity anomalies appear under three midocean ridges: the mid-Atlantic ridge, the South Indian Ocean ridge and the East Pacific Rise. At that depth magma seems to be fed toward the axis of the ridge by lateral transport from a few broad thermal anomalies deep in the upper mantle. In order to make computer simulation of mantle convection feasible most numerical models have assumed that the cyclic flow represented by a plate's outward motion from a midocean ridge, its subduction and the return of the material to the ridge are all in the same vertical plane. The tomographic analysis suggests this simplifying two-dimensional assumption is off the mark.

The new data are clear evidence of the importance of lateral transport of both cold and hot material in the mantle. At the same time investigators have found that many surface features have surprisingly large and deep "roots." Volcanic "hot spots" such as Iceland and the Hawaiian Islands, for instance, seem to be the surface expression of large upwelling mantle plumes anchored at a depth of at least 300 to 400 kilometers.

Robert W. Clayton of Cal Tech thinks the low-velocity seismic anomalies related to some hot spots, including Iceland, can be traced as deep as 1,500 to 2,000 kilometers. Clayton and his colleagues have mapped the lower mantle, which ends where the core begins at 2,900 kilometers, from the velocities of millions of compressional body waves registered at 1,600 standard seismic stations around the world. Their findings bear on one of the most sharply debated issues in geophysics: the nature of the boundary between the upper and the lower mantle. Some workers have argued that the seismic discontinuity at 650 kilometers (no earthquake focus has been observed at a greater depth) marks the boundary between materials that are chemically different and immiscible. If this were true, convection in the mantle would occur in two distinct layers. If, on the other hand, surface hot spots can be conclusively linked to anomalies below the boundary, it may simply be that the lower mantle is heating the upper mantle without material flow across the boundary; it may also be, however, that the convective flow is mantle-wide, with the two layers differing only physically, in their crystalline structures.

Clayton's group has found many large inhomogeneities in the lower mantle, particularly near the core. Not all of them are associated with hot spots, but according to Bradford H. Hager of Cal Tech, they do have an observable effect at the surface. Hager has calculated that the density anomalies in the lower mantle, when they are incorporated into a particular model of mantle convection, can account for most of the large-scale variation in the earth's gravity field, such as the slightly stronger gravitational force over Africa and the central Pacific. One of the interesting features of his model is its requirement that there be some convective transfer of material across the 650-kilometer discontinuity.

Although it will not be the last word on the nature of convection in the mantle, Hager's work illustrates well how tomographic analyses have already changed the debate. Until recently geophysicists had to infer the presence of anomalous mantle masses primarily from variations in the gravity field without being able to specify the depth of the density anomalies. They are now in a position to observe the anomalies more directly, to derive from them models of mantle convection and to use gravity data as a check on those models.

Hot Cars

Automobile engines could be substantially better if critical parts could be made of ceramic materials. Ceramics can tolerate higher temperatures than steel, and engines that run at higher temperatures are more efficient and emit smaller quantities of pollutant. Ceramics can also outdo steel in resistance to wear and corrosion, so that engines would last longer.

The potential of ceramic parts has been recognized for many years, but little had been done to realize it because the technology of steel engines was highly developed and the abundance of cheap fuel meant there was little economic incentive to design more efficient engines. Now that fuel is more expensive and thought is being given to alternatives such as coal-based fuels that are more corrosive than gasoline, a substantial amount of work is afoot to develop ceramic parts for automotive engines. The work is reviewed in *Mechanical Engineering,* the journal of the American Society of Mechanical Engineers.

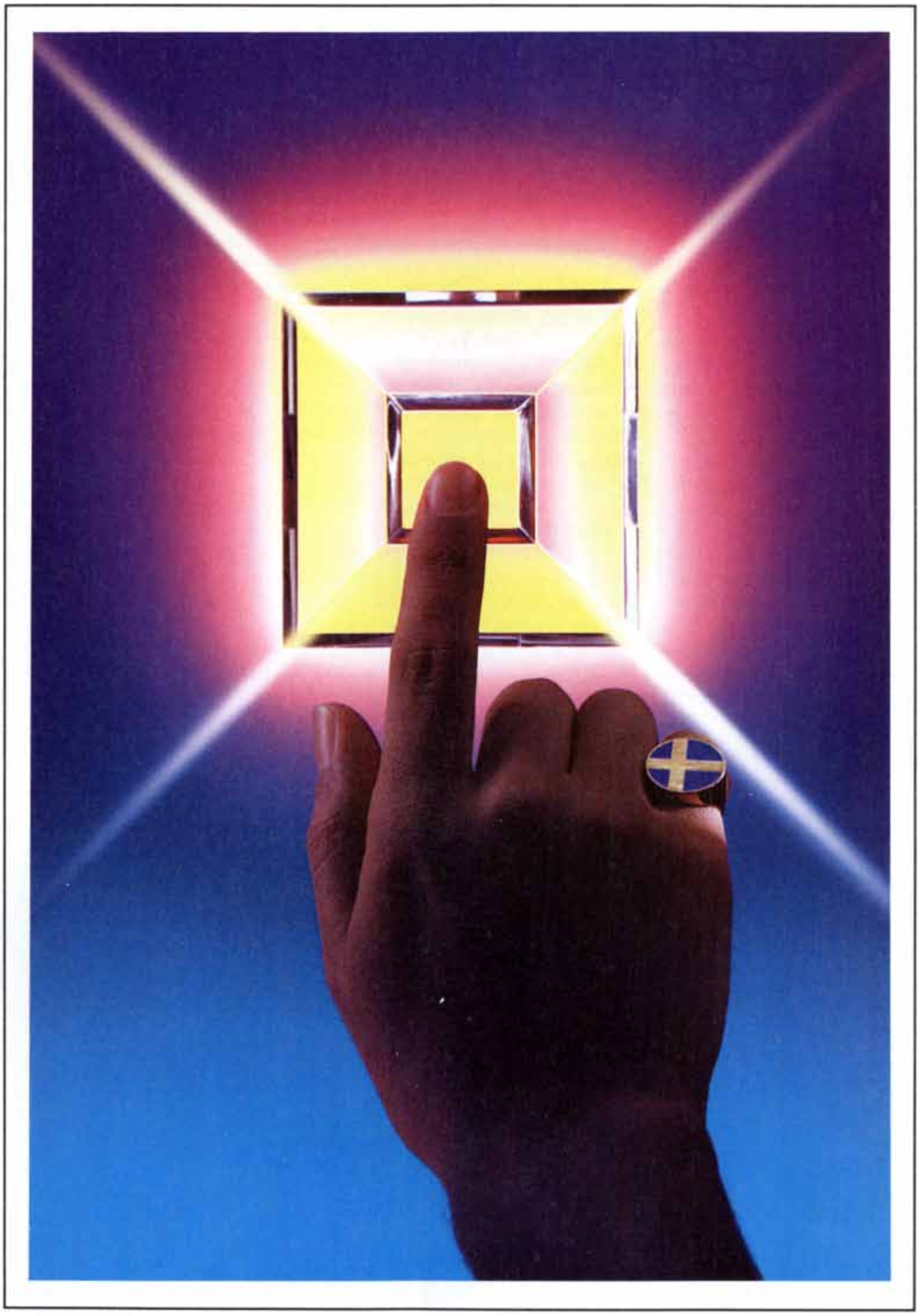
In the standard automobile engine the parts that might most usefully be made of ceramic materials are valves, the "hot plate" where the valves enter the engine, piston crowns, cylinder liners and liners for the exhaust manifold. Some or all of them may soon appear in the U.S. as an outgrowth of work being done on an adiabatic engine: an engine that would run at a stable temperature and in particular would not lose significant amounts of heat to the air or to a liquid coolant. The objective is a "minimum-cooling engine." Ceramic parts would be crucial in such an engine because of their ability to retain heat and to function at high temperatures.

Somewhat farther along is work on ceramic parts for turbochargers, particularly the turbine rotor. A ceramic rotor has about a third the density of steel, and its lower thermal expansion makes for operation at closer clearances and higher efficiency. Ceramic parts have been turning up in turbochargers for heavy-duty trucks, marine and industrial engines and more recently in automobile engines. In Japan ceramic rotors for turbochargers are on the market, along with small parts for diesel engines (glow plugs and prechambers, where fuel is injected before combustion).

Ceramic parts will also be crucial in the hot sections of the advanced gas-turbine engines that are under development in the U.S. A key part is the turbine wheel. Among the ceramic materials under consideration for such engines are silicon carbide, silicon nitride, lithium aluminum silicate, aluminum silicate, magnesium aluminum silicate and zirconia. The parts are made from a powdered form of the material by casting or molding (with the powder suspended in a liquid) or by sintering: compressing the powder.

Several problems must be overcome before ceramic parts become commonplace. Ceramics are brittle and therefore break easily. They cost more than steel parts, mainly because the manufacturers of ceramics have had little experience making parts that must meet rigorous standards of performance and seldom have had to make parts in the volume that would be required for the automotive industry. Lubrication is also a problem for a ceramic part that moves against other parts. The research now in progress is aimed at overcoming these problems.

SWEDISH ENERGY INNOVATION



Sweden's Move Away from Oil

This Special Report was researched and written by Don Hinrichsen

Sweden's industrial base has traditionally rested on a "tripod" structure of relatively cheap energy, abundant raw materials and an experienced, highly qualified work force. All three of these "legs" have been seriously weakened over the past decade. And the country has been forced to do some rather hasty reinforcing work in order to keep its industrial base intact.

After the 1973 oil crisis, Sweden found itself wallowing in the same wake of inflated oil bills as the rest of the world's oil-importing developed countries. But with two important exceptions: unlike the others Sweden was heavily dependent on oil, which accounted for nearly 70 percent of the country's energy consumption up to 1979-80; and second, also unlike the rest of Europe, Sweden decided to alter its energy course dramatically.

In 1975 the Parliament (Riksdag) introduced comprehensive legislation, known as the Energy Management Bill, that outlined, in part, how the country could begin to reduce its almost total dependency on oil. The program consisted of the following elements: deliberate conservation schemes, a broad energy R&D program, increased international cooperation, an active oil substitution policy, and guaranteeing an energy supply by using a variety of sources, including domestic "renewables."

In the mid-1970's the situation was critical. High energy prices began to cut into Sweden's industrial productivity and the economy slid into a tailspin from which it has still not fully recovered.

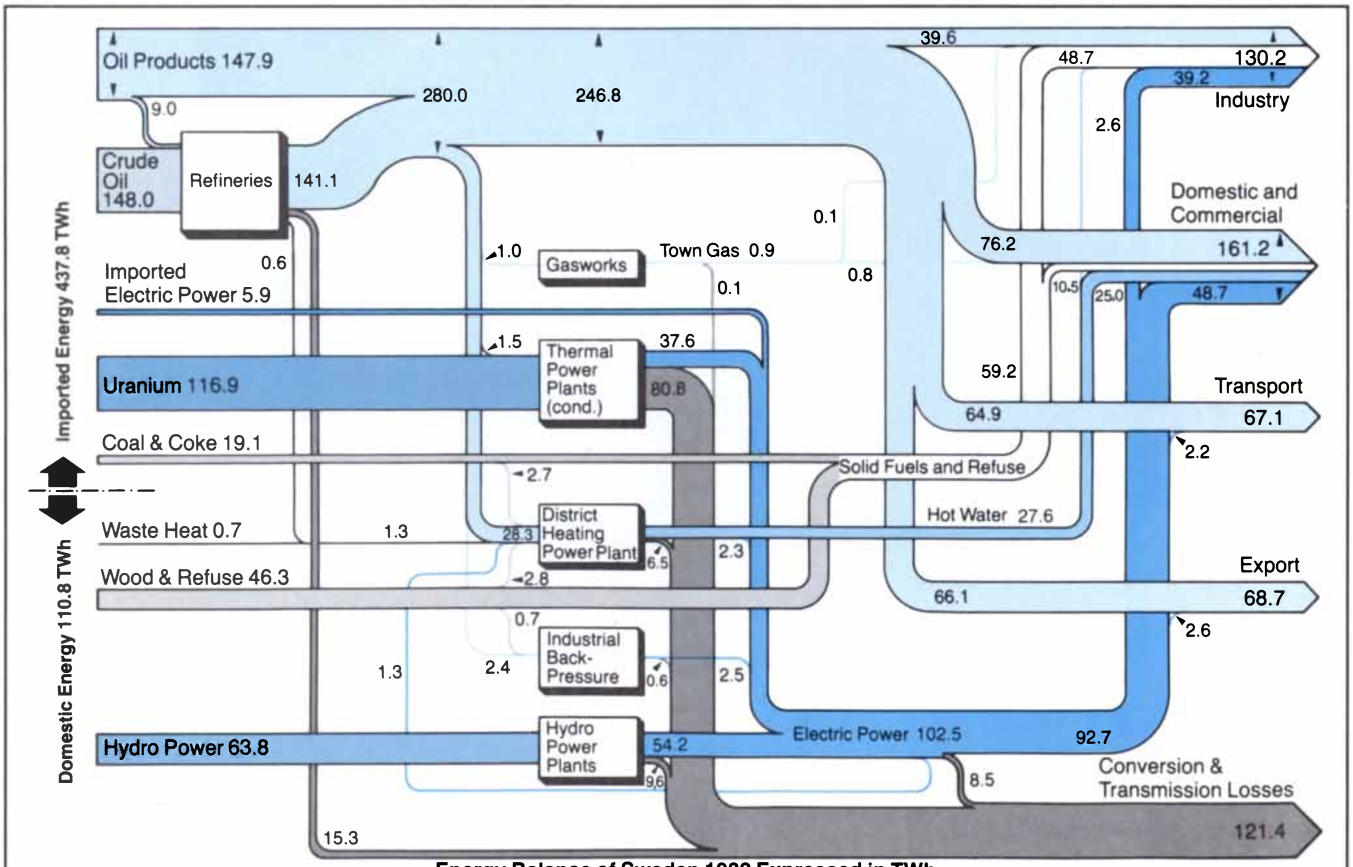
A simple comparison illustrates Sweden's energy dilemma. Whereas the average Swede in 1850 consumed the equivalent of two barrels of oil a year, mostly in the form of wood, coal or coke, the average Swede of 1975 used about 35 barrels of oil, most of it in the form of oil. In 1979, for example, the country consumed 26 million metric tons of oil, or approximately four per person, the highest per capita consumption of imported oil in the world. Today, Sweden's consumption of oil has dropped to about 16 million tons.

But a lot has happened since 1979. In 1981, one year after the national referendum on nuclear power that called for its eventual phasing out by the year 2010, Parliament drew up another ambitious piece of legislation, known as the Energy Bill. This package is nothing less than a blueprint for moving the country away from oil and keeping it away. The most striking feature is that it calls for a significant reduction in oil consumption, from the present 1983 level of around 50 percent to 40 percent by 1990. It also outlines how this can be done by stressing measures that have short- and medium-term impacts. Ironically, in order to do this, the bill calls for a return to the same sources of energy used by the average Swede over a century ago—wood and coal (although other factors such as energy conservation measures figure prominently in the equation).

A total of 2.58 billion Skr ($344 million)* has already been spent on energy R&D programs since 1975, but more is needed. The Energy Bill allocates 1.7 billion Skr ($226 million) for the oil substitution fund. Characteristically the money for the fund comes from a tax on petroleum products.

In the longer term, Sweden intends to develop indigenous biofuels, hopefully sometime during the 1990's, but in the meantime coal has been designated as the major transition fuel in the move away from oil.

At the time of writing 1 U.S. dollar equals approximately 7.5 Swedish Krona (Skr).



Energy Balance of Sweden 1982 Expressed in TWh
(Total Energy Turnover 549 TWh)

Source: AF Energikonsult

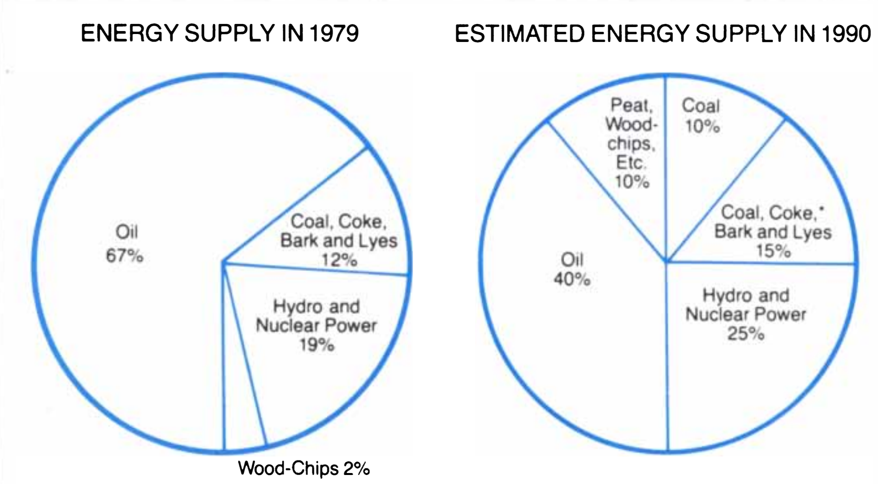


Figure 2

Figure 1 shows Sweden's total energy balance for 1982. Figure 2 compares the year 1979 with projections for 1990 and shows where energy planners think they can get the alternatives to oil. Figure 3 provides the same information in more detail.

The Energy Bill of 1981 also restructured a major part of the government's energy-related activities. Figure 4 shows how the various energy organizations were reshuffled: the National Industrial Board's Energy Section (SIND-E), the National Board for Energy Source Development (NE), the Oil Substitution Fund (OEF), the Delegation for Oil Substitution (OED) and the Energy Conservation Committee (ESK) were all placed under the newly created National Energy Administration with Carl Tham as its director general. Figure 4 also shows how the rest of the government bureaucracy concerned with energy fits into the overall picture. The reorganization was long overdue, according to some specialists, especially since it streamlines the government's channels and concentrates more of its essential activities under one roof.

The situation, however, is far from cut and dried. The course may be set, but getting there is going to be a major challenge for the country over the next decade.

Sweden's blueprint, spelled out in the Energy Bill of 1981, seems realistic enough, but there are complications that make its ultimate implementation problematic. Sweden not only has to switch fuels, but must do so with a number of options already closed to it, as if it were playing chess with some of the pieces missing.

Thanks to the 1980 referendum on nuclear power, the country's 12 nuclear reactors (10 at the moment) are slated to be phased out by the year 2010. Whether or not it will actually happen remains to be seen. But at this point, the government is not planning on continuing with its nuclear program. A second caveat: Parliament has placed a solid lid on developing further hydropower potential—estimated to be about 30 TWh/yr. According to Carl Tham, only an extra 3 TWh/yr will be on stream by 1990, bringing hydro total installed capacity up to 66 TWh/yr.

This gives the situation some urgency and makes the transition not only exciting but more difficult. Furthermore, there is really no time to waste, since the cheap electricity of the 1980's will be replaced by increasingly expensive electricity in the 1990's as the nuclear plants begin to be shut down.

The Swedish State Power Board (Vattenfall) predicts that there will be no hikes in electricity rates until the 1990's. This will buy the country valuable time—time it will need to make that most difficult of all energy transitions: the one away from oil toward more reliance on domestic energy sources, mainly biomass, with marginal contributions from solar energy and wind.

Fuel Switching

Of the 26 million metric tons of oil used by Sweden in 1979, 5.5 million metric tons went to fuel the transportation sector, 5.2 million metric tons were used in industry, 1.8 million metric tons were burned to generate electricity and 11.2 million metric tons—about 40 percent of the total—went for space heating and hot water, much of it provided by oil-fired district heating plants.

By the end of this decade, the government expects to replace nine million metric tons of that total with other sources: coal will replace three million metric tons of oil; peat and wood wastes will replace an additional three million metric tons; solar energy, along with alternative biofuels for transportation, should replace one million metric tons; increased use of electricity for heating (and in connection with heat pumps) should replace another one million metric tons; and a revised district heating network for Stockholm, using a large-scale heat pump plant with an output of 150 MW (thermal) and coal will save a final one million metric tons of oil a year.

Those district heating plants are a key element in the reduction of Sweden's oil bills. A major target in the government's oil substitution program is the conversion of the country's estimated 10,000 "block centrals" from oil to coal and biomass. Block centrals are mini district heating plants that generally serve up to 50 apartments. Most of these small-scale plants burn oil. Replacement of the burners in the boilers with newer ones capable of burning a variety of fuels, such as coal and wood wastes, will have a significant impact on Sweden's fuel-switching policies.

The Second Coming of Coal

It is somewhat ironic to see Sweden switching from oil back to coal when just a couple of decades ago the country was frantically making the switch from coal to oil. In 1979, 400,000 metric tons of coal were imported. This figure now stands at 1.2 million tons and is expected to increase to between four and six million metric tons by 1990—about the same level as in the early 1950's. Critics complain that this policy merely substitutes one form of imported energy for another, without reducing in any substantial way Swedish dependence on foreign sources of energy.

The inevitable environmental objections have been overcome, so far, by promises to use only low-sulfur coal—with a maximum sulfur content of 1.5 percent—imported mostly from the U.S., but with small shipments coming from Poland, the U.K. and the U.S.S.R. The coal will also be treated before being burned and the emissions carefully monitored with new environmental control technologies in combination with new combustion technologies being developed that permit the use of cheaper, high-sulfur coal.

One large-scale project already operating indicates the extent to which coal is coming back into fashion, despite environmentalist misgivings.

The Igelsta Plant

Billed as Europe's most modern district heating plant, the 360-MW coal-fired plant at Igelsta, near Södertälje (south of Stockholm), will burn 250,000 metric tons of low-sulfur coal a year in three 120-MW boilers. The plant will supply heat and hot water for 20,000 dwellings, a number of single-family houses and several large industrial enterprises in the area.

The plant incorporates the most modern pollution-control equipment and will reduce pollution loads by 60 percent compared with the previous oil-fired boiler. The project has brought together some of Sweden's most advanced companies in the energy field: Skanska, Fläkt and Götaverken.

In the final analysis, Sweden's brave attempt to alter its consumption patterns and to tap new energy sources is one way of getting off the roller coaster of oil prices that keep fluctuating wildly.

"All of our efforts," states Carl Tham, director general of the National Energy Administration, "are based on the assumption that the price of oil will keep going up in the future." Indeed, it rose 45 percent in two years (1981–82), sending shudders through the economy. "Even when the price of oil went down in general terms, it actually rose for us, because the Swedish krona is so devalued in relation to the dollar," points out Tham. "So in effect, the price went up for Sweden."

Sweden stands every chance of succeeding at getting away from an oil-based economy. If it does accomplish the task, the path it took will no doubt be well-worn soon after. Well-worn by the rest of us who will follow in their footsteps. ■

Sweden's Brave New World of Energy

by Brigitta Dahl, Minister for Energy

Energy is one of our most basic needs. The quest for energy to cook our food, to heat our homes, to make our work easier and more efficient has always been one of mankind's major priorities.

The technological level has increased even more rapidly: from the simple wood fire, which is still the only source of energy for nearly two billion people (half the world's population), to nuclear power plants and large-scale electricity distribution networks that link together entire continents.

The increasing use of energy is playing a significant role in efforts to improve the material well-being of people in the industrialized countries. But, unfortunately, our selection of specific energy sources as well as how we obtain and use energy can have serious disadvantages, impairing our quality of life on one hand, while enhancing it on the other.

This century has seen the quest for energy and especially its end uses embroiled in controversy. But our need for increasing amounts of energy has been so great and of such urgency that other interests—often more important ones—have been given a back seat.

Fuel wood exploitation in the Third World has devastated forests and laid the land open to erosion and impoverishment. Lakes and forests in the Northern Hemisphere have died from exposure to acid rains created by the burning of fossil fuels. Rivers have changed their courses and local climates have been affected as a result of the construction of hydroelectric power plants. Nuclear power has posed new risks—like the disposal of radioactive wastes—which have fueled public fears that the "peaceful atom" is an unsafe source of energy.

If we want to improve our quality of life, we still have to consider very carefully both the material and the psychological aspects of energy production and energy consumption. Economic realities have to be weighed against psychological factors and short-term interests balanced against more long-range considerations.

Before the "energy crisis" decade of the 1970's, Sweden's energy policy was largely nondescript and uncontroversial. With low oil prices, there was little competition from other energy sources and, as a result, Sweden's dependence on oil grew steadily after World War II.

The country then decided to take two steps to balance off the overwhelming dependence on imported energy. In broad political agreement, we decided to make increased use of our large hydropower resources and we launched a full-scale program to develop our own nuclear technology for the production of electricity.

But the turbulent 1970's changed this "consensus" on energy policy into one of confrontation. It became increasingly difficult to expand our hydroelectric power. The controversy focused on the role of nuclear power in the country's energy mix.

A compromise was reached through a national referendum, held in 1980, on the future role of nuclear power in Sweden. The people voted for the continuing use of those nuclear reactors already in use or being built, but for termination of the program after the 12th plant had been finished. Parliament then decided that all nuclear reactors would have to be closed down and phased out of operation by the year 2010 at the latest.

As a result of the Parliament's decisions taken on Sweden's future energy course, we will have to build a completely new energy system during the next 30 years. This is an unprecedented and difficult task, but also an inspiring one.

In order to achieve this goal we will have to develop new technologies, to explore overlooked possibilities in energy conservation and to utilize energy sources not yet developed except perhaps on a very small scale.

City heating systems are being built that will enable us to distribute the heat from our new energy sources such as peat, heat pumps and waste heat from industries or sewage plants, and perhaps later solar power plants. Municipal heating systems will also make use of the heat produced in "back-pressure" power plants, which we will have to build when all the nuclear reactors have been decommissioned.

At the same time, there are exciting developments going on in the area of small-scale combustion units, like fluidized beds, as well as in conservation technologies for private homes.

We are confident that our brave new energy future will work and that it will be completed according to plan. We also believe that our efforts may be of use to other countries since our energy situation is in no way unique. And just as we have taken up ideas from other countries and made them work in a Swedish context, we believe that our own "home-grown" solutions can be applied in varying climates and under varying industrial and economic conditions to help others tackle their own energy problems. ■

Total energy balances, TWh, in 1979 and 1990

	Year 1979	Year 1990 (Lower usage)
Oil and oil products	295	160–140
Coal	3	31–45
Coal and coke for metallurgical uses	18	17
Natural gas	–	4–9
Motor alcohols	–	1–3
Bark and Lyes	36	42
Forest energy, etc.[1]	7	25–30
Peat	0	6–11
Solar heat[2]	–	1–3
Waste heat[3]	2	3–4
Hydroelectric power	60[4]	65
Wind power	–	0–1
Nuclear power (electricity)[5]	22	56
Nuclear power (heating)[6]	–	9
Total supplies for energy purposes	**443**	**428**
Of which conversion and distribution losses of electricity and in district heating	27	28
Total final consumption	**416**	**400**

[1] Forest waste, wood, chips, energy forest, straw, etc.
[2] Added to which heat pumps are used in combination with naturally stored solar energy
[3] Incl. burning of refuse
[4] 61–62 TWh with normal precipitation
[5] Incl. the stations' energy consumption (gross production)
[6] If heat tap-off from the Forsmark 3 nuclear reactor is *not* adopted, the electricity production from nuclear power will increase by about 3 TWh. This electricity may possible be used for operation of heat pumps in Stockholm. The district heating supplies for Greater Stockholm may also be based on coal. The quantity of coal in the figures for 1990 wil then increase by about 11 TWh.

Figure 3 Source: Ministry of Industry

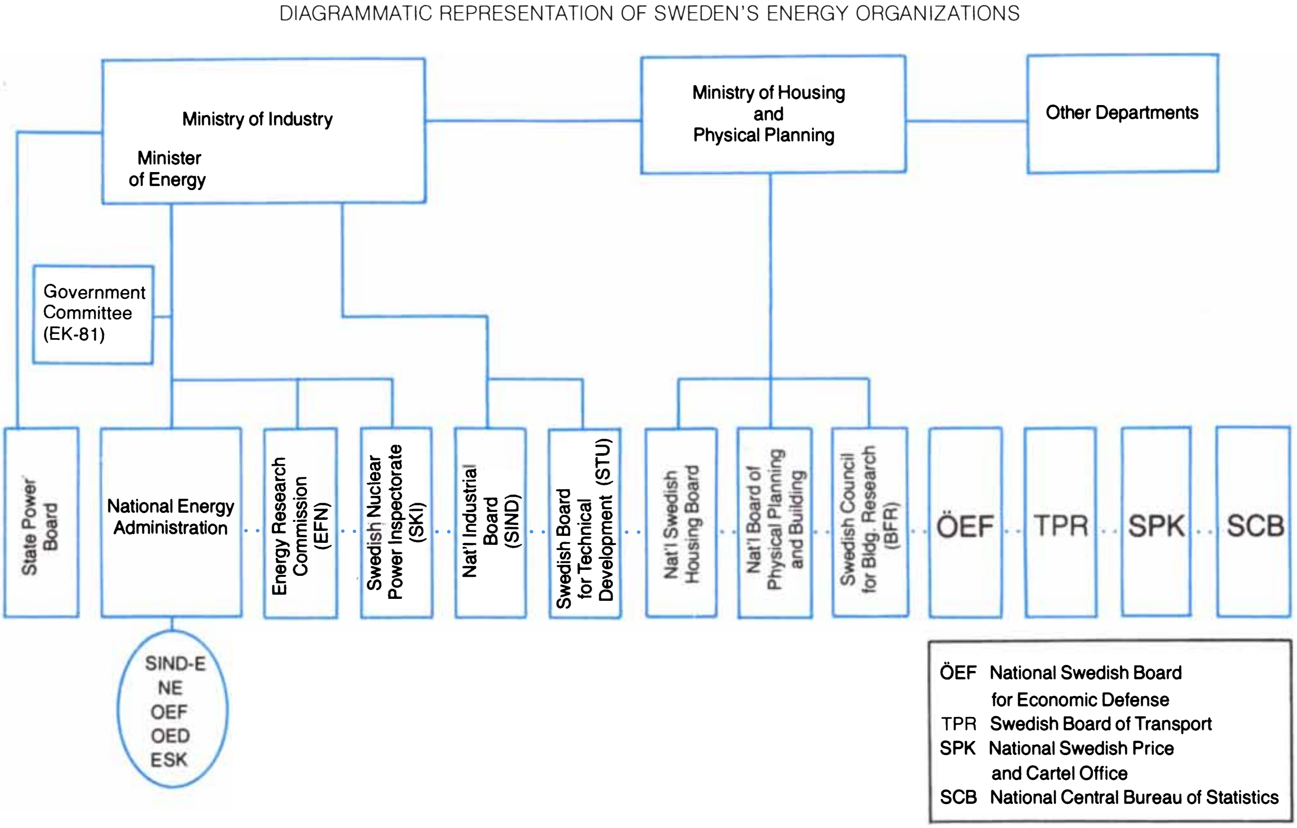


Figure 4.

(EK-81 is a government committee set up to study the phasing out of nuclear power, as called for in the Energy Bill of 1981.)

Energy Conservation: Waste Not, Want Not

Following the energy crisis of 1973, the one thing that all of Sweden's squabbling political parties could agree on, virtually without reservations, was the need to conserve energy. The politicos may have disagreed on the fine print, but not with the overall strategy. The country's energy planners were ordered to draw up a broad array of conservation schemes. Even a cursory look at Sweden's energy picture is enough to understand the importance of conserving energy. After all, 40 percent of Sweden's total energy goes for space heating and hot water for buildings and another 40 percent goes to keep Sweden's industrial machine turning. Obviously, by concentrating on these two heavy energy consuming sectors, the planners reckoned that they could achieve both the most immediate and long-term savings.

Since 1975 (when the major conservation strategies were launched) Sweden has invested roughly $1.2 billion in conservation measures. Over the same period nearly 4 billion Skr ($533 million) was invested in energy research, development and demonstrations. Reflecting this, the Swedish Board for Technical Development's (STU) Energy Department has earmarked most of its 120 million Skr budget ($16 million) for the current fiscal year (1983/84) for energy saving technological processes or systems in industry and for substituting domestic energy sources for imported oil. Furthermore, most of STU's requested budget of 433 million Skr ($57.7 million) for the next three-year period (1984-87) will go into energy conservation technologies or processes for industries.

Industrial Energy Saving

STU, reflecting government policy, provides risk money for industry in the form of outright grants or loans negotiated usually on a 50-50 basis. "Our main job," says Staffan Ulvönäs, head of STU's Energy Department, "is to take technical risks for industry."

STU's energy program covers a wide range of activities but concentrated under three main umbrellas: in saving energy in existing processes, reducing oil dependence and promoting a flexible approach to energy sources and end use. According to Ulvönäs, STU achieves these objectives when it comes to industry, by supporting "tightening up processes in existing operations" and in helping industry to develop different processes that either save energy or use domestic sources instead of imported oil.

This last area is where most of the new R&D money is directed. As part of the overall strategy to get Sweden off the oil hook, STU is also expending efforts to develop transportation systems using alternative fuels such as methanol as well as in promoting the development of electric vehicles.

Most of the basic R&D is being carried out in cooperation with technical research institutes such as the Chalmers Institute of Technology in Gothenburg and the Royal Institute of Technology in Stockholm. Over the course of the next three years (1984-87) STU will also give continued support to R&D for fluidized beds, more efficient solar collectors and energy storage in the ground.

Sweden's real energy conservation "success story" is the pulp and paper industry, which gobbles up 40 percent of all the energy used in the industrial sector. The Swedish Pulp and Paper Association reports that oil consumption decreased from 2.2 million cubic meters in 1973 to 1.45 million cubic meters by 1981, despite a 20 percent growth in paper production over the same period. This figure is expected to drop to 0.91 million cubic meters by 1986 (half the volume of 1978).

Göran Wohlfahrt of the Swedish Pulp and Paper Association writes: "Of course, most of the oil saved will be replaced by other energy sources: forest fuel, 180,000 cubic meters of oil equivalent; peat and coal each about 100,000 cubic meters of oil equivalent; more spent liquors and improved yield from bark by drying before burning, another 275,000 cubic meters of oil equivalent. Since both pulp and paper production are forecast to rise, and total use increases only marginally, there is also an improvement in overall fuel efficiency of 8 percent." Given all this activity, Wohlfahrt concludes that oil consumption might even be halved again after 1986. In 1982 alone the industry saved 500 million Skr ($66.6 million) in energy costs.

These significant reductions were made possible by a system of government grants—up to 35 percent of the total investment costs of new technologies, processes or plants—that were instituted by Parliament in the 1970's. This close cooperation initiated between government and industry continues on many levels.

The major technological fixes that have been introduced include: the widespread use of chemical recovery boilers for black liquors (black liquors are essentially dissolved wood from the pulping process); whole tree utilization techniques (stumps, roots, needles, branches, etc); replacing oil-fired boilers with ones that burn peat, wood chips, bark etc; and the general introduction of more efficient production processes requiring less energy.

"Chemical recovery from pulp and paper is the biggest chemical industry in Sweden," comments STU's Ingmar Fastmark. Not only have such energy conservation measures generated more jobs in other sectors as well, but in the final analysis, the pulp and paper industry is the first to admit that by reducing energy costs and developing more efficient production processes, the industry as a whole has become more competitive internationally.

Plasma Technology

The iron and steel industry has also made significant savings in energy, mostly by developing energy-saving processes. But in some cases entirely new technologies have been developed. SKF Steel Engineering has recently introduced a series of industrial applications of plasma technology, developed by the company during the last decade.

In most of these applications a cheap fuel, for example, steam coal, is converted to a fuel gas which is directly used in an industrial process, *e.g.*, PLASMARED for production of so-called sponge iron, and PLASMADUST for recovery of metals from various wastes. For thermodynamic reasons, all gas cannot be utilized in the process. A clean energy-rich off-gas is thus obtained. This gas is perfect as a fuel gas for the generation of steam or hot water, cogeneration of electric energy or other purposes.

An example is the SCANDUST project, a commercial PLASMADUST plant being constructed by SKF Steel Engineering for a customer in southern Sweden. The off-gas from the plant will be used for generation of hot water for the local district heating system.

Plasma technology for the gasification of fuels does not have to be used in combination with metal production. Practically any kind of fuel can be converted to a fuel gas, e.g., coal, oil, peat, wood and some household waste. The energy authority in the city of Gothenburg is studying a detailed proposal from SKF Steel Engineering concerning the conversion of a butane-based town gas works to a coal-based operation, using plasma coal gasification. The coal gas produced in this case would be converted to a standard town gas with the aid of a conventional methanation process. Excess heat will be used in the local district heating system.

Plasma gasification of coal or other fuels has big advantages from an environmental point of view. The gasification temperature is sufficiently high to exclude formation of hazardous or unpleasant hydrocarbons, and practically all ash in the fuel ends up as a molten slag, where all elements after cooling and granulation are firmly bound. In addition, sulfur from the fuel is easily absorbed in a dolomite or lime filter.

For all plasma processes, plasma generators developed and manufactured by SKF Steel are being used. Generators with power levels of 1 to 10 MW are available. The PLASMADUST plant under construction, will use three 6-MW plasma generators. ■

Energy Saving in Buildings

Over the past eight years this sector has seen amazing improvements in the energy efficiency of buildings, both commercial and residential. Some of these improvements were predictable if not inevitable. Sweden's dour climate is infamous. Long dark winters, followed by short summers—with indoor and outdoor temperature differences sometimes reaching 40 degrees C.—means that space heating (including hot water) requires considerable energy inputs, even more so if the buildings are not tight.

Many of Sweden's older buildings, especially apartment complexes, were plagued with faulty insulation materials (or none at all), lacked double glazed windows, and had bad air circulation systems. The government tackled the problems with a host of R&D activities, tightened up the building codes, and instituted generous loans and grants to encourage home owners (as well as multifamily dwellings) to improve the energy efficiency of their buildings.

The programs have paid off dramatically. Lee Schipper, an American scientist commissioned by the Swedish government to compile a comprehensive report on residential energy use and conservation in the country, points out that between 1974 and 1979, 320,000 single-family dwellings and 840,000 multifamily dwellings were given conservation grants and loans. During this period the government paid out 1 billion Skr in grants and 2.5 billion in loans ($133 million and $333 million respectively). Schipper writes that the "expected savings (all fuels mixed) were about 10.5 Petajoules (PJ) per year, giving an investment cost of 360 Skr per Gigajoule per year (GJ/yr) saved and a savings of about 10 GJ per dwelling.

Says Ingrid Munro, Director General of the Swedish Council for Building Research, "Compared to 5 years ago, we have saved 70 percent of the energy used in normal houses."

Sweden now holds the record for having the lowest energy use and the lowest heat losses for centrally heated dwellings of any country in the OECD.

"Normal one-family houses in Sweden use only 50 percent as much energy as a comparable house in North America," concludes Ingrid Munro. ■

SKF STEEL ENGINEERING:
A New Generation of Energy-saving Plasma-based Technologies

SKF Steel Engineering is part of the well-known SKF Steel Group, the largest producers of specialty steels in Scandinavia. The parent company supplies nearly 20 percent of the Western world's ball and roller bearings (their European share is an impressive 35 percent).

SKF Steel Engineering concentrates in two areas: specialty steel engineering know-how and plasma technology.

SKF Steel owns and operates two steel plants in Sweden with a combined yearly output of some 750,000 ingot tons. In the 1960's, SKF Steel together with ASEA, developed the ASEA-SKF ladle furnace, considered a big advance in ladle technology. Today, there are over 50 ASEA-SKF ladle furnaces operating worldwide.

The ladle refining process was later combined with a twin-shell furnace for more efficient melting of scrap and other raw materials. The combination of the twin-shell furnace and the ladle refining process is now known as the SKF/MR (Melting and Refining) process. Today, all of SKF Steel's output is based on SKF/MR technology.

However, it is in the area of plasma technology where SKF Steel Engineering has made the most significant breakthrough. Back in the early 1970's, SKF Steel launched a long-range R&D program aimed at expanding and perfecting the applications of plasma technology. This R&D program had four basic goals: energy efficiency, energy flexibility (i.e. able to use different types of fuel), environmentally sound and small scale.

This necessitated the development of a new generation of plasma generators: the basic tool which has enabled the company to forge ahead in plasma technology. SKF Steel's plasma generators transform nearly 85 percent of the electrical energy supplied into heat energy. Furthermore, there are no waste gases with a plasma generator and the total gas volume is smaller than with other processes. Thanks to the extremely high energy concentration in the plasma, the reactions can take place within a very small space. These two factors make it possible to use much smaller equipment than with other conventional processes, reducing considerably the capital investment.

The range of plasma applications can be separated into the following categories:

1) The PLASMARED process for producing sponge iron.

2) The PLASMASMELT family of processes for smelting and reduction of various metal oxides to metals. The PLASMADUST process to be used in the Landskrona plant falls within this category (see below). It also includes PLASMAZINC for the production of zinc from primary raw materials and PLASMACHROME for the cost-efficient production of ferro-chromium. Other variations on this theme include the production of ferro-manganese and other ferro-alloys.

3) PLASMABLAST, which involves the use of plasma generators to reduce the energy costs of existing blast furnaces.

4) PLASMAWASTE, which is a process for the incineration of household and industrial waste.

SKF Steel Engineering has supplied a turnkey PLASMADUST plant to ScanDust AB at Landskrona, in southern Sweden. The PLASMADUST process will recover zinc, lead, chromium, molybdenum, iron and nickel from baghouse dust—the wastes generated by steel mills. The dust will come initially from Scandinavian steelmakers, but will also be shipped to Landskrona from the continent.

The PLASMADUST plant begins commercial operations this year and will recover some 36,000 metric tons of metals from 70,000 metric tons of waste per year. And as a result of the substantial environmental benefits inherent in this process, the National Swedish Environment Protection Board invested some 25 million Skr ($3.3 million) in the project. Recovering metals in waste dust from the steelmaking process will mean that the heavy metals in the dust can be reused instead of being dumped in landfills, where they often leach into the groundwater.

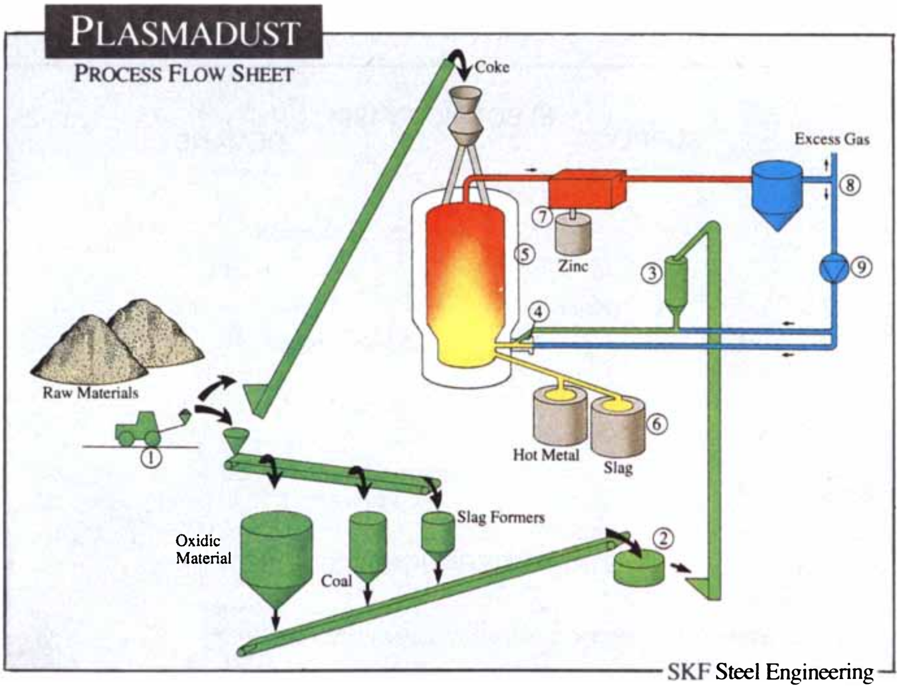
In addition, the Landskrona plant will generate 65 GWh per year of off-gas for the community's district heating network, saving about 1.7 million gallons of oil a year.

The baghouse dust, supplied to the plant in bags or containers, is mixed with coal, sand and water in a slurry mixer. The mixture is then fed into a filter to remove most of the water, which is used primarily as a medium to obtain a homogeneous mixture in the slurry. The filter cake formed in the filter is further dried, crushed and fed into the shaft furnace.

The plasma itself is formed in the plasma generators which develop temperatures of 3,000-5,000 degrees C. With injected coal powder, the metal oxides are reduced during the smelting process. Zinc and lead vaporize, rise in the shaft and condense to a liquid in the condensers. The iron forms a hot metal that is tapped from the shaft furnace. When the plant is using dusts from furnaces producing stainless or other high-alloy steels, the result is an alloyed hot metal usually containing rich amounts of chromium, molybdenum, iron, nickel, etc.

The exhaust gas, containing a substantial amount of heat energy, is cooled and cleaned. A certain amount of the gas is recirculated as process gas, but most of the off-gas is used in the city's district heating network.

SKF Steel Engineering is convinced that the commercial breakthrough for the PLASMADUST process is just the beginning for plasma-based technologies. And the company is continuing to plow substantial investments back into plasma technology R&D. ■



Hydropower and the National Electricity Grid

It is no exaggeration to say that Sweden's rapid industrialization at the end of the 19th century would not have been possible without cheap and abundant hydropower. The first large-scale hydro plants went on stream in the 1910's, followed by hundreds more. Today, there are around 1,000 hydropower stations scattered about Sweden (mostly in the North) with a total installed generating capacity of 15,215 MW. Sweden's vast reservoirs hold another 32 billion kilowatt hours per year (in years with normal rainfall).

In 1982, the country's hydro plants produced 54,191 Gigawatt hours of electricity, accounting for about 55 percent of Sweden's electricity production. Presently, about 63 TWh/yr of hydropower is available. The Swedish State Power Board (Vattenfall) points out that Sweden's remaining untapped rivers in the northern part of the country contain an extra 30 TWh/yr of hydro potential that could be technically and economically exploited. However, the country's strong and vocal environmental movement has managed, so far, to prevent further hydro projects from being developed. And Parliament has decreed that the remaining four free-flowing rivers—the Torne, Kalix, Pite and Vindelälven—shall remain inviolate.

Sweden's Grid

Vattenfall and some 16 other major producers (including municipal-owned utilities and big companies) generate about 100 TWh/yr of electricity which is distributed in a nationwide grid (*see Figure 5*). In 1982, 55 percent of this electricity was produced from hydropower, 40 percent from nuclear, and 5 percent from other sources like coal, oil and cogeneration plants. All told, some 100 billion kWh of electricity were consumed in Sweden in 1982. This figure is expected to increase to 130 billion kWh by 1990. Until the very end of this decade, nearly all of Sweden's electricity needs will be met with hydro and nuclear.

Vattenfall is by far the largest utility in the country, accounting for about 50 percent of all electricity generated. This government-owned public corporation also controls and operates the national electricity grid, which shuttles power throughout the country.

With most hydro in the northern half of the country and the bulk of Sweden's 8.3 million people in the southern half, Vattenfall, in close cooperation with ASEA, developed high voltage transmission systems capable of shunting the power from North to South, with minimal line losses. In 1952 ASEA, collaborating with Vattenfall, built the world's first 400-kV (kilovolt) high voltage transmission line, running over 1,000 km between Sweden's largest hydropower plant—the 940 MW Harsprånget plant on the Lule River—and the town of Hallsberg in the south.

Today, the 400-kV network contains seven North-South lines with a total length of roughly 9,000 km.

ASEA is responsible, in large measure, for the technological breakthroughs that have made the long distance transmission of both AC and DC high voltages possible. And they have developed reactive power compensation systems that are the most advanced in the world for controlling high voltage transmission as well as reducing power losses in the lines. In the late 1960's and early 1970's, ASEA introduced series capacitors on all of Sweden's 400-kilovolt lines.

What this means in practical terms is that ASEA's technicians were able to optimize the power transmitting capacity of the existing lines, greatly increasing the amount of power transmitted. It also meant that Vattenfall did not have to build more, expensive transmission lines.

"There are nine series capacitors on all of Sweden's 400-kV lines running from the northern hydropower plants to the South, where the customers are," says Mr. Sture Torseng of ASEA's Reactive Power Compensation Department. "We have been able to transmit 100 percent more power on these 400-kV lines due to the series capacitors and shunt control compensators."

The Swedish grid is, in turn, plugged directly into the national electricity grids of neighboring Norway, Finland and Denmark—making the entire network one of the largest in the world. This gives each country great flexibility when buying and selling electricity. If shortfalls suddenly arise, Vattenfall, for example, can buy cheap hydropower from Norway (whose grid is 100 percent hydro), instead of burning oil in back-up plants. "This way," points out Vattenfall's Managing Director Jonas Norrby, "we can get Norway's hydro for perhaps 8.5 öre per kilowatt hour, instead of spending 15 öre/kWh on oil."

The overall cost of electricity in Sweden to the distributor is about 15 öre per kilowatt hour, one of the lowest electricity prices in the world (only Norway's 100 percent hydro system is cheaper).

Comments Norrby, "We consider the nuclear, hydropower mix very economical. This is why we can offer such cheap electricity."

Vattenfall is unique among state-owned enterprises in that it is free to act like a private corporation. In 1982–83 its total turnover from sales was 10 billion Skr ($1.33 billion), predominantly from the sale of electricity, with a small amount from heat (district heating) as well.

All things considered, Vattenfall insists that the prices for Sweden's electricity will not rise above normal inflation levels until the beginning of the next decade, aiding the country's transition from oil to coal and domestic renewables. ■

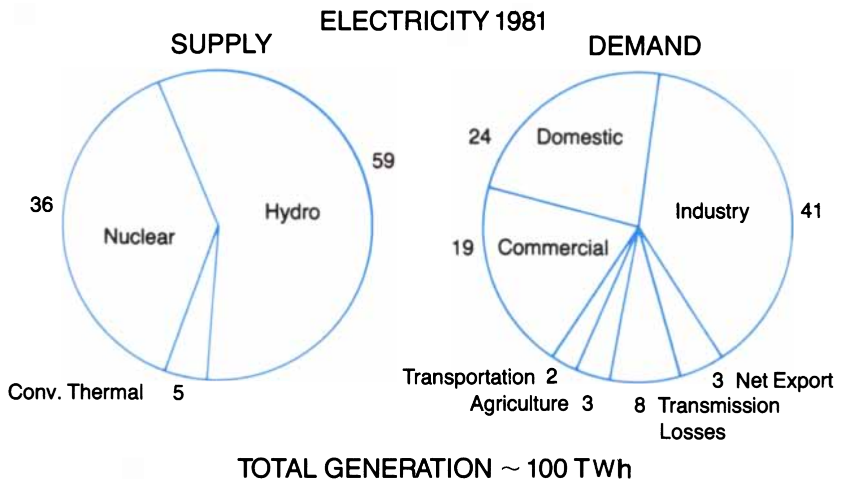


Figure 5. Sweden's total electricity output and consumption in 1981.

Nuclear Power: What Future?

All around the world, utilities and energy planners are asking themselves the question: What is the future of nuclear power? As the price of nuclear-generated electricity continues to rise, especially in relation to that of fossil fuels, its future looks grim indeed.

In Sweden, probably no other issue in recent memory has proved so disruptive as the nuclear debate to the middle-of-the-road, consensus politics that has been this country's hallmark. It prompted a convulsive and bitter "civil war" within the scientific community, which quickly cut through the social fabric of the country, dividing families and ripping apart traditional political alliances. The nuclear question (as it came to be called) became so politically hot that it caused the downfall of two governments: the Social Democrats in 1976 and a Center Party–led coalition in 1978.

In order to resolve the issue, a national referendum was held in March, 1981 to determine the course of nuclear power in Sweden's energy future. Typically, perhaps, the referendum itself touched off a nasty row, since it offered three alternatives and so could not be settled with a simple "yes" or "no." On top of this, many felt it was silly to have a referendum nearly 10 years *after* the first nuclear plant was commissioned.

To cut a long story short, the Social Democratic alternative, called the "parenthesis" line, which called for the phasing out of all nuclear power stations by 2010, won a majority. That, at least, left the government with a clear goal. The blueprint for reducing oil consumption to 40 percent by 1990 and dismantling the country's nuclear infrastructure was outlined in a major piece of legislation—the Energy Bill of 1981. But blueprints are one thing, reality often another.

Nearly three years after the referendum, the scars have still not healed. Major segments of Swedish industry remain skeptical that nuclear energy can be done away with at all, and certainly not within the time frame called for. The government, on the other hand, seems determined to push ahead according to plan.

In all fairness, the Swedish public increasingly perceived nuclear energy in terms of its problems and uncertainties, rather than its benefits. Right or wrong, many Swedes faulted nuclear power on three counts: its potential for proliferating nuclear know-how (for nuclear weapons), the question of safety and the disposal of radioactive wastes.

SWEDISH REACTORS

REACTOR		POWER	COMMERCIAL OPERATION
Oskarshamn 1	BWR	440	1972
Oskarshamn 2	BWR	580	1974
Barsebäck 1	BWR	580	1975
Ringhals 2	PWR	820	1975
Ringhals 1	BWR	760	1976
Barsebäck 2	BWR	580	1977
Forsmark 1	BWR	900	1980
Ringhals 3	PWR	915	1980
Forsmark 2	BWR	900	1981
Ringhals 4	PWR	915	1982
Oskarshamn 3	BWR	1050	1985
Forsmark 3	BWR	1050	1985

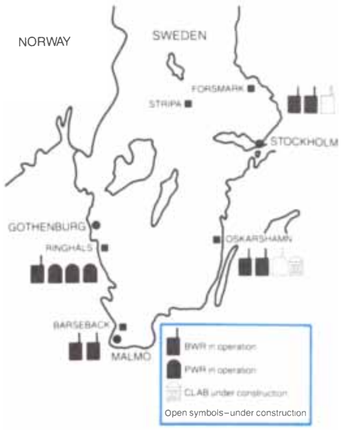


Figure 6. The Swedish nuclear program.

Nuclear Scorecard

At the moment there are 10 operating nuclear reactors in Sweden with an installed capacity of 7,330 MW, producing nearly 40 percent of the country's electricity output. Seven are boiling water reactors (BWR's), all designed and manufactured by ASEA-ATOM, while the other three are pressurized water reactors (PWR's) based on Westinghouse technology. By 1985–86 the two final plants, both built by ASEA-ATOM (1,060 MW each), will be operating, thereby increasing the total output of Sweden's reactors to 9,450 MW. At the end of this decade nuclear power should be producing about 45 percent of the country's electricity. This percentage will begin to decline sometime in the late 1990's if the older plants are decommissioned (*see Figure 6*).

In 1982, Sweden's 10 reactors produced 39 TWh of electricity, which amounted to around 4,600 kWh (kilowatt-hours) of nuclear generated electricity per person, the highest in the world.

Since Sweden does not produce its own uranium, the nuclear industry buys its enrichment uranium from foreign sources: mainly from the U.S. government and the Soviet Union.

Ranstad

It is not a very well known fact that Sweden has 70 percent of Europe's uranium deposits. The gently rolling hills of western Sweden around Ranstad contain some one million tons of uranium in a low-grade orebody—although not all of it recoverable. The catch is that it takes one ton of ore to get 300 grams of yellow cake. Uranium is now selling for about $25 per pound. Says ASEA-ATOM's Arto Kaipainen, "The price of uranium would have to triple before Ranstad would be economical. And it would take five to 10 years before Ranstad could have a commercial operation." For the foreseeable future, Ranstad's uranium reserves will stay where they are, in the ground.

Nuclear Industry

Sweden's nuclear industry essentially consists of one company—ASEA-ATOM, formed in 1969 and part of the ASEA group. In close cooperation with the government and the State Power Board, ASEA began to develop a boiling light water reactor back in the 1950's and 1960's. They delivered their first BWR—a 440-MW turnkey unit—to OKG utility in Oskarshamn in 1971 and it began commercial operations in 1972.

Now with a firm lid placed on any further nuclear development in Sweden, the company has turned abruptly to other fields of power generation, such as coal and fluidized bed technologies. Despite the prospects that the company may never again have a domestic market for their reactors, ASEA-ATOM has pushed ahead with novel reactor designs with the export market in mind.

ASEA-ATOM began work on its SECURE reactor for district heating in 1973–74. SECURE, which means Safe and Environmentally Clean Urban Reactor, is based on ASEA's BWR design, but with one important switch. The reactor core is at the bottom of a large pool containing cold, highly borated water. The pool is enclosed in a cylindrical prestressed concrete tomb and the entire plant is below ground level. It is covered with a prestressed concrete lid and is slightly pressurized. ASEA-ATOM puts the costs of getting SECURE at $80 to $100 million for a unit of up to 400 MW thermal. ■

Green Power: Biofuels Are a Growing Concern

In order to reach the goals spelled out in Parliament's comprehensive Energy Bill of 1981, Sweden has embarked on a fairly substantial biomass program: everything in fact from energy plantations and utilizing logging waste and junk timber to harvesting agricultural residues such as straw. But whether or not Sweden will actually be able to run in part on "plant power" depends to a large extent on the outcome of a variety of test projects, both small- and large-scale, that are being carried out now. No doubt Sweden will use more of its wood resources, but how much is difficult to predict.

When one tries to nail down the amount of exploitable fuel wood in Sweden's vast forests, the going gets very thick indeed. Sweden has often been called "Europe's last great wilderness," since the middle half of the country is almost solid woods—the area of productive forest land covers 235,000 square kilometers (or 53 percent of Sweden's total land area). However, the actual amount of marketable timber cut every year in Sweden's forests amounts to nearly 70 million cubic meters (solid volume), most of it going into exports worth 25 billion Skr ($3.33 billion) annually. About 50 million cubic meters of potential fuel wood rots at the cutting sites after private and commercial logging operations have finished. This waste consists of stumps, roots, tops, branches, leaves, needles, trimming wastes and deciduous "junk" trees. Of this 50 million cubic meter potential, the Energy Administration estimates that close to 28 million cubic meters of it could be economically harvested if whole tree utilization techniques were introduced on a large scale. But even this modest figure is chopped down by the forest industry to between eight and 13 million cubic meters, given the present state of harvesting technologies. The real figure probably lies somewhere between 11 and 28 million cubic meters.

The Swedish Pulp and Paper Association is quick to point out that the country's forestry industry is already beginning to utilize nearly every stick of wood in order to reduce wastage and cut down on energy costs. Many pulp and paper companies, such as Stora Kopparberg and MoDo, use stumps and roots, converting this "waste" into extra wood chips for the pulping process, and bark-fired boilers are replacing oil-fired ones in many sectors of the industry. Furthermore, as the demand for wood-based products increases, companies will exploit their own resources more efficiently.

As of 1982 there were around 46 large-scale plants fired with biomass (some burning peat and refuse-derived fuel as well) scattered throughout the country. Most were industrial boilers, but a number of the plants were producing hot water for district-heating. In Mora, central Sweden, a 4-MW district heating plant burns 30,000 cubic meters of wood chips a year, saving the community around 30 million Skr ($4 million) annually in heating costs. In northern Sweden, the city of Östersund has one wood-fired district heating plant and another will soon be operating on peat.

For Sweden to reduce its dependency on oil to 40 percent by 1990, the government is counting on using at least 13 million cubic meters of wood-derived fuels by the beginning of the next decade. Given the potential amount of wood in the forests, this seems like a reasonable enough target. Coupled to this is the likelihood that Sweden will also have full-scale energy forest plantations operating by the 1990's and will also have begun to exploit its vast reserves of peat. Despite these bright prospects, there are several important caveats.

Eighty years ago the country virtually ran on wood and coal, but the switch to oil in the 1950's meant heavy restructuring. Before Sweden can go back to being "green," some obstacles have to be overcome. Not the least of these obstacles is the creation of a new infrastructure that can economically use biomass-based fuels. This means that growing "green energy" is only the first step, although an important one. Harvesting machinery must be developed and demonstrated, boilers must be able to burn a variety of biomass fuels and transportation systems must be established. In all of these areas, the country is muscling ahead. The question is: Will they make it on time?

Energy Plantations

To date, Gustaf Sirén from the University of Agricultural Sciences in Uppsala, and his researchers have experimentally cultivated over 6,000 clones, but so far only 20 to 30 varieties seem able to stand up to his stringent 16-point criteria of acceptance (*see Figure 7*).

Research has shown that the most suitable trees for short rotation energy forests in Sweden are willows (Salix),

Sweden's energy forest program has a rigorous 16-point criteria of acceptance for developing fast-growing clones.

- Growth Rate
- Growth Rhythm
- Rooting Propensity
- Resistance to Parasites and Pathogens
- Hardiness to Frost
- Tolerance of Physiological Stress
- Dry Matter Production with Different Rotations
- Sprouting Propensity
- Response to Fertilizers, Ion Absorption Capacity
- Tolerance to Biotope Variability
- Trunk and/or Branch Type
- Wood Attributes (Wood Density, Moisture Ratio, *etc*.)
- Bark Thickness
- Leaf Attributes
- Attractiveness to Wild Animals
- Sociability (Tolerance of Screening/Spacing Variability)

Figure 7 Source: G. Siren

One of Sweden's experimental energy forests near Stockholm sprouts fast-growing clones of willows and poplars.

SKANSKA: Creating Custom-made Energy Storage Beneath the Earth's Crust

Barely 10 years after the world's first gas streetlight lit up Pearl Street in New York City, Skanska built its first hydroelectric power plant in northern Sweden. The year was 1890. Since then, this giant Swedish construction company has built over 200 hydropower stations, large and small, in a dozen countries around the world, close to 100 of them in Sweden.

With its vast experience in hydro construction, which involves a significant amount of tunneling work, Skanska gravitated, as it were, further underground. Today the company, which claims to be able to "build everything, anywhere," has either completed or is in the process of completing some 80 different underground storage facilities. Most of these projects—carried out in the Nordic countries and France—involve energy storage of one kind or another. Their underground facilities accommodate everything from crude oil, propane, butane, petrol, gas-oil, diesel oil, jet fuel, kerosene and coal to hot water for district heating and the intermediate storage of highly radioactive spent nuclear reactor fuel in underground concrete water tanks. According to Skanska, the total volume of all this storage space amounts to about 15 million cubic meters.

Petroleum storage in solid rock is a relatively recent development, given a needed nudge by the "energy crisis" of the 1970's.

"The construction techniques vary, of course, depending on what is to be stored, but two general rules apply in all cases of liquid energy storage: the products must be stored *below* the water table and must be nonsoluble in water," points out underground project manager Hans Pilebro.

The reasons for this are readily apparent. All petroleum products committed to the earth must be sunk below the groundwater table because 1) the product must be kept out of the groundwater and, more importantly, 2) the natural groundwater pressure in the rock can be used to reinforce the cavern, keeping the product locked firmly inside its underground "cocoon."

Groundwater, of course, finds its wet way into the caverns and must be pumped out via a separator. But then, underground caverns, such as mines, are not naturally waterproof. And when it comes to energy storage caverns, they are *designed* not to be waterproof. Crude oil (like most other petroleum products) floats on a "water bed," the level of which is constantly regulated by pumps.

More volatile petroleum gases, such as propane, can be stored safely and securely in the ground too, but they must be placed deep enough so that the groundwater pressure in the surrounding rock is higher than the pressure of the gas. "In practice this means that the roof of the propane storage cavern is 100 meters (300 ft.) *below* the water table," explains Gösta Jansson, Skanska's business development manager for international projects.

With projects getting larger, deeper and more complicated, Skanska's wealth of underground experience has netted them about 60 percent of all energy storage projects in Sweden.

Recently, Skanska finished one of its largest underground projects, a mammoth cavern for storing 800,000 cubic meters of crude oil, in Gothenburg on Sweden's storm-tossed west coast.

By contrast, one of the company's smaller projects (in terms of volume) happens to be one of its most interesting. The Uppsala suburb of Lyckebo is the test site of a unique 100,000 cubic meter underground cavern that will store solar-heated hot water from summer to winter to provide space heating and hot water for 550 single-family homes.

The unusual shape of the cavern—circular with a pillar of solid rock in the middle—was designed to minimize heat losses to the surrounding rock (*see photo*). Skanska had to keep the volume large but the surface area relatively small.

However, Skanska's senior vice-president, Folke Eneroth is quick to point out that this is only a test project: "The total cost of the project is set at 39 million Skr ($5.2 million), but most of that goes for the flat plate solar collectors and auxiliary equipment. The cavern and related hardware cost only about 17 million Skr ($2.26 million)."

Skanska is also involved in another, much larger, district heating project at Igelsta, just south of Stockholm. The Igelsta project has brought together a number of Sweden's most innovative companies.

One of the conditions for using coal, however, was that the plant should be provided with a completely closed coal handling and storage system. Skanska got the contract to design and construct two deep silos capable of storing 100,000 cubic meters of coal safely in the ground. The company also built in a number of safety devices that will more or less eliminate the possibility of spontaneous ignition, always a nagging problem with stored fuels.

Skanska, with its unique store of practical knowledge, expects to make many more constructive contributions, offering solutions to some of the troublesome technological snags encountered in storing and moving energy. ■

alders (Alnus) and poplars (Populus).

At five experimental stations scattered around Sweden—Bogesund, Studsvik, Ultuna, Köping and Sösdala and at Jädraas in Norrland—researchers were able to produce very high-yielding clones using normal doses of agricultural fertilizers. The plots were carefully selected for different microclimates and soil conditions. The results, according to Sirén, are impressive. Some of the plots produced up to 60 cubic meters of wood per hectare annually, or in other words, 20 tons of dry matter per hectare per year. For those who like to fiddle with figures, 7.5 cubic meters of Salix wood equals one ton of crude oil. In this context, the energy content in a ton of biomass is approximately equivalent to the energy needed to drive a truck about 5,000 km. Dr. Lars Hansson, from the Biomass Division at the National Energy Administration, calculates that a trained farmer will be able to produce five to 10 tons of oil-equivalent per hectare per year. Projections also indicate that at such a yield, production costs are in the brackets of present oil prices.

The amount of land actually available for energy forests has been hotly contested—with the government insisting on higher figures and the forestry industry and others reducing them. In any event, Sirén figures that about 1.5 million hectares—taking into account abandoned farmland, previously exploited peat bogs and derelict "waste" land—could realistically be employed to grow energy. This adds up to about 3.5 percent of Sweden's total land area, or about 50 percent of the land currently put to the plough.

Growing high-yielding energy plantations might very well be likened to a large-scale successful agribusiness operation. Machines will plant the clones only after the soil has undergone proper weeding and structural and nutritional preparations, as well as a water regime control. The plots will then be harvested mechanically every two to four years. Right now, there are no commercially available harvesters for working energy plantations, but two prototype harvesters will be tested out in the winter of 1984-85.

"Our fertilization program is carefully monitored and is based on the idea of recirculation of the nutrient elements, as well as avoiding leakage to the environment," comments Sirén. After a high yield has been obtained from a carefully fertilized stand, the harvest is used as fuel and the ash is returned to the soil. "Because of the ash return," explains Sirén, "only minor quantities of nutrients other than nitrogen are needed, in addition to the litter. Nitrogen losses to the soil are made up for by adding appropriate quantities of calcium and ammonium nitrate, or by using nitrogen-fixing alders as a mix or in pure stands. In short, we are applying intensive agricultural techniques on crops of woody plants."

For energy forests to work, the stumps must continue to sprout after every harvest year after year. So far, so good. Sirén's clones seem to be among the fastest in the West. The growth record was captured in 1982 when one shoot shot up to 4.7 meters. According to Sirén, "sample plots with an average height of six to seven meters are frequent in two-year-old stands."

Within the framework of Sweden's overall energy picture, fast-growing forests will possibly contribute up to 15 terawatt-hours (TWh) per year to the budget by the turn of the century.

This view is based on the assumption that the test results so far obtained can be duplicated commercially. Four large-scale test projects are under way to prove that they can. But a number of problems must be overcome before biomass becomes commercially viable. Very little is known, for example, about the potential threats to energy monocultures from insects, moose, field mice and a wide variety of pest organisms, including fungi, parasites, bacteria and viruses.

The energy forest program has also come under fire from environmentalists claiming that the amount of fertilizers used in some stands was excessive. Complains Professor Lars Emmelin, Course Director of the Environmental Studies Program at Lund University, "If nitrogen fertilizer has to be applied in amounts of about 200 kilograms per hectare, the risks of contamination of ground and surface water is considerable."

Sirén, aware of the problem, is attempting to overcome it by using nitrogen-fixing alders, among other things. Alders are also attractive because they seem to grow well on waterlogged soils, precisely those most susceptible to contamination.

As for diseases, Sirén calculates on an attrition rate of 15 percent per stand. "Pesticides will not be used to control either pests or diseases," he affirms.

The government will attempt to sort out fact from fiction and hopefully give a go-ahead to the further exploitation of Sweden's "green oil."

Peat: Back to the Bogs

What Sweden lacks in fossil fuels—the country has none, with the exception of small deposits of a practically useless claylike coal—is richly compensated for by large tracts of wooded peat land. Dr. Hugo Sjörs, Professor of Ecological Botany at Uppsala University, sets the total area of peat land in Sweden at 5.4 million hectares, or about 14 percent of the entire country.

However, out of this impressive figure only a slice of it may prove to be economically exploitable. Many of Sweden's peat bogs are simply too inaccessible at any price, some are found within nature preserves or other protected regions and still others contain such thin, poor-quality peat that it is not worth excavating.

In a controversial report released in 1980, the Swedish Geological Survey put peat resources at about 10 billion tons of dry matter (eight billion tons of this is in northern Sweden), which figured out to about four billion tons of oil equivalent. It has been calculated that if all the peat could be utilized for energy production, it would free the country from oil for 100 years.

The harsh reality is that out of this reserve of potential peat power, perhaps only 5 percent of the total can be dug up and used as fuel. Nevertheless, the Energy Administration estimates that there are about 50,000 hectares of easily exploitable bogs which should produce around 10,000 TWh/y.

Peat-power proponents such as Dr. Olle Lindström at the Department of Chemical Technology of the Royal Institute of Technology in Stockholm are convinced that peat can be a major source of energy by the turn of the century.

"Because of more favorable and realistic cost-benefit analyses, peat remains an attractive alternative source of energy," maintains Lindström. "Peat is already economically competitive with coal in Sweden, but the problem is, we have a long tradition of burning coal and very little experience with peat for energy purposes."

The number one problem in utilizing peat as a fuel is how to dewater it. Even processed milled peat still contains about 50 percent water. New processes are being developed that dry the peat and form it into pellets or bricks for easier transportation and end-use. PLM Environmental Systems in Malmö and Fläkt in Stockholm, for example, have developed flash driers and other processes that dewater biomass-based fuels, forming them into chips, powder, bricks or pellets.

In order to push peat R&D activities a little harder, the government intends to establish a Peat Research Institute in Umeå, northern Sweden, with funds from both the public and the private sectors. ■

Today, the screws are being tightened on environmental standards in many countries, and the U.S. is no exception. While virtually all major fossil fuel-fired power plants are equipped with filters for collecting fly ash from the flue gases, many still lack desulfurization equipment. The result is an increasing blight of acid fallout, both wet and dry, across large areas of North America (especially along the eastern seaboard) and in Northern Europe. Acid deposition kills freshwater fish like salmon and trout, and damages crops and forests. Corrosion damage to buildings and metal surfaces caused by sulfur gases from power plants and industrial processes costs western Europe about $40 billion a year and pollution damage to U.S. crops amounts to an estimated $2 billion a year.

Fläkt, Inc., a subsidiary of the giant Swedish company Fläkt AB, offers a wide range of pollution control equipment: from sulfur scrubbers, electrostatic precipitators and fabric filters to air conditioning and ventilation equipment, waste handling systems using compressed air, heat recovery and heat exchangers, and anti-stall industrial fans, among many others.

In order to reduce these tremendous costs the installation of proper environmental control technologies, such as Fläkt's Dry Flue Gas Desulfurization (FGD) is necessary. This system is based around two primary items for cleaning the boiler off-gases. This would consist of spray dryer reactor(s) followed by a particulate collector of either an electrostatic precipitator or a fabric filter.

The first stage is a dry SO_2 scrubber (reactor) where absorbent lime is sprayed into the hot flue gases as they pass through the reactor. As the lime slurry is absorbed by the exhaust gases, the calcium in the lime binds with the sulfur dioxide, is dryed by the hot gases and passed on with the fly ash to the particulate collectors. To facilitate proper sulfur removal and to insure a dry product the temperature within the reactor is controlled via the water content of the lime slurry.

According to Fläkt's engineers, "for coal with a sulfur content of up to 1.5 percent this method will pick up 80-90 percent of all sulfur emitted in the combustion process."

The second stage can consist of an electrostatic precipitator to collect the combined fly ash lime dry product prior to exit to the atmosphere. In this process the dust particles are passed through strong electrical fields thereby charging the particles. These particles then migrate to oppositely charged collecting electrodes where they are deposited.

Alternatively a fabric filter may be used for the second stage. In this case the particles are collected on the insides of thousands of large (12″x30′ long) cylindrical fabric bags housed in various compartments within a massive steel structure.

In either case the particles are eventually deposited into hoppers when the dust is transmitted to storage areas for later disposal or selling. Both devices provide 99 percent plus dust removal from the gas stream thereby insuring clean gas is emitted from the power plant stack.

Fläkt's Dry FGD plant at the Grand River Dam Authority's 520 MW coal-fired electricity generating station near Pryor, Oklahoma will use four dry scrubbers and two electrostatic precipitators when it is completed in 1985. The $70 million project is the largest single contract ever awarded to Fläkt's U.S. subsidiary. It also includes a lime unloading, storage, transfer, slaking and slurry pumping system and a complete microprocessor-based instrumentation and control system. The system will treat an incoming gas stream at 300 degrees F., remove 85 percent of the sulfur, 99.7 percent of the particulate and exit a 170 degrees F. cleaned gas to the atmosphere.

Other Products

Fläkt's new modular KDA air handling unit with heat recovery, is another product that has seen dramatic sales increases in the U.S. market. KDA has an air flow range of up to 72,000 cubic feet per minute and it comes in 12 sizes with over 24 different unit sections available. The units are produced in Fläkt's modern manufacturing facility in Winston-Salem, North Carolina.

Fläkt's breakthrough with its new anti-stall axial flow industrial fan, is yet another example of innovative development. The novel design provides stable fan performance regardless of system resistance and prevents "stalling"—a problem which invariably limited the operating range of axial flow fans. Stalling increases energy consumption, noise levels and the risk of mechanical breakdown.

The company also excels in ventilation and air conditioning equipment, designed for rigorous environments.

Fläkt's DIRIVENT ventilation system has been installed in over 2,000 work environments around the world; from machine shops and department stores to sports facilities and warehouses. In buildings with high ceilings, the heated air rises to the top, leaving pockets of cooler air close to the floor where people work. Fläkt's system uses small jets of air to direct pretreated ventilation air horizontally as well as vertically to wherever it is needed.

Additionally, Fläkt supplies and installs up-to-date ventilation and air conditioning systems for the textile, tobacco, and nuclear utility industry through its Bahnson Division. Bahnson further has a specialized line of textile cleaning equipment that has been the recognized leader for many years in the textile industry. ■

Fläkt's DRYPAC system is being employed at the Grand River Dam Authority's coal-fired power station near Pryor, Oklahoma.

Sweden's Solar Drive

In their epic book *Solar Sweden,* the Swedish physicists Peter Steen and Thomas B. Johansson noted that Sweden could be run completely on renewable energy sources by the year 2015. They also calculated that Stockholm, for example, receives each year 1,500 to 2,000 hours of sunlight, or about 40 percent as much as the Sahara Desert. The problem, of course, is obvious. Whereas North Africa has constant sun the year round, Sweden's incoming solar energy per square meter is 30 times greater in the summer than in the winter. This means that the country must be able to store solar energy from summer to winter.

Sweden, in attempting to tap solar radiation directly by using solar panels to provide space heating and hot water for homes, offices and apartments, has launched a surprising array of solar-related energy storage projects, most of which are under the direction of the Swedish Council for Building Research (BFR).

During the three-year period 1981–84, BFR devoted almost $7.4 million for solar heating technology out of a total budget for "energy use in buildings" of $51.2 million. When one keeps in mind that 40 percent of Sweden's total energy bill goes for space heating and hot water, reducing dependence on oil in this sector is a very important part of the overall energy plans.

"We have three times as much money now as we did five years ago," comments Ingrid Munro, director general of BFR. "About half of our money goes into developing new energy techniques (e.g., solar heating) along with energy-saving technical fixes, such as better insulation materials."

BFR is focusing its solar storage projects in three areas: storage in the ground (rock or clay), in water and in underground caves.

The government-owned company Studsvik Energiteknik AB, in cooperation with BFR, announced in 1978 that they had developed the first solar energy system for central heating that featured (at that time) season-to-season storage. The prototype plant, operating since 1979, supplies a 500-square-meter office building at Studsvik with its total yearly heating requirements (about 22.5 MWh/yr).

Sweden's geology grants the country favorable rock conditions for energy storage in the ground. Methods for such storage are now known, and assessment is being carried out in several experimental installations. Borehole heat stores in rock and clay, excavated pit heat stores, rock cavern heat stores and aquifer heat stores are all under consideration. Using present-day methods, it should be possible to build large-scale heat stores at a capital cost of less than 1 Skr/kWh/yr of storage capacity.

The Lyckebo Project

In the Uppsala suburb of Lyckebo, one of Sweden's most unique underground energy storage projects is taking shape. When it is finished in 1984, 550 dwellings will be heated by hot water stored in a 100,000 cubic meter cavern sunk 30 meters into solid rock. The water will be heated by 4,300 square meters of solar collectors and a 6-MW electric boiler.

The project contains some curious technological twists. Since the temperature of the solar energy "captured" by the fluid in the collectors varies greatly over the course of the day, it must be fed into the rock cavern at different temperatures. In order to do this, a sophisticated computer-controlled system was devised that will channel the heated water from the collectors into the corresponding temperature layer in the cavern, where water temperatures will vary from 90 degrees C. at the top to 40 degrees C. at the bottom. With such definite stratified temperature layers, the water can then be tapped at different temperatures, depending on the season. In summer, for example, the water will be drawn from the 60 degrees C. layer, whereas in winter it will be taken from the 75 to 90 degrees C. layers.

Not surprisingly, the 4,300 square meters of solar collectors are only sufficient to provide about 15 percent of all the hot water needs of the community (at a peak supply of 2.5 MW). The remaining 85 percent will be generated "simultaneously" in the 6-MW electric boiler during summer nights (when electricity is very cheap). This way the whole system acts like a 100 percent solar heated energy system.

Studsvik Energiteknik

Studsvik Energiteknik's experimental solar-heated hot water basin. The solar panels sit atop an insulated tank filled with solar-heated water for heating the office building in the background.

Conclusion

BFR predicts that by 1990 solar energy, most likely in combination with heat pumps and electric boilers, will provide nearly 1 TWh/yr of energy for space heating and hot water.

On the optimistic side, Ingrid Munro believes that up to one third of all newly constructed multifamily homes in Sweden will have solar-heated tap water by the end of this decade, corresponding to an installed capacity of 25,000 square meters per year of collectors. According to BFR, the same estimate applies to houses already built. To this can be added "5,000 to 10,000 square meters of collectors for villas," states a BFR report, giving a grand total of 50,000 to 60,000 square meters of installed solar collectors every year, or 25 to 27 GWh/yr.

This may sound like an overly optimistic assessment, but in reality solar is still a very small part of the overall picture. Solar has potential in Sweden because—in combination with other technologies—it can help to make reductions in the highest oil-consuming sector, space heating, where every calorie of energy saved or provided by a domestic source means cheaper oil bills. ■

Fluidized Beds: Boilers With More Bang

Conventional fluidized bed technology relies on what is termed a "bubbling bed." Now for the sake of simplicity, a bubbling bed can be compared to a pan of water—only a quarter filled—which is set on the stove to boil. Like the water in the pan, the bubbling bed has a well defined "bed" surface or area. When the water in the pan reaches the boiling point it bubbles violently and the water is rapidly converted into steam. The same principle applies to the bubbling bed, except the water is replaced by air and the bed material consists of sand, ash, limestone and a small amount of fuel. Combustion air is blown into the boiler through the base plate and loosens up the bed material until it "floats" on the turbulent air. The more violent this "boiling" in the bed becomes, the more turbulent the fluidization and gas bubbles shoot through the bed material (hence the term "bubbling bed.")

Continuing the analogy a little further, as the water in the pan will eventually boil away unless more is added, so too the bubbling bed can only be maintained if a small but constant amount of combustable fuel is added to the bed material. The height of the bed can be controlled by either removing some ash or adding more of it. The fluidized bed, like the water in the pan, must "boil" in order to perform its function—producing energy (as opposed to cooking food).

Bubbling bed processes burn fuel very efficiently, maintain constant combustion temperatures (usually around 850 degrees C.) and are much less polluting than standard boilers, since sulfur and nitrogen oxides are removed *in situ* as part of the combustion process. But conventional bubbling beds do have limitations: they are normally designed to burn only one type of fuel, like coal or oil (fuel switching is complicated and expensive); on the average, 10 to 15 percent of the fuel remains unburnt; they have poor "turn-down" properties (in other words, once in operation it is difficult to alter the fuel/combustion air ratio); and last, they are not as robust as pulverized coal burning boilers for industry or district heating, which is a well-established technology.

Three Swedish companies—Götaverken Energy Systems of Gothenburg, ASEA-PFBC (a joint company with ASEA-ATOM and Stal-Laval) and Stal-Laval, both headquartered in Finspäng—claim to have overcome these major commercial stumbling blocks with three important technological advances: Götaverken's Circulating Fluidized Bed (CFB), ASEA-PFBC's Pressurized Fluidized Bed Combustion (PFBC) and Stal-Laval's Multi-Bed Combustion (MBC).

All three systems have the following advantages over typical bubbling beds: With the exception of PFBC (which is designed to burn any kind of coal), each system burns a variety of fuels—high- or low-grade coal, wood chips or sawdust, peat and even refuse-derived fuels (RDF). The heat transfer to the tubes containing hot water or steam is much more efficient; their environmental improvements are manifold—sulfur retention is 85 to 95 percent, as compared with a maximum of around 80 percent for bubbling beds, and all three have significantly reduced NO_x emissions because of staged combustion; and finally, since the combustion process is more efficient, less fuel is used.

Here their main similarities end. In attempting to overcome the commercial limitations of conventional bubbling beds, each company has approached the problem from different angles.

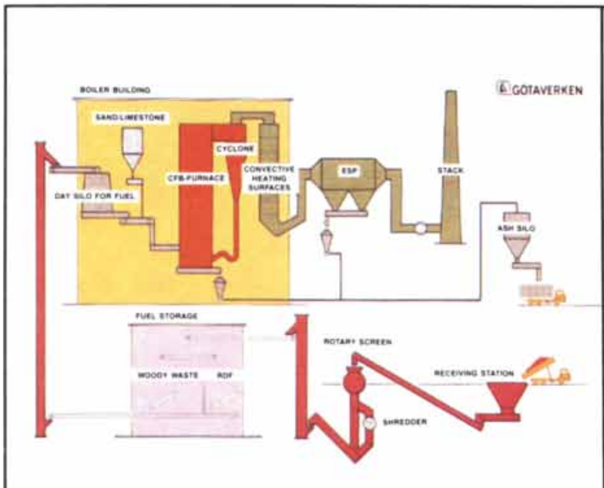
Circulating Fluidized Bed (CFB)

In Götaverken's CFB process there are two bed "regions": a bubbling bed at the bottom and a "suspension bed" on top of it. The bed particles, consisting of sand, limestone (dolomite) ash and a small amount of fuel, are constantly circulating between the two beds. This circulation results in better fluidization in the beds and a combustion efficiency of 99 percent. The excellent combustion properties in the CFB are due to the very intensive mixing of air and fuel and the internal recirculation of the unburnt fractions of the fuel by the cyclone.

The process can be likened to a Ferris Wheel at an amusement park. Combustion air is pumped in through the bottom of the boiler, while the limestone and fuel is fed in through the sides. The bed particles then move from the bottom bed to the upper one and are circulated around again via the cyclone.

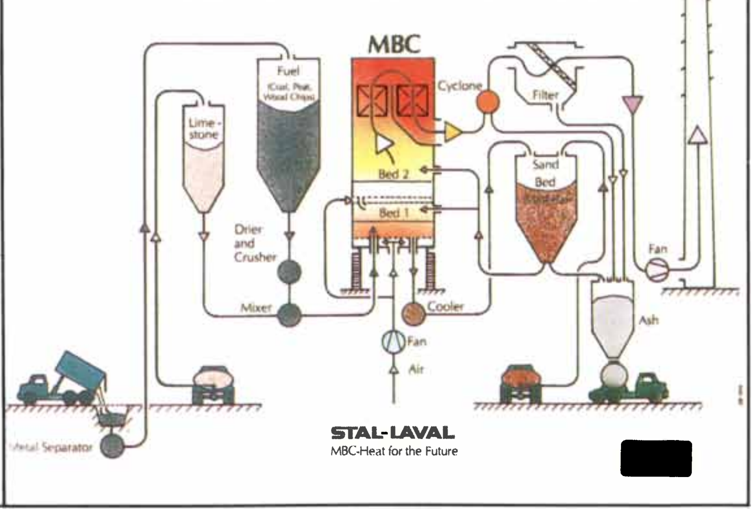
Through this dual process, the density of the particles in the suspension bed can be altered according to the type of fuel burned. Furthermore, the mean particle size in the bottom bed can be changed according to fuel type, which allows for better control of the firing temperature in the boiler, and in turn gives a better thermal efficiency and less pollution, especially from nitrogen oxides (NO_x) and sulfur. The combustion is easily controlled over a wide load range.

By using staged combustion, a reducing atmosphere is maintained in the



CFB-BOILER PLANT FOR RDF/WOOD WASTE
AT SUNDSVALL, SWEDEN

MBC Boiler Plant with Fluidized Bed Combustion



lower part of the boiler, thus reducing NO_x emissions from fuel-bonded nitrogen. An oxidizing atmosphere is maintained in the upper part of the boiler to complete the combustion process. Sulfur is removed by adding limestone or dolomite to the bed material, as mentioned previously. In this way, sulfur is removed as part of the combustion process and no other energy inputs are needed. The chemical bonding of sulfur to the calcium in the limestone forms gypsum, which can be sold to the construction industry as building material or as roadbed filler.

The flow diagram shows Götaverken's new CFB boiler plant, which is being built at Sundsvall, Sweden. This turnkey 20-MW plant will be attached to an existing cogeneration district heating/electricity generating plant. The CFB unit will burn refuse-derived fuel, wood waste and peat, and the resulting superheated steam will be used to increase the output of the cogeneration plant.

The Multi-Bed Combustion (MBC) Boiler

Stal-Laval's MBC boiler has a capacity of from 5 to 50 MW and consists of two bubbling beds, stacked one on top of the other, like a bunk bed. The bottom "bunk" is the primary combustion bed, and the top bunk acts as the secondary one. Just as with Götaverken's CFB technology, the main reason for the two-bed system is that more calories of energy can be tapped per kilo of fuel compared with more conventional methods. Another reason is its small, compact size. A 20-MW MBC unit takes up an area no larger than 25 × 25 meters, with a height of only 14 meters.

It works on a simple design principle: cooling pipes, which produce hot water or steam are embedded in sand and ash in the first bed. The fuel, say coal, along with limestone, is then fed into the lower bed either from the bottom, sides or even the top (depending on the type of fuel used). Like the other fluidized beds discussed in this section, only 1 percent of the total bed consists of fuel; the rest is sand, ash and limestone (or in some cases dolomite, although the MBC uses only limestone). The combustion air enters the boiler through a perforated plate under the bed material of the first bed and mixes with the coal and limestone, creating a very efficient heat transfer to the cooling pipes. In theory, at least, what combustible particles escape the first bed are burned up in the second. Comments Lars-Olof Ingesson, head of market research and product development for the MBC, "The amount of unburned particles is only 2.5 percent, and even if the boiler is running at only 30 percent of its load capacity, the combustion efficiency is 95 to 96 percent" (*see diagrams*).

Like most fluidized beds, it must be preheated with oil or gas before the ignition temperature of 850 degrees C. is reached. However, it can be fine-tuned to take different fuels, including peat and wood wastes. According to Stal-Laval, three different techniques can be used to adjust the bed characteristics to control the energy output: taking out some of the bed material, controlling the temperature and turning the boiler on and off. The boiler load can be adjusted by taking out bed material; controlling the temperature in this way, the load can be adjusted down to 30 percent. The load can be adjusted further—down to 5 percent—by turning the boiler on and off.

Flue gases from the primary and secondary combustion beds are finally cooled 150 to 200 degrees C. in devices called economizers placed above the beds. Stal-Laval says that the high degree of heat transfer in the beds results in the need for smaller heating surfaces and permits a more modest boiler volume.

When coal is incinerated, it produces polluting oxides of sulfur. The limestone binds the sulfur oxides forming calcium sulfate, which can then be removed together with the dust and ash via a cyclone and a filter. The process captures 85 to 90 percent of all the sulfur released during combustion, and only around 200 ppm (parts per million) of nitrogen oxides (NO_x) are released into the atmosphere, half the permitted level.

The new boiler operates automatically in a completely closed system. A 10-MW reference unit for district heating purposes—to be fueled with coal, peat and wood wastes—has been ordered by the town of Nyköping, south of Stockholm.

Pressurized Fluidized Bed Combustion (PFBC)

ASEA-PFBC technology is nothing less than a breakthrough. For the first time, a fluidized bed can actually be used on a large scale to generate electricity, as well as hot water or steam for district heating. PFBC is perhaps most promising as a technology that will be used in cogeneration plants, producing both electricity and hot water for district heating, as well as in condensing power plants.

All fluidized beds are designed for optimum combustion efficiencies, but scaling bubbling fluidized beds up to the sizes required for electricity production was considered uneconomical, too complicated and too expensive. Indeed, it took a truly multinational effort by Stal-Laval involving ASEA-ATOM, Deutsche Babcock and the American Electric Power Corporation to develop and demonstrate the PFBC boiler.

Essentially, Stal-Laval engineers have put a bubbling bed inside a chamber pressurized to 16 bar (atmospheres). A Stal-Laval GT-120 gas turbine, which runs on cleaned flue gas from the boiler, drives a compressor that supplies the fluidized bed with pressurized combustion air. Coal and dolomite are fed into the bed through the bottom plates, along with combustion air (no sand is needed for the bed). The coal particles along with granular inert ash particles and dolomite are kept in a fluidized suspension by the turbulent upward flow of air. Thanks to the pressurization, the combustion takes place over a large volume and is very efficient—99 percent of the fuel is consumed. Most of the thermal energy released is transferred to the water and steam in the tube coils distributed throughout the fluidized bed. The steam that is generated drives a steam turbine (also made by Stal-Laval) that produces electricity. (The turbine can also produce hot water or steam for district heating.)

The flue gases are cleaned in three cyclone stages to remove most of the fly ash before it goes to the gas turbine. Apart from the compressors, the gas turbine also drives an alternator. The heat remaining in the flue gases is transferred to the feed water. Finally, the exhaust gases are led through a fabric filter before being discharged up the stack.

The PFBC unit described here is designed to produce up to 330 MW of electricity. (However, the unit could just as well be plugged into a district heating network.)

Since dolomite is added with the fuel, most of the sulfur is captured by the calcium in the dolomite. "Tests carried out in our Component Test Facility (CTF) in Malmö demonstrated that the PFBC could even capture 93 percent of the sulfur in coal from the eastern U.S. with a sulfur content of 4.7 percent," exclaims ASEA-PFBC's president, Carsten Olesen. "And the NO_x emissions are also well below permitted levels." ■

For further information on energy developments in Sweden contact:
M. Höjeberg
National Energy Administration
S-117 87 Stockholm
Tel: (8) 74 49 500
Telex: 12870 energy

Today:

The great name in optics and electronics for microsurgery.

Microsurgery. The very word evokes images of medical miracles. Corneal transplants that restore vision. Intricate procedures that reinstate hearing, re-attach severed digits and limbs, even help "barren" couples conceive.

Yet microsurgery as a viable discipline is a mere thirty years old. Zeiss launched it, and virtually every microsurgical procedure in use today has been developed with a Zeiss Operating Microscope—an OpMi.

Recently, in an eminently logical extension of OpMi utility, Zeiss integrated a CO_2 Surgical Laser with a Zeiss Floor Stand to be used in conjunction with a Zeiss OpMi. The system introduces new versatility, compactness, and mobility to this rapidly expanding surgical technique.

Nor is Zeiss' involvement with medicine restricted to the operating room. Zeiss also manufactures an extensive line of instruments for medical examination, diagnosis and research.

The Zeiss Fundus Camera, for example, enables the physician to photograph the back of the eye to detect diabetes and vascular diseases early on. With the Zeiss Slit Lamp, the front of the eye can be examined and photographed, an invaluable advantage in diagnosing corneal diseases and in fitting contact lenses.

Zeiss optical and electron microscopes are indispensable to the clinical pathologist in examining blood, tissue, and urine samples, and in the detection of infections and cancer.

Zeiss' triple skill in optics, mechanics, and electronics is tapped in allied sciences and industry for the tools and techniques needed in diverse disciplines. Currently representative are: computer numerical controlled measuring equipment, computerized surveying and mapping, microscopy and image analysis, electro-optical engineering and design, and products for sight.

And if that's Zeiss today...

Who knows what tomorrow will bring?

Carl Zeiss, Inc.
One Zeiss Drive
Thornwood, NY 10594
(914) 747-1800

Who knows what tomorrow will bring?

ZEISS

The Proteins of Oncogenes

The genes that cause cancer are altered versions of normal genes encoding proteins with important roles in normal cells. Oncogenic proteins induce cancer by partially mimicking the normal proteins

by Tony Hunter

Oncogenes are genes that cause cancer. They are altered versions of ordinarily benign genes present in normal cells. What makes them oncogenic? Like other genes they encode proteins, the fundamental substances that provide the structure of the living cell and catalyze its biochemistry, thus determining its form and function. The proteins encoded by oncogenes function abnormally and somehow ordain the transformation of a normal cell into a cancer cell. What are these oncogenic proteins and how do they transform cells? A decade's intensive study is beginning to reveal common principles underlying the functions of transforming proteins, suggesting that a limited number of mechanisms may account for carcinogenesis and that one day the mechanisms may be understood.

Oncogenes have been brought to light by two disparate but converging lines of investigation. The first was a study of retroviruses: viruses whose genetic material is not DNA but RNA, which is transcribed "backward" into DNA when the virus infects an animal cell. Some retroviruses carry among their genes a single gene responsible for rapid oncogenesis. Such a gene was first identified in the early 1970's in Rous sarcoma virus (RSV), which causes cancer in chickens; the gene was named *src,* for sarcoma. (Other oncogenes have since been designated by analogous three-letter abbreviations.) Soon the protein the *src* gene encodes was isolated by Raymond L. Erikson and Joan S. Brugge, who were then at the University of Colorado School of Medicine; it was designated p60*src* because its molecular mass is 60,000 daltons.

In 1975 J. Michael Bishop and Harold E. Varmus of the University of California School of Medicine in San Francisco found that *src* is not truly a viral gene. It is a nearly exact copy of a gene found in all chicken cells. The normal "proto-oncogene" was picked up by a weakly oncogenic retrovirus in the course of infection, and in the process it somehow became a cancer gene [see "Oncogenes," by J. Michael Bishop; SCIENTIFIC AMERICAN, March, 1982]. By now a score of oncogenes have been isolated from retroviruses that variously cause carcinoma, sarcoma, leukemia or lymphoma in chickens, other birds, rats, mice, cats or monkeys. In each case the oncogene has been found to be closely related to a normal gene in the host animal and to encode an oncogenic protein similar to a normal protein.

In the past five years a second line of investigation has also begun to uncover oncogenes, but in human and animal tumor cells rather than in retroviruses. Robert A. Weinberg of the Massachusetts Institute of Technology, Geoffrey M. Cooper of the Dana Farber Cancer Institute, Michael Wigler of the Cold Spring Harbor Laboratory and Mariano Barbacid of the National Cancer Institute and their colleagues independently found genes in the DNA of various kinds of tumor cells that, on being introduced by "transfection" into normal cultured cells, transform them into cancer cells. These oncogenes too are virtual copies of proto-oncogenes. Presumably (although this has not yet been shown directly) each oncogene was instrumental in generating the tumor from which it was isolated.

A Single Pool

The human tumors that yielded oncogenes are not induced by retroviruses, and so there was no reason to expect the tumor oncogenes to be the same as the retroviral ones. Yet it has turned out there is some overlap between the two groups. The most frequently isolated tumor oncogene has proved to be a homologue of the retroviral oncogene *ras,* which is carried by the Harvey strain of murine sarcoma virus. The cellular genes homologous to several other retroviral oncogenes have been found in an altered form in individual human tumors, although in these cases the altered genes have not been shown directly to be oncogenic. It may be that retroviruses and transfection experiments have exposed to view members of the same pool of normal genes that are sometimes activated to become cancer genes.

The pool may not be large. Several oncogenes have been isolated more than once, suggesting that the number still to be discovered may be small. Of the total complement of human genes, estimated to be some 30,000, perhaps fewer than 100 will prove to be proto-oncogenes. What they have in common is that they appear to have been highly conserved in the course of evolution. For example, the proto-*src* gene is found not only in all vertebrate animals but also in the fruit fly *Drosophila;* the proto-*ras* gene is present in species ranging from yeasts to man. This high degree of conservation implies that the proto-oncogenes play crucial roles in the cell.

How do proto-oncogenes become oncogenes? There is evidence for several mechanisms. In some cases an oncogene has undergone mutation, presumably by a carcinogenic agent (most carcinogens are known to be mutagens), so that the protein encoded by the mutant oncogene is slightly different from the normal protein. For example, Weinberg and Barbacid found that the sequence of nucleotides (the subunits of DNA and RNA) in the *ras* oncogene they isolated differs by only a single critical nucleotide from the sequence of the corresponding proto-oncogene [see "A Molecular Basis of Cancer," by Robert A. Weinberg; SCIENTIFIC AMERICAN, November, 1983]. Each three-nucleotide codon in a DNA sequence specifies a particular one of the 20 amino acids that are constituents of proteins. A protein's shape and function are determined by the precise order of its amino acids, and so changing just one of them can have dramatic consequences. The mutation in the *ras* gene alters one amino acid of the gene's product and thereby converts a normal protein into an oncogenic one.

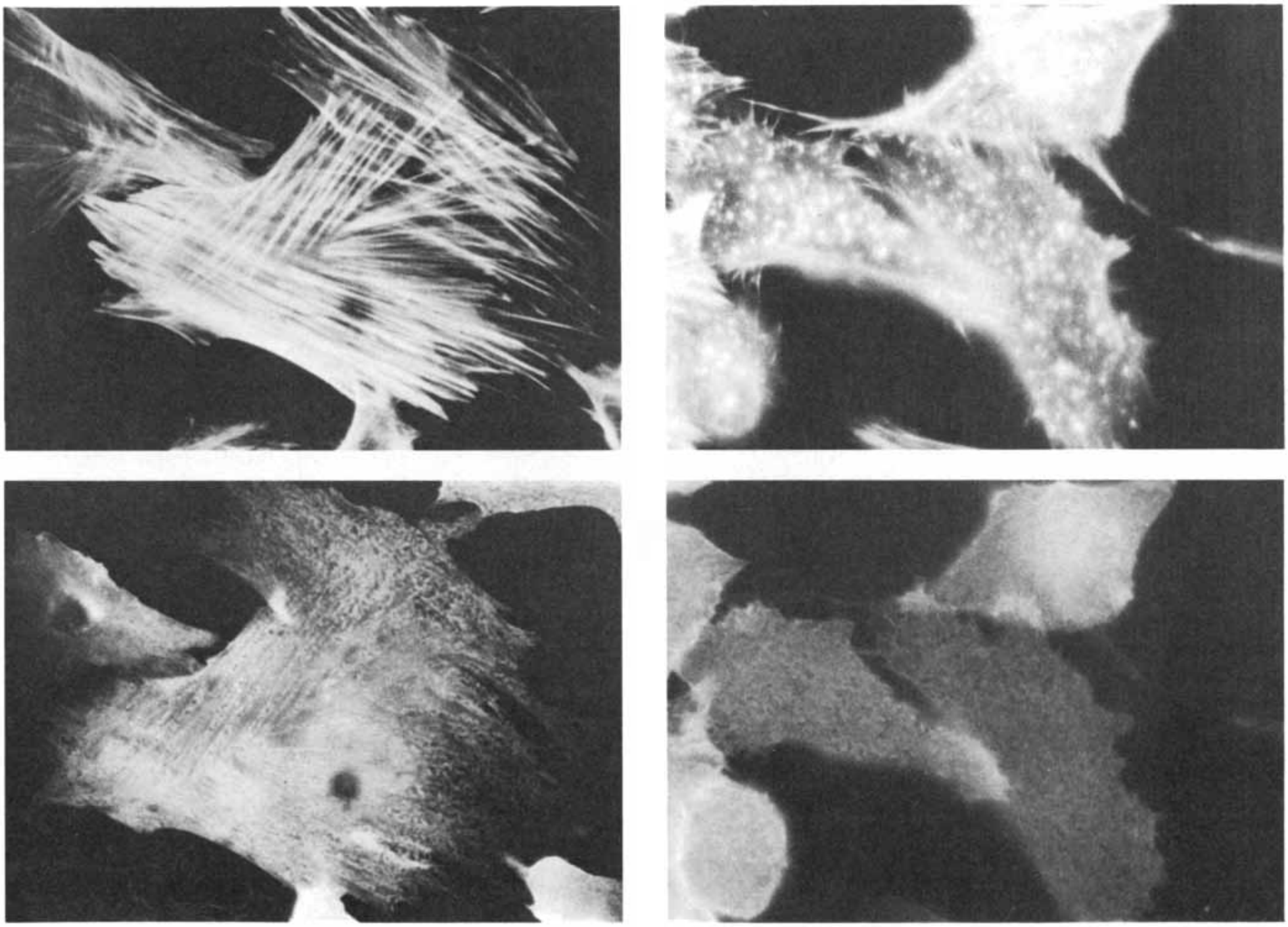
Activation may also involve a "dosage" effect. The quantities of some oncogenic proteins are very high in transformed cells compared with the levels of their counterparts in normal cells. In tumor cells this can result from a failure of gene regulation or from the presence of multiple copies of the gene. In virally transformed cells the oncogenic protein is made in large amounts because the oncogene functions like an active viral gene. In other cases a gene is expressed (translated into protein) inappropriately, either in the wrong kind of cell or at the wrong time in the cell's life cycle.

Protein Kinases

Whatever the specific mechanism converting a proto-oncogene into an oncogene may be, an oncogene must finally have its effect by way of the protein it encodes. Much of what is known about these proteins stems from the work of Erikson and Marc S. Collett. When they isolated p60*src* from RSV-infected cells in 1978, they found it displayed enzymatic activity: it catalyzed the addition of a phosphate molecule to other proteins. This was a highly suggestive finding because phosphorylation was already known to be an important regulator of protein function. The enzymes that phosphorylate proteins are called protein kinases (from the Greek *kinein,* "to move"). They transfer the energy-rich terminal phosphate group of adenosine triphosphate (ATP), the cell's major energy carrier, to the protein being modified.

When p60*src* was found to be a protein kinase, several other such enzymes were already known. It was assumed p60*src* would be like them in that it would add phosphate to one of two amino acids, serine or threonine, that have hydroxyl (OH) groups to which the phosphate can be linked. It therefore came as a surprise when Bartholomew M. Sefton and I discovered that p60*src* instead phosphorylates tyrosine, the only other amino acid that has a hydroxyl group. At the same time Owen N. Witte and David Baltimore at M.I.T. reported that P120*abl,* the transforming protein of a mouse leukemia retrovirus, also adds phosphate to tyrosine. At first the phosphorylation of tyrosine seemed to be a unique property of the two oncogenic proteins, but we soon learned that all normal cells have small amounts of phosphotyrosine in their protein, implying that normal cells too have tyrosine-specific protein kinases. The first such enzyme we identified in a normal cell was none other than the product of the proto-*src* gene.

The finding that p60*src* and its analogue in normal cells have a similar enzymatic activity illustrates an important principle: a protein encoded by a retroviral oncogene almost invariably has a function similar to that of its cellular counterpart, and it may therefore act by mimicking the normal protein. The fidelity with which p60*src* imitates its normal counterpart should be a clue to



ONCOGENIC PROTEINS transform normal cells into tumor cells. These fluorescence photomicrographs made by G. Steven Martin and Celeste Carter of the University of California at Berkeley show the effect on chick-embryo fibroblasts of the protein p60*src,* the product of the *src* oncogene carried by the Rous sarcoma virus. Normal cells (*left*) and cells transformed by *src* (*right*) were rendered permeable and then treated with phalloidin, a mushroom toxin that binds to filaments of the cell-skeleton protein actin (*top*), and with an antibody to the protein p36 (*bottom*); the phalloidin and the antibody were labeled with different fluorescent dyes. Whereas bundles of actin filaments crisscross the normal cells in regular arrays (*top left*), in the transformed cells (*top right*) the actin bundles have broken down, accounting in part for the changes in shape that are characteristic of tumor cells, and the actin is redistributed into small aggregates. The protein p60*src* is a protein kinase: an enzyme that adds a phosphate group to other proteins. One of its targets is p36. The p36 is here shown to be localized in a sheet just under the plasma membrane in both normal cells (*bottom left*) and transformed cells (*bottom right*).

whether it is essentially an unaltered cellular protein (present, perhaps, in inordinately large amounts) or is significantly changed and acts differently. I shall return to this question below.

A cancer cell differs in many ways from a normal cell of the same type. How might the addition of phosphate to tyrosine in a cell's proteins cause the varied effects of transformation? In principle p60*src* could link phosphate to a number of substrate, or target, proteins, altering the function of each one. The multifaceted nature of the malignant state might then be due to the presence of a substrate protein in each of several cellular pathways. What role might be played by the phosphorylation of tyrosine in particular? As will become clear below, there are reasons to think tyrosine phosphorylation may be intimately involved in complex regulatory systems that maintain cellular shape and control the growth of cells. Subtle alterations in those systems could give rise to the cancerous state. So far, however, it is not clear why the phosphorylation of serine or threonine could not have similar effects.

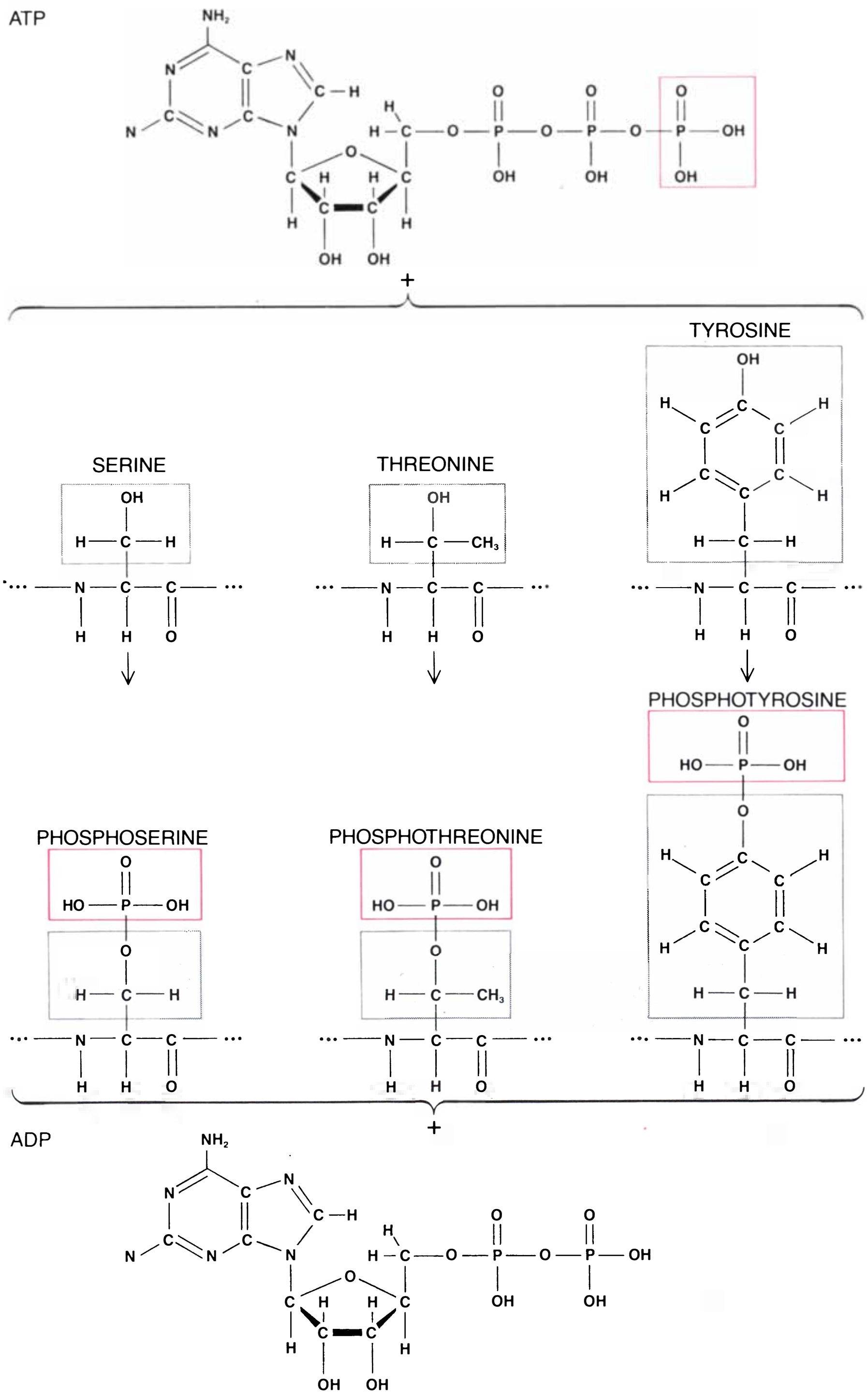
The Protein-Kinase Domain

The protein products of the known oncogenes have all been tested for tyrosine-specific protein-kinase activity. Six of them, in addition to p60*src,* were shown to have this enzymatic property (the products of the *yes, fgr, abl, fps, fes* and *ros* oncogenes), whereas the others did not display any detectable tyrosine-specific protein-kinase activity. Molecular-cloning techniques have made it possible to determine the nucleotide sequence of almost every viral oncogene and, by applying the genetic code, to predict the amino acid sequence of its protein product. A comparison of these sequences reveals interesting similarities. A stretch of some 250 amino acids near one end of p60*src* is known to be the protein-kinase domain responsible for catalyzing the transfer of phosphates. A recognizably similar domain of about the same length is found in each of the other six oncogenic proteins shown to have tyrosine-specific protein-kinase activity.

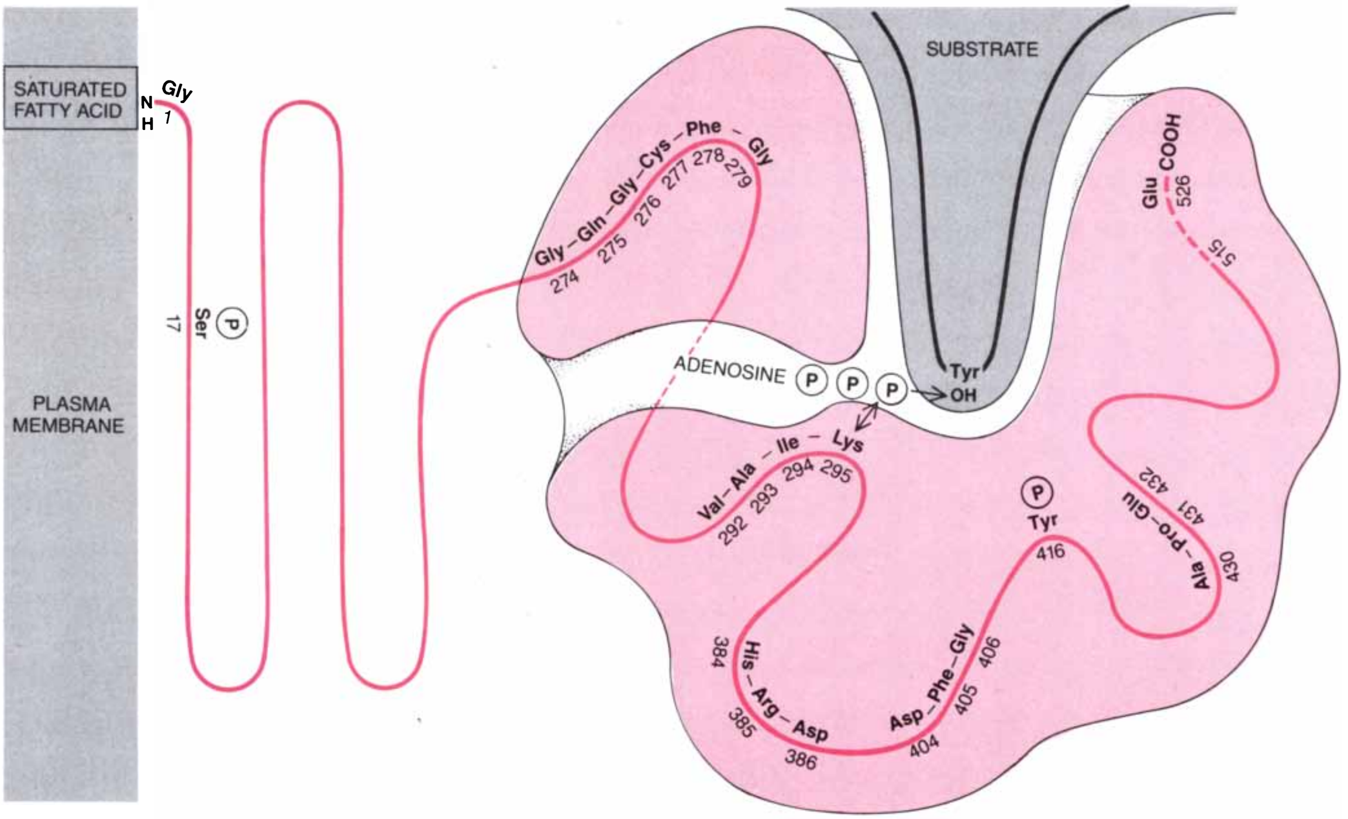
The striking similarity of the catalytic domains in the various enzymes suggests they may have had a common evolutionary progenitor. Evidence that the domains are part of a still larger family came from a broad computer search for homologies between the amino acid sequence of p60*src* and the sequence of all other proteins on file. The search showed that the only serine-specific protein kinase whose sequence is known (cyclic-AMP-dependent protein kinase) is related to p60*src*. In other words, not only is there a family of tyrosine protein kinases but also there may be a superfamily, comprising all protein kinases, encoded by genes descended from a single antecedent gene.

When the sequence of oncogene products other than the seven with clear-cut protein-kinase activity is examined, five of them prove to have sequences related to that of the tyrosine protein-kinase domain. The meaning of this kinship is not yet fully understood, but we surmise that these proteins (the products of the *erb-B, fms, raf, mil* and *mos* oncogenes) have functions similar in some sense to protein kinases. Already there is experimental evidence that the *erb-B* protein may be able to phosphorylate proteins on tyrosine.

As I mentioned above, phosphoryla-



PROTEIN KINASES transfer the third phosphate (*colored box*) of the cellular energy carrier adenosine triphosphate (ATP) to a protein and thus alter the protein's activity. Most protein kinases phosphorylate proteins on one of two amino acids, serine or threonine, having a hydroxyl (OH) group on their side chain (*gray boxes*). The enzymes encoded by oncogenes such as *src* were found instead to add phosphate to tyrosine, the third amino acid carrying a hydroxyl group. The illustration shows the reaction catalyzed by a protein kinase: ATP plus one of the three amino acids yields a phosphorylated amino acid plus adenosine diphosphate (ADP).



PROTEIN ENCODED BY *src* ONCOGENE is a protein kinase composed of 526 amino acid subunits, designated p60*src* for its molecular mass of 60,000 daltons. The drawing shows how its main chain (*colored line*) is anchored at the amino (NH_2) end in a cell's plasma membrane and how the globular enzymatic domain might bring the third phosphate (*P*) of ATP into place for transfer to a tyrosine on a substrate protein. The positively charged lysine at position 295 is thought to be attracted to the negatively charged third phosphate. The named amino acids appear to be highly conserved, that is, they are essentially the same in analogous domains of p60*src* and six other oncogene-encoded protein kinases. There is one major difference between p60*src* and its normal counterpart: the amino acid sequence from position 515 to the carboxyl (COOH) end of the chain replaces a different sequence of 19 amino acids seen in the normal protein.

tion on tyrosine is a rare event in normal cells. Phosphotyrosine accounts for only one in 2,000 of the phosphate molecules linked to proteins. (Some 90 percent of the phosphate is coupled to serine and 10 percent to threonine.) In cells transformed by the viruses carrying oncogenes whose products have tyrosine protein-kinase activity there is a dramatic tenfold increase in the phosphotyrosine level—compelling evidence that the function the oncogene proteins display in the test tube is also manifested in the living cell. On the seemingly justifiable assumption that phosphorylation of cellular proteins on tyrosine is part of the mechanism whereby the viruses transform cells, a paramount requirement for understanding the mechanism is identification of the proteins that are targets for the protein kinases.

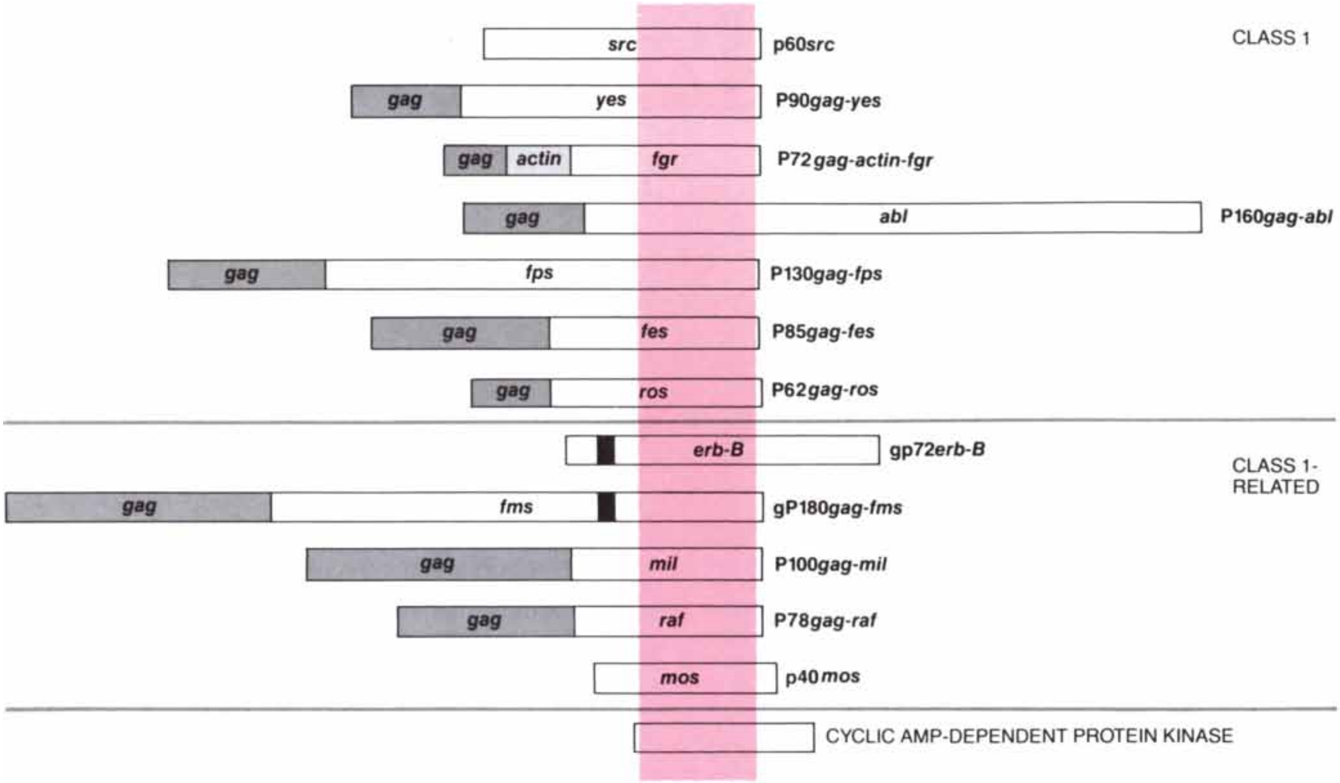
Target Proteins

What proteins in cells infected by the relevant retroviruses contain phosphotyrosine? Jonathan A. Cooper, Sefton and I have searched for such proteins in two ways. First we looked at proteins with a specific function whose alteration might explain some property of malignant cells. For example, the shape of a cell transformed by a tumor virus is very different from that of its uninfected parent cell. This suggests that the cytoskeleton, or internal framework, of the cell has been altered. A number of cytoskeletal proteins have been purified in the past decade. Might one of them be a substrate for the transforming tyrosine protein kinase? With S. Jonathan Singer of the University of California at San Diego we isolated 10 cytoskeletal proteins from RSV-transformed cells. Only the protein vinculin could be shown to contain any phosphotyrosine. Vinculin contains some phosphotyrosine in normal cells too, but the amount of phosphorylation is increased twentyfold in transformed cells.

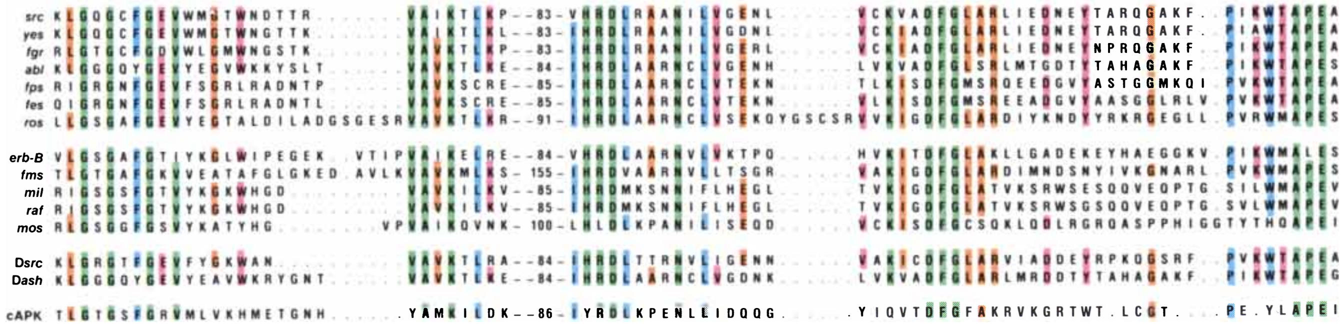
Vinculin is localized in adhesion plaques: small patches on the bottom of a cell that adhere closely to the surface on which it grows. Bundles of cytoskeletal filaments composed of the protein actin crisscross the cell and terminate at adhesion plaques. Vinculin is seen to be interposed between the cell's plasma membrane and the ends of the actin bundles; it may connect the bundles to a hypothetical anchor protein in the membrane. Larry R. Rohrschneider of the Fred Hutchinson Cancer Research Center in Seattle found that p60*src* is localized in adhesion plaques, in position to phosphorylate vinculin. The greatly increased phosphorylation of tyrosines in the vinculin of transformed cells might reduce its tenacity as a linker and lead to the release of the actin filaments, which are indeed radically disorganized in such cells. Some recent findings by Rohrschneider suggest, however, that vinculin phosphorylation alone may not be enough to cause such disruption. Phosphorylation of other proteins in the adhesion plaque may also be necessary.

Another characteristic of tumor cells is their tendency to secrete a large amount of lactic acid as a result of the increased breakdown of glucose by glycolysis, an anaerobic pathway with 11 sequential enzymatic steps. The activity of several of the glycolytic enzymes is increased in RSV-transformed cells, suggesting that the enzymes might be phosphorylated by p60*src*. Collaborating with Robert J. Schwartz and Nachum A. Reiss of Baylor University, Cooper and I isolated eight of the glycolytic enzymes from transformed cells. Three of them did contain phosphotyrosine: enolase, phosphoglycerate mutase and lactate dehydrogenase. Only a small fraction of the enzyme molecules, however, had a phosphorylated tyrosine. Moreover, none of these enzymes is rate-limiting for glycolysis; the key controlling enzyme, phosphofructokinase, does not appear to be phosphorylated on tyrosine. We therefore think tyrosine phosphorylation of the three enzymes may have little to do with the overall increase in glycolysis but rather reflects a breakdown in the specificity of phosphorylation resulting from an overabundance of p60*src* in the cell. Some more general cause than phosphorylation by p60*src* may explain the increase in glycolysis not only in RSV-transformed cells but also in other tumor cells. Glycolysis may simply be driven faster by an augmented supply of glucose, which is taken up more rapidly by tumor cells than by other cells.

Our second strategy for identifying target proteins is unbiased by any preconception of which proteins "ought" to be phosphorylated in tumor cells. We grow both normal cells and virally transformed cells in the presence of phosphate incorporating a radioactive isotope of phosphorus and spread out

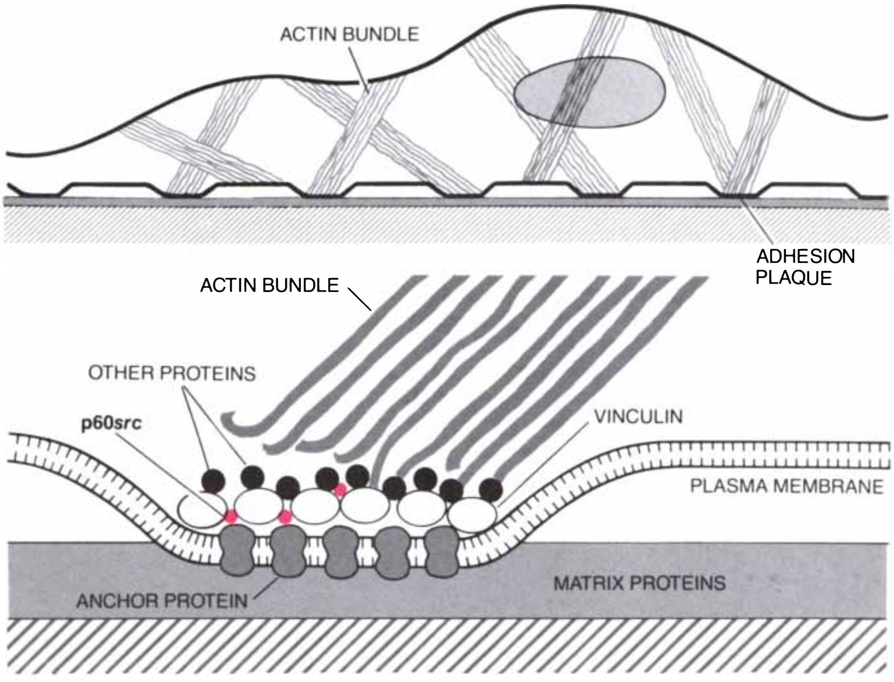


STRUCTURES of two groups of viral oncogene products and of a normal serine-specific protein kinase (*bottom*) are given by the bars, whose length is proportionate to the number of amino acids in each protein; the NH_2 end of each chain is at the left. The Class 1 products display clear tyrosine protein-kinase activity; the Class 1–related products do not. Analysis of the amino acid sequence of the proteins (predicted from the nucleotide sequence of their DNA) shows they have in common a 250-amino-acid region related to the protein-kinase domain of p60*src* (*color*). In most cases part of the protein encoded by the viral gene *gag* (and in one case part of actin) is synthesized with the oncogene protein as single product. Two proteins extending outside cell have identifiable transmembrane domain (*black*).



A ALANINE
C CYSTEINE
D ASPARTIC ACID
E GLUTAMIC ACID
F PHENYLALANINE
G GLYCINE
H HISTIDINE
I ISOLEUCINE
K LYSINE
L LEUCINE
M METHIONINE
N ASPARAGINE
P PROLINE
Q GLUTAMINE
R ARGININE
S SERINE
T THREONINE
V VALINE
W TRYPTOPHAN
Y TYROSINE

AMINO ACID SEQUENCES from within the protein-kinase domain of 12 oncogene products as well as from the *src* gene of the fruit fly *Drosophila*, an "*abl-src* homology region" (*ash*) of the *Drosophila* genome and the cyclic-*AMP*-dependent protein kinase are arrayed to show their similarity. The amino acids are represented by their one-letter codes (*left*). A stretch in the middle of each sequence has been omitted; the number of subunits omitted is indicated. Each dot represents a one-amino-acid gap introduced to attain the best alignment. The colored bars indicate the degree of similarity at the most highly conserved positions. A green bar means that the same amino acid appears at a given position in 14 or 15 of the 15 proteins; a blue bar indicates 12 or 13 identities, an orange bar 10 or 11 and a red bar eight or nine.



ADHESION PLAQUES anchor cells to surfaces and serve as internal anchors for bundles of actin filaments (*top*). The protein vinculin is localized in adhesion plaques, where it may link actin bundles to an anchor protein (*bottom*). In cells transformed by the Rous sarcoma virus, p60*src* (*color*) is also detected primarily in adhesion plaques and the vinculin is found to be phosphorylated on tyrosine. Phosphorylation by p60*src* may disrupt the vinculin link (*left*) and so contribute to the characteristic disorganization of actin bundles in transformed cells.

their proteins in a gel, exploiting a method developed by Patrick H. O'Farrell of the University of California School of Medicine in San Francisco. The technique separates the proteins in one dimension on the basis of their electric charge and in a second dimension on the basis of size. The 500 or so phosphoproteins in a cell are labeled by the incorporation of radioactive phosphate and can be detected as spots by autoradiography when an X-ray film is exposed to the gel.

To distinguish the few proteins that are phosphorylated on tyrosine from the plethora of other phosphoproteins we take advantage of the fact that a strong alkali tends to break the link between phosphate and serine but not the one between phosphate and threonine or tyrosine. Soaking the gel in alkali before autoradiography cleaves the labeled phosphate from many of the proteins containing phosphoserine, revealing a pattern of a manageable number of alkali-stable phosphoproteins. Some of them are present in higher amounts in virally transformed cells. When these proteins are cleaved to release their individual amino acids, a few of them are found to contain phosphotyrosine.

One phosphotyrosine-containing protein, so far known only as p36 (for its molecular mass of 36,000 daltons), is particularly prominent in gel analyses. It was first detected in RSV-transformed chick-embryo fibroblasts (connective-tissue cells) by G. Steven Martin and Kathryn Radke of the University of California at Berkeley, and it seems to be the major substrate for all the viral protein kinases. (As in the case of all substrates phosphorylated on tyrosine, only one particular tyrosine of the many in the protein is phosphorylated.)

p36 Function

The function of p36 in the cell remains a mystery, however; unfortunately one cannot assign a function to a protein on the basis of its size and charge alone. What is known is that p36 is associated with the inner face of the plasma membrane of many types of cultured cells, but it is not in all cell types in the body. Kathleen L. Gould and I have found, for example, that it is abundant in the intestine but essentially absent from brain tissue.

The distribution of p36 in intestinal cells may provide a clue to its function. The protein is concentrated in the columnar epithelial cells covering the surface of the villi, the fingerlike projections arrayed on the folded lining of the gut. The tip of a columnar cell is covered with a "brush border" of microvilli, each of which has a core bundle of actin filaments. The base of each bundle is anchored to a transverse filamentous structure called the terminal web; p36 is found predominantly in the terminal web, suggesting the protein may have a structural function. In other cell types it could have a similar role in skeletal structures just inside the plasma membrane. The membranes of RSV-transformed cells commonly have a more ruffled appearance than those of normal cells and are altered in other ways. The phosphorylation of p36 could be responsible for such changes.

The total amount of phosphotyrosine in an RSV-transformed cell is larger than the sum of the amounts present in the identified target proteins, indicating that yet other substrates for p60*src* remain to be discovered. Even at this stage, however, one can ask two pertinent questions. First, which of the target proteins is phosphorylated on tyrosine in cells transformed by oncogenes other than *src?* Target proteins common to many transforming protein kinases may be the most important ones for transformation. Second, which of the targets identified in transformed cells is phosphorylated on tyrosine in normal cells? This knowledge might indicate whether or not the viral tyrosine-specific protein kinases have their effect because they phosphorylate inappropriate substrates.

When we compare cells transformed by viruses carrying the *src, yes, fgr, abl, fps* and *fes* oncogenes, we find that by and large the same spectrum of proteins is phosphorylated in all of them, including vinculin, p36 and the three glycolytic enzymes. That is not very surprising considering the probable common ancestry of the catalytic domain in all six protein kinases, but it does reinforce the idea that the viruses in this group transform cells by a common mechanism involving phosphorylation on tyrosine. Nevertheless, it remains possible that the transforming proteins of these viruses may have some other capacity as critical for transformation as protein phosphorylation. Erikson, Lewis C. Cantley, Jr., and their colleagues at Harvard University have recently found that p60*src* can phosphorylate the lipid phosphatidylinositol, a major component of the plasma membrane. The phosphorylation is thought to accelerate the breakdown of phosphatidylinositol to diacylglycerol, which is known in turn to activate a serine-specific protein kinase. This might be another way p60*src* can alter cells.

With the exception of vinculin, the proteins phosphorylated on tyrosine in RSV-transformed cells are not detectably phosphorylated on tyrosine in normal cells. This suggests that p60*src* may generally choose inappropriate proteins as substrates because of either the large amount of p60*src* in an infected cell or a structural difference between p60*src* and its normal counterpart. There is indeed a significant difference at the carboxyl, or COOH, end of the two proteins. Experiments are under way to test whether the difference accounts for p60*src*'s transforming capacity. Preliminary results suggest it does. The viral tyrosine-specific protein kinases other than p60-

modern times
PC with you.

9″ monitor with easy-to-read amber characters. Text and graphics capability. And an 83-key keyboard.

All fitted into a sturdy, transportable case that's easy to handle. And park.

The IBM® of portable personal computers. Make no mistake about it, this is a true IBM PC.

Which means it is part of the same dependable family as the IBM Personal Computer, the IBM PC/XT and the new IBM PC*jr.* And that means you can use many IBM Personal Computer Software programs to help you reach your goals.

All this and five expansion slots, ready to accept expanded memory, printers and other useful IBM Personal Computer options. Which should keep you rolling far into the future.

Pick one up at a store near you. You can see the new IBM *Portable* Personal Computer at any authorized IBM PC dealer or IBM Product Center.

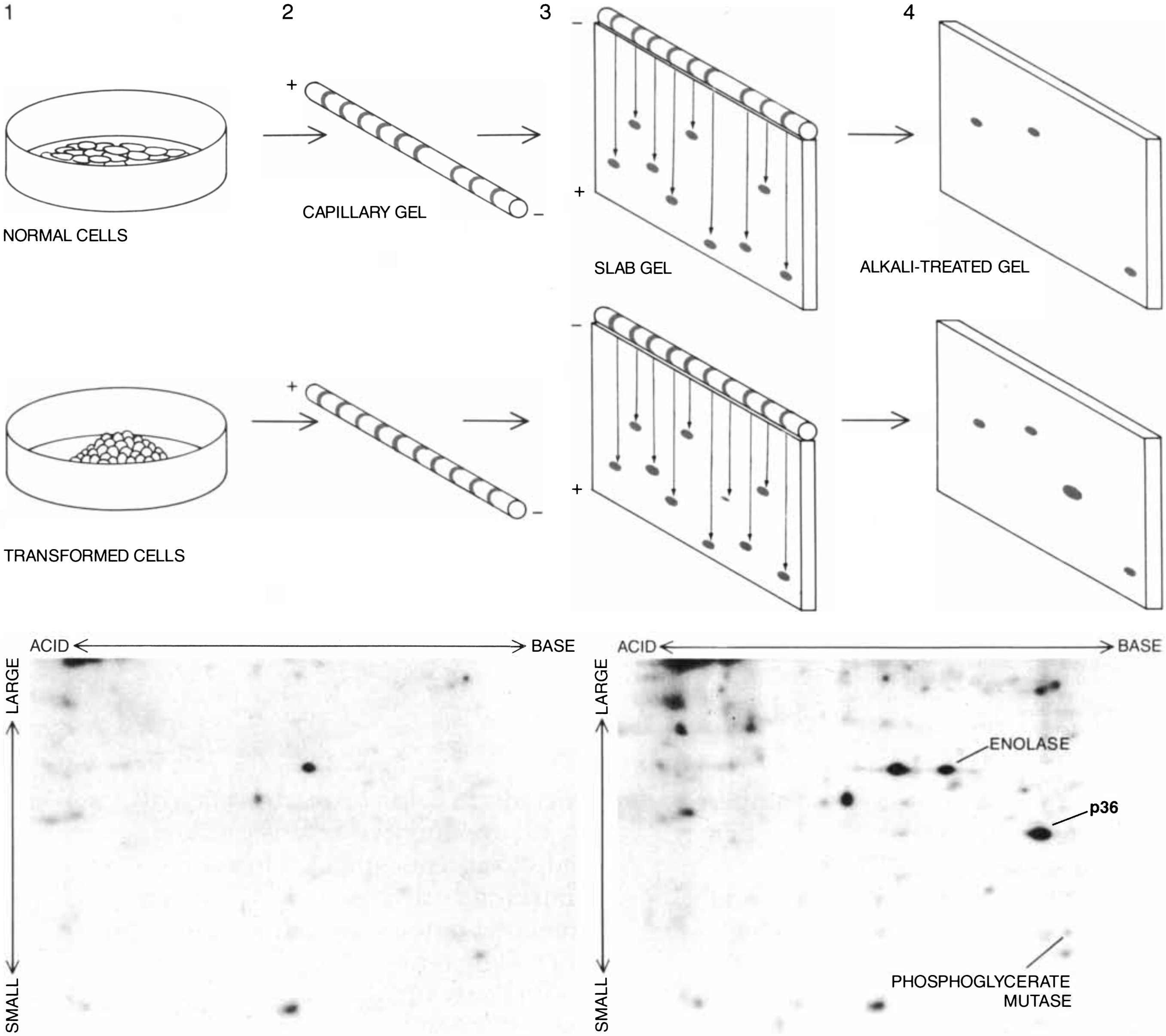
To find the store nearest you, call 1-800-447-4700. In Alaska or Hawaii, call 1-800-447-0890. IBM®

src are all even more radically altered with respect to their cellular counterparts. At least for this major group of viral oncogenes, then, it seems likely that aberration in protein structure is of oncogenic significance.

Growth Factors

Perhaps the most striking property of tumor cells is their incessant proliferation. They somehow evade a number of control systems that keep normal cells from dividing too often or continuing to divide indefinitely. Some control systems are mediated by extracellular growth factors and inhibitors that circulate in the blood. Several protein growth factors have been isolated and characterized in the past few years. Two of the most intensively studied ones are epidermal growth factor (EGF) and platelet-derived growth factor (PDGF), which deliver their signal by binding to specific protein receptor molecules embedded in the cell's plasma membrane. When either growth factor is added to a culture of nondividing cells, the cells are stimulated to divide a single time.

When the receptor protein for EGF was identified and purified by Stanley Cohen of the Vanderbilt University School of Medicine and his colleagues, it was found to be associated with protein-kinase activity; intriguingly, the activity is tyrosine-specific and is stimulated when an EGF molecule binds to the receptor. Recently both the binding of EGF and the phosphorylation of cellular proteins on tyrosine have been shown to be activities of this single receptor molecule. It is a protein that spans the membrane; the part of the molecule outside the cell recognizes EGF and a catalytic domain inside the



TWO-DIMENSIONAL GEL ANALYSIS identifies proteins phosphorylated on tyrosine in transformed cells. Normal cells and transformed cells are grown with phosphate incorporating a radioactive isotope of phosphorus (*1*). Their radioactively labeled proteins are separated according to charge in a capillary gel placed in an electric field (*2*). The gel is then soaked in a negatively charged detergent, which is bound by the proteins (*gray bands*) in proportion to their mass. When the capillary gel is laid along an edge of a slab gel in an electric field, the proteins migrate toward the positive pole at a rate inversely proportional to their mass (*3*). To enrich the gel for proteins containing phosphotyrosine, the gel is soaked in alkali, which cleaves the label from many proteins containing phosphoserine; the presence of unique phosphoproteins in the transformed cells is revealed (*4*). X-ray films (*bottom*) exposed to the alkali-treated gel show spots representing alkali-resistant phosphoproteins from normal cells (*left*) and RSV-transformed cells (*right*). Some of these proteins are shown to be phosphorylated on tyrosine; of these certain ones, including the ones labeled in the illustration, appear only in the transformed cells.

cell accomplishes phosphorylation. The binding of EGF must transmit a signal across the membrane that increases the protein-kinase activity of the catalytic domain. The PDGF receptor has now been shown to have a similar enzymatic function, which may indeed be a general property of growth-factor receptors.

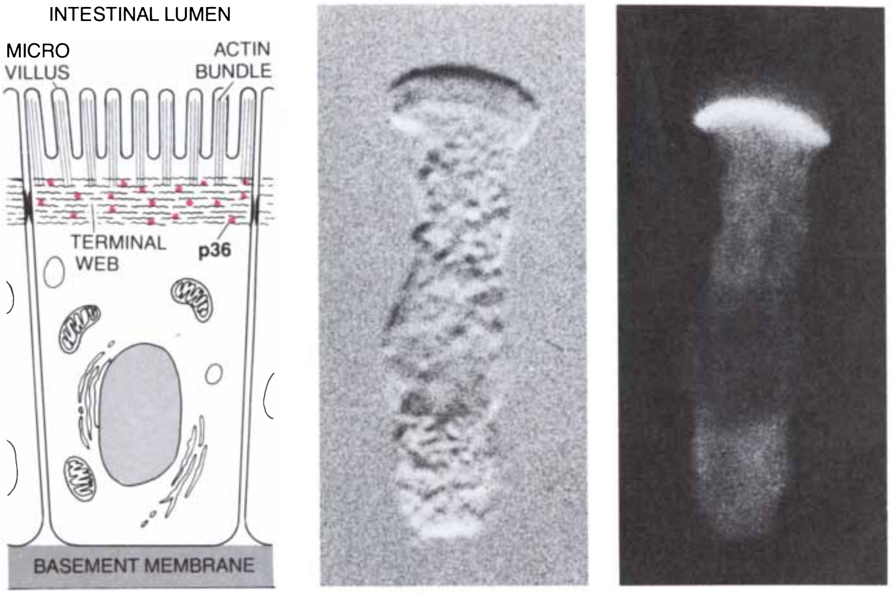
Cooper and I have observed that when EGF or PDGF is added to cells bearing the relevant receptors, there is an increase in the phosphotyrosine level in cellular proteins. This means the signal to divide could be transmitted from the occupied receptor to the interior of the cell by the phosphorylation, on tyrosine, of a target protein or proteins inside the cell. The fact that an increase in tyrosine phosphorylation is a characteristic held in common by normal cells exposed to growth factors and by certain virally transformed cells is highly suggestive. Perhaps the protein kinases encoded by many oncogenes can usurp some cellular growth-control pathway through phosphorylation of a target protein normally recognized by a receptor's protein kinase and thereby drive infected cells to unabated growth.

The *erb-B* and *sis* Oncogenes

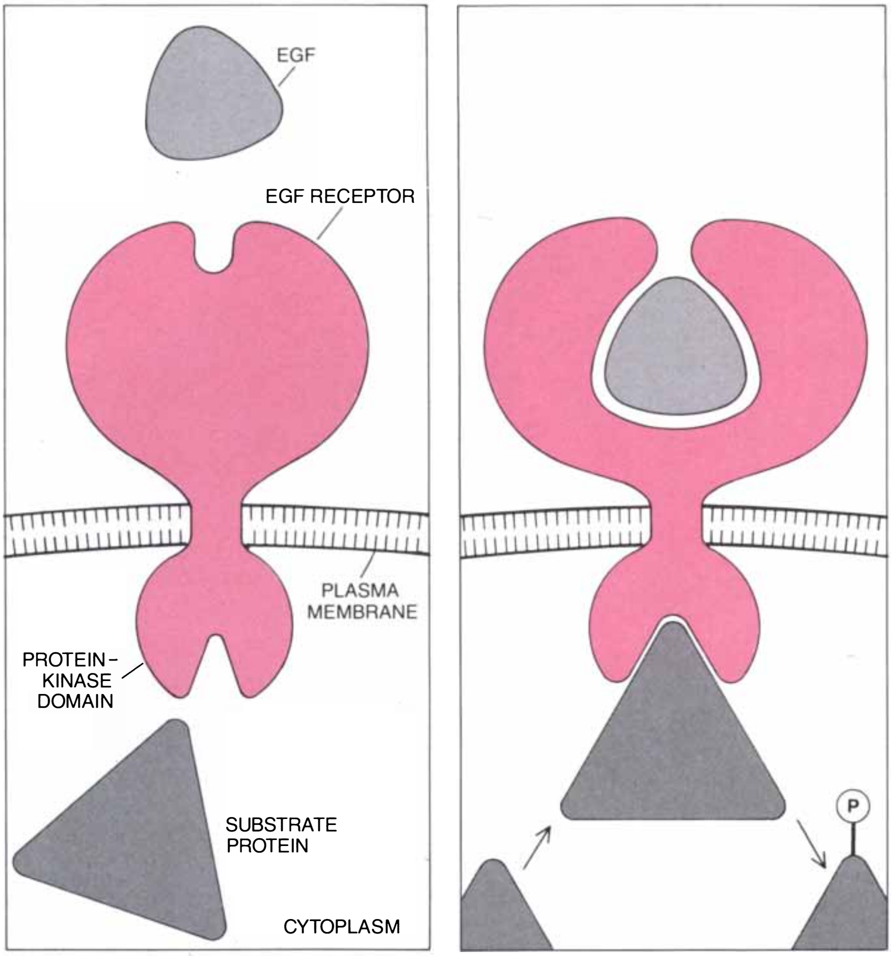
Is the functional analogy between the growth-factor receptors and viral protein kinases reflected in a structural homology? A group headed by Michael D. Waterfield of the Imperial Cancer Research Fund in London, Joseph Schlessinger of the Weizmann Institute of Science and Axel Ullrich of Genentech, Inc., has recently found that part of the EGF receptor is closely related in amino acid sequence to the product of the *erb-B* oncogene. In fact, the *erb-B* protein looks just like a truncated form of the EGF receptor: it lacks the receptor's EGF-binding region but retains the transmembrane segment and the catalytic domain that extends inside the cell.

Perhaps the *erb-B* protein, which is reported to be associated with cellular membranes, mimics the action of the EGF receptor but does so in an unregulated way. The fact that phosphorylating activity has not yet been shown to be associated with the *erb-B* protein seems to argue against such a possibility, but intense efforts now under way may yet reveal that the oncogene product is indeed a protein kinase.

The connection between growth control and viral transformation makes it important to identify the proteins phosphorylated by the tyrosine protein kinases of growth-factor receptors. Cooper and I, collaborating with Russell Ross and his colleagues at the University of Washington School of Medicine in Seattle, have identified a phosphotyrosine-containing protein with a molecular mass of 42,000 daltons in cells treated with EGF or PDGF. The protein (p42)



COLUMNAR EPITHELIAL CELLS cover the lining of the intestine. The phosphotyrosine-containing protein p36 (*color*) is concentrated in the terminal web, where the actin bundles of the microvilli are anchored. The localization of p36 in the terminal web is documented by the juxtaposition of two photomicrographs of the same isolated rat epithelial cell: an optical (Nomarski) image and a fluorescence micrograph showing where labeled antibodies bind to p36.

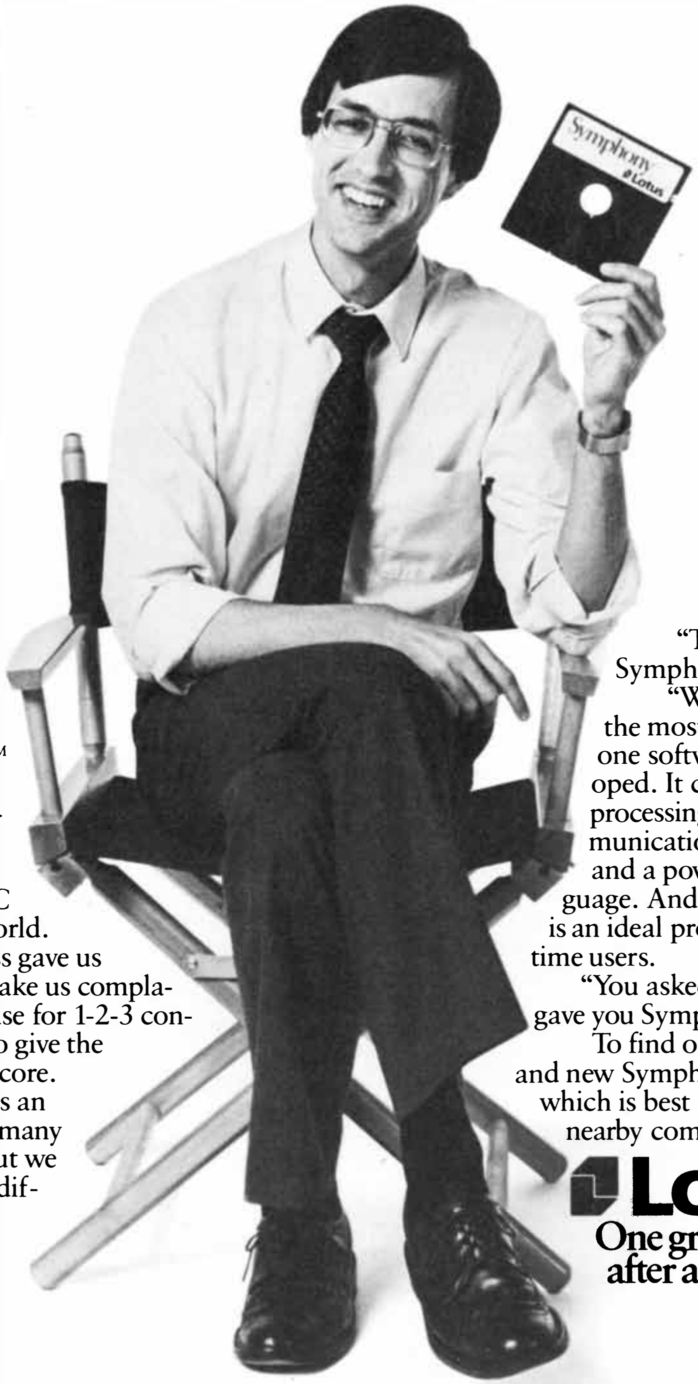


RECEPTOR for epidermal growth factor (EGF) has an extracellular domain that binds EGF and a tyrosine-specific protein-kinase domain inside the cell (*left*). The binding of an EGF molecule presumably changes the receptor's conformation so that the intracellular domain catalyzes the phosphorylation of a cellular substrate protein (*right*). Protein kinases encoded by some oncogenes may mimic the receptor's enzymatic activity without an appropriate signal from EGF and thus drive a cell to the uncontrolled growth characteristic of tumor cells.

“You asked for an encore.

Brian Stains is the Symphony Program Manager at Lotus Development Corporation. He was previously involved with the creation of 1-2-3 software and is one of the original members of the company.

"We gave you a Symphony."



"It took a dedicated team to create what the business community termed 'an overnight success.' But that's just what 1-2-3™ software became.

"When people discovered the incredible speed and power of 1-2-3, they made it the number one PC business software in the world.

"Naturally this success gave us confidence, but it didn't make us complacent. So even as the applause for 1-2-3 continued, we were working to give the business community an encore.

"We realized 1-2-3 was an analytical tool that solved many people's business needs. But we also talked to people with different requirements.

"We listened and we learned.

"The result is new Symphony™ software.

"We believe Symphony is the most comprehensive, all-in-one software system ever developed. It combines word processing, spreadsheet, communications, database, graphics, and a powerful command language. And like 1-2-3, Symphony is an ideal program even for first time users.

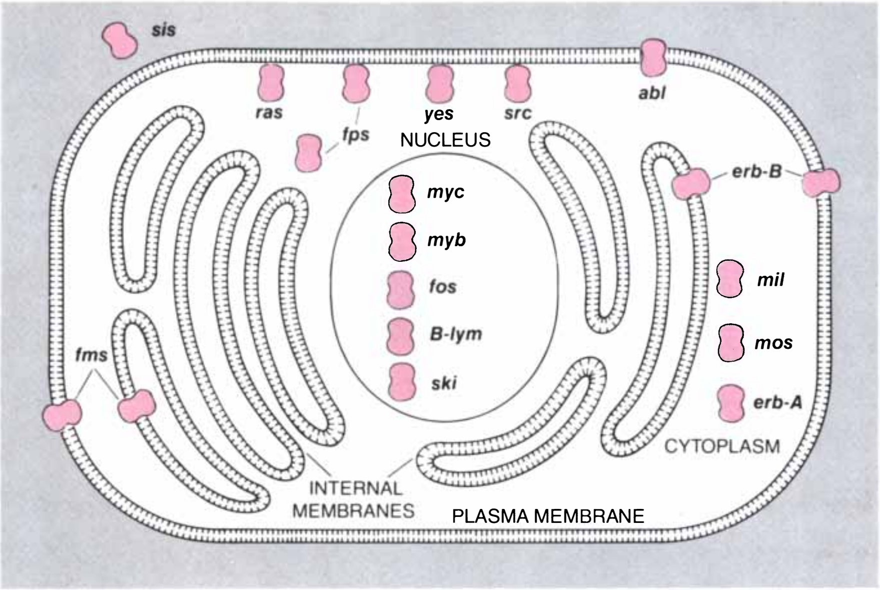
"You asked for an encore, we gave you Symphony."

To find out more about 1-2-3 and new Symphony software—and which is best for you—visit your nearby computer dealer.

Lotus

One great idea after another.™

Symphony, 1-2-3 and Lotus are trademarks of Lotus Development Corp.



SITE OF ACTION is known in the case of a number of oncogene products. The sites were determined by staining with fluorescently labeled antibodies to the various proteins or by fractionating cells in a centrifuge and then identifying the proteins found in particular fractions.

seems to be the major substrate for the receptor protein kinases; its phosphorylation may be the first step in the pathway delivering to a normal cell the signal to divide. Might such a step be mimicked in some virally transformed cells? Phosphorylated p42 has indeed been detected in some types of cells transformed by oncogenes encoding a protein kinase, but not in all such cells. Whether or not the enzymes intervene at this step in a cellular growth-control pathway remains an open question.

Still another interesting link between the protein kinases of viral oncogenes and those of growth-factor receptors has recently been reported by Russell F. Doolittle of the University of California at San Diego and by Waterfield and his colleagues. They observed independently that the protein encoded by the *sis* oncogene of simian sarcoma virus is almost identical with PDGF itself. Apparently the proto-*sis* gene is none other than the gene for PDGF. The strong implication is that a cell infected by simian sarcoma virus proliferates because it synthesizes and secretes a PDGF-like growth factor. The autogenous growth factor presumably mimics the legitimate one, interacting with the cell's PDGF receptors and thereby stimulating tyrosine phosphorylation. Whereas the real PDGF delivers its signal only on appropriate occasions, however, the *sis*-encoded protein may do so continually. In support of this notion Stuart A. Aaronson of the National Cancer Institute and Thomas F. Deuel of the Washington University School of Medicine and their colleagues have independently reported evidence that cells infected by simian sarcoma virus secrete a PDGF-like growth factor that may be responsible for their unregulated growth. One would expect to find an increase in tyrosine phosphorylation in such cells, but so far this has not been detected.

The *ras* Product

Much less is known about the remaining transforming proteins than about the protein-kinase groups I have been discussing, but there are a number of interesting findings on the products of the *ras* and *myc* oncogenes. The *ras* proteins are important because they have been implicated as one of the causes of some human cancers. Edward M. Scolnick and his colleagues, then working at the National Cancer Institute, identified the product of the Harvey sarcoma virus *ras* oncogene: a protein with a molecular mass of 21,000 daltons. When Thomas Y. Shih and Scolnick tested p21*ras* for protein-kinase activity, they found it could catalyze its own phosphorylation (on threonine rather than tyrosine) but could not phosphorylate other proteins. Moreover, the "autokinase" reaction did not depend on ATP as the phosphate donor but instead took its phosphate from a related molecule, guanosine triphosphate (GTP).

It turns out that the p21*ras* proteins encoded by the viral *ras* oncogenes are atypical: neither the normal cellular p21*ras* nor the proteins encoded by human *ras* oncogenes display autokinase activity. Like the viral p21*ras*, however, they do bind GTP tightly, forming a stable complex. The binding of GTP by p21*ras* is intriguing. The protein is localized (like p60*src* and several other oncogene-encoded protein kinases) on the inner face of the plasma membrane. A number of GTP-binding proteins are known to be in the same location in normal cells; on binding GTP they function as "coupling factors" in systems relaying signals from hormones such as adrenalin to the interior of the cell. Now Scolnick and his colleagues have found that p21*ras*, like coupling factors, can break down GTP to guanosine diphosphate and phosphate. This GTP-ase function is reduced fivefold in the case of the oncogenic p21*ras*, implying that the altered protein may remain active longer. It is tempting to speculate that the normal p21*ras* has a coupling-factor function and that the mutant oncogenic versions may transmit a continuous signal rather than a regulated one.

Nuclear Proteins

None of the oncogenic proteins discussed so far functions in the cell nucleus. Considering the number of alterations in tumor cells that require changes in the expression of genes, one might have thought most transforming proteins would be found in the nucleus, where they could exert their effects directly on the genetic material and on its expression. Presumably some of the membrane-associated and cytoplasmic proteins affect nuclear events indirectly, perhaps by interfering in pathways through which the cell normally responds to external stimuli. There is, however, a small group of nuclear transforming proteins. It includes the products of the *myc*, *myb* and *fos* oncogenes.

A number of human cancers show evidence for abnormal expression of the *myc* gene as a result of amplification or of inappropriate regulation. Some tumor cells have multiple copies of the gene (which is apparently the proto-oncogene rather than a mutant version) instead of the normal cell's complement of two; these tumor cells make abnormally large amounts of *myc*'s protein product. (Gene amplification seems to be quite common in tumor cells. In one human tumor cell line the proto-*ras* gene is amplified thirtyfold, and another line carries multiple copies of the proto-*abl* gene.)

A different abnormality is found in the lymph-gland cancer called Burkitt's lymphoma. In many cases the transformed cells, which are a type of *B* lymphocyte, carry a chromosomal translocation: a segment of DNA near the tip of one chromosome is transposed and fused to the end of another chromosome, with the result that the *myc* gene ends up close to part of a gene encoding an antibody molecule. Antibody genes are extremely active in *B* lymphocytes. As a result of the translocation the expression of the *myc* gene is increased because it is treated by regulatory elements in the chromosome as if it were an antibody gene.

Not much is known about the protein

encoded by the proto-*myc* gene. It is presumably similar to the product of the viral *myc* oncogene, which binds to DNA and is found in the nuclear matrix, a skeletal structure thought to have a role in the replication of DNA. Perhaps, then, *myc* proteins somehow make DNA replication a continuous process in affected cells. In keeping with this idea Philip Leder, Charles D. Stiles and their colleagues at the Harvard Medical School have recently shown that the proto-*myc* gene product appears in growing cells shortly before they start duplicating their DNA. The presence in a cell of a large amount of the proto-*myc* protein may cause cells to become immortal and replicate indefinitely, in contrast to normal cells, which have a finite capacity for dividing.

Immortalization alone would not necessarily convert a normal cell into a tumor cell. The activation of a second oncogene—an event that would not itself immortalize the cell—might be needed to give rise to the many other aspects of the cancerous state. Weinberg's group at M.I.T. and H. Earl Ruley of the Cold Spring Harbor Laboratory have shown that cells cultured after being taken directly from animals, unlike "established" cells that have been maintained for a long time in a culture medium, are not transformed by the introduction of a single oncogene. Certain pairs of oncogenes—*myc* and *ras,* for example—can, however, apparently cooperate to transform such cells. Epidemiological evidence and studies of tumor cells have made it clear that carcinogenesis is a multistep process, and now there is confirmation of that multistep nature at the molecular level.

Cancer research is at an exciting stage. Genes that can cause cancer have been identified and one begins to understand what the products of those genes do to transform cells to the malignant state. The products of the proto-oncogenes from which oncogenes are derived seem to have roles that must be crucial in the control of cell growth and differentiation and in embryonic development. Transforming proteins may have their profound effects on cells because they perturb these fundamental cellular processes. Even though the protein product of an oncogene and that of its proto-oncogene are in many cases essentially similar, the normal and abnormal proteins are nonetheless often distinguishable. This opens up the possibility that rational approaches to cancer therapy will one day be developed based on a detailed knowledge of the structure and function of oncogenic proteins.

NAME OF ONCOGENE			ONCOGENIC PROTEIN		
	RETROVIRUS	TUMOR	CELLULAR LOCATION	FUNCTION	CLASS
src	CHICKEN SARCOMA	—	PLASMA MEMBRANE	TYROSINE-SPECIFIC PROTEIN KINASE	CLASS 1 (CYTOPLASMIC TYROSINE PROTEIN KINASES)
yes	CHICKEN SARCOMA	—	PLASMA MEMBRANE (?)		
fgr	CAT SARCOMA	—	(?)		
abl	MOUSE LEUKEMIA	HUMAN LEUKEMIA	PLASMA MEMBRANE		
fps	CHICKEN SARCOMA	—	CYTOPLASM (PLASMA MEMBRANE?)		
fes	CAT SARCOMA	—	CYTOPLASM (CYTOSKELETON?)		
ros	CHICKEN SARCOMA	—	?		
erb-B	CHICKEN LEUKEMIA	—	PLASMA AND CYTOPLASMIC MEMBRANES	EGF RECEPTOR'S CYTOPLASMIC TYROSINE-SPECIFIC PROTEIN-KINASE DOMAIN	CLASS 1–RELATED (POTENTIAL PROTEIN KINASES)
fms	CAT SARCOMA	—	PLASMA AND CYTOPLASMIC MEMBRANES	CYTOPLASMIC DOMAIN OF A GROWTH-FACTOR RECEPTOR (?)	
mil	CHICKEN CARCINOMA	—	CYTOPLASM	(?)	
raf	MOUSE SARCOMA	—	CYTOPLASM	(?)	
mos	MOUSE SARCOMA	MOUSE LEUKEMIA	CYTOPLASM	(?)	
sis	MONKEY SARCOMA	—	SECRETED	PDGF-LIKE GROWTH FACTOR	CLASS 2 (GROWTH FACTORS)
Ha-ras	RAT SARCOMA	HUMAN CARCINOMA, RAT CARCINOMA	PLASMA MEMBRANE	GTP-BINDING	CLASS 3 (CYTOPLASMIC, GTP-BINDING)
Ki-ras	RAT SARCOMA	HUMAN CARCINOMA, LEUKEMIA AND SARCOMA	PLASMA MEMBRANE		
N-ras	—	HUMAN LEUKEMIA AND CARCINOMA	PLASMA MEMBRANE		
fos	MOUSE SARCOMA	—	NUCLEUS	(?)	CLASS 4 (NUCLEAR)
myc	CHICKEN LEUKEMIA	HUMAN LYMPHOMA	NUCLEUS	DNA-BINDING	
myb	CHICKEN LEUKEMIA	HUMAN LEUKEMIA	NUCLEUS	(?)	
B-lym	—	CHICKEN LYMPHOMA, HUMAN LYMPHOMA	NUCLEUS (?)	(?)	
ski	CHICKEN SARCOMA	—	NUCLEUS (?)	(?)	
rel	TURKEY LEUKEMIA	—	(?)	(?)	UNCLASSIFIED
erb-A	CHICKEN LEUKEMIA	—	(?)	(?)	
ets	CHICKEN LEUKEMIA	—	(?)	(?)	

PROTEIN PRODUCTS of known oncogenes are characterized and classified according to function. The second column names the animal from which each viral oncogene was first isolated (now usually the normal host) and the cancer it induces. Some oncogenes given different names (such as *fps* and *fes*) may in fact be equivalent genes in birds and mammals. The third column lists human and other animal tumors not caused by viruses in which an oncogene (*ras*) or an amplified or inappropriately expressed proto-oncogene has been identified.

The Great Temple of Tenochtitlán

This place of worship for the Aztecs was described and destroyed by the Spanish. Accidental encounters with its lost remains have led to their large-scale excavation in the heart of Mexico City

by Eduardo Matos Moctezuma

On the night of February 21, 1978, workers of the Mexico City electric-power company were digging at the corner of Guatemala and Argentina streets in the heart of the city. At a depth of two meters below the street level they encountered hard stone that blocked further excavation. On removing the dirt that covered the stone and seeing carved reliefs, they suspended work for the day.

A call to the Office of Salvage Archaeology of the National Institute of Anthropology and History led to a team of archaeologists' being dispatched to the site. They verified that the discovery was a relief sculpture on which was visible a face in profile with adornments on the head. Rescue work continued until February 27 under the archaeologists' direction. It uncovered an enormous stone disk, 3.25 meters in diameter. On the upper surface of the disk was sculptured the representation of a female deity: nude, decapitated and with her arms and legs separated from her torso. This was without doubt Coyolxauhqui, sister of the Aztec god Huitzilopochtli and herself a lunar deity, slain and dismembered by her brother after a battle on the hill of Coatepec.

The chance discovery renewed interest in excavating the ancient Great Temple of the Mexicas: the people of the Aztec city of Tenochtitlán. Parts of the temple were already known. The southwest corner was uncovered by Manuel Gamio in 1913–14 and was excavated further by Hugo Moedano and Elma Estrada in 1948. A little earlier, in 1933, Emilio Cuevas had discovered part of the wall and a short section of the stairs of the platform that supported the structures built in one of the last construction epochs of the Great Temple. The monumental sculpture of Coyolxauhqui formed part of the temple itself, and any attempt to excavate the temple completely had to take into account the modern urban surroundings of the site. It was then that my colleagues and I at the National Institute of Anthropology and History began to plan three phases of investigation.

The first phase consisted of bringing together all the available information, both archaeological and historical. We had the data from previous excavations in the temple area, dating back to 1790, when the world-famous Calendar Stone and the monumental sculpture of Coatlícue were found. We also had the 16th-century narratives of Bernal Díaz del Castillo and Hernán Cortés, who saw the Great Temple and described it, and the later writings of Bernardino de Sahagún, Diego Durán, Hernando Alvarado Tezozomoc and others.

The written sources tell us that the Great Temple was a large platform of four or five stepped levels, facing toward the west, with two stairways giving access to the top level. On the platform were two structures: the sanctuaries of Tlaloc (the god of water, rain and fertility) and of Huitzilopochtli (the god of war and of the sun). The prominence of these two deities reflects the fundamental needs of the Mexicas: their economy was based on agricultural production (hence the importance of water and rain) and on tribute collected by conquest (hence the importance of war). Thus we expected that all the elements associated with the Great Temple, such as offerings and sculptures, would in some way be related to these two fundamental themes.

The second phase of the investigation consisted of an effective plan of excavation, since we had to consider the temple's urban environment. It was necessary to analyze the existing structures in the areas to be excavated and to make sure those of the Spanish colonial period would not be affected. It was decided that the Advisory Council on Monuments would examine each case to see whether a particular building could be demolished and the excavation continued. Of the 13 structures eventually removed only two included any elements from the colonial period. They were carefully photographed and their parts were numbered for transfer to the national Bureau of Monuments.

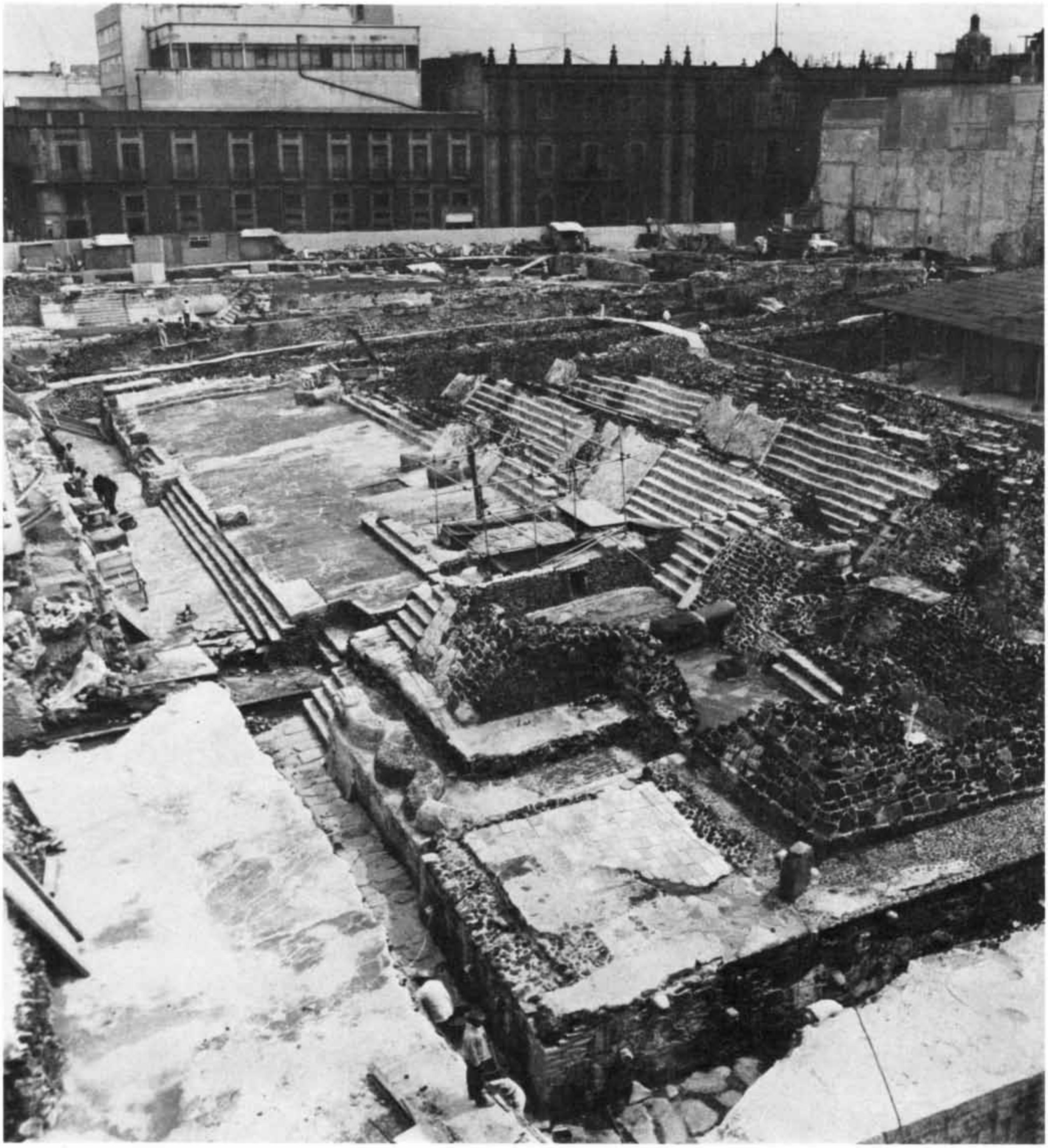
We can now confirm that the last Mexica construction epoch of the Great Temple had been razed to its foundations by the Spanish. We found only traces of the edifice on the stone-slab pavement of the great plaza. On the north side of the plaza only about a meter of the platform wall remains. The earlier construction epochs, however, were better preserved: they were older, smaller and had sunk farther below the present street level.

This gave us another problem to consider. Usually the water level in Mexico City is four or five meters below the street level. The earliest excavated epoch of the Great Temple was one we designated Epoch II; it was found almost intact and dates from about A.D. 1390. We could excavate only the uppermost parts of this building and the remains of the two sanctuaries on top of it. Lowering the water level to make it possible to do further work would have seriously threatened the stability of the colonial and modern buildings around the temple and of the temple itself. Perhaps in time, when a method is developed to control the ground water, the temple can be excavated in its entirety.

The work began with the establishment of a two-meter grid, with each square identified by letter and number. The depth below street level was recorded in the same way, with measurements of elevation from a fixed point. We foresaw the discovery of objects that would require immediate preservative measures, and so we established a conservation workshop whose staff members would help the excavators immediately when such perishable things as ceremonial offerings or the remains of murals were uncovered. We also depended on the collaboration of such specialists as biologists, chemists and geologists from the Department of Prehistory of the National Institute of Anthropology and History, and on the help of a consulting firm of engineers familiar with soil mechanics in order to avoid putting the excavation pits too close to buildings adjoining the site.

In what follows I shall place our archaeological findings against the background of historical accounts. For example, we know from historical sources that the Great Temple had on its top level two structures, one dedicated to Tlaloc and the other to Huitzilopochtli.

Sahagún writes of Tlaloc: "This god called Tlaloc Tlamacazqui was the god of rain. They said he gave them the rains to irrigate the earth and that these rains caused all the grasses, trees, fruits and grains to grow. It was he who also sent hail and thunder and lightning and storms on the water and the dangers of the rivers and sea. The name Tlaloc Tlamacazqui means that he is the god



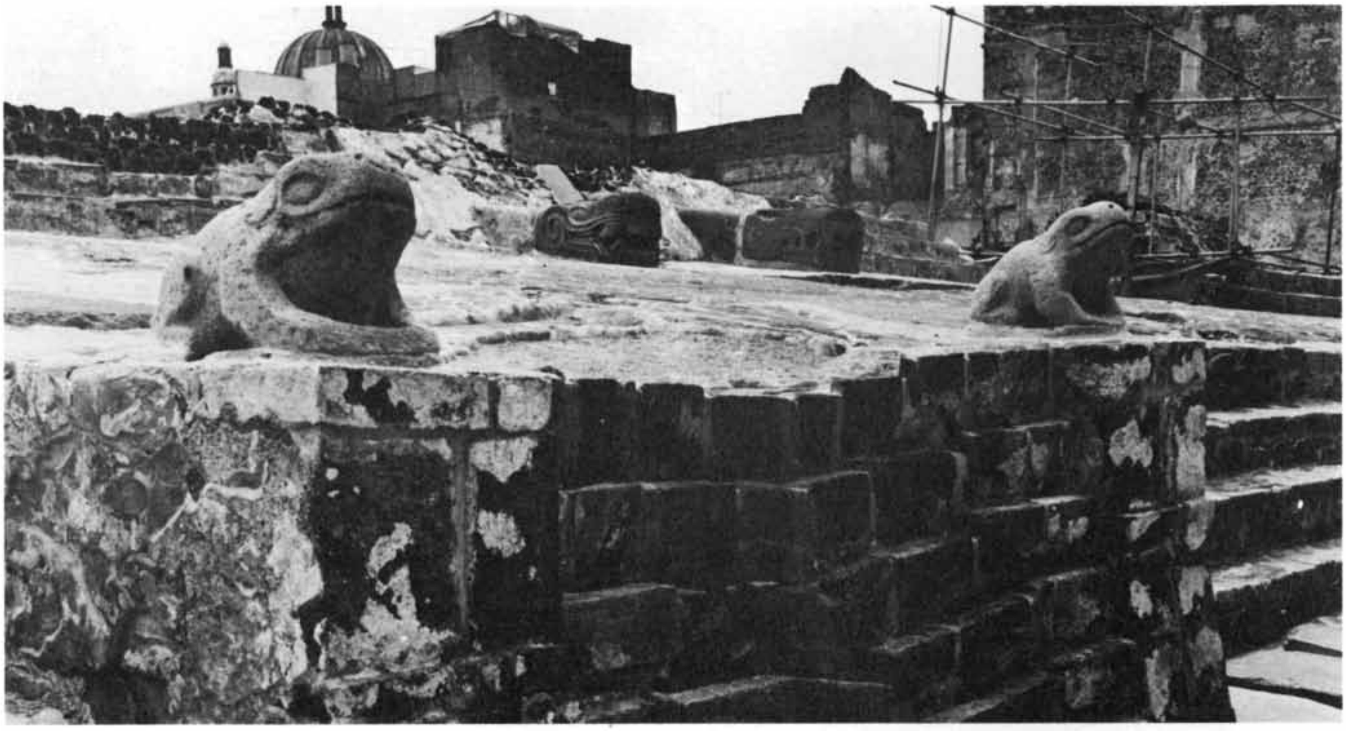
PARTIALLY CLEARED RUINS of the Great Temple of Tenochtitlán are seen in this photograph looking up Mexico City's Argentina Street, to the left, made in 1982. The fact that the summit of the Great Temple pyramid bore two sanctuaries, that of Huitzilopochtli, the god of war (to the right from the front of the temple), and that of Tlaloc, the god of water (to the left), is known from historical sources. The excavation has confirmed this fact archaeologically. In the foreground at the front of the temple (*left*) is a sculptured snake, in line with the stairway to the sanctuary of the war god. Farther down the front are two sculptured frogs (*not clearly visible in photograph*), in line with the stairway to the sanctuary of the water god. Snakes were often symbolic of Huitzilopochtli; frogs were symbolic of Tlaloc.

who resides in the terrestrial paradise and gives to men the subsistence necessary for life."

Sahagún writes of Huitzilopochtli: "The god called Huitzilopochtli was another Hercules, exceedingly robust, of great strength and very bellicose, a great destroyer of towns and killer of people. In warfare he was like living fire, greatly feared by his enemies.... While he lived this man was highly esteemed for his strength and prowess in war."

Bernal Díaz describes what he saw: "On each altar were two giant figures, very tall and very fat. They said that the one on the right was Huichilobos [*sic*], their war-god."

Díaz adds: "At the very top of the [pyramid] there was another concavity, the woodwork of which was very finely carved, and here there was another im-



PAIR OF FROGS at the Great Temple flank an altar that stands above the flagstones of the ceremonial precinct in front of the pyramid. This monument in honor of Tlaloc was built in about A.D. 1469, in the epoch of periodic temple construction designated Epoch IVb.

UNDULATING SNAKE was also sculptured in Epoch IVb. The layer of rubble fill visible behind the sculpture marks a subsequent epoch of temple enlargement. The builders had reached Epoch VII by the time the Spanish arrived and conquered the Aztec empire.

age, half man and half lizard.... They said that the body of this creature contained all the seeds in the world, and that he was the god of seedtime and harvest."

Of the temple itself Sahagún writes: "In the center and higher than the [other temples of the city] the principal [temple] was dedicated to the god Huitzilopochtli, or Tlacauepan Cuexcotzin. This [pyramid] was divided at the top so that it looked like two; it had two [sanctuaries] or altars, each of which was topped by a [tall roof], which bore distinct insignia or devices. In one of those [sanctuaries], the principal one, stood the statue of Huitzilopochtli... also called Ilhuicatl Xoxouhqui; in the other was the image of the god Tlaloc. In front of each one of these statues was a round stone like an executioner's block, called *techcatl,* where they killed all those whom they sacrificed in honor of that god. From the block to the ground there was a pool of blood from those who were killed on it, and this was true of all the [temples]. They all faced west and had very narrow and steep steps leading to the top."

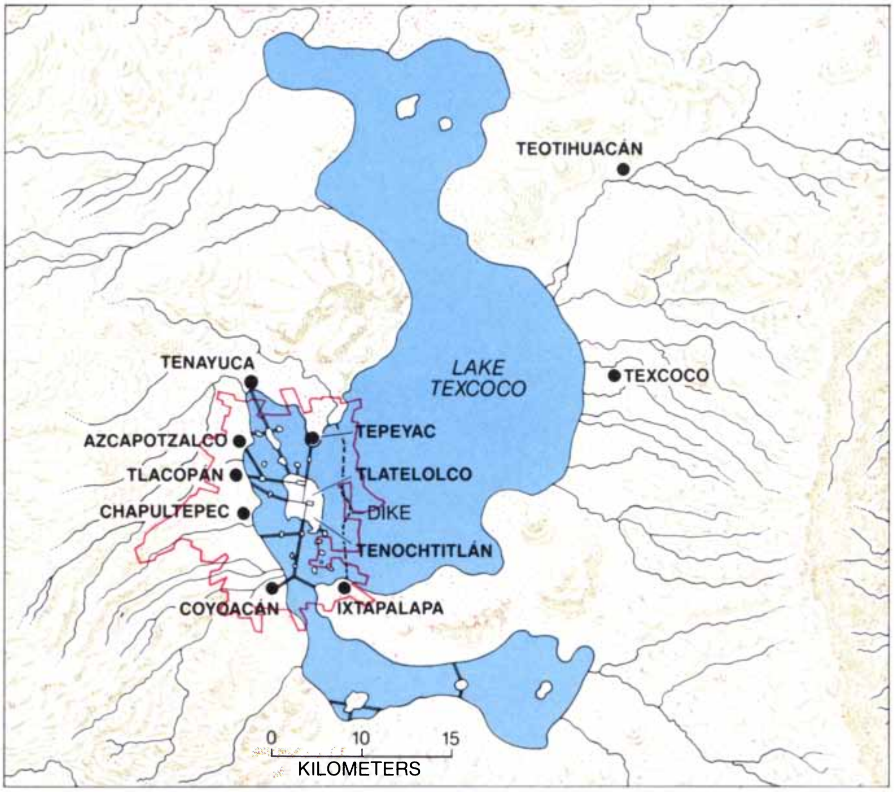
Discussing sacrifices in honor of Xipe Totec, Sahagún adds: "Having brought them to the sacrificial stone, which was a stone three hands in height or a little more, and two in width, or almost, they threw them on it on their backs."

It is interesting that we found a sacrificial stone, the obvious symbol of Mexica power and where captives of war were sacrificed, in front of the sanctuary of Huitzilopochtli. It is a slab of black volcanic rock, and its dimensions conform with those given by Sahagún. The stone, which was found in situ embedded in the floor near the stairs (two meters away), measured 50 centimeters by 45 (20 inches by 18).

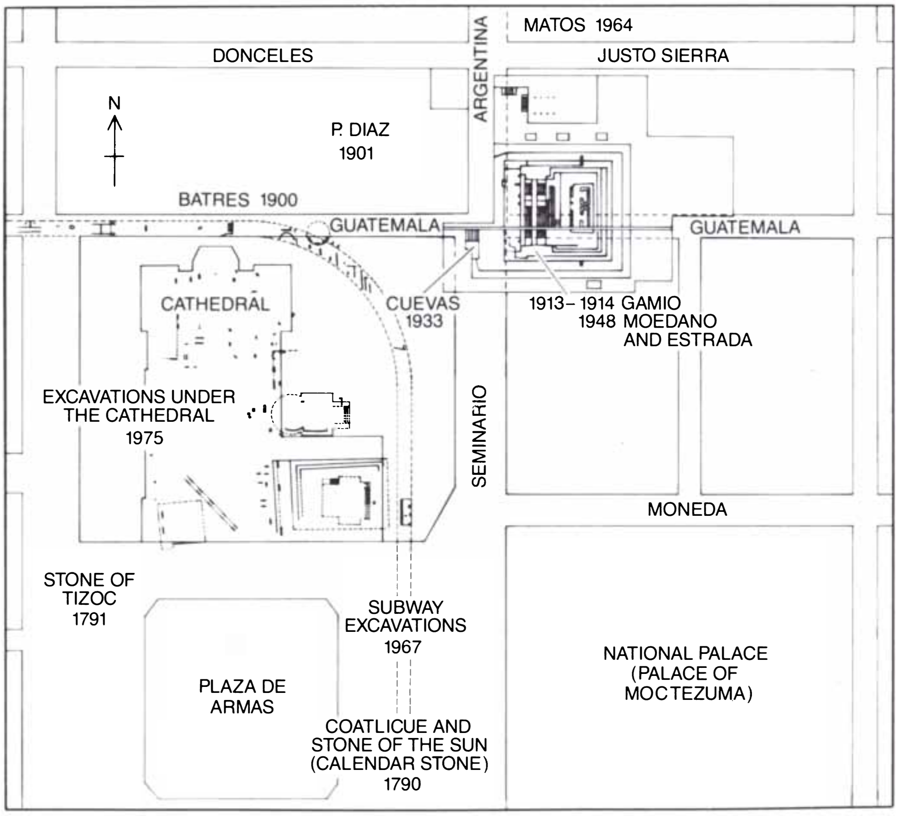
On the Tlaloc side of the top level we found a polychromed statue, known as a *chacmool,* also in situ and in the same position as the sacrificial stone in relation to the sanctuary of Huitzilopochtli. The find confirms the historical interpretation of the role of the *chacmool:* it is an intermediary between the priest and the god, a divine messenger. Both elements—the sacrificial stone and the *chacmool*—in front of the sanctuaries can be considered as dual symbols, the first symbol related to war and the second to a more "religious" idea: the divine messenger.

We discovered that two large stone piers framed the entrance of the sanctuary of Tlaloc. The surfaces of the piers that faced outward were painted with a row of black and white circles representing the eyes of Tlaloc; just below the circles were three horizontal bands, one blue and two red. The lower halves of the piers were decorated with alternating vertical bands of black and white.

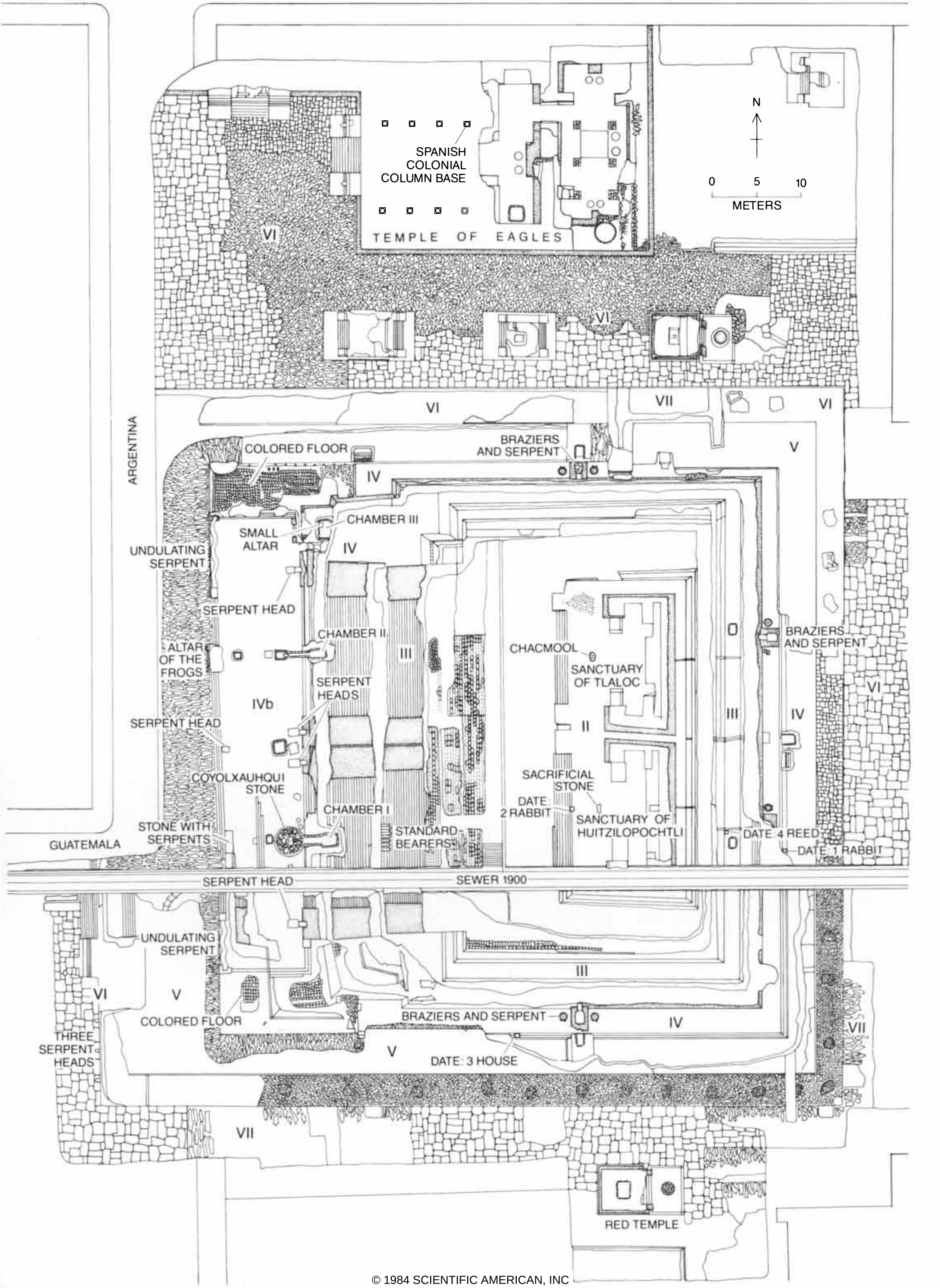
On the inner surface of each pier is a



VALLEY OF MEXICO was dominated by five adjoining lakes that became one body of water in the summer rainy season. Tenochtitlán, the Aztec capital, occupied an island in the central lake, Lake Texcoco. The area outlined in red indicates the extent of Mexico City today.



ARCHAEOLOGICAL ACTIVITIES in downtown Mexico City in the 20th century are indicated by the dates on this map. The southwest corner of the Great Temple was exposed in 1913–14 and was further examined in 1948. The two 18th-century dates indicate when and where two sculptured monuments, the Calendar Stone (1790) and the Stone of (the Aztec ruler) Tizoc (1791), were accidentally uncovered centuries after the Spanish had destroyed the temple.


SPANISH COLONIAL COLUMN BASE
TEMPLE OF EAGLES
N
0 5 10
METERS
VI
VI
VI
VII
VI
ARGENTINA
COLORED FLOOR
BRAZIERS AND SERPENT
V
IV
CHAMBER III
SMALL ALTAR
UNDULATING SERPENT
IV
SERPENT HEAD
CHAMBER II
III
CHACMOOL
SANCTUARY OF TLALOC
BRAZIERS AND SERPENT
ALTAR OF THE FROGS
IVb
SERPENT HEADS
VI
III
IV
SERPENT HEAD
II
COYOLXAUHQUI STONE
SACRIFICIAL STONE
DATE: 2 RABBIT
SANCTUARY OF HUITZILOPOCHTLI
CHAMBER I
STONE WITH SERPENTS
STANDARD-BEARERS
DATE: 4 REED
DATE: 1 RABBIT
GUATEMALA
SERPENT HEAD
SEWER 1900
UNDULATING SERPENT
III
VI
V
COLORED FLOOR
BRAZIERS AND SERPENT
IV
VII
THREE SERPENT HEADS
V
DATE: 3 HOUSE
VII
RED TEMPLE

painting of a standing personage; the body is yellow and is adorned with wristlets and anklets of blue and black. The left hand extends forward and holds a baton or spear. The interior of the sanctuary appears to have been divided into two parts. A platform stood against the back wall, evidently to support an image of the deity that once occupied the sanctuary. A small altar rises above the middle of the platform in the sanctuary of Huitzilopochtli. It is aligned with the sacrificial stone outside. Apparently the altar itself once supported an image of the god.

We found pieces of wood embedded in the piers of both sanctuaries. They must be the remains of decorated panels. Richly carved wood panels were seen in Aztec sanctuaries by both Bernal Díaz and Cortés.

All these remains were part of the early construction of the Great Temple designated Epoch II. This epoch probably dates to the period before A.D. 1428, the year the Mexicas of Tenochtitlán began their war of liberation from their overlords of neighboring Azcapotzalco. Diverse kinds of information lead us to this conclusion, among them the modest size of the edifice and the fact that Epoch II is the earliest yet excavated. Further, on the top step of the Huitzilopochtli side, in line with the sacrificial stone, is carved a face with two glyphs above it. One of them gives the date "2 Rabbit"; the other is unreadable. "2 Rabbit" can be interpreted as representing the year A.D. 1390.

As many as 10 more epochs of construction, each epoch enlarging the temple, were built over this early structure. Only five of them were enlargements of the entire pyramid, that is, expansions on all four sides. The other five only enlarged the temple's main facade. It is clear that the fill covering the Epoch II temple was far greater in volume than the fill covering any of the subsequent facades. This leads to the conclusion that once the Mexicas were liberated from the yoke of Azcapotzalco they wanted to build a larger temple and had the means to do so, including forced labor to transport volcanic rock and dirt to the site for fill and to supply other building materials.

CHACMOOL STATUE, a reclining figure, was found in situ in front of the sanctuary of Tlaloc. The statue had been emplaced in Epoch II. Much of its polychromed surface has survived.

I shall now briefly describe the subsequent construction epochs of the Great Temple and also discuss the other structures found near the temple during the excavations. To begin with Epoch III, the stairways of this epoch are well made and the walls of the different levels of the pyramid stand vertically. Eight sculptures representing standard-bearers were found on the stairway of the Huitzilopochtli side. These figures must once have adorned the Epoch III temple, but when the next construction epoch began, they were collected and placed on the stairway where they were found. On the back wall of the platform of the Huitzilopochtli side was carved the glyph "4 Reed," which refers to the year A.D. 1431.

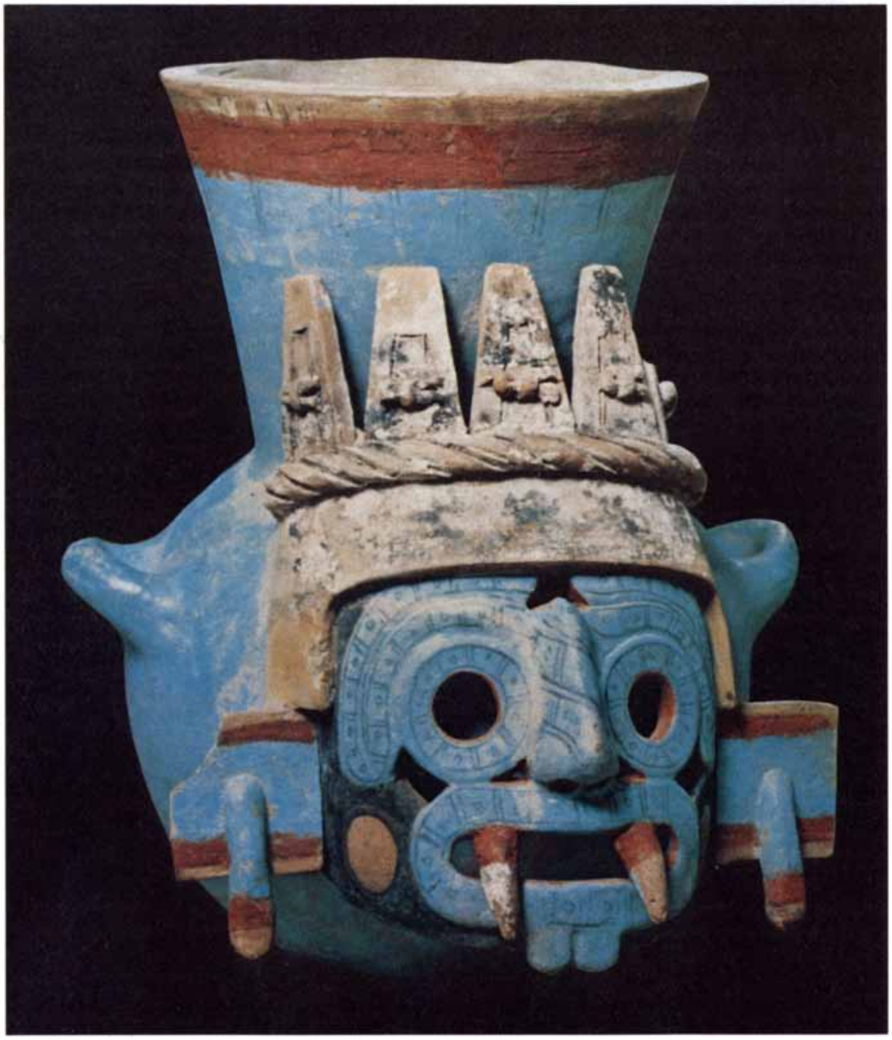
The Epoch IV construction brought additions that made the temple particularly splendid. The platform was decorated on all four sides with braziers and serpent heads. The braziers on the Tlaloc side are adorned with the face of the water god; those on the Huitzilopochtli side are decorated with a knot, a symbol of the warrior deity. Just below the braziers and serpents we unearthed various offerings enclosed in cists, or stone-walled boxes.

Epoch IVb involved an addition to the main facade (on the west): a vast platform extending in front of the temple. At both ends of this wide, unbroken area we uncovered enormous stone serpents with an undulating body and a huge head, still bearing some of the paint that had originally covered them. A small altar with two frogs stands in line with the middle of the stairs that once led to the part of the temple dedicated to Tlaloc. In the corresponding place on the Huitzilopochtli side a stone two meters long and decorated with serpents forms part of the top of the platform.

The remains of the temple pyramid itself include the base of the stairways with four large stone serpent heads, one at each end and two in the middle, marking the meeting place between the two halves of the temple. The platform in front of the stairs on the Huitzilopochtli side is where the street excavators uncovered the monumental sculpture of Coyolxauhqui in 1978. This was the goddess, the sister of Huitzilopochtli, who conspired against him before he was born and on whom he took revenge by dismembering her on the hill of Coatepec. The temple pyramid is in fact

REMAINS OF THE GREAT TEMPLE attributable to the seven main epochs of construction are combined in this partial reconstruction. The sanctuaries of Tlaloc and Huitzilopochtli, respectively facing a *chacmool* statue and a sacrificial stone, are shown as they appeared in Epoch II. The second set of steps descending westward is from Epoch III. The next-lower set of steps is from Epoch IV. The platform below them, the one with the huge snakes, the Coyolxauhqui stone shown on the cover of this issue and the altar of the two frogs, is from Epoch IVb. Epoch V is represented by the lower part of the platform on three sides of the temple. Epoch VI construction is represented by flagstones around all four sides and several smaller temples: to the north the Temple of the Eagles and three platforms, and to the south the Red Temple. The destruction of the final temple by the Spanish conquistadors was so thorough that Epoch VII is represented only by small areas of slab flooring and a small area of platform.

OFFERING CIST (a stone-walled box) was one of several cists discovered under the paving of the Epoch IVb platform. Designated Chamber III, it held figurines, masks and two pots.



EFFIGY POT, representing Tlaloc, is one of many kinds of pottery found in offering cists at the Great Temple. Many objects at the temple came from other parts of Mexico under Aztec control, but this image of the water god appears to be the work of a resident of Tenochtitlán.

a symbolic representation of the mythical hill.

In the platform we found various offerings, some around the massive Coyolxauhqui relief, some between the two middle serpent heads and some aligned with the midpoint of the Tlaloc stairway. All were buried under the floor of the platform except for chambers I and II, which were found under the temple stairs. At the north and south ends of the platform were rooms with colored stone floors, and at the north end was a small altar with stairs. Within it were two offerings dedicated to Tlaloc, one consisting of more than 42 skulls and bones of children.

We think Epoch IV generally corresponds to the reign of the Aztec ruler Moctezuma I, because the date "1 Rabbit" (equivalent to the year 1454) is on the back part of the temple on the Huitzilopochtli side. The addition of Coyolxauhqui and the serpents in Epoch IVb may well date to the reign of the ruler Axayacatl. A glyph on the south side of the pyramid represents the date "3 House" (1469), the year Axayacatl became Tlatoani (the Aztec word for speaker or ruler).

We found that the temple platform associated with Epoch V was covered with stucco. We also uncovered part of the floor of the ceremonial precinct consisting of stone slabs. Epoch VI, which was the next to the last, is represented by the remains of part of the temple platform. A wall on the west side is decorated with three inset serpent heads. Also belonging to Epoch VI are the small temples on the north side of the Great Temple, the Platform of the Eagles, and a great plaza of stone slabs on which these structures stand. On the south side of the Great Temple is a structure we call the Red Temple, which corresponds to the small temple on the north side.

The last epoch of the Great Temple, the one the Spanish saw, was Epoch VII. Its remains are generally limited to part of the slab floor of the ceremonial precinct and traces that remained when the temple was razed. On the north side, however, part of the temple platform can be discerned; it corresponds to the platform of the preceding epoch. Apparently in Epoch VII the small adjacent edifices of Epoch VI were covered with fill and a floor was laid over them. The Great Temple platform was built on top of the preceding one but was not enlarged.

We unearthed more than 100 offerings, including more than 7,000 artifacts. Many were of Mexica origin but many others came from tributary areas. The large majority of them had come from what is now the state of Guerrero, from the Mixtec area in Oaxaca, from the Gulf of Mexico coast, from Puebla and from other parts of

EAGLE WARRIOR, a life-size pottery figure nearly two meters high, is one of the finest discoveries at the Great Temple. The figure may originally have been covered with an eagle costume. The wings on the arms, the talons below the knees and the oversize bird mask identify the figure as representing one of the two main orders of Aztec soldiery. The soldiers of the other order wore a jaguar costume.

CEREMONIAL KNIVES, painted with red ocher and bearing seashell teeth and eyes with obsidian pupils, resemble those that were used by the Aztecs to cut out the heart of a captive in a sacrificial rite. At the bottom of each of the two knives at the left is a lump of copal incense.

Mexico. Not a single object, however, is attributable to the neighboring Tarascan or Mayan cultures.

The imports from the tributary areas included many interesting stone objects. Among them were masks and figures of various types and sizes, done in the Mezcala style of Guerrero, and alabaster pieces from the Puebla region, including representations of deer heads and seated deities. Two magnificent funerary urns of orange ceramic are of Gulf Coast origin. Inside them were the remains of burnt bones, necklaces and other materials. In other offerings were seashells, fishbones, sawfish bills and corals from both the Gulf coast and the Pacific coast.

Little more can be added for the present, since the material is still under study. This work constitutes the third phase of our research: interpretation. It can be said, however, that the majority of the materials—for example all the objects of marine origin—represent Tlaloc or symbols associated with the water god. Huitzilopochtli himself was not actually represented among the artifacts, but there are symbols of the war god, for example skulls and ceremonial knives decorated with eyes and teeth of conch shell.

To help the reader understand the importance of these offerings, I can do no better than to further quote Durán and Bernal Díaz. Durán writes: "Seeing the speed with which the temple was being built and wishing to honor his god further, King Moctezuma ordered all the sovereigns of the land to gather from all the cities a great number of precious stones, strings of green stones, which they called *chalchihuites,* transparent stones, bloodstones, emeralds, rubies and carnelians, in short every type of rich stone and precious jewel and many treasures. Every *braza* [about six feet] that the building grew these precious stones and rich jewels were to be thrown into the mortar.

"And thus giving tribute to the leaders each city in its turn lent support and gave the jewels and stones to throw into the foundations, so that at every *braza* of the building they tossed in such a quantity that it was something to marvel at. And it was said that since god gave these riches, it was not inappropriate to employ them in his service, because they were his."

Referring to the time when a similar temple in the neighboring city of Tlatelolco was destroyed, Díaz writes: "Some curious readers may ask how we came to know that they had thrown gold and silver and precious *chalchihuites* and seeds into the foundation of the [pyramid] and watered them with the blood of Indian victims, seeing that the building was erected 1,000 years ago. My answer is that after we conquered that great and strong city and divided the ground we decided to build a church to our patron and guide St. James in place of Huichilobos' [pyramid], and a great part of the site was taken for the purpose. When the ground was excavated to lay a foundation, gold and silver and *chalchihuites,* and pearls, seed pearls and other precious stones were found in great quantities, and a settler in Mexico [City] who built on another part of the site found the same."

What concept underlay the fact that two deities shared the summit of the Great Temple? The economic aspect of this duality has already been mentioned; it remains to discuss the political implications. The double temple as a type of monument was present in central Mexico as early as the beginning of the 13th century A.D., for example at Tenayuca and later at such other Mexica sites as Teopanzalco, Santa Cecilia, Tlatelolco and of course at Tenochtitlán itself. We think the pattern emerged as the result of a struggle for power between two factions in the societies of the Classic Period (A.D. 300–900). For example, it seems possible that in the city-state of Teotihuacán, which many scholars consider to have been a theocratic polity, control was exercised by priests in collaboration with a coercive military. The struggle between two such opposing groups could have resulted finally in the union of both state systems, the ideological priesthood and the repressive military, to control the whole of society. This evidently happened among the Mexicas, whose leader was both the high priest and the military commander. The political alliance of priests and warriors was symbolized in the double sanctuaries of Tlaloc and Huitzilopochtli on top of the city's largest temple. Moreover, both aspects were embodied in the person of the Mexicas' single ruler, the Tlatoani.

In short, the Great Temple of Tenochtitlán symbolized to the Mexicas water and war, life and death, Tlaloc and Huitzilopochtli. It was a place of glory for the Mexicas and a place of suffering for those under their power.

DEDICATION STONE of the Great Temple shows at the top two members of the Aztec royalty, a ruler and his successor, facing each other on opposite sides of a "grass ball of sacrifice." Each is letting blood from his ear. The exaggerated streams of blood splash on the grass ball before entering the mouth of a monster representing the earth. The glyph at the bottom signifies "8 Reed" (1487), the year of the ceremony observing the completion of a major building epoch.

Fiber Optics in Plants

The tissues of plant seedlings can guide light through distances as great as several centimeters. The cells of a plant may thus exploit "light pipes" to coordinate aspects of their physiology

by Dina F. Mandoli and Winslow R. Briggs

Plants employ light in two quite different ways. In one of them, photosynthesis, the light is a source of energy: it powers the synthesis of organic molecules. In the other the light is a source of information; in a word, the light is a signal. Thus the light may determine whether or not a seed will germinate; it may determine the angle a growing root will make with respect to gravity; it may determine the angle a growing shoot will make with respect to the direction of incident light; it may determine how rapidly a stem will elongate or a leaf will expand; it may determine when a plant will bloom. In each case the energy the signal requires to trigger the response is orders of magnitude less than the energy needed to drive the response itself.

Now a further wrinkle enters: it is beginning to appear that the tissues of a plant, and in particular an etiolated plant—a seedling in the darkness below the soil line—can guide light. In essence the seedling acts like a bundle of fiber-optic strands. At the Department of Plant Biology of the Carnegie Institution of Washington we have detected the guiding of light through columns of cells in the shoots of etiolated seedlings over distances as great as 4.5 centimeters. This means the light has passed through not only the interior of dozens, even hundreds, of cells but also the complex, light-scattering interfaces between successive cells. One conceivable result of such "light-piping" is that parts of the seedling buried underground may begin to respond to light as soon as the tip of the seedling pierces the surface of the soil, or even before if the soil transmits a significant amount of light. A more speculative possibility is that the piping of light helps to coordinate the activity of cells as the plant develops. Plants, after all, have nothing that closely resembles a nervous system. Perhaps in some instances they rely on fiber optics instead.

The place to begin is with the signals plants get from light and the ways the signals are measured. Both are remarkably complex. Basically plants get four distinct types of information from the light impinging on them. They are the amount of light, the quality of the light (that is, its spectral distribution), the direction of the light and the duration of the light. The measurements are made by pigment molecules inside the cells of the plant, and notable among the various pigments is the one called phytochrome. It has been studied in many laboratories, including our own.

Phytochrome is uniquely well suited to measure the quality of light. In particular it measures the ratio of energy between two rather narrow parts of the spectrum: the red part (wavelengths near 660 nanometers) and the far-red part, just beyond the visible (wavelengths near 730 nanometers). What suits it for making that measurement is its photochemical behavior: the absorption of red light converts it into a form that makes the molecule preferentially sensitive to light in the far red. Conversely, the absorption of far-red light converts it into a form that is preferentially sensitive to red. In short, the spectral quality of the light absorbed by phytochrome determines what form it is in.

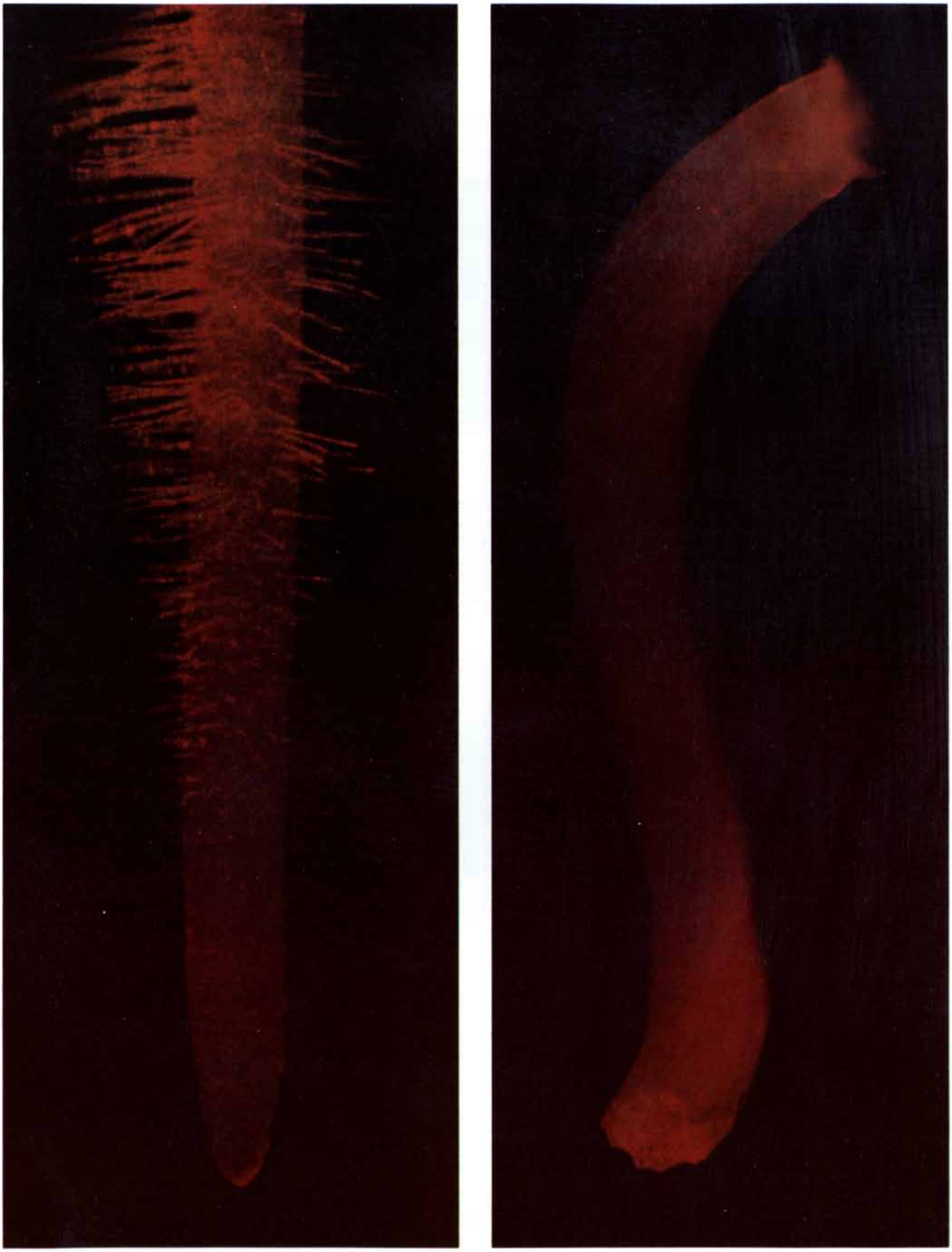
Phytochrome is customarily abbreviated P and the state of the pigment is denoted by a subscript. Hence P_r signifies the red-absorbing form of phytochrome, which predominates in plants growing in the dark or in deep shade. (In the environment of a shaded plant there is more far-red light than red because the leaves above the plant selectively absorb the red. The far red keeps the shaded plant's phytochrome in the form that is preferentially sensitive to red.) P_{fr} signifies the far-red absorbing form, which predominates in plants growing in unfiltered daylight. P_{fr} is thought to be the active form of the pigment: the one that potentiates the plant's responses.

The responses fall into four classes; phytochrome participates in all of them. Low-fluence responses are those that show an increasing degree of response with increasing fluence, or quantity, of red light. Most can be saturated by a few minutes of red. Examples include seed germination, the directionality of root development (in particular the alteration of the directionality dictated by gravity) and the elongation of the shoots of etiolated seedlings. The low-fluence responses depend on both the quantity and the quality of light. Lettuce seeds kept in the dark but exposed to sequential flashes of light furnish a spectacular example. The seeds will germinate only if the last flash in the sequence is red. They remain dormant if the last flash is far red.

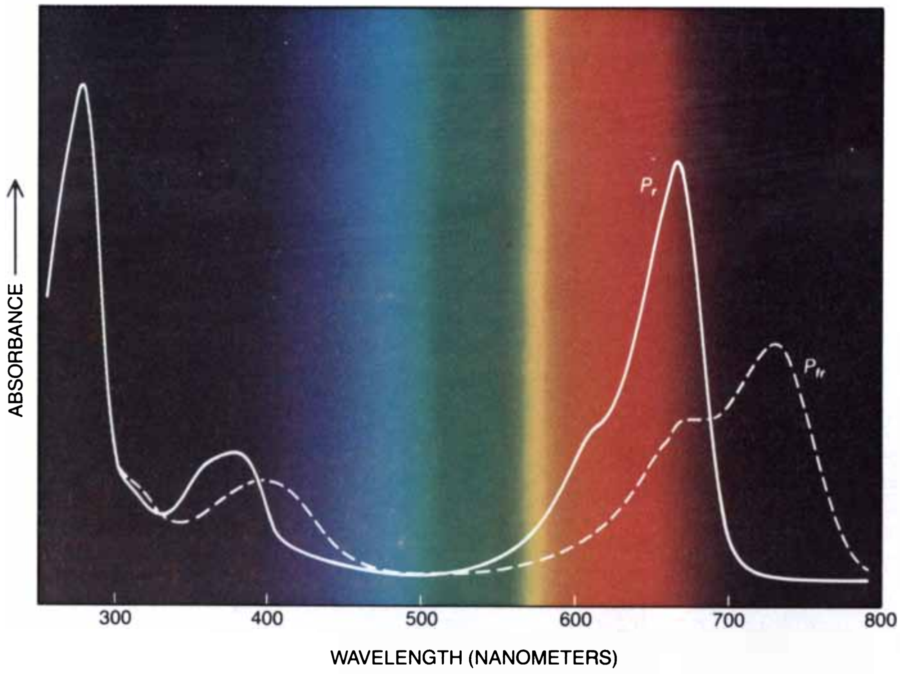
End-of-day responses are superficially similar to low-fluence responses. Like the low-fluence responses, they require no more than a small quantity of light, they can be stopped by light (that is, the initial chemical processes underlying them can be reversed) and they measure both red and far-red light. On the other hand, they depend on the light at a particular time: the end of the day. The elongation of the stem of the sunflower plant is an example. The elongation, and hence the height of the plant, is greatest if far red dominates the end-of-day illumination.

High-irradiance responses require far more light than the low-fluence responses. Moreover, they require exposures of hours, not minutes, to reach saturation. Stem elongation again provides an example. The far-red environment of a shaded plant can cause the stem of the plant to elongate rapidly, so that the top of the plant can poke through an overlying canopy of vegetation.

Finally, some responses depend on the length of the day. The most notable example is the flowering of a plant. It is perhaps the most familiar response of a plant to light, and yet it is perhaps the least understood of all the responses mediated by phytochrome. Some plants require long days to flower; others require short days; still others seem not to require one or the other. Evidently all day-length-sensitive flowering plants, long-day and short-day alike, detect the duration of both the day and the night. If a



LIGHT-GUIDING PROPERTIES of plant tissues are responsible for these photographs, which show red laser light guided through tissues of seedlings. At the left is a corn root less than a millimeter in diameter. From it come root "hairs," which are extensions of individual cells. Each hair is a few micrometers in diameter. Light has been "piped" to the tips of the hairs. At the right is the curved stem of an oat seedling a millimeter in diameter. Light has been guided through the curves. The stem acts like a bundle of glass fiber-optic strands.



LIGHT-ABSORBING MOLECULE phytochrome mediates many of the responses of a plant to light. The molecule has two forms, which give it the patterns of light absorption shown in the illustration. In one form, labeled P_r (*solid line*), the molecule is preferentially sensitive to red light. In the other, labeled P_{fr} (*broken line*), it is preferentially sensitive to far-red light, that is, to light at the transition from red to infrared wavelengths. The absorption of red light changes P_r into P_{fr}; the subsequent absorption of far-red light changes P_{fr} into P_r. The absorption at the left in both spectra is from several amino acids in the protein component of the molecule.

plant receives a single flash of red light during the night, it behaves as if the days were long. The reason is not known; perhaps the response involves the interaction of phytochrome with a circadian rhythm, or daily cycle, of the plant. Nevertheless, horticulturists exploit the phenomenon: they apply artificial light to short-day plants such as chrysanthemums in order to delay their flowering in the winter and extend the season in which fresh blooms are available.

Our own investigations into phytochrome and the fiber-optic properties of plants began with some surprises among the low-fluence responses. We were growing oats, which have a characteristic pattern of germination. First a root emerges from the seed; this is by definition the onset of germination. The root heads downward, sending out the projections called root hairs. Then a shoot emerges. It is a rather complicated structure that includes the mesocotyl, or first stem, and above it the coleoptile, a cylinder of leaf tissue shielding the primary leaves, the first true photosynthetic ones. The mesocotyl and the coleoptile are joined by the structure called a node, which appears as a slight bulge in the shoot. Compressed at the node are a number of groups of cells destined to become leaves. In grasses such as the oat, nodes are the chief places where cells proliferate. (This is why a lawn can be mowed without killing the grass.) Elsewhere in the stem, growth proceeds largely by cell elongation, not by cell proliferation.

The mesocotyl serves to position the node just below the soil line, regardless of the depth at which the seed was buried. Accordingly the mesocotyl grows rapidly in the dark but is progressively inhibited by the increasing amount of light it encounters as it grows toward the soil surface. The coleoptile acts in much the opposite way. It grows slowly in the dark but grows faster as the amount of light increases. When the plant breaks through the soil, the coleoptile soon stops growing, and the first of the true photosynthetic leaves, which have kept pace with the growth of the coleoptile, break through its tip. They are beginning to expand and develop into fully competent photosynthetic organs.

The light-induced inhibition of the growth of the mesocotyl and the concomitant promotion of the growth of the coleoptile are among the most intensively studied of the low-fluence responses mediated by phytochrome. Thus we found it puzzling that the responses of oat plants varied markedly from our expectations when we duplicated other investigators' protocols for growth and irradiation. An explanation soon emerged: a group of plants that germinated in absolute darkness acted differently from ones that had occasionally been illuminated by a green safelight. The plants exposed to the safelight tended to inhibit the growth of the mesocotyl and promote the growth of the coleoptile. They acted exactly as if they had "seen" a minute amount of red light.

The explanation is not unreasonable. P_r absorbs green light only poorly. Still, it does absorb some of the light. In fact, green light is 10 percent as efficient as red light at converting P_r into P_{fr}. Moreover, a plant's response to the formation of P_{fr} occurs over several orders of magnitude of the P_{fr} concentration. Even the conversion of a small proportion of P_r into P_{fr} by red or green light should elicit a detectable change in the plant's low-fluence responses. Several of the initial experiments were compromised by the light from luminescent watch dials. In other words, the seedlings were exquisitely sensitive to light.

What this meant for our experiments was that the further growing of oat seedlings would have to be done in absolute darkness. We therefore proceeded in what we called "reagent-grade darkness" to measure the response of the mesocotyl and the coleoptile of etiolated oat seedlings to increasing fluences of red light. The data showed that both organs had pronounced responses to minute amounts of light. In fact, we could measure responses when the fluence was four orders of magnitude less than the amount low-fluence responses are thought to require near their threshold. It appeared that the low-fluence responses were actually a composite of low-fluence responses and a second class of responses, which we call very-low-fluence responses. For one thing, the very-low-fluence responses proved not to be reversible by far-red light. The classical low-fluence responses (such as the germination of lettuce seeds) are quite reversible. Moreover, a very-low-fluence response could be elicited by the amount of energy in a single flash of firefly light. The classical low-fluence responses would require at threshold 10,000 such flashes, or the energy in one second of moonlight.

Finally, we calculate that the very-low-fluence responses begin when as little as .01 percent of the phytochrome in the plant is in the form of P_{fr} and are saturated when about .5 percent is in that form. Since even the purest available far-red light leaves about 3 percent of phytochrome in the P_{fr} form, it is hardly surprising that the very-low-fluence responses are not far-red reversible, and indeed that they can be brought about by far-red light itself. The threshold for low-fluence responses is about 2 percent, and the saturation of low-fluence responses comes at about 87 percent.

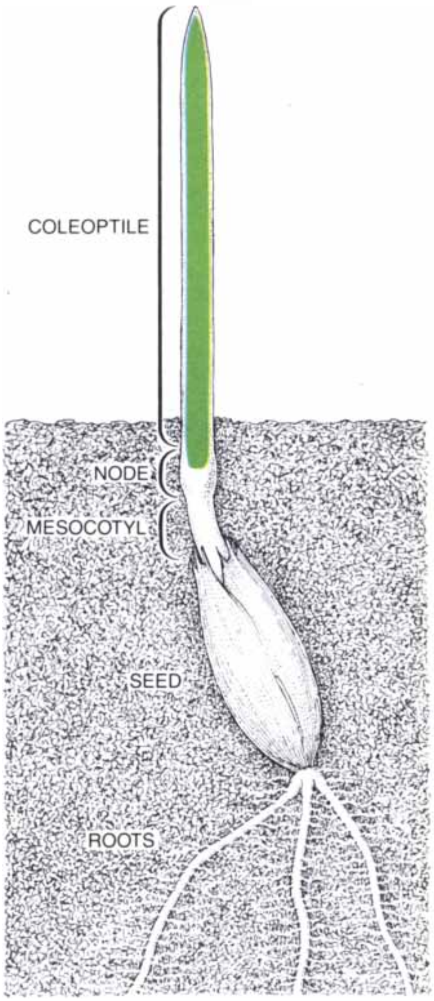
In the wake of these findings came the first suggestion that plant tissues act as fiber optics. The idea we had was simply to locate the site of photoperception in the etiolated plant. From the 1920's through the 1940's several groups of workers had made such an effort: they had shielded parts of plants from light and then had assessed the subsequent growth of either the entire seedling or parts of it. The responses to light in the low-fluence range always proved to be greatest when the node was irradiated. This result accords quite well with the distribution of phytochrome in the plant. As was shown in 1963 by one of us (Briggs), working with Harold W. Siegalman, and later by Robert E. Hunt and Lee H. Pratt of Vanderbilt University, phytochrome is found in all the tissues of oat seedlings but is most abundant in the node and at the tip of the coleoptile. The cells there contain about 10 times as much phytochrome as the cells in adjoining tissue.

Of course, no information was available for the very-low-fluence range. We therefore repeated the experiment, irradiating small regions along the length of a row of oat seedlings grown in the dark. In both the low-fluence range and the very-low-fluence range the results were unexpected. We thought we would find a maximum response when the light impinged on the node or the tip. Instead the maximum response came when the light impinged on tissue adjacent to the node. A possible explanation was that light was traveling to areas that were not intentionally irradiated. Yet given our method of masking, the only available path was through the plant itself.

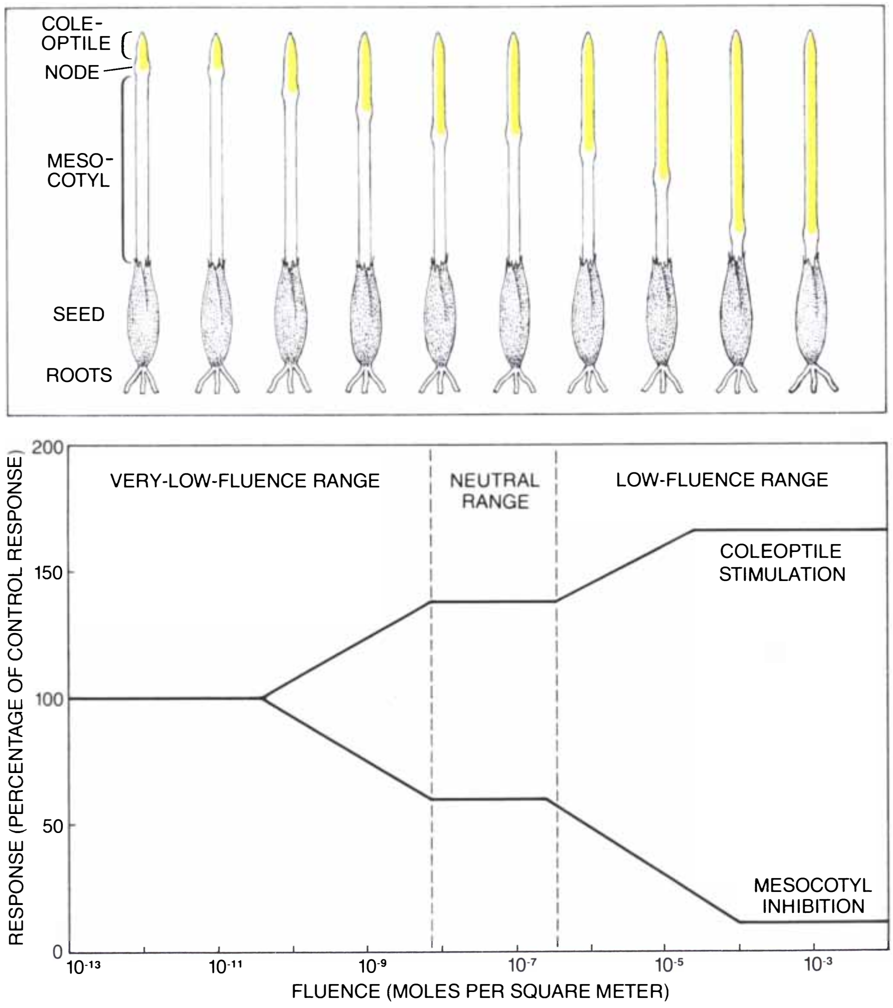
For the most part living tissues are poor transmitters of light. On the other hand, light-guiding in living organisms is not unknown. In the eye of flies the rod cells and the cone cells are known to guide light; they contribute to both the sensitivity of vision and the perception of color. In the leaves of plants the layer of cells called the mesophyll may guide light; it may thus augment the light available for photosynthesis. In the inner ear of vertebrate animals the hair cells of the organ of Corti guide light, but this guiding is surely unrelated to their function. (The cells serve the sense of hearing.)

Do the tissues of multicellular plants such as oats act as true light guides, that is, as fiber optics? The oat plant does grossly resemble a coaxial bundle of commercial fiber-optic strands. For one thing, the cells in the seedling tend to be arrayed in columns. Moreover, when the seedling is illuminated, the cell walls remain dark but the interior of each cellular column glows brightly.

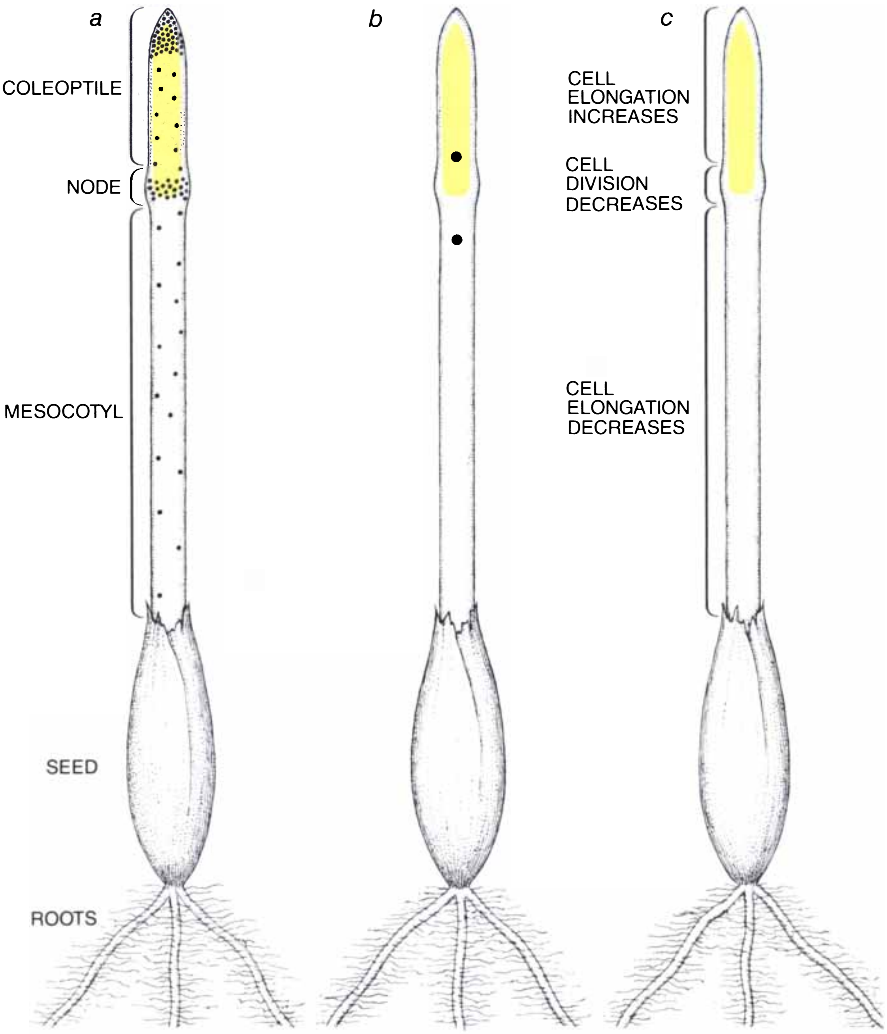
Nevertheless, the hypothesis was dif-



STRUCTURE of an oat seedling typifies the structure of many grasses. The shoot emerging from the seed includes the mesocotyl, or first stem, and the hollow cylinder called the coleoptile, which sheathes immature photosynthetic leaves. Connecting the mesocotyl to the coleoptile is the slight bulge called the node, which will give rise to additional leaves.



LOW-FLUENCE RESPONSES of a seedling demonstrate the modulation of the pattern of growth of the plant induced by small amounts of light; part of the modulation results from very small amounts of light and so represents what the authors call very-low-fluence responses. Two responses are diagrammed: the growth of the mesocotyl, which is inhibited by light, and the growth of the coleoptile, which is promoted by light. The vertical scale plots each such response as a percentage of the response of a plant grown entirely in the dark. The horizontal scale plots quantities of red light. The quantities are minute. For example, 10^{-9} mole of photons per square meter corresponds to the energy a firefly would emit in a single flash.



FIRST HINT OF LIGHT-GUIDING, that is, of fiber-optic properties, in plant tissues came when the distribution of phytochrome in oat seedlings (*a*) was compared with the sites in the plant that seem to be most sensitive to light (*b*): the sites whose illumination elicits the maximal low-fluence responses. It thus appeared that the plant was piping light. The responses of seedlings to light (namely changes in cell division and cell elongation) occur at still other sites (*c*).

ficult to prove. Our first attempts were hampered because we had to bend the seedlings. Otherwise part of the beam of light we directed into one end of the shoot would pass directly into the photomultiplier we positioned at the other end. The seedlings, which are only 1.2 millimeters in diameter, break easily when they are bent. We simply had to go on trying until we succeeded at keeping bent seedlings intact. Then a second problem arose: the handling and illumination of the seedlings tended to dry them out, and with dehydration came a diminution of the light-guiding capacity of the tissue. This diminution is a useful and accurate measure of water losses in living plant tissue, as John S. Boyer of the University of Illinois at Urbana-Champaign and we later documented. Initially, however, the dehydration was a nuisance.

Lasers producing light at a visible wavelength generate none of the infrared radiation that is partly responsible for drying out the tissue. Hence a low-wattage helium-neon laser suited our needs admirably: it provided a non-damaging, high-fluence beam at a wavelength near the peak of phytochrome's absorption in the red. Under such illumination oat, corn and mung-bean shoots all proved capable of light-guiding, but with varying efficiency. Still, we had no trouble detecting light-guiding through lengths of dark-grown seedling tissue as great as 4.5 centimeters.

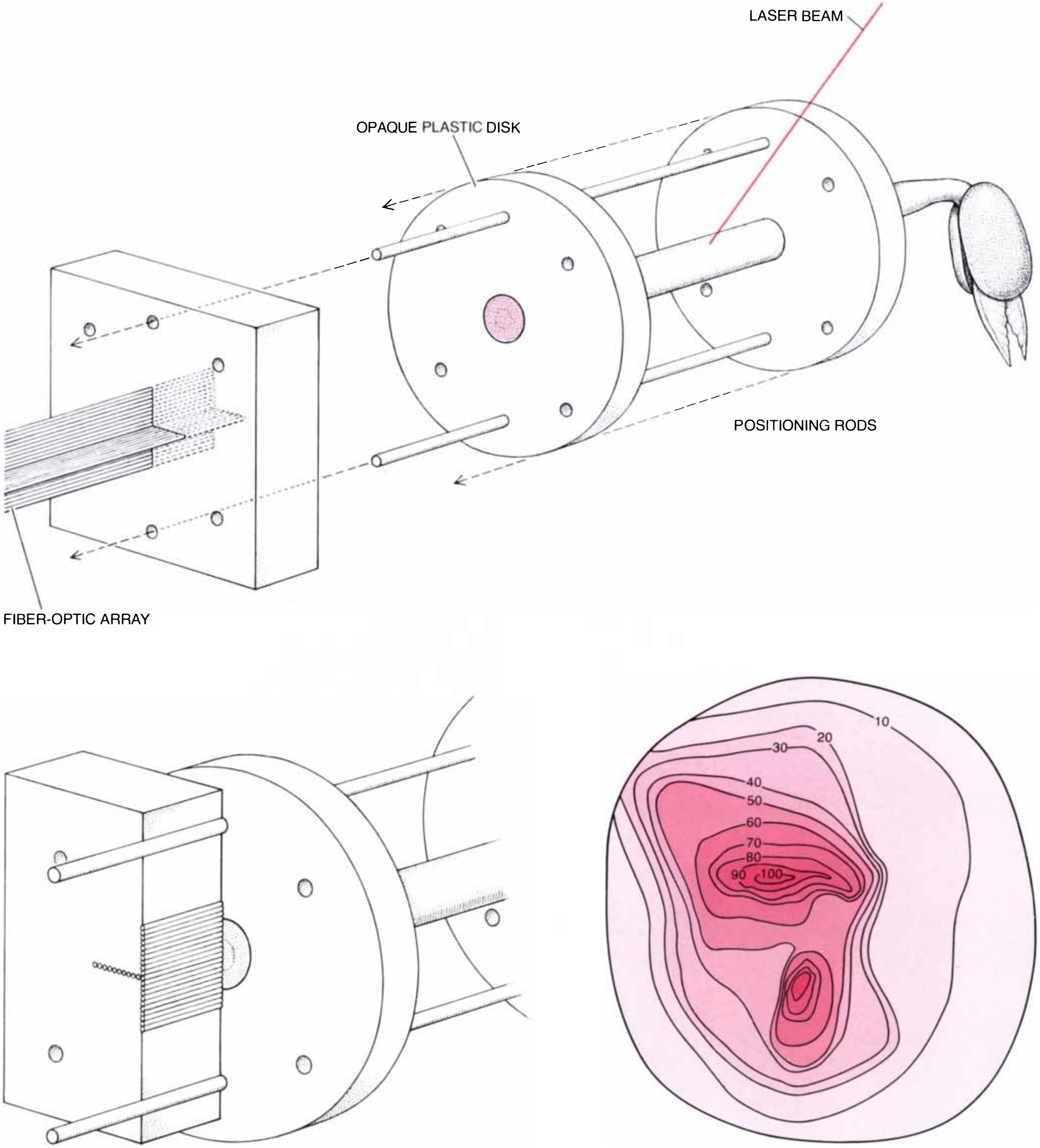
Now we could compare the properties of the seedlings with the properties of commercial optical fibers. Fundamentally a fiber-optic strand continues to guide light if the fiber is bent not because the light bends inside the fiber but because the light caroms repeatedly off the inner surface of the fiber, thus describing a zigzag path as it travels. At each reflection point the light encounters the interface between the fiber and the surrounding medium. Under normal circumstances part of the light would be reflected there and part would escape, although at an altered angle. In brief, part would be reflected and part would be refracted, but for a particular angle of encounter, which is called the critical angle, the refracted part of the light is directed precisely along the interface. Hence at angles exceeding the critical angle no light escapes from the fiber. All of it gets reflected. The fiber is then a light guide; it has the property of "total internal reflection."

Total internal reflection requires first that the refractive index of the light-guiding medium (a measure of the speed of light in the medium) be greater than that of the surrounding medium and second that the interface between the two mediums be smooth. Consider, then, a bundle of individual fiber-optic strands, and suppose the refractive index surrounding the bundle is changed to match that of the light-guiding medium of each of the strands. Only the strands at the perimeter of the bundle will lose their light-guiding property. The others, deeper in the bundle, will retain a working interface. If the bundle were a single, large-diameter fiber, the equalization in refractive index would eliminate all its light-guiding. Now suppose a bundle of fiber-optic strands is cut lengthwise into two equal bundles. Each will conduct half the amount of light conducted by the entire bundle. In contrast, the cutting of a single, large-diameter fiber roughens the only interface in the system, and so it eliminates light-guiding. All the plant tissues we tested acted in both respects as if they were multifiber bundles.

A further property of fiber optics is that the amount of light trapped by a light-guiding medium depends on the angle at which light arriving from outside the fiber first impinges on the interface. Light gets trapped in the fiber only if it arrives at the interface at an angle that equals or exceeds the critical angle. For manufactured fiber optics the experimenter's strategy is plain: one can tease a fiber out of the bundle and measure its critical angle directly. For tissues in which columns of cells appear to be the light-guiding units this strategy is useless. A given cell, say an oat cell, is only about .08 millimeter in diameter and as little as .1 millimeter long. Furthermore, the cut end of a plant consists of damaged cells, which scatter light.

Since a measurement of the critical angle was impossible for individual fiber-optic elements, we decided to try to measure the angular relation between light and the tissues of a plant, a relation more relevant to the physiology of the plant than the relation for individual light guides would be. We expected a complex result, but plants illuminated at different angles with a laser showed a strong angular preference. The pattern varied from species to species. Remarkably, however, the pattern seemed to be independent of the geometry and size of the tissue. For example, the roots of a corn seedling are covered by root hairs, the mesocotyl is a solid regular cylinder and the coleoptile is a hollow ellipsoidal cylinder, yet all three proved to have a range of acceptance angles with a maximum at an angle of 47 degrees. (The long axis of the seedling is taken to establish a reference angle of 90 degrees.) The factors other than refractive index that may be responsible for determining the angular specificity are still unexplored.

One final property of a fiber-optic bundle is that it transmits an image faithfully if the fibers in the bundle are coaxial. Otherwise (that is, if the fibers intertwine) the image is scrambled in transit. The first of these situations is

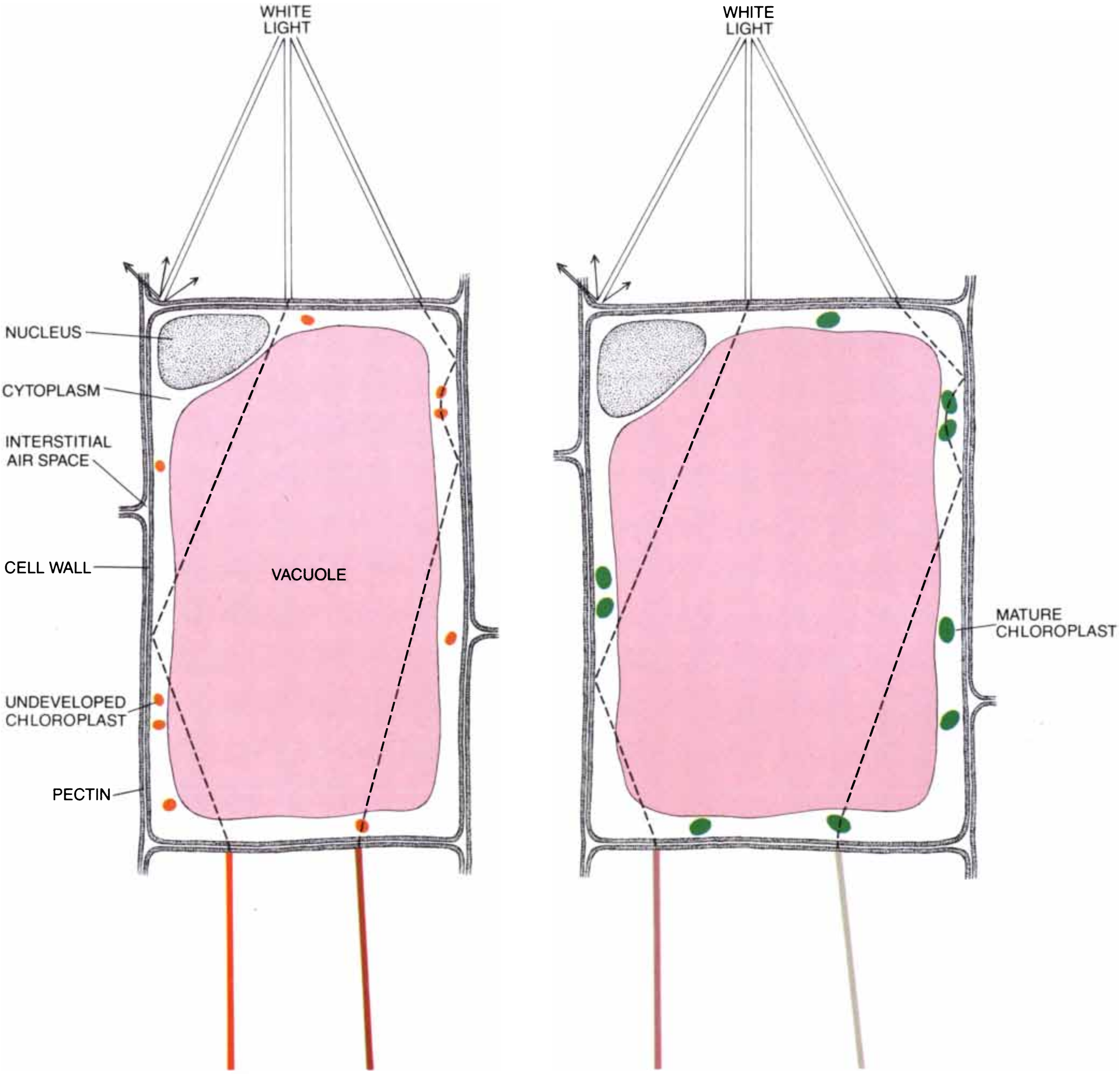


LIGHT-GUIDING IS MEASURED in the stem of a mung-bean plant by means of a laser beam and an array of commercial fiber-optic strands. The apparatus aligns the cut end of the stem with a cross of optic fibers. Each such fiber is .25 millimeter in diameter. A closer view of the alignment is shown at the bottom left; the pattern of laser light emerging from the tissue is mapped at the bottom right. The light is conducted coherently, that is, with its geometric pattern more or less preserved as it travels away from the point of irradiation.

referred to as coherence; the second is incoherence. Our measurements show that dark-grown plant tissues whose cells are long and fairly regular are often coherent image transmitters: they are capable of transmitting simple patterns of laser light.

Light that is guided away from the site of irradiation by tissues configured as coherent fiber-optic bundles may have physiological effects independent of those the light has at the site of irradiation. Consider phototropism, and in particular the fact that blue light impinging on the tip of the coleoptile of oat and corn seedlings causes the plants to grow toward the light. Since the early part of this century it has been assumed that the tropism occurs because there is more light impinging on the surface of the plant that faces the source of the light. Actually, however, our measurements show that the amount of light conducted down the opposite side of the plant can be two or three times greater than the amount conducted down the side the light actually hits. The tip of the coleoptile is composed of arches of cells. We reason that light tends to enter these arches at an angle fairly close to the maximum acceptance angle of 47 degrees. Perhaps the greater part of the light is guided along the arches to the opposite side of the plant.

The spectral quality of the light guided through plant tissues must also be important. For example, a change in the red component of the light that enters a given cell may yield a change in the ratio between the red and the far-red wavelengths in the light, and that in turn can



PASSAGE OF LIGHT through various compartments of columns of cells in beet plants was deduced from the absorption of the light at particular wavelengths by pigments sequestered in various parts of the cells. In an immature beet seedling (*left*) the light proved to pass through both the cytoplasm and the vacuole of the cells. The vacuole, an intracellular cistern, contains the pigments anthocyanin and betacyanin. In seedlings exposed to light (*right*) the development of chloroplasts reduced the transmission of light and changed its spectrum.

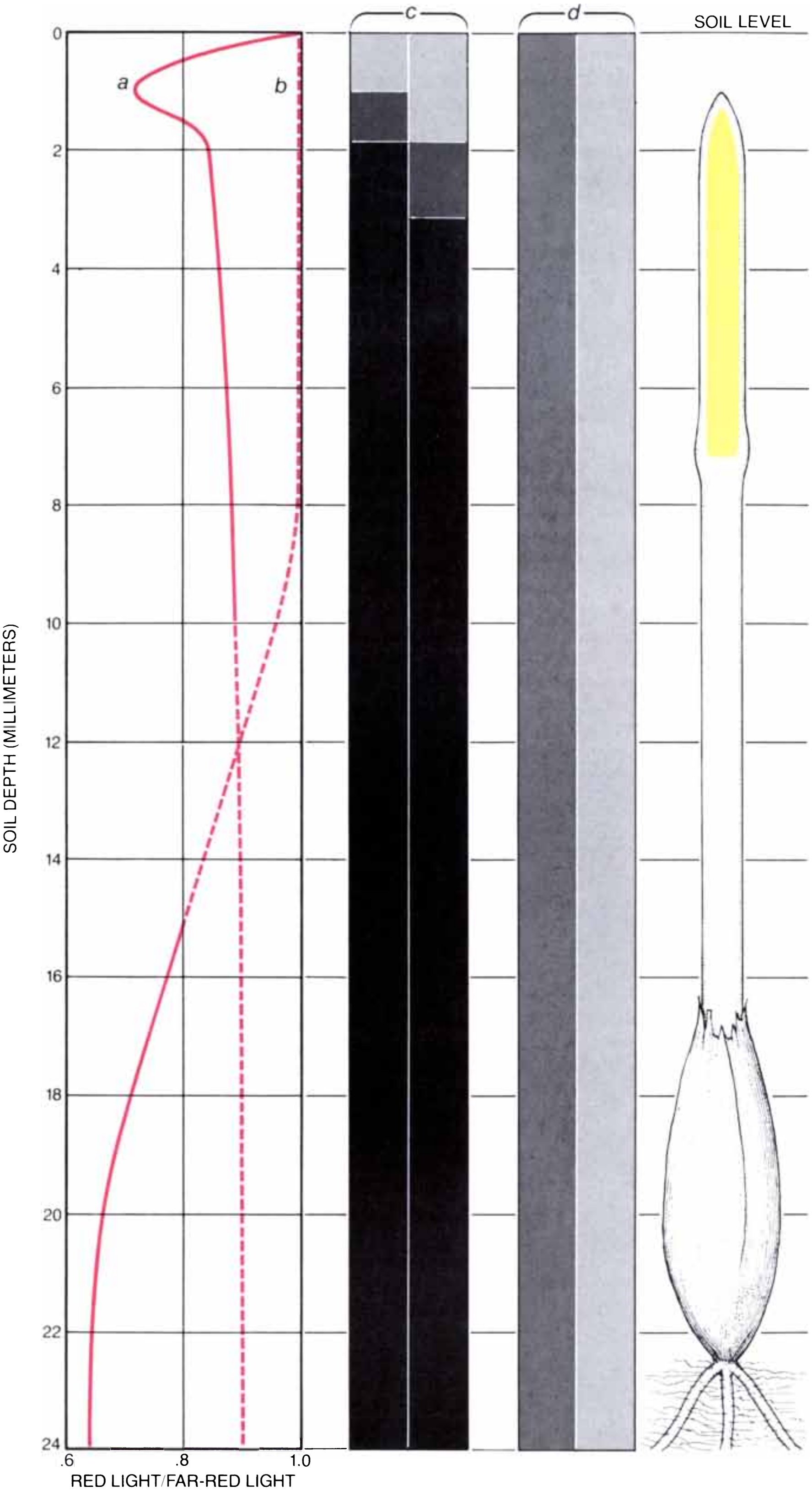
change the state of the phytochrome in the cell. The result is presumably a modulation of the growth of the cell; for example, the lower the ratio is, the faster the stem of a plant tends to elongate.

What changes the spectra of the light guided through a plant tissue? For one thing, they are changed by the pigments in plant cells. In cells from maple leaves and the ribs of beet leaves, red and purple pigments such as anthocyanin and betacyanin are found in the vacuole, a large, cisternlike intracellular structure. The green chlorophyll and the orange carotenoids, however, are found in organelles in the surrounding cytoplasm. A concentration of any such pigment would be in effect a filter. Thus measuring the spectrum of the light conducted through plant tissue would show where the light had traveled. We measured the spectrum of light guided through dark-grown beet seedlings; it showed the absorption of light at wavelengths characteristic of the anthocyanins and betacyanins. Some of the light had passed through the intracellular vacuoles.

Then we exposed the seedlings to light. They soon developed photosynthetic organelles; they manufactured chlorophyll and turned green. The spectrum of the light guided through the plant now showed additional absorptions due to chlorophyll and the carotenoids. Some of the light had passed through the cytoplasm. In fully greened plants the absorption of light sharply reduced the quantity of light-guiding and limited the wavelengths of the light to the far-red part of the spectrum. The significance of this phenomenon for the phytochrome in mature plants remains to be determined.

Let us return to the remarkable ability of grasses to position their node just below the soil line regardless of the depth at which the seed was buried. A full understanding of the phenomenon will doubtless require an understanding of the light-transmission properties not only of the plant but also of the soil. Surprisingly little is known about the latter. Accordingly Lawrence Waldron and M. Nemson of the University of California at Berkeley joined with us to gather light-transmission data for various types of soil. In sands and loams the quantity of light transmitted turned out to depend on hydration and on the size of the soil particles. The quality changed as well: the peak of the transmission spectrum underwent a "blue shift," and with it a complex change in the ratio between red and far-red light, with increasing depth or with decreasing particle size.

Consider, for the sake of simplicity, a buried seed that (unlike a lettuce seed) does not require light in order to germi-

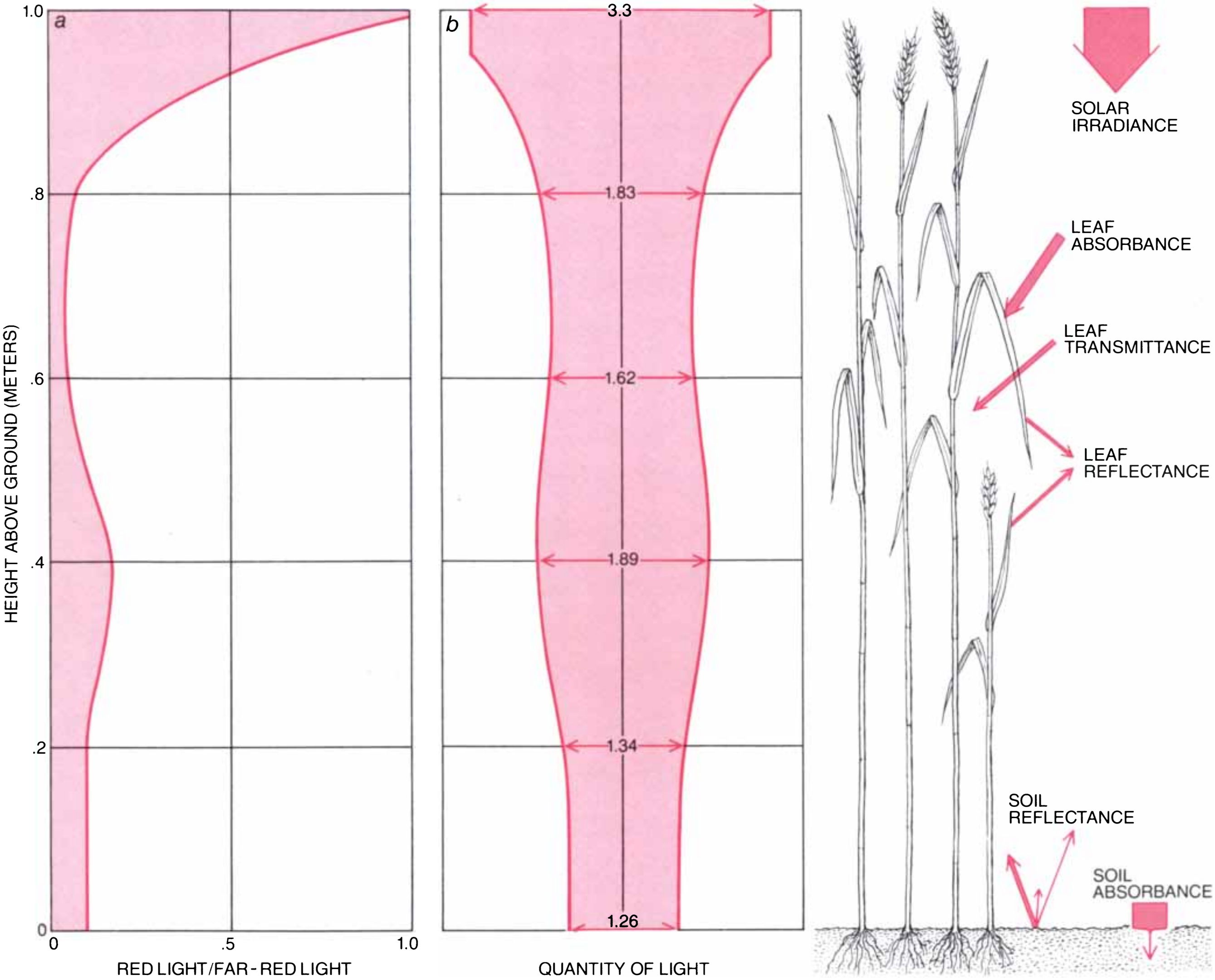


UNDERGROUND LIGHT ENVIRONMENT encountered by a seedling depends on the extent to which sunlight penetrating the soil gets piped inside the plant itself. At the left the spectral quality of sunlight (in particular the ratio of red to far-red light) is plotted for increasing depth in a sandy soil (*a*) and in an oat seedling (*b*). In addition the quantity of the light is shown as a set of grays: light gray where the quantity is enough to induce low-fluence responses, darker gray where the quantity is sufficient for very-low-fluence responses. One set of bars (*c*) shows conditions in the soil; another set (*d*) shows conditions inside the seedling. Each set is divided vertically: the left half corresponds to one second of sunlight at the surface, the right half corresponds to 10 minutes. Apparently one second of sunlight is enough to initiate very-low-fluence responses in an oat seedling even if it germinates at a depth of 20 millimeters or more.

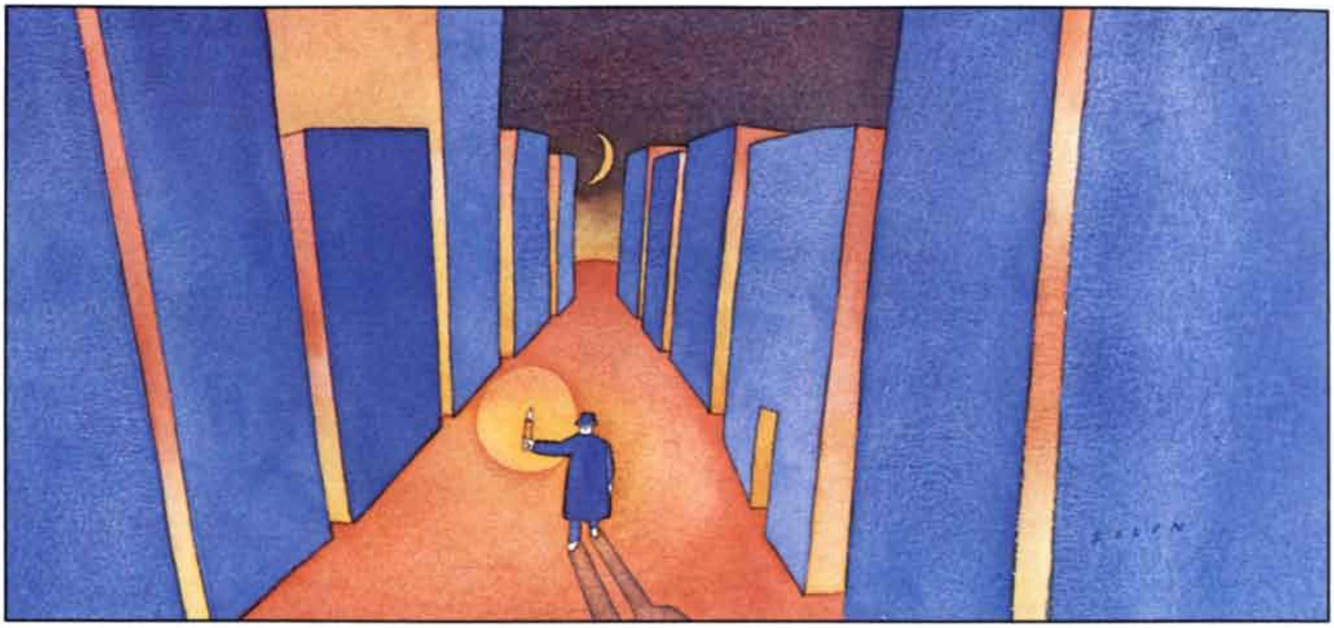
nate. Gravity guides its development; thus the shoot grows upward, entering in sequence the varying conditions created by the soil. Each cell is now affected not only by the local environment but also by the light guided to it inside the plant. Suppose the seed was buried in sandy California loam to a depth of about 20 millimeters. In the environment first encountered by the tip of the coleoptile the quantity of light is in the very-low-fluence range, and the quality of light, expressed as the ratio between red and far-red, is about .9. Light can be guided inside the plant with little attenuation and with essentially no change in quality over a distance of at least eight millimeters. In principle, therefore, the plant is aglow to the ends of its roots. Even under excellent growing conditions it will be many hours before the tip of the plant reaches the surface and has the prospect of being bathed directly by sunlight. Light-guiding may therefore give the plant some time in which to begin to develop the leaf tissues and the biochemical machinery it will need for photosynthesis.

The light environment of a dark-grown seedling is rather simple compared with that of the fully mature green plant. In fact, the mature plant lives in a range of environments. It gets sunlight and partial shade at its top; then in its own canopy it gets shade and bursts of sunlight, together with light and infrared radiation reflected from the ground. Finally, underground its root system, which can account for more than half of the mass of the plant, may get no direct light at all. Even there the mature plant may respond to light. We have found, for example, that some of the roots of the corn plant are almost horizontal in the dark but grow downward at an angle of 30 to 50 degrees if we expose them to red light. We think it is possible that the quantity and quality of the light transmitted through the soil is sensed by roots, which utilize the information to avoid the soil surface as they radiate outward, obtaining water and nutrients.

The half century that has passed since the use of light as a signal was discovered in plants has been spent on efforts to understand the complex physiology mediated by plant pigments such as phytochrome. The findings we have made are spurring an awareness that the optical properties of plants can also influence plant physiology. There are undoubtedly further surprises in store. In coming years we hope to see connections made among three aspects of the matter: the physics of light in plants (for example the physics of fiber optics), the transduction of light into chemical signals and the effects of such phenomena on plant biology.



ABOVEGROUND LIGHT ENVIRONMENT is actually a complex range of environments created by interactions of light with the soil and light with the plant canopy. Here a wheat field is shown. The quality of light (*a*) varies markedly with height above the ground. Again the quality is represented by the ratio of red to far-red light. The quantity (*b*) varies as well. The latter is measured on a logarithmic scale in which 3.3 signifies the sunlight at the top of the plants. In particular, it signifies 2×10^{-3} mole of photons per square meter per second. The data for the charts are taken from measurements made by M. G. Holmes and H. Smith of the University of Leicester.



THE NEXT ENERGY CRISIS
Will we have enough electricity?

With our economy growing, we're using more electricity. Unless plans are made now to satisfy this demand, we could have electricity shortages in the early 1990s.

Although some utilities have recently had more than the minimum generating reserves, these reserves are shrinking. They could become inadequate as our need for electricity grows faster than the utilities can provide it.

For decades, our use of electricity has grown along with our GNP. This year, they're growing at about 8 percent; at just 3 percent growth per year, we'll need 50 percent more electricity by the year 2000—and twice as many power plants as we're building now.

The high cost of running short

Electricity shortages would have a serious impact on our lifestyles and our economy, reducing industrial output, driving industries into other countries, putting people out of work, and limiting our ability to compete worldwide.

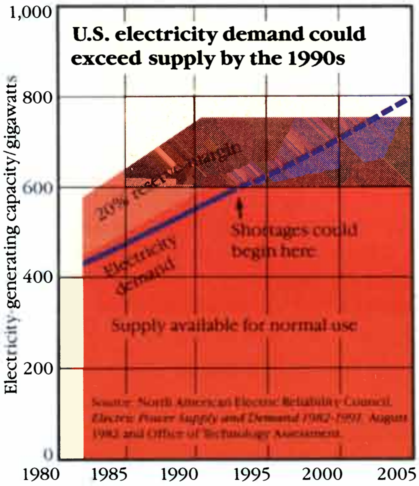
Some measures can postpone shortages, but they're costly:

- Delaying the retirement of older, less efficient plants
- Relying more on oil- and gas-fired turbines
- Importing more electricity

To *avoid* shortages, we need to finish the plants now under construction and plan for more new power plants—primarily coal and nuclear plants, since these use the economical domestic fuels that provide the capacity, economy, and security we need.

We're in a planning crisis

Due to energy crises, recessions and inflation since 1973, scores of coal and nuclear plants were canceled in the planning stage. Financial problems caused many utilities to cancel partially completed plants.



If demand for electricity continues growing at an average 3% per year (this year it's around 8%), in the 1990s the utilities' reserve margin could drop below the 20% necessary to assure a reliable electricity supply.

These plants and more will likely be needed to avoid electricity shortages in the 1990s. Yet utilities are faced with many obstacles and uncertainties.

Borrowing money for plant construction is increasingly difficult; state rate-setting procedures often discourage new capacity; opposition groups continue to block construction; Federal regulations continue to multiply, often causing the repeated tearing down and rebuilding of plant systems; and high interest rates and prolonged construction periods force electricity bills up.

These problems need to be solved through the understanding and cooperation of the industry, government, and the public. Only then will the utilities be able to take the steps needed to prevent electricity shortages and sustain a growing economy.

Free booklet tells more

Write: U.S. Committee for Energy Awareness, P.O. Box 37012 (D6), Washington, D.C. 20013.

Information about energy America can count on today
U.S. COMMITTEE FOR ENERGY AWARENESS

The Mechanical Manipulation of Randomly Oriented Parts

It is one of the main obstacles to the broader application of robots in industry. A computer system can now "see" an object at the top of a bin of mixed parts and direct a mechanical arm to pick it up

by Berthold K. P. Horn and Katsushi Ikeuchi

Consider the fine coordination between the eye and the hand of a young child who picks a cookie out of a jar. Although the cookies are roughly uniform in size and shape, the pile of cookies at the top of the jar is a jumble of visual cues, a rugged topography from which the child must extract enough information to determine what part of the visual or tactile field can be ascribed to the single, target cookie. As the child learns to take a cookie without crushing or breaking the ones around it, the child comes to realize that not every orientation of the hand can be successful. For example, seizing the edge of the cookie between thumb and forefinger works only if the center of the cookie is on or near the line connecting the opposing points of pressure. A much more reliable strategy is to determine the attitude, or orientation, of the cookie visually and then turn the hand to one of the positions best suited for picking it up. Finally, having grasped the cookie in one attitude or another, the child must transform the spatial coordinates of the cookie that pertain to the hand into the coordinates that pertain to the mouth.

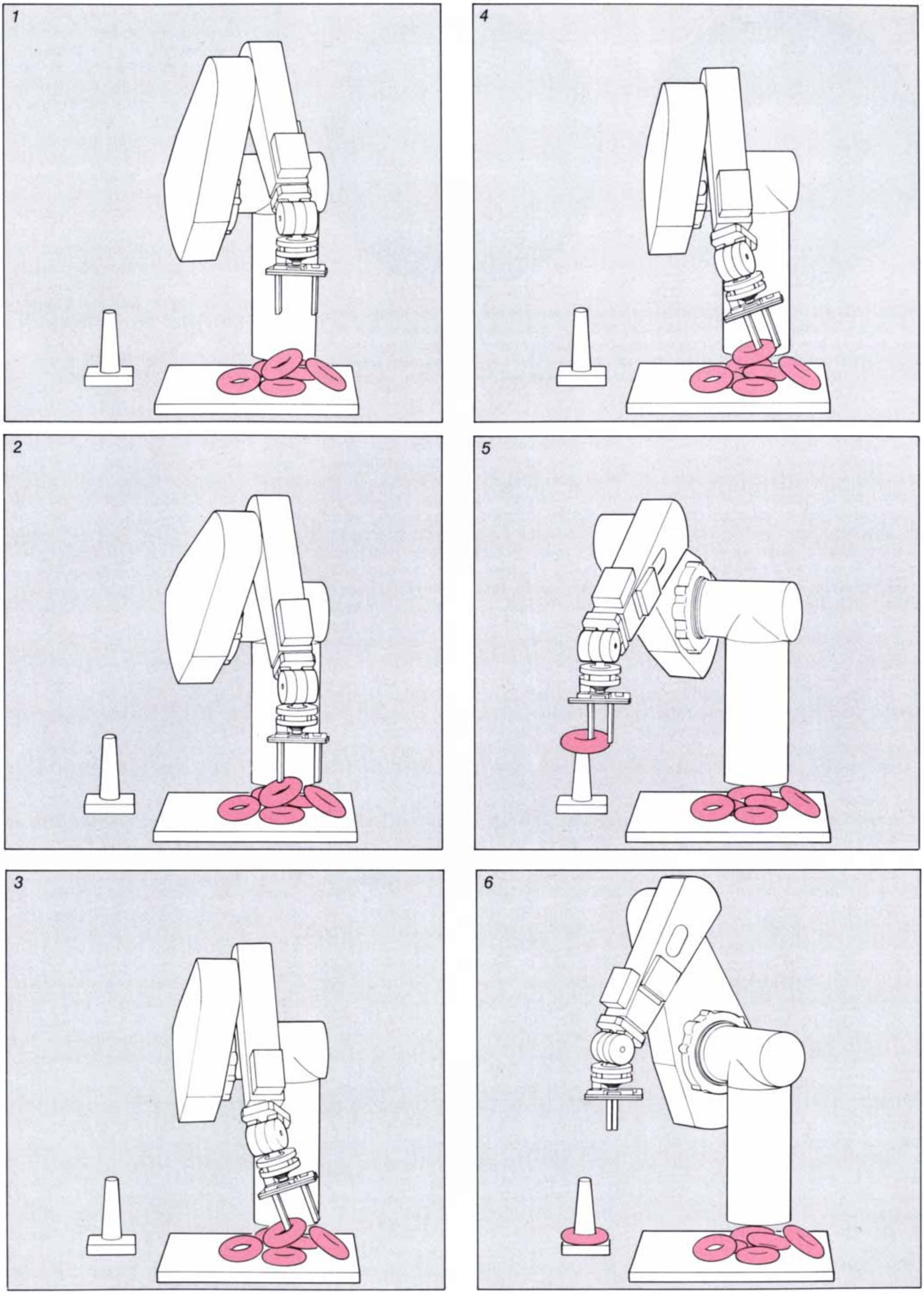
Until recently such a complex set of coordinated actions was beyond the capability of mechanization that seeks to replicate some of the functions of factory workers. The robot now working in the factory is fundamentally a playback machine for motions in space. To carry out a task the robot must first be "trained" by a person already skilled in the task. The "arm" of the robot is guided through a series of motions, and the sequence of robot configurations needed to follow the trainer is recorded on a tape or other memory device. When the tape is played back, it directs the robot to execute the same sequence of motions. The ability of the robot to record spatial motion has been exploited by choreographers to make a permanent record of dance movements, but without notable success. Nevertheless, the playback robot has found a niche in the factory because many industrial tasks are so highly repetitive that they can be done as a sequence of fixed motions. Mechanical manipulators have therefore been applied to spot welding, machine loading, painting, deburring, seam welding, sealing and other tasks that are boring or hazardous.

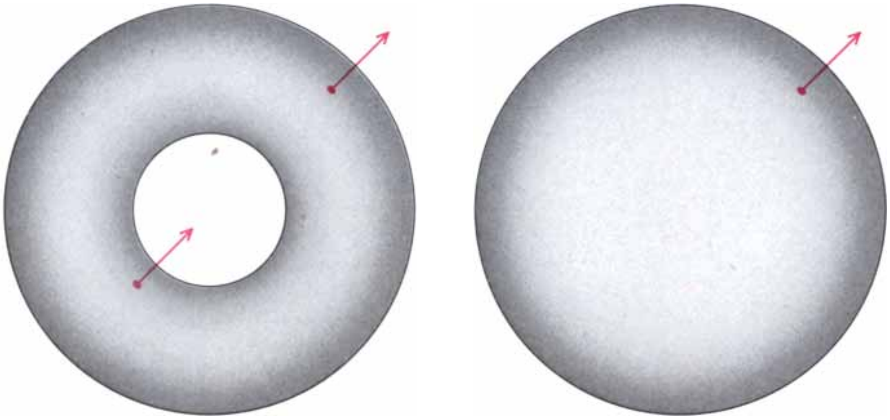
There is much factory work that cannot readily be adapted to a fixed routine of movement. In manual assembly, for example, it is common to have parts stored in trays or bins surrounding the work station. There the blind playback robot is virtually useless because it can tolerate very little uncertainty in the position of a part it must handle. An obvious solution to the problem is to avoid jumbling the parts together in the first place, or in other words to maintain a controlled orientation from the time they are made. There is a trend among manufacturers in favor of this solution: parts can be organized on carriers or attached to pallets on which they can be mechanically manipulated without the need for sensing. Nevertheless, the solution has its costs. The carriers or pallets must be designed and manufactured, often to close tolerances. Moreover, the pallets are usually heavy, they take up a large amount of space and they often have to be redesigned when the part they carry is modified. Indeed, the design of the part itself may have to be altered for the sake of automatic feeding. Suffice it to say there are many circumstances in which the volume of production has not presented enough economic incentive for the manufacturer to depart from more traditional, manual methods.

We have now developed a computer system that can determine the position of a part with an arbitrary shape in a randomly arranged pile. The system requires only a few electronic images of the pile of parts. The images are mathematically transformed by the computer into a form that is readily compared with a mathematical model of the part stored in the computer memory. The mathematical model is rotated by the computer until it closely matches the attitude of the object to be grasped. The results are applied to direct a mechanical arm to pick up the part. Such a flexible sensing system may be able to substantially extend the range of applications of industrial robots.

We are, of course, not the first to develop a sensing system that can be employed to guide the motions of a machine. Indeed, the first stage in our procedure is common to many other kinds of machine vision: we record a digitized image of the object on the image plane of an electronic camera. The image plane is made up of a large number of pixels, or picture elements, arranged in a regular pattern. The brightness of the object, which is called the gray level, is measured for each area that corresponds to a pixel in the image plane. The brightness values are quan-

MOTION OF A ROBOT that selects an object from a small pile of similar objects is depicted in a series of drawings based on photographs made by the authors. The object is a torus, or doughnut-shaped solid, which is difficult for most computer-controlled systems to recognize and pick up. The command that directs the arm of the robot along a ray in space is based on information provided by three images made by an electronic camera. A computer program determines the identity and orientation of an object and then finds the region that corresponds to the object in the image plane. The program also selects the points at which the robot is to grip the object. When an infrared beam passing from one side of the gripper to the other is interrupted, the motion of the arm along the ray is stopped. The gripper then maneuvers into position for the pickup, closes on the object and lifts it free. The object can be set down in any orientation.


1
2
3
4
5
6



ORIENTATION OF A SURFACE at a point is given by the direction perpendicular to a plane that is tangent to the surface at the point (*left*). It can be represented by the coordinates of a point with the same orientation on the surface of a unit sphere called the Gaussian sphere, after Carl Friedrich Gauss (*right*). Orientation can be defined for any point not on a crease or at a vertex of the surface. On the torus more than one point can have the same orientation.

tized, or rounded off, to one of as many as 256 gray levels.

In special cases it is sufficient to calculate certain properties of the object directly from the quantized image on the array. For example, in certain situations it is possible to distinguish points in the image that correspond to the object of interest from points that do not. Such a segmentation into object and background is usually based on differences in brightness. The resulting image is called a binary image because each pixel represents one of two states of the object: its presence or its absence. The binary image of an object is conceptually much like the pictures that are formed from an array of lights on a theatrical marquee or a stadium scoreboard.

Binary-image processing can be done with high-speed equipment of moderate cost. Unfortunately the binary image is often too crude a representation for it to serve as a guide to automatic manipulation. If the shape of the binary image is to conform even roughly with the silhouette of the real object, the contrast in brightness between object and background must be quite strong. If there is more than one object within the field of view, they must not overlap or touch; if the objects are not separate on the image plane of the camera, the silhouettes can change in unpredictable ways and the outline of the binary image may have little to do with the actual shape of a single object. Furthermore, unless the object has some rotational symmetry, the silhouette of the object can change in a complicated way when it is rotated in any plane except one that is parallel to the image plane. The information carried by the binary image of an object in an arbitrary configuration is in general too variable to be matched reliably with the representation of the object stored in the memory of a computer.

There has been substantial progress in machine vision since the first binary-image processors were demonstrated in the laboratory about 15 years ago. Nevertheless, the same strategic question about the design of such a system must still be faced: How can a symbolic description of the three-dimensional world be recovered from the quantized, gray-scale image recorded by an electronic camera? The form and detail required in such a description depend on its application. For picking randomly arranged objects out of a bin the description need give only the identity, position and attitude of the objects in space.

It is often thought that the identity, position and attitude of a part can readily be derived if the three-dimensional topography of the top of the bin of parts is known. It turns out the derivation is not straightforward, but the topography of the parts is still a first step in determining the description. The best-known cue for recovering three-dimensional topography from two-dimensional images is the depth perception afforded by stereoscopic vision. We can see in depth partly because we have two eyes that form images from slightly different viewpoints. A number of machine-vision systems attempt to exploit stereoscopic vision, but they are slow, complex and expensive, and they can deal only with certain kinds of images.

For practical applications machine vision does not have to emulate the admirable capabilities of biological vision. We have chosen instead to adopt a method invented at the Massachusetts Institute of Technology by Robert J. Woodham, which is called photometric stereo. The method determines the surface orientation of each small patch on the surface of an object but does not give the absolute distance to a point on the object. It turns out that for segmenting, identifying and finding the attitude of an object in space only local surface orientation is necessary.

The orientation of a surface at any point (except a point on a crease or at a corner) is defined by the direction of a line perpendicular to the plane tangent to the surface at that point. Every possible orientation of a surface corresponds to the orientation of some point on a sphere, and every separate point on the sphere has a different orientation. The line that gives the orientation of a point on an arbitrary surface is therefore parallel to the line that gives the orientation of some point on the sphere. It also follows that any spatial orientation can be specified by giving two coordinates, say the latitude and the longitude, of a point on a unit sphere called the Gaussian sphere, after the mathematician Carl Friedrich Gauss.

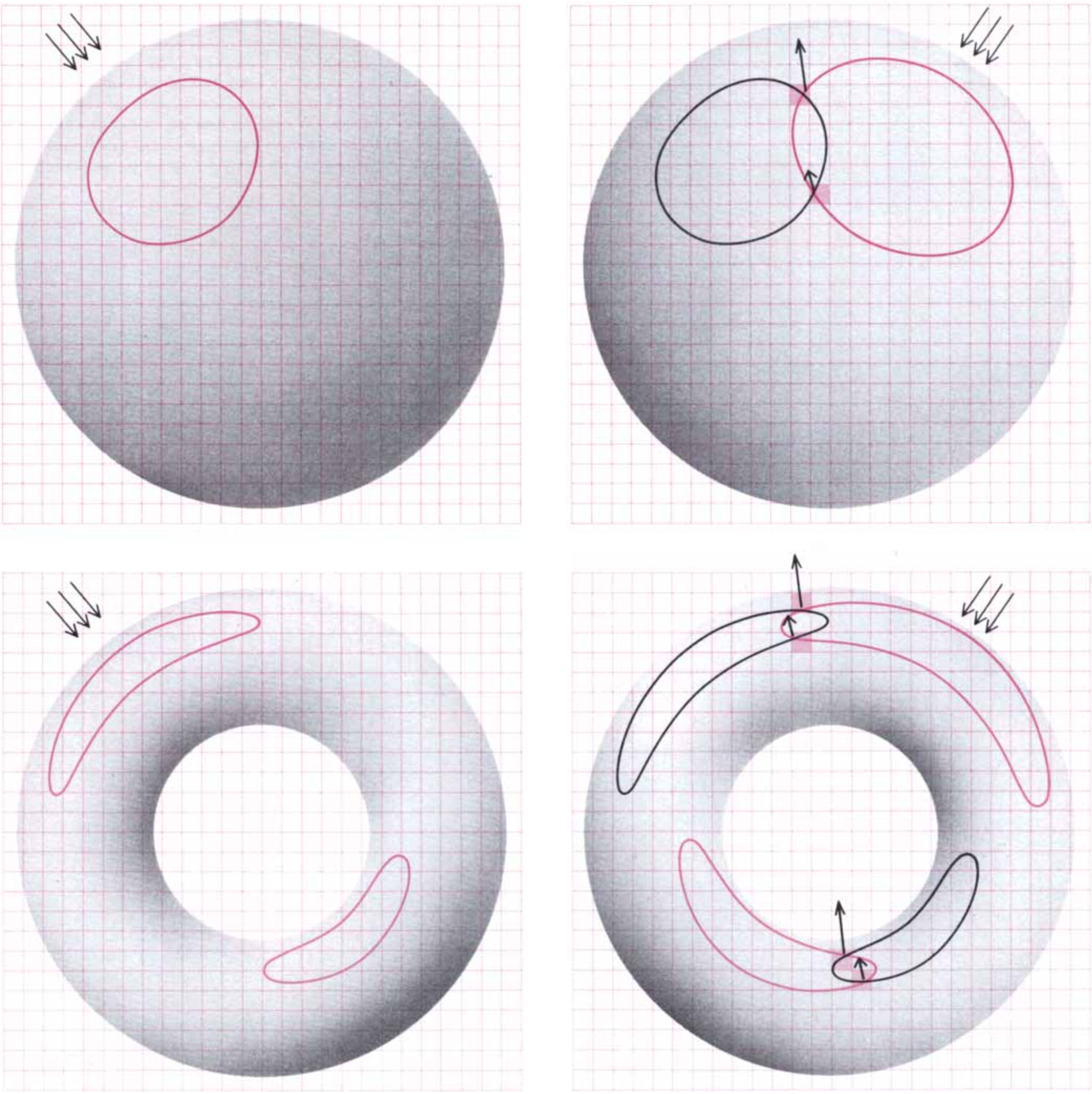
Suppose a Gaussian sphere is illuminated by a distant source of light. Furthermore, suppose the material on the surface of the sphere reflects all incident light and appears equally bright from all viewing directions. Since the light source is far away, the distance between the light and a point on the sphere does not vary significantly with the position of the point. The amount of light captured and reflected by a small patch on the surface of the sphere therefore depends only on the apparent area of the patch as seen from the light source. The apparent area depends in turn on the inclination of the patch with respect to the light.

Since the brightness of the spherical surface is assumed not to change with viewing direction, the brightest part of the surface for any viewer is the small patch around the point where the surface orientation matches the direction of the incident light, or in other words the point for which the source of light is directly overhead. The brightness of the surface decreases with the distance, measured on the spherical surface, from the brightest point. Patches of equal brightness form concentric rings around the brightest point because they are all inclined at the same angle to the light.

If the first light is turned off and a second distant light is turned on, the visible patches on the spherical surface are grouped into thin rings of equal brightness centered on a second point. Any small patch on the surface of the sphere that can be illuminated by both lights is thereby assigned two brightness values, one value for each light source. The first value limits the possible positions of the patch to a circle centered on the point directly under the first light; the second value assigns the patch to a second circle with a different center. The circles can intersect at no more than two points on the surface of the sphere. Hence for a given pair of gray levels there can be at most two corresponding points on the sphere, or in other words two orientations.

Suppose a table of values is constructed in which the brightness measurements made on the sphere are matched with the orientations to which they correspond. If a new object of arbitrary shape is put in place of the sphere, its surface orientations can be determined directly from the table. For each small patch on the new object a pair of brightness measurements are made, one measurement for each light source that was previously turned on to calibrate the surface orientations of the sphere. The orientations that correspond to each brightness measurement are then simply read from the table of values. The procedure is fast because the brightness measurements for all the surface patches of the new object can be obtained simultaneously from two images and because the data manipulation needed to determine the orientation of each surface patch from its brightness values is trivial. Moreover, the method can work for almost any object, no matter how complicated its surface or how strange the arrangement of the lights.

One obvious problem with the procedure is that the surface orientation of a patch is not uniquely determined. A



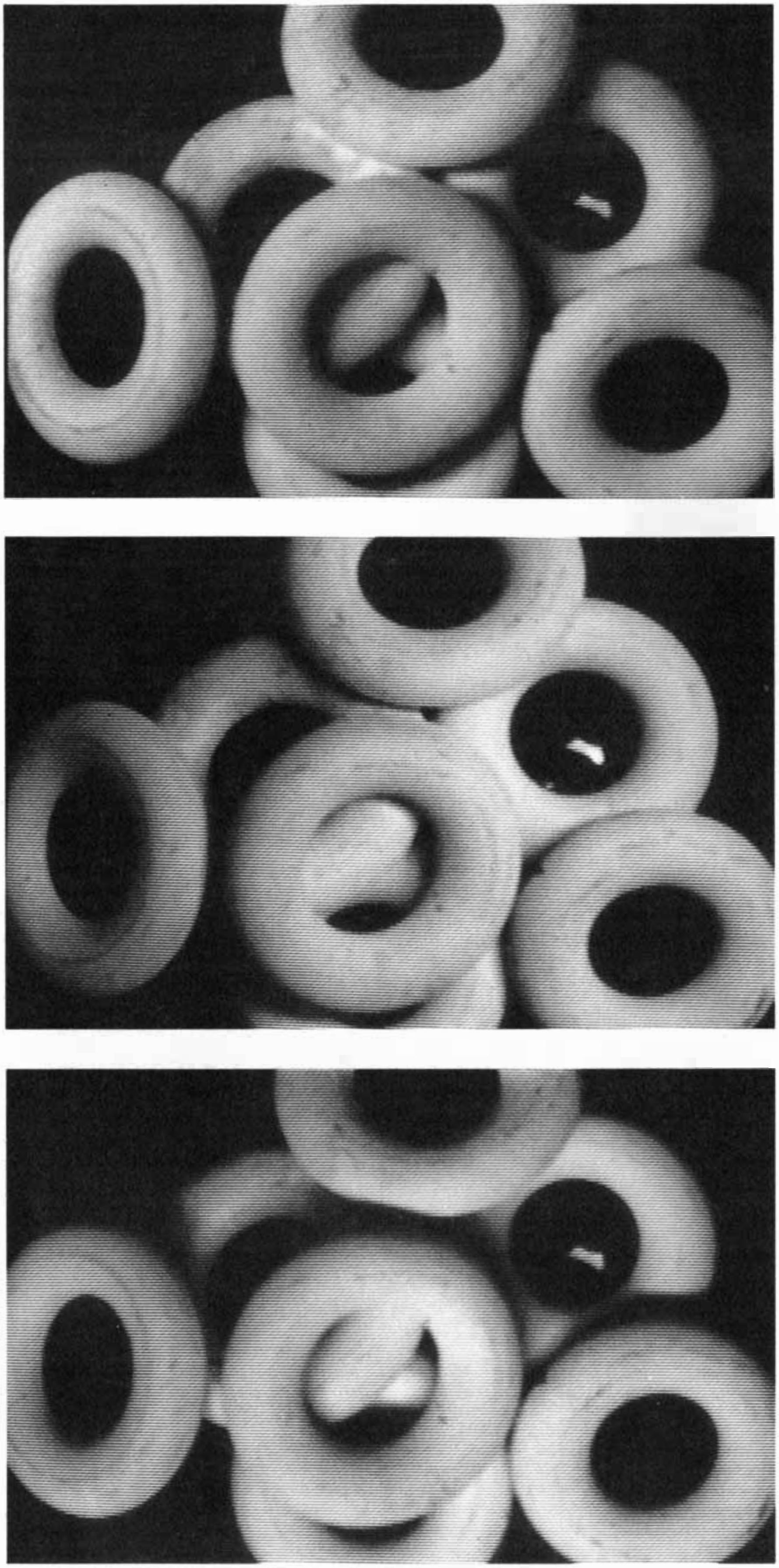
BRIGHTNESS MEASUREMENTS of the light reflected from any small patch of a surface can specify the orientation of the patch. A sphere is placed in the field of view of an electronic camera in order to calibrate the computer system. The orientations of the surface points of the sphere are known, and so each brightness measurement, or gray level, recorded by a picture element in the image plane of the camera can be associated with a known orientation. When the sphere is illuminated by one source of light, the contours of constant brightness on the sphere are concentric circles, one of which is shown in color at the upper left. A second brightness measurement with a light source from a different direction gives a second set of concentric circles (*upper right*). The two measurements specify a pair of gray levels for each picture element. Only two patches on the sphere, which correspond to the two picture elements where two circles intersect, can have a given pair of gray levels. A third measurement gives a unique orientation of a patch on the sphere. When the same measurements are made for an unknown object, the orientation of a surface patch is determined from the calibrated gray levels (*lower left, lower right*).

third light source can remove the remaining ambiguity, but the information it provides is far greater than that needed to distinguish two orientations. Instead of being content with overkill, one can exploit the three sources of light to derive additional information about surface properties. For example, if a surface reflects only a fraction of the incident light, and if that fraction, which is called the albedo of the surface, varies from point to point, each of the three brightness measurements gives rise to an equation with three variables. The variables are the two coordinates and the albedo of each point on the surface of a sphere. The system of three equations can be solved for the variables, provided the three lights and the illuminated object do not all lie in a plane.

If the brightness measured by each pixel of the camera is rounded off to one of, say, 16 values, there are 16^3, or 4,096, possible combinations of brightness values for each pixel when the brightness is measured for three sources of light. Most of the combinations, however, are not to be found in the lookup table. For example, no surface orientation of the sphere would correspond to the combination in which all three brightness values are maximum, unless all three sources of light were to impinge on the surface from the same direction. In that case, however, surface orientations could not be uniquely defined by the variations in lighting. Brightness combinations absent from the lookup table are nonetheless detected by the camera at some pixels, and such "impossible" combinations can be quite valuable in segmenting the image, or dividing it into regions that correspond to different objects.

One cause of anomalous brightness combinations is the shadowing of one object by other objects in the pile. A crude way to detect shadows is to assume that gray levels darker than a certain threshold in at least one image indicate a shadow. A second cause of anomalous brightness combinations is mutual illumination, the reflection of light from one object onto another; it is particularly common when objects of high albedo face one another. We assume that if the gray levels are brighter than the shadow thresholds, most observed combinations that are not found in the lookup table are caused by mutual illumination. The effect is generally seen near the edges of objects and along boundaries where objects tend to obscure one another; it can therefore be exploited for image segmentation. We also look for discontinuities in surface orientation and for high surface inclination, both of which tend to mark regions where one object obscures another.

Once a connected object of interest has been tentatively identified in the



REFLECTED LIGHT from a random pile of objects is shown for sources of light from three directions. The differences in shading for a given region of the surface are subtle to the eye, but they can readily be detected by electronic sensors. The photographs were made by the authors.

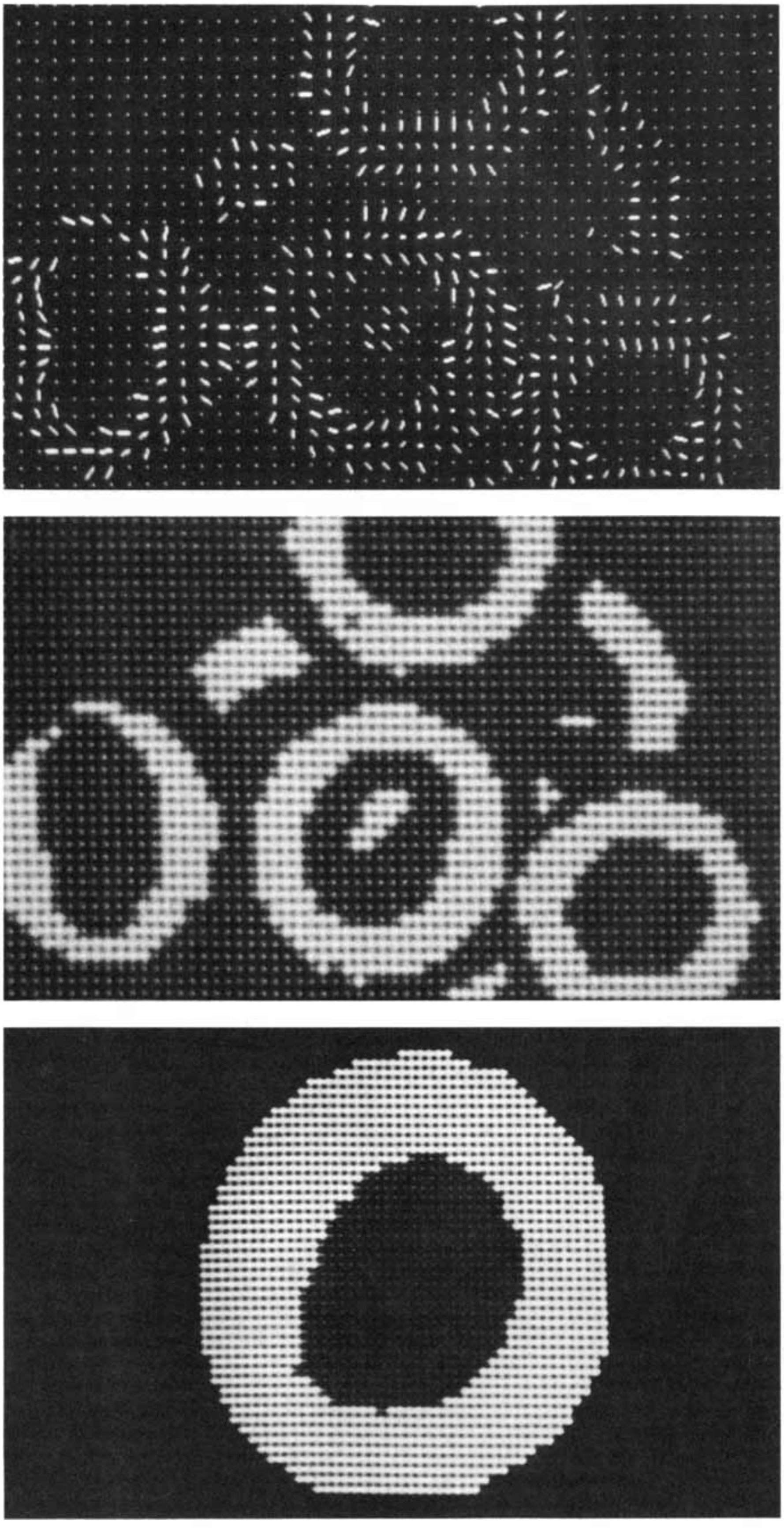
field of view, our goal is to match the observed object with one of the prototype objects that is abstractly represented in the memory of the computer. When the match is made, the observed object is identified. The only data available for making the match, however, are the position and orientation of surface patches on the object in view. That information can be represented by constructing a line perpendicular to the surface of the object at each point that corresponds to the center of one of the pixels into which the image is divided.

Suppose all the perpendicular lines have the same length. The shape of the observed object can be represented by the length and direction of the perpendicular lines as seen in perspective: the lines on the surface patches that face the viewer are represented as points, and the lines on the surface patches that slope away from the viewer vary in length with the sine of the inclination of the patch. The resulting figure resembles a surface covered with the quills of a porcupine; it is called a needle diagram.

It is costly and computationally inefficient to compare the needle diagram of the observed object directly with a needle diagram of a prototype object. Oddly enough, it is much more efficient to temporarily disregard the information that gives the relative position of various surface patches and focus instead on the surface orientations alone. A mathematical representation of the surface orientations called the extended Gaussian image, or EGI, is constructed from the needle diagram. The prototype objects are stored in computer memory in a similar mathematical form.

The EGI of any object is a sphere on which are plotted the relative contributions of each orientation of the surface of the object to the area of the surface as a whole. In order to identify the object selected in the field of view, the EGI of a prototype is abstractly rotated within the computer until it matches the EGI of the observed object as closely as possible. The same procedure is repeated for each prototype stored in memory. The observed object is assumed to be the prototype that gives the best overall match; the match simultaneously gives the attitude of the EGI for the object.

To understand how the EGI of an object is constructed, remember that any point on the surface of the object can be associated with a point having the same orientation on the Gaussian sphere. Similarly, a patch on the surface of the object can be associated with a patch on the surface of the Gaussian sphere by matching each point on the object with its corresponding point on the sphere. For example, wherever the surface of the object is relatively level like the flat side of an egg, the corresponding patch on the Gaussian sphere



NEEDLE DIAGRAM (*top*) represents the orientation of surface patches on the random pile of objects shown in the photographs on the opposite page. The orientation corresponding to each picture element in the camera is given by the direction of a needle, or line segment of constant length. The needles are shown as if they were attached to the surface at right angles like the quills of a porcupine and viewed from the camera. The computer divides the image into connected segments (*middle*), and one of the segments is isolated for further processing (*bottom*).

encompasses a relatively small fraction of the surface of the sphere. On the other hand, wherever the surface of the object curves relatively sharply like the end of an egg, the corresponding patch on the Gaussian sphere is relatively large.

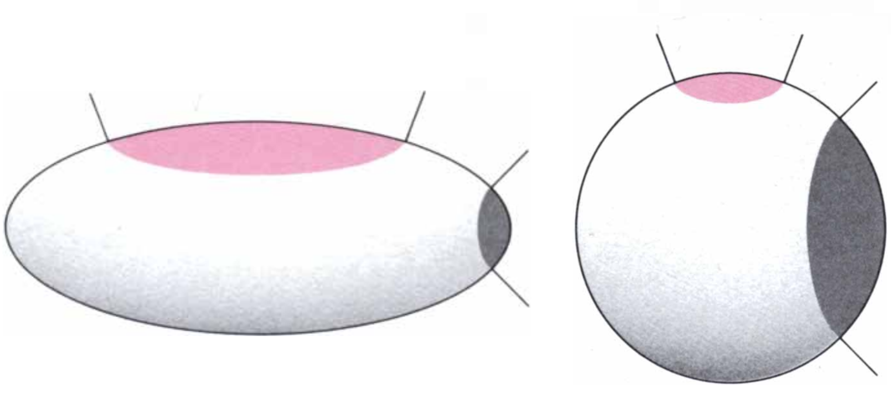
Imagine now that the egg is covered by a material of uniform density. To construct the EGI of the egg, the material from every patch on the surface of the egg is compressed or spread out in such a way that it fits exactly into the corresponding patch on the Gaussian sphere. The material on the flat region of the egg is compressed like a lump of clay in order to fit into a relatively small region on the Gaussian sphere. The material on the end of the egg must be spread out so that it fills a relatively large region on the Gaussian sphere. As the patches on the object of interest become progressively smaller the density of the material on the Gaussian sphere can vary continuously over its surface.

The visible hemisphere of the EGI, which corresponds to the visible surface of the observed object, can be numerically approximated from the needle diagram. The surface of the Gaussian sphere is tessellated, or divided into cells, and each cell corresponds to some small range of possible orientations. Every pixel of the needle diagram whose orientation falls within the range of orientations corresponding to one of the cells is assigned to that cell.

In determining the mass of material that is to be assigned to the cell, one must remember that the surface area of the observed object projected onto a pixel depends on the inclination of the surface with respect to the viewer. A surface that is steeply inclined away from the viewer is foreshortened and appears smaller than it would if it were viewed head on; one can correct for the effect because the angle of inclination is known from the needle diagram. The mass on each cell on the Gaussian sphere is then equal to the total mass of the parts of the observed surface that are visible in the pixels assigned to the cell. Since the material that covers the observed surface has a uniform density, its mass over any patch of the surface is directly proportional to the area of the patch. Hence the mass on each cell is also equal to the area of the parts of the observed surface to which the cell corresponds. The mass distributed over all the cells is equal to the total area of the observed surface. The tessellated Gaussian sphere is a quantized, or discrete, approximation of the EGI. It is called the orientation histogram.

The distribution of mass on the orientation histogram and, for that matter, the distribution on the EGI lead to a number of mathematical results that are useful in matching the observed object with a prototype. It is straightforward to calculate the center of mass for any visible hemisphere of the orientation histogram. (Note that this quantity has nothing to do with the center of mass of the visible part of the real object.) Since the orientation histogram of each prototype object in the memory of the computer is known over the entire Gaussian sphere, the center of mass can be calculated for any visible hemisphere and stored in memory. We generally do the calculation for each hemisphere that is visible when one of the cells in the tessellation is viewed head on.

Consider the plane that divides the Gaussian sphere into a visible hemisphere and an invisible one. The center of mass of the visible hemisphere lies at some distance D above the plane in the direction of the viewer. The product of the mass of the visible hemisphere and D is called the first moment of the mass about the dividing plane. Since the mass of the visible hemisphere is equal to the surface area of the object to which the hemisphere corresponds, the first moment is equal to the area of the visible surface of the object times D.



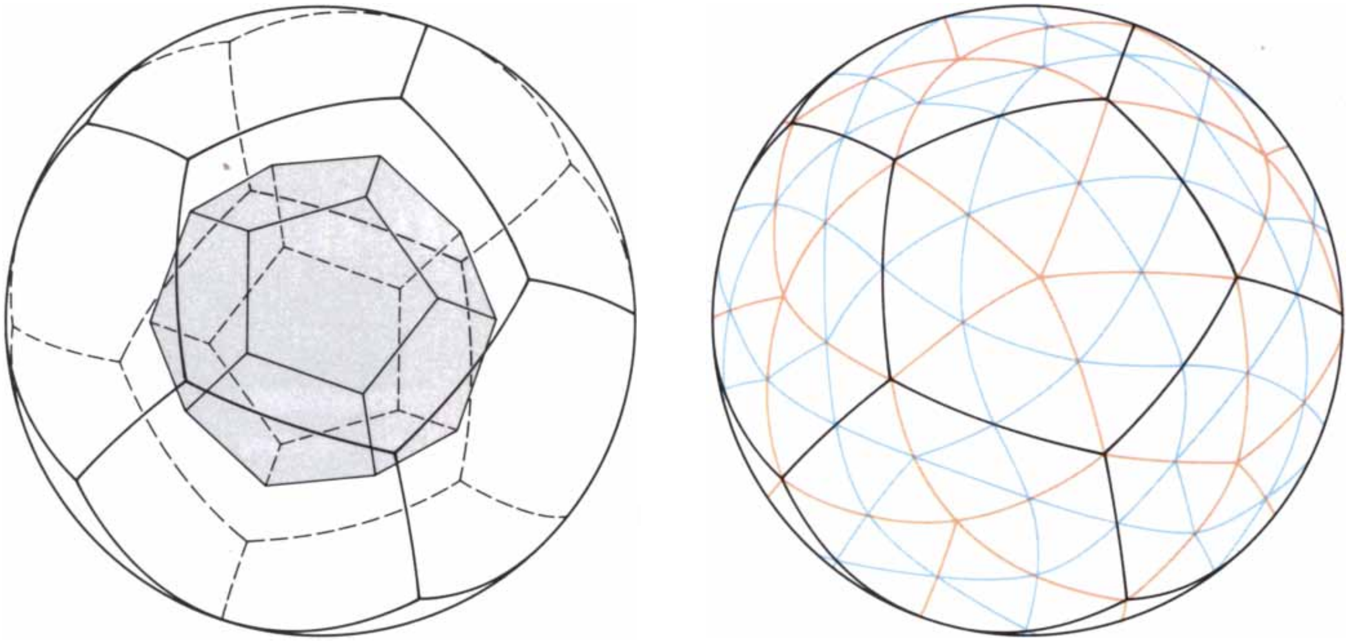
PATCH ON THE SURFACE of an object can be associated with a patch on the surface of the Gaussian sphere. Every point in the patch on the object is matched with the point on the sphere that has the same orientation. The patch on the sphere is a large proportion of the total surface area when the corresponding patch on the object is strongly curved (*gray*); the patch on the sphere is small when the corresponding patch on the object is relatively flat (*color*).

There is another way to represent the first moment of the mass of the visible hemisphere. Consider the mass assigned to any cell in the tessellation of the Gaussian sphere. The individual cell's first moment about the dividing plane is the product of its mass and its distance from the plane. Since the cell lies on the surface of the unit sphere, its distance from the plane is readily calculated. If the cell directly faces the viewer, its first moment is equal to its mass. If the cell is inclined from the viewer, its first moment is reduced by a factor that depends on its inclination; the factor is equal to the cosine of the angle between the orientation of the cell and the viewing direction. Remember that the mass of the cell is equal to the area of the parts of the surface to which it corresponds. When those parts of the surface are viewed, their actual area is also reduced by the cosine of the angle between their orientation and the viewing direction. It follows that the first moment of the cell about the dividing plane is equal to the cross-sectional, or apparent, area of the surface to which it corresponds.

The first moment of the mass of the entire hemisphere is equal to the sum of the first moments of all the visible cells; in other words, the first moment is equal to the cross-sectional area of the visible surface of the object. As we have shown, however, the first moment is also equal to the product of the actual area of the surface and D. The result is that D, which is a number defined for the orientation histogram, is equal to the ratio of the cross-sectional area of the observed object to its actual surface area. The ratio can be calculated directly from the needle diagram of the observed object.

The position of the center of mass for any hemisphere of the orientation histogram determines the value of D. Hence the observed ratio of the cross-sectional area of an object to its actual surface area can be compared with the values of D associated with various attitudes of the prototype objects. Although the value of D does not unambiguously give the attitude of the prototype matching that of the observed object, it does save computation. Any hemisphere of the orientation histogram for which the center of mass is not at least approximately in the right position need not be scrutinized further.

In general the attitude of an object can be specified by giving the direction of some axis that passes through it and the amount of rotation of the object about that axis. Since the number of directions for the axis and the number of rotations for the object are both infinite, one cannot compare the EGI of the observed object with all possible attitudes of the EGI of a prototype. Our matching procedure depends on sampling a finite number of the attitudes of the EGI that



TESSELLATION OF THE GAUSSIAN SPHERE can be done by projecting a regular dodecahedron onto the surface of the sphere (*left*). The dodecahedral tessellation is ideal for matching a representation of the unknown object with a prototype stored in the memory of the computer, except that the 12 pentagonal cells in the tessellation are too large. Smaller cells can then be generated (*right*) by dividing each pentagon into five triangles (*red*); the subdivision could be continued indefinitely by dividing each large triangle into smaller ones (*blue*).

can be derived from the tessellation of the Gaussian sphere.

The tessellation is constructed to meet several independent criteria. In order to distribute the surface orientations represented by the needle diagram in an unbiased way, the cells of the tessellation should ideally have the same area and the same shape. If the cells are relatively "rounded" polygons, such as pentagons or hexagons, instead of sharply angled ones, such as triangles, the range of orientations assigned to each cell can be minimized. Moreover, it is desirable that when one cell is rotated into the former position of another, the rest of the cells on the Gaussian sphere are permuted from their initial positions. In this way the rotation of the sphere can be represented in the computer simply by permuting the masses associated with the cells. All these criteria can be met by projecting the regular dodecahedron, which is a polyhedron with 12 pentagonal faces, onto the sphere.

Unfortunately, with only 12 pentagonal cells on the Gaussian sphere the tessellation is too coarse for the comparison with an unknown object. A finer tessellation can be constructed by dividing each of the 12 cells into five triangles; each triangle can in turn be subdivided into four smaller triangles. The resulting tessellation has 240 cells, and so about 120 cells cover any hemisphere of the sphere that is to be matched with the orientation histogram of the object.

There are many technical refinements to the matching process that we shall not discuss here. In general we align the orientation histogram of the observed object with the orientation histograms of the various prototypes. One way to measure the success of the match is to find the square of the difference in mass for each corresponding cell. The best match is the one for which the sum of the squares for all pairs of corresponding cells is a minimum. In practice we find that about 720 trials are needed to find the attitude of the EGI of an unknown object. The accuracy is between five and 10 degrees of arc.

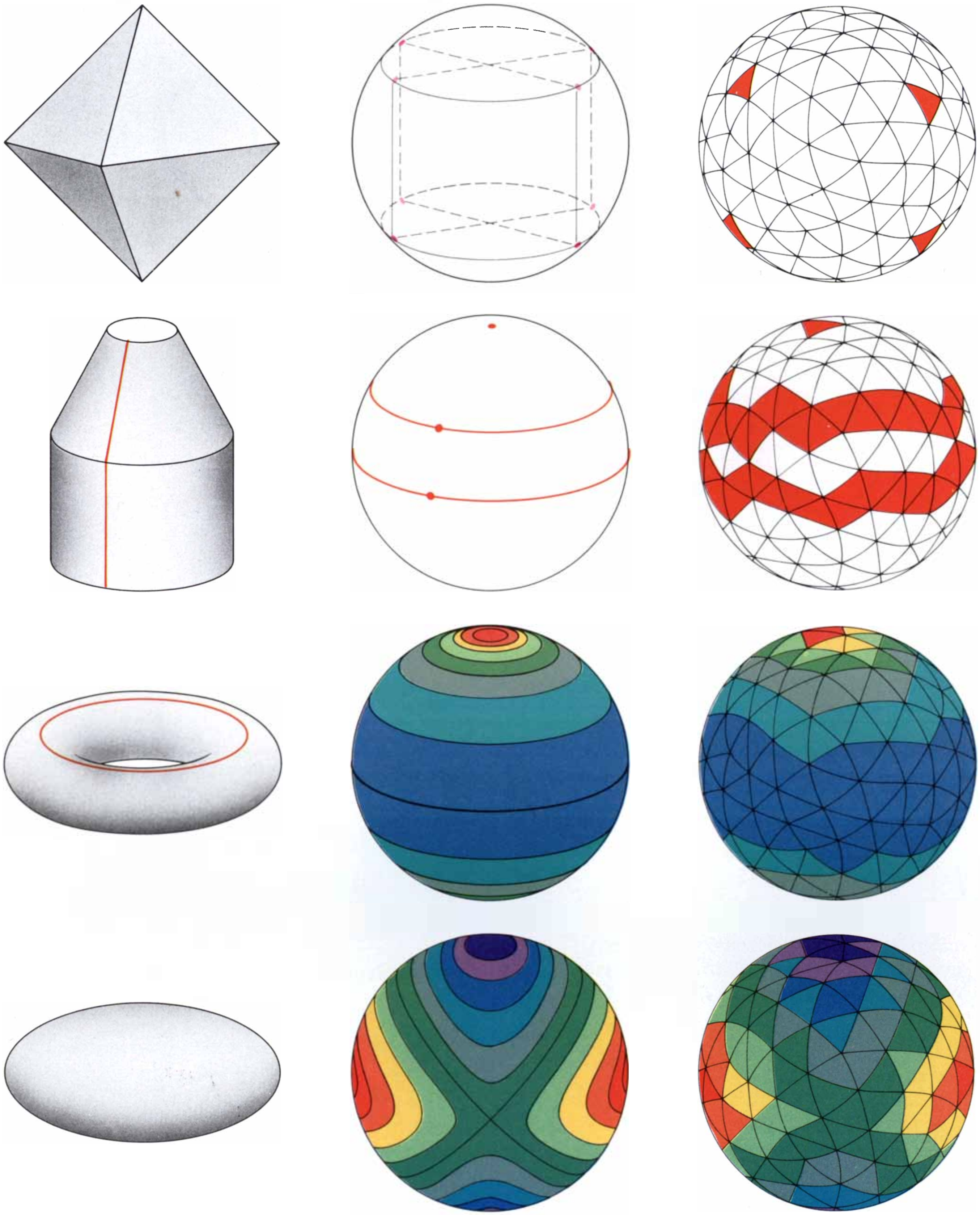
Once the identity of the observed object and the attitude of its EGI are known, the attitude of the object in space is also determined; the control of the robot arm is then relatively straightforward. The computer must determine which points on the surface of the object are most suitable for grasping. In part the decision is dictated by the shape of the object, but it is also desirable to choose points for grasping that are high on the object in order not to interfere with neighboring objects in the bin.

It is worth noting that the calculations we have described do not give the position of the object. Position can be roughly defined by the point in the center of the object region in the image plane of the camera. We find the position more accurately by calculating a needle diagram from the known orientation of the prototype. The calculated needle diagram can then be matched with the observed needle diagram.

The position of the object in the camera image defines a ray, or direction, from the camera. In order to command the arm of the robot to move along the ray, it is necessary to transform the spatial coordinates measured with respect to the camera into spatial coordinates measured with respect to the arm. We establish a general rule for the transformation by calibrating a few fixed points. The gripper of the robot moves a surveyor's mark, or two-by-two checkerboard, to several fixed points in two planes parallel to the image plane of the camera. For each point the spatial coordinates of the arm, which are determined by the sensors of the robot, are matched with the spatial coordinates measured by the camera. After the calibration is made each point in the image can be associated with a point in each of the two planes. The two points define a ray in the spatial coordinates of the arm.

The arm of the robot begins moving along the ray from some convenient height above the objects in the bin. Since photometric stereo does not give information about the absolute distance to the object along the ray, we installed a sensor on the gripper of the robot. The sensor is actuated by a modulated infrared beam of light that propagates from one side of the gripper to the other. When the beam is interrupted, the arm is stopped. The hand is then reoriented if necessary to match the attitude of the object, the gripper is closed and the object is lifted free.

Our system takes about a minute to switch the lights on and off, record the images, match the observed data with the prototypes and send the proper commands to the manipulator. There is no inherent reason the cycle time could not be much shorter. The calculations are



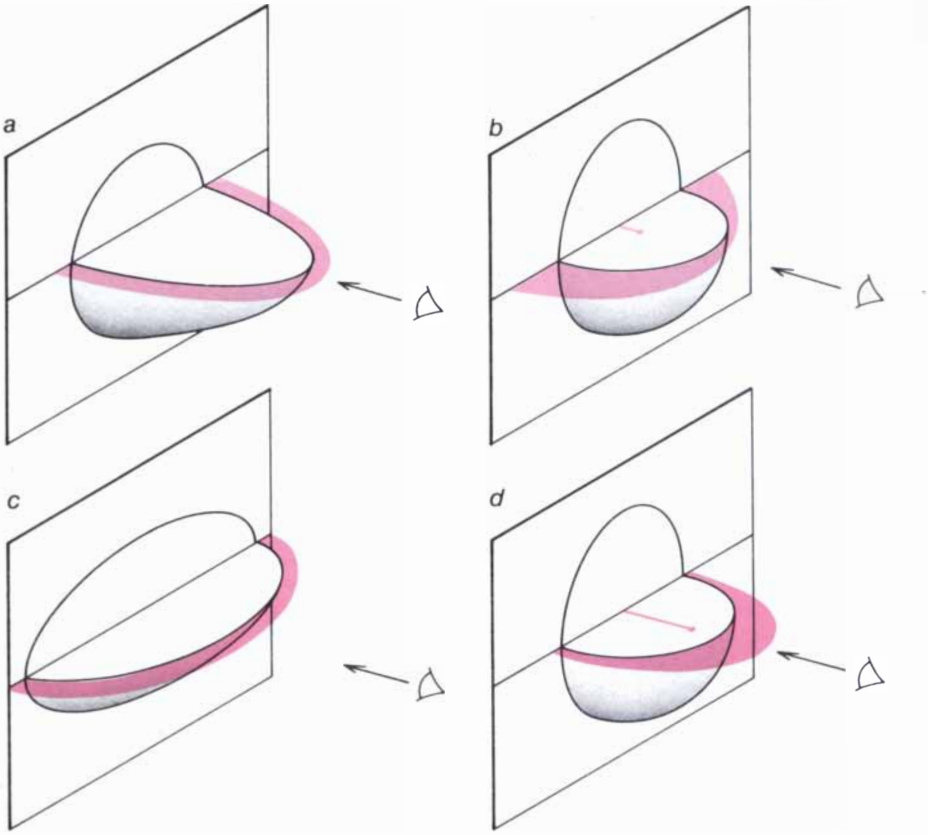
EXTENDED GAUSSIAN IMAGE (EGI) of an object can be pictured as a distribution of material over the surface of the Gaussian sphere. The material is initially spread evenly over the surface of the object. Each patch of material on the surface is then moved onto the sphere and compressed or spread out like clay to fit into the corresponding patch on the sphere. The EGI is shown in the middle column of the illustration for various objects. The regions of highest density are shown in red, and regions of lower density are shown in orange, yellow, green, blue and purple. For example, all the points on a face of a polyhedron have the same orientation, and so all the material from that face is concentrated at one point on the Gaussian sphere. The surfaces of a cone and a cylinder are each mapped into a circle on the Gaussian sphere; a line on the cone and a line on the cylinder parallel to the axis of rotation are each mapped into a point. The computer "perceives" the objects as they are shown in the column at the right; there the EGI is quantized on a tessellated Gaussian sphere.

simple, and special-purpose hardware could be built to speed the matching. Our aim was solely to demonstrate the feasibility of our approach to the problem, not to show how fast it can work. Most of the time the robot picks up one of the objects from the pile on its first approach. Occasionally the fingers of the gripper bump into another object before they reach the target; the arm then backs out of the field of view and the process is started again from scratch.

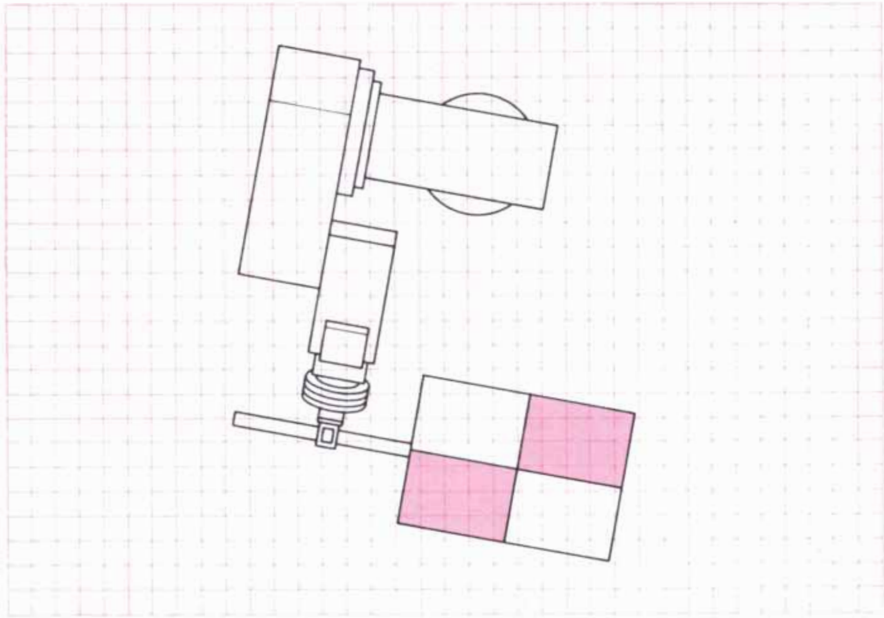
There are numerous ways our system could be modified, and many improvements will undoubtedly be made before it is adopted by industry. We have recently added a system designed by H. Keith Nishihara of M.I.T. that simulates stereo vision for determining the topography of a surface. The advantage of the additional system over photometric stereo used alone is that it gives some information about absolute depth. In another experiment we substituted laser range sensors for the lights and camera employed in photometric stereo. Both methods enable the robot to avoid moving along rays that intersect objects in the foreground of the target.

For many industrial applications, of course, the robot is too slow and its versatility is not needed. In such cases "hard" automation is the rule: special-purpose machinery is designed to orient a part. For example, small parts such as screws and other objects with a cylindrical geometry can be dumped into a vibratory bowl that can reject all configurations of the objects except the configuration needed. Large or heavy parts, however, as well as parts with complex shapes are not well suited to vibrational sorting. Moreover, a huge production volume may be necessary to justify the cost of such machinery.

We believe the system we have described is flexible and robust enough to be adapted to industrial tasks. It can reliably recognize objects and determine their attitude in space. The cameras and other necessary hardware are relatively inexpensive because only a few thousand pixels are scanned for each field of view. The computer program is largely devoted to pattern matching, and the patterns for the prototypes can be derived directly from data already present in programs for computer-aided design. Photometric stereo can readily be applied in the factory because it requires no special lighting conditions; extended sources of light can be placed in almost arbitrary positions, provided the gray-level calibrations are made after the lights are fixed in place. Moreover, the method is not limited to materials with particular light-reflecting properties. It remains to be seen what additional improvements will be made before machine vision is extensively applied to the mechanization of work.



CENTER OF MASS of the material (*color*) that covers a hemisphere of the extended Gaussian image depends on the area of the cross section of the object to which the EGI corresponds. For example, when an egg is viewed end on (*a*), the center of mass of the corresponding hemisphere of its EGI is relatively close to the plane that divides the Gaussian sphere in half (*b*). When the relatively flat surface of the egg is viewed (*c*), the center of mass of the corresponding hemisphere of its EGI is farther from the dividing plane (*d*). It can be proved that the distance from the center of mass to the dividing plane is equal to the ratio of the area of the cross section to the area of the visible surface of the object. Because that ratio is known, many orientations of the EGI of the prototype can be eliminated from further comparison with the unknown object.



COORDINATION BETWEEN EYE AND HAND of the robot is arranged by calibrating the spatial coordinate system of the robot arm with the spatial coordinate system of the electronic camera. A surveyor's mark is moved to a series of fixed points in the coordinate system of the arm, and the images of the points are given a second set of coordinates measured with respect to the camera. The robot and the surveyor's mark are shown as they appear to the camera through a coordinate grid. The computer calculates a transformation whereby the coordinates of a point in one system can be determined from the coordinates of the point in the other.

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.™

Cooperative Breeding in the Acorn Woodpecker

The birds share mates and raise their young in groups. Study of the acorn woodpecker's unusual social system shows how natural selection yields both cooperation and competition

by Peter B. Stacey and Walter D. Koenig

In the northern temperate regions of the earth almost all species of birds breed in monogamous pairs. The two birds, either singly or together, choose the site for a nest, build the nest, gather food, incubate the eggs, feed the young and guard them. Thus in most species each monogamous pair of birds constitutes a more or less self-contained social unit. A notable exception to this pattern is the acorn woodpecker (*Melanerpes formicivorus*), a common and conspicuous resident of oak and pine-oak woodlands in the southwestern U.S., Mexico and Central America. The social unit of the acorn woodpecker is a territorial group that can include more than a dozen members. In winter the group's main source of food is mast: acorns or other nuts stored in holes drilled in a tree that serves as a storage facility. Such a tree, which is called a granary, is held collectively: the drilling of new holes, the storing of mast and the defending of the tree against intruders are done by all the group members.

Furthermore, in the acorn woodpecker group both mating and the rearing of young are also collective tasks. The sexually mature birds in each group are divided into breeders and nonbreeding "helpers." Among the breeding adults mate-sharing is common. For example, in a group with three breeders of each sex all three males apparently can breed with any female. Thus acorn woodpecker young are truly progeny of the group, and the adult members of the group all contribute to the raising of the young birds. The degree of mate-sharing and the ratio of helpers to breeders depend partly on the climate of the region and partly on the available food resources. As a result there is considerable variation in group structure from one region to another.

The overall mode of breeding of the acorn woodpecker raises difficult questions for the evolutionary biologist. Two central questions are: Why do the helpers remain in the group during the breeding season instead of leaving to reproduce on their own elsewhere? Why do the breeding adults cooperate in raising nestlings that may not be their own offspring? Most students of evolution agree that in general selective pressure operates at the level of the individual organism or at the level of the genes and not at the level of the social group or the species. Evolution encourages reproductive strategies that enable an organism to contribute copies of its genes to the next generation. A bird's serving as a nonbreeding helper or raising young that are not its own offspring appears to run counter to evolutionary pressure. Why, then, does such behavior survive in the acorn woodpecker? The answer can illuminate how underlying selective pressures are manifested in social systems that appear to be paradoxical.

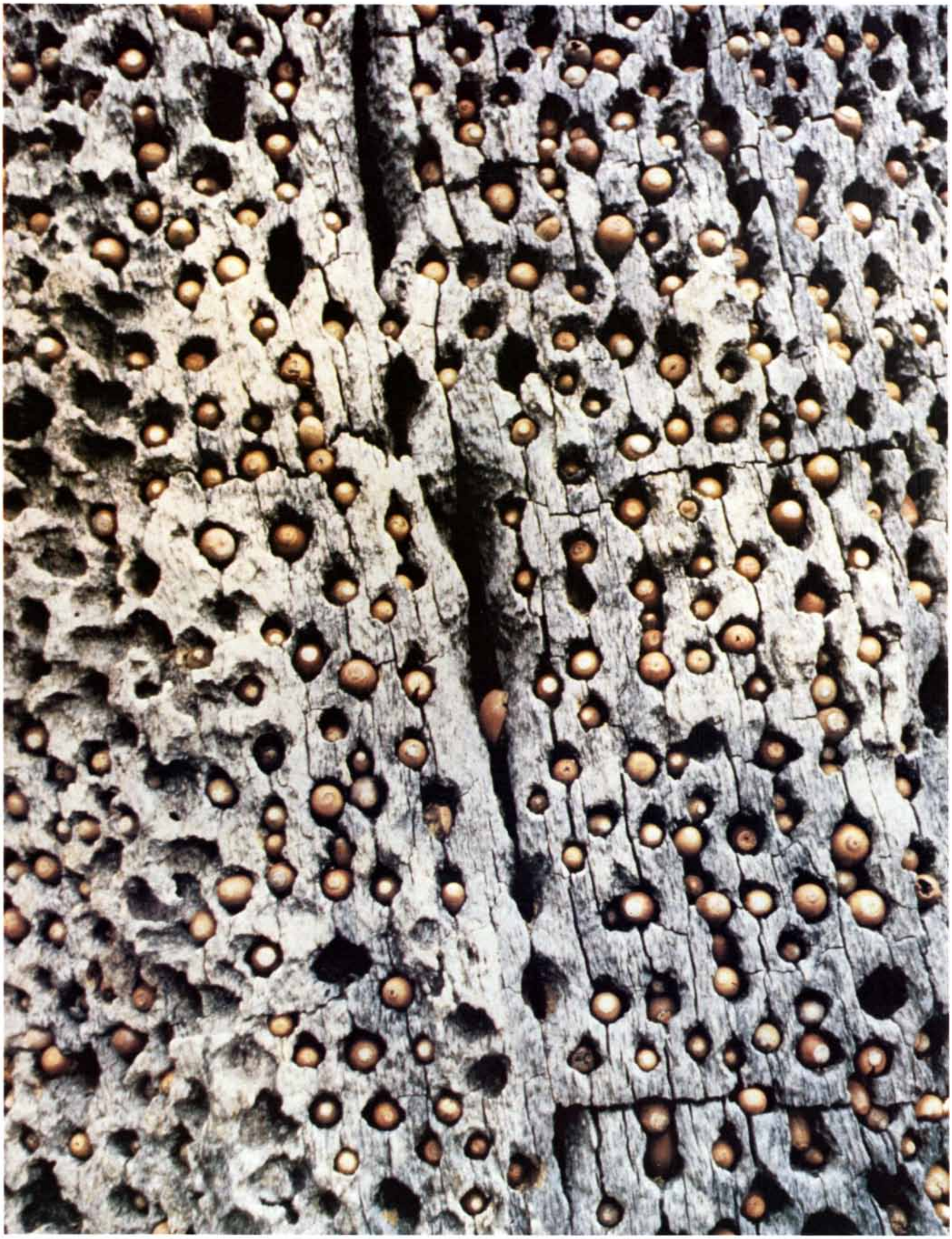
Over the past decade we and our colleagues studied the cooperative breeding of the acorn woodpecker in three different U.S. habitats. One of us (Stacey) worked mainly at Water Canyon in the Magdalena Mountains of central New Mexico and also at the Appleton-Battell Research Ranch in southeastern Arizona. The other (Koenig) worked at the Hastings Natural History Reservation in central coastal California. In most areas where acorn woodpeckers have been studied they live in permanent social groups consisting of as many as 15 members. Each group occupies a territory that is vigorously defended from all other acorn woodpeckers. There is generally only one active nest in a group at any one time.

The feature of acorn woodpecker behavior that first attracted the attention of ornithologists was not the fact that the birds live in groups but their habit of storing large quantities of acorns, piñon nuts and other mast in a granary tree. In the fall the group members harvest the mast and wedge it into holes drilled in bark or in a dead tree. Some granary trees have more than 30,000 holes. Storage facilities with this kind of capacity enable the group to protect large quantities of food from the effects of weather and from other animals.

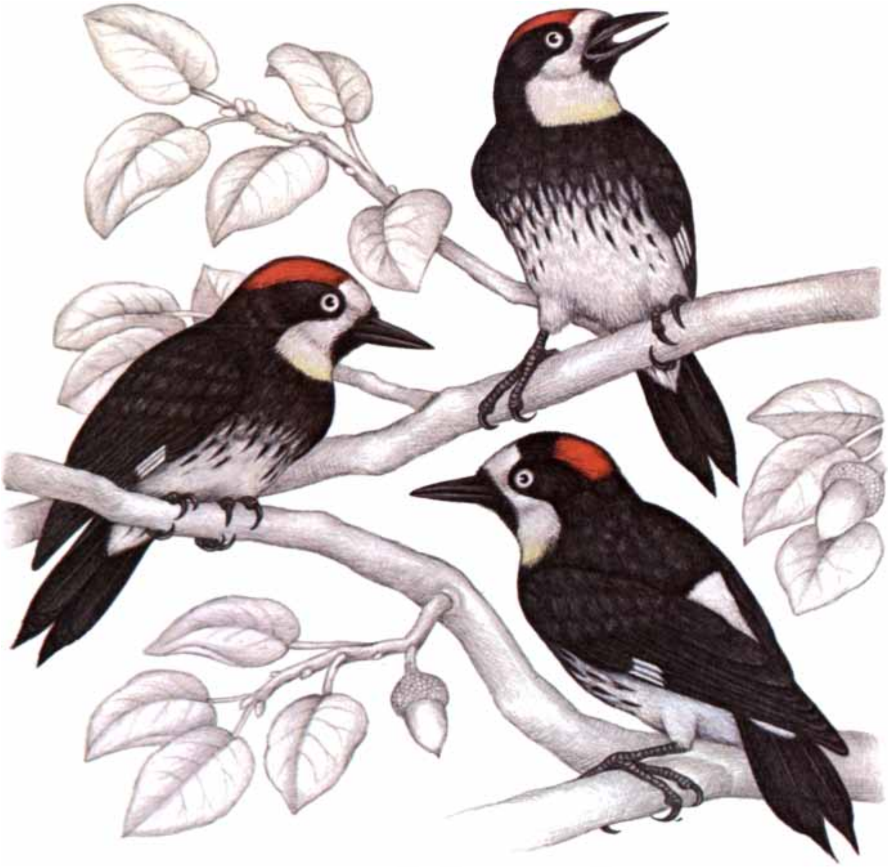
When ornithologists first noted how the acorn woodpecker stores mast, some of them suggested the birds could be employing the mast to farm insects. It was argued that as the mast decomposed over the winter it would be invaded by insect larvae, and the woodpeckers would have the insects as a ready food source. Closer field observations showed, however, that the birds regularly move the mast from hole to hole to keep it from rotting or falling out of the tree as the nuts dry and shrink. The mast itself serves the woodpeckers as food, and mast that has been invaded by insects is rejected.

Any bird that is a member of the group can feed on the mast at any time. Over the winter the stored mast is a highly significant resource for the survival of the group. In the spring and summer the woodpeckers eat insects, tree sap, leaf catkins and flower nectar. In most temperate habitats, however, such foods are either scarce or completely unavailable in winter. An acorn woodpecker group that exhausts its stored mast in the winter months is often forced to abandon its territory.

Many granary trees are easily accessible to the student of the acorn woodpecker, and with patience the stored acorns or nuts can be counted. In this way the energetic value of the stores and of the area's total woodpecker population can be estimated. The estimates show that stored mast is critical not only for survival but also for reproduction. For example, in a 10-year period at the Hastings reservation in California, of the groups that possessed stored mast in the spring, 83 percent successfully raised young birds to the stage of fledging, or leaving the nest. Of the groups



GRANARY TREE is utilized by the acorn woodpecker group for the storage of mast: acorns and other hard-shell nuts. The tree in the photograph, which was made by one of the authors (Koenig) at the Hastings Natural History Reservation on the coast of central California, is an oak. The holes are drilled by successive generations of the members of a single acorn woodpecker group; some granary trees are perforated with as many as 30,000 holes. The nuts are inserted in the holes in the fall as the group attempts to accumulate enough food stores to last through the winter. The granary tree is a central feature of the ecology and social structure of the woodpecker group.



ACORN WOODPECKERS are highly visible residents of oak and pine-oak woodlands in the southwestern U.S., Mexico and Central America. The birds are about nine inches long. The only visible difference between the male and the female is that the female has a black band separating the red and white areas on the head; here a female is with two males. The acorn woodpecker lives in social groups consisting of as many as 15 members, including several breeders of each sex and additional nonbreeding helpers that are offspring from previous years.

that had exhausted their stores only 20 percent successfully raised young to fledging. In years when the acorn crop is particularly large groups sometimes breed in the fall as well as the spring, a phenomenon rare in temperate climates.

The effect of mast storage was also observed at Water Canyon in New Mexico. Over a nine-year period groups that had mast in the spring fledged an average of 2.7 young per year and groups that had no mast fledged an average of 1.3 young. The results are particularly striking because the mast is rarely fed to the nestlings. Instead it functions as a food reserve that enables the adults to search for insects, which are then fed to the young birds.

By noting at regular intervals the amount of mast stored by different groups and correlating the counts with observations of the groups it was shown that mast storage has a strong influence on the structure of the acorn woodpecker population in a particular area. In years when the quantity of stored reserves is comparatively high the fraction of adults that survive the winter is also high. As a result the total woodpecker population of the area is large and more groups are able to remain in their territories until spring. Since most territories are continuously occupied, few young birds can leave to colonize territories of their own. Therefore a relatively large fraction of each group is made up of nonbreeding helpers.

The amount of mast a woodpecker group can store is determined partly by the yield of acorns or nuts from the trees in its territory. It is also influenced by the number of storage holes the group has drilled in the granary tree. Since the amount of stored mast has such a strong influence on the structure and continuity of the group, it would appear that group members should spend much of their time drilling new storage holes.

One of us (Stacey) found, however, that the birds drill storage holes only when they have a supply of mast. When the food stores are exhausted, no drilling is done. The reason is that in the absence of accumulated mast the members of the group spend much of their time looking for other food. If a store of mast is available, less time is spent in foraging. Furthermore, some members of the group must be near the granary to guard the stores, and it is the guards that drill the new holes; if there are no stores to guard, the guards too are off looking for food.

The construction of storage facilities by the acorn woodpecker group is an example of a biological positive-feedback system. Groups with many storage holes can accumulate enough mast to enable them to drill new holes over the winter. The enlarged capacity makes it possible for the group to store more nuts the following fall. Over a long period the number of storage holes could be limited only by the fact that the granaries do not last forever.

The granaries are often constructed in trees that are dead or partly so. Even the sturdiest storage trees, which are oaks, typically last for only a few decades from the time the woodpecker group begins to utilize them. Eventually the granary tree rots and falls. The fall of the tree leads to a reduction in the group's size, in its reproductive success and in the probability that it will remain resident in its territory throughout the year. In some instances the fall of a large granary tree can even lead to the dissolution of the group.

Research in recent years has shown there is a considerable diversity in the reproductive organization of species of birds that breed cooperatively. In some species, including the green woodhoopoe of Africa and the Florida scrub jay, the birds form groups made up of a single breeding pair and nonbreeding helpers; the helpers are generally offspring from previous years. Other species, including the Tasmanian native hen and the Galápagos hawk, show cooperative polygamy, or mate-sharing. Mate-sharing entails the presence of several male or female breeders; in both the Tasmanian and the Galápagos species the multiple breeders are always males. In such species it is thought the brood can receive genes from all the male breeders. Even among such unusual species the acorn woodpecker stands out because the woodpecker groups have both nonbreeding helpers and mate-sharing. Moreover, the mate-sharing often involves males and females simultaneously.

In the acorn woodpecker, courtship behavior between male-female pairs is absent. In addition the entire group tends a single nest. As a result determining the reproductive roles in a cooperative group is not an easy task. Copulation is rarely observed, and so it is particularly difficult to know the reproductive status of male birds. It had long been suspected that more than one male in a group can breed, but proof was lacking. It has recently been obtained by genetic analysis of enzymes in the blood of members of several woodpecker groups; this work was done in collaboration with Nancy E. Joste and J. David Ligon of the University of New Mexico and Ronald L. Mumme and Robert M. Zink of the University of California at Berkeley.

The reproductive status of females is easier to determine because egg laying can be observed if the group is watched carefully during the breeding season, which is generally in late spring and early summer. At Water Canyon woodpecker groups with more than one breeding female are rare. At the Hastings reservation, however, 20 percent of the groups include two or more females that lay eggs in the same nest. In both areas the groups are supplemented by males and females that do not breed.

With both the Water Canyon and the Hastings reservation woodpecker populations one of the first lines of work was to unravel the reproductive relations in the group and find out what determines the reproductive status of individual birds. It developed that there are intriguing differences in structure between groups in one of the areas and those in the other. In both areas helpers are group offspring from previous years. At the Hastings reservation, however, the males that share mates are frequently brothers and less frequently father and son; the females that share mates can be sisters or mother and daughter. At Water Canyon the birds that share a mate can be genetically related or unrelated.

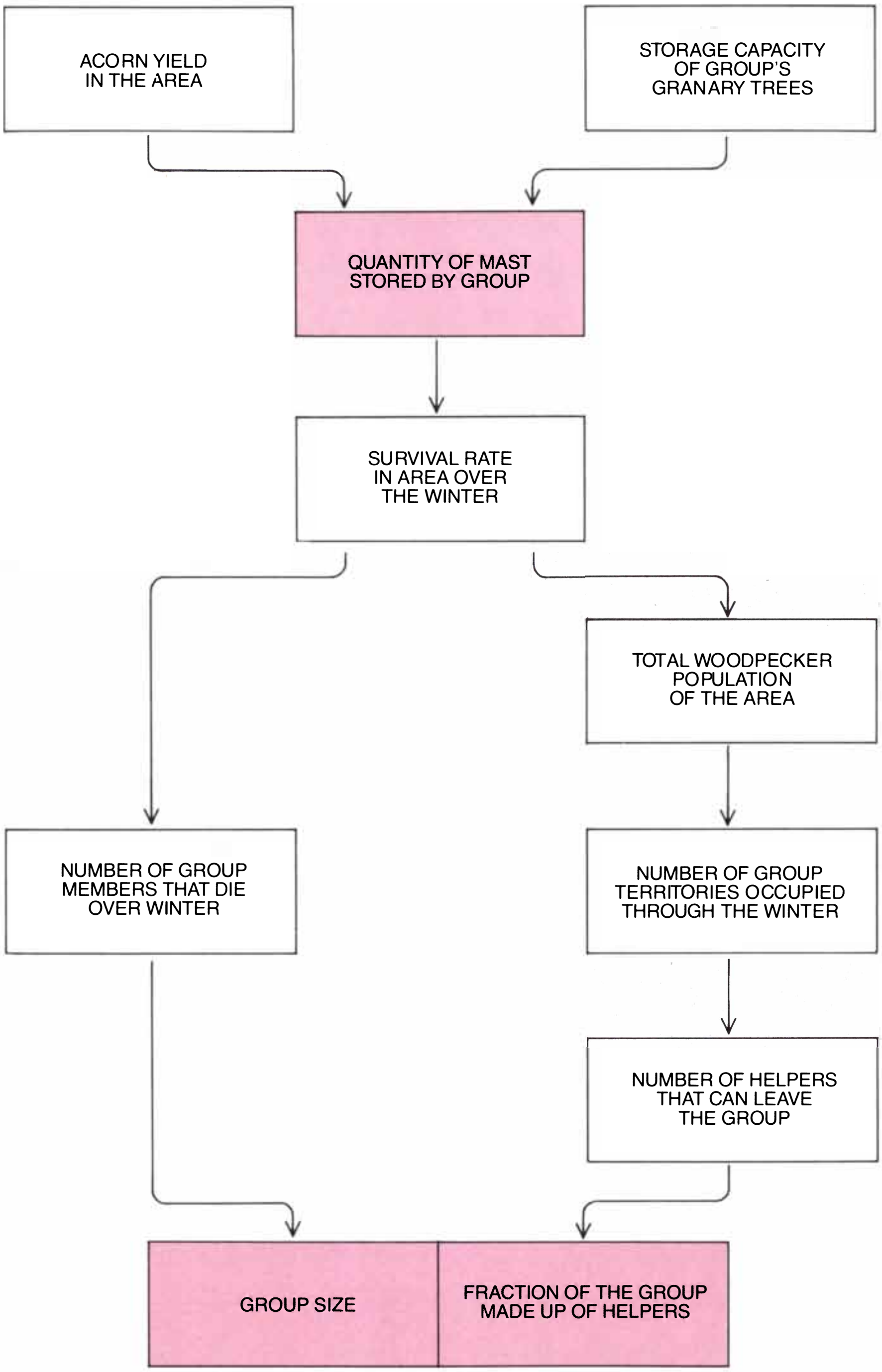
To understand how such differences could arise the history of a typical group must be considered. A group can be founded by a male-female pair or by the joining of two "coalitions" from different groups. One coalition consists of from one male to four males, usually siblings, drawn from an established group. The other consists of from one female to three females, also usually siblings, drawn from another group.

Once the coalitions have merged, all the mature adults can breed as a unit. The young birds that are subsequently fledged remain in the group at least until their first spring and sometimes longer. With one notable exception, which will be discussed below, the group offspring do not breed as long as they remain in their natal group. The group varies in size as young birds are fledged and as they and other birds leave. The unit maintains its continuity as long as at least one of the original breeders of each sex is present. As the original breeders die or leave, however, there comes a time when all the founders of one sex or the other have gone.

The disappearance of all the original male breeders or all the original female breeders creates what is known as a reproductive vacancy, which is of fundamental significance in the history of the group. The reproductive vacancy is not filled by helpers from the group; a daughter does not replace her mother as a breeder, nor does a son replace his father. Instead the gap is filled by mature birds from outside the group.

Contrary to what was thought only a few years ago, young acorn woodpeckers do not remain in their natal territory waiting to inherit it from parents of the same sex. Once the reproductive vacancy has been filled, however, group offspring of the sex opposite that of the new breeder can mate. For example, if a female breeder in a group has been replaced, the group offspring that are male can breed in spite of the presence of the original male breeders; the converse is true for the females.

Explaining the features of reproduc-

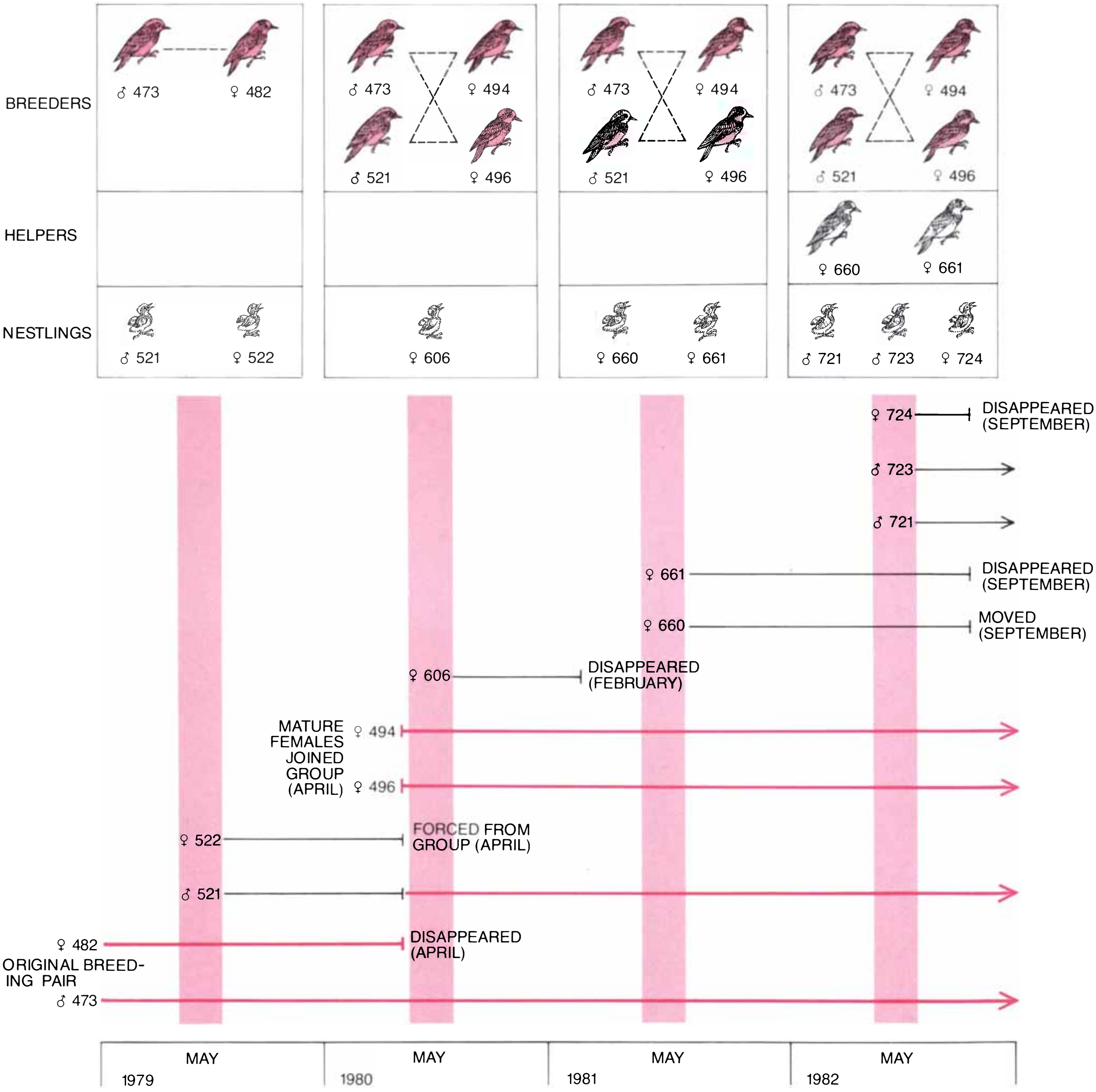


STRUCTURE OF THE ACORN WOODPECKER GROUP is strongly influenced by the quantity of mast the group can store in the fall. The quantity of stored mast is itself determined by the yield from the trees in the group's territory and the storage capacity of the granaries held by the group. The quantity of mast affects how many birds survive the winter. The survival rate influences the total woodpecker population of the area and also the number of desirable territories that remain occupied all year. If the bird population is large and many territories are occupied, few young can leave their parental groups to breed on their own. As a result the groups are large and a substantial fraction of each group consists of nonbreeding helpers. The quantity of stored mast also influences the number of young raised successfully by the group.

tion among acorn woodpeckers is harder than describing how the group forms. The sharing of mates does provide one explanation for why a group of birds will cooperate in feeding the young. The reason is that in some groups all the cooperating adults are actually or potentially the parents of the young birds. The hypothesis does not, however, explain the role of the helpers or the willingness of the breeders to share mates.

Indeed, it might be supposed that mate-sharing runs counter to the reproductive interest of each breeding adult. The number of young that can be cared for by the group is limited. Hence when the breeder shares a mate, there is a reduction in the fraction of the group young that are the offspring of that breeder. In most species of birds that cooperate in raising offspring there is no mate-sharing; the dominant male and the dominant female exclude all other adults from breeding and only a single pair reproduces.

Recent work by one of us (Stacey) suggests, however, that under certain circumstances mate-sharing is to the advantage of the individual breeder. In some environments cooperation makes a significant contribution to the individual's reproductive success and even to its long-term survival. The reduction of



REPRODUCTIVE VACANCY, the disappearance of all the original male breeders or all the original female breeders, is an event of high significance for the acorn woodpecker group. The illustration shows the history from 1979 through 1982 of the "Plaque" group at the Hastings reservation. Color indicates a potential breeding adult. In the boxes at the top the broken lines show the group members that could mate; since copulations are rarely observed, it is not easy to tell whether all potential breeders actually do mate. In May, 1979, the Plaque group consisted of male No. 473 and female No. 482. Their first offspring, male No. 521 and female No. 522, were born in 1979; as long as their parents were in the group the offspring could not breed. Before the 1980 breeding season the original female breeder (No. 482) disappeared. Two sisters from a nearby group (Nos. 494 and 496) joined the Plaque unit after a power struggle and forced female No. 522 out of the group. With his mother gone, male No. 521 could begin to breed with the immigrant females, as could his father.

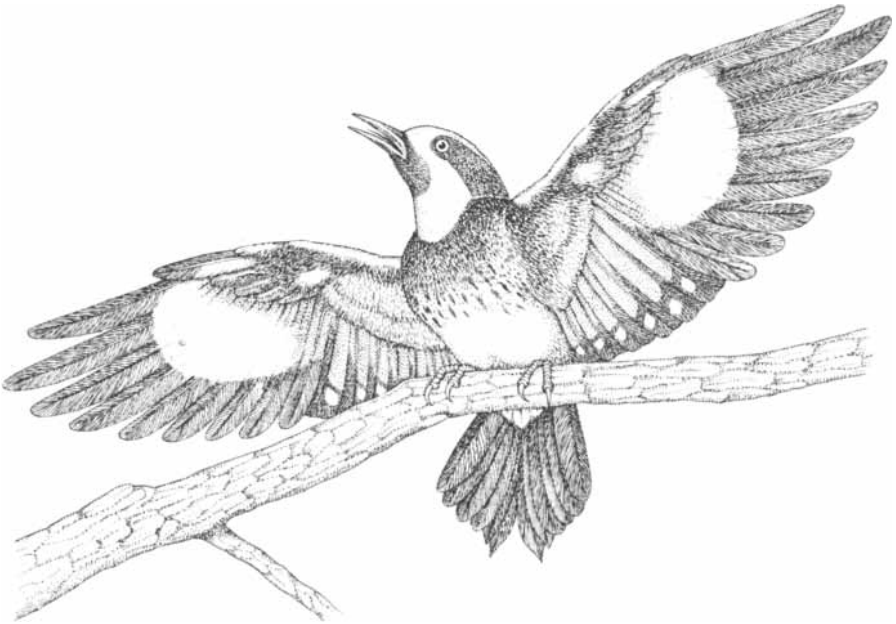
group size would put the individual at a disadvantage. If subordinate birds that were excluded from breeding were likely to leave the group, the dominant birds could enhance their own reproductive success by sharing the breeding.

Data collected at Water Canyon and at the Hastings reservation support the idea that such factors are operating in acorn woodpecker groups. At Water Canyon the number of young that are fledged per breeding male is higher in groups consisting of two males and one female than it is in breeding pairs. In the groups of three there were 1.16 young per adult male on the average; in the pairs there were .92 young per male. Thus when a male shares paternity with another male in a threesome, he will fledge more of his own young than he would if he evicted the second male and bred by himself.

At the Hastings reservation the reproductive success of the individual decreases as the group expands but the fraction of birds that survive the year increases. A male or a female that is the only breeder of its sex in a group has an annual survival rate of 70 percent. Among females sharing a mate the survival rate is 79 percent and among males sharing a mate the rate is 86 percent. Although the rates were lower at Water Canyon, the pattern was similar: survival was 47 percent for birds breeding in pairs and 65 percent for those breeding in larger groups. When the data for reproductive success and survival are combined, it becomes clear that in both the Water Canyon and the Hastings reservation populations a bird that begins its reproductive career sharing a group with other individuals of the same sex has an advantage over one that breeds in a pair.

Although at first mate-sharing and cooperative breeding appear to contradict the principle that natural selection operates on the individual organism, close examination shows that cooperative activity benefits the individual breeder. Not all the activities in the acorn woodpecker group, however, are harmonious. Selective pressure has also encouraged intense competition. Two forms of reproductive competition are notable: egg-tossing and the killing of young birds.

Egg-tossing in acorn woodpeckers was first observed by one of us (Stacey) and Mumme at the Hastings reservation. As we have mentioned, many of the breeding females in a group are closely related genetically. A typical egg-tossing episode could involve two sisters. In the breeding season the sisters lay their eggs in periods that nearly overlap but are not precisely synchronized. Each female lays some three to five eggs at the rate of one egg per day. The eggs are laid in the same nest cavity and the sisters share the incubation and the feeding of the young.



WAKA DISPLAY is a gesture made by an acorn woodpecker to other members of its group while guarding the granary and also in the power struggle over a reproductive vacancy within a group; the raising of the wings constitutes a "greeting" or "recognition" display among group members. The power struggle can involve as many as 50 birds that engage in several days of continuous chasing and display. The winners of the struggle remain to breed in the empty territory; the losers return to the group where they were born and resume the function of helper.

The first egg to be laid triggers a competition between the females. When one of the females lays the first egg, it is almost invariably removed from the nest by the sister. The sister takes the egg to a nearby tree and puts it in a depression in the bark of a limb. Then the egg is pecked open and the contents are eaten by several group members, often including the female that laid the egg.

When the female that laid the first egg lays another on the following day, the egg-tossing is repeated. Observations have been made of as many as five eggs being consecutively tossed from a nest. The egg-tossing stops only when the second female begins to lay. At that point the bird apparently cannot distinguish her eggs from her sister's and therefore refrains from further destruction.

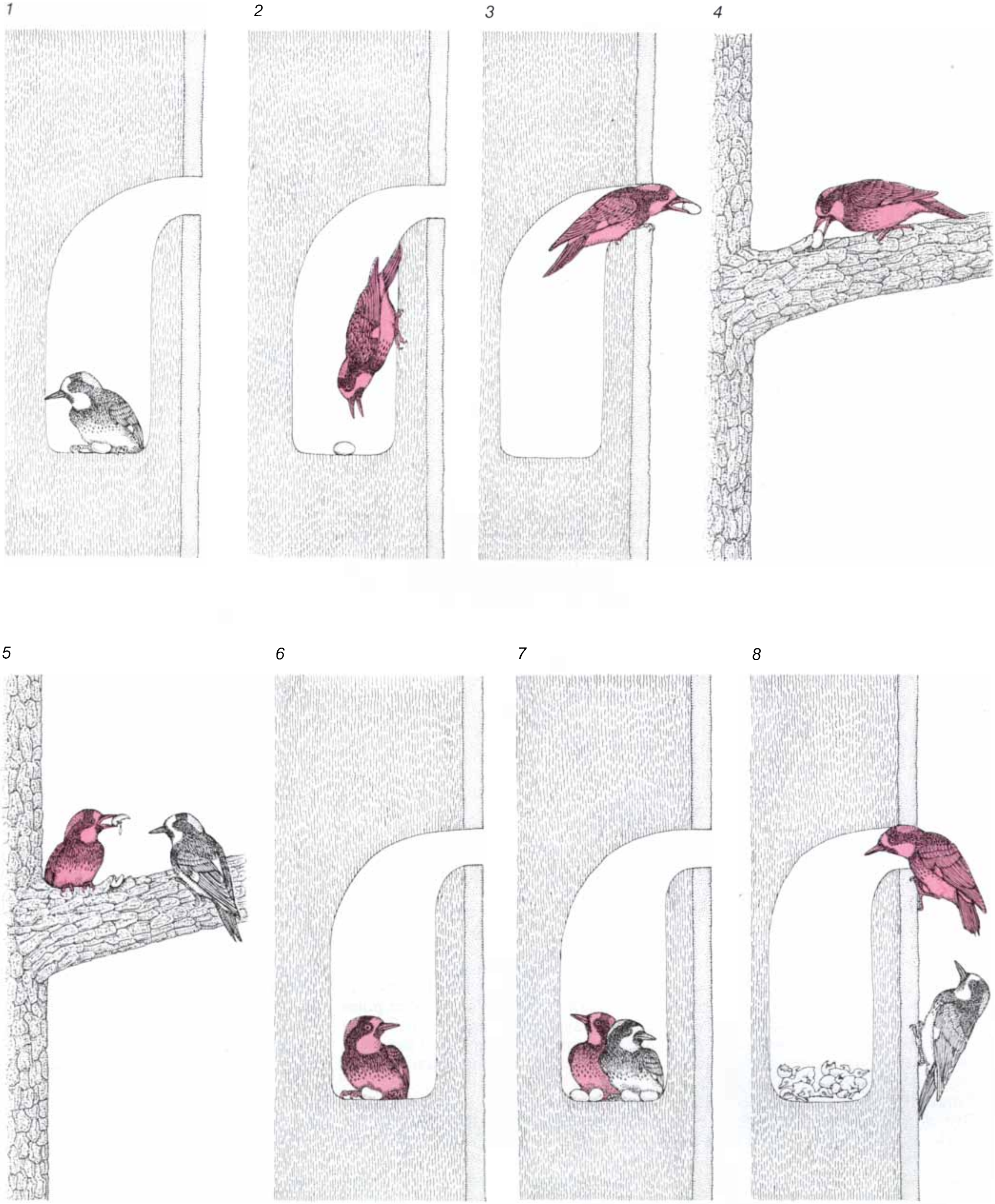
The adaptive value of egg-tossing is clear. The female that destroys her sister's eggs gains a genetic advantage in the clutch of eggs incubated by the group. If the egg-tossing had not taken place, each sister would have contributed an equal number of genes to the next generation. After the eggs are tossed the bird that did the tossing contributes a majority of the genes passed to the offspring. A pair of sisters may live together in a group for several years, however, and the female that has her eggs tossed one year may toss her sister's eggs the next year. The reproductive success of the sisters over their life span could be approximately equal.

An even more extreme form of reproductive competition is the killing of young birds by recent immigrants to the group. Among animals that live in groups the killing of young by recent arrivals had been noted only in primates and mammalian carnivores such as lions. It was first observed in the New Mexico acorn woodpecker population in 1981.

The killing of young birds generally takes place in the spring or early summer when a woodpecker joins a group with an active nest. A precondition for the killing appears to be that the immigrant is the only mature bird of its sex in the group, either because the previous breeders of its sex have left or because the immigrant has driven them away. Once the new bird is accepted by the group members of the opposite sex, it immediately begins to visit the nest and eventually destroys the eggs or kills the nestlings. The bodies of the young birds are removed from the nest and can be eaten by all the group members.

There is controversy among evolutionary biologists over whether the killing of young in a cooperative group is always an adaptively advantageous move. In the acorn woodpecker group, however, the immigrant clearly benefits. Since there is only one nest, a bird that joins the group after the eggs are laid cannot contribute its genes. If enough time remains for the group to breed again, an immigrant that kills the young forces the group to breed and thereby ensures that its genes are passed on.

Acorn woodpecker groups are com-



EGG-TOSSING is a form of genetic competition between breeding females in the acorn woodpecker group. The two females shown here are sisters. Each bird lays one egg per day for about three to five days. The periods of egg laying overlap but are not exactly synchronized. Soon after one female (*white*) lays the first egg (*1*) her sister (*color*) retrieves the egg from the egg cavity (*2, 3*). The sister puts the egg in a depression on a nearby branch (*4*). She is joined by other group members, often including the female that laid the egg; together they peck open the egg and eat its contents (*5*). The destruction of the eggs is repeated until the second female lays her first egg in the same cavity (*6*). When that happens, she stops tossing her sister's eggs and after both females have laid eggs on several successive days (*7*) the eggs are incubated by the group until they hatch (*8*). The female that does the tossing ends up with more eggs in the clutch than her sister has. The perpetrator that tosses the eggs in one year, however, could be the one whose eggs are tossed the next year. Hence over the history of the group the two females might contribute an approximately equal number of copies of their genes to the members of succeeding generations.

plex units in which members cooperate and compete to maximize their genetic contributions to the next generation. Why the intricate group structure evolved is not known with any precision. Some of the environmental factors that influence the group, however, are fairly well understood, including the factors that determine the status of the nonbreeding helpers.

Most group offspring remain in the natal group as helpers because they are ecologically constrained from leaving; if the opportunity to leave arises, the young birds generally disperse to form a new group. The constraint on the helpers appears to be habitat saturation, which means simply that there are no empty territories suitable for occupation by new groups. Habitat saturation does not imply that every square inch is occupied by an established acorn woodpecker group. In any given area, however, the optimum habitats are continuously occupied. Furthermore, in most regions where acorn woodpeckers live there is a shortage of marginal habitats to be taken over by a new group made up of birds that had not previously been breeders. Since group offspring cannot breed independently, they might benefit from remaining in their parents' territory and helping to raise the offspring of their genetic relatives.

Several lines of evidence support the conclusion that habitat saturation has a significant role in keeping young acorn woodpeckers in their natal group. One striking item of evidence is the dramatic power struggle that takes place over reproductive vacancies. Frequently sets of siblings from groups several miles apart converge on a territory where there is a vacancy; the birds fight for the right to remain. Such struggles, which can engage as many as 50 birds and can last for several days, consists of uninterrupted chasing and territorial displays.

The winners in the struggle fill the vacancy and the losers return to their parental group. The ecological significance of these chaotic struggles is that nonbreeding helpers are willing to fight intensely for the opportunity to leave the natal group and breed on their own. Such struggles are logical only if good habitats are at such a premium that the young are constrained from leaving the parental group.

The lack of marginal territories suitable for colonization by newly formed groups is apparently due to the fact that the woodpecker relies heavily on the granary tree. The granary is the result of work done by generations of woodpeckers, with each bird drilling a few holes per year. A threshhold of from several hundred to a thousand storage holes must be reached before a territory can be permanently occupied. When enough holes have been drilled, the territory can be occupied continuously until the granary tree falls. Individual birds cannot create a new granary in a short period, and therefore they cannot disperse into a new area unless it already has a granary tree. Marginal habitats are scarce because all mature granaries are in occupied territories.

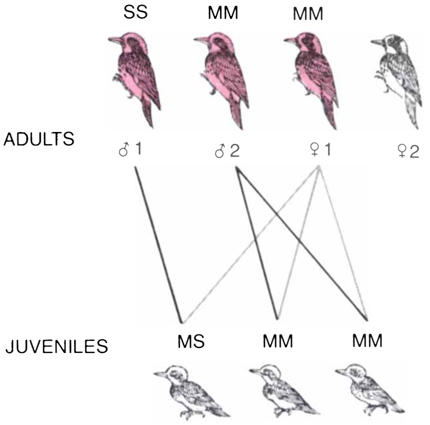
Several other workers, including Jerram Brown of the State University of New York at Albany and Stephen T. Emlen of Cornell University, have suggested that ecological constraints are one of the main causes of cooperative breeding in birds. The study of the acorn woodpecker in several different regions has made it possible to carry out the first quantitative test of that hypothesis. The level of habitat saturation can be measured according to the number of territories that become vacant over the winter, which is the time young birds attempt to leave the natal group.

At the Hastings reservation most groups can store enough mast to last through the winter and hence most territories are occupied continuously. In Water Canyon, however, the yield of acorns or other nuts is more variable; many groups exhaust the stored food and abandon their territory. The vacant territory is then available for colonization. As is predicted by the habitat-saturation hypothesis, fewer young birds remain in their original group for a year or more at Water Canyon than remain at the Hastings reservation.

Additional evidence that habitat saturation and the construction of mast-storage facilities are among the causes of cooperative breeding in the acorn woodpecker group comes from a study done in southeastern Arizona by one of us (Stacey) in collaboration with Carl E. Bock of the University of Colorado at Boulder. In this area the annual acorn yield is sparse and variable; there are few years when the woodpeckers can store enough mast to support the group over the winter. As a result groups do not construct granaries.

Instead acorns are gathered and stored in natural cracks and holes in the bark of trees. The capacity of the storage system is small, and the mast is generally exhausted in the fall soon after the oaks stop producing acorns. When the food store is used up, the birds leave their territory and migrate to Mexico for the winter. In the spring many of the same birds return and establish new breeding territories. Because of the annual migration most of the breeding habitat is unoccupied in the winter and therefore much territory is available for colonization in the spring.

The absence of permanent territories has a profound effect on the birds' reproductive behavior. In contrast to the young birds in California (Hastings reservation) and New Mexico (Water Canyon), young birds in southeastern Arizona leave their natal territory in the fall and do not migrate to Mexico as part of a family group. In the spring the adults and yearlings return to Arizona individually and often breed in a new territory with a new partner every year. Even more striking is the fact that in Arizona the migratory acorn woodpeckers do not breed cooperatively. There are no nonbreeding helpers and mating is in isolated pairs.



FIRST DEMONSTRATION of multiple paternity in an avian communal breeder was obtained for the acorn woodpecker by the authors and their co-workers. The paternity relations in a group at Water Canyon, N.Mex., are shown. The letters *S* and *M* stand for slow band and medium band. Each letter designates an allele, or variant of a gene, for a blood enzyme. Each pair of alleles was identified by gel electrophoresis, in which enzymes can be separated according to their mobility in a gel. Female No. 2 was a nonbreeding helper.

From the work done on the acorn woodpecker in the past decade have come new perceptions of the ecology and social organization of cooperative breeders. Such studies have given a detailed picture of the birds' reproductive behavior and their mating system. Above all, the work has illuminated how intricate the relations are among members of an avian social group.

Biological evolution favors individuals that consistently maximize their genetic contribution to the next generation relative to the contribution of other breeders. Yet the strategies by which the genetic contribution is increased are not simple. The strategies adopted by the acorn woodpecker range from the straightforward genetic competition entailed in destroying a sister's eggs or killing the young of other group members to the apparently self-denying behavior of mate-sharing or acting as a nonbreeding helper. Further work on this unusual bird will undoubtedly reveal other unexpected means whereby a woodpecker in a social group can increase its reproductive advantage and hence its evolutionary fitness.

THE AMATEUR SCIENTIST

Deep think on dominoes falling in a row and leaning out from the edge of a table

by Jearl Walker

Lorne A. Whitehead of Vancouver, B.C., offers a remarkable variant on a row of falling dominoes. Instead of having dominoes all the same size he has each domino 1.5 times larger than the preceding one. In a row of 13 dominoes a tap on the first domino, which is tiny, soon brings down the last, which is 64 times larger. If Whitehead had been able to continue scaling up the dominoes in the same way, the 32nd domino would have been as big as one of the twin towers of the World Trade Center in New York.

I began looking into how dominoes in a row fall using a row of 50 regular dominoes. At a separation of one domino width the chain reaction took two seconds. With about half that separation the reaction took half as long. I could hear the difference in the timing of the collisions in the two trials. The timing also changed noticeably after the chain reaction had passed the first five or six dominoes, presumably because in the early stages only a few dominoes were moving, whereas later more were falling at any one time.

By freezing the action with a snapshot or a strobe light I could see the pattern of the chain reaction. When the leading domino of the moving subgroup is falling toward the next domino, which is still stationary, the leading domino has the preceding domino leaning against it. Several other dominoes in the moving group are also leaning and falling. In a theoretical analysis of the chain reaction D. E. Shaw of Villanova University found that at any one time about five dominoes are in motion.

I next examined how the chain reaction ascends and descends a smooth ramp. The incline must of course be a shallow one or it will not be possible to stand the dominoes on end. When the chain reaction traveled up a properly inclined ramp, the dominoes had to be spaced farther apart than dominoes in a level chain. A chain reaction going down the ramp was much faster and succeeded with any spacing less than the height of a single domino. The chain reaction could also be made to climb a small flight of shallow stairs, but the reaction was quite slow.

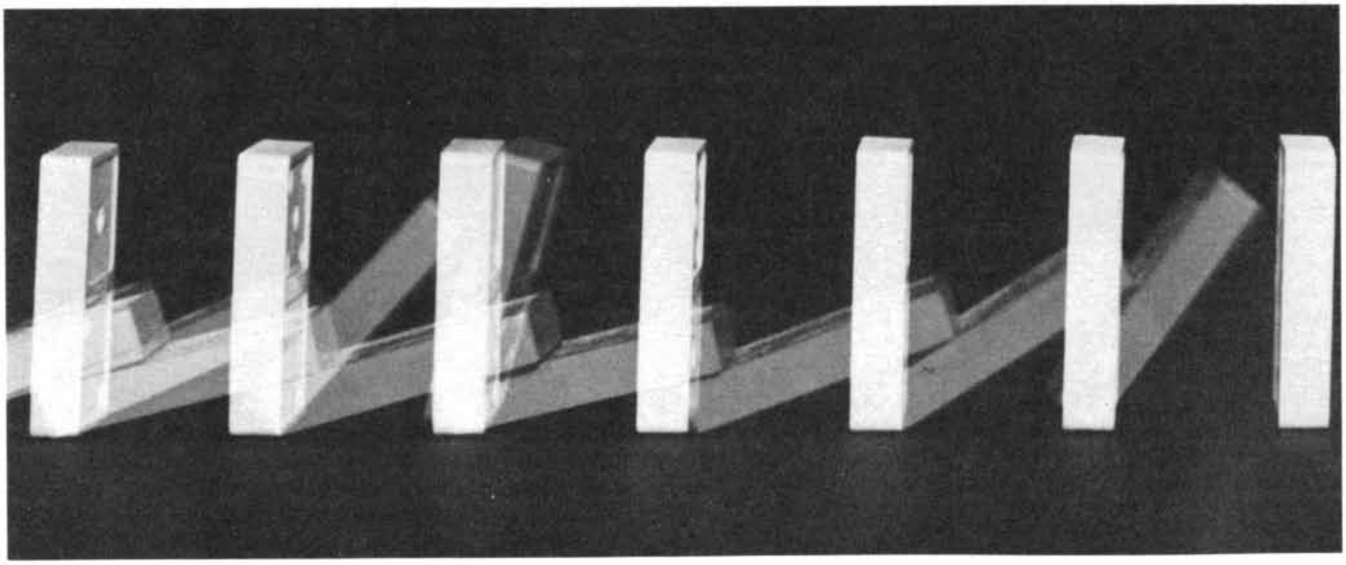
A chain of dominoes can also be made to split into two chains. At the junction the last domino of the original chain falls against the first dominoes in the new chains. A chain reaction can also be made to turn, indicating that it is not necessary for one domino to hit the next one squarely.

To understand the physics of domino toppling I considered a conventional domino of height h, width w and depth d. What factors are responsible for the stability of the upright domino? If I am to topple it, how hard must I strike one of its broad faces in order to rotate it about an edge? Why are deeper (thicker) dominoes harder to topple? Why cannot a line of cubes (such as a child's blocks) be toppled in a chain reaction?

Consider an upright domino. Gravity pulls downward on every small volume of it. A simpler way of looking at the situation is through the concept of the center of mass, which is at the geometric center of the domino. The total pull of gravity (the weight of the domino) is said to operate there. The upright domino is stable because its weight vector points toward the support region.

In principle I can put the domino in another stable arrangement by tilting it on one of its bottom edges until the center of mass is over that edge. Although the weight vector again points toward the support region, the balance is precarious because even a small force destroys it, toppling the domino. To knock over a domino I strike it so that it is tilted completely through its position of precarious stability. (I make the assumption that the friction from the table is large enough to keep the domino from slipping. Could a chain reaction take place on a frictionless table?)

As the domino tilts it embodies two forms of energy. The kinetic energy depends on the speed of tilting. The poten-



A chain of falling dominoes

Mary Ann Sanford makes over 20 critical checks on seat belt systems in GM cars.

She is only one of over 1,800 GM people who make more than 6,000 safety inspections on GM cars built in America.

Checks are made on many things from seat belt retractors to anchor bolts. Latch buckles. Warning lights and buzzers. Every seat belt in every car gets checked.

We do this because it's all part of building safer cars. That's important to you. And to us.

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.

tial energy has to do with the height of the center of mass above the table. When I strike the domino, I impart kinetic energy, but as the center of mass rises the energy is converted into potential energy. What is the minimum amount of energy I could impart to the domino so that it just barely passes through the position of precarious stability? Initially the center of mass is at a height $h/2$ above the table. When it is at its highest point during the rotation, it is a distance r from the table (the distance between the center of the domino and one of the domino's bottom edges).

If the domino has a mass m, it has a weight mg (g representing the acceleration of gravity). The potential energy of the domino at any instant during rotation is the product of the domino's weight and the height of the center of mass above the table. Hence to lift the center of mass from a height $h/2$ to a height r I must provide the domino with an energy of $mg(r - h/2)$. If my blow delivers less energy, the domino cannot pass through the position of precarious stability and will fall back to its original orientation. If I deliver more energy, the domino falls over.

The key to the instability of an upright domino lies in its shallow depth. Consider a domino of fixed height and width. If the domino is narrow, r and the weight are low. Only a little energy is needed to lift the center of mass through the position of precarious stability. A light tap of the finger is enough. If the domino is wider, all the other factors (r, the weight and the energy needed to lift the center of mass) are larger. A tap of the finger does not provide enough energy and does not topple a thick domino.

The rotation of a domino is caused by a torque that is the product of a force and a lever arm. A blow against the domino creates such a torque. The second illustration from the top on page 127 demonstrates how to determine the lever arm. Extend the vector representing the force. Draw a line from the rotation edge to intersect perpendicularly the extension of the vector. This line is the lever arm. In order to deliver a large torque that will make the domino rotate rapidly you should hit high on the face of the domino to have the benefit of a large lever arm.

As the domino rotates, its weight vector creates another torque that attempts to return the domino to its initial orientation. To find the lever arm for this torque extend the vector representing the weight. Draw a line from the rotation edge to the extension of the vector, intersecting the vector extension perpendicularly. This line is the lever arm associated with the weight of the domino. As the domino rotates upward the lever arm is shortened, decreasing the torque imparted by the weight.

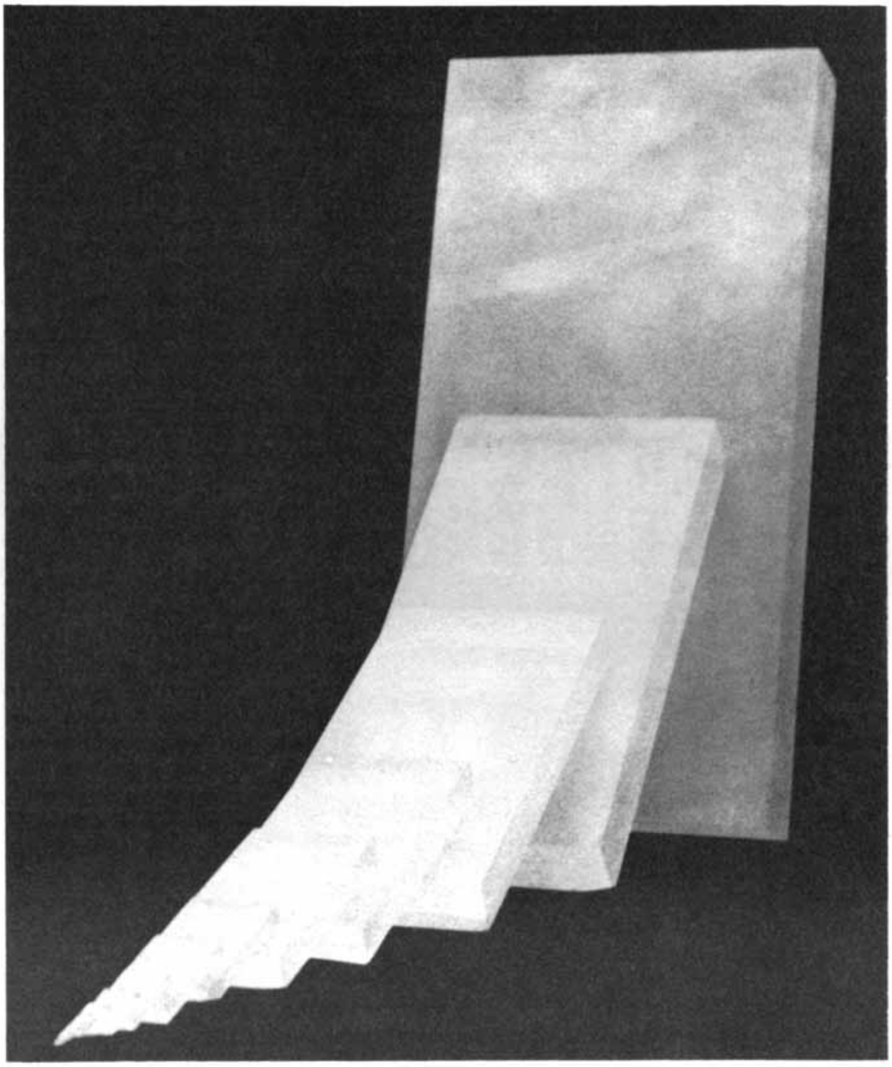
In the illustration the torque from the striking force acts to generate a clockwise rotation of the domino, whereas the torque from the weight acts in the opposite direction. Since the striking force is brief, its torque is also brief. During the ensuing rotation the only torque on the domino is from its weight. If you are to topple the domino, the torque you apply must provide enough rotation to prevent the weight's torque from stopping the movement before the center of mass passes through the position of precarious stability.

A tap of a finger on a child's block is unlikely to topple it. Since the block is deep, the initial lever arm for the weight's torque is large. Moreover, the block weighs considerably more than a domino of the same width and height. For these reasons the torque from the weight easily overpowers the torque from the striking force.

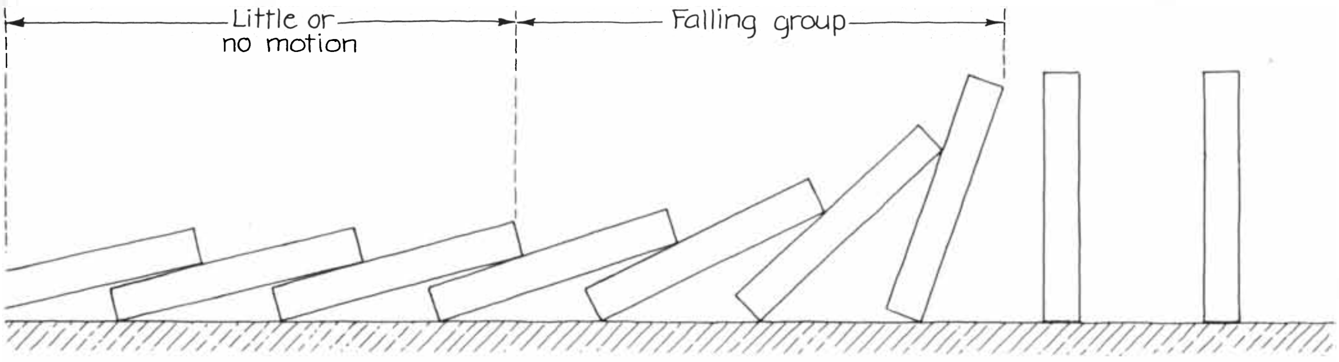
I then examined the matter of knocking over a domino so that it would collide with a second one. Obviously the distance between the dominoes must be less than their height or they will not collide. Is there a minimum spacing? If you strike the first domino vigorously enough, the second domino is bound to fall as long as the separation is less than the height of one domino. Suppose your blow barely pushes the first domino through its position of precarious stability. The second domino will fall only if it is beyond a certain distance (call it the least distance) from the first one.

I did experiments with dominoes of two sizes. One set consisted of standard plastic dominoes 4.4 centimeters high, 2.2 centimeters wide and .7 centimeter deep. The other set consisted of large wood dominoes 13.9 centimeters high, seven centimeters wide and 1.9 centimeters deep. The width of a domino does not matter except as it contributes to the weight. The ratio of height to depth, which does matter, was about the same for the two sets.

I taped a sheet of fine-grain sandpaper to a table and taped a ruler to the sandpaper. The sandpaper provided enough friction to keep a domino from slipping. Along the edge of the ruler I stood two of the smaller dominoes upright. Leaving the position of the second dom-



Lorne A. Whitehead's variation with scaled-up dominoes

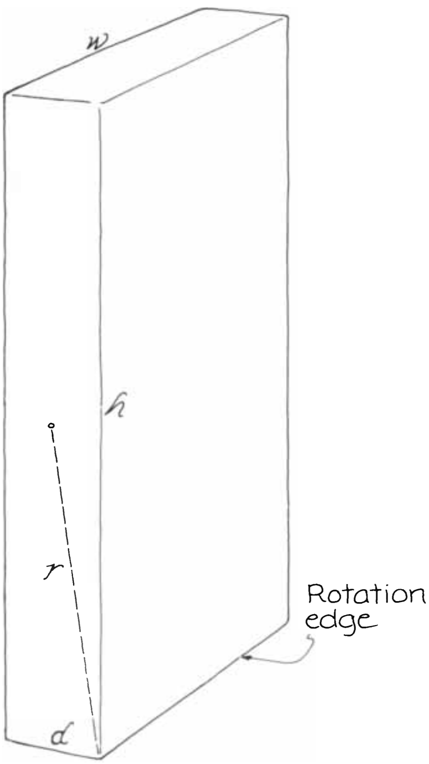


A detail of the chain reaction

ino unchanged, I varied the distance between the two by moving the first one. In each trial I released the first domino from its position of precarious stability with essentially no kinetic energy. To put the domino into that orientation the separation between the dominoes had to be at least .7 centimeter. To make the first domino knock over the second the separation had to be at least 1.2 centimeters. At separations between .7 centimeter and 1.2 centimeters the first domino ended up simply leaning against the second, which remained upright.

The results with the larger dominoes were similar. To release the first of two dominoes in its position of precarious stability the separation had to be at least 1.9 centimeters. To topple the second domino the separation had to be at least 2.3 centimeters.

The least distance is determined by energy. To topple the second domino the collision must impart enough energy to lift the domino's center of mass through the position of precarious stability. This crucial amount of energy is $mg(r - h/2)$. Assume an ideal collision, in which at the instant of collision all the kinetic energy of the first domino is transferred to the second. How far must the first domino fall from its highest point in order to impart enough energy to the second domino to topple it? The first domino must fall just as far as the center of mass of the second domino must rise. Thus it must fall by a height of $(r - h/2)$.



Terms employed in calculations

Assuming a complete transfer of energy in the collision, this requirement sets the value of the least distance between the dominoes. When the spacing is too small, the first domino cannot fall far enough to provide the necessary energy. In such cases the second domino rotates slightly but then rocks back to its initial orientation.

How does the value of the minimum distance determine the success of a chain reaction in a row of dominoes? Suppose the spacing in the row is so close that the first domino cannot be put into its position of precarious stability before it hits the next domino. The initial blow must be quite vigorous, since in essence it must make the entire chain lean as one.

If the dominoes are farther apart but still are separated by less than the least distance, a sound blow on the first domino is needed to give it enough kinetic energy after it has passed through its position of precarious stability. The additional energy is required because the domino will not be able to fall far before it collides with the next one. The second domino topples not because of the energy in the fall of the first one but because of the extra energy in the original striking force.

The least distance is important when the first domino has just enough energy to pass through its position of precarious stability. Then the second domino must be beyond the least distance so that the first one can fall far enough. Only with such a fall can enough energy be delivered to the second domino to topple it. A weak blow on the first domino can initiate a chain reaction when the dominoes are separated by more than the least distance and less than the height of a domino. If the separation is less than the least distance, the initial blow must deliver substantially more energy to the first domino to create a chain reaction.

With my home computer I estimated the values of the least distance for dominoes of various shapes. I assumed that the collision completely transferred the kinetic energy of one domino to the next. I also assumed that afterward the lean of the first domino helped to push the second one through the position of precarious stability. To find the least distance I had the first domino pass through its position of precarious stability with no kinetic energy. To impart enough energy to the second domino the center of mass of the first has to fall to a height of $h/2$.

My estimate of the least distance for my small dominoes was .7 centimeter and for the larger ones 1.9 centimeters. These estimates were below my experimental results of 1.2 and 2.3 mainly because my calculations assumed a perfect transfer of kinetic energy in the collisions. In a real chain reaction some of the energy is lost to vibration of the dominoes and to friction as they touch each other. When I lubricated the faces of the dominoes, the experimental results decreased only slightly, implying that vibration contributes more than friction to the loss of energy.

I glued some of the smaller dominoes together to make new dominoes two or three times deeper than the originals. Repeating my experiments, I found that the doubly deep dominoes required a least distance of separation of 2.2 centimeters. (My computer estimate was 1.5 centimeters; energy losses in the real collisions again accounted for the difference.) The triply deep dominoes almost toppled one against the next, but friction held them at the last instant, leaving them at a tilt.

My calculations showed that the necessary least distance in spacing increases as the depth of the dominoes increases. The reason is that each domino must fall farther in order to gain enough kinetic energy to topple the next one. When the required least distance is almost equal to the height of a domino, the energy transfer is inefficient and the chain reaction stops. A line of cube blocks cannot topple in a chain reaction because the required least distance exceeds the height of the blocks.

Next I returned to some of my earlier demonstrations. Normally a chain reaction of dominoes is initiated with a vigorous rap on the first domino. Each domino in the chain passes through its position of precarious stability with more than enough energy. Hence even when the dominoes are spaced closely and have little chance to fall, the collisions move them safely through their position of precarious stability. The kinetic energy in the wave is large enough for the wave to split at a junction and then travel along two chains.

A chain reaction can be set off on a ramp provided the ramp is not too steep. The weight vector must point toward the lower face of the domino. You will also find that the dominoes must be separated more than they need to be on a level surface.

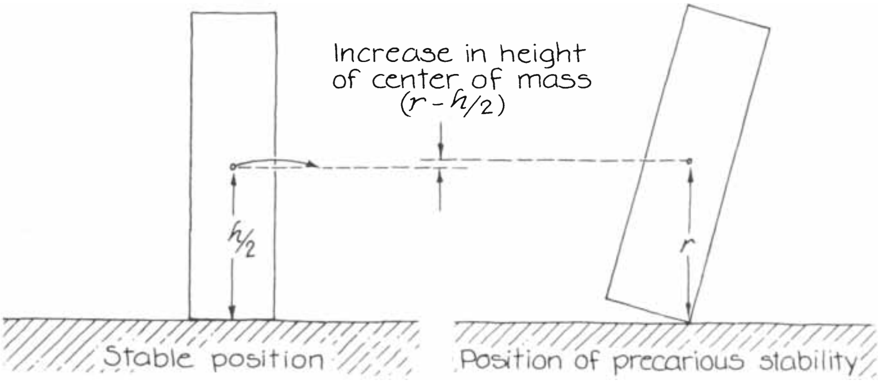
Consider the second domino in the chain. Since the ramp slants, the center of mass must be raised quite far in order to pass through the position of precarious stability. Thus the domino is slower to topple than one on a level surface. Toppling requires more energy. A domino must fall a substantial distance for there to be a significant transfer of energy when it strikes the next domino. Therefore the least separation between dominoes is large and the chain reaction moves slowly.

A chain reaction moves faster down a ramp. Each domino needs only a small amount of energy in order to pass through its position of precarious stability. The least separation is essentially zero. Moreover, each domino does not have to fall far to topple the next one, so that the time from one toppling to the next is quite brief.

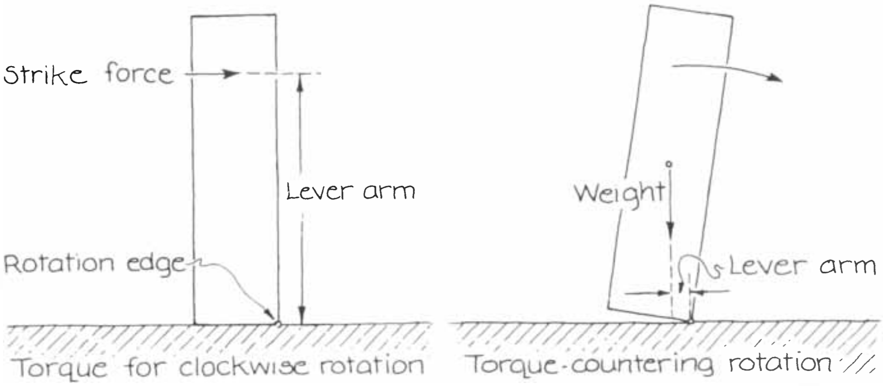
The dominoes on a staircase topple slowly. Since each domino strikes the next one low on a broad face, the lever arm for the collision is small, providing a small torque. The transfer of energy is also inefficient. The struck domino rises slowly through its position of precarious stability, keeping the speed of the reaction low.

Whitehead's demonstration of consecutively scaled-up dominoes is stunning. The first domino is so small that I can hardly stand it upright. The last one (the 13th) is so heavy that I have difficulty lifting it into place. Yet the chain reaction initiated with the tiny domino easily drops the heavy one.

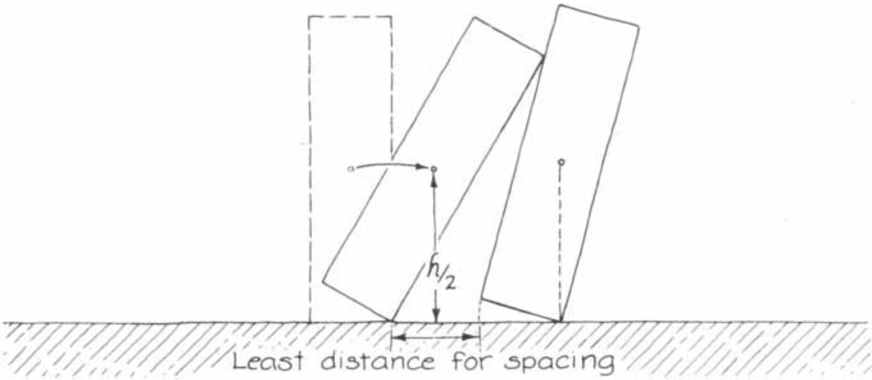
Since each domino is 1.5 times larger



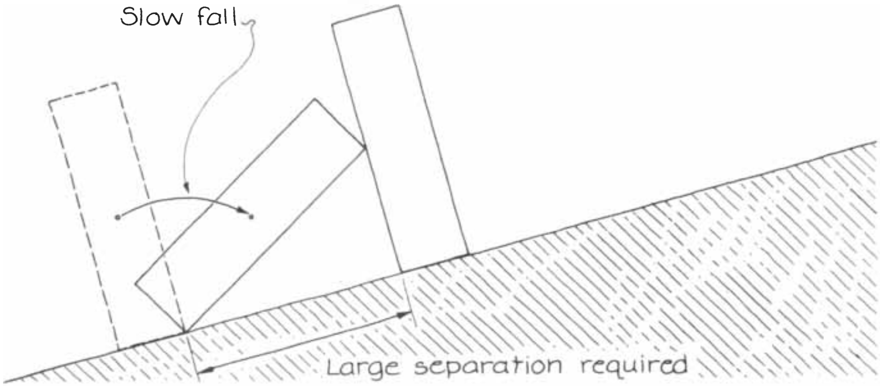
How the center of mass rises



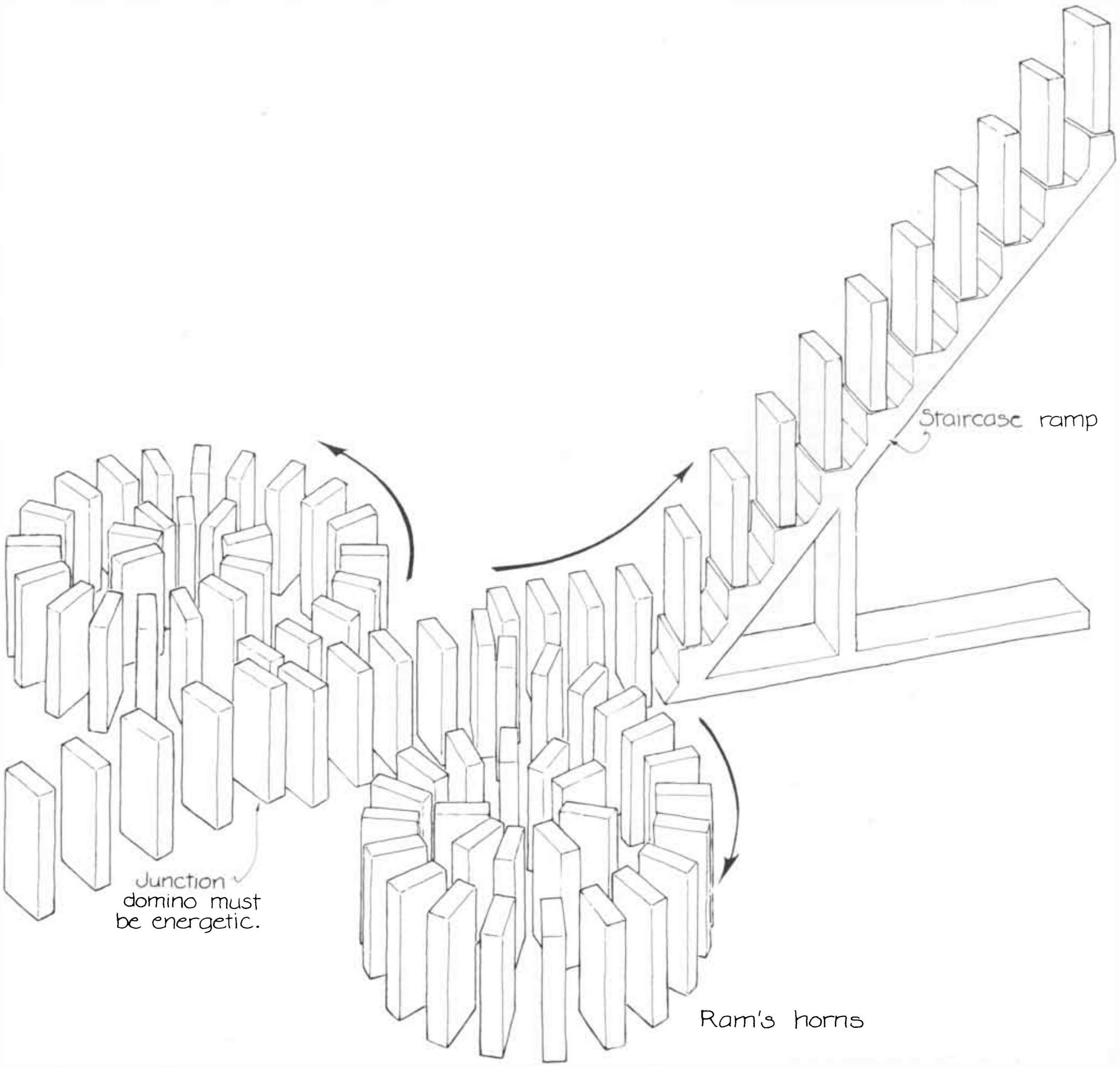
The torques on a domino



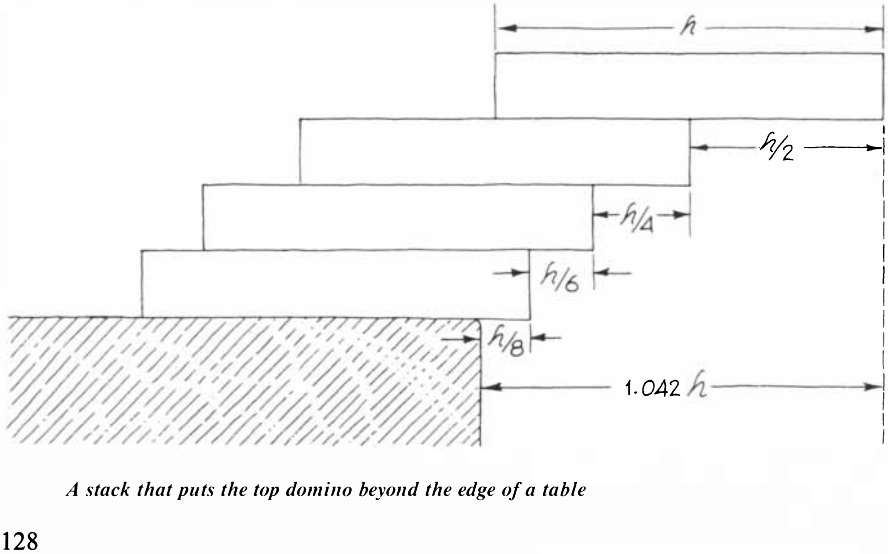
Data on the minimum spacing for dominoes to fall



An arrangement for making dominoes fall up a ramp



A ram's-horn chain and dominoes on steps



A stack that puts the top domino beyond the edge of a table

Overhang in domino lengths	Number of dominos required
.5	1
1	4
2	31
3	227
4	1,674
5	1.2×10^4
10	2.7×10^8

Data on stacking

in every dimension than the preceding one, the last is not only 64 times larger than the first but also 262,144 times heavier. With a center of mass 64 times higher than that of the first domino its potential energy is almost 17 million times greater. A slight nudge on the first domino gives it about .024 microjoule of kinetic energy. The kinetic energy of the last domino as it completes its fall is about 51 joules, which is two billion times the amount imparted to the first domino.

Whitehead made his dominoes out of acrylic sheets, sandblasting them to make them smoother. He laminated several thin sheets to make the larger dominoes. When he lines up the group of 13 dominoes for a chain reaction, each domino is separated from the next-larger one by about its own width.

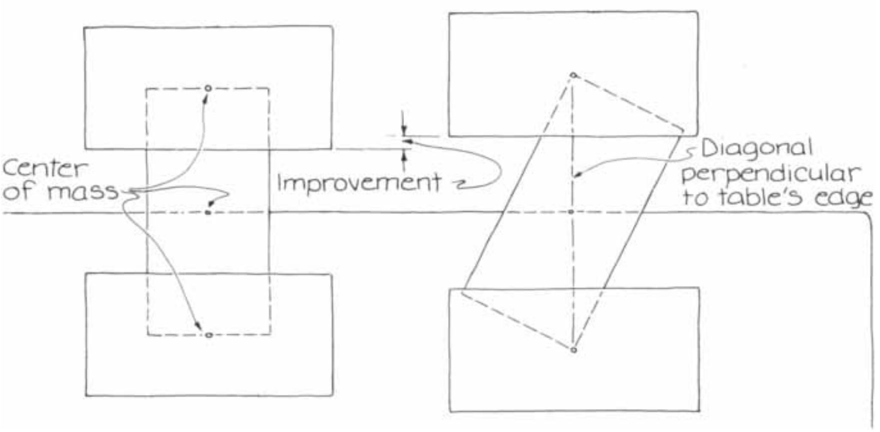
Could a chain of dominoes be designed with a scale factor larger than 1.5? Assuming an ideal transfer of energy in the collisions, I think a scale factor of about 2.5 is the limit for dominoes shaped like Whitehead's. Larger scaling factors are possible if the ratio of height to depth is increased. You might like to experiment with this problem.

Here is a puzzle. The object is to stack dominoes at the edge of a table so that the stack leans out from the edge as far as it can without falling to the floor. Each domino is positioned with a broad face down and its long dimension at right angles to the edge of the table. How should the dominoes be laid so that the top one is displaced from the table's edge by the maximum amount? What is the smallest number of dominoes needed to make the top domino completely clear the edge of the table? Is there a limit to how far the stack can extend from the table?

A problem of this kind was posed in 1955 by Paul B. Johnson. His solution was later simplified by Leonard Eisner, who also applied it in a prank. One night Eisner and another graduate student constructed a leaning stack of a library's volumes of *The Physical Review,* leaving the disconcerting overhang as a surprise for the librarian. (They were presumably too young to have learned that if you need help from librarians, or even if you want to avoid being fired from the university, you had better not fool around with the library volumes.)

Consider the first domino put down. It is stable as long as its center of mass is over the table so that the weight vector cannot supply a torque. Its maximum stable overhang is achieved when its center of mass lies almost over the edge.

Now you put a second domino under the first. The outer edge of the lower domino acts in place of the edge of the table to keep the top domino from rotating. How should this pair be placed on the table? The combined center of mass of the pair, which lies midway between the centers of the dominoes, must be placed over the table's edge to achieve the maximum stable overhang. No computation is needed. A few trials will reveal the proper balance.



How to make one of three dominoes clear the edge of a table

When a third domino is put under the first two, the combined center of mass of the three dominoes must be above the edge of the table. This procedure can continue indefinitely. With a stack of n dominoes the distance between the edge of the table and the outer edge of the top domino is given by the series $(h/2)$ $(1 + 1/2 + 1/3 + 1/4 \ldots 1/n)$, where h is the long dimension of a domino and the expression in the second set of parentheses is the harmonic series.

I programmed my computer to sum the expression to any wanted number of dominoes. At least four dominoes are required if the overhang is to exceed h, making the top domino fully clear the table. To have the top domino clear the table by two domino lengths you must stack 31 dominoes. Thereafter the number goes up rapidly. You will need 12,367 dominoes to make the top domino clear the table by five domino lengths and 1.5×10^{44} to make it clear by 50 domino lengths.

Evaluating the expression for the maximum overhang is simple enough for a computer until about a million dominoes are being considered. A computer's normal accuracy (called single precision) is usually limited to seven significant figures; six figures are printed out after the computer has rounded off the calculation. To continue the overhang expression past one million dominoes leads to error unless the program calls for double precision (usually an accuracy of 14 significant figures, with 13 figures printed out after rounding off). Computation with a large number of dominoes at double precision is slow. Even with this precision the computer's arithmetic functions may yield only approximate results.

Steve Wallin of Laramie, Wyo., has shown me a fast technique for approximating the overhang expression. It employs as a benchmark the 1,674 dominoes needed for an overhang of four domino lengths. The number of dominoes needed for a given overhang (expressed in domino lengths) is equal to 1,674 multiplied by the exponential of twice the difference between the wanted overhang and four lengths. For example, if your objective is an overhang of 10 domino lengths, subtract 4 from 10, multiply by 2 and take an exponential of the result. Finally multiply by 1,674. The answer is the number of dominoes you need.

Here is another puzzle. What is the maximum amount by which a domino can be made to clear the edge of the table when you are working with only three dominoes? Two solutions are shown in the illustration above. In the first solution the center of mass of the lower domino is just above the edge of the table. The second domino is put on the inner end of the first; the upper domino's center of mass is just over the edge of the lower domino. The third domino is now put on the outer end of the first domino. The combined center of mass lies just above the edge of the table, so that no torque is applied to the pile by the weight.

A better solution is to rotate the first domino so that a diagonal line running between two of its corners is at right angles to the edge of the table. The second domino is balanced at a corner distant from the table's edge. The third domino is balanced with its center of mass over the opposite corner. Again the combined center of mass lies at the edge of the table, giving the assembly stability. With this arrangement the third domino is farther from the edge of the table than the third domino in the other arrangement because the diagonal line across the domino is longer than a domino length. You might try standing the third domino on one of its narrow side faces. The center of mass is in the same relation to the table as before but the nearest face of the domino is farther from the edge of the table.

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

COMPUTATION: FINITE AND INFINITE MACHINES. Marvin L. Minsky. Prentice-Hall, Inc., 1967.

TURING MACHINES. John E. Hopcroft in *Scientific American,* Vol. 250, No. 5, pages 86–98; May, 1984.

THE CLIMATIC EFFECTS OF NUCLEAR WAR

LONG-TERM BIOLOGICAL CONSEQUENCES OF NUCLEAR WAR. Paul R. Ehrlich, John Harte, Mark A. Harwell, Peter H. Raven, Carl Sagan, George M. Woodwell, Joseph Berry, Edward S. Ayensu, Anne H. Ehrlich, Thomas Eisner, Stephen J. Gould, Herbert D. Grover, Rafael Herrara, Robert M. May, Ernst Mayr, Christopher P. McKay, Harold A. Mooney, Norman Myers, David Pimentel and John M. Teal in *Science,* Vol. 222, No. 4630, pages 1293–1300; December 23, 1983.

NUCLEAR WINTER: GLOBAL CONSEQUENCES OF MULTIPLE NUCLEAR EXPLOSIONS. R. P. Turco, O. B. Toon, T. P. Ackerman, J. B. Pollack and Carl Sagan in *Science,* Vol. 222, No. 4630, pages 1283–1292; December 23, 1983.

NUCLEAR WAR AND CLIMATIC CATASTROPHE: SOME POLICY IMPLICATIONS. Carl Sagan in *Foreign Affairs,* Vol. 62, No. 2, pages 257–292; Winter 1983/1984.

GLOBAL ATMOSPHERIC EFFECTS OF MASSIVE SMOKE INJECTIONS FROM A NUCLEAR WAR: RESULTS FROM GENERAL CIRCULATION MODEL SIMULATIONS. Curt Covey, Stephen H. Schneider and Starley L. Thompson in *Nature,* Vol. 308, No. 5954, pages 21–25; March 1, 1984.

THE MICROBIAL ORIGIN OF FOSSIL FUELS

THE HOPANOIDS: PALAEOCHEMISTRY AND BIOCHEMISTRY OF A GROUP OF NATURAL PRODUCTS. Guy Ourisson, Pierre Albrecht and Michel Rohmer in *Pure and Applied Chemistry,* Vol. 51, No. 4, pages 709–729; April, 1979.

PREDICTIVE MICROBIAL BIOCHEMISTRY: FROM MOLECULAR FOSSILS TO PROCARYOTIC MEMBRANES. G. Ourisson, P. Albrecht and M. Rohmer in *Trends in Biochemical Sciences,* Vol. 7, pages 236–238; 1982.

PETROLEUM FORMATION AND OCCURRENCE. B. Tissot and D. H. Welte. Springer-Verlag, 1984.

A SUPERLUMINOUS OBJECT IN THE LARGE CLOUD OF MAGELLAN

CENTRAL OBJECT OF THE 30 DORADUS NEBULA, A SUPERMASSIVE STAR. Joseph P. Cassinelli, John S. Mathis and Blair D. Savage in *Science,* Vol. 212, No. 4502, pages 1497–1501; June 26, 1981.

THE NATURE OF R136A, THE SUPERLUMINOUS CENTRAL OBJECT OF THE 30 DORADUS NEBULA. Blair D. Savage, Edward L. Fitzpatrick, Joseph P. Cassinelli and Dennis C. Ebbets in *The Astrophysical Journal,* Vol. 273, No. 2, Part 1, pages 597–623; October 15, 1983.

THE MOST LUMINOUS STARS. Roberta M. Humphreys and Kris Davidson in *Science,* Vol. 223, No. 4633, pages 243–249; January 20, 1984.

THE PROTEINS OF ONCOGENES

PROTEIN KINASE ACTIVITY ASSOCIATED WITH THE AVIAN SARCOMA VIRUS GENE PRODUCT. Marc S. Collett and R. L. Erikson in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 75, No. 4, pages 2021–2024; April, 1978.

TRANSFORMING GENE PRODUCT OF ROUS SARCOMA VIRUS PHOSPHORYLATES TYROSINE. Tony Hunter and Bartholomew M. Sefton in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 77, No. 3, pages 1311–1315; March, 1980.

FUNCTIONS AND ORIGINS OF RETROVIRAL TRANSFORMING GENES. J. Michael Bishop and Harold Varmus in *RNA Tumor Viruses,* edited by R. Weiss, N. Teich, H. Varmus and J. Coffin. Cold Spring Harbor Laboratory Monograph Series, 1982.

REGULATION OF CELL GROWTH AND TRANSFORMATION BY TYROSINE-SPECIFIC PROTEIN KINASES: THE SEARCH FOR IMPORTANT CELLULAR SUBSTRATE PROTEINS. Jonathan A. Cooper and Tony Hunter in *Current Topics in Microbiology and Immunology,* Vol. 107, pages 125–161; 1983.

THE GREAT TEMPLE OF TENOCHTITLÁN

THE BASIN OF MEXICO: ECOLOGICAL PROCESSES IN THE EVOLUTION OF CIVILIZATION. William T. Sanders, Jeffrey R. Parsons and Robert S. Santley. Academic Press, 1979.

THE TEMPLO MAYOR OF TENOCHTITLÁN: ECONOMICS AND IDEOLOGY. Eduardo Matos Moctezuma in *Ritual Human Sacrifice in Mesoamerica,* edited by Elizabeth H. Boone. Dumbarton Oaks, 1984.

FIBER OPTICS IN PLANTS

PHOTORECEPTOR OPTICS. Edited by A. W. Snyder and R. Menzel. Springer-Verlag, 1975.

PLANTS AND THE DAYLIGHT SPECTRUM: PROCEEDINGS OF THE FIRST INTERNATIONAL SYMPOSIUM OF THE BRITISH PHOTOBIOLOGY SOCIETY, LEICESTER, 5–8 JANUARY, 1981. Edited by H. Smith. Academic Press, 1981.

FIBER OPTIC PLANT TISSUES: SPECTRAL DEPENDENCE IN DARK-GROWN AND GREEN TISSUES. D. F. Mandoli and W. R. Briggs in *Photochemistry and Photobiology,* Vol. 39, pages 419–424; 1984.

THE MECHANICAL MANIPULATION OF RANDOMLY ORIENTED PARTS

DETERMINING SHAPE AND REFLECTANCE USING MULTIPLE IMAGES. B. K. P. Horn, R. J. Woodham and W. M. Silver. M.I.T. A.I. Laboratory, Memo 490; August, 1978.

A ROBOT SYSTEM WHICH ACQUIRES CYLINDRICAL WORKPIECES FROM BINS. Robert B. Kelley, John R. Birk, Henrique A. S. Martins and Richard Tella in *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. SMC-12, No. 2, pages 204–213; March/April, 1982.

COOPERATIVE BREEDING IN THE ACORN WOODPECKER

SOCIAL ORGANIZATION AND BEHAVIOR OF THE ACORN WOODPECKER IN CENTRAL COASTAL CALIFORNIA. Michael H. MacRoberts and Barbara R. MacRoberts. The American Ornithologists' Union, Ornithological Mono graphs No. 21, Allen Press, 1976.

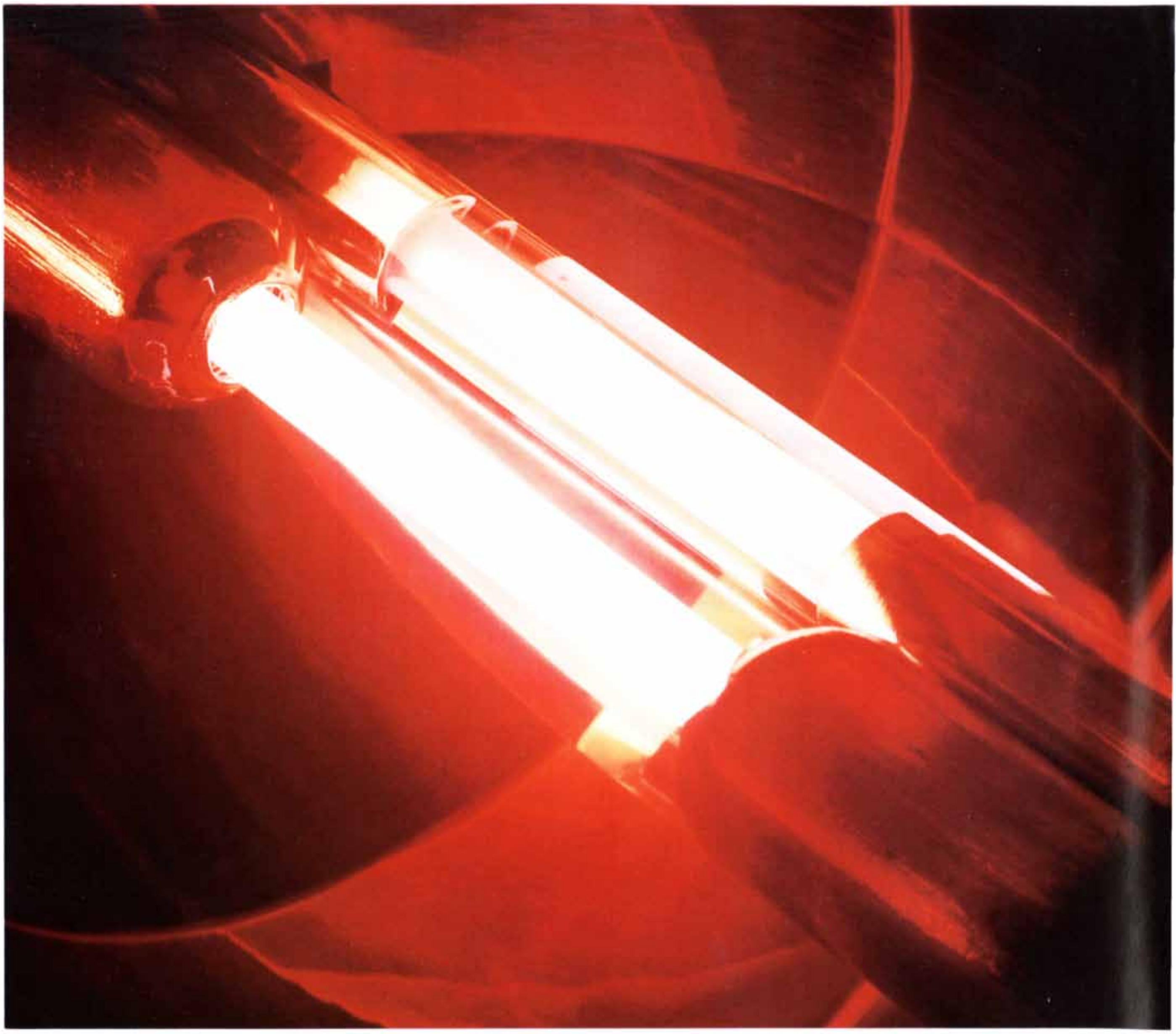
AVIAN COMMUNAL BREEDING SYSTEMS. Jerram Brown in *Annual Review of Ecology and Systematics,* Vol. 9, pages 123–155; 1978.

COOPERATIVE BREEDING STRATEGIES AMONG BIRDS. Stephen Emlen and Sandra Vehrencamp in *Perspectives in Ornithology: Essays Presented for the Centennial of the American Ornithologists' Union,* edited by Alan Brush and George A. Clark. Cambridge University Press, 1983.

THE AMATEUR SCIENTIST

LEANING TOWER OF LIRE. Paul B. Johnson in *American Journal of Physics,* Vol. 23, No. 4, page 240; April, 1955.

DOMINO "CHAIN REACTION." Lorne A. Whitehead in *American Journal of Physics,* Vol. 51, No. 2, page 182; February, 1983.



In a new technique developed by Hitachi researchers, crystals are excited by light from an alkali vapor lamp to produce a highly efficient "optically pumped" laser.

One drive is worth a thousand words.