SCIENTIFIC AMERICAN

AUGUST 1990
$2.95

The threat of Third World ballistic missiles.

Fighting the opportunistic infections of AIDS.

Why some ceramics are superconductors.

Reconstructions of Maya murals: *the artist's paint box provides invaluable aid to scientists studying these fast-fading pictures.*

The World Championship-winning Mercedes-Benz team now uses only one make of tires: Goodyear Eagles.

A Goodyear Eagle "contact patch." Where an Eagle demonstrates its superiority.

Mercedes-Benz is known throughout the world for quality, technical innovation and engineering brilliance.

Nowhere are all of those qualities more apparent than in the highly exotic and very sophisticated Championship-winning Mercedes-Benz race cars.

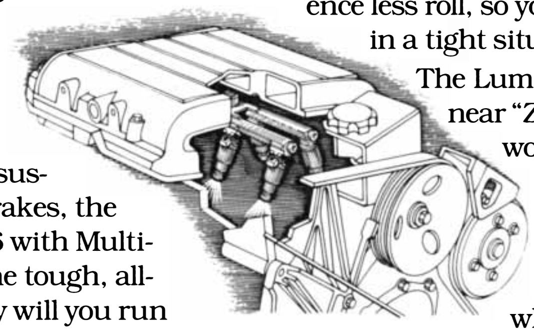
These are the cars that, in 1989, won seven of the eight endurance races that constitute the FIA

Lumina Sedan. It's designed to steer clear of trouble.

Granted, most of us will never need to understand the science behind near Zero Scrub Radii, drag coefficients, heat dissipation in braking, or, for that matter, modern day algebra. But everyone can understand the benefits of a car designed to help them respond to trouble on the road.

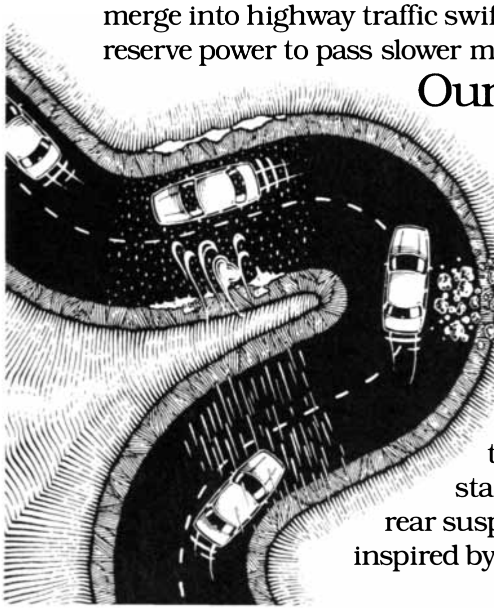
So, we'd like to explain in some detail some very technical reasons why you should choose the Lumina Sedan to be your next family car. Because when you put all those benefits together, they help make the Lumina Sedan one of the most innovative and responsive Chevrolet automobiles ever.

The engineering that went into the engine.



Can you see the safety aspects of an engine? Our engineers can. By taking into account the weight of the car, suspension geometry, tires and brakes, the Lumina's available 3.1 Liter V6 with Multi-Port Fuel Injection becomes one tough, all-around performer. So, not only will you run your family around the city efficiently,* you can merge into highway traffic swiftly and have enough reserve power to pass slower moving vehicles.

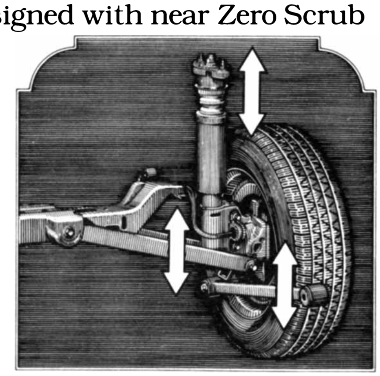
Our suspension won't keep you in suspense.



You'll appreciate Lumina's suspension most when the road takes a turn for the unexpected. For starters, the Lumina's rear suspension design was inspired by the Corvette. Its leaf spring is made of fiberglass, not steel. And instead of one steel spring per wheel, as in many cars, one fiberglass rear leaf spring is all that's needed in the Lumina. And with its advanced suspension, your Lumina is designed to help you recover quickly should you have to swerve to avoid something in the road.

Here's how. In a swerve, when you turn the wheel hard to the left, the right side of your car will push down. Straighten the wheel, and as the right side comes up, the left side will go down. This is called roll. But thanks to the tuning of Lumina's fiberglass rear leaf spring suspension, you'll experience less roll, so your car will respond more quickly in a tight situation.

The Lumina's front suspension also boasts near "Zero Scrub Radii." Unfortunately, it would take pages to explain it. So suffice it to say that in some cars when you drive through a puddle at 35 miles an hour your car will slow down and your steering wheel will pull. To lessen this, the Lumina was designed with near Zero Scrub Radii which reduces the sensitivity of your front wheels to changes in the road to help you stay in control without having to overcompensate at the steering wheel. A very helpful feature.

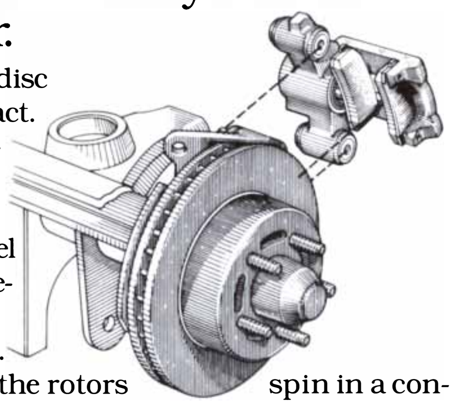


We'd like to clear the air about aerodynamics.

Admittedly, the technical aspects of aerodynamics, such as crosswind stability and drag coefficients, are less than thrilling. We'd much rather describe for you how the Lumina's aerodynamics can contribute to your safety. For instance, Lumina's smooth shape allows for a low drag coefficient of 0.34 which contributes to impressive fuel economy* and acceleration when passing or merging. But those aren't the only rewards of smooth aerodynamics.

Believe it or not, the Lumina is not only aerodynamic when the wind is rushing over the front end, it's aerodynamic when the wind is coming at you from the side. And you'll appreciate that on very windy days or when you get caught in the wake of a passing semi on the interstate.

Here's the brake you've been looking for.

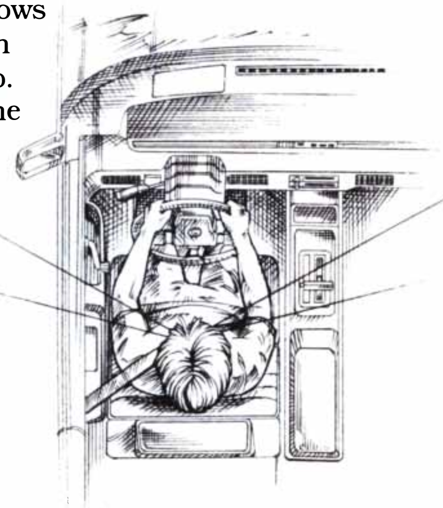


Four-wheel disc brakes, to be exact. They come standard on the Lumina Sedan. Lumina's 4-wheel disc brakes benefit from greater heat dissipation. That's because the rotors spin in a constant air flow to keep them cool. And the cooler the brakes stay, the more efficiently they work. Now you could buy a Ford Taurus and not get these standard. Or you can buy a Lumina and get standard 4-wheel disc brakes. The choice is yours.

Know what you're getting into.

Innovation isn't limited to exterior features of the Lumina Sedan. Here are just two examples of some thoughtful interior design that may not be evident at first glance.

Notice all the glass. The windows seem deep and wide, don't they? While providing a very comfortable feeling of expansiveness, this also serves a very practical function: better visibility. And while all that glass helps eliminate blind spots, standard tinted windows all the way around can cut down on glare, too.



Next, look at all the buttons, knobs and controls on the dash. Notice how logical their placement, how easy they are to reach. A simple thing, true, but consider its importance and what can happen in that split second when your eyes leave the road and search for a button.

Now we hope we didn't bore you with all these explanations, but we thought the more you knew about the Lumina Sedan before getting into it, the more you and your family would be able to get out of it.

To find out even more, come in and test drive a Lumina for yourself or, better yet, ask a friend who's driving one. Either way, you're sure to see why Lumina is the fastest-selling 1990 new car name. And why more Americans are winning with The Heartbeat of America.

THE Heartbeat
OF AMERICA IS WINNING.
TODAY'S CHEVROLET™

*EPA estimated MPG city 19/highway 27. Chevrolet, the Chevrolet emblem, Lumina and Corvette are registered trademarks of the GM Corp. ©1990 GM Corp. All Rights Reserved. Let's get it together... buckle up. GM®

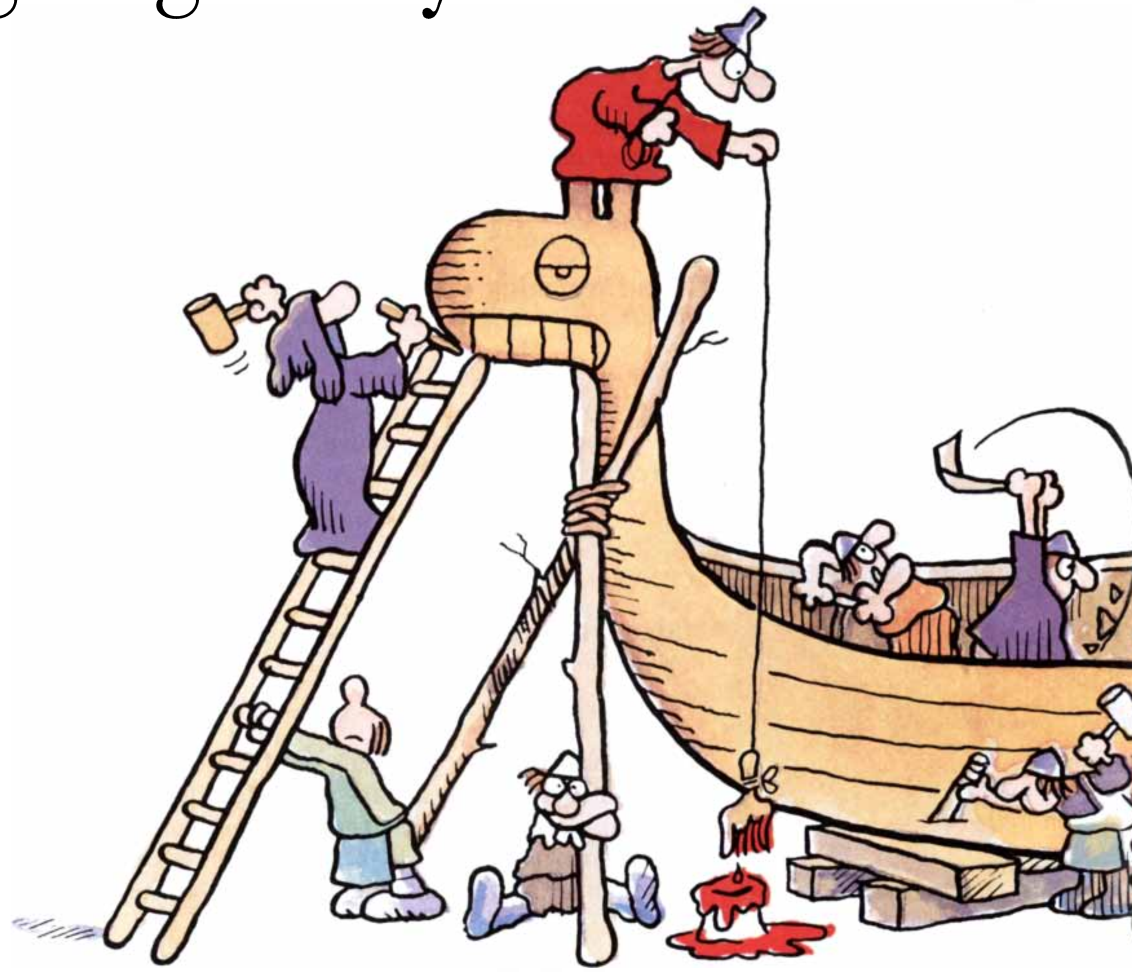
SCIENTIFIC AMERICAN

August 1990 Volume 263 Number 2

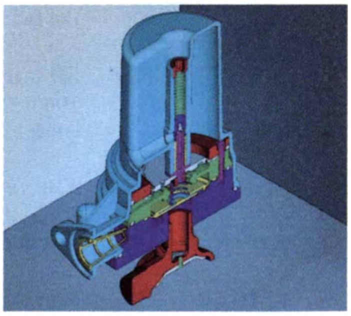
Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017. Copyright © 1990 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. 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 rates: one year $36, two years $64, three years $88 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. only 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537.

DEPARTMENTS

The IBM RISC System/
Designing on any other workstation



Whatever you're creating, you'll sail into a whole new age with any of the four POWERstations in the RISC System/6000 family. Because POWER (Performance Optimization With Enhanced RISC) processing can give you performance you've probably only dreamed about:

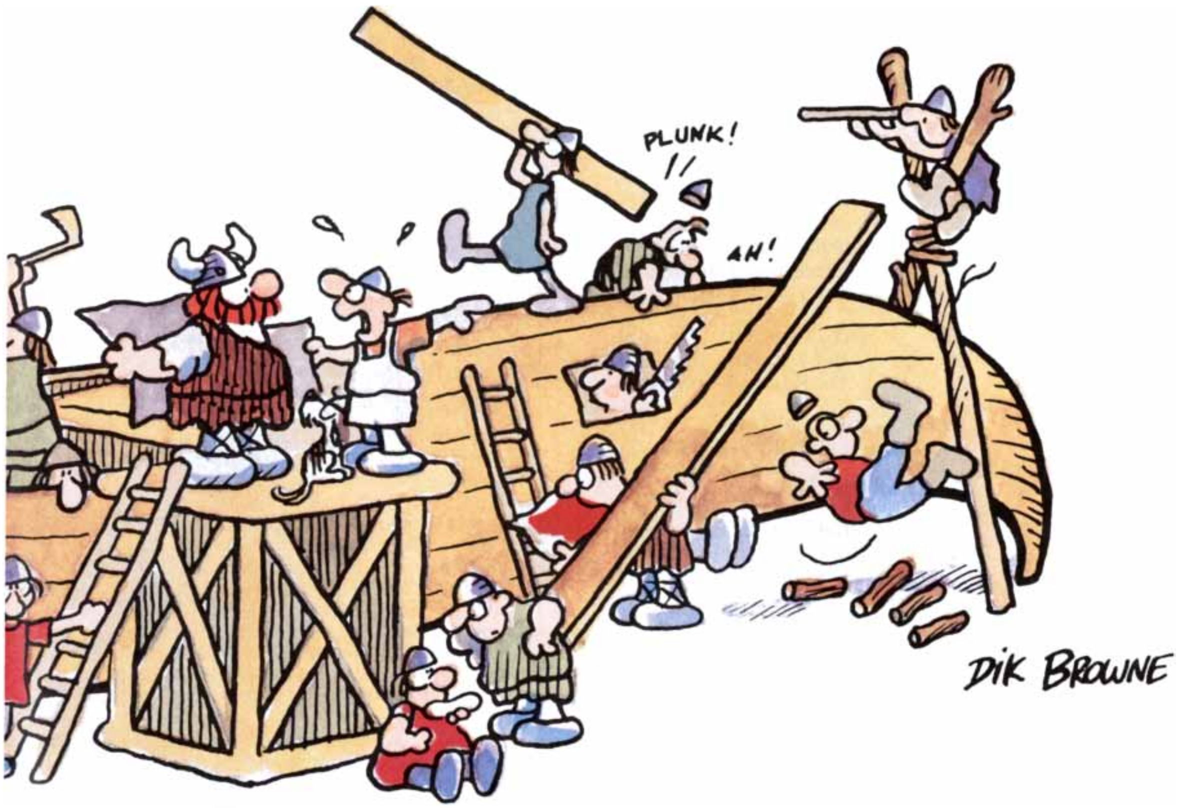


up to four instructions per machine cycle, 42 MIPS and 13 MFLOPS. Suddenly, complex designs don't take eons anymore.

The four RISC System/6000 POWERstations feature a range of graphics processors from grayscale to Supergraphics to satisfy any graphics demand. Great news for Power Seekers working on animation, scientific visualization, medical imaging and engineering solutions like CADAM,™ CAEDS™ and CATIA.™ And for electrical design automation, there's IBM's all new CBDS™ and an arsenal of over 60 EDA appli-

IBM is a registered trademark and RISC System/6000 and CAEDS are trademarks of International Business Machines Corporation. CADAM is a trademark of CADAM INC. CATIA is a trademark of Dassault Systèmes. CBDS is a trademark of Bell Northern Research Corporation.

6000™ family.
will seem downright primitive.

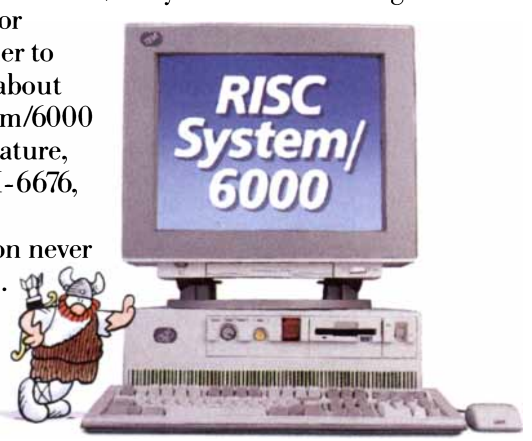


cations from more than a dozen vendors.

With every POWERstation, you can get an almost unimaginable palette of 16 million colors, which gives you 3D images so realistic, they fairly leap off the screen, with super sharp resolution of 1,280x1,024 pixels. And when it's time to call in the heavy artillery, the POWERstation 730 draws nearly one million 3D vectors per second. Like all POWERstations, it can come complete with its own graphics processor, freeing the POWER processor to rapidly create and analyze your designs. All at prices that won't sink anybody's budget.

So if you're tired of paddling upstream with yesterday's performance, call your IBM marketing representative or Business Partner to find out more about the RISC System/6000 family. For literature, call 1 800 IBM-6676, ext. 991.

Civilization never looked so good.



For the Power Seeker.

IBM®

THE COVER painting of a Maya priest is part of a mural at Bonampak in Mexico. It was painted in the mid-1970's by Felipe Dávalos for the Florida Museum of Natural History in Gainesville. The work of Dávalos and other artists may soon be the only record of rapidly fading Maya paintings and decorated reliefs. The original murals at Bonampak, for example, are falling prey to fluctuating temperatures, humidity and microorganisms. Some attempts to preserve them did more harm than good (see "Maya Art for the Record," by June Kinoshita, page 92).

THE ILLUSTRATIONS

Cover photograph by Merle Greene Robertson

Page	Source
35	Impact Photo via Jane's Information Group
36-37	Gabor Kiss
38	Johnny Johnson
39	George Retseck
40	Johnny Johnson
43-47	Michael Goodman
48	Lynn F. Schneemeyer, AT&T Bell Laboratories
49	Johnny Johnson
51	William Travis, National Cancer Institute
52-53	Ian Worpole
54	Erwin Feurstein, National Institutes of Health (*left*), Henry Masur (*right*)
55-56	Ian Worpole
57	Johnny Johnson
60-61	H. O. Peitgen, H. Jürgens, D. Saupe and C. Zahlten, *Fractals: An Animated Discussion*, with Edward Lorenz and Benoit B. Mandelbrot (video)
62	Andrew Christie (*top and bottom left*), Hartmut Jürgens, Heinz-Otto Peitgen and Dietmar Saupe (*bottom right*)
63	Edward Bell (*left*), Hartmut Jürgens, Heinz-Otto Peitgen and Dietmar Saupe (*middle and right*)
64-65	Hartmut Jürgens, Heinz-Otto Peitgen and Dietmar Saupe
66	Hartmut Jürgens, Heinz-Otto Peitgen and Dietmar Saupe (*left*), Richard F. Voss, International Business Machines Research (*right*)
67	H. O. Peitgen and P. Richter, *The Beauty of Fractals*
69	Thomas L. Beck, University of Cincinnati
70-71	Ian Worpole
72	Hai-Ping Cheng, University of Chicago (*top*), Ian Worpole (*bottom*)
74	Ian Worpole
77	Patricia J. Wynne
78	Steven N. Handel
79-80	Patricia J. Wynne
82-83A	Steven N. Handel
84-85	George Retseck
86-88	Patricia J. Wynne
89-91	John Deecken
92	City of Bristol Museum and Art Gallery
93	Norman Hammond, Boston University
94	Merle Greene Robertson
95	Mary Miller, Yale University
96	Norman Hammond, Boston University
97	Norman Hammond, Boston University (*top left*), Justin Kerr (*top right*), Merle Greene Robertson (*bottom*)
107-109	Michael Goodman

SCIENTIFIC AMERICAN®

Established 1845

LETTERS

To the Editors:

The article "Test Negative" ["Science and the Citizen," SCIENTIFIC AMERICAN, March] concluded that virtually no reliable research supports the nationwide increase in workplace drug-testing programs. The author, John Horgan, badly misrepresented a study I directed involving 198 (not 116) employees of the Georgia Power Company who tested positive for drugs.

Although final results of the study were released in September, 1989, Horgan used extremely early findings issued one year before. This error was brought to his attention while he was still gathering information, but he apparently chose to ignore it.

Instead of finding no significant behavior or performance decrements among drug users, we found that drug users had filed for 8.4 times as much in medical benefits as demographically matched co-workers and had been absent 3.5 times as often. Those testing positive had also been involved in almost twice as many on-the-job accidents. Clearly, our findings did show significant differences between drug users and nonusers.

HOWARD WINKLER

Georgia Power Company
Atlanta

John Horgan responds:

I chose to focus on the data published by the National Institute on Drug Abuse rather than those still undergoing review. Nevertheless, all the findings of the Georgia Power study are flawed for at least three reasons.

First, the Georgia researchers acknowledged in their first report that "the primary subject for the data base is the problem employee." Indeed, subjects were tested because of high absenteeism or involvement in accidents. So of course they displayed those traits (as well as high use of medical benefits, which one would expect to be associated with high absenteeism and accidents) more often than their "demographically matched co-workers."

Second, none of the primary subjects were tested for alcohol, which is by far the most widely abused substance and may have contributed more to their poor performance than drugs did.

Finally, none of the control subjects were tested for drugs. Winkler has absolutely no idea whether they were nonusers, as he claims.

To the Editors:

The profile of Noam Chomsky ["Science and the Citizen," SCIENTIFIC AMERICAN, May] wrongly states that Robert Faurisson argued "most accounts of the Holocaust are exaggerated." What Faurisson actually argues is that there were no gas chambers, that the gas chambers and genocide are one and the same thing, and that, therefore, there was no genocide and the allegation that there was is a Zionist lie.

Holocaust revisionism is an ugly and deeply disturbing phenomenon. It is distressing to find it so breezily dismissed in a responsible journal.

ANNABELLE SHECKTER

Pittsburgh, Pa.

To the Editors:

"Chaos and Fractals in Human Physiology," by Ary L. Goldberger, David R. Rigney and Bruce J. West [SCIENTIFIC AMERICAN, February], describes the fractallike structures of dendrites on certain neurons. The dendrites have branches, these branches can also be seen to branch and, in an idealized fractal, this branching could go on to infinity.

To me, this sounds suspiciously like our federal bureaucracy.

WILLIAM P. BOLAND, Jr.

Nicasio, Calif.

To the Editors:

Contrary to Karen Wright's implication in "The Road to the Global Village" ["Trends in Communications," SCIENTIFIC AMERICAN, March], worldwide deployment of fiber-optic cables is not related to the emergence of advanced technologies with intensive data-transmission requirements. Rather these fiber-optic facilities are used chiefly to support basic telephony. (There are, after all, hundreds of millions of telephones.) Those who primarily benefit from these facilities are telephone companies and carriers and large corporations, which, by leasing their capacity, can route internal telephone traffic at a cost far below the public tariffs.

Most integrated-services digital networks (ISDN's) to date have been implemented privately rather than publicly. Here again, the principal users and beneficiaries have been large corporations. Universal availability of ISDN's would permit home users to make phone calls, receive high-speed fax transmissions and have meters read by local utilities simultaneously. Yet it is unclear whether there is a pervasive demand for such capabilities, however attractive they may seem.

Capitalism will probably treat information technology much as it does other technologies, as a vehicle for extracting and maximizing profits efficiently. People who see only a rosy future in the "global village" should ponder some of the other possible consequences, such as global bandwidth cheap enough to prompt U.S. corporations to relocate their computer-processing and data-entry operations to, say, the Philippines or India.

The telecommunications infrastructure of a country is expensive to build, maintain and upgrade. In advanced capitalist countries, the potential for profit offers a powerful inducement to invest in telecommunications. Countries where such incentives are weak or absent risk exclusion from the global village altogether. Left to itself, the global village could become a community of the economically privileged.

MARTIN MORELL

Lynx Technologies, Inc.
Little Falls, N.J.

ERRATA

The profile of Noam Chomsky incorrectly states that a group at Columbia University claimed its experiments in ape communication contradicted Chomsky's views. Although numerous other animal researchers have taken this position, the Columbia group was not among them.

Many readers have written to point out that in "The Early History of Indo-European Languages," by Thomas V. Gamkrelidze and V. V. Ivanov [SCIENTIFIC AMERICAN, March], the tree diagram omitted mention of Portuguese, Yiddish and many other languages. The illustration was only meant to show relations among some of the hundreds of Indo-European languages and not to be comprehensive.

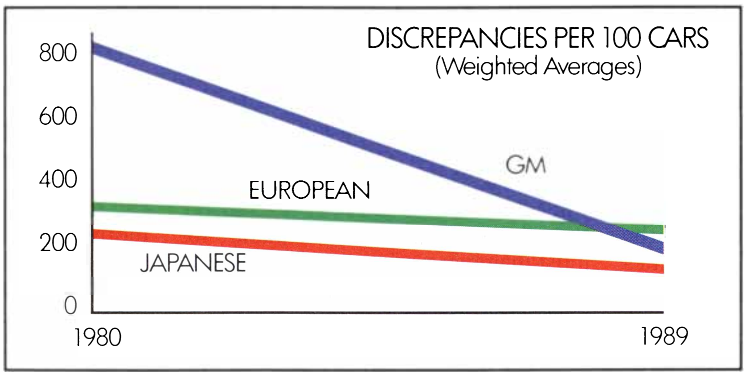
In "Mathematical Recreations" [SCIENTIFIC AMERICAN, March] the financial losses attributed to inflation were underestimated. During a period of 5 percent inflation, $1,000 tucked under a mattress loses half of its value in 14 years—not 20 years as stated.

YOU ASKED US TO CHANGE MORE THAN JUST THE LOOKS OF OUR CARS AND TRUCKS. AND WE HEARD YOU.

We've introduced 77 new models since 1985, dramatically improving almost all the cars and half the trucks we make. The looks are new, of course, but the big improvements are in the heart of our cars and trucks—in the engines, transmissions, and electrical systems.

Quality That Lasts

According to one quality survey by Harbour & Associates, GM vehicles are better built than the average European import, and the difference between GM and the average Japanese import is less than one-half of one discrepancy per car. Other studies confirm GM's improvement.



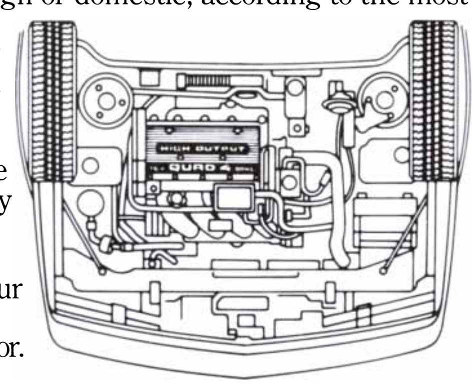
You can count on GM cars for the long term. When J.D. Power studied the dependability of 1985 models, GM was ranked highest in vehicle dependability among all American manufacturers.* No other U.S. carmaker has done the job better in the last five years.

What does this kind of quality mean to our customers? Just ask. After six months of ownership, at least 95% of all Chevrolet, Pontiac, Oldsmobile, Buick, Cadillac, or GMC Truck owners would recommend a vehicle from that division to a friend.

(*J.D. Power and Associates Vehicle Dependability Index Study.SM In a ranking of the three domestic manufacturers, based on things gone wrong to 4-to-5-year-old 1985 model vehicles in the past 12 months.)

Dependable Engines

GM's engines are more dependable than those of all other domestic carmakers. Our 3800 V-6 is at the top in engine quality among engines from all makers, foreign or domestic, according to the most comprehensive customer-based survey in the auto industry.



In the latest evaluation of 1989 engine quality, the 2.3-liter Quad 4 was as problem-free as 2-liter engines from Toyota or Honda after 3 months' ownership. In the latest survey of a full year of ownership (for 1988 models), Cadillac owners reported fewer engine problems than Toyota or Mercedes-Benz owners, a tribute to the 4.5-liter V-8 engine. In 1990, we're dedicated to delivering the quality you demand.

Problem-free Transmissions

Any car or truck is only as reliable as its transmission. Our automatic transmissions are more problem-free than those of most imports, and more problem-free than any domestic competitor.

All of GM going all out for you.

We Care About You

Today there is a new pride at GM. A new commitment to quality. A clear focus on our customers and on their needs.

People throughout General Motors, the GM Quality Network, are dedicated to continually improving our vehicles. Year after year. Until every model is as good as the best in the world.

PUTTING QUALITY ON THE ROAD

GM

MARK OF EXCELLENCE

Chevrolet Pontiac Oldsmobile Buick Cadillac GMC Truck

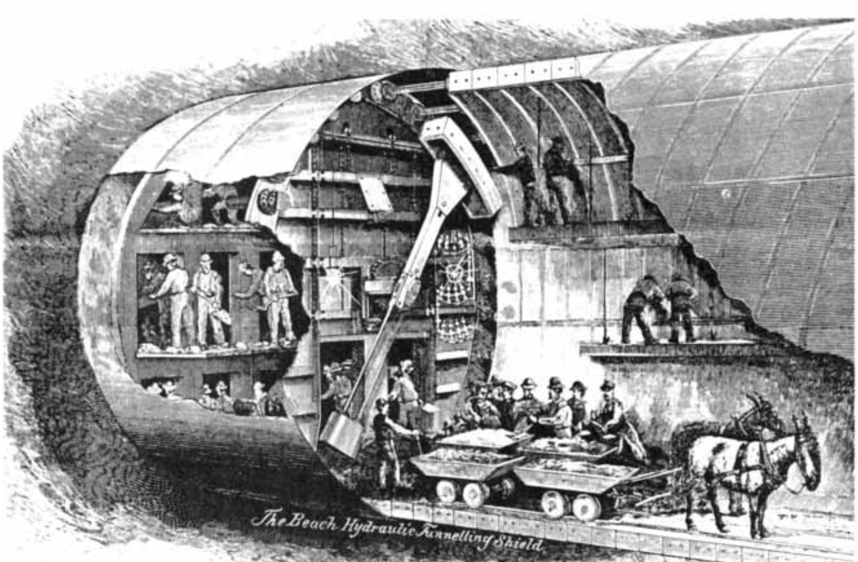
50 AND 100 YEARS AGO

AUGUST, 1940: "Ernest Lawrence's next cyclotron will be tagged for fundamental research, to unfold the secrets of atomic power and transmutation. It will generate 100,000,000 electron volts, perhaps much more. So powerful an engine of atomic disintegration is dangerous as well as useful. Even the present cyclotron—one twentieth the size of the one to come—is treated with respect by scientists. If you should stick your hand into the deuteron beam which is sometimes released as a spectacular stream of bluish-lavender light you would be burned as if you had fooled with a blow torch."

"The New York World's Fair offers a great spectacle of illumination. Take just one item—the fluorescent lamp. Here is an efficient and practical source of daylight lighting. Invisible ultra-violet energy, produced by an arc, is converted into visible light through the medium of fluorescent powders lining the tubes. The new lamp produces about 40 lumens per watt, the household filament only 15 lumens."

"The President is right when he asks for 50,000 airplanes for our national defense, and advises building 20,000 military and naval aircraft in one year. We can do so, provided: Methods of selecting prototypes and placing production orders are greatly simplified and accelerated; Methods of inspection and approval are accelerated; Trade schools make a great effort to train immense numbers of mechanics; Trade unions remove their restrictions on the means of entry of new men; Universities train the engineers that will be required; The Allied commissions immediately impart all their war-gained information to our Air Services; The Chief of Air Corps and the Chief of the Bureau of Aeronautics use the dictator's methods in boldly selecting certain types, and in boldly constructing those types; Constructors follow the German plan of building planes and engines which will last 100 hours instead of 5,000 hours because the war-time life of a plane is far shorter than the peace-time life of a plane—we must maintain performance, maneuverability, and gun fire at the sacrifice of durability."

"Every honest and upright person, citizen and alien alike, should welcome a universal fingerprint law. Through its enforcement would come several personal as well as national advantages. Fingerprinting has made possible positive identification of amnesia victims, unknown dead, missing persons, and so on. The records would give a continuously corrected census of the nation and a complete record and control of all persons in the United States, whether resident or entering or leaving. It would prevent the deported from re-entering and the fugitive from leaving the country. The records would make possible a speedy call to arms and would prevent draft dodging."



SCIENTIFIC AMERICAN

AUGUST, 1890: "Mr. Thomas A. Edison claimed that an epidemic of yellow fever could be prevented if on the discovery of the first case or cases the ground about the house and streets about the block were saturated with a germicide. If the theory of the propagation of yellow fever along the surface of the earth be correct, and there is every reason to believe it is, the plan of Mr. Edison is well worthy of trial."

"In a recent lecture before a scientific club, Professor Elihu Thomson declared that much higher speeds than can now be obtained with steam locomotives are to be expected by means of electricity. He believed that if we could come back after another hundred years, we would find 150 miles an hour to be the speed of traveling."

"A wretch named Kemmler, whose crime had been the atrocious murder of a woman, was appointed to be the first to suffer electrical death. The doomed man was strapped to a stout chair, electrodes were placed so as to make contact with top of head and base of spine, an alternating electrical current from a powerful Westinghouse generator was joined, a switch was moved, and the criminal was struck dead—instantly killed by lightning. The execution of a criminal, whether by the guillotine, the garrote, the gallows, the gun, or the dynamo, is a ghastly business; and it is not surprising that the sensational newspapers, aided by the electrical opponents of the law, should have made the most of such an occasion to fill their columns with revolting details."

"A tunnel to extend between Port Huron, Mich., on the American side, and Sarnia, on the Canadian side, is being bored by means of the Beach hydraulic shields, the invention of Mr. Alfred E. Beach, of the SCIENTIFIC AMERICAN. The object of the shield is to protect the workmen while excavating the earth and building the tunnel. It consists of a strong cylinder somewhat resembling a huge barrel with both heads removed. By means of a system of hydraulic jacks capable of either combined or separate action, it is possible to govern the direction of the tunneling shield with the utmost precision."

This will help you spend money overseas.

© Visa International 1987.
Reproduced with permission.

This will help you save money overseas.

Calling home from overseas can be a lot less expensive when you use *AT&T* ***USADirect***® Service. Just dial the ***USADirect*** number for the country you're calling from, and you'll be connected to an *AT&T Operator* in the U.S. within seconds. It's fast. It's easy. It's available in over 50 countries. And it helps you minimize hotel surcharges. You save with AT&T's economical international rates, whether you use your *AT&T Card* or call collect. For your information card, call **1-800-874-4000, Ext. 301.**

AT&T
The right choice.


PRODUCT OF ENGLAND
BOMBAY
SAPPHIRE
Distilled
LONDON
DRY GIN
1 Litre 94° Proof
BOMBAY
IMPORTED
ENGLISH GIN
INTRODUCING BOMBAY SAPPHIRE.

SCIENCE AND THE CITIZEN

Radioactive gaggles, kuru in Europe, cane toads get licked, Mr. Gene Therapy

Dark Days

Eastern Europe brings to mind the West's polluted past

The lungs tighten, eyes water, skin stings with the touch of corrosive gases. An occasional tang of sulfur is almost welcome. It is hard to believe that anything can grow here. This is not some imagined hell of apocalyptic environmentalists; it's just another spring day downwind of East Germany's brown-coal district. Cities once famous for their cultural and political history—Leipzig, Dresden, Weimar—are now gaining fame as environmental disaster areas. Magdeburg, Cottbus, Plzeň, Teplice, Krakow, Katowice, Copsa Mica: the litany goes on.

These places are stark reminders of the danger of industrialization run wild. Statistics speak of significantly reduced life expectancy; birth defects afflict as many as 10 percent of infants in northern Bohemia, according to the Czech Green party. The anecdotal evidence is clear, too: a middle-aged man coughs his lungs out at a roadside refreshment stand in Teplice. The blue, cloud-dappled sky lets only a bleak, gray light down to the ground.

Is there really something unique about the environmental devastation in Eastern Europe? Some observers suggest that enthusiasm for bemoaning the region's fate reflects the West's deliberate amnesia about its own environmental crimes of the not-too-distant past. It was only in the late 1940's, for example, that darkness at noon in Pittsburgh stopped being considered a sign of robust industry, in the early 1960's that lethal fogs no longer enveloped London (the fog of 1953 killed more than 2,000 people).

Before its shutdown in the late 1970's, the Inco smelter at Sudbury in Ontario pumped into the air 1 percent of all the sulfur emitted in the world annually. No trees grew within a 10-mile radius. Even today Eastern pollution may have a match in the brown cotton-wool smog that envelopes Los Angeles or Denver.

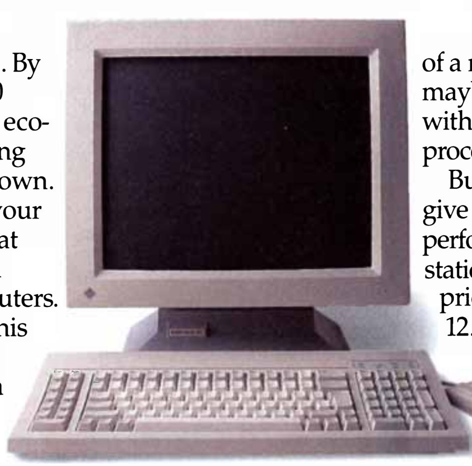
Totalitarian industry need not in principle have caused more pollution than its democratic counterpart, but it did. Executives in the West may have wished for secret police to jail the activists who dumped toxic sludge on their desks, stopped up outflow pipes and gathered grassroots support for reform. Executives in the East got that wish. Environmental data behind the Iron Curtain were classified top secret.

Centrally planned socialism played a role in the debacle more insidious but no less compelling than that of the secret police. Western governments also jailed the odd environmentalist (and

ROMANIAN TOWN of Copsa Mica has gained a reputation as the most polluted place on the earth. Photo by Gad Gross.

WHY DID WE
A 12.5 MIPS W
FOR $

BECAUSE



SPARCstation™ SLC. By breaking the $5,000 barrier, it turns the economics of computing absolutely upside down. And lets you give your people a system that runs circles around PCs and minicomputers.

Up till now, for this price you'd have to settle for a 386 with one or two little extras. Or a share of a minicomputer. Or maybe an X terminal with an underpowered processor.

But now, you can give everyone a high-performance RISC workstation for the same price. Complete with 12.5 MIPS of processing power. The UNIX® operating system. 8 MB of memory. Even built-in Ethernet®.

But more than a hot machine, you can give them all the strategic advantages that come with our SPARC™ technology.

First, our workstations were meant to be networked. So you can have whole groups working on the same project. In perfect synch. And every user has serious processing power. Right at the desktop level. Second, there's the OPEN LOOK™ user interface. Which lets people learn to use UNIX in no time flat. Finally, there's

still do, as the Earth First! tree spikers can attest), but Western companies focused on the bottom line. When government regulations made polluting unprofitable, businesses tended to stop doing it. Behind the Iron Curtain, however, people not only suffered the destruction of their environment, they subsidized the process heavily. In an economy where money did not matter, neither did the costs of pollution. East German factories, for example, paid environmental fines totaling more than $50 million in 1988. So what? They weren't run to make a profit.

Nowhere are the perverse incentives of centralized socialism more evident than in the energy business: expensive energy might typically spur conservation but not in a country where the state foots the bill and boasts of the yearly increase in coal production. East Germans, for example, use 25 percent more energy every year to support a much lower standard of living than their Western cousins. Insulation, central heating and double-glazed windows are virtually unknown. Residents of Prague discharge almost twice as much water every day (to an overburdened sewage system) as do residents of equally picturesque Vienna.

The sulfur- and dust-laden air, however, is only partly socialism's fault. The region has little oil or natural gas to speak of and no hard currency to import the oil that has powered free-market development. People must make do mostly with lignite, a mediocre excuse for fuel that contains about 90 percent carbon, up to 5 percent sulfur and the rest silicates and other minerals. The closed economy has wrought other peculiar catastrophes: in Czechoslovakia, for example, cheap phosphate fertilizer from Gabon has poisoned the land with toxic levels of cadmium. (Fertilizers have contributed to general pollution of land and water as well, but that is hardly a socialist problem.)

The cramped scale of Central European geography intensifies the pollution problems. The rivers and aquifers into which factories dump their wastes are the same ones that cities tap for drinking water. The land that is strip-mined is the same land on which people must live and grow food. Although the region most afflicted is about the size of Ohio and western Pennsylvania, proportionately it is as if the entire U.S. east of the Mississippi were industrial wasteland.

So what do the newly liberated nations intend to do to clean up their acts? Now that the state no longer owns the factories, some Eastern and Western officials lament, no one can give a single order to install pollution controls. That argument is specious, but questions do persist about the ability of Eastern governments to enforce any new environmental laws. They must build environmental protection agencies essentially from scratch.

Then, of course, there is the matter of money. Cost estimates for decontaminating East German air, water and soil, for example, run from $40 to $150 billion. Of course, not all of that pollution is the East's fault. Since the 1970's more than 10 million tons of West German toxic waste have found their way to a belt of landfills just east of the soon-to-vanish border. The new, united Germany will have to find a new home for its continuing flow of waste. A single landfill in Ketzin may cost as much as $600 million to clean up.

The Czechs have yet to finish their environmental analysis. What are the worst problems? "Everything. The air, the water and the land." Jiří Nechvátal of the state environmental committee shrugs his shoulders. At least his country was largely too poor to buy pesticides. In Poland, officials concede they do not have the $400 million it would take to halt environmental damage, much less the $25 billion they believe is needed to reverse it.

Some money may be forthcoming from the newly organized European Bank. And the shift to free markets may also help the environment: many of the most polluting plants are antiquated and unprofitable even without the expense of antipollution equipment. They will probably shut down on economic grounds alone. Other plants—such as those in East Germany eyed by West German investors—may receive dispensations to continue operating during a transition period that could last as long as 10 years.

What responsibilities will Western companies take on when they buy Eastern plants? Any exports to the European Community will have to meet environmental standards there, but the E.C. will not regulate the environmental quality of the production lines themselves. "That would be presumptuous," says Hans Hermann of the West German environmental ministry.

The East bloc's few ecologically beneficial practices may not survive the free-market style of economic development. West German Environment Minister Klaus Töpfer has said he admires the East German recycling system, which reuses 40 percent of its paper, glass and metals, but other officials decry the high prices paid for scrap as distortions of the marketplace.

Mass transit is another chance for development to go awry. It is heavily subsidized in East bloc countries, not to mention that no one can afford a car. The two-lane main highways could not accommodate more vehicles in any case. There are obvious financial incentives for laying down a Los Angeles-style network of roads from Stettin to Sofia—but regrets loom farther down the line. The Easterners will have trouble resisting the mistakes the West has made. *—Paul Wallich*

Big country, big problems

The Soviet Union does not have the cramped spaces that make industrial pollution so much of a residential problem for most of Eastern Europe. Instead, reports Nikolai N. Vorontsov of the State Committee for Environmental Protection, the nation's vast territories contain equally vast tracts of polluted land, air and water.

Lake Baikal: The largest volume of fresh water in the world, once pristine, is now polluted by paper and chemical plants.

Azov Sea: Polluted by pesticides from rice farms on the Kuban River, its salinity is rising as fresh water is diverted for irrigation. "The ecosystem of the whole area has been destroyed," Vorontsov says.

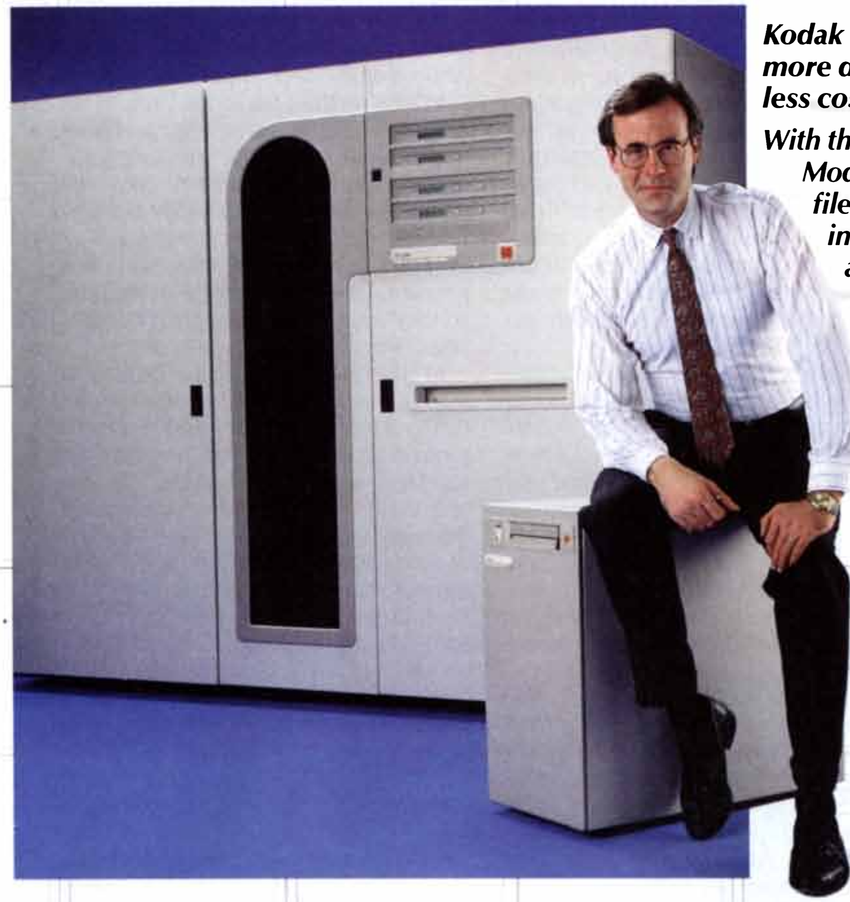
Aral Sea: Diversions for irrigation lowered water levels by 15 meters, creating "a new desert." Winds pick up dust laden with salt and pesticides and strew it over thousands of square kilometers.

Byelorussia: The Chernobyl nuclear disaster rendered almost a quarter of all farmland unusable. Data suggest a fifth of the population should leave.

Ural Basin, eastern Ukraine, south-central Siberia: Air pollution is "a problem," and "we have found bad levels of clarity in the air and water."

"Our main problem," Vorontsov says, "is that there is a very small level of cultural understanding of ecology in the Soviet Union. There is a lack of money, but money is not the main problem.

"We used to think that this country was very rich, that it had big resources. We thought we didn't have to pay attention to the environment." *—Fred Guterl*



Hot Geese

Contaminated wildlife roam nuclear reservations

Those pesky Canadian geese. First they took over the Northeast, laying waste on suburban lawns. Now they have moved south to Oak Ridge, Tenn., where some of them have become too hot to handle. In fact, they have become radioactive—as have some white-tailed deer and mallards, gadwalls, and other waterfowl on the 35,300-acre Oak Ridge Reservation.

Oak Ridge is not the only nuclear-weapon facility beset by wildlife woes. Department of Energy sites throughout the country are studded with basins and ponds where radioactive waste was dumped, and contamination has found its way into flora and fauna. Now there is increasing concern that "hot" wildlife could become hunters' dinners, moving the radiation to humans. "It is a very serious concern," says Kenneth Bossong of Public Citizen, a watchdog group in Washington, D.C.

At the Savannah River Plant in South Carolina former DOE engineer William Lawless remembers when radioactive turtles were found two miles from the site on a commercial hog farm. "It showed that there was a really serious problem with contaminants going off-site that were not being monitored."

One Oak Ridge goose had 3,950 picocuries of radioactive cesium 137 per gram of breast meat. In terms of Christmas dinner, just one pound would deliver almost 100 millirems, which is the generally accepted standard for annual exposure. Officials for Martin Marietta Energy Systems, the DOE's Oak Ridge contractor, insist the situation is not dangerous. "There is no way that would cause any health effects," says James G. Rogers, coordinator of environmental protection activity. Rogers says handling such a goose would expose someone to "much less than a chest X ray."

But eating contaminated game may be another matter. The low-level waste contains radionuclides that can remain in the body for many years. Cesium 137, with a half-life of 30 years, concentrates in muscle tissue; strontium 90, with a half-life of 27 years, is deposited in bone. "The prevailing consensus is that there is danger of damage at any level," says Scott Saleska, staff scientist at the Institute for Energy and Environmental Research.

Despite the DOE's claims that wandering wildlife are not worrisome, contamination has necessitated the killing of between five to eight geese each year since 1985, according to James W. Evans of the Tennessee Wildlife Resources Agency. And the DOE is studying geese and other Oak Ridge waterfowl to determine their radiation levels and tagging them to see how far they roam. "We are very concerned about anything moving off the reservation," says Gordon Blaylock, senior researcher in Oak Ridge's environmental sciences division.

It's not just gaggles of geese and sords of mallards—it's herds of deer. An exploding deer population pushed roadkills in the Oak Ridge area to an all-time high five years ago. So the reserve initiated a public hunt with the understanding that bounty must be tested for radiation and may be confiscated. "If hunters get a nice eight- or 10-point buck they get frustrated. But we take a picture of them with the deer so they can show their hunting buddies later," Rogers says.

At the Hanford, Wash., facility—a 570-square-mile reserve where the DOE recycles uranium and removes plutonium—radioactivity has infused tumbleweed, ducks, rabbits, coyotes and other animals. Sodium-rich waste dumped in pits since the 1940's served as a salt lick for animals, says Timothy Connor, now at the Energy Research Foundation and formerly of the Hanford Education Action League. Rabbits ate the salt, coyotes ate the rabbits and "next thing you know you've got teams of workers at Hanford patrolling with Geiger counters and picking up radioactive coyote droppings."

Similarly, at the Savannah River facility, whose reactor was closed in 1988 because of safety and managerial problems, some fish and turtles have been contaminated. Michael H. Smith, director of the University of Georgia's ecology laboratory at the site, reports levels ranging from 10 to 100 picocuries of cesium 137 per gram in minnows, darters and bass. "I wouldn't want to eat those fish," Smith says, also citing a host of other contaminants.

During the past several years, as information about the extent of contamination has finally reached the public, the DOE has been pressured to clean up the sites. Still, critics worry that unless more enforcement power is given to the EPA, weapon facilities could slip out of regulatory control. A bill passed by the House of Representatives and pending in the Senate would create "real incentives to comply, such as EPA enforcement and penalties," says Dan W. Reicher, a Natural Resources Defense Council attorney.

Under the 1976 Resource Conservation and Recovery Act and Superfund, the DOE is required to decommission seepage basins at Oak Ridge and other laboratories. But the extent to which wildlife has been contaminated is unknown. In addition to birds, deer, turtles, fish and burrowing gophers at Los Alamos National Laboratory in New Mexico, Lawless says, "you've got insects. They have not been looked at yet." —*Marguerite Holloway*

OAK RIDGE NATIONAL LABORATORY is one of three facilities on the 35,300-acre Department of Energy reserve where radioactive waste has contaminated wildlife.

Oravske Kuru

A human dementia raises the stakes in mad cow disease

Eva Mitrová, an epidemiologist at the Research Institute for Preventive Medicine in Bratislava, Czechoslovakia, is worried. Since 1976, 22 cases of a rare, fatal dementia have been diagnosed in Orava, a sparsely populated sheep-rearing region of the Slovakian republic. The incidence is accelerating: 12 of the cases have occurred within the past three years. Another 19 cases are clustered around Lučenec, 80 miles to the south. "The people are extremely afraid," Mitrová says.

The outbreak in Slovakia has been identified as Creutzfeldt-Jakob disease (CJD). It kills within seven months after the first symptoms appear. CJD normally occurs in about one person per million every year throughout the world: the incidence in Orava is several hundred times that, according to Mitrová. CJD resembles kuru, a fatal infection that was common among Papua New Guinean tribes that ate and handled the brains of their dead during mourning rites. In addition, CJD has close similarities to so-called mad cow disease, which has caused 15,000 cattle to be destroyed in Great Britain since 1986.

Mitrová's concern is shared by other researchers. Most investigators think that CJD, kuru and mad cow disease—bovine spongiform encephalopathy, or BSE—are all unconventional slow infections that are variants of scrapie, a disease of sheep that causes spongy degeneration and deposits of a fibrous "scrapie protein" in the brain. D. Carleton Gajdusek, chief of the laboratory of central nervous system studies at the U.S. National Institute of Neurological and Communicative Disorders and Stroke, who is collaborating with Mitrová, calls the Slovakian CJD outbreak "Oravske kuru." Gajdusek suggests that BSE and Oravske kuru—as well as the rapid spread of scrapie in the U.S.—indicate that a worldwide epidemic of "kuru virus" started during the 1970's. "We have a major problem in human disease," cautions Gajdusek, who won a Nobel prize in 1976 for establishing that kuru can be transmitted to chimpanzees.

What has most alarmed epidemiologists is that Oravske kuru may be the result of scrapie jumping from sheep to humans. Many of the Orava villagers coming down with CJD had frequent and close contact with sheep, Mitrová says. Moreover, she has recently found the characteristic brain degeneration of scrapie in sheep from Orava. "If [Mitrová's] numbers are accurate, I would have difficulty being reserved about that," comments Robert G. Rohwer, who is a scrapie microbiologist at the University of North Carolina at Chapel Hill.

Scrapielike diseases are promiscuous in their host range. Within the past few months a BSE-like disease has been reported in zoo antelopes and in domestic cats in Great Britain. There is near hysteria there because of the fear that BSE might be transmitted to people, and beef prices have slumped. The U.S. Department of Agriculture has inaugurated a surveillance program for BSE and is tracking down cattle imported from the U.K.

Yet in spite of the sudden appearance of BSE in Great Britain, scrapie and its variants, including CJD, are not easily transmitted in nature. Even though scrapie has long been present in that country, CJD is not common there. But what is difficult in nature can be helped with technology. "If you try hard enough, you can eventually transmit these agents to almost any animal" by injecting infectious extracts directly into the brain, remarks Richard F. Marsh, a scrapie researcher at the University of Wisconsin at Madison. "You may have to use 40 or 50 animals, and then you'll get it," adds Paul W. Brown, a co-worker of Gajdusek's and an expert on CJD.

At least 25 cases of CJD in Western countries can be traced to what Gajdusek refers to as high-technology cannibalism. Fourteen resulted from injections of growth hormone or gonadotropin extracted from human pituitary glands, according to Brown. And more cases are likely because of the long incubation period of the disease. Three cases arose in patients given grafts of dura mater, skull-lining tissue taken from cadavers that is used for patching in surgery. At least six were transmitted by unsterile instruments or electrodes employed during neurosurgery; one case was transmitted through a corneal graft and one by root-canal dentistry, Gajdusek says.

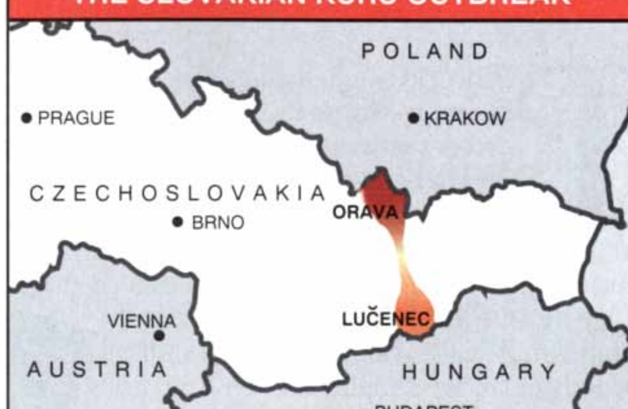
BSE, in a similar fashion, spread silently in Great Britain during the early 1980's by technological misfortune. Scrapie-infected sheep carcasses were rendered into protein supplements fed to cattle. Most likely, changes in rendering practices allowed the scrapie agent to infect cattle via their feed. The government has now banned any use of ruminant products in ruminant feed, as well as all use of cattle offals thought to be capable of harboring the scrapie agent.

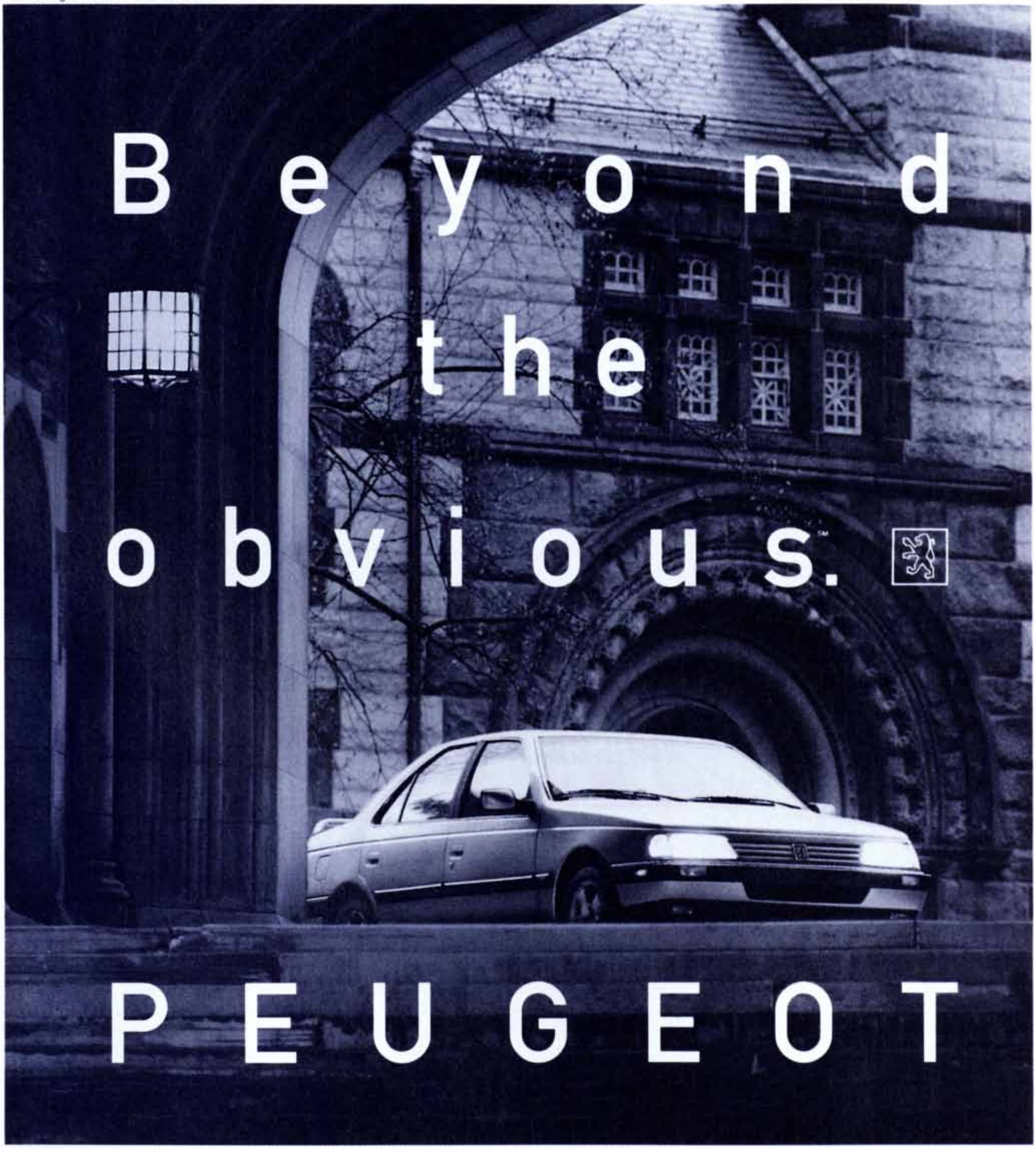
The nature of the causal agent is bitterly controversial. The agent certainly has extraordinary properties: it elicits no immune response and is resistant to treatments that kill most other infectious agents. Brown has four times repeated an experiment apparently showing that scrapie survives—just barely—a temperature of 360 degrees Celsius, hot enough to break down

SLOW BRAIN INFECTIONS

Scrapie	Sheep, goats
Bovine spongiform encephalopathy—mad cow disease	Cattle
Transmissible mink encephalopathy	Mink
Creutzfeldt-Jakob disease Kuru Gerstmann-Sträussler syndrome	Humans
Chronic wasting disease with spongiform encephalopathy	Captive Rocky mountain elk and mule deer
Spongiform encephalopathy	Nyala Gemsbok Domestic cat

THE SLOVAKIAN KURU OUTBREAK





1991 405 models from $15,300 to $21,700. †Call 1-800-447-2882. †MSRP. Excludes tax, title, options, registration and destination charges. *Based on R.L. Polk & Co. owner retention study of 'MY 1984-1986.

After more than a century of building fine automobiles, Peugeot creates cars so well-conceived that some of the best qualities of the car may not be apparent at a glance.

But those willing to take the time to look more closely will find themselves richly rewarded. With a distinctive European automobile whose rare combination of intelligent engineering, legendary driving comfort and enduring style has won the acclaim of automotive enthusiasts the world over.

All of which only begins to explain why people who own Peugeots keep them longer than most import cars on the road.*

Evidently, once you've looked beyond the obvious, it is difficult to see anything less.

amino acids and the base pairs that make up DNA and RNA. "That's rather distressing," he says.

The scrapie agent is also unusually resistant to ultraviolet radiation and to other treatments that destroy nucleic acids. One view, entertained by Stanley B. Prusiner, a scrapie researcher at the University of California at San Francisco, is that the agent may be the aberrant fibrous protein that is deposited in infected brains. If the rogue protein could somehow catalyze its own formation from a healthy version of the protein that is present in all brains, it might not need any nucleic acid. Gajdusek has even speculated that a mineral could play a role.

Rohwer thinks that the protein-only theory of scrapie is "increasingly implausible." His analysis of heat and radiation experiments leads him to conclude the agent is probably a small virus. Other experiments demonstrate the existence of different strains of scrapie, suggesting some genetic component. "The Andromeda-strain stuff is based on an overinterpretation of disinfection experiments," Rohwer surmises. He fears contamination may explain some results.

The most common view seems to be an intermediate one: the agent is a protein protecting a tiny amount of nucleic acid too small to code for protein but big enough to carry a little genetic information. Marsh is a supporter of that idea. He has found a fragment of mitochondrial DNA in scrapie proteins that have been extracted from infected brains. He has not yet been able to prove its role by identifying a sequence that is present only in scrapie extracts.

With no obvious technological catastrophe to blame, why is CJD so common in Orava? Genes may play a part. Experiments show that mutations in the gene that codes for the healthy scrapie protein can affect an animal's susceptibility to the disease. Lev G. Goldfarb, Brown, Gajdusek and their co-workers have found a specific mutation in the human scrapie-protein gene in five unrelated individuals who were stricken with CJD, including some cases from Orava.

So Oravske kuru could be the result of genetically susceptible people being exposed to the scrapie agent. "I think both factors are important," Mitrová says. Half of the Orava cases are in families that have other cases, which lends support to a role for genes. Yet many CJD cases do not carry the mutation. A lot could be hanging on the efforts of Gajdusek and Mitrová to solve the mystery. —*Tim Beardsley*

BUFO MARINUS, ***or cane toad, secretes venom from glands on its upper back.***

Bufo Abuse

A toxic toad gets licked, boiled, teed up and tanned

Plump, greenish-yellow and pebbly in texture, it is not much to look at. It can be a nuisance, too, poisoning dogs and squishing noisily under automobile tires. But *Bufo marinus,* also known as the cane toad, has become an international celebrity of late, inspiring drug-war hysteria in the U.S. and trade talks in the Far East. Here is its tale, warts and all.

This large toad—some attain the girth of Frisbees—once lived quietly in the warmer regions of the Americas, ranging as far north as central California. In the 1930's it was exported to Australia in an attempt to control beetles infesting cane fields (hence its common name). Like many amphibians, *B. marinus* wards off predators by secreting a toxic goo from glands in its skin. The secretion contains a compound called bufotenine, which resembles the neurotransmitter serotonin and also occurs in certain toadstools and plants. Although these bufotenine-containing substances can be lethal, they have reportedly been used as intoxicants by some "primitive" societies.

Intrigued by these accounts, U.S. researchers synthesized bufotenine and began testing it in humans—along with many other psychoactive drugs—in the 1950's, according to Stephen Szara, chief of the biomedical branch of the National Institute on Drug Abuse. These investigators, he says, hoped to gain insights into schizophrenia and other mental disorders. The Pentagon and the Central Intelligence Agency also supported the work as part of an effort to develop brainwashing agents.

In one experiment, researchers injected bufotenine into four inmates of an Ohio prison. The prisoners experienced hallucinatory effects "reminiscent of LSD and mescaline" as well as nausea and chest pains, the scientists reported in the May 18, 1956, issue of *Science.* The investigators also observed that "if the color of an eggplant were diluted, it would approximate the unique purple hue of the faces of the subjects." These side effects, Szara says, discouraged further studies of bufotenine—of an official nature.

The U.S. Drug Enforcement Administration outlawed bufotenine in the late 1960's. Ironically, the DEA's action inspired a few people to try licking live toads, says Darryl S. Inaba, director of drug programs at the Haight-Ashbury Free Medical Clinic in San Francisco, ground zero of the 1960's drug culture. But these adventurers became sick rather than high, he adds, and toad licking never caught on.

For the past two years, however, newspapers have been filled with lurid accounts of cane toad abuse. In April, 1988, *USA Today* reported that Australian "hippies" were "forsaking traditional drugs for cane toads, which they boil for a slimy, potentially lethal cocktail." Although Australian authorities have denied the story, it apparently primed the media for more. A few months later Inaba gave a lecture on drugs in which he mentioned—for comic relief, he says—the rare 1960's practice of toad licking. Soon reporters all across the world were calling to inquire about "crazed hippies licking toads in the mountains," says Alex Stalcup, director of the clinic.

Inaba and Stalcup assured the reporters that there is no evidence that anyone is ingesting toad secretions—an assertion that law-enforcement officials confirm. "It's not anything the Drug Enforcement Administration is worried about," says Cornelius Dougherty, a spokesperson in Washington, D.C. But the story persisted. "Toads take a licking from desperate druggies," exclaimed the *New York Post* this past January. "How low will people stoop to get high?" sneered the Santa Clara (California) *Press Democrat.*

Inevitably, reality imitated fiction. Last year, Stalcup says, two teenagers in New Mexico ingested cane toad toxin after reading stories about the "fad" and had to be hospitalized. An Australian youth died after eating cane toad eggs. "This rumor has caused a lot of misery," Stalcup says.

Meanwhile Australia is coping with other toad-related problems. Cane toads have become so abundant in the northeast of the country that they have driven out indigenous amphibians and have poisoned other wildlife and pets. Some dogs have even become compulsive toad lickers, according to Glen Ingram, a herpetologist at the Queensland Museum. "It gives them a kick, perhaps like alcohol," he explained to a newspaper. Angry Australians have organized "toad eradication" outings; golf clubs and cricket bats, while officially frowned on, are the favored instruments of execution.

Yet the citizens of Brisbane may have found a silver lining in the toad glut. They recently persuaded China, whose traditional medicines include compounds from other *Bufo* species, to explore the therapeutic potential of *B. marinus* venom. In April Brisbane's Office of Economic Development shipped 100 grams of the stuff to the Shanghai Industry Foundation. Other Brisbanians are marketing products made of toad leather. Although single toad skins are somewhat small, says Mark C. Underhill of the development office, they make "quite attractive" wallets, purses and even outerwear if artfully sewn together. If you can't lick them, join them. —*John Horgan*

that built the computers. Support that helps you get more from your Dell™ 486 than you would from other 486 computers.

Best of all, since we sell direct, cutting out the retailer and his markup, you can buy a complete Dell 425E for just $6,399. That's $5,954 less than Compaq's 33 MHz 386™ PC and $7,855 less than Compaq's 25 MHz 486†. Or you can lease our system for as low as $232 a month.**

Just call us. You'll get fast delivery of a computer with the works. Including a one-year limited warranty and next-day deskside service by the Xerox Corporation.△

Not to mention the full attention of a company that's been voted number one for overall customer satisfaction in all six *PC Week* polls of corporate volume buyers for personal computers.

To order one of the most powerful PCs around, call us now.

And soon we'll have another word for people like you. Satisfied.

TO ORDER, CALL NOW.
800-873-1120
IN CANADA, CALL 800-387-5752.

FOR NETWORKING/UNIX INFO
800-678-UNIX
HOURS: 7 AM-7 PM CT M-F 9 AM-2 PM CT ON SAT.

DELL COMPUTER CORPORATION

optional extras. All prices and specifications are subject to change without notice. Dell cannot be responsible for errors in typography or photography. **Payment based on 36-month, open-end lease. Leasing arranged by Leasing Group, Inc. In Canada, configurations and prices may vary. DELL SYSTEM is a registered trademark and monitor; and DeskPro 486/25 Model 120, 120 MB hard drive and video graphics monochrome monitor. UNIX is a registered trademark of AT&T in the United States and other countries. Other trademarks and trade names are used to identify the entities claiming the marks and names or their products. Dell Computer Corporation disclaims any Arboretum Boulevard, Austin, Texas 78759-7299, Attn: Warranty. ©1990 Dell Computer Corporation. All rights reserved.

ple notes, pieces of bowls left beside the track indicate that the round-bowl tradition in Neolithic England began no later than about 3796 B.C.

Even more astonishing, a jade ax hailing from the Alps shows that items produced in the heart of Europe were making their way to western England by 3796 B.C. "That's a date few people would have believed before," John Coles says. *—Ricki Rusting*

ANCIENT ROAD, found in a peat bog in 1970, is now literally showing its age.

A Clock in the Trees

Tree rings reveal the precise age of the oldest road

In 1970, when archaeologist John M. Coles began excavating an 1,800-meter-long wooden road buried in a peat bog in southwestern England, he suspected the structure was ancient. Radiocarbon analysis proved him right. The wood dated to about 4000 B.C., which meant the find was a relic of the Neolithic period (the late Stone Age) and was the oldest road yet discovered.

Coles was still not satisfied, however, because the radiocarbon results had a potentially large margin of error. Now Jennifer Hillam of the University of Sheffield and Mike G. L. Baillie of Queen's University of Belfast and their colleagues have come up with a much more precise date. The road, known as the Sweet Track, was built from trees felled in the winter of 3807–3806 B.C.

Hillam and Baillie, who report their findings in the journal *Antiquity*, found the answer by measuring and counting tree rings. Each ring on a tree represents a year's growth, and the thickness of a ring reflects the growing conditions of that year. Therefore, when two trees show the same sequence of thin and thick rings, the trees can be assumed to have laid down the matching patterns at the same time under the same environmental conditions.

The first step was to construct a master tree-ring chronology for the track. The dendrochronologists made complete plots of sequences from individual wood samples in the track and then aligned the patterns of one plot with the same patterns from other samples to produce a composite sequence. They then matched this master chronology to a segment of a reliably dated, over 7,000-year-long chronology derived from trees that grew in Europe between 5289 B.C. and A.D. 1983. The correspondence indicated the trees used for the Sweet Track grew between 4202 and 3807 B.C.

Closer examination of the wood showed the trees were axed after forming virtually complete rings in 3807 B.C., which means the track builders actually felled them after the autumn of 3807 B.C. and before growth resumed in the spring of 3806 B.C. Coles and his wife, Bryony, who collaborated in excavating the Sweet Track, surmise that the builders processed the wood in the cold months and then constructed the road, which traversed a swamp, as water began to subside from the swamp in the early spring of 3806 B.C.

The Sweet Track chronology fills a gap in a larger chronology now being developed for England. It also greatly sharpens the dating of many Neolithic relics in England. The Coleses have shown that the road served for only about 10 years before being overcome by the swamp and buried in peat [see "The World's Oldest Road," by John M. Coles; SCIENTIFIC AMERICAN, November, 1989]. It is clear, then, that any artifacts found beside the structure were made before or within a decade or so after 3806 B.C.—as, by implication, were many similar prehistoric artifacts found elsewhere. Thus, the cou-

Unsolved Mysteries

New data suggest a solution to the solar-neutrino problem

Preliminary results from an experiment in the U.S.S.R. seem to confirm that the sun is producing far fewer of the elusive particles known as neutrinos than current theories predict. The finding, if it holds up, could lead to revisions in particle physics, astrophysics and cosmology. "It is devastatingly exciting," says John Ellis of the European laboratory for particle physics (CERN) near Geneva.

There are three kinds of neutrinos, corresponding to the three particles known collectively as leptons: the electron, tau and muon. Since neutrinos were first detected in 1956, they have become integral to models of particle interactions and of stellar evolution.

Physicists first suspected that something might be wrong with these models in the 1970's, when a detector in a mine in South Dakota recorded less than a third as many neutrinos coming from the sun as had been expected. (Placing the detector underground reduces false alarms from other kinds of cosmic radiation.) When that finding was corroborated in the late 1980's by a detector in Japan, the apparent deficit of solar neutrinos emerged as a major mystery in physics [see "The Solar-Neutrino Problem," by John N. Bahcall; SCIENTIFIC AMERICAN, May].

Some investigators, however, noting that the detectors in South Dakota and Japan are sensitive only to certain high-energy neutrinos, have questioned whether a deficit also exists among more common low-energy neutrinos. To resolve this issue, Soviet and U.S. scientists recently joined to build a low-energy neutrino detector in the U.S.S.R.'s North Caucasus region. Known as the Soviet-American Gallium Experiment, or SAGE, it began gathering data in January.

In June researchers reported results from the first two months of observation at Neutrino 90, a meeting held at CERN. The shortfall of solar neutrinos


© 1990 British Airways

observed by SAGE so far is even greater than that reported by the other detectors. According to the standard model of solar physics, SAGE should have detected about 14 neutrinos from the sun over a 60-day period, but it recorded none.

If the trend continues, says Alexy A. Starobinsky of the L. D. Landau Institute of Theoretical Physics in Moscow, it will spur a "drastic change in our notions about neutrinos." One tantalizing proposal is that electron neutrinos, the only type currently detectable, are somehow transformed into tau or muon neutrinos before they reach the earth. Such an effect is predicted by some grand unification theories, which attempt to explain all the forces controlling the behavior of matter. A profound consequence of these theories, Starobinsky notes, is that at least one type of neutrino has mass—perhaps even enough to account for the "dark matter" that many cosmologists think dominates the evolution of the universe. *—John Horgan*

The Sea Turtle's Tale

An enigmatic reptile reveals sophisticated homing

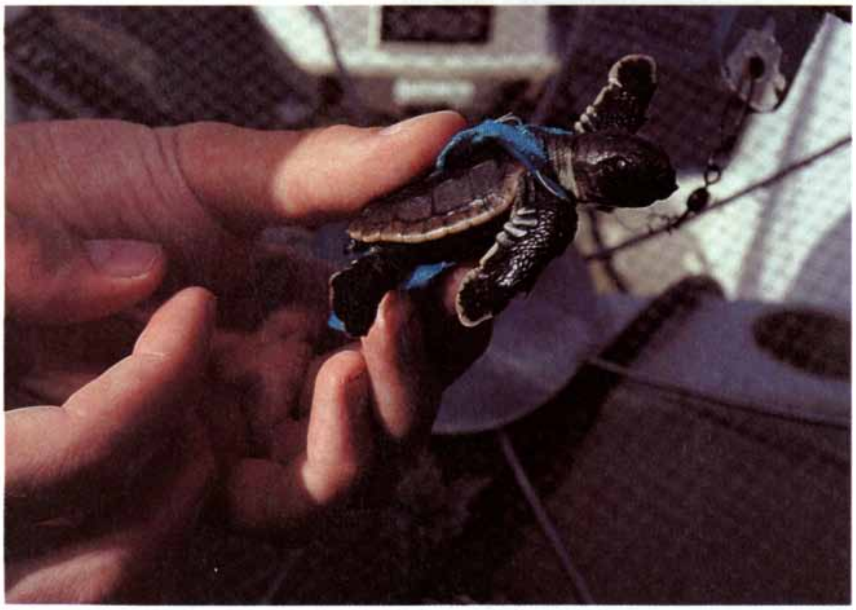
Sea turtles, which plied the oceans eons before mammals evolved, are renowned for their prodigious migrations. As soon as they hatch, they scramble down their nesting beaches toward the surf and strike out for the open ocean, arriving years later at offshore feeding grounds thousands of miles away. When fully grown—which may take three decades—the turtles make a return migration to the nesting grounds to breed. They then repeat the epic round-trip journey every few years.

Folk wisdom has long held that adult females always return to the nesting sites where they hatched, but proof has been lacking. Now one group seems to have provided that proof, and another has gone far toward explaining how the animals navigate on their long voyages. The new reports also raise troubling questions about the prospects of the imperiled reptiles.

The evidence that female green turtles return to their natal beaches to nest was presented in *Science* by Brian W. Bowen, a geneticist at the University of Georgia, and his co-workers, Anne B. Meylan and John C. Avise. They analyzed turtle mitochondrial DNA. Animals inherit these subcellular bodies only from their mothers. Mutations accumulate slowly in mitochondrial DNA, as in any other DNA.

The workers found that samples of mitochondrial DNA taken from different breeding populations were quite distinct. That means females must rarely switch from one population to another. If females often nested far from where they themselves hatched, the DNA from the different populations would be mixed up. "Geographic structuring is really strong evidence for the natal-homing hypothesis," Bowen says.

Confirming that sea turtles do indeed return to their birthplaces deepens the question of how they navigate.



SEA-TURTLE HATCHLING is fitted with a harness for an orientation experiment by Michael Salmon and Kenneth J. Lohmann. Photo by Tom Smoyer.

Marine biologists Michael Salmon of Florida Atlantic University and Kenneth J. Lohmann and Jeanette Wyneken of the University of Illinois at Urbana-Champaign have found clues. Sea turtles hatch at night, and it is known that the hatchlings' first instinct is to head toward the lightest part of the horizon—which is normally out to sea.

How the turtles keep to a course once they have left land far behind was still a mystery, however. By putting hatchlings into the sea at various places and tracking them, Salmon and his colleagues found that green, leatherback, and loggerhead turtles all use wave motion as a "compass." The clincher: loggerhead hatchlings head into artificial waves even in a dark laboratory. "They prefer to use the wave compass over anything else when they first leave the beach," Salmon says.

Yet there is apparently more to turtle navigation than that. By artificially reversing the magnetic field in a laboratory tank, Lohmann has shown that loggerhead hatchlings can also use magnetic fields as a compass. Like some other animals that respond to magnetic fields, loggerheads have a magnetic material—magnetite—in their brains. Lohmann believes that sea turtles, like pigeons, probably use both magnetic and other compass cues.

Lohmann also suspects that sea turtles might use magnetic fields to detect more than just direction. Because the angle of dip of the earth's magnetic field varies with latitude, and because longitudinal "stripes" in field strength exist across the oceans, the turtles could use magnetic fields to get an idea of where they are as well as to decide which way is north. Experiments to test the idea are under way.

The recent findings, Bowen says, suggest that conservation efforts must intensify if sea turtles are to survive. Because colonies never mix, overharvested turtle populations—in Peru and particularly in Mexico, where Bowen witnessed rampant poaching for eggs and skins—are unlikely to be replenished from protected populations.

In addition, the National Research Council estimates that 40,000 sea turtles perish each year in shrimp trawls. The five species found in U.S. waters are all on the endangered-species list. A law passed last year requires drift nets to be fitted with devices that allow turtles to escape, which should reduce the carnage. But for one species, Kemp's ridley, it may be too late. At the species' only major nesting beach in Rancho Nuevo, Mexico, numbers are down to less than 1 percent of their 1947 levels. *—T.M.B.*

PROFILE: GENE DOCTOR

W. French Anderson pioneers gene therapy

The idea of understanding genetic disease at the molecular level came early to W. French Anderson. He wrote about that topic on his Harvard undergraduate application in 1953, the year Francis Crick and James D. Watson discovered the structure of DNA. He was 16 years old at the time. Within four years he would conceive of gene therapy: supplementing defective genes with healthy ones.

In his cramped office at the National Institutes of Health in Bethesda, Md., where he is chief of the molecular hematology branch of the National Heart, Lung and Blood Institute, Anderson is now preparing for the first attempt to cure a genetic disease. This fall, barring unexpected hitches, the parents of one or more extremely ill children will sign consent forms for a new experimental treatment. The children suffer from severe combined immunodeficiency, or SCID, a rare, fatal condition known as bubble-boy disease because of the famous case of David, the Houston SCID patient who lived most of his life in a sterile, plastic bubble.

Even during an interview, Anderson makes himself constantly busy. He speaks quietly and deliberately about the background that led him to pioneer gene therapy. He grew up in Oklahoma in a bookish family—his mother was a journalist, his father an engineer—and his parents' attempts to persuade him to play sports and "be normal" were not notably successful. Anderson remembers spending much of his youth reading texts on astronomy, mathematics, chemistry and, significantly, a volume called *Heredity and You*.

Once at Harvard University, he decided that "it didn't make any sense to do astronomy because the earth's atmosphere prevents you from seeing very much." That left math or medicine. Although a natural mathematician, he ruled that field out as a career after observing that "real mathematicians were a lot smarter than I was." When he took a course on the structure and function of biological molecules, the die was cast.

Anderson did not forget math entirely. His first published paper, in 1956, established how to do arithmetic in Roman numerals—thus disproving the popular wisdom that computation is impossible in systems that have no place value. He came up with the same trick for Minoan Linear A and Minoan-Greek Linear B, which prompted *Time* magazine to feature him as a prodigy. Anderson laughs out loud when recalling a contest in which he easily beat a challenger by using Roman numerals to add together 10 four-digit numbers.

At the same time, the young undergraduate was scavenging scientific apparatus from around the Harvard campus. With help from one professor, George B. Kistiakowsky (who later became science adviser to President Dwight D. Eisenhower), Anderson and his collaborators succeeded in putting the record straight on the isomerization energy of *cis*-butene-2.

Soon after hearing Paul M. Doty lecture on biological molecules in 1957, Anderson began investigating how ultraviolet light affected the biological properties of a more significant molecule—DNA. By then the goal of gene therapy had become "a clear vision." The next year he went to work with Crick in Cambridge, England, where he met his future wife, Kathryn, in the dissecting room. (Kathryn was, he admits, better than he at dissection.)

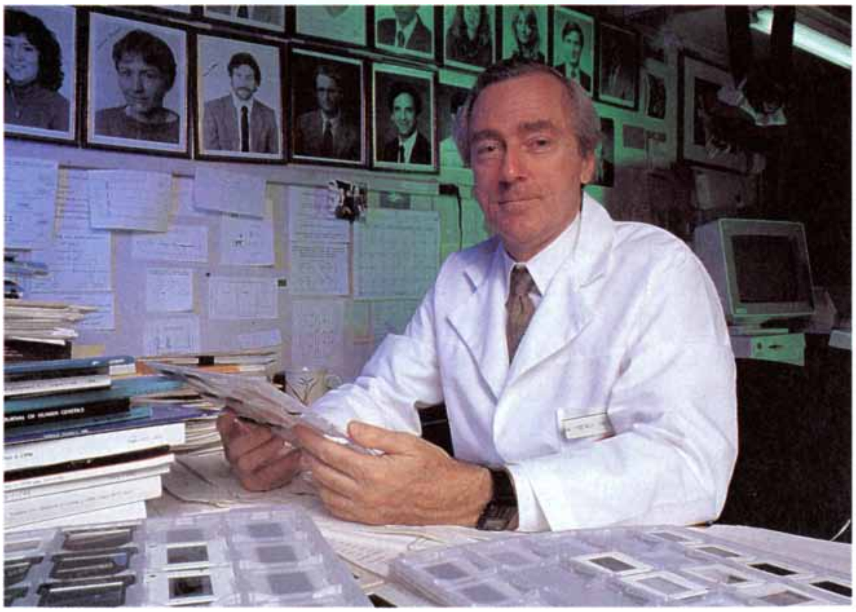
Anderson's research with Crick helped to provide evidence that the genetic code is organized by triplets, an insight that underlies all of molecular genetics. Anderson's time in Cambridge also coincided with one of the early triumphs of molecular biology, Sir John Kendrew's discovery of the structure of the muscle protein myoglobin. Anderson added one atom to the model.

One benefit of the two years he spent in England, Anderson says, is that he became accustomed to working in small quarters. That experience has proved useful at the NIH, where his windowless office is crowded with stacks of paper, much of it documentation for regulatory committees. Above the door are photographs of the two thalassemia patients, now dead, on whose disease Anderson built his scientific reputation.

The NIH has been Anderson's base of operations since 1965, after working his way through an M.D. from Harvard. Although salaries are "lousy" at the NIH, Anderson is committed to staying there. "If your primary interest is in doing research, this is the best place in the world." The reason: All NIH patients are admitted as research subjects. He has no other teaching or treatment obligations.

At the NIH, Anderson first worked with Marshall W. Nirenberg, who cracked the genetic triplet code. Anderson's own scientific star entered the ascendant in 1970, when he showed how universal factors promoting gene translation can be used to make hemoglobin from messenger RNA. "That paper made me," he recalls. "After that, the field was ours." A year later he used the technique to reproduce the molecular defects underlying two hereditary blood diseases, sickle cell anemia and thalassemia.

Around 1980 Anderson lost confidence in his research on blood diseases when it became clear that using the hemoglobin gene to correct anemias and thalassemias would be impractical



W. FRENCH ANDERSON frequently puts in 90 hours a week in his cramped office at the National Institutes of Health. Photo by Randy Santos.

Programming in FORTRAN no longer requires heavy machinery.

For all too many people, programming in FORTRAN means relying on a department mini or corporate mainframe.

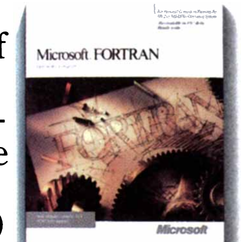
But with Microsoft® FORTRAN version 5.0, this no longer has to be the case.

Because now you can develop as well as maintain applications and subroutines. For the mainframe, the mini, and, of course, the PC.

All on that relatively light piece of equipment residing right on your desk. Otherwise known as your PC.

Now you're probably asking yourself how is this possible? Well, for starters, it supports FORTRAN 77 ANSI-standard syntax. Secondly, it has major VAX® extensions. And finally, it's completely SAA compliant.

And unlike "heavy" equipment, with FORTRAN 5.0 you can expect to find a range of high-quality development tools. Like our CodeView® debugger and Microsoft Editor. Plus, you can use Dynamic Link Libraries to link in portions of your code at runtime.



Giving you time to concentrate on other things. Like your work, perhaps.

To learn more, call (800) 541-1261, Department L88. You'll find that programming in FORTRAN doesn't have to be a real heavy experience.

Microsoft®
Making it all make sense™

Customers inside the 50 United States, call (800) 541-1261, Dept. L88. In Canada, call (416) 673-7638. Outside the U.S. and Canada, call (206) 882-8661.

for the foreseeable future. To stave off depression, he eased off from research for a few years so that he could also practice sports medicine. He became physician to the national tae kwon do team (he holds a fourth-degree black belt) and attended the team at the Seoul Olympics in 1988.

Not until 1984 did Anderson return to his dream of gene therapy. It was then that he decided gene therapy might be feasible for a disease that required only a small amount of gene activity to be restored. SCID emerged as the most likely candidate. Some SCID patients have a defective gene that causes them to lack the enzyme adenosine deaminase, or ADA. They fail to develop an immune system and so fall prey to every passing infection.

Since then, Anderson has devoted much of his time to winning regulatory approval for his clinical experiments. His first experiment involving the transfer of genes into patients, approved in 1989, was a collaboration with R. Michael Blaese and Steven A. Rosenberg of the National Cancer Institute. They used foreign genes to track tumor-killing cells in cancer patients. The proposed SCID therapy will be the first designed to benefit patients.

Sometimes, when members of a committee overseeing his research question him skeptically, Anderson loses his patient delivery and breaks into an exasperated stammer. Yet he claims not to mind the heavy burden of regulation that falls on him more than on most. "I've gone for so many years with people telling me I'm wrong," he says. "Now that people are taking me seriously, it's great."

Anderson does object when committees that are supposed to ensure compliance with ethical guidelines start to second-guess the science. The members who are particular thorns in his side, he says, "want beautiful data so we can be certain a clinical protocol will work perfectly." But to wait for those data would be to delay gene therapy too long for his patients. "As long as it's safe and as long as it is reasonably probable it will work, it ought to be done," he declares.

Approval for the gene therapy protocol was complicated by the recent approval of a new therapy for SCID, which consists of administering the missing ADA bound up with polyethylene glycol (PEG). PEG-ADA has benefited some patients, and it is far simpler than gene therapy. Anderson and his collaborators sidestepped the ethical complication by proposing to treat patients who are already receiving PEG-ADA. Only if gene therapy brings additional benefits will the investigators seek permission to withdraw PEG-ADA.

Final approval of the protocol for Anderson and his coinvestigators, Blaese and Kenneth W. Culver, was expected from the NIH's Recombinant DNA Advisory Committee on July 30; the first experimental treatment could come by autumn. There are risks. The procedure could conceivably initiate cancers. The reconstituted immune system might be ineffective. But the accumulated results of scores of experiments finally persuaded an NIH review committee in June that there was a reasonable chance that the infused cells will thrive and improve immune function, if only temporarily.

Assuming that permission is granted, blood will be drawn from SCID patients at the clinical center of the NIH. Their white cells will be cultured and then exposed to an engineered retrovirus containing a working version of the gene the patients lack. The virus will infect cells and transfer the gene to them, but it is designed to be self-limiting. If tests confirm that cells have started to produce the missing enzyme and that no infectious virus is present, the cells will be infused back into the patients.

If it works, the gene therapy could be a lifesaver for children who cannot receive the principal alternative treatment, bone-marrow transplantation. It could also usher in gene therapies for other diseases. After getting this far, Anderson is not likely to give up. He meticulously charts how he spends his time, and colored-in squares show that 90-hour workweeks are common.

Anderson and his wife, who is now vice-chairman of surgery at the Children's Hospital National Medical Center in Washington, D.C., decided early not to have children, preferring to dedicate themselves to their careers. But the couple has been "more than godparents, less than foster parents" to nine young people during the past 25 years. The youngsters, Anderson says, "found each other and found us." Photographs of his surrogate children decorate his office, along with pictures of students.

It's Saturday afternoon. As I take my leave, Anderson accompanies me outside to see the sun briefly before returning to his office, where he has been since eight o'clock this morning. He will take the evening off, he says, for his regular Saturday night date with Kathryn. Tomorrow he will be back at work. Doesn't he feel the need to relax sometimes? "People say to me, 'Go out and have some fun,'" Anderson says. "For me, this is fun." —*Tim Beardsley*

SCIENTIFIC AMERICAN August 1990 Volume 263 Number 2

Third World Ballistic Missiles

The spread of strategic weapons to unstable regions greatly increases the likelihood that they will be used. We cannot reverse the process and so must learn to manage its consequences

by Janne E. Nolan and Albert D. Wheelon

This spring the news media offered the public a rare glimpse into a shadowy, frightening world. On April 2 Iraqi leader Saddam Hussein announced that he had the means of killing half of the population of Israel. The next day Israel launched its second satellite into orbit with its powerful Jericho II rocket. On April 11 British customs agents impounded a shipment of metal tubes bound for Iraq, ostensibly for a petrochemical complex. The agents maintained—and one of the British suppliers later confirmed—that the meter-wide steel tubes were actually intended for the barrel of a huge gun capable of firing rocket-assisted shells more than 3,000 miles. Such a gun had in fact been designed in the 1960's by Gerald V. Bull, a former consultant to the U.S. Department of Defense who later became an international arms broker. On March 22, 11 days before Hussein made his pronouncement, Bull was found shot to death at the door of his home in Brussels.

JANNE E. NOLAN and ALBERT D. WHEELON have collaborated in the study of missile proliferation. Nolan, a senior fellow at the Brookings Institution, holds a doctorate from the Fletcher School of Law and Diplomacy at Tufts University. In the Carter administration she was a delegate to U.S.-Soviet negotiations on the limitation of arms transfers. From 1983 to 1986 she was the senior designee to the Senate Armed Services Committee for Senator Gary Hart and managed the senator's foreign and defense policy staff during his 1984 presidential campaign. Wheelon retired as chief executive of the Hughes Aircraft Company in 1988 and taught at the Massachusetts Institute of Technology in 1989. He has been a member of the President's Foreign Intelligence Advisory Board and the Defense Science Board. He earned a Ph.D. in physics from M.I.T. in 1952 and for 10 years designed ballistic-missile guidance systems. He then joined the Central Intelligence Agency, where as deputy director for science and technology he established the agency's broad program for the collection and analysis of technical intelligence.

Good may yet come of this bizarre episode if it draws attention to the newest phase in the arms race. Ballistic missiles and other means of long-range destruction, traditionally limited to a handful of industrialized nations, are fast becoming a fixture in many regional conflicts. The Third World military buildup is perhaps even more worrisome than its First World prototype, for it is far more likely to find expression in war.

There are several reasons why this should be so. In the past decade the number of countries in the missile club has more than doubled, to 18. Many of the new members have been at war or are embroiled in disputes. Unlike the major powers, these countries have not had time enough to perfect systems of command and control over their new strategic forces. They have had little time to learn to manage the complexities of military brinksmanship. Finally, because many regional conflicts overlap, an escalation in the arms race tends to convey itself from one area of tension to another.

For many years the big industrialized countries ignored the proliferation of ballistic missiles and sought political advantage by arming their clients. In doing so, they presumed that the bipolar alignment of power would restrain regional conflicts. The preoccupation with East-West issues overshadowed problems in the Third World. Smaller industrialized powers sold missiles to generate revenues to support their own military industries. Meanwhile the developing countries eagerly acquired missiles for the same reasons that had motivated their predecessors: to deter attack, intimidate enemies, build a technological base and win prestige.

Industrialized countries began to realize how missile proliferation threatened their interests during the 1982 war over the Falkland Islands, when Argentina used Exocet missiles supplied by France to sink a top-of-the-line British cruiser. Five years later, during the Iran-Iraq War, the Iraqis nearly sank a U.S. destroyer with the same system.

A coordinated strategy began to appear only in 1987, however, when the U.S., Great Britain, France, West Germany, Canada, Italy and Japan agreed not to export ballistic missiles capable of delivering a payload of more than 1,100 pounds over a range greater than 182 miles. Had this agreement come earlier, it might have been more effective. But now that developing states are themselves exporting missile technology, it seems difficult to put the

IRAN'S SCUD B missile, supplied by the Soviet Union, is seen being paraded through the streets of Teheran.

genie back in the bottle. In any case, arms control cannot be imposed by a few countries on all the rest—such great-power prerogatives are no longer sustainable.

The roots of the problem reach back several decades. The U.S. and the U.S.S.R. first began to develop intercontinental ballistic missiles (ICBM's) in 1953, when lightweight thermonuclear warheads became practical. Both countries tested liquid-fueled ICBM's in 1957 and deployed modest numbers of them shortly afterward. Significant deployments began only in the 1960's, when reliable, quick-firing solid propellants and inertial guidance systems became available.

A family of U.S. nuclear missiles with a range of about 1,500 nautical miles was deployed in the U.K., Italy and

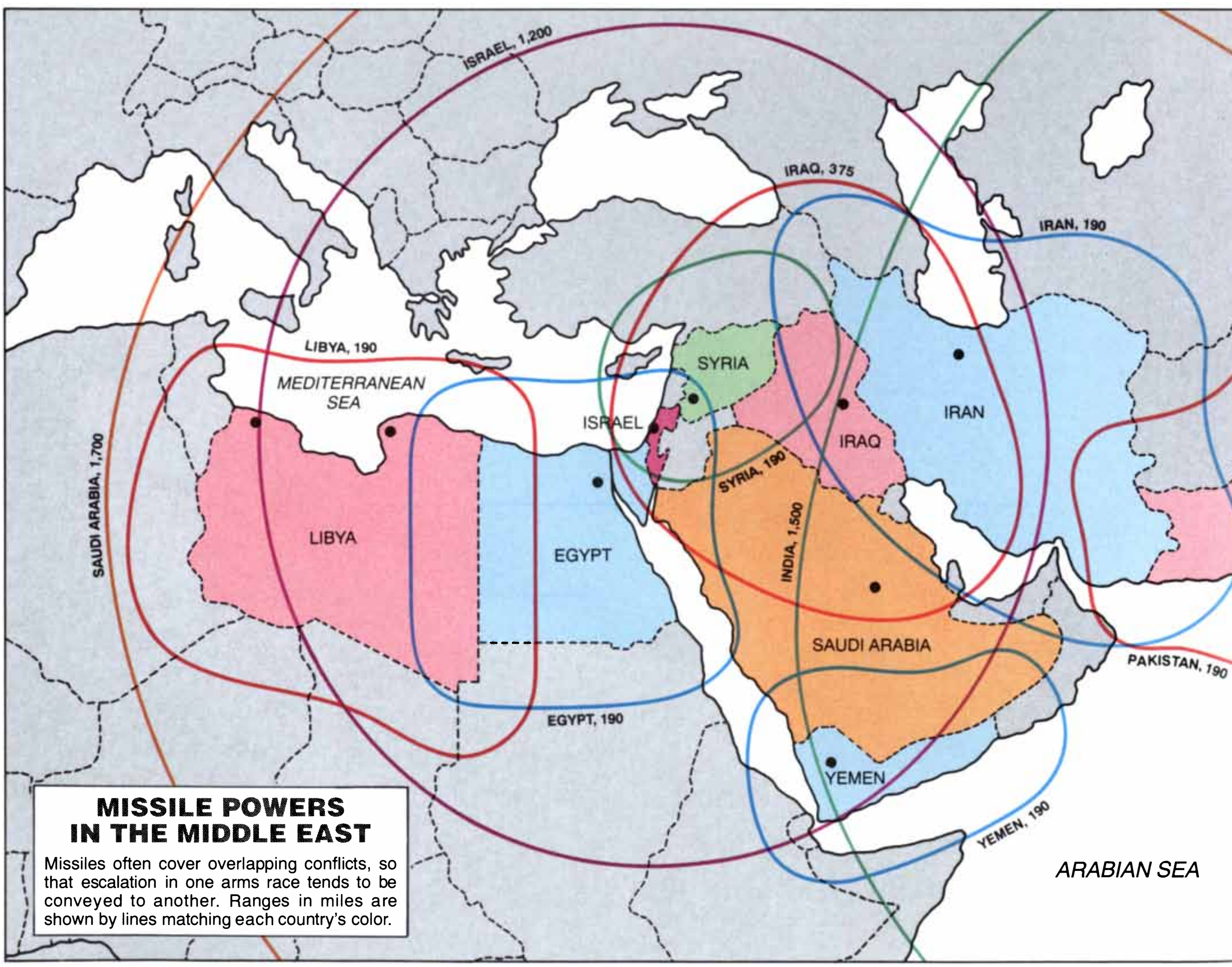
Turkey between 1958 and 1963, then withdrawn after the Cuban Missile Crisis. Such missiles were also deployed on nuclear submarines; later versions attained intercontinental range. The Soviet Union developed a greater variety of intermediate- and long-range nuclear missiles and deployed them in large numbers only during the 1970's. They also developed a broad range of short-range missiles capable of carrying conventional, chemical and nuclear warheads. They provided these to a large number of client states. The U.S. also built short-range missiles but released very few to other countries.

France, rebelling against the U.S. export policy, became the third country to develop strategic ballistic missiles that could deliver nuclear warheads. (The British exploited their special relationship with the U.S. by purchasing its sea-launched ballistic missiles and nuclear submarines, then adding nuclear warheads of their own manufacture.) China became the fourth missile manufacturer, quickly establishing independent nuclear and ballistic missile capabilities after breaking with the Soviet Union in the late 1950's. China is now beginning to export a variety of medium-range ballistic missiles. Six European countries have jointly developed the Ariane space launcher for commercial and scientific applications. Although the Ariane is unsuitable as a strategic weapon, it has technologies that are pertinent to missile development, such as guidance systems, which have since found their way to other countries.

The U.S. has been the most restrained missile exporter among the Big Four. In the 1950's it transferred Honest John missiles having a range of only 20 miles to Greece, Turkey and South Korea. In 1972 it provided Israel with 160 Lance missiles, which are capable of bearing either conventional or nuclear warheads but limited to a range of 70 miles. The U.S. rejected Israel's longstanding request for the Pershing Ia, which can carry a larger payload up to 450 miles.

France has declined to export long-range missiles. In the mid-1960's, however, when it still maintained a tacit alliance with Israel, it allowed Dassault to provide Israel with the technology to build Jericho I missiles. This mobile system has a range of from 300 to 400 miles and a payload of approximately 1,200 pounds.

In March, 1988, China agreed to supply Saudi Arabia with an undisclosed number of its CSS-2 ballistic missiles, putting most of the Middle East and part of the Soviet Union within range of that nation. In its original version this missile carries a one-megaton warhead 2,200 miles, with an accuracy of about a mile. The Saudis maintain, however, that the version they are purchasing is fitted for a 4,500-pound conventional warhead, which it can carry only 1,700 miles. To allay the concerns of the U.S., the U.S.S.R. and Middle East-



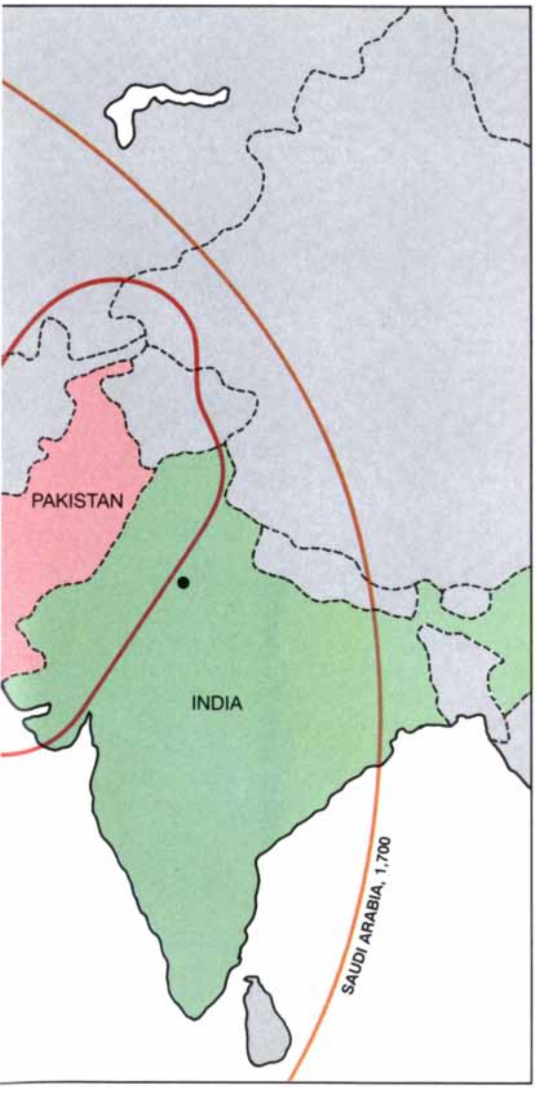
MISSILE POWERS IN THE MIDDLE EAST

Missiles often cover overlapping conflicts, so that escalation in one arms race tends to be conveyed to another. Ranges in miles are shown by lines matching each country's color.

ern countries, the Saudis have agreed to sign the 1970 Nuclear Non-Proliferation Treaty.

The Soviet Union has supplied five short-range missiles, primarily to client states in the Middle East. The earliest Soviet export missiles, the Frog 4 and 5, had ranges of 30 miles; they were shipped to North Korea, Algeria and Egypt beginning in 1957. The next wave, which began in 1965, was by far the most important because it primarily involved the Scud B. This missile has a range of 180 miles and reportedly is accurate to about 1,000 yards. The Scud B has been sold to North Korea, Egypt, Libya, Syria, Yemen, Iraq and Iran. Meanwhile the Soviets also sold the Frog 7—with a range of 40 miles—to many of these same countries and to Kuwait; the SS-21—with a range of 70 miles—has been sold by the Soviets to Syria and Yemen.

Once missiles are in the hands of sovereign nations, they can be modified, as was shown during the Iran-Iraq War. Both belligerents had Scud B missiles, but Baghdad (which is closer to the border) was within range, whereas Teheran was not. By lightening the warhead with help from Western Europe and possibly East Germany, Iraq doubled the range of its missiles to about 375 miles. South Korea has managed a rather more complicated conversion, turning the two-stage, Nike-Hercules antiaircraft missiles into ballistic missiles for use against ground targets. They have a range of as much as 160 miles, enough to counter North Korea's Scud B.



Space launch vehicles can also be adapted to military purposes because they are much alike in their propulsion and guidance requirements. The U.S., France and the Soviet Union have provided such technology to India, Brazil, Argentina and Japan. It should be no surprise that of the four, only Japan is not exporting ballistic missiles.

A pattern by which missile technology reaches hands for which it was not intended can be discerned. First the clients modify missiles to achieve capabilities that the original suppliers would not have countenanced. Next they produce copies of the improved version. Then they design new missiles from scratch, seeking export markets to defray some of the expense. The emergence of such Third World missile exporters has helped create a buyer's market, in which a remarkable range of missile technology is available to anyone with cash. A study conducted at the Massachusetts Institute of Technology in 1981 identified 10 solid-propellant rocket motors that were then available on the open market. Such motors could propel a 1,000-pound payload 600 miles. Two of the motors, arrayed in separate stages, could greatly extend both the payload and the range.

Although advanced ballistic missile guidance systems are apparently not now exported, they are not needed for short-range weapons. That is because accuracy is inversely related to the distance a ballistic missile must travel. A relatively crude guidance system might control the velocity cutoff error (the speed attained when the motor switches off) to an accuracy of one foot per second. Such an error would cause the warhead to miss its target by a mile at a range of 5,000 miles but by only a tenth of a mile at a range of 500 miles.

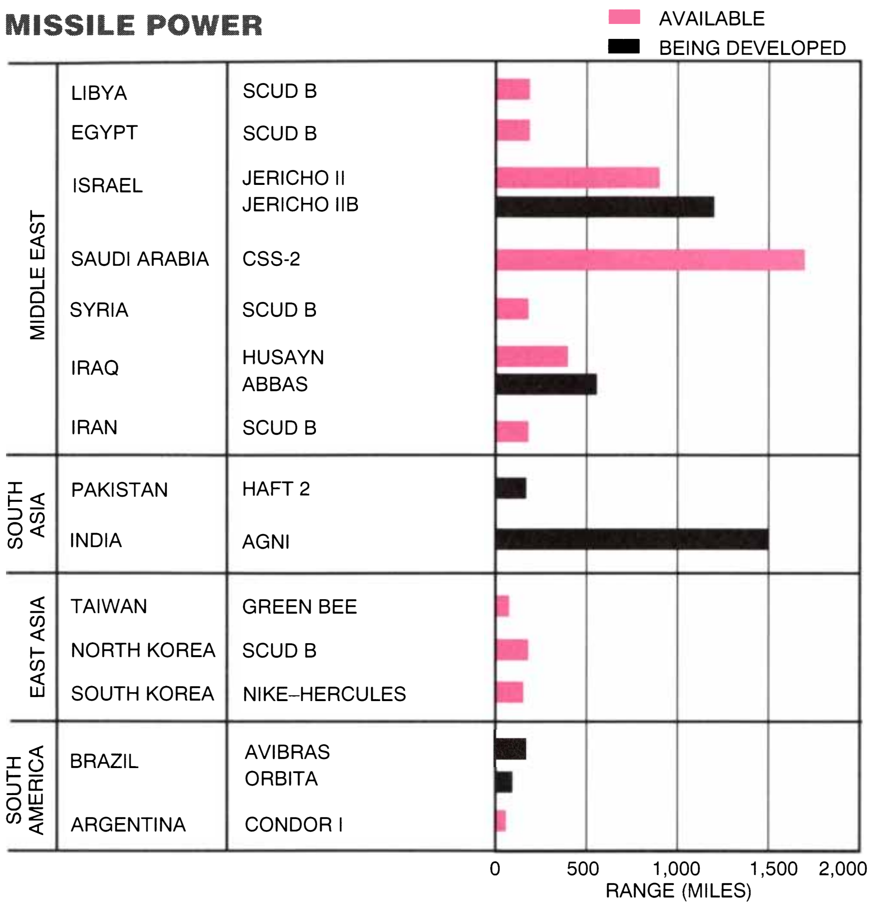
Typically a crude guidance system can be cobbled together from widely available aircraft navigation systems. To be sure, an aircraft's navigation functions differ significantly from those of a ballistic missile: the aircraft need steer only in two dimensions, but the missile requires three. With the proper technical assistance, such problems can be overcome: technical assistance is itself a commodity, and a growing number of engineers and organizations have shown themselves all too willing to be bought.

During the past 30 years these political, economic and technological factors have produced an array of nations armed with missiles that can deliver conventional, nuclear or chemical warheads (the last has been called the poor man's nuke). We offer a status report, based on the open literature and other unclassified materials.

Israel clearly has the most advanced missile program outside the Big Four, having deployed its first model, the Jericho I, in 1973. The original version had sufficient range to hit targets in Egypt and Syria but not in Iraq, Iran, Libya or Saudi Arabia. The Israelis apparently set out to develop their own long-range missile in 1975. The result was the Jericho II, a two-stage, solid-propellant model that was successfully tested in 1987. The same rocket equipped with a third stage (called the IIB) seems to have launched Israel's first satellite in 1988. Analysis of the launch suggests that the Jericho IIB can propel a payload of 2,200 pounds up to 1,500 miles or one of 1,100 pounds up to 2,200 miles.

The Jericho missiles are presumed to be armed with nuclear warheads, a capability that Israel has neither confirmed nor denied. With a range of 900 miles, the Jericho II can reach all of Egypt, Syria, Jordan, Iraq, Iran, the Gulf States and parts of the U.S.S.R. By any standard, the Jericho II represents an intimidating deterrent to Israel's enemies, which have not yet attained its level of nuclear or missile capability. The country may have continued to receive assistance from abroad, and there are indications that it has collaborated with Iran, Taiwan, South Africa and China in missile development.

Egypt seems determined to build long-range missiles of its own design; indeed, it was the first country apart from the Big Four to make the attempt. But that program, ordered in the early 1960's by President Gamal Abdel Nasser and manned partly by German rocket scientists, did not succeed. More recently Egypt collaborated with Argentina and Iraq to produce a solid-fueled, inertially guided missile with a range of 500 to 600 miles but reportedly ended its participation in Septem-



ber of 1989. It has, however, begun to manufacture a rocket with a range of 50 miles and is developing a more advanced version of the Scud B, possibly with North Korea's help. Egypt's older, Soviet-supplied Scud B's can reach Libya and, if posted in the Sinai, much of Israel and Jordan as well. In 1989 the U.S. filed criminal charges against five people, including two colonels in the Egyptian army, for allegedly attempting to acquire U.S. missile technology for Egypt.

Libya is negotiating a large arms deal with Brazil, apparently involving the purchase of missiles having ranges of up to 500 miles. Last year it may have tried to buy the Chinese CSS-2 but apparently without success. There are unconfirmed reports that Libya is developing a missile, the Fatih, having a range of 300 miles.

Iraq's quest for strategic weapons may be the most ambitious in the Arab world. Its program seems motivated not merely by its conflicts with Iran, Israel and Syria but also by a more general drive for primacy in the Middle East. Iraq's diverse missile development force is based largely on modifications of the Scud B. Its Husayn more than doubles the Scud's reach, to 375 miles, and the Abbas, still under development, is expected to fly up to 550 miles. Among the systems nearing the production stage are the Tammuz, with a range of 1,250 miles, the three-stage Abid, which was successfully tested in December of 1989, and a tactical antiballistic missile system called the Faw. Iraq recently signed an agreement with Brazil, which is to train Iraqi missile engineers. The Iraqi secret service is known to be trying to acquire technology for nuclear weapons.

Iran launched a considerable number of short-range missiles against Iraq during their recent war, many of them Scud B's purchased from North Korea or China. Iran is reportedly receiving Chinese, Israeli and North Korean help in building missiles having ranges of 25, 80 and 200 miles.

India startled the world in May of 1989 by launching a two-stage ballistic missile on a flight of 1,500 miles. The missile, called the Agni, was derived from space-launch technology provided by France and the Soviet Union. Ironically, the U.S. was preparing to provide additional space technology on the eve of the launch and may still do so.

The Agni is the result of a development program, begun in 1967, which has produced successively more capable rockets. India has funded a broadly based program to develop propulsion and guidance systems. The country can cast its own solid-propellant rocket motors and is now building several missile test sites. There is no reason to doubt that India can develop missiles of still longer range than the Agni, should it decide to do so.

India, which has refused international supervision of its nuclear program, demonstrated its capability by openly detonating a nuclear device in 1974. The government denies having turned the program to military purposes, but the testing of the Agni makes such assertions incredible.

Pakistan asserts that it has tested two of its own ballistic missiles, the Haft 1 and 2, and says that the Haft 1 can carry a payload of 1,200 pounds up to 62 miles. Press reports indicated that the Haft 2 can loft a 1,100-pound warhead 180 miles. If that is so, the missile must surely have been developed with outside assistance, probably from China. In October of 1989, Pakistan announced plans to develop a missile having a range of 375 miles.

Pakistan denies that it seeks to develop or stockpile nuclear weapons but refuses to allow international controls on its uranium-enrichment program. Moreover, it continues to try to obtain nuclear-related technology through espionage and outright purchase. Its former leader, Zulfikar Ali Bhutto, reportedly once said his countrymen would "eat grass," if necessary, to buy or build a counter to India's nuclear capability. The Carnegie Endowment for International Peace has concluded that Pakistan either has working nuclear warheads or need only turn a few screws in order to obtain them.

South Africa revealed that it tested its own ballistic missile system in July of 1989. It is reported to have established two missile facilities: one on Marion Island, 1,200 miles southeast of Cape Town, the other in the Kalahari Desert. Israel has provided South Africa with technical assistance, possibly in exchange for enriched uranium. South Africa refuses international controls on its nuclear program, although it has never openly tested nuclear weapons. Should South Africa choose to develop such weapons—which it asserts it could do at any time—it would then have great incentive to build missiles, since its aging bomber fleet does not constitute an effective strategic delivery system.

Taiwan developed its ballistic missile, the Green Bee (range: 60 miles), in the 1970's, apparently using Lance missile technology provided by Israel. The country is reportedly developing a 600-mile-range missile called the Sky Horse, and last September it announced plans to launch a satellite into low orbit by 1994. Taiwan has signed the Nuclear Non-Proliferation Treaty, which is certainly most welcome given its growing nuclear power program.

North Korea developed an enhanced version of the Scud B, apparently with Chinese technology, and is now exporting the missile. It is reportedly helping Egypt and possibly Iran to establish facilities to manufacture the Scud B.

Argentina and Brazil have developed missiles mainly for export. Argentina's industry grew out of the space program it founded 30 years ago. In 1980 it developed a single-stage rocket with a range of 60 miles, and in 1982 it began working on a two-stage missile called the Condor II, which was to carry a 1,000-pound warhead to a range of between 500 and 600 miles. The Condor II at first enjoyed the technical assistance of firms in Western Europe and the financial support of Egypt and Iraq. But the 1987 Missile Technology Control Regime has helped slow the transfer of technology, raising the projected cost of the missile to an estimated $3.2 billion, a high price to pay for a production run of only 400 missiles. The two Arab states ended their participation in late 1989, and the Argentines suspended the project recently.

Brazil's space program has developed four generations of sounding rockets with assistance from Europe, Canada and the U.S. Two firms, Orbita and Avibras, are now converting the technology embodied in these rockets to military applications.

China is providing Brazil with missile technologies, probably including guidance systems. Orbita is developing an inertially guided missile capable of lifting a payload of 1,100 pounds up to 90 miles. The Avibras missile will have twice the payload and range. Libya and Iraq are reportedly interested in buying it. Avibras also is reportedly developing a missile that can fly 740 miles.

This comparative list of national strategic power should be regarded as no more than a rough index by which one may rank the military and political significance of each country. Yet in considering these armories, it is important to remember that strategic weapons cannot be defined in absolute terms. What matters is the weapons' ability to reach beyond the front lines to threaten an enemy's depots, factories and cities. Missiles that the superpowers relegate to battlefield targets could easily devastate the capitals of small countries in the Levant or the Persian Gulf, as the Iran-Iraq War demonstrated.

Another lesson of that bloody conflict was the efficacy of chemical weapons. Iraq used mustard gas against massed Iranian troops with deadly effect and apparently has now acquired a variety of more advanced chemical weapons. Libya drove this lesson home with its widely publicized efforts to acquire chemical-warfare technology.

Of course, the superpowers never quite forgot about chemical weapons. The Soviets designed the Scud B to carry chemical as well as nuclear and high-explosive warheads. The following countries are known to be capable of producing chemical weapons: Iran, Iraq, Libya, Syria, China, the Koreas, Taiwan and possibly Israel. Egypt may also be developing such a capacity. The first four have particularly strong incentives to equip their Scud B missiles with chemical warheads, because they are all threatened by the Israeli nuclear capability. It is reported that Syria has already done so.

How effective would existing Third World ballistic missiles be in war? Even without nuclear warheads they can be surprisingly destructive. A Scud B releasing 1,200 pounds of the chemical VX agent 4,000 feet above an airfield will kill half of the people in a strip .3 mile wide and 2.5 miles long, oriented in the direction of the wind. That is enough to disable an airfield and certainly enough to devastate a city. An equal weight of high explosives can destroy buildings more than 150 feet away and kill people more than 400 feet away. The 4,500-pound warhead planned for the Saudi missiles would more than double this lethal range. Cluster bombs further enhance these effects by detonating many submunitions over a large area, thus producing a blanket rather than a point explosion.

Accuracy is as important as destructive power. It is generally measured by the "circle error probable" (CEP), an ellipse whose major and minor axes describe the maximum misses along and across a missile's trajectory. The most primitive missile guidance programs, which use inertial instruments taken from aircraft navigation systems, would typically have velocity cutoff errors of from two to four feet per second at burnout. That yields a CEP of approximately 1,000 feet at a range of 200 miles.

Much more accurate guidance technology is available. Twenty years ago an unclassified report described inertial guidance systems that could control velocity cutoff errors to .4 foot per second; that technology has surely diffused to many Third World countries by now. If so, it would allow them to achieve a CEP of 200 feet at a range of 500 miles and a CEP of 120 feet at 200 miles. That is close enough for high-explosive warheads to destroy most point targets and for cluster bombs or poison gas to wipe out troop concentrations and cities.

There is also a psychological dimension to the military effectiveness of ballistic missiles: they sap the will to resist. The thousand V-2 rockets that Germany launched against Great Britain late in World War II were not partic-

MAJOR REGIONAL TENSIONS

REGION	COUNTRY	POTENTIAL TARGETS	WARHEAD CAPABILITIES: HIGH EXPLOSIVE	CLUSTER	CHEMICAL	NUCLEAR
MIDDLE EAST	LIBYA	TEL AVIV, CAIRO	•		•	
	EGYPT	BENGHAZI, TEL AVIV	•		•	
	ISRAEL	BENGHAZI, CAIRO, DAMASCUS, BAGHDAD, RIYADH, TEHERAN	•	•	•	•
	SAUDI ARABIA	TEHERAN, TEL AVIV	•			
	SYRIA	BAGHDAD, TEL AVIV	•		•	
	IRAQ	TEHERAN, TEL AVIV	•		•	
	IRAN	BAGHDAD, TEL AVIV	•			
SOUTH ASIA	PAKISTAN	NEW DELHI	•		•	•
	INDIA	KARACHI	•		•	•
EAST ASIA	PEOPLE'S REPUBLIC OF CHINA	TAIPEI	•		•	•
	TAIWAN	BEIJING	•		•	
	NORTH KOREA	SEOUL	•		•	
	SOUTH KOREA	PYONGYANG	•		•	

ularly accurate. But their 1,000-pound warheads wreaked as much terror among the British as the German Air Force had in the more extensive bombing raids of several years before. A thousand missiles also were launched during the Iran-Iraq War—the only other time such weapons have ever been employed—with similar effect on morale. Iraq killed only a fraction as many Iranians in its missile attacks on Teheran as it had in repelling Iran's human-wave assaults at the front, but the missiles are thought to have played a decisive role in forcing the Iranian leadership to accept an armistice in 1988.

For all these reasons, one must conclude that missiles do offer advantages in the waging of war. That does not mean, however, that the race to acquire missiles is a coolly calculated game of nations. An irrational, destabilizing element is at work as well. The very attempt to acquire weapons of mass destruction could provoke preemptive attack from the enemies one would deter. Moreover, a country that diverts scarce resources to missile programs may starve its forces of other weapons, munitions, fortifications and training, thereby reducing military readiness.

The transition of new states into the missile age will have political repercussions before their missile forces achieve commensurate military significance. At least 18 nations evidently consider that the intangible values of missiles outweigh their cost and danger. Missiles are symbols of technical prowess and political prestige, serving today much as battleships did 80 years ago—as the sine qua non of great-power status.

What are the prospects for controlling this troubling situation? It is hard to be optimistic. We believe that as long as ballistic missiles are the touchstone of national sovereignty in the Third World, it will be difficult to reverse their proliferation.

The major powers can still retain part of their accustomed technical superiority by agreeing quickly to stem the flow of still more advanced military technologies. But more than anything else, they must learn to adapt to the situation as they find it.

U.S. policy has focused on preventing missile proliferation rather than containing its most adverse consequences. The assumptions underlying this policy have obviously changed in the past decade. We are in need of new approaches. As a first step, the U.S. must recognize that it is no longer so easy to intervene in regional conflicts. It must therefore work harder than ever to forge a consensus by which regional conflicts can be managed, if not settled. Such a consensus might lead to the establishment of institutions for building confidence and security, including sharing of intelligence, reciprocal visits to defense plants and space-launch facilities and prior notification of missile tests. These so-called confidence and security building measures should ease suspicions.

Such instruments of peaceful intent are mere signs of political will and can be violated at any time. Nevertheless, they may help lay foundations for a more lasting peace. Examples include the recent agreement between India and Pakistan not to attack each other's nuclear facilities, the on-site visits and exchange of information between Argentina and Brazil concerning their nuclear capabilities (which are not under international control), and the informal discussions on missile proliferation the U.S. has had with various Middle Eastern countries.

More ambitious mechanisms involving international safeguards and monitoring cannot be imposed, but they might be actively sought if they constituted tokens of membership in a sophisticated club. Such a club might provide a launch facility, or "common carrier," for legitimate space missions. It might provide stabilizing satellite reconnaissance data to all parties.

Realism need not be a pretext for fatalism. The proliferation of ballistic missiles cannot be reversed, but some of its greatest dangers can be mitigated. Above all, the U.S. should recognize the waning of its influence on the global arms race so that it can wisely wield the influence it retains. The stakes have never been higher.

FURTHER READING

BALLISTIC MISSILE GUIDANCE. David Hoag. Symposium on New Technology and Arms Race. The Draper Laboratory, 1970.

ASSESSING THE COMPARABILITY OF DUAL-USE TECHNOLOGIES FOR BALLISTIC MISSILE DEVELOPMENT. Mark Balaschak et al. The MIT Center for International Studies, June, 1981.

THE UNDECLARED BOMB: THE SPREAD OF NUCLEAR WEAPONS, 1987-1988. Leonard S. Spector. Ballinger Publishing Co., 1988.

MISSILE PROLIFERATION: SURVEY OF EMERGING MISSILE FORCES. Congressional Research Service Foreign Affairs and National Defense Division. CRS Report, No. 88-642 F. Revised February 9, 1989.

Ford Taurus SHO Think of it as the official pace car of the *Fortune 500.*

The 24-valve, 220 horsepower Ford Taurus SHO.

With the coming of Ford Taurus SHO, the road to success has become much more exhilarating to drive.

Consider, for example, its fuel-injected Super High Output (SHO) engine, a 24-valve, 220 horsepower V-6. Engineered for thoroughbred performance and ably assisted by aerodynamic ground effects, anti-lock brakes, and a special sport suspension, it allows you to manage traffic as effectively as you manage your business.

The Taurus SHO's specially-designed interior is also dressed for success. Surrounding you with appointments like fully-articulated sport seats (leather trim optional), and new performance instrumentation, it offers the proper seating for those of your standing.

In short, those in positions of power should find themselves right at home in the driver's seat of a Ford Taurus SHO. It could just be one of the easiest executive decisions you'll ever make.

Taurus SHO for 1990 comes equipped with a driver air bag supplemental restraint system.

Buckle up—together we can save lives.

Have you driven a Ford...lately? Ford

Superconductors beyond 1–2–3

Yttrium1-barium2-copper3 oxide is only one of many new high-temperature superconductors. All have planes of copper and oxygen atoms, providing the perfect highways for electrons

by Robert J. Cava

More than a decade ago Bernard Raveau and his collaborators at the University of Caen in France began synthesizing a series of compounds composed of lanthanum, barium, copper and oxygen. If some of these ceramic materials had been cooled to a temperature near 40 kelvins, workers might have discovered that the ceramics offered no resistance to the flow of electric current. Yet few physicists, chemists or materials scientists had reason to suspect that the materials were superconductors, and fewer still would claim that superconductors with such high transition temperatures would ever be made. The ceramics languished in relative obscurity for several years. Then in 1986 physicists K. Alex Müller and J. Georg Bednorz of the IBM Zürich Research Laboratory realized that these ceramics were the key to developing a new class of high-temperature superconductors that would revitalize superconductivity research and technology. Their discovery began one of the most startling revolutions in solid-state physics.

Spurred on by the possibility of finding the ultimate high-temperature superconductor, nearly everyone who had a periodic table of the elements and a furnace in 1987 began furiously fabricating exotic compounds based on copper and oxygen. Nature, however, is very picky about which copper oxides are allowed to form. Although the theory of solid-state chemistry guides the choice of elements that might be combined to make new compounds, the materials are so complex that no theory can yet reliably predict the structure or behavior of the new materials. The trick is to employ chemistry, intuition and luck to find just the right combination of elements to drive the transition temperature even higher.

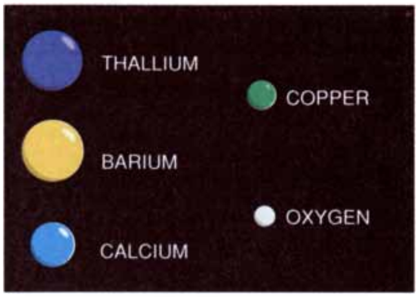
In February, 1987, Maw-Kuen Wu of the University of Alabama, Ching-Wu (Paul) Chu of the University of Houston and their co-workers replaced lanthanum in the Bednorz-Müller compound with yttrium—a smaller rare-earth element—to form a superconductor whose transition temperature was 90 kelvins. Soon after, R. Bruce van Dover, Bertram Batlogg and I at AT&T Bell Laboratories were the first to determine that the superconductor was the chemical compound now known as 1-2-3 because of the ratio of yttrium to barium to copper atoms. The 1-2-3 material was the first superconductor with a transition temperature above 77 kelvins and therefore could be cooled easily and inexpensively in a bath of liquid nitrogen. The best conventional metallic superconductors, such as alloys of niobium and tin, required the more costly liquid helium as a coolant, as did the first high-temperature superconductor discovered by Bednorz and Müller [see "The New Superconductors: Prospects for Applications," by Alan M. Wolsky, Robert F. Giese and Edward J. Daniels; SCIENTIFIC AMERICAN, February, 1989].

During the past four years the efforts of thousands of investigators around the world have produced about a dozen superconductors that have transition temperatures above 40 kelvins as well as a handful above 77 kelvins. Almost all high-temperature superconductors were found by workers who were trying out one idea but serendipitously created something else, outsmarted (or rather helped) by nature. So far the highest transition temperature—which was achieved in a compound composed of thallium, barium, calcium, copper and oxygen—is an astounding 125 kelvins.

Although investigators do not know the limits of superconductivity or even understand the fundamental interactions that give rise to it in ceramic superconductors, we can identify the molecular features that enhance or inhibit superconductivity. Most important, the crystal lattices of all high-temperature superconductors contain planes of copper and oxygen atoms sandwiched between layers of other elements. When a high-temperature superconductor is cooled below its transition temperature, the copper-oxygen planes provide the perfect highways for electrons to travel. The other elements in the crystal lattice can be chosen and arranged to increase or decrease the transition temperature to the superconducting state.

The materials discovered recently by Bednorz, Müller and others behave in some ways like a conventional metallic superconductor. When a voltage is applied to one of the materials at room temperature, electrons be-

ROBERT J. CAVA, who is a materials scientist at AT&T Bell Laboratories, has been a leader in the search for high-temperature superconductors. In 1978 he earned his Ph.D. in ceramics from the Massachusetts Institute of Technology. For the past 10 years he has worked in the department of solid-state chemistry at AT&T Bell Laboratories. He is an avid amateur astronomer and a lover of Italian opera.

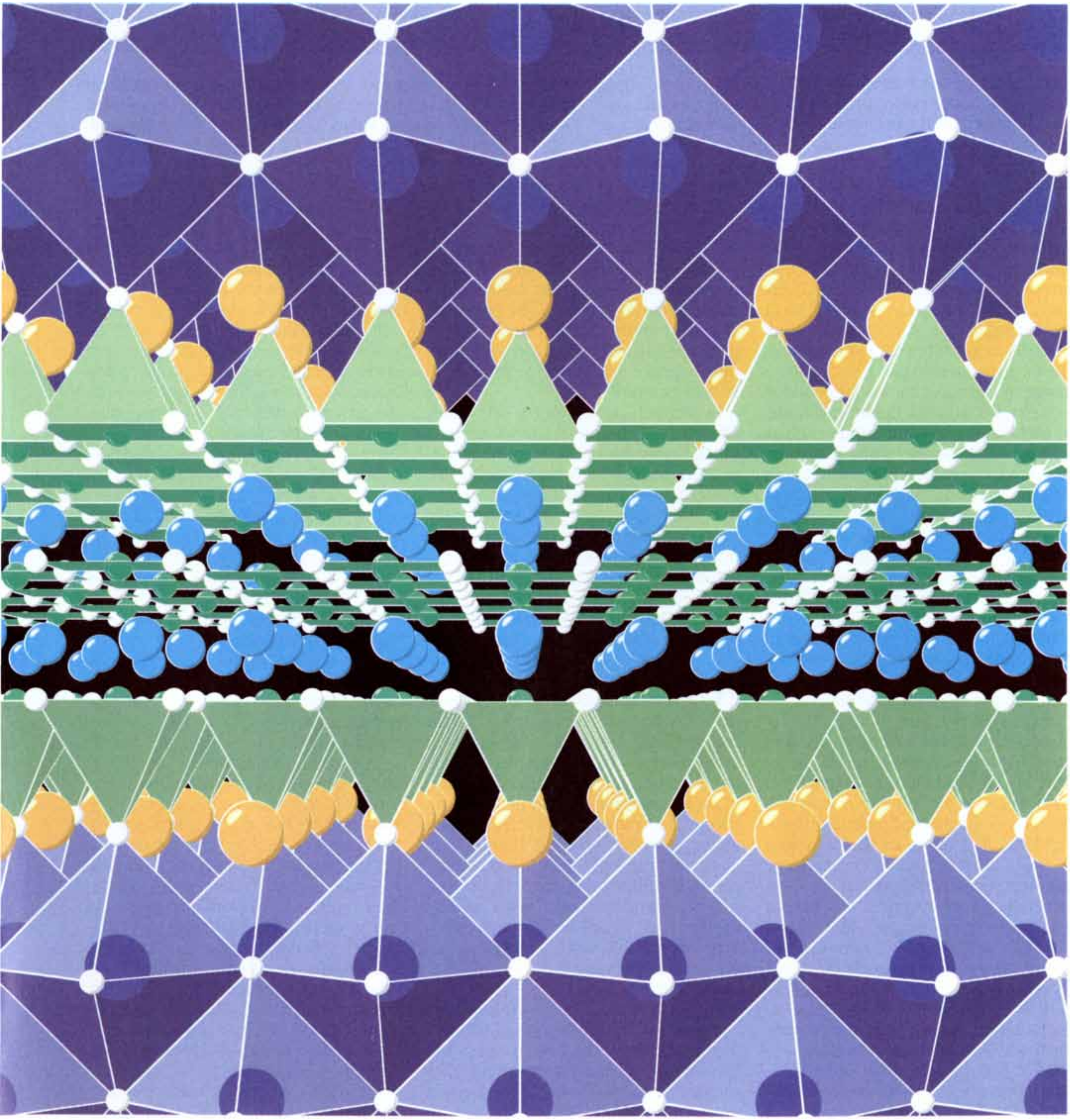


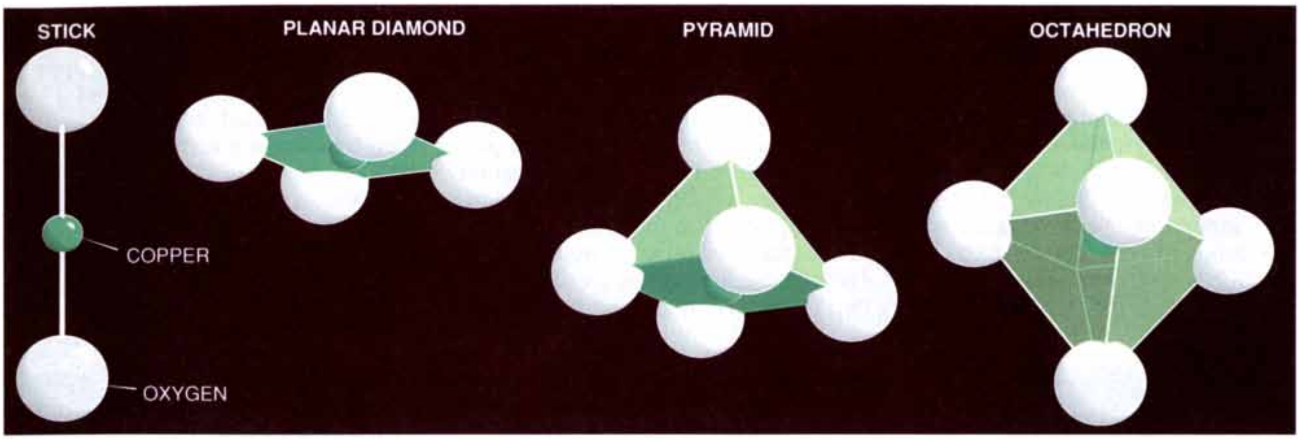
SUPERCONDUCTOR composed of thallium, barium, calcium, copper and oxygen ($Tl_2Ba_2Ca_2Cu_3O_{10}$) exhibits the highest transition temperature discovered thus far. It offers no resistance to the flow of electricity when it is cooled below 125 kelvins. Like other ceramic superconductors, the thallium compound contains planes of copper and oxygen atoms that allow conduction of electrons through the material.

gin to move through it. Because the electrons scatter off the atoms that make up the compound, they lose energy, a process that generates electric resistance. As the material is cooled to low temperatures, the resistance and energy loss decrease. When the temperature drops just below the transition temperature, the resistance suddenly drops to zero. The electrons undergo a phase transition and can no longer be scattered by the atoms.

The Bednorz-Müller compounds are among a relatively small number of ceramics that are good conductors. (After all, ceramics, such as the dinner plates and commodes in our homes, are known for their insulating properties.) The metals lanthanum, barium and copper react readily and individually with oxygen to form lanthanum oxide, barium oxide and copper oxide. These three substances are insulators, that is, they conduct electricity poorly at room temperature. When Bednorz and Müller combined lanthanum, barium, copper and oxygen in the appropriate proportions, they created a ceramic that conducts electricity well at room temperature and superconducts at 28 kelvins.

Whether a ceramic is an insulator or a conductor depends on how the electrons, the carriers of negative charge, behave within the compound. Barium oxide (BaO), for example, is an insu-





ELECTRONIC HEART of nearly all high-temperature superconductors is the chemical bond between copper and oxygen. A copper atom can combine with oxygen atoms in four different ways, which can be represented as a coordination polyhedron: a geometric figure that encloses the copper atom and the nearest oxygen atoms to which it is directly bonded.

lator because the electrons are confined to the regions around the nuclei of the barium and oxygen atoms. The electrons are localized because of the charge and energy requirements of barium and oxygen. The barium nucleus has a net positive charge of 56, so in the neutral, atomic state it is surrounded by 56 negatively charged electrons. The oxygen nucleus has a net positive charge of eight and thus is surrounded by eight electrons in the atomic state. To form chemical compounds such as barium oxide, the electrons distribute themselves between the barium and oxygen atoms to achieve the lowest possible energy state.

Electrons are distributed around the atoms in orbitals. One can think of these orbitals as concentric shells that contain a certain number of electrons. Electrons in the inner shells are tightly bound to the nucleus. Electrons in the outermost shell are more easily removed and hence play a greater role in chemical and electrical phenomena.

Oxygen is the glue that holds ceramic compounds together. An oxygen atom has six electrons in its outer shell, which can hold up to eight electrons. Because a filled outer shell is a much more stable configuration than a partially filled outer shell, an oxygen atom can achieve a much lower energy state if it gains two electrons. An oxygen atom therefore prefers to have a negative charge of two; in other words, it has a "valence" of 2–. A barium atom has two electrons in its outer shell, but it can achieve a much lower energy state if it loses two electrons. The valence of barium then is 2+. To provide the lowest-energy configuration of electrons in BaO, each barium nucleus donates two electrons to each oxygen atom. Because significant energy is then required to add or subtract an electron from the filled orbital, the electrons are strongly localized and are not available to carry electric current.

This exchange of electrons also forms the basis of the bond that holds the ceramic material together. The positively charged barium atoms (valence 2+) attract the negatively charged oxygen atoms (valence 2–). This type of bond is known as ionic.

In virtually all ceramics, the outer electrons of the metal atoms are in a much higher energy state than the outer electrons of the oxygen atoms. To achieve a low-energy, stable configuration, therefore, metal atoms typically shed one or more electrons from their outer shells. These electrons are captured by the oxygen atoms in the ceramic. Because electrons remain localized around the metal and oxygen atoms, the ceramic is an insulator.

Many copper oxide ceramics are good conductors because some of the electrons within the material are free to move from atom to atom. In this type of ceramic the electrons are not localized because of an unusual interaction between the copper and oxygen. In materials such as Cu_2O, copper readily donates one electron to oxygen, which leaves copper with a filled outer shell. Hence, the two copper atoms (each with valence 1+) form an ionic bond with the oxygen (valence 2–). Because the electrons are localized around the atoms, the ceramic is an insulator.

For oxides such as CuO, however, copper does not give up a second electron as easily as the first, because the loss of the second electron creates a vacancy in the outer shell. The oxygen must "fight" for a second electron to complete its outer orbital. It turns out that the oxygen "wins" because a filled oxygen orbital is a slightly more stable configuration than a filled outer copper orbital. If copper and oxygen are mixed with other elements in a crystal lattice, however, the delicate energy balance can be shifted so that copper and oxygen share electrons to complete their outer shells. (This sharing of electrons forms a "covalent" bond between the oxygen and copper.) Because the electrons are free to move between the copper and oxygen atoms, materials containing copper, oxygen and other elements can be good conductors.

Copper can donate one, two or three electrons in a chemical bond with oxygen. The most stable of the three valence states is 2+, and hence the 1+ state (found, for example, in Cu_2O) is called "reduced" copper and the 3+ state (found, for instance, in sodium copper oxide, $NaCuO_2$) is known as "oxidized" copper.

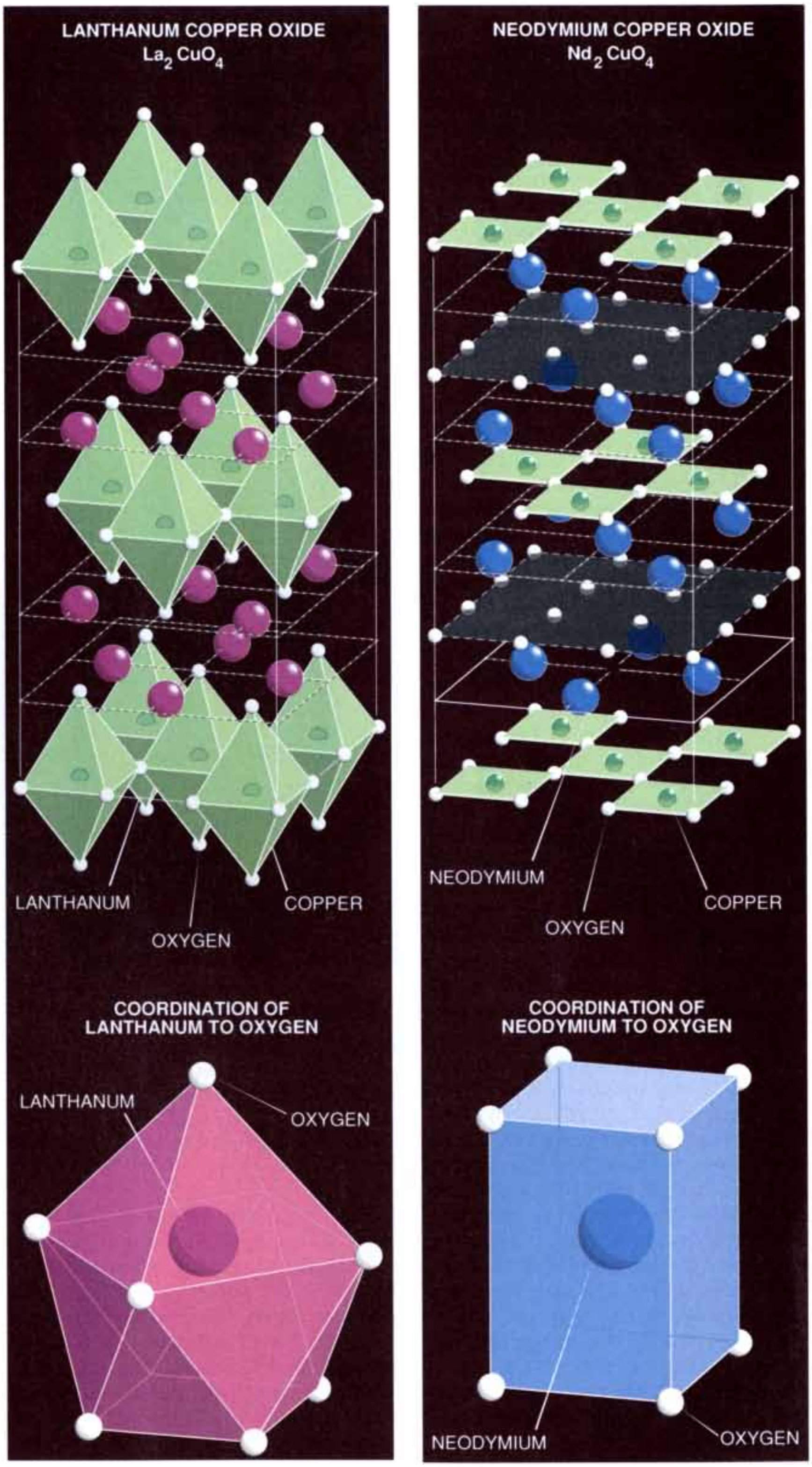
In ceramic superconductors the copper appears to act as if it has a fractional valence. The valence of copper depends on the influence of oxygen and other atoms in the superconductor. When the valence of copper is 2+, the electrons are localized in a copper-oxygen bond. In some cases, when additional oxidizing atoms such as lanthanum and barium are present in the compound, more than two electrons can be pulled away from some of the copper atoms, driving the valence toward 3+. In other cases, "reducing" atoms can add electrons to some of the copper atoms, driving the valence from 2+ toward 1+. In both cases the electrons are no longer localized and can participate in electric conduction.

Although this kind of valence counting is essential in understanding ceramic superconductors, it is a simplification. In superconducting ceramics the oxygen and copper atoms share electrons to meet their energy requirements. These shared electrons move from copper atom to oxygen atom, forming a "sea" of electrons or, more specifically, a conduction band. If the copper atoms are reduced to valences of less than 2+ —that is, if they donate fewer than two electrons on average—then only a few electrons move about the conduction band. If the copper atoms are oxidized to valences greater than 2+, donating more than two electrons on average, the conduction band swarms with electrons, leaving regions of positive charge known as holes. The concept of fractional valences makes sense only because electrons are actually being removed from or added to a conduction band.

Copper is not the only metal that can achieve fractional valences in the presence of oxygen. Bismuth and lead have electron orbitals that, like copper, have energy requirements similar to those of oxygen. Under the proper circumstances both metals will form covalent bonds with oxygen and will allow electrons to move freely within a conduction band. Workers have discovered both lead oxides and bismuth oxides that superconduct at relatively high temperatures. Whether these compounds are superconducting for the same reason that copper oxides are superconducting is still a matter of considerable debate.

The family of superconductors discovered by Bednorz and Müller is based on the chemical modification of the compound lanthanum copper oxide (La_2CuO_4), whose crystal structure is shown in the illustration at the right. The structures of ceramic superconductors are typically described in terms of their coordination polyhedra, one of the cornerstone concepts of solid-state chemistry. A coordination polyhedron is a geometric figure that encloses a metal atom and the nearest oxygen atoms to which it is directly bonded. The polyhedra reveal the number of oxygen atoms that an ion prefers in its immediate bonding environment. The number of preferred oxygen neighbors can vary, for instance, from two neighbors for copper (valence 1+) to 15 neighbors for cesium (valence 1+). The actual number of oxygen atoms to which an ion is coordinated depends both on the ion's size and the particular requirements of the other metal atoms in the oxide.

In La_2CuO_4 the copper atoms are co-



CRYSTAL STRUCTURE of lanthanum copper oxide (La_2CuO_4) (*top left*) is the basis for many superconductors. If some of the lanthanum atoms are replaced by either calcium, strontium or barium atoms, a superconductor is formed. The coordination polyhedron for oxygen and lanthanum is shown (*bottom left*). Superconductors based on La_2CuO_4 are *p*-type conductors because they support positive-charge carriers known as holes. *N*-type superconductors can be derived from neodymium copper oxide (Nd_2CuO_4) (*top right*) if cerium or thorium atoms substitute for some of the neodymium atoms. Superconductors based on Nd_2CuO_4 are *n*-type conductors because they support negative-charge carriers called electrons. A neodymium atom is coordinated to eight oxygen atoms (*bottom right*).

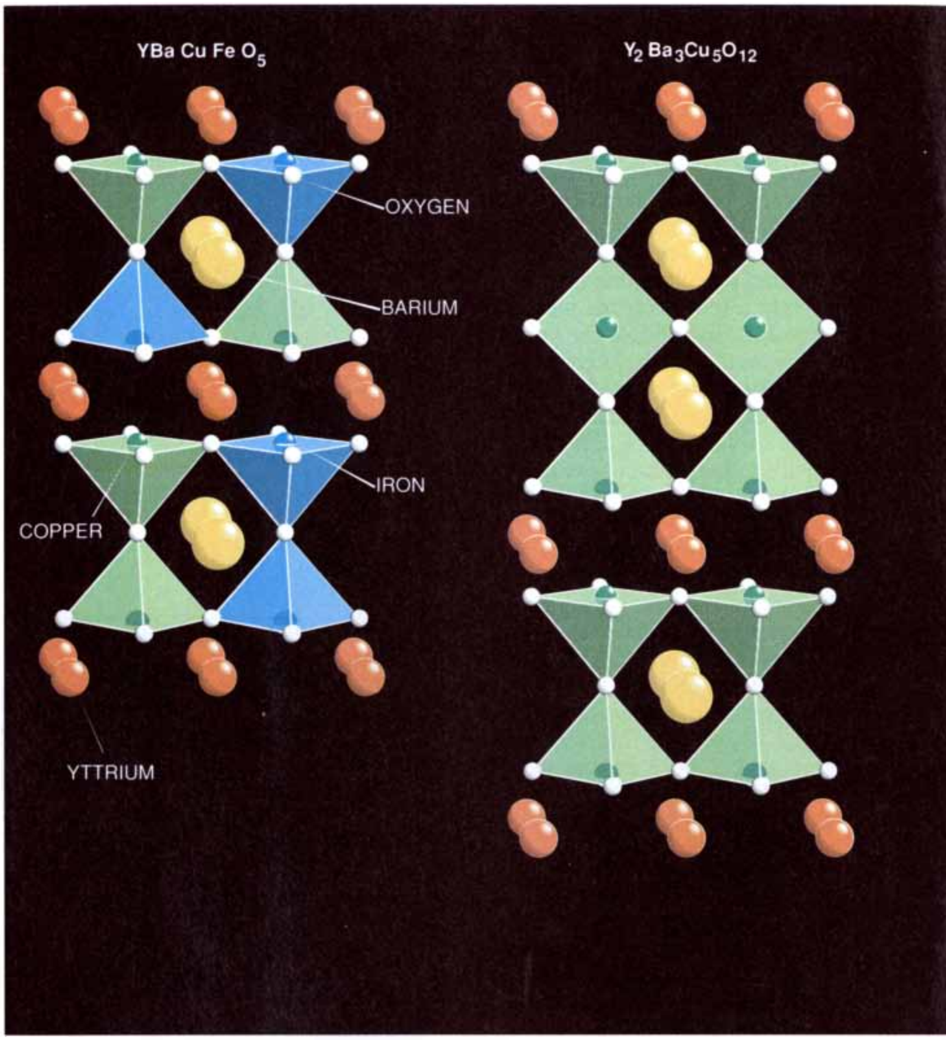
ordinated to six oxygen atoms, which sit at the corners of an elongated octahedron. Because of the energetics of copper's outer shell in its valence 2+ state, oxygen atoms at two opposite corners of the octahedron will always be farther from the copper than oxygen atoms at the four other corners. This structural distortion, known as the Jahn-Teller effect, is what first led Bednorz and Müller to focus on copper oxides as possible superconductors. The effect suggests that the electrons interact strongly with the positions of copper and oxygen atoms in the crystal lattice, an interaction considered very important for the occurrence of superconductivity.

The copper-oxygen octahedra in La_2CuO_4 are joined to one another at the four corner oxygen atoms closest to the copper. The copper atoms and their closest oxygen neighbors all lie in a plane. This copper-oxygen plane is the electronic heart of all the known copper oxide superconductors. It is the microscopic region of the crystal structure where the superconducting charge carriers originate.

The crystal structure of La_2CuO_4 is remarkable because the coordination polyhedra of lanthanum and copper combine in a way that allows three-dimensional space to be filled in a highly two-dimensional manner. As a result, the planes of copper and oxygen interspace the inert double layer of lanthanum and oxygen.

Lanthanum copper oxide itself is not a superconductor because of an interaction among the electrons in outer shells of neighboring copper atoms. Each copper donates two electrons to the oxygen and retains nine electrons in its outer shell. Each of the nine electrons has a magnetic moment, that is, the electrons behave as if they were tiny bar magnets with north and south poles. The energy arising from this magnetic interaction is minimized when eight of these nine electrons become paired: their magnetic moments (the poles of the tiny magnets) cancel when they are aligned in opposite directions. When the remaining unpaired electrons on neighboring copper atoms interact with one another and align, their magnetic moments point in opposite directions. This type of alignment, known as antiferromagnetism, pins the electrons to the crystal lattice, eliminating the possibility of superconductivity and indeed any conductivity at all.

Superconductivity appears only if antiferromagnetism is fully destroyed. For this reason, many investigators believe that the origins of magnetism and high-temperature superconductivity are intimately linked. A few have implied that the connection is sinister, calling magnetism and superconductivity the "Jekyll and Hyde" of high-temperature superconductivity.

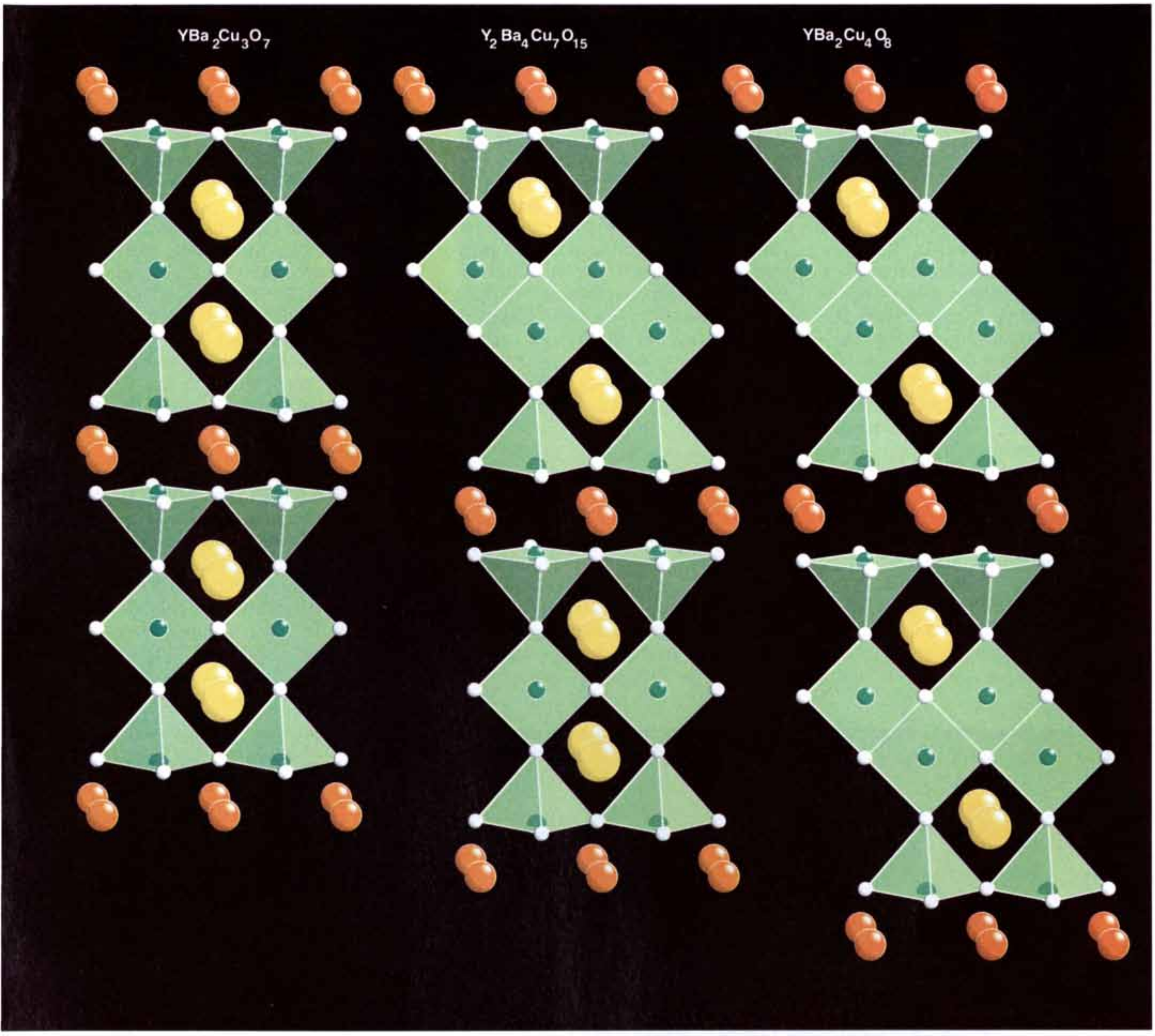


1–2–3 COMPOUND ($YBa_2Cu_3O_7$) is only one member of what chemists call a homologous series. Each member of the series can be derived from another by either the addition or subtraction of simple structural components. Three members of the series—$YBa_2Cu_3O_7$, $Y_2Ba_4Cu_7O_{15}$ and $YBa_2Cu_4O_8$—are superconductors whose transition temperatures exceed 77 kelvins. The compounds $YBaCu_2O_5$ (not shown) and $Y_2Ba_3Cu_5O_{12}$ are not yet known to exist. The structural features of $YBaCu_2O_5$ are predicted to be similar to $YBaCuFeO_5$, which is not a superconductor.

The crystal structure of La_2CuO_4 can be modified, however, to eliminate antiferromagnetism and to introduce superconductivity. Bednorz and Müller created their famous high-temperature superconductor by replacing some of the lanthanum atoms with barium atoms. The formula for the compound is $La_{2-x}Ba_xCuO_4$.

Charge neutrality is maintained in $La_{2-x}Ba_xCuO_4$ if one copper atom is oxidized from 2+ to 3+ for every barium atom substituted for a lanthanum atom. The extra electron donated by copper is not localized but instead goes into the conduction band. When the copper atoms reach a critical valence, near 2.2+, antiferromagnetism disappears and superconductivity appears.

In $La_{2-x}Ba_xCuO_4$ copper is coordinated to six oxygen atoms that form the corners of an elongated octahedron. Because the coordination number increases roughly with the size of the atom, the lanthanum atoms, which are larger than copper atoms, are coordinated to nine oxygen atoms. The barium atoms, which are similar in size to the lanthanum atoms, are also coordinated to nine oxygen atoms. Within the crystal lattice, therefore, barium atoms can occupy the same sites as the lanthanum atoms, and, in fact, they are distributed at random through the crystal lattice. Such a dis-



tribution of atoms in fixed crystallographic sites is known as a solid solution. (In contrast, a liquid solution consists of a random distribution of ions in random positions.)

The idea of a solid solution is central to the occurrence of superconductivity in copper oxides. As a rule, solid solutions do not form unless the sizes of the ions are relatively close and the types of chemical bonds formed are similar. Thus, barium, strontium or calcium can form solid solutions in La_2CuO_4 by replacing some of the lanthanum. When strontium replaces one of nine lanthanum atoms in a crystal of La_2CuO_4, the result is $La_{1.8}Sr_{.2}CuO_4$, where the copper valence is 2.2+. The superconducting transition temperature of this compound is 40 kelvins, the highest among solid solutions based on La_2CuO_4.

Unlike the solid solutions that make La_2CuO_4 superconducting, the superconductor $YBa_2Cu_3O_7$ (the 1-2-3 compound) has a fully ordered crystal lattice consisting of planes of yttrium, barium and copper ions [*see illustration above*]. There is no mixing of the metal atoms among the sites in the crystal lattice. The small yttrium 3+ ions are always bonded to eight oxygen atoms; the large barium 2+ ions to 10 oxygen atoms. In between the barium and the yttrium layers, copper atoms are coordinated with oxygen in pyramids. The bases of the copper-oxygen pyramids face one another across a plane of yttrium atoms. The bases form the copper-oxygen plane necessary for superconductivity. In between two consecutive layers of barium, copper atoms are coordinated with four oxygen atoms in a flat, diamondlike shape. The corners of the diamonds are connected to form a chain of diamonds.

The peculiarities of copper-oxygen bonding have conspired to fill three-dimensional space with a combination of one-dimensional chains and two-dimensional pyramidal planes. An Egyptian scientist once told me that he thought his fellow citizens should be very good at finding new superconductors because the magic of pyramids is such an ancient part of their culture.

In $YBa_2Cu_3O_7$ the copper has an average valence of 2.33+ (seven of the 14 electrons needed by oxygen are provided by one yttrium atom and two barium atoms; the remaining seven electrons are donated by three copper atoms). Workers have demonstrated that the copper atoms in both the chains and the bases of the pyramids have been oxidized to approximately the same valences.

In the compound $YBa_2Cu_3O_7$, the seven oxygen atoms are particularly crucial for superconductivity. When the oxygen content is reduced from seven to six, the insulator $YBa_2Cu_3O_6$ is formed. Oxygen is removed from one crystallographic site only, thereby transforming the copper-oxygen diamonds into sticks. The sticks are a stable coordination geometry for copper in the 1+ valence state. The decrease in oxygen content does not directly affect the copper-oxygen pyramids, but, as counting the charges reveals, the copper atoms associated with the pyramids are now in the 2+ valence state. In this configuration the electrons are localized by the effect of antiferromagnetism.

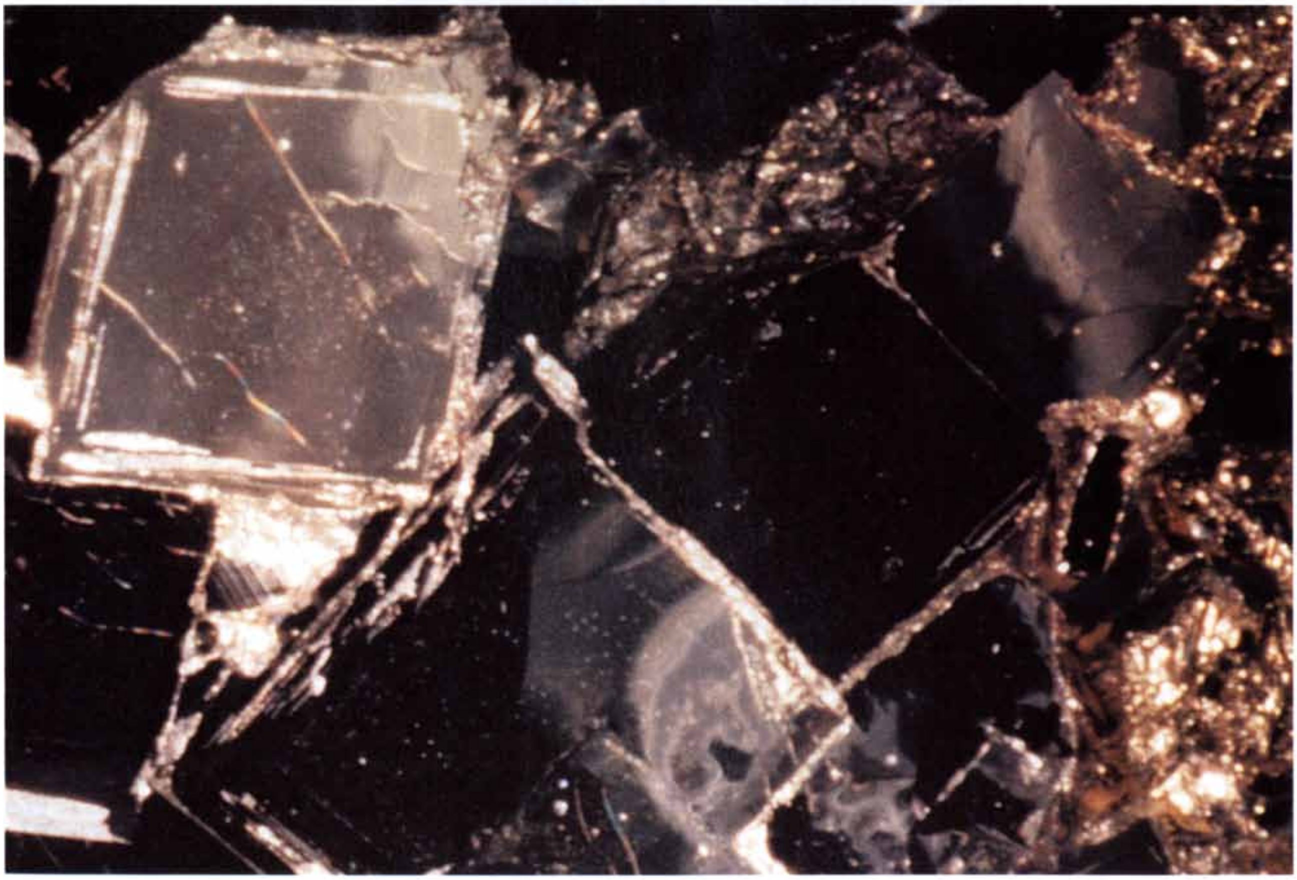
As the oxygen content of $YBa_2Cu_3O_6$ is increased, oxygen is added directly into the bonding environment of the sticks, oxidizing the valence 1+ copper to valence 2+. Oxygen does not attach randomly to the copper-oxygen sticks but rather tries to form as many copper-oxygen diamonds as possible.

As the oxygen content reaches a level where on average 6.5 oxygen atoms are available for every three copper atoms, oxygen pulls enough electrons away from the copper atoms in the copper-oxygen pyramids to yield superconductivity. The resulting compound, $YBa_2Cu_3O_{6.5}$, has a transition temperature of 60 kelvins. This compound has a distinct crystal structure in which chains of diamonds and sticks form an ordered array. As the oxygen content increases even further, the diamond chains continue to grow. When the full chain structure is obtained, the superconducting transition temperature jumps to 90 kelvins.

In both YBa_2CuO_7 and $La_{1.8}Ba_{.2}CuO_4$, the copper atoms are oxidized to valences greater than 2+. As a result, positively charged holes are introduced into the conduction band. Hence, these materials are called *p*-type superconductors. Superconductors whose charge carriers are electrons are described as *n*-type. Until 1988 all the known superconductors were *p*-type, leading many to postulate that there would never be an *n*-type copper oxide superconductor.

In 1988 Y. Tokura, H. Takagi and Shin-ichi Uchida of the University of Tokyo discovered the first and, so far, the only *n*-type ceramic superconductor. The new superconductor was based on neodymium copper oxide (Nd_2CuO_4). The structure of Nd_2CuO_4 is similar to La_2CuO_4. But whereas lanthanum prefers nine oxygen neighbors, the smaller neodymium ion is usually coordinated to eight oxygen atoms. In the compound Nd_2CuO_4 the oxygen atoms form a square prism around the neodymium atoms. As a result, the copper atoms are coordinated to four oxygen atoms in a square planar geometry.

In the superconductors based on Nd_2CuO_4, some of the neodymium atoms are replaced by cerium or thorium, forming a solid solution. Both cerium and thorium are valence 4+ and are of an appropriate size to substitute for some of the neodymium of valence 3+. Hence, they can form the solid solutions $Nd_{2-x}Ce_xCuO_4$ and $Nd_{2-x}Th_xCuO_4$. When x equals .17, these ma-



HIGH-TEMPERATURE SUPERCONDUCTORS, such as this one composed of lead, strontium, dysprosium, calcium, copper and oxygen ($Pb_2Sr_2Dy_{1-x}Ca_xCu_3O_8$), typically form rectangular crystals. Here the crystals are about 100 microns thick.

BASIC COMPOSITION OF COPPER OXIDE SUPERCONDUCTORS AND SOME OF THEIR VARIATIONS

PARENT STRUCTURE	SUBSTITUTIONS	TRANSITION TEMPERATURE RANGES (KELVINS)	COMMENTS
$La_{2-x}Ba_xCuO_4$	Ca or Sr	20–40	A CERAMIC COMPOSED OF La, Ba, Cu AND O WAS THE FIRST COPPER OXIDE SUPERCONDUCTOR DISCOVERED
$Nd_{2-x}Ce_xCuO_{4-y}F_y$	Pr, Sm or Eu Th	10–25	THIS FAMILY CONTAINS THE ONLY *n*-TYPE CERAMIC SUPERCONDUCTORS
$La_{1.8-x}Sm_xSr_{.2}CuO_4$	Eu, Gd, Tb or Dy	20	
$YBa_2Cu_3O_7$	La, Nd, Sm, Eu, Gd, Dy, Ho, Er, Tm, Yb or Lu Cu_4O_8 or Cu_7O_{15}	80–93	$YBa_2Cu_3O_7$ WAS THE FIRST SUPERCONDUCTOR FOUND WHOSE TRANSITION TEMPERATURE EXCEEDS 77 KELVINS
$Bi_2Sr_2CuO_6$	$CaCu_2O_8$ or $Ca_2Cu_3O_{10}$	0–110	
$Tl_2Ba_2CuO_6$	$CaCu_2O_8$ or $Ca_2Cu_3O_{10}$	80–125	$Tl_2Ba_2Ca_2Cu_3O_{10}$ HAS THE HIGHEST TRANSITION TEMPERATURE OF ANY KNOWN SUPERCONDUCTOR
$TlBa_2Cu_2O_5$	$CaCu_2O_7$, $Ca_2Cu_3O_9$ or $Ca_3Cu_4O_{11}$	0–122	
$Pb_2Sr_{2+x}Pr_{1-x}Cu_3O_8$	Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er or Tm	70–85	
$Bi_2Sr_2Sm_{2-2x}Ce_{2x}Cu_2O_{10}$	Eu or Gd	20–25	
$Ba_{1.33}Nd_{.67}Sm_{1.33}Ce_{.67}Cu_3O_9$	Sm, Eu or Gd Nd, Eu or Gd	40	
$La_{2-x}Sr_xCaCuO_6$	NONE KNOWN	60	

terials reach their highest transition temperatures, near 25 kelvins. These materials remain the subject of active research because they have subtle chemical features that are not yet understood and that are relevant to the nature of the charge carriers.

The superconductors with the highest known transition temperatures were discovered in 1988 by Allen M. Hermann and his colleagues at the University of Arkansas. The superconductors are composed of thallium, barium, calcium, copper and oxygen. The toxicity of thallium oxide makes handling the materials under anything but carefully controlled laboratory conditions highly inadvisable, and so their potential for commercial use is not yet clear. Nevertheless, the crystal structure of these compounds confirmed that the key component necessary for high transition temperatures is planes of copper and oxygen atoms.

In these materials, thallium is coordinated to oxygen in a large octahedron. The thallium-oxygen octahedra form a plane that rests on a plane of copper-oxygen octahedra or pyramids. A general formula for the thallium superconductors is written as $Tl_mBa_2Ca_{n-1}Cu_nO_{m+2n+2}$. The subscript m describes the number of layers of thallium-oxygen octahedra. Only compounds that have one or two layers are known. The number of copper layers (n) can vary from one to four. All compounds in the series are superconducting at high temperatures. The highest, at 125 kelvins, is the superconductor $Tl_2Ba_2CaCu_3O_{10}$.

Although the copper oxide superconductors that have been discovered recently have more complex structures than their predecessors, all are basically built of electronically active planes of copper and oxygen. The planes are sandwiched between other layers, which act as spacers and, most important, as reservoirs of positive and negative charge. The electronic state of the layers determines the amount of charge on the copper-oxygen planes and the transition temperature of the compound.

More than 15 years ago Arthur W. Sleight and his collaborators at Du Pont fabricated a ceramic composed of barium, lead, bismuth and oxygen. Although the ceramic had a transition temperature of only 12 kelvins, it was the compound that inspired Bednorz and Müller to search for ceramic superconductors. In 1988 my colleagues and I at AT&T Bell Laboratories discovered that a compound based on barium, bismuth and oxygen was superconducting at a surprising 30 kelvins. These materials have a great deal in common with copper oxides and are an interesting story in themselves. Perhaps these or other now obscure materials will provide new routes to high-temperature superconductivity—only time will tell.

FURTHER READING

INTRODUCTION TO CERAMICS. W. D. Kingery, H. K. Bowen and D. R. Uhlmann. Second Edition. John Wiley & Sons, Inc., 1976.

SOLID STATE CHEMISTRY AND ITS APPLICATIONS. Anthony R. West. John Wiley & Sons, Inc., 1984.

HIGH-TEMPERATURE SUPERCONDUCTIVITY. Reprints from *Physical Review Letters* and *Physical Review B*. The American Physical Society, 1987.

CHEMISTRY OF HIGH-TEMPERATURE SUPERCONDUCTORS. A. W. Sleight in *Science*, Vol. 242, No. 4885, pages 1519–1527; December 16, 1988.

STRUCTURAL CHEMISTRY AND THE LOCAL CHARGE PICTURE OF COPPER OXIDE SUPERCONDUCTORS. R. J. Cava in *Science*, Vol. 247, No. 4943, pages 656–662; February 9, 1990.

CHEMISTRY OF SUPERCONDUCTOR MATERIALS. Edited by Terrell Vanderah. Noyes Publications (in press).

AIDS-Related Infections

Until the AIDS virus is defeated, hope for improving survival time and the quality of a patient's life will rest to a great extent on improved therapies for the opportunistic infections of AIDS

by John Mills and Henry Masur

The AIDS-causing human immunodeficiency virus (HIV) does not by itself produce most of the illness and death associated with the acquired immunodeficiency syndrome (AIDS). HIV can damage organs directly, but by progressively crippling the body's defenses, the virus also sets the stage for the development of opportunistic infections: invasions of microorganisms that proliferate wildly only because the immune system is defective. Such infections, which rarely cause disease—impairment of organ function—in people with a healthy immune system, account for as much as 90 percent of the mortality of AIDS (the end stage of HIV infection).

Antiviral drugs that are effective against HIV, such as zidovudine (AZT), do not eradicate HIV completely, but they do slow the HIV-related decline in immunity. In so doing, they help to delay the onset of opportunistic infections. Until there is a cure for HIV infection, however, the survival time and comfort of patients will depend greatly on therapies that specifically prevent individual opportunistic infections or treat them more effectively.

Research into such therapies has expanded markedly in the past five to 10 years and is already bringing improvements in patient care. For instance, at the start of the AIDS epidemic, there was no way to combat severe infection by cytomegalovirus, a herpesvirus that in immunocompromised patients can cause blindness or damage to the digestive tract or lungs. As of 1990, one drug is licensed and a second is widely available for acute therapy and prevention of recurrences. Until recently the disease *Pneumocystis carinii* pneumonia struck 85 percent of all HIV-infected patients at least once. Today two thirds or more of these episodes can be prevented with medication.

HIV-related opportunistic infections are quite varied, yet they share several features. For example, most of them also arise in people with immunity impaired for other reasons, such as those who take immunosuppressive drugs to prevent rejection of organ transplants. (There are exceptions, though. For unexplained reasons, a few disorders found frequently in AIDS patients are rare in other immunosuppressed individuals.)

It is true as well that HIV-related infectious diseases are caused by organisms that are quite common. Indeed, the incidence of such diseases varies from region to region because the microbes endemic to the areas vary. The diseases usually represent the reactivation of quiescent infections that were held in check by the patient's immune system before the acquisition of HIV.

As a rule, the infections that beset HIV patients are those normally controlled by the cell-mediated arm of the immune system, which consists of such white blood cells as *T* lymphocytes and macrophages and is decimated by HIV. In contrast, infections (including many caused by bacteria) that are eradicated primarily by the other major—antibody-mediated—arm of the immune system or by nonspecific defenses are normally managed fairly well, at least for a time.

Just how cellular immune function is impaired by HIV is only partly understood. The virus is known to eliminate gradually the subset of *T* cells that bear what is called the CD4 receptor. Such a loss would be expected to debilitate the cell-mediated immune response, because CD4-positive (CD4+) *T* cells, also known as helper *T* cells, control the activity of many other components of the cellular immune response. Nevertheless, some immune dysfunction is often present even before a decline in CD4+ *T* cells becomes evident, which means the loss of helper *T* cells is not the sole cause of the impairment.

Investigators have found, too, that the development of particular opportunistic infections is related to the level of CD4+ *T* cells in the blood. Healthy individuals have about 1,000 such cells in every cubic millimeter of blood. In HIV-infected individuals the number declines by an average of about 40 to 80 every year.

When the helper *T* cell count falls to between about 400 and 200 per cubic millimeter, the first infections usually appear: relatively benign but annoying infections of the skin and mucous membranes. Among these may be thrush (painful sores of the mouth caused by the fungus *Candida albicans*), shingles (infection of the nerves and skin by varicella-zoster virus), unusually severe athlete's foot (caused by several types of fungi) and oral hairy leukoplakia (whitish patches on the tongue caused by Epstein-Barr virus).

Once such symptoms appear, a person is often said to have the AIDS-related complex (ARC). The same is true for individuals who suffer from chronic, unexplained fevers, diarrhea, night sweats or weight loss. As immunity wanes still further, serious, AIDS-defining opportunistic infections usually develop. These often include three major killers: *P. carinii* pneumonia, cryptococcal meningitis (caused by a fungus) and toxoplasmosis (a parasitic infection of the brain). In the past, those three infections alone accounted for 50 to 70 percent of the

JOHN MILLS and HENRY MASUR work on separate coasts but have a shared mission: developing better therapy for the opportunistic infections associated with AIDS. Mills, who earned his M.D. in 1966 from the Harvard Medical School, is professor of medicine and microbiology at the University of California School of Medicine, San Francisco, and chief of the division of infectious disease at San Francisco General Hospital. Masur, who received his medical degree from Cornell University in 1972, where he was on the faculty for several years, is chief of the department of critical-care medicine at the National Institutes of Health.

deaths in patients with AIDS in the U.S.

John P. Phair of Northwestern University and his colleagues have shown that the risk of developing *P. carinii* pneumonia increases dramatically after the CD4+ *T* cell count drops below 200 per cubic millimeter. Without preventive therapy, more than a fifth of patients with a count of less than 200 will have a first bout of *P. carinii* pneumonia within a year. The risk for disease from certain other infections, such as those caused by cytomegalovirus, does not soar until the CD4+ *T* cell count has fallen below the 100 mark.

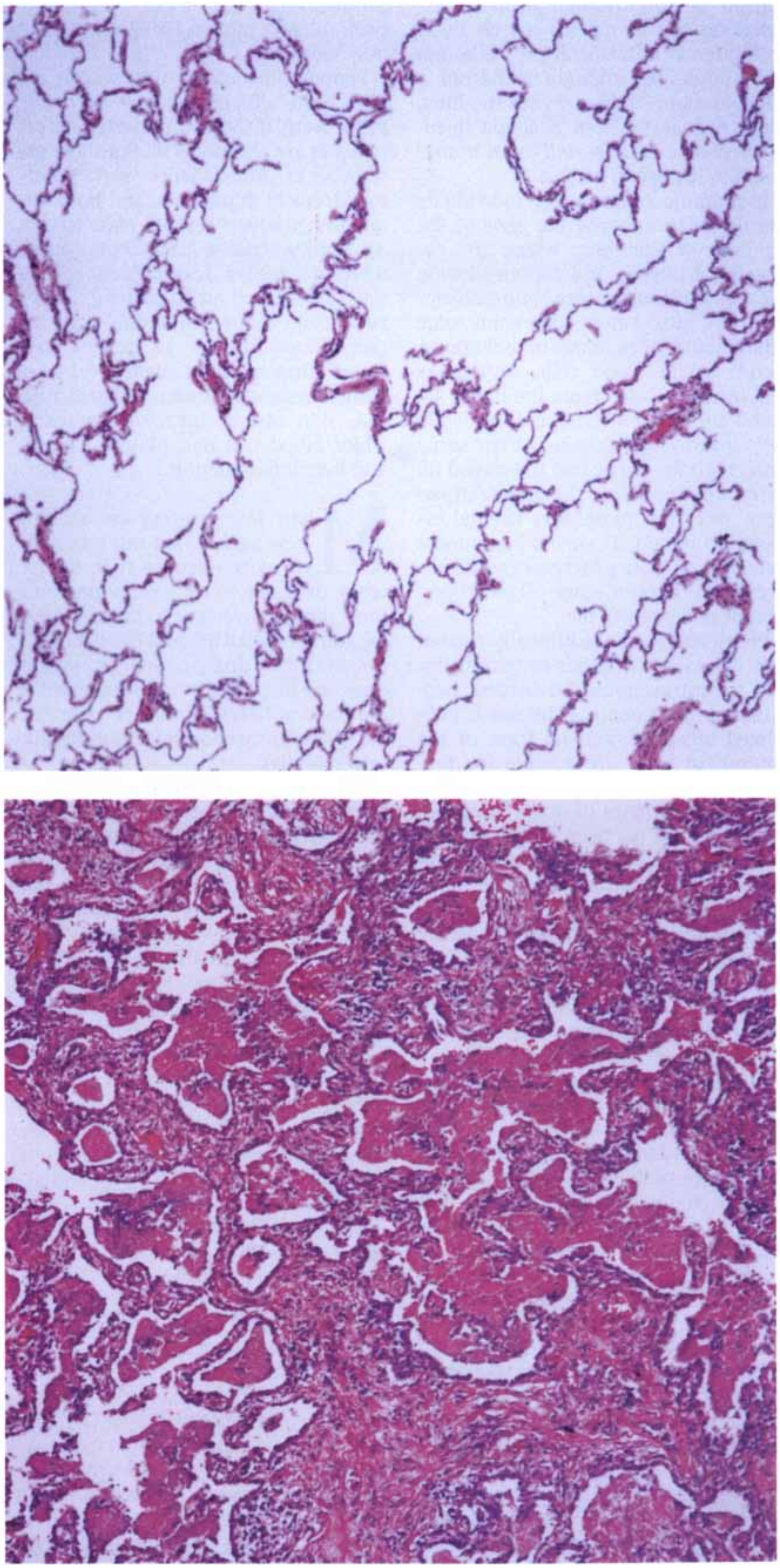
The ideal approach to the management of the opportunistic infections of AIDS has three parts: prevention of the infection (primary prophylaxis), treatment of any active infection and prevention of recurrences (secondary prophylaxis). These days a good deal of attention is being focused on secondary prophylaxis because, without it, HIV-associated opportunistic infections almost always return.

The infections recur because antimicrobial therapies rarely eliminate the causative microorganisms completely. In healthy individuals such treatments are effective because they give the body time to marshall its defenses against a pathogen. In patients infected by HIV, the weakened immune system cannot muster the strength required to eradicate the pathogen.

Generally, a drug that is effective against an active infection is likely to be of value for primary or secondary prophylaxis. Unfortunately, many drugs that can be tolerated for a short time have drawbacks that militate against their indefinite administration as preventive therapies—namely, potential toxicity or the need for intravenous delivery, which is inconvenient and expensive and may cause infection. These problems underscore the importance of developing a range of drugs for each of the major opportunistic infections of AIDS. Choices are needed so that a patient who does not respond to one compound or who cannot tolerate its side effects or its interaction with other drugs (such as zidovudine) will have recourse to other therapies.

Three-part management of the major opportunistic infections of AIDS has been most successful for *P. carinii* pneumonia, even though the survival strategy of the organism is rather a mystery. This pathogen, which usually infects the lungs (but may affect other sites), is often classified as a parasite, although new studies indicate it may be a fungus.

P. carinii probably passes from per-



ALVEOLI, air sacs of the lung, contain only gases (*clear areas*) in a healthy individual (*top*) but become filled with fluid and debris (*pink*) in patients with *Pneumocystis carinii* pneumonia (*bottom*), one of the major opportunistic infections of AIDS.

son to person through droplets exhaled into the air, much as tuberculosis or influenza is transmitted. Although most adults are thought to harbor a small amount of *P. carinii* in the lung, there has never been a single documented case of *P. carinii* in an immunologically normal person.

In immunocompromised individuals the organisms invade the alveoli, the air sacs of the lung, where the exchange of oxygen and carbon dioxide occurs. Despite severe immunosuppression, AIDS patients do mount some inflammatory response to infections. Hence, white blood cells, antibodies and other proteins from the blood accumulate in the alveoli, and the white cells invade lung tissues. At the same time, both the tissue and the alveoli fill with fluid, impeding the uptake of oxygen and thus its delivery to vital organs. Untreated, *P. carinii* pneumonia almost inevitably progresses to death. Even with therapy, some 10 to 20 percent of patients will die.

Physicians have traditionally treated this disease with either of two drugs. One is intravenously delivered pentamidine isethionate. (The much publicized inhalable, aerosol form of the compound is of great value for prevention and shows promise for the nontoxic treatment of acute infection, but it has not yet been shown conclusively to be effective for acute therapy.) The other drug is oral or intravenous co-trimoxazole, a combination of two other drugs, sulfamethoxazole and trimethoprim.

In certain organisms, and presumably in *P. carinii*, pentamidine is thought to prevent microbial replication by inhibiting the synthesis of DNA, although exactly how it does so is not clear. Co-trimoxazole inhibits DNA synthesis as well [*see illustration on page 55*]. The sulfamethoxazole component inhibits the activity of an enzyme—dihydropteroate synthase (DHPS)—that helps to produce a derivative of folic acid. This folate is essential to the production of nucleotides, the building blocks of DNA. DHPS inhibitors are attractive because they impair the synthesis of microbial DNA but do not interfere with the synthesis of human DNA. Human and other mammalian cells do not need and indeed do not possess DHPS; they can make use of preformed folate in food.

The trimethoprim component of co-trimoxazole complements sulfamethoxazole by inhibiting strongly the activity of a related enzyme—dihydrofolate reductase (DHFR)—that converts the folate synthesized by microbes into forms that are directly involved in the production of nucleotides. Trimethoprim inhibits human DHFR as well but only minimally.

Pentamidine and co-trimoxazole are each quite effective against *P. carinii*, particularly if therapy is started before patients are extremely ill. But there are drawbacks. Both agents work poorly in extremely ill patients, and both frequently produce serious side effects. As many as half of all patients cannot tolerate the first drug prescribed and must be offered an alternative. For its part, intravenous pentamidine can impair kidney function or cause a profound drop in blood sugar levels. Co-trimoxazole can produce severe nausea, skin rashes, fever, decreases in white blood cell and platelet counts, and liver inflammation.

Many investigators are seeking new agents that are less toxic or more effective than the existing drugs or, in the case of pentamidine, more convenient to take. A growing number of DHFR and DHPS inhibitors, both new drugs as well as existing ones, are in clinical trials. Among these are the new DHFR inhibitors trimetrexate (given intravenously) and piritrexim (given orally).

Other approaches being studied in the clinic, with promising results, are the combination of two standard antimicrobial compounds—clindamycin (which inhibits bacterial protein synthesis) and primaquine (an antimalarial drug)—and the combination of trimethoprim and dapsone (a DHPS inhibitor in the sulfone class). Also under study clinically are difluoromethylornithine, or DFMO (an inhibitor of DNA synthesis in certain parasites), and an agent provisionally called compound 566C80, which impairs the generation of energy by mitochondria. Researchers are also screening thousands of DHPS and DHFR inhibitors for activity against *P. carinii* in the test tube.

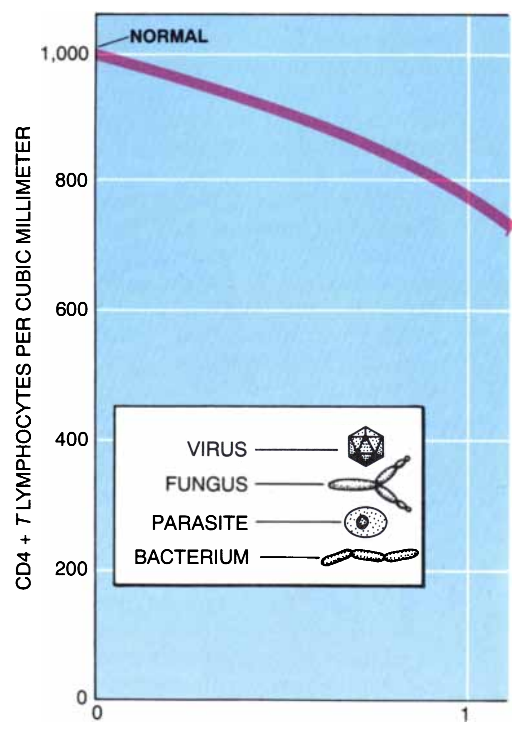
Therapies that do not attack the microbe itself are being considered to improve the care of acute infection. According to several recent clinical trials, prompt delivery of glucocorticoids reduces inflammation of the lungs and can improve the chance of survival for severely ill patients. Because patients with *P. carinii* pneumonia suffer from a deficiency of the lung surfactant responsible for keeping the airways open, aerosol replacement therapy with an artificial surfactant may be tested in the future.

As for primary and secondary prophylaxis, inhaled pentamidine, dapsone and co-trimoxazole all appear to be highly effective. Their relative efficacy and toxicity are not known, however, and are under study. On the basis of Phair's findings, it seems that *P. carinii*-specific primary prophylaxis should begin shortly before the CD4+ *T* cell count falls to 200 or earlier if a patient has ARC or AIDS.

The value of secondary preventive therapy becomes strikingly clear when one looks at recent data. Not long ago about half of all patients who survived a first attack of *P. carinii* pneumonia had a recurrence within a year, whether or not they took zidovudine. Today only 10 to 20 percent of patients taking both zidovudine and a *P. carinii*-specific prophylactic suffer a relapse within a year. Moreover, those who receive no specific preventive therapy are likely to live for an average of only a year after the first episode of pneumonia; those who receive secondary prophylaxis and zidovudine appear to be surviving for an average of about three years.

Such data are encouraging. Still, we and others hope that as new drugs become available for the treatment of acute *P. carinii* infection, some of them will also form the basis of effective, nontoxic, convenient and economical preventive therapies.

Progress toward developing new therapies for other AIDS-related opportunistic infections has been slower. Toxo-



OPPORTUNISTIC INFECTIONS frequently found in people infected with the AIDS-causing human immunodeficiency

plasmosis is a case in point. This disease is caused by *Toxoplasma gondii*, a one-cell parasite that lives in cats and other animals as well as in humans. It can be passed to humans by contact with feces from infected cats and can also be transmitted in undercooked meat from other infected animals. In the U.S. perhaps 20 percent or more of adults harbor toxoplasma organisms, commonly in the brain or muscle.

This latent infection is activated, usually in the brain, in some 5 to 15 percent of AIDS patients. The rapidly dividing parasites evoke an inflammatory response that destroys brain cells, causing seizures or localized abnormalities of function, such as weakness or decreased sensation on one side of the body.

So far the only effective therapy for toxoplasmosis is a combination of pyrimethamine (a DHFR inhibitor) and sulfadiazine (a DHPS inhibitor), which can be given by mouth. Most patients improve within one to three weeks, as growth of the organism is halted and the associated swelling of the brain decreases. Sadly, most of the brain damage itself is irreversible. And, as is true for *P. carinii* pneumonia, secondary prophylaxis is required because the drugs prescribed for toxoplasmosis suppress the infection rather than cure it. The therapy causes many of the same side effects as co-trimoxazole. Indeed, about 30 percent of patients cannot tolerate a full course of treatment.

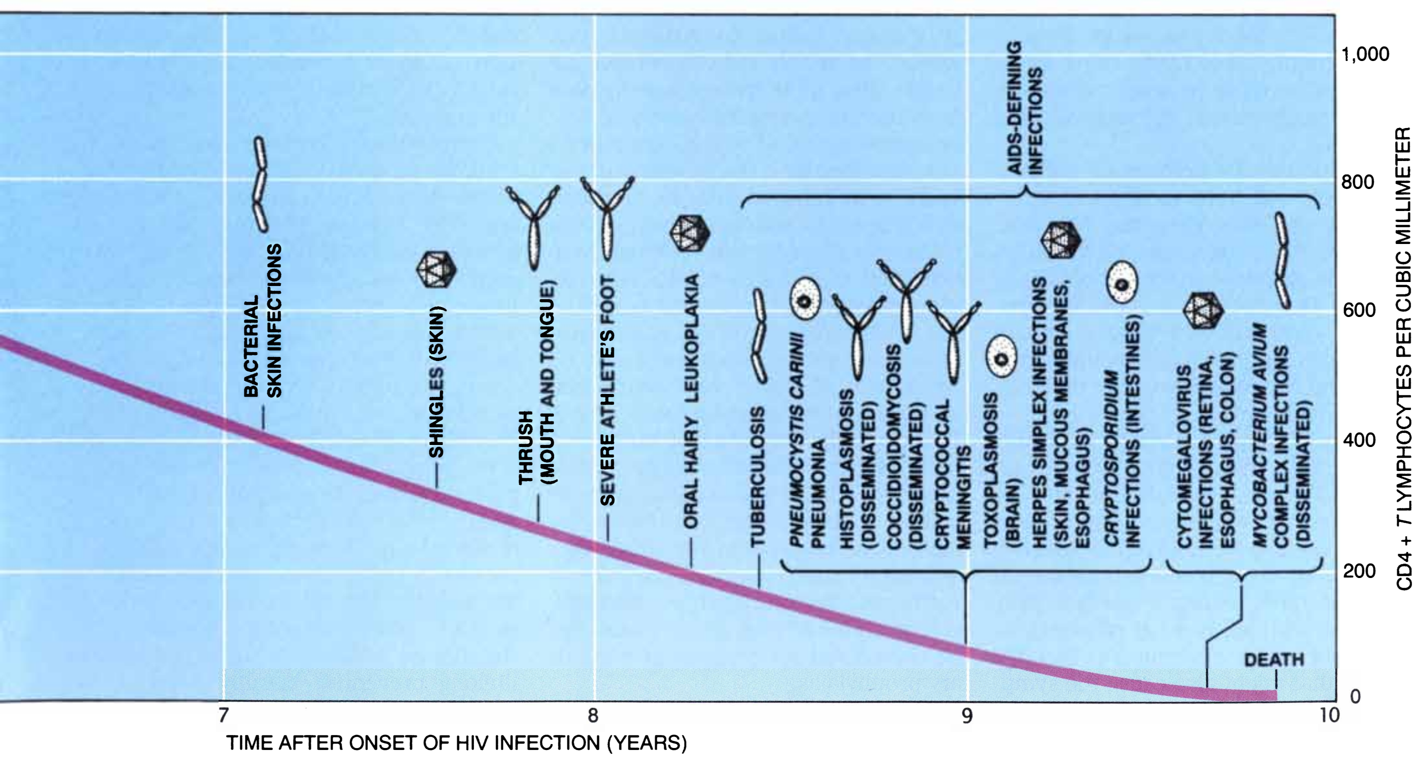
Fortunately, some possible alternatives are showing promise. In one small clinical trial (still awaiting replication), a combination of the DHFR inhibitor pyrimethamine and the antibacterial clindamycin worked nearly as well as the standard pyrimethamine-sulfadiazine combination. Moreover, screening of other antimicrobial drugs for activity against *T. gondii* in the test tube or in mice has identified a number of potential treatments. These include clindamycin and tetracycline analogues as well as certain DHFR and DHPS inhibitors. None of the compounds has yet been well studied in people, however.

In a different strategy, some workers are capitalizing on the fact that *T. gondii* can survive only within cells and favors macrophages. In normal individuals, macrophages ingest invading microbes and then are activated to destroy them. They are activated by gamma interferon and other so-called cytokines secreted by *T* cells. But, because of *T* cell damage, gamma interferon is in short supply in patients infected with HIV, and so the parasite is ingested but not destroyed. As a result, researchers are exploring the value of adding gamma interferon to the standard pyrimethamine-sulfadiazine regimen. In cell-culture and animal studies, this approach has increased the effectiveness of the two-drug therapy. Clinical trials are expected to begin shortly.

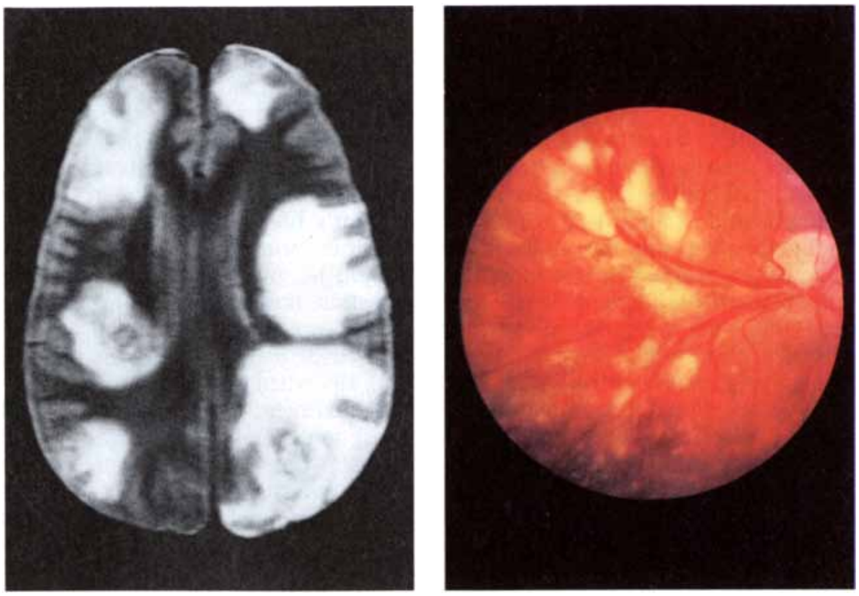
Pyrimethamine and sulfadiazine are usually continued for the rest of a patient's life, to prevent recurrences. Whether primary preventive therapy should be given routinely is an open question, however. Many physicians in the U.S. hesitate to administer potentially toxic drugs to everyone infected with HIV when they consider that, without treatment, only a fifth or fewer of those individuals will ever fall ill with toxoplasmosis. Yet, in France, where toxoplasmosis is much more common, primary prophylaxis is recommended frequently. Studies to assess the possible benefits of such therapy have begun in the U.S. and Europe.

In contrast to toxoplasmosis, the major bacterial infections of adults with AIDS—those caused by organisms of the *Mycobacterium avium* complex—rarely involve the brain. They do, however, affect many other organs—notably, the lungs, liver, spleen, lymph nodes, bone marrow, intestines and blood. Indeed, infection is usually disseminated throughout the body.

These bacteria, which are closely re-



virus (HIV) typically arise in a sequence related to the concentration in the blood of cells known as CD4+ *T* lymphocytes. The cell count, which declines over time, reflects the degree of immune impairment caused by HIV. The infections are charted on the basis of data collected in part by Suzanne M. Crowe of Fairfield Hospital in Melbourne, Australia.



CERTAIN INFECTIONS are identifiable by imaging. For instance, toxoplasmosis of the brain (*left*), which can cause seizures, produces discrete lesions that show up as light areas on magnetic resonance images. In photographs of retinas damaged by cytomegalovirus (*right*), yellowish areas represent dead tissue, and bright-red splotches (shown obliterating blood vessels) reflect bleeding into the retina. (The black area has detached from the back of the eye and accounts for some bleeding.)

lated to the agent responsible for tuberculosis (*M. tuberculosis*), are everywhere in the environment—such as in dust, soil and poultry and dairy products. Presumably, they make their way into the body in inhaled dust or droplets and in food and water. Despite their ubiquity, they rarely cause active infection, even in people who are immunocompromised by causes other than HIV.

Yet almost 50 percent of patients with AIDS will have widespread infection by the time they die. Mycobacteria evoke a vigorous cell-mediated immune response in immunologically healthy individuals but not in AIDS patients. Their infected tissues contain enormous numbers of *M. avium* organisms, but the tissues are only minimally inflamed.

The proliferating mycobacteria may well be the cause of the symptoms that often accompany the infection, such as fever, sweats, weight loss and fatigue. But no one has determined conclusively whether these mycobacteria are, in fact, at fault. Patients are generally infected with many other microorganisms (such as cytomegalovirus and HIV itself) that might also cause the symptoms. Likewise, decreases in red cells, white cells and platelets that can accompany *M. avium* infection could be caused by infection of the bone marrow by the mycobacteria. On the other hand, some patients have huge amounts of mycobacteria throughout their body but exhibit few or no signs of specific organ dysfunction.

When a patient is symptomatic, it obviously would be desirable to have an effective treatment. Yet the mainstay of therapy against tuberculosis—isoniazid—invariably fails to control the proliferation of *M. avium*. Investigators have therefore tested a battery of other agents, some of which, alone or in combination, limit the growth of these bacteria in cultured cells or immunodeficient mice. Among the compounds often prescribed for tuberculosis, those classified as aminoglycosides (such as streptomycin and amikacin) as well as the rifamycin-group drugs and ethambutol have all demonstrated some effect against *M. avium*. The same is true of certain other antimicrobials: ciprofloxacin, imipenem and clofazimine.

Two separate clinical trials have demonstrated that combinations of four or five of these drugs can reduce the amount of bacteria in the blood and ameliorate such symptoms as fever. If further studies validate those findings, combinations of such agents could also be considered for primary or secondary prophylaxis.

Work on treatments for the major fungal and viral infections of AIDS is advancing as well. For example, a new treatment has been developed for cryptococcal meningitis—which arises in about 10 percent of AIDS patients in the U.S. And additional experimental therapies are in development.

The organism responsible for cryptococcal meningitis, the yeast *Cryptococcus neoformans*, closely resembles the fungus that causes thrush, with one important difference: *C. neoformans* spreads more readily beyond the skin and mucous membranes. It is enveloped in a slippery outer coating that impedes macrophages and other immune cells, such as granulocytes, from ingesting the cryptococci. The coating also helps the ingested fungi resist damage by enzymes in those cells. What is unknown is just how *C. neoformans* produces disease. So far no one has identified any toxic products of the organism.

The initial infection is thought to begin in the lungs, where it probably remains confined in people with normal immune defenses. In immunocompromised patients, however, the fungus invades the bloodstream and then usually lodges in the meninges, the membranes covering the brain. The potentially lethal invasion of the meninges announces itself by causing severe headaches, a stiff neck and, often, fever.

For the past 30 years up until recently, amphotericin B was the only treatment. This drug binds to ergosterol, a cholesterollike substance unique to the cell membrane of fungi. Its binding action damages the membrane, yielding a fatal leakage of the internal contents of the organism.

Unfortunately, the drug must be given intravenously, damages human red blood cells (causing anemia) and can harm the kidneys. Moreover, the formulation required to make the active compound suitable for intravenous delivery frequently induces strong flulike symptoms, such as shaking chills and high fever. Even with a full course of treatment, some 20 percent of AIDS patients who come down with cryptococcal meningitis will die from it.

In theory, amphotericin B can actually eradicate fungal infections, not merely slow their growth. Yet the high doses required would probably kill the patient as well. Because the doses that are in fact given are suppressive rather than curative, patients who survive the disease must, once again, receive lifelong preventive therapy. They require intravenous treatments once or twice a week.

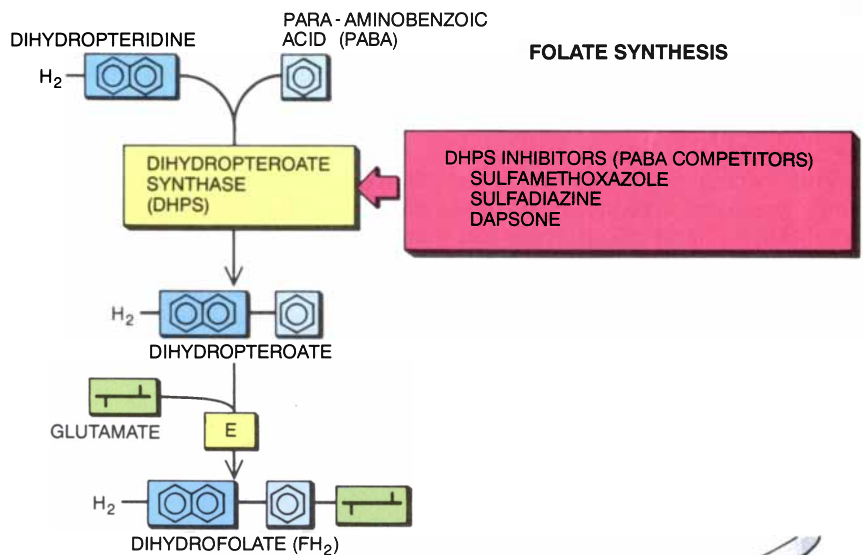
Clearly, an alternative that could be taken by mouth would be a major advance for preventive therapy. The most promising agents in clinical trials belong to a group of antifungal compounds known as imidazoles, which inhibit the synthesis of ergosterol. Two of these agents—itraconazole and fluconazole—are particularly exciting. Early clinical trials with itraconazole indicate it may be nearly as effective as amphotericin B but less toxic. Analysis of a recently completed, large prospective study suggests that fluconazole may be as effective as amphotericin B, and it is now licensed in many countries, including the U.S. Another study is under way to evaluate that compound as a preventive therapy.

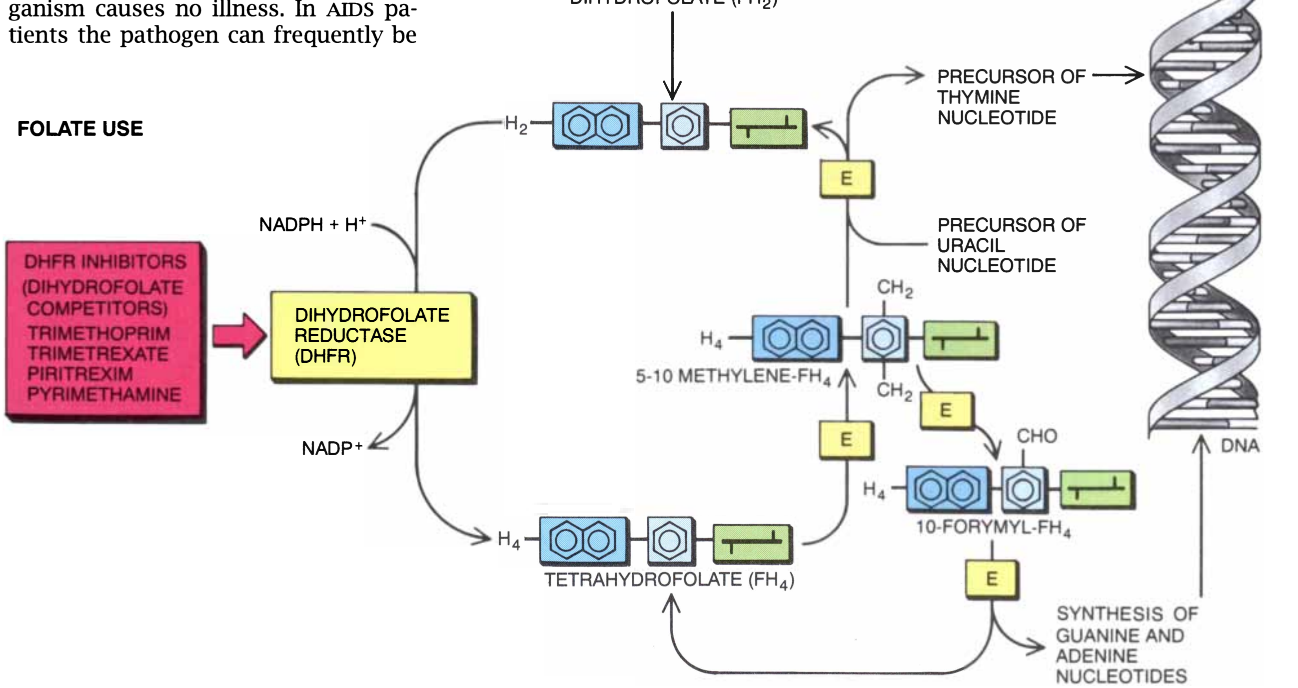
Oral medications are also badly needed for the prevention and treatment of cytomegalovirus infection. By adulthood, more than half of all Americans will have been infected by this virus, whether by sexual contact or as a consequence of spending time in close quarters with other people, particularly where hygiene is poor.

In healthy adolescents and adults, cytomegalovirus infection can cause symptoms that resemble mononucleosis, but most often patients are unaware they have acquired the virus. As with all herpesviruses, an initial infection is followed in such healthy individuals by the establishment of a lifelong, quiescent one, in which the microorganism causes no illness. In AIDS patients the pathogen can frequently be recovered in quantity from the blood and many other tissues.

Functional damage, however, is usually confined to the eye and the gastrointestinal tract. In some 7 percent of those with AIDS, the virus invades the rod and cone cells of the retina, which sense light and color. The infection ultimately kills the affected cells, irreversibly damaging vision. Worse, dead areas of the retina are likely to tear and detach from the eye. The detachment and the bleeding that ensues may result in sudden, complete blindness.

In another 5 percent of AIDS patients, the virus takes root in the digestive tract, anywhere from the mouth to the rectum, although it most often affects the esophagus or colon. Ulcers in the esophagus can make swallowing painful and may also cause chest pain. In the colon, they often give rise to ab-





MANY DRUGS for *P. carinii* pneumonia or toxoplasmosis (*red*) interfere with activity of the microbial forms of the enzymes dihydropteroate synthase (DHPS) or dihydrofolate reductase (DHFR). These enzymes participate in the production (*top*) and use (*bottom*) of folates (folic acid derivatives) for the production of nucleotides, the building blocks of DNA. DHPS inhibitors resemble para-aminobenzoic acid (PABA), a subunit of the major end product of folate synthesis: dihydrofolate. By binding to DHPS, the medications prevent the enzyme from interacting with PABA itself and thus from linking the PABA to dihydropteridine, another subunit of dihydrofolate. DHFR inhibitors are analogues of dihydrofolate. By binding competitively to DHFR, they prevent it from converting the true folate into a form that can itself be converted into the varieties directly needed for the synthesis of nucleotides. The boxed letter E represents other enzymes.

dominal pain and diarrhea. At times, the uncontrollable bleeding from damaged esophageal or colonic tissues can result in death.

The drugs that are now widely available for treating cytomegalovirus disease are toxic and must be given intravenously. One, ganciclovir, is approved in the U.S. and several other countries. The other, foscarnet, is still experimental but is expected to be licensed soon. Both compounds selectively inhibit the viral DNA polymerase, an enzyme essential to the construction of new DNA from nucleotides and thus essential to viral replication. In other respects, however, the two drugs are quite different.

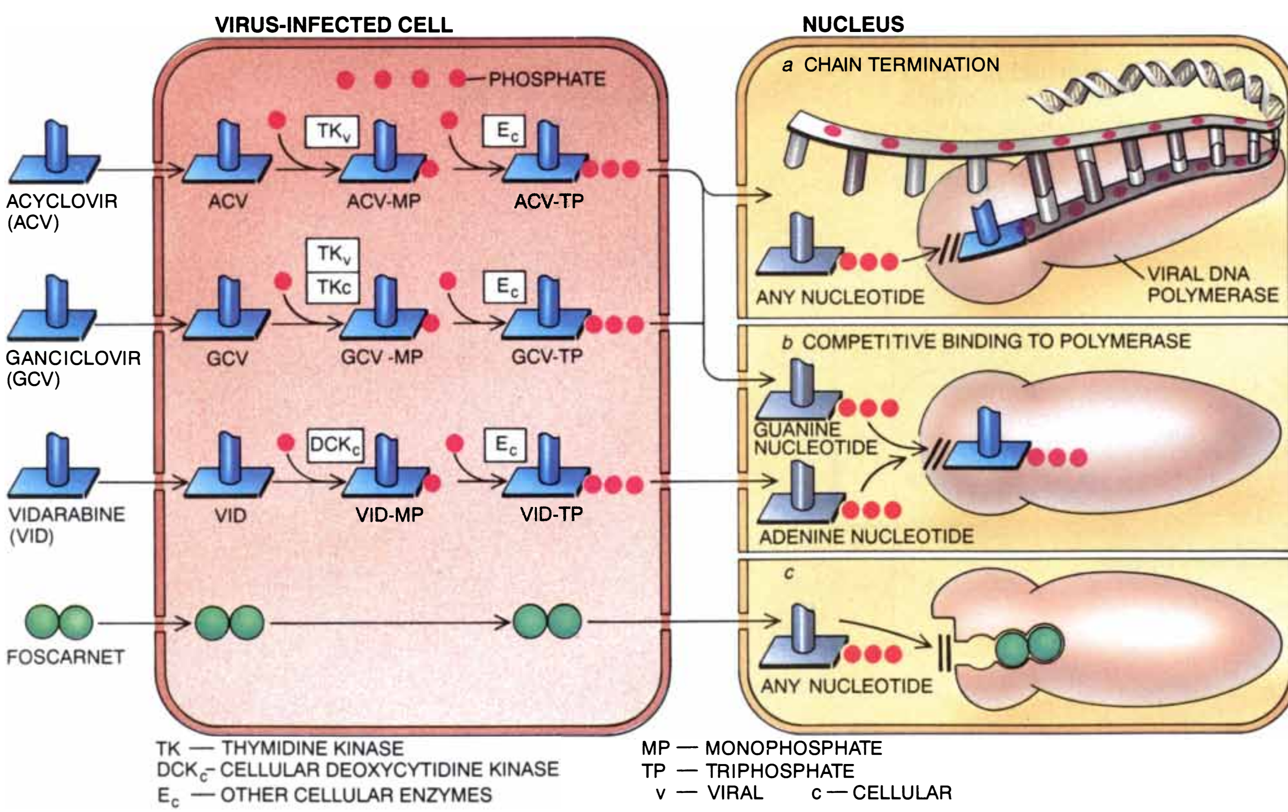
Ganciclovir, which works against many herpesviruses, is an analogue of the nucleoside guanosine (a nucleotide precursor), and the drug's ability to mimic guanosine accounts for its antiviral activity. Virus-infected cells add phosphate to the compound as if the drug were guanosine, forming ganciclovir triphosphate, an ersatz nucleotide. This imposter then binds competitively to the viral DNA polymerase, inactivating the polymerase by blocking it from taking up true nucleotides. Unfortunately, ganciclovir is also phosphorylated to some extent in uninfected cells, where it can interfere with the activity of the human DNA polymerase.

The other drug, foscarnet, is an analogue of two phosphates that the DNA polymerase cleaves from triphosphate nucleotides before adding the nucleotides to a growing chain of DNA. It, too, is taken up by the viral DNA polymerase and blocks the enzyme from taking up nucleotides. Foscarnet is active against the DNA polymerases of all herpesviruses that infect humans and also against the DNA polymerase (reverse transcriptase) of HIV itself; it is not active against the human enzyme.

By interfering with viral replication, both drugs halt the progression of retinal disease caused by cytomegalovirus and apparently halt the spread of infection at other sites, too. But because the drugs do not actually kill the virus, they must, as usual, be taken continually. For secondary prevention, ganciclovir and foscarnet have to be administered once or more a day, which requires the implantation of a catheter leading to a vein in the chest. Although such a shunt makes therapy possible, it also makes the patient susceptible to infections in the tissue around the catheter.

Another drawback is the toxicity of ganciclovir to white blood cells, particularly granulocytes. In 10 to 20 percent of patients, the decline in granulocytes is so severe that treatment has to be discontinued (because a low granulocyte count leaves patients vulnerable to sometimes fatal bacterial infections). Foscarnet can cause reversible kidney failure and reversible metabolic and neurologic abnormalities.

In a particularly troubling development, cytomegalovirus has been shown capable of developing resistance to ganciclovir. The problem is not yet widespread, but several patients who were treated at one point then later failed to respond to the drug. Fortunately, in laboratory tests and early clinical trials, ganciclovir-resistant strains have proved sensitive to foscarnet. Yet, in



FOUR ANTIHERPES MEDICATIONS (*far left*) interfere with the ability of viral DNA polymerases to construct DNA from nucleotides. These enzymes normally remove two of the three phosphate groups on a nucleotide and then connect the remainder of the nucleotide to an OH group on the previously joined nucleotide. Acyclovir, ganciclovir and vidarabine are all phosphorylated by enzymes in virus-infected cells, after which they resemble normal triphosphate nucleotides (*left panel*); foscarnet resembles the diphosphate group usually cleaved from nucleotides by a polymerase. Acyclovir triphosphate can halt DNA production *(right)* by chain termination (*a*): if it is added to a growing strand of DNA, it halts chain growth because it lacks the OH needed for the addition of the next nucleotide. Acyclovir triphosphate—like phosphorylated ganciclovir and vidarabine—can also bind strongly to the polymerase, blocking its nucleotide binding site and thus preventing the enzyme from acting on actual nucleotides (*b*). Foscarnet, too, clogs the nucleotide binding site (*c*).

the laboratory, investigators have succeeded in producing mutants of cytomegalovirus that are resistant to foscarnet as well. No such variants have yet been identified in a patient, but it seems only a matter of time before they begin to show up.

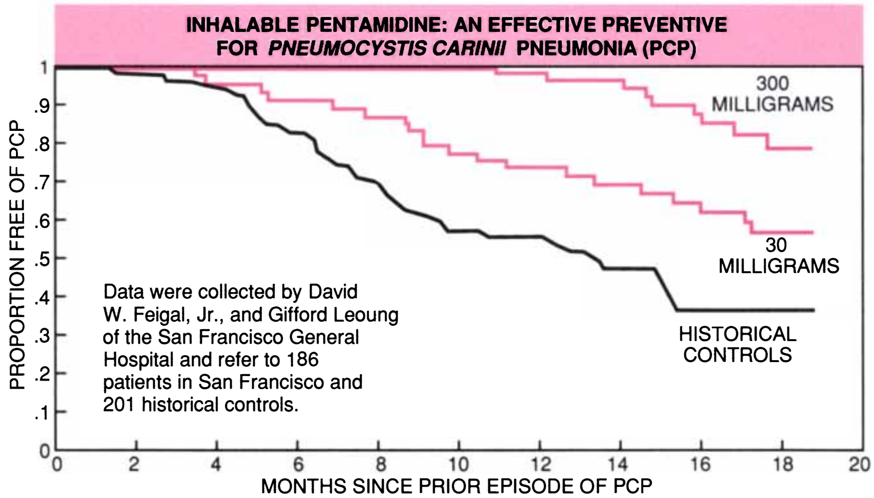
Because of the phenomenon of resistance and the problems posed by intravenous delivery, alternative therapies are urgently needed. We have every hope that in the near future one or more orally delivered relatives of ganciclovir will be developed. One such drug, named HOE-602 (after its manufacturer, Hoechst), is being evaluated in clinical trials in West Germany. The substance is readily absorbed from the digestive tract and is converted by the body to ganciclovir.

Another nucleoside analogue—acyclovir—may prove useful as well. Like ganciclovir, it inhibits the viral DNA polymerase. Acyclovir is appealing because it is significantly less toxic than ganciclovir and because it can be taken orally. On the other hand, acyclovir is also much less active than ganciclovir toward cytomegalovirus. Nevertheless, high doses can suppress the activation of cytomegalovirus infection in organ-transplant recipients and might do the same in AIDS patients.

The search continues for drugs that not only are less toxic than ganciclovir and foscarnet but also are equal or more effective at inhibiting the viral DNA polymerase. At the same time, investigators are trying to develop products that interfere with viral activities other than DNA replication in the hopes that these agents will be effective against ganciclovir- and foscarnet-resistant viruses.

In contrast to cytomegalovirus, two other herpesviruses important in AIDS—herpes simplex type 1 and herpes simplex type 2—do produce recurrent infection in many immunologically normal individuals. The viruses most commonly cause, respectively, self-limited cold sores around the mouth and recurring genital ulcers. In the context of HIV infection, these lesions are not self-limited. They tend not to heal, and they enlarge progressively. Because they are painful, they can interfere with the desire to eat if they arise in the mouth or rectum.

Although acyclovir is not overly effective against cytomegalovirus, it is an excellent therapy for herpes simplex viruses. Even patients who have large ulcers have their lesions controlled within only a few weeks. Unfortunately, the resistance of herpes simplex viruses to the drug seems to be on the upswing.



Most of the drug-resistant strains can escape harm by acyclovir because they do not produce significant amounts of an enzyme called thymidine kinase. The viral version of the enzyme phosphorylates acyclovir, a step crucial to the activity of all nucleoside analogue drugs. The viral kinase is not, however, essential to the replication of herpes simplex viruses, which have ways to compensate for the loss of the enzyme.

A logical strategy for coping with thymidine kinase–deficient strains is to deliver a nucleoside analogue that can be phosphorylated by human enzymes in infected cells. One such drug is vidarabine, which has served for more than 10 years to treat life-threatening herpes simplex and varicella-zoster virus infections. At least in laboratory tests, many acyclovir-resistant strains of herpes simplex viruses have proved susceptible to vidarabine. This finding suggests that vidarabine may help when acyclovir fails, although it has to be given intravenously.

Because foscarnet does not have to be phosphorylated to become active, it too might work, and both laboratory and preliminary clinical studies have been encouraging. Reports of herpes simplex strains that are resistant to foscarnet are already appearing, however.

As the problem of treating drug-resistant strains of viruses illustrates, the challenges ahead are formidable. There must be effective and varied treatments for each of the major opportunistic infections affecting AIDS patients, as well as easily administered, nontoxic and economical drugs for prophylaxis. Otherwise, infections and other HIV-related disorders (such as certain cancers) that cannot be controlled will simply take the place of the infections that are managed effectively.

Therapies are needed as well for opportunistic infections that are less common (such as infection by the parasite cryptosporidium, which can cause diarrhea) or are important only in selected pockets of the globe. Examples are histoplasmosis and coccidioidomycosis, which are disseminated fungal infections generally confined, in the first case, to the central U.S. (in particular the Mississippi River valley) and scattered areas elsewhere in the world and, in the second case, to deserts of the southwestern U.S. and Central and South America.

Clearly, research into new therapies is far from complete. Still, we are pleased to see growing recognition that HIV-related opportunistic infections must be a major target of drug development and study. Each improvement will make a significant difference in the life of someone who contends each day with the direct and indirect effects of the deadly AIDS virus.

FURTHER READING

AIDS: ETIOLOGY, DIAGNOSIS, TREATMENT AND PREVENTION. Edited by Vincent T. DeVita, Jr., Samuel Hellman and Steven A. Rosenberg. Second Edition. J. B. Lippincott Co., 1988.

OPPORTUNISTIC INFECTIONS IN PATIENTS WITH THE ACQUIRED IMMUNODEFICIENCY SYNDROME. Edited by Gifford Leoung and John Mills. Marcel Dekker, 1989.

PULMONARY INFECTIOUS COMPLICATIONS OF HUMAN IMMUNODEFICIENCY VIRUS INFECTION. John F. Murray and John Mills in *American Review of Respiratory Disease*. Part I, Vol. 141, No. 5, pages 1356–1372; May, 1990. Part II, Vol. 141, No. 6, pages 1582–1598; June, 1990.

The Language of Fractals

These unimaginably detailed structures are more than mathematical curiosities. Fractal geometry succinctly describes complex natural objects and processes

by Hartmut Jürgens, Heinz-Otto Peitgen and Dietmar Saupe

"Nature has played a joke on mathematicians. The 19th-century mathematicians may have been lacking in imagination, but nature was not. The same pathological structures that the mathematicians invented to break loose from 19th-century naturalism turned out to be inherent in familiar objects all around us."

—FREEMAN DYSON
"Characterizing Irregularity,"
Science, May 12, 1978

The "pathological structures" conjured up by 19th-century mathematicians have, in recent years, taken the form of fractals, mathematical figures that have fractional dimension rather than the integral dimensions of familiar geometric figures (such as one-dimensional lines or two-dimensional planes). The current fascination with fractals is largely a result of the work of Benoit B. Mandelbrot of the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y. Mandelbrot coined the term fractal in 1975; he derived the word from the Latin *fractus*, the adjectival form of *frangere*, or "to break." The concept of fractals exploded into the consciousness of mathematicians, scientists and the lay public in 1983, when Mandelbrot's ground-breaking book, *The Fractal Geometry of Nature*, was published.

Fractals are much more than a mathematical curiosity. They offer an extremely compact method for describing objects and formations. Many structures have an underlying geometric regularity, known as scale invariance or self-similarity. If one examines these objects at different size scales, one repeatedly encounters the same fundamental elements. The repetitive pattern defines the fractional, or fractal, dimension of the structure. Fractal geometry seems to describe natural shapes and forms more gracefully and succinctly than does Euclidean geometry.

Scale invariance has a noteworthy parallel in contemporary chaos theory, which reveals that many phenomena, even though they follow strict deterministic rules, are in principle unpredictable. Chaotic events, such as turbulence in the atmosphere or the beating of a human heart, show similar patterns of variation on different time scales, much as scale-invariant objects show similar structural patterns on different spatial scales. The correspondence between fractals and chaos is no accident. Rather it is a symptom of a deep-rooted relation: fractal geometry is the geometry of chaos.

Another parallel between fractal geometry and chaos theory lies in the fact that recent discoveries in both fields have been made possible by powerful modern computers. This development challenges the traditional con-

HARTMUT JÜRGENS, HEINZ-OTTO PEITGEN and DIETMAR SAUPE collaborate on the mathematics of complex dynamical systems, fractals and computer graphics at the Institute for Dynamical Systems at the University of Bremen. They have created a laboratory for experimental mathematics within the institute. Peitgen received a Ph.D. in mathematics from the University of Bonn in 1973. He is professor of mathematics at the University of Bremen and also visiting professor at the University of California, Santa Cruz. His work has focused on nonlinear analysis and differential equations, numerical methods, dynamical systems and fractals. Saupe and Jürgens earned their Ph.D.'s in mathematics at the University of Bremen in 1982 and 1983, respectively. Saupe is assistant professor of mathematics. Jürgens is director of the group's laboratory.

THREE-DIMENSIONAL renderings of the Mandelbrot set have been used to study this fascinating, complex fractal structure. The map shows the electric potential surrounding an electrically charged Mandelbrot set. The eerie similarity between the Mandelbrot set and features in the real world points to the prevalence of fractallike structures throughout nature. The image is taken from a computer-animated videotape produced by the authors and their colleagues.

ception of mathematics. Many mathematicians have greeted computers with a sense of rejuvenation and liberation, but others view them as a rejection of pure mathematics.

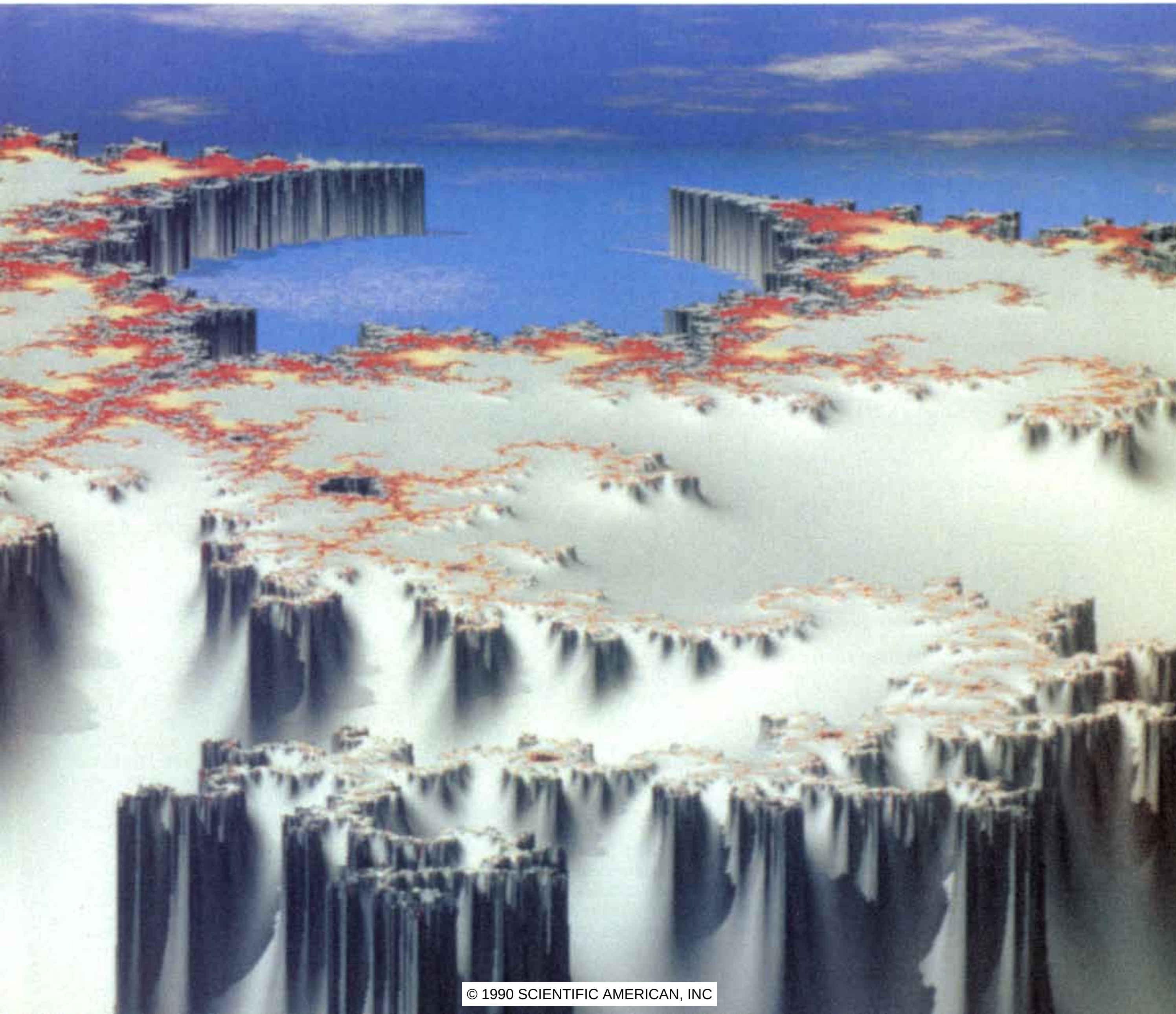
Fractals are first and foremost a language of geometry. Yet their most basic elements cannot be viewed directly. In this aspect they differ fundamentally from the familiar elements of Euclidean geometry, such as the line and circle. Fractals are expressed not in primary shapes but in algorithms, sets of mathematical procedures. These algorithms are translated into geometric forms with the aid of a computer. The supply of algorithmic elements is inexhaustibly large. Once one has a command of the fractal language, one can describe the shape of a cloud as precisely and simply as an architect might describe a house with blueprints that use the language of traditional geometry.

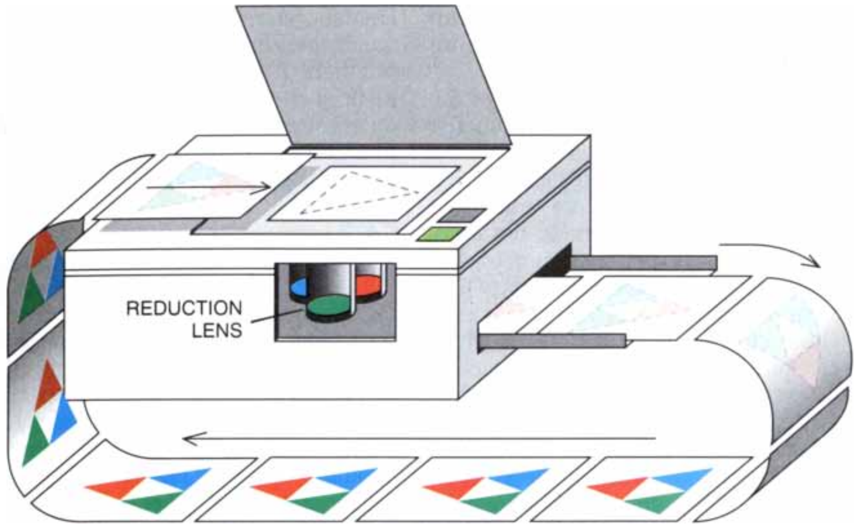
Language is an apt metaphor for the ideas that underlie fractal geometry. Indo-European languages are based on a finite alphabet (the 26 letters from which English words are constructed, for instance). Letters do not carry meaning unless they are strung together into words. Euclidean geometry likewise consists of only a few elements (line, circle and so on) from which complex objects can be constructed. These objects, in a sense, only then have geometric meaning.

Asian languages such as Mandarin Chinese are made up of symbols that themselves embody meaning. The number of possible symbols or elements in these languages is arbitrarily large and can be considered infinite. Fractal geometry is constructed in much the same way. It is made up of infinitely many elements, each complete and unique. The geometric elements are defined by algorithms, which function as units of "meaning" in the fractal language.

There are two main fractal language groups: linear and nonlinear. Both are spoken using an infinite number of algorithms and therefore encompass an infinite number of possible fractal images. The language of nonlinear fractals is far richer and more varied, however. Most dialects follow a deterministic set of rules (analogous to spelling and grammar). One family of fractals, called random fractals, differs in that it is assembled by incorporating controlled randomness.

Linear fractal geometry is the most

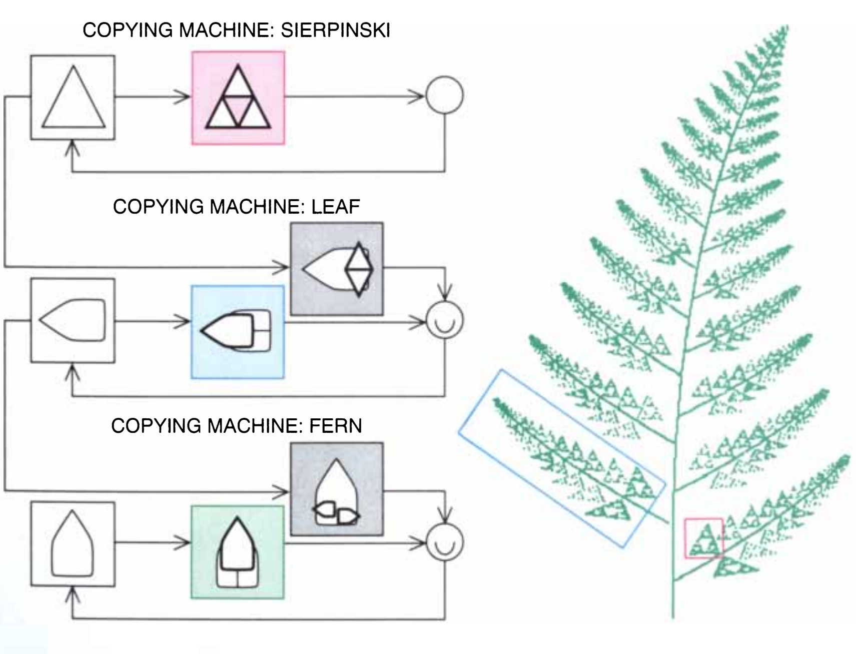




MULTIPLE-REDUCTION COPYING MACHINE performs a feedback loop that creates a fractal form. Several lenses transform an arbitrary initial image (input) into a new image (output) that is a collage of reduced copies of the input. The output image is itself run through the machine over and over, producing a final image.

basic dialect of the fractal language. These fractals are described as linear because their algorithms have the same form as those that define lines in a plane (in mathematical jargon, they incorporate only first-order terms).

The linear algorithms can be explored with the aid of an imaginary image duplicator—the multiple-reduction copying machine [*see illustration above*]. This is a metaphor for the beautiful work of John E. Hutchinson, a mathematician at the Australian National University in Canberra. The machine functions much like a normal copier that has a reduction option but differs in that it has several reduction lenses, each of which can copy the original image placed on the machine. The lenses can be set for different reduction factors, and the reduced images can be positioned at any desired location. Thus, the image can be moved, stretched, shrunk, reflected, rotated or transformed in any way, as long as the straight lines of the image remain straight.



NETWORKED COPYING MACHINE can create composite fractal images such as this fern composed of Sierpinski triangles. Several machines are linked together in parallel; one creates Sierpinski triangles, the second arranges the triangles into leaves and the third produces the overall fern shape (*left*). Note that leaves branch alternatingly off the main stem; the triangles on the leaves stand opposite (*right*).

The manner in which the image is repositioned and reduced is determined by the algorithm. A feedback loop processes the image over and over, gradually building up a fractal form. One example of a fractal produced by a feedback (recursive) algorithm is the Sierpinski triangle, named after the Polish mathematician Waclaw Sierpinski, who first described it in 1916. The Sierpinski triangle is self-similar: every part of the figure, no matter how small, contains an image that can be magnified to produce the entire Sierpinski triangle.

The Sierpinski triangle is created in a multiple-reduction copying machine in the following manner. An image is placed on the machine, reduced by one half and copied three times, once onto each vertex of an equilateral triangle. The result is a triad configuration. When the procedure is repeated, the previous triad image is again reduced by one half and copied three times and so on. After just six copies, or iterations, a final shape begins to appear. This shape is called the limiting image because it is the limiting result of an infinite number of cycles of the copying machine. The limiting image can be quickly approximated but can never be completely achieved.

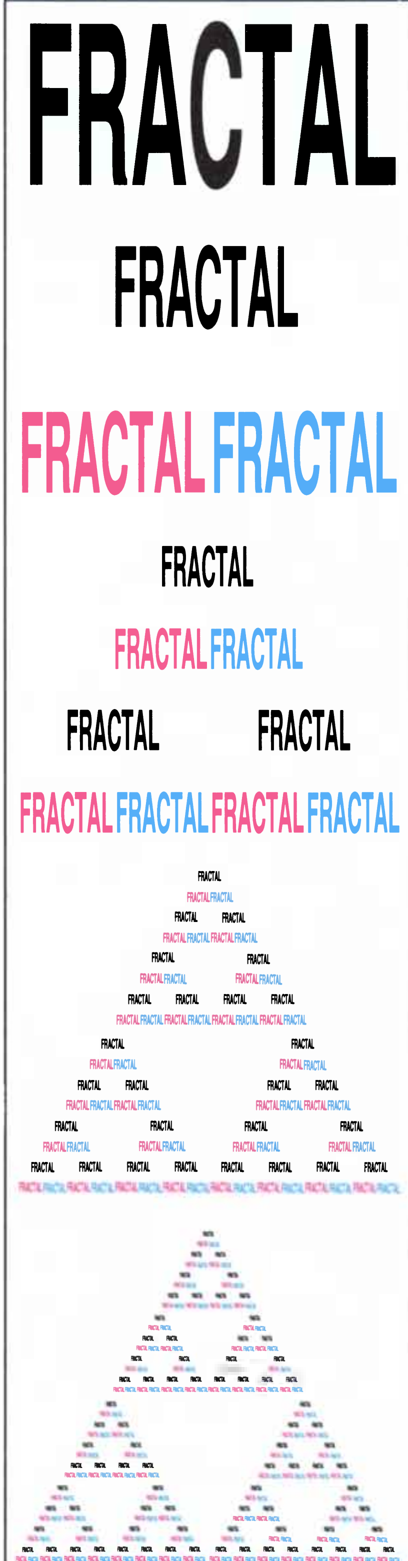
The limiting image is independent of the initial image. A distinctive initial image, such as the word FRACTAL, can be placed on the copying machine. After six copying runs in the machine the initial image is nearly invisible, and the Sierpinski triangle form dominates. Traces of the initial FRACTAL become increasingly obscure with each copy made.

Small reconfigurations of the copying machine can produce entirely different limiting images: a fractal tree or a frond-shaped fractal [*see illustration on opposite page*]. The limiting image depends only on the reduction and displacement rules (algorithms) programmed into the machine.

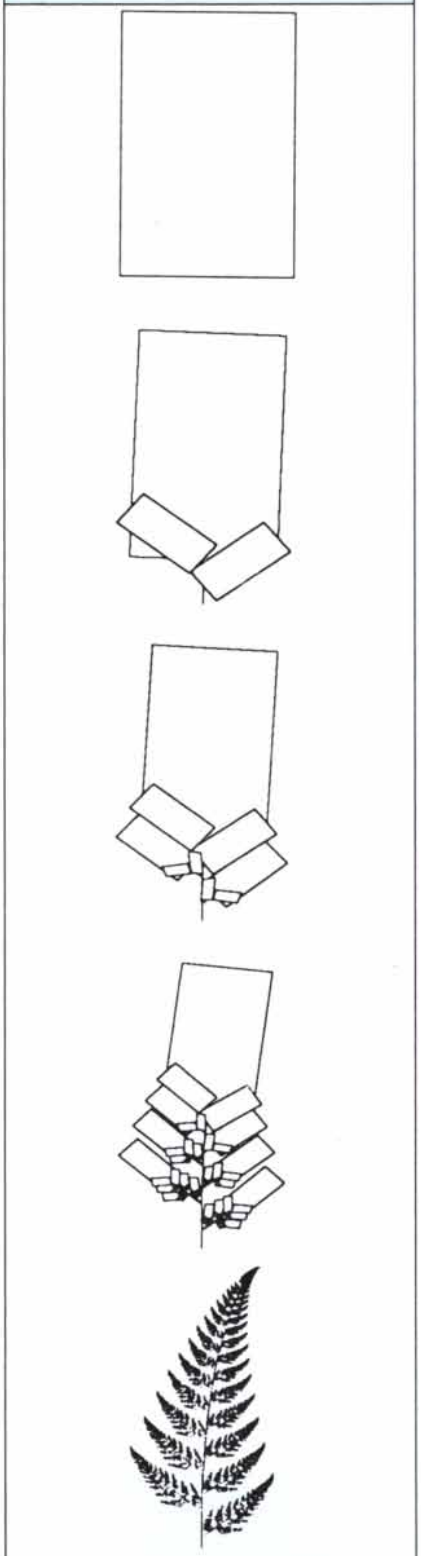
These rules are part of the general concept that mathematicians call affine-linear plane transformations: transformations that maintain the straightness of lines but alter their positions, scale and overall orientation. The rules for a linear fractal dialect can be completely described by a number (n) of transformation functions, denoted as $\{f_1, f_2, \ldots f_n\}$ [*see top of illustration on opposite page*].

Herein lies one of the great practical potentials of fractal geometry. Describing suitable objects by means of

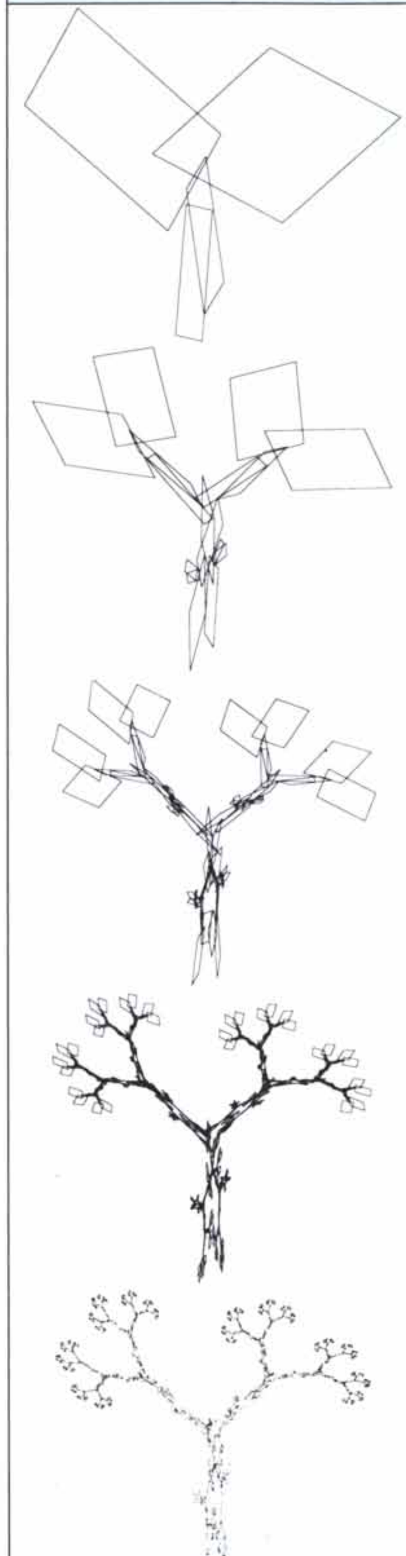
a_{11}	a_{12}	a_{21}	a_{22}	b_1	b_2
0.5	0.0	0.0	0.5	0.0	0.0
0.5	0.0	0.0	0.5	0.5	0.0
0.5	0.0	0.0	0.5	0.25	0.5



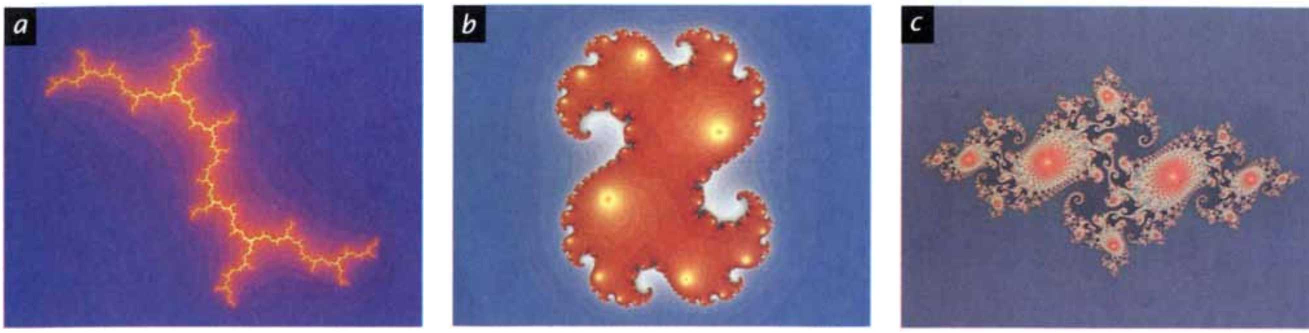
a_{11}	a_{12}	a_{21}	a_{22}	b_1	b_2
0.0	0.0	0.0	0.17	0.0	0.0
0.84962	0.0255	−0.0255	0.84962	0.0	3.0
−0.1554	0.235	0.19583	0.18648	0.0	1.2
0.1554	−0.235	0.19583	0.18648	0.0	3.0



a_{11}	a_{12}	a_{21}	a_{22}	b_1	b_2
0.195	−0.488	0.344	0.443	0.722	0.536
0.462	0.414	−0.252	0.361	0.538	1.167
−0.058	−0.070	0.453	−0.111	1.125	0.185
−0.045	0.091	−0.469	−0.022	0.863	0.871



FRACTAL IMAGES from the feedback loop of the copying machine depend only on the programmed copying routine. The word FRACTAL is transformed by a program that reduces an image to one half its size and copies it three times, once at each corner of an equilateral triangle. The resulting image is known as a Sierpinski triangle (*left*). Slightly more elaborate transformations of this kind result in a fern-shaped fractal (*center*) or a fractal tree (*right*). Any initial image would produce the same result when put through the copying machine. A few numbers defining the copying rules (*top*) are sufficient to describe an image that would require hundreds of thousands of numbers to describe by conventional means.



JULIA SETS are fractal boundaries that emerge from the iteration of the quadratic transformation $z^2 + c$. They take on a bewildering variety of shapes that depend only on the number c, called a control parameter. Some values of c lead to Julia

a linear fractal dialect can significantly reduce the amount of data necessary to transmit or store an image. This is convincingly demonstrated by the fern frond. A complex form like a fern can be completely described by a linear algorithm that is based on only 24 numbers! In contrast, representing the image of the leaf point for point at television-image quality would require several hundred thousand numerical values. In principle, any image may be coded using the appropriate set of linear transformation functions.

The time, complexity and cost of transmitting satellite images to the earth could be drastically reduced by converting them into codes using fractal algorithms. Such a possibility raises a crucial and still largely unsolved problem. How does one obtain the smallest possible family of transformation functions $\{f_1, \ldots f_n\}$ necessary to define an image to the desired precision? The problem currently is the focus of intense study. More general applications of such a procedure would include codes to create halftone or even color images.

Fractal-image coding is only useful if some efficient method exists for extracting the image locked away in the fractal algorithms. The fractal fern offers a good opportunity to examine how the image is produced. The copying-machine rules for this fractal specify that every transformation results in four reductions and repositionings of the previous image. One transformation performs a particularly rapid reduction that squeezes the image into a vertical line; this line forms the stem.

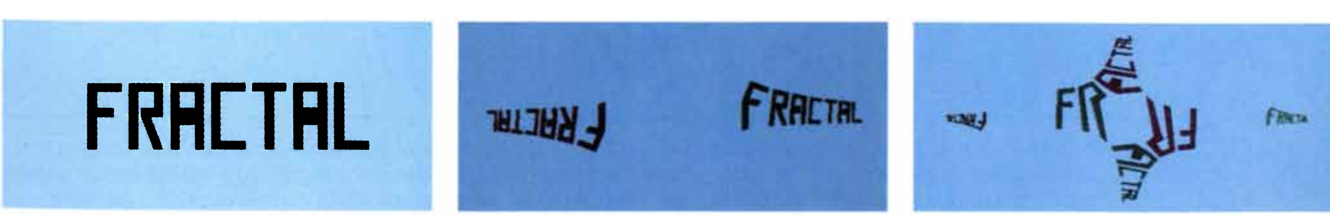
If one begins with a single rectangle, the number of rectangles increases fourfold from copy to copy, totaling 4^m after m transformations. After four iterations the initial image (in this case, a rectangle) is still easily identifiable. For the rectangle to be small enough so that the limiting image (the fern) is visible, one would have to perform roughly 50 iterations and therefore calculate and draw 4^{50} (approximately 10^{30}) rectangles. The task would exceed the capacity of any existing computer.

Given this difficulty, one might wonder how these limiting images are produced. The trick that makes the creation of these images possible is an algorithm we call the chaos game, which was suggested by Michael E. Barnsley and Stephen Demko of the Georgia Institute of Technology. The game begins with the selection of an arbitrary point on a plane. Next a four-sided die, each side of which corresponds to one of the four transformations that create the fern figure, is thrown. The roll of the die randomly calls up one of the transformations $\{f_1, f_2, f_3, f_4\}$, which is then applied to the marked point, moving it to a new point on the plane. Another roll selects another transformation, which is applied to the point previously obtained and so on. The points produced by successive throws of the die soon settle down and densely fill up the limiting image. The problem with this technique is that it may take an extremely long time to obtain this image.

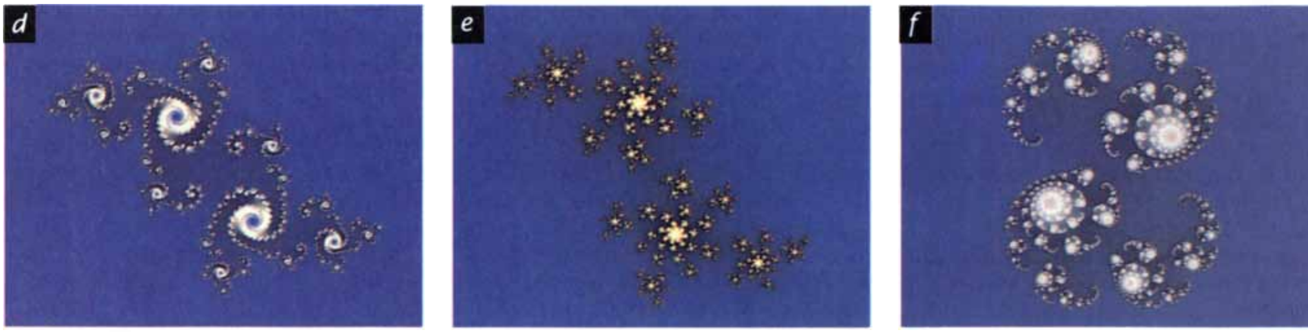
In the above example the roll of the die gives equal probability to each f_k (k simply represents one of the possible functions). The limiting image can be obtained much faster if each f_k is assigned a probability P_k with which it is thrown in the chaos game so that certain f_k's become more likely than others. An image can be formed fastest by assigning the greatest probability to the functions that reduce the image the least. This alteration causes the chaos game to hit each spot of the limiting image equally often, with the result that all parts of the image fill in equally quickly.

Altering the chaos game makes it possible to describe halftones simply by linking the frequency with which an image point is hit to a gray-scale value. Through the appropriate choice of P_k, a desired gray-scale value (in other words, the desired frequency of hits) can be obtained for every image point. Applying this technique to the additive primary colors (red, green and blue) permits color images to be coded. Thus, the value of fractal data compression is enhanced even further.

So far there is no satisfactory method for automatically generating fractal encodings of a given picture or image. For self-similar images such as the Barnsley fern there is a semiautomatic procedure that involves interaction be-



NONLINEAR FRACTALS such as Julia sets also can be created by an appropriately designed multiple-reduction copying machine. The lenses no longer perform simple reductions; they now bend and distort the images in addition to shrinking them.



sets that are in one piece (*left*). Other values yield Julia sets that are disconnected and dustlike (*right*). The Mandelbrot set consists of all points *c* associated with connected Julia sets; it also functions as a table of contents of Julia sets.

tween computer and viewer. First one breaks down the image into parts that are similar to the whole image. In the case of the fern the two lower leaflets are similar in form to the whole, as is the upper part of the fern that is left when the lower leaflets are removed. A multiple-reduction copying machine can be designed to incorporate transformations that reduce the whole image into these parts. This can be done easily by trial and error using an interactive computer program.

The method's underlying concept suggests that only strictly self-similar images can be coded in fractal form. This restriction can be overcome by means of a promising extension of the method, which is currently being researched. The central idea is to operate several copying machines simultaneously in parallel in a hierarchical network. Such a network can control individual self-similar features or mix more than one. It becomes possible, for instance, to create a fernlike leaf composed of Sierpinski triangles [*see illustration on page 62*].

We now turn to a different set of fractal dialects, the nonlinear dialects. One of these, the quadratic dialect, has garnered particular attention. It produces a great wealth of geometric forms from a fairly simple algorithm, and it is closely related to current chaos theory.

The theory behind the quadratic dialect was first described in 1918 by the French mathematician Gaston Julia while he was in a military hospital recovering from injuries suffered during World War I. Both his work and the contemporaneous work of his ardent competitor, Pierre Fatou, were soon all but forgotten, but recently Mandelbrot's work has revitalized interest in their theory. The intellectual achievements of Julia and Fatou are particularly remarkable because these men had no computers at their disposal and therefore had to rely solely on their own inherent powers of visualization.

Julia and Fatou were interested in complex numbers, each of which consists of a real number and a multiple of i, the imaginary number defined as the square root of -1. Complex numbers are commonly plotted in a plane along two perpendicular axes, one of which represents real numbers, the other imaginary ones. The two men were trying to resolve what happens to a sequence of points z_k in the complex-number plane that are generated by the transformation $g(z) = z^2 + c$. A new point, z_{k+1}, is obtained by putting the preceding point, z_k, through the transformation. The complex number c is a control parameter that can be chosen at will. This seemingly simple feedback process forms the basis for a dazzling array of shapes.

When an initial point z_0 is put through the transformation, the resulting sequence behaves one of two ways. Either it roams freely, increasing toward infinity, or it is trapped within a certain region of the complex-number plane. The unfettered points are known as the escape set; those that remain confined are known as the prisoner set. An initial point z_0 chosen from the prisoner set generates a sequence that remains in a numerical prison no matter how many generations of the sequence are computed. The shape of the prison depends on the value of c that is chosen. For a point z_0 outside the prisoner set, the sequence z_k moves away from the center of the plane and escapes toward infinity. The prisoner set and the escape set are separated by an infinitely narrow boundary known as the Julia set [*see illustration above*].

Amazingly, the Julia set can also be obtained using the multiple-reduction copying machine by refitting it with special lenses that reverse the action of $g(z)$. The inversion of $g(z) = z^2 + c$ consists of two transformation functions, $f_1(u) = +(u - c)^{1/2}$ and $f_2(u) = -(u - c)^{1/2}$. (In these functions, c is the familiar control parameter, and u is the chosen input value.) These two functions can be considered the "reductions" performed by the copying machine. Repeated operations of the machine then cause randomly chosen points to move toward the Julia set.

The presence of the square root in the equations means that the copying machine no longer applies a uniform reduction factor. Moreover, because this transformation is nonlinear, straight lines are now imaged onto curved lines. From an initial image two

Two lens systems graphically reverse the quadratic transformation that defines the Julia set. The lenses perform the transformations $+(z - c)^{1/2}$ and $-(z - c)^{1/2}$, the inverses of $z^2 + c$. The limiting image of the copying machine is a Julia set.

smaller images emerge, then four, then eight, until the limiting image gradually takes shape [*see bottom illustration on preceding two pages*]. As with linear fractals, the limiting image does not depend on the particular initial image but is completely determined by f_1 and f_2 or, equivalently, by the choice of parameter c.

Now comes one of the most difficult but fascinating problems of fractal geometry. Returning to the metaphor of language, the problem can be expressed as a question: What are the grammatical rules of the quadratic dialects? In mathematical terms the question is, Does an order principle underlie the infinite variety of Julia sets?

The search for the answer has led to one of the most beautiful discoveries of experimental mathematics. The solution lies in the fact, known to Julia and Fatou, that for every control parameter c the resulting fractal image falls into one of two categories. The Julia set may be a single connected piece, or it may consist of an infinite number of disconnected points, like dust.

Suppose one plots a dot for every control parameter c in the complex plane that belongs to a connected Julia set and leaves blank every c that belongs to a disconnected one. The result is the now famous Mandelbrot set, a fractal of truly remarkable richness.

Obviously, one needs to know if a given Julia set is connected in order to decide whether a point c belongs to the Mandelbrot set. One of the great successes of Julia and Fatou was their discovery that this difficult problem can be solved by a simple calculation. Consider the sequence of values of z_k produced by $g(z) = z^2 + c$ when the initial point z_0 equals zero. In this way, attention is focused on the crucial factor, the control parameter c. The resulting sequence is 0, c, $c^2 + c$, $(c^2 + c)^2 + c$,.... If this sequence does not escape toward infinity, the associated Julia set is connected, and the point c belongs to the Mandelbrot set.

Each portion of the Mandelbrot set characterizes a family of related Julia sets. For example, the heart-shaped main body of the Mandelbrot set characterizes Julia sets that look like crumpled circles. Although the Mandelbrot set is not exactly self-similar as are the Sierpinski triangle and fractal fern, it has a related characteristic: magnifying the boundary of the Mandelbrot set reveals an endless number of tiny copies of the set. The wealth of shapes and structures in the Mandelbrot set can be appreciated only when one inspects them in such minute detail.

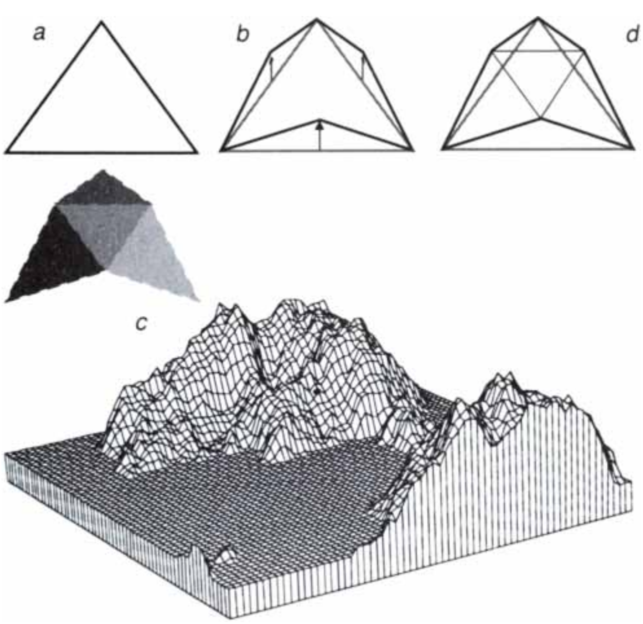
Perhaps the most fascinating aspect of the Mandelbrot set is that it functions as an endlessly efficient image storer: besides classifying Julia sets as connected or not connected, the Mandelbrot set also functions as a direct, graphic table of contents for an infinite number of Julia sets. Enlarging the Mandelbrot set around a point c on the edge of the set reveals forms that are also the building blocks of the Julia set associated with that point c. The mathematical rigor of this finding has not yet been settled, however. Tan Lei, a noteworthy young researcher now at the University of Lyon in France, has shown that the Mandelbrot set functions in this manner for most parameter values c that lie exactly on the boundary of the Mandelbrot set.

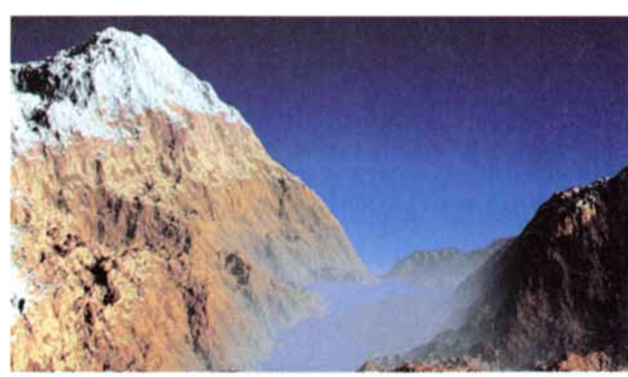
The properties of the Mandelbrot set have been and continue to be a great challenge to mathematical research. Enormous progress has been made through the fusion of mathematical theory and computer-graphics experiments, particularly in the fundamental work of Adrien Douady of the École Normale Supérieure in Paris and John H. Hubbard of Cornell University.

By far the most successful work in this area has been the so-called electrostatic potential of the Mandelbrot set. Imagine the Mandelbrot set being equipped with an electric charge. One could measure the potential by placing a point charge outside the set and measuring the electrostatic force on that point. It turns out that the computation of the potential is closely related to the series 0, c, $c^2 + c$, $(c^2 + c)^2 + c$,... that is used to determine whether or not a point c belongs to the Mandelbrot set.

Producing a three-dimensional representation of the potential has proved cumbersome, particularly in animations used to study the Mandelbrot set. Closer examination of the computer-graphics properties of the potential recently has made it possible to reduce the cost in computer time by an order of magnitude. As a result, researchers, including us, are increasingly exploring the Mandelbrot set by means of computer-animated video. Related work is also progressing on three-dimensional potential renditions of other fractals.

All fractals discussed up to this point can be considered deterministic. Although random processes (such as the roll of a die) may help generate fractal images, they do not have any impact on the final fractal form. The situation is completely dif-





FRACTAL LANDSCAPES can be created by random fractals using the midpoint-displacement method. The midpoints of the edges of a triangle (*a*) are connected by line segments and moved up or down, out of the image plane (*b*). This creates four smaller triangles on which the procedure is performed repeatedly. A distribution law assigns the amount of displacement and so defines the roughness of the fractal terrain. A computer-graphics program then fills the triangles with appropriate shading (*c*). The result can look startlingly real (*d*).

MANDELBROT SET reflects the order underlying the infinite variety of Julia sets. All points of the Mandelbrot set represent values of the parameter *c* that yield connected Julia sets. If the point *c* lies outside the Mandelbrot set, the associated Julia set is unconnected. The Mandelbrot set contains an unimaginable wealth of detail. Three zooms into the set reveal similar, repeating structures, including miniature copies of the overall Mandelbrot set, along with many new and different forms. If the entire set were shown at the scale of the image on the right, it would be the size of 100 football fields.

ferent for another class of fractals, the so-called random fractals.

One fractal of this type may begin with a triangle lying in an arbitrary plane. The midpoints of each side of the triangle are connected, breaking the triangle into four smaller triangles. Each midpoint is then shifted up or down by a randomly selected amount. The same process is applied to each of the smaller triangles, and the process is repeated ad infinitum. Over repeated iterations, an increasingly detailed surface begins to form.

In this midpoint-displacement method, the random amounts that the midpoints are raised or lowered are guided by a distribution law, which can be adjusted to obtain a close approximation of the surface to be modeled. For a model of a relatively smooth surface, the transformations should invoke a rule in which the amount of midpoint displacement becomes very small after only a few iterations. Such a rule adds only small bumps onto the overall contour. To represent a rough surface, such as the topography of a mountain range, it makes more sense to allow the displacement amounts to drop slowly with each iterative step.

This method of building surfaces has many applications. It has been used to model soil erosion and to analyze seismic patterns with a view to understanding changes in fault zones. Richard F. Voss, one of Mandelbrot's colleagues at IBM's research center, has employed the concept to create images of planets, moons, clouds and mountains that look remarkably real [*see illustration on opposite page*].

Regardless of their origin or method of construction, all fractals share one important feature: their roughness, complexity or convolutedness can be measured by a characteristic number, the fractal dimension. The various conceptual definitions of fractal dimension more or less date back to the 1919 work of Felix Hausdorff, a mathematician at the University of Bonn.

Following Mandelbrot's ideas, fractal dimension can be determined by a box-counting scheme. Imagine a complex shape that is masked with a lattice of squares plotted on graph paper. Some squares will contain part of the shape; others will be empty. The number N of nonempty squares depends on the given form and the mesh, or square size, E of the lattice. N is postulated to be proportional to $1/E^D$ (the finer the mesh, the more nonempty squares). The exponent D is the dimension. For a planar figure such as a circle, reducing the mesh by one half should multiply the number of nonempty squares by four (two squared) because the figure has a dimension of two. For a fractal, the number of nonempty squares would be multiplied by a slightly larger, or smaller, fractional value.

The above process is not restricted to mathematical objects or forms contained within a plane. One can also calculate the fractal dimension for real entities such as rivers, clouds, coastlines, trees, arteries, or villi of intestinal walls. Human arteries, for example, have a fractal dimension of about 2.7.

In addition to its usefulness for describing the complexities of natural objects, fractal geometry offers a welcome opportunity for the revitalization of mathematics education. The concepts of fractal geometry are visual and intuitive. The forms involved have a great aesthetic appeal and a wide variety of applications. Fractal geometry therefore may help to counter the perception that mathematics is dry and inaccessible and may motivate students to learn about this puzzling and exciting realm of study.

Scientists and mathematicians themselves have experienced a childlike wonder at the new and rapidly evolving language of fractals. As Mandelbrot himself wrote:

"Scientists will...be surprised and delighted to find that not a few shapes they had to call *grainy, hydralike, in between, pimply, pocky, ramified, seaweedy, strange, tangled, tortuous, wiggly, wispy, wrinkled,* and the like, can henceforth be approached in rigorous and vigorous quantitative fashion.

"Mathematicians will...be surprised and delighted to find that [fractal] sets thus far reputed exceptional...should in a sense be the rule, that constructions deemed pathological should evolve naturally from very concrete problems, and that the study of Nature should help solve old problems and yield so many new ones."

FURTHER READING

THE FRACTAL GEOMETRY OF NATURE. B. B. Mandelbrot. W. H. Freeman & Co., 1983.

THE BEAUTY OF FRACTALS. H. O. Peitgen and P. Richter. Springer-Verlag, 1986.

FRACTALS EVERYWHERE. M. Barnsley. Academic Press, Inc., 1988.

THE SCIENCE OF FRACTAL IMAGES. H. O. Peitgen and D. Saupe. Springer-Verlag, 1988.

FRACTALS FOR THE CLASSROOM. Heinz-Otto Peitgen, Hartmut Jürgens and Dietmar Saupe. Springer-Verlag, 1989.

FRACTALS: AN ANIMATED DISCUSSION, with Edward Lorenz and Benoit B. Mandelbrot. H. O. Peitgen, H. Jürgens, D. Saupe and C. Zahlten (video). W. H. Freeman & Co., 1990.

When the Melting and Freezing Points Are Not the Same

Atomic clusters can offer clues to freezing and melting. These aggregates of atoms or molecules—numbering from a few to a few hundred—can coexist as solids and liquids

by R. Stephen Berry

Almost every solid has a melting point, and almost every liquid has a freezing point. These two points are one and the same, just viewed from different perspectives: ice melts at zero degrees Celsius—the highest temperature at which it can be a stable solid—whereas water freezes at zero degrees C—the lowest temperature at which it can be a stable liquid. Hardly anything could appear simpler.

But appearances can be deceiving. Small clusters of atoms or molecules are offering new clues about melting and freezing points—namely, that they are not so simple to pinpoint. These clusters—aggregates of atoms or molecules numbering from four or five to perhaps 100 or 200—can coexist as solids and liquids over a finite range of temperature and have distinctly different melting and freezing points.

It is the unique character of clusters that has allowed investigators, including myself, to probe the secrets of their freezing and melting points. Clusters are bigger than individual molecules yet smaller than bulk matter—which consists of so many atoms that the number can be treated as infinite—and so they exhibit the properties of both. Because of their intermediate size, clusters can be studied in almost as precise detail as the atoms or molecules constituting them, while simultaneously illustrating some features of bulk matter.

In the future, atomic clusters may even provide ways to make new kinds of materials and carry out new kinds of chemical reactions. Because clusters can exist in many different stable forms—as, for example, 55-atom icosohedra or 60-atom "soccer balls"—they could be condensed to form solids unlike any we now know. Materials scientists hope that it will be possible to design such materials to have certain desired microelectronic, mechanical or catalytic properties.

From the time of English scientist John Dalton in the early 19th century, when the atomic theory became generally accepted, the study of how matter behaves has fallen into two camps. Reductionists have concentrated on the properties of individual atoms and molecules. In the 1930's this focus led to nuclear physics and then to particle physics. Other researchers have dealt with bulk matter, emphasizing the properties of large aggregates of atoms or molecules. Clusters, however, offer a bridge between the two areas of research. Their exploration had to await the development of suitable experimental and theoretical techniques, which only became available in the early 1970's.

Unlike molecules, which are characterized by definite compositions and, in most cases, definite structures, clusters are not so tightly restricted. A cluster of silicon atoms, for example, may contain three or 10 or 100 atoms. Moreover, most clusters may exhibit a variety of stable structures, with one more stable than another.

Clusters differ from bulk matter not only in the variability of the number of atoms or molecules they contain but also in the number of atoms or molecules that reside on the surface. In bulk matter only a small fraction of all the atoms are found on the surface. For clusters, that fraction can be quite large. In a cluster of 55 argon atoms, for example, at least 42 atoms are, in some sense, on its surface.

The theoretical basis for different freezing and melting points in clusters is found in several concepts, in particular one called the potential well. This idea derives from generalizing the everyday experience of living with gravity. Think of the earth with its hills and valleys. The bottom of a valley—or well—is the point of lowest potential energy for any object with weight because of the attractive force of gravity. A ball placed at the top of a hill rolls to the bottom. A ball resting at the bottom stays there. A force pulls or pushes an object in the direction that reduces its potential energy.

A molecule or a cluster exactly in equilibrium can be thought of as sitting at the bottom of a potential well, but in this case electric forces, not gravity, do the pulling and pushing to reduce the potential energy. The pulling is done by the attractive force between pairs of electrons and protons of the constituent atoms, and the pushing is done by the repulsive force between pairs of protons and pairs of electrons. The interplay of pulling and pushing is what determines the diameters of atoms: the

R. STEPHEN BERRY is the James Franck Distinguished Service Professor at the University of Chicago, where he holds appointments in the department of chemistry, the James Franck Institute and the School of Public Policy Studies. Berry has studied the dynamics of atomic and molecular processes and thermodynamics, and he has worked in the area of energy efficiency and resource policy. His interest in atomic clusters grew out of a long-standing fascination with nonrigid molecules and the relation between nonrigidity and the liquid phase. He is coauthor of *Physical Chemistry* (with Stuart A. Rice and John Ross) and of *TOSCA: The Total Social Costs of Coal and Nuclear Power* (with Linda Gaines and Thomas V. Long). A MacArthur fellow (1983–1988), Berry received his Ph.D. from Harvard University in 1956 and taught at the University of Michigan and at Yale University before he joined the faculty at Chicago in 1964.

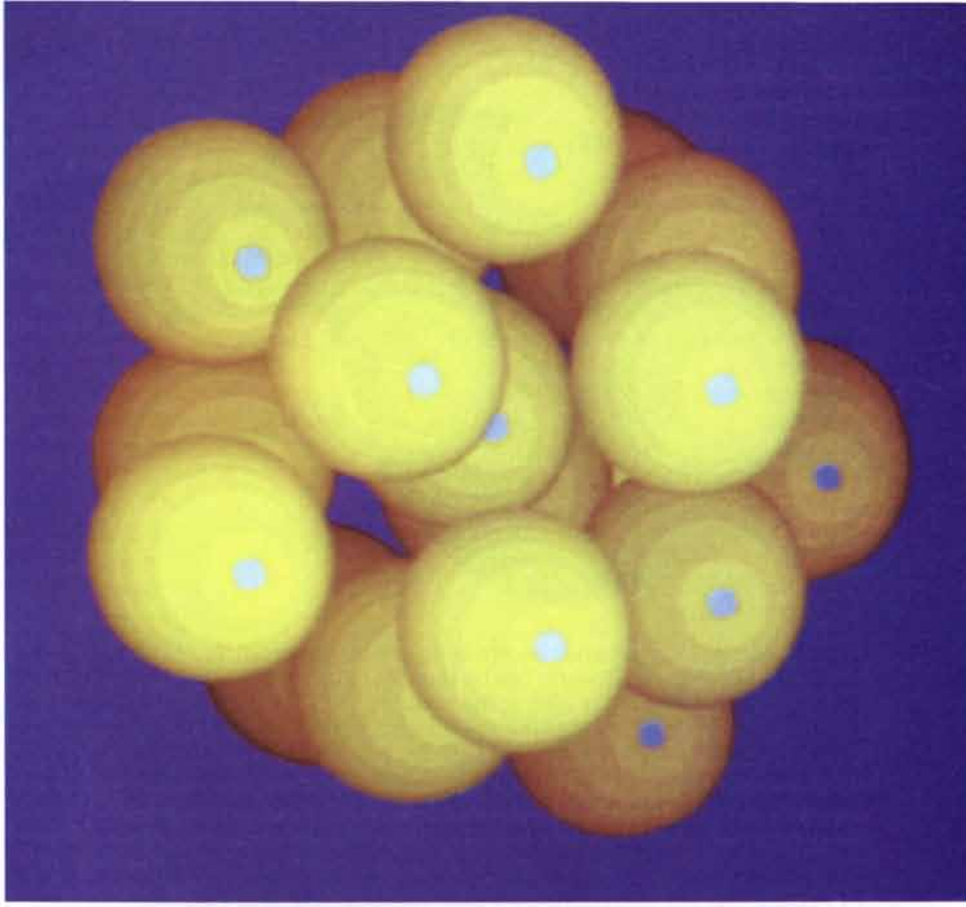
separations at which atoms "bump" into one another. Atoms are not as hard as billiard balls or even tennis balls; they are more like the soft pad of your hand below the thumb.

The potential well of two interacting atoms forming a bond has a constant value when the atoms are far apart, a decreasing value because of electron-proton attraction as the atoms approach each other, and then a sharply rising value as the atoms begin to bump one another. Although it is relatively easy to extend the concept of the potential well to an abstraction that represents the mutual interaction of three or more atoms, drawing such a well would be difficult because it has too many dimensions. In other words, the well depends on too many independent variables to be depicted on paper or even to be made into a three-dimensional model. Nevertheless, the language and imagery associated with the concept of the potential well remain useful in thinking and talking about how molecules and clusters of several atoms behave.

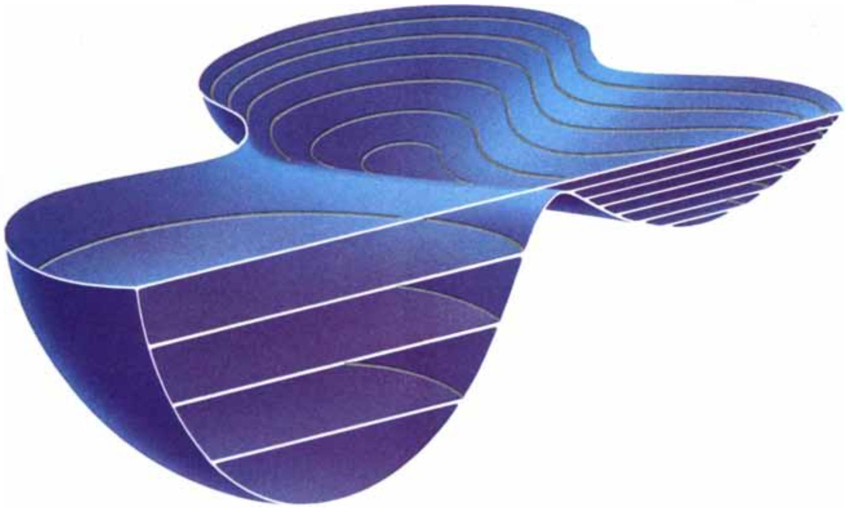
So far clusters have been described in terms of classical mechanics. The real world is better described by quantum mechanics. The most important change here is that the allowed energies of a cluster are quantized. Classically, of course, a cluster can assume any energy in its potential well, but quantum mechanically, the energy of a cluster is restricted to a discrete set of levels. The energy levels form a kind of irregular ladder in the potential well.

The uneven spacing of the rungs—or quantized energy levels—in this ladder plays an important part in causing the different melting and freezing points of clusters: a deep, steep potential well has widely spaced rungs, whereas the rungs in a broad potential well are closely spaced. Wells may be deep and steep at the center but have broad, shallow sides. In that case, the levels at low energies are widely spaced, and those at high energies are close together.

The deep wells are the homes of sol-



CLUSTER of 19 argon atoms can exist in a solid state (*top*) or a liquid state (*bottom*). Clusters in the solid state exhibit stiff, quivering vibrations, whereas those in the liquid state have no definite geometric structure and so rearrange themselves easily. The computer images were made by Thomas L. Beck, now at the University of Cincinnati.



POTENTIAL WELL is a useful concept in describing the behavior of clusters. A molecule or cluster in equilibrium can be thought of as sitting at the bottom of a potential well, just as a ball resting at the bottom of a valley remains stationary. Classically, a cluster can assume any energy in its potential well, but quantum mechanically, the energy is restricted to a discrete set of levels (horizontal planes) that form the rungs of a kind of irregular ladder in the well. Deep, steep wells have widely spaced energy levels and are the homes of solid clusters, whereas broad, shallow wells have closely spaced levels and are the ranges for liquid clusters.

id clusters. These clusters have stiff, quivering vibrations and can only rotate like Tinkertoy models made of balls and sticks. The undulating plains are the ranges of liquid clusters, which have no definite geometric structure and so rearrange themselves easily. Solid clusters have relatively widely spaced energy levels; liquid clusters have densely spaced energy levels. In the situation best suited to causing different melting and freezing temperatures, the deep, narrow valleys of the potential well are separated from its broad, undulating plains by moderately high or inaccessible passes.

The final concept needed to explain the different melting and freezing points of clusters is free energy. Free energy is defined as the energy of the system—in this case, a molecular ensemble—minus the product of its temperature and entropy, or randomness. The cluster will rearrange itself in a way that reduces its free energy, either by minimizing its energy or maximizing its entropy, or some combination thereof: matter tends toward low energy and high entropy. Free energy is simply a scorecard of the competition between energy and entropy.

Now imagine an ensemble, or collection, of clusters held at a temperature low enough so that all the clusters are in a solid state. At such a temperature the entropy of each cluster is low, because the energy levels are spaced fairly wide apart and only a few levels are occupied. In this case, the free energy of each cluster is at a minimum because the energy is low.

Suppose the temperature is raised somewhat. Each cluster now has access to higher-energy levels and thereby a means of increasing its entropy—and potentially decreasing its free energy. But with the temperature raised only slightly, the entropy gained from roaming over the higher-energy levels is offset by the energy saved from remaining at the low-lying levels. And so the clusters remain in the solid phase, venturing occasionally into the more wide-ranging states.

As the temperature is raised even further, each cluster gains increasing access to the high-energy levels of the liquid phase, levels that are more and more densely packed along the scale of increasing energy. Therefore, the randomness associated with an individual cluster goes up enormously as it moves virtually unencumbered among the levels. This time, the large gain in entropy is more than enough to counteract energy required to reach the liquidlike levels, and so a minimum in the free energy takes place in clusters in the liquid phase.

Under these conditions some startling results occur. The minimum in the free energy for clusters in the solid phase also continues to persist (albeit typically less deeply). Consequently, because there are two minima in the free energy, both solid and liquid clusters can coexist.

With further increases of temperature, the energy-entropy balance tips more in favor of the liquid clusters, and the minimum in the free energy for clusters in the solid phase becomes less and less pronounced. Finally, at a high enough temperature, that minimum disappears, leaving only the liquid-side minimum, and the energy-entropy balance tips completely in favor of the liquid clusters.

Now only liquid clusters can exist. From the first appearance of the stable liquid up to the disappearance of the stable solid, then, in a certain range of temperature, both solid and liquid clusters can coexist. Within the range of coexistence the fraction of clusters that are solid or liquid depends on the difference in the free energies of the two forms, and those free energies vary with temperature. In particular, the fraction of clusters that are liquid follows a sort of flattened S-shaped curve as the temperature is increased.

Thus, the density of energy levels and the interplay between energy and entropy have led to the conclusion that a cluster may have a distinct freezing point, below which only the solid phase is stable and a distinct melting point, above which only the liquid phase is stable. Between them lies a finite range of temperature within which both phases are stable. In other words, the melting and freezing points are not the same! The melting and freezing points have now been disconnected; nothing dictates that they need occur at the same temperature.

The unlinking of the melting point and the freezing point from each other presents an apparent paradox because in everyday experience these two values are the same. Clusters, however, reconcile this seeming contradiction through the special properties that accompany their intermediate size. The behavior of a small system—10, 100 or even 10,000 atoms or molecules—is very different from that of a large system—a trillion atoms or molecules. For a small cluster of, say, 10 to 20 atoms, the solid and liquid phases are expected to be detectable over a range of several degrees. For a "supercluster" of a million atoms the temperature range over which the solid and liquid phases can be observed to coexist is thought to be less than a thousandth of a degree. For bulk matter the limits of this range of coexistence are so close that they cannot possibly be distinguished. The indistinguishability completely justifies using the terms melting point and freezing point interchangeably for bulk matter.

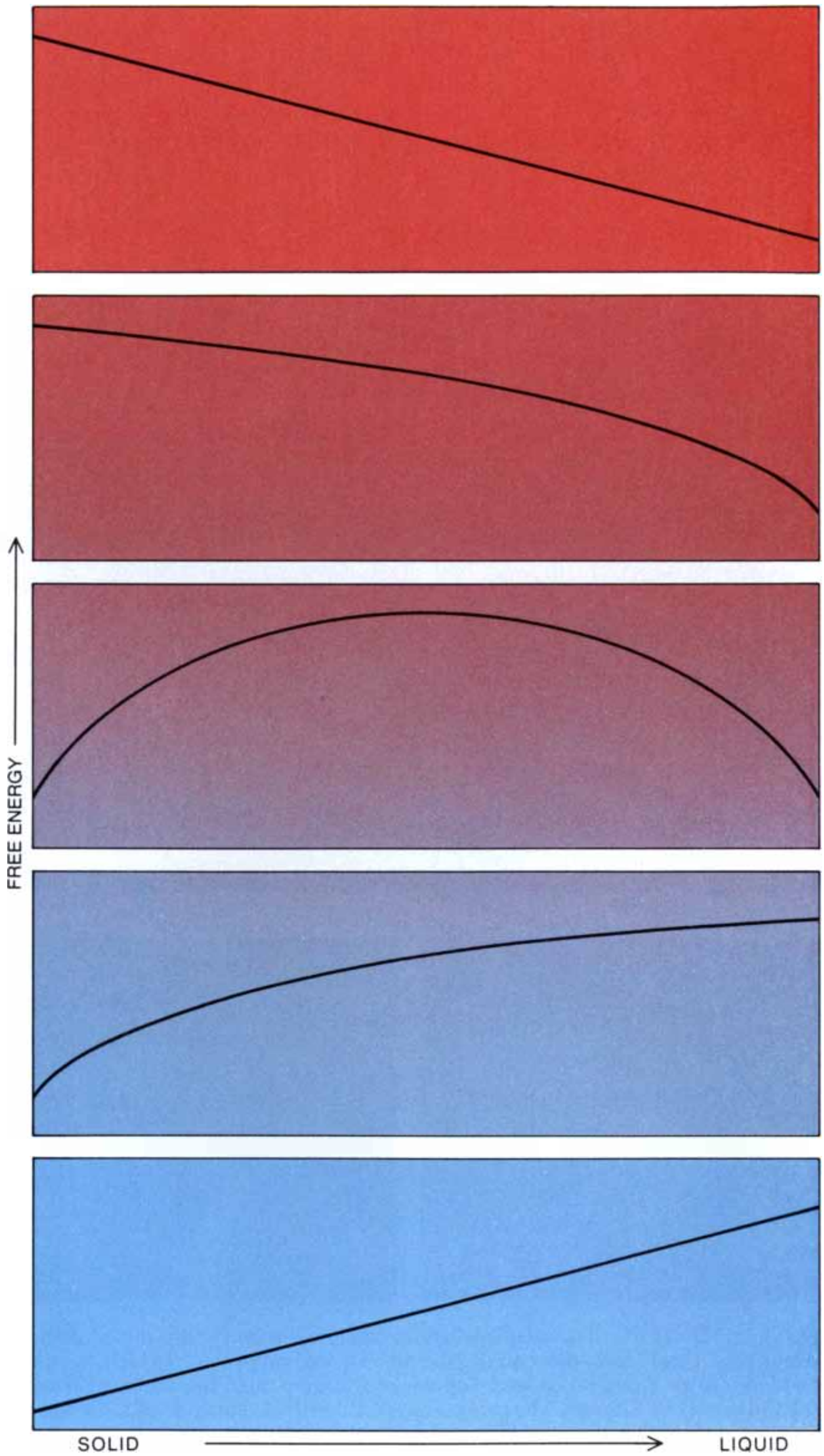
Both laboratory experiments and computer simulations provide evidence that supports the theory for the different melting and freezing points of clusters. Clusters are made in the laboratory either by generating a vapor of the component atoms or molecules and letting the components aggregate or by simply knocking the clusters out of a bulk solid. Either way the clusters can be studied as a vapor or as a trapped species in an inert matrix.

To distinguish between solid and liquid clusters, one must be able to "see" the clusters behave like solid and liquid forms of matter. Solids are stiff and uncompliant; they can be pushed, for example. Liquids, on the other hand, are soft and compliant; I cannot put my finger on water, only in water, because it yields to the slightest pressure. (Of course, if a force is applied suddenly enough, even a liquid can seem solid—as anyone who has inadvertently belly-flopped into a swimming pool well knows.)

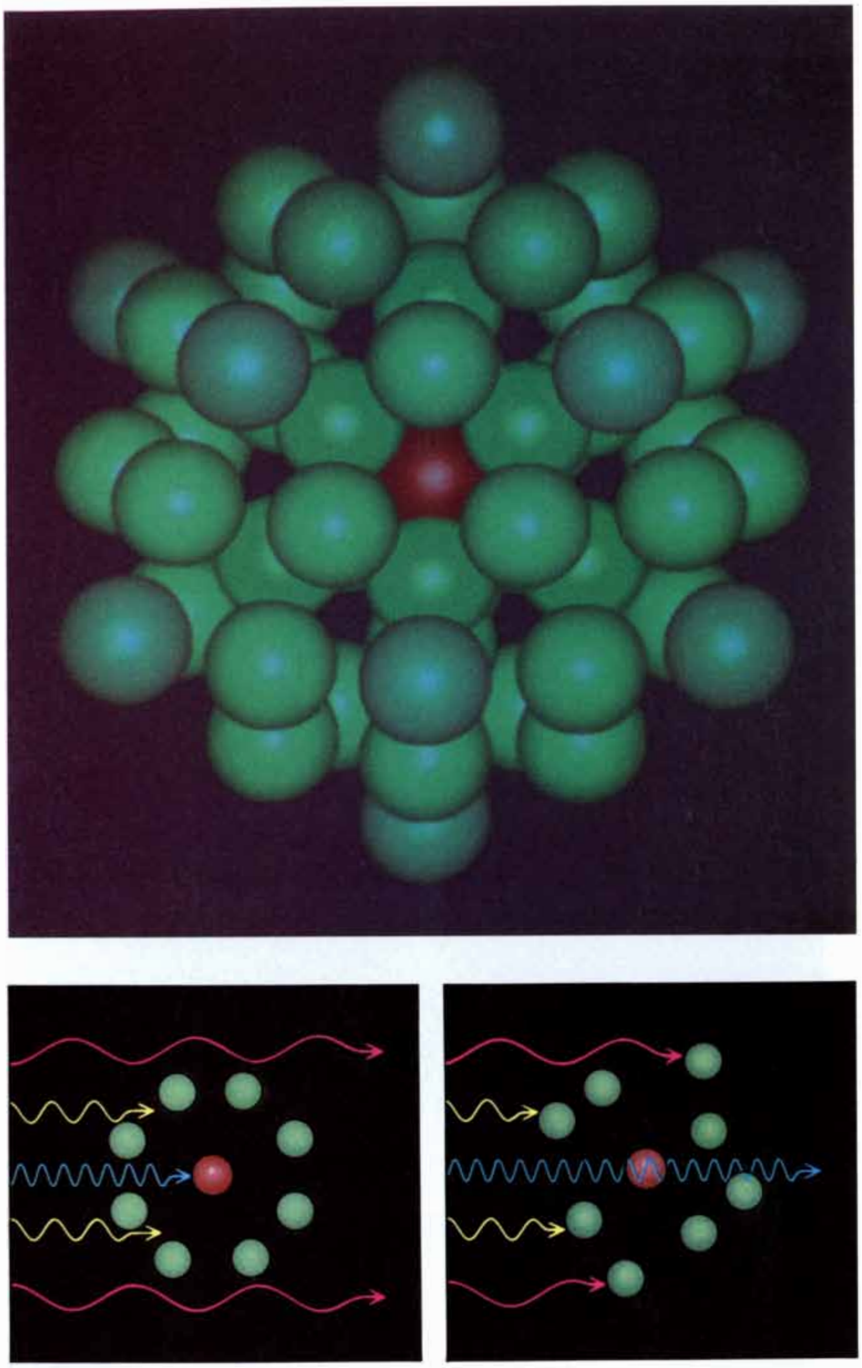
A solid cluster is uncompliant in that it is almost completely restricted to vibrating about a single geometric configuration (or particular valley). The cluster may undergo a million or even a trillion vibrations about one configuration before it passes to another configuration, if it has the energy to do so. (A typical cluster vibrates about 10 trillion times per second.) A liquid cluster is compliant in that it moves easily from one configuration to the next (from one valley to another). The cluster may undergo a few vibrations or possibly as many as a few hundred, before it passes to its next configuration.

For investigators to see solid and liquid clusters, the clusters themselves must spend a sufficient interval of time in each form. The liquid clusters must remain liquidlike long enough to display the characteristic properties of a liquid, notably the explorations of many geometric configurations through the rearrangement of atoms in a floppy way. The solid clusters must exhibit the stiffness and noncompliance characteristic of that phase.

The best laboratory evidence at this time seems to indicate that solid and liquid clusters can in fact be distinguished from one another. Most of what has been done relies on incorporating into each cluster a foreign molecule that is different from all the other atoms or molecules in the cluster. The foreign molecule serves as a probe. The



COEXISTENCE of solid and liquid clusters is illustrated over a finite temperature range. The temperature represented by each of the five panels increases from bottom to top. The vertical axis of each panel is the free energy of the clusters: the energy of the clusters minus the product of their temperature and entropy, or randomness. The horizontal axis of each panel is a measure of the nonrigidity of the clusters; solid clusters are on the left, and liquid clusters are on the right. At the coldest temperature (*bottom panel*) there is a minimum in the free energy for the solid clusters only, and at the highest temperature (*top panel*) there is a minimum for liquid clusters only. For a range of intermediate temperatures (*middle three panels*), however, there are minima for both solid and liquid clusters: the two forms coexist. In other words, the melting and freezing points are not the same.



PROBE MOLECULES embedded in clusters can allow scientists to distinguish between liquid and solid states. In the top panel a sodium probe (*red*) has been embedded into an argon 54 cluster. The image was generated by Hai-Ping Cheng of the University of Chicago. The probe emits different wavelengths of radiation depending on whether the cluster is a solid (*bottom left*) or a liquid (*bottom right*).

spectrum of the probe molecule, or the characteristic wavelengths of radiation emitted and absorbed by the molecule, depends somewhat on the surrounding atoms, as well as on the probe molecule itself. Consequently, the spectrum of a probe molecule embedded in a solid cluster should be different from that of a probe molecule embedded in a liquid cluster.

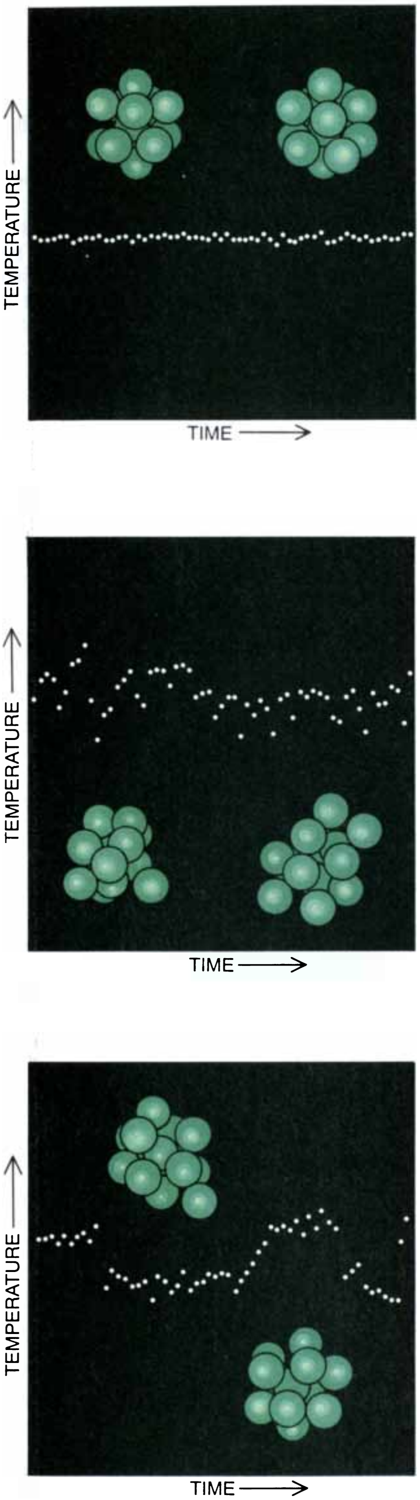
One powerful way to identify a cluster is to use a laser to excite the probe molecule to a known energy level. (The wavelength of the laser radiation needed to induce such an excitation depends sensitively on the type of probe molecule and the surrounding atoms of the cluster.) Then a second laser is tuned to a wavelength that will ionize the already excited probe molecule. That is, the second laser strips the probe molecule of one of its outermost electrons. The free electron escapes, leaving the probe molecule and the cluster in which it is embedded with a net positive charge. The remaining charge provides a way to apply a force and accelerate the cluster (by means of an electric field) and hence a way to remove the cluster from a mixture. The cluster is thereby identified according to its mass, that is, according to the number of atoms or molecules that it comprises [see "Detecting Individual Atoms and Molecules with Lasers," by Vladilen S. Letokhov; SCIENTIFIC AMERICAN, September, 1988].

Jürgen Bösiger and Samuel Leutwyler of the University of Bern and Mee Hahn and Robert L. Whetten of the University of California at Los Angeles have followed such an approach in studying clusters of argon atoms. The Swiss investigators used a large, flat probe molecule called carbazole. The California team used benzene, a smaller, pill-shaped probe. The probe molecules embedded in the solid clusters display a sharp, distinct spectrum, whereas the probe molecules embedded in the liquid clusters have a broad spectrum. (In spite of subtleties concerning the interpretation of the experimental data, the arguments for such characteristic spectra are persuasive.)

Both sets of experiments seem to show that at a temperature of about 20 to 30 kelvins (–253 to –243 degrees C) small clusters are liquidlike (having broad spectra) and large clusters are solidlike (having sharp spectra). Clusters within a range of intermediate sizes exhibit spectra indicating a mixture of coexisting solid and liquid clusters.

The data from the Swiss and California laboratories represent an important step in the experimental program to understand clusters. The coexistence of solid and liquid clusters over a finite range of size but at a fixed temperature is highly suggestive of different freezing and melting points, although the evidence is of course not completely conclusive. It would be ideal to keep the size of the clusters fixed and the temperature varied, but the precision needed to perform such an experiment is difficult to attain. Additional work will be needed to make the inference absolutely irrefutable.

Another way to test the theory is through computer simulations of clusters. Naturally, simulations are only that and not real experiments. Yet as long as there are ways to test the validity of a simulation, the approach can



COMPUTER-SIMULATED time histories of mean temperatures of argon 13 clusters at various constant energies show the coexistence of liquid and solid clusters. At a low energy (*top*) the clusters are solidlike, at a high energy (*middle*) the clusters are liquidlike and at an intermediate energy (*bottom*) the clusters behave like both solids and liquids. The temperature of a cluster is directly proportional to its kinetic energy. Each point represents 5×10^{-12} second.

reveal new phenomena and permit "experiments" that are far cheaper than their laboratory counterparts. Computer models also can help in designing laboratory experiments.

Massive computation can most appropriately be used to simulate a system that has a trustworthy theoretical model whose application is too complicated for pencil and paper—for example, the behavior of a cluster of atoms. For many substances, the forces between any two atoms are well understood, but it is impossible to solve the equations of motion even for three or more interacting atoms in analytical terms, to say nothing of more than three. The equations for the motion of several atoms (or even very many) can, however, be solved numerically, step by step, with the aid of a computer. That way, the behavior of each "atom" in the simulation can be followed.

Most of my attention here will focus on simulations of clusters of argon atoms that Thomas L. Beck, Hai-Ping Cheng, Heidi L. Davis, Julius Jellinek, David J. Wales and I have done at the University of Chicago. Of the information these simulations yield, thé easiest to understand is the history of a single cluster evolving in time. One useful quantity that can be readily extracted is the average temperature of a cluster whose total energy is constant—or, its complement, the average energy of a cluster whose temperature is constant. (The mean temperature is, apart from a numerical conversion factor, the average kinetic energy.)

What "average" signifies in these experiments is crucial. Averaging the temperature only once over an entire history can be as misleading as averaging the temperature over such a short time that each atom in the cluster has little or no time for encounters with the other atoms. Typically we take averages over about five millionths of a millionth of a second, time enough for just a few vibrations of the argon atoms.

We have found that a cluster of 13 argon atoms held at a low energy shows very small temperature fluctuations around its mean temperature. Such a cold cluster is solidlike. A much more energetic cluster has a significantly higher average temperature, and the fluctuations around that temperature are correspondingly larger. Several aspects of its behavior indicate that the cluster is in the liquid phase. If the energy of the cluster is made high enough, the simulation shows atoms "evaporating" from the cluster, as one would expect.

Between the two extremes is a range of energies for clusters at constant energy in which some clusters exhibit a remarkable phenomenon: they spend long intervals of time as solids, then more or less at random jump into the liquid form for rather long periods, then turn back into solids and so on. Other clusters that display such coexistence include argon clusters composed of seven, 15 or 19 atoms.

I should point out, however, that not all clusters give a clear indication of a two-phase coexistence. For example, argon clusters composed of six, eight or 17 atoms do not. Clusters of six copper atoms show both stiff and soft solid forms. Some large clusters of sodium chloride, according to simulations by Daphna Scharf and Joshua Jortner of Tel-Aviv University and Uzi Landman of the Georgia Institute of Technology, can exhibit solid and liquid regions simultaneously on different sides of the same cluster. In simulations done by Cheng, the surfaces of argon clusters composed of 52 or 55 atoms melt at temperatures well below the melting temperatures of the clusters themselves—a finding that illustrates the importance of the surface characteristics of clusters as opposed to those of bulk matter.

All in all, small clusters may exhibit a variety of behavior as they are given energy and become floppier and less rigid. Certainly this richness will serve as a valuable tool in probing still deeper the phenomena of melting and freezing.

FURTHER READING

THE THERMODYNAMICS OF SMALL SYSTEMS. T. L. Hill. Parts 1 and 2. Benjamin Co., Inc., 1963.

LARGE FINITE SYSTEMS in *Proceedings of the Jerusalem Symposia on Quantum Chemistry and Biochemistry*, Vol. 20. Edited by Joshua Jortner, Alberte Pullman and Bernard Pullman. Reidel Publishing Co., 1987.

MICROCLUSTERS. Edited by S. Sugano, Y. Nishina and S. Ohnishi. Springer-Verlag, 1987.

PHYSICS AND CHEMISTRY OF SMALL CLUSTERS. Edited by P. Jena, B. K. Rao and S. N. Khanna. Plenum Press, 1987.

ELEMENTAL AND MOLECULAR CLUSTERS. Edited by G. Benedek and Martin Pacchioni. Springer-Verlag, 1988.

EVOLUTION OF SIZE EFFECTS IN CHEMICAL DYNAMICS, Part 2 in *Advances in Chemical Physics*, Vol. 70. Edited by I. Prigogine and Stuart A. Rice. Wiley-Interscience, 1988.

THE CHEMICAL PHYSICS OF ATOMIC AND MOLECULAR CLUSTERS in *Proceedings of the CVII Course of the International School of Physics "Enrico Fermi,"* Varenna, Vol. 289. Edited by Giacinto Scoles and S. Stringari. Societá Italiana di Fisica (in press).

Seed Dispersal by Ants

Thousands of plant species rely on ants to disperse their seeds. With special food lures and other adaptations, a plant can induce the insects to carry away its seeds without harming them

by Steven N. Handel and Andrew J. Beattie

To distribute their seeds far and wide, plants often enlist the help of animals. Some produce seeds that temporarily attach themselves to the hairs of mammals or the feathers of birds. Commercial orchard trees and some other plants have seeds that fruit-eating animals ingest and that germinate when excreted or regurgitated. Vertebrates are not the only animals that disperse seeds, however—ants, too, are important.

Biologists are just beginning to recognize the specialized mechanisms that make ants a major force in the spread of plants around the world. Plants dispersed by ants are, in fact, found in diverse habitats on every continent except Antarctica. Currently more than 3,000 species belonging to more than 60 families of flowering plants are known to be distributed in this way, and the list is growing.

Between many plants and the ants that transport their seeds exists a true mutualism, or mutually beneficial relationship. This mutualism arose independently in so many plant groups that there must be strong, frequently recurring pressures that favor it. This article will concentrate on the natural selection and evolution of this mutualism and its ecological advantages.

Ants disperse seeds through two general mechanisms. One depends on inefficiencies in the activities of harvester ants, which gather large quantities of seeds, transport them to their nests and then eat them. Some seeds are dropped or lost along the way, and others may be cached in the soil and later ignored by the ant workers. These seeds may germinate and establish plants at a new site.

Because the ants eat more seeds than they drop or misplace, this dispersal mechanism seems to benefit the ants much more than the plants, which lose a large fraction of their seeds. Seed dispersal by harvester ants is therefore best described as an outcome of seed predation, not as a mutualism. Its impact is limited largely to arid habitats.

The second dispersal mechanism, and the one that interests us, is quite different and has a far wider significance. It involves plants that produce an elaiosome, a fat body that is near or attached to a seed. Elaiosomes lure the ants, which carry both the seed and its elaiosome back to their nest. There the colony eats the elaiosome and discards the seed unharmed.

This interaction does not sacrifice any seeds to ant predation. Termed myrmecochory, from the Greek for "ant" (*myrmēx*) and "dispersal" (*kōrē*), it seems to be a true mutualism—that is, it benefits both the transporting ants and the plants that produce the elaiosome.

Elaiosomes have repeatedly emerged as lures for ants during the evolution of diverse taxonomic families of plants. They are commonly found among plants in the moist woodlands of eastern North America and of Europe, in the dry shrub communities of eastern Australia and in plant communities of southern Africa.

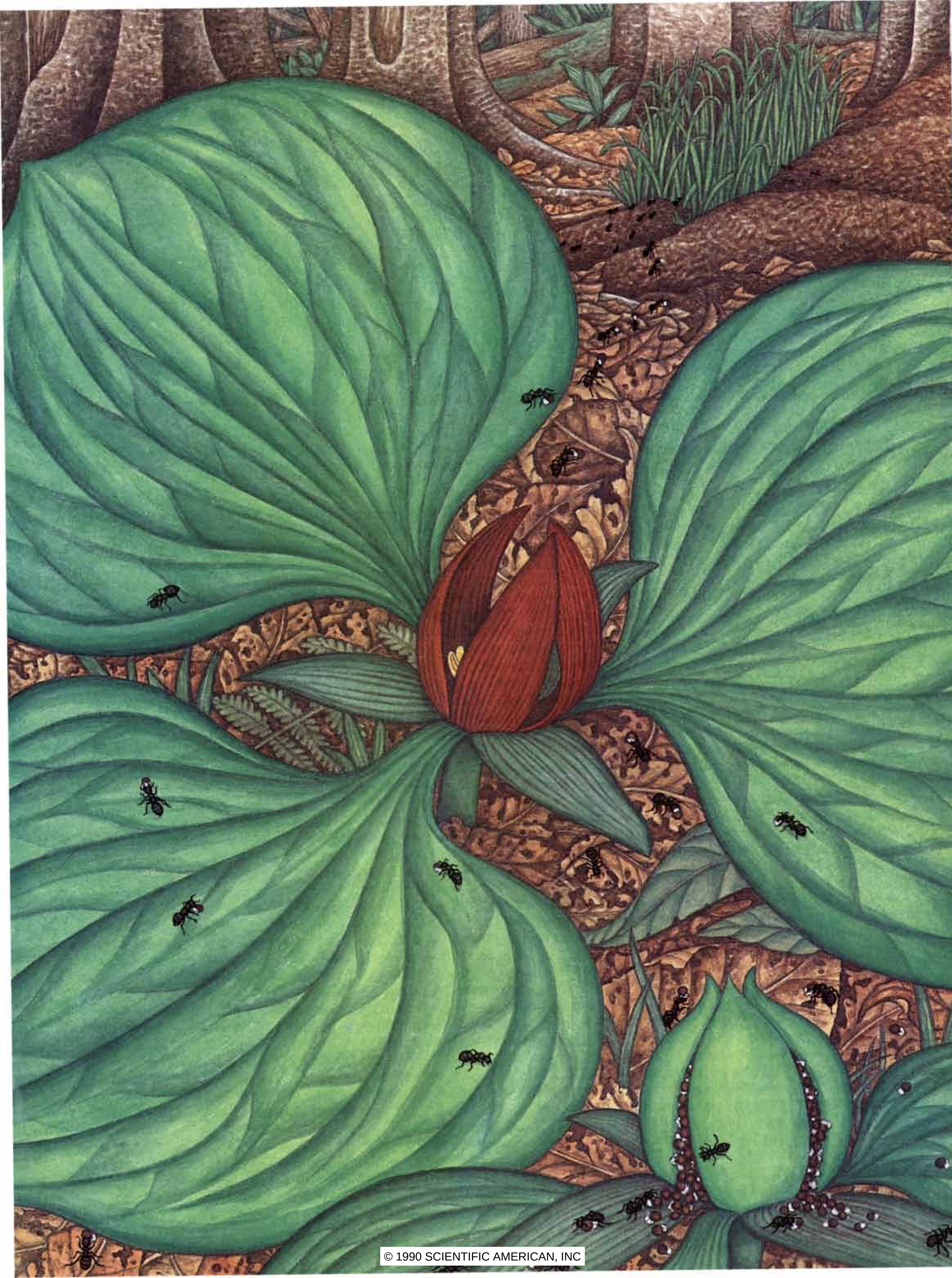
Most often only some species in each family are dispersed by ants. In the large genus of sedge called *Carex*, for example, only a handful of species bear elaiosomes that have been shown to elicit dispersal by ants. Many other species in the genus are dispersed by water or by vertebrates. Similarly, in the large-flowered woodland genus *Trillium*, some species have elaiosomes on their seeds and are dispersed by ants, but others have fleshy fruits and are dispersed by vertebrates. As these examples drawn from highly different phylogenetic groups illustrate, dispersal of seeds by ants can appear independently within a genus.

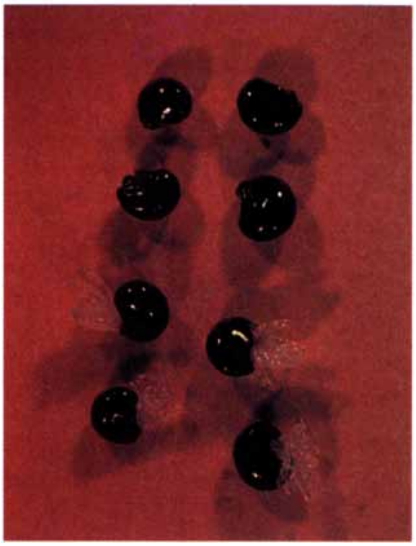
Myrmecochory was first studied comprehensively by Johan Rutger Sernander, a botanist at the University of Uppsala in Sweden, who published a survey of European myrmecochores in 1906. His experimental, quantitative approach was remarkable, and his book established the importance of myrmecochory to much of European vegetation. Sernander performed a long series of field experiments with many plant species to show that when offered a variety of seeds, ants prefer to carry away those with elaiosomes.

Although myrmecochores were first investigated in Europe, botanists soon explored vegetation in other continents as well. North and South American plants were slowly added to the list of

STEVEN N. HANDEL and ANDREW J. BEATTIE share an interest in plant-animal interactions. Handel, associate professor at Rutgers University and associate editor of the journal *Evolution*, received his doctorate from Cornell University in 1976. He has taught at the University of South Carolina, Yale University, the Rocky Mountain Biological Laboratory in Colorado and the University of Virginia's Mountain Lake Station. From 1983 to 1985 he was chair of the genetics section of the Botanical Society of America. Beattie became professor at Macquarie University in Sydney, Australia, in 1987. Previously he had chaired the department of ecology and evolutionary biology at Northwestern University. He has also directed the Rocky Mountain Biological Laboratory and has been a visiting research fellow at the University of Melbourne. He earned his doctorate from the University of Liverpool.

WILDFLOWER *Trillium petiolatum* has evolved with a distinctive morphology that encourages ants to disperse its seeds. The plant grows its flower unusually close to the ground, which makes it easy for ants to reach the seeds. As an added inducement, the seeds bear fat bodies called elaiosomes that ants find attractive. Ants carry elaiosome-bearing seeds to their nests, where they eat the elaiosomes and discard the seeds, which later germinate.





DUTCHMAN'S BREECHES (*left*) is a woodland herb that drops elaiosome-bearing seeds from seed pods (*center*) in the late spring. Four seeds with elaiosomes are seen in a close-up, along with four other seeds taken from an ant nest's garbage dump (*right*). Although their elaiosomes have been eaten by ants, the seeds can germinate in the nutrient-rich soil.

known myrmecochores. In Europe and North America most myrmecochores are herbs found in moist, deciduous forests. (Sernander was the first to notice this correlation.) In Latin America herbs, epiphytes and vines of the tropical rain forest are frequently dispersed by ants.

Myrmecochores are also particularly abundant in Australia and southern Africa, where they typically are tough-leaved shrubs growing in arid soil with few nutrients. In 1975 Rolf Y. Berg of the University of Oslo published his findings that approximately 1,500 species of Australian plants belonging to 87 genera were dispersed by ants. Vegetation called fynbos in South Africa also contains more than 1,000 such species. Tropical research now in progress will undoubtedly add many more to the list.

The taxonomic diversity of plants that have elaiosomes is mirrored by the wide variety of plant tissues that evolution has modified into ant-attracting bodies. In some species, such as Dutchman's Breeches, a section of tissue on the seed coat has been expanded and modified into an elaiosome. In others, including the spring flowering hepaticas of eastern North America, the elaiosome is derived from a portion of the ovary wall that surrounds the seed. Elaiosomes arise from the bract tissue that envelopes the ovary in species of *Carex*. Several other structures of flowering plants are also known to become elaiosomes.

Collectively, the diverse origins of elaiosomes are a good example of convergent evolution, showing that structures with different forms and functions can be reshaped by natural selection to play the same ecological role. In the case of elaiosomes, plant tissues that were originally protective against herbivorous insects and other threats were changed biochemically and structurally into food lures for ants.

Elaiosomes consist of greatly modified cells that contain large vacuoles, or membrane-enclosed regions, filled with a rich mixture of nutrients. After surveying a wide selection of myrmecochores, Andreas Bresinsky of the University of Munich reported that elaiosomes contain many different fats, fatty acids and other common nutrients required by animals. These substances could be consumed as food by ants.

Most ants are omnivores that eat insects and other animal and plant parts found on the soil surface. Elaiosomes and their attached seeds may chemically mimic animal tissue to trick ants into picking them up.

Elaiosomes may also contain other biochemical cues that trigger collecting behavior in ants. Diane L. Marshall of the University of New Mexico and her co-workers identified an attractant in the elaiosomes of the European violet, *Viola odorata*, as a specific chemical called 1,2 diolein, a polar lipid compound. A similar lipid has been discovered in elaiosomes from two Australian shrubs, *Acacia myrtifolia* and *Tetratheca stenocarpa.*

The significance of these chemicals to ants is not yet clear, but their appearance in myrmecochores on opposite sides of the world suggests that evolutionary convergence is at work. Even more intriguing, this similarity leads to the hypothesis that, beyond food gathering, the elaiosome may elicit other innate behaviors in ants. For example, oleic acid is known to stimulate necrophoresis, or corpse-carrying behavior, in some ants. Conceivably elaiosomes containing this substance are carried off for the same reason.

In addition to elaiosome food bodies, myrmecochorous plants sometimes display other morphological adaptations that facilitate the placement of seeds near ant trails. In some species the stems and fruit supports are thin and weak and bend when the seeds are mature. The spindliness causes the seeds to be presented close to or on the ground, where ants forage.

In other plant species the typical morphology is altered more profoundly. For example, in *Carex umbellata,* a sedge species that we have studied, the flowering stem is exceptionally short, and the seeds and surrounding tissues mature close to the ground. As a result, instead of growing on long stems that bend toward the ground at maturity, the seeds remain near the ants' foraging level throughout their seasonal development.

Even more dramatically altered is the morphology of *Trillium petiolatum,* a wildflower of western North America. Most trilliums have a flower and three leaves that grow atop a long stem, which is often 30 centimeters high. In *T. petiolatum,* however, the large and showy flower forms very close to the ground, where the elaiosome-bearing seeds mature and are eventually presented to ants.

Moreover, the leaves of trilliums usually grow in a whorl below the flower, but in *T. petiolatum* that architecture would put the leaves on the forest floor. Although the leaf blades of *T. petiolatum* originate in the normal place beneath the flower, they grow at the ends of long petioles, or stems, that raise them above the flower to a position more favorable for photosynthesis. In short, the usual structure of a trillium has almost been inverted. The reasonable evolutionary explanation for *T. petiolatum*'s form must lie with the remarkable advantages of the distribution of seeds by ants.

To facilitate seed dispersal, myrmecochores may also have altered the times at which their seeds mature. In temperate climates, for example, most myrmecochores bear seeds and elaiosomes that mature by early spring. Insect carcasses, often the primary foodstuff of ants, are not as abundant then as they are in summer, after insect populations have expanded. Plants that bear ripe elaiosomes in the spring should therefore face less competition for the attention of food-gathering ants, and their seeds may be transported more frequently than seeds presented in the summer or fall.

In this way natural selection for the early ripening of seeds and elaiosomes could explain the predominance of vernal myrmecochores. Of course, other factors may favor high metabolic activity of woodland herbs in the spring, such as the availability of light on the forest floor before the canopy trees leaf out. The economics of ant foraging may be only an additional selection pressure reinforcing the development of myrmecochores in early spring.

The ants that collect the seeds of myrmecochores are a diverse lot. Many are clearly adapted for an existence that is primarily carnivorous. For example, Carol C. Horvitz of the University of Miami has shown that in southern Mexico the ants that transport the seeds of *Calathea* belong to the genera *Odontomachus* and *Pachycondyla,* which have powerful stingers and large mandibles for subjugating prey. The ants nonetheless harvest the seeds excitedly and take them back to the nests, where they remove the elaiosomes and feed them to larvae. Perhaps chemicals in the elaiosomes stimulate the ants in the same way that prey would.

Many other genera of ants also disperse seeds: *Formica, Myrmica* and *Aphaenogaster* are commonly involved in temperate forests of Europe and North America; species of *Rhytidoponera, Pheidole* and *Iridomyrmex* are prominent in the heath vegetation of southeastern Australia. Even seed harvesters such as *Messor, Pogonomyrmex* and *Veromessor,* which ordinarily eat seeds, have been identified as seed dispersers under certain conditions.

Strategically it makes sense that myrmecochory would strive to encompass as many ants as possible. Many different species can usually be found in any setting, and a plant that evolved to attract only one of them would probably be at a disadvantage. In fact, of the thousands of known myrmecochorous plant species, not one has been shown to rely on a single ant species.

Similarly there is no evidence that any ant species has evolved to take advantage of a particular myrmecochore. This lack of specialization contrasts sharply with the many species-specific relationships between insects and plants that have been observed in the tropics, where such relationships are often important for pollination.

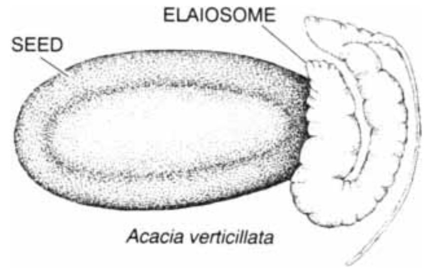
The phenomenon of myrmecochory is therefore best described as a result of plant evolution, not of the coevolution of a plant and an insect. To the ant world, an elaiosome may be just another package of nutrients to be carted to the nest.

Why have ants been singled out for dispersing seeds? Certainly many other groups of insects are common in the habitats where myrmecochores are found. The effective distribution of seeds, however, requires an insect that moves the seeds a significant distance and does not harm them. Only social insects, which carry food back to their nests rather than eating it at the point of discovery, meet both criteria. Workers typically scavenge an area around a nest and then carry foodstuff back to it to provision their larvae. The evolution of social behavior in ants had therefore preadapted, or predisposed, them to become efficient dispersers of seed.

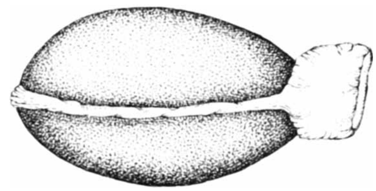
Ants also have other characteristics that increased the likelihood that selection would favor them as seed distributors. They are among the most abundant insects in most habitats. They forage on the surface of the soil efficiently during the long growing season of plants. When ants find new food resources, they recruit workers to collect as much of the food as possible. Ants will even move their entire nests to areas that seem particularly rich in food. All these behaviors benefit the myrmecochorous plants whose seeds are being dispersed.

Because myrmecochory occurs in very

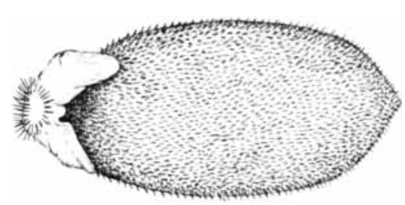
SOME SEEDS WITH ELAIOSOMES

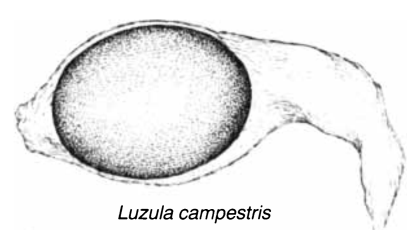


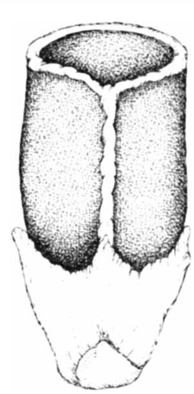
Acacia verticillata


Viola nuttallii


Corydalis aurea


Polygala vulgaris


Luzula campestris


Lamium album

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, Utah 84104. 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 $1.50 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, Mass. 01970. Specify CCC Reference Number ISSN 0036-8733/89. $1.50 + 0.00.

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

Advertising correspondence should be addressed to Advertising Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Address subscription correspondence to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 3187, Harlan, IA. 51593. Telephone inquiries: 1-800-333-1199, U.S. only; other 515-247-7631/32. The date of the last issue on subscriptions appears on each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, a mailing label of a recent issue) as well as the new one.

Name ____________________

New Address

Street ____________________

City ____________________

State and ZIP ____________________

Old Address

Street ____________________

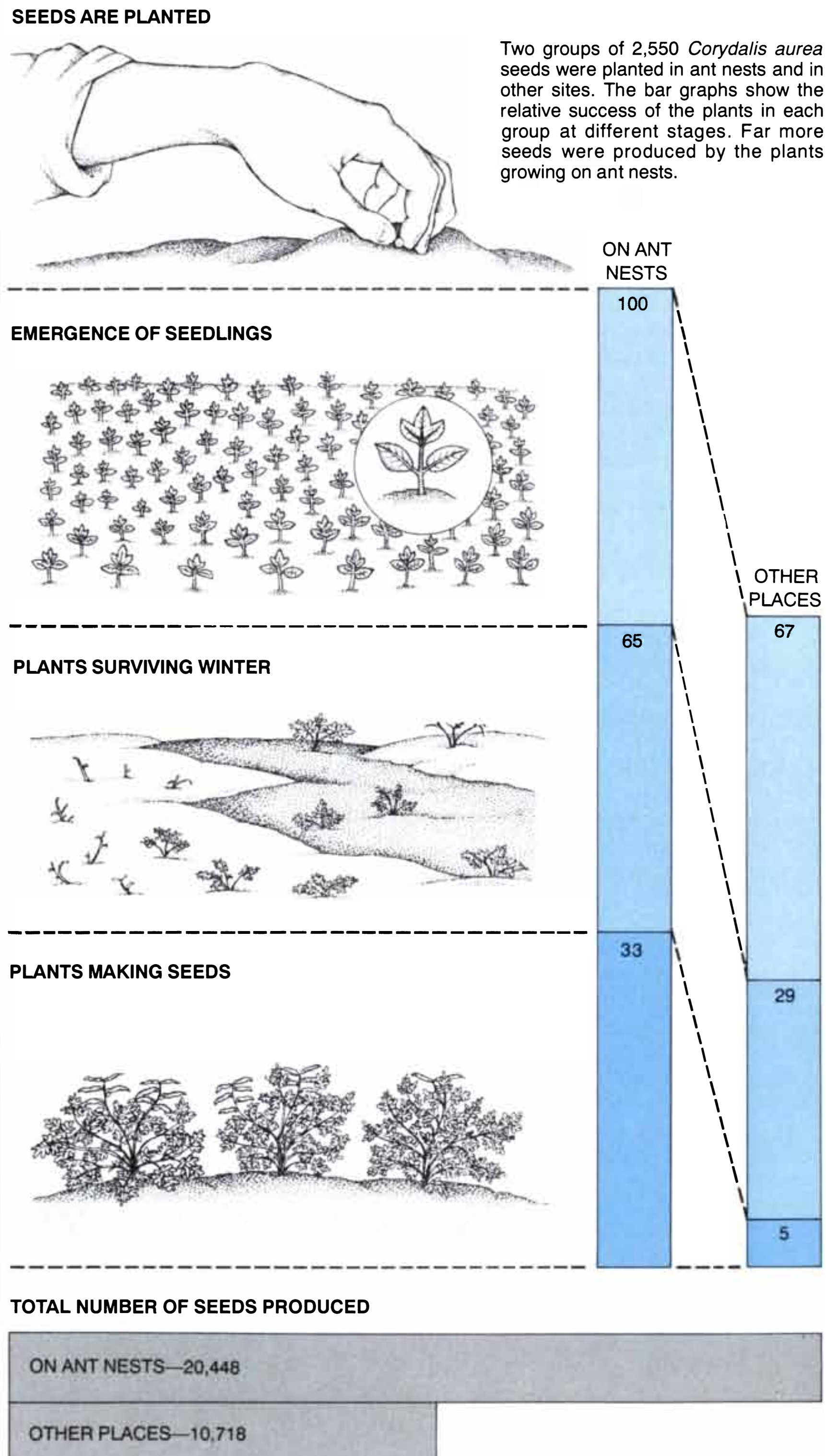
City ____________________

State and ZIP ____________________

different habitats around the world, ecologists have wondered whether generalizations can be made about the evolutionary advantages of this phenomenon for plants. Only recently has a series of field and laboratory experiments uncovered how the attractiveness of seeds to ant transporters could increase the survival and fecundity of a plant species.

One fundamental benefit of the dispersal of seeds by ants is the expansion of a plant's range. Ants have often been observed carrying seeds one or

ADVANTAGES OF GROWING ON ANT NESTS



Two groups of 2,550 *Corydalis aurea* seeds were planted in ant nests and in other sites. The bar graphs show the relative success of the plants in each group at different stages. Far more seeds were produced by the plants growing on ant nests.

two meters, but trips of 70 meters have also been witnessed. Ants therefore make it possible for plants to colonize new areas. By dispersing its population, a plant species lessens the risk it will become extinct because of local changes in habitats. Any ant species, regardless of nesting habits, can confer this benefit.

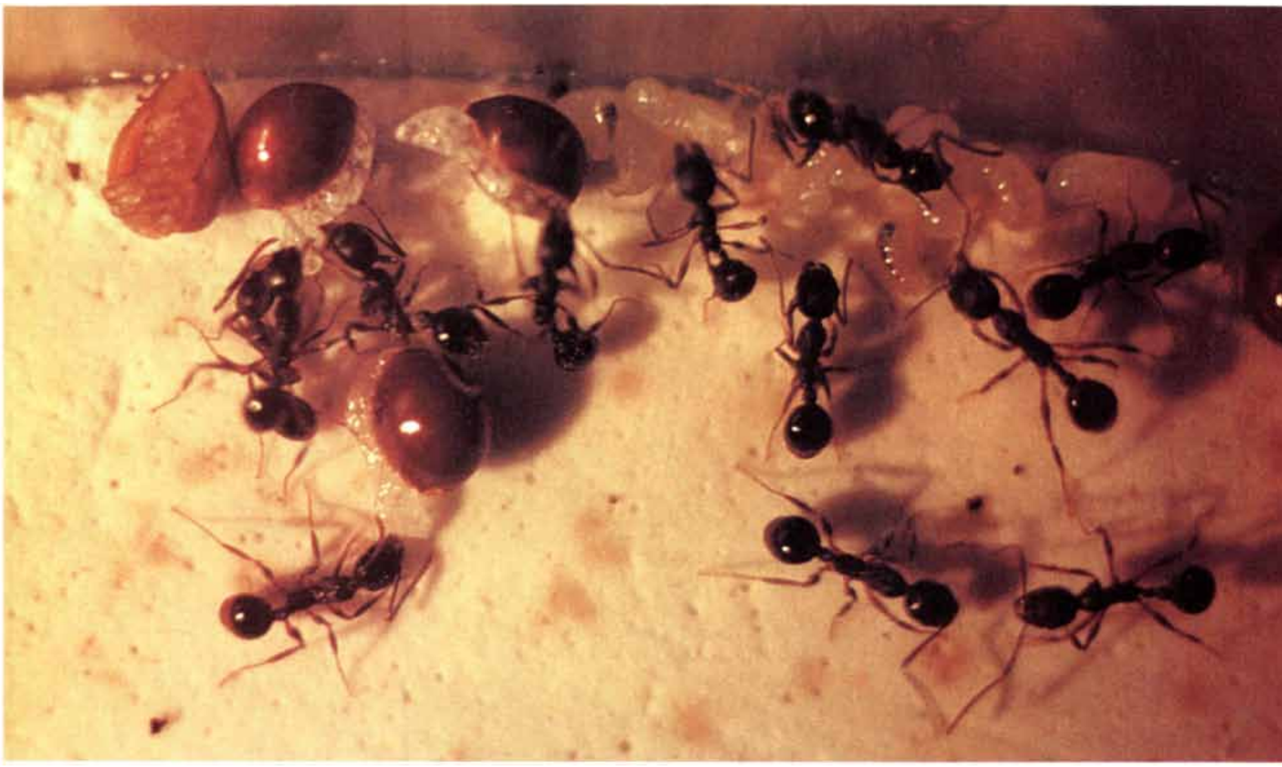
Ants can also improve the odds of a seed's survival by transporting it away from the parent plant, which may shade and inhibit the growth of the seedling. In an experiment conducted by one of us (Handel), seeds of the sedge *Carex pedunculata* that were left under the parent plant grew into seedlings with only three leaves. During that same period, seedlings from seeds that had been moved away from their parent grew an average of 89 leaves. Moreover, the transported seeds were far more fecund—only they produced plants that flowered the next summer.

Aside from helping to minimize competition between seedlings and parent plants, the movement of seeds by ants may also reduce competition among different plant species. For example, one of us (Handel) conducted experiments with three species of *Carex*, only one of which was a myrmecochore, that grow in the same habitat. The myrmecochore did not grow well in the presence of the others; yet it flourished when isolated.

Because the local ants were interested only in seeds with elaiosomes, they naturally isolated the myrmecochore's seeds in their nests. The myrmecochore was therefore able to monopolize certain sites in the habitat where ant nests were common, such as rotting wood. The plants did not have to compete with the seedlings of other *Carex* species for space, light, nutrients and other basic resources. Myrmecochory could easily be effective in the presence of many other genera whose seedlings compete for growth sites.

Even more than competition, predation is a major cause of attrition among seeds and seedlings. Many animals, including birds and small rodents, primarily eat seeds. And every gardener also knows that slugs and snails spell death for seedlings.

Research in several parts of the world has focused on the possibility that seeds brought into ant nests avoid at least some predators. Studies in the forests of West Virginia and the subalpine meadows of Colorado revealed that seeds placed in small areas from which ants had been excluded were almost invariably eaten within 24 hours. Conversely, when ants were present,



WORKER ANTS of the species *Aphaenogaster rudis* carry larvae to seeds from bloodroot (*Sanguinaria canadensis*) so that the larvae can eat the elaiosomes. The special advantages of elaiosomes as food for ants are not yet well understood.

seeds with elaiosomes were taken underground rapidly. In the case of *Viola nuttallii* in Colorado, Christine Turnbull of Macquarie University showed that the seasonal and daily release of seeds coincides with the maximum level of ant activity.

Australia may have a particularly intriguing case of predator avoidance. In the Australian heath and woodland, where vegetation is dominated by sclerophyllous (tough-leaved) shrubs, myrmecochorous species are abundant. So, too, are predators. Ironically, the major predatory species are ants. Recent work by Lesley Hughes, also of Macquarie, suggests that seeds released into this environment await the winners of a "race" between beneficial seed-dispersing ant species and harmful seed-eating species. The fate of the seed depends on which kind of ant finds it first. An elaiosome increases the probability that a disperser ant will reach a seed before a predatory ant.

Fires are another threat, particularly in the shrub-dominated vegetation of Australia and southern Africa. Adaptations for surviving fires are common among these plants, however. Many species, including some myrmecochores, are not merely fire-resistant but actually depend on fires to help their reproduction.

Research by several Australian investigators strongly suggests that the removal of seeds to ant nests protects the seeds from the lethal heat of brushfires. Yet, paradoxically, some seeds dispersed by ants cannot germinate unless they are exposed to lower, nonlethal levels of heat. Excavations of ant nests have revealed that seeds are buried at many different levels. This arrangement may be advantageous for the plants, because it allows at least some seeds to be buried deep enough so that they are safe from dangerous heat yet warmed enough to germinate.

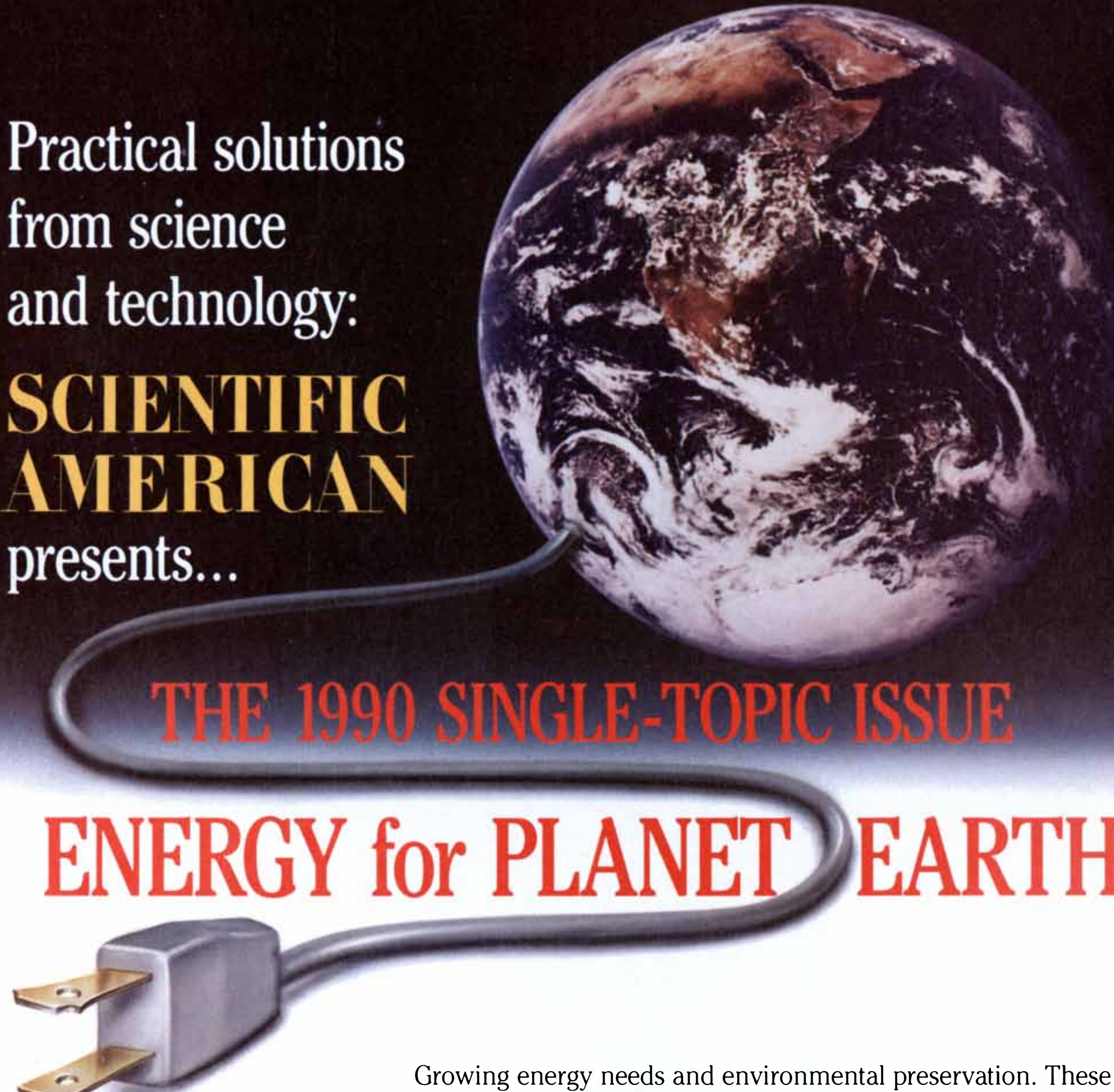
Unlike birds and mammals, which scatter the seeds they pick up widely in the habitat, ants bring seeds to carefully sited colonies—another behavior pattern that aids seed survival. In moderately moist forests, for example, ants often nest in rotting logs and stumps that are elevated above the soil. Such locations are less susceptible to flooding in the spring, which makes them attractive to both ants and seeds.

Ant colonies, like all other animal and human societies, have a propensity to accumulate waste. Their garbage dumps contain the remains of prey, fecal material, ant corpses and a host of other items (many of no apparent use) that ants collect and invariably bring home. Germinating seeds and seedlings, particularly those of myrmecochores, may benefit from being placed in these dumps.

Organic waste often contains high levels of nutrients critical for plant growth, which is why gardeners maintain compost heaps and farmers spread muck from feedlots over their fields. The levels of organic material, nitrogen, potassium and phosphorus are frequently greater in ant nests than in surrounding soils. Ant-colony wastes may therefore provide a small but ready-made cache of compost for seedlings that nourishes them through their delicate early stages.

Other physical characteristics of the soil in and around ant nests may also enhance the survival of seedlings. Nest building often makes the soil crumbly and better aerated. It also increases the soil's water-holding capacity. Some researchers now suspect that the fundamental benefit provided by many nests is to guarantee a modest water supply for a seedling until its roots are large enough to find water on their own.

Clearly, then, ants can profoundly affect the ecological conditions for the growth of seeds. To investigate the impact myrmecochory has had on evolution, researchers have followed the fates of seeds in field experiments. These studies have taken the basic



ROTTING LOGS are often the sites of ant nests. The plants growing out of the log shown above came from seeds brought into the nests. Most of the visible plants are a sedge, *Carex pedunculata*, but violets and hepaticas are also present.

form of comparing two cohorts of seeds, one of which has been taken into nests by ants and the other of which has been planted by hand at random in the same general habitat. One example of an experiment of this type was an early study of two myrmecochorous species of violets in southern England. It showed that after three years, when the seeds had germinated and the seedlings had emerged, the only survivors were from the ant-manipulated cohort.

A more recent study was conducted with a plant known as golden smoke (*Corydalis aurea*), a biennial that produces seeds during its second year. Frances M. Hanzawa of Grinnell College discovered that survival rates did not differ between the seedlings that had grown from ant nests and those that had not. More nest-derived seedlings survived the winter, however, and developed into reproducing adults. There was consequently a striking difference in the total number of seeds for a second generation produced by the two cohorts: the ant-manipulated plants yielded twice as many as the controls.

Because the two cohorts started with exactly the same number of seeds in their first generations, it is clear that a population of golden smoke plants used by ants will grow much faster than one that is not. And a fast-growing population has a better chance of competing with other plants for nutrients, space and other resources. Hanzawa's work therefore showed that the ecological conditions for seed dispersal, such as the availability of ants, affected the evolutionary potential of the plant populations.

Myrmecochory, then, obviously holds many advantages for some species of plant. At present, however, we do not know exactly how the ants gain from the interaction. We do know that elaiosomes are avidly sought by foragers, quickly gnawed off seeds and fed to larvae. How this behavior affects the growth rates of ant colonies has yet to be measured.

It is also worth pointing out that not all ants engage in seed dispersal. When seeds are shed by a plant, only a small subset of the ant species in a habitat takes any interest in the elaiosomes. Some specialization must be present in these ants, but no one yet knows whether it is behavioral, morphological, dietary or otherwise.

Seed dispersal by ants is therefore an important model for studying a wide variety of plant-animal interactions that in a sense appear to be asymmetric. Plants have clearly adapted themselves for the interactions with ants—elaiosomes are the most obvious of their adaptive features—but the adaptations made by the ants are much less evident.

As a dispersal system for seeds, myrmecochory is good but not foolproof. An elaiosome is an attractive prize for many different kinds of ants. The golden smoke experiments showed that no seedlings ever emerged from the nests of certain ant species. These ants appear to be, in effect, parasitic "elaiosome thieves" that may destroy the attached seeds or seedlings.

Aside from such thief ants, a dozen other factors in any habitat can contribute to the success or failure of myrmecochory as a seed-dispersal mechanism. Rainfall can flood out ants, and colonies can suffer from epidemics of fungal diseases or heavy predation. If other sources of food are abundant, elaiosomes may seem less attractive to ants. If plant species compete for the attention of ants, the seeds with the smallest elaiosomes may be ignored.

Because the success of seed dispersal by ants can vary considerably, Hall Cushman of Macquarie and John F. Addicott of the University of Alberta have suggested that myrmecochory is a conditional mutualism. At any given time and place, depending on the prevailing conditions, the interaction may function suboptimally. When all the conditions are right, however, the advantages of myrmecochory for both the plants and the ants are very significant. These advantages are strong enough to reinforce powerfully the pressures to maintain the traits that make the behavior possible.

As the long list of known myrmecochores grows, biologists can expect to learn more about the importance of this seed-dispersal mechanism around the world. Further studies of the benefits of myrmecochory for plants and ants should also help to elucidate details of mutualistic relationships and their evolutionary consequences.

FURTHER READING

MYRMECOCHOROUS PLANTS IN AUSTRALIA AND THEIR DISPERSAL BY ANTS. Rolf Y. Berg in *Australian Journal of Botany*, Vol. 23, No. 3, pages 475-508; June, 1975.

DISPERSAL ECOLOGY OF CAREX PEDUNCULATA (CYPERACEAE): A NEW NORTH AMERICAN MYRMECOCHORE. Steven N. Handel in *American Journal of Botany*, Vol. 63, No. 8, pages 1071-1079; September, 1976.

THE COMPETITIVE RELATIONSHIP OF THREE WOODLAND SEDGES AND ITS BEARING ON THE EVOLUTION OF ANT-DISPERSAL OF *CAREX PEDUNCULATA*. Steven N. Handel in *Evolution*, Vol. 32, No. 1, pages 151-163; March, 1978.

THE EVOLUTIONARY ECOLOGY OF ANT-PLANT MUTUALISMS. Andrew J. Beattie. Cambridge University Press, 1985.

SEED ECOLOGY. M. Fenner. Routledge, Chapman & Hall, Inc., 1985.

DIRECTED DISPERSAL: DEMOGRAPHIC ANALYSIS OF AN ANT-SEED MUTUALISM. Frances M. Hanzawa, Andrew J. Beattie and David C. Culver in *American Naturalist*, Vol. 131, No. 1, pages 1-13; January, 1988.

Global Warming Trends

Analysis of land and marine records confirms that our planet has warmed half a degree Celsius in the past century. Future warming trends, however, remain uncertain

by Philip D. Jones and Tom M. L. Wigley

Meteorological data compiled over the past century suggest that the world is warming. Is it? There are many factors, ranging from changes in thermometer design to the growth of "urban heat islands," that may bias the data and so create a spurious warming trend.

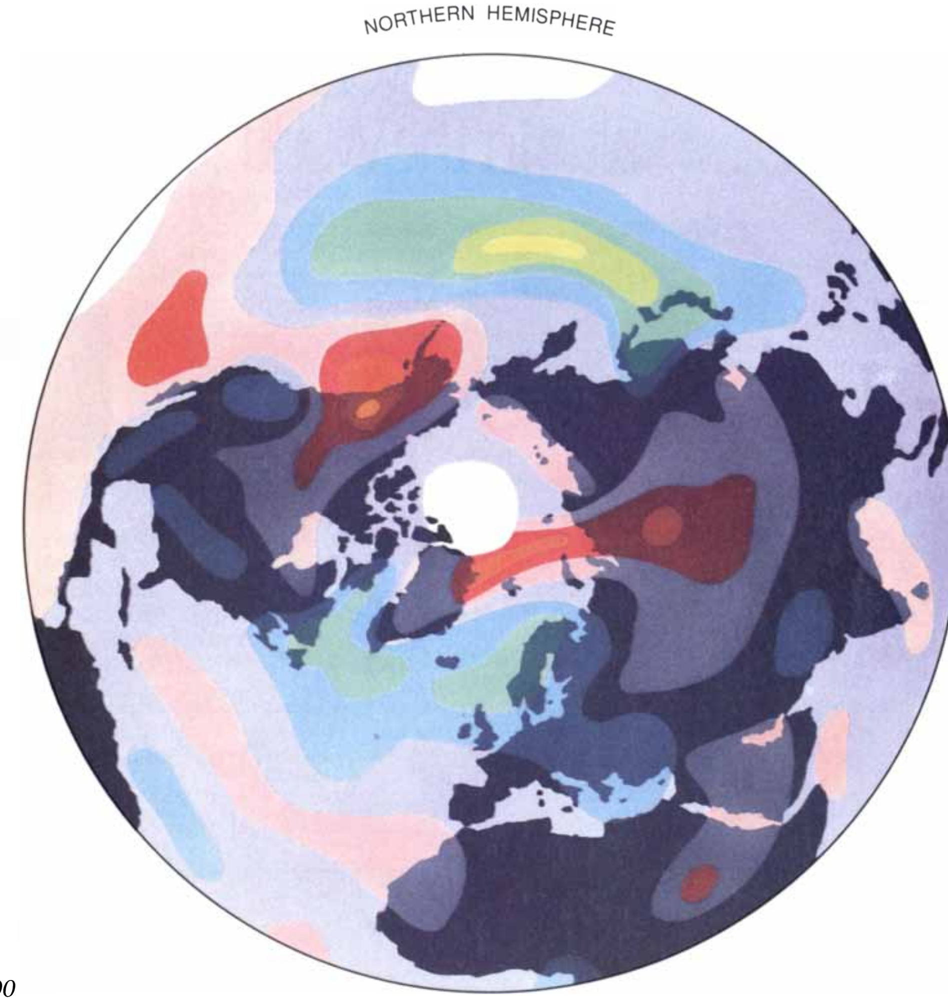
Even if temperatures have increased during the past 100 years or so, is there reason to believe that the warming will continue? Computer models predict that heat-trapping gases emitted into the atmosphere during the past two centuries will cause the world's average temperature to rise by from one to four degrees Celsius over the next 50 to 75 years. Such models, however, are relatively crude simplifications of the myriad complex physical processes taking place in the atmosphere and oceans; they cannot prove that emissions of greenhouse gases will significantly alter the earth's climate.

We have recently completed a 10-year analysis of global temperature trends, systematically correcting the data both from land-based and from marine observations to eliminate potential sources of bias. Our work shows conclusively that the world's climate, although highly variable over periods of decades or less, has become generally warmer during the past century. The rising temperature trend was briefly interrupted by a cooling spell from about 1940 to 1970, but since then it has returned to an upward slope and shows no signs of abating.

The causes of global warming are less certain than the trend itself. Although observed warming is consistent with the greenhouse effect, other factors, ranging from volcanic eruptions to ocean currents, influence climate and blur the greenhouse signal. Data gathered in the next decade or two (during which warming is predicted to intensify substantially) may settle some of these ambiguities.

Historical temperature records are crucial to quantifying just how much the northern and southern hemispheres have warmed since the industrial revolution touched

PHILIP D. JONES and TOM M. L. WIGLEY are both climatologists at the University of East Anglia in Norwich, England. Jones's research interests include long-term climate monitoring and the comparative analysis of climate data and computer-generated models. He has a B.A. from the University of Lancaster and a Ph.D. from the University of Newcastle upon Tyne. Currently he is secretary of the International Commission for Climatology of the International Union of Geodesy and Geophysics. Wigley is director of the Climatic Research Unit of East Anglia, where he models past and future climate change. He received his Ph.D. from the University of Adelaide, trained as a meteorologist at the Australian Bureau of Meteorology and taught engineering at the University of Waterloo in Ontario before joining the CRU in 1975. A youthful interest in spelunking led first to studies of carbonate geochemistry and then to a more general interest in carbon dioxide and the greenhouse effect.



off a massive increase in atmospheric levels of carbon dioxide and other greenhouse gases. Unfortunately, such records are difficult to come by. Today global temperature measurements are compiled through the World Weather Watch System, a global cooperative network of national meteorological services. Observations in previous centuries were largely compiled by individual observers working without coordination.

About 10 years ago, in the face of growing global climate concerns, we and our colleagues at the Climatic Research Unit of the University of East Anglia initiated a project to collect and analyze, once and for all, every available historical temperature record. In this effort we had the sponsorship of the U.S. Department of Energy and the collaboration of Raymond S. Bradley of the University of Massachusetts at Amherst and Henry F. Diaz of the Environmental Resources Laboratory of the National Oceanic and Atmospheric Administration (NOAA).

The task was not simple. Few of the workers who established routine meteorological observing networks in the 18th and 19th centuries could have anticipated how important their data would be in helping subsequent generations of scientists unravel the mysteries of climate change. Not surprisingly, the records they left behind are sparse, often incomplete and inconsistent. Nevertheless, by quantifying and then eliminating uncertainties in the data, we have been able to assemble a fairly accurate picture of what has happened to the earth's climate in the roughly 300 years since instrumental meteorological recording began.

The earliest records are of little use today because their coverage is confined to western Europe. Moreover, many of the oldest temperature records either have been lost or survive only in summary form. Even more would have been lost but for the efforts of Heinrich Wilhelm Dove, a German meteorologist who collected, mostly by correspondence, as many data as possible.

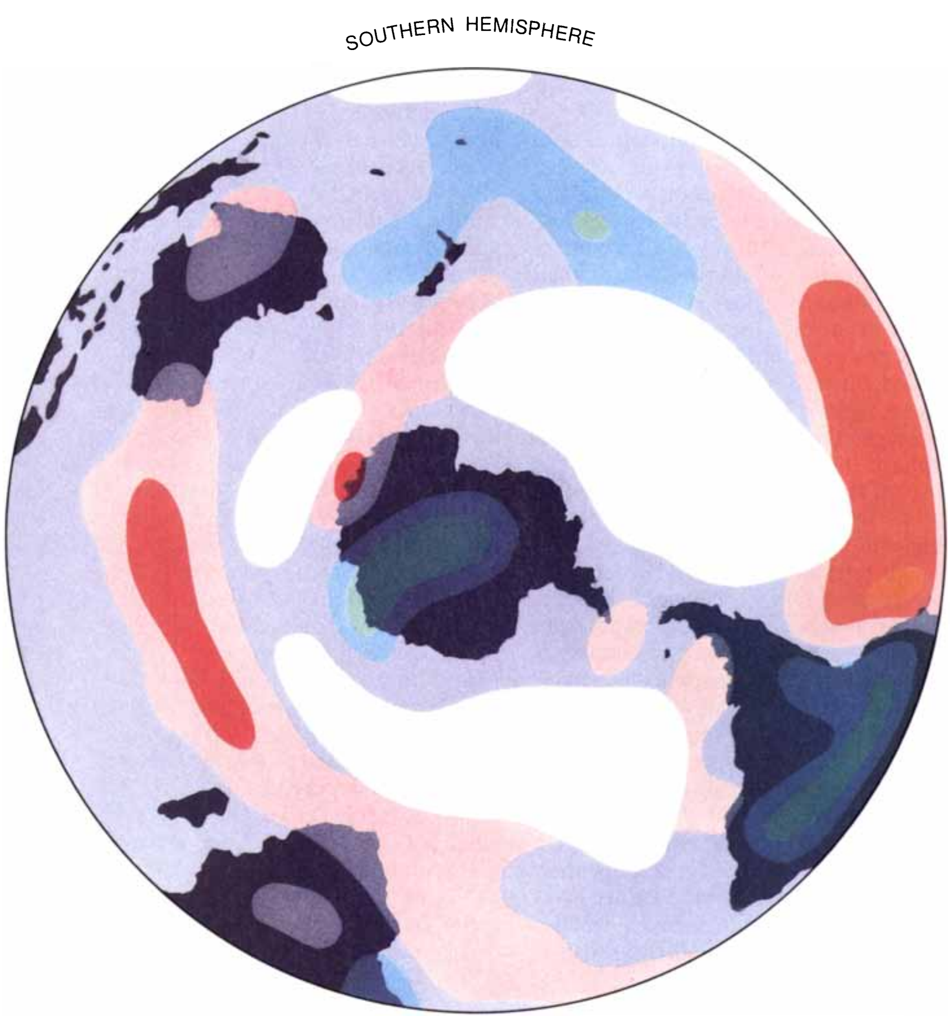
Dove's compilations are of great importance, and the early analyses of his data were far ahead of their time, but they have limited value today because they do not include the interiors of Africa, Asia, South America or Australia. Since 1850 national meteorological agencies have made a concerted effort to collect and archive temperature data; temperature records compiled since then are thus far more complete.

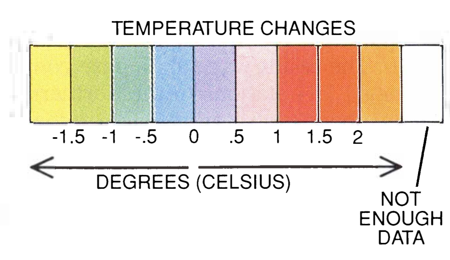
Observing stations were gradually established throughout most of the rest of the world; by the late 1950's the observing network had even reached Antarctica. (There are a few notable pauses in this growth—the freezing of mercury, for example, stymied early measurements in northern parts of the Soviet Union and Canada.)

Modern attempts to determine global temperature trends date back roughly 30 years. Early efforts, however, were limited by the paucity of locations—a few hundred or so—for which data had been published. Our search turned up more than 3,000 sets of records, many unpublished, that could be analyzed.

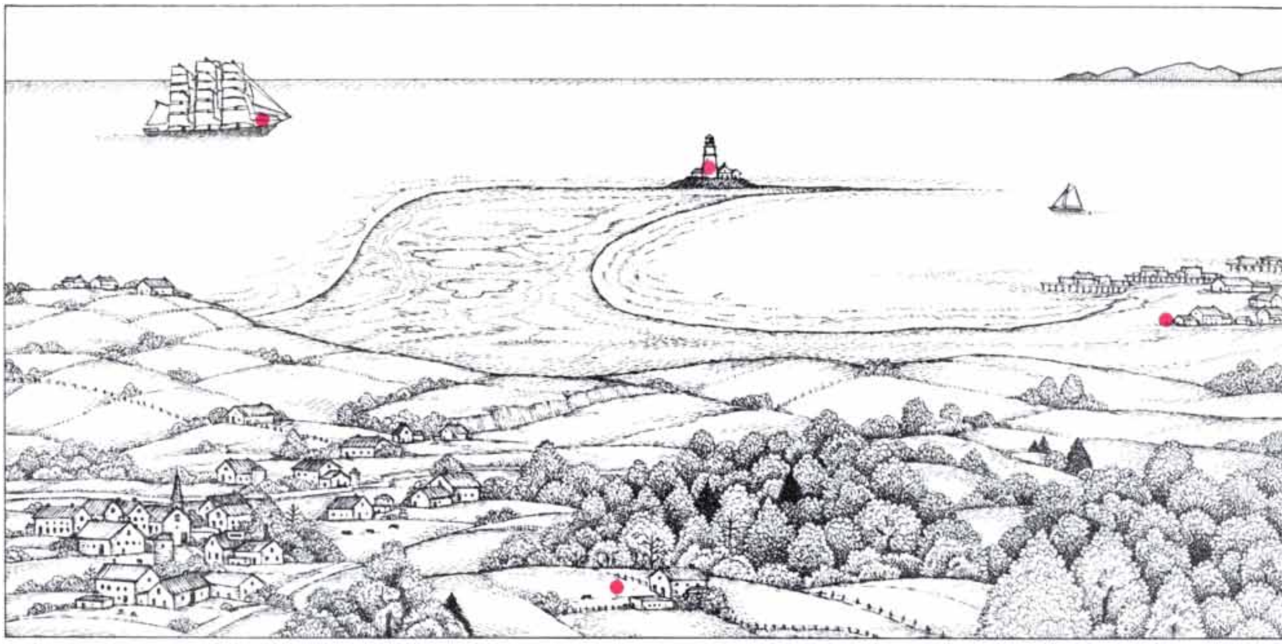
Having built up a large list, we then proceeded to pare it down. The essential requirement was homogeneity: temperature records had to reflect only day-to-day weather changes and longer-term shifts in climate. Fluctuations that could be attributed to other causes—such as the relocation of a temperature-measurement station, inconsistent calculations of mean monthly temperature, or changes in instrumentation, observation times or local conditions—can seriously affect the reliability of the data. Environmental effects are the most damaging source of inhomogeneity because they may introduce a consistent bias rather than merely decreasing overall accuracy. The growth of towns and cities, for example, leads to urban heat islands, which typically increase measured temperatures at nearby stations.

We assessed homogeneity by comparing each station's records with those of stations from a few tens of





WARMING TRENDS vary from region to region around the world as indicated by data collected from 1967 to 1986. In both the northern (*left*) and southern (*right*) hemispheres, most regions have warmed, some markedly, but a few—notably in the northern Pacific and Atlantic oceans—experienced cooling. If predictions are correct, global warming will accelerate in the coming decades.



CHANGING LANDSCAPES affect temperature readings in ways that may produce spurious temperature trends. During the 19th century, for example (*left*), most cities and towns were relatively small and so had little effect on the climate of surrounding regions. Today (*right*) urban heat islands have a direct impact on climate, inflating temperatures at nearby me-

kilometers to a few hundred kilometers distant. Jumps or trends in the temperatures recorded by one station that are not reflected at the others are generally signs of inhomogeneity. The homogenizing process works well except in regions where the nearest stations are more than a few hundred kilometers away or where all stations are affected by related factors such as similar rates of urban growth.

On the basis of these comparisons we accepted some station data, adjusted others for jumps by adding or subtracting appropriate correction factors and rejected about 10 percent because they could not be reliably corrected. We also rejected some stations whose temperature records ended prior to 1950. The result was a data base containing 1,584 stations in the Northern Hemisphere (down from 2,666 initial records) and 293 in the Southern Hemisphere (down from 610).

From the station data we calculated regional and hemispheric temperature averages. The task is not as easy as it might seem. One important source of error is the turnover in stations; over time new ones appear as others are lost. If stations located in warmer places (such as valleys) are substituted for ones at colder spots (such as hillsides), for example, the result would be a spurious overall warming trend.

The simplest way to overcome such spurious trends is to express the temperatures of all stations in terms of deviations from their average temperatures during a reference period (1950 to 1970) for which reliable global data are available. (We estimated averages for some of the stations that had no records covering the reference period from those of nearby stations.)

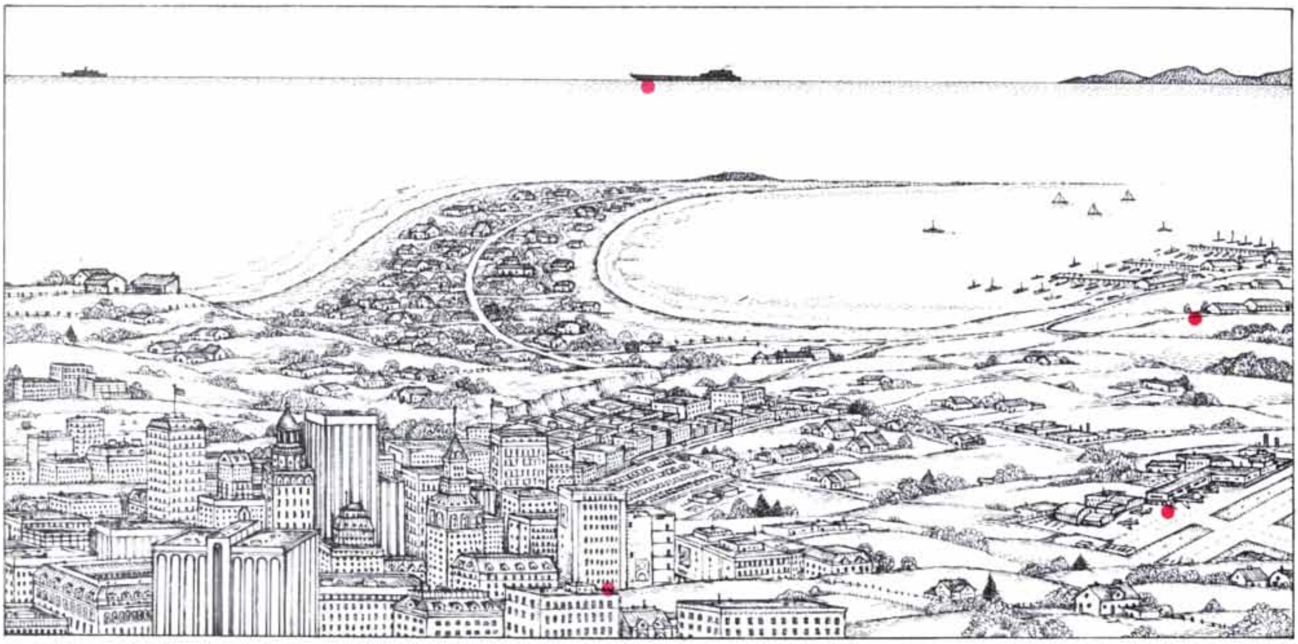
We then processed the station data to produce area averages—the average temperature for each point on a grid of five degrees latitude by 10 degrees longitude (about 550 by 1,100 kilometers at the equator). Some areas contained many stations whose data were averaged to provide the value for a single grid point; in others the average was based on only a single record. From these figures we finally produced the average temperatures for each hemisphere. (This method eliminates the bias of the hemispheric average toward areas with many temperature-recording stations.)

Our calculations yielded two immediate results: first, it was clear that the global climate varies considerably from year to year, and second, we confirmed that the earth has experienced an overall warming trend of half a degree since the late 19th century.

Despite our painstaking efforts to ensure that the records we analyzed were homogeneous, a number of questions concerning their reliability remain. Are the earlier averages, compiled from a much smaller set of stations, comparable to those of more recent times? Can we be certain that we have removed the biases associated with urban warming? And finally, how well do land-based temperatures represent the climate of a planet that is two thirds covered by water?

To determine whether the sparseness of older records made 19th-century averages unreliable, we compiled revised averages for a subset of weather stations. These averages mimicked the sparser coverage typical of the 19th century. The size of the discrepancies between the subset averages and the true figures implies that temperature estimates for years before 1880 are only about half as accurate as those for years since 1920. The same data show, in contrast, that estimates of the average temperature over the course of a decade—more important for determining long-term trends—are quite accurate. Sufficient data exist to pinpoint the 10-year averages to within .1 degree C from the 1850's onward in the Northern Hemisphere and the 1880's in the Southern Hemisphere.

Next, to test our methods for removing the urban-heat-island bias from the record, we compared our annual averages for the entire U.S. with those calculated by Thomas R. Karl of the National Climatic Data Center in Asheville, N.C., who used data recorded predominantly in rural areas. Our U.S. data show only about a tenth of a degree C more warming than do Karl's, so clearly our careful screening has had



teorological stations. The relocation of urban stations at outlying airport sites, in contrast, may generate apparent trends in the other direction. The reliability of ocean data has also changed over time. As ships increased in size, air-temperature readings were taken farther above the ocean surface (where the air is warmer), creating a spurious cooling trend.

some effect in removing any urban bias.

Although the remaining tenth of a degree C could represent a slight urban-warming bias, it could also be associated with other factors. Careful comparisons of our data with rural records from the Soviet Union, eastern China and eastern Australia show only an excess warming of between zero and .05 degree C per century, indicating that factors other than urban-warming bias may be responsible for at least some of the differences between Karl's averages and ours.

The temperature measurements we have discussed thus far were all made on land, which covers only about a third of the earth's surface. Nevertheless, there are reasons to think that land data can provide good estimates of hemispheric temperature fluctuations, at least over decades and centuries. The upper layers of the oceans have a far greater heat capacity than either the atmosphere or the thin layer of the earth's crust that participates in temperature changes over the course of a century or less, and so one might expect that temperature changes over land should closely reflect temperature changes over the oceans. Because the winds that blow between oceans and land ensure good thermal communication between these two parts of the globe, any differential between the two should quickly disappear.

Indeed, the strong parallels between land and marine temperatures offer an opportunity to verify and strengthen confidence in warming estimates based on land records alone. In this regard, the climatological world is indebted to a U.S. naval captain, Matthew Fontaine Maury. Through his pioneering efforts in the 1830's and 1840's, he helped to standardize the methods by which ships at sea make meteorological observations, including measurements of water and air temperature. Largely as a result of his work, an international agreement to take, collect and exchange marine meteorological observations was signed in Brussels in 1853.

Since that time the various maritime nations—the U.S., Great Britain, France and others—have archived logbooks containing marine weather observations. During the past 20 years the information in these logs (amounting to some 80 million observations of sea-surface temperature alone) has been transferred to two computerized data banks, the Comprehensive Ocean-Atmosphere Data Set compiled by NOAA and other U.S. agencies and a separate compilation by the Meteorological Office in Great Britain.

Like land-based records, marine data must be adjusted for changes in measurement techniques and other effects. Before the 1940's sea-surface temperatures were taken by throwing a bucket over the side, hauling seawater onto the deck and then waiting a few minutes for the thermometer to respond before recording the water's temperature. Since the early 1940's, however, most sea-surface temperature measurements have been taken in the intake pipes that draw seawater in to cool ships' engines. Although many bucket measurements are still made today and some intake measurements were made before 1940, for the most part the switch was rather abrupt.

Comparative studies show that intake measurements are generally about .3 to .7 degree C warmer than uninsulated canvas-bucket measurements—an increment roughly as large as the total warming seen in the land record. Correcting this inhomogeneity is therefore imperative. Yet logs have specified whether temperatures were taken by bucket or by intake pipe only since the 1970's, and so corrections are somewhat problematic.

In addition, not all bucket measurements are equal. Wet buckets cool by evaporation as they are lifted from the ocean and placed on deck. The precise amount of cooling depends on weather conditions and the insulating qualities of the bucket. In an attempt to standardize temperature readings, the 1853 Brussels agreement specified the use of wooden buckets, whose insulating qualities are good, but 19th-century fleets continued to carry buckets made of canvas, tin and other materials. During the period from 1900 to 1940 most ships carried canvas buckets, which are poor insulators and allow the seawater to cool appreciably in the interval between sampling and measurement. Buckets used since World War II are made of plastic and properly insulated; their readings agree well with intake measurements.

Marine air-temperature measurements, meanwhile, are also subject to homogeneity problems. The most critical one is caused by the increase in the average size of ships over time and the resulting increase in the height of the deck (and thus the thermometer) above sea level. Air temperature typically falls sharply with height, and this height increase should produce a spurious cooling trend. In addition, it is almost impossible to know whether thermometers on particular ships were exposed to direct sunlight or placed near warm structures, either of which would cause abnormally high readings.

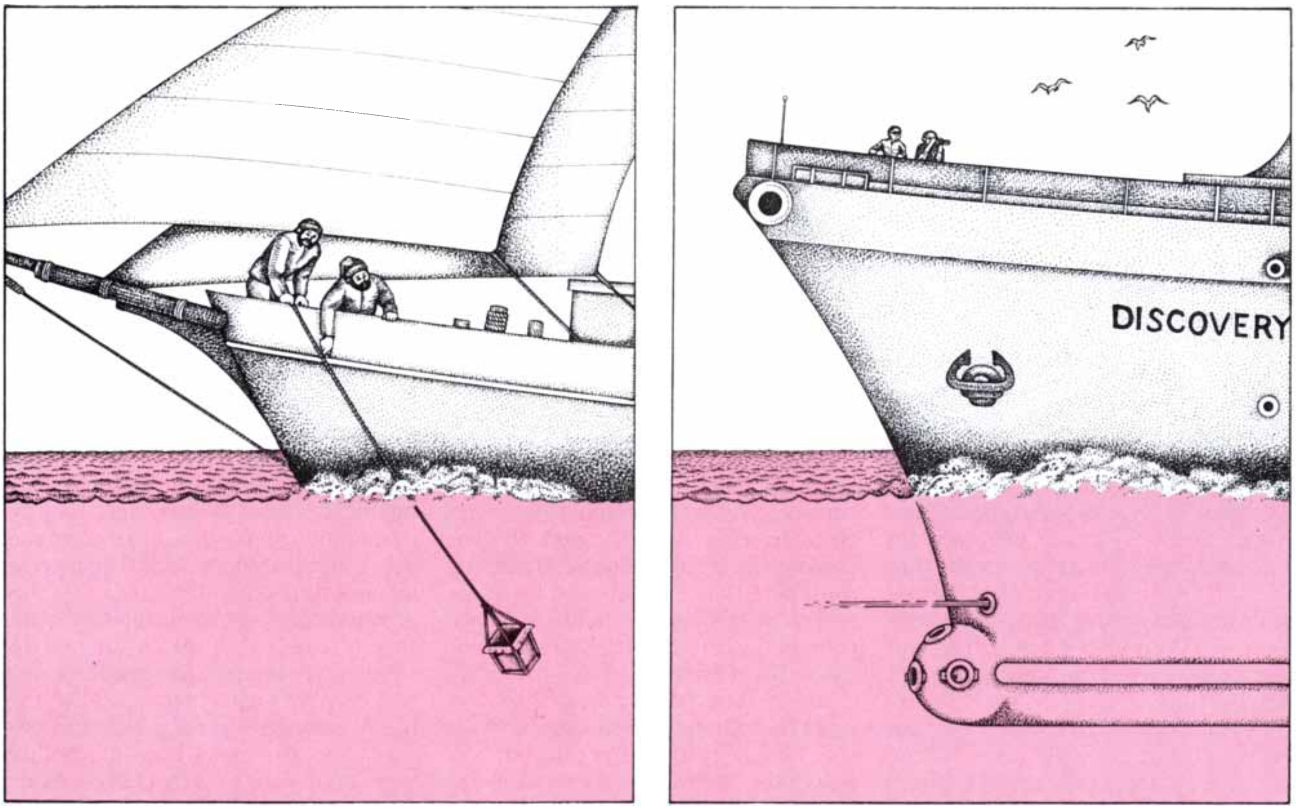
Records compiled during World War II furnish classic examples of various causes of inhomogeneity. During that period most thermometers were moved to ships' bridges where readings could be taken more safely (but where the heated structure warmed the air). Furthermore, the vast majority of readings were taken during the day because a light to read the thermometer at night would have been forbidden. As a result, temperatures recorded during the war were consistently about one degree C warmer than in the years just before or after.

Early on in our analysis we corrected the marine data by comparing marine records made near land with land-based observations made on islands and coasts. We attributed any differences between the land and marine temperatures to marine-measurement inhomogeneities and derived correction factors by averaging these differences over many regions. This approach assumes that land data are homogeneous; nevertheless, the corrections it yields are remarkably consistent. The averaging required, however, limits the technique to corrections only of hemisphere-wide, annual-mean data rather than local averages or month-by-month time scales.

We have since adopted a more sophisticated correction technique, developed by Chris K. Folland and David E. Parker of the U.K. Meteorological Office, which adjusts older measurements for the evaporative cooling of buckets. The amount of cooling depends on two factors: weather conditions and the exposure time between hauling the bucket on deck and measuring the seawater temperature. Whereas prevailing weather conditions depend on the time of year and the location at which the measurement is taken, exposure time is generally unknown and must be estimated from the data.

To estimate the exposure time, we express sea-surface temperatures for any given month in terms of their deviation from the average for that month during the 1950–1979 reference period. Because seasonal cycles of ocean temperature have remained relatively stable during the past 100 years, these deviations should display no discernible summer-winter pattern—the difference between December, 1890, temperatures and the December reference-period average, for example, should be neither higher nor lower than the difference between June, 1890, temperatures and the June reference-period average. Any seasonal pattern in the data can be blamed on the buckets, which cool by an amount that depends on the time of year. From whatever seasonal pattern appears, therefore, we can estimate an exposure time and can correct all the bucket data accordingly.

The corrections derived from land data and from cooling models agree quite well during the period from 1900 to 1940, when almost all ships used canvas buckets. The agreement is also



SEA-SURFACE TEMPERATURES must be corrected to account for measurement technique. Before about 1940, readings were taken by hauling water on deck in a bucket (*left*). Since then thermometers have generally been inserted in engine-cooling water-intake pipes. Because water cools by evaporation, bucket readings may be as much as .7 degree C colder.

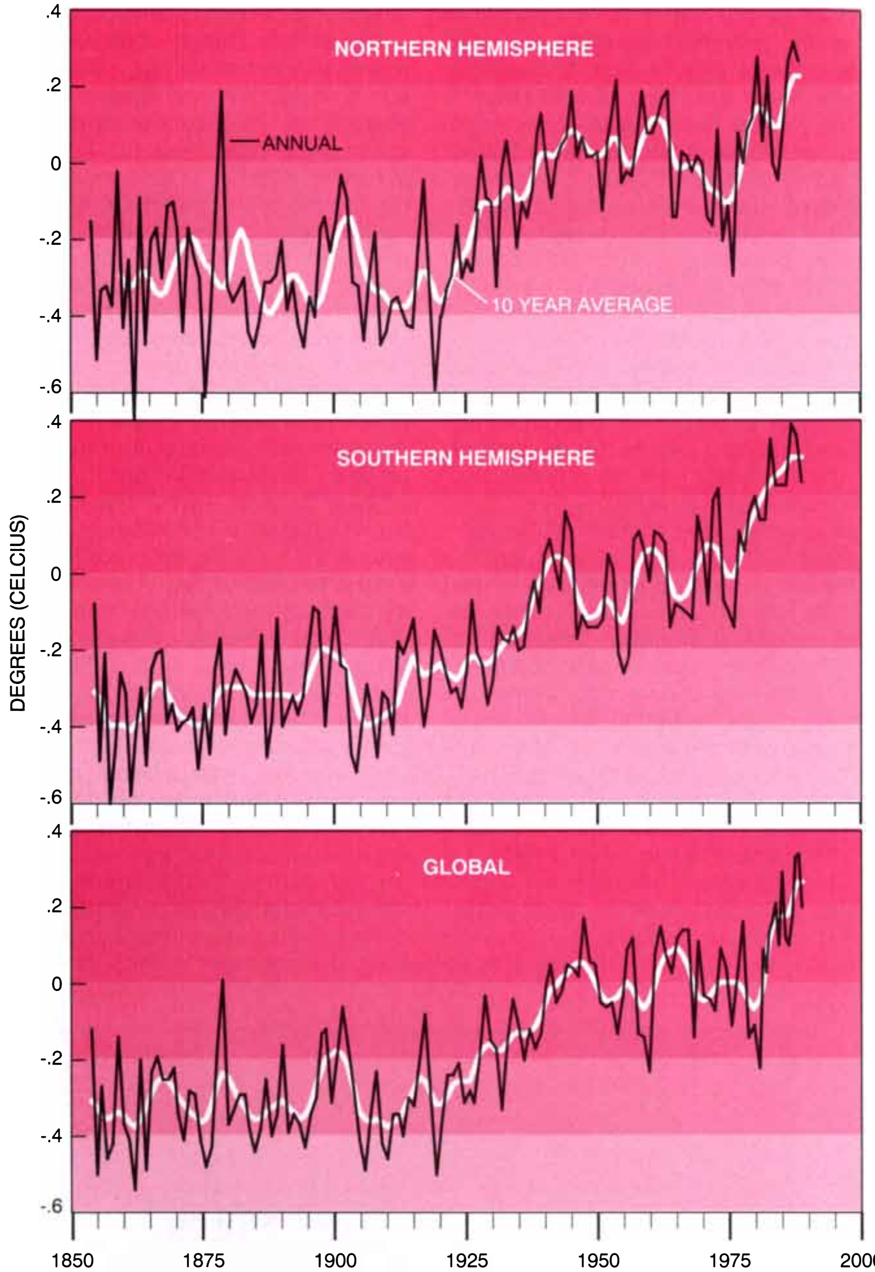
fairly good for 19th-century data if one assumes that most samples were taken with wooden buckets. If one assumes that canvas buckets were used, the hemispheric averages of marine data are consistently about .2 degree C warmer than the land averages.

Nevertheless, overall comparisons of land and marine data demonstrate a remarkable parallelism. Even on an annual time scale the hemispheric averages over land and ocean are strongly correlated; for longer-period fluctuations the two almost completely agree. Any remaining uncertainties must be blamed on poor global coverage. Some areas, particularly the southern oceans, are poorly measured even today. The results of recent satellite surveys, however, suggest that incomplete coverage is not a serious problem. Roy W. Spencer of NOAA and John R. Christy of the University of Alabama produced temperature estimates from satellite data for the period from 1979 to 1988 that agree very closely with our combined land and marine series.

After our 10 years of tracking down temperature records and correcting them for potential sources of error, we can state unequivocally that global temperatures have risen over the past century. But many questions persist: How strong is the warming trend? What is its cause? Is it related to the greenhouse effect, and, if so, why is it punctuated by a period of cooling? How significant is it that 1987 and 1988 were the two warmest years on record?

Most of these questions can be answered definitively only by decades of additional data gathering. Even without such data, however, current models of climate change may yield some understanding. Climate change is determined by both internal and external factors; among the internal ones are changes in planetary albedo (resulting from natural fluctuations in cloud cover or surface characteristics) and in patterns of atmospheric or oceanic circulation. Because atmospheric circulation patterns determine the horizontal and vertical fluxes of heat through the atmosphere, they influence how heat is taken up from land or the oceans and returned to them. Oceanic circulation patterns strongly influence the temperatures at the bottom of the atmosphere and the rates of heat exchange between the oceans and the atmosphere. Fluctuations in both may lead to long-term fluctuations in temperature.

External factors that affect climate include natural ones, such as changes in the sun's luminosity, and others



CLIMATE VARIABILITY is evident in historical records of annual and 10-year average temperatures over the land and ocean areas of the Northern Hemisphere (*top*), the Southern Hemisphere (*middle*) and worldwide (*bottom*). In all three graphs, however, an upward trend in temperature is apparent; even the coldest years of the past decade are warmer than all but the hottest years of a century ago.

that may be either natural or artificial. Changes in the fraction of short-wavelength radiation reaching the troposphere (the portion of the atmosphere that takes part in weather), for example, may result from a buildup of industrial emissions or from the natural injection of dust and sulfates into the stratosphere by volcanic eruptions, or both. Industrial emissions can also affect climate indirectly by altering the albedo of clouds. Increasing concentrations of greenhouse gases, meanwhile, can affect climate by changing the absorption of outgoing long-wavelength radiation in the troposphere.

Most of the variations in climate from year to year stem from internal factors involving the circulation of the atmosphere. On longer time scales—between two and eight years—changes in the vertical circulation of the oceans and in sea-surface temperatures drive climate variation. For example, the El Niño/Southern Oscillation phenomenon (a breakdown of the predominant pattern of easterly surface winds, westerly high-altitude winds and upwelling of cold water in the eastern Pacific) leads to a worldwide increase in storms and a temporary drop in global mean temperature. Indeed, the warming trend of the late 1980's stands out even more clearly after correcting is done for El Niño; 1989 becomes the warmest year in the record, and 1988 and 1987 are second and third, respectively.

Considerable climate variation is also expected to occur on time scales of decades or longer as a result of the large thermal inertia of the oceans, which interacts with short-period fluctuations and accentuates the longer time scales. The effect of the oceans' thermal inertia can be estimated by forcing an appropriate climate model with random noise to reproduce the observed high-frequency (year-to-year) variations in global mean temperature.

It turns out that over the course of a century the resulting low-frequency temperature variations can be as large as .2 to .3 degree C. In other words, perhaps 50 percent of the observed warming trend in this century could be attributable to a natural internal fluctuation. An equally plausible interpretation of the data, however, might be that a much larger warming of .7 to .8 degree C has been partially offset by an internally generated cooling fluctuation.

The earth's climate also responds to various external factors. Solar variation is one such factor. Recent satellite observations have confirmed that the sun's output varies by about .1 percent in parallel with the 11-year sunspot cycle—a variation of about .24 watt per square meter in the radiation striking the upper atmosphere. If the climate system were able to respond immediately to changes in solar output, the globe would warm (or cool) by between .08 and .24 degree C over the course of the sunspot cycle. The oceans' thermal inertia prevents such a rapid response; the actual global temperature change is probably less than .03 degree C.

There has been speculation that the sun's luminosity may vary by larger amounts on longer time scales. Prolonged periods of low sunspot activity, such as the Maunder, Spörer and Wolf minima (which occurred during the periods of 1645–1715, 1450–1550 and 1280–1350, respectively), correlate with intervals of widespread mountain-glacier advances (the last one was the Little Ice Age, which lasted roughly from the mid-16th to the mid-18th century). It has been suggested that the sun's output declined by from .2 to .6 percent during such events and that these declines were responsible for the climate shifts. No prolonged periods of anomalous sunspot activity, however, have been observed since the end of the Maunder minimum.

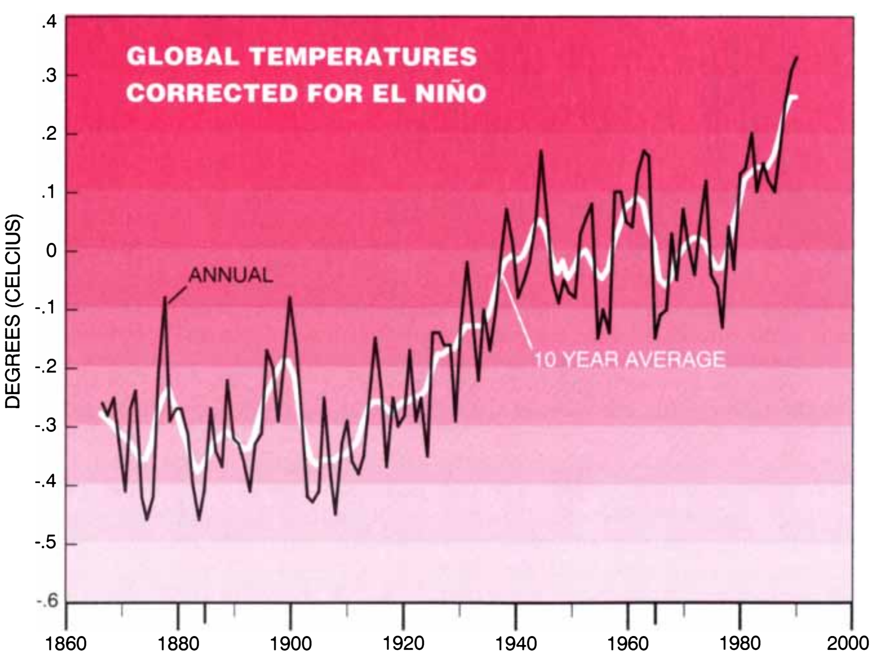
Changes in the sun's radius, which increases and decreases over a roughly 80-year cycle, may also affect luminosity, but the strength of the link between radius and luminosity is unclear. Although the effect could be significant, it could also be negligible. Satellite observations during the coming decade should resolve these uncertainties.

At present the impact of the sun on global temperature trends over the past 100 years or so is uncertain, but it appears to be small. Even the more speculative suggestions involve phenomena whose effects are insignificant compared with those of greenhouse-gas emissions. The most likely magnitude of the luminosity decline responsible for the Little Ice Age, although far larger than any variation observed recently, is only one watt per square meter, equivalent to about 40 percent of the greenhouse change that has occurred to date.

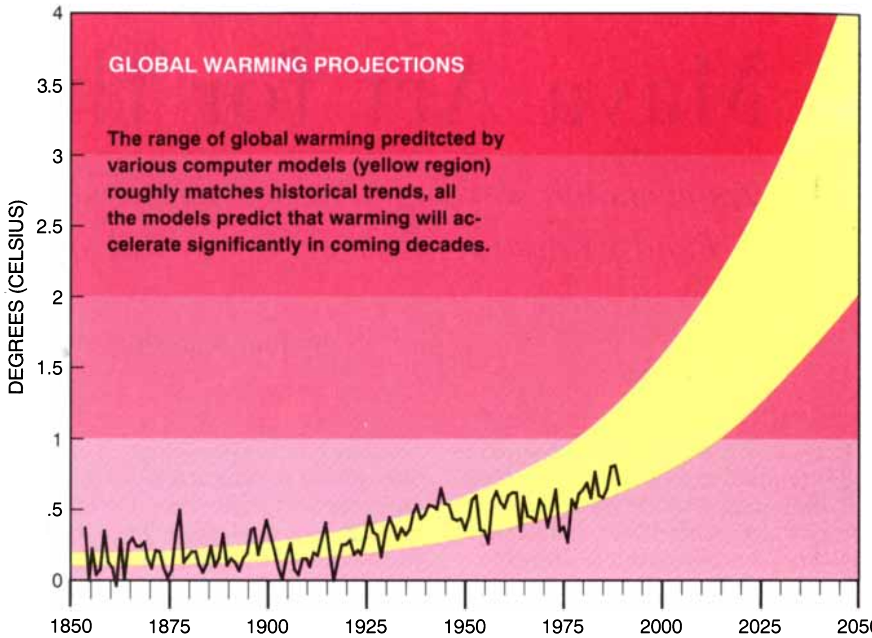
The effect of volcanic activity on climate change is more clear-cut, at least on short time scales. Volcanic eruptions that inject large amounts of dust and sulfate aerosols into the stratosphere can cause significant short-term cooling. The eruption of Krakatoa, near Java, in 1883, for example, appears to have cooled the lower atmosphere by a few tenths of a degree C. The effect began within a few months of the eruption and was measurable for nearly two years thereafter. The eruption of Agung in Bali in 1963, although it was less violent and injected less dust into the stratosphere, produced large amounts of sulfur dioxide and had a similar climatic effect.

The long-term effects of such eruptions are much more debatable. Even if volcanic aerosols precipitate out of the stratosphere within a few years, as seems likely, it could still be argued that the oceans' thermal inertia modulates the eruptions' effects in such a way as to give rise to long-term climatic effects. It might be possible, for instance, to attribute some of the warming that took place between 1920 and 1940 to the dearth of large eruptions during the period.

There are no continuous records of volcanic-aerosol concentrations—particularly sulfates—in the stratosphere, and so it is impossible to estimate reliably their long-term effect on climate. Although there are numerous proxies, including records of eruptions, of atmospheric transparency and sulfate concentrations in the Greenland and Antarctic ice cores, estimates derived



NATURAL FLUCTUATIONS in the temperature record may mask global warming. Here global temperatures are adjusted for the effects of El Niño, a disruption of Pacific weather patterns that lowers the average global temperature. Other natural events, such as volcanic eruptions (shown here by tick marks), may also have a short-term cooling effect on climate and thus temporarily hide global warming.



from these sources do not correlate well with one another. It is difficult therefore to draw firm conclusions about the impact of volcanic activity on long-term climate change.

What, then, of the greenhouse effect? At least in this case a good record exists of the changes in concentrations of greenhouse gases over the past few centuries. Since 1765, levels of atmospheric carbon dioxide have increased from around 280 parts per million by volume to more than 350 parts per million. Methane concentration has more than doubled, from 800 parts per billion to 1,700, and nitrous oxide has increased by about 10 percent, from 285 parts per billion to 310. During the past 30 years concentrations of chlorofluorocarbons have risen from essentially zero to about one part per billion.

Computer models predict that the change in the global radiation balance caused by increases in greenhouse gases is roughly equivalent to a 1 percent increase in the luminosity of the sun. The changes in the levels of greenhouse gases that have already occurred should eventually cause a global mean temperature increase of between .8 and 2.6 degrees C (the uncertainty can be attributed to a lack of understanding of various kinds of feedback in the climate's response to external forcing).

The thermal inertia of the oceans, however, prevents the climate from responding immediately to greenhouse forcing and has reduced the expected warming during the past century to somewhere between .5 and 1.3 degrees C. The observed warming of about .5 degree C is thus just barely consistent with that expected from the greenhouse effect.

The consistency between historical warming trends and the predictions of greenhouse models does not mean, however, that the greenhouse effect has been conclusively detected or that it is relatively small. Given the magnitude of natural climate variability and the other external factors that may affect the climate, the observed warming could still be attributed to factors other than the greenhouse effect, or, of course, a more substantial greenhouse-related warming could have been offset by other climate fluctuations.

Many characteristics of the historical temperature record, in fact, appear to conflict with the greenhouse hypothesis. The earth warmed more rapidly between 1920 and 1940 than greenhouse models would predict, and it cooled between 1940 and 1970 even as concentrations of greenhouse gases were increasing rapidly. Furthermore, the hemispheric records also conflict with expectations. Because the Southern Hemisphere has more ocean than the Northern Hemisphere, it should warm more slowly; however, the Southern Hemisphere has been leading slightly in the warming race. These discrepancies can be resolved at least qualitatively: the rapid warming during the early 20th century could be caused by internal factors or in part by lessened volcanic activity or changes in solar output, and the cooling between 1940 and the early 1970's could just be the result of natural variability again superposed on the greenhouse effect.

Most of the uncertainties surrounding the causes of recent climate change will never be resolved because the necessary historical data are lacking. As a result, it is impossible as yet to interpret accurately the undeniable global-scale warming that has occurred during this century.

The observed warming is at the lower end of the range predicted by greenhouse models, which may suggest that the greenhouse effect is smaller than predicted by current models. It is also possible, however, that the greenhouse effect is stronger than the models suggest and has merely been partially offset by natural variations in climate or other anthropogenic influences.

Together with advances in modeling, data to be gathered in the coming decades will reduce the uncertainties associated with the greenhouse effect and lead to better predictions of climate change. In the meantime, however, attempts to explain past variations in global mean temperature are bound to be frustrated by the lack of information about underlying causes of natural climatic variability.

Although this is obviously an unsatisfactory situation as far as policy implications are concerned, these uncertainties must not be used as excuses to delay formulating and implementing policies to reduce temperature increases caused by greenhouse gases. The longer the world waits to act, the greater will be the climate change that future generations will have to endure. A policy of inaction would be justified only if researchers were sure that the greenhouse effect was negligible.

FURTHER READING

ANALYSIS OF THE TEMPERATURE RECORD. T.M.L. Wigley, James K. Angell and P. D. Jones in *Detecting the Climatic Effects of Increasing Carbon Dioxide.* Edited by M. C. MacCracken and F. M. Luther. U.S. Department of Energy, Carbon Dioxide Research Division, DOE/ER-0235, December, 1985.

EMPIRICAL CLIMATE STUDIES: WARM WORLD SCENARIOS AND THE DETECTION OF CLIMATIC CHANGE INDUCED BY RADIATIVELY ACTIVE GASES. T.M.L. Wigley, P. D. Jones and P. M. Kelly in *The Greenhouse Effect, Climatic Change, and Ecosystems.* Edited by Bert Bolin, B. R. Döös, Jill Jäger and Richard A. Warrick. John Wiley & Sons, Inc., 1986.

GLOBAL TEMPERATURE VARIATIONS BETWEEN 1861 AND 1984. P. D. Jones, T.M.L. Wigley and P. B. Wright in *Nature*, Vol. 322, No. 6078, pages 430–434; July 31, 1986.

SUN AND DUST VERSUS THE GREENHOUSE. J. E. Hansen and A. A. Lacis in *Nature* (in press).

Maya Art for the Record

As economic development accelerates damage to Maya murals, hand-painted replicas serve as vital archaeological sources

by June Kinoshita

It was in 1900 that Adela Catherine Breton first rode into the archaeological camp at the great Maya temple city of Chichén Itzá. Bundled up in a Victorian riding costume with her severe, girlish face protected from the Yucatán sun by a low-brimmed hat, the 50-year-old Breton cut a striking figure. Over the next eight years, the Englishwoman time and again braved heat, ticks and foot-dragging bureaucrats to return to Chichén Itzá to sketch and paint the stone reliefs and colorful murals that enlivened the limestone walls of the ruins.

Today Maya scholars are grateful for her efforts. In the 80-odd years since her visits, many of the murals that Bret-

on so painstakingly copied at Chichén Itzá and other Maya sites have been lost. The steamy Mesoamerican climate and burning sun have exacted a heavy toll on the colors, particularly the intense, indigo-based azure known as Maya blue. Vandals and tourists have spirited away many fragments, and the surviving paintings are prone to attack by algae and vegetation. In many instances Breton's watercolors are the only records of these Maya relics. Her work was donated to the Peabody Museum at Harvard University and to the City of Bristol Museum and Art Gallery in England, which commemorated her achievements in a special exhibit this past winter.

Breton was not the only eccentric Victorian to explore the world armed with a sketchbook, but she was unusual for her rigorous attention to color. "The murals are packed with information carried by color," says Arthur G. Miller, professor of art history at the University of Maryland at College Park.

Ablaze in Color

Consider the now badly defaced battle scene on the south wall in the Upper Temple of the Jaguars at Chichén Itzá (A.D. 900–1100). Breton's watercolors show the circular shields of the combatants rimmed with red or blue to indicate the opposing armies. "Color is part of the cosmic view of the ancient Mesoamericans," says National Geographic Society archaeologist George E. Stuart.

The cardinal directions, for example, are color-coded: east is painted red, west black, north white, south yellow and center blue-green. "The color of deities' costumes show directional associations," Stuart explains. "For humans, color may signify rank or where they're from. The shapes of objects can be similar: without color it's hard to tell if something is jade or maize." Maryland's Miller notes that "most sculpture was also painted, but today we see only bare limestone. The murals give us a glimpse of what the sculpture looked like."

Not only murals and sculpture but whole cities were "ablaze in color," says Merle Greene Robertson, director of the Pre-Columbian Art Research Institute in San Francisco. Robertson should know. In 1973 she embarked on a thorough survey of color at Palenque. This classic-period city, located amid the lush rain forests at the foot of a chain of hills in northern Chiapas state, reached its apex in the reign of Pacal

***FESTOONED WARRIORS pay homage to a masked figure framed by the coils of a feathered serpent god in a stone relief from the Lower Temple of the Jaguars at Chichén Itzá* (left). *Adela Breton recorded the pigmentation by applying watercolors over a black-and-white photograph. At Chichén Itzá, Breton also painted watercolor replicas of the battle scenes on the interior walls of the Upper Temple of the Jaguars* (right). *Shown here is a detail of the south wall. The murals have been almost obliterated, as a 1984 photograph shows* (above).**

the Great (A.D. 615–683). It is renowned for delicate stucco reliefs and numerous hieroglyphic inscriptions.

Using a standard color chart, Robertson, whom fellow Mayanists describe affectionately as "the proverbial little old lady in tennis shoes," set out to document the vestiges of paint that once adorned the buildings and bas reliefs. According to Robertson, the walls were once painted in intense red, inside and out. Human skin was rendered in red, that of gods in blue.

When Robertson began her project, most of the painted surfaces were still clean, she recalls. But by the early 1980's they had become encrusted with "black scab," which forms when limestone reacts with acidic moisture. The unsightly encrustations obscured what little color was left. She attributes the black scab to acid rain, which she says wafts in from uncapped oil wells and smokestacks near the Gulf Coast cities of Coatzacoalcos and Carmen, 125 kilometers north of Palenque.

Robertson sounded the alarm on acid rain last year in a study for the National Geographic Society. In the famed Temple of the Inscriptions at Palenque, she reported, the black scab was "so bad that upon looking at them [the inscriptions], one seemed to be looking at a piece of flat, black sculpture." Seymour Z. Lewin, a chemist at New York University who has examined weathering patterns at several Maya sites, agrees that "the evidence is typical of the weathering produced by acid rain." But he adds that "paint suffers more from microorganisms and the leaching of salt" than from acid rain.

Robertson's survey of Palenque was nearing completion in 1982 when disaster struck: the volcano El Chichón, which had been dormant for 600 years, suddenly erupted, ejecting a dense cloud of ash and sulfuric acid droplets. The eruption dumped hundreds of tons of abrasive ash on Palenque, leaving it looking as though a freak blizzard had struck. Mary E. Miller, a Maya scholar from Yale University (no relation to Arthur Miller), says: "When the rains started, it scoured off the black goop, but it also took off the paint." Six weeks after the eruption, she recalled, "the whole site looked like it had been scrubbed with Bon Ami."

Fortunately, Robertson's records by that time were complete enough to enable her to reconstruct the original appearance of Palenque's halls in a series of vivid paintings. "I think Robertson's reconstructions are very good," Mary Miller says. "The colors seem hard, flat and excessively bright—and that's how they were."

Just as urgent as retrieving the lost color of the ruins is the task of recording the murals that still survive. Even the most celebrated of Maya paintings—the murals at Bonampak—have not yet been satisfactorily recorded, experts say. This remote, late classic site tucked away in the southeast corner of Chiapas state was first brought to the attention of the outside world by the English explorer Giles G. Healey, who was led there by local villagers in 1946.

Unwitting Obliteration

Dated to A.D. 800, the Bonampak murals cover the interiors of three stone chambers. In brilliant colors the realistic images depict a battle, its gory aftermath and the victory celebrations. At the time of Bonampak's discovery,

TEMPLE OF THE SUN at Palenque was colored red with hematite (iron oxide) and adorned with colorful friezes, as shown in this reconstruction by Merle Greene Robertson (left). *Today the seventh-century temple stands starkly white against the rain forest* (bottom). *In the photograph above, Robertson holds a swatch of color against traces of pigment on the cheek of Lady Zac Kuk, a queen of Palenque.*

the established view held that the classical Maya were a peaceful race who had no written history. Bonampak overturned that view and revealed the Maya for what they were: a warlike people who recorded their story in hieroglyphs. "Bonampak was a watershed in Maya studies," Mary Miller notes.

Because of their importance, the murals' precarious condition alarms many scholars. "The site was shrouded in deep tropical rain forest and was kept dark and damp for 1,000 years," Mary Miller explains. "The first thing they did after the murals were found was to cut down the trees and put a galvanized tin roof over the site. This made it very hot in the day and cold at night." In the 1960's conservators injected the walls with silicone, which held the paintings up "just as it would a woman's face," she observes. "But then the silicone flowed down, and the murals ended up being worse off." Conservators also unwittingly obliterated parts of the mural when they sealed several large cracks with cement.

Urgent Need

A few years ago the murals were cleaned. A milky layer of calcium carbonate had hardened over the painted surface, and when it was removed, the underlying scenes were revealed in their full glory. Ironically, the cleaning could hasten their destruction. "The reason they survived in the first place is because they were behind calcium carbonate," Mary Miller says. Now that the paintings are exposed, they remain subject to gyrating temperatures, humidity and microorganisms. "It's dangerous to have revealed them without providing better protection," she fears. "I'm concerned about their future."

There is consequently an urgent need to make a modern record of the Bonampak murals. Several have been made, in fact, including photographs taken in 1946 by Healey, another set by Hans Ritter and four reconstruction paintings, one a spectacularly detailed reconstruction of the first chamber. It was painted in the mid-1970's by Felipe Dávalos for the Florida Museum of Natural History in Gainesville.

Yet as good as some of these reconstructions are—particularly those by Dávalos—none is complete, scholars say. Except for the Dávalos reconstruction, "the hieroglyphs look like spaghetti," Stuart says. "It's hard to draw that stuff. It took me a year to learn to draw in the Maya art style." Breton put it this way in a letter to her colleague Alfred M. Tozzer: "Making drawings of them [Maya art] would require not

BONAMPAK MURALS as they appeared before their recent cleaning are shown in the photograph above. The murals had become obscured by a white crust of calcium carbonate, seen on the ceiling, and could be viewed only by wetting the crust with water or kerosene. White-robed nobles stand in the top row. Beneath them, musicians beat drums and shake rattles to celebrate victory in battle. A vivid replica (below) *painted by Felipe Dávalos for the Florida Museum of Natural History in Gainesville reconstitutes exquisite details of Maya vestments and hieroglyphs.*

modern artistic skill, but the very different capacity of seeing them as ancient Americans did."

According to Stuart, "the Bonampak murals in their entirety have never been properly photographed or drawn." That situation may soon change. Roberto Garcia Moll, director of Mexico's National Institute of Anthropology and History, has expressed interest in such a project. Mary Miller supports this move, saying that "records are more important than restorations."

New murals continue to be discovered throughout Mesoamerica. Last year Mexican archaeologists dug down to a buried staircase at Cacaxtla, a ruined palace complex 100 kilometers east of Mexico City, and uncovered a pair of breathtakingly fresh-looking murals on the two walls flanking the stairs. Along with two other murals that were found in the mid-1970's at Cacaxtla, the new find makes the group one of the most significant in Mesoamerica. The murals were created between A.D. 655 and 835 by the Olmeca-Xicalanca, a people of probable Gulf Coast origin, says Ellen T. Baird of the University of Nebraska at Lincoln. The

***BLACK-PAINTED WARRIOR** garbed in magnificent avian plumage glares from a painted palace wall at Cacaxtla* **(top).** *A turtle adorns another section of wall* **(bottom).** *The murals are thought to have been painted by Maya artisans hired by the lords of this Olmeca-Xicalanca site, 500 kilometers west of the Maya domain.*

murals are "a unique mix of Maya style and central Mexican motifs," she says.

The styles apparently became mixed, according to the late Donald Robertson of Tulane University, because the lords of Cacaxtla hired Maya artists; the artists painted in their native style but put in Highland Mexican glyphs in accordance with their patrons' wishes. Mary Miller marvels at the explicit rendering of the maize god in the newly found murals: heads of yellow maize bearing human faces poke out between green leaves on a stalk. "This is iconography for dummies," she says.

The Mexican government has built a large shelter over the entire site. Still, the murals unearthed in the mid-1970's have already faded perceptibly, observers say. Fortunately, they have been copied by a Mexican team. The newly discovered paintings at Cacaxtla were photographed this past winter under the auspices of National Geographic. These murals are in such superb condition that photographic records may be adequate, Mary Miller says. Less pristine murals may require hand copying as well. "This strikes many people as terribly old-fashioned, yet the eye can see things the camera can't pick up," she notes.

Nothing Romantic

Most important is the recording of "what is extant before it fades," Arthur Miller agrees. Few people, however, are willing to expend the time and effort to make such records. He notes that his own work on the murals at Tulum, a postclassic (15th- to 16th-century) complex perched above the sea on the eastern coast of the Yucatán, consumed four years. "It takes commitment, experience and painstaking effort," he says. "There's nothing very romantic about it."

Adela Breton would have sympathized. Copying murals is "very trying to brain and nerves as well as to eyes and hand," she complained to an acquaintance. Indeed, one reason the Bonampak murals have never been replicated completely may be that the artists simply ran out of steam. Dávalos's efforts on the second and third rooms do not begin to compare with his reconstruction of the first room. Another artist working on a Bonampak replica in Mexico City started with the third room but appears to have flagged by the time she got to the others.

Good replicas are increasingly important because many Maya sites are vulnerable to economic development. At Tulum the walls are black from soot belched by tour buses. Tourists, who bring in much needed foreign exchange, raise the humidity inside the stone chambers and exhale plant spores, which can take root on the walls. The oil wells and smokestacks operated by Pemex, Mexico's national oil company, may be contributing to acid rain. "There's a tension between conservation and economics," Arthur Miller observes. The solution, he suggests, is to make replicas and install them in place of the originals, which would be relocated to museums.

Even with safeguards, degradation is inevitable for ancient artifacts, especially painted ones. "Once you've found them, you may as well write them off," Stuart observes. "That's why we want to record them— for all time."

***COLORS FADE** as soon as paintings are brought to light. Two photographs of a battle scene mural at Cacaxtla, one made in 1978* **(top left)** *and the other in 1989* **(top right)**, *show the deterioration of indigo-based "Maya blue," whereas the red of hematite has remained stable. At Tulum, portions of a mural have been rubbed away by hands and blackened by air pollution, as documented in two photographs, one from 1966* **(bottom left)** *and the other from 1985* **(bottom right)**.

A catalogue of Adela Breton's watercolors is available from the City of Bristol Museum and Art Gallery in England.

SCIENCE AND BUSINESS

A Thin Line

Chip makers look to X rays to create denser circuits

Drive north from New York City along the leaf-decked Taconic State Parkway and eventually the road winds past East Fishkill, population about 23,000. There are no impressive historic monuments here; the biggest shop in town is the Grand Union grocery. But nestled in East Fishkill may be one of the U.S. semiconductor industry's brightest hopes for the future: an X-ray lithography research center being built by IBM. By the year's end IBM intends to have in place an electron-storage ring that will produce X rays. Researchers hope to harness those beams to produce the microscopic features of a new generation of integrated circuits by mid-1991.

Other semiconductor makers will be avidly watching IBM's progress. X-ray lithography is increasingly touted as essential for making breathtakingly complicated integrated circuits, such as dynamic random access memories (DRAM's) that will pack more than 256 million bits (megabits) of data onto a silicon chip somewhat larger than a thumbnail. Current lithographic techniques, which use visible light, will barely enable manufacturers to reach 64 megabits.

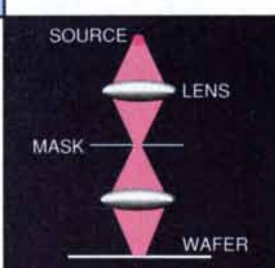
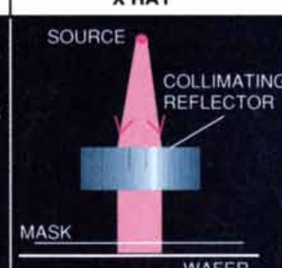
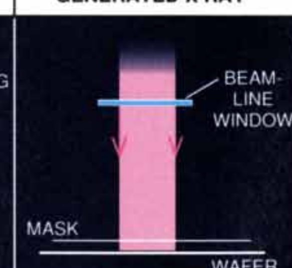
Animals to order, container reclaimer, flexible flyer

Although U.S. researchers have dabbled with X-ray lithography for almost 20 years, the technology is still embryonic. "X-ray [lithography] has to make it on process ease, simplicity, reliability, cost" and other characteristics, says Moshe J. Lubin, president of Hampshire Instruments in Rochester, N.Y., one of the few U.S. companies trying to make a go of selling X-ray lithography equipment.

For now X-ray lithography technology is neither cheap nor simple. And U.S. total investment in the field has been falling behind Japan's. During the past decade IBM has poured about $500 million into X-ray lithography. The rest of the U.S. funding for the technology during this time does not approach the IBM mark. (Funding for a research program initiated two years ago by the Defense Advanced Research Projects Agency [DARPA] amounts to $30 million in fiscal 1990.) In contrast, Japan's spending has been double that of the U.S., says Aubrey C. Tobey, a consultant in Burlington, Mass. Virtually all Japanese semiconductor makers are working with X rays. A recent Commerce Department report warned that based on current investment in semiconductor manufacturing technology, including X-ray lithography, Japan will be "the world's No. 1 electronics producer and trader by the early 1990's."

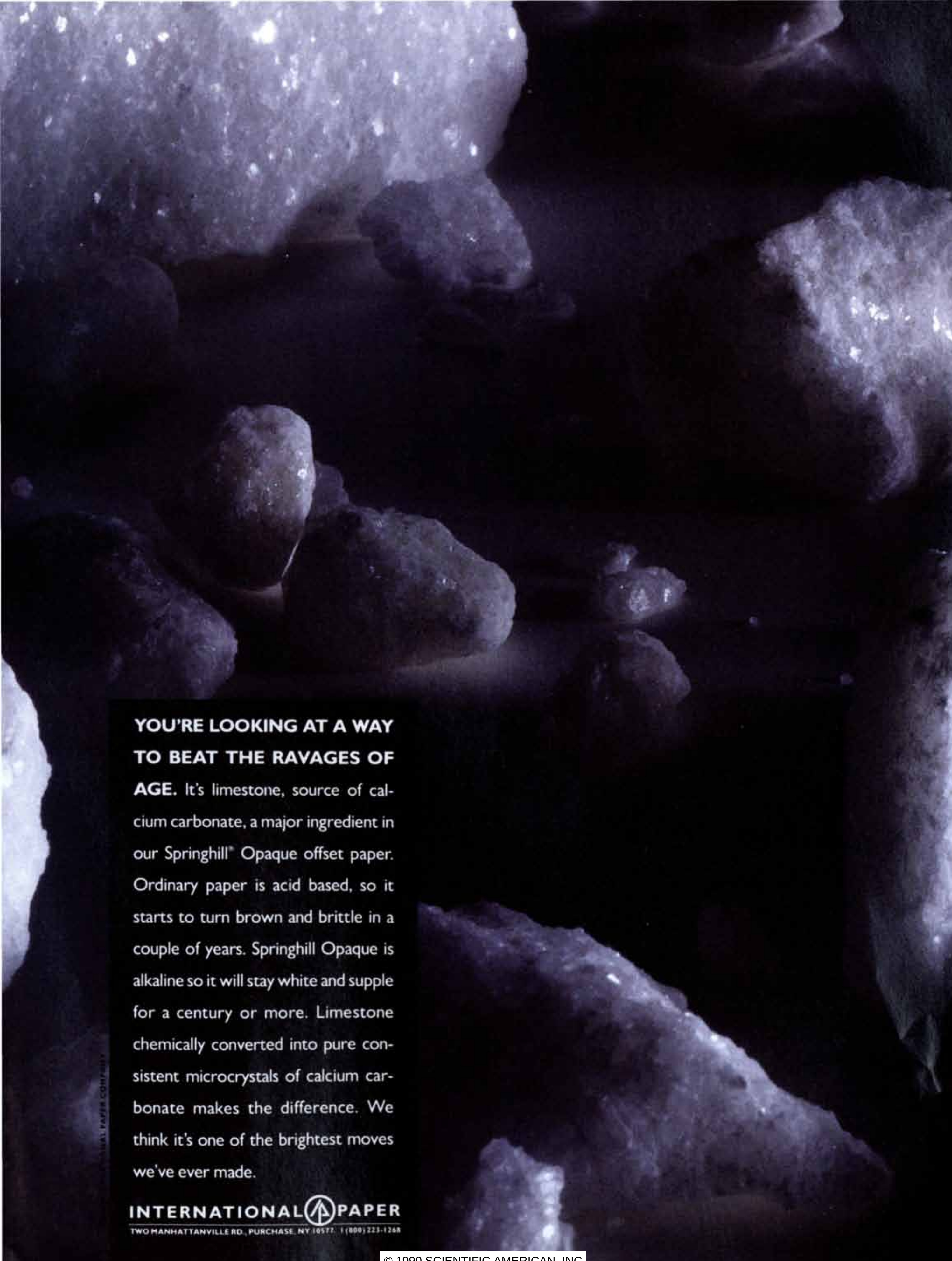
Optical technology and X-ray lithography share some common elements. Both require that manufacturers employ "masks" etched with the desired circuit design. Circular silicon wafers, which are eventually diced up into 100 or so chips, are coated with a light-sensitive material, or "resist." Workers then shine visible light or X rays through the mask onto the wafer, ex-

INTEGRATED-CIRCUIT MICROLITHOGRAPHY TECHNIQUES

	OPTICAL PROJECTION	ELECTRON BEAM (DIRECT WRITE)	LASER-GENERATED X RAY*	SYNCHROTRON-GENERATED X RAY*
EXPOSURE SOURCE SYSTEM				
SOURCE WAVELENGTH (NANOMETERS)	436 (g-line) 365 (i-line) 248 (deep ultraviolet)	20–50 THOUSAND ELECTRON-VOLT BEAM	2.2–.8	1.1–.8
MINIMUM FEATURE SIZE ON WAFER (MICRONS)	.65 .50 .35	.15	.15	.15
THROUGHPUT (WAFERS/HOUR)	40 40 20	4	20–40	40

Source: Hampshire Instruments

*prototype techniques


YOU'RE LOOKING AT A WAY TO BEAT THE RAVAGES OF AGE. It's limestone, source of calcium carbonate, a major ingredient in our Springhill® Opaque offset paper. Ordinary paper is acid based, so it starts to turn brown and brittle in a couple of years. Springhill Opaque is alkaline so it will stay white and supple for a century or more. Limestone chemically converted into pure consistent microcrystals of calcium carbonate makes the difference. We think it's one of the brightest moves we've ever made.
INTERNATIONAL PAPER
TWO MANHATTANVILLE RD., PURCHASE, NY 10577 1 (800) 223-1268

posing areas of the resist. Washing the wafer with a solvent dissolves the unwanted resist and leaves a copy of the mask design on the wafer. A finished chip will be a sandwich of as many as 20 circuit layers.

As circuit designs become tighter, the wavelength of visible light becomes as troublesome for printing circuits as a dull crayon is for drawing lines. Several years ago workers expected optical lithography to give out around .5 micron. (Four-megabit DRAM's, which can store four million bits of data and are the most advanced memory chips on the market, have circuit lines measuring about .8 micron.) But by tweaking resists and turning to excimer lasers for shorter wavelengths in the deep-ultraviolet range, researchers have pushed optical technology to .35 micron—tight enough for a 64-megabit DRAM, such as the prototype chip unveiled by Hitachi in June.

X rays, given their shorter wavelengths, should work well at .25 to .2 micron—the anticipated feature size of 256-megabit chips. Moreover, workers can print crisper lines with X rays than with visible light because the shorter wavelength means that light stays focused over a longer distance. "Look at the physics," says David O. Patterson, who manages the DARPA program, "and it looks like X rays are going to take us into the future."

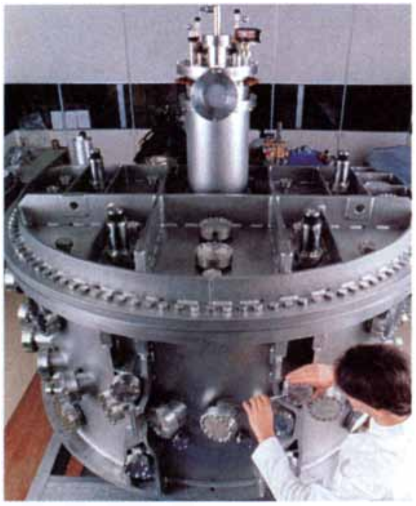
Generating X rays bright, or powerful, enough to print circuits is tricky. Researchers have typically employed electron-storage rings, or synchrotrons, that cost between $25 million and $50 million.

Swapping a laser for a synchrotron may be one alternative—if the laser proves sufficiently powerful. Hampshire Instruments, for instance, prices its units at roughly $3.5 million apiece. Hampshire's approach may be the most economical, particularly for small chip houses, says G. Dan Hutcheson, president of VLSI Research, a consulting firm in San Jose, Calif. Nevertheless, because chips have dozens of circuit layers, a manufacturer would need many laser units to produce adequate supplies of chips.

Laser-based systems can be technically fickle, too. To convey as much power as possible, the X-ray source must be near the mask and wafer. But because the X rays emanate from a single point, the light waves travel through the mask at skewed angles, casting distorting shadows on the resist. Hampshire has consequently developed a reflecting optic for straightening, or collimating, the light.

Even the shortness of the X-ray wavelengths themselves can cause headaches. In optical systems the circuit etched on the mask is often five times larger than the final chip design; a series of lenses focuses and shrinks the image. Because X rays cannot be focused by a lens (they are absorbed instead), manufacturers must use masks with the same circuit dimensions that they want on their final chips. Carving such tiny features onto a mask requires enormous precision. Mask development has consequently become one of the thrusts of the DARPA program.

Farther down the road, researchers believe there are ways around the mask difficulty. Workers at AT&T Bell Laboratories, for instance, are trying to focus X rays by bending the waves with specially coated mirrors and masks. Richard R. Freeman, who heads Bell Labs' electronics research department, has shown a 20-fold reduction in mask-to-final-circuit dimensions by using a particular arrangement of mirrors. In one experiment, his team printed features as small as .05 micron.



SYNCHROTRON X-ray source for making integrated circuits is being built by Oxford Instruments for IBM.

Still, the work has not yet reached the prototype stage, Freeman emphasizes, and researchers must conquer other challenges. For instance, at such short wavelengths, X rays do not penetrate a resist coating. As a result, Freeman's group must use two or three layers of different resist materials and bathe the wafer in a series of solvents to transfer images progressively deeper into the substrate. Although the technique neatly fixes the problem, it is a long way off from being "economical and bulletproof," Freeman adds.

Apart from X-ray lithography, other techniques for patterning dense circuits are emerging. Manufacturers rely on electron beams, for instance, to carve the masks used in lithography. Directly writing on a wafer with such a beam is feasible and precise—but excruciatingly slow because the electron beam must serially write every feature on a chip. Nevertheless, one U.S. electron-beam equipment manufacturer, Lepton in Murray Hill, N.J., is hoping to convince chip makers to supplement their optical lithography systems with an electron beam. Manufacturers could use the optical techniques for most circuit layers and turn to electron beams for particularly demanding and intricate layers, says Martin P. Lepselter, the company's president.

At this time, researchers agree that X-ray lithography seems the most alluring means for making future chips. The question is when will it gel into a practical technique. "We've had forecasts since the late 1970's saying that by the mid-1980's X-ray lithography would come on strong," Hutcheson notes.

Chip producers hope IBM's experimental manufacturing center in East Fishkill will finally usher in X-ray lithography. It will be the only such facility in the U.S. (Previous research on synchrotron-generated X-ray lithography has taken place in ad hoc workshops set up near existing synchrotrons, such as the one at Brookhaven National Laboratory.) Even Sematech, the semiconductor-manufacturing research consortium supported by industry and government, is putting most of its efforts into optical lithography.

IBM is also taking part in the DARPA program and offering to give U.S. semiconductor makers access to the East Fishkill center, says Martin C. Peckerar, who heads the nanoelectronics processing facility at the Naval Research Laboratory. So far only Motorola is actively working at the facility. Other companies, including Hampshire, Lepton and two new companies created from the ashes of Perkin-Elmer's semiconductor equipment business, are keeping tabs on IBM through the DARPA program. "Even IBM realizes that it can't stand alone," Peckerar says.

With work, X-ray lithography may become a useful tool at the tail end of the 64-megabit DRAM generation, sometime in the late 1990's, says Robert W. Hill, who manages IBM's advanced lithography system group. Its development will depend, however, on how quickly demand grows for more powerful memory chips. Nevertheless, Hill asserts: "If you want to stay competitive, you have to use the best tools to do it." —*Elizabeth Corcoran*

Plastic Goes Green

Suppliers, users tout a brave new era of recycled plastics

Plastics were created with the promise of better living through technology. Now the question of disposal has cast them in the role of villains. Incinerating them risks degrading air quality. Burying them takes up dwindling landfill space. More than 600 communities have proposed banning plastic packaging materials entirely. So plastics manufacturers are trying to put on an environmentally friendly face by touting a now familiar concept: recycling.

At this point, demand for recycled plastic far outstrips supply. In principle, plastics can be recycled essentially to their original form. Sam Hartwell, research associate at the Natural Resources Defense Council in New York City, doubts there would be a glut "even if 100 percent were recycled."

Plastic materials make up 7 percent by weight of the municipal waste stream in the U.S. The figure is comparable to percentages for glass and steel and far greater than that for aluminum. It translates into about 29 billion pounds of municipal waste plastic every year. Industrial activities contribute another 19 billion pounds.

By far the most extensively recycled plastic is polyethylene terephthalate. PET is used in soda bottles; industry sources claim a 12 percent recycling rate for it. PET can be recycled into a wide variety of items, including carpeting, road barriers, tennis-ball felt, twine, thermal stuffing, even automobile parts. Procter & Gamble is already marketing some products in recycled PET containers.

To tap into this promising market, Du Pont has formed a joint venture with Waste Management, Inc., in Oak Brook, Ill., to recycle PET and high-density polyethylene (HDPE, used for milk jugs). The venture has two plants running and plans three more by 1994. Altogether these plants will be able to produce 200 million pounds of recycled plastic per year. This may signal a significant trend, but that amount is still a drop in the bucket compared with the 1.5 billion pounds of PET and 6.5 billion pounds of HDPE disposed of every year.

Even more ambitious are attempts to begin large-scale recycling of polystyrene, the plastic commonly used in foam packaging. Some five billion pounds of polystyrene land in the garbage each year. A few years ago the recycling rate for polystyrene was essentially zero. Despite environmental and economic concerns, the number has grown only to a dismal 1 percent. The polystyrene industry predicts that the situation is about to change, however.

PLASTICS RECYCLING efforts remain small scale. But some manufacturers are turning soda and milk containers into a range of products, from carpets to fences.

In 1989 seven companies—including giants like Amoco Chemical, Dow Chemical and Mobil Chemical—formed the National Polystyrene Recycling Company (NPRC), whose goal is to achieve a 25 percent recycling rate for polystyrene by 1995. Chris Vanden Heuvel of the Polystyrene Packaging Council in Washington, D.C., is optimistic because the supply of polystyrene is concentrated at large fast-food restaurants and institutional dining halls. Harvard University, Boston College, Wang Corporation, McDonald's and a number of hospitals and public schools are cooperating with the NPRC's pilot recycling center in Leominster, Mass.

Manufacturers are also starting to respond. Rubbermaid sells office products composed of recycled polystyrene. Dolco Packaging received approval from the Food and Drug Administration to manufacture egg cartons from recycled polystyrene—the first use of recycled plastic for a food container. Other products ready to be marketed include foam insulation and plastic trash receptacles.

But large obstacles stand in the way of more extensive recycling. Polystyrene containers are often covered with food waste that can contaminate the plastic and create emissions during recycling. Furthermore, the bulkiness of the material tends to make transportation costs high.

As a result, Hartwell thinks "polystyrene recycling is unlikely to ever reach a large scale." Adds Jan Beyea, senior staff scientist of the National Audubon Society: "It is hard to imagine recycling growing more than 5 percent a year. But the rate of garbage production is growing faster than that in many places." He also notes that companies hit an unpleasant "capital hump" when they begin creating the infrastructure for recycling.

For now the driving force behind plastics recycling is political, not economic. The environmental effect of bans on plastic packaging is debatable—in Portland, Ore., a ban on polystyrene fast-food containers has led to the use of nonrecyclable polyethylene-coated cardboard—but the public mood has undeniably forced corporations to make plastic more politically palatable. Beyea likens the situation to that now in Eastern Europe: "The public pressure is so strong that companies just can't hold back." —*Corey S. Powell*

CUSTOM COWS: the first calves genetically engineered in the U.S. contain implanted genes that may make them grow faster and leaner. Photo by Carter Smith.

Down on the Pharm

Animal developers seeking better beasts breed controversy

The four calves frolicking on a ranch near Houston, Tex., look and act like any others. Appearances can be deceiving. Each contains genes from other species intended to make the animals grow faster and leaner. The genes were inserted into embryos by a team headed by Bert W. O'Malley, a cell biologist at the Baylor College of Medicine.

Support for the research, however, came from Granada BioSciences, a Houston livestock breeder. To Granada and a growing number of companies, such genetically altered, or transgenic, animals spell future profits. Although the effort is still in the early stages—and the work is controversial—a group of small companies are pursuing the technology: Transgenic Sciences, DNX, GenPharm International and Embrex. In addition, Du Pont, Merck, Genentech and Genzyme are building transgenics programs.

Manipulating mammalian genes has broad potential. Livestock breeders believe that the ability to implant new genes in animals will provide a faster way of accomplishing what traditional breeding programs do—namely, creating animals with better-quality meat and improved resistance to disease. Other companies are adding human genes to laboratory animals to produce "models" of human diseases. Transgenic goats, rabbits and mice are also serving as living drug factories, producing pharmaceutical proteins in their milk.

The most immediate impact of transgenic animals has been on modes of medical research. Instead of finding naturally occurring animal counterparts of human disease or inducing such conditions with surgery or drugs, transgenics researchers can breed animal models with the very genes that cause an illness. Mice have been turned into subjects for studying many cancers as well as viruses such as hepatitis. Mouse models have also been developed for gene-defect disorders such as Duchenne's muscular dystrophy and multiple sclerosis.

Most of the near-term applications will be slightly better models of diseases in which a single gene is responsible or at least critically involved, as in breast or colon cancer. What medical researchers are most anxious to have, however—small-animal models of diseases such as Alzheimer's or AIDS—will require implementation of complex, multiple gene changes, something beyond the reach of scientists now. "You can't dial a gene. It's still a somewhat hit-or-miss process," notes James P. Sherblom, chief executive officer of Transgenic Sciences in Worcester, Mass.

To help medical researchers develop transgenic laboratory animals, the National Institutes of Health in May awarded a five-year contract to DNX in Princeton, N.J., to establish and operate a National Transgenic Development Facility. The company, which owns the patent on the most widely used technique for injecting genes into animal embryos, will aid scientists in creating their own transgenic research mice.

A hundred researchers a year will each pay $750 for DNX's best efforts to create transgenic mice containing their single genes. The NIH will subsidize the rest of the real cost of $7,000 to $10,000 per gene. "Until now people have been relying on friends and colleagues at other institutions that have facilities," says National Institute of Child Health and Human Development project manager Joel M. Schindler. "We felt it was appropriate to invest public dollars to support a facility for this common enterprise."

The government is also involved in an effort to use transgenic mice in carcinogenicity testing. The National Institute of Environmental Health Sciences (NIEHS) in Research Triangle Park, N.C., is testing mice carrying genes known to be associated with cancers to see if they are more sensitive to carcinogens than their unaltered brethren. "We're creating a much larger target for the action of chemicals," explains Raymond W. Tennant, chief of cellular and genetic toxicology at the NIEHS.

The mice in question were developed in Philip Leder's laboratory at Harvard University and, in April of 1988, became the first higher life-forms to receive a U.S. patent. Du Pont, which licensed the mice from Harvard, already sells them for cancer research for about $50 apiece.

Several biotechnology companies are also putting transgenic animals to work producing human proteins that can be used as drugs. Most efforts so far have focused on directing genes to mammary tissue so that the desired substances are secreted in the animals' milk. The idea behind turning large animals into drug "pharms" is not to be first to make a drug but to be able to make it at lower cost than companies using bacteria or cell cultures. Pharmaceutical Proteins in Edinburgh, Scotland, uses sheep to make blood-clotting factors that hemophiliacs rely on. Genzyme in Boston plans to use goats to manufacture growth hormone and interferon. Transgenic Sciences favors rabbits, in part because Food and Drug Administration regulators are familiar with them as laboratory animals.

Probably farthest off, and raising the most ethical issues, are altered livestock. Some of the early experiments on inserting new genes into animals done at the University of Pennsylvania resulted in utterly miserable creatures. Descendants of the first pigs born six years ago with a gene to make excess growth hormone now sprawl in stables at the U.S. Department of Agriculture's research station in Beltsville, Md. Their meat is leaner, as expected, but the pigs suffer from arthritis, lethargy and low sex drive.

The problem can be solved, Transgenic's Sherblom declares, if the gene is regulated and not simply left turned on to crank out hormone continuously. He says experiments his firm has done in mice suggest the mechanism can be controlled, and Transgenic plans to repeat the work in pigs. "We're likely as a society to continue to eat meat. If we're going to do that, it should be good for us," Sherblom declares.

Even so, the experiments gone awry raise important ethical questions. Can a mouse that twitches all the time from Parkinson's disease lead a reasonably comfortable life? Is it right to riddle an animal with tumors? "There's a lot of gray in this stuff. It ain't all black and white," comments Franklin M. Loew, dean of the Tufts University School of Veterinary Medicine. If a mouse or pig is changed substantially enough, does it become deprived of its intrinsic "mouseness" or "pigness"? "At least now we're not able to do enough genes to make that a possibility," Loew observes, but cynics say it won't be long before that possibility is very real.

Ethicists and animal rightists are not the only groups criticizing gene transfer. Some family farm groups fear the expense of engineered animals could force them out of business. Wildlife experts foresee trouble if animals with a modified genome escape into the wild. Research physiologist Vernon G. Pursel, who studies the so-called super swine at the USDA, characterizes the current state of transgenics technology in this way: "We're three or four years past the development of the first airplane, but we want to apply what we know to make a 747."

To Granada, those new calves look like a step in the right direction. If they remain healthy and produce higher-quality meat, the company hopes within a decade it will be clear that the benefits of transgenic breeding outweigh the drawbacks. *—Deborah Erickson*

Solar Sells

Flexible, lightweight solar cells spark competition

Sunnyside up, please, Sanyo might have said if ordering weather for the planned July 1 takeoff of its solar-powered midget airplane. The cross-country flight from San Diego to Kitty Hawk was a publicity stunt to flaunt the company's new flexible, lightweight solar cells.

Sanyo plans to incorporate these solar cells into all manner of consumer goods, beginning with portable power packs that could be for sale in a year or two, says Yasuo Kishi, general manager of the firm's functional materials development center in Osaka. Beach umbrellas, camping tents, sails and even clothing could all be made to generate power, Sanyo predicts.

The solar cells are made from amorphous—or noncrystalline—silicon, a less costly but less efficient alternative to cells cut from single crystals of silicon. Until now the light-converting silicon coating has been applied to rigid, rather heavy backings, including glass, stainless steel and thick plastic. The cells have been put to use mainly in low-power hand-held calculators and watches. In the new material, however, the silicon is applied to both sides of a flexible plastic film, which is then slipped between two other pieces of plastic to protect it. The result is a flexible yet lightweight solar-cell bank.

Although Sanyo insists that its technology does not infringe any patents, others are not so sure. Credit for pioneering amorphous silicon rests firmly with Stanford R. Ovshinsky, founder of Energy Conversion Devices (ECD) in Troy, Mich. ECD and its subsidiary, Sovonics Solar Systems, hold 128 patents on amorphous silicon and techniques for making it. Moreover, Sovonics has been developing its own flexible solar cells, about the thickness of two plastic garbage bags, for four years. The company says it is interested in examining Sanyo's offering.

But Sovonics' marketing skill does not measure up to its technological expertise. Sovonics has not yet shown a profit—nor has its parent after 30 years of operation. And its flexible material is far from being commercially available. Instead of aiming the stuff at the consumer market, Sovonics decided to pitch it to aerospace companies and charge a premium price of $100 to $200 per watt. So far Sovonics has delivered just one batch of the thin, golden sheets to Lawrence Livermore National Laboratory. ECD is hopeful that its latest joint venture, formed with Canon, Inc., of Japan on the very day of Sanyo's test flight, will help it compete in the marketplace.

Success will require both companies to pick their applications with care. Amorphous silicon cells are not very efficient at converting sunlight into electricity. In fact, exposure to light makes the material's efficiency drop off by 10 percent in the first week of use and another 5 percent within a month. Sanyo says the flexible cells in its plane have a 4 percent efficiency—compared with about 20 percent for crystalline silicon cells. Sovonics claims its premium product is 5.5 to 6 percent efficient. So even though the new cells are lightweight, a large area is required to generate adequate power. Sanyo's 198-pound glider has a 57.4-foot wingspan, covered with solar cells. These will generate the 2,000 watts of power needed to keep the lightweight composite airplane aloft at cruising speeds of from 40 to 99 miles per hour.

Consumers don't care about efficiency, insists Byron L. Stafford of the Solar Energy Research Institute: "They're more interested in power. If it gives enough, great." And Sanyo agrees: it is convinced it can find enough applications that could be served by manageable areas of cells even given the low efficiency.

Another potential obstacle to commercialization is durability. "Nobody puts calculators in a rainstorm, and when they stop working you throw them out and get another. But if you go to more expensive products, you want them to last longer. That requires a lot of engineering," Stafford says. Sanyo claims its product works fine at temperatures ranging from –40 degrees C to 100 degrees C, but the company has not had it long enough to know how it will stand the test of time.

Still, Sanyo is betting that the strategy that popularized its first generation of rigid solar cells—introducing low-end, mass-market consumer products—will work again. If higher production volume lets it drop prices, the company may be able to move into more demanding applications. Amorphous silicon solar cells could yet have their day in the sun. Maybe what it takes is flexibility. *—D. E.*

SOLAR AIRPLANE is powered by sheets of flexible, lightweight solar cells.

THE ANALYTICAL ECONOMIST

Competing advantages

"What is good for the country is good for General Motors, and what's good for General Motors is good for the country."
—CHARLES E. WILSON, to the Senate Armed Forces Committee, 1952

Almost 40 years ago Wilson's logic seemed unassailable. Today economists and policymakers aren't so sure. They wonder if ostensibly "American" corporations have become so international that their strategies work against the interests of the nation as a whole. They also ask what the U.S. government should do to improve national competitiveness.

At one end of the political spectrum are the technowarriors fighting on Capitol Hill for all manner of government support for domestic industries. Among the most recent efforts are bills aimed at freeing joint manufacturing projects from the treble damages that usually accompany antitrust suits. The bill passed by the House of Representatives in June would grant that privilege only to groups in which more than 70 percent of the companies are American; the Senate is considering a broader version.

On the opposite end of the debate is Michael E. Porter, a professor at Harvard Business School. His recent book, *The Competitive Advantage of Nations,* is an 855-page discussion of what makes companies internationally competitive. Porter's message: the role of governments has been vastly overstated. "Governments do not control national competitive advantage; they can only influence it."

Porter begins with the tenet that companies—not countries—compete. "It is far from clear what the term 'competitive' means when referring to a nation," he declares. Governments worry about the success of their national industries only because the fate of those firms directly affects the standard of living at home.

Precisely what contributes to the success of firms takes up most of Porter's pages. He postulates four primary influences on companies: *production factors,* namely, the availability of smart workers and useful equipment; *home demand,* or how sophisticated local buyers are; *related and supporting industries,* which will spur the firm to produce better or cheaper products; and *firm strategy, structure and rivalry,* or essentially how the company is managed. Influencing these elements is the company's homeland.

Consider, for instance, the thriving patient-monitoring equipment industry built in the U.S. after World War II. The industry sprung from a collection of national characteristics: a growing electronics industry that produced bright and energetic engineers, increasing funding for medical research, calls for high-quality hospital care and a welter of small firms scrambling for customers, to name a few.

The role of government has been indirect at best, Porter insists. And sometimes the results are surprising. Regulatory hurdles, rather than costly burdens on firms, can be the stuff of spring training—and force companies to innovate in ways that eventually make them more competitive. "Sweden's tough standards for product safety and environmental protection, for example, have been a significant source of competitive advantage in a variety of industries," he writes.

Porter is no fan of easing antitrust restrictions, either. Allowing erstwhile competitors to work together can blunt their will to compete, he argues, and in this way hurt companies' ability to compete globally.

He agrees with those who suggest that the most compelling advantages a nation can offer are highly educated researchers and a well-schooled work force. Yet companies, he asserts, must lend a hand in improving local conditions. Firms must train workers. They should also demand high quality from domestic suppliers, provide top-notch tools to others and be good competitors. "The firm has a stake in making its home base a better platform for international success," Porter states.

But if a company takes a hard look at what advantages its home country offers and decides the factors are not helping it compete, the firm should move, Porter says. Putting research and development centers near smart workers, savvy customers and top competitors makes sense.

Government leaders keen to keep or attract high-technology companies must look at their nation's business environment in a new light, Porter argues. Instead of shutting out foreign investment, nations should welcome tough, competitive companies that will challenge local businesses. The U.S. will gain from such "globalization" if foreign firms find advantages for their businesses here and establish vibrant research and manufacturing facilities.

So far direct foreign investment for acquiring or establishing businesses in the U.S. has been strong; annual outlays increased an average of 55 percent between 1983 and 1988, reaching a high of $72.7 billion in 1988, according to the Bureau of Economic Analysis. That figure slipped to $64.6 billion last year, however, possibly because of the government's scrutiny of foreign takeovers and calls for extensive financial disclosures by foreign firms.

Although not all economists agree with Porter's views, many are beginning to see the need for a fresh look at how U.S. regulations treat "U.S." and "foreign" firms. "Public policymakers need to recognize that Honda U.S. can serve U.S. interests as well as Ford can," says Proctor P. Reid, who is directing a study on globalization at the National Academy of Engineering, due out this fall.

Asking such questions could lead to a paradigm shift—a radical revision of the relations between governments and corporations. Yet at least one economic adage still holds from the days of David Ricardo, who penned the principles of comparative advantage in the early 1800's. As Porter puts it: "No nation can be competitive in (and be a net exporter of) everything."

—Elizabeth Corcoran and Paul Wallich

THE COMPARATIVE ADVANTAGE OF NATIONS

GERMANY	ITALY	JAPAN	SOUTH KOREA	U.S.
AUTOMOTIVE 16.1*	AUTOMOTIVE 5.0	AUTOMOTIVE 23.8	CARGO SHIPS 16.1	AUTOMOTIVE 10.3
CHEMICALS 3.3	SHOES 4.7	TV AND VCR 8.1	SYNTHETIC-FIBER TEXTILES 7.2	AIRCRAFT 8.0
MACHINE TOOLS 2.2	JEWELRY 2.9	ELECTRONICS 6.6	IRON AND STEEL 6.4	COMPUTERS 6.3
PRINTING PRESSES .5	WORKED STONE .9	CAMERAS .9	OTHER TEXTILES 3.7	CORN AND SOYBEANS 4.3

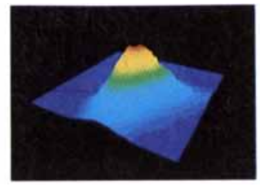
[shaded] SIGNIFICANT SHARE OF WORLD MARKET (shaded: Germany—Printing presses; Italy—Shoes, Jewelry, Worked stone; Japan—all; South Korea—Cargo ships, Synthetic-fiber textiles; U.S.—Aircraft, Computers, Corn and soybeans)

* percentage of national exports, 1985 data

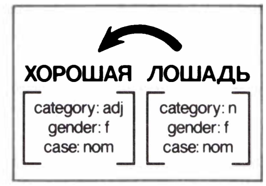
Last Year, This Is How They Won The IBM Supercomputing Competition.

1st Place: Gas dynamics and star formation in merging galaxies. Kevin M. Olson, University of Massachusetts.

1st Place: 3-D reconstruction of cochlea structure. Carl S. Brown and Alan C. Nelson, University of Washington.



1st Place: Stretching and bending of material surfaces in turbulence. Stephen B. Pope and Sharath Girimaji, Cornell University; Pui-Kuen Yeung, Pennsylvania State University.



1st Place: A dependency parser for variable word order languages. Michael A. Covington, University of Georgia.

This Year, This Is How You Can.

For information and an entry form, write to:

Supercomputing Competition Administrator
IBM Corporation
36 Mill Plain Road, Suite 404
Danbury, CT 06811

Name ______________________

Affiliation ______________________

Address ______________________

City ______________ State ________ Zip ________

Phone ______________ FAX ______________

SAM

Call For Papers

Last year, the response was overwhelming. So this year, the IBM Supercomputing Competition, with a new division, returns.

Created to discover, acknowledge and reward gifted American and Canadian scientists and engineers, this prestigious competition honors those whose papers describe innovative approaches to large-scale analysis and modeling.

Each paper must describe work using an IBM 3090™ Supercomputer and/or the paper must describe distributed or cooperative processing in which IBM technical workstations are linked to IBM mainframes.

The papers will be judged in five divisions: Physical Sciences and Mathematics; Engineering; Life and Health Sciences; Social Sciences, Humanities and the Arts; and Computer Sciences—Distributed and Cooperative Processing (new division). Each division will receive a first, second and third prize of $25,000, $15,000 and $10,000, respectively. Also, a "PROCEEDINGS" of selected papers will be published.

Contestants must register an abstract of their paper by October 16, 1990. Those who have registered their abstracts must forward their papers by January 15, 1991.

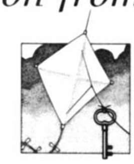
To select winners, IBM will retain a panel of independent authorities in each division. Winners will be announced by March 30, 1991.

For more information, contact your IBM marketing representative or call the IBM Supercomputing Competition Administrator at (203) 794-1355, FAX: (203) 792-7507 in the U.S.; (416) 758-4136 in Canada.

IBM®

THE AMATEUR SCIENTIST

How to monitor ultraviolet radiation from the sun



by Forrest M. Mims III

Most ultraviolet radiation that bombards the earth from the sun never reaches the planet's surface because of a thick, vacuous blanket of pale-blue toxic gas known as ozone. Were it not for this chemical shield, ultraviolet rays would strike with such intensity that most organisms would perish. As human activities and such natural events as volcanic eruptions alter the composition of the atmosphere, it would be wise to monitor the ozone layer and ultraviolet radiation to observe possible changes.

Thanks to many ground stations and several satellites, workers have learned much about the density and distribution of ozone in the atmosphere. Yet investigators lack a comparable network to observe the ultraviolet radiation that seeps through the ozone layer. Aside from instruments operated by the Smithsonian Institution and several other organizations, the only network of ultraviolet monitoring stations in the U.S. comprises fewer than two dozen Robertson-Berger meters.

These devices are designed to detect the wavelengths of ultraviolet radiation that cause erythema, or reddening of the skin, and eventual sunburn. Erythema develops most rapidly when the skin is exposed to ultraviolet radiation with a wavelength near 300 nanometers. This wavelength falls within the ultraviolet-B spectrum, which extends from 280 to 320 nanometers.

Since 1974 the average flux of ultraviolet B has been measured with a network of eight Robertson-Berger meters. From 1974 to 1985 the average flux fell some .7 percent per year. Because stratospheric ozone over the network decreased about .3 percent per year from 1978 to 1985, an increase in ultraviolet B would have been expected.

The decrease in ultraviolet-B flux was probably related to the fact that all the Robertson-Berger meters were located in urban areas. An independent investigation had shown that meters in some rural areas received from 5 to 7 percent more ultraviolet-B flux than urban meters. Another study has revealed that since 1981 ultraviolet-B flux has increased in the remote regions of the Swiss Alps. Were the results of the 1974–1985 study skewed by urban air pollution, which often contains gases and particles that absorb or scatter ultraviolet rays?

Here is where the amateur scientist can make a valuable contribution. With a little effort, you can construct an ultraviolet-B radiometer to record daily the flux of radiation. Comparing your observations with those of others from different regions would provide important information about how air pollution affects ultraviolet B.

Before building a radiometer for this purpose, one needs to understand how ultraviolet-B radiation travels through the atmosphere. Some ultraviolet-B rays scatter off air molecules; the remainder penetrate directly through the atmosphere. The sum of scattered and direct ultraviolet radiation is called global radiation.

Global radiation is of high interest in studies of the deleterious effects of ultraviolet on both living systems and materials such as paints and plastics. Global measurements are also helpful in determining how clouds affect ultraviolet B. (Robertson-Berger meters measure global ultraviolet B.)

Measurements of direct ultraviolet radiation yield valuable information about the presence and effect of absorbing and scattering agents in the earth's atmosphere. Because of the unpredictable nature of clouds and the presence of such barriers as buildings and trees, measurements of direct radiation are preferred to global ones for comparing the effects of air pollution on the relative magnitude of ultraviolet B at two or more locations.

So should the amateur observer monitor global or direct ultraviolet B? I recently discussed this question with John E. Frederick of the department of geophysical sciences at the University of Chicago. Frederick has devised a computer model that predicts the levels of ultraviolet B at the earth's surface for a range of conditions. He suggests that the amateur first concentrate on measuring direct ultraviolet B, since it is less affected by the many variables that can impair global measurements. The instruments described below, therefore, are both designed to detect direct ultraviolet B.

An ultraviolet-B radiometer requires a detector and a means for selecting the wavelength to be detected. The detector signal is amplified and transmitted to a digital voltmeter, an analog chart recorder or a computerized data-acquisition system.

Wavelength can be selected either with a monochromator or an optical interference filter. A monochromator provides a convenient but expensive way to measure ultraviolet B across a wide range of discrete wavelengths. An optical interference filter offers a much cheaper and more compact method for selecting a reasonably narrow band of ultraviolet-B wavelengths. Interference filters also allow considerably more radiation to reach the detector.

Ultraviolet-B interference filters, however, transmit a slightly wider band of wavelengths than do monochromators. Moreover, interference filters transmit low but detectable levels of radiation outside their specified bands, which can cause significant errors in measurement. An ultraviolet detector that eliminates a filter's secondary bands is said to be solar-blind.

Some of the various radiometers I have designed and assembled are solar-blind. One of them is relatively easy to build because its detector incorporates both an interference filter and amplifier. The detector is a DFA-3000, made by EG&G Judson (221 Commerce Drive, Montgomeryville, PA 18936). A single detector costs $125. A calibrated detector, which costs another $75, will enable you to make absolute measurements of ultraviolet B. But even with an uncalibrated detector you can monitor relative trends in ultraviolet B.

Although the DFA-3000 detector greatly simplifies the assembly of a radiometer, its ultraviolet filter transmits a low-amplitude band of red light, which is much nearer the detector's peak spectral response than the ultraviolet wavelengths it is intended to detect. When the sun is high in the sky, the red response is perhaps 10 percent of the detector's signal. (Later I will describe a simple method for correcting this effect.)

The only electronic components necessary to transform a DFA-3000 detector into a 300-nanometer radiometer are a resistor, a potentiometer (variable resistor) and two nine-volt batteries [*see illustration at left below*]. The DFA-3000 contains a silicon photodiode and an operational amplifier. The diode generates a small electric current when radiation strikes its active surface. The amplifier transforms the current into a voltage, which equals the product of the current and the resistor labeled R1 in the diagram.

Because the current produced by the photodiode can be less than 10 millionths of an ampere, R1 must provide a resistance of at least 10 million ohms (or 10 megohms) to provide an output of a volt or so. I have found that 30 megohms (three 10-megohm resistors connected in series) provides a satisfactory resistance for R1, but some readers may want to increase it because the ultraviolet-B flux is weaker north of my latitude (29 degrees 35 minutes north). EG&G Judson suggests that up to 200 megohms can be used. At very high resistances, however, considerable care must be exercised to preclude false signals caused by leakage currents between the detector's inverting input pin (9) and ground (1). A thin film of dust, moisture or oil may provide a path for an error-generating current. Eltec Instruments, Inc. (P.O. Box 9610, Daytona Beach, FL 32020), manufactures miniature resistors of from 10 to 100 megohms.

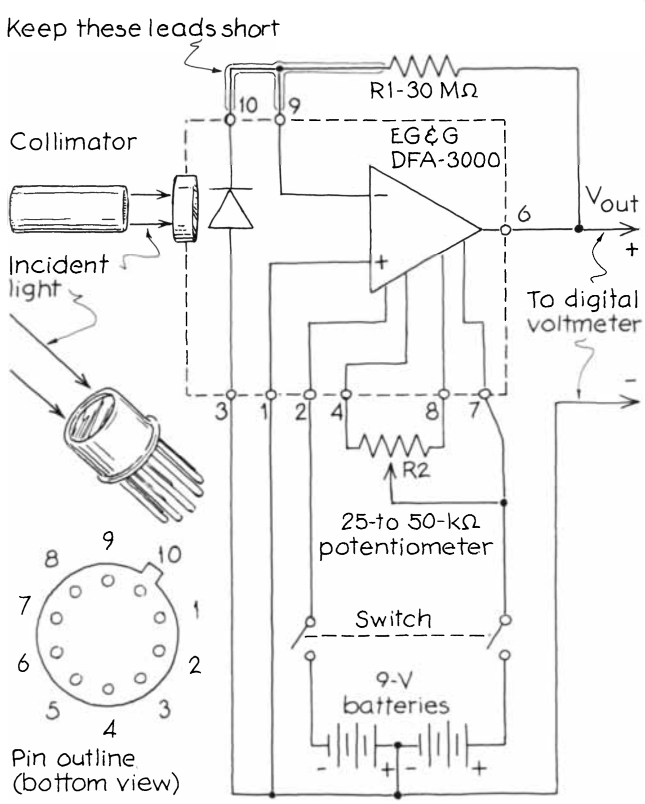
I installed the radiometer in a pocket-size plastic enclosure purchased from Radio Shack (part number 270-291). If you are inexperienced at assembling electronic circuits, you will find it easier to use a larger enclosure. Keep in mind that the enclosure should be as light-tight as possible. Light leaking through the enclosure may cause a false signal, because the base of the detector may not be totally opaque.

Because I installed the detector in a small enclosure, it was necessary to incorporate a potentiometer that can be adjusted with a miniature screwdriver. I connected the resistors to the detector with wrapping wire, a thin wire wound around component pins or leads by a special tool. If you prefer, you can solder standard wires to the pins of the components. It is important to keep the connections between the input of the amplifier (pin 9), the photodiode (pin 10) and R1 short and direct. To provide electrical contact with the nine-volt batteries, use a pair of battery clips. Inspect the wiring carefully before connecting the batteries to the circuit.

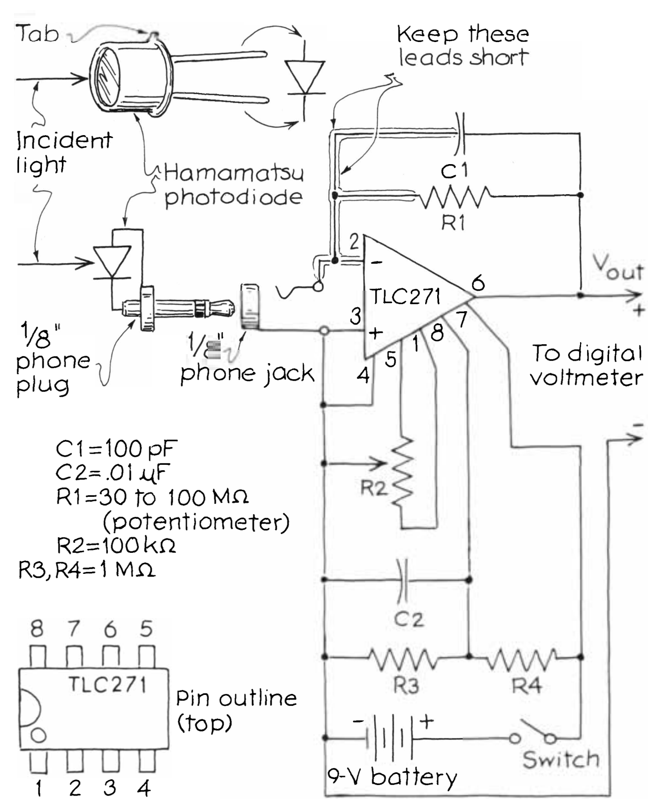
The unit I assembled includes a pair of output leads equipped with pin jacks that receive the probes from a miniature digital Multimeter. You may prefer to install both the radiometer and a digital voltmeter module inside a larger enclosure.

Because the radiometer is designed to measure the direct radiation from the sun, a collimator is required to restrict the detector's field of view. Thin-walled brass tubing from a hobby shop works well. A tube with an outside diameter of one centimeter should slip over the detector. Wrap a layer of tape around the detector if it fits too loosely in the tube. Coat the inside of the tube with flat, black enamel. A tube around 90 millimeters long will provide a field of view of approximately four degrees when the tube is pushed all the way to the detector's base.

Before installing the collimator, you should clean the surface of the detector's filter since dust and oil absorb

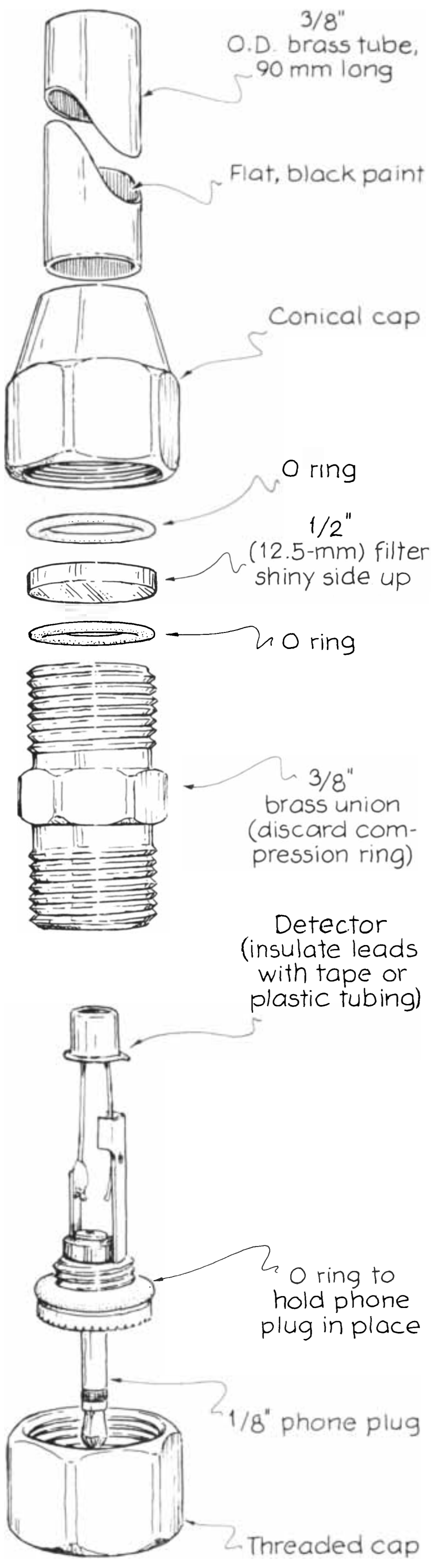


How the EG&G DFA-3000 Ultraviolet Sensor is connected



Circuit for a solar-blind ultraviolet-B radiometer ultraviolet B. Remove fingerprints by swabbing the surface of the filter with ethyl alcohol and wiping away the residue with lens cleaning paper. Blow away dust with clean, compressed air.

The assembled radiometer is simple to operate. First, look down the collimator tube. If you see a reflection of the pupil of your eye, the detector is perfectly centered. If not, realign the tube. After the voltmeter is connected and the radiometer's power switch is toggled on, block the opening of the collimator tube and adjust the potentiometer until the output voltage is zero. (Repeat this procedure before each measurement session.) Then point the tube toward the sun and align the tube until its shadow disappears. The detector will now be aimed directly at the sun. Record the voltage and make another measurement. You will soon discover that even on a clear day the signal level fluctuates, sometimes considerably—especially around noontime and whenever the atmosphere is obscured by clouds, smoke or dust.



A detector in a brass union coupling

Your readings will include an error factor because the detector responds to the red light that leaks through its filter. You can eliminate the error simply by following each reading with a second one during which you block the ultraviolet rays by placing a filter over the entrance of the collimator tube. An ultraviolet filter intended for a camera works well, and so does a WG-345 clear glass filter.

If you have an uncalibrated detector, subtract the second reading (B) from the first (A) to get a voltage that will be correct with respect to measurements you make at other times.

If you have a calibrated detector, you can compute the absolute spectral irradiance at 300 nanometers in terms of watts per square meter. Ultraviolet-blocking filters typically reflect about 8 percent of the incident nonultraviolet radiation. The nonultraviolet radiation without the filter is therefore approximately equal to the B reading divided by 92 percent. Because the active area of the DFA-3000 is about 9.9 square millimeters, the detector signal must be multiplied by 101,000 to find the signal per square meter. The formula that results is

$$\frac{A - (B/.92)}{R1 \times Dr} \times \frac{101{,}000}{F},$$

where Dr is the detector's calibrated responsivity and F is the filter's band pass. (The band pass is the number of nanometers between the two points where the filter's transmission falls to half the maximum.) The ideal filter would have a band pass of less than a nanometer. Real filters have a wider band pass.

My detector has a responsivity of .04 ampere per watt and a filter band pass of 10.4 nanometers. Readings at noon on a clear August day are typically 1.50 (A) and .116 volt (B). Inserting these values into the formula gives

$$\frac{1.50 - (.116/.92)}{30{,}000{,}000 \times .04} \times \frac{101{,}000}{10.4},$$

or .011 watt per square meter per nanometer. Remember, this is direct ultraviolet. The diffuse contribution from radiation scattered by molecules in the atmosphere adds at least 30 percent to this value at my latitude.

You can write a computer program to solve the formula. An even better approach is to program a computerized spread sheet to solve the formula, record the time and date, graph the measurements and save them.

An ultraviolet-B radiometer based on a gallium phosphide diode can also be built. These detectors, unlike silicon photodiodes, do not respond to red light and thus provide true solar-blind operation. Gallium phosphide diodes are made by Hamamatsu Corporation (P.O. Box 6910, Bridgewater, NJ 08807). The G1961 ($27.45 plus shipping and tax) is housed in a TO-18 package and has an effective surface area of 1.0 square millimeter. The G1962 ($35) is housed in a larger TO-5 package and has a surface area of 5.2 square millimeters. For a limited time and for readers of this department, Hamamatsu will calibrate the G1962 at 300 nanometers for an additional $65.

The most expensive component of the solar-blind radiometer is the optical filter. High-quality filters are made by Barr Associates (P.O. Box 557, Westford, MA 01886). Barr makes filters only on a custom basis. Therefore, unless you are connected with an institution that can afford to place a custom order, you will need to go elsewhere.

MicroCoatings (One Lyberty Way, Westford, MA 01886) makes a 12.5-millimeter-diameter filter that transmits 300-nanometer radiation and has a band pass of 10 nanometers (catalogue number ML3-300). The price is $77, a very reasonable amount for an interference filter. Twardy Technology, Inc. (P.O. Box 2221, Darien, CT 06820), sells a 25-millimeter-diameter filter with the same specifications for $210.

Most important in building the solar-blind radiometer is to install the detector and the filter in a light-tight housing. If you have access to a machine shop, you can make one. Or you can install a 12.5-millimeter filter and detector in a brass compression fitting or union coupling [*see illustration on this page*]. The coupling and the required O rings are available from hardware and plumbing stores. A two-conductor phone plug is inserted into one of the union's caps and secured in place with a rubber O ring. The leads of the detector are inserted into a light-emitting diode socket soldered to the plug's terminals. You can, however, solder the detector directly to the terminals. The cathode lead should be soldered to the terminal that is common to the tip of the plug. In either case, some couplings will accept only detectors in miniature TO-18 packages.

The filter, protected by a pair of O rings, is installed in the second end cap. A conical cap works best but may be hard to find. If the filter and O rings do not leave sufficient space for the end cap's threads to engage those of the union, replace one of the O rings with a paper spacer. Screw the end cap down so that it stays in place but does not apply pressure to the filter. If necessary, cement the end cap in place with a drop of removable glue. Be sure the filter remains clean during the installation procedure.

Depending on the detector's dimensions, a conical end cap will give a field of view of around 10 degrees. You should therefore attach a collimator tube to the opening in the end cap to reduce the field of view to four degrees or less. Brass tubing can be soldered or cemented to the opening in the end cap. Coat the inside of the tube with flat, black paint.

I store filter-detector assemblies in an airtight plastic refrigerator container along with a package or two of silica gel desiccant. The desiccant helps to protect the filters from deterioration caused by long-term exposure to water vapor. Ask a pharmacist or a salesperson at an electronics or camera store to save silica gel packets for you.

Because the solar-blind radiometer includes a separate detector, filter and amplifier, it is more difficult to assemble than an instrument in which these components are combined in a single package. Nevertheless, its circuit is functionally identical to that of the first instrument, as can be seen by referring to the circuit diagram on page 107.

An important advantage of this radiometer is that it consumes very little current. Because it can be powered by a single nine-volt battery, it easily fits inside a compact housing such as Radio Shack's part number 270-257. The operational amplifier specified in the accompanying circuit diagram (TLC271CP) can be purchased from major electronics distributors or Texas Instruments (P.O. Box 225012, Dallas, TX 75265). Other operational amplifiers can also be employed if they have a very low input bias current. Check the manufacturer's data sheet, since they may require different connections.

The TLC271CP can be damaged by static electricity. Therefore, do not touch the pins of the operational amplifier while installing it. You can avoid this problem by soldering an eight-pin integrated-circuit socket to the circuit. Insert the operational amplifier into the socket after assembling the circuit but before applying power.

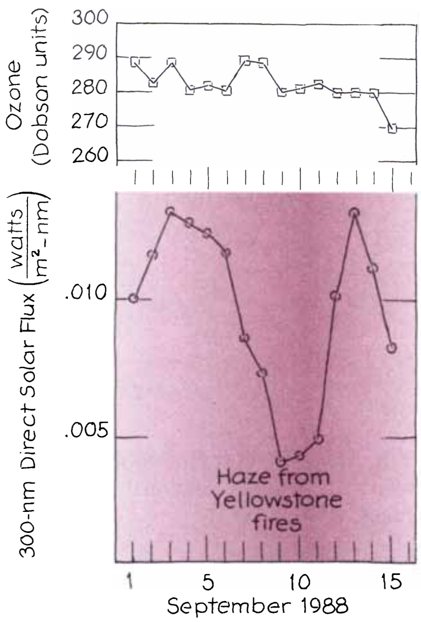
The solar-blind radiometer operates much like the previous instrument except that its output will not fall below zero, because only a single battery powers the circuit. Therefore, adjust the potentiometer until the output just reaches zero when the collimator is blocked. The formula for the DFA-3000 detector system will help you calculate the ultraviolet level. The filter manufacturer should provide the filter's peak transmission wavelength and its transmission in percentages. Literature available from Hamamatsu states the detector's area and a graph of typical responsivity versus wavelength.

Regularly measuring direct solar ultraviolet B with either a calibrated or an uncalibrated detector can yield significant data. Always try to take a measurement at solar noon. For more information about determining solar noon, consult any standard reference on astronomy or sundial construction.

Rarely does the peak ultraviolet-B reading occur precisely at solar noon. Instead the signal fluctuates constantly as the 300-nanometer radiation is attenuated and scattered by the atmosphere and its constituent gases. For this reason, allow at least five minutes to make a single measurement.

Virtually every day for two years I have measured the direct solar flux at four ultraviolet-B wavelengths and six additional wavelengths. I have found that direct solar irradiance at 300 nanometers is significantly attenuated by fog, haze, clouds and aircraft contrails. During the summer of 1988 I observed that ultraviolet-B radiation decreased when smoke from the Yellowstone fires drifted over Texas.

The passage of a cold front, which raises barometric pressure, is more often than not followed by a reduction in ultraviolet B, even when the atmosphere is exceptionally clear and dry. Satellite data and my daily ozone measurements confirm that this phenomenon is caused by an increase in ozone of as much as 15 percent and sometimes more. Low barometric pressure, at least where I live in central Texas, is typically accompanied by a decrease in ozone. For example, when Hurricane Gilbert passed nearby in



Solar ultraviolet observations

September of 1988, my radiometer registered a significant increase in ultraviolet B when the sun was visible. Several weeks later a packet of data from Arlin Krueger of the Goddard Space Flight Center revealed a simultaneous, pronounced decrease in ozone.

My observations may not hold true for your location, which is reason enough for you to conduct your own measurements. For example, can you explain and measure an increase in ultraviolet B caused by snow or a decrease caused by grass fires or a rapidly approaching squall line? If you make measurements at two closely spaced ultraviolet-B wavelengths, you can compute the ozone in a column through the atmosphere. For the relevant formulas and a list of references, send me a stamped, self-addressed envelope in care of SCIENTIFIC AMERICAN.

FURTHER READING

BIOLOGICALLY EFFECTIVE ULTRAVIOLET RADIATION: SURFACE MEASUREMENTS IN THE UNITED STATES, 1974 TO 1985. Joseph Scotto et al. in *Science*, Vol. 239, No. 4841, Part 1, pages 762-764; February 12, 1988.

THE BUDGET OF BIOLOGICALLY ACTIVE ULTRAVIOLET RADIATION IN THE EARTH-ATMOSPHERE SYSTEM. John E. Frederick and Dan Lubin in *Journal of Geophysical Research*, Vol. 93, No. D4, pages 3825-3832; April 20, 1988.

INDICATION OF INCREASING SOLAR ULTRAVIOLET-B RADIATION FLUX IN ALPINE REGIONS. Mario Blumthaler and Walter Ambach in *Science*, Vol. 248, No. 4952, pages 206-208; April 13, 1990.

BOOKS

Ants at long last, human evolution, television's spread and smart machines

by Philip Morrison

THE ANTS, by Bert Hölldobler and Edward O. Wilson. Harvard University Press, 1990 ($65).

This magnificent and long-awaited volume is the definitive work on an animal group that is at once household-familiar and vastly underappreciated in most human-inhabited areas of the earth. Since William Morton Wheeler's classic book, *Ants,* was published nearly a century ago, general works on ants have been both thin and routine. The publication of this tour de force by Höildobler and Wilson now remedies that situation.

Every imaginable area of interest to a biologist, a sociologist, even a curious citizen, is covered: evolution, taxonomy, life history, chemical ecology, kin recognition, community organization, symbiosis—the list is a long one. Army ants, fungus growers, harvesting ants and weaver ants (all of particular interest to the authors) are each given a chapter of their own. Appropriately, the book's last chapter tells the reader how to collect, culture and observe live ants (members of the family Formicidae) and so is a source of essential information for avid yet inexperienced ant watchers as well as for the professional entomologist.

LEAF-CUTTING ANTS,* Atta sexdens, *are slicing through a plant stem, which they will then transport back to the nest.

Monumental in size as well as in scope, the book weighs more than three kilograms (almost seven pounds) and spans 732 large, two-column pages filled with pictures and diagrams. At once remarkably exhaustive and parsimonious, the book does not stint on expansive detail wherever such detail is required. Summarized in its pages are the contents of more than 15,000 books and papers published over the course of three centuries, plus current knowledge that has not yet appeared elsewhere in print. More than 2,500 of these titles can be found in the book's 64-page bibliography; some of them are also referenced in tables of technical data. The book thus serves as both an invaluable encyclopedia of knowledge and as an essential guide to primary sources of literature. On yet another level it is a work of art: the illustrations, which are found throughout the text, range from carefully selected reproductions to original paintings, photographs and scanning electron micrographs, most of them of the highest quality.

The authors, Wilson of Harvard University and Hölldobler of the University of Würzburg (previously of Harvard), blend their respective talents well. Hölldobler details the micromorphology, physiology and attendant behavior of ants and their associated guests; Wilson gives special attention to the systematics and ecology of the ant family in addition to the evolutionary and social underpinnings of their lives.

Despite such accolades and the emergence of *The Ants* as the authoritative handbook on its subject, I worry about its long-term effect on ant research. In the areas of morphology, physiology, behavior and sociology, it should serve as a benchmark and a stimulus for investigators now entering (and undoubtedly about to enter) this field of study that the authors have helped to make so attractive. But for the investigator interested in ant systematics, the effect may be less than salutary over the next few years.

The reason, ironically, is related to the size of the book rather than the depth or scope of its coverage. Ant taxonomy has recently entered an era of unprecedented revisionary activity, and the book's keys to the ant tribes and genera of the world incorporate a considerable body of the latest advances.

AFRICAN WEAVER ANTS, **Oecophylla longinoda,** *cooperate in nest construction. The workers are shown here along the edge of a leaf, which will eventually be joined to another leaf with silk produced by the larvae.*

The taxonomy is an unquestionable and immeasurable improvement over Wheeler's well-known 1922 classification of the ant family (derived from the Italian entomologist Carlo Emery), which has frustrated identifiers for more than six decades.

But the new keys, nicely illustrated as they are, are encased in a monster volume that is anything but handy to consult, whether one is seated at a microscope in the laboratory or crouched over ant specimens in the field. Will the tome's sheer mass intimidate would-be ant enthusiasts? And will those brave enough to wrestle with its pages find that the handsome volume deteriorates quickly in the face of taxonomic thumbing?

An obvious solution is to publish the keys and their accompanying illustrations as a separate book, say, with smaller pages and a spiral binding, which would allow the book to be folded flat and thus be read more easily. Harvard University Press recently announced plans to publish such a companion volume; they are to be congratulated for doing so. Nevertheless, I would urge them to delay the volume's publication until several major, impending revisionary studies are completed. The American Ornithologists' Union Checklist and its paralyzing effect on avian taxonomy should stand as a warning of what can happen when a classification is published prematurely and so becomes frozen in a contemporary standard. *The Ants* is indeed a masterpiece; its taxonomic contents should not be anything less.—*Reviewed by William L. Brown, Jr., Department of Entomology, Cornell University*

THE HUMAN CAREER: HUMAN BIOLOGICAL AND CULTURAL ORIGINS, by Richard G. Klein. University of Chicago Press, 1989 ($39.95).

In this up-to-date and engaging summary of our hominid family, a University of Chicago paleoanthropologist exerts a refreshingly cool and credible tone of voice. The 500-page text is exceptionally well illustrated, not for show (there are no photographs) but for understanding. Maps, charts and detailed drawings are found throughout, and many are visually keyed to help a novice reader interpret the image presented. The author avoids conjecture and resists drawing verdicts based on limited and poorly controlled data; instead he invites the reader to evaluate the available evidence according to his or her own criteria. Indeed, because experiments are rare and fossil finds depend more on chance than design, paleoanthropology is said to resemble a court of law more than, say, a physics laboratory.

The methods are few: skeletal remains bear witness to physical form, artifacts reveal traces of ancient behavior and extranuclear DNA extracted from the cells of living people offers genetic insight into ancestral lineages.

One chapter, a tale of dry bones, sets the early stage. The branch of mammals leading to primates begins at the end of the Cretaceous period (some 80 million years ago), with the appearance of tropical squirrellike animals. These small, arboreal mammals had teeth and jaws that were unmistakably primate; much later came lemurs and tarsiers. Eventually the fossil record bears evidence of "protopeople [that] probably diverged from protochimpanzees" between 10 and five million years ago, in the late Miocene, although the sparse record in Africa "refuses to illuminate the event."

The oldest hominid fossils (the Australopithecines) are some 3.8 million years old. Their remains have been excavated from limestone pits and caves at a few rich sites in southern Africa as well as from layered beds of volcanic ash that line the Rift Valley all the way to the Horn of Africa. Although scholars cannot agree on a family tree, even so it is obvious that human evolution is marked by two adaptive pathways: one leads the way to big teeth and heavy jaws fit for grinding coarse vegetal foods, the other to expansion of the brain case and the appearance of tools. Both pathways indicate a meatier diet, won not only by hunting but also by scavenging the kills of big carnivores.

Animal bones and hominid tools are often found together at streamside sites in Olduvai Gorge and around Lake Turkana in Kenya, which by itself does not suggest causality. Carnivores and hominids alike would have gathered at such water holes; only a few of the many kills may therefore have been the work of our hominid ancestors. By interpreting the manner in which flakes were chipped from stones, paleoanthropologists have concluded that the occupants of the Oldowan Industrial Complex at Olduvai were mainly right-handed and so had already diverged from chimpanzees, which show no strong laterality. Such a dangerously seductive clue to incipient human brain asymmetry is also suggested by various skull casts.

The climate in which our early ancestors evolved seems to have been one of progressive aridity. Over the course of some two million years, grasslands replaced forests, and arboreal primates came down from the trees to walk on two feet across the emerging savanna. These were tropical beings; neither in Africa nor anywhere else are early hominids found at higher latitudes. Although a site far south—near the Cape of Good Hope—has yielded hundreds of thousands of vertebrate fossils of about the right age, not one hominid has emerged from that locality.

The first indisputable evidence that hominids migrated beyond Africa exists in the form of stone tools (estimated to be more than a million years old), found in Jordan. From then on, clear signs of hominids can be found throughout the Old World, as far west as Morocco and as far east as Beijing, Thailand and Java. Java Man and Peking Man and several other prototypes are now grouped—with some uncertainty—into a single, far-flung and variable species, *Homo erectus.* The oldest fossils are from Lake Turkana and are about 1.8 million years old; the most recent and best specimens (estimated to be from 400,000 to 500,000 years old) are from Java and China. The picture, however, is sketchy: "In total, the fossil sample... comes from perhaps one hundred individuals [in] bits and pieces,... [and] fewer than fifty... archeological sites."

Yet throughout that entire span of time and space there is a remarkable conservatism among the stone artifacts, which suggests that human culture progressed slowly. Eventually, however, the species did spread and even adapted to cooler climates. The past few 100,000 years of human history (covered by two chapters in the book) are complicated by the coming and going of the enigmatic Neanderthalers across Europe and by the discovery of the remains of many half-modern humans, which looked more like their descendants than those who preceded them, both within Africa and without.

The final coherent act in the evolution of our species commenced about 40,000 years ago, during the time of Cro-Magnon Man in southwestern France. It was an explosive period of change, marked by a swift, steady growth in the complexity of artifacts. Traps, harpoons, tally sticks and images—worked in bone, antler and even flint—appeared for the first time. Hearths and houses were built. Some 15,000 years ago the magnificent murals that adorn the walls of the Franco-Cantabrian cave were painted. Ponder the site of Dolní Vestonice in Czechoslovakia. It is about 27,000 years old and was, of course, settled by our own sapient kin. A kiln in which many fire-hardened clay figures of people and animals were apparently made has been found there. Among the figures is one with a face that droops on the left side, while buried nearby are the skeletal remains of a real woman with a temporomandibular bone deformation, known to induce partial facial paralysis. How could we avoid being moved by empathy for this sad figurine, arguably the first of all human portraits?

Such an amazing legacy is entirely contrived by our own kind: the clever, big-brained subspecies *H. sapiens sapiens.* Exponential change thus had its origins long ago. It persists still: our population has increased by three or even four orders of magnitude during the past 40,000 years. We do not yet know why or how or when that pattern of growth will plateau, but we do know that such leveling is inevitable.

INDIA'S INFORMATION REVOLUTION, by Arvind Singhal and Everett M. Rogers. Sage Publications, Inc., 1989 ($28; paperbound, $14).

The television revolution in India began bravely in 1975. Lively black-and-white images were beamed to the earth from a U.S. satellite positioned over India for a year. A crude chicken-wire dish picked up the signals, which were then fed to ruggedized public TV sets, one in each of 2,400 villages across six language areas. The programs being broadcast covered such subjects as health, farming, primary school matters and the rights of farmers and women. Resident observers certified that interest in the programs was highest among poor villagers.

Carefully planned growth along frugal lines of development was supplanted with one stroke in 1983 when the first government-owned satellite brought its color signal to a wide area of India. Now 90 percent of the population lives within range of one of 400 government-built local relay transmitters. A color TV is not yet for everyone; although the cost can be spread among some six or eight family wage earners, a television set costs half a year's wages for a factory worker. Instead of owning their own, many Indians rent a TV and a videocassette recorder for occasional entertainment. Still, the number of sets is growing: 15 million can now be found in Indian households (about one in six), and mushrooming interest suggests an audience of close to 400 million by the year 2000.

At the same time that individual interest is growing, the government television network has gone commercial on a grand scale; its income from the sale of commercials has increased fortyfold since 1980, making it the one government bureau that pays for itself. Consumerism and its homogenizing trends are certainly spreading rapidly across kaleidoscopic India. A survey made by the authors in 1987 found that Maggi 2-Minute Noodles, which had been advertised on TV ("fast to cook, and good to eat"), were being eaten by a third of city viewers, even though "in India, noodles were perceived as a very Chinese product."

Most TV viewers in India today do not live in villages but in towns and cities where they form an urban middle class. The programs they watch are diverse, although most are entertainment shows, produced by the Indian equivalent of our Hollywood. Versions of ancient epics in the bejeweled song-and-dance style of the Bombay films, for example, gather a huge audience. Vignettes of these films, which touch on significant individuals and events, are found throughout the text. The reader meets the scriptwriter of the first Hindi soap opera, as well as the Indian-American magnate of the disk drive, along with other heroes and rascals.

The book offers a savvy look at the role of communication in development. It was once thought that a mass-information source like TV could influence development at all levels of society right down to the rice paddy. In the 1980's a new view emerged: uniform mass communication would be superseded by more interactive schemes focused on the individualization of messages in time and space.

Some examples can be cited. For three generations, farm women sold their home-grown vegetables on the bustling sidewalks of a growing city district. But police harassed them, claiming they were traffic hazards. Videotapes of the encounters, made by the Self-Employed Women's Association (SEWA), however, indicated that the women were not interfering with traffic patterns. Municipal leaders, on seeing the tapes, were moved to political action, and the farm women were once again allowed to sell their goods.

Hardware and software growth, along with VCR's and personal computers, are described in fascinating detail. The growth in well-paying informational jobs and the brain drain that has poured 6,000 Indian-trained electrical engineers into California's Silicon Valley are also examined. The book offers no sharp predictions, but its candor and catholicity divert readers at the same time it helps them to see farther, into both the developing and developed countries of the world.

NATURALLY INTELLIGENT SYSTEMS, by Maureen Caudill and Charles Butler. The MIT Press, 1990 ($19.95).

In much the same way that a halftone image is tediously built of dots, the mathematical calculations of Alan M. Turing are based on a logic of yes-no bits. Indeed, the similarity gives rise to the intriguing view that all finite knowledge is mappable as a string of decisions, linked to one another by repetitive step-by-step computation.

All the same, no artist ever consciously scanned a scene bit by bit in the manner of a video camera in order to paint it. Reductionism, however universal, is not all there is to knowledge. Although everyday reasoning and judgment can be theoretically reduced to vast strings of information, they never explicitly disclose their hidden details any more than a glass of water reveals its arrangement of atoms.

The lighthearted book at hand, which offers plenty of mathematical reasoning but not one equation, is a timely witness to the rise of a new and already flourishing discipline. I refer to the emerging class of information-processing systems that tolerate error and fuzzy logic and have the ability to learn by example. Such systems are "naturally intelligent," smart in the same ways human beings are.

Metaphorically called neural networks, intelligent systems bear a general but vague resemblance to neural nets in the brain. Yet the term is something of a misnomer: the networks are not intended to have real similarity to biological systems. They are not engineered as brain models, which lie in the domain of neurobiologists, but as machines constructed with the help of broad hints provided by living systems.

The processing units that form the machine's nodes are termed neurodes (a useful neologism coined by the authors to connote artificial neurons). Like neurons, the neurodes are simple yet numerous and multiply interconnected. Their interconnections (synapses) can be influenced by a number of conditions, including the frequency of previous use. Memory, for example, resides mainly in the synapses. The neurodes are not necessarily digital or highly serial; they do not respond simply—step by step—but adjust themselves globally to the intricate flow of space-time patterns.

The authors provide careful explanations of key concepts, offer a look at the history behind neural networks and describe recent applications of the technology. Their open informal style is typical of good technical journalism; a number of agreeable cartoons are included as well as supplementary passages that either amplify the text or provide gossipy yet insightful anecdotes. Nevertheless, abstract arguments are not always presented in a way that can be easily grasped by a lay reader, and the book unfortunately lacks internal references.

The 19 chapters are grouped into five parts. Part I provides an overview of the subject and introduces the three-step behavioral repertoire of a neurode. Neurodes, it seems, receive input signals from many places and then form a net input, which more or less summarizes the signals. The summary number might either be a simple linear sum of the signals received or something much more complicated, which would be computed by various manipulations of assigned weights that take into account various thresholds and delays. Individual neurodes eventually lose track of each signal, passing only its overall effect on to the next neurode. The receiving neurode in turn generates an appropriate output signal of its own, which it transmits to its correspondent neurodes. A neurode may be designed in many ways, yet until recently most have carried out mathematical operations normally associated with digital computing.

Part II presents the ingenious structures of neural-network memories, which have been worked out over the past 15 years. These are associated memories, in which entries are correlated with other entries by content rather than address. There are several types of memory including ones built of competitive filters; in these, neurodes compete with one another to increase their own input at the expense of others, teaching themselves to be more competitive by trial and error.

Part III, which is about learning, describes a remarkably clever set of network ideas. Presented are idealized models derived from studies of behavioral conditioning in animals, some of which provide entire sequences of learning patterns. There is also a brief but fine account of the first neural learning machine, constructed in the late 1950's by Frank Rosenblatt of Cornell University.

Part IV covers network design. Some intelligent systems, for example, contain three layers of neurodes: a layer for input, a layer for output and at least one hidden layer sandwiched between the two, which transmits information in both directions. All the designs are cannily described—in both written and diagrammatic form—but require close attention; the section is not for a reader on the run.

Part V focuses on the practical applications of neural networks, most of which were developed in the late 1980's. Four of them are discussed at length. They include a system that can evaluate mortgage loans, a robotic arm able to adapt to varying weight loads, a sonar-echo classifier and a network that can be taught to "drive" on a computer-simulated freeway. The driving abilities of the network, it seems, depend on the teacher. A person who drives cautiously might teach the network to change lanes slowly, in contrast to a more aggressive driver who might teach the network to be a hot-rodder. The sonar classifier, however, requires no such instruction. After many trials it can distinguish between various objects, say, a rock and a cylinder, by deriving an ingenious set of echo categories.

None of these so-called intelligent technologies are apt to shake the world, but the practical role for inexpensive expert systems is real and growing. More important, neural networks have transformed our attitude toward intelligent machines. Unlike rule-based expert systems that justify their decisions by producing a list of criteria adhered to, neural nets follow no such criteria. In the same way that people are known to judge first and to reflect on and rationalize their actions later, perhaps a neural network can be teamed with a rule-based system to provide a set of relevant rules.

Nevertheless, it can be stated with some degree of certainty that even if true intelligence is someday embodied in an advanced neural network, no one will be able to explain with surety just how it came to think that way, any more than one can be certain how an imaginative friend might respond in a given situation. Such machines may be held to think, but before they do they will no longer seem machinelike. Thus, a distasteful paradox is evaded. Although burdened with detail, this small book is a superb nontechnical introduction to the tantalizing prospect of machine-based intelligence.

ESSAY

A born naturalist's keen insights

by Miriam Rothschild

Looking back on my childhood, I am convinced that naturalists are born and not made. My father was a keen naturalist, but although we all shared the same upbringing, my siblings were not hooked on plants and insects in the way I was. They soon abandoned their pressed flowers and butterfly nets. Unlike them, I had no ear for music and languages, I was not an early reader and I could not do mental arithmetic. But I had a retentive memory and was deeply, passionately fond of animals and keenly interested in everything connected with them.

Before World War I we spent our summer holiday in Hungary. I have only the vaguest memory of the house we lived in, of my room and my toys. Nor do I recall clearly the large family of grandparents, aunts, uncles and cousins who were assembled next door. Nor can I remember what my mother wore or how she did her hair. But I do have a crystal-clear recollection of the dogs, the farm animals, the local birds and, above all, the insect fauna. My father gave me a net, but I required assistance to "box" my captures.

After dark a fair number of moths found their way into my bedroom, attracted by the night-light, and in the morning were seen settled on the white walls and the ceiling. I noticed they were all different, and I was greatly puzzled by the strange way the lappet moth hunkered down. Where was its head? How did its wings fit?

Since we had no television or radio, a gramophone provided our evening entertainment. It required a constant change of needle and persistent winding up. Living in the country meant the only mode of transport was a pony and trap. This was, in fact, a tremendous boon for it forced us to seek our amusement in the fields and woods—and I found it. No thrill later in life ever equaled the sudden discovery of a ditch lined with white violets or a nest belonging to a long-tailed tit. I was happy from morning till night.

Had my father lived—he contracted a mortal illness when I was nine—he would have appreciated my passionate interest in flowers and butterflies, astronomy and the weather and would have organized some lessons for me. I found a letter from him written to me from a Swiss sanitorium, apparently in answer to one of mine: "How do you know *Chrysis ignita* is a parasitic wasp?" he asked. My father disliked schools, examinations and pedagogues, and so I was "educated" at home—by one dear old governess, with a love of the Romans and Walter Scott but with absolutely no knowledge of the natural world. At the age of 17, I insisted on attending evening classes at the Chelsea Polytechnic, but up to then I was entirely self-taught. I had no guidance regarding the books I should read.

When I was in my teens, my brother returned from boarding school with a new type of holiday task: he was to dissect a frog and draw the internal organs. Could I give him a hand? The lovely and luckless frog was chloroformed in a glass jar. I had never before seen fresh, internal organs, blood vessels and nerves. Their extreme beauty was a revelation. That experience shocked me back into natural history, which I had abandoned after my father's death. It was my road to Damascus.

Subsequently, I had the great piece of good fortune to get into a conversation with G. C. Robson in the library of the British Natural History Museum. For a reason I cannot now recall, I asked him how the mantle of a snail produced a shell. Robson was a great naturalist, who unfortunately suffered from severe nervous breakdowns. Looking back, I feel uneasy about his state of mind at that moment, for he instantly took an enormous and quite unmerited interest in me and my future as a research worker.

Robson obtained for me the London University Table at the Biological Station in Naples and told me to work on a rather primitive bivalve called *Nucula*, to see if by chance it contained traces of a long-lost radula. He overlooked not only my complete ignorance of mollusks but also the virtual absence of *Nucula* in the coastal waters around Naples. In a way, this was also fortunate, for while the staff scoured the sea for the absent *Nucula*, I took the opportunity of becoming acquainted with some of the most beautiful marine animals in existence.

Back in England I decided to have a look at *Nucula*, which was easily obtained on the southern coast at Plymouth. During the first day of dissecting the bivalve, I found a specimen infested with larval trematodes. It proved to be a hitherto undescribed species and a most extraordinary and unusual one from every point of view. My fate was sealed. I was completely hooked on this world of relative transparencies and fantastic life cycles such that evolution seemed an impossible explanation. One could only fall back on the idea of a humorous deity.

I lost seven years' work in the Plymouth blitz in World War II (I also got married) and so was forced to turn my attention to a group of animals I could study inland at home: fleas. I began to wonder how they jumped and spent two years trying to tease apart the puzzle. I had a flash of insight—curiously enough while I was on an airplane to Israel—and I solved the problem.

During my 75 years as a passionate naturalist, I think I spent at least half my time proving that my blinding flashes of insight were rubbish. What is insight anyway? I think if you are a dedicated naturalist you are scanning your material all the time without being aware of it and storing the information. Then something happens that brings these observations together in your conscious mind. Suddenly you fancy you see the answer to the riddle, because it all seems to fit together.

I am a naturalist, not a scientist. I have a vivid emotional love of the natural world from rain clouds to tapeworms, and my enthusiasm has lured me into varied investigations. Furthermore, I love discussing my favorite topics, be they birds or flowers, trees or fleas, and burning the midnight oil, enthralled, while reading about other people's observations and discoveries. This has resulted in my publishing 300 papers and books, which some might glorify with the title of scientific research.

But have no illusions—curiosity, a keen eye, a good memory and boundless delight and enjoyment of the animal and plant world, with perhaps a little intuition thrown in, do not make a scientist; one of the outstanding and essential qualities required is self-discipline, a quality I lack. A scientist requires not only self-discipline but hard training, perseverance, energy, determination, judgment and a goal. A scientist, up to a point, can be made. A naturalist is born. If you can combine the two, you get the best of both worlds.

MIRIAM ROTHSCHILD, a self-taught investigator in the amateur tradition, is best known for her research on fleas.


All buttoned up in Seoul.
You arrived for your meeting with all your ducks in a row. You took United to Seoul.
Getting you to the Far East in top form is top priority for United. So we give you more flights there than anyone. All provide the best in international travel, including, for First Class passengers, sleeper seats and our exclusive Concierge Service.
Come fly the friendly skies.
UNITED
AIRLINES
TOKYO • OSAKA • HONG KONG • SEOUL • TAIPEI • SYDNEY • MELBOURNE • BEIJING • SHANGHAI • AUCKLAND • SINGAPORE • MANILA • BANGKOK