SCIENTIFIC AMERICAN

JANUARY 1991
$3.95

The race to build the fastest supercomputer yet.

What makes aspirin a wonder drug.

A widening search for worlds around distant stars.

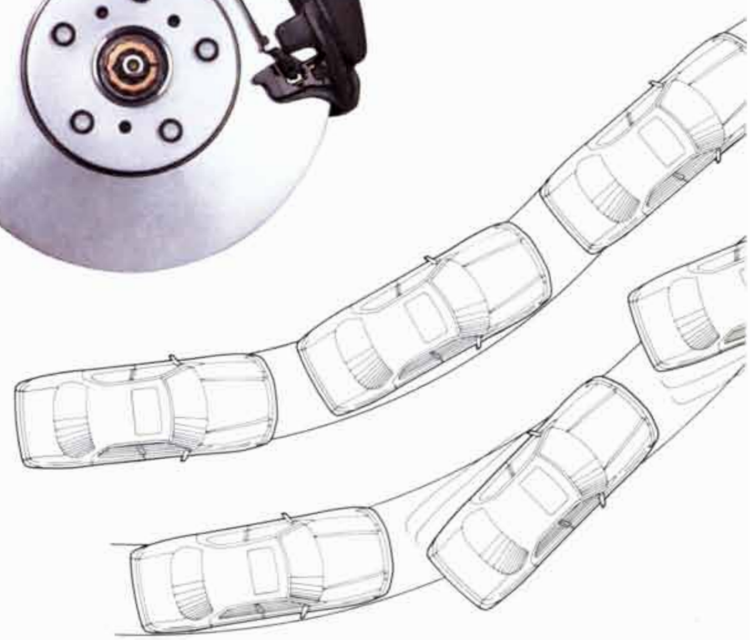
Falling dominoes *show how a small event can trigger a catastrophe. Earthquakes, economies and ecosystems may follow similar rules.*

For greater reliability, we went to the effort of gold-plating the electric terminals in our driver's-side airbag Supplemental Restraint System. Nothing succeeds like excess.

For evasive action, thrust is provided by a Four-Cam, 32-valve V8 engine.

You've Heard Of Th
Now Meet The

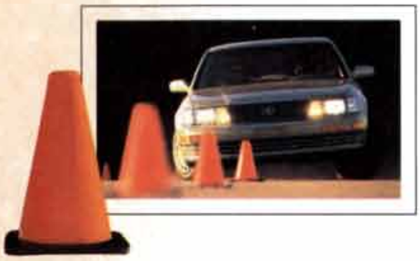
Performance is important at red lights as well as green. All four vented disc brakes are activated by one of the world's most advanced anti-lock braking systems (ABS).



Every Lexus comes with a fifth shock absorber: an advanced driver's-side airbag SRS* activated by our redundant crash-sensors.

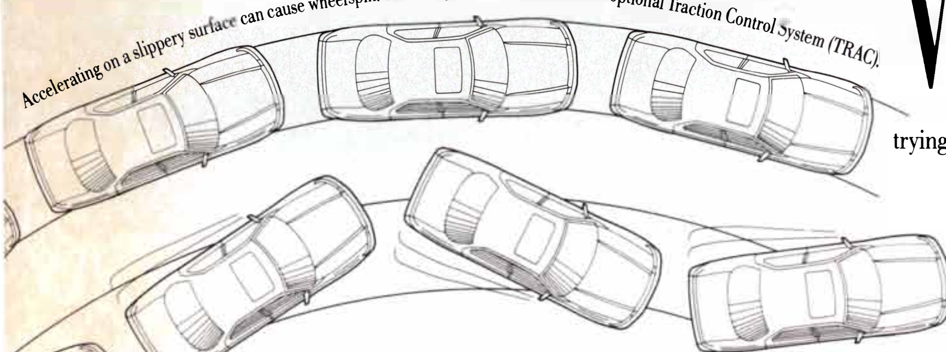
*The driver's-side Supplemental Restraint System (SRS) will inflate only i … n. the three-point seat belt provides primary protection; the airbag will not inflate.

Our Grand Prix-inspired double-wishbone suspension is responsive and surefooted.



Accidents can't always be avoided. Fortunately, we spent years preparing for such a mishap. Countless prototypes were crash-tested and, as you can see, special shock-absorbing crumple zones were developed.

e Defensive Driver. Defensive Car.



In the blink of an eye, TRAC can limit wheelspin by both throttling back the engine and pulsing the rear brakes.

While no car can guarantee your safety in every conceivable situation, this has not stopped us from trying. Pictured here is only a small fraction of the safety features on the LS 400.

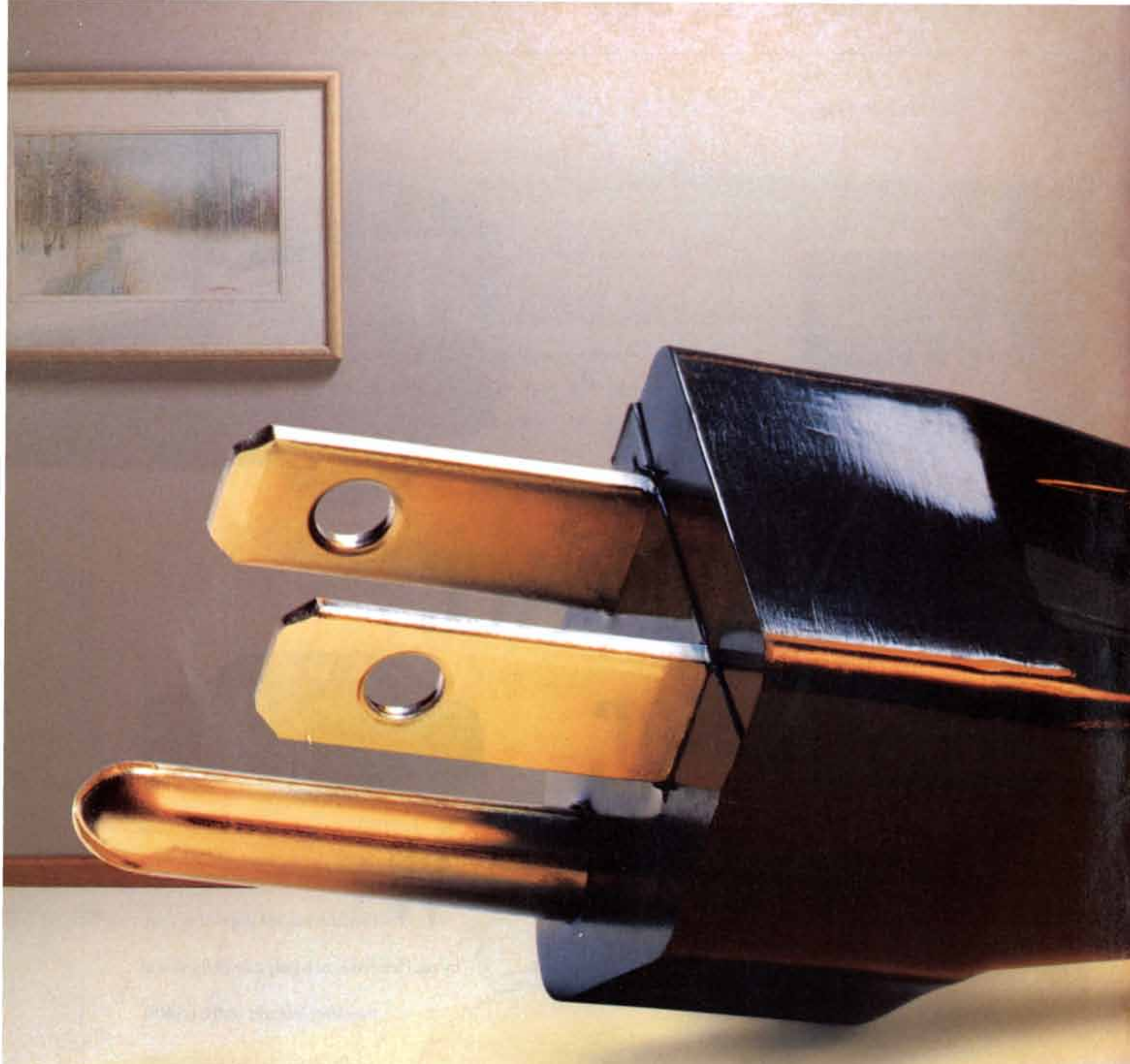
But what others may think superfluous, we consider judicious.

So visit your nearest Lexus dealer.

But please drive extra carefully. You're not in a Lexus yet.

LEXUS

The Relentless Pursuit Of Perfection.



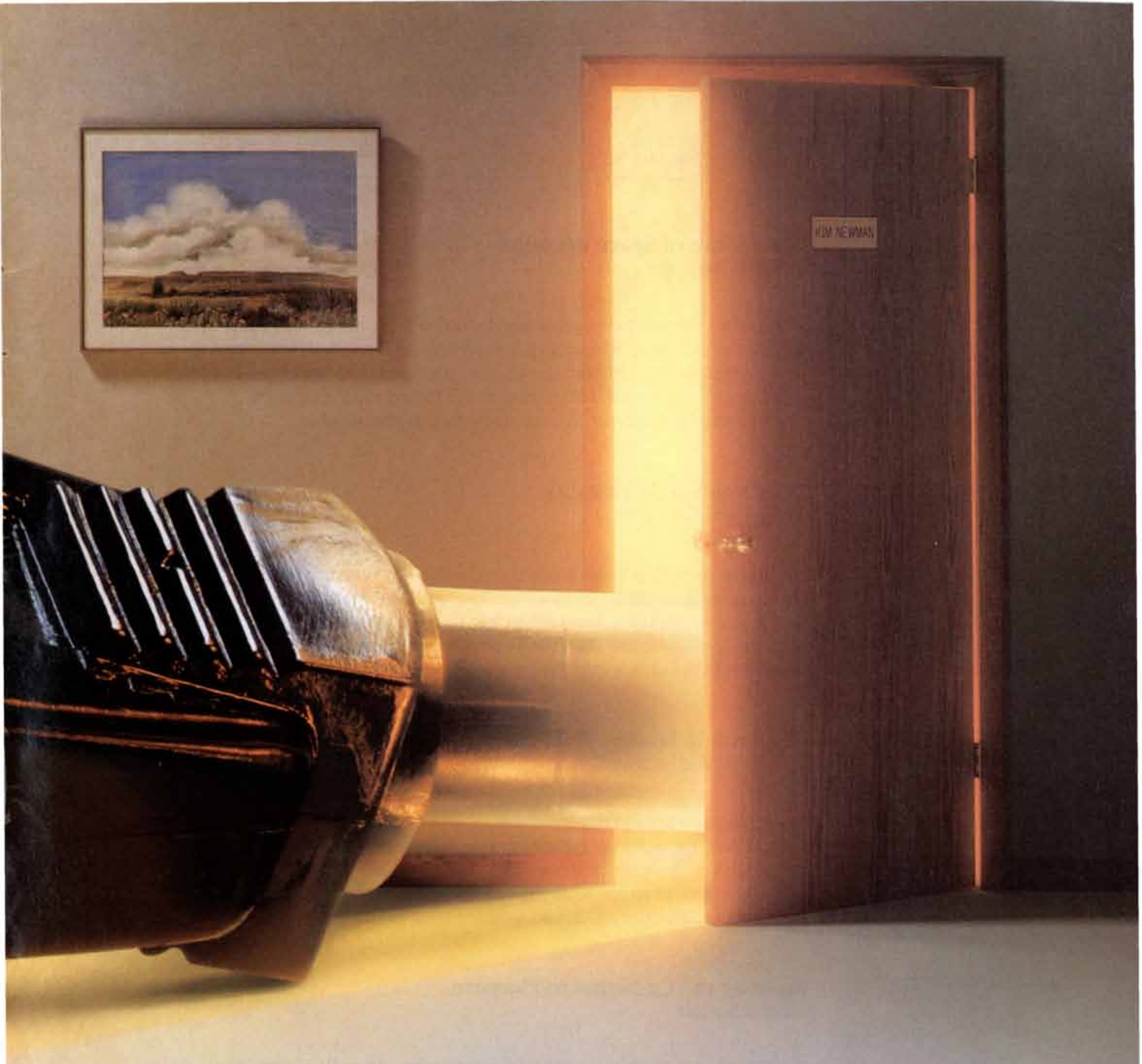
The 486™ PC. It may be a little

Never before has this much power been plugged into a business PC.

Presenting the Intel 486 microprocessor—a veritable powerhouse that's been harnessed for business.

A 486 microprocessor-based PC has everything it takes to run today's high-powered applications. And run them the way you need to—simultaneously and at lightning speed.

Plus, it's compatible with the hard-



more power than you're used to.

ware and business applications you already own, so you won't spend any extra time or money on training.

The 486 PC. Plug it in and start shocking the corporate world.

For additional information, call 1-800-548-4725 and ask for "The 486 Microprocessor Means Business" brochure.

intel®

The Computer Inside.™

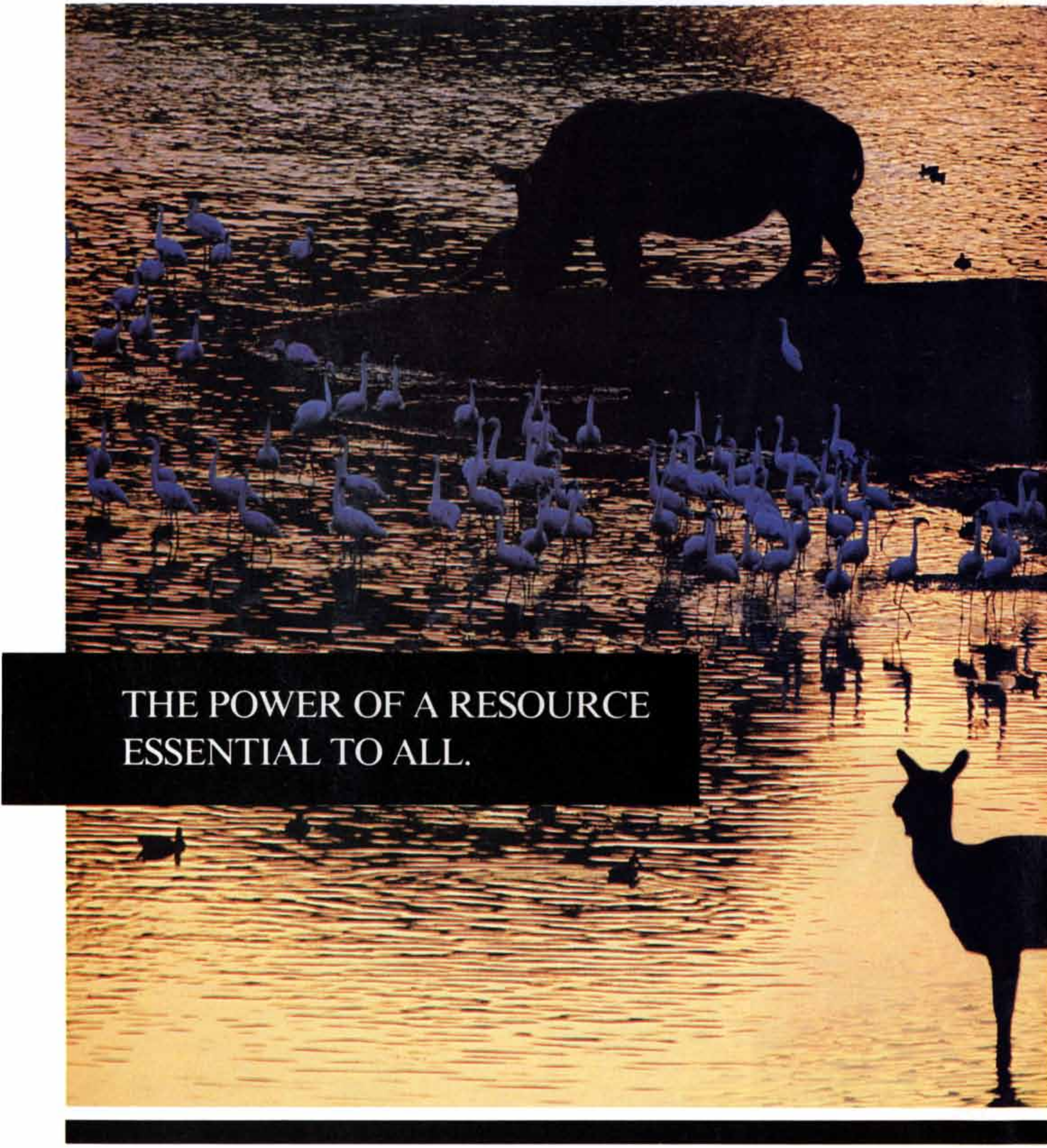
486 is a trademark of Intel Corporation.

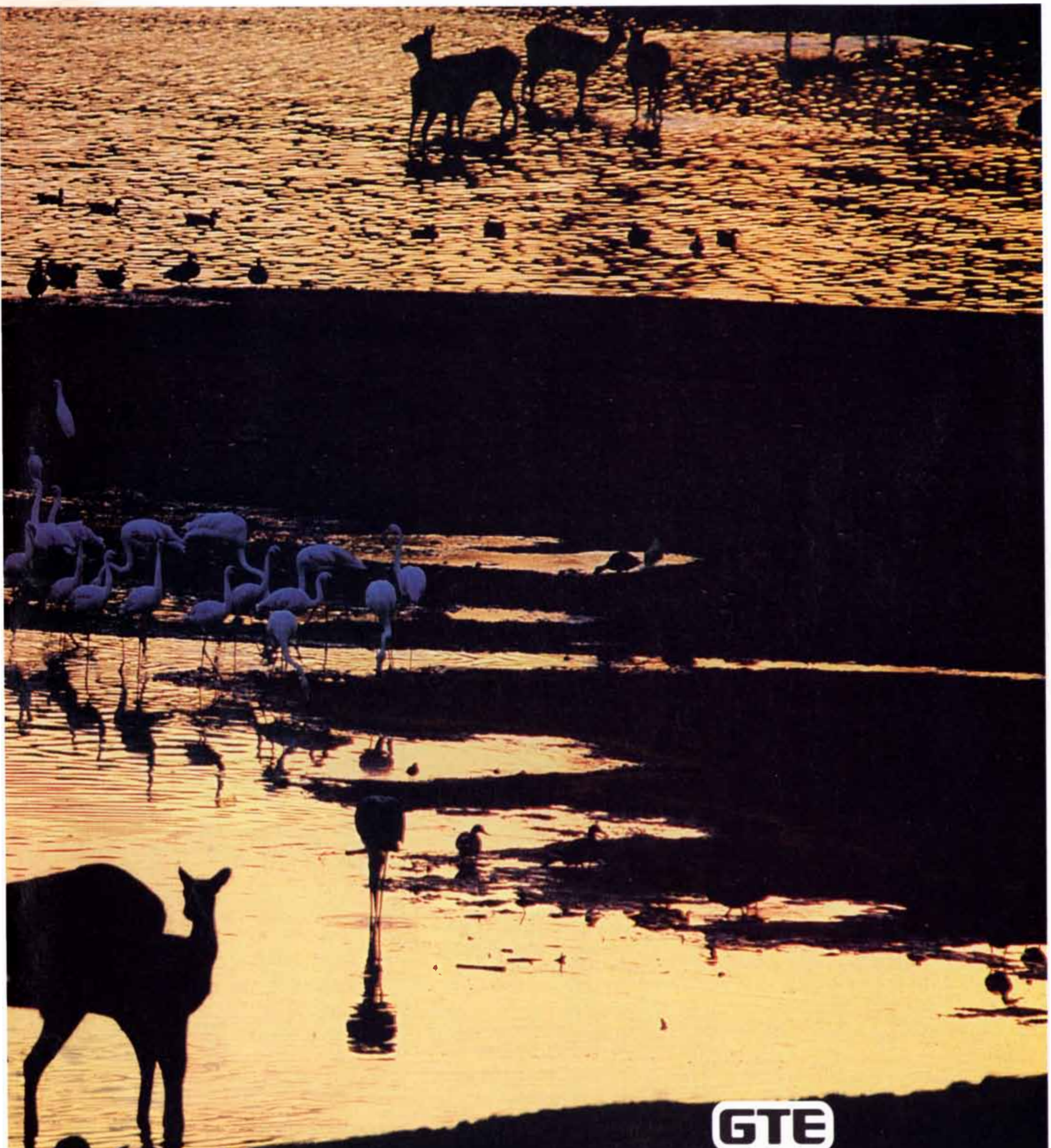
SCIENTIFIC AMERICAN

January 1991 Volume 264 Number 1

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 yearfor 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 POWER OF A RESOURCE
ESSENTIAL TO ALL.



THE COVER photograph represents part of a system of more than 2,000 dominoes that were toppled to demonstrate subcriticality, criticality and supercriticality. In a subcritical system of dominoes, a large chain reaction can occur only when the dominoes are disturbed by a tremendous force. In a critical system, catastrophe can sometimes strike when a single domino falls. A supercritical system almost always produces huge chain reactions. It took three workers 15 hours to set up the dominoes (see "Self-Organized Criticality," by Per Bak and Kan Chen, page 46).

THE ILLUSTRATIONS

Cover photograph by Quesada/Burke

Page	Source
39	Loral Corp.
40–43	Joe Lertola
44	William E. Burrows, New York University
47	Quesada/Burke and Domino F/X
48	Eric A. Myers, University of Texas at Austin
49	John R. Deecken
50	Glenn A. Held, IBM Thomas J. Watson Research Center
51	John R. Deecken
52	Galen Rowell/Mountain Light
54–55	Frederic M. Richards and Paul E. Vogt, Yale University
56	George V. Kelvin
57	George V. Kelvin; omega turn courtesy of George Rose, Pennsylvania State University
60	George V. Kelvin; courtesy of Jane S. Richards, Duke University
61–62	George V. Kelvin
63	Jay W. Ponder, Washington University
66–67	Mike Kagan, Monkmeyer Press
68	Gaetano Barone, Monkmeyer Press
69	*Zibaldone,* Buonaccorso Ghiberti, unpublished manuscript BR 228, Biblioteca Nazionale, Florence
70	George Retseck (*top*), *Brunelleschi*, by Peiro Sanpaolesi, plate X, Club del Libro, Milan (*bottom*)
71	George Retseck
72	*Leonardo da Vinci: Engineer and Architect*, Montreal Museum of Fine Arts, detail of plate III, from the Bartolini Salimbeni Collection, Florence
76–80	Andrew Christie
81	Anneila I. Sargent and Steven Beckwith, Owens Valley Radio Interferometer
82	Andrew Christie
85	Pamela J. Harper
87–90	Dana Burns-Pizer
93	Michael Leach, Oxford Scientific Films/Animals Animals
94–96	Patricia J. Wynne
97	Oxford Scientific Films/Animals Animals (*top*), Ian Wyllie, Monk's Head Experimental Station, Huntingdon, England (*bottom*)
98	Patricia J. Wynne
100–101	Robert B. Wilhelmson, National Center for Supercomputing Applications; Joe Towers, WeatherStock (*inset*)
102–103	Peter Ray, Florida State University (*top left*), Joseph B. Klemp, National Center for Atmospheric Research (*top middle*), Robert B. Wilhemson, National Center for Supercomputing Applications (*top right*), Jason Kuffer (*bottom*)
104	Joe Lertola
106	David Hessler, University of California, San Diego
107	Thinking Machines Corp.
109	Michael Goodman
119	Reprinted from *Computers, Pattern, Chaos, and Beauty*, by Clifford A. Pickover

SCIENTIFIC AMERICAN®

Established 1845

EDITOR: Jonathan Piel

BOARD OF EDITORS: Alan Hall, *Executive Editor;* Michelle Press, *Managing Editor;* Timothy M. Beardsley; Elizabeth Corcoran; Deborah Erickson; Marguerite Holloway; John Horgan; Philip Morrison, *Book Editor;* Corey S. Powell; John Rennie; Philip E. Ross; Ricki L. Rusting; Russell Ruthen; Gary Stix; Paul Wallich; Philip M. Yam

ART: Samuel L. Howard, *Art Director;* Joan Starwood, *Design Director;* Edward Bell, *Associate Art Director;* Johnny Johnson

COPY: Maria-Christina Keller, *Copy Chief;* Nancy L. Freireich; Jonathan Goodman; Daniel C. Schlenoff

PRODUCTION: Richard Sasso, *Vice President Production and Distribution; Managers:* Carol Albert, *Prepress;* Tanya DeSilva, *Projects;* Carol Eisler, *Manufacturing and Distribution;* Carol Hansen, *Composition;* Madelyn Keyes, *Systems;* Leo J. Petruzzi, *Manufacturing and Makeup;* Carl Cherebin; William Sherman

CIRCULATION: Bob Bruno, *Circulation Director;* Lorraine Terlecki, *Business Manager;* Cary Zel, *Promotion Manager*

ADVERTISING: Robert F. Gregory, *Advertising Director.* OFFICES: NEW YORK: William Buchanan; Peter Fisch; John Grant; Michelle Larsen; Meryle Lowenthal. CHICAGO: 333 N. Michigan Avenue, Chicago, IL 60601; Patrick Bachler, *Advertising Manager.* DETROIT: 3000 Town Center, Suite 1435, Southfield, MI 48075; Edward A. Bartley, *Detroit Manager;* William F. Moore. WEST COAST: 1650 Veteran Avenue, Suite 101, Los Angeles, CA 90024; Kate Dobson, *Advertising Manager;* Joan Berend, San Francisco. CANADA: Fenn Company, Inc. DALLAS: Griffith Group.

ADVERTISING SERVICES: Laura Salant, *Sales Services Director;* Diane Schube, *Promotion Manager;* Mary Sadlier, *Research Manager;* Ethel D. Little, *Advertising Coordinator*

INTERNATIONAL: EUROPE: Roy Edwards, *International Advertising Manager,* London; GWP, Düsseldorf. SEOUL: Biscom, Inc. TOKYO: Nikkei International Ltd.

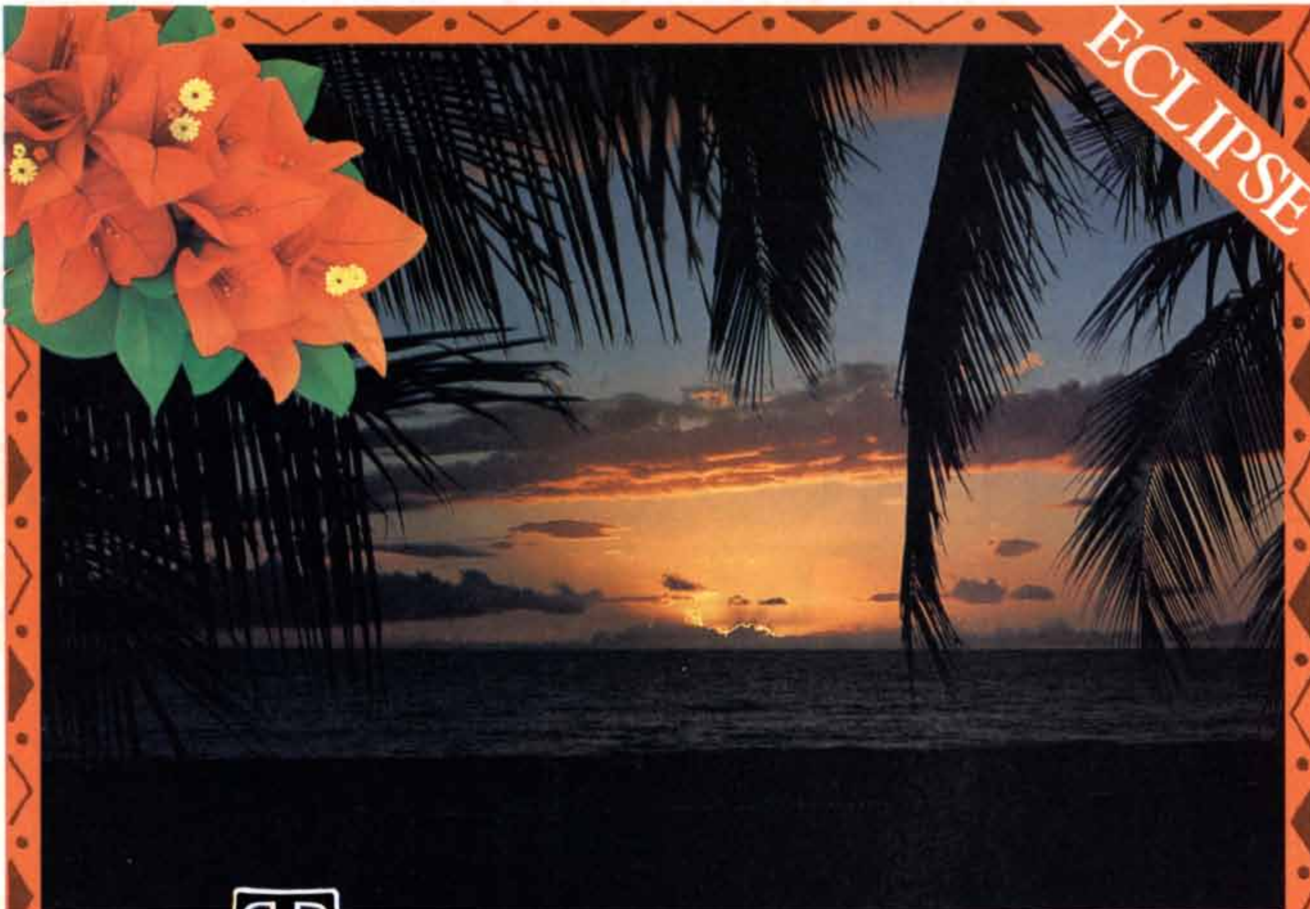
ADMINISTRATION: John J. Moeling, Jr., *Publisher;* Marie D'Alessandro, *Business Manager*

SCIENTIFIC AMERICAN, INC.
415 Madison Avenue
New York, NY 10017
(212) 754-0550

PRESIDENT AND CHIEF EXECUTIVE OFFICER: Claus-Gerhard Firchow

CHAIRMAN OF THE BOARD: Dr. Pierre Gerckens

CHAIRMAN EMERITUS: Gerard Piel



THE BIG ONE IS COMING. . .

(and it won't be back until 2,017 A.D.!)

Don't miss your last chance to see the great TOTAL ECLIPSE OF THE SUN from the comfort of your luxury hotel on the magnificent Big Island of Hawaii. WorldWide demand for this event is tremendous!

MOMENTUM TRAVEL has exclusively reserved several deluxe hotel packages that include air travel and vacation activities for this once in a lifetime phenomenon.

While on the Big Island, visit an active volcano, enjoy deep sea fishing for marlin, or explore one of the island's hidden waterfalls, lush valleys or black sand beaches. Day trips to Honolulu or Neighbor Islands can also be arranged.

Call MOMENTUM TRAVEL for reservations today! **1-800-346-0276**

Momentum TRAVEL

75-5751 Kuakini Hwy., Suite 103, Kailua- Kona, Hawaii 96740
Toll-Free 1-800-346-0276 (Hawaii Standard Time) • FAX: (808) 329-8000

SUN PROUDLY ANNOUNCES THE ULTIMATE STRIVING MACHINE.

SPARCstation™ 2. If you think limits were made to be exceeded, this is your kind of machine.

After all, it exceeds all our own limits. Last year, SPARCstation 1 broke every record for price and performance. And became the best-selling workstation in history. By far. But we went right back to the drawing board. And created the entire SPARCstation 2 line.

POWER YOU CAN ACTUALLY USE.

To begin with, you get twice the performance. For about the same price. 28.5 MIPS. 21 SPECmarks. And 4.2 MFLOPS. You can even have up to 96MB of RAM. And as much as 7.6GB of mass storage.

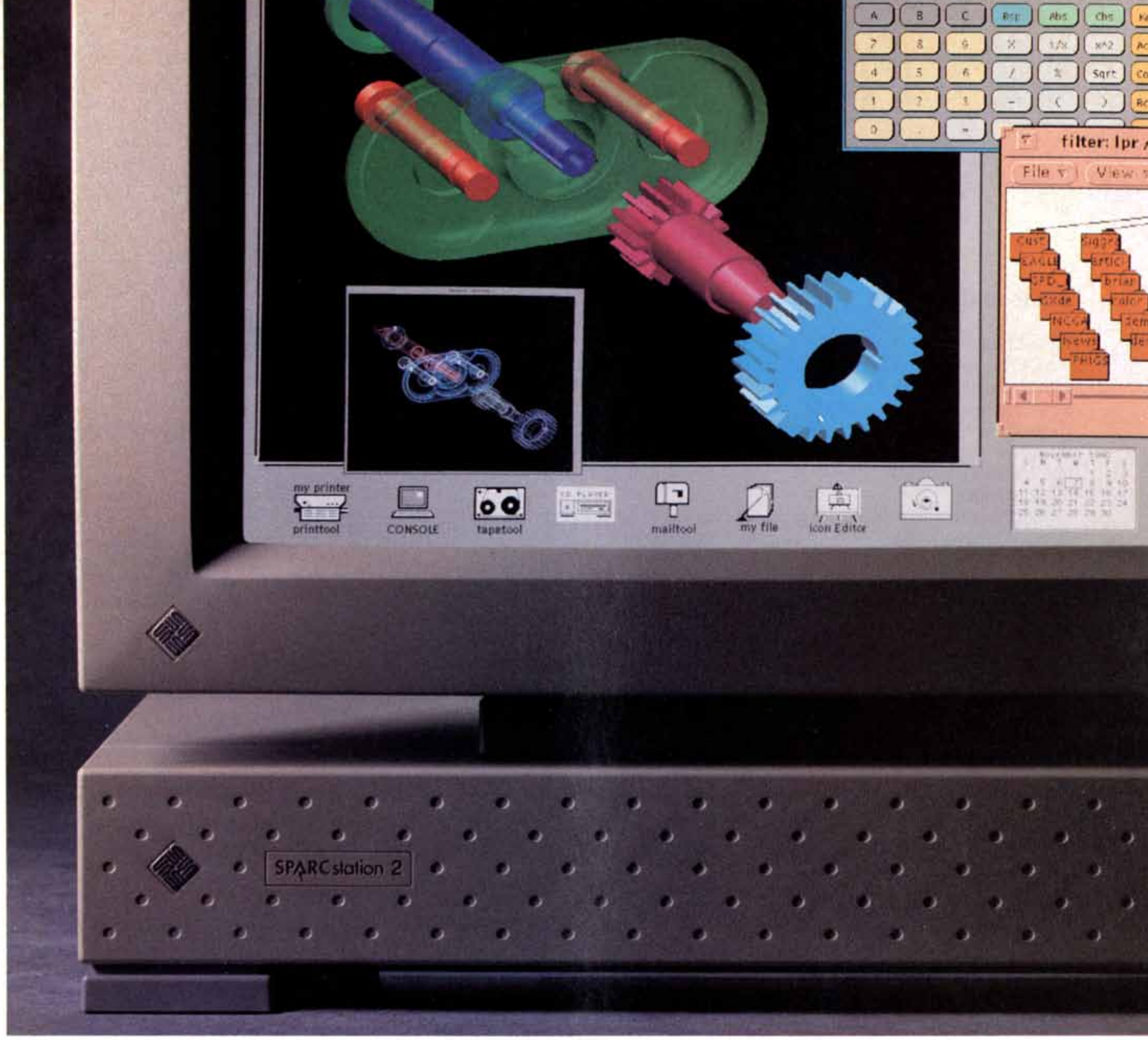
But more than just a hot engine, you get everything else you need to do your job. Unbelievably real graphics. Easy networking. A huge selection of software. And complete expandability.

Put all that together, and you get the kind of power you can actually use.

THE WHOLE LINE IS AWESOME.
THE PRICES AREN'T.

Just look at SPARCstation 2GX. It gives you ultra-high speed at no extra cost. And brings a whole new level of performance to X-window applications. So it's ideal for electronic publishing. Financial analysis. And for anyone who has to work with 2-D and 3-D wireframe applications.

And that's just the most basic color model. We've also built SPARCstation 2GS. It lets



you create 3-D solid images in 24-bit true color. It's the kind of machine you hate to share. And from now on, you won't have to.

At the high end, there's SPARCstation 2GT. It does all the above, but it's been tuned especially for PHIGS, which is the highest standard for 3-D graphics on the planet. So it runs five times faster than the GS. With all this, it gives you a level of image quality you've never seen at anywhere close to its price.

THE WHOLE THING MAKES PERFECT STRATEGIC SENSE.

At Sun, we make a full line of SPARC-based systems. From the lowest-cost RISC/UNIX® workstation in the world to servers that support hundreds of users. They're all binary compatible. And they're built to run the most widely accepted standards for workstations.

On the subject of software, there are more than 2100 SPARCware™ applications. Including all the most popular solids modeling programs. And the most popular PC software. And with our OPEN LOOK® interface, you'll spend less time learning the system. And more time on your real job. If you'd like to know more, call us at 1-800-821-4643. (From California, 1-800-821-4642.)

And we'll give you a better machine to strive with.

sun microsystems

Computers that network people.™

LETTERS

Costs and Climate

To the Editors:

In "The Great Climate Debate" [SCIENTIFIC AMERICAN, July 1990], Robert M. White has estimated that the cost of controlling greenhouse gas emissions would be "hundreds of billions of dollars." Many economists believe the costs to be negative.

I believe the costs of reducing emissions, if there are any, must be undertaken in any event. If we continue to increase emissions at the current rates, something will stop us: global warming, acid rain or something no one has yet imagined. The only question is how fast to make the change. Waiting until we have proof and then making crash changes will do more environmental damage and be more expensive than making steady reductions over a moderate length of time.

AARON C. BROWN
New York City

DNA Fingerprinting

To the Editors:

The readership of *Scientific American* has been badly misinformed by Peter J. Neufeld and Neville Colman in "When Science Takes the Witness Stand" [SCIENTIFIC AMERICAN, May 1990]. Their claim that DNA-based forensic identification technology is untested is patently untrue. The technical basis of forensic DNA identification is the Southern blot, a technology established in 1975 and cited more than 2,400 times a year. Its application to forensic identification has been discussed extensively in the scientific literature. Lifecodes' staff alone have published over a dozen peer-reviewed articles.

The authors' argument that crowding in gel lanes can cause similar alleles to be called identical is wholly wrong. Hypervariable loci may have as many as 100 different alleles, but individuals carry only two alleles. Therefore, there is no need to resolve between 100 alleles, only between two. Furthermore, genetic variability is so great that a chance coincidence at one gene locus would not be repeated at two, let alone three or more. The information concerning population genetics and band shifts has also been misrepresented.

Perhaps the authors' most reprehensible suggestion is that the defense bar and the scientific community have abdicated their consciences and principles. It impugns the integrity of the scientists and lawyers who have spent untold hours examining and preparing DNA-typing results for both the prosecution and the defense.

No one wants standards more than the laboratories doing identity testing. With or without those standards, however, DNA-based identification remains the most accurate and reliable method available. These are facts that no amount of distortion can change.

KEVIN C. MCELFRESH
Director of Forensics
Lifecodes Corporation
Valhalla, N.Y.

Neufeld and Colman respond:

We emphasize that at the time our article appeared, DNA analyses had been used in courts for more than two years without a single peer-reviewed publication on the two most important novel methodologies applied to forensic identification. The first is the use of numerical criteria to define a match. Each of the three primary forensic laboratories (including Lifecodes) applies quantitative rules to construct reports that can determine life-or-death issues. Unfortunately, none of the rules are the same, and none have been published.

The second distinguishing feature of forensic identification is the calculation of allele-prevalence probabilities based on modest data bases. Forensic laboratories blithely disregard the compelling evidence of major ethnic differences in gene frequencies. Considering the weight that an expert's opinion is given in a court of law, expression of these numbers without independent scientific verification is unconscionable.

Honorable Failures

To the Editors:

In "Death Watch" ["Science and the Citizen," SCIENTIFIC AMERICAN, June 1990], John Horgan wrote disparagingly about four scientific investigations that he tells us are now failing. I think that he does a disservice to public understanding of the scientific method by implying that only winners are worthy of respect. In fact, negative results may be of great importance.

Horgan should read General Jan C. Smuts's address as president of the South African Association for the Advancement of Science in 1925: "The value of a hypothesis often depends not so much on its correctness as on its fruitfulness.... The actual truth is a far-off goal to which any hypothesis may be only a step and often a faulty step." The hypothesis to which Smuts was referring was Wegener's theory of continental drift.

KEITH RUNCORN
Department of Physics
Imperial College, London

Giving Credit Where Due

To the Editors:

Your readers should know that in "A Volcanic Eruption," by Vincent E. Courtillot [SCIENTIFIC AMERICAN, October 1990], the table comparing the ages of mass extinctions and flood basalts was based on our article "Flood Basalt Volcanism During the Past 250 Million Years" in the August 5, 1988, issue of *Science*. The volcanic periodicity of roughly 30 million years was also derived there. A very similar periodicity in impact cratering was discovered and submitted by us to *Nature* two months before the equivalent discovery and submissions by Walter Alvarez and Richard A. Muller, a point that Alvarez and Frank Asaro neglected to mention in "An Extraterrestrial Impact," which accompanied Courtillot's article.

MICHAEL R. RAMPINO
RICHARD B. STOTHERS
Goddard Institute for Space Studies
National Aeronautics
and Space Administration
New York City

Editor's note:

According to Courtillot, Rampino and Stothers were among several groups of researchers whose work contributed to the table in his article. Because *Scientific American* is a magazine for general readers, it is not our policy—nor is it even possible in most cases—to cite the contributions of all the scientists participating in research. The papers by Alvarez and Muller and by Rampino and Stothers appeared simultaneously in the April 19, 1984, issue of *Nature*.

ERRATUM

The diagram appearing on page 125 of the October 1990 issue of SCIENTIFIC AMERICAN should have been credited to the *Harvard Business Review*.

50 AND 100 YEARS AGO

JANUARY 1941: "Nobel Prizes have been awarded since 1901, and Germany has won the largest number. But a rating based on population, and limited to the most recent decade, gives Switzerland a 465 percent attainment; next comes Holland with 217 percent, then Great Britain with 175, Austria 162, Germany 135, the United States 78, France 44 and Italy 22. Our score of 78, for the years of the last decade, is *more than four times* the 18 we received as an attainment score for the earlier years."

"Long-standing cases of schizophrenia, commonest form of insanity, have responded in encouraging manner to a new operation in which fibers of the frontal lobe of the brain are severed. The surgical procedure is known technically as pre-frontal leucotomy, referring to the cutting of white matter in the pre-frontal lobes of the brain. It was devised by Dr. Egas Moniz, of Spain, and was introduced into this country by Dr. Walter Freeman and Dr. James W. Watts, of the George Washington University School of Medicine."

"Hunters and skeet shooters 'lead,' or aim ahead of their targets, but their guns are fired from a nearly stationary rest. Aerial huntsmen have to bag 400-mile-per-hour pursuits from ships traveling at comparable speeds. A gunner firing from a plane in a side-slip, for example, must allow for the fact that his weapons are traveling through *three directions* simultaneously; forward, sidewise, and downward. To facilitate aim, gunners employ sights which swing off center to compensate for the effect of the slipstream. There is, however, another phenomenon called the 'Magnus Effect' for which no sights will compensate. It is induced by the clockwise spin that is imparted the bullets as they leave the gun barrel. When the gunner swings his armament for a left broadside and fires, the spin causes the bullets to drop in their flight just as top-spin drops a tennis ball. Contrariwise, from a right broadside, the bullets would rise in flight like a golf ball given backspin."

"The sound system for Walt Disney's motion picture 'Fantasia' cannot be used with the regular movie-theater equipment of today. The range of the best conventional sound-on-film production is about 35 decibels. The range of a symphony orchestra is about 70 decibels. In 'Fantasound,' as the new system is known, the range is about 75 decibels. In the conventional sound film, a narrow longitudinal strip carries the sound record. This track is present in 'Fantasia' also, but only for emergency use. The quality depends on a second strip of standard movie film that runs in a specially designed sound reproducing unit. It operates simultaneously with the picture film and in precise synchronism with it. On the second or sound-only film are *four* individual sound tracks."

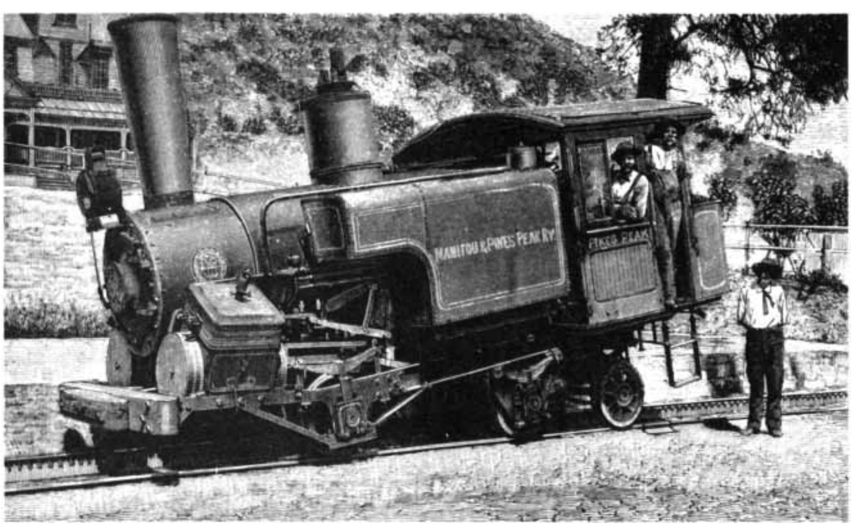
SCIENTIFIC AMERICAN

JANUARY 1891: "Inventors have for several years been setting their wits to work to devise small apparatus for allowing amateurs to take photographs without any one seeing them do it. Here we have something designed to meet with great success among practicians: it is a question of a necktie provided with a pin. The latter is an objective, and the necktie is a camera. When any one approaches you and speaks to you at a distance of 2 or even 3 ft., you press a rubber bulb concealed in your pocket, and you have the portrait of your interlocutor."

"The Panama Canal is actually a thing of the past, and Nature in her works will soon obliterate all traces of French energy and money expended on the Isthmus. Reports of October 25 say that the late heavy rains have caused vast slides into the canal from the hilltops near Obispo, and the canal excavation at Circaracha is entirely filled up. Only one dredge now remains."

"Dr. Henry Schliemann was born in 1822 and early received a strong bent toward the study of Greek, and it is said that when a mere boy he determined to discover Troy. His commercial business prospered, and he eventually acquired a fortune. He devoted the rest of his life to investigations among the ruins of Greece and Asia Minor. In 1874 appeared his 'Troy and its Remains,' giving the results of his excavations on the site of ancient Troy. Dr. Schliemann's home life was influenced by his classic tastes. His wife was a Grecian. Greek was the language of his household, even his servants receiving Greek names. His two children were named respectively Agamemnon and Andromache. In the course of his wanderings he found himself in California when that State was admitted to the Union. He became himself a United States citizen. His death was announced on December 27, in Naples."

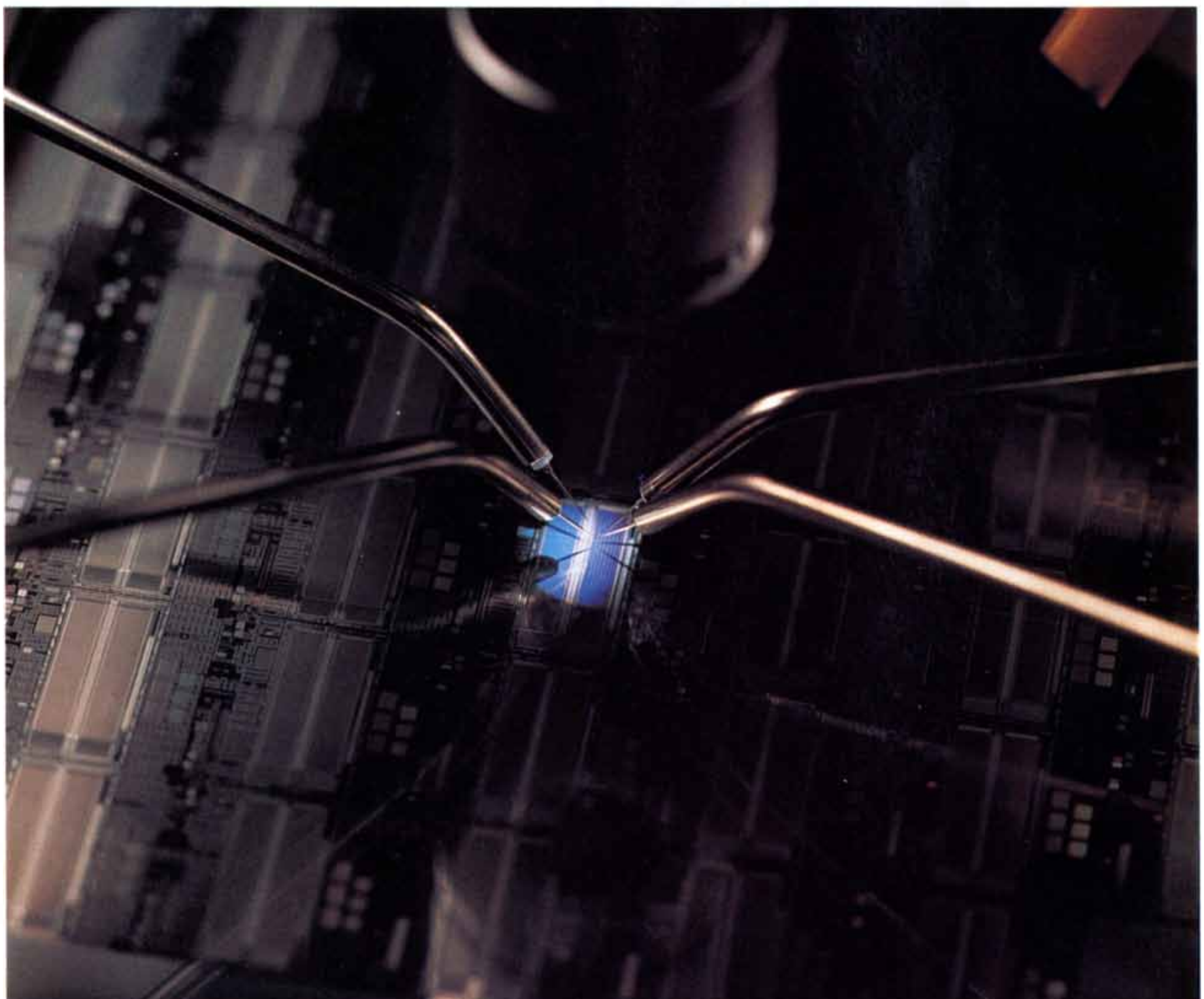
"The engines were built by the Baldwin Locomotive Works, of Philadelphia, and present a very different appearance from the ordinary locomotive. The three engines each weigh 26 tons. They are built to stand level on a 16 per cent grade. The power is transmitted by a main drum, which has teeth similar to the pinions on the track."



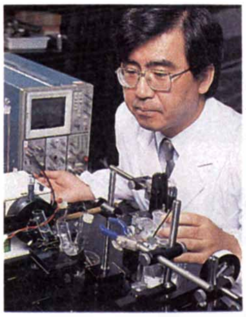
Pike's Peak locomotive

TOSHIBA RESEARCH

COMBINING THE PRACTICAL AND THE EXOTIC



BY EDWARD ZUCKERMAN



At the Toshiba R&D Center in Kawasaki (bird's-eye view, right), Dr. Hiroshi Nakanishi (above) developed an artificial retina that emulates the human sight process. Previous page: Toshiba's EEPROM chip paves the way for memory cards that may replace the magnetic memories of computers.

In the Tokyo suburb of Ome, a Toshiba Corporation plant manufactures laptop computers, minicomputers, disk drives, word processors, and a wide array of peripheral equipment for sale in Japan and throughout the world. Four thousand people work there. Five hundred of them are blue-collar employees on production lines. Three thousand are engineers and scientists.

That ratio might seem unbalanced for what is, after all, a factory. But it works for Toshiba, where 21% of the 70,000 employees are engineers and scientists, and 1,000 more are hired every year. For 116 years, the company has been propelled by research and development. In the 19th century, Toshiba was a pioneer in the fields of telegraph equipment and electric lamps. Today the company is an innovative leader in technologies ranging from semiconductor devices to superconductivity, from nuclear fusion to artificial intelligence, and other areas both practical and exotic.

At the sprawling Toshiba Research and Development Center in Kawasaki, Dr. Hiroshi Nakanishi has developed an artificial molecular membrane model of the retina, designed to emulate the sight process as it occurs in the human eye and brain. When light bearing an image reaches a retina, it strikes cells containing molecules called retinals. The retinals change

For 116 years, R & D has helped Toshiba pioneer fields ranging from semiconductors to nuclear fusion.

shape, initiating a chain reaction that culminates in the sending of electric pulses to the brain, which recreates the image. Dr. Nakanishi and his colleagues, searching for a substance to mimic that behavior, focused on an azobenzene molecule, a component of dyes used to color kimonos. When light strikes an azobenzene molecule, it changes its shape.

After confirming the special properties of azobenzene, Toshiba's scientists needed to fabricate an artificial membrane to house the substance. They produced a 10-micron-thick nickel plate into which they bored holes with diameters of 100 microns. Then they dipped the plate into an ionized solution, on the surface of which floated a lipid membrane containing azobenzene. As the plate moved down, the membrane filled in the holes.

Next, the researchers passed an electric current through the liquid and suspended an electrode behind the nickel plate; its output was read on an oscilloscope. Finally, they presented light bearing an image to the nickel plate Azobenzene molecules exposed to the light bent. This changed the electrical resistance of the membrane and thus the shape of the oscillation signal on the oscilloscope.

Particular shapes on the oscilloscope correspond to particular images presented to the nickel plate, just as electrical patterns in the brain correspond to images registering on the retina. The device thus "sees" the image. More significantly, its sight is fundamentally different from the process employed by existing computer recognition systems, which break an image into a large number of pixels, register the signals one at a time, and then assemble them into an image—a time-consuming approach. The artificial membrane system recognizes images all at once.

"Data processing in the brain is an interaction of molecules," says Dr. Nakanishi. "In principle, the same kind of molecular detector can be built for taste. Our dream application is a biocomputer that will use molecular devices to gather multiple information and perhaps even make judgments. The field will take years to develop."

Dr. Seiichi Takayanagi says Toshiba's R&D commitment is reflected in an R&D budget equal to 8% of total revenues.

In the short term, Dr. Nakanishi's work may lead to such commercial applications as improved ZIP code readers. Or maybe not. "We're not sure it will be practical," Dr. Nakanishi says. "But we must do it to find out what will come next. Molecular electronics may dominate the 21st century." Or, again, maybe not. "But this is the kind of risk Toshiba is pleased to take."

PERSPECTIVE ON R&D

Dr. Seiichi Takayanagi, senior vice president in charge of corporate technology, says the commitment to R&D has been an essential part of Toshiba's growth over a century. "Look at the roots of Toshiba," says Dr. Takayanagi. "The company was founded by an inventor and a professor of electrical engineering." The inventor was Hisashige Tanaka, who founded the Tanaka Engineering Works in 1875 to manufacture telegraphic and electrical equipment. Tanaka was well known for creating ingenious mechanical dolls and clocks. The professor was Dr. Ichisuke Fujioka, who in 1890 founded Hakunetsusha & Co. to produce electric lights

The two companies merged in 1939 and the resulting business was renamed Toshiba in 1978. The founders' legacy is honored by a long list of research firsts. In 1915, Toshiba produced the first X-ray tube in Japan. In the 1920s, its scientists achieved two breakthroughs in light bulb technology that are still used today—coiled-coil incandescent lamps and inside frosted bulbs.

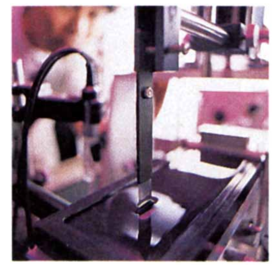
"Our active commitment to R&D is reflected in our R&D expenditure," says Dr. Takayanagi. "The amount has been steadily increasing every year, both in real terms and as a percentage of total sales." For example, for the fiscal year ended March 1990, parent Toshiba Corporation spent some ¥233 billion (US$1.68 billion), or 7.6% of total sales for R&D. By the end of the current fiscal year, the figure will increase to ¥260 billion (US$1.88 billion), or 8% of total revenues. Toshiba pays virtually the entire R&D cost itself, with only 3% coming in the form of subsidies from the Japanese government and other public institutions.

Now, how does Toshiba organize research and development? "Toshiba's R&D activities are organized on a hierarchical and functional basis to achieve optimum efficiency," continues Dr. Takayanagi. "We have institutionalized a three-tier system of laboratories."

The first tier of this network is the four corporate-level laboratories—the Research and Development Center, Ultra Large Scale Integration (ULSI) Research Center, System and Software Engineering Laboratory, and Manufacturing Engineering Laboratory. They conduct basic research with a medium- and long-range perspective of five to ten years.

In addition, the company has eight development laboratories connected to its business groups. These laboratories are oriented to practical applications for research conducted in the corporate labs, with a short and medium-range perspective.

Finally, each business group's department of engineering works on practical commercialization. They focus on implementing specific product designs, and on enriching finished products with



Unlike existing computer recognition processes which break images into a number of pixels, Toshiba's artificial retina "sees" images all at once. A short-term application may be an improved ZIP code reader.

enhanced performance, functionality and convenience. Most of the Ome plant engineers are in this category; their tasks include designing Toshiba's technologically advanced—and immensely successful—line of laptop computers.

VERY LARGE SCALE RESEARCH

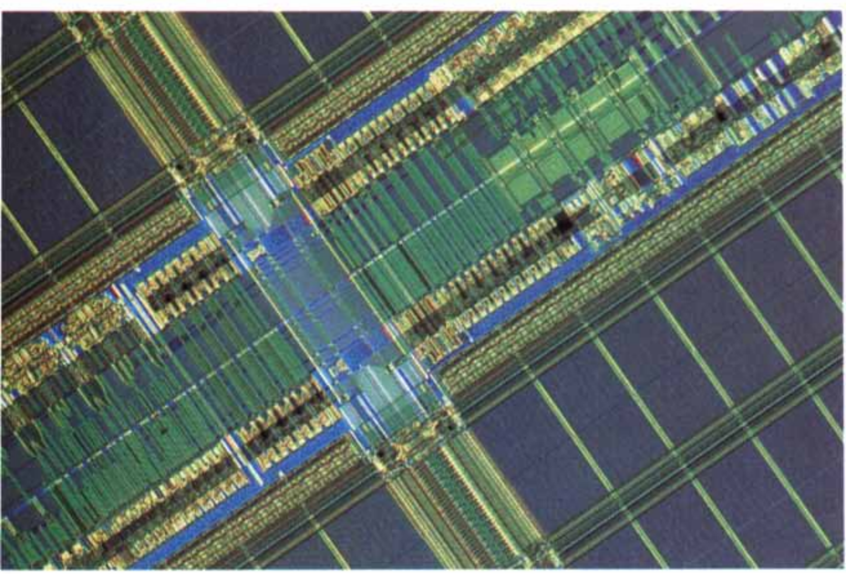
One of the projects on which Toshiba has been placing major emphasis is called "Project W." Recognizing that semiconductor chips are the essential building block of computers and other electronic devices, in 1983 Toshiba initiated the project to enhance the expertise and the scope of its operation in semiconductors.

A case in point is the company's involvement with the DRAM (dynamic random access memory) chip. A decade ago, when 64-kilobit DRAM chips (capable of storing about 64,000 bits of information) were the common currency of the electronics world, Toshiba was an also-ran in the field

But, recognizing that larger memory capacities were key to more compact and more reliable electronic equipment, company executives made a decision: Go for it. Toshiba invested ¥20 billion (US $145 million) to build a new Very Large Scale Integration (VLSI) Research Center with an ultra clean room and advanced hardware and software aids to produce experimental DRAM prototypes.

With no moving parts, flash EEPROM memory cards are smaller, cheaper, and more reliable than disk drives.

Dr. Fujio Masuoka and his team (right) developed programmable read-only memory (PROM) chips that can be erased and reprogrammed. Toshiba's prototype 16-megabit DRAM chip (below) can store text equivalent to a 64-page newspaper.



Meanwhile, DRAM chips with 256 kilobits of memory came to market, and Toshiba was still not a major player. But the company was looking ahead to the next generation: 1-megabit DRAM chips with the capacity to store one million bits of information. Toshiba researchers went on a crash schedule. And in 1986 Toshiba was first to market with a 1-megabit DRAM chip. The company moved to the front of the DRAM chip market and remains there to this day.

But not by standing still. With the 1-megabit DRAM wrapped up, the VLSI Research Center (now renamed the ULSI Research Center) turned its attention to a 4-megabit DRAM chip. Researchers squeezed more memory cells into the same surface area as earlier chips by folding capacitors into trenches etched in the chip's surface instead of lying them flat. Toshiba shipped the world's first 4-megabit DRAM commercial samples in November 1988, and now is a leading supplier.

Toshiba has now developed a prototype 16-megabit DRAM chip, which can store all the information contained in a 64-page newspaper. This supermicrochip is comprised of 35 million transistors and capacitors on a surface 7.87 millimeters wide and 17.4 millimeters long. Circuits are etched with separations of 0.6 microns. (A human hair is 100 microns thick; bacteria range in size from 1 to 10 microns.)

Researchers experimenting with electron-beam lithography hope to etch even finer circuits. Under development is a three-dimensional structure for transistors and capacitors. Work on a 64-megabit DRAM prototype is well underway, with a 1-gigabit chip—with a one billion-bit storage capacity—predicted by 2002, and a 16-gigabit chip by 2008.

While the push to develop the 1-megabit DRAM was still underway, Dr. Fujio Masuoka, then manager of memory design technology in the semiconductor group, had a new idea. Unlike DRAM chips, which lose their memories when power is cut, programmable read-only memory (PROM) chips maintain memory in the absence of power. Even more useful are PROM chips that can be erased and reprogrammed.

The two existing types of PROM chips had problems: the UVEPROM, an ultraviolet-erasable PROM, is slow in reprogramming; the EEPROM, an electrically-erasable PROM, is large and expensive. Dr. Masuoka conceived a "flash EEPROM." Its design was based on the compact UVEPROM model, with the addition of an element that allowed its memory to be erased by the faster electric method.

Flash EEPROM devices have now been developed that may eventually replace computer magnetic memories—hard and floppy disk drives. With no moving parts, flash EEPROM memory cards and modules will be smaller, lighter, cheaper, and more reliable than disk drives, and they will require far less power. The advantages for laptop computers alone are significant.

CASCADING PRODUCTS

Toshiba's encouragement of personal initiative, combined with the vast resources allotted to research and development, has resulted in a cascade of innovative applications. The Research and Development Center alone obtains some 150 American patents every year.

The ExpressReader optical character recognition system is one recently introduced product. A plug-in board that works with IBM-compatible computers, the ExpressReader converts scanned printed material into word processing text with 99.7% accuracy. The system recognizes thousands of type fonts, and can follow newspaper and magazine articles as their columns wend their way around headlines, picture captions, and other design elements.

Another recent product introduction is a digital still camera. Images are captured by a ⅔ inch-format charge-coupled device microchip with 400,000 pixels and stored on a credit-card-sized semiconductor memory card or digital audiotape. Since the images are digital, they can be displayed on computer screens, transmitted over telephone lines (with no loss of picture quality), and combined with documents or sound in multimedia databases

Looking slightly ahead, Toshiba, in cooperation with IBM Japan, is currently building a plant to produce large color liquid crystal display (LCD) screens. The factory comes after several years of development, culminating in 1988 when Toshiba and IBM Japan produced a prototype 14-inch color LCD, then the largest in the world.

LCD screens are made by injecting a layer of liquid crystal material five-microns thick between two glass plates. To maintain the uniform

THREE TIMES FASTER THAN A BULLET TRAIN

One of Toshiba research's most crowd-pleasing applications will float out of Tokyo's central railway station one day near the turn of the century. For more than 20 years, Toshiba has joined with other manufacturers and Japanese railways in developing a

Toshiba magnets will levitate the MAGLEV train.

magnetically levitated and propelled train. In 1979, a test vehicle attained a speed of 321 miles per hour. When the "MAGLEV" train goes into service on the Tokyo-Osaka run in about 10 years, it will make the 344-mile trip in an hour, easily outpacing Japan's famed Shinkansen "bullet trains," which now travel the route at a pokey 114 miles per hour

Toshiba's contribution has come in the design and building of superconducting magnets that are the core of the vehicle

'UNDER THE TABLE' STUDY

Not all of Toshiba's successes can be traced to careful implementation of the corporate agenda. Individual creativity is part of the Toshiba system, too. "Every member of the Research and Development Center staff is encouraged to spend 10% of regular working hours and budget pursuing independent projects," says Dr. Naohisa Shimomura, director of the Research and Development Center. "We named the program 'under-the-table study' not because of any shady undertones, but because it proceeds without formal project plans, approval requests, and other paperwork lying 'on the table.'"

Developing language analysis techniques for a Japanese word processor propelled Toshiba toward the goal of computers that understand human language. In 1985 Toshiba rolled out a machine that translates documents from English to Japanese with 90% accuracy.

Under-the-table study has, over the years, produced significant results. In 1970, Dr. Kenichi Mori began spending his under-the-table hours pursuing an idea for a Japanese word processor. The project was far trickier than may be immediately obvious to Westerners with languages written in 26 phonetic characters. The Japanese language is largely written with *kanji*, thousands of ideographic characters of Chinese origin that signify ideas or objects. Japanese also includes phonetic *kana* characters, which are used to modify kanji and spell out words for which kanji do not exist.

When Dr. Mori started thinking about a word processor, many Japanese documents were simply written by hand. The only alternative was to type them on a 3,000-character kanji typewriter that could only be operated—slowly—by a specialist.

Dr. Mori set out to produce a practical kana-kanji conversion program with word processing capabilities. He studied grammar and sentence structure, and compiled a dictionary of modern business terms since existing classically oriented Japanese dictionaries lacked such words. Dr. Mori's project soon gained official corporate support and, in 1978, Toshiba marketed the first Japanese word processor. It cost ¥6.3 million and was the size of a desk. Today laptop models cost less than ¥100,000 (US$720). The invention has changed the way a nation writes and works.

One operates a Japanese word processor by typing in phonetic kana characters that represent kanji sounds, and the appropriate kanji characters appear on the screen with speed that belies the complexity of the conversion process. Consider, for example, what the word processor must do when someone types in the five-kana phrase, "ha-shi-ra-na-i." The computer's task is complicated by the fact that there are no spaces between words in Japanese. It is possible that the first word of the phrase is "ha," for which there are no less than four kanji characters, meaning "tooth," "blade," "leaf," and sect."

Applying grammatical rules and consulting a dictionary stored in read-only memory, the computer determines that none of these words makes sense with any possible meaning of the kana that follow "ha." So the computer moves on to consider the first two kana together: "hashi." The three kanji pronounced "hashi" mean "bridge," "edge," and "chopsticks." Again, none of those words makes sense with the kana that follow. So the word processor moves on to interpret the first three kana. "Hashira" can mean "pillar" or "will run." "Pillar" is another reject, but "will run" does make sense with the rest of the phrase, which means "not." The word processor promptly displays the characters signifying "will not run" as the typist moves ahead.

Toshiba took another step toward the widely sought goal of computers that understand human language in 1985 when it began marketing a workstation-based system that translates technical documents from English to Japanese with 90% accuracy. Translations must be edited by a human translator, but the process is speeded considerably by software that presents the human with likely word choices. Documents can be fed via a scanner into the system, which recognizes, for example, that the sentence "I saw a girl with a telescope" has two possible meanings and that "I saw a girl with a bag" does not.

With IBM Japan, Toshiba developed a 14-inch liquid crystal display (above). But technical accomplishments will mean little if the Earth's environment deteriorates. So Toshiba established the Environmental Engineering Lab (left) as its newest R&D center.

thickness of the liquid crystal layer in a large screen, researchers had to improve the mechanism that applies heat and pressure to the glass "sandwich." They studied the physical properties of the glass to insure the heat would be spread evenly. Also, since the material used in smaller LCDs developed excessive resistance in the large panels, Toshiba had to find a new material (a molybdenum-tantalum alloy) to make the electrodes that carry current to the liquid crystals. The new color LCDs will roll off the production line at the new joint venture this spring as compact, lightweight, low-power displays for Toshiba and IBM computers and workstations.

The Electron Devices Laboratory of Toshiba's Research and Development Center also envisions near-future applications for a new type of visible light-emitting semiconductor laser diode. The laser diode has a room-temperature wavelength of 636 nanometers (billionths of a meter), the world's shortest for such a device. It is expected to lead to improved compact and video disk players, bar code readers, and high-speed laser printers.

PROTECTING THE ENVIRONMENT

Toshiba realizes that all its technical accomplishments will provide little comfort if the earth's environment deteriorates. So in 1989 the company established its newest research facility, the Environmental Engineering Laboratory in the Research and Development Center. Staffed by 60 physicists and chemists, the facility consults with other laboratories developing new materials and processes in order to nip potential environmental problems in the bud.

The laboratory has already taken steps to reduce Toshiba's internal use of chlorofluorocarbons (CFCs), which have been implicated in the destruction of the ozone layer. CFCs are mainly used for cleaning critical electronic components in a company like Toshiba. After a meticulous study of the cleaning process, the laboratory developed a line of alternative, non-CFC cleaning fluids They are being used internally by Toshiba and are marketed as new products. Toshiba has pledged to eliminate its use of CFCs by 1995. Once again, scientific research has led to an important application.

For more than a century, this has been the story of Toshiba: identifying society's needs and providing practical and exotic solutions. "Research and development is our tradition," emphasizes Dr. Takayanagi. He likes to quote Kisaburo Yamaguchi, president of Toshiba Corporation from 1927 through 1946. Yamaguchi once remarked that "a manufacturer without R&D facilities is like an insect without antennae."

Toshiba's antennae are still extended.

Edward Zuckerman's book, Small Fortunes, *will be published in January by Viking.*

SCIENCE AND THE CITIZEN

Cosmic cannibalism, gene therapy in the brain, muscular dystrophy clues, cholesterol controversy

Chunnel Vision

An undersea link between Great Britain and France

In 1751 a French farmer dreamed of building an undersea tunnel from France to England. The project received Napoleon's approval in 1802; it was proposed again by a French engineer in 1830 and actually attempted in 1975. Finally, on October 30, 1990, the first of three of the world's longest undersea tunnels linked Great Britain and France for the first time since a land bridge joined the countries during the last ice age, 8,000 years ago.

This coup d'etat of engineering prowess was accomplished by Eurotunnel, an Anglo-French company that was chartered five years ago. Although Eurotunnel set out to run the Channel Tunnel Project like clockwork, more than a bit of dirt jammed its gears.

Eurotunnel initially raised $9 billion for the privately financed project. But investors nearly halted the project in early 1990 when the company asked for an additional $5 billion. Nor did Eurotunnel foresee all the technical obstacles to plowing through the 52.5 kilometers of chalk between terminals on either side of the Pas-de-Calais, the narrowest part of the English Channel.

Still, the effort by Eurotunnel and its partner, Transmanche Link, a consortium of French and British construction companies, is the landmark project of 20th-century civil engineering. More than 14,000 workers overcame enormous challenges in removing several million metric tons of earth, rock and "muck" from the ground. They accurately maneuvered giant drills, or tunnel-boring machines, in wet conditions under extraordinary pressure. They repaired jammed conveyors and debugged untested tunneling technologies. They patched segments of tunnel where jets of water showered high-voltage equipment. So far nine have died.

This month Eurotunnel is scheduled to complete the construction of the "service" tunnel, which will serve as an access for rubber-tired utility vehicles. The company has already excavated 70 percent of two flanking "running" tunnels that will carry high-speed trains. If progress continues at the current pace, the railroad tunnels will be finished by the end of this year, and by June 1993 passengers and vehicles will travel between terminals at Calais and Folkestone in 30 minutes.

From the beginning the Channel Tunnel, or Chunnel, project has been a curious hybrid of old and new, a grab bag of ancient engineering practices and high-tech wizardry. Although surveyors standing on opposite shores precisely measured the distance separating them with radar waves, they calculated their exact positions by triangulation and dropped plumb lines to fix their starting points.

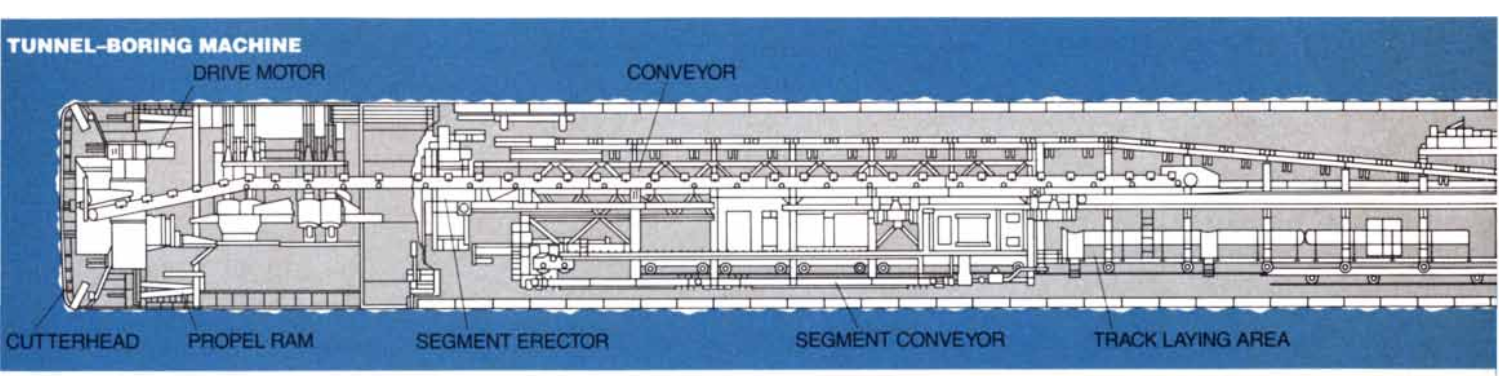
Underground, technicians had no better option than dead-reckoning to steer their high-technology tunnel-boring machines, or moles. Yet each machine was kept on a straight path by a laser beam. As the moles bored from opposite sides of the Channel, on-board computers calculated the beam's position and angle, compared the progress with a three-dimensional map of the tunnel held in memory, and displayed instructions for the machine's operator.

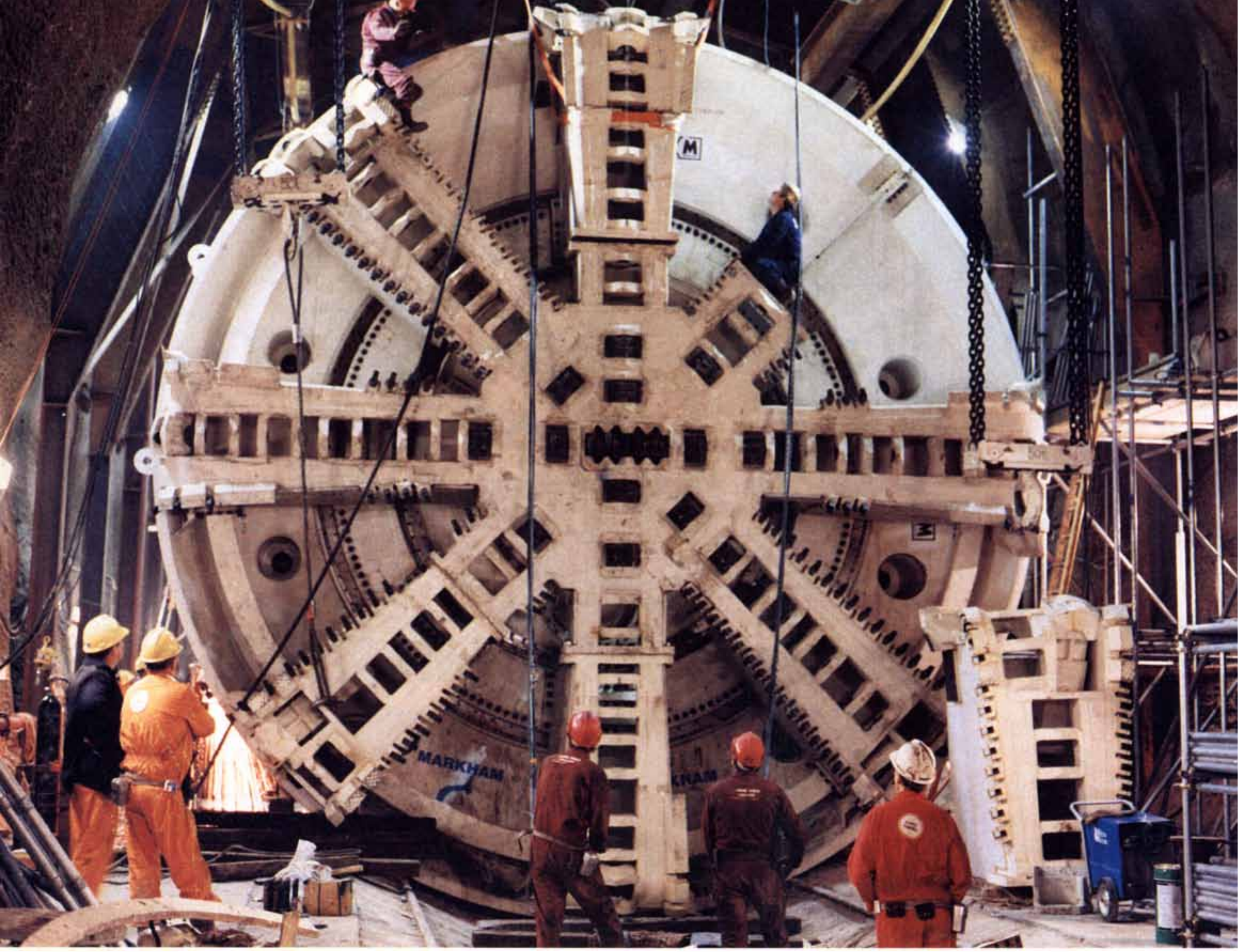
On each coast, two boring machines were positioned in each tunnel, one digging out under the channel, the other headed inland to the site of future train terminals. British tunnelers had a head start over their French counterparts because the tunneling attempt of the 1970s left two access tunnels in Shakespeare Cliff, one of the white cliffs south of Dover.

Each of the boring machines used for the project is a huge, cylindrical monster that bores like a giant worm, first scraping off chalk with tungsten picks and cutting disks mounted on its slowly rotating head and then digesting the muck through its body of pumps and conveyor belts.

The most severe logistical problem of the project has been transporting workers, equipment and supplies to the cutting faces and carting away the spoil, or "muck." Since the Channel Tunnel has no access shafts along its 37.9-kilometer underwater leg, the only way to get in or out is through the two work sites.

The transportation system has grown to a network of more than 100 locomotives operating on 190 kilometers of track. The locomotives are powered by overhead electric lines and intermittently by batteries. (Diesel locomotives were ruled out at first because of poor ventilation.) The batteries power the locomotives for up to one kilometer—for instance, to cover the last few hundred meters of the tunnel, where work in progress makes power lines too dan-



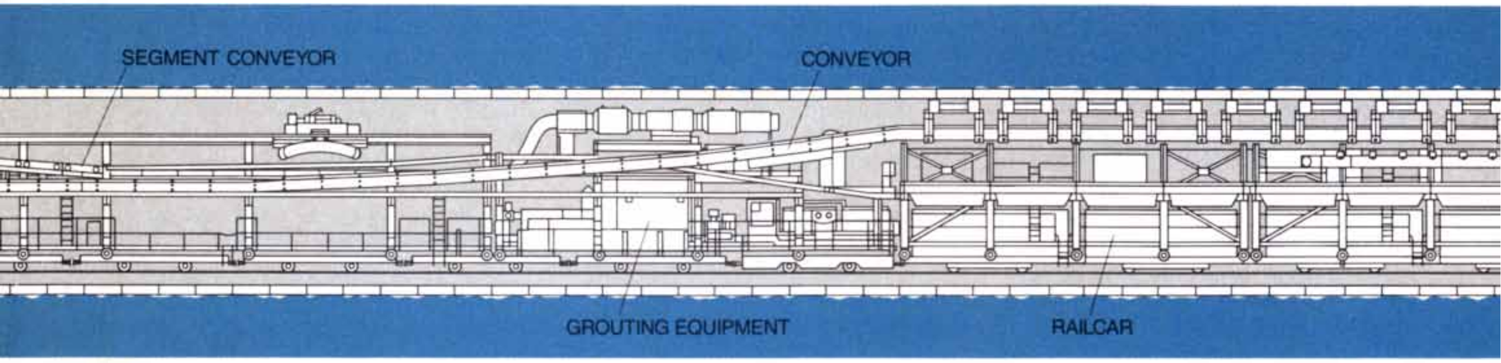


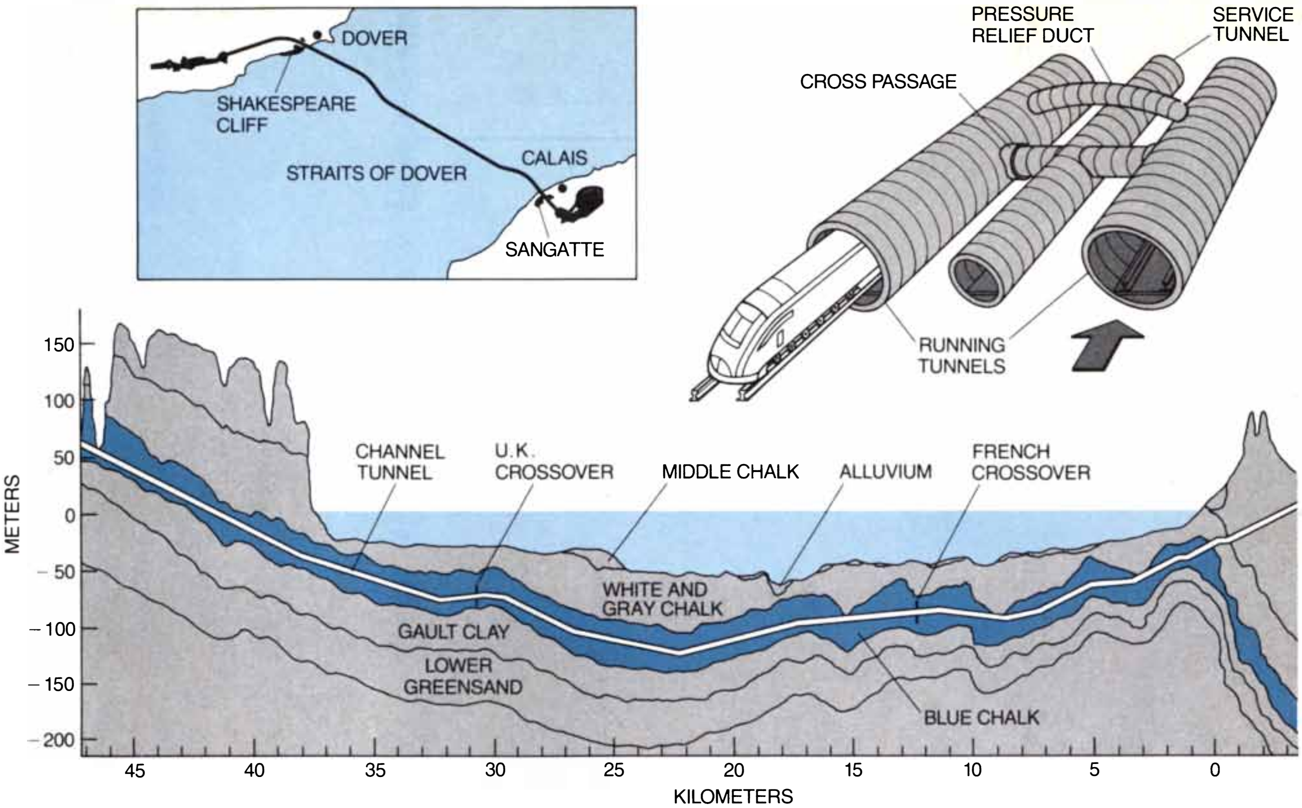
***CUTTING EDGE** in tunnel-boring technology is an 8.4-meter machine that is currently digging through chalk underneath the English Channel. QA Photos.*

gerous to install. On both sides of the Channel, the power lines initially proved unreliable, and trains had to rely more than expected on batteries.

On the English side, the trains caused problems from the start. Because the boring machine for the service tunnel weighed 1,500 tons and spanned 300 meters, it was assembled piecemeal in the access tunnels under Shakespeare Cliff. Trains brought additional parts as the boring machine dug itself more space. The access tunnels, however, turned out to be too narrow, and the boring machine parts too big, to leave any space for power lines.

As a result, engineers were forced to operate the trains solely on battery power. On the surface the exhausted batteries were removed and replaced with fresh ones for the next trip down the





1.5-kilometer tunnel. So the tunnel-boring machine often sat idle, waiting for parts and supplies. Because of these delays, it took six months to put the whole machine together.

Once the digging was started in earnest, however, engineers on the British side thought the drilling would go smoothly. The bulk of the tunnel runs through a layer of chalk marl—a soft rock that is impermeable to water and that retains its shape for many years after it is dug. The team was also fortunate because the stratum of chalk was thick and almost horizontal, allowing the team to plot a rather straight course.

These conditions allowed British engineers to use an exceptionally quick method for lining their tunnels. Each time the mole advanced 1.5 meters, tunnelers working in the mole's tail arranged blocks of precast concrete to form a ring. The blocks were held in place with a small wedge at the top of the ring. Once a film of concrete grout was injected between the tunnel wall and the ring, the blocks formed a classical double arch, redirecting continuous pressure from the surrounding chalk laterally along the concrete walls.

The British tunnelers had trouble with some patches of wet chalk between the first and the fifth kilometer. In the service tunnel, since they had prepared themselves for the worst, they were able to switch to a cast-iron lining, which was more expensive and difficult to install but watertight. In the running tunnels, however, they had equipped themselves only with precast concrete lining segments, which, until they could be grouted at a later stage, leaked like sieves. Electric trains were running over tracks fully submerged in water. The electric system failed repeatedly. The electrohydraulic rams used to drive the mole through the chalk were in constant danger of breaking. Workers were pelted with water. "If you've ever been in a hailstorm, it was like that," recalls Andrew Copperwheat, project design engineer at Hunslet GMT Ltd., which supplied the locomotives.

"We didn't know enough about the leakage of electric systems in these moist saline conditions. It took 12 months before we pioneered our way out of the technical difficulties," comments John R. J. King, tunnels director at Transmanche Link. In the end the company was forced to convert 10 of its electric locomotives to diesel power.

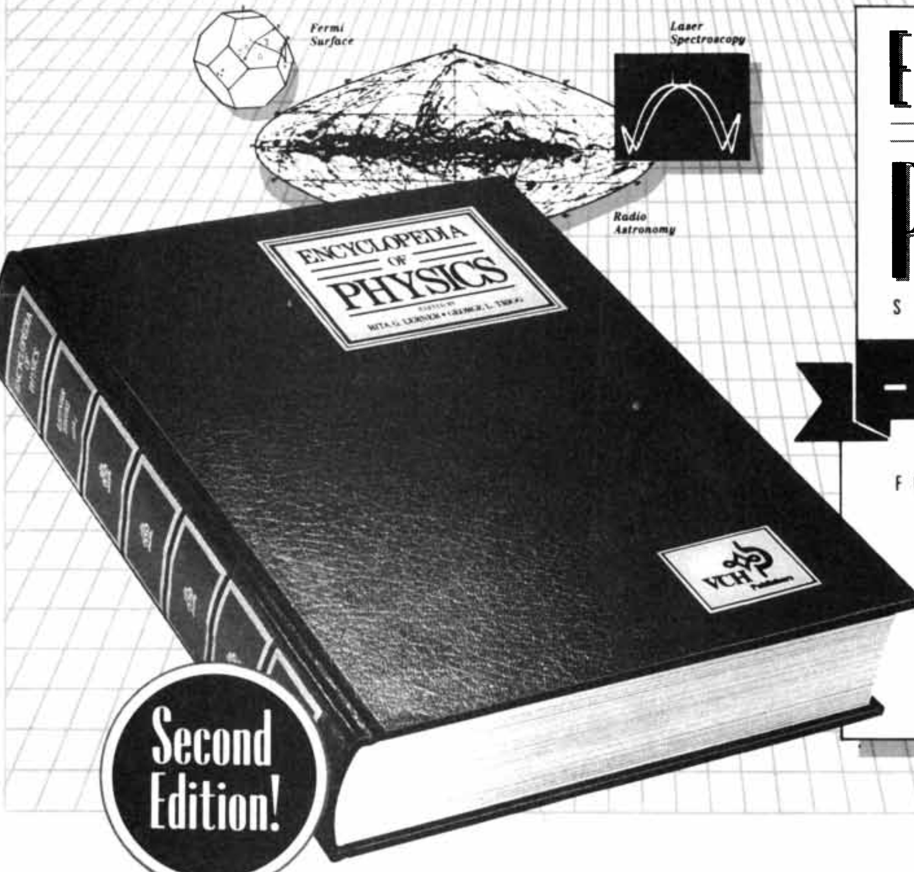
The French sandhogs had an even tougher time. Not only were their work trains plagued with similar problems, they also had geology working against them. Because the chalk marl takes a sharp dip downward just as it reaches France, engineers first had to dig through a layer of gray chalk—a brittle material highly permeable to water—before reaching the impermeable layer below.

To get their tunnel-boring machine in place, the French dug a huge cylindrical shaft, 55 meters in diameter and 66 meters deep, at Sangatte. Once the shaft was waterproofed, four cranes were installed over the opening to lower parts of the machine.

Because the French tunnelers expected to encounter a great deal of water, they commissioned a boring machine that could withstand almost 10 atmospheres of water pressure—far more than the three or four atmospheres usually encountered in underwater tunnels. The machine had a sealed, watertight head from which the wet, mushy muck was pumped with an Archimedes screw—a rotating shaft with spiral grooves—from the pressurized chamber onto a conveyor belt in the work area. The machine also provided a watertight workspace where workers assembled a precast concrete tunnel lining. Unlike the lining used on the British side, the French lining had to be laboriously bolted together and made watertight with rubber gaskets.

Hopes of a quick start on the French side were dashed when a financially troubled company in the north of France canceled its contract to manufacture the boring machine for the service tun-

COMPLETELY REVISED ❑ UPDATED ❑ EXPANDED



ENCYCLOPEDIA OF PHYSICS

SECOND EDITION

– A $150.00 VALUE –

FOR ONLY **$4.95!**

when you join the
LIBRARY OF SCIENCE

You simply agree to buy 3 more books—at significant savings—over the next 12 months.

- **Large 8½" x 11" format—over 1,500 pages**
- **More than 500 entries by eminent scientists**
- **Over 750 photos, graphs, diagrams and charts**
- **Completely cross-referenced and fully indexed**
- **Edited by Rita G. Lerner and George L. Trigg**

When first published, the *Encyclopedia of Physics* won international acclaim for its broad and authoritative coverage of state-of-the-art physics. *American Scientist* called it "...clearly the definitive encyclopedia of physics." *Physics Today* recommended it as "an open-ended invitation to a self-paced course in any and all of physics."

Now in its second edition, the Encyclopedia keeps pace with the very latest developments in all branches of physics—from radio astronomy, elementary particles, black holes, and unified field theory to semiconductors, biophysics, geophysics and physical chemistry.

More than 500 world-renowned scientists have prepared special entries for their respective fields, including such Nobel Prize winners as PHILIP W. ANDERSON (Quasi-particles), OWEN CHAMBERLAIN (Nucleon), SHELDON GLASHOW (SU^3 and Higher Symmetries), MELVIN SCHWARTZ (Leptons), BURTON RICHTER (Positron-Electron Colliding Beams), and KLAUS VON KLITZING (Quantum Hall Effect).

To reflect dramatic breakthroughs and discoveries, the second edition of the *Encyclopedia of Physics* features many all-new entries. Here is a selected listing:

- NEUTRON ASTRONOMY
- CHAOS
- COSMIC STRINGS
- FRACTALS
- GRAVITATIONAL LENSES
- HIGH FIELD ATOMIC STATES
- JAN-TELLER EFFECT
- QUARKONIUM
- SCANNING TUNNELING MICROSCOPY
- SOLITONS
- SUPERSYMMETRY AND SUPERGRAVITY

With its new additions and revisions, the *Encyclopedia of Physics* remains the reference of choice not only for physicists, engineers and scientists in related fields, but for anyone with an avid interest in science.

This outstanding volume, published at $150.00, can be yours for just $4.95. Don't delay. Fill in the convenient reply card and send for your copy today.

THE LIBRARY OF SCIENCE is the nation's oldest and most respected science book club. Dedicated to serving both professionals and serious amateurs, it surveys the entire field of published sources on scientific subjects to make the very finest works available to club members at substantial savings.

MEMBERSHIP BENEFITS • In addition to getting *The Encyclopedia of Physics* for only $4.95 when you join, you'll also receive discounts on books you choose to purchase. Discounts generally range from 15% to 30% off the publishers' prices, and occasionally even more. • Also, you will immediately become eligible to participate in our Bonus Book Plan, with savings of 60% off the publishers' prices. • At 3-4 week intervals (15 times per year), you will receive the Library of Science News, describing the coming Main Selection and Alternate Selections, together with a dated reply card. • In addition, up to two times a year, you may receive offers of Special Selections which will be made available to a group of select members. • If you want the Main Selection, do nothing, and it will be sent to you automatically. • If you prefer another selection, or no book at all, simply indicate your choice on the card and return it by the date specified. • You will have at least 10 days to decide. If, because of late mail delivery of the News, you should receive a book you do not want, we guarantee return postage.

If reply card is missing, please write to Library of Science, Dept. 2-DC4/44575, 3000 Cindel Drive, Delran, NJ 08075, for membership information and an application.

Scientific American 1/91 © Newbridge Communications, Inc.

Newly excavated section of the running tunnel. QA Photos.

The access shaft in Sangatte, France. QA Photos.

nel. The machine's designer, the Robbins Company, scrambled to build the machine itself in Portland, Ore., delivering it to Sangatte three months late. A faulty discharge pump and a jammed screw conveyor further delayed the start. "All kinds of things didn't work at first, mainly because of the novelty of the design problems," says Richard Robbins, president of the company.

As a result, the French engineers logged only one kilometer in their first year of drilling. After digging through the first 500 meters of gray chalk, where the friendlier chalk marl began, the French moles had to weave their way through peaks and valleys in the stratum, occasionally nicking the permeable gray chalk above and passing through several fissures.

Despite the slow starts on both sides of the Channel, progress picked up considerably after the first year. The British team avoided water leakage problems in its two running tunnels by improving its grouting technique. As a result, by mid-1990 it was regularly logging more than 300 meters of tunneling per week.

The French team, its luck taking a turn for the better, made even more striking improvements. Its boring machines were soon exceeding speed expectations, logging 290 meters in their best week and pushing the meeting point for the running tunnels slightly toward the English side. As of November, 14.6 kilometers remained to be dug in the tunnel running north and 18 kilometers in the one running south. "Despite all of the problems we've encountered," King says, "it seems we will be able to meet our mid-1991 deadline" for completing the digging.

A major technical challenge for both the French and British teams was to dig the two underground caverns—each 60 meters high and 20 meters wide—where the two running tunnels merge. These so-called crossover caverns are designed to house a diamond switch so that trains can change from one tunnel to the other. They will provide a way to close down a segment of the tunnel for maintenance and inspection without interrupting service. Digging the crossover caverns required considerable finesse. Workers first excavated several small tunnels and then sprayed concrete into them to make temporary retaining walls. Later they dug through the concrete and chalk that remained between the tunnels, leaving one huge cavern.

Now that Transmanche Link and Eurotunnel are well acquainted with the geology of the Channel and the worst of the digging is over, engineers can start the second phase of the Chunnel proj-

ect: the installation of fixed equipment such as rails, ventilation equipment, high-intensity lights and cooling pipes. Passenger safety has imposed rigid design constraints. One of the biggest threats is fire. An emergency escape route is provided by the service tunnel, which is connected to the running tunnels by scores of large ventilation shafts and pumping stations.

With the link-up completed and the celebrations over, the prospect of more traffic, higher real estate prices and noisy trains rumbling toward London or Paris are becoming more controversial than safety and engineering issues. And the rest of the job is beginning to look like routine slogging. But even when the first trains begin running, it will be hard to forget that the Chunnel realizes a dream that has persisted for more than 300 years.

—Fred Guterl, London, and Russell Ruthen

Short-Order Cook

A stellar cannibal yields insights to pulsar evolution

Near the center of our galaxy, a kind of feeding frenzy for celestial objects is taking place. An international team of astronomers has discovered a pulsar that apparently "cooks off" parts of a companion star. The ablated gas then spirals into the pulsar, occasionally eclipsing the radio signals. The observation provides evidence for the reigning theory that describes how a new class of pulsars evolves.

Andrew G. Lyne of the University of Manchester, Richard N. Manchester of the Australia Telescope National Facility and their colleagues detected the binary system by pointing their radio telescopes to the globular cluster Terzan 5, about 21,000 light-years from the earth. There they found a pulsar, dubbed according to its coordinates as PSR 1744-24A, that spins once every 11.56 milliseconds.

Such rapidly spinning, or millisecond, pulsars have apparently taken an evolutionary track unlike that of other pulsars. Most pulsars have periods greater than 0.1 second and are thought to have resulted directly from the process that forms a supernova. Outer parts of the star explode to form a nebula. Inner parts collapse, becoming so dense that electrons combine with protons to form neutrons. As the neutron star spins, it acts as a lighthouse, sending out a beacon of radio waves.

That does not seem to be the case for millisecond pulsars. Astronomers suspect that such pulsars begin life as part of a stellar binary. They develop slowly by stealing matter from the companion star. As the pulsar continues to accrete material, its rotational speed increases.

In 1982 Donald C. Backer of the University of California at Berkeley, Shrinivas R. Kulkarni of the California Institute of Technology and their colleagues obtained evidence for the scenario. They discovered the fastest-rotating pulsar, with a period of 1.56 milliseconds—about 100 times faster than typical pulsars. The record-setting period and the weak magnetic field suggested that millisecond pulsars formed slowly, perhaps over tens of millions of years. Then, in 1988, Andrew S. Fruchter and his colleagues at Princeton University found an eclipsing binary pulsar system. In this system, which consists of a star in orbit around a millisecond pulsar, the companion periodically blocks signals from the pulsar.

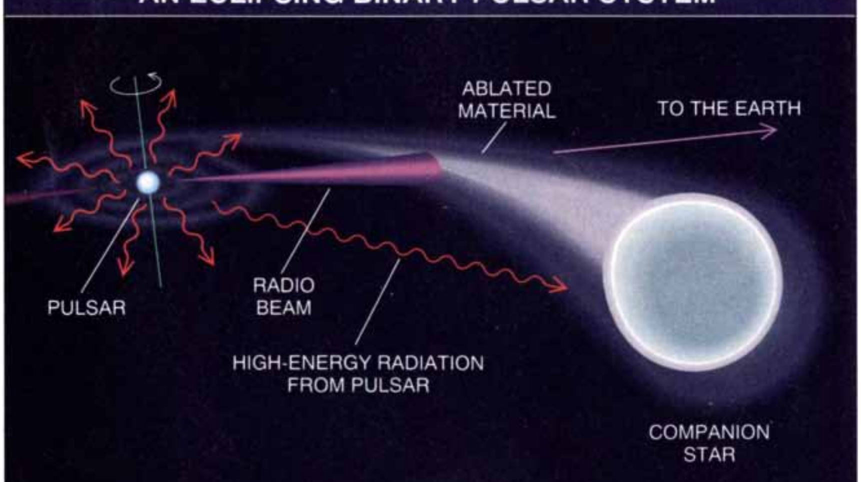
The eclipsing pulsar found last year by Lyne and his colleagues may be a missing link in the evolutionary chain. Based on the sporadic eclipses and slower rotational period, PSR 1744-24A seems to be younger than the eclipsing system discovered by Fruchter.

According to Lyne, the new pulsar was probably once a neutron star on its deathbed, spinning slowly and not emitting much radiation. But in globular clusters, "it is not unlikely that it could have a collision of some sort," Lyne says. The gravity of the neutron star captured a small star and began sucking up the companion's outer material when the companion star reached the giant stage in its evolution. Such systems are called low-mass X-ray binaries because of the small size of the stars and the emission of X rays that occurs when matter falls into the neutron star.

As the neutron star continued to gain mass, it began spinning faster, eventually rotating fast enough to begin a new life as a pulsar. As a pulsar, it no longer has to rely solely on its own gravity to draw in sustenance. The rapid rotation causes pulsars to emit energetic radiation capable of vaporizing matter. "What we're seeing now is the pulsar cooking its companion," Lyne says. As the gas comes off the companion, it spirals into the pulsar, sporadically blocking the radio signals. The material from the companion—now suspected to have evolved into a white dwarf—seems to engulf the pulsar on occasion. The pulsar may continue to draw material until its white dwarf companion has been "accreted to death."

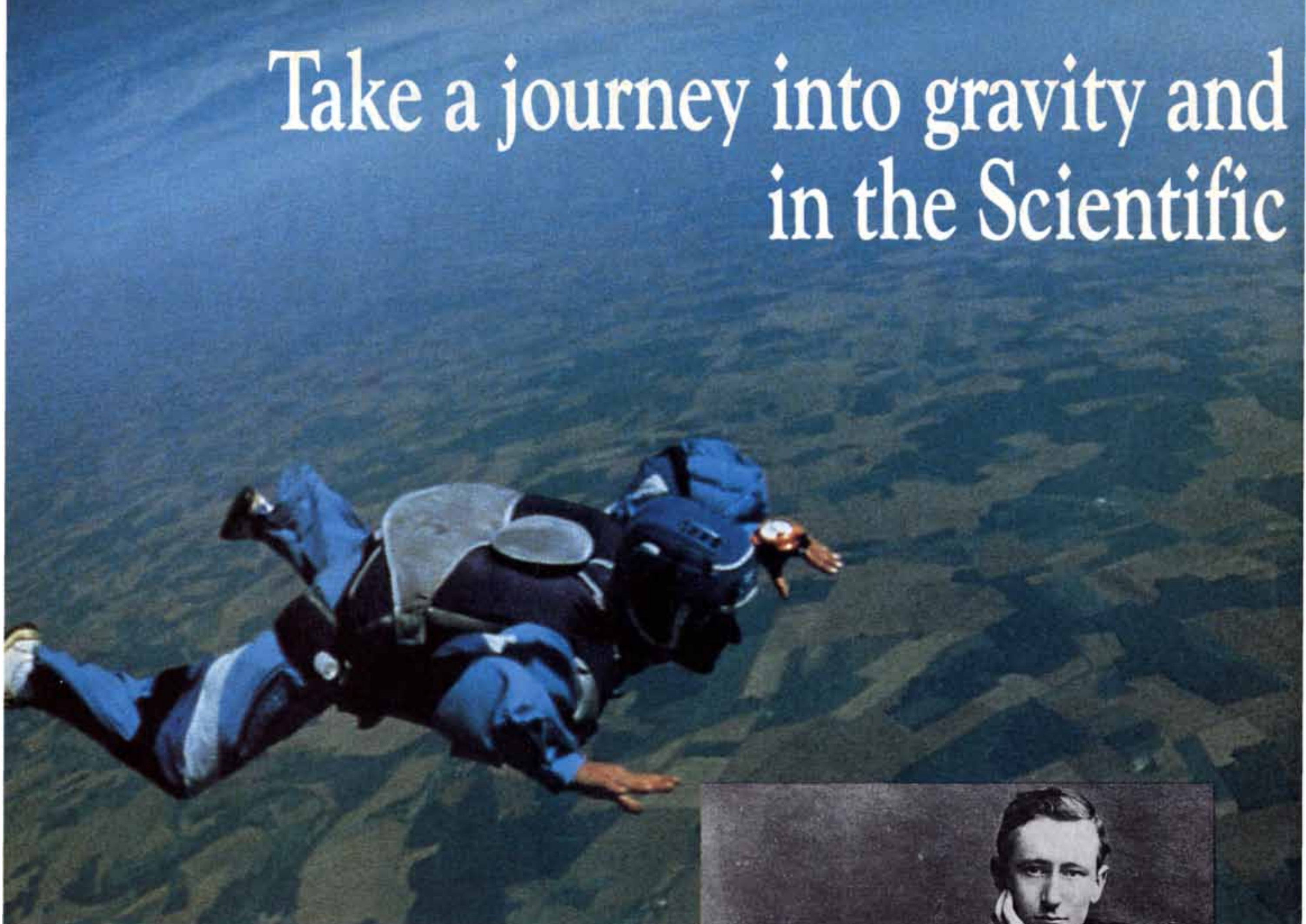
Binary pulsars are often considered ideal sources by which to test general relativity. Astronomers think binary systems lose energy in the form of gravity waves, and the slight change in the orbit and pulse period should be detectable. Unfortunately, the latest find lies too close to the galactic center. "It may be possible to detect relativistic effects, but measurements will be prone to uncertainty," Lyne remarks. "There's a lot of rubbish around." *—Philip Yam*

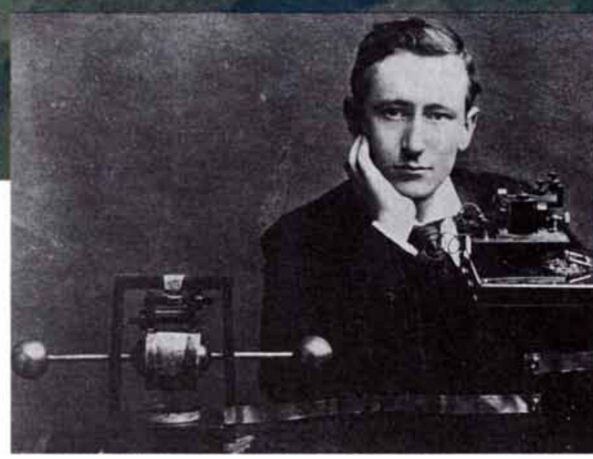
AN ECLIPSING BINARY PULSAR SYSTEM



Energetic radiation from a recently discovered pulsar apparently "cooks" the surface of a companion—probably a white dwarf star. The ablated material then spirals into the pulsar, occasionally eclipsing the radio beam.

Take a journey into gravity and in the Scientific



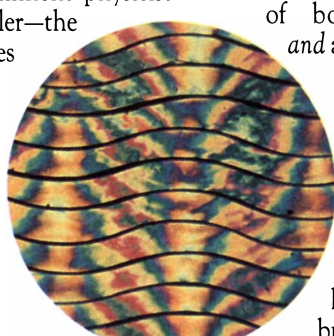


From Signals, *Guglielmo Marconi in 1896. Courtesy Marconi Company Ltd.*

Join the Scientific American Library now by previewing *A Journey into Gravity and Spacetime* for 15 days risk-free.

Gravity controls the movement of sky divers, stars, and entire galaxies. It has been studied by the most renowned scientific minds in the world yet it remains a profound mystery of nature.

Now you can see how the latest thinking on this intriguing force is shaping our view of the universe. Join the SCIENTIFIC AMERICAN LIBRARY and receive *A Journey into Gravity and Spacetime* by the preeminent physicist John Archibald Wheeler—the first in a remarkable series of books that will take you on a wondrous journey to the frontiers of scientific discovery.



From The Science of Structures and Materials, *a polarized light view of a stressed graphite fiber-epoxy resin composite. Courtesy Tsu-Wei Chou.*

Written by leading scientists in accessible, compelling prose

SCIENTIFIC AMERICAN LIBRARY books are created in a unique collaboration between our editors and some of the world's most renowned scientists (including many Nobel laureates). The result is a series of books that are readable *and* authoritative.

Each book:

- Covers a topic of major interest with current relevance to ongoing research.
- Presents the firsthand views of a scientist who has made significant contributions to the field.
- Delivers clear, compelling prose that even the non-expert will find fascinating.

For sheer editorial quality, no other science books can match this fine collection.

Beautiful, richly illustrated volumes that are made to endure

In addition to entertaining and informing, SCIENTIFIC AMERICAN LIBRARY books provide you with an excellent resource for study and research. Each volume features abundant photos, paintings, and illustrations, most in full color. Helpful diagrams, charts, and tables clarify points made in the text.

And SCIENTIFIC AMERICAN LIBRARY books are produced to exacting

Neural Vector

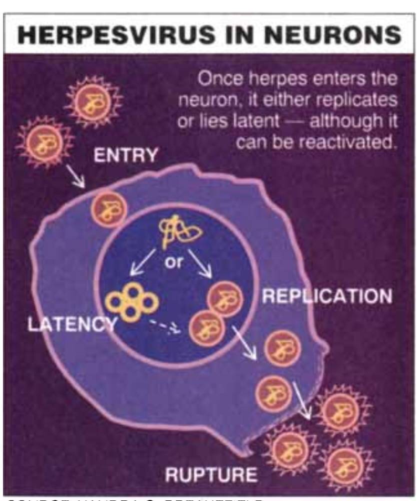
Herpes may open the way to gene therapy in neurons

By turning a deadly virus to their advantage, scientists hope to make possible gene therapy in the brain. The virus—herpes simplex—is naturally attracted to nerve cells, making it an ideal carrier for foreign genes that could correct diseases of the nervous system. Although little has been published on this novel use of herpes simplex, researchers are racing to enhance the benefits of the virus—and to keep its dangers at bay.

"This is really exciting because neurobiologists are dying for a way to do gene therapy in the brain," says Joseph C. Glorioso, chairman of molecular genetics and biochemistry at the University of Pittsburgh. "And this, I think, is the only way to do it." Although human tests may be years away, scientists believe the technique could open the way to treat such diseases as Alzheimer's, Parkinson's and Huntington's.

Some pioneers of gene therapy are inserting genes into cells with a type of virus called a retrovirus, which can incorporate the corrective gene into the DNA of the cell it infects. For the virus to achieve this integration, however, the infected cell has to be reproducing. Because mature neurons no longer reproduce, retroviruses cannot work in the nervous system.

In contrast, the herpes simplex virus not only infects neurons, it expresses foreign genes incorporated into its DNA—without requiring that the neuron itself reproduce. The virus remains quiescent in these cells, meaning that it does not replicate and kill the nerve cells unless triggered by stress or perhaps some element in the environment.



SOURCE: XANDRA O. BREAKEFIELD

Although 90 percent of humans carry herpes simplex, it usually resides quietly in peripheral neurons—not those in the brain, explains Xandra O. Breakefield, a neurogeneticist at Harvard Medical School. "It is the only virus that we know of that can go into a neuron and stay latent," she says.

Latency is crucial because even though the virus does not reproduce, it can express certain genes—ones that are inserted or controlled by researchers. "The groundwork is there," says Breakefield, whose laboratory is one of several that has successfully expressed genes that can dye the neuron blue—a colorful test that proves the virus is working as planned. "Now we need to put in a meaningful gene," she adds.

The herpes simplex virus has other advantages as well. Because it is very large—composed of some 70 genes—it can shuttle large foreign genes, or more than one smaller gene, into neurons. Certain strains of the virus can also enter the peripheral nervous system and travel to the central nervous system, obviating the need to inject the virus directly into the brain.

Despite these benefits, it remains potentially deadly. "The issue of gene therapy has to be treated very carefully," Breakefield cautions. "We are talking about a virus that can kill the brain." At any point—which scientists cannot identify—the virus can become active, reproduce and destroy the neuron.

To counteract its toxicity, some laboratories use weakened or defective strains of the virus that are not always lethal. (Some such strains, however, do lead to brain infections in laboratory animals.) Scientists also alter the virus so that it no longer replicates.

Howard Federoff, a molecular biologist at the Albert Einstein College of Medicine, and his colleagues deleted more than 80 percent of the viral DNA, producing a harmless virus that cannot reproduce. Federoff also inserted a promoter, which activates certain genes, specific to neurons. Therefore, the genes are expressed only in nerve cells.

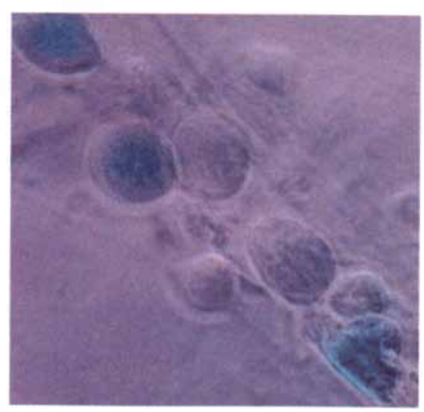
Because Federoff greatly shortened the virus to make it benign, some argue that it will not express genes in live animals over a long period. The virus has to resemble its original state as much as possible to be useful for gene therapy, Breakefield says. But Alfred I. Geller, a research associate at the Dana-Farber Cancer Institute in Boston, who collaborated with Federoff, notes that their form of the virus does express genes in adult rats for at least six weeks.

Other laboratories use a different system to prevent the virus from reproducing. Rather than greatly altering the virus, they repress its replication and capacity to reactivate by eliminating two genes necessary for these functions, says Jack G. Stevens, professor and chairman of microbiology and immunology at the University of California at Los Angeles. "You can get this agent to express any gene you want to," says Stevens, who is currently studying the introduction of nerve growth factor to neurons via the viral shuttle.

At Pittsburgh, Glorioso also uses a repressor system. He says that in a few months he will use the virus as a carrier of the tyrosine hydroxylase gene—which transcribes an enzyme leading to the production of the neurotransmitter dopamine—to treat Parkinson's disease in primates. (The tremor and muscular rigidity experienced by people with Parkinson's is caused, at least in part, by the shortage of dopamine.) Some investigators caution that the long-term gene expression Glorioso has reported could be interpreted as an infection; however, Glorioso says he finds no evidence to support this contention.

Aside from its potential for replication or reactivation, other aspects of the virus can be dangerous to the cells. "Some of its protein coat is toxic," Breakefield says, although she adds that only neurons in culture dishes appear to be bothered by this viral packaging. The virus can also undergo mutations, and "many mutations are toxic," says Paul A. Johnson of the University of California at San Diego. Johnson is constructing different herpes mutants in an effort to reduce cytotoxicity.

Despite the hurdles, researchers working with herpes simplex are excited. "The virus seems absolutely tailored for gene therapy," Stevens says. "I am normally quite conservative," he adds, "but this has a tremendous amount of potential." —*Marguerite Holloway*



NEURONS infected by the herpesvirus express a gene causing the infected cells to turn blue. Photo: courtesy Alfred I. Geller.

Trans Fat

Does margarine really lower cholesterol?

Butter, as anyone who has not been living in a cave for the past 10 years has probably heard, contains a lot of saturated fat, which increases the levels of cholesterol in the blood. Margarine, on the other hand, is made from vegetable oils, which contain cholesterol-lowering polyunsaturated fat. So switching to a diet with only vegetable fats should lower cholesterol levels, right?

Wrong, says Margaret A. Flynn, a nutritionist at the University of Missouri. When she performed the experiment with a group of 71 faculty members—switching in both directions—she found that "basically it made no difference whether they ate margarine or butter." The reason, according to a growing group of nutritionists, could be partially hydrogenated fats. Recent studies suggest that such fats might actually alter cholesterol levels in the blood in all the wrong ways, lowering the "good" high-density lipoprotein and increasing the "bad" low-density lipoprotein.

Partially hydrogenated fats are made by reacting polyunsaturated oils with hydrogen. The addition of hydrogen turns the oils solid, and some of their polyunsaturated fat is turned into *trans* monounsaturated fats. Monounsaturated fat is generally perceived as good, but things are not so simple. "*Trans* monounsaturates act in the body like saturated fats," says Fred A. Kummerow, a food chemist at the University of Illinois at Urbana-Champaign. Almost all naturally occurring monounsaturated fat is of the *cis* variety, which is more like polyunsaturated fat.

Flynn's study is not the first to raise questions about *trans* fatty acids. Ten years ago a Canadian government task force noted the apparent cholesterol-raising effects of *trans* fats and requested margarine manufacturers to reduce the amounts—which can easily be done by altering the conditions of the hydrogenation reaction.

Last August two Dutch researchers, Ronald P. Mensink and Martijn B. Katan, published a study in the *New England Journal of Medicine* that showed eating a diet rich in *trans* fats increased low-density lipoprotein and decreased levels of high-density lipoprotein. In an editorial accompanying the study, Scott M. Grundy, a lipid researcher at the University of Texas Southwestern Medical Center at Dallas, wrote that the ability of *trans* fatty acids to increase low-density lipoprotein "in itself justifies their reduction in the diet." Grundy called for changes in labeling regulations so that cholesterol-raising fatty acids, including *trans* monounsaturates, are grouped together.

James I. Cleeman, coordinator of the National Cholesterol Education Program, disagrees. "To raise a red flag is premature," he says. "Mensink's audience is the research community—the public needs usable simplifications." Cleeman points out that the subjects in Mensink and Katan's study ate relatively large amounts of *trans* fats. He believes more typical consumption levels should be investigated before any change in recommendations is warranted.

Furthermore, Cleeman notes that studies like Flynn's are hard to interpret because subjects were allowed to eat as they pleased. Flynn's study, published this month in the *Journal of the American College of Nutrition*, found considerable variability among subjects in their blood lipid profiles. "The only way to study the question properly is in a metabolic ward," Cleeman says. "*Trans* fats are a wonderful example of an issue that's not ready for prime time."

Edward A. Emken, a specialist on *trans* fats at the Agricultural Research Service in Peoria, Ill., also downplays the concern but for different reasons. Although Mary G. Enig, a nutritional researcher at the University of Maryland, has estimated American adults consume 19 grams of *trans* fat per day, Emken thinks that figure is too high. According to his calculations, eliminating *trans* fatty acids from the diet will for most people make only a tiny change in lipoprotein levels. "If you're hypercholesterolemic, it could be important, but if you're not, then it is not going to affect risk at all," he concludes.

Emken, together with Lisa C. Hudgins and Jules Hirsch, has performed a study, to be published in the *American Journal of Clinical Nutrition*, that finds no association between levels of *trans* fats in fat tissue in humans and their cholesterol profiles. To Emken, that suggests *trans* fats are not a major threat for most people.

Nevertheless, *trans* fats seem destined for more limelight. "How can one defend having cholesterol and saturated and unsaturated fats listed on food labels but not allow public access to *trans* information when such fats behave like saturates?" asks Bruce J. Holub, a biochemist at the University of Guelph in Ontario. "At the very least, one has to ask whether cholesterol-free claims should be allowed on high-*trans* products." —*Tim Beardsley*

Wavelet Theory

An analysis technique that's creating ripples

It analyzes fractals, compresses images and sounds, tracks stock market fluctuations and recovers information from the murkiest pictures. It's a wavelet—a small mathematical fluctuation typically containing a single central spike and two small negative spikes on either side. Researchers at university and industrial laboratories are feverishly working out the implications of wavelet theory, and a Massachusetts start-up has already announced special-purpose wavelet transform microchips.

The creators of wavelet theory, meanwhile, are already backpedaling, trying not to oversell their work and create a backlash. "It's not going to solve the problems of the universe," comments mathematician Ingrid Daubechies of AT&T Bell Laboratories. "You have to know your problem." Adds Gregory Beylkin, a mathematician at Schumberger-Doll Research: "It will take several years before the impact is felt." Although the essential theory is there, he says, efficient algorithms for all but the most basic wavelet operations have yet to be developed.

For practical purposes, wavelets are analogous to Fourier transforms, which are used to analyze signals in terms of their frequency components or characteristic wavelengths. Fourier transforms can reconstruct sounds, for example, by summing different proportions of various pitches. But wavelets surpass Fourier transforms in analyzing signals that change rapidly. Wavelets provide information about the location of features: they can detect and highlight the onset of a spoken syllable or the attack of a trumpet note.

Daubechies has helped develop an algorithm that computes wavelet transforms suitable for analyzing sound and video significantly faster than even the best Fourier algorithms can. Her method, whose complexity is proportional only to the number of data points analyzed, is being adapted to microchip form by Aware, Inc., in Cambridge, Mass. And that is just the beginning of uses for this "big bag of tools," says mathematician Ronald Coifman of Yale University. Speech and image recognition, statistical analysis and simulations of turbulent fluid flow could all make use of wavelets.

Coifman's own favorite is a new adaptive "wavelet packet transform" that can represent signals in an extraordinarily compact fashion. The transform is good for compression, but, even more important, it elucidates the structure of a signal. "When you have done the transform, now you have said something important," he says. Adaptive wavelet transforms of electrocardiograms, for example, might help clinicians determine exactly what characteristics distinguish the beating of healthy and diseased hearts, he predicts.

Wavelets do not merely transform data. According to Beylkin, they can be used to simplify the mathematical transformations that physicists and others use to represent the world. Adopting wavelet representations, he says, could reduce a problem requiring 10,000 operations to one requiring only 100—or reduce a trillion operations to only a million.

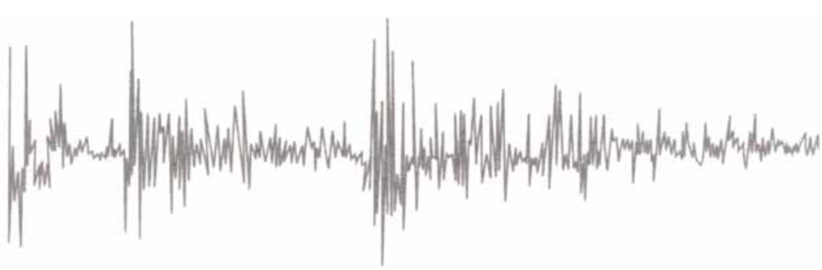
Yet adapting current wavelet techniques to such new problems is not straightforward, Beylkin says. Computers run out of memory on all but the smallest data sets. Daubechies concurs: the major challenge for wavelet theorists is to extend the success they have had on one-dimensional signals—such as sound—to two, three or more dimensions. "Simple" multidimensional transforms, she says, are constrained by "preferred directions"; they may work to compress images containing mostly horizontal and vertical lines, for example, but diagonals elude them.

Daubechies, Beylkin, Coifman and other mathematicians responsible for wavelet theory are more than a little bemused to see their work so suddenly taken up in so many disparate directions. The initial insights, Coifman notes, were known in pure mathematics by the early 1950s but never spread to other fields.

In the mid-1980s mathematician Yves Meyer of the University of Marseille told Jean Morlet, a petroleum engineer at Elf-Aquitaine, that Morlet had merely rediscovered the mathematicians' work. Only then did researchers begin to build bridges between disciplines. Since then, progress has been rapid, and new theoretical results may be incorporated in software or chips almost as fast as they are derived.

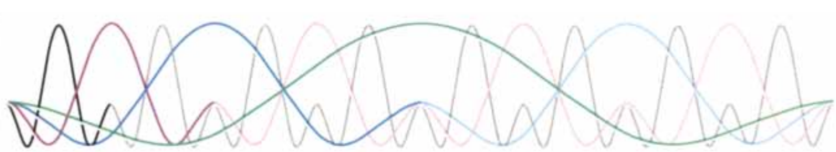
All told, Coifman comments, the delay in transferring wavelet expertise from mathematical theory to application is minimal. Carl Friedrich Gauss invented the fast Fourier transform in the late 18th century, and it lay unused until a pair of engineers at Bell Labs reinvented it in the early 1960s. Applied mathematics and physics are still almost entirely based on the pure mathematics of the 19th century, he asserts. Wavelets may be the beginning of developments based on the math of the past 50 or 60 years. *—Paul Wallich*

HOW WAVELETS ANALYZE INFORMATION



Signals such as speech and video display structure on many different scales. Small-scale structure in one area of the signal is often unrelated to structure in a different area.

A Fourier transform breaks a signal into continuously repeating components at various scales, and so it is unsuitable for representing signals that change suddenly over time.



Wavelets isolate the location as well as the scale of features in a signal. As a result, they can encode rapidly changing signals in a compact form.

Leaky Channels

Calcium influx may cause muscular dystrophy

Duchenne's muscular dystrophy (DMD), the most common of the inherited muscle disorders, is still untreatable and lethal. Two research groups in the San Francisco area are nonetheless hoping that their recent discoveries have improved the odds of one day treating the illness. They independently claim to have found evidence that DMD is the result of improperly guarded "doors" in muscle membranes that allow calcium ions to enter inappropriately. If these calcium channels are central to the cause of DMD, they will be good targets for drug therapies.

According to the Muscular Dystrophy Association, one out of every 3,400 boys is born with DMD. (The disease is caused by a genetic defect on the X chromosome, of which males have only one copy.) Sometime after their first year, boys with DMD begin to feel weakness in their limbs—a condition that gradually worsens and spreads until they are disabled. DMD patients rarely live past their mid-twenties.

Although the gene at fault in DMD was pinpointed in 1986, details of how the illness arises have remained mysterious. The damaged gene in DMD patients is supposed to produce minute quantities of dystrophin, a large protein associated with the membrane enclosing each muscle cell. Lack of dystrophin clearly causes DMD, but "nobody could figure out what on earth dystrophin was doing," remarks Richard A. Steinhardt, a DMD researcher at the University of California at Berkeley.

Steinhardt has found evidence that apparently ties dystrophin to the regulation of the flow of calcium ions into muscle cells through specialized channels. These calcium ions, in turn, may activate destructive proteases, or protein-snipping enzymes, that harm the cells. For Steinhardt, the first piece of the puzzle fell into place in 1988, when he and his research team reported that in muscle cells lacking dystrophin the free calcium ion concentration was twice as high as in normal cells. They also showed that by lowering the calcium levels outside the cells, they could lower the internal calcium concentrations and slow the rate at which the cells' proteins break down.

During the past year Steinhardt and Jeffry B. Lansman of the University of California at San Francisco both reported the discovery of previously over-looked calcium channels in muscle cells. (These calcium channels are distinct from the well-understood voltage-regulated ones that flutter open when nerves stimulate muscles to contract.) Last April, Lansman described a class of channels in mouse cells that opens or closes to varying degrees depending on whether the muscle fiber is stretched. Then, in November, Steinhardt announced that his group had found calcium "leak" channels on mouse and human muscle cells. Both research teams showed that in cells lacking dystrophin, these calcium channels are open much more of the time.

Whether Steinhardt and Lansman are looking at two different sets of calcium channels or examining the same set in different ways is fuel for debate. Despite their disagreements, however, the scientists are united in their belief that it may be possible to treat DMD by blocking the unregulated channels.

"Obviously, it would be beautiful if you could just find that magic bullet that targets itself to the channel in the muscle," Lansman says. His group thinks it may have identified a class of drugs that will block the channels. He adds that agents known to block the voltage-regulated channels were tried on DMD patients several years ago, but they were not helpful. "And now we know there's a good reason for that," he explains, "because those aren't the right channels."

Gene therapy, in which a functioning copy of the dystrophin gene would be inserted into DMD-affected cells, is usually considered the best hope for a true cure. Steinhardt maintains that the new calcium channel discoveries offer a means of testing potential gene therapies in cell cultures. "You're not just looking to see whether dystrophin is present. You can check whether the channel activity is regulated," he explains. He and his co-workers are now collaborating with other laboratories on those types of genetic experiments.

Yet many researchers are not persuaded that Lansman and Steinhardt have put their finger on a crucial, specific defect caused by a lack of dystrophin. Instead they suspect that abnormalities in calcium levels may be only one result of a generalized instability in the outer membranes of the diseased cells. "Almost the entire guts of these cells are leaking out, and all sorts of ions and huge enzymes are passing in and out," comments Eric P. Hoffman of the University of Pittsburgh Medical School, one member of the team that identified dystrophin. "I don't think you can explain all that with these channels." —*John Rennie*

PROFILE: AIDS DISPUTE

Robert Gallo toughs out controversy

Robert C. Gallo, chief of the laboratory of tumor cell biology at the National Cancer Institute, is angry. Preeminent in the field of human retrovirology, he has published some 900 papers and received every major award in biomedicine save the Nobel Prize. But the public knows him primarily because of the rancor surrounding the discovery of the cause of AIDS. "I've been through the mill," Gallo says.

The controversy dates back to September 1983, when French virologist Luc Montagnier of the Pasteur Institute in Paris sent Gallo a sample of a French isolate of the AIDS virus, called LAV. Seven months later Gallo held a press conference at which he announced he had identified a virus, which he called HTLV-III, that causes AIDS. Gallo also said he had developed a blood test based on it. In 1985 France sued the U.S. for breach of contract, implying that Gallo had used the French virus sample to create the test.

The two governments settled their dispute in 1987, agreeing that Gallo and Montagnier should share equal credit. The two scientists then published an official history of their work on AIDS in *Nature* and wrote an article in these pages [see "AIDS in 1988," by Robert C. Gallo and Luc Montagnier; SCIENTIFIC AMERICAN, October 1988].

There things rested until the *Chicago Tribune* published an article by correspondent John Crewdson in November 1989 once again asserting Gallo had unfairly claimed credit. That article prompted Congressman John D. Dingell to ask the National Institutes of Health to examine the affair. Last February the NIH initiated an inquiry into the role of Gallo's laboratory in isolating AIDS virus samples, growing the virus in cell cultures and proving it to be the principal cause of the disease.

After the panel had taken hours of testimony and examined Gallo's laboratory records, William F. Raub, acting director of the NIH, announced last October that the group had determined Gallo had "a substantial number" of detections and isolations of the AIDS virus "from several different sources" during the crucial period when he and his co-workers were isolating and growing HTLV-IIIB. The finding, Raub says, "does remove the principal motive" for Gallo or one of his colleagues to have stolen the French virus.

At the same time, however, Raub said that in one key research report by Mikulas Popovic and others in Gallo's group there was "an apparently significant difference between the work as described and the work that was done." Raub launched a formal investigation to focus on missing data and some questionable statements in the paper.

No one likes to wait for a jury. Gallo seems to find the experience particularly uncomfortable. Wounded, he has to have his say, and he has it. "I'm confident nobody in this laboratory committed any wrongdoing," Gallo declares. "Obviously, Mika [Popovic] is not a good notekeeper, but he made a great contribution: he discovered how you can grow the virus. I don't know how anybody could interpret wrongdoing in these things. I don't know what the purpose of such wrongdoing would be."

Even though both the NCI and his own lawyer try to keep him out of the public eye, Gallo's drive to gain vindication thwarts them. Recently Gallo got on the phone to lambaste author Bruce Nussbaum, who was being interviewed on a Washington, D.C., radio talk show about his recent book on AIDS, *Good Intentions*. "The people who risked their lives on a daily basis to make the openings that proved the cause of the disease, and developed the blood test allowing the epidemic to be studied, and saved tens of thousands of lives, in addition to developing the first systems to study drugs for it, are the people you're slandering," he told Nussbaum.

Gallo's voluble and volatile personality is a fact of life in biological science. He raises hackles, his colleagues say, with his penchant for sarcasm and his sometimes derisive attitude toward the work of other scientists, including Montagnier's group. But by common consent, Gallo is a creative thinker who makes strenuous efforts to dominate intellectually. "I think his reputation is very mixed," says Nobel laureate virologist Howard Temin of the University of Wisconsin. "He has made many people dislike him, but I also think he has many supporters."

Gallo was born in Waterbury, Conn., in 1937. Science did not attract him, he says, although he was used to seeing his father, a self-made metallurgist, reading technical journals. Then, when Gallo was 12, his younger sister, Judith, fell ill with leukemia. Before she died a year later, Gallo's father had taken her to Children's Hospital in Boston, where she was one of the first patients to receive chemotherapy.

Gallo admits that the tragedy, which devastated his father, may have influenced his career path. Father and son eventually became close friends of Marcus Cox, the pathologist who diagnosed Judith's disease. Gallo describes Cox as a gleeful skeptic and cites him as a powerful early influence. "I still didn't want to be a physician, mainly because my father wanted me to be one," Gallo says. "If anybody wants me to do something, there's an antibody reaction. But he never pushed, he'd back off. He wasn't like me—he was diplomatic."

As a youth, Gallo was by his account more interested in sports and girls, but by the time he entered Providence College science had captured his imagination. His journey toward success—and notoriety—began with an ambitious and disastrous attempt to discover the action of the murine thymus gland: all 200 of his mouse subjects died, and Gallo suffered multiple bites. Later, at Jefferson Medical College in Philadelphia, he skipped gross anatomy and histology classes. "They had named everything after somebody who had already done it," he explains. Instead he spent much of his time in the laboratory of the late Allan J. Erslev, who was doing research on blood cells. "I liked thinking about biochemical reactions that were occurring in the body."

After an internship at the University of Chicago, Gallo joined the NIH in 1965 for his military service. The combination of teaching, patient care and research was "the noblest thing you could do," he says. He worked at first on a childhood leukemia ward—"the last place I wanted to be." He soon accepted a posting to a white blood cell laboratory and devoted himself to leukemia research.

Gallo remembers the NIH as "a paradise" in those days. He and his colleagues discovered the *T* cell growth factor interleukin-2 in 1976 and the first human retrovirus, HTLV-I, in 1979. Those discoveries alone would probably have earned him a Nobel Prize, some researchers feel, if Gallo had not subsequently become embroiled in the dispute over the AIDS virus. He started to get interested in AIDS in 1982 and proposed that the disease was caused by a retroviral variant of HTLV-I. The idea "stimulated the right search," he asserts. "It turned out to be part right and part wrong, but that's the way science goes."

In May and June of 1983 Gallo says he "greatly intensified efforts" on AIDS, redirecting staff and coordinating an assault on the disease. "To my knowledge, we were the first laboratory in the world to regularly bring AIDS samples to culture," Gallo says. The annual meetings hosted by Gallo's laboratory became a mecca for AIDS researchers from all over the world. "I feel my role is to criticize, to hypothesize, to stimulate and to try to bend the direction of the laboratory. I think—I hope—that the right thing the leader of a group should try to do is intellectually dominate to a degree."

Gallo denies that the clamor has caused lasting damage. "I don't think it's destroyed the morale of the laboratory." But he concedes that it has done some harm. "It's hurt badly five or eight people, in terms of time and effort and morale." In addition, he says, "I lost a lot of creative thought time and peace." Press coverage of the dispute most dismays Gallo: "The worst part is watching and doing nothing about things that are printed that are lies or half-truths or misrepresentations. I have read repeatedly in some of the continental European press that this was for money."

AIDS researcher Gallo has "been through the mill."

They couldn't be more wrong, Gallo insists. When he developed the blood test, he says officials of the Department of Health and Human Services told him he "had to have a patent to protect against fraud, or else anybody could claim to have a blood test." Although several of his previous discoveries have been turned into medical technologies that "make money for other people," Gallo maintains that "when I was told to patent I was never told that I could make a dime. In fact, we could only make $10,000, and I didn't know that."

President Ronald Reagan changed patent royalty rules for government employees in 1986. Gallo notes that he now earns the maximum of $100,000 a year from his several patents. "But I've read that we could get millions of dollars," he says, becoming visibly irate. "Those are lies."

Gallo says he has some theories about why he is the victim of so many accusations—and he does not hesitate to name a former government employee who he contends is orchestrating a conspiracy. As evidence, he cites the rapid distribution of Congressman Dingell's letter requesting the inquiry. "The same day it came to the NIH, it was in the hands of my collaborators all over the place," he states. So, too, he says, were copies of Crewdson's article. "The conspiracy has been to block us from being able to work with people," Gallo says.

Crewdson gave an account of how Gallo wrote the abstract of a paper published by Montagnier's group so as to reflect Gallo's own thinking about the AIDS virus. "What is left out is that I was asked to; they forgot to write the abstract. What is left out is that they saw the galleys and approved them," he snaps. "They made a mistake, they had an immunologic reaction with HTLV-I in their paper. Everybody forgets that—why do they forget?"

Finding out just who forgot what is the object of the current NIH investigation. The NIH is also using new techniques to trace the origins of the virus Gallo used to develop the U.S. blood test. Gallo has acknowledged that his virus may have arisen from accidental contamination from LAV, and so it is unclear whether the effort will prove anything about culpability.

Nevertheless, Gallo predicts his name will be cleared early this year. He says that his lab is still turning out important discoveries in its continuing work on AIDS and other human disease viruses. "Somehow or other things are going to come out," he promises. Whether they will be to the satisfaction of the Nobel selection committee remains to be seen. —*Tim Beardsley*

SCIENTIFIC AMERICAN January 1991 Volume 264 Number 1

The Future of Space Reconnaissance

As the superpowers continue to launch spy satellites, many nations are planning to orbit their own. Such extensive proliferation will complicate international politics into the next century

by Jeffrey T. Richelson

During the past three decades the U.S. and the Soviet Union have exploited successive advances in launch capability, materials science and electronics to develop fleets of reconnaissance satellites. At present, some 20 of these spacecraft swarm in low or geosynchronous orbits around the earth. One might reasonably expect that the waning tension between the superpowers, the reunification of Germany and the crumbling Warsaw Pact would presage a thinning out. Yet there are indications that the number of intelligence satellites in orbit will proliferate—perhaps dramatically so—in the next decade.

Part of the explanation lies in the fact that despite the warming of U.S.-Soviet relations, the U.S. will still need to conduct extensive satellite reconnaissance, as will the Soviets. Besides monitoring advances in military technology and compliance with arms-control treaties, satellites have additional targets to examine. As demonstrated by recent events in the Persian Gulf, regional hot spots present constant threats. Targeting weaponry or listening in on an enemy's military communications from space is feasible for any nation operating a spy satellite. But, at the same time, satellites will also enable nations to gauge threats accurately and thus possibly circumvent potential hostilities. In any event, a multitude of orbiting eyes and ears from various countries—hostile, friendly and neutral—will affect international affairs for some time to come.

Much of the surveillance technology other countries will use, however, will not match that of the U.S. Unclassified documents, military experts and former intelligence officials reveal that U.S. satellite reconnaissance, having been an established and accepted component of intelligence operations for more than 30 years, has now reached a pinnacle of high technology. Indeed, analysts think the U.S. may budget as much as $5 billion on space reconnaissance each year; the Department of Defense has already spent an estimated $100 billion since 1960, when the U.S. began launching its photoreconnaissance satellites.

These early satellites produced photographic images. After photographing the target, the satellite ejected the capsule containing the film. Equipped with a parachute, the capsule descended until it reached the earth or until Air Force transport aircraft such as the C-130 snatched them out of midair. Intelligence agencies then developed and analyzed the film.

Although film-return satellites provided excellent information on Soviet and Chinese strategic forces, such as military installations and missile silos, delays in recovery and processing hampered their usefulness in fast-breaking crises. For example, a film pack returned shortly before the August 20, 1968, invasion of Czechoslovakia by the Soviet Union showed no signs of the upcoming battle. A second film pack returned after the invasion that carried images taken just before the attack showed unmistakable signs of Soviet troops massing along the border. Few in the photoreconnaissance program forgot that experience.

The inability of the film-return satellites to provide immediate data led the U.S. to phase them out—the last operated in 1984—in favor of a more sophisticated kind. These satellites, which the U.S. began launching in 1976, are designated KH-11. They use charge-coupled devices. Such devices transform the varying light levels in a scene into digital signals, which are transmitted to a receiving station at Fort Belvoir, Va., via a relay spacecraft. Fort Belvoir then sends the real-time imagery data to the Central Intelligence Agency's National Photographic Interpretation Center in Washington, D.C., as well as to other U.S. intelligence agencies for analysis.

The KH-11's electro-optical capability has had a dramatic impact on intelligence. The speed with which the digital signals are returned to the earth allows

JEFFREY T. RICHELSON is currently a consultant to the National Security Archive in Washington, D.C. He received his Ph.D. in political science from the University of Rochester in 1975 and has taught at the University of Texas at Austin and at American University. Richelson has authored six books on intelligence and espionage, including *America's Secret Eyes in Space: The U.S. Keyhole Spy Satellite Program*. The views expressed here do not necessarily reflect those of the Archive.

RADAR IMAGE of San Diego Bay, taken from an aircraft, simulates the ability of Lacrosse spy satellites to form images that resemble visible-light photographs.

a picture to be in the hands of an interpreter within minutes of a KH-11's pass over a target. The president or secretary of defense no longer has to wait days or weeks to obtain photographic evidence of events. Instead the image can be obtained, analyzed and set on a desk within an hour.

In addition, the KH-11 made it possible for the U.S. to increase dramatically the number of targets. Twenty years ago, when the U.S. used only film-return satellites, the supply of film limited the target base to about 20,000 (about 80 percent of which lay in the Soviet bloc and China). Charge-coupled device technology and rapid signal processing enable the KH-11 to behave like a television camera. The U.S. can now monitor 42,000 targets (fully half of which lie outside the Eastern bloc countries and China).

Currently the U.S. operates two KH-11s, whose orbits range from 150 to 250 miles above the earth. The satellites have expected lifetimes of three to four years. One will probably cease operating soon (it was launched in October 1987). The other is expected to last until late this year. Many experts believe that, under ideal conditions, these two satellites can probably resolve objects six inches across; exact performance statistics remain classified.

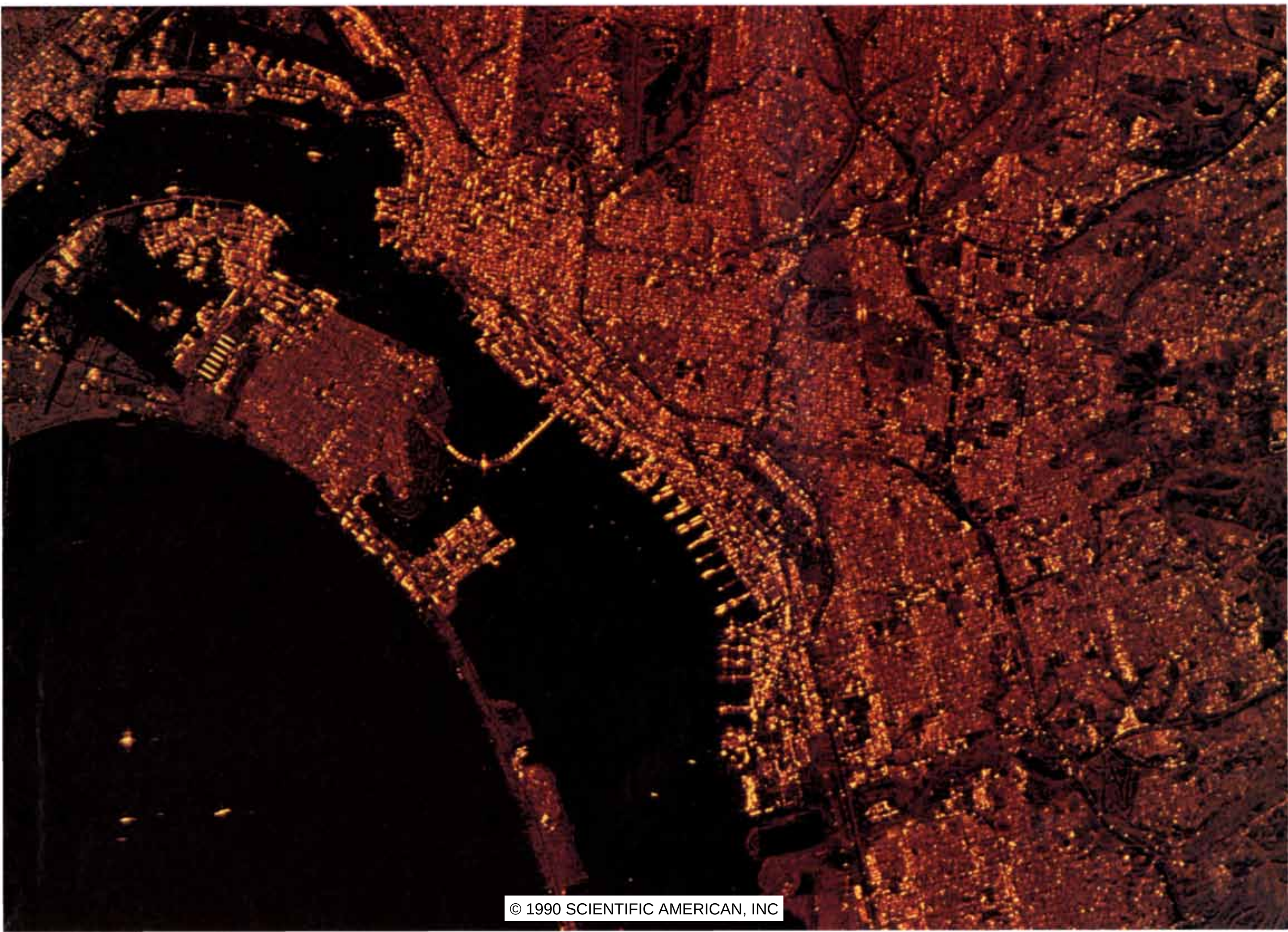
Since they began operating, the KH-11s have monitored the Soviet Union's construction of a nuclear-powered aircraft carrier, the Abalakova early-warning radar system and the testing of the Blackjack bomber. The satellites have also been used to identify landing and departure sites for the attempted hostage-rescue mission in Iran, to detect Libya's chemical warfare plant at Rabta and to examine nuclear weapon facilities across the globe.

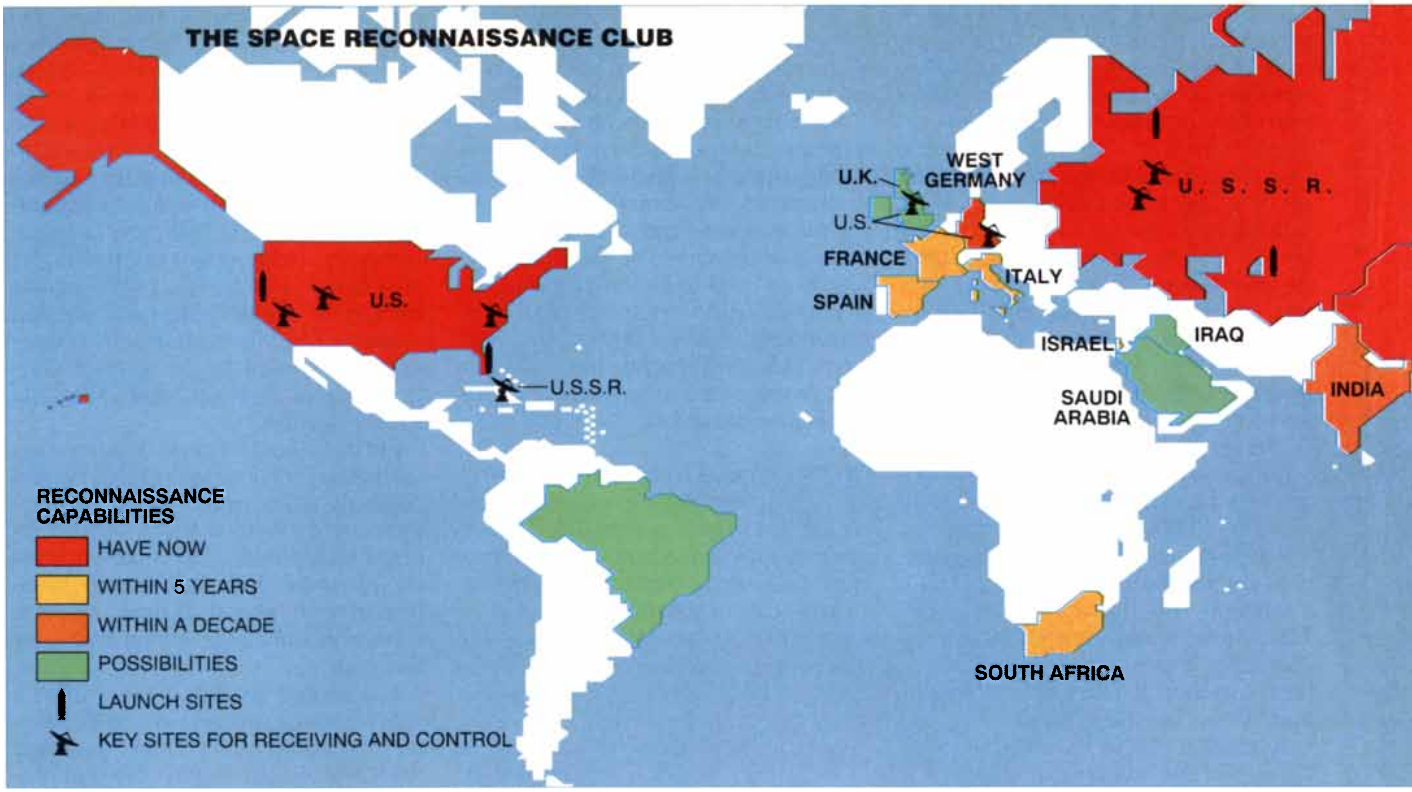
Despite its capabilities the KH-11 has two significant limitations. The optical system functions efficiently only in daylight. It cannot produce detailed photographs at night, nor can its sensors penetrate cloud cover—a particular problem with regard to the Soviet Union and Eastern Europe, where clouds blanket many regions as much as 70 percent of the time.

One solution was implemented in December 1988, when the space shuttle *Atlantis* deployed a satellite first called Indigo, later code-named Lacrosse. (The U.S. alters the designations to confuse the Soviets or to restore cover when a code name leaks to the public.) The Lacrosse operates in the same way *Magellan* does, the interplanetary probe now mapping the surface of Venus. Rather than producing images from existing light, the Lacrosse transmits radio waves to the target area below, which reflects the signals back to the satellite. The Lacrosse then converts the information into images and relays them in the form of electronic signals to White Sands, N.M., via a relay satellite.

Close reading of news reports suggests that such a radar-imaging system probably has a resolution of between three and 10 feet—as much as 10 times better than *Magellan*'s resolution. Present plans call for two Lacrosse satellites to be in orbit at all times. The U.S. expects to launch the second craft later this year.

The second solution to the KH-11's shortcomings was implemented when the space shuttle *Columbia* launched the world's most sophisticated photoreconnaissance satellite in August of





1989. Known as the Advanced KH-11 in intelligence circles, the craft has a larger fuel capacity, which enhances its ability to avoid antisatellite weapons and enables it to change orbits and thus examine new targets.

The Advanced KH-11, mistakenly referred to as KH-12 in early accounts, also reportedly has an infrared-imaging capability. This technology enables the satellite to provide some night photography. The device can also render images in false color to reveal hidden details. For example, cut vegetation used for camouflage differs in false color from healthy, uncut flora. The system's resolution, however, is probably similar to or just slightly better than that of the previous KH-11.

In February 1990 the space shuttle *Atlantis* orbited a second Advanced KH-11. The satellite, augmented with antennas that can intercept communications and electronic signals, apparently malfunctioned. A Titan IV rocket reportedly launched a replacement last June.

Currently the U.S. uses two KH-11s, two Advanced KH-11s and one Lacrosse. According to a former CIA official, future plans call for at least three Advanced KH-11s and two Lacrosse satellites to make up the imaging portion of the U.S. space reconnaissance program. Complementing this surveillance capability will be signals intelligence satellites, designed to intercept a wide variety of foreign signals, including military communications and telemetry data.

The U.S. began its signals intelligence satellite program in 1962 when it launched into low earth orbit the first "ferret" satellites. Reportedly, the exclusive mission of these satellites was the interception of radar signals from the Soviet Union, China and other potentially hostile countries. Among other advantages, the interception enabled the U.S. to locate the radar stations and target them for destruction or disruption (by jamming) during war.

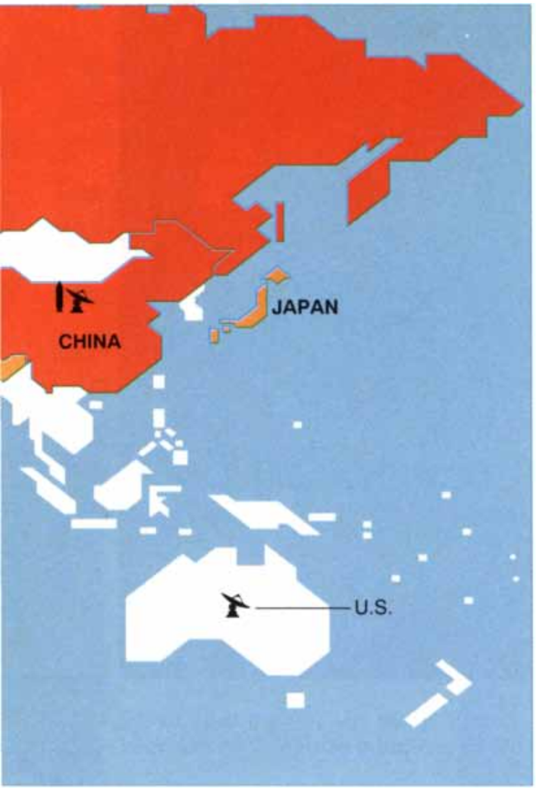
Of far greater utility to peacetime reconnaissance are signals intelligence satellites that occupy geosynchronous orbits. At 22,300 miles above any point over the Equator, the speed of the satellite orbiting around the earth matches the speed of the earth's rotation. The satellite effectively hovers above a single point.

From such a vantage, the satellite's listening antennas can intercept signals from well over a third of the earth's surface and monitor a given set of frequencies or communications links continuously. Through such constant vigilance, geosynchronous satellites can witness brief, unexpected events such as missile tests, which last only 20 or 30 minutes.

In 1970 the U.S. launched its first signals intelligence satellite, code-named Rhyolite (subsequently renamed Aquacade), that operated in geosynchronous orbit. Since then the U.S. has launched four more Rhyolites, four or five Vortexes (another kind of signals intelligence satellite, previously code-named Chalet) and two Magnums (advanced versions of the Rhyolite).

Ground stations at Pine Gap in Australia, Menwith Hill in Britain and Bad Aibling in Germany have enabled the U.S. geosynchronous satellites to monitor the signals from many nations. The information includes walkie-talkie traffic from military exercises, telemetry signals from missile tests and military, government and economic communications. Currently one or two Vortexes and two Magnums monitor the globe at all times.

The combination of photoreconnaissance and signals intelligence satellites provides the U.S. with formidable surveillance capabilities. During the cold war, the satellites have helped the U.S. determine the size of Soviet weaponry, to monitor treaty compliance and to detect swift military moves. More recently news reports indicate that U.S. photoreconnaissance satellites have been monitoring Iraq's troop strength in the Persian Gulf and have shown the movement of military hardware, such as chemical munitions and mobile missile launchers. The sig-



nals intelligence satellites have been monitoring Iraq's air defense systems and military communications. In fact, newspaper accounts stated that satellite photography helped to convince King Fahd of Saudi Arabia to allow U.S. troops onto Saudi soil.

The U.S. has also used satellites to monitor nonmilitary operations, such as disaster relief. During the Chernobyl incident the Vortex satellite monitoring the western Soviet Union intercepted the two-way radio and telephone communications of the military, government and security forces within several hundred miles of the incident. Combining the intercepted communications with photography obtained by a KH-11 of the damaged reactor and clean-up operations, the U.S. made detailed assessments of the catastrophe as it was unfolding.

The U.S. consequently determined the extent of the damage, including the burning of the graphite reactor, well before the Soviets admitted the severity of the accident. The data also reassured the U.S. that only one reactor melted down, not two as some early reports indicated.

Reconnaissance satellites have also monitored domestic upheaval in other nations, including secessionist activity and ethnic conflict in the Soviet Union. They recorded events in Beijing during the days and weeks surrounding the Tianamen Square demonstrations and thus circumvented the news blackout imposed by China's leaders.

Since the beginning of satellite surveillance the Soviet Union has trailed the U.S. in space reconnaissance capabilities. In recent years, however, the Soviets have made significant advances that have narrowed the technology gap between the two countries. Until late 1982, the Soviet Union relied solely on film-return spacecraft for their photoreconnaissance activities. These included satellites that provide a wide field of view as well as high-resolution satellites for close-look reconnaissance.

Since December 1982 the Soviets have been launching a limited number of satellites that could provide real-time data. Like the KH-11s, these satellites, part of the Cosmos series, transmit their digital signals to the earth via a relay satellite in geosynchronous orbit. The Soviets launched the latest versions in 1989: the *Cosmos 2007* in March and the *Cosmos 2049* in November. The main receiving station may be in Vatutinki, 50 kilometers southwest of Moscow, at a site belonging to the Soviet General Staff's chief intelligence directorate (GRU).

Experts in Soviet satellite operations believe the GRU has used the satellites to monitor military conflicts throughout the world, including the Indo-Pakistani War, the Yom Kippur War and the U.S. invasion of Grenada. In addition, based on its orbital changes, a third-generation, high-resolution spacecraft called *Cosmos 1343* apparently monitored the White Sands, N.M., shuttle landing site just before the touchdown of the third shuttle mission in 1982. The Soviets have offered for sale to the general public some medium-resolution photographs, including those of the U.S. army base at Fort Riley in Kansas, the Strategic Air Command in Omaha and a U.S. intercontinental ballistic missile field.

The Soviets usually maintain three photoreconnaissance satellites (primarily film-return types) in orbit during most of the year. During the summer months, two or three more craft may join the fleet to enable the Soviets to take advantage of the long summer light. Soviet satellite engineers, who prefer many cheap satellites to a few, highly sophisticated ones, designed their film-return satellites to last from 14 to 44 days. Their real-time imaging satellites have much longer lifetimes (more than six months).

Thus, to maintain year-round coverage, the GRU must launch a greater number of satellites than does the U.S. In 1988 the GRU launched 32 photoreconnaissance spacecraft (only one of which provided real-time images); in 1989 it launched 31 satellites.

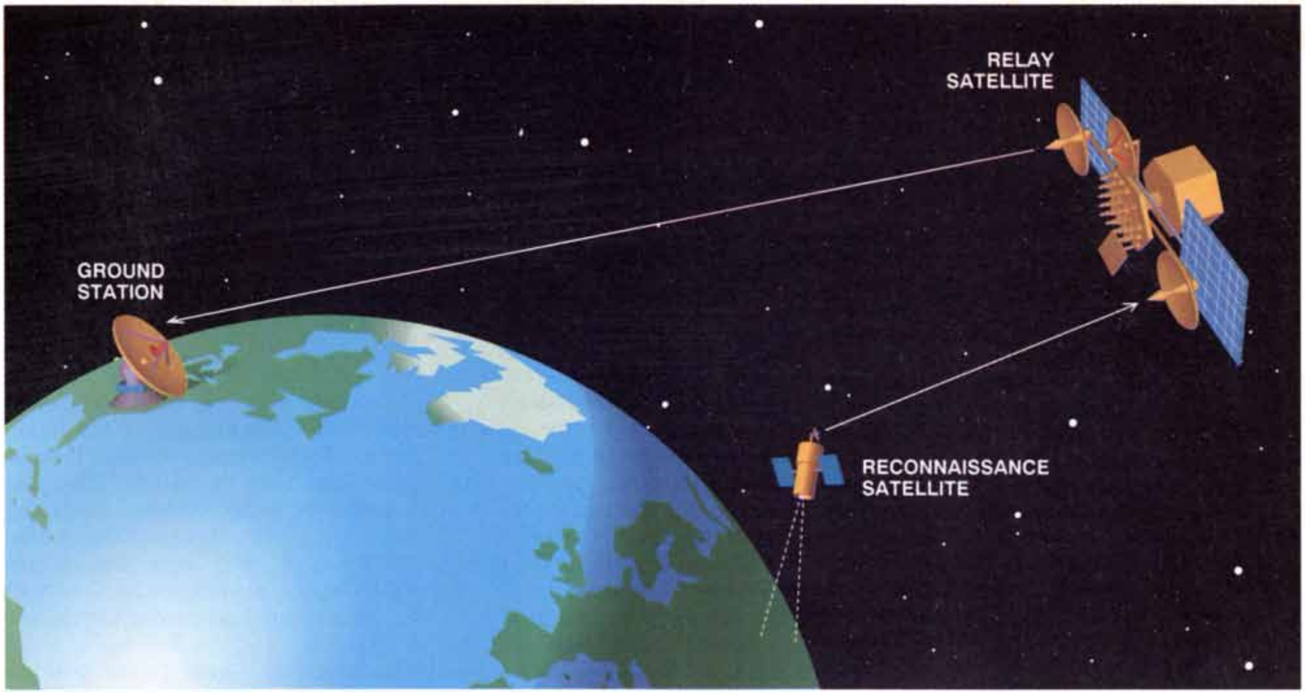
Although the Soviet Union has almost matched the photoreconnaissance technology of the U.S., a dramatic gap exists in signals intelligence satellite capability. In 1967, five years after the U.S. began launching its ferret satellites, the Soviets began placing their signals intelligence satellites in low earth orbit, at altitudes of approximately 400 to 500 miles. In such low orbits, the satellites remained in range over a given target for a brief period and thus could play only a limited role in communications and missile-test monitoring. Their primary mission instead must have been the identification and location of radar systems, including those covered by the Antiballistic Missile Treaty of 1972. The Soviets are currently operating one constellation of six signals intelligence satellites in low earth orbit.

In 1985 the Soviet Union sent up *Cosmos 1738,* its first geosynchronous-orbiting signals intelligence satellite, fully 15 years after the U.S. launched its own. The U.S. found that the Soviet satellite, along with its two successors, *Cosmos 1961* and *Cosmos 2054*, orbit over the Western Hemisphere at 14 degrees west longitude. Apparently, the Soviets' signals intelligence complex in Lourdes, Cuba, serves as the ground control station, which ultimately relays the signals to the GRU.

The data these satellites can obtain originate throughout most of the U.S. and South America. The satellites probably pick up telemetry data from tests conducted in the Carribbean of the new Trident D-5 submarine-launched ballistic missile as well as military signals related to the Brazilian and Argentine ballistic missile and nuclear power programs.

Despite the changing world situation, both superpowers will continue to photograph each other's missile silos, bomber fields, submarine facilities and military test sites, bases and installations. The signals intelligence satellites will keep monitoring radar signals, missile-test telemetry and political, military and economic communications. Such ongoing reconnaissance should reveal major advances in military technology as well as any treaty violations that may elude on-site inspectors.

Both superpowers can be expected to devote increased attention to the growing problem of weapon proliferation. The more frequently known military facilities are monitored, the greater the chance is of detecting an interesting



RELAY SATELLITES, such as the Tracking and Data Relay Satellite System, act as the intermediaries between satellites and ground stations. Relays speed the flow of data to the earth and complicate the interception of signals by enemies.

event. Such persistence was rewarded when a KH-11 spotted Argentina's Condor II missile on its launch pad, awaiting testing.

Satellite sensors can also spend more time searching for new facilities. U.S. satellites discovered, for instance, the chemical weapon plant near Rabta, Libya. Frequent reconnoitering will help detect such installations before they become operational.

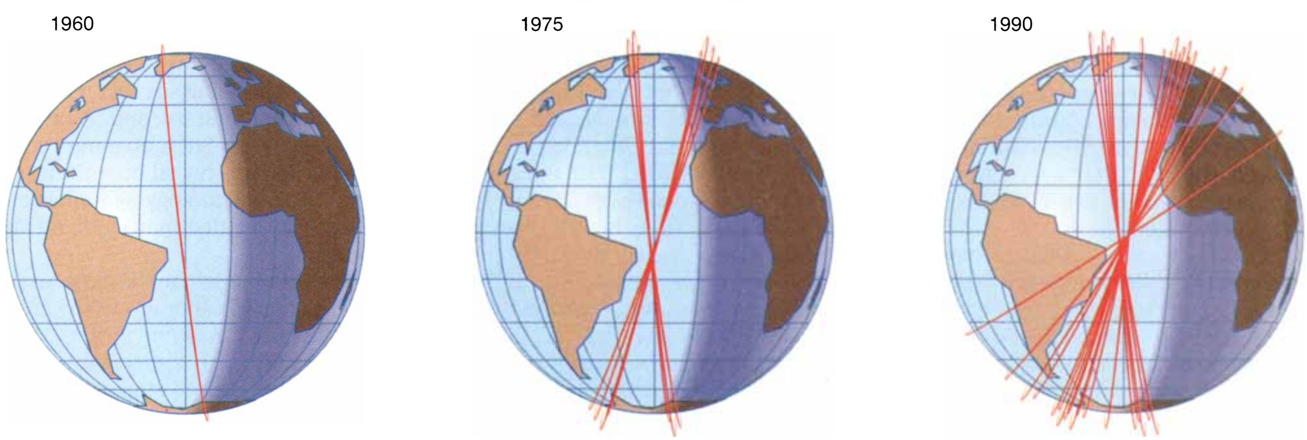
As tensions have ebbed, the U.S. and Soviet Union have begun sharing some of the information produced by their reconnaissance systems. For example, in 1989 the U.S. informed the Soviets that KH-11 photography had turned up evidence of a new nuclear reactor outside Pyongyang, North Korea, and even specified the exact geographic coordinates of the plant, enabling Soviet satellites to photograph it.

Such national assets can serve another facet of international security: monitoring the continuing degradation of the earth's environment. The U.S. and Soviet satellites will be able to devote an increasing amount of time to producing images of oil slicks, sources of pollution and damage from natural disasters such as earthquakes.

Although the U.S. and Soviet Union have been the dominant players in the space reconnaissance arena, they have not been alone. Since 1970 the People's Republic of China has been a participant. Progress there has been uneven. China has yet to orbit a signals intelligence satellite. Its photoreconnaissance systems have made slow but steady gains. China began producing images from space on April 24, 1970, when it placed a 354-pound imaging satellite in low earth orbit. By the mid-1970s China could launch photoreconnaissance satellites weighing almost 10,000 pounds. Although this payload equals about a third of the weight of a KH-11, it gives engineers considerable opportunity to construct a fairly large and sophisticated optical system.

Between 1970 and December 31, 1989, China orbited 12 satellites associated with military photoreconnaissance. It usually launched a satellite no

PROLIFERATION OF SATELLITE RECONNAISSANCE



more than once a year and only during the late summer or early winter. From their orbits the satellites covered the entire world except for the most northerly parts of the Soviet Union, Canada and Scandinavia.

The satellites primarily returned the film to the earth for development and analysis. On one occasion, instruments in the satellite developed the film on board. A scanner read the images and then transmitted them to the earth via radio signals. U.S. intelligence agencies believe China is currently developing a more advanced system, which will apparently be similar to France's SPOT commercial observation satellite. Scheduled for launch in the early 1990s, the satellite will use charge-coupled devices to form images.

The Soviet Union, particularly its border with China, has been and will continue to be a high-priority target of the latter country's satellites. Vietnam and India have been adversaries in the past. As potential future enemies, they will undoubtedly remain prime targets. China also will probably survey Cambodia and Afghanistan, two countries whose antigovernment guerrilla groups it has supported.

MILESTONE LAUNCHES IN SPACE RECONNAISSANCE HISTORY

COUNTRY	YEAR	NAME	DESCRIPTION
U.S.	1960	*Corona/ Discoverer 14*	First operational photoreconnaissance satellite. Its film capsule was successfully recovered in midair.
U.S.S.R.	1962	*Cosmos 4*	First Soviet photoreconnaissance satellite. The Soviet Union began launching Cosmos satellites with near clocklike regularity in 1965 and currently operates three to six photoreconnaissance satellites at all times.
U.S.	1962	"Ferrets"	Signals intelligence satellites in low earth orbit to monitor foreign radar systems.
U.S.S.R.	1967	*Cosmos 148*	Signals intelligence satellite similar to U.S. ferrets.
U.S.	1970	Rhyolite	First geosynchronous-orbiting satellite for signals intelligence. From such an orbit, a satellite can monitor one third of the earth's surface.
CHINA	1970	*China 1*	First Chinese photoreconnaissance satellite.
U.S.	1976	KH-11	Satellite with charge-coupled device to form real-time images of objects as small as six inches on any dimension.
U.S.S.R.	1982	*Cosmos 1426*	Photoreconnaissance satellite similar to KH-11 but with lower resolution.
U.S.S.R.	1985	*Cosmos 1738*	Geosynchronous-orbiting satellite for signals intelligence.
U.S.	1988	Lacrosse	Radar-imaging satellite capable of resolving objects three to 10 feet across. It currently operates under a new code name.
ISRAEL	1988	*Offeq 1*	Experimental satellite built with the aid of South Africa.
FRANCE ITALY SPAIN	1993	Helios	System of four photoreconnaissance satellites to be launched over a period of 10 years.

Smaller powers seem inclined here as in other technological areas to follow the trails that the superpowers have blazed. Practical considerations as much as hubris supply motivation. Even nations aligned with one of the superpowers often do not wish to rely on a patron for satellite data. For instance, Israel's entrance into the satellite reconnaissance field stemmed from the dissatisfaction Israeli officials felt with being dependent on U.S. satellite data. Israel's cabinet minister and former Chief of Staff Mordechai Gur complained that the U.S. withheld satellite data immediately before the 1973 Yom Kippur War. Meir Amit, the former chief of the Mossad, the Israeli secret service, said his country was receiving the "crumbs" of satellite intelligence. Satellite reconnaissance would also help an aligned nation make more independent assessments of foreign affairs as well as provide it with a measure of prestige in the elite world of high technology.

Commercial satellites would not completely fulfill intelligence requirements, either. Besides the comparatively low resolution (ranging from 30 to 100 feet) of such satellites, the client could not exercise significant control over the operation. A client would also have to compete with other customers for use and could not directly prevent disclosure of areas of interest.

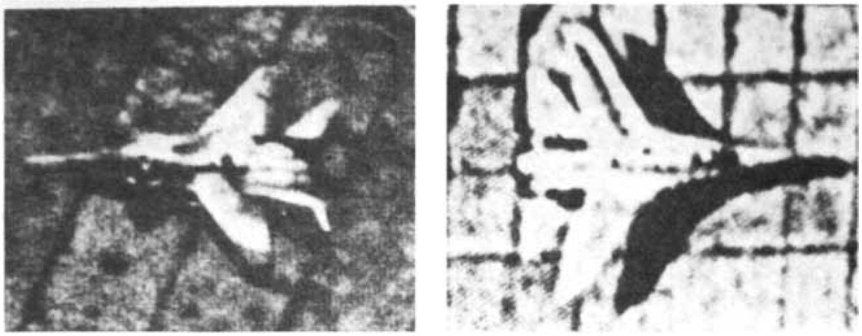
High-flying aircraft do not make a good alternative to satellites. Overflights tend to be a politically sensitive issue. They can also be decidedly dangerous. Many countries now have sophisticated tracking devices and effective ground-to-air weapons. Aircraft suffer technological deficiencies as well. They cover only limited areas, and fuel supply renders many target nations distant enough to make aerial reconnaissance risky. Israel's air force, for example, would have difficulty monitoring its enemies' movements deep in the Middle East.

Consequently, several countries have decided that the benefits outweigh the costs, which run into hundreds of millions of dollars for a single satellite. Many governments are on the verge of joining the space reconnaissance club or are actively considering it. France currently operates the SPOT system, which provides color images that have 66-foot resolution and black-and-white images of a 33-foot resolution. Although the system is designed for commercial purposes (geology, urban planning, agriculture), the French military exploits its abilities for intelligence. France, however, will soon have its own dedicated reconnaissance system, called Helios, to gather intelligence.

The Helios system calls for the launching of four satellites over 10 years, beginning in 1993. These high-resolution satellites, developed with the participation of Spain and Italy, will be able to resolve objects smaller than a baseball bat, according to intelligence analysts.

Jacques Bosquet, head of guided-weapon and space programs for the French Ministry of Defense, stated that the different needs of the three countries will severely tax the capabilities of Helios. France will undoubtedly reconnoiter the now unified Germany and parts of the Middle East and Africa, such as Chad, where France has had traditional interests. The system should also provide accurate targeting information for France's nuclear strategic missiles. In addition, Helios will independently verify arms-control treaties by the Soviet Union and the U.S. Italy, wary of terrorist attacks similar to the *Achille Lauro* incident, will probably monitor areas in the Middle East. For what purpose Spain will use Helios remains unclear.

Threatened by many of its Middle East neighbors and unhappy with U.S.-supplied data, Israel began its space reconnaissance program in 1988 by launching *Offeq 1*. Intelligence sources believe Israel probably developed the experimental satellite in conjunction



SATELLITE IMAGES used in 1984 congressional hearings show a Soviet MiG-29 (*left*) and an SU-27 (*right*). They were mistakenly published in the proceedings.

with South Africa. In early March of 1990 Israel lofted *Offeq 2* into an orbit that ranges from 125 miles at perigee to 923 miles at apogee.

Although Israeli officials deny that the 374-pound *Offeq 2* carries any kind of optical system, it is clear to space and intelligence experts that the satellite is at least the predecessor of a photoreconnaissance satellite. Israel would target activities throughout the Middle East, including Libya's construction of a chemical warfare complex and Saudi Arabia's deployment of CSS-2 missiles.

India, which has an active space program, will be ready to use its Polar Satellite Launch Vehicle in the next few years. This delivery vehicle will be able to place a 7,000-pound spacecraft in low earth orbit. Such a launch capability will be adequate to orbit a photoreconnaissance satellite. Satellite images should prove quite useful to India, which has long-standing disputes with two of its border antagonists, Pakistan and China.

For instance, in 1987 India engaged in a large-scale military exercise, called Operation Brass Tacks, near the Pakistani border. Pakistan, unsure of India's motives, then massed its own troops, threatening to turn the exercise into a full-scale war. Satellite data would have provided an early warning and an opportunity for diplomats to resolve the situation. Similarly, when Indian and Chinese troops faced off in the Sumdorong Valley in Arunachal Pradesh in 1986, satellite data would have given India's leaders the exact size, location and kind of units that Beijing had deployed.

If India orbits photoreconnaissance satellites, targets will include Pakistan's military bases, suspected terrorist training camps and undoubtedly the nuclear facility in Kahuta. India will also use satellite reconnaissance to provide targeting information for its intermediate-range ballistic missile, Agni [see "Third World Ballistic Missiles," by Janne E. Nolan and Albert D. Wheelon; SCIENTIFIC AMERICAN, August, 1990].

In 1988 Britain was reportedly developing a signals intelligence satellite, code-named Zircon, for launch in geosynchronous orbit over the Indian Ocean. A satellite in such an orbit will give Britain an electronic view of the eastern half of the Soviet Union. The Zircon would also be close enough to Soviet communications satellites to intercept some of their traffic.

Early in 1990, West Germany announced studies to determine the desirability of a space-based earth-observing system. West German officials indicated that they were primarily concerned with the role the satellites would play in arms-control verification. That role will probably not change significantly with the unification of the two Germanys.

Japan has reportedly just started development of a photoreconnaissance satellite for defense and environmental purposes. The craft may go up within the next 10 years. Iraq, Brazil and Saudi Arabia, all of which have ballistic missile programs, may also have space reconnaissance systems ready sometime after the year 2000.

The potentially massive explosion in reconnaissance systems will force nations to adjust their foreign policy. Military operations, including hostage-rescue missions and other actions considered justifiable, will be harder to keep secret from prying eyes—including hostile ones. If two nations have nonnegotiable differences, then reconnaissance may help increase the efficiency with which war is conducted. The Israeli attacks on the nuclear reactor in Osirak, Iraq, and on the headquarters of the Palestine Liberation Organization in Tunis were planned with the help of satellite images obtained (according to news reports), legally and illegally, from the U.S.

But a proliferation of space reconnaissance systems may have positive consequences as well. Military secrecy and undetected arms-control violations will become more difficult to achieve. As the number of satellite eyes increases, the harder it will become to hide weapon systems, radar installations and military production facilities.

For instance, when only a few reconnaissance satellites orbited the earth, a country could conceal its secret aircraft or suspend training exercises once or twice a day. With many satellites, however, such practices become costly and inconvenient. The U.S. alone, with five pairs of eyes in space at all times, imposes a heavy cost on any nation that would seek to keep targets out of view.

Such extensive surveillance will make it more difficult for nations to prepare for war or mount a surprise attack. Assuming that countries truly prefer to settle their differences peacefully, satellite reconnaissance can provide reassurance that another nation is not preparing to attack. Satellite data can also reveal the adversary's military capabilities. Intelligence officials can then decide whether to deploy new offensive weapons.

The continued improvement of space reconnaissance systems by the U.S. and Soviet Union and the development of new systems by other nations will have a significant impact on international affairs. Some countries may use satellites to wage war, others to avoid it. No one can accurately predict all the changes that proliferation will bring, and it is impossible to say whether spy satellites will prevent military conflicts. But history seems to testify that satellite reconnaissance has helped moderate the arms race and keep the peace between the U.S. and Soviet Union for 30 years. Perhaps space reconnaissance will limit hostilities between other nations as well.

FURTHER READING

DEEP BLACK: SPACE ESPIONAGE AND NATIONAL SECURITY. William E. Burrows. Random House, 1986.

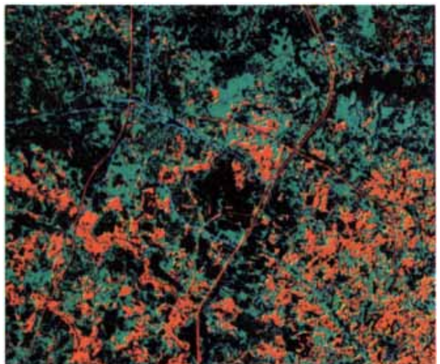
AMERICA'S SECRET EYES IN SPACE: THE U.S. KEYHOLE SPY SATELLITE PROGRAM. Jeffrey Richelson. HarperBusiness, HarperCollins, 1990.

COMMERCIAL OBSERVATION SATELLITES AND INTERNATIONAL SECURITY. Michael Krepon, Peter D. Zimmerman, Leonard S. Spector and Mary Umberger. St. Martin's Press, 1990.

THE SOVIET YEAR IN SPACE 1989. Nicholas L. Johnson. Colorado Springs, Teledyne-Brown Engineering, 1990.

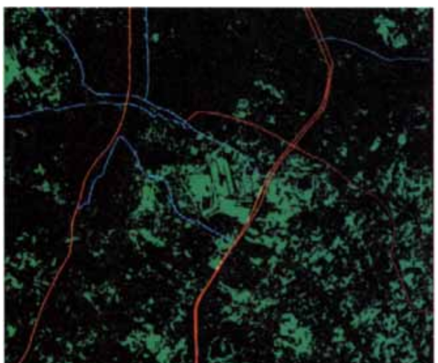
No two businesses see things alike.

Millersville, Maryland as it would appear to the human eye.

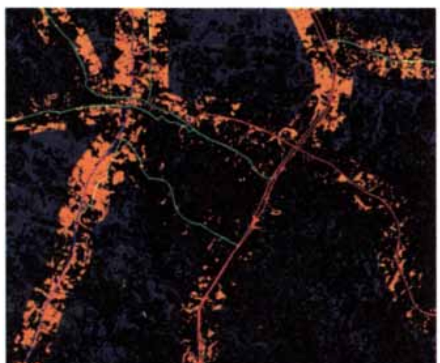


Coniferous trees are shown in orange. Deciduous in green.

We've laid down a buffer-zone that highlights in red all developed regions within 570 meters of a road.



All 5,188 acres of crops and similar vegetation are highlighted in green.



This buffer-zone highlights in yellow all wetlands within 570 meters of a road.

Here, virtually everything is identified. Trees in green. Water in blue. Grasslands in orange. Wetlands in brown. Suburban development in black.

Every business has its own way of seeing things.

And more and more businesses are seeing things through the eyes of a Landsat satellite.

Because unlike anything on earth, a Landsat satellite gives them a custom-made view of their world.

Often mistaken for a photograph, an image from a Landsat satellite is, in reality, a visual form of data captured by sophisticated on-board sensors.

And when these data are interrelated, they can isolate and reflect precisely the information a business needs.

For foresters, we help to identify harvestable trees that are close to haul roads.

We lead real estate developers to properties that contain a minimum of ecologically-sensitive wetlands.

We also point retailers to intersections that'll be busy not only tomorrow, but ten years down the road.

We can detect damage in seemingly healthy crops. We can find hidden geological faults. We can even see water where it appears to be arid.

Now, if you're wondering what all this costs, you'll be happy to know that we're quicker and far more cost effective than conventional aerial photography and surveying.

And since we've been in orbit for nearly twenty years, we've accumulated a vast archive that can be used to compare the past to the present.

If you'd like to know more, simply give us a call.

We'll introduce you to a consultant who can work with our data to give you precisely what you need.

You'll quickly understand why so many businesses have come to see things our way.

1-800-545-3570 U.S.
FAX 1-301-552-5476
TELEX 277685 LSAT UR

EOSAT

4300 FORBES BOULEVARD, LANHAM, MARYLAND 20706 · 1-301-552-0537

Self-Organized Criticality

Large interactive systems naturally evolve toward a critical state in which a minor event can lead to a catastrophe. Self-organized criticality may explain the dynamics of earthquakes, economic markets and ecosystems

by Per Bak and Kan Chen

When catastrophe strikes, analysts typically blame some rare set of circumstances or some combination of powerful mechanisms. When a tremendous earthquake shook San Francisco, geologists traced the cataclysm to an immense instability along the San Andreas fault. When the stock market crashed on Black Monday in 1987, economists pointed to the destabilizing effect of computer trading. When fossil records revealed the demise of the dinosaurs, paleontologists attributed the extinction to the impact of a meteorite or the eruption of a volcano. These theories may well be correct. But systems as large and as complicated as the earth's crust, the stock market and the ecosystem can break down not only under the force of a mighty blow but also at the drop of a pin. Large interactive systems perpetually organize themselves to a critical state in which a minor event starts a chain reaction that can lead to a catastrophe.

Traditionally, investigators have analyzed large interactive systems in the same way as they have small, orderly systems, mainly because the methods developed for simple systems have proved so successful. They believed they could predict the behavior of a large interactive system by studying its elements separately and by analyzing its microscopic mechanisms individually. For lack of a better theory, they assumed that the response of a large interactive system was proportional to the disturbance. It was believed that the dynamics of large interactive systems could be described in terms of an equilibrium state that is disturbed now and then by an external force.

During the past few decades, it has become increasingly apparent, however, that many chaotic and complicated systems do not yield to traditional analysis. In 1987 one of us (Bak)—in collaboration with Kurt A. Wiesenfeld, now at the Georgia Institute of Technology, and Chao Tang, now at the Institute for Theoretical Physics in Santa Barbara—developed a concept to explain the behavior of composite systems, those containing millions and millions of elements that interact over a short range. We proposed the theory of self-organized criticality: many composite systems naturally evolve to a critical state in which a minor event starts a chain reaction that can affect any number of elements in the system. Although composite systems produce more minor events than catastrophes, chain reactions of all sizes are an integral part of the dynamics. According to the theory, the mechanism that leads to minor events is the same one that leads to major events. Furthermore, composite systems never reach equilibrium but instead evolve from one metastable state to the next.

Self-organized criticality is a holistic theory: the global features, such as the relative number of large and small events, do not depend on the microscopic mechanisms. Consequently, global features of the system cannot be understood by analyzing the parts separately. To our knowledge, self-organized criticality is the only model or mathematical description that has led to a holistic theory for dynamic systems.

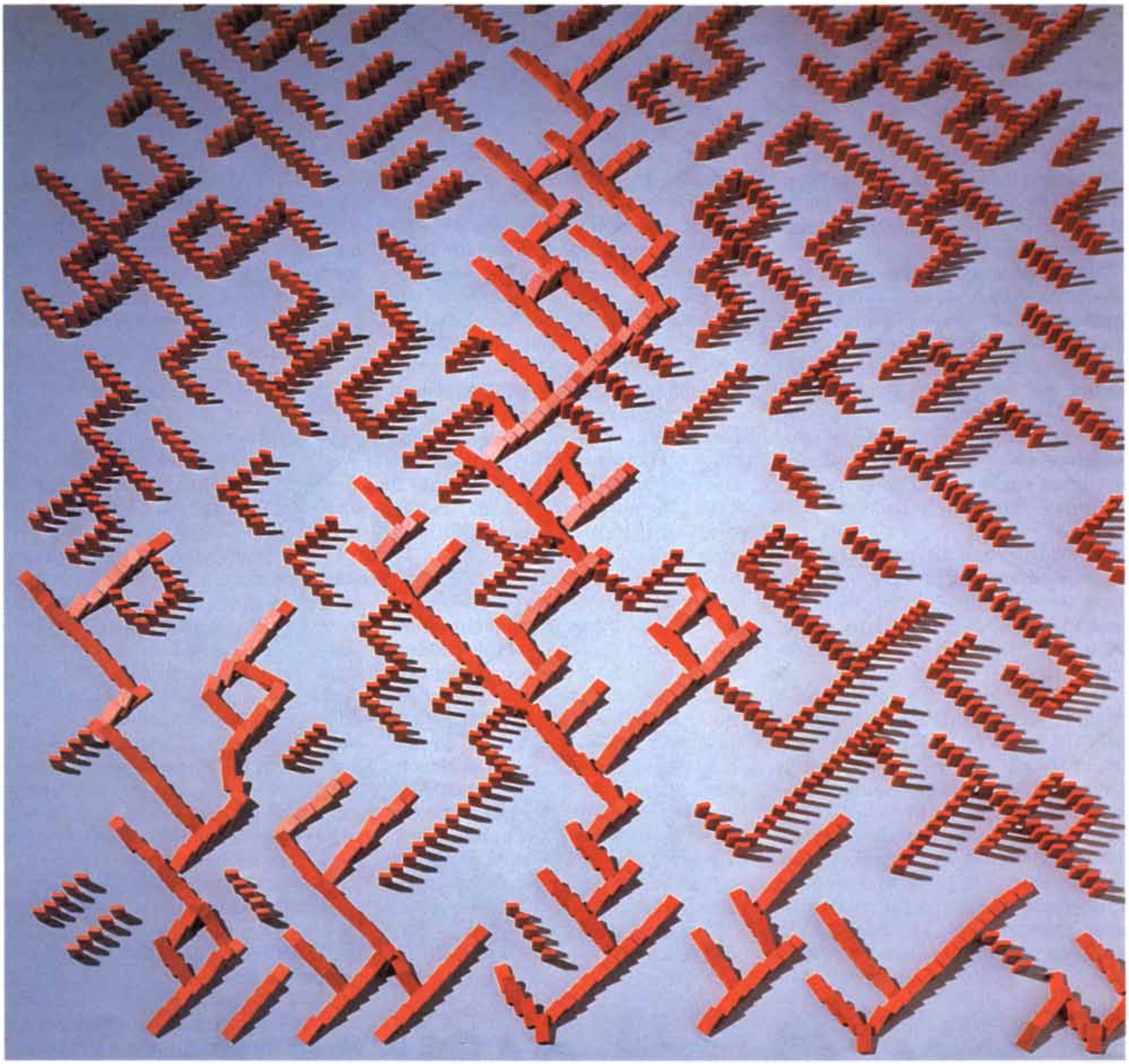
During the past four years, experiments and models have demonstrated that many composite systems at the heart of geology, economy, biology and meteorology show signs of self-organized criticality. These insights have improved our understanding of the behavior of the earth's crust, stock markets, ecosystems and many other composite systems.

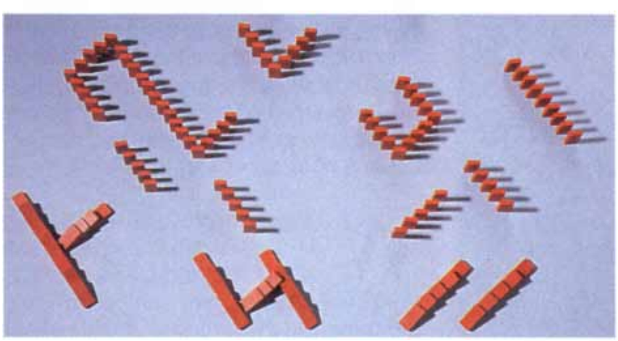
Because composite systems contain many components and are governed by many interactions, analysts cannot possibly construct mathematical models that are both totally realistic and theoretically manageable. They therefore have to resort to simple, idealistic models that capture the essential features of real systems. If the simple models are robust with respect to various modifications, they might be able to extrapolate the findings to real situations. (This approach has been very successful in equilibrium statistical mechanics, where universal phenomena in systems with many degrees of freedom can be understood from studies of simple models.)

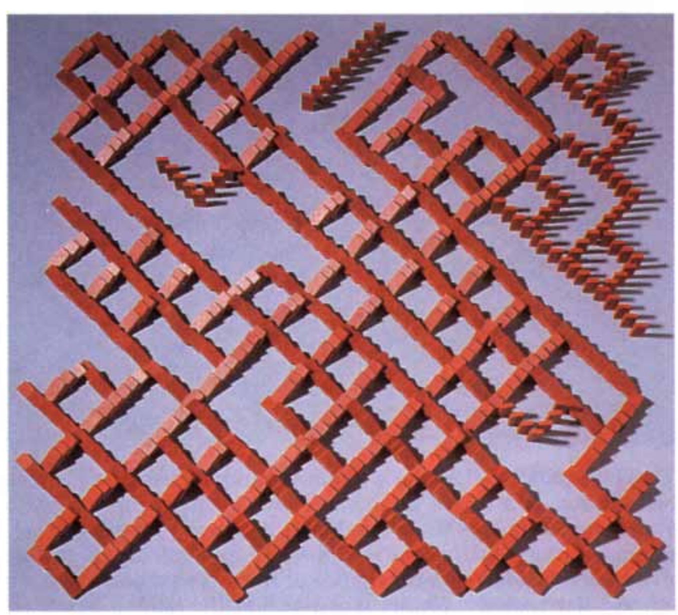
A deceptively simple system serves as a paradigm for self-organized criticality: a pile of sand. Some investigators have simulated the dynamics of sandpiles with computer programs; others such as Glenn A. Held and his colleagues at the IBM Thomas J. Watson Research Center have performed experiments. Both models and experiments reveal the same features.

Held and his co-workers devised an apparatus that pours sand slowly and uniformly, one grain at a time, onto a flat, circular surface. At first the grains stay close to the position where they land. Soon they rest on top of one another, creating a pile that has a gentle slope. Now and then, when the slope becomes too steep somewhere on the

PER BAK and KAN CHEN have studied self-organized criticality together since 1988. Bak is a senior scientist at Brookhaven National Laboratory. In 1974 he received his Ph.D. in physics at the Technical University of Denmark. Working for NORDITA Denmark, the IBM Thomas J. Watson Research Center and the University of Copenhagen, he pursued his interests in statistical physics, dynamic systems (chaos), low-dimensional conductors and quasicrystals. Bak was recently elected to the Danish Academy of Sciences. Chen is a postdoctoral fellow at Simon Fraser University, British Columbia. In 1983 he graduated from the University of Science and Technology in China, and in 1988 he obtained his Ph.D. in physics from Ohio State University. As a research associate at Brookhaven, he investigated self-organized criticality and dynamic optimization processes.







DOMINOES demonstrate criticality, subcriticality and supercriticality. In the critical system (*top*), dominoes were randomly placed on about half of the segments in a diamond grid. When the dominoes in the bottom row were tipped over, the critical system produced many sizes of chain reactions. The subcritical system (*bottom left*)—in which the density of dominoes was much less than the critical value—produced small chain reactions. The supercritical system (*right*)—in which the density was much greater than the critical value—exploded with activity.

pile, the grains slide down, causing a small avalanche. As more sand is added and the slope of the pile steepens, the average size of the avalanches increases. Some grains begin to fall off the edge of the circular surface. The pile stops growing when the amount of sand added is balanced, on average, by the amount of sand falling off the edge. At that point, the system has reached the critical state.

When a grain of sand is added to a pile in the critical state, it can start an avalanche of any size, including a "catastrophic" event. But most of the time, the grain will fall so that no avalanche occurs. We have found that even the largest avalanches involve only a small proportion of grains in the pile, and therefore even catastrophic avalanches cannot cause the slope of the pile to deviate significantly from the critical slope.

An avalanche is a type of chain reaction, or branching process. By simplifying the dynamics of the avalanche somewhat, one can identify the major features of the chain reaction and develop a model.

At the beginning of an avalanche, a single grain slides down the slope because of some instability on the surface of the pile. The grain will stop only if it falls into a stable position; otherwise, it will continue to fall. If it hits grains that are almost unstable, it will cause them to fall. As the process continues, each moving grain may stop or continue to fall, and it may cause other grains to fall. The process will cease when all the active particles have stopped or have moved off the sandpile. To measure the size of the avalanche, one can simply count the total number of fallen particles.

The pile maintains a constant height and slope because the probability that the activity will die is on average balanced by the probability that the activity will branch. Thus, the chain reaction maintains a critical state.

If the pile is shaped so that the slope is less than the critical value—the subcritical state—then the avalanches will be smaller than those produced by the critical state. A subcritical pile will grow until it reaches the critical state. If the slope is greater than the critical value—the supercritical state—then the avalanches will be much larger than those generated by the critical state. A supercritical pile will collapse until it attains the critical state. Both subcritical and supercritical piles are naturally attracted to the critical state.

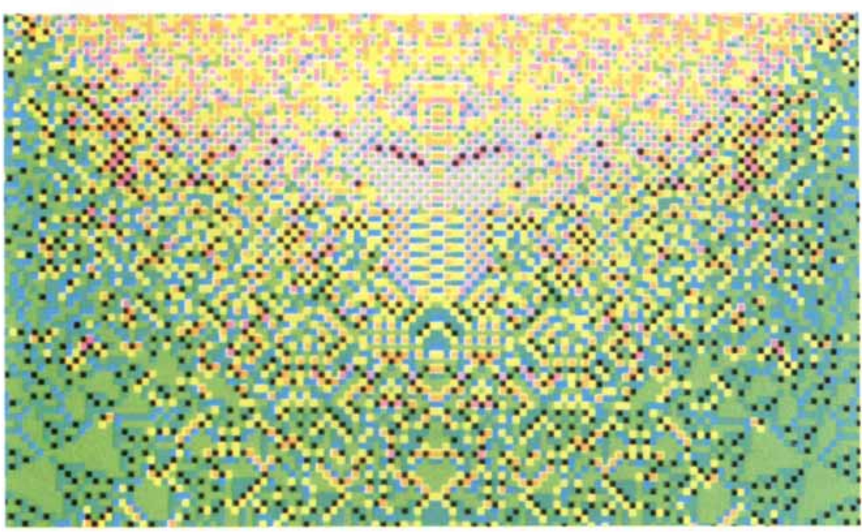
What will happen if one uses wet sand instead of dry or if one tries to prevent avalanches by building "snow screens"? At first the wet pile produces smaller avalanches at a slower rate than those of a comparable dry pile. After a while the wet pile builds up to a state steeper than the dry pile. In that state the wet pile supports avalanches of all sizes; it has evolved to a critical state. Similar dynamics can be observed for a pile that has snow screens. In general the critical state is robust with respect to any small change in the rules for the system.

The sandpile has two seemingly incongruous features: the system is unstable in many different locations; nevertheless, the critical state is absolutely robust. On the one hand, the specific features, such as the local configurations of sand, change all the time because of the avalanches. On the other, the statistical properties, such as the size distribution of avalanches, remain essentially the same.

An observer who studies a specific area of a pile can easily identify the mechanisms that cause sand to fall, and he or she can even predict whether avalanches will occur in the near future. To a local observer, large avalanches would remain unpredictable, however, because they are a consequence of the total history of the entire pile. No matter what the local dynamics are, the avalanches would mercilessly persist at a relative frequency that cannot be altered. The criticality is a global property of the sandpile.

Even though sand is added to the pile at a uniform rate, the amount of sand flowing off the pile varies greatly over time. If one graphed the flow versus time, one would see a very erratic signal that has features of all durations. Such signals are known as flicker noise, or $1/f$ noise (pronounced "one over 'ef' noise"). Scientists have long known that flicker noise suggests that the dynamics of a system are strongly influenced by past events. In contrast, white noise, a random signal, implies no correlation between the current dynamics and past events.

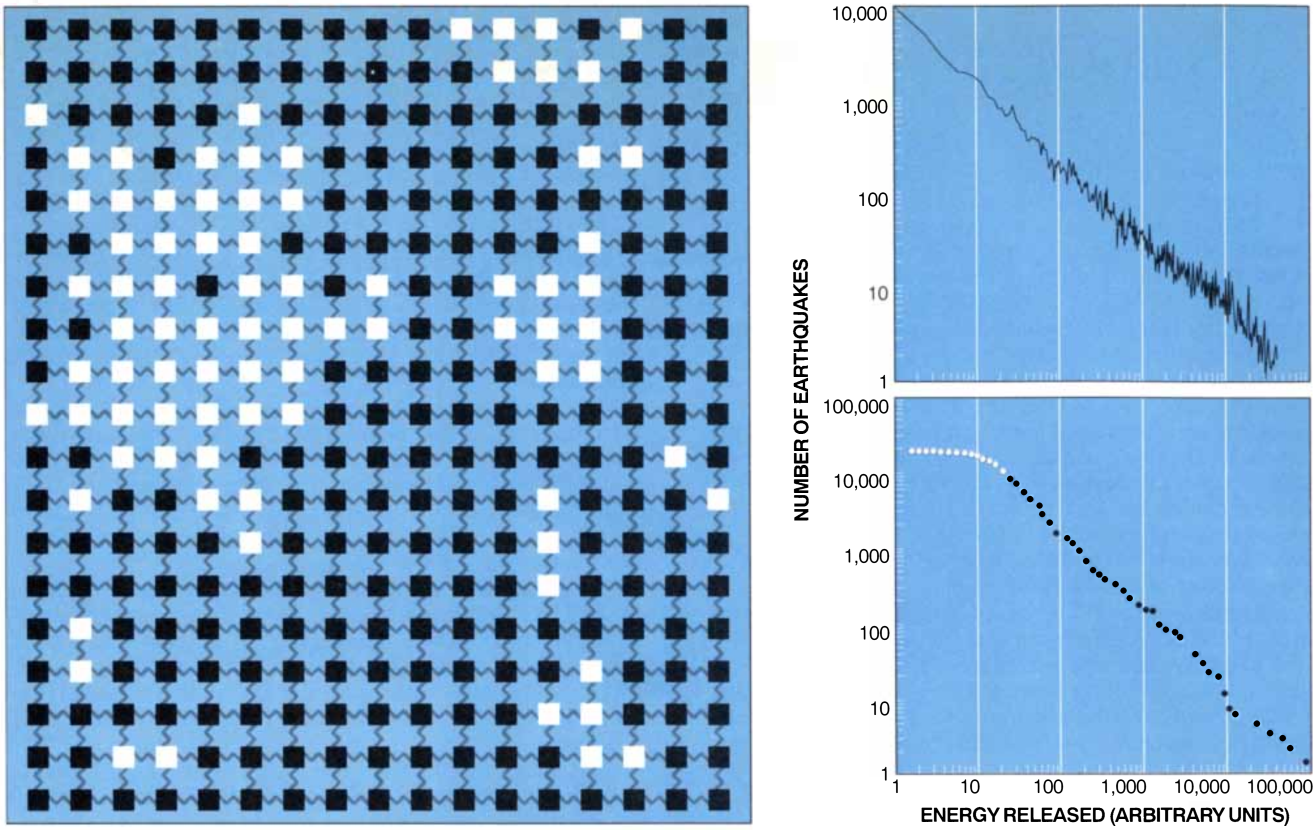
Flicker noise is extremely common in nature. It has been observed in the activity of the sun, the light from galaxies, the current through a resistor and the flow of water through a river. Indeed, the ubiquitousness of flicker noise is one of the great mysteries in physics. The theory of self-organized criticality suggests a rather general interpretation: flicker noise is a superposition of signals of all sizes and durations—signals produced when a dynamic system in the critical state produces chain reactions of all sizes and durations.



COMPUTER SIMULATION of a sandpile naturally evolves to a critical state in which the addition of a single grain can cause avalanches throughout the system. As grains were added to the pile (along the top row of the image), the computer determined where each grain would move and calculated the slope of the pile at several points. Pink squares represent the steepest parts of the pile; black squares indicate flat regions. Grains that reached the sides or bottom fell off the pile.

We and our colleagues have built many computer models that exhibit self-organized criticality. These models have helped us understand the dynamics of earthquakes, ecosystems and turbulence in fluids.

Models of earthquakes have perhaps been the most successful. In 1956 geologists Beno Gutenberg and Charles F. Richter, who is famous for devising the Richter scale, discovered that the number of large earthquakes is related to the number of small ones (known as the Gutenberg-Richter law). The number of earthquakes each year that re-



EARTHQUAKE MODEL simulates the forces on blocks of the earth's crust. Whenever the force on a block exceeds a critical value, the block slides, and the force is transferred to neighboring blocks. Each white square represents sliding blocks; each cluster represents an earthquake. The model produces earthquakes of all sizes, ranging from a single sliding event to "catastrophic" clusters extending throughout the system. The total number of sliding events in a cluster is a measure of the energy released during that earthquake. The graph at the top left shows the results accumulated after 10,000 model earthquakes. For comparison, the graph at the bottom right displays real earthquake measurements collected by Arch C. Johnston and Susan Nava at the New Madrid seismic zone in the U.S. The results from both the earthquake model and the measurements can be described by a power law: the number of earthquakes of energy, *E*, is proportional to *E* to the power of some constant. Power laws can be taken as evidence of self-organized criticality.

lease a certain amount of energy, E, is proportional to one divided by E to the power of b where the exponent b is about 1.5. The exponent b is universal in the sense that it does not depend on the particular geographic area. Hence, large earthquakes are much more rare than small ones. For example, if an area is hit each year, say, by one earthquake of energy 100 (in some units), it will experience approximately 1,000 earthquakes of energy 1 each year.

Because the number of small earthquakes is systematically related to the number of large earthquakes, one might suspect that small and large events arise from the same mechanical process. We and our co-workers believed the power-law distribution was evidence of self-organized criticality. Before we could test the theory, however, we needed to understand how we could simulate the process that produces earthquakes.

It is generally assumed that earthquakes are caused by a stick-slip mechanism: regions of the crust stick and then slide against other regions, creating faults. When one region slides against another, stress is released and propagates to adjacent regions.

To replicate this mechanism in the laboratory, Vladimir Bobrov and Mihail Lebyodkin of the Institute of Solid State Physics in Chernogolovka, Moscow, performed an experiment in which they applied pressure to an aluminum rod, representing a region of the earth's crust. The pressure caused a transition from elastic flow (at which point the rod would return to its original shape once the pressure was released) to plastic flow (at which point the deformation was irreversible). In the plastic phase the rod developed a "fault" region where two parts of the rod would slide against each other. Bobrov and Lebyodkin observed "earthquakes" whose size and frequency were related by a power law. When they conducted the experiments with niobium bars instead of aluminum, they obtained the same results, even though the microscopic mechanisms for the two materials are different.

We have constructed a simple computer model of the earth's crust that reproduces important features of earthquakes. Our model is composed of one elastic plate and one rigid plate, for simplicity's sake. The elastic plate is represented by a two-dimensional array of blocks, each connected to four neighbors by springs. As the array of blocks is squeezed, the springs exert a force on the blocks proportional to the compression. (We later included other types of forces in the model, resulting in little change in the dynamics.) The blocks of the elastic plate interact with the rigid plate through friction.

Whenever the spring force exerted on a particular block exceeds a critical value, the block slides. It continues to move until the spring force has been reduced below the critical value. The force lost by the block is transferred equally to its four neighbors. (During this process, the potential energy stored in the springs is first converted to kinetic energy and then dissipated when the blocks are decelerated by frictional forces.) The model describes the force distribution before

Self-Organized Criticality and Sandpiles

The theory of self-organized criticality makes a simple prediction about sandpiles: when a single grain of sand is dropped on a pile, it usually causes a few grains to fall, but every so often it will initiate a large avalanche. To test that prediction on real sandpiles, Glenn A. Held and his colleagues at the IBM Thomas J. Watson Research Center recently devised an ingenious experiment.

The most difficult challenge for the IBM group was constructing an apparatus that would add sand to a pile slowly, essentially one grain at a time. Held and his co-workers mounted a variable-speed motor on a laboratory stand and clamped a funnel to the motor's drive shaft. The funnel consisted of a 250-milliliter leveling bulb and a capillary tube 23 centimeters long and 2.0 millimeters in inner diameter. They then filled the funnel with sand and angled it about two degrees lower than horizontal so that sand grains slid into the capillary tube but did not fall through.

As the motor rotated the funnel about its axis at approximately 60 revolutions per minute, the sand grains lined up in the capillary tube and traveled single file to the end. By adjusting the angle of the funnel and the speed of the motor, the team could tune the apparatus so that one grain would fall every 15 seconds. They positioned the mouth of the capillary tube about 10 centimeters above the pan of a high-precision balance.

The balance had a precision of .0001 gram and a capacity of 100 grams. Each grain of sand weighed about .0006 gram; a sandpile whose base was four centimeters in diameter weighed approximately 15 grams.

To support the sandpiles, Held and his colleagues built circular plates ranging from one to eight centimeters in diameter. They attached each plate to a post 2.5 centimeters long and .5 centimeter in diameter. The post, in turn, was connected to a circular base four centimeters in diameter. The whole assembly—which looked like a tiny cake stand or a spool—rested on top of the balance pan. They built a metal skirt around the post of the "spool" to keep the sand that fell off the pile away from the balance pan so that the balance would record only the weight of the pile itself.

The balance was enclosed by a shield to prevent drafts from blowing the sand around. The equipment sat on a heavy table to damp vibrational disturbances. It took Held about 10 hours to put the whole apparatus together.

In the first experiments Held and his co-workers used aluminum oxide particles, but they found that sand from Long Island Beach worked just as well. They prepared the sand by drying it in an oven and filtering the grains through a coarse and then a fine sieve. They kept the grains that passed through the coarse mesh (eight cross wires per centimeter) but removed the grains that flowed through the fine mesh (10 cross wires per centimeter).

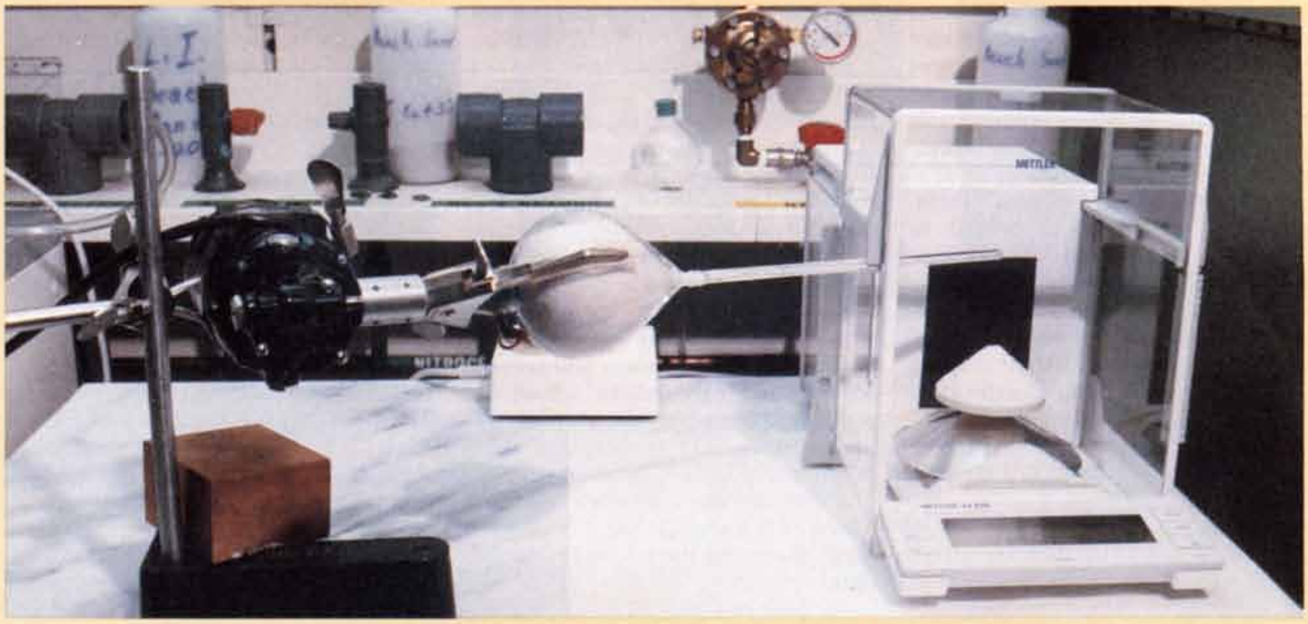
They filled the funnel with sand and crudely molded a pile on a four-centimeter circular plate. To ensure that the sandpile would settle into its natural form, they allowed grains to fall on the pile continuously for a few hours. They then watched avalanches of sand cascade down the pile. As sand fell off the edges of the plate, they measured the fluctuations in the mass of the pile.

The group used a personal computer to control the motor and to monitor the balance. When the computer detected a change in mass comparable to a grain of sand, it stopped the rotation of the funnel and thus the flow of sand. After the balance stabilized, the computer recorded the mass. It then restarted the motor to resume the dropping process.

Held and his colleagues ran the system for two weeks, dropping more than 35,000 grains on the four-centimeter plate. They observed avalanches in a range of sizes. The mass of the pile fluctuated by one to several hundred grains when as few as one and as many as several thousand grains were added to the pile. This result strongly suggested that the sandpile had indeed organized itself to a critical state.

When the team increased the diameter of the sandpile's base by using an eight-centimeter plate, however, they found that the pile produced only large avalanches (about four grams). They concluded that sandpiles of this size do not exhibit self-organized criticality. They do not yet understand why only small piles naturally evolve to a critical state.

More information about the sandpile experiment can be found in "Experimental Study of Critical-Mass Fluctuations in an Evolving Sandpile," by Glenn A. Held et al., in *Physical Review Letters*, Vol. 65, No. 9, pages 1120–1123; August 27, 1990.



and after each event but not the motions of the block or other details of the dynamic process.

When the model is driven by increasing the force at a uniform, low rate on all blocks in the same direction, the model begins to produce a series of earthquakes. At first the model produces only small earthquakes, but eventually it evolves to a critical state in which both small and large earthquakes are generated. The uniform increase in force is balanced by the release of force at the boundary.

We studied the model most intensely after the system had evolved to the critical state. We assume that the crust of the earth has already evolved to the stationary, critical state, and, therefore, real earthquakes can be simulated by the critical state of the model.

In the model, the energy released during an earthquake is related to the number of sliding events that follow a single instability at some "epicenter." Indeed, if one counts the number of earthquakes of each size during an extended period, one obtains a power-law distribution similar to the Gutenberg-Richter law [*see illustration on page 49*]. The catastrophic earthquakes are represented by the high-energy part of the power-law curve, which can be extrapolated smoothly from the low-energy part, which represents small earthquakes. There is no separate mechanism for the large earthquakes.

We have created models in two, three and four dimensions in which four, six or eight springs, respectively, are connected to each block. The dimension determines the exponent *b* of the power law. In the critical chain reaction picture, different *b* values correspond to different couplings between the individual branching processes. Sergei P. Obukhov of the Landau Institute in Moscow has shown that in four or more dimensions the individual branching processes are essentially independent, and the *b* value can be determined to be 1.5 by analytical methods.

Of course, the active regions of real earthquakes are three-dimensional, and computer simulations are so far the only way to predict the real *b* value. One cannot expect a grossly simplified model to yield the correct exponents for the distribution of real earthquakes. Nevertheless, the model suggests that power laws should arise from self-organized critical systems, and, conversely, the Gutenberg-Richter law can be taken as evidence that the earth's crust is indeed locked in a perpetually critical state.

Investigators around the world have applied the theory of self-organized criticality to account for many other features of earthquakes. Keisuke Ito and Mitsuhiro Matsuzaki of Kobe University were able to explain the spatial distribution of epicenters with a slightly modified model. They also explained a simple empirical law for the number of aftershocks of a given magnitude known as Omori's law. Anne and Didier Sornette of the University of Nice studied the time intervals between large earthquakes and found patterns that may have important implications for long-term forecasting of earthquakes. Jean M. Carlson and James S. Langer of the Institute for Theoretical Physics created a one-dimensional model that simulated the motions of the earth's crust along a fault. They found that the model did indeed evolve to a critical state.

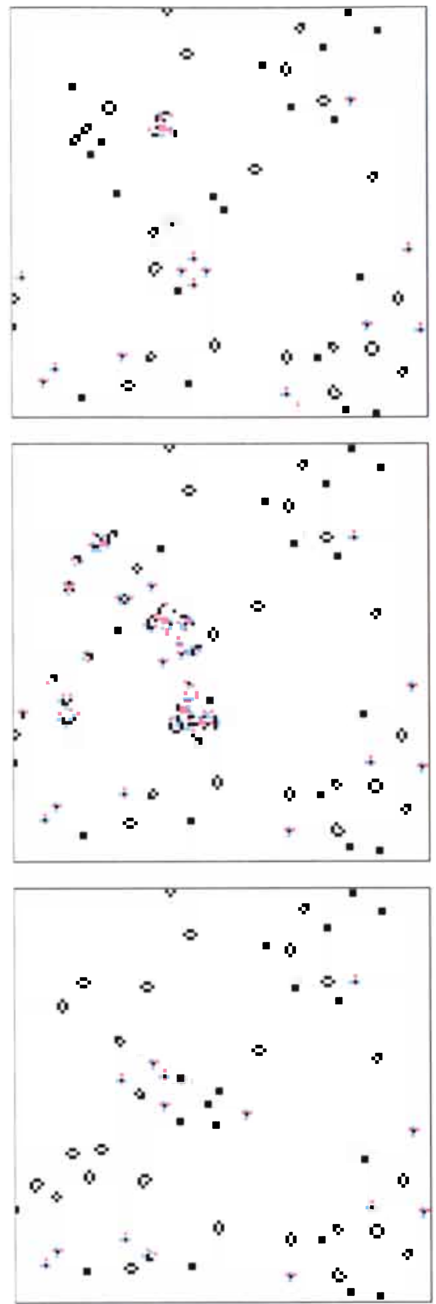
The theory of self-organized criticality has been successful not only in explaining the evolution of earthquakes but also in describing the distribution of the epicenters of earthquakes. For more than a decade, workers have known that power laws can describe the distribution of objects such as mountains, clouds, galaxies and vortices in turbulent fluids. The number of objects within, say, a sphere of radius r is proportional to r to the power of some constant D. Such a distribution of objects is generally called a fractal [see "The Language of Fractals," by Hartmut Jürgens, Heinz-Otto Peitgen and Dietmar Saupe; SCIENTIFIC AMERICAN, August 1990]. We find that fractals describe the distribution of the epicenters of earthquakes.

Although fractals appear throughout nature, investigators have only begun to understand the dynamics that create fractals. We and our colleagues suggest that fractals can be viewed as snapshots of self-organized critical processes. Fractal structures and flicker noise are the spatial and temporal fingerprints, respectively, of self-organized criticality.

Earthquake prediction still remains a difficult task. The stability of the earth's crust appears to be quite sensitive to the initial conditions of the system. Sometimes conditions far away from the epicenter can affect the evolution of an earthquake.

To evaluate the accuracy of predictions for a dynamic system, one must know the initial conditions with some precision as well as the rules of the dynamics. In nonchaotic systems, such as the earth orbiting around the sun, the uncertainty remains constant at all times: one can determine the earth's position in one million years with almost as much precision as one can know the earth's position today.

In chaotic systems, a small initial uncertainty grows exponentially with time. Furthermore, as one attempts to make predictions further and further into the future, the amount of information one needs to gather about the initial conditions increases exponen-



GAME OF LIFE simulates the evolution of a colony of organisms and suggests that the theory of self-organized criticality can explain the dynamics of ecosystems. Black squares indicate living organisms, and red squares show dying ones; blue squares represent the imminent birth of an organism. The first frame shows the colony a short time after a single organism was added to a stable configuration. In the second and third frames, the colony evolves to a new, stable state.

AVALANCHE DYNAMICS may be explained by the theory of self-organized criticality, which states that snow piles and other large complex systems naturally evolve to a critical state in which minor events can cause chain reactions of many sizes. If the theory proves correct, analysts may improve predictions of catastrophes.

tially with time. For the most part, this exponential growth prevents long-term predictions.

To check the accuracy of predictions in our earthquake model, we conducted two simulations of the critical state. The simulations differ by a small random force on each block, representing a small uncertainty about the initial conditions. When we run the two simulations, the uncertainty grows with time but much more slowly than it does for chaotic systems. The uncertainty increases according to a power law rather than an exponential law. The system evolves on the border of chaos. This behavior, called weak chaos, is a result of self-organized criticality.

Weak chaos differs significantly from fully chaotic behavior. Fully chaotic systems are characterized by a time scale beyond which it is impossible to make predictions. Weakly chaotic systems lack such a time scale and so allow long-term predictions.

Because we find that all self-organized critical systems are weakly chaotic, we expect weak chaos to be very common in nature. It would be interesting indeed to know whether the inaccuracy of earthquake predictions, economic forecasts and weather forecasts generally increases with time according to a power law rather than an exponential law.

For example, if weather is chaotic and if 100 observatories gather enough information to predict the weather two days in advance, then 1,000 observatories might allow predictions four days in advance. If weather is weakly chaotic, then 1,000 observatories might allow predictions 20 days in advance.

By a change of language (and a little bit of imagination), one can transform the sandpile or the earthquake model into many situations. It has been shown, for instance, that traffic on a highway is flicker noise. The stop-and-go pattern can be thought of as critical avalanches propagating in the traffic.

Models of traffic, sandpiles and earthquakes are similar in the sense that the number of elements is always conserved. For instance, the number of grains of sand in the pile always equals the number that was dropped on the pile minus the number that fell off. The conservation of elements is an important feature of many systems that naturally evolve to a critical state. The theory of self-organized criticality is not limited to systems that have local conservation laws, however, as can be shown by the "game of life."

In 1970 mathematician John H. Conway invented the game, which was subsequently popularized by Martin Gardner in a series of articles in these pages [see "Mathematical Games," SCIENTIFIC AMERICAN, October 1972]. The game of life simulates the evolution of a colony of living organisms and mimics the generation of complexity in nature.

To start the game, the pieces, or organisms, are placed at random on a board composed of square sites. Each site is occupied by at most one organism and is surrounded by eight neighboring sites. To determine the status of a site at each turn, one must count the number of organisms that occupy the eight neighboring sites. If one counts two live sites around an empty or an occupied site, then the status of the site does not change. If one counts three live sites around a site, the live sites give birth to a new organism or sustain the life of an old organism. In all other cases, an organism will die from overcrowding or loneliness.

The game continues according to the rules until it comes to "rest" in a simple periodic state, containing stable colonies. When the game is perturbed by adding an extra "live" cell, the system often sustains long transients of activity.

We and Michael Creutz of Brookhaven National Laboratory recently studied the game of life to determine whether the number of live sites would fluctuate over time, like the size of the avalanches in the sandpile model. Once the system settled into its rest state, we added a single organism at a random position, waited until the system settled and then repeated the procedure. Next we measured the total number of births and deaths in the "avalanche" after each additional perturbation. Indeed, the distribution was found to be a power law, indicating that the system had organized itself to a critical state.

We also found that the distribution of live sites is a fractal that can be described by a power law. The average number of active sites within a distance r from a given active site was proportional to r to the power of D, where D turned out to be about 1.7.

But is the criticality accidental in the sense that it occurs only for the very particular rules invented by Conway? To find the answer, we have constructed models that are variations of the game of life. Some variations were three-dimensional; in others, organisms were added to the system as it was evolving, or they were introduced at specific rather than random sites. All variations evolved to a critical state and could be described by power laws that seem to depend only on the spatial dimension.

We speculate that our models may have important ramifications in real biology. One may view the game of life as a toy model of a coevolutionary system. Each site can represent a gene of a very simple species. This gene can assume the values of 1 or 0. The stability of each value depends on the environment as expressed by the values of the genes of nearby species. The coevolutionary process then takes the system from an initial random state to the highly organized, critical state with complex static and dynamic configurations. The complexity of the global dynamics is intimately related to the criticality of the dynamics. In fact, the theory of complexity and the theory of criticality may generically be one and the same thing.

Biologist Stuart Kauffmann of the University of Pennsylvania has suggested a model for evolution in which species are represented by strings of numbers (genes). The genes interact both within and between species. The fitness of individual species is thus coupled to the fitness of other species. Kauffmann speculated that the complexity of life might be intimately related to the existence of a critical state. Our studies indicate that evolution may indeed automatically lead a simple, more or less random interactive dynamic system to precisely such a critical state. If this scenario is correct, then evolution operates at the border of chaos. The extinction of the dinosaurs, for instance, may be thought of as an avalanche in the dynamics of evolution, and an external force, such as a meteorite or volcano, would not be necessary.

Philip W. Anderson of Princeton University, Brian W. Arthur of Stanford University, Kauffmann and one of us (Bak) became aware that fluctuations in economics might indeed be avalanches in a self-organized critical state of that system. Benoit B. Mandelbrot of IBM has analyzed indicators such as the Dow Jones index and found fluctuations similar to flicker noise. The various metastable stationary states of economics might correspond to the various metastable states of a sandpile or the earth's crust.

Conventional models assume the existence of a strongly stable equilibrium position for the economy, whereby large aggregate fluctuations can result only from external shocks that simultaneously affect many different sectors in the same way. Yet it is often difficult to identify the reasons for such large-scale fluctuations as the depression of the 1930s. If, on the other hand, the economy is a self-organized critical system, more or less periodic large-scale fluctuations are to be expected even in the absence of any common jolts across sectors.

To check the viability of those ideas, we and José A. Scheinkman and Michael Woodford of the University of Chicago have created a simple model in which companies that make different products occupy positions in a two-dimensional lattice, or economic web. Each company buys supplies from two companies located at adjacent positions. Each then produces new products, which it tries to sell on the open market. If the demand for each company's product varies at random by a small amount, many companies may experience an "avalanche" in sales and production. Simulations indicate that such a model tends toward a self-organized critical state in the way the sandpile model does. Large fluctuations are intrinsic and unavoidable properties of the dynamics of this model economy.

The theory of self-organized criticality may also find applications in fluid dynamics. It has long been assumed that energy in turbulent fluids is stored in swirling vortices of all sizes. Mandelbrot has suggested that the energy dissipation would be confined to a tiny part of the space, located on a complex fractal structure. Although this seems to be in accordance with experiments, no theory or calculation that supports this picture has been performed.

In collaboration with Tang, we have constructed a simple "toy" model of turbulence that operates at the self-organized critical state. The model simulates forest fires where "trees" grow uniformly and burn (energy dissipated) on a fractal. The energy dissipation may be viewed as caused by a sequence of fires propagating as avalanches. In the critical state, there is a distribution of fires and forests of all sizes, corresponding to the fact that during turbulence energy is stored in vortices of all sizes. Although the model is quite unrealistic for turbulence in liquids, we nevertheless speculate that turbulence may indeed be a self-organized critical phenomenon. One consequence of this hypothesis (which can be readily studied experimentally) is that fully developed turbulence is not a "strong" chaotic phenomenon, as often assumed. Turbulence would then be only weakly chaotic, as is the case in the earthquake model.

One might think of more exotic examples of self-organized criticality. Throughout history, wars and peaceful interactions might have left the world in a critical state in which conflicts and social unrest spread like avalanches. Self-organized criticality might even explain how information propagates through the neural networks in the brain. It is no surprise that brainstorms can be triggered by small events (for instance, we hope, reading this article).

FURTHER READING

SELF-ORGANIZED CRITICALITY. Per Bak, Chao Tang and Kurt Wiesenfeld in *Physical Review A*, Vol. 38, No. 1, pages 364–374; July 1, 1988.

EARTHQUAKES AS A SELF-ORGANIZED CRITICAL PHENOMENON. Per Bak and Chao Tang in *Journal of Geophysical Research*, Vol. 94, No. B11, pages 15635–15637; November 10, 1989.

THE PHYSICS OF FRACTALS. Per Bak and Kan Chen in *Physica D*, Vol. 38, pages 5–12; 1989.

SELF-ORGANIZED CRITICALITY AND EARTHQUAKES. Anne Sornette and Didier Sornette in *Europhysics Letters*, Vol. 9, No. 3, pages 197–202; 1989.

SELF-ORGANIZED CRITICALITY IN THE 'GAME OF LIFE.' Per Bak, Kan Chen and Michael Creutz in *Nature*, Vol. 342, No. 6251, pages 780–782; December 14, 1989.

A FOREST-FIRE MODEL AND SOME THOUGHTS ON TURBULENCE. Per Bak, Kan Chen and Chao Tang in *Physics Letters*, Vol. 147, Nos. 5–6, pages 297–300; 1990.

The Protein Folding Problem

In theory, all one needs to know in order to fold a protein into its biologically active shape is the sequence of its constituent amino acids. Why has nobody been able to put theory into practice?

by Frederic M. Richards

In the late 1950s Christian B. Anfinsen and his colleagues at the National Institutes of Health made a remarkable discovery. They were exploring a long-standing puzzle in biology: What causes newly made proteins—which resemble loosely coiled strings and are inactive—to wind into specifically shaped balls able to perform crucial tasks in a living cell? In the process the team found the answer was simpler than anyone had imagined.

It seemed the amino acid sequence of a protein, a one-dimensional trait, was fully sufficient to specify the molecule's ultimate three-dimensional shape and biological activity. (Proteins are built from a set of just 20 amino acids, which are assembled into a chain according to directions embedded in the genes.) Outside factors, such as enzymes that might catalyze folding, did not have to be invoked as mandatory participants.

The discovery, which has since been confirmed many times—at least for relatively small proteins—suggested that the forces most responsible for proper folding in the cell could, in theory, be derived from the basic principles of chemistry and physics. That is, if one knew the amino acid sequence of a protein, all that would have to be considered would be the properties of the individual amino acids and their behavior in aqueous solution. (The interior of most cells is 70 to 90 percent water.)

In actuality, predicting the conformation of a protein on the basis of its amino acid sequence is far from simple. More than 30 years after Anfinsen made his breakthrough, hundreds of investigators are still at work on that challenge, which has come to be widely known as the protein folding problem.

The solution is of more than academic interest. Many major hoped-for products of the developing biotechnology industry are novel proteins. It is already possible to design genes to direct the synthesis of such proteins. Yet failure to fold properly is a common production concern.

For a time, those of us working on the folding problem despaired of ever finding an answer. More recently, however, advances in theory and experiment, combined with growing interest on the part of industry, have kindled new optimism.

Most of the detailed information available so far comes from studies on small, water-soluble, globular proteins containing fewer than 300 or so amino acids. The relative importance of various rules of folding and assembly may be somewhat different for those proteins than for others—notably long fibrous proteins and varieties residing in cellular membranes. Indeed, some large proteins have recently been shown to need folding help from other proteins known as chaperonins. The balance of the article will not consider such complexities but will focus entirely on the unassisted folding reaction undergone by a great many proteins.

It would be wonderful if researchers had an atomic-level microscope that could take a movie of individual protein molecules folding up from their extended, unstable state to their final, or native, state, which is more stable. From a collection of movies, all aspects of the reaction pathways could be seen directly. Unfortunately, no such instrument exists; investigators must fall back on much less direct measurements and very careful reasoning.

One can gather helpful clues to the rules of folding by examining the three-dimensional structures of unfolded and

FREDERIC M. RICHARDS is Sterling Professor of Molecular Biophysics and Biochemistry at Yale University. Richards joined the Yale faculty in 1955, three years after earning his doctorate from Harvard University.

fully folded proteins and by analyzing the properties of individual amino acids and small peptides (linear chains of amino acids). Fortunately, the architecture of hundreds of native proteins has been determined by such imaging techniques as X-ray crystallography and, more recently, nuclear magnetic resonance (NMR). Both techniques have advanced dramatically in the past decade, as has theoretical work attempting to predict folding mathematically by computer.

Isolated amino acids consist of a central carbon atom—called the alpha carbon—bound to an amino group (NH_2), a carboxyl group (COOH) and a side chain. The differences among amino acids, then, stem from differences in their side chains, namely, in shape, size and polarity. Shape and size affect the packing together of amino acids in the final molecule. Polarity (or lack of it) determines the nature and strength of interactions between amino acids in a protein and between the protein and water.

For instance, polar amino acids interact strongly with one another in what are called electrostatic interactions. The molecules are considered polar if they carry a formal charge (owing to the loss or gain of one or more electrons) or if they are electrically neutral overall but have localized regions where positive or negative charges predominate. (Positive charges are contributed by protons in atomic nuclei, negative charges by electrons surrounding the nuclei.) Molecules are attracted when their oppositely charged regions are close; they are repelled when like charged regions are close.

Nonpolar amino acids can also attract or repel one another, albeit more weakly, because of what are called van der Waals forces. Electrons and protons vibrate constantly, and the vibrations result in attractions between substances that are near one another. The attraction turns into repulsion when the substances are about to touch.

In aqueous solution, polar amino acids tend to be hydrophilic; they attract water molecules, which are quite polar. In contrast, nonpolar amino acids, which generally include hydrocarbon side chains, tend to be hydrophobic: they mix poorly with water and "prefer" to associate with one other. Alternatively, one can think of them as being squeezed out of the water as a consequence of the strong attraction between polar substances.

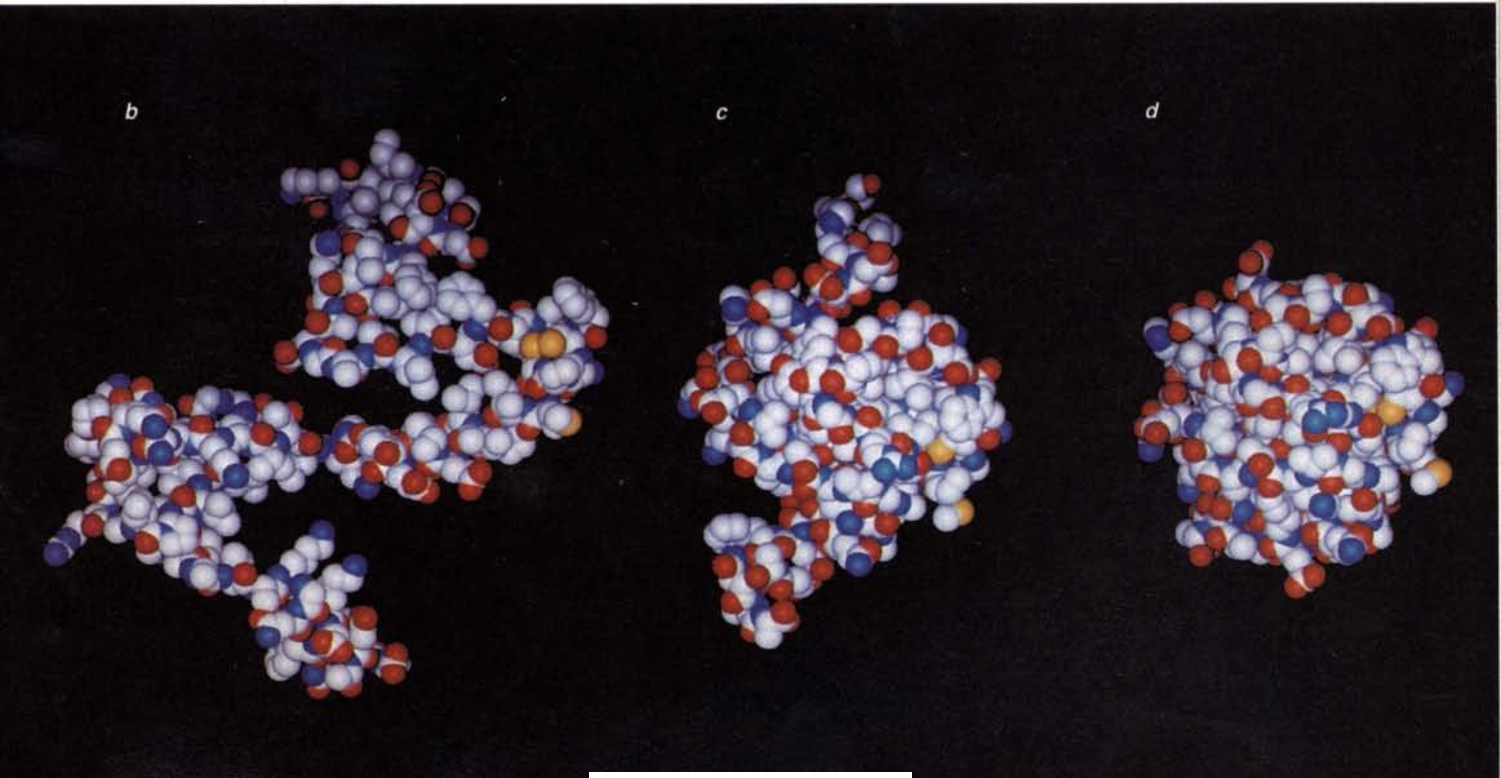
The peptide bond linking one amino acid to the next in a sequence influences folding as well; it markedly constrains the universe of possible conformations that can be taken by the protein backbone (the repeating series of alpha carbons, carboxyl carbons and amino nitrogens in a peptide chain). A peptide bond forms when the carboxyl carbon of one amino acid binds with the amino nitrogen of the next, releasing a molecule of water. The resulting strong linkage between the connected amino acids—called residues once they are joined—is quite rigid.

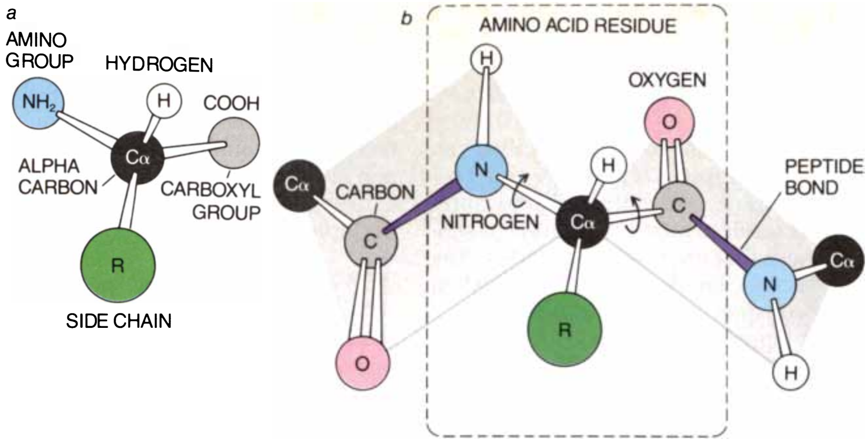
Consequently, rotation about the peptide bond is severely limited. Indeed, the atoms lying between alpha carbons are held in a single plane, so that they essentially form a stiff plate. Folding of the peptide backbone is therefore accomplished mainly by rotation of the plates around other bonds—namely, those connecting the plates to the alpha carbons.

Examination of the peculiarities of denatured, or unfolded, proteins has added still other hints to how folding is accomplished. Unfolded or newly formed proteins are often called random coils, implying that no region of the backbone looks significantly different from any other region. In fact, the chains are probably never truly without some regions that are twisted, associated or otherwise different from the rest of the molecule. Certain of these substructures, which are probably unstable and fluctuating, might well serve as "seeds," around which stably sculpted regions eventually form.

Significantly more is known about

FOLDING of the protein thioredoxin is generally representative of how other small proteins fold: the initially open, unstable chain (*a*) becomes increasingly compact (*b* and *c*), ultimately adopting a spherical shape (*d*). The intermediate stages shown are hypothetical, because their shapes—and those of the intermediates of most proteins—are not known fully. White represents carbon; red, oxygen; blue, nitrogen; and yellow, sulfur.





AMINO ACIDS (*a*) are linked together in a protein (*b*) by a strong bond that forms between the carboxyl carbon of one amino acid and the amino nitrogen of the next. Because the resulting linkage, which is known as a peptide bond, holds the joined atoms rigidly in a plane, the bond limits the conformational options of the protein. Folding is accomplished mainly by rotation about the axes of the bonds connecting the central alpha carbon with the amino nitrogen and carboxyl carbon.

the folded than the unfolded state. For instance, most of the backbone of the compact, native molecule can be divided into regions of secondary structure, which are distinct segments having characteristic shapes. (The amino acid sequence is the primary structure.) The secondary elements fall into three main categories: helices (mainly the so-called alpha helix), beta strands or beta sheets, and turns connecting the helices and strands [*see illustration on opposite page*]. In beta strands the backbone is extended, or stretched out; in beta sheets two or more parallel or antiparallel strands are arranged in rows.

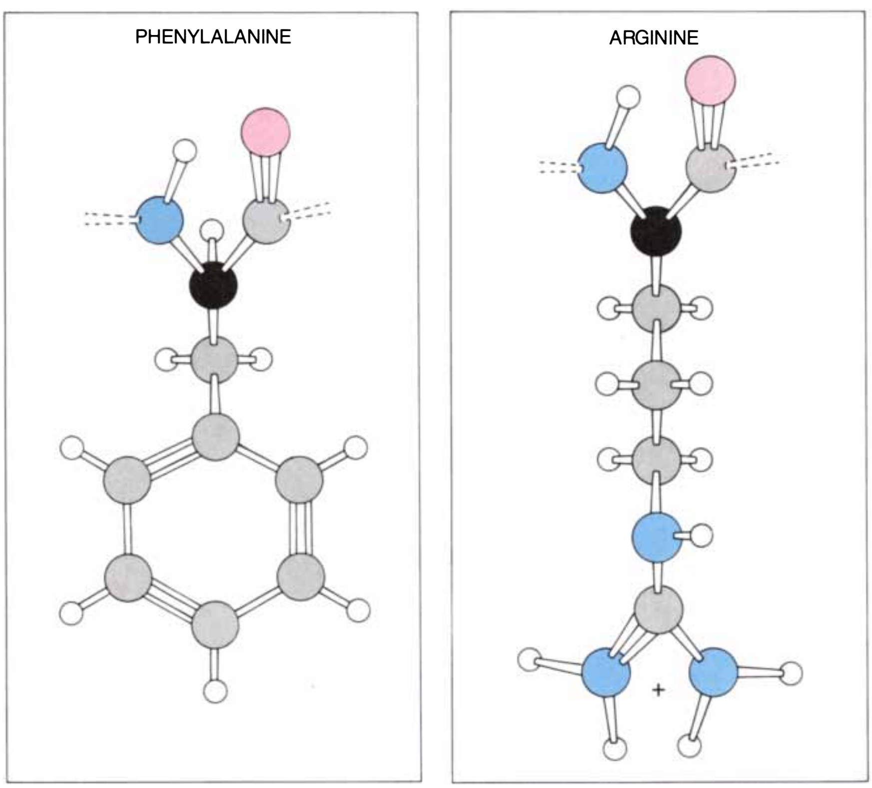
Secondary elements can combine with one another to form motifs, or supersecondary structures, and the final assembly of all secondary elements is the tertiary structure. Several tertiary classes have been identified, such as the all alpha-helix class, the all beta-strand class and particular arrangements of combinations of helices and beta strands.

The presence of different secondary elements raises the possibility that certain amino acids favor development of specific secondary arrangements. For example, some amino acid residues are found more often in helices than elsewhere, whereas others tend to be found in beta sheets. On the other hand, none of these or other similar statistical correlations are strong.

Several other discoveries show that, as might be expected from the hydrophobic and hydrophilic properties of the amino acids, the tendency of water and nonpolar residues to avoid one another has a profound effect on the final shape of a protein. The interior of native proteins is largely free of water and contains mainly nonpolar, hydrophobic amino acids. Conversely, residues with formal charges almost invariably reside at the surface, in contact with water. Polar residues are found on both the outside and the inside, but in the interior they are invariably joined to other polar groups by hydrogen bonds. Such bonding, in which two atoms (usually nitrogen and oxygen) are joined through a shared hydrogen atom, apparently enables the residues to remain comfortable in the interior, away from water.

Hence, one rule of folding seems to be that contact between water and hydrophobic amino acids must be limited as much as possible, although this general rule is not sufficient to predict which specific residues will appear where. For example, it is not possible to identify which nonpolar residues will remain at the surface, as some fraction of them invariably do.

Another general rule, based on other analyses, posits important steric constraints. The final product has to be packed efficiently, that is, the space must be filled without having neighboring atoms overlap. Structural studies show that the amino acids in folded proteins are generally packed about as tightly as other small organic molecules pack together. Computer modelers can safely assume that in the final protein (with a few rare exceptions), the lengths of bonds between atoms and the angles between consecutive bonds will be identical to those that have been found in smaller organic molecules.



DIFFERENCES in the shape, size and polarity of amino acids derive from differences in their side chains. In phenylalanine, for example, the side chain is nonpolar and cyclic, whereas the side chain of arginine is both strongly polar and linear.

Researchers agree on details of the structure of folded proteins, but they diverge on many other points. There is, for instance, little agreement on the nature and number of folding pathways.

At one extreme is the doubtful suggestion that a newly made protein tries out all possible conformations until it finds the unique, stable structure of the native protein. This proposal assumes that all conformations are equally likely to be tested; yet they are not. Also, as was pointed out years ago by Cyrus Levinthal, then at the Massachusetts Institute of Technology, no molecule would have the opportunity to test anywhere near all the possible conformations in the time it takes for proteins to fold—a few seconds at most.

At the other extreme is the notion that proteins follow a single, defined pathway: every molecule of a given protein becomes compacted by following one defined sequence of steps. Considering the great number of conformations an unfolded molecule can adopt, that idea seems improbable as well. This hypothesis is akin to the proposition that everyone will enter New York City via Interstate 95, regardless of where they start.

A third suggestion, which admits of one or more pathways, assumes the hydrophobic effect is all important initially, much more so than electrostatic interactions or space-filling concerns. This idea holds that the chain collapses rapidly to approximately its final density in order to remove hydrophobic amino acids from water. Then, in this much reduced space, it rapidly reorganizes itself into the correct secondary and tertiary structures. From a mechanical point of view, this scenario seems unlikely because the chain would have to open up somewhat to permit the required movements. Nevertheless, the model has some experimental support.

The best guess today is that secondary structure forms before most proteins are able to compact extensively. Molecules of the same protein can follow different pathways to the same end, but the choice of pathway is limited. Various models along these lines have been proposed, including what is called the framework model of Robert L. Baldwin of Stanford University and Peter S. Kim, now at the Whitehead Institute for Biomedical Research.

In general, such models suggest that the unfolded chain rapidly forms marginally stable bits of secondary structure. Some of these segments interact. If they pack together particularly well, or form bonds readily, they stabilize one another, at least for a time. The stabilized units, or microdomains, lead the molecule toward greater structural organization by associating with other segments or helping to bring distant segments into contact, or both.

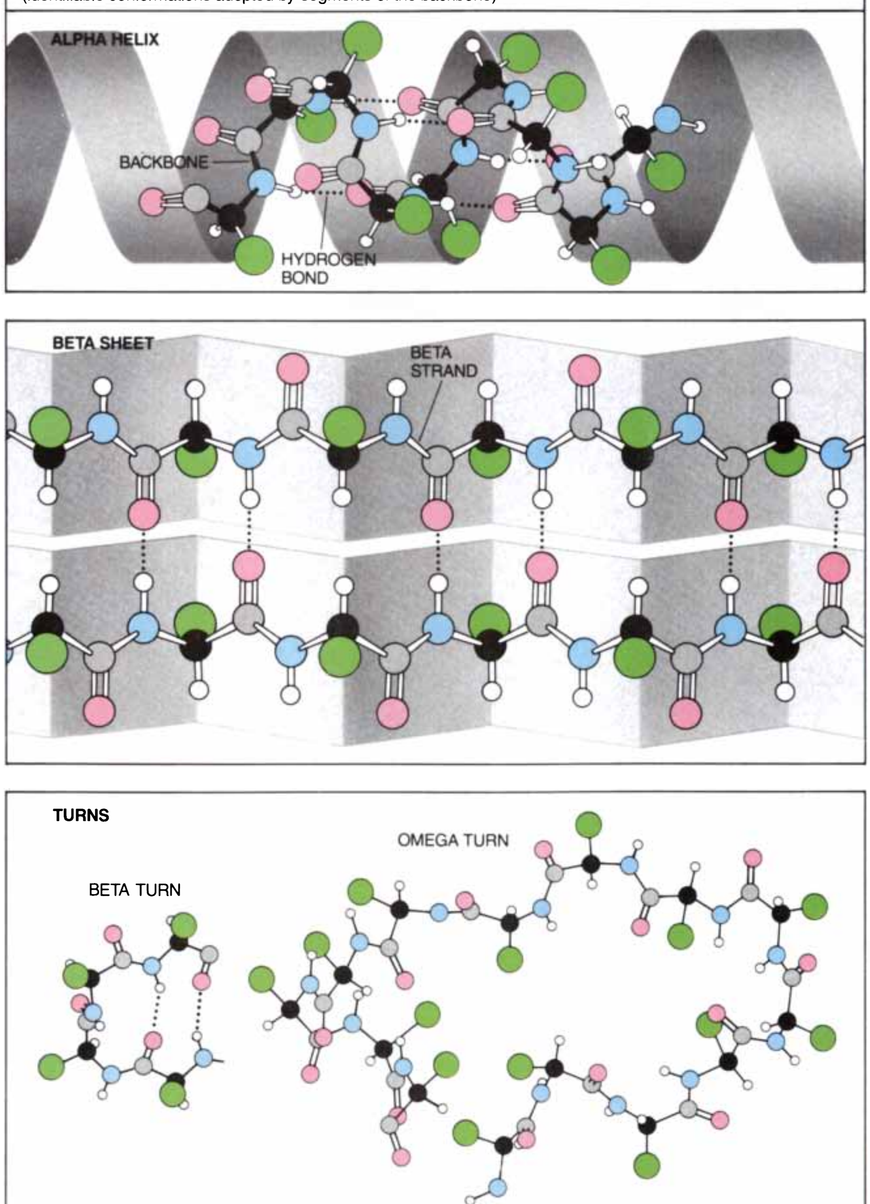
Inherent in this kind of model is the assumption that the hydrophobic effect is large but can be spent incrementally. Some fraction of its energy is expended to influence the formation of secondary elements, and the rest promotes the association of those elements into the tertiary conformation.

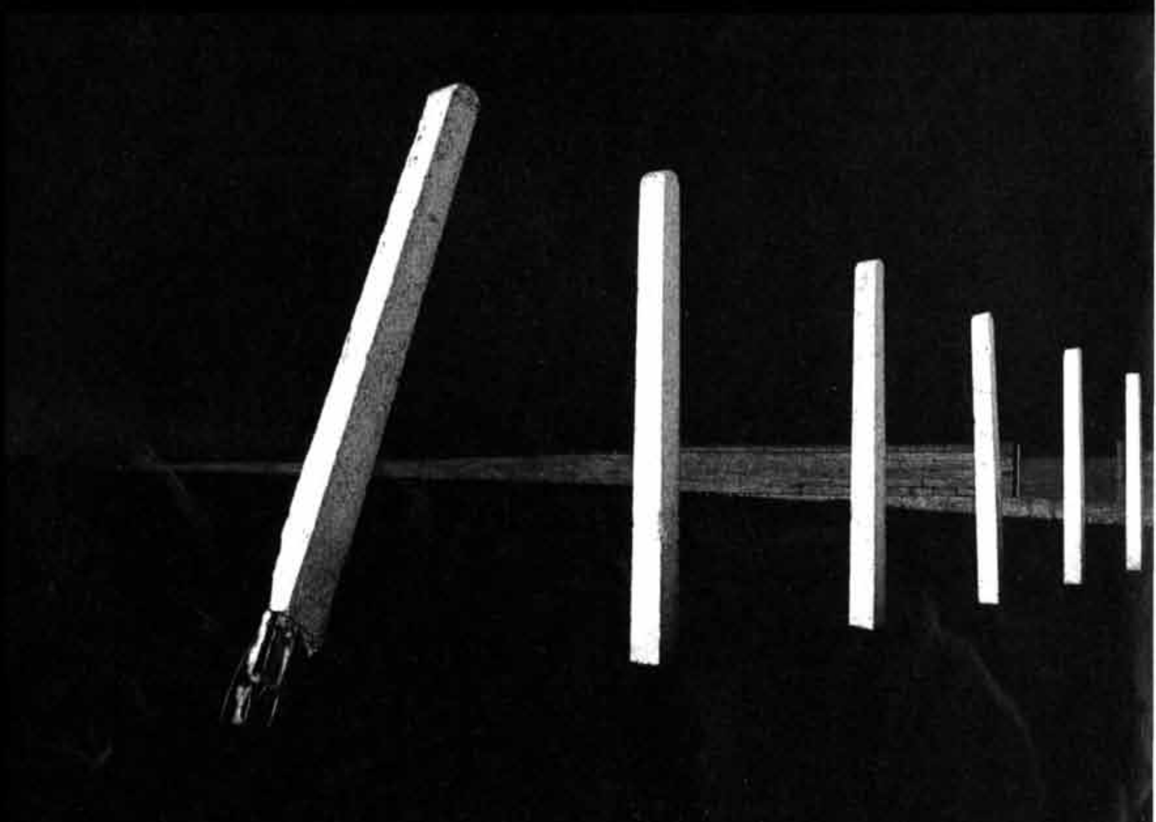
Knowing something about the structures that repeatedly appear en route to the native state would help to clarify the rules of folding. Regrettably, trapping intermediates is difficult, in part because folding is a highly cooperative process. Interactions that promote folding by one part of the protein also promote folding elsewhere in the molecule; hence, intermediate shapes do not persist for long. Nevertheless, clever techniques have captured or identified some characteristics of a number of intermediates.

There is now firm evidence, for example, that certain proteins form an intermediate that is larger than the native form of the protein and has its secondary structure intact. Oleg Ptitsyn of the Institute for Protein Chemistry at Pustshino in the Soviet Union calls this structure the "molten globule." The existence of such a structure is puzzling, however.

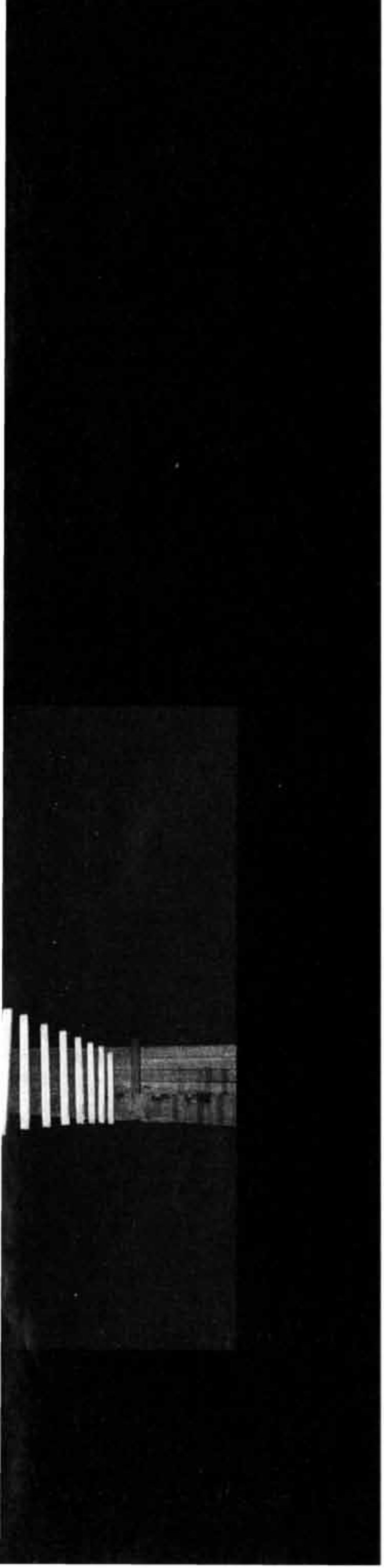
Because the globule is larger in vol-

COMMON SECONDARY STRUCTURES IN PROTEINS
(identifiable conformations adopted by segments of the backbone)





Michael Kenna, *Tilted Poles*



"The field cannot well be seen from within the field."

Ralph Waldo Emerson

To know what is possible tomorrow you must be willing to step outside of what is possible today.

The long view is all.

At Toshiba, we've been taking that view to help create a tangible future for over one hundred years.

Sustaining it now with our commitment to manufacturing in the U.S. And with a wide diversity of products of the highest quality.

All of which have helped make us one of the leading electrical and electronic manufacturers in the world.

A nice perspective to have. When you work in as many fields as we do.

In Touch with Tomorrow

TOSHIBA

Color TVs & VCRs

Computers & Printers

Copiers

CT Scanners

Digital Instrumentation

Disk Drives

Facsimile Systems

Home Appliances

Industrial Motors & Controls

Liquid Crystal Displays

Medical Tubes

MR Imaging

Nuclear Medicine

OPTO Electronic Devices

Picture Tubes

Portable Audio

Professional Video Systems

Room Air Conditioners

Semiconductors

Telecommunication Systems

Turbines & Generators

Ultrasound

Uninterruptible Power Supply

Variable Speed Drives

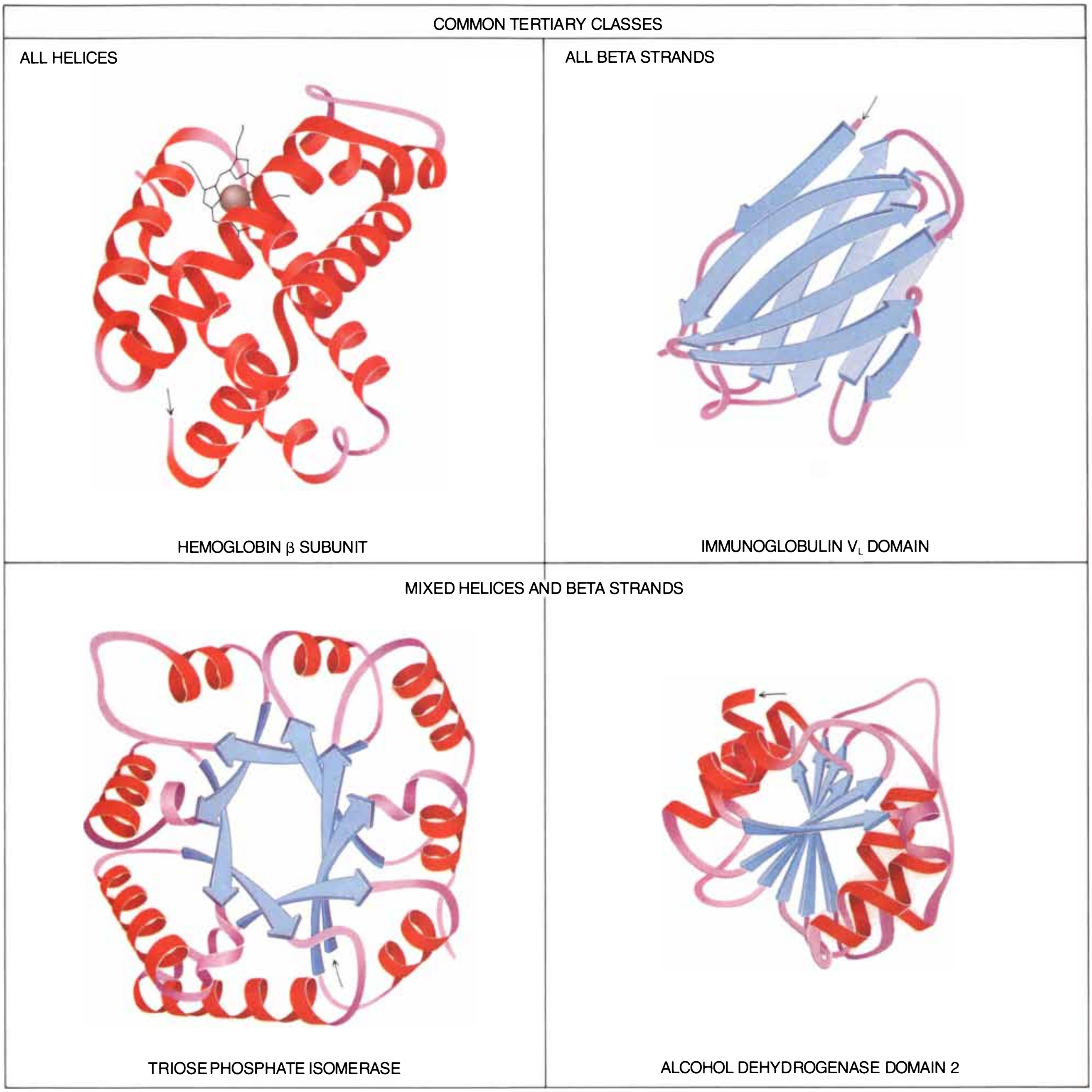
X-Ray Systems

ume than the native molecule, it must contain a considerable amount of water, and many of the side chains in the globule do seem to be in contact with water. Yet the force of the hydrophobic effect should be squeezing this water out. How can one have a stable, observable intermediate under these conditions? What can its structure actually be? These intriguing questions cannot yet be answered.

In other experiments, Thomas E. Creighton and his colleagues at the Medical Research Council Laboratory of Molecular Biology in Cambridge, England, have studied the folding of the protein pancreatic trypsin inhibitor (PTI), which, like many proteins, forms internal disulfide bonds as it folds. A disulfide bond is a sulfur-sulfur (S-S) linkage between the side chains of two cysteine amino acid residues. Creighton and his co-workers unfolded the native product and then started the folding reaction, interrupting it at different intervals. They thereby captured intermediates that could be identified by a particular disulfide bond. In this way, a folding pathway was tracked for the first time.

The complete structures of the intermediates are not yet known in detail, but the work has revealed that folding does not necessarily proceed along a single, direct track. As PTI folds, intermediates having disulfide bonds that do not exist in the final molecule appear and then disappear. In other words, parts of the molecule apparently act something like a party host who brings two well-matched strangers together and then, when the two are engrossed in conversation, leaves them and mixes with other guests.

Studying a major proposed intermediate of the same PTI protein, Kim and Terrence G. Oas, also at the Whitehead Institute, have found evidence that even though some parts of a molecule associate only transiently, other parts probably do form structures that remain stable. From Creighton's work, they knew that two specific stretches of the molecule become connected by a disulfide



bond early in the folding process and that the bond persists. They wondered whether the supersecondary structure in the region around the bond also formed early and persisted.

To answer the question, they chemically synthesized two separate fragments of the protein, each including one of the two cysteines that participate in the stable disulfide bond. The small peptides had no discernible structure of their own, but when they joined in solution, they adopted a conformation closely resembling that seen in the native chain.

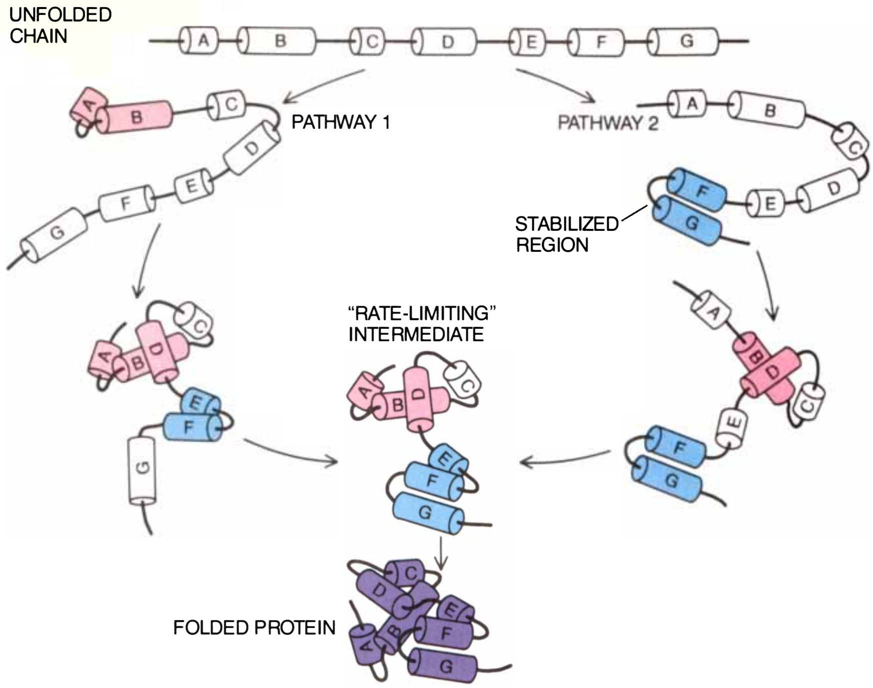
This finding confirms that nativelike structures can indeed form early, and it suggests that certain parts of the molecule may be more important than other parts in initiating folding. The result also indicates that interactions between apparently unstructured segments of a protein may facilitate the development of secondary structure.

Intermediates are being studied by another ingenious method that capitalizes on the many internal hydrogen bonds found in all native proteins. First normal hydrogen atoms bound to the nitrogen involved in peptide bonds are exchanged with a related atom—the hydrogen isotope deuterium (D)—by placing the chains in heavy water, D_2O. Then folding is initiated.

As folding proceeds, what would have been hydrogen bonds become "deuterium" bonds (N-D-O) instead. At some chosen time, normal water (H_2O) is substituted for the heavy water. When that happens, any deuterium atoms not protected by being in deuterium bonds trade places with hydrogen from the water. Folding then continues to completion.

By identifying the regions of the compacted molecule that contain protected deuterium, one can determine which parts folded before the others. Moreover, a series of tests that progressively lengthen the time of transfer to water can potentially reveal the order in which several different intermediates form.

With this technique, Heinrich Roder of the University of Pennsylvania was able to show that a first step in the folding of the protein cytochrome c is the association of two helices at opposite ends of its chain. In a study of ribonuclease, Baldwin and his Stanford colleague Jayart B. Udgaonkar showed that the beta-sheet part of that enzyme—found in the middle of the molecule—forms early. Such findings by themselves have yet to yield general rules of folding, but they do highlight the power of the method for identifying folding pathways.



PLAUSIBLE MODEL of how proteins fold allows for several energetically favorable pathways, although only two possibilities are shown. First the chain forms regions of unstable structure (*uncolored cylinders*). By associating, certain regions become stabilized (*color*). These stabilized microdomains then facilitate the association of other regions and thus lead the molecule toward increasing structural organization. Eventually, all pathways lead to one or more "rate-limiting" intermediates, which all give rise to the same final conformation for the protein.

Experimentation is not by any means limited to studies of intermediates. A number of scientists are approaching the folding problem by wielding genetic engineering technology as a tool to examine the effects of amino acid substitutions, deletions and insertions, both on protein structure and on the folding process.

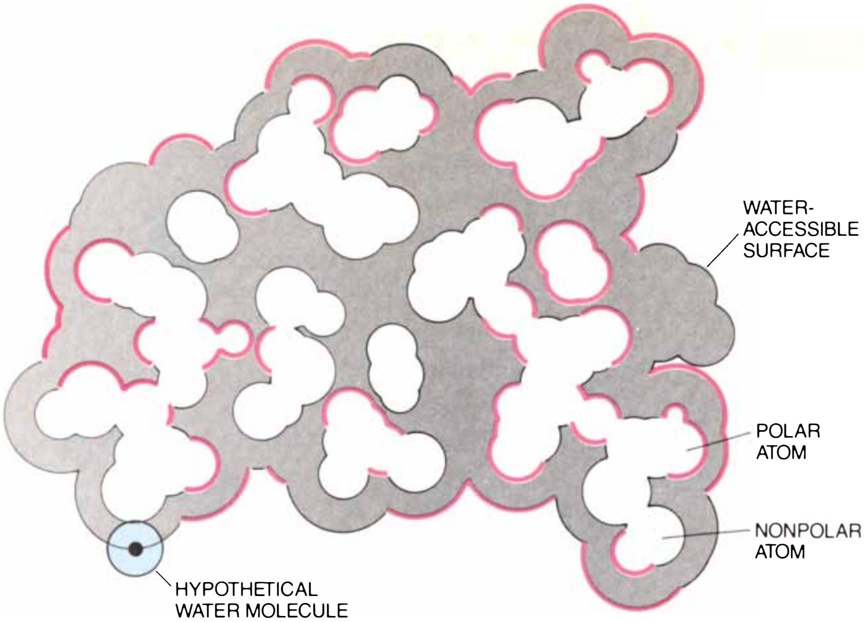
So far the experimental data, as well as computer simulations of amino acid substitutions, indicate that replacement of one or even several residues usually does not interfere with the development of proper architecture. In other words, the answer to the folding problem lies not with a few key amino acids but with some more global aspect of the amino acid sequence. In contrast, only a small part of the entire molecule may be responsible for a protein's activity. Single amino acid substitutions in the active region can dramatically affect biological behavior even when the overall structure of the protein seems unaffected.

In a different kind of experiment, Siew Peng Ho and William F. DeGrado of the du Pont Company and, separately, Jane S. and David C. Richardson of Duke University are trying to design fully original proteins that will fold into selected conformations. In this way, they are testing various hypotheses, such as the proposal that certain sequences of hydrophobic and hydrophilic amino acids are likely to form an alpha helix and then a cluster of interdigitating helices (a helix bundle).

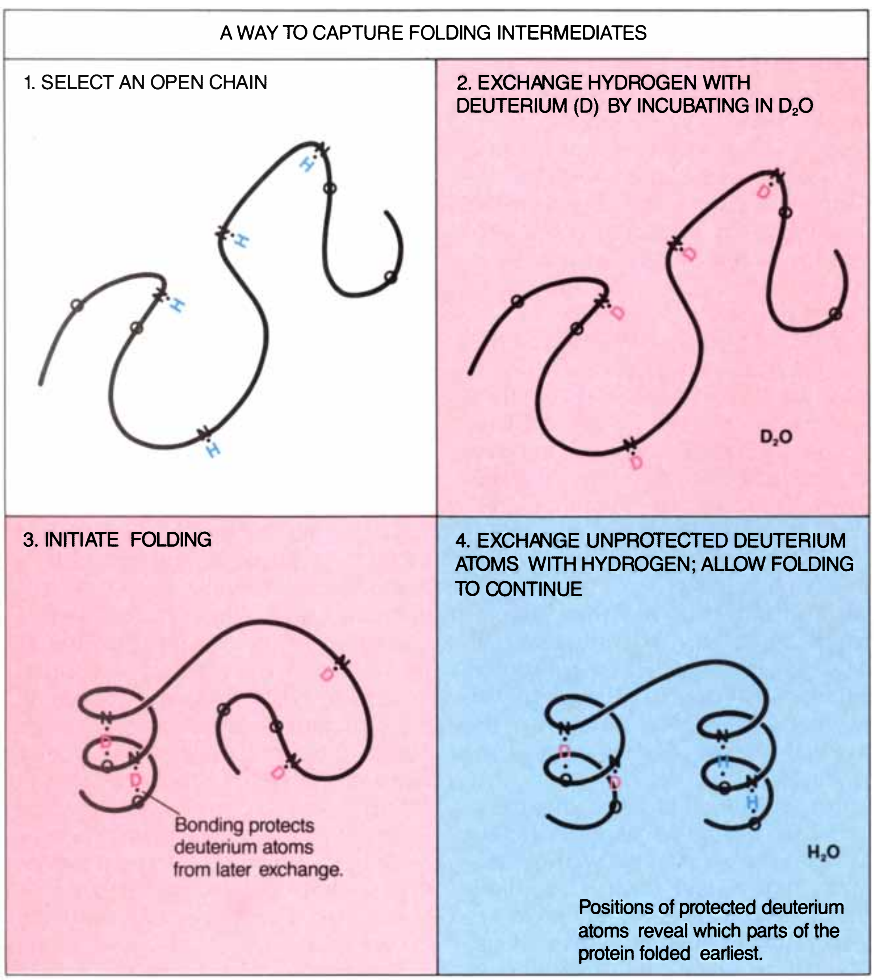
They have succeeded in making proteins of a specified architecture. Nevertheless, researchers are still far from being able to predict the tertiary structure of any given protein on the basis of its amino acid sequence if they lack other information about the substance.

Theoretical endeavors complement the experimental work. For example, the shape of a folded protein might in principle be determined by a mathematical formula known as the potential-energy function. One feeds a computer a host of numerical values that describe the strength and other aspects of the attractions between all pairs of atoms in the protein chain. Then the computer adjusts the coordinates of the atoms so that the overall energy is lowered until a minimum is found—that is, until all further changes result in an increase in energy. (The final structures of proteins are generally assumed to represent the minimum-energy state.)

The function takes into account such



CONCEPT OF WATER-ACCESSIBLE SURFACE enables one to estimate the force exerted by water on a molecule in aqueous solution. The accessible surface (*outer line*)—shown for a thin slice through a protein—is determined by tracing the path of the center of an imaginary water molecule as it rolls along the protein's external atoms. The surface is conceptually divided into water-loving (*colored line*) and water-hating (*black line*) parts, depending, respectively, on whether the atoms are polar or nonpolar. A large water-hating surface corresponds to a strong compressive force, and a large water-loving surface corresponds to a strong expansive force. The compressive force dominates when protein molecules are folding.



factors as the influence on energy of the length, stretching and twisting of bonds and the strength of electrostatic interactions, hydrogen bonds and van der Waals forces. The approach has been valuable for confirming or improving models of structures that were determined experimentally.

For molecules whose final structures are a complete mystery, however, problems arise. Certain of the numbers plugged into the equations may have large margins of error. Furthermore, there is no way to know whether the reported energy minimum represents the absolute minimum or simply an intermediate low-energy state. At the moment, theory does not provide any way of ascertaining what the absolute minimum value ought to be.

In a related approach that might eventually yield motion pictures of proteins in the act of folding, Martin Karplus of Harvard University applies Newton's laws of motion to the atoms in a protein. The forces on the atoms of a molecule in a given state are derived from the potential-energy function. Then the computer calculates the acceleration of each atom and its displacement at the end of an extremely short interval.

By repeating the process over a period controlled by the available computing power, the program can reveal movements of the individual atoms. Consequently, it is now becoming possible to identify the effects of small mutations on protein stability and dynamic behavior. Yet limits on computing power make it impossible to track more than a few nanoseconds in a molecule's life, a span too short to directly reveal much about protein folding.

In spite of its limitations, theoretical work based on the potential-energy function holds much promise. Studies involving the function should make it possible to discern the relative importance of various forces acting on a protein, such as electrostatic interactions and van der Waals repulsions. Teasing apart the influences is critical because a folded protein is only marginally more stable than an unfolded one. Hence, the factors that make the difference are likely to be subtle. (The slight energy differential between the stable and unstable state might reflect the need for a cell to inactivate proteins rapidly as its needs change.)

In the long run, calculations involving the potential-energy function may well succeed in predicting the tertiary structure of any protein from its amino acid sequence. In the meantime, other less fundamental but still useful ways of thinking about the folding problem have emerged.

Any resolution of that puzzle will have to include a way of defining the force exerted on a protein molecule by water. In principle, estimates of the hydrophobic effect are, or can be, embedded in the potential-energy function, but exactly how best to accomplish that step is far from clear.

One method for analyzing the effect of water has emerged from work done by Byungkook Lee at Yale University in 1971. Lee developed an algorithm to calculate the solvent-accessible area of a protein of known structure—that part of the complex surface in direct contact with water. On the basis of preliminary findings, he and I suggested the algorithm would be useful in studying protein folding.

We divide the accessible area of an extended protein chain (or any selected molecule) according to the nature of the atoms that contribute to the area. Are they nonpolar and therefore hydrophobic (mainly carbon and sulfur atoms), or are they polar and therefore hydrophilic (mainly nitrogen and oxygen atoms)?

The surface tension of water in contact with such atoms is known. This tension is, as Cyrus Chothia of the Medical Research Council has pointed out, a direct measure of the force exerted on the molecule by the solvent. Surface tension is high when nonpolar molecules and water are in contact, just as it is when oil is mixed with water—that is, a strong force tends to reduce the area of contact between the water and the oil, and to squeeze a protein chain into a ball. Tension is low when polar atoms and water are in contact, and the hydrophobic effect is not seen.

Summation of the nonpolar accessible areas of an unfolded chain yields a measure of the potential hydrophobic effect. In general, as might be expected from structural analyses, the net force acting on most protein chains is large and positive, tending to reduce contact with the solvent and thus to compact the chain.

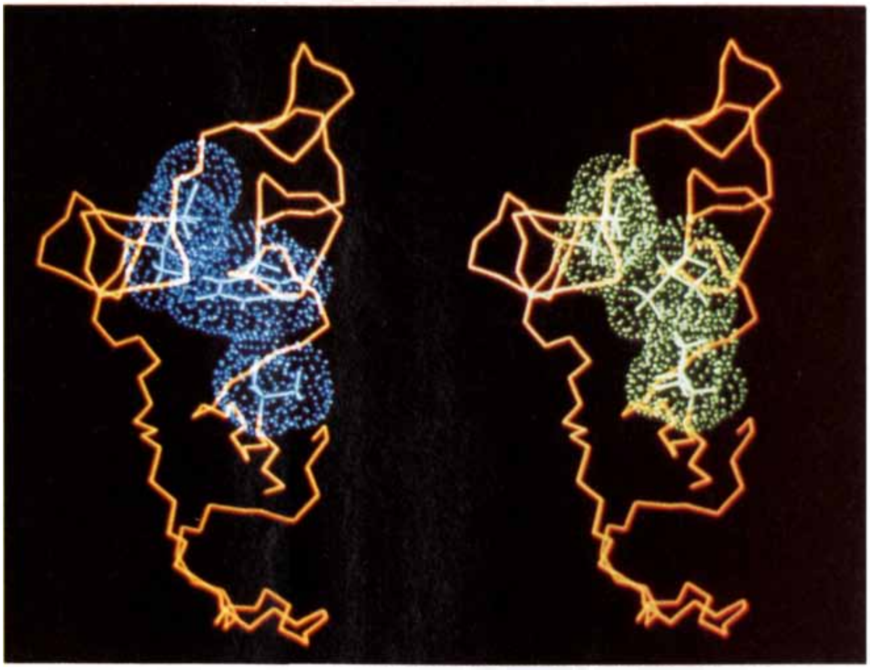
Various investigators are also examining the extent to which packing considerations direct folding. In one approach, lists have been made of the amino acid sequences of molecules that adopt essentially the same three-dimensional conformation. On the basis of the steric properties of the amino acids in the molecules—such as shape and volume—Jay W. Ponder of Yale has generated other lists of amino acid sequences that theoretically should adopt the same conformations.

Just how well those sequences actually fit their assigned classes is still being determined experimentally, but many do seem to fit. This finding, together with the profound influence of water, makes it conceivable that the hydrophobic effect and steric considerations by themselves determine how a protein folds.

If that is the case, what is the role of long- and short-range electrostatic interactions in protein folding? Undoubtedly, the contribution of such interactions varies from protein to protein. For many proteins, large changes in the formal charges can be made without significantly affecting the final overall structure. Hence, it may be that electrostatic interactions are often more important for stabilizing the final conformation than for forming it in the first place.

Determining whether this possibility is correct requires an ability to gauge the strength of electrostatic interactions. Yet the mathematics is complicated by the fact that atoms in a folding protein are often separated by water, which can mute the long-distance attractions or repulsions in ways difficult to estimate in the absence of detailed structural information. Moreover, as the protein folds, the distances between the atoms constantly change, which adds further complexity.

The precise effects of hydrophobic, steric and electrostatic interactions, then, remain a matter of conjecture. Research into protein folding, however, is proceeding enormously faster today than in the past. Those of us involved in the effort still cannot "play the music," but we are rapidly learning certain of the notes. That progress alone is heartening, as is knowing that a solution to the folding problem will resolve a question of deep scientific interest and, at the same time, have immediate application in biotechnology.



CONFORMATION of a fold in the interior of the protein crambin (*left*), depicted mainly as a chain of alpha carbons (*orange*), derives from the tight packing of five nonpolar amino acids (*blue spheres*). That conformation is maintained in a computer-generated "mutant" (*right*) even when four of the five amino acids are replaced with others. Indeed, many combinations of amino acids can be accommodated if the substitutes resemble the originals in shape and volume. Knowledge of how amino acids pack may go a long way toward predicting the shape of a protein.

FURTHER READING

PROTEIN FOLDING. Charis Ghélis and Jeannine Yon. Academic Press, 1982.

SPECIFIC INTERMEDIATES IN THE FOLDING REACTIONS OF SMALL PROTEINS AND THE MECHANISM OF PROTEIN FOLDING. Peter S. Kim and Robert L. Baldwin in *Annual Review of Biochemistry,* Vol. 51, pages 459–489; 1982.

PROTEINS, STRUCTURES AND MOLECULAR PROPERTIES. T. E. Creighton. W. H. Freeman and Company, 1983.

FOLDING AND ASSOCIATION OF PROTEINS. Rainer Jaenicke in *Progress in Biophysics & Molecular Biology,* Vol. 49, No. 1, pages 117–237; 1987.

PROTEIN FOLDING: DECIPHERING THE SECOND HALF OF THE GENETIC CODE. L. M. Gierasch and J. King. American Association for the Advancement of Science, 1990.

Building the Cathedral in Florence

The octagonal dome of Santa Maria del Fiore, crowned by a lantern, orb and cross, is architecturally unique. Building the ensemble advanced the engineering and technology of the early Renaissance

by Gustina Scaglia

In 1294 the religious and civic officials of Florence decided to enlarge the small church of Santa Reparata and dedicate it to Saint Mary. More than 100 years later Florentine citizens could see the walls of the cathedral Santa Maria del Fiore rising from the center of their city—absent a cupola, or dome. Covering, or vaulting, the open space of the structure posed a technological challenge, and by 1418 the administrators of the cathedral's workshop, called the Opera del Duomo, had reached an impasse.

The architects who had designed the cathedral over the years—notably Arnolfo di Cambio, Francesco Talenti and Giovanni di Lapo Ghini—had constructed the eight walls of the tribune where the high altar would be situated. The walls were to support a majestic cupola. But vaulting an octagon—indeed an enormous one (the dome would measure 182.5 feet in diameter)—was unprecedented. Arnolfo's model, complete with cupola, which had served all builders since his time, had been lost.

Traditional building methods and machines could not overcome the difficulties presented by this structure. One of those techniques entailed filling the interior space of the massive octagon with a forest of timber to support the vault of the stone cupola while its mortar hardened. But wood was expensive, and there were differences of opinion about the method of construction. So the Opera del Duomo and the Woolworkers' Guild—an influential guild that had been in charge of the project since 1331—announced an architectural competition, as was customary for guilds in the republic of Florence.

In 1420 two master architects were chosen. Filippo di Ser Brunelleschi (1377–1446), a Florentine sculptor and goldsmith, had described but not yet presented a model for a novel if not revolutionary means to construct the cupola. His arch rival, Lorenzo Ghiberti (1378–1455), another sculptor and goldsmith, was jointly appointed. Ghiberti was popular among leaders of the community and of the guilds because he had won the commission to execute the bronze doors of the Baptistery in front of the cathedral in 1402. Brunelleschi had lost that commission.

Brunelleschi proposed the construction of a double-vaulted cupola, one having an inner and an outer shell—a design that has never been duplicated since. This cupola would be built without using costly centering (the temporary wooden structure designed to support a vault while it is being constructed). His innovation would make each level of construction strong enough to support itself while the workmen built the next level. To achieve this new and safe technique, Brunelleschi had invented various machines such as hoists and cranes with load positioners.

GUSTINA SCAGLIA has studied drawings of Renaissance machines designed by Filippo di Ser Brunelleschi, Leonardo da Vinci and Francesco di Giorgio Martini. Until recently, she taught history of art and archaeology at Queens College of the City University of New York. Born in Glastonbury, Conn., she received a Ph.D. in 1958 from the Institute of Fine Arts at New York University. After rediscovering drawings of Brunelleschi's inventions in 1955, Scaglia wrote her dissertation and later a book analyzing them. She has recently completed a book, to be published by Lehigh University Press, about the engineering drawings and treatises of Francesco di Giorgio Martini (1439–1501). He is the Sienese engineer who will be featured in the celebration this year of the 750th anniversary of the University of Siena.

Brunelleschi desired to be the sole architect of the project since he alone knew how to do it. He devised a scheme to eliminate Ghiberti's role in the construction of the cupola. According to the biographer of the Renaissance artists, Giorgio Vasari (1511-1574), who based his information on a biography of Brunelleschi written by Antonio Manetti, Brunelleschi feigned illness, leaving Ghiberti alone to direct the workmen. Ghiberti did not know how to proceed, so work came to a halt until Brunelleschi returned to the site, having more than adequately established his authority.

Ironically, the most detailed drawings of Brunelleschi's machines for the work come from the hand of Ghiberti's grandson, Buonaccorso Ghiberti (1451-1516). In 1955 I found his sketchbook preserved in the National Library in Florence. These records, in conjunction with the published archives of the Opera del Duomo (kept since 1293), which include the names of workers, administrators and the dates and amounts of payments for labor and supplies, provide a relatively clear picture of the work in progress and of Brunelleschi's role. Using these graphic and literary sources, I deciphered how Brunelleschi used the machines to build the two vaults and the lantern and how later workers positioned the 3,276-pound gilded orb and the cross on the tip of the lantern.

Brunelleschi—considered to be the principal founder of Renaissance architecture—reached into the architecture of the Roman past for elements of the classical orders (Doric, Ionic and Corinthian), thereby altering the design of later buildings. He also applied contemporary principles of mathematics to derive a system of architectural proportions that was scaled to human dimensions. Brunelleschi's technological accomplishments, including his innovative machines and subsequent designs of new buildings, advanced the architectural methods of the Renaissance.

SANTA MARIA DEL FIORE towers over the city of Florence. The cathedral was built in stages between 1293 and 1470.

PANTHEON in Rome was one of the buildings that Filippo di Ser Brunelleschi, architect of the cupola of Santa Maria del Fiore, studied. It has a circular dome and an oculus, or eye, at the apex, the only source of natural light for the interior.

Brunelleschi overcame the technological constraints that troubled some of his contemporaries. Vasari describes him as fascinated by "time, movement, weights and wheels, how wheels can be turned and moved, so that he built some very good and beautiful clocks." He also executed paintings that illustrate perspective—introducing the concept of the vanishing point in his famous painting of the Baptistery. In so doing, he revolutionized some of Renaissance painting by introducing the illusion of three-dimensionality for objects portrayed on a flat surface. Brunelleschi's use of proportions and geometry was influenced by the Florentine mathematician and astronomer Paolo del Pozzo Toscanelli (1397–1482).

In addition to his knowledge of mechanics and mathematics, Brunelleschi spent some time in Rome measuring and studying the ruins. These structures provided him with material evidence of the Roman techniques for the construction of great vaults. According to Vasari, Brunelleschi had explored the foundations of every ancient building in Rome before 1428, noting the means of binding bricks and of cutting stones and hoisting them, which he could discern by their grooves and marks. Some features of the cupola of Santa Maria del Fiore are the result of his discoveries. In particular, his studies helped him resolve the challenge of vaulting the cathedral.

At the time of the contest for the design of the cupola in 1418, Brunelleschi's idea was controversial because the method of construction he proposed had no precedent. All extant cupolas were small, semicircular domes, such as those of the cathedrals of Pisa and Siena or the Pantheon of Rome. The Pantheon has a single-shell dome with an oculus, or opening, at the crown, providing the only source of natural light to its interior. This dome and the Roman vaults in ruins may have guided Brunelleschi toward his design. The Florentines finally allowed Brunelleschi to begin work, but the Opera del Duomo's records of its procedural meetings indicate that its members remained cautious.

Vaulting without centering was perhaps Brunelleschi's most famous achievement. The Gothic arches of Santa Maria del Fiore's octagonal tribune (which are pointed as opposed to the semicircular Roman arch) were not built as high as those of the cupola. In addition, those of the central tribune had utilized wooden supports. For the cupola, however, Brunelleschi determined that some bricks could be placed at angles to others, providing a structural backbone that could absorb the weight of the rising vault. The weight of each new layer of bricks was diffused down into the piers and walls of the octagon.

Brunelleschi also invented a system of stone chains to hold the vaults together. These stones, joined at the ribs of both vaults, are reinforced by metal cross pins. The enlocked stones bind each stratum of the vaults horizontally and vertically, so they strengthen the construction. Without them, the ribs would push outward, rupturing the cupola.

Brunelleschi incorporated in his model other practical features. Rainwater spouts were built on the exterior, and openings in the outer vaults dissipated the forces of the wind and could have helped to withstand the stress of earthquakes. He integrated iron bars that could support scaffolding for artists who would paint frescoes or arrange mosaics. (Vasari wrote that Brunelleschi installed a kitchen on the upper vaults and scaffolding so the workmen would not lose time descending to street level to eat.)

He designed the outer vault to protect the inner, stronger structure from the assault of wind and rain. The space between the shells allowed for a path and a stairway, which provided the workmen with access to build both shells and keep them in repair.

At the apex of the cupola, where the eight vertical ribs between the inner and outer vaults converge, Brunelleschi designed a circular stone enclosure, called a seraglio. The seraglio surrounds the walls of the oculus, which measures 20 feet in diameter and is 12 to 15 feet deep. The oculus serves as the keystone for the eight arches of the cupola. Each of the eight sides of the oculus has three windows, allowing light and air into the seraglio.

The windows served another purpose as well, illustrating the genius with which Brunelleschi conceived design and function. Long wooden beams could be lain through the windows across the oculus. When enough beams were inserted there, they formed a temporary floor on which cranes could stand during the construction of the marble lantern and the installation of the orb and the cross. All the materials needed for the lantern were hoisted up through an opening in the platform.

Although the lantern was Brunelleschi's critical element for the double-shell cupola designed in 1418, a competition for its design was announced only when the cupola was completed. In 1436 Brunelleschi's model won. That model is preserved today. The eight buttresses of the lantern work to thrust the weight onto the ribs of the cupola, then downward vertically on the eight great piers of the octagon.

Brunelleschi's political deftness had a counterpart in his architectural design. One of the eight columns of the lantern is hollow and contains a steep staircase providing access to the cone, the structure that supports the orb and the cross. When Brunelleschi presented his model of the lantern to the authorities, he kept the stairway carefully hidden, unveiling it only when the judges wondered how the workers would reach the top to build the cone. Cathedrals in Italy are often crowned with an orb and a cross, symbols of a cardinal's domain and of Saint Peter's in Rome.

To achieve each innovative aspect of Santa Maria del Fiore, Brunelleschi invented machines capable of lifting huge weights great distances so that they accurately reached the workers who would fit them into place. The devices, including forms of a hoist, cranes and load positioners, as well as tools such as specially designed hooks, are evidence of his knowledge of mechanics and his visionary approach to construction. The machines were also safe and allowed the work to be completed relatively quickly.

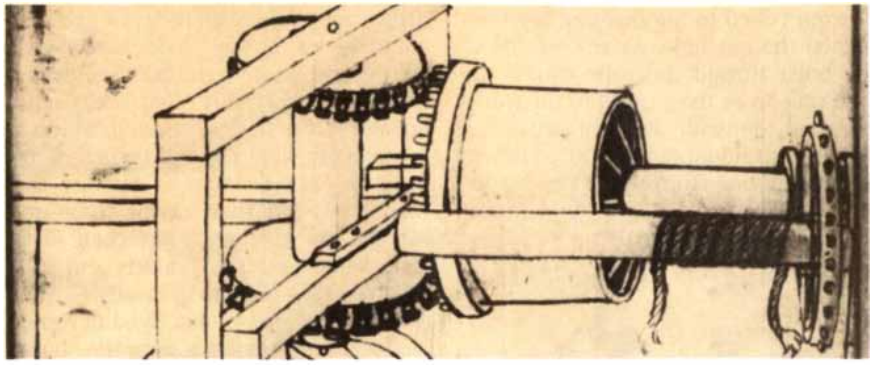
Because a competitive artistic spirit then infused Florence, Brunelleschi was very secretive about his model for the cupola and about his machines. He feared that his ideas would be pirated, and consequently he never committed them to paper (patents were unknown in Florence and just coming into use in Venice). Indeed, he went so far as to have each piece of a hoist made by a different carpenter, ironsmith, foundryman or other artisan. Each craftsman lived outside of Florence because Brunelleschi was concerned about the shifting allegiances and rivalries within the city. He gave an artisan only a simple measured profile drawing of a part; he assembled the parts to form the whole machine himself.

Brunelleschi's machines were integral to his techniques for constructing the cupola. The hoist clearly demonstrates Brunelleschi's masterly command of mechanics and his skill as an inventor. Some parts of the hoist were unlike anything that had been built previously, and the crane and the load positioner were entirely new.

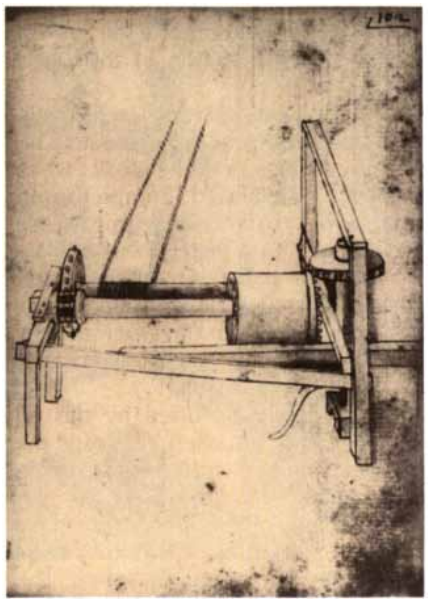
The hoist [*see top and middle left illustrations on this page*] was embedded or fixed in the ground in the center of the tribune, directly under the oculus. There it remained for some 50 years as first the cupola, then the lantern and finally the orb and the cross rose above it. The hoist raised building materials—stone, marble blocks, mortar and long chestnut beams—to the workers.

The machine was relatively simple to operate. One end of the rope was attached to a winch drum. When the load of stone, beams or masonry was ready for hoisting, it was attached to the other end of the rope by means of stone

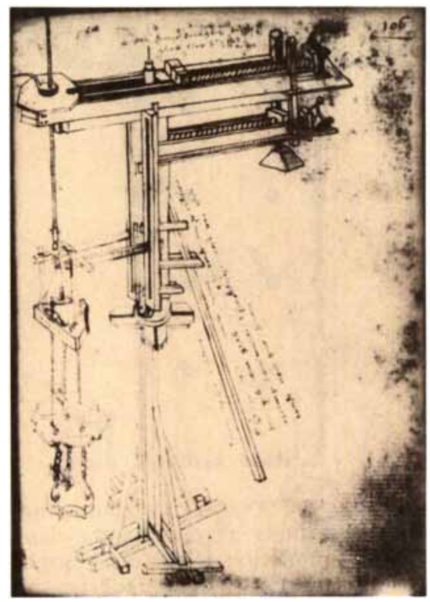
BRUNELLESCHI'S MACHINES DRAWN BY BUONACCORSO GHIBERTI



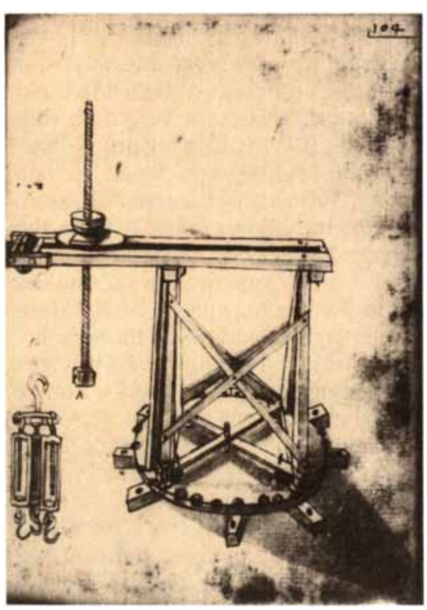
REVERSIBLE HOIST (DETAIL)



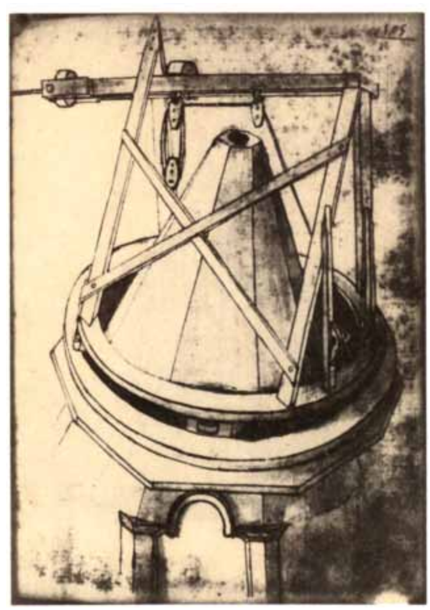
REVERSIBLE HOIST (ANOTHER VIEW)



CRANE

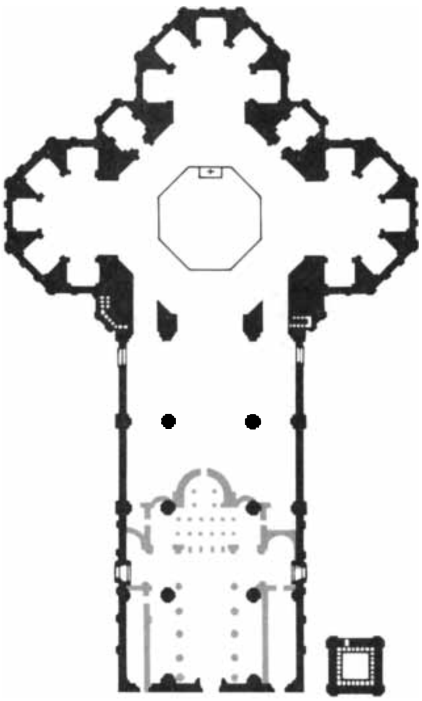


LANTERN-BUILDING CRANE

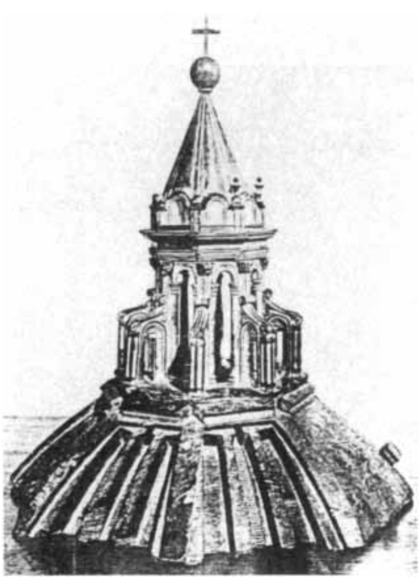


CONE-BUILDING CRANE

hangers (a Roman device similar to stirrups that entered the stone, which was rediscovered by Brunelleschi). Horses or oxen yoked to the pivoting bar then turned the machine. As the wheels of the hoist turned the rope drum, the load rose from the ground to the workplace on the vault. When it arrived, a signal was undoubtedly shouted down, and the driving stopped.



CATHEDRAL FLOOR PLAN shows the octagonal tribune surrounding the altar and the foundation of the old church of Santa Reparata buried underneath.



LANTERN MODEL was designed by Brunelleschi. The lantern's weight falls on the eight ribs of the cupola, thereby strengthening the entire structure.

Brunelleschi ingeniously designed the wooden hoist to be reversible. The hoist had two wheels, one above the other; one could raise the rope and one could lower it. The wooden screw in the central post of the hoist could be adjusted so that only one wheel would engage at a time. This reversible clutch system allowed the work animal to walk in one direction to operate the machine rather than having to be unhitched, reversed and reattached. (The rope weighed 1,106 pounds and was made by shipbuilders in Pisa.)

Brunelleschi also conceived new mechanical components for the hoist called *palei*—rotating wooden rollers fitted on U-shaped irons—which acted as low-friction gear teeth. There were 91 *palei* on the main wheel. These rollers may have reduced the loss of energy through friction in the wheel drive.

For his machines, Brunelleschi adapted the stone hangers, or lewises, that fitted into sockets in stone blocks when they were transported through the air. Turnbuckles with hooks gripped these lewises and lifted them to ensure safe hoisting. The iron lewis was formed of three loose units of wedge-shaped irons that fit into a slot in the stone so they could not fall out, just like a mortise and tenon.

Once a load was lifted by the reversible hoist, Brunelleschi's crane positioned it on the work site. This crane [*see middle right illustration on preceding page*] rotated, and its carriage transferred materials where they were needed on the rising masonry. The load positioner slid the material horizontally while a screw descending from its far end lowered the load vertically near its intended position. Both the crane and its load positioner were coordinated with the reversible hoist.

Another crane was designed to build the lantern. This crane rested on rollers under its circular platform, allowing it to pivot like a lazy Susan [*see bottom left illustration on preceding page*]. The lantern-building crane includes the carriage of Brunelleschi's load positioner. The entire apparatus was installed across the oculus, and its beams were inserted in the windows of the seraglio.

Some understanding of the way the lantern-building crane works comes from the notes of Leonardo da Vinci, who copied all of Buonaccorso's drawings of Brunelleschi's design. Leonardo describes it as having "four wooden screws that raise this scaffold, and once it is raised a strong platform can be built under it."

Yet another crane was created to build the cone. Its design can be definitively attributed to Brunelleschi because the lantern is his concept, and he would have prepared every stage of the work. Moreover, Brunelleschi's model of the lantern includes the orb and the cross, so it follows that he would also have designed the devices by which they would be installed.

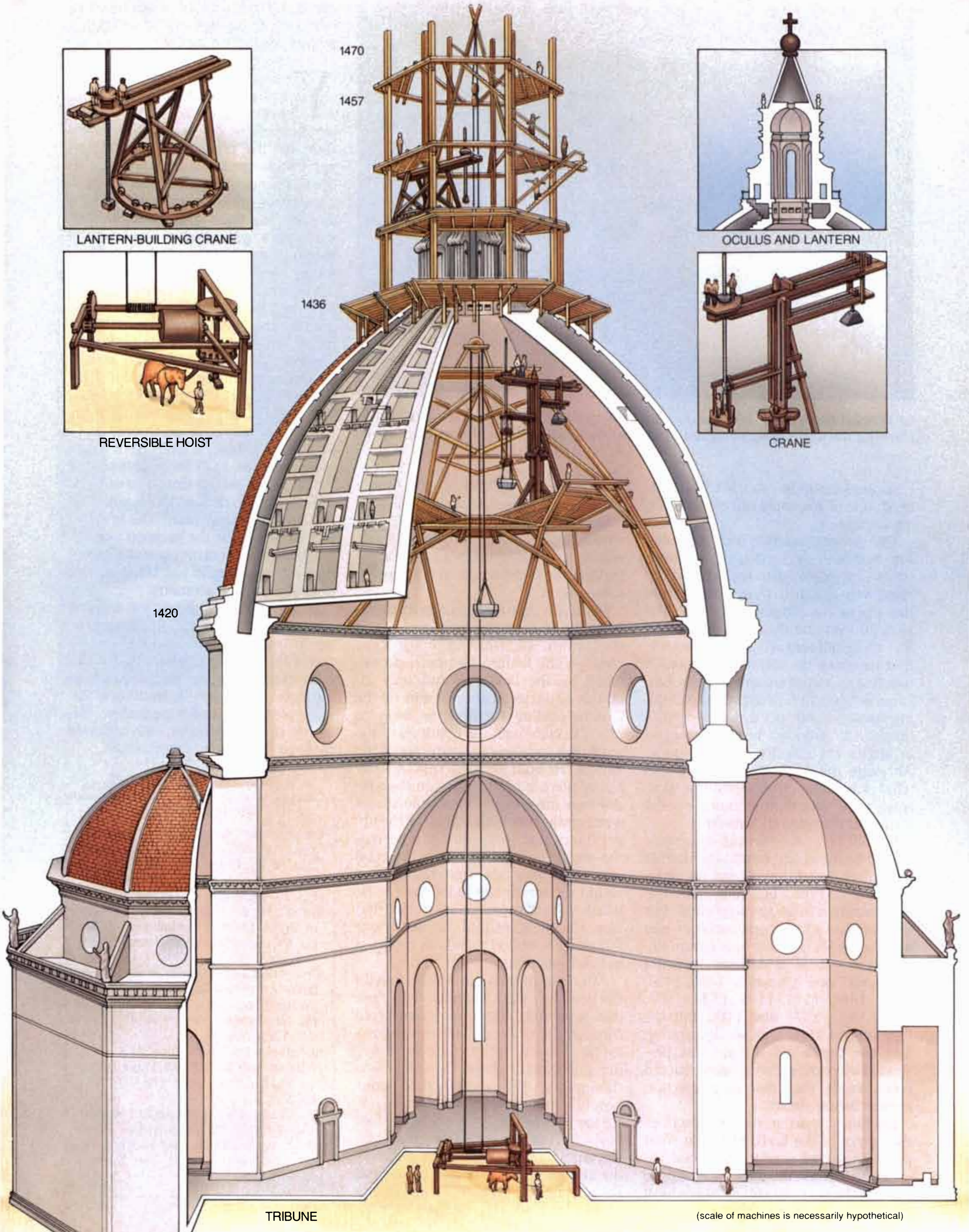
The carriage of the cone-building crane consists of a horizontal trolley that suspends a rope pulley system. Buonaccorso's drawing [*see bottom right illustration on preceding page*] shows only the basic components of the cone-building crane, without the massive scaffold needed to secure it in place and without any indication of how the circular platform could rotate. His drawing, however, does show the entire apparatus erected on the protruding marble cornice of the lantern. It also depicts the opening in the top of the cone into which the orb and the cross would be affixed and held by four wooden beams inside.

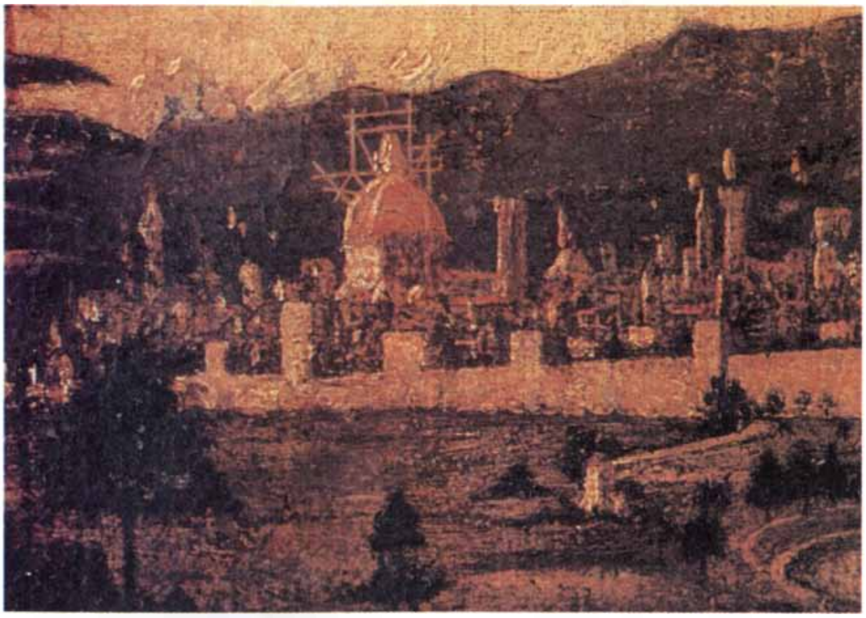
When Brunelleschi died in 1446, the lantern was not completed, and the orb and the cross had yet to be installed. Because there are so few written documents and illustrations from which to deduce Brunelleschi's techniques, the Opera del Duomo's sketchy records about the arduous task of installing the lantern and the orb and the cross—and reinstalling them after a natural disaster—become important for reconstructing the artist's original method. These descriptions supplement Buonaccorso's and Leonardo's drawings.

In 1468 the Florentine sculptor Andrea del Verrocchio (1435–1488) was awarded the commission for the orb and its copper framework. (The cross was made by other artisans, who fitted it on the orb at a later stage.) The Opera del Duomo specified that Verrocchio should make the orb in eight sections with a bronze neckpiece. The central pin of the neckpiece would fit into the hole in the hollow cone—following Brunelleschi's model of the lantern. Later that year, after the eight sections were soldered together outside the lantern on its nine-foot-wide terrace, the seven-foot orb was lifted into place by a hoist described in the documents as pyramidal. Its shape recalls that of the cone-building crane.

Although there are no further details, the apparatus described in Buonaccorso's drawing of the cone-building crane must have been used to transfer the pieces of the orb laterally once it was lifted to the right height. (As a point of interest, Leonardo must have witnessed this installation, since he was appren-

A PROPOSED VIEW OF BRUNELLESCHI'S MACHINES IN OPERATION





GIOVANNI BATTISTA UTILI painted the Florence Cathedral around 1470, clearly showing the scaffolding designed by Brunelleschi spreading out over the cupola.

ticed to Verrocchio at that time. Indeed, one of his notes refers to soldering a sphere.)

The wooden scaffold that Brunelleschi had built to construct the lantern was still in place when the orb and the cross were installed. (A document from the Opera del Duomo says that the structure was costly, difficult to make and to install and would have no further use once the orb was in place.) A painting executed around 1470 by Giovanni Battista Utili depicts the scaffolding spreading out beyond the cupola's terrace, its wooden beams arranged at angles set directly on the cupola. Although the whole scaffold is somewhat schematic, Utili's painting portrays what was then standing on the terrace and around the lantern.

In 1602 Gherardo Mechini designed a similar scaffold and rebuilt the marble cone after it and the orb and the cross were toppled by a lightning bolt. The bolt smashed them to the ground, but the sections of the orb survived and were reused. Lightning was not an uncommon source of damage to Santa Maria del Fiore. It struck in 1492, 1494, 1495, 1498, 1511, 1536, 1542, 1561, 1570, 1577, 1578 and 1586. But the lightning in 1600 was so devastating that the Opera del Duomo was persuaded by papal decree to place sacred relics inside the cross as protection against future storms.

Mechini's written report describes the height of the scaffold and at what point the cone was to be rebuilt in order to install the orb. "The upper floor of the scaffold comes to the level of the orb, so that to set the orb in place it will be necessary to build another floor about 24 feet high," he writes. In all probability, Brunelleschi would have built his scaffold for installing the lantern and the orb to the same height.

Mechini's drawing also attests to the function of Brunelleschi's oculus and the lantern. Since three quarters of the cone on the lantern had been demolished by the lightning and only 20 of the 32 steps in the column of the lantern leading to the cone were in place, Mechini built the first level of the scaffold to rest on the buttresses of the lantern. He built a floor across the oculus by placing beams through the windows in the seraglio. A simple tripod winch crane was built on one of the upper floors, and an opening in it permitted small marble blocks for the cone and some long chestnut beams that would secure the orb in its place to be lifted from the pavement to this terrace. Mechini's scaffold was hardly different from Brunelleschi's as depicted in Utili's painting.

After the artisans in 1602 wrought the new orb's eight sections, each piece was separately lifted and transferred through the lantern's narrow windows to the terrace, where they were soldered to form a sphere. The sphere was then gilded. The workers were somehow able to house the fire used in soldering so that the scaffold would not burn, but there were some casualties. According to the Opera del Duomo's documents, eight workmen died from the effects of the mercury they used to gild it. (Mercury is commonly used to bind gold to the underlying metal and is then evaporated by fire.)

Years after his death, Brunelleschi's lantern continued to yield surprises. In 1511 the administrators of the Opera del Duomo authorized the creation of a one-inch aperture framed by a bronze ring. The opening would permit the sun's rays to sometimes fall on a gnomon, or sundial, on the floor of the tribune. The sun's rays generated a shadow, the length or the position of which indicated the summer and winter solstices.

This astronomical feature of the lantern and the gnomon were rediscovered by the Jesuit astronomer P. Leonardo Ximenes in 1755. He wished to observe the summer and winter solstices and recalibrated the gnomon. Until recently, historians believed the aperture was his creation. But in 1979 the archivist of the Opera del Duomo, Enzo Settesoldi, published a record of a payment made in 1475 for a bronze ring "to be placed on the lantern in order to see where the rays of the sun fall on certain days of the year." The inventor of this device for the gnomon was the Florentine mathematician and astronomer Toscanelli, who had inspired Brunelleschi to learn geometry.

Brunelleschi's mathematical and mechanical abilities made him a forerunner in modern structural design. Although his immediate successors who built domes did not use his machines but rather returned to traditional devices of winches and rope pulleys, elements of his machines foreshadowed those of the Industrial Revolution.

FURTHER READING

DRAWINGS OF BRUNELLESCHI'S MECHANICAL INVENTIONS FOR THE CONSTRUCTION OF THE CUPOLA. Gustina Scaglia in *Marsyas: Studies in the History of Art*, Vol. 10, pages 45–68; 1960–61.

BRUNELLESCHI'S CLOCK? Frank D. Prager in *Physis. Rivista Internazionale di Storia della Scienza*, Vol. 10, fasc. 3, pages 203–216; 1968.

BRUNELLESCHI'S DOME OF S. MARIA DEL FIORE AND SOME RELATED STRUCTURE. Rowland Mainstone in *Transactions of the Newcomen Society*, Vol. 42, pages 107–126; 1969–70.

BRUNELLESCHI: STUDIES OF HIS TECHNOLOGY AND INVENTIONS. Frank D. Prager and Gustina Scaglia. The MIT Press, 1970.

BRUNELLESCHI'S HORIZONTAL ARCHES AND RELATED DEVICES. Thomas B. Settle in *Annali dell'Istituto e Museo di Storia della Scienza*, Vol. 3, fasc. 1, pages 65–80; 1978.

INTRODUCING THE GREAT AMERICAN BEAUTY.

The 1991 Buick Park Avenue.

The elegant, highly aerodynamic shape of this new Buick Park Avenue is certain to make a beautiful first impression.

But its true beauty is also apparent in Park Avenue's highly refined engineering. To quote *Motor Trend:* "Buick engineers have come up with one of the quietest, tightest and smoothest operating luxury sedans we've ever driven."

EPA estimated miles per gallon:
18/27
City Hwy.

As you'd expect, all Park Avenue's accomplishments are underscored by Buick quality.

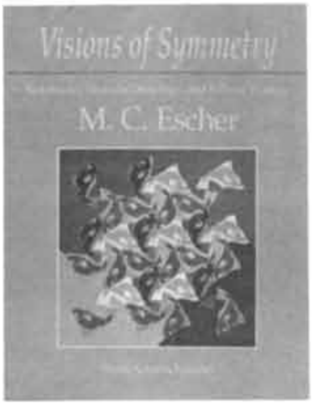
We invite you to visit your Buick dealer today. Test-drive Buick quality as evidenced in the all-new 1991 Park Avenue.

For more information, please call 1-800-3PARK-AVE.

BUICK®

The New Symbol For Quality In America.

SAVE 20%—An Exclusive Offer

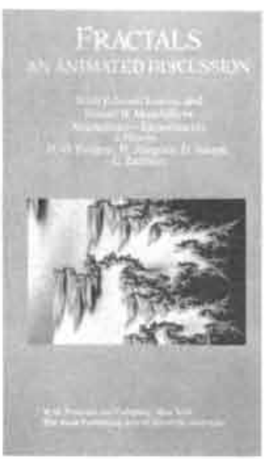


VISIONS OF SYMMETRY
Notebooks, Periodic Drawings, and Related Work of M. C. Escher

Doris Schattschneider

Visions of Symmetry is a visual and intellectual presentation of M. C. Escher's theory of symmetry and the interplay of art and science. Ideal for Escher enthusiasts, the book features the artist's complete set of symmetry drawings—about 80 of which have never before been published—reproduced in full color. It also includes the two notebooks in which Escher outlined his theory of symmetry, providing a fascinating glimpse into the creative mind of this renowned graphic artist.

352 pages, 370 illustrations
Cloth: 2126-0 $39.95 SPECIAL PRICE $31.99



FRACTALS:
An Animated Discussion with Edward Lorenz and Benoit B. Mandelbrot

H. O. Peitgen, H. Jürgens, D. Saupe, and C. Zahlten

A visual exploration of fractals and contemporary mathematics, this new 63-minute video is a clever combination of full-color animated sequences and intriguing interviews. The film turns the Mandelbrot set and the Lorenz attractor into visible and easily comprehensible objects as their discoverers, Benoit B. Mandelbrot and Edward Lorenz, discuss the background, history, and details of their work. It also features new computer-graphic illustrations of chaos and self-similarity.

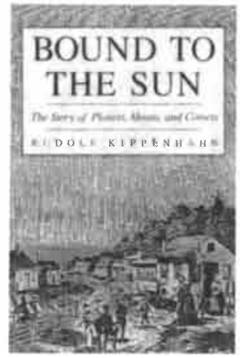
63-minute VHS video
2213-5 $59.95 SPECIAL PRICE $47.95

AARON'S CODE
Meta-Art, Artificial Intelligence, and the Work of Harold Cohen

Pamela McCorduck

Aaron's Code tells the story of Aaron, a computer program created by artist Harold Cohen. Aaron makes creative drawings autonomously, without human intervention, in the first profound connection between art and computer technology. Includes 16 pages of color illustrations.

225 pages, 77 illustrations plus color inserts
Cloth: 2173-2 $24.95
SPECIAL PRICE $19.95



BOUND TO THE SUN
The Story of Planets, Moons, and Comets

Rudolf Kippenhahn
Translated by Storm Dunlop

With a storyteller's charm, the former director of the Max Planck Institute for Astrophysics tells the fascinating story of how the outer world of planets, moons, and comets was discovered. Kippenhahn describes how early astronomical observations led to strange beliefs and practices, presents an interview with Galileo, and offers a modern exploration of the planetary system through space probes, high-powered telescopes, and information from the recent Voyager II flight.

304 pages, 104 illustrations
Cloth: 2124-4 $19.95
SPECIAL PRICE $15.95

THE EPICUREAN LABORATORY
Exploring the Science of Cooking

Tina L. Seelig

The art of cooking is also a fascinating science, whose principles explain why gelatin works in a salmon mousse; why bread rises; and why the tongue burns from red peppers. Each principle discussed is followed by a mouth-watering, kitchen-tested recipe that demonstrates the chemistry behind the dish. Includes 22 recipes.

163 pages, 120 illustrations
Paper: 2162-7 $15.95
SPECIAL PRICE $12.95

SCHIZOPHRENIA GENESIS
The Origins of Madness

Irving I. Gottesman

"For anyone who has ever had an interest in understanding schizophrenia, [this] is *the* definitive source book. It is a clearly written book that is, nevertheless, so packed with information about all aspects of schizophrenia ...that readers will find themselves going back again and again to absorb the full fare that is offered."—L. Erlenmeyer-Kimling, Columbia University

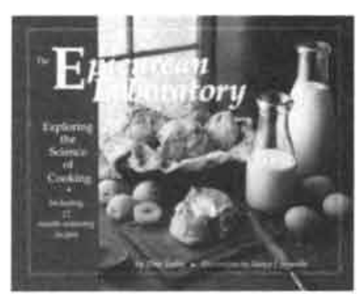
297 pages, 18 illustrations
Cloth: 2145-7 $24.95
SPECIAL PRICE $19.95
Paper: 2147-3 $14.95
SPECIAL PRICE $11.95

MIND SIGHTS
Original Visual Illusions, Ambiguities, and Other Anomalies, with a Commentary on the Play of Mind in Perception and Art

Roger N. Shepard

Part artist's portfolio and part essay on perception and the psychology of art, *Mind Sights* proves both stimulating to the eye and provocative to the mind. Renowned psychologist Roger N. Shepard introduces his wonderfully original drawings of visual tricks (many never before published), discusses the origin of his scientific and artistic work, and shares his reflections on the nature of art, perception, and the mind.

228 pages, 100 illustrations
Cloth: 2134-1 $24.95
SPECIAL PRICE $19.99
Paper: 2133-3 $14.95
SPECIAL PRICE $11.95



to Scientific American Readers

OUR UNIVERSE
An Armchair Guide

Michael Rowan-Robinson

From one of Britain's leading astronomers comes a pictorial survey of all that is known about the universe today. Twenty famous astronomical objects, about half of which can be seen with the naked eye, illustrate the full range of theories and insights of modern astronomy linking the familiar night sky with the most up-to-date ideas and discoveries at the frontier of research.

192 pages, 328 illustrations
Cloth: 2156-2 $24.95
SPECIAL PRICE $19.95

ISLANDS OF TRUTH
A Mathematical Mystery Cruise

Ivars Peterson

Ivars Peterson presents another exciting itinerary of Mathland, introducing strange vibrations in the shadows of chaos, new twists in knot physics, and the straight side of circles. His engaging exploration—from astonishing applications of number theory to new developments in fractal geometry—gives readers renewed appreciation for the dynamic world of modern mathematics.

336 pages, 119 illustrations
Cloth: 2113-9 $19.95
SPECIAL PRICE $15.95

KEEP YOUR EYE ON THE BALL
The Science and Folklore of Baseball

Robert G. Watts and A. Terry Bahill

Could Sandy Koufax's curve really have acted like it "fell off a table?" Why does a well-pitched knuckle ball silence so many great bats? Were the hitters of yesterday really better than those of today? In *Keep Your Eye on the Ball*, engineers Watts and Bahill put some of baseball's most cherished myths to the test of scientific scrutiny. The result is a highly informative and entertaining guide that provides a scientific slant on some classic baseball conundrums.

224 pages, 23 illustrations
Cloth: 2104-X $18.95
SPECIAL PRICE $15.95

ECOLOGY
Third Edition

Robert E. Ricklefs

Lively, comprehensive, and clear, *Ecology* draws examples from vertebrates, invertebrates, plants, and microorganisms and from marine, freshwater, and terrestrial environments, to reflect the latest findings in the field. "Ricklefs writes with authority on the broadest range of ecologically relevant subject matter so far published in one book."—*Ecology*, on the first edition

912 pages, 520 illustrations
Cloth: 2077-9 $44.95
SPECIAL PRICE $35.95

MOLECULAR CELL BIOLOGY
Second Edition

James Darnell, Harvey Lodish, and David Baltimore

From three pioneers in molecular and cellular research comes the lavishly revised second edition of this groundbreaking book, the first volume to unify the presentation of cell biology, genetics, and biochemistry into one cohesive framework. The new edition incorporates important late-breaking research and stunning new full-color art.

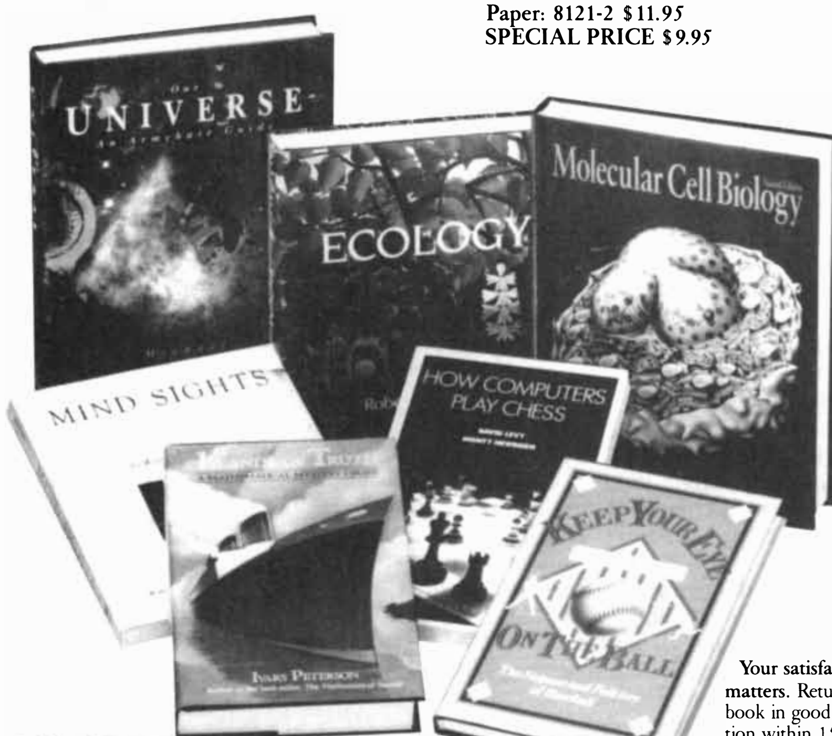
Scientific American Books
1100 pages, 1095 illustrations
Cloth: 1981-9 $54.95
SPECIAL PRICE $43.95

HOW COMPUTERS PLAY CHESS

David Levy and Monroe Newborn

Computer chess is more intriguing than ever, following the much-touted playoff between Kasparov and the Deep Thought program. Here, Levy and Newborn explore the world of computer chess, from its history to how the computer decides where to move. The book examines most aspects of programming computers to play chess, including an in-depth description of the basic principles. The definitive book for the chess aficionado and computer buff alike.

Computer Science Press
264 pages, 160 illustrations
Cloth: 8239-1 $23.95
SPECIAL PRICE $19.95
Paper: 8121-2 $11.95
SPECIAL PRICE $9.95



Your satisfaction matters. Return any book in good condition within 15 days of receipt and we'll send you a prompt refund.

To order, mail this coupon or a copy to:

W.H. Freeman and Company
4419 West 1980 South
Salt Lake City, UT 84103

All orders must be prepaid in U.S. dollars only.
☐ I enclose my check or money order made payable to W.H. Freeman and Company.

☐ Visa ☐ MasterCard Exp. date ______

Acct. no. ______

Signature ______
(Credit card orders must be signed)

Name ______
(Please print)

Address ______

City/State/Zip ______

Qty.	Author	ISBN	Special Price	Total

Postage & handling* (Add $1.95 for first book, $1.25 for each additional book) $ ______

NY, CA, and UT residents add sales tax $ ______

TOTAL $ ______

*For Canadian residents, postage and handling is $2.25 for first book, $1.50 for each additional book. Please allow four weeks for delivery. Items ordered together may be shipped separately

4893

Worlds around Other Stars

Theory and observation imply that planetary systems like our own should be common. Astronomical searches are closing in on planets that may orbit some nearby stars

by David C. Black

Do planets like our own orbit around other stars? The answer links intimately with the question of whether life exists beyond the earth. The most promising place for life to arise is on the surface of a planet, which functions as a "cosmic petri dish" where life can begin and be nurtured as it evolves to more complex states. A search for other planetary systems should therefore be a major component of any effort to understand the prospects for extraterrestrial life. Moreover, the results from such a search are essential for understanding the origin of the earth and the surrounding solar system.

Over many years, astronomers have built up a detailed picture of the origin of the sun and its retinue of planets. Unfortunately (but until now inseparably), the specific elements of that picture are based entirely on features found in a single planetary system—our own. Finding statistically significant evidence regarding the plentitude and nature of other planetary systems undoubtedly would provide many unexpected details about the process by which stars and planets are born.

Prospects for such a discovery have recently grown brighter. A number of researchers have possibly detected planetary companions circling other stars, although all the sightings remain highly tentative. A new generation of detectors and telescopes, along with some innovative detection techniques, should improve the situation.

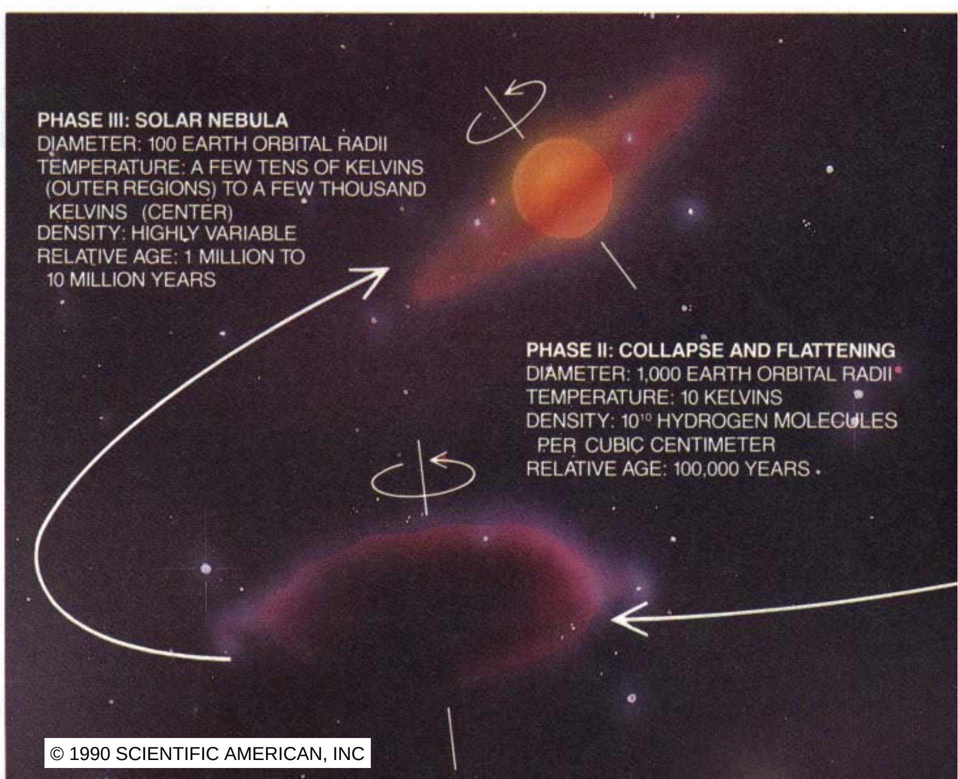
The current view of the origin of the solar system has its roots in concepts developed by Immanuel Kant and Pierre Simon Laplace late in the 18th century. They proposed a nebular hypothesis of the solar system's origin, wherein the sun and planets condensed out of a large, lumpy cloud. Over the decades, the nebular hypothesis has been greatly refined and modified, but the basic concept remains. It shapes current thinking on how and where planets form and, therefore, how one might go about searching for them.

Some notable regularities and variations in the structure of the solar system hint at its nebular origin. The planets all orbit in nearly the same plane: the orbits of all planets except Mercury and Pluto lie within three degrees of the ecliptic, the plane of the earth's orbit. The mean orbital plane of the planets also sits within six degrees of the equatorial plane of the sun. These properties suggest that the planets formed from a common disklike structure, known as the solar nebula.

If viewed from above the earth's North Pole, the planets all revolve about the sun counterclockwise, the same direction in which the sun rotates on its axis. The planets also travel in nearly circular orbits (Mercury and Pluto again are mild exceptions). Such orderly movements fit with the notion that the ancestral disk was dynamically ordered, not chaotic or irregular, and that motions within the disk were dominated by rotation about the sun.

Another noteworthy pattern in the solar system is that planetary composition varies according to distance from the sun. The gaseous outer planets (Jupiter, Saturn, Uranus and Neptune) primarily contain relatively light, volatile elements. This is particularly true for Jupiter, whose composition, being dom-

DAVID C. BLACK has done extensive research in theoretical astrophysics and planetary science, focusing on studies of star and planetary system formation. Black received his Ph.D. in physics from the University of Minnesota. He worked for the National Aeronautics and Space Administration's Ames Research Center from 1972 to 1988, when he assumed his current position as director of the Lunar and Planetary Institute in Houston. Black has studied the composition of noble gases in meteorites; he was the first to determine that meteorites contain material that originated from beyond the solar system.

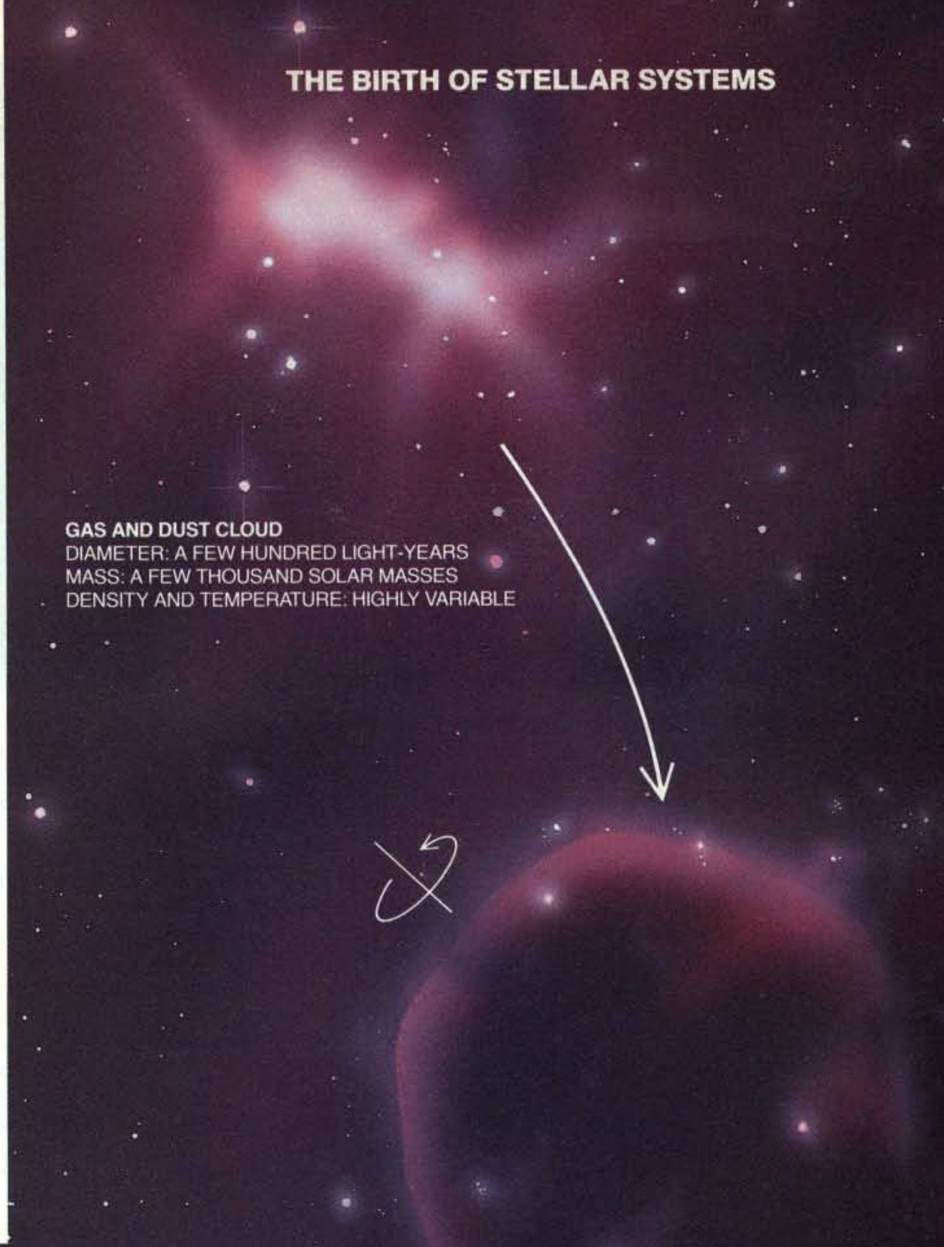


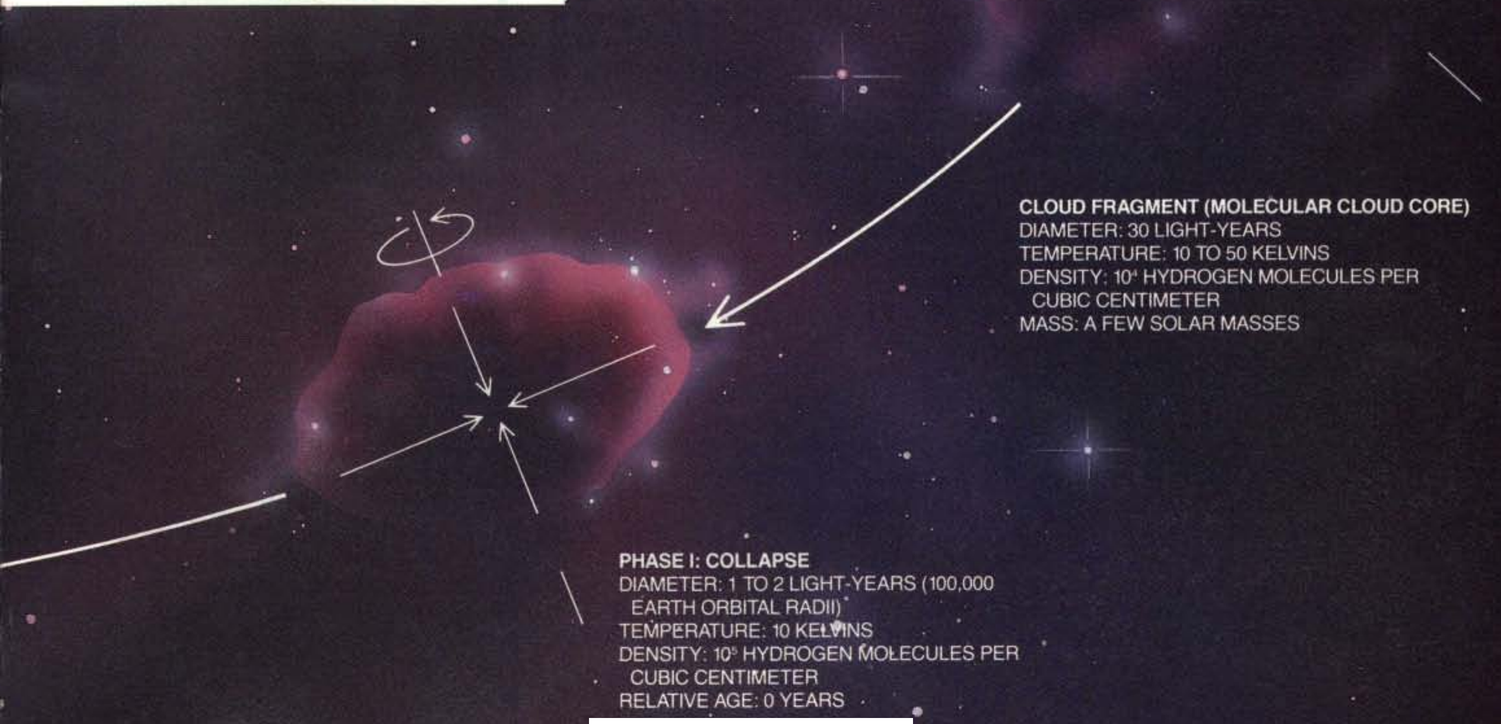
inated by hydrogen and helium, is similar to that of the sun. The other outer planets contain less hydrogen but are rich in hydrogen compounds, such as ammonia.

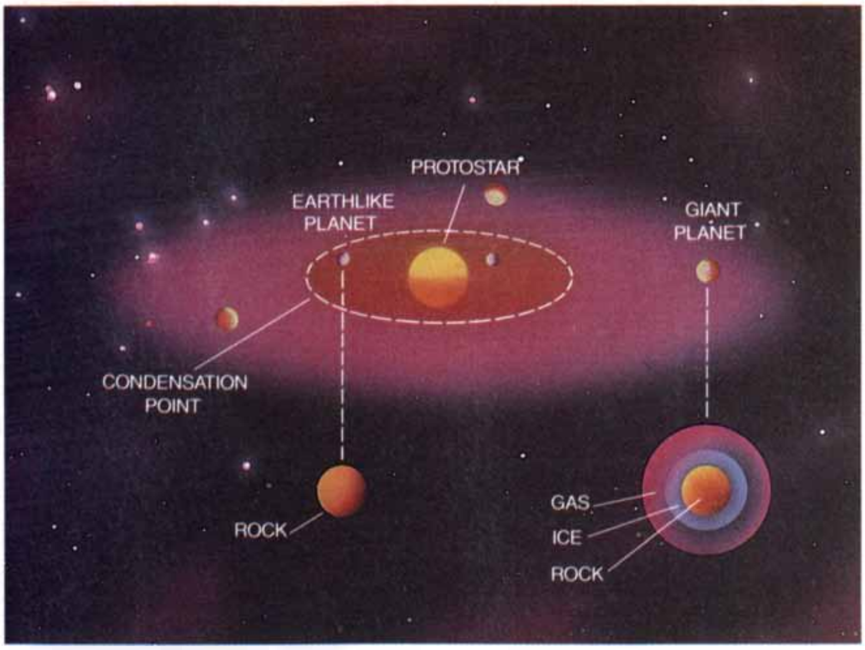
In contrast, the inner, or terrestrial, planets (Mercury, Venus, Earth and Mars) consist mostly of heavier elements, such as silicon and iron, which cosmically are far less abundant than hydrogen. If enough light elements were added to the earth to make its composition match that of the sun, its mass would be comparable to Jupiter's. It seems that the outer planets are more massive because they were able to hold on to more light, volatile elements and compounds in the solar nebula when the planets formed.

The generally accepted explanation for this difference is that the inner part of the solar nebula was sufficiently hot that volatile elements existed only in gaseous form. Critical early stages of planet formation are most likely controlled by the accumulation of solid material into progressively larger objects. Water is cosmically abundant and condenses at high temperatures compared with the average nebular temperatures, so it probably played an important role in planetary formation. Regions of the nebula where the temperature was 170 kelvins or lower (the condensation point for water ice) should have contained an adequate supply of solid material for giant planets to form.

Therefore, one might reasonably assume that the largest planets always form in the cool, outer regions of a circumstellar disk. This hypothesis helps to define where, relative to the central







GIANT PLANETS probably form only in cool, distant parts of the solar nebula where water condenses into ice, providing abundant material for planetary growth. Planets in hotter regions closer to the forming star accumulate from less common silicon and iron compounds; they end up as smaller, rocky bodies like the earth.

star, astronomers should look for giant extrasolar planets.

Circumstellar disks seem to be a natural outgrowth of the manner in which stars are born. Stellar formation begins with the gravitational collapse of dense, cool cores of material within clouds of molecular gas and dust. As the core shrinks, random motions cancel out, and the core's overall rotation creates a flattened gaseous disk. Evidence suggests that such disks are highly stiff, or dissipative, causing angular momentum to transfer outward in the disk and mass to flow inward into the infant star.

Planets coalesce during the brief interval—no more than a few million years—between when the disk forms and when it disappears, either swallowed or expelled by the newborn star. Various evidence hints that the planets did not simply collapse out of the solar nebula. For instance, the rotational axes of the planets are not generally perpendicular to the ecliptic, which suggests that the planets grew through a complicated, chaotic process involving the accretion of smaller units. The implication is that a fundamental difference exists between stars and planets in the way that they form, an important distinction to keep in mind when searching for planetary systems.

If current understanding of stellar birth is broadly correct, then disks should be found in association with many young stars. Some fraction of these disks may then pass through evolutionary stages similar to those inferred to have occurred in the young solar system. The diffuse, low temperature and often heavily obscured nature of disks around forming stars make them difficult to study visually, and so astronomers have concentrated on observations made at radio or infrared frequencies. These longer wavelengths of electromagnetic radiation can penetrate thick clouds of dust far more effectively than visible light.

The first convincing indication that disk structures do indeed surround young stars was the discovery that these objects are sources of energetic jets or winds of gas and dust. The outflows presumably represent material ejected from an evolving disk, but they provide only circumstantial evidence of disks around forming stars.

More recent and direct studies have bolstered this evidence. When dust particles absorb light from a star, they grow hot and reradiate the light as less energetic infrared radiation. The wavelength of the infrared rays depends on the size of the particles. A number of young stellar objects (most notably the objects known as HL Tauri, R Monoceros and L1551/IRS 5) emit abnormally large amounts of infrared radiation, indicating the presence of dust particles. Infrared observations of the above objects, both from earthbound telescopes and from the National Aeronautics and Space Administration's *Infrared Astronomy Satellite* (*IRAS*), imply that these dust particles are one to a few tens of microns across. Small as the particles may seem, they are significantly larger than those that exist in interstellar clouds, hinting that the building-up process that leads to planets may have begun around these stars. Some researchers argue that the dust is produced by collisions between cometlike bodies—the precursors of planets, perhaps—revolving around these stars.

A particularly powerful observation technique, known as speckle interferometry, freezes out blurring from the earth's atmosphere by using very short exposures from which the undistorted image is mathematically reconstructed. Even this technique cannot produce a clear picture of the dust clouds surrounding young stars, but it does permit astronomers to determine that the shape and size of the clouds is consistent with those expected for a disk. Images from *IRAS* confirmed that many stars, including older stars such as Vega and Beta Pictoris, are surrounded by disklike dusty structures. Observations of radio emission indicate the presence of gas in the disks, particularly carbon monoxide, which radiates prominently at radio wavelengths of a few millimeters.

Additional evidence for condensing disks around other stars comes from infrared observations of young stars known as T Tauri stars. As they age, these stars radiate less at short infrared wavelengths, as if the hottest particles—those closest to the star—are being swept up or vaporized.

Efforts to search for other planetary systems employ either direct or indirect methods. Direct methods involve detecting reflected light or infrared radiation from the planets themselves. The primary difficulty with this approach is that emission from a planet tends to be drowned out by the vastly brighter emission of its nearby parent star.

Indirect methods involve scrutinizing a star for signs that it is responding to the gravitational tug of an orbiting planet. As the planet moves from one side of the star to the other, it pulls the star back and forth. This pull manifests itself as a slight wobble superposed on the star's overall motion across the sky. It can also be detected as a slight, periodic change in the star's velocity with respect to the earth.

Any motion toward or away from the earth causes the star's light to be slightly compressed or stretched. When light is compressed, it becomes slightly bluer; when stretched, it becomes slightly redder, a phenomenon known as the Doppler effect. Careful measurement of absorption lines in a star's spectrum can in principle reveal any periodic changes in its motion.

Indirect searches can be facilitated by some simplifying assumptions. Many researchers have guessed that giant planets around other stars will have orbital periods similar to Jupiter's, about one decade. Fixing the orbital period allows one to determine the size of the orbit as a function of stellar mass and thus to calculate an expected angular or velocity perturbation (given the distance to the star and assuming a certain mass for the planet).

An alternative possibility, which I think is more logical, is that giant planets form at distances from their stars where the temperature is at or below the condensation point of water. In this case, the star's luminosity determines the size of the orbit for giant planets. One can then proceed as before to determine how much the planet is likely to affect the star's velocity and apparent path [*see table on this page*].

The size of the angular disturbance, or wobble, grows with increasing distance between planet and star but shrinks with increasing stellar mass. If planetary systems form in such a manner that giant planets always tend to have more or less similar orbital periods, then faint low-mass stars should be most strongly perturbed by planets, because giant planets would orbit far from these stars relative to the stars' mass. If, however, giant planets normally form at distances where temperatures are below the condensation point of water, the opposite should be true because giant planets would orbit far from bright, massive stars relative to the stars' mass. (In either case, low-mass stars experience the largest velocity perturbation.)

If temperature is the determining factor, then typical orbital periods of giant planets may be much shorter than those commonly expected. The average star is cooler and much less luminous than the sun, so giant planets could orbit in relatively close, fast orbits. For a typical nearby star with a mass of about 0.3 solar mass, the orbital period for a giant planet would be less than one year, compared with Jupiter's orbital period of about 12 years.

Even with the use of simplifying assumptions, detecting other planetary systems remains an extremely difficult task. Current telescopes and instruments operate at the limits of their capabilities when they scan for stellar companions having masses comparable to or less than the mass of Jupiter. As a result, the history of planetary searches is full of false leads and phantom discoveries.

Perhaps the best-known star to those who search for other planetary systems is the faint, cool, low-mass (M-type) star known as Barnard's star, named after the American astronomer E. E. Barnard, who noticed its unusual attributes in 1916. Barnard's star has the highest-known apparent motion across the sky; it is also, after the Centauri triple-star system, the second closest star system to ours, lying only six light-years away.

Peter van de Kamp, working at the Sproul Observatory, realized in 1937 that these two properties make Barnard's star ideal for indirect searches because any apparent wobble would be relatively large and well defined. Van de Kamp examined positional data on Barnard's star dating back to 1916 and began collecting his own data using the telescope at Sproul.

In the 1960s van de Kamp concluded that Barnard's star had two Jupiter-mass companions, one completing an orbit every 12 years, the other every 24 years. Yet studies during the past decade by George D. Gatewood of Allegheny Observatory in Pittsburgh and, independently, by Robert S. Harrington of the U.S. Naval Observatory find that the motion of Barnard's star is inconsistent with the existence of van de Kamp's planets. Indeed, the observations by these two researchers exclude the possibility that Barnard's star has any companion much in excess of Jupiter's mass, although lower-mass companions could still be present.

Another widely reported substellar object was the companion to the nearby star Van Biesbroeck 8, dubbed VB8-b, which was independently discovered in 1984 by Harrington and by Donald McCarthy and co-workers at the University of Arizona. This sighting seemed especially persuasive because two different techniques were used to find the object: Harrington used astrometric observations to infer a kink in the motion of VB8-b, and McCarthy observed what appeared to be a faint object close to it by means of infrared speckle interferometry techniques.

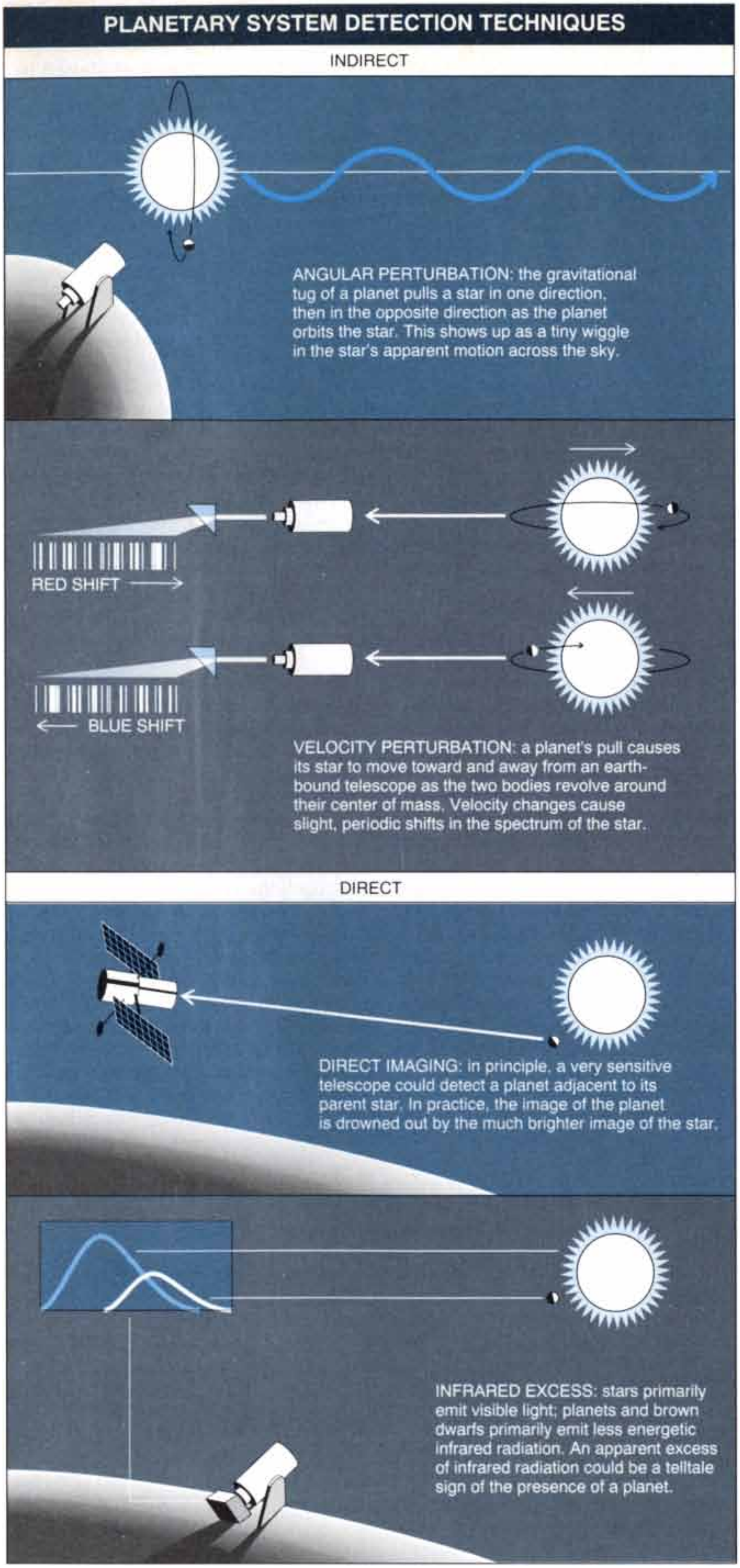
After considerable debate a consensus emerged that VB8-b was an intermediate-class object having about one twentieth the mass of the sun, or about 50 times that of Jupiter. Such an object, known as a brown dwarf, would be too small to trigger the nuclear reactions that power stars, and so it would be nearly impossible to observe at visible wavelengths. A few years after the discovery of VB8-b, several research teams undertook more sensitive observations to confirm and characterize better its properties. Despite considerable effort, they found no trace of the much publicized object. Astronomers now generally think that VB8-b does not exist, at least not as described by the early observations.

An extremely precise technique for detecting tiny velocity perturbations has been developed recently by Bruce T. E. Campbell of the University of Victoria and his co-workers. This technique involves comparing the spectrum of stars with a very high resolution reference spectrum from an unlikely and rather nasty compound, hydrogen fluoride. In principle, Campbell's instrument shows far smaller Doppler shifts, and hence far slighter velocity disturbances, than previously could be detected.

Campbell and his colleagues at the Canada-France-Hawaii telescope located on Mauna Kea have measured the radial velocities of 15 stars six times a year with a stated accuracy of about 10 meters per second. (This accuracy is

SIGNS OF PLANETS AROUND STARS

STELLAR MASS (SUN=1)	ORBITAL DISTANCE (A.U.)	TEMP. (KELVINS)	ORBITAL PERIOD (YEARS)	ANGULAR DISTURBANCE (MICROARC-SEC)	VELOCITY DISTURBANCE (METERS/SEC)
EQUAL ORBITAL PERIODS					
3	7.6	293	12	76	0.17
1	5.2	135	12	157	0.39
0.3	3.5	57	12	351	0.88
CONSTRAINED BY TEMPERATURE					
3	22.6	170	61.6	226	0.1
1	3.3	170	6.0	99	0.5
0.3	0.4	170	0.5	40	2.6



remarkable given that convective currents at the surfaces of sunlike stars typically move on the order of 1,000 meters per second.) Campbell's group has stopped short of overtly claiming to have discovered another planetary system, but they have found signs of long-term accelerations—a potential indicator of the presence of planets—in nearly half of the stars studied.

The most intriguing object studied by Campbell's group is the star Gamma Cephei, an old, orange subgiant star (class III–IV, spectral type K1) whose mass is estimated to be slightly greater than that of the sun. Spectral measurements collected since 1981 led Campbell to conclude that Gamma Cephei experiences a cyclic velocity variation having an amplitude of 25 meters per second and a period of 2.6 years. This period, combined with the star's estimated mass, implies that the companion orbits roughly 300 million kilometers from Gamma Cephei, or about twice the earth's distance from the sun. To produce the observed velocity perturbation, the companion must have at least 1.5 times the mass of Jupiter.

It would be surprising to find a giant planet orbiting so close to a relatively luminous star, but surprises are the brood stock of scientific inquiry. A more troubling aspect of Campbell's finding is that the more recent and accurate data from the study of Gamma Cephei show significant deviations from the motions that would be expected as the putative planet orbits its star. Not until Gamma Cephei has been observed through several 2.6-year "orbits" will the data appear truly convincing. Nevertheless, this work and similar, equally sensitive and accurate efforts being conducted at the University of Arizona by Robert S. McMillan and co-workers hold great potential.

David Latham and his collaborators at the Harvard-Smithsonian Center for Astrophysics, using less accurate but more conventional velocity measurement techniques, have collected some of the strongest evidence yet for the existence of a substellar companion. Latham's group observed the solar-type star HD114762 for more than 12 years and found that it undergoes periodic velocity variations. In this case, measurements made at European observatories have confirmed the data.

Latham and his colleagues find a consistent velocity variation having a period of about 84 days and an amplitude of about 550 meters per second. Assuming the star has the same mass as the sun (which is consistent with its temperature and luminosity), the perceived period implies that a companion orbits at a mean distance of about 60 million kilometers, similar to the distance of Mercury from the sun. The lower limit to the mass of the companion is 11 times the mass of Jupiter. Recent, more accurate observations by W. D. Cochran and co-workers at the University of Texas at Austin confirm the period and imply that the orbit is eccentric. Their work also suggests that the companion's mass is significantly greater than previously estimated.

Many researchers understandably tend to view this companion as a true extrasolar planet, but I have doubts. The companion is far more massive than any planet in the solar system. It may represent not a planet but the very low mass end of the population of starlike objects, and HD114762 may be an extreme type of binary star. Although the process by which binary stars form is not well understood, it seems unlikely that the process depends on whether the objects are large enough to ignite fusion reactions in their cores. There is no reason to suppose that nature would not make binary systems in which one member is a star and the other a brown dwarf.

Objects more than 10 or 20 times the mass of Jupiter probably form like stars, not like planets—that is, they condense directly from a gas cloud rather than from the disk around a star. In my view, the different modes of formation are manifested in the fact that planetary systems (at least the one known) contain a multitude of bodies, whereas stars are predominantly double or triple systems. The reason for the cutoff between planets and brown dwarfs is mysterious but seems to be real. This information leads to an important principle of search efforts for planetary systems: the object of these searches is to discover planetary systems, not brown dwarfs. Systems composed of a star and a single substellar companion are not planetary systems.

Nevertheless, studies of brown dwarfs may elucidate how stars and planets form. Direct searches currently lack the sensitivity to detect Jupiter-size or smaller planets around other stars, but they can provide significant information about larger substellar companions. Most of these searches are conducted in the infrared part of the spectrum.

Studies using sophisticated new infrared array detector systems have revealed faint, cool companions around a small number of stars. Ben Zuckerman and Eric E. Becklin of the University of California at Los Angeles, among others, have detected dwarf companions by looking for excess infrared radiation that cannot be explained by a normal star's emission. All the objects found so far can be modeled as brown dwarfs weighing several tens of Jupiter masses. Sightings of brown dwarfs remain highly uncertain, however.

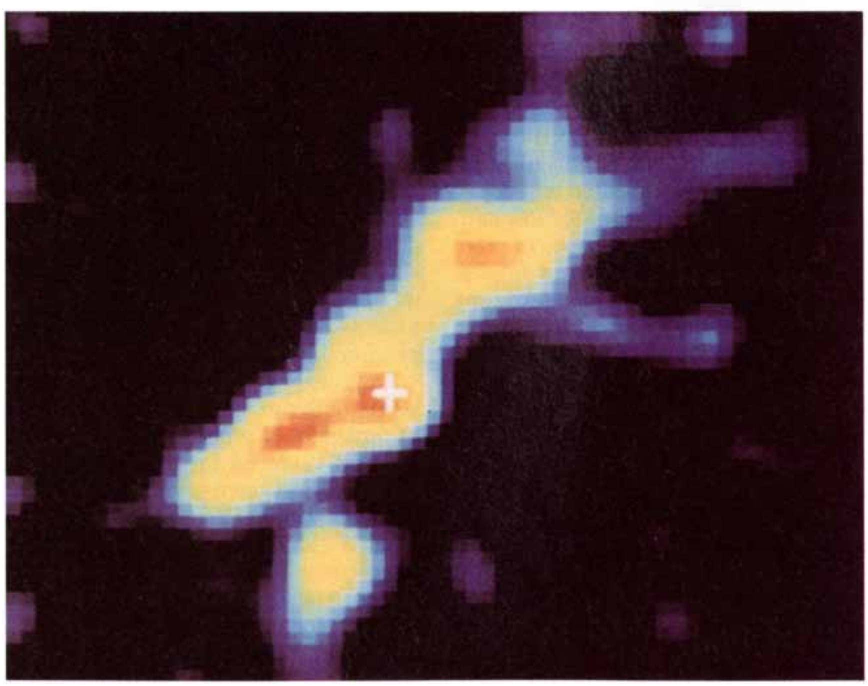
Various groups, such as McCarthy and his co-workers at Arizona and Geoffrey Marcy and colleagues at San Francisco State University, are compiling statistics on the abundance of brown dwarfs. The Arizona group has examined 27 nearby M-type, red dwarf stars. The researchers used speckle interferometry at infrared wavelengths of 1.6 and 2.2 microns (known as the H and K bands) to search for companions too dim to be detected at visible wavelengths.

The number of stars in the infrared survey grew greater with increasing faintness (larger magnitudes denote fainter stars) up to a K-band magnitude of about 10.0, at which point the numbers fell off abruptly. The Arizona group found no objects with K-band magnitudes between +10.0 and +11.5. A brightness of +11.5 (the limit of the survey) corresponds to the expected infrared luminosity of a brown dwarf 70 to 80 times the mass of Jupiter and several billion years old, as seen from a distance of five parsecs (16 light-years). The apparent absence of such objects suggests that no continuous population of objects extends all the way from stars to planets.

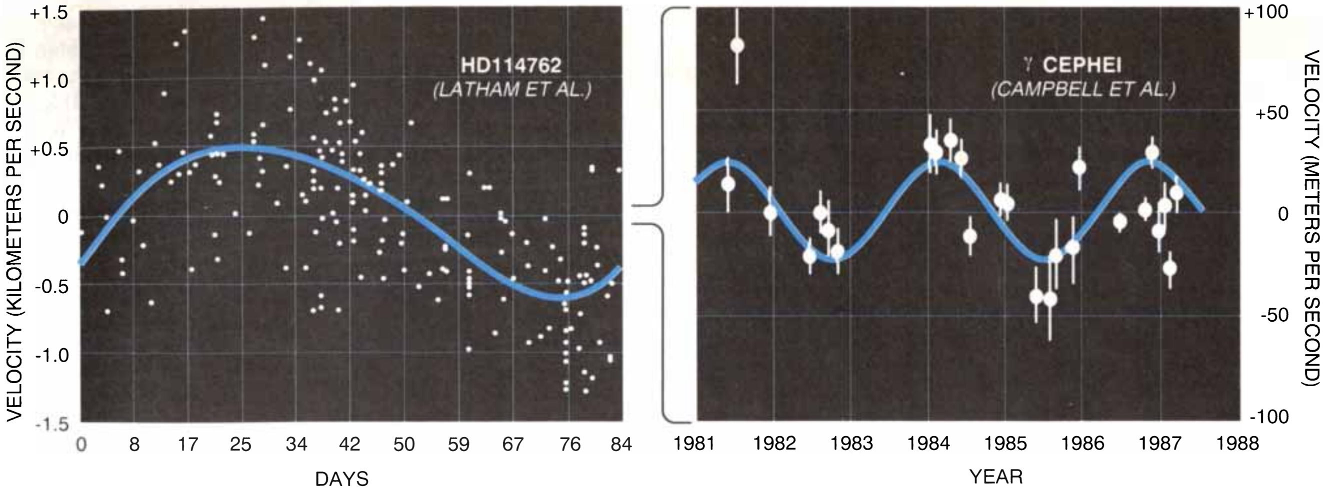
This view was confirmed by Marcy and K. Benitz, a student at San Francisco State, who recently surveyed 70 low-mass stars using radial velocity techniques precise to 230 meters per second. The measurements should have revealed any substellar companions down to a mass of about seven Jupiter masses, so long as the companions' orbital periods were less than four years. The survey uncovered six hitherto unknown stellar companions to the stars studied but gave no evidence for the presence of substellar companions. Indeed, the radial velocity survey, combined with results from long-term astrometric studies of stellar positions, indicates that less than 2 percent of all stars have substellar companions whose mass is greater than 10 times that of Jupiter.

In some instances, the absence of a discovery can itself be an important discovery. I have long felt that such is the case with substellar objects. As a rule, lightweight stars are more common than massive ones. Extrapolating downward, many theorists expected that brown dwarfs would be sprinkled liberally throughout the galaxy, like dust. The paucity of substellar companions is nature's way of telling us that it is time to rethink some of the physics associated with the formation of extremely low mass stars.

It is clear that current facilities are



FLATTENED DISK of gas and dust orbiting the infant star HL Tauri (located at the white cross) is revealed in this false-color map of radio emission from carbon monoxide molecules. The disk is about 30 times the diameter of Pluto's orbit. Similar disks surround other young stars; they are the likely birthplace of planets.



SPECTROSCOPIC OBSERVATIONS of the star HD114762 show the gravitational signature of a substellar companion, probably a brown dwarf. Evidence for a Jupiter-mass planet orbiting the star Gamma Cephei is enticing but less convincing.

inadequate to conduct a comprehensive search for other planetary systems. The detailed observations necessary to conduct this project in a scientifically adequate manner will require a new set of advanced instruments. In general, dedicated searches should be conducted with sufficient accuracy and sensitivity that a null result unambiguously advances the understanding of planetary-system formation. This principle should help with assessing facilities that might be used to search for other planetary systems. Many promising design concepts have already emerged, some ground based, others based in space.

Radial velocity (velocity perturbation) measurements are relatively impervious to the blurring caused by the earth's turbulent atmosphere. Larger, dedicated telescopes permit a careful survey of tens to hundreds of stars for the presence of planetary companions. Telescopes now in operation at the University of Arizona and the University of Texas could serve as prototypes for the next generation of radial velocity systems.

Another class of ground-based systems would use very large aperture (seven to 10 meters) telescopes incorporating active optics that physically manipulate the mirror to compensate for atmospheric distortion. These instruments, which would be used to search for other planetary systems directly at infrared wavelengths, could revolutionize understanding of the structure and evolution of the disks that surround young stars.

Most of the techniques for planetary searches would function best in space. NASA's *Space Infrared Telescope Facility,* currently planned for launch in the late 1990s, will provide infrared observations of circumstellar disks and newborn planetary systems that will greatly improve on the images returned by *IRAS* in 1983.

Astrometry—the ultraprecise measurement of stellar position—will gain greatly by being conducted in space. The goal is to be able to measure stellar angular deflections with an accuracy of ten millionths of an arc second—the angular extent of a dime on the moon as viewed from the earth! Such accuracy would permit the detection of companions as small as 10 earth masses around any star within 10 parsecs (30 light-years) of the sun.

Telescopes with precisely ground mirrors and masks to blot out the bright light of a central star might be able to capture direct visible-light images of planets around other stars. An exciting possibility is that both astrometry and imaging could be done by a single telescope. A study of the feasibility of building such a combined instrument is now under way at the Jet Propulsion Laboratory in Pasadena, Calif.

The next generation of planetary search instruments undoubtedly will be optical interferometers, networks of telescopes placed far apart that combine their images to create, in effect, a single enormous telescope. Such devices could offer thousands of times the resolving power of existing instruments. The moon would be an ideal location for an optical interferometer and many other astronomical instruments. It is conceivable that the lunar far side will be the site of a suite of scientific facilities that will search not only for other planetary systems but also for signals from intelligent life elsewhere in the universe [see "Observatories on the Moon," by Jack O. Burns, Nebojsa Duric, G. Jeffrey Taylor and Stewart W. Johnson; SCIENTIFIC AMERICAN, March 1990].

Throughout the ages, humans have wondered about the possible existence of worlds other than their own. The search for other planetary systems has been going on in earnest for more than half a century, during which time astronomers have edged tantalizingly close to their goal. An explosion of interest in this field of research, along with the development of more accurate instruments, promises that the next few decades will be especially exciting. The first positive sighting of a planetary system other than our own will be a landmark, completing the revolution in thought begun some 450 years ago by Nicolaus Copernicus.

FURTHER READING

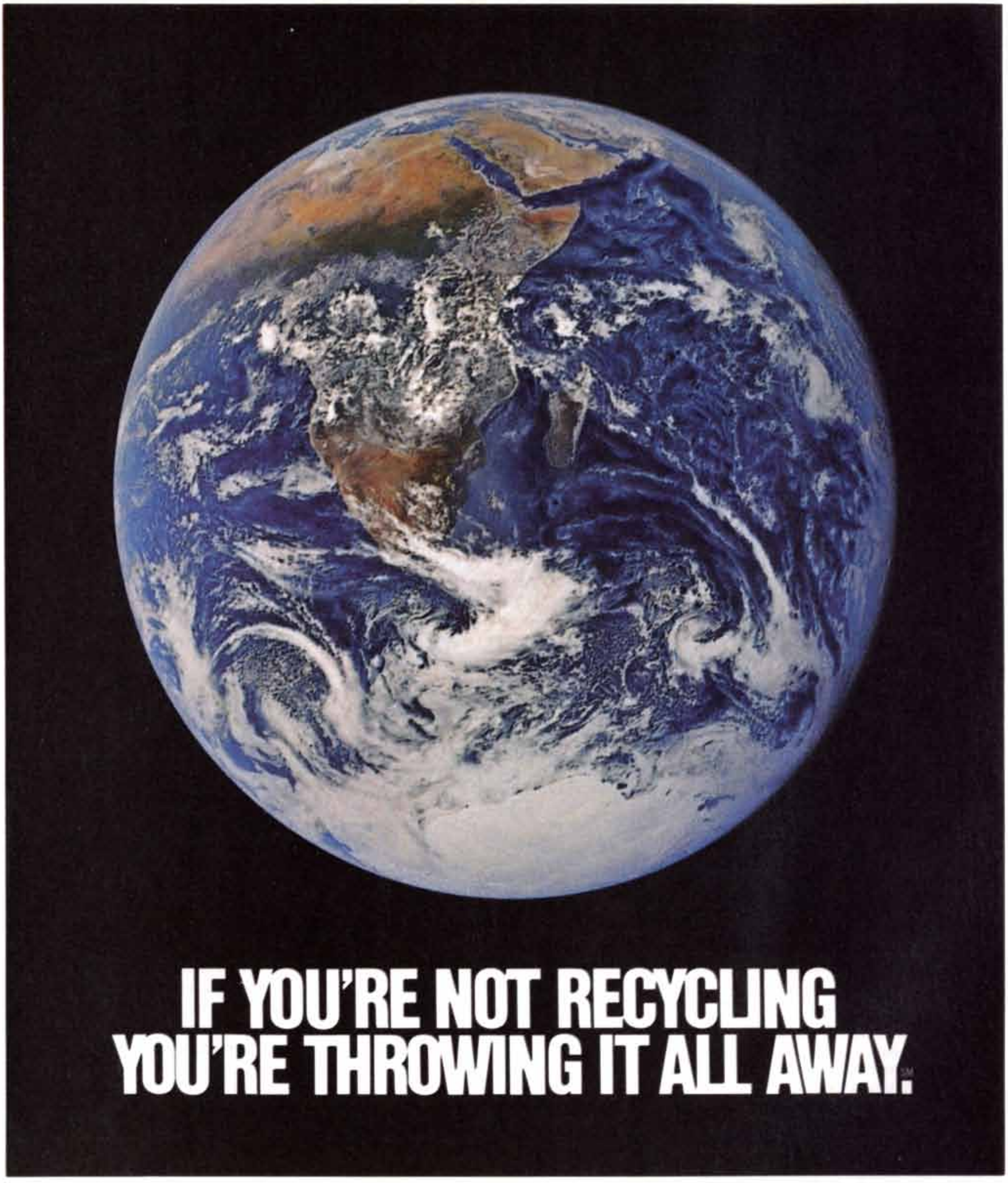
IN SEARCH OF OTHER PLANETARY SYSTEMS. David C. Black in *Space Science Reviews,* Vol. 25, No. 1, pages 35-81; January 1980.

PROTOSTARS AND PLANETS II. Edited by David C. Black and Mildred Shapley Matthews. University of Arizona Press, 1985.

LIFE IN THE UNIVERSE. Donald L. DeVincenzi in *Planetary Report,* Vol. 7, No. 6, page 3; November/December 1987.

CIRCUMSTELLAR MATERIAL ASSOCIATED WITH SOLAR-TYPE PRE-MAIN-SEQUENCE STARS: A POSSIBLE CONSTRAINT ON THE TIMESCALE FOR PLANET BUILDING. K. M. Strom, S. E. Strom, S. Edwards, S. Cabrit and M. F. Skrutskie in *Astronomical Journal,* Vol. 97, No. 5, pages 1451-1470; 1989.

A PROFUSION OF PLANETS. David C. Black and Eugene H. Levy in *The Sciences,* pages 30-35; May/June 1989.



A little reminder from the Environmental Defense Fund that if you're not recycling, you're throwing away a lot more than just your trash.

You and your community can recycle. Please write the Environmental Defense Fund at: EDF-Recycling, 257 Park Avenue South, New York, NY 10010, for a free brochure that will tell you virtually everything you need to know about recycling.

EDF™ Ad Council

Aspirin

After more than 200 years, the mechanisms by which this venerable drug and its relatives achieve their wide range of effects have yet to be fully elucidated

by Gerald Weissmann

There is a bark of an English tree, which I have found by experience to be a powerful astringent, and very efficacious in curing aguish and intermitting disorders.
—The Rev. Mr. Edmund Stone of Chipping-Norton in Oxfordshire, in a letter to the Right Honourable George, Earl of Macclesfield, president of the Royal Society, April 25, 1763

What Stone had discovered, although he did not know it, was that salicylates—the general term for derivatives of salicylic acid—reduced the fever and relieved the aches produced by a variety of acute, shiver-provoking illnesses, or agues. The bark of the willow tree (*Salix alba*) is astringent because it contains high levels of salicin, the glycoside of salicylic acid.

Today the salicylate most commonly used is acetylsalicylic acid, better known by its first trade name, "Aspirin." Americans consume 16,000 tons of aspirin tablets a year—80 million pills—and spend about $2 billion a year for nonprescription painkillers, many of which contain aspirin or aspirinlike drugs.

As Stone observed, these compounds exhibit a wide range of effects: At the lowest doses—less than one tablet a day—aspirin can be used to treat and prevent heart attacks and to prevent cerebral thrombosis. Two to six tablets a day (one to three grams) are useful for reducing pain or fever. And much higher doses (four to eight grams a day) reduce the redness and swelling of joints in diseases such as rheumatic fever, gout and rheumatoid arthritis.

Aspirin and the salicylates also have many other biological effects, only some related to their clinical use. Salicylates can dissolve corns on the toes, provoke loss of uric acid from the kidneys and kill bacteria in vitro. Aspirin inhibits the clotting of blood, induces peptic ulcers and promotes fluid retention by the kidneys.

Cell biologists use aspirin and salicylates to inhibit ion transport across cell membranes, to interfere with the activation of white blood cells and to derail the production of the energy-storage compound adenosine triphosphate by isolated mitochondria. Molecular biologists use the compounds to activate genes that code for so-called heat-shock proteins in the lampbrush chromosomes of the fruit fly, *Drosophila*. And botanists use salicylates to induce flowering of such plants as *Impatiens* and the voodoo lily.

The enormous range of effects that aspirin can produce has made it a complex task to pin down the biochemical mechanisms involved. Not until the early 1970s did biologists find a hypothesis to explain the action of aspirin and such related drugs as ibuprofen, indomethacin and piroxicam. That hypothesis was based on the ability of these drugs to block the synthesis of prostaglandins, cellular hormones involved in pain and inflammation.

More recently it has become clear that the prostaglandin hypothesis explains only some of the effects of aspirin and related compounds. Their crucial anti-inflammatory power appears to stem not only from prostaglandin inhibition but also from their ability to disrupt interactions within cell membranes. Recent work in my laboratory, for example, has shown how aspirin-like drugs prevent the activation of cells that mediate the first stages of acute inflammation.

The story of how willow extracts found their way from the agues of Oxfordshire to the bench side of molecular biologists can be told in terms of the four Gs of Paul Ehrlich: *Geduld, Geschick, Geld und Glück* (patience, skill, money and luck). In the case of Reverend Stone, luck seemed to have shown the way.

In about 1757 Stone tasted willow bark (already a well-known folk remedy) and was surprised at its extraordinary bitterness. The resemblance to the taste of Peruvian bark (*Cinchona*)—a rare and expensive remedy for the ague—excited Stone's suspicions. Six years of careful clinical observation culminated in his letter to the Royal Society. Stone offered a skillful rationale for the use of willow bark in febrile disorders, based on the traditional doctrine of signatures, which held that "many natural maladies carry their cures along with them, or their remedies lie not far from their causes." Willows, as do feverish illnesses, abounded in moist shires.

Half a century later, driven by national rivalry, French and German pharmacologists competed to find the active principle of willow bark. By 1828, at the Pharmacologic Institute of Munich, Johann A. Buchner isolated a tiny amount of salicin in the form of bitter-tasting yellow, needlelike crystals. One year later, H. Leroux in Paris improved on the extraction procedure and obtained one ounce of salicin from three pounds of the bark. In 1833 the pharmacist E. Merck of Darmstadt obtained a clean preparation of salicin that was cheaper by half than the impure willow extracts. Not until 1838 did Raffaele Piria of Pisa, working in Paris, give the compound the name by which it is

GERALD WEISSMANN is professor of medicine at New York University School of Medicine and director of the division of rheumatology at the university's medical center, where he studies the molecular biology of inflammation. After completing his M.D., he received postdoctoral training in biochemistry at N.Y.U. and in cell biology at the Strangeways Research Laboratory in Cambridge, England. Weissmann is a past president of the American College of Rheumatology and the Harvey Society, and he is editor in chief of *MD Magazine*, *Inflammation* and *Advances in Inflammation Research*. He is also a frequent contributor to *Hospital Practice* and the *New York Times*. His latest book of essays is *The Doctor with Two Heads*.

known today: *l'acide salicylique,* or salicylic acid.

Other plants were also rich natural sources of salicylates. Meadowsweet (*Spiraea ulmaria*) yielded ample quantities of an ether-soluble oil from which the Swiss chemist Karl Jakob Löwig crystallized "*Spirsäure*" in 1835. In 1839 Dumas demonstrated that Löwig's *Spirsäure* was none other than Piria's *acide salicylique*. Another Gallic pharmacologist, Auguste Cahours, showed in 1844 that oil of wintergreen—a traditional remedy for aguish disorders—contained the methyl ester of salicylic acid.

As was to be the case in much of 19th-century chemistry, French and British scientists were slightly ahead of the Germans in the study of natural products, but Germans held the edge in synthetic know-how. Forced to compete with the French and British dye industries, which supplied the textile mills of Lyons and Macclesfield with pigments imported from overseas colonies, the Germans replied by inventing cheap aniline dyes, creating in their train such giant enterprises as I. G. Farben. By the mid-1870s German synthetic chemistry was preeminent worldwide. Whereas in the 1860s no German dyes were exported, by 1888 they supplied more than 80 percent of the world's needs.

Germans also began to dominate the willow business. In 1860 Hermann Kolbe and his students at Marburg University synthesized salicylic acid and its sodium salt from phenol, carbon dioxide and sodium. In 1874 one of those students, Friedrich von Heyden, established the first large factory devoted to the production of synthetic salicylates in Dresden. Whereas the price of salicylic acid made from salicin was 100 *Thaler* per kilogram in 1870, by 1874 the price of the synthetic product was only 10 *Thaler* per kilogram.

The availability of cheap salicylic acid spread its clinical use far and wide. In 1876 Franz Stricker and Ludwig Riess, writing in the *Berliner Klinische Wochenschrift,* and T. J. MacLagan, writing in the *Lancet,* reported the successful treatment of acute rheumatism—the disease now known as acute rheumatic fever—with salicylates at doses of five to six grams a day. Unfortunately, only the disease's acute symptoms, not its long-term consequences, respond to salicylates. Patients with rheumatic fever mount an inflammatory response against their own joints as if out at any cost to eradicate bacteria lodged there. The most severe and lasting damage is inflicted on the heart; fully one third of the victims suffer scarring of the valves.

The following year, in Paris, Germain Sée introduced salicylates as effective

WILLOW TREE is a source of salicin, a bitter compound whose derivative, acetylsalicylic acid (aspirin), is now the world's most common remedy for pain and fever. Other plants, such as meadowsweet and wintergreen, also contain salicylates.

therapies for gout and chronic polyarthritis. The latter category includes rheumatoid arthritis, an often crippling, inflammatory disorder of the middle years that affects more women than men, and degenerative osteoarthritis, a painful ailment that affects the knees of football players or the toes of young ballet dancers—and various joints of most people over 60.

Aspirin, now the most common salicylate, entered the competition fairly late. Its discovery, in 1898, began with the arthritic parent of an aniline dye chemist at the Bayer division of I. G. Farben. Felix Hofmann's father could not tolerate sodium salicylate because of chronic and acute stomach irritation. (No wonder, six to eight grams of salicylate a day will predictably irritate almost anyone.) Hofmann searched the chemical literature for less acidic derivatives of sodium salicylate and hit on acetylsalicylic acid. It proved more palatable and—Hofmann claimed—more effective at helping his dad. (I have a hunch that Hofmann *père* had osteoarthritis and got away with lower, analgesic doses of acetylsalicylic acid rather than anti-inflammatory quantities of sodium salicylate.)

Bayer called the new drug aspirin, the "a" from *acetyl* and the "spirin" from the German *Spirsäure* (the French root would have yielded *asalicylin*). By 1899 there was no chemical industry on earth that could compete with the German cartels. The Germans had won the aspirin war and could dictate the terms of victory.

Competitors entered the field as markets expanded for drugs that could reduce fever and pain. Some contenders, in fact, had been invented before aspirin but gained acceptance as over-the-counter remedies in Europe and the U.S. only after aspirin's success at the turn of the century. Based on anecdotal accounts from the Alsace that a product formed from aniline treated with vinegar made a useful febrifuge, or fever treatment, Karl Morner synthesized acetanilide in 1889—essentially an aniline version of acetylsalicylic acid.

Acetanilide itself caused bone marrow depression and anemias, and so other derivatives were sought. The most widely accepted aniline-derived product turned out to be phenacetin: the "P" in the APC that since 1939 has been so beloved by U.S. Army physicians, who have given every recruit with a fever a pill compounded of aspirin, phenacetin and caffeine. Both acetanilide and phenacetin are broken down in the body to form *N*-acetyl-*p*-aminophenol, which by various anagrammatic combinations yielded the generic names "acetaminophen" in the U.S. and "paracetamol" in Great Britain, as well as the trade name "Tylenol."

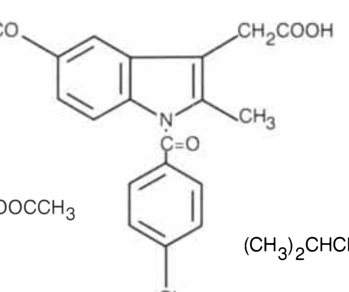
Neither acetanilide nor phenacetin, however, proved as useful as aspirin in treating rheumatic fever or rheumatoid arthritis, and for half a century (between 1900 and 1950) clinicians appreciated that there was something unique about high-dose salicylates. Doses of more than about four grams of aspirin a day did not just relieve fever and pain; they also reduced swelling and diminished the objective signs of inflammation. In addition, they brought under control such laboratory markers of the disease as the erythrocyte sedimentation rate (the rate at which red blood cells fall through plasma) and the levels of C-reactive protein, a substance produced by the liver in response to infection. Although later such drugs as ibuprofen, indomethacin and piroxicam also acted against inflammation, the mechanism by which they did so was unclear.

By the early 1970s no useful hypothesis had yet explained how salicylates exerted their various effects. Renal physiologists found that low doses of salicylates blocked the excretion of uric acid by the kidneys, thus raising levels of this acid in the blood, but, paradoxically, high doses of salicylates promoted renal excretion and so lowered uric acid levels. The latter property explained the utility of salicylates in both acute and chronic gout.

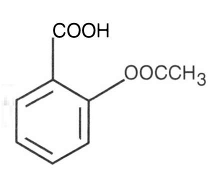
Pharmacologists had shown that salicylates reduce pain by acting on tissues and associated sensory nerves—as opposed to morphine, which acts on the brain. But physiologists main-

NONSTEROIDAL ANTI-INFLAMMATORY DRUGS

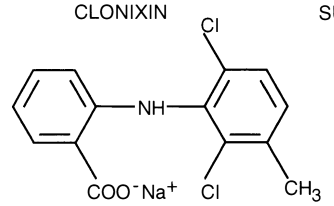
CARBOXYLIC ACIDS				PYRAZOLES	OXICAMS
SALICYLATES	*ACETIC ACIDS*	*PROPIONIC ACIDS*	*FENAMATES*		
ACETYLSALICYLIC ACID (ASPIRIN)	INDOMETHACIN	IBUPROFEN	MEFENAMIC ACID	PHENYLBUTAZONE	PIROXICAM
SALSALATE	ACEMETACIN	FLURBIPROFEN	FLUFENAMIC ACID	FEPRAZONE	ISOXICAM
DIFLUNISAL	CINMETACIN	NAPROXEN	MECLOFENAMATE	APAZONE	TENOXICAM
FENDOSAL	SULINDAC	KETOPROFEN	NIFLUMIC ACID	TRIMETHAZONE	
	TOLMETIN	FENOPROFEN	TOLFENAMIC ACID	MOFEBUTAZONE	
	ZOMEPIRAC	BENOXAPROFEN	FLUNIXIN	KEBUZONE	
	DICLOFENAC	INDOPROFEN	CLONIXIN	SUXIBUZONE	
	FENCLOFENAC	PIRPROFEN			
	ISOXEPAC	CARPROFEN			



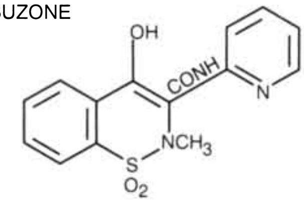
INDOMETHACIN



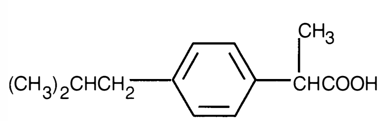
ACETYLSALICYLIC ACID



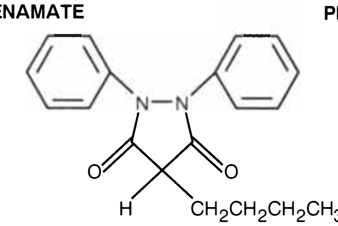
MECLOFENAMATE



PIROXICAM



IBUPROFEN



PHENYLBUTAZONE

tained that salicylates reduced fever by working directly on the fever centers of the hypothalamus, not by peripheral action.

It was even more difficult to explain how aspirin inhibited platelet function, caused salt and water retention and provoked indigestion. And why did some patients develop the nasal polyps accompanied by sniffles and wheezes known as aspirin hypersensitivity?

The first satisfactory mechanism for the action of aspirin was proposed in 1971 by John R. Vane (now Sir John and 1982 Nobelist) and his colleagues at the Royal College of Surgeons in London. That hypothesis moved aspirinlike drugs into the forefront not only of pharmacology but also cell biology and, eventually, clinical medicine.

Vane had been impressed by the fact that many forms of tissue injury are followed by the release of prostaglandins, a group of ubiquitous local hormones produced by the enzymatic oxidation of arachidonic acid, a fatty acid contained in cell membranes. (Prostaglandins have a host of regulatory functions, including regulation of blood vessel tone, uterine contraction and platelet function.) Unlike hormones such as insulin, prostaglandins are not stored within cells but are released when cells are injured or stimulated by other hormones. Moreover, sensitive chemical and biological assays had permitted investigators to show that two particular groups of prostaglandin, E_2 and I_2, caused several of the signs of inflammation, including redness (vasodilation) and heat (fever).

Vane then used radioactively labeled arachidonic acid to demonstrate that aspirin and related drugs inhibited the synthesis of prostaglandins E_2 and $F_{2\alpha}$. Moreover, platelets taken from volunteers given aspirin and indomethacin failed to make prostaglandins in response to the clotting factor thrombin. Finally, indomethacin inhibited the normal release of prostaglandins from dogs' spleens stimulated by the neurotransmitter catecholamine. There was no question that aspirinlike drugs blocked prostaglandin synthesis.

At long last the salicylate story seemed to have found a beginning, middle and end. All that remained, it appeared, was to fill in the details: to show how and when prostaglandins caused redness and swelling with heat and pain and to study the means by which aspirinlike drugs inhibited the enzyme—now known as prostaglandin H synthase—that transformed arachidonic acid to prostaglandins. That enzyme produces stable prostaglandins (those of the E, I and F series) via the unstable intermediates PGG_2 (prostaglandin G_2) and PGH_2 (first discovered in the 1970s by another 1982 laureate, Bengt Samuelsson of the Karolinska Institute in Stockholm).

By 1974 Vane and Sergio Ferreira had amassed convincing evidence for the prostaglandin hypothesis. Almost all aspirinlike drugs (by then generally called nonsteroidal anti-inflammatory drugs, or NSAIDs) inhibited prostaglandin synthase, and the potency of these drugs pretty much paralleled their effectiveness. Aspirin was anywhere from 1/40 to 1/200 as active as indomethacin and from one fifth to 1/50 as active as ibuprofen. Furthermore, such centrally acting analgesics as morphine or codeine did not inhibit prostaglandin synthase; neither did antihistamines, antiserotonin drugs or cortisone and its analogues.

Vane and his colleagues argued not only that prostaglandins were produced at sites of inflammation but also that they could, alone or in concert with other mediators, provoke the cardinal signs of inflammation. Indeed, prostaglandins E_2 and I_2 do induce vasodilation and promote swelling when dilated blood vessels have been made leaky by histamine. They also produce fever when injected either into the cerebral ventricles or into the anterior hypothalamus, and they sensitize pain receptors of the skin to such other pain-provoking substances as bradykinin or histamine.

Perhaps the most persuasive aspect of the prostaglandin hypothesis was its explanation of the clinical side effects of NSAIDs. Their most troublesome side effect is stomach irritation and ulceration—aspirin is the worst offender in this regard. The drugs cause this irritation because they block the synthesis of prostaglandins that the stomach lining needs to regulate overproduction of acid and to synthesize the mucus barrier that prevents its self-digestion.

In addition, most NSAIDs prevent the body from excreting salt and water properly, particularly when heart or liver disease compromises blood flow to the kidneys. When NSAIDs prevent the kidneys from synthesizing PGI_2 (a prostaglandin that causes blood vessels to dilate), the renal blood supply is reduced even further; patients sometimes accumulate enough fluid to choke their circulation.

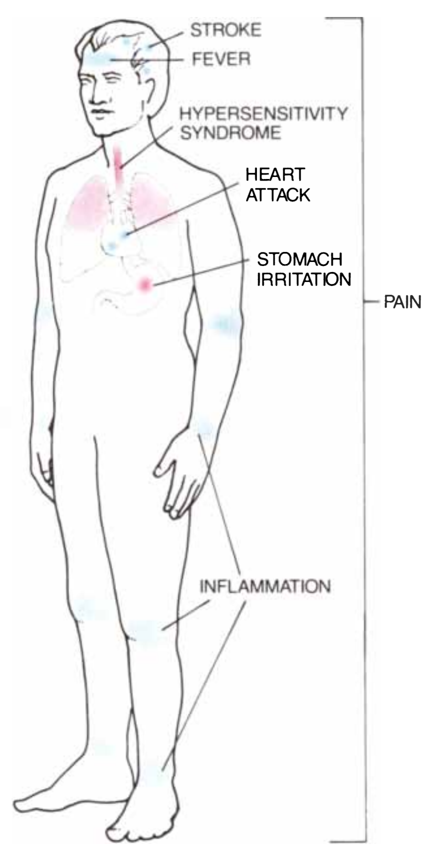
Another side effect of NSAIDs—but not sodium salicylate—is aspirin sensitivity syndrome in genetically susceptible patients. It turns out that blocking prostaglandin synthase diverts arachidonic acid to another pathway that transforms it to a host of substances—notably leukotriene B_4-, C_4 and D_4—that exceed in irritative potency the products of PGH synthase.

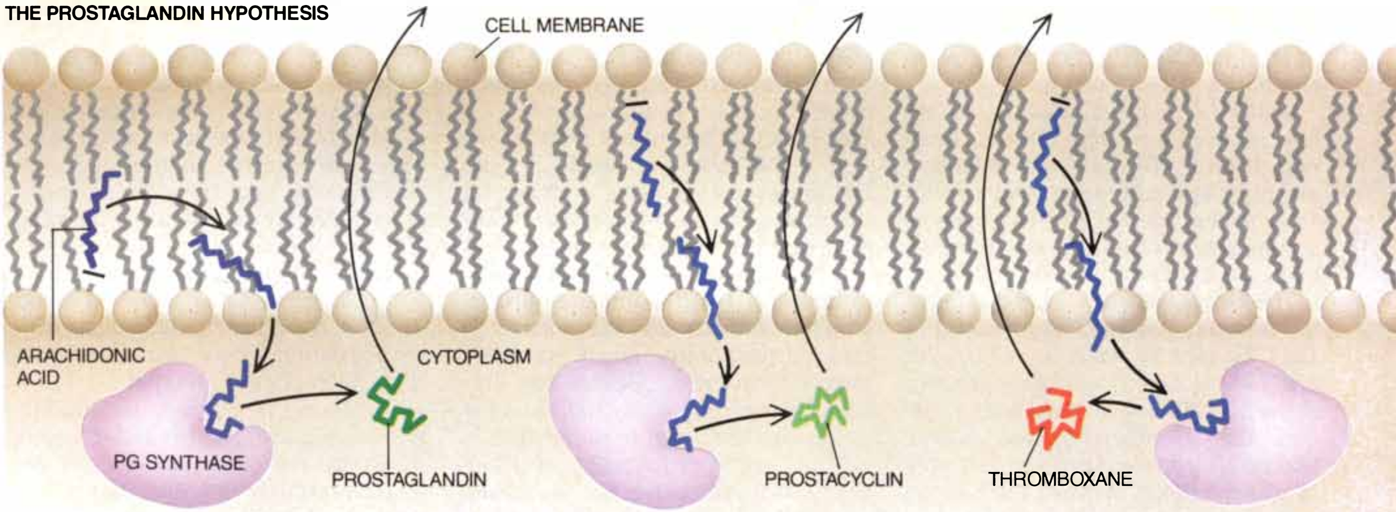
Finally, the most common side effect of NSAIDs, and especially aspirin, is interference with the clotting of blood. Patients who take these drugs sometimes suffer from untoward bleeding after tooth extraction, minor surgery or trauma. Aspirin inhibits platelet aggregation (the cellular aspect of blood clotting), and other NSAIDs—again with the exception of sodium salicylate—also inhibit platelet function.

NSAIDs act by blocking the production of both PGG_2 and PGH_2. Platelets transform the latter into a most potent vasoconstricting and platelet-aggregating substance, thromboxane B_2. Meanwhile the endothelial cells that line blood vessels use those same prostaglandin intermediates to make a potent vasodilator, prostacyclin, or PGI_2.

These seemingly arcane discoveries form the basis for using aspirin to prevent strokes and heart attacks: carefully calibrated doses can interfere with thromboxane production while leaving prostacyclin synthesis unaffect-

WHERE ASPIRIN WORKS





Swelling, heat and pain are caused by prostaglandins that cells synthesize in response to injury. The enzyme PG synthase transforms cell membrane component arachidonic acid into unstable prostaglandin intermediates and then into stable prostaglandins.

Endothelial cells turn prostaglandin intermediates into prostacyclin, a vasodilatcr, whereas platelets use them to synthesize thromboxane, a potent vasoconstrictor and clotting agent. As a result, selective inhibition of PG synthase in platelets can reduce the risk of heart attack and stroke.

ed. Aspirin irreversibly inactivates PGH synthase. Platelets cannot make more synthase and so make no more thromboxane. Endothelial cells, however, can make new synthase, so prostacyclin synthesis is inhibited only a few days.

Furthermore, Garret A. FitzGerald and John Oates of the Vanderbilt University School of Medicine have shown that less than one tablet a day can irreversibly block the PGH synthase activity of platelets in the portal circulation (the circulation that drains the intestines via the liver), thus reducing the risk of dangerous clots, before appreciable amounts of aspirin appear in the general circulation, where they might disrupt prostacyclin production.

No discovery that has evolved from Vane's hypothesis has had more of an impact on public health: the hundreds of thousands of patients around the world who now take aspirin to treat or prevent strokes and heart attacks owe Sir John a not inconsiderable debt.

The prostaglandin hypothesis certainly explained the effects of very low (antithrombotic) and intermediate (analgesic and antipyretic) doses of aspirin. Although there were some troubling discrepancies, Vane and his colleagues gave sound explanations for them. For example, acetaminophen does not keep prostaglandins from synthesizing, but it is effective against the synthase from the brain (perhaps explaining its antipyretic effect). And although nonacetylated salicylates are roughly one tenth as potent against prostaglandin synthase in vitro as aspirin—indicating that they should be ineffective analgesics—studies of prostaglandin metabolites show that sodium salicylate may indeed inhibit prostaglandin synthesis in the body.

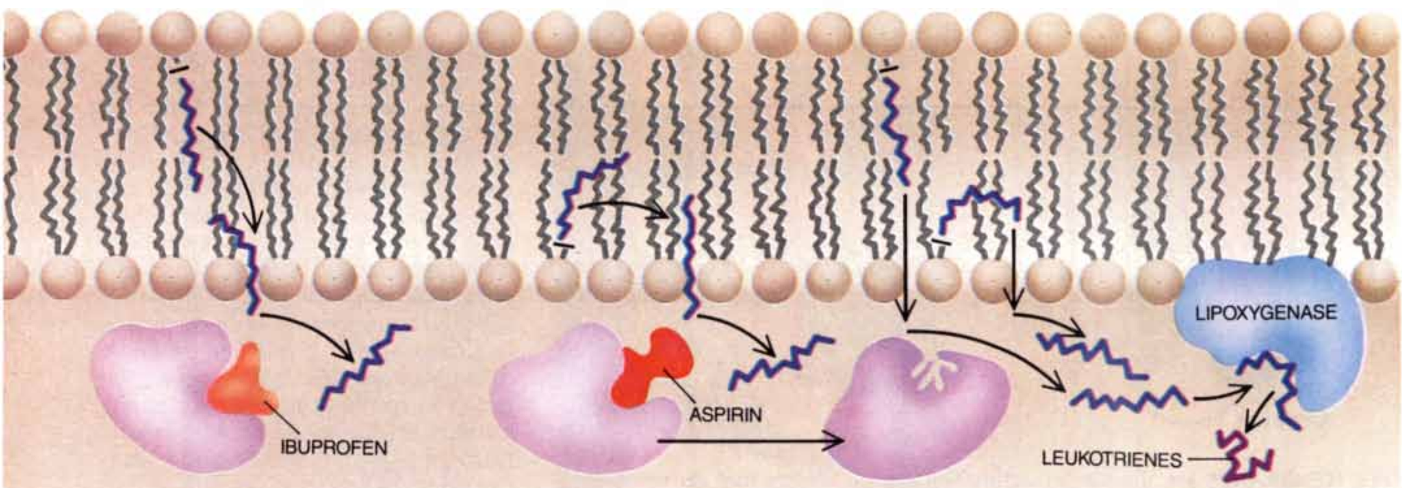
Vane's theory that the local production of prostaglandins leads to inflammation has only partly been substantiated, however. The anti-inflammatory effect of salicylates requires much higher doses than does analgesia. This discrepancy suggests either that the prostaglandin synthase of cells provoking inflammation is relatively insensitive to aspirin or that aspirin owes its anti-inflammatory property at higher concentrations to a mode of action distinct from its ability to inhibit prostaglandin synthesis.

The properties of sodium salicylate and acetaminophen provide further evidence that aspirinlike drugs exert clinical effects that do not depend on inhibiting prostaglandin synthesis. Sodium salicylate shares many of the analgesic properties of aspirin, but it fails to inhibit prostaglandin synthesis in disrupted cell preparations at concentrations that can be achieved in the body. It also does not inhibit platelet function or cause bleeding. And today's most widely used analgesic and antipyretic drug, acetaminophen, does not inhibit prostaglandin synthesis; nor does it keep blood from clotting or reduce inflammation. Pain and fever, it seems, can effectively be reduced without inhibiting the synthesis of prostaglandins at all.

The broad spectrum of NSAID effects most likely results from their physical properties, which permit them to disrupt interactions within biological membranes. NSAIDs are planar, anionic molecules that have an affinity for lipid environments such as the lipid bilayers of plasma membranes. Moreover, the more acidic the environment (as at inflammatory sites), the greater the lipophilicity of NSAIDs. It is therefore not surprising that these drugs interfere with many functions of inflammatory cells.

For example, aspirin alters the uptake of fatty acids and their insertion into the membranes of cultured human monocytes and macrophages. Salicylates also inhibit anion transport across a variety of cell membranes. Finally, NSAIDs inhibit bone metabolism and the synthesis of proteoglycan, a substance that forms the matrix of cartilage, by mechanisms that do not depend on the inhibition of the prostaglandin synthase. The last point does not merely weaken the prostaglandin hypothesis; it is also a matter of significant clinical concern.

Recent work in my laboratory has uncovered an alternative mechanism for the effects of aspirinlike drugs: interference with stimulus-response coupling in neutrophils, the most abundant cells of acute inflammation. These cells are the first line of defense against foreign intruders and among the first to cause injury in autoimmune diseases like rheumatoid arthritis. They damage tissues by releasing enzymes



Aspirin can selectively inhibit PG synthase in platelets because it permanently deactivates the enzyme, whereas some other aspirinlike drugs block the enzyme only temporarily. (Endothelial cells continue producing prostacyclin because they can make new synthase to replace that destroyed by the aspirin.)

Hypersensitivity syndrome afflicts some individuals exposed to aspirin. Once the action of PG synthase is blocked, a competing enzyme, lipoxygenase, transforms arachidonic acid into substances that have even more irritative power than prostaglandins.

that break down proteins (proteases), as well as inflammatory peptides, reactive oxygen species such as O_2^- and H_2O_2 (peroxide), and lipid irritants such as platelet-activating factor and leukotriene B_4.

Within five seconds after coming into contact with substances that provoke inflammation (immune complexes, complement components—a cascade of enzymes and bioactive peptides that interact with antibodies to provoke immune response—and other chemoattractants), the neutrophil is transformed into a secretory cell capable of provoking tissue injury. One of the first steps in tissue injury is neutrophil aggregation, or cell-cell adhesion. Both "homotypic" adhesion between neutrophils and "heterotypic" adhesion of neutrophils to the walls of blood vessels are required for cells to make their way out of the circulatory system and to cause inflammation.

Therapeutic concentrations of salicylates and NSAIDs, however, inhibit the cell-cell adhesion of human neutrophils. Furthermore, similar concentrations of sodium salicylate and aspirin have the same effect on neutrophils, even though they have widely divergent effects on prostaglandin synthase. It is therefore likely that the anti-inflammatory effect is related to the ability of both compounds to inhibit homotypic and heterotypic adhesion in neutrophils rather than to their unequal effect on prostaglandin synthesis.

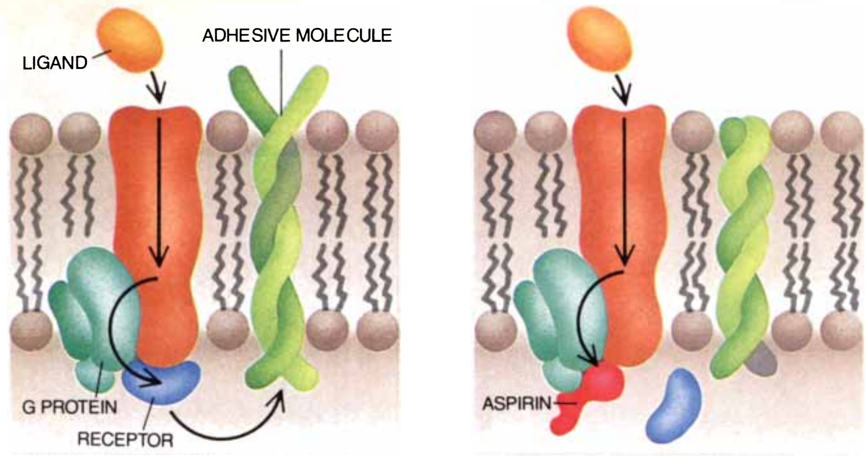
Inhibitory effects of NSAIDs on neutrophil activation can also be demonstrated in the clinic. The function of neutrophils derived from individuals given therapeutic doses of indomethacin, piroxicam or ibuprofen is significantly reduced. Neutrophils from the synovial fluid in the joints of patients with rheumatoid arthritis produced less superoxide anion—derivatives of molecular oxygen that can damage cells—after 10 days of therapy with piroxicam. And cells from normal volunteers given ibuprofen or piroxicam for three days failed to aggregate normally in response to chemoattractants.

All NSAIDs inhibit the homotypic adhesion of neutrophils, but they differ in their effects on other functions of the neutrophil. Piroxicam inhibits generation of superoxide anion by neutrophils exposed to various chemoattractants, but ibuprofen does not. Similarly, piroxicam and indomethacin inhibit the generation of superoxide anion by preparations of disrupted cells, but sodium salicylate, ibuprofen and the NSAID meclofenamate do not.

Vane's hypothesis is further weakened by findings from many laboratories, including mine, that the stable prostaglandins, E_1, E_2 and I_2, possess anti-inflammatory properties as well as proinflammatory ones. Robert B. Zurier of the University of Pennsylvania School of Medicine and others have shown that high doses of stable prostaglandins inhibit inflammation in animals with arthritis, and much lower doses inhibit similar inflammation induced by local skin irritants. It has also been clear since the early 1970s that PGI_2 and prostaglandins of the E type inhibit the activation in vitro not only of platelets but also of cells involved in inflammation, such as neutrophils and mononuclear phagocytes.

Paradoxically, both NSAIDs and prostaglandins of the E series similarly inhibit effects on the activation of neutrophils or platelets. Adding piroxicam to human neutrophils exposed to a chemoattractant can reduce superoxide anion generation by about 40 percent. Yet adding PGE_1 or PGE_2 at nanomolar to micromolar concentrations does not override the inhibition caused by piroxicam, as would be expected from the prostaglandins' proinflammatory capabilities. Instead it reduces superoxide anion generation by another 40 percent. More recent studies with the clinically useful PGE_1 derivative misoprostol also show synergistic rather than antagonistic effects between NSAIDs and prostaglandins.

Moreover, NSAIDs and prostaglandins have similar effects on the generation of such second messengers as calcium and cyclic adenosine monophosphate (cAMP), both of which transmit signals within cells. Increases in intracellular calcium induced by chemoattractants in human neutrophils are diminished not only by indomethacin but also by pretreatment of the cell with PGE_2. Chemoattractants increase levels of cAMP in the neutrophil slightly in order to provoke their effect. But prosta-



CELL MEMBRANES are the site of action for aspirin effects that do not depend on prostaglandin inhibition. The drug blocks transmission of chemical signals through the membranes by binding to a key regulatory protein. This prevents the first step in inflammation: adhesion of white blood cells to vessel walls.

glandins raise intracellular levels of cAMP by a larger amount, an effect that antagonizes cell activation. NSAIDs also enhance the increases of cAMP provoked by chemoattractants.

Some of the effect of NSAIDs on cells arises from interference with the binding of chemoattractants and other stimuli. These drugs inhibit the binding of at least some of these ligands with their receptors in the cell membrane, whereas acetaminophen, which fails to inhibit cell aggregation, has no effect on ligand binding.

The effects of NSAIDs on the binding of chemoattractants, however, are insufficient to explain their effects on neutrophils. NSAIDs inhibit cell activation in response to ligands whose binding they do not affect, such as $C5_2$ (a chemoattractant peptide), platelet-activating factor and leukotriene B_4. NSAIDs also inhibit activation in response to other stimuli. H. Daniel Perez of the University of California at San Francisco has shown that meclofenamate, for example, inhibits $C5_2$-induced neutrophil functions, without inhibiting binding of radiolabeled $C5_2$.

Because NSAIDs are acid, lipophilic molecules, they would be expected to alter membrane processes that depend on the overall mobility of membrane lipids. Salicylates at concentrations as low as 100 micromoles per liter decrease the viscosity of neutrophil membranes, whereas piroxicam and indomethacin increase viscosity at concentrations of 10 and 50 micromoles per liter, respectively. Acetaminophen, the analgesic, affects neither membrane viscosity nor the passing of chemical signals across the membrane.

Studies of purified membrane preparations and intact neutrophils show that NSAIDs interfere in particular with signals that depend on so-called G proteins for transduction through cell membranes. The evidence for this hypothesis begins with experiments that expose cells to pertussis toxin. This bacterial toxin interferes with signal transduction in a variety of cells, including the neutrophil, by altering certain G proteins in the plasma membrane. Neutrophils treated with toxin produce less superoxide anion when later exposed to chemoattractants.

Sodium salicylate similarly inhibits superoxide production—although far more modestly. Cells incubated with both pertussis toxin and sodium salicylate, however, regain their toxin-inhibited ability to generate superoxide anion. This paradoxical effect of salicylate suggests that salicylates interfere with the action of pertussis toxin near the site of its interaction with the G protein; they too must interact with the G protein in the cell membrane.

In addition, NSAIDs such as salicylate, piroxicam and indomethacin block the pertussis toxin-induced alteration of the G protein in purified neutrophil membranes. And salicylates and piroxicam inhibit in part other pertussis toxin-sensitive activities that follow cell activation.

All these NSAID effects have nothing to do with prostaglandin synthesis. A final blow to the generality of the prostaglandin hypothesis comes from one of the most primitive and ancient creatures, the marine sponge *Microconia prolifera*. This sponge, whose ancestry stretches back more than a billion years, offers a unique model for investigating the anti-inflammatory effects of NSAIDs.

The activation of sponge cells in the course of their aggregation is not influenced by stable prostaglandins, nor do the cells contain enzymes that could synthesize prostaglandins. Nevertheless, NSAIDs (but not acetaminophen) inhibit the aggregation of these cells just as they do that of neutrophils. Dispersed cells aggregate in response to a species-specific molecule weighing about 20 million daltons, called MAF. NSAIDs inhibit cell aggregation in response to MAF at concentrations similar to those that inhibit neutrophil aggregation. Because marine sponges cannot make prostaglandin, it is clear that these effects of NSAIDs—like those on insects, plants or human cells (neutrophils)—are unlikely to result from their inhibition of prostaglandin synthesis.

Vane's prostaglandin hypothesis explains a good part of the action of aspirin and related drugs, but much remains to be learned about the biology of these compounds as they interact with crucial systems of the cell. Nevertheless, it is reassuring to know that they have already helped to unravel biochemical pathways shared by creatures from which humans have been separated by a billion years of evolution.

FURTHER READING

INHIBITION OF PROSTAGLANDIN SYNTHESIS AS A MECHANISM OF ACTION FOR ASPIRIN-LIKE DRUGS. J. R. Vane in *Nature-New Biology,* Vol. 231, No. 25, pages 232–235; June 23, 1971.

NEW ASPECTS OF THE MODE OF ACTION OF NONSTEROIDAL ANTI-INFLAMMATORY DRUGS. S. H. Ferreira and J. R. Vane in *Annual Review of Pharmacology,* Vol. 14, pages 57–73; 1974.

DOSE-RELATED KINETICS OF ASPIRIN: PRESYSTEMIC ACETYLATION OF PLATELET CYCLOOXYGENASE. Anders K. Pederson and Garret A. FitzGerald in *New England Journal of Medicine,* Vol. 311, No. 19, pages 1206–1211; November 8, 1984.

MODES OF ACTION OF ASPIRIN-LIKE DRUGS. Steven Abramson, Helen Korchak, Rocio Ludewig, Henry Edelson, Kathleen Haines, Richard I. Levin, Robert Herman, Lisa Rider, Steven Kimmel and Gerald Weissmann in *Proceedings of the National Academy of Sciences,* Vol. 82, No. 21, pages 7227–7231; November 1985.

THE MECHANISMS OF ACTION OF NONSTEROIDAL ANTIINFLAMMATORY DRUGS. Steven B. Abramson and Gerald Weissmann in *Arthritis & Rheumatism,* Vol. 32, No. 1, pages 1–9; January 1989.

Coevolution of the Cuckoo and Its Hosts

The cuckoo reproduces at the expense of other birds by laying its eggs in their nests. The strategy provokes an evolutionary arms race between parasite and host

by Nicholas B. Davies and Michael Brooke

One of the most remarkable sights for a naturalist in summer is that of a little warbler feeding a young cuckoo (*Cuculus canorus*). When the nestling cuckoo is full grown, it overflows the tiny nest, and the warblers seem to risk being devoured themselves as they bow deep into the enormous gape with food. Once the cuckoo leaves the nest, the situation becomes even more extraordinary as the warblers perch on the cuckoo's back to reach the mouth of a fledgling eight times their own weight.

Although the parasitic habits of the cuckoo have been known at least since Aristotle wrote about them 2,300 years ago, it is only recently that biologists have undertaken detailed studies of how the cuckoo tricks hosts into rearing its young. The subterfuge costs the hosts dearly, in evolutionary terms, for although the cuckoo lays just one egg in a nest, the cuckoo hatchling ejects the host's own eggs or young and so becomes the sole occupant of the nest.

Given that the hosts gain no reproductive reward for their nurturing efforts, natural selection should therefore favor traits that help hosts thwart cuckoos. The evolution of such traits, in turn, should select for more sophisticated trickery by the parasite. The result is an evolutionary "arms race" between cuckoo and host, leading to ever more intricate adaptations and counteradaptations.

In our discussion we shall often have recourse to turns of phrase that might appear to suggest that evolution involves planning: strategies and counterstrategies, the seeking of benefits, the avoidance of costs. Such phrases can be useful provided they are recognized as metaphors. Natural selection lacks all intention; it works by favoring the reproduction of individuals possessing advantageous and heritable traits.

Many characteristics of organisms are the result of coevolution, that is, a process of reciprocal interaction with another group in which each party has adapted to selection pressures imposed by the other. Some examples of coevolution are mutualistic, with each party gaining a benefit. Thus, many hedgerow berries turn red in the fall to signal to birds that they are ripe for eating. The birds gain a nourishing meal, and the plant gains when its seeds are dispersed to new pastures by the birds. In other cases, such as cuckoos versus hosts or predators versus prey, one party gains a benefit while the other bears a cost. We were particularly attracted to the relationship between the cuckoo and its hosts not only because it provides an intriguing case study in natural history but also because it presents an excellent opportunity to unravel a case of coevolution by experiments in the field.

In Britain, cuckoos parasitize four main species of host: the meadow pipit (*Anthus pratensis*) in moorland, the reed warbler (*Acrocephalus scirpaceus*) in marshland, the dunnock (*Prunella modularis*) in woodland and farmland, and the pied wagtail (*Motacilla alba*) in open country. Individual female cuckoos are thought to specialize on one particular host species, and in Britain, therefore, there may be four genetically distinct strains of female cuckoo, referred to as gentes (singular, "gens"). Females of each gens lay a distinctive egg. Pipit-cuckoos lay brownish eggs, matching the pipit's own eggs. Reed warbler–cuckoos lay greenish eggs, mimicking the warbler's greenish eggs. Wagtail-cuckoos lay pale, grayish-white eggs, again matching those of their host. The exception is the dunnock-cuckoo, whose pale, spotted egg contrasts markedly with the immaculate blue eggs of the dunnock. Statistics compiled by the British Trust for Ornithology over the past 50 years show that in Britain as a whole, the proportions of host nests parasitized are 2.7 percent for meadow pipits, 5.5 percent for reed warblers, .4 percent for pied wagtails and 1.9 percent for dunnocks.

We considered two questions: first, how does the cuckoo trick the host and, second, how does natural selection adapt the host's behavior to meet the challenge? To answer the first question, we closely observed the reed warbler–cuckoos on Wicken Fen, one of the few remaining reserves of ancient wetland near Cambridge.

Reed warblers begin to build toward the end of May, weaving their nests

NICHOLAS B. DAVIES and MICHAEL BROOKE collaborate in the study of cuckoos at the University of Cambridge. Davies is a university lecturer in zoology and a fellow of Pembroke College. He previously served as demonstrator in the Edward Grey Institute at the University of Oxford. He has also published studies of dunnocks, wagtails and toads. Brooke served as warden of the Skokholm Bird Observatory, off the coast of Wales, where he studied the Manx shearwater, a native seabird. He managed a wildlife reserve in the Seychelles islands, took a research post at Oxford and then moved to Cambridge, where he is a senior research associate.

CHANGELING CUCKOO dwarfs its host, a dunnock. The dunnock apparently does not distinguish the stranger from its own young, which the cuckoo, on hatching, quickly ejects from the nest.

among vertical reed stems over water. We monitored 274 nests during egg laying and found that 44 of them (16 percent) were parasitized by cuckoos. In eight cases, the warblers rejected the cuckoo egg, either by throwing it from the nest (four cases) or by deserting the whole clutch (four cases). Although the hosts sometimes rebelled, in most cases the deception succeeded.

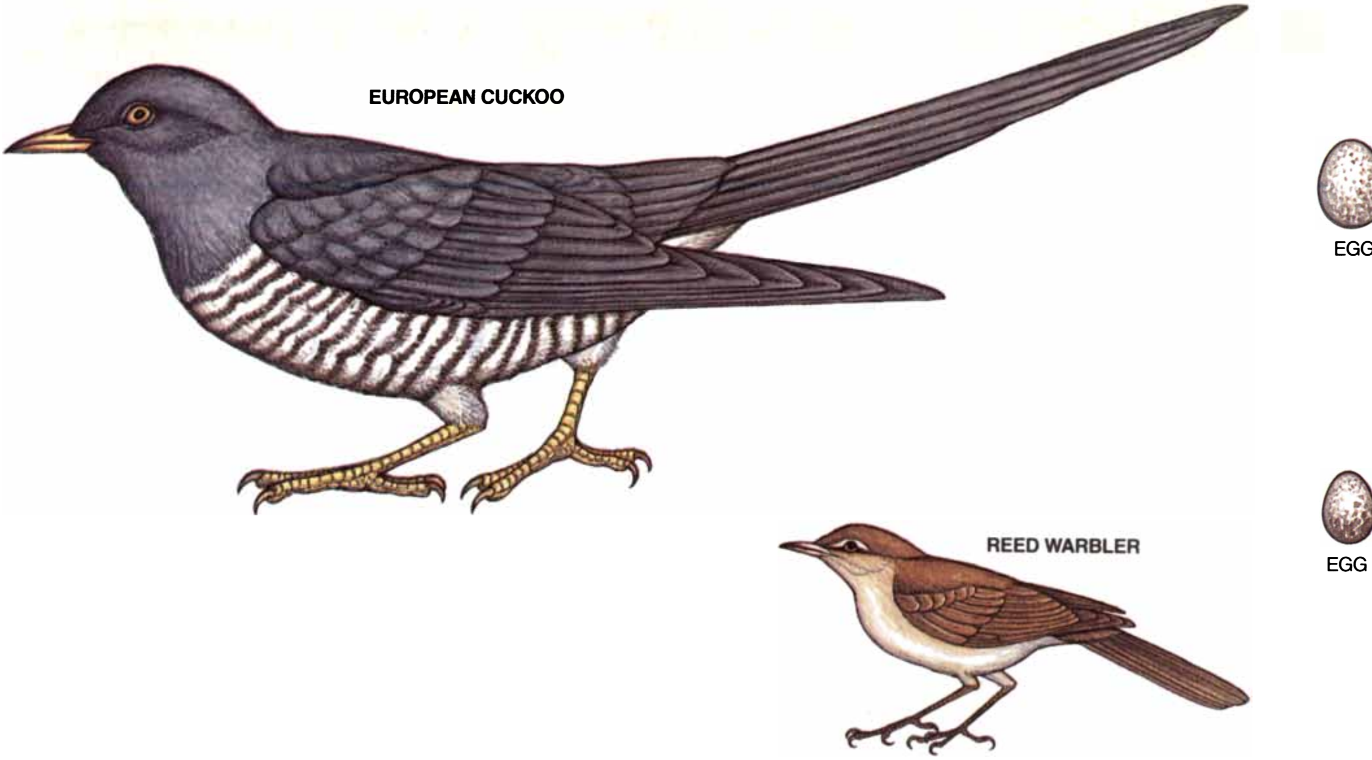
Cuckoo deception involves surveillance, stealth, surprise and speed. The procedure was first described 70 years ago in the pioneering study of pipit-cuckoos made by Edgar P. Chance, an English ornithologist, and it was recently shown to be the same for reed warbler-cuckoos by Ian Wyllie of Monk's Head Experimental Station in Huntingdon.

The female usually finds a nest by watching the hosts build. Then a few days later, during the host's laying period, she parasitizes the nest, usually between midday and late afternoon. Before laying, she remains quietly on a perch nearby, sometimes for an hour or more, waiting for the two host birds to absent themselves from the nest. She then suddenly glides down to the nest. She removes one host egg, sometimes more, lays her own directly into the nest and then flies off carrying the host egg in her bill. The trespass lasts less than 10 seconds, and it is difficult to believe she can have laid in such a short time. Yet a visit to the nest will reveal the cuckoo egg lying among the warblers' eggs, matching them well in color and pattern and identifiable only by its slightly larger size.

To investigate the extent to which natural selection has molded the cuckoo's behavior in relation to hosts' defenses, we used a simple but powerful experimental technique; we ourselves acted the part of the cuckoo by "parasitizing" reed warbler nests with model cuckoo eggs. Our idea was to vary the different aspects of the cuckoo's procedure and note how this affected the hosts' behavior. The model eggs, of the exact size and weight of real cuckoo eggs, were made of resin and painted to resemble the different eggs laid by various cuckoo gentes [*see illustration on page 96*].

We first copied the procedure of real cuckoos by removing an egg from a reed warbler nest one afternoon during the laying period and substituting a model cuckoo egg. The warblers accepted all the model eggs that resembled their own in color but rejected two thirds of those that did not, usually by pushing them out of the nest. Clearly, host discrimination against strange-looking eggs has selected for a mimetic cuckoo egg.

Host discrimination may not, however, be the only selective pressure favoring the evolution of mimetic cuckoo eggs: the parasites may discriminate, too. At six out of the 44 nests parasitized by cuckoos, a second cuckoo later visited the nest and laid an egg. (We were able to tell one reed warbler-cuckoo egg from another because each female lays eggs of a characteristic, identifiable color.) It would clearly pay a second cuckoo to remove the first cuckoo egg because, as the older egg, it is likely to hatch first, whereupon the chick will eject the other eggs. Some cuckoos visited nests in which we had placed a model egg, and indeed they tended to remove the model egg if it was unlike the host eggs in color. Discrimination by cuckoos themselves



therefore may have played a part in the evolution of egg mimicry.

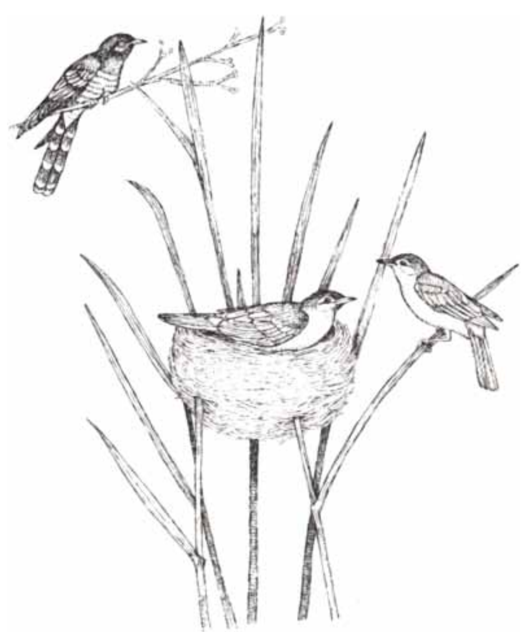
To test whether other features of the cuckoo's behavior were also tailored to circumvent the hosts' defenses, we used mimetic model eggs and varied each part of the laying procedure in turn. When a model was placed in the nest at dawn, during the host's laying period, the reed warblers often rejected it. Afternoon laying is therefore an important part of the cuckoo's trickery, perhaps because later in the day the warblers are less likely to be in attendance at the nest. Model eggs placed in nests before the hosts themselves had begun to lay were all rejected, so the hosts seem to adopt the rule "any egg appearing in the nest before I start to lay cannot be mine." This explains why cuckoos wait until hosts begin to lay before they parasitize a nest.

The remarkable speed of laying is also important. When we placed a stuffed cuckoo in the warblers' nest to simulate a female that was slow to lay, the warblers mobbed it vigorously and were subsequently more likely to reject even well-matching model eggs from their nest. Finally, the size of the cuckoo's egg is significant, for although it is a little larger than the hosts' eggs, it is still extraordinarily small for a bird the size of a cuckoo. Typically, a cuckoo-size bird would be expected to lay an egg about three times heavier. But when we placed such giant models in the warblers' nests, they were often rejected, even when painted to match the color of the warblers' eggs. Still, host discrimination may not be the only selective pressure in favor of a small cuckoo egg—large eggs may simply be difficult for small hosts to incubate efficiently.

All aspects of the cuckoo's strategy discussed so far are adapted to increase the chances that the host will accept the foreign eggs. One feature, however, apparently has nothing to do with the hosts' defenses: the cuckoo's habit of removing an egg before she lays. Our experiments show that the warblers accepted all mimetic models, and nonmimetic models stood the same chance of rejection, regardless of whether a host egg was removed or not. This finding shows the hosts are not alerted to the presence of a parasitic egg by counting an extra egg in their clutch.

The female cuckoo seems to derive two advantages from the removal of a host egg: she improves the efficiency with which her egg is incubated, and she gains a free meal. But why then does she not remove all of the eggs and thereby gain even more nourishment? Host responses again provide the answer: when a reed warbler clutch is reduced too drastically, the birds desert it. On the other hand, although the hosts always desert a single egg, they always stay to tend a single chick. This pattern of behavior explains very neatly why it is the cuckoo chick that takes on the task of ejecting the rest of the nest's contents, rather than its mother, earlier on during the egg stage. The chick does the work because it alone can eject all the hosts' eggs without penalty.

Reed warblers quickly note and reject strange eggs but show marked tol-



SURVEILLANCE of up to an hour or more opens the reed warbler–cuckoo's campaign against her prospective hosts.

CUCKOO FLEDGLING

REED WARBLER FLEDGLING

CUCKOO MIMICS its host in the egg phase but not in the chick. The drawings appear at one half the actual size.

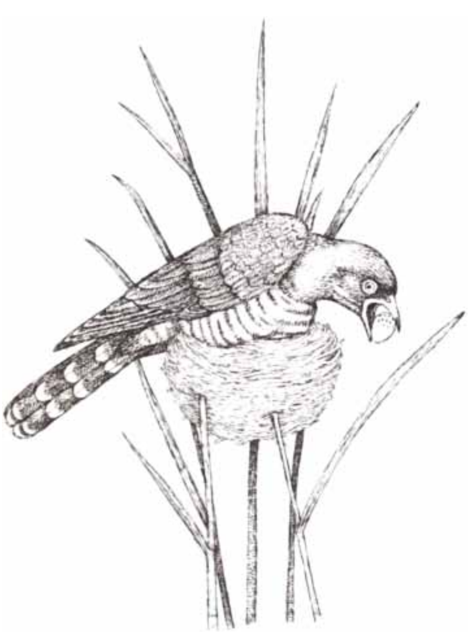
erance for strange chicks. Indeed, there is no evidence that hosts ever reject cuckoo chicks. Yet the newly hatched cuckoo's pink body and vivid orange gape is quite unlike the warbler chick's black skin and yellow gape with tongue spots. One might suppose that host parents tolerate cuckoo chicks because they have nothing to compare them with—the cuckoo chick sees to that by hatching first (it needs less incubation) and doing away with its rivals.

We tested this supposition by giving warblers the opportunity to compare a cuckoo chick with their own young. We strapped two nests side by side, one holding a cuckoo chick, the other holding a reed warbler chick. Faced with this choice, the hosts fed the contents of both nests indiscriminately. Not only did this experiment show that warblers do not favor their own young over cuckoo chicks even when they are given the chance to do so, it also disproves the hypothesis that cuckoo chicks somehow provide a supernormal stimulus that makes them more attractive than their hosts' own young. Moreover, further experiments showed that warblers accepted other kinds of different-looking chicks (such as the reed bunting, *Emberiza schoeniclus*) in among their own brood. Taken together, these results demonstrate that warblers simply show no discrimination at the chick stage and that the cuckoo has therefore not been forced, by natural selection, to evolve a mimetic chick.

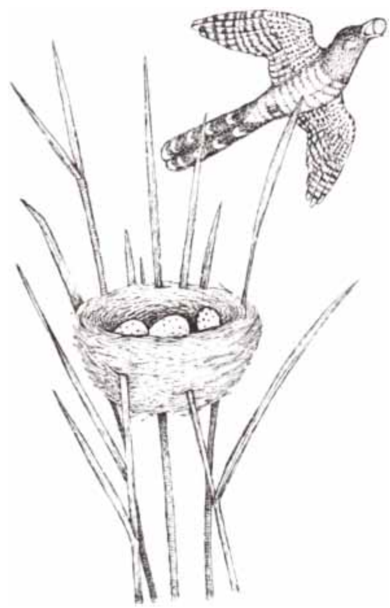
But why haven't the hosts evolved chick discrimination? Perhaps discrimination of young is more prone to error because, unlike eggs, the young change drastically in appearance from day to day. Moreover, chick discrimination may provide less of a selective advantage than egg discrimination because the host that spots a parasite egg can still save its brood—by the chick stage, however, it may be too late. Nevertheless, the absence of chick discrimination is still puzzling because in hosts of some other brood parasites, where the parasitic young are reared alongside the hosts' young (such as *Vidua* finch parasites of estrildid finch hosts), the parasitic young closely mimic the appearance of the host young [see "Mimicry in Parasitic Birds," by Jürgen Nicolai; SCIENTIFIC AMERICAN, October 1974]. These findings suggest that some host species do show chick discrimination.

The cuckoo chick is fed on the same types of prey as the reed warblers bring to a brood of their own young, namely, a variety of flies, bugs, caterpillars and other arthropods. The cuckoo chick is also provisioned at about the same rate as an average warbler brood (three or four chicks), suggesting that its development is adapted to a feeding rate the hosts can normally sustain.

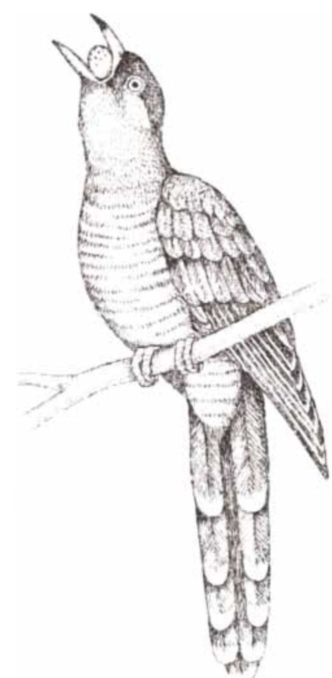
Female cuckoos defend territories in which they lay, on average, eight eggs a year. They lay on alternate days and usually in two batches, separated by several days' rest. If nests of the favorite host are in short supply, then the female may turn to alternative hosts. For



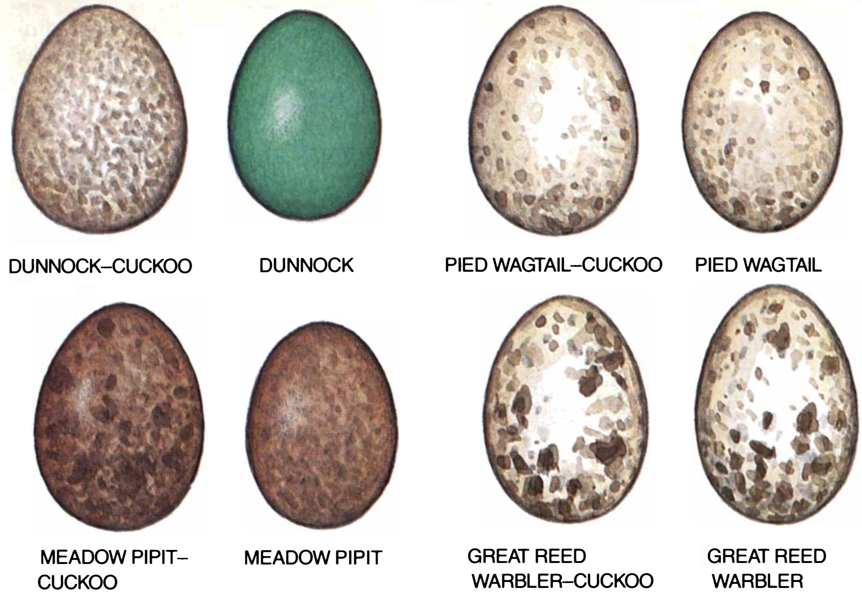
INVASION occurs when the hosts are absent from their nest. The cuckoo steals an egg and lays one of her own.



FLIGHT carries her from the scene of the trespass to safety. Elapsed time since the invasion: less than 10 seconds.



LUNCHING off her stolen egg, the cuckoo enjoys the first of the many free meals the hosts are doomed to provide.



VARYING MIMICRY is seen in these eggs of cuckoos and their hosts, drawn at about life size. All but the great reed warbler and the cuckoo that parasitizes it are taken from populations in Britain. Cuckoo eggs are only slightly larger than the eggs they mimic, although the cuckoo is generally more than twice the weight of its hosts.

example, reed warbler-cuckoos sometimes lay in the nests of sedge warblers (*Acrocephalus schoenobaenus*). The female cuckoo is also known to plunder nests that she finds at the incubation or chick stage, which are too advanced for parasitism. This attack induces the hosts to start a replacement clutch and so increases the availability of suitable nests.

We suspected that cuckoos were responsible for much of the predation of reed warbler nests on Wicken Fen. For one thing, many clutches were taken without the kind of obvious disturbance to the nest that normally marks the work of large avian predators such as crows; for another, most of the nests were over water and so probably inaccessible to mammalian predators. If cuckoos themselves depredated reed warbler nests, it would obviously pay them to remember the location of the nests they had parasitized and to leave these nests untouched. Our prediction was, therefore, that if the cuckoo is itself a major predator, then nests containing a cuckoo egg ought to suffer *less* predation than the average. Our data supported this prediction: only 22 percent of the parasitized nests suffered predation at the egg stage as compared with 41 percent of the unparasitized nests—a statistically significant difference.

We then turned to consider the three other major cuckoo hosts in Britain. Like reed warblers, both meadow pipits and pied wagtails tended to reject model eggs unlike their own in color and were more likely to accept a well-matching model representing the egg type laid by their own cuckoo gens. The mimetic eggs of pipit- and wagtail-cuckoos have thus been selected through host discrimination.

How then are we to explain the total absence of mimicry by dunnock-cuckoos? As Gilbert White noted in *The Natural History and Antiquities of Selborne* (1770), "You wonder...with good reason, that the hedge-sparrows [dunnocks]...can be induced at all to sit on the egg of the cuckoo without being scandalized at the vast disproportioned size of the supposititious egg; but the brute creation, I suppose, have very little idea of size, colour, or number."

Our experiments found a good reason why the dunnock-cuckoo has not evolved a mimetic egg: it does not need one, because the dunnock accepts eggs of any color. We wondered, therefore, if the dunnock was color-blind but found that even black or white model eggs were accepted. Moreover, dunnocks accepted not merely one strange egg but whole clutches of them. These conclusions leave unresolved the problem of why dunnock-cuckoos lay a distinctive egg type, intermediate in darkness between the eggs of reed warbler–cuckoos and pied wagtail–cuckoos and no more variable than the eggs laid by the other cuckoo gentes.

An interesting question for future research is how the different gentes remain distinct. One possibility is that daughters lay the same egg type as their mother, so that a female cuckoo hatching from a greenish egg, for example, will likewise lay greenish eggs. The cuckoo's problem would then be how to select the correct host, in this case the reed warbler, for whom its egg will be the appropriate match. A likely mechanism is that the young cuckoo imprints on the host that rears it and then, when adult, chooses to parasitize that same host species. Imprinting would thus ensure that the match between cuckoo and host egg is maintained over the generations. We are now testing this idea by transferring newly hatched cuckoos from one gens into the nest of a species normally parasitized by another gens, to see whether the swapped cuckoo, when adult, then prefers to parasitize the new host species.

If both male and female cuckoos imprint on their host, one might expect there to be four genetically distinct races of cuckoo in Britain; however, no differences in appearance can be cited to support this view. But if only the females imprint, there would be four distinct female lines, and interbreeding with males would maintain the unity of the species. In this scenario the gentes should differ with respect to the maternally inherited DNA of the mitochondria (intracellular organelles involved in energy metabolism) but not the DNA of the nuclei, which is inherited from both parents. We are studying such genetic differences among the gentes with Lisle Gibbs of Queen's University in Ontario.

Now we can consider the second main question posed at the start—namely, how do hosts evolve in response to cuckoos? Has the egg discrimination shown by reed warblers, meadow pipits and pied wagtails evolved specifically in response to cuckoo parasitism? If so, could the lack of discrimination by dunnocks mean that these birds are relatively recent victims, lagging behind in their counteradaptations to a new selective pressure?

We sought the answers by comparing the responses to model cuckoo eggs of a wide variety of passerine birds (the songbirds that constitute more than half of all living bird species). First we tested species suitable as hosts for cuckoos—those that have open nests accessible to a laying female cuckoo and that feed their young on invertebrates held in the bill. It turned out that these species exhibited varying degrees of rejection of model eggs unlike their own. Some, such as the reed bunting and spotted flycatcher (*Muscicapa* *striata*), were even more discriminating than the current favorite hosts. This unexpected discrimination may provide a clue to the evolutionary past of these species, a subject to which we shall presently return.

If rejection evolves only in response to cuckoo parasitism, then, we predicted, species unsuitable as hosts would show no rejection of eggs unlike their own because they have had no history of interaction with the cuckoo. Unsuitable hosts include species that, although their diet is suitable, nest in small holes or burrows and those that feed their young on seeds. Our prediction was strongly supported: eight of the nine unsuitable host species we tested showed little if any rejection of eggs unlike their own.

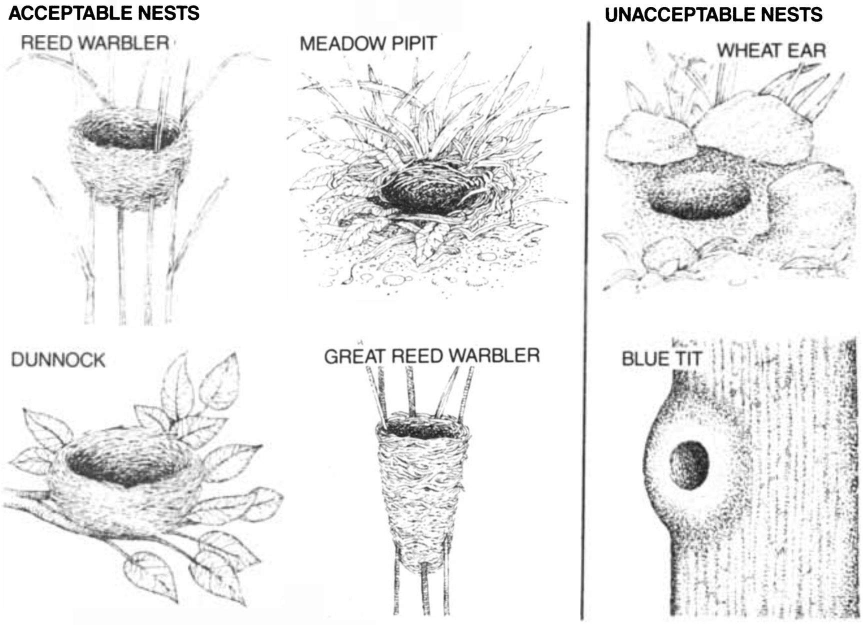
Some comparisons between closely related species were particularly revealing. Of the four finches we tested (family Fringillidae), only the one that feeds its young predominantly on invertebrates and is therefore suitable as a cuckoo host (the chaffinch *Fringilla coelebs*) showed strong rejection. The three unsuitable hosts, which feed their young mostly on seeds (the greenfinch *Chloris chloris*, the linnet *Acanthis cannabinna* and the bullfinch *Pyrrhula pyrrhula*), showed no rejection. Of the two flycatchers (family Muscicapidae), the spotted flycatcher—the open nests of which are exploitable by cuckoos—showed strong rejection, whereas the pied flycatcher (*Ficedula hypoleuca*), whose hole nests are inaccessible to cuckoos, showed no rejection at all. These data indicate that rejection is related not to the taxonomic group to which a species belongs but rather to the species' evolutionary experience with cuckoos.

Our comparative study strongly suggests that before cuckoos parasitized meadow pipits, pied wagtails and other current favorite hosts, these species showed no rejection of strange-looking eggs. Of course, we cannot go back in time to see if this was the case, but we could conduct the next best experiment. The cuckoo breeds from western Europe to Japan but not in Iceland, where it is only a rare vagrant and has never been known to breed. Iceland does, however, have isolated populations of meadow pipits and white wagtails (of which the pied wagtail is a subspecies), both of which are unparasitized. We therefore took our model eggs to Iceland.

Because the Icelandic populations have low breeding densities, we had to work hard to find nests, but the effort was more than repaid by the results.

ODD EGG of the cuckoo stands out among those of the dunnock, yet the dunnocks accept it as their own. Such poor mimicry by the parasite and poor discrimination by its host suggest that their evolutionary relationship is recent. This hypothesis is supported by reconstructions of the ecological history of the British Isles.

REED WARBLER-CUCKOO CHICK, still blind, balances the eggs of its hosts on its back one by one, then heaves them from the nest, so becoming the sole occupant.



EGG DISCRIMINATION in birds corresponds with how well their diets and domiciles suit cuckoos, which eat invertebrates and need open nests. Species that live this way tend to reject strange eggs. Species that eat seeds or nest in holes accept any egg. The findings imply that rejection evolves in response to parasitism by cuckoos.

Both pipits and wagtails showed much less discrimination against eggs unlike their own than did members of the parasitized populations of these same species in Britain. They did, however, reject some of the pure blue egg models, and so, unlike the unsuitable host species, they were not completely naive. Possibly the Icelandic populations were derived from parasitized populations from other parts of Europe and have inherited some of their ancestors' egg discrimination.

The experiments with model eggs allowed us to reconstruct the likely stages in the evolutionary arms race between the cuckoo and its hosts: A species at first shows little if any rejection of strange eggs. Cuckoo parasitism then selects for host discrimination against eggs unlike the host's. This host discrimination selects for egg mimicry by cuckoos, which, in turn, selects for hosts that are better at seeing through the egg's disguise, producing a finer host discrimination that induces still closer mimicry. We suggest that the varying degrees of rejection and of egg mimicry shown by different host species and cuckoo gentes may represent different stages in this arms race.

Dunnocks, with their complete lack of discrimination, may provide a snapshot of a particularly early stage—early, that is, in the evolutionary time scale. Shakespeare referred to the parasitism suffered by dunnocks, known more commonly as hedge sparrows, 385 years ago in *King Lear* (act 1, scene 4), where the Fool warns Lear that his daughters may prove to be his ruin, as when "the hedge-sparrow fed the cuckoo so long, that it's had it head bit off by it young." Even older is Chaucer's reference in "The Parlement of Foules" (circa 1382), where the cuckoo is chastised as "thou mordrer of the heysugge on the braunche that broghte thee forth"; *heysugge* being Middle English for hedge sparrow.

Although we know the dunnock has been a victim for at least 600 years, only 2 percent of British dunnock nests are now parasitized, and at this meager rate it could take thousands of generations—and years—for egg discrimination to spread through the host population. Most of Britain was covered in primeval woodland from 6,000 to 8,500 years ago, and the dunnock is not common in such an environment. Possibly it did not become a victim of the cuckoo until extensive forest clearance occurred between 2,500 and 6,500 years ago, within the time calculations suggest it would take for discrimination to spread.

Reed warblers, meadow pipits and pied wagtails in Britain may represent a more advanced stage in the arms race, where hosts show rejection of unlike eggs and the cuckoo gentes have evolved quite good mimetic eggs. In central Europe, where there are larger tracts of undisturbed habitat, one finds examples of even better egg mimicry by cuckoos, perhaps reflecting even finer host discrimination. These include the beautiful blue mimetic eggs laid by the cuckoo gens that specializes on redstarts (*Phoenicurus phoenicurus*) and the wonderfully detailed mimicry shown by the great reed warbler (*Acrocephalus arundinaceus*) gens, where every spot on the host egg seems to have been copied to perfection. In central Asia and Africa, where cuckoos probably originated and where several species coexist, each specializing on different hosts, there are examples of such perfect mimicry that the only sure way to recognize the parasite's egg is to weigh the shell. Here interactions between parasite and host have presumably gone on longest of all.

In conclusion, our experiments reveal a true case of coevolution. Cuckoos have obviously responded to their hosts: their egg-laying procedure seems largely designed to defeat host defenses, and the refinement of the egg mimicry of the different gentes reflects the degree of discrimination shown by their respective hosts. Hosts, in turn, have responded to cuckoos: species with no history of parasitism show no egg rejection, and populations of favorite host species that have long been isolated from cuckoos show less egg discrimination than do parasitized populations.

Perfection of parasite mimicry and host discrimination may not be the only outcome of coevolution. Our experiments suggest another possibility. Among suitable hosts, some species now rarely if ever parasitized show stronger rejection of strange eggs than do the hosts currently favored by the cuckoo. This finding implies that those species bear the scars of an arms race their ancestors ran against the cuckoo long ago. Maybe they evolved such strong egg discrimination that they drove their cuckoo gentes to extinction—or into the nests of other species as yet unhardened by the struggle.

FURTHER READING

THE CUCKOO'S SECRET. Edgar P. Chance. London, Sedgewick & Jackson, 1922.

AN EXPERIMENTAL AND TELEONOMIC INVESTIGATION OF AVIAN BROOD PARASITISM. Stephen I. Rothstein in *Condor*, Vol. 77, No. 3, pages 250–271; Autumn 1975.

THE CUCKOO. Ian Wyllie. London, Batsford, 1981.

CUCKOOS VERSUS REED WARBLERS ADAPTATIONS AND COUNTERADAPTATIONS. N. B. Davies and M. de L. Brooke in *Animal Behaviour*, Vol. 36, Part 1, pages 262–284; February 1988.

Will this become the only way to save the rain forests?

Fifty-four irreplaceable acres of tropical rain forest are wiped out every minute. Yet we need every acre desperately.

Rain forests house more than half the world's animals and plants. And they're crucial to maintain the earth's fresh water supply. Without rain forests, we could be living in a burning, barren desert.

That's why World Wildlife Fund needs your help to save the rain forests and thousands of species that call them home. We're fighting for their survival and, ultimately, our own.

So send for information now. Help us save life on earth. Otherwise, better save this picture.

World Wildlife Fund

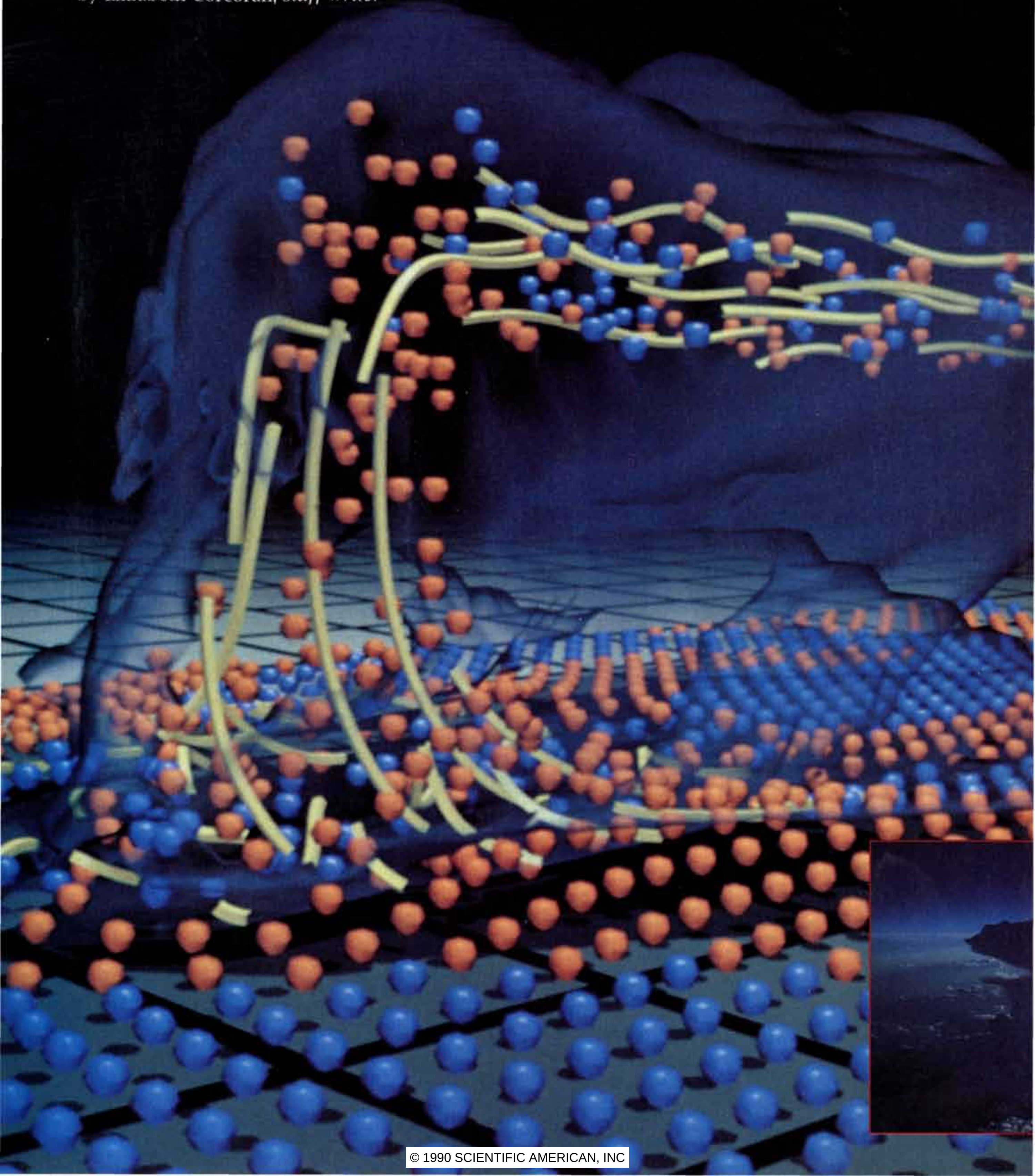
Dept. A, 1250 24th St. NW, Washington, D.C. 20037

Prepared as a public service by Ogilvy & Mather.

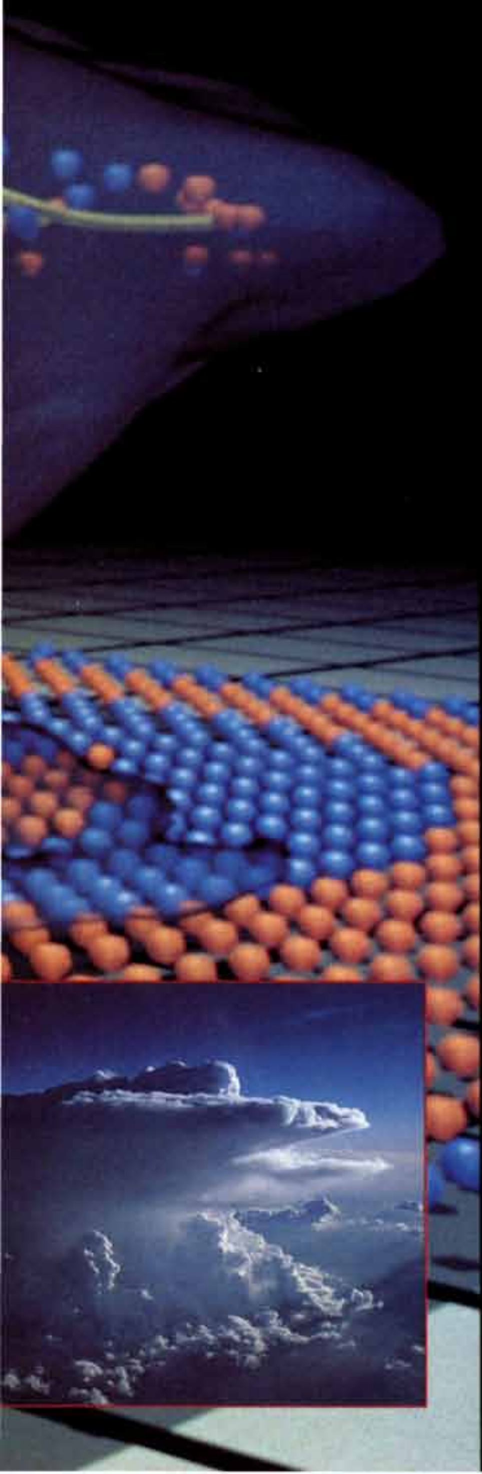
WWF

CALCULATING REALITY

by Elizabeth Corcoran, *staff writer*



Defying traditional designs, supercomputer architects are racing to build machines that are powerful enough to transform science.



MOMENT IN THE LIFE OF A SEVERE STORM simulated on a supercomputer shows the turbulence of air within a cloud mass, which is about 45 minutes old.

To map the flow of air, the simulation releases weightless tracer particles in the path of the storm from a horizontal plane about one kilometer above the ground. Particles are colored orange when rising, blue when sinking. Yellow ribbons attached to some particles show the path of air during the previous 500 seconds.

Researchers at the National Center for Supercomputing Applications and the department of atmospheric sciences at the University of Illinois at Urbana-Champaign hope these simulations will improve understanding of storms and lead to more accurate predictions of the behavior of actual storms (inset).

A clear, late autumn sky stretches above the flat Wisconsin farmlands like a blank slate. From the windows of Supercomputer Systems, Inc., in Eau Claire, little else is visible to computer architects inside the office as they endeavor to fill their design boards with plans for what they hope will be the fastest supercomputer in the world.

SSI is only a fledgling challenger, though. A few miles away in Chippewa Falls is the established master of the supercomputer art, Cray Research, Inc. There workers assembling the next generation of Cray supercomputers proudly wear sky-blue jackets with embroidered patches that declare: "Cray: World's Fastest Computers." Even now, however, Cray's hold on the title is slipping.

Several hundred miles south, investigators at the University of Illinois at Urbana-Champaign see a different challenge in that midwestern sky. They are preoccupied with its perturbations—namely, convulsive storms. With the enormous computation power of various supercomputers in hand, the researchers are trying to untangle why some storms grow severe. Yet their computers still lack the agility to portray tornados, those small vicious storms that cause billions of dollars in damage and kill as many as 80 people in the U.S. every year.

The Illinois researchers nurture a hope shared by supercomputer users everywhere: that by the end of the decade, the competing teams of supercomputer designers will have produced machines with enough speed and memory to illustrate more nearly precise snapshots of some of the most complicated phenomena in nature. Such computers would no longer be just dizzyingly fast engines of computation. Instead they would become the cross-disciplinary equals of the superconducting supercollider or the scanning tunneling microscope—portals to new insights and questions in virtually every scientific field.

From the perspective of supercomputer designers, their decathalon is best described by three *T*s—a trillion operations a second, a trillion bytes of memory and data communications rates of a trillion bytes per second. In each category, that is almost 1,000 times the capability of existing supercomputers. To reach these "tera" levels of performance, most believe they must abandon traditional designs, which treat problems in a step-by-step, or serial, fashion. In this way, tera computers will take on some flavor of parallel computation, in which the massive problems are resolved by being broken first into pieces, then reassembled.

Yet not since the first modern computer was constructed in the 1940s has the organization, or architecture, of computers been so open to reinterpretation and radical rethinking. Including SSI and Cray Research, about a dozen competitors are laying plans for the next generations of the highest-performance computers.

It is a neo-Darwinian struggle in which survival will depend on far more than just building the fastest machine. And there is a perverse factor afoot: those who could make best use of the vast computation power of the emerging supercomputers are not necessarily computer gurus. Rather they are experts in their own scientific disciplines but are often disinclined to devote enormous time to trying to program a computer. Building a computer that can achieve a teraflops (a trillion floating-point operations per second) is simple compared with designing a system that researchers can easily use.

Robert B. Wilhelmson, an atmospheric scientist at the National Center for Supercomputing Applications at the University of Illinois, is one such researcher. Currently at his command is a Cray 2, theoretically capable of churning through almost two billion floating-point operations per second, or giga-

EVOLVING MODELS

Over the past decade, storm simulations have revealed a wealth of information. In 1982 Peter S. Ray (Florida State University) and Robert B. Wilhelmson (University of Illinois) portrayed wind and precipitation with a physical model. They strung intersecting wires between layers of Plexiglas, then glued on molded plastic arrows (*left*). In 1985 Joseph B. Klemp of the National Center for Atmospheric Research in Boulder, Colo., used a computer and color-graphics display to create static storm pictures (*center*). Two years later, working with a supercomputer and a graphics animator, Wilhelmson compiled a short movie of the evolution of a storm (*right*).

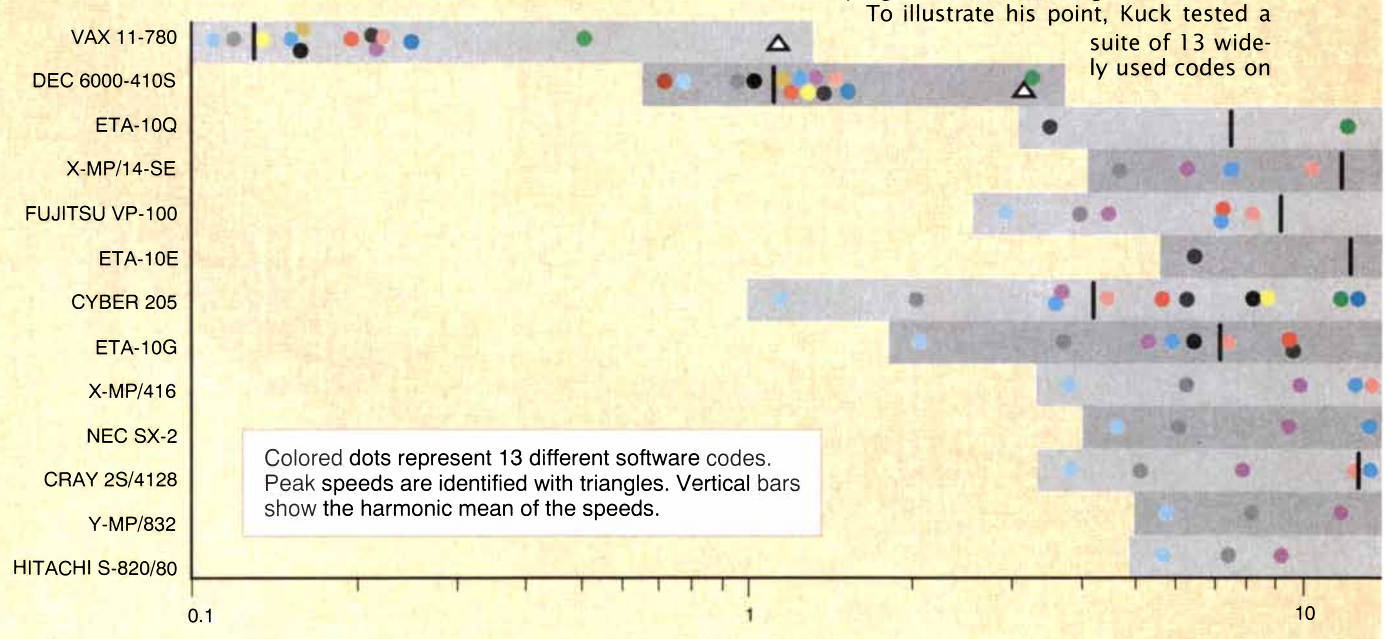
HOW FAST IS FAST?

Asking a dozen supercomputer users and manufacturers to rank the fastest machines guarantees a quarrel.

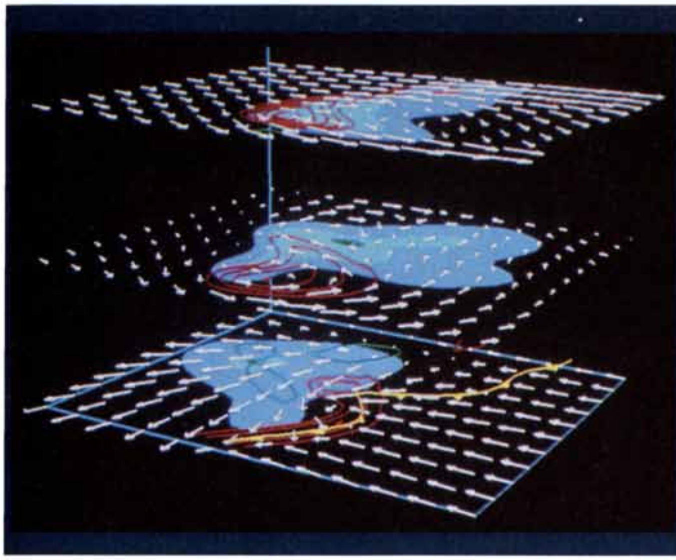
There is good reason for the controversy. The most useful way to measure such speeds is to run a program, or algorithm, and calculate how many floating-point operations the machine churns through in a second. The problem is that subtle differences in codes will make various supercomputers seem significantly faster or slower.

Such variability in computer performance irks David J. Kuck, who directs the Center for Supercomputing Research and Development at the University of Illinois. "There aren't any 100-megaflops machines when it comes down to ordinary serious users sitting down and running one program and other," he grumbles.

To illustrate his point, Kuck tested a suite of 13 widely used codes on



SOURCE: CENTER FOR SUPERCOMPUTING RESEARCH AND DEVELOPMENT



flops. With it, Wilhelmson built a short, animated movie of the evolution of a storm. He and his colleagues had to spend a tedious year making the video.

Give him a teraflops supercomputer, Wilhelmson says, and he could make such a movie in a few hours, then change parameters and run the simulation again. Wilhelmson is not alone in his thirst for more computational brawn. Almost a decade ago Nobel laureate Kenneth Wilson suggested a collection of scientific questions that presented "grand challenges" for the research community. Unraveling the human genome and predicting global climate changes over decades or even centuries are two of them. To take on such questions, scientists need tera levels of computer performance, declares Justin R. Rattner, director of technology at Intel Corporation's supercomputer systems division in Beaverton, Ore.

Such computing power is not readily available. Sometime in 1992 Cray Research will begin shipping its next generation machine, the Y-MP/16, which will boast a peak speed of about 16 gigaflops and may regularly achieve 10 gigaflops. That performance is still far short of a teraflops.

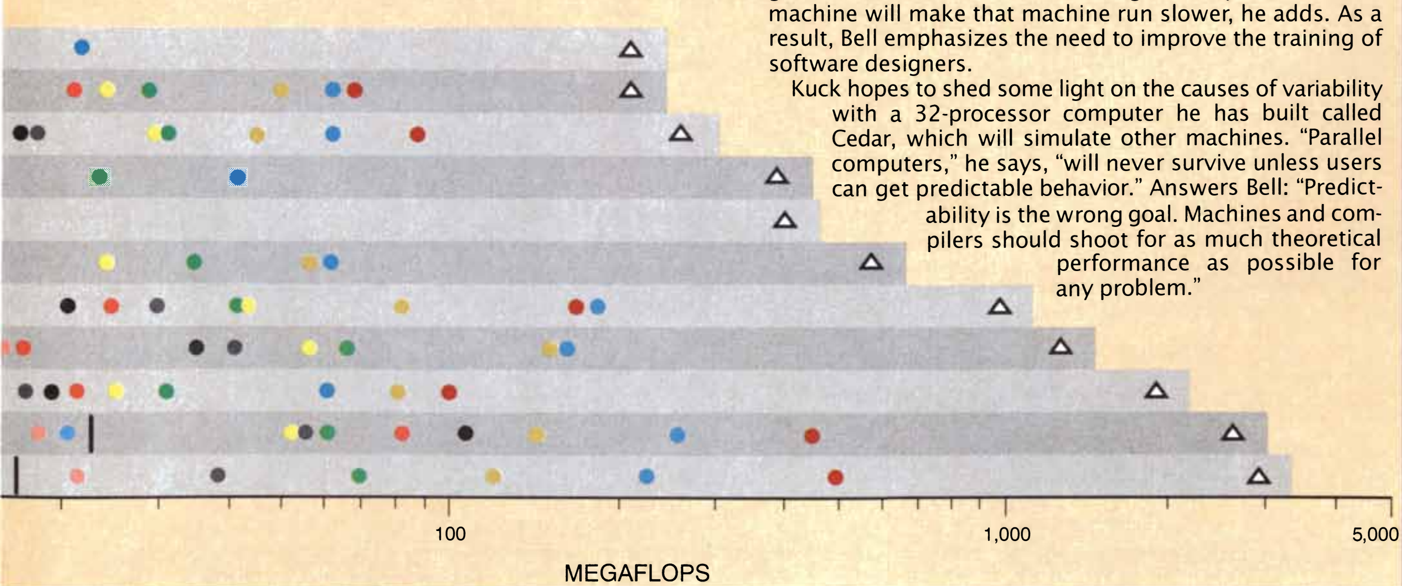
The Y-MP/16 will mark the end of an era for Cray Research, the company

a variety of computers, including the 11 supercomputers and two workstations charted below. The data show great variability: most of the supercomputers run some codes up to 100 times faster than others. Comparable supercomputers run the same code at different rates. And none of the supercomputers approached its supposed peak.

"These problems are largely ignored by industry and make for a disaster from the user's point of view," Kuck charges. "First companies deny it. Then after a lot of discussion they say, 'True, but why does it bother you? These are difficult machines.'" Kuck's answer: no one outside the research community has the patience to cope with such unpredictable behavior. Instead commercial users want the predictability of workstations such as the VAX 11-780 and DEC 6000 series. Although few of Kuck's "perfect 13" codes have been adapted to run on massively parallel computers, he expects that those architectures would demonstrate even more variability.

"That's the nature of engineering," retorts C. Gordon Bell, chief architect of the VAX. Computers are designed with differing degrees of parallelism, he says, as is software. Programs that do not take full advantage of the parallelism of a machine will make that machine run slower, he adds. As a result, Bell emphasizes the need to improve the training of software designers.

Kuck hopes to shed some light on the causes of variability with a 32-processor computer he has built called Cedar, which will simulate other machines. "Parallel computers," he says, "will never survive unless users can get predictable behavior." Answers Bell: "Predictability is the wrong goal. Machines and compilers should shoot for as much theoretical performance as possible for any problem."



Teraflops Design in Eight Easy Steps

Designing a supercomputer that will carry out a trillion operations a second forces a researcher to confront a maze of issues. At *Scientific American*'s request, David A. Patterson, a professor at the University of California at Berkeley and coauthor of *Computer Architectures—A Quantitative Approach*, put together one pass at building a fictitious "TF-1."

1. What large questions cannot be addressed with existing supercomputers?
The designer should have some problems in mind. There are also "grand challenge" questions, including mapping the human genome and predicting global climate changes. Focus on a few such applications.

2. How much money is available?
No one will pay $1 trillion for a teraflops machine; no one can build a TF-1 for $1,000. Current supercomputers cost between $10 million and $30 million and perform about 0.1 to 0.5 percent of a teraflops. As long as the development budget is in the neighborhood of $50 million to $500 million, the designer need not worry about money—yet.

3. How many applications will customers be willing to pay more than $25 million to solve?
If this list is empty, then the designer is doomed. If there is only one potential application, the designer should consider building a special-purpose architecture to solve that specific problem at a much lower cost. Assuming the designer can imagine a collection of worthy applications, the effort to build a general-purpose TF-1 should proceed.

4. Characterize the degree of parallelism in the likely applications.
Characterizing parallelism broadly means uncovering repetition in a problem. For instance, modeling the flow of air can be a highly parallel problem because the programmer need apply only one force to many air particles. A series of nonlinear equations, each of whose results must be fed into the next equation, exhibits little parallelism.
Using those characteristics, answer the following:

5. What is the average parallelism in the problems?
Suppose an application models the behavior of four particles pushed by the same force. There will be little reason to use a machine that has, say, eight or 16 processors to solve this problem. Building such a machine may cost twice as much as a simpler machine and not solve the problems twice as fast.

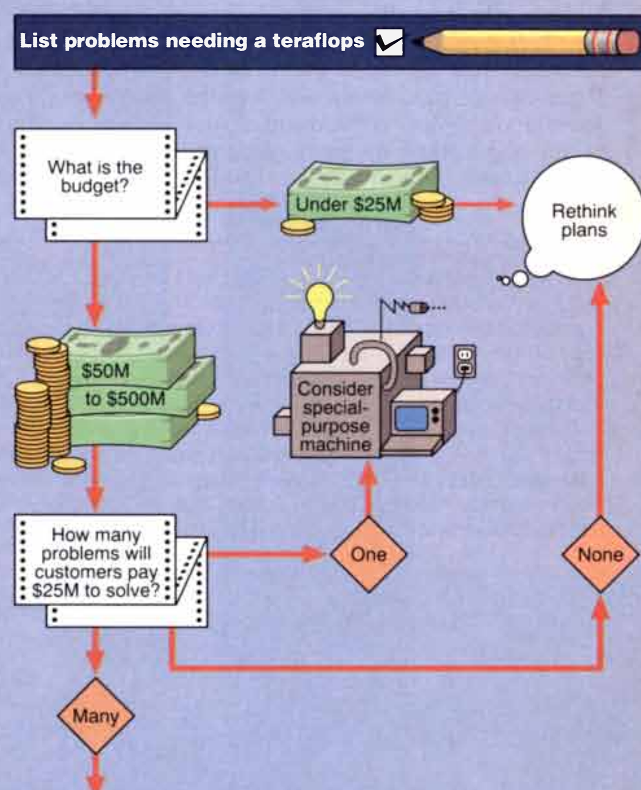
If problems can be readily divided among 10 processors, then each processor will have to run its portion of the application at about 100 gigaflops to solve the problem at a teraflops rate. This speed is between 25 and 100 times faster than the fastest computer today. Processors are unlikely to achieve such blistering speeds before the end of the decade. Applications with some 10,000 parallel elements can be solved with thousands of 100-megaflops processors. Such speeds should be possible for single-chip processors within the next few years.

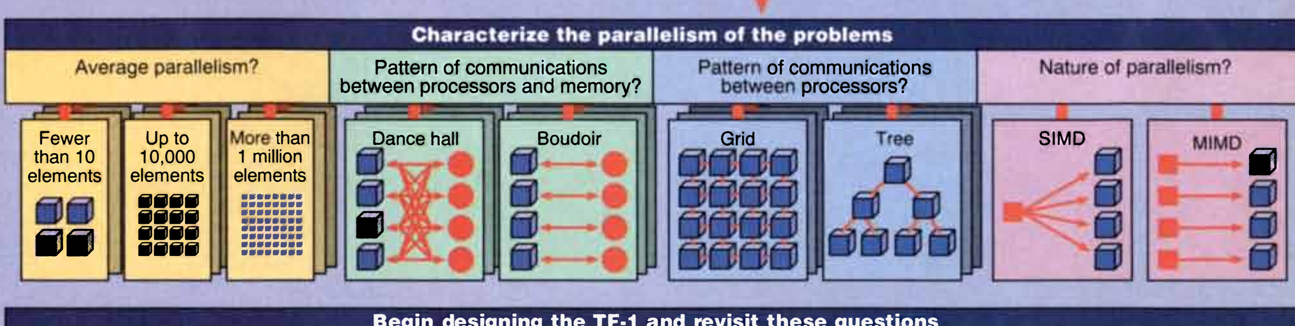
To treat more than a million parallel elements, the designer must decide whether it would be more cost-effective to use a million tiny processors packed 100 to a chip or 10,000 more powerful units, each located on one chip and handling many parallel tasks.

6. How frequently are processors likely to need to retrieve data from particular banks of memory?
Computing requires that processors reach into memory both for data and for instructions. A machine that is 1,000 times faster than current supercomputers should have a memory that retrieves information 1,000 times faster and is about 1,000 times larger than current memories. But the larger the memory, the slower the computer. So the TF-1 designer may well turn to 1,000 independent memories that work simultaneously.

If the amount of communication, or sharing, between parallel activities is likely to be irregular, the designer could use a "dance hall" approach. All the processors line up on one side, the memory units line up on the other, and they pair off. But as the number of processors increases, this design becomes exceedingly complex.

If communication needs are likely to be minimal, the designer can divide the memory into as many pieces as there are processors and associate a piece of memory with every processor. In this "boudoir" configuration, most of the traffic of data and instructions takes place between a processor and its local memory.





7. What is the pattern of communications between processors?

Processors must also be linked in a way that enables them to reach one another and swap information or coordinate their activities.

The simplest and cheapest interconnection mechanism is a bus—a set of wires that connects all elements together. Because only one device at a time can transfer data or instructions along a bus, this approach is too limited when linking many processors. Currently the most expensive scheme is a crossbar switch, which provides an explicit path between every communicating device. This becomes prohibitively expensive when connecting thousands of processors.

Between these two solutions are many imaginative connection schemes that trade off the number of simultaneous transfers of information, the speed of those transfers and the cost of the connections. One example is a grid, in which information is sent to a processor's nearest neighbor. (Such a design is ideal when processors need only communicate with neighbors one step away.) Extending this idea to multiple dimensions leads to a hypercube design. A tree design conveys information along various branches.

The problem with such restricted topologies is that the programmer may have to understand how the processors on the parallel machine are linked to solve a problem efficiently. The goal is to select a topology that can efficiently handle a wide variety of applications with different amounts and patterns of communication.

8. What is the nature of parallelism? How should the programmer control the processors?

The character of parallelism in potential problems may help the designer decide how programs should control the operations of the machine. If the parallelism of the application can be controlled by a single sequence of instructions that operates on many sets of data, then the designer can use a single-instruction stream, multiple-data stream, or SIMD, architecture. In SIMD machines, a single, separate memory supplies all instructions—much like a conductor directing an orchestra.

SIMD synchronization has both strengths and weaknesses. Such implicit synchronization can simplify the task of the programmer, making SIMD look like a serial-processing machine. But there may be too much synchronization if some pieces of applications need different processors to carry out different tasks.

If the applications will demand many independent sets of instruction sequences, then the architecture should take the form of a multiple-instruction stream, multiple-data stream, or MIMD, computer. This machine is clearly more general than the SIMD variety, but MIMD processors must periodically ensure they are working in harmony with other processors. Here, too, the driving issue is the cost-to-performance ratio of the choices.

where supercomputing lore began. Disputes over schemes for the next supercomputer generations have splintered the firm. Seymour Cray, supercomputer patriarch and architect of the Cray 1 and 2, decided his next machine (the Cray 3) would reach high speeds by relying on a faster electronic material, gallium arsenide. Building gallium arsenide chips, however, became exceedingly time-consuming and costly. So, less than two years ago, Cray closed his Wisconsin workshop and started another firm, the Cray Computer Corporation in Colorado Springs.

Cray's departure followed that of Steve S. Chen, the young designer of the X-MP and Y-MP series. Chen marched out to found Supercomputing Systems when Cray Research executives decided not to pursue his next architecture, which they described as too ambitious. The designers who have remained at Cray Research, meanwhile, have set out on a course that veers radically from the Cray tradition. They are committed to building a massive parallel system that relies on thousands of processors instead of just a few.

This abrupt parting of ways reflects some of the divergence in the fundamentals of supercomputing architectures. When Cray began designing computers in the 1950s, he, like many others, built on the skeleton developed a decade earlier by the architects of the ENIAC, one of the earliest computers. Programs and data were stored as numbers in a central memory. A single processor interpreted the instructions one at a time and processed the data. An internal clock metered the pace of the operations.

For decades, building a fast computer meant speeding up the clock—literally, reducing the time between independent instructions. At first, it was easy: simply build faster components. Then, by the early 1970s, improvements in serial speed slowed dramatically. So Cray and his research team became wizards at exploiting detours around this bottleneck. They minimized the distances electronic signals had to travel by packing components more tightly together and developed innovative techniques for flushing heat from circuits to prevent them from burning up.

Following the lead of others, Cray also began taking modest steps toward parallelism. Rather than letting the central processor carry out every step of a task in sequential order, Cray divided up a task. A method called pipelining split the central processor into an assembly line of cooperative subunits. One subunit would carry out the first step in a task, then pass its results to another. As the second subunit acted on its step, the first would begin the next task. Pipelining led to vector processing, in which similar operations were applied simultaneously to every number in an ordered array, or vector.

Cray Research's other star designer, Chen, pushed for more speed on a different front: adding more processors. The Cray X-MP, unveiled in 1982, was essentially two Cray 1 processors tied together. Although both processors could access data stored in a common memory, each carried out a separate stream of instructions. Chen was not the first to use multiple processors. But when both processors worked jointly on a problem, Chen's X-MP became the first supercomputer to outpace Seymour Cray's designs.

This architecture—many vector processors tied to a central memory—became the predominant supercomputer design. It also heralded a time when supercomputers moved out of a few specialized centers, namely, the U.S. national laboratories, and became more widely available. Spurring this change was the National Science Foundation's decision to fund five university-based centers devoted to supercomputer applications. Thousands of graduate students and dozens of companies first tasted supercomputing at these centers.

Most important, computer users began "visualizing" their results by transforming data into arresting and informative graphics. With these techniques, scientists could peer into previously opaque realms: inside storms, along chains of molecules, on top of the wings of planes. Visualization has consequently begun turning computation into a legitimate scientific method and full partner to experimentation and theorizing; it also demands even greater computational power.

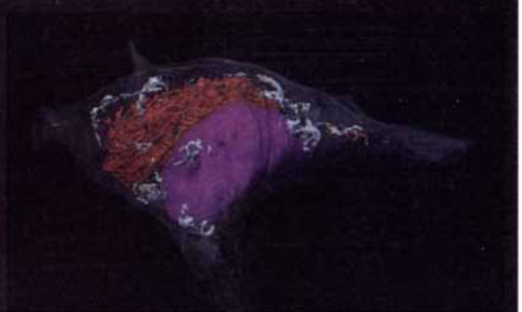
Meanwhile competition among the builders of supercomputers began to heat up, particularly as three of Japan's largest electronics houses—Fujitsu, Hitachi and NEC—plunged into the arena. Masters of complex circuit design, the Japanese manufacturers have leaned even more heavily on vector processing and pipelining than did Cray and so have produced faster processors. NEC is the current champion. Each processor in its SX-3 supercomputer has a clock cycle of 2.9 nanoseconds and a peak speed of 5.5 gigaflops—noticeably faster than the Cray 2 processors, which have 4.1-nanosecond clocks.

Yet even NEC designers doubt they can push the speed of their processors much further. "To be frank, we are reaching a crossroads" in supercomputer design, says Tadashi Watanabe,

PORTRAYING NEURONS

The progress of Alzheimer's disease in a neuron becomes readily apparent when cell data gathered with a high-voltage electron microscope are portrayed in a three-dimensional picture. The image shows precisely where paired helical filaments (*red*), characteristic of the disease, are attached to the cell nucleus (*blue*). The cell also contains the Golgi apparatus (*white*), which is where sugars are added to newly synthesized proteins.

Researchers at the University of California at San Diego used a collection of tools and algorithms to visualize the neuron. Working together, David Hessler, Steve J. Young and Mark H. Ellisman first created a simplified image of the cell on a graphics workstation. After examining many orientations of the image, they reprocessed a few select views in finer detail using a Cray Y-MP supercomputer. The images could then be reexamined and manipulated via a workstation.



a manager at NEC nicknamed "the Seymour Cray of Japan" for his work on the SX series. "Probably in the next [SX] generation we can get higher-density bipolar chips. But I don't know about the next, next generation," he adds.

Designers at Fujitsu, which dominates the Japanese supercomputer market, are equally pessimistic about the future of single processors. Fujitsu's current offering, the VP-2600, boasts a 3.2-nanosecond clock and peak performance of five gigaflops. "It will be very difficult to achieve a one-nanosecond system clock cycle" using conventional packaging and circuit integration design, concedes Keiichiro Uchida, who manages Fujitsu's logic design effort.

Abandoning the classic design carries a heavy cost, however. Adapting software written for vector processors to massively parallel architectures can be agonizing. Although specialized software, called compilers, can help users mold programs into shape for vector processors, compilers for more massively parallel architectures barely exist. As a result, massive parallelism "is a compromise between speed and ease of use," Watanabe declares. "From the user's viewpoint, for a general-purpose machine the number of processors should be small."

From Seymour Cray's vantage, giving up the vector-processor design would be a severe handicap. He is not especially comfortable with massively parallel designs. According to Neil Davenport, chairman of Cray Computer, "Seymour would say that he wants to make a unique contribution and that others have more interest and capability than he does in parallelism."

As a result, Cray and the Japanese manufacturers—separately—hope to eke more speed from the old design by swapping silicon processors for gallium arsenide ones. (Because electrons travel more easily through gallium arsenide than they do through silicon, logic chips made from the material should be both faster and cooler than conventional ones.) Cray intends to tie 16 gallium arsenide processors together.

But gallium arsenide is notoriously finicky—and so a precious gamble. Cray had to set up a gallium arsenide foundry alongside his new company. Gallium arsenide willing, the company hopes to ship its first machine late this year—at least two years behind schedule. "The need for patience is profound here," Davenport says, sighing.

The Japanese supercomputer makers may be among the very few organizations with enough financial and technical stamina to pursue gallium arsenide and other novel materials. Yet even they are investigating new materials cautiously. Under the auspices of a 10-year program sponsored by Japan's Ministry of International Trade and Industry (MITI), Fujitsu, NEC and a handful of other computer makers built a collection of fast, prototype devices.

Incorporating such components into supercomputers will take at least another decade, researchers say. Even so, "inevitably, to get higher performance, the number of processors must increase," Watanabe observes.

Few know for sure, but the most ambitious effort to tie together many vector processors may well be under way at SSI. Although Chen has not yet made the leap to a massively parallel architecture, "every engineering decision was made in favor of parallelism," says Robert A. Walan, an SSI director. Chen is tight-lipped about his project, an attitude reflected in the fortune cookies served at the local Chinese restaurant, which he also owns. A recent visitor to Eau Claire got this one: "A person of words and not of deeds is like a garden full of weeds."

In spite of appreciable financial and technical support from IBM, SSI workers remain acutely conscious of the need to beat other supercomputer makers to market. Competitors "might not be surprised at what we're doing but at how we're doing it," one SSI engineer says. According to Walan, the company is still on target for finishing the SS-1 in 1993. "The main thing we wish," the engineer adds, "is that Seymour would keep delaying the Cray 3."

In a long race for speed the Cray 3 design may not have much staying power. Company chairman Davenport says the gallium arsenide supercomputer will be about 10 times faster than the Cray 2—say, about 20 gigaflops. Even if Cray finds the funding and the customers to support the development of a Cray 4, that future machine would likely be only another tenfold faster—still far from a teraflops.

Indeed, within the past year or two it has almost become dogma that only massively parallel machines—those with several hundreds or thousands of processors—will reach teraflops performance. Whereas the first two supercomputer conferences sponsored by the Institute of Electrical and Electronics Engineers featured as keynote speakers Seymour Cray and John A. Rollwagen, chairman of Cray Research, the third and most recent meeting starred Danny Hillis, designer of the most massively parallel computer on the market. When Hillis asked the audience who disagreed with his claim that the fastest machines would be parallel, only a few arms were raised.

Massively parallel architectures gain speed by ganging together large numbers of microprocessors, which put logic, memory and communications mechanisms onto one chip. Individual microprocessors are still typically slower than the processor in the Cray 1. But by divying up a problem into pieces, the assortment of processors can quickly solve many problems.

Proponents see many advantages. Microchips are briskly picking up speed, points out Intel's Rattner. Microprocessors can also be relatively cheap to use because many types are produced in large quantities for other products. Still, teaming up so many processors raises a host of design questions.

Although there is no standard massively parallel design, Hillis and his Connection Machine have become emblematic of the massively parallel wave. Hillis is also a fountain of the kind of stories that the media love best. As a graduate student at the Massachusetts Institute of Technology, Hillis drove to class in an antique fire engine. More relevant, he turned his Ph.D. dissertation from a research project funded partly by the Defense Advanced Research Projects Agency into a product.

The Connection Machine was—and remains—a highly innovative architecture. Taking the brain and its massive population of relatively slow neurons as his model, Hillis designed a computer that could accommodate more than 64,000 relatively simple processing elements. Data are distributed throughout by pairing each processing element with a memory unit.

The operations of the processors are controlled by broadcasts of identical instructions. All processors then carry out the orders simultaneously on their own data. Like the Rockettes' chorus line at Radio City Music Hall, the processors need no additional coordination, because they are doing the same thing at the same time. This approach is a single-instruction stream, multiple-data (SIMD) design.

Most of Hillis's competitors are designing multiple-instruction stream, multiple-data (MIMD) architectures. A MIMD configuration more closely resembles a ballet. The multiple processors execute different parts of a single task—just as many dancers executing individual movements perform one dance together.

MIMD schemes can be both more flexible and more complex than SIMD designs. For instance, MIMD machines can have physically shared or distributed memories. (A vector processor is a shared-memory MIMD machine because a programmer can divide up a problem among all the processors.) But MIMD designs also require designers to connect the components in a way that lets processors frequently check whether they are still properly synchronized with other processors.

Hillis is confident that his approach will eventually win more converts—as well as the race to a teraflops. "I'm comfortable that we'll have the first machine really capable of a teraflops," he says, referring to the CM-3, the next model of the Connection Machine. The design will enable the company to build a family of machines for different-size problems. "It will be a teraflops on specific things that need a teraflops, like quantum chromodynamics," Hillis promises.

Precisely how the internal mechanisms will scale up to achieve the increased speed is still proprietary, Hillis says. Indeed, company designers even refuse to comment on whether the CM-3 will sport the flashing red lights that make the current model resemble a prop from a *Star Trek* set.

In any case, the CM-3 will not lack for competition. Just a few miles up Memorial Drive from Thinking Machines is Bolt Beranek and Newman (BBN), founded in the late 1940s. Several years ago BBN built an experimental architecture called Monarch with funding from DARPA. Now the company hopes to transform Monarch into what it calls "3T." Rather than linking all processors to their nearest neighbors, BBN uses a web of switches between processors and memory.

Another popular connection network is a hypercube. In this design, processor-memory pairs, called nodes, are situated along the corners of a cube and attached along the edges. Developed first at the California Institute of Technology, hypercube networks have been built by nCUBE in Beaverton, Ore., by Intel and by Thinking Machines.

Intel's Touchstone project aims to produce a teraflops by the mid-1990s by using a mesh connection scheme. In this topology, nodes are connected to their nearest neighbors; the nodes communicate by passing messages. And Tera Computer Company in Seattle, a fledgling start-up, is pulling that mesh into three dimensions. Burton J. Smith, Tera's chief scientist, sprinkles nodes of processors and memory units throughout the mesh. Processors can execute instructions for other parallel activity when waiting for data from memory. As a result, the computer acts like a shared-memory machine. So far the company has simulated a 256-processor version.

Many others are reluctant to describe their designs. "You'll notice we don't

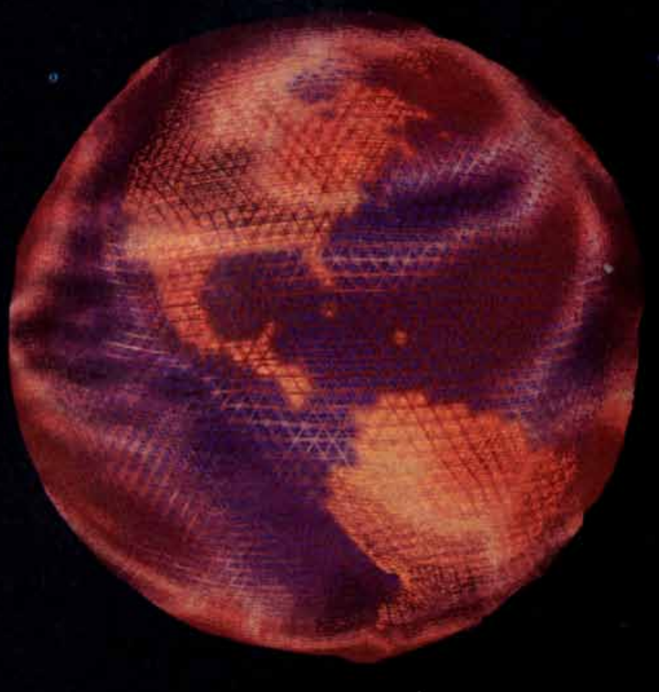
FROM EXPLOSIONS TO CLIMATIC CHANGE

Researchers at Los Alamos National Laboratory have turned a technique developed for modeling explosions into one that simulates climatic change.

This portrait of the atmosphere's planetary boundary layer relies on meshes made of half a million tetrahedrons. Land-masses are colored yellow. Areas of higher kinetic energy are shown in red; lower and lowest energy regions are white and purple, respectively. Because this atmospheric layer is no more than 1,500 meters from the earth's surface, mountains create a bumpy surface. Every tetrahedron covers an area no wider than 30 kilometers.

To create the flexible meshes, the Los Alamos scientists used the Connection Machine, a massively parallel computer (and graphics software developed by Delany Enterprises in Cambridge, Mass.). In the event of a disturbance such as a hurricane, these meshes would twist. Conventional models, which use rigid meshes of rectangular bricks, typically lack the resolution to portray such comparatively local phenomena.

To calculate perturbations, a programmer can assign one processor to every tetrahedron (software creates the illusion of as many "virtual processors" as needed). "We only have to do it right for one processor," says Harold Trease, a researcher in the computational physics group at Los Alamos. "The computer does the rest."



even have a sign on the door," points out Henry Burkhardt III, chairman of Kendall Square Research in Waltham, Mass. University projects abound both in the U.S. and abroad. The German government is sponsoring a parallel architecture project called Suprenum. The prototype architecture relies on 256 nodes—teams of vector processors, memory and communications units. Even the Japanese supercomputer makers are quietly working on a range of highly specialized machines that are largely parallel. Still, the question all must address is not how fast their machines will run test codes but whether they will efficiently solve users' problems. For hints, designers are looking to the algorithms, or mathematical descriptions and equations, being developed for parallel machines.

Few existing algorithms were designed to be broken up into separate components and solved simultaneously. New, parallel algorithms are still largely under construction. Moreover, a symbiotic relationship exists between algorithms and the machines that run them. As mathematicians and computer scientists flesh out algorithms, machine designs change in turn, says David B. Salzman, who directed research at the now defunct John von Neumann National Supercomputer Center at Princeton University. "An architect designs a machine, a user maps an algorithm to it and then the next machine has an improved design so it can run that algorithm better. Then faster algorithms are developed, stretching the machine again, and so on."

Hillis and other advocates of parallelism argue that many problems are naturally parallel—they simply have not been expressed that way in the past. Charles Peskin, a professor at New York University's Courant Institute, agrees. He has spent almost two decades building algorithms that describe the dynamics of a beating heart.

Modeling the flow of blood through the valves is a computationally exhausting task. "They used to call me 'the man with the two-dimensional heart,'" Peskin says with a rueful laugh. Within the past year, using a Cray 2 to run the calculations, he has transformed his model into a three-dimensional, multicolored portrait. Even so, generating one heartbeat demands more than an entire day of Cray processor time.

Although Peskin has worked primarily on vector processors, he points out that his problem seems highly parallel: it involves solving the same sets of equations for many fibers in the heart. As a result, he is optimistic that a massively parallel architecture will enable him to speed up the calculations and add more realism to his model.

Yet even if many such problems are intrinsically parallel, they are unlikely to feature the same degree of parallelism. Computer architects must consequently anticipate the degree of parallelism in future problems and then design hardware that should—in principle—run those algorithms quickly.

"You have to envision the future," says Tera's Smith. "You must design a system based on a model of computation that is as general as the uniprocessor model, one that lets its prospective users program it quickly and well."

For that reason, Cray Research is paying more attention to the needs of its potential customers than it ever has in the past. "I'm giving our users a homework assignment," says Steve Nelson, who is leading the Cray Research team working on a highly parallel design. "What should the network [between processors and memory] look like?" Nelson is even considering putting full-time Cray employees at several user sites to work on applications. At this point he estimates that Cray's design will call for several thousand processors in a mesh or an "Omega" network, which uses an intermediate series of switches between processors.

Intel is following another path. A recently formed consortium of U.S. research institutes has established a new supercomputing center at Caltech, which will receive the first of Intel's next Touchstone machines. The consortium wins a discount on what could easily be a $20-million price for the 32-gigaflops machine. In turn, Intel benefits when researchers develop software for its machine.

In addition to speed and software, supercomputer designers must face a smorgasbord of other issues when building commercial products. For instance, supercomputers need enormous memory capacity to handle large problems and an ability to shuffle vast amounts of data in and out of the machine rapidly. All supercomputers are "out of balance," complains Michael P. Burwen, a supercomputer analyst at the Superperformance Computing Service in Mountain View, Calif. Parallel ones are often the worst. "Do you know how long it takes to load eight gigabytes of memory? Hours."

Even if massively parallel architectures win ribbons for speed, will most researchers—particularly in industry—be willing to use the machines? "I have a bet with Danny Hillis," says C. Gordon Bell, the lead designer of Digital Equipment's VAX workstation in the 1970s and chief scientist at Stardent Computer, Inc., which makes graphics workstations. Bell is wagering that by 1995 vector-processor supercomputers will still run most of the problems.

In fact, parallel architectures may well coexist with those traditional machines. "There is no single solution to realizing a single best [computational] system," points out Nobuhiko Koike, a research manager at NEC. "There will be different architectures for different systems. If you look at computational fluid-dynamics problems, a vector-style highly parallel machine is best suited to it. Other applications may need neural networks or special-purpose parallel computers."

As a result, computer vendors may wind up selling packages of hardware and software designed to tackle specialized problems, such as building chemical engineering models. Parallel com-

Modeling on Multiple Machines

At supercomputer centers across the U.S., the dream of connecting many different computers into a seamless network is turning into a reality. Such a complex fabric would give scientists working at personal computers in their offices the flexibility to solve parts of problems on specialized architectures (including workstations, special-purpose processors and supercomputers), regardless of where the machines are located.

The key to such networks are both the high-speed links between components and the hardware and software for managing the network. Few sites have yet installed 100-megabit-per-second links. Only a handful of companies and organizations boast 1,000-megabit-per-second links.

Although there is no ideal network, here is how three scientists working in their offices (*black arrows*) might use a network:

YELLOW: An atmospheric physicist is studying how quickly a cloud becomes a severe storm under a variety of wind, temperature and pressure conditions. Because creating a three-dimensional model of such a storm involves many coupled, nonlinear partial-differential equations, the scientist develops a program on his workstation that will run efficiently on one or a few processors in a vector-processing supercomputer.

The model requires that the supercomputer tap a large data base stored in a separate file server, or data vault. A high-speed data link between the file server and supercomputer shuttles the data back and forth.

Once results are available, the information is downloaded to a specialized graphics workstation that turns the puters may never become general-purpose tools, says Sidney Karin, who directs the San Diego supercomputing center. Instead "the future may hold more 'shrink-wrap' supercomputer solution boxes alongside the traditional systems," he says.

Aiding the move in this direction may be the evolution of networks that convey data at very high speeds. In the past the linkages between computers were too slow for dividing computational tasks (rather than just file transfers). The past year or so has seen much progress in formalizing standards and interfaces that will enable computer users to pump hundreds of megabytes per second over networks.

Such networks will give researchers the option of breaking a single problem into distinct parts and assigning the sections to the most appropriate architectures—different supercomputers, highly specialized processors, workstations, personal computers and so on. The links will extend beyond a single site. Government-funded testbeds are in the works.

Also on the horizon in the mid-1990s: networks of powerful workstations that achieve gigaflops performance. Provided the pieces of the problems need not communicate too frequently with one another, "a lot of people will find it worthwhile not to go to a Cray," declares Pat Savage, who directs Shell Oil's supercomputing efforts in Houston. "It is a different class of software problem. Now you are one node in a large distributed computer."

Teams of computers will undoubtedly accelerate another trend: teams of researchers. Karin is convinced that the massive problems addressed by supercomputers will be managed and interpreted by many collaborators. Just as hospitals have teams of specialized medics, surgeons and nurses working together on an operation, so too may massive computer simulations require group efforts.

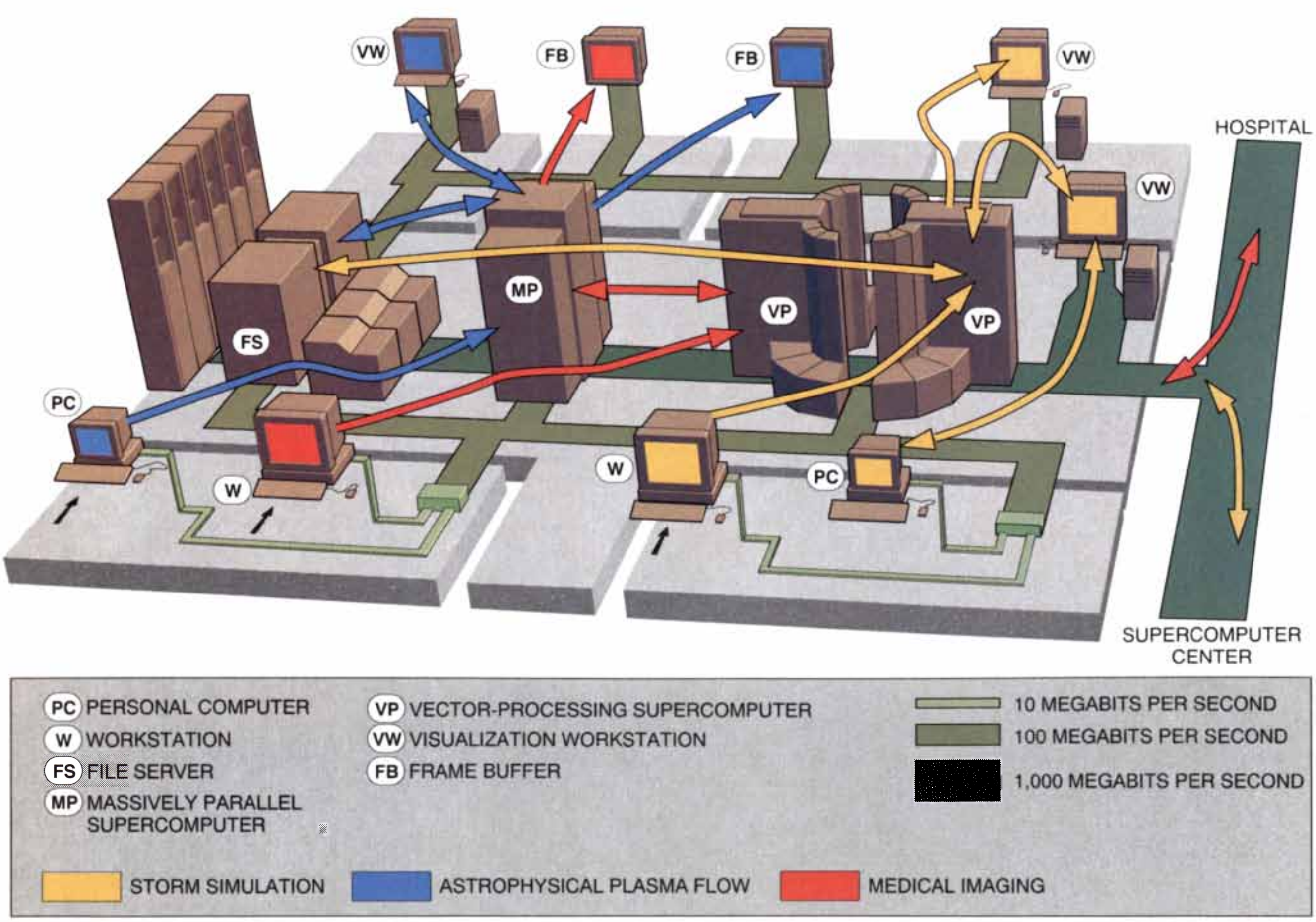
In this way, the shift to new paradigms of computing is already beginning to change the nature of science. "When they look back at the late 20th century," Hillis suggests, "they'll say computers were the things that affected people's lives in a big way—just as the automobile did." The results may not be all good, he concedes, "but they will be pretty fundamental."

With reporting by correspondent Tom Koppel in Tokyo.

numbers into three-dimensional images. The scientist may call up the model on such workstations and examine slices of the images working from a personal computer. He may also explore a subset of the problem with assistance from a specialized processor at a distant supercomputer center.

BLUE: Another scientist aims to animate the flow of astrophysical plasma jets in three dimensions. Because that work involves calculating the local interaction of particles in a volume of space, the researcher finds it faster to run the calculations on a massively parallel computer. Colleagues can watch the results on frame buffers, or terminals, in separate offices.

RED: A medical researcher receives planar projections of a brain, generated by a nuclear magnetic resonance imager, from a distant hospital. The researcher first uses a vector processor to help construct a three-dimensional model from the slices. Then a parallel computer animates the data. Images are displayed locally and returned to the hospital.



Second-Generation Silicon

"Band-gap engineering" may keep silicon a winner

Name that cluster, scientific stage sets, space on the cheap, mapping for dollars

What does the future hold for silicon, the semiconductor that brought you the Cray supercomputer and greeting cards that play "Silent Night"? Materials scientists often wonder whether silicon can hold its own in a not so distant time when laptop supercomputers might be sold at the local electronics store.

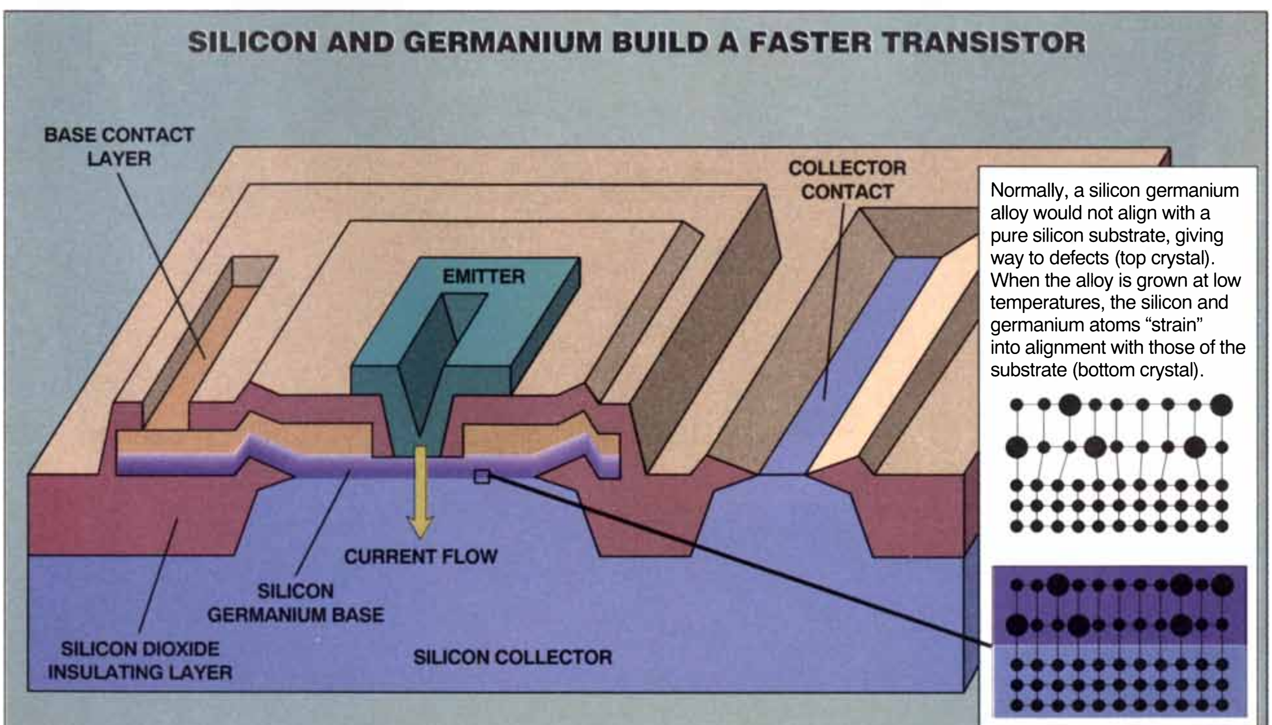
The reason is that the speed of electrons moving through a silicon crystal is slower than that in gallium arsenide and other compounds drawn from elements in the third and fifth columns of the periodic table. Silicon is also cursed with properties that make it exceedingly difficult for the element to emit light—giving an advantage to gallium arsenide in a world in which photons may serve as the primary vehicle for transmitting and eventually processing information.

But even if it is a tortoise, silicon still has a chance to win the race because of the ease with which it can be manufactured into transistors and other electronic devices. Silicon's superiority in the clean room comes from its ability to be easily sliced into large wafers and, when heated, to form silicon dioxide, a natural insulator that serves as a mask when circuit patterns are photolithographically etched onto silicon wafers. Neither gallium arsenide nor any of the other III-V materials are as easy to use as silicon in fabricating electronic devices, a fact that has kept them out of widespread use despite their attractive electronic and optical properties. "We know so much about silicon, and it is so manufacturable," says Leonard C. Feldman, head of thin-film physics research at AT&T Bell Laboratories. "That's why a group of us are trying to extend it beyond its present capabilities."

One way to continue silicon's lease on life is by combining it with other materials. Researchers at Bell Labs, IBM and elsewhere have shown that despite the usual mismatch between silicon and the crystal lattices of most other materials, an alloy of silicon and a small amount of germanium is elastic enough to be grown on a pure silicon substrate without defects.

The most pragmatic demonstration so far of this idea came early last year, when Gary L. Patton and co-workers at the IBM Thomas J. Watson Research Center reported that they had fabricated a transistor using an alloy of silicon and germanium that switches at a rate of 75 billion cycles a second, almost double the previously recorded speeds for silicon—and approaching performance of experimental transistors made with III-V materials. In December, IBM was expected to present a technical paper detailing how these transistors could be integrated into simple circuits.

The addition of germanium enabled IBM to "engineer" the band gap of silicon, as is widely done with III-V compounds. The band gap is the amount of energy needed to move the outermost (valence) electrons of atoms in a crystal lattice into a higher-energy state—the conduction band—where they can move freely and carry current. Making the gap between the valence and conduction bands larger or smaller can tweak higher performance out of a device.



With this technique, IBM made a bipolar transistor similar in design to the ones in its most powerful mainframe computers by using the silicon germanium alloy for the device's base, the component that turns current flow off and on. Germanium, whose band gap is smaller than that of silicon, enables more electrons to flow from the emitter on one side of the transistor's base to the collector on the other. By gradually increasing the amount of germanium across the base, researchers created a difference in electric potential that acts like a slope down which the electrons flow at breakneck speed.

Advances in materials engineering open the way to using not just germanium but a handful of elements and compounds that can be combined with silicon to create custom-designed band gaps. "If we can do this with silicon germanium, it expands our horizons," enthuses James W. Mayer, director of Cornell University's Microscience and Technology Program. "We can adapt to silicon all the techniques the gallium arsenide people have been working on for the past 15 years."

Materials engineers may even enable silicon to meet the more challenging problem of becoming an ingredient of optoelectronic circuits. Because of its band gap, silicon usually emits more heat than light. But its ubiquitous presence in electronics would make it an ideal candidate for combining optical and electronic transmission either on a chip or between chips or circuit boards.

One approach that could make silicon a primary material for optoelectronics is wedding it to gallium arsenide or another III-V material. Depositing gallium arsenide on silicon is no simple feat, however. Early attempts to deposit gallium arsenide on silicon failed because the alignment of the individual gallium and arsenic atoms with silicon is off by 4 percent—a huge mismatch according to materials scientists. As the two crystal lattices try to line up, defects are introduced into the gallium arsenide.

Although some researchers continue to pursue this approach, Eli Yablonovitch and his colleagues at Bell Communications Research (Bellcore) found a new way to cope with the problem. Three years ago they used hydrofluoric acid to lift off a thin film from a gallium arsenide wafer and graft it onto silicon. Because the film had been grown in native gallium arsenide, the crystal structure was free of the usual defects. In 1989 the Bellcore team announced it had made lasers, transistors and other devices using the process, which can also place gallium arsenide on glass.

Other investigators are trying to coax silicon itself to emit light. The effort confronts a seemingly insurmountable technical hurdle. The III-V compounds are what are called direct band-gap materials. When an electron in gallium arsenide falls from the higher-energy conduction band to the valence band, the electron combines with a positively charged hole to emit a photon. In silicon the electron's transition is made via an indirect band gap. Instead of spitting out a photon, vibrations in the crystal eventually shake an electron loose, dissipating its energy through the lattice as heat.

For a decade, researchers have been considering alternative ways to light up silicon. One method is to supplement silicon with complexes of sulfur, beryllium or other atoms. This addition does not change silicon's conductive properties. But the impurity atoms can still trap a roving electron or hole nearby. The electron may then attract a hole. The electron-hole combination, called a bound exciton, may then emit a photon.

In 1989 researchers at the University of Rochester's Institute of Optics built a light-emitting diode using silicon doped with sulfur, which gives off light in the near infrared zone of the spectrum typically used for fiber-optic communications (1.3 microns). The strongest light emissions occurred, however, at temperatures between about –370 and –315 degrees Fahrenheit. The ability of impurity atoms to bind the electron-hole pair is small at room temperature, and the crystal lattice jiggles at such a rate that little light is emitted.

There may be a better way. Dennis G. Hall, a professor of optics at the University of Rochester, is collaborating with Joze Bevk of Bell Labs to enhance silicon's light-generating properties by building a so-called superlattice made of atomic layers of silicon alternating with those of a silicon germanium alloy. The relatively small band gap of the silicon germanium layers tends to catch electrons and holes between the higher band-gap "walls" of the silicon layers. Thus ensnared, electrons and holes may emit a photon. "Superlattices tend to keep electrons and holes in the same neighborhood," Hall says.

Even so, researchers in silicon optoelectronics know they are taking a long shot. "Where we are now is sort of a beginning. We've had an encouraging sign or two," says Hall of his own research. "But I don't think we have the basis for an industry on what we've found to date." —*Gary Stix*

Future Shock

Fear of the "big one" fuels sales of insurance and Velcro

In August of 1987 a series of purported astronomical events combined with worldwide human celebration—the Harmonic Convergence—was to have ushered in a new era of peace and prosperity. Since then—eerily enough—the Berlin Wall has fallen, democracy has swept through Eastern Europe and Mikhail Gorbachev has won the Nobel Prize for Peace. Now the moon's proximity to the earth and its alignment with the earth and sun have led to predictions of natural disaster.

A New Mexico climatologist, Iben Browning, forecast an even chance that a major earthquake would rock the New Madrid fault in the Midwest on or around December 3, 1990. To scientists, Browning's earthquake "projection" had little more credence than the prophesies of the New Agers. Still, it was enough to cause some schools to plan to close and residents to stock up on canned goods.

The day fated for the Midwest's "big one" nearly coincides with an extended wave of media attention on earthquakes. Last October news coverage marked the anniversary of California's Loma Prieta earthquake. Close behind followed an NBC two-part television series, "The Big One: The Great Los Angeles Earthquake." The media may have generated the most acute public awareness of earthquake hazards in recent memory. Walter W. Hays, a U.S. Geological Survey deputy chief who has coordinated about 60 workshops on earthquake preparedness during the past 13 years, says the level of anxiety over earthquakes has reached near hysteria levels. "Nothing has ever stirred people up like these events," he says.

A measure of the level of concern was the more than 50,000 calls registered by an American Red Cross hot line that NBC viewers were told to call for information on how to prepare for an earthquake. In addition, sales of earthquake insurance in the heartland have boomed. State Farm Insurance, the nation's largest property and casualty insurer, found in September of 1990 that dollar sales of earthquake coverage in the seven midwestern states nearest the fault tripled from those registered in February. "Sales have gone through the roof," says David A. Jewell, a spokesman for the Earthquake Project, a lobbying organization for property and casualty insurers.

Indeed, the Earthquake Project, organized three years ago to persuade the government to set up a federal earthquake insurance program, has seized the moment to press its case in the halls of Congress and elsewhere. A day after the NBC special, AIG, a New York insurer that is an Earthquake Project member, ran an advertisement in the *New York Times* explaining why federal earthquake insurance is "vital for all Americans." The timing with the NBC special was coincidental, a company official insists.

Without a federal safety net, property and casualty insurers could be financially crippled if confronted with a catastrophic earthquake that wreaks more than $50 billion in property damage in a major city, asserts Edwin A. Manton, a senior AIG adviser. "The government can print money if it has to; we can't do that," Manton says. "We go to jail if we do that."

In the aftermath of the 1989 California earthquake, Congress has been particularly responsive to the issue of earthquake preparedness. Last year two bills to provide federal earthquake insurance were introduced in the House of Representatives and are expected to be revived in the current session. In addition, the National Earthquake Hazards Reduction Program, an interagency effort that has steadily received $65 million year after year, saw its funds increased by more than 50 percent in the current budget.

With or without federal priming, earthquake preparedness may be moving beyond cottage-industry status. Companies are peddling everything from P-wave detectors that warn of imminent shocks to Velcro straps that secure computer monitors.

One retailer, the Center for Emergency Preparedness and Provisioning (CEPP) in Los Angeles, got its start eight years ago and claims that it has grown to become the nation's largest seller of earthquake survival kits, featuring items ranging from drinking water to toilet paper. In recent months, CEPP's sales in the previously untapped midwestern market have skyrocketed. The company, moreover, was looking forward to a prosperous holiday season.

While an Emergency Fannypak may make an ideal stocking stuffer, various federal agencies are wondering how to keep people thinking about what to do about earthquakes when the holidays have come and gone. If they can keep their message fresh, then they may have an easier time focusing on the very real task of bringing about changes in building codes to ease the impact of the big one that one day is sure to come. —*Gary Stix*

Gene Rush

Companies seek profits in the genome project

The genome project, under way officially since October 1990, is as close to "big science" as biology has ever been. Mapping and eventually sequencing the 100,000 or so genes in the human body could cost the federal government a total of $3 billion, an average of $200 million annually, over the next 15 years. The gene hunt is on at national laboratories, research institutes, universities and companies.

Business opportunity leaps in every step of the process—from the massive chore of charting where genes exist on chromosomes to the tedious task of constructing copies of those genes for use in drugs and diagnostic tests. All the information and methods developed with federal funds (administered jointly by the Department of Energy and the National Institutes of Health) are to be made available for "aggressive commercialization."

The best strategy to follow is anyone's guess. "The questions of niche versus generic market and small company versus big company aren't nailed down yet," observes Mark L. Pearson, director of molecular biology, central research and development at the du Pont Company. The reason, he adds, is that "nobody knows what the technology of 1995 will truly be."

Even so, companies are already positioning their products and technologies. Some, like start-ups Genmap and Transkaryotic Therapies (TKT), are developing laboratory techniques aimed at speeding the project's first five-year goal: mapping. Mapping breaks the genome into fragments of DNA that scientists can analyze systematically. A variation in fragment size between diseased and healthy individuals may indicate that the fragment contains a disease-causing gene. Not all deleterious mutations will be found so easily, however. Some defective genes may be the same length as normal ones yet have an important change that mapping will help to detect.

HUMAN GENOME PROJECT

PHYSICAL MAPS

Major centers

NIH: Human chromosomes 4, 7, 11, X
DOE: Human chromosomes 16, 19, 21

Smaller-scale efforts

NIH: Human chromosomes 1, 4, 9, 14, 17, 18, 21, Y
DOE: Human chromosomes 4, 5, 7, 11, 17, 22, X

SEQUENCING

NIH: Pilot projects at the megabase level
Sequencing research organisms
Novel sequencing technologies

DOE: DNA sequencing
Novel sequencing technologies

MODEL ORGANISMS

NIH: Expression of genes in research organisms such as bacteria and mice

INFORMATICS

NIH / DOE: Data management, analysis, accessibility

SOURCE: National Center for Human Genome Research

"We would like to be considered a gene-finding company," notes William T. Carroll, research manager for Genmap in New Haven, Conn. The tiny firm, founded in 1988, plans to offer collaborators or contractors its laboratory technique for cloning pieces of DNA that are as long as five million base pairs. With such lengthy segments, researchers can cover ground quickly—for instance, along chromosome 17, where a gene involved in breast cancer is thought to lie somewhere among 80 million or 90 million base pairs. The firm also makes what it calls reduced chromosome maps, with probes that highlight the repetitive sequences of DNA where some studies suggest active genes are often found.

Workers at TKT in Cambridge, Mass., believe the most difficult step in gene mapping is moving, or "walking," from a known chromosomal marker to a disease-causing gene. It is certainly one of the most time-consuming laboratory procedures, observes Richard Selden, the company's chief scientific officer. Consequently, researchers tend to pursue genes from rare diseases with clear inheritance patterns rather than less well characterized conditions such as heart disease. "They don't want to end up saying, 'Damn, we were walking to the wrong gene,'" Selden says.

So TKT has developed a method of speed-walking to isolate target genes. Selden declines to describe the proprietary technique but says it takes just a few hours instead of the usual four to six weeks. The firm will put the trick to work identifying common, yet genetically complex diseases, such as hypertension, manic-depressive illness and various forms of cancer.

Other companies are following what is sometimes called the Gold Rush strategy—selling instruments and supplies to those who would sift for gold. Start-ups such as Genomyx in South San Francisco and Bios in New Haven are going up against heavy hitters the likes of du Pont, Pharmacia and Applied Biosystems for a share of the market in automated DNA sequencers. The existing generation of machines detects gene sequences by reading the patterns made by pieces of DNA as they travel through a gel. Improvement stands to come incrementally, with better enzymes and gels.

New robotic machines that break apart DNA promise to automate the now laborious manual process that is the "front end" of gene sequencing. Applied Biosystems in Foster City, Calif., will introduce a robotic workstation early this year that it says will handle smaller sample sizes than does the Beckman Instruments model already on the market. A robotic arm uses a syringe-based system to move chemicals from storage to a reaction vial. The real engineering achievement is preventing a minuscule microliter of liquid from evaporating as it cycles back and forth between room temperature and just below boiling. "The reduced quantity of chemical required keeps costs down," explains Michael Hunkapiller, vice president of science and technology at Applied Biosystems.

Perhaps the biggest business opportunity, and the one with which biologists are least familiar, is creating the technology to deal with the huge amounts of data produced by the project. How the information is managed and made electronically accessible to the user—the so-called informatics of the genome project—is the domain of a joint DOE/NIH task force headed by Yale biophysicist Dieter Söll. "We need a flexible structure that even computer-illiterate molecular biologists would use like a book and index," Söll says.

No matter what technology a firm is backing, "investing in this area is about as risky as you can get," says Alan Walton, partner at the venture capital firm of Oxford Partners in Stamford, Conn. The risk is not merely a matter of competition and rapid change, he notes, but an issue of "whether you find a money gene or not." It is also difficult to tell if someone else has used your method to develop their product.

Some firms speak of giving royalty-free technology licenses to universities to speed research. If the academic work were ever commercialized by another company, a percentage would flow back to the technology originator. Patent attorney S. Leslie Misrock, senior partner at Pennie & Edmonds in New York City, warns against the practice because of the "doctrine of laches," which recognizes that if someone is allowed to do something for a time without paying for it, that party comes to believe it has the right to continue. For his part, Misrock continues to believe genes are patentable. "They're new and useful compositions of matter discovered by fishing them out of a lot of random DNA," he says. But so many biotechnology patents have been challenged that companies are bound to start keeping genes secret—at least until forced to tell all to the Food and Drug Administration. Finders, keepers. —*Deborah Erickson*

Yale set designers turn to finite element analysis

This spring a group of students at Yale School of Drama will learn how to use a mathematical technique that helped to build the Stealth bomber. Instead of making low-profile planes, however, these graduates will move on to Broadway or scenery studios to erect 35-foot-high robotic reptiles or construct superstructures that weigh 50,000 pounds when loaded with lights and equipment.

The formal introduction of a course in finite element analysis (FEA)—it was taught once before on a trial basis—is an attempt to move beyond the methodology for technical set design that has prevailed since stagehands cobbled together balconies at Shakespeare's Globe Theater. "You put something in, it broke, then you put in something bigger or stronger," says Bronislaw J. Sammler, production supervisor and chairman of the technical design and production department of the Yale drama school.

A more incisive look at stage construction has been called for because blockbuster Broadway shows, Las Vegas–style extravaganzas, opera and even regional theater must be loaded with special effects and labyrinthine sets to entice audiences jaded by movies and television.

Finite element analysis is used today to analyze point stresses in automobiles, jet fighters and even knee braces for professional football players. In set design the technique uses matrices of simultaneous equations to derive deflections and stresses when a load—a 250-pound opera singer, for example—is applied to a series of finite elements: the tubing, I-beams and other components that make up the steel frame for an overhanging ledge on which the diva stands.

Prospective technical designers will not be required to wrestle with equation solving to learn the technique, however. A personal computer software program, called Algor, will do the calculations. The students will learn the program, in part, from a tutorial written as a master's thesis by David L. Sword, the assistant professor who will teach FEA as an elective course.

The software simplifies the process by requiring a user to trace out a three-dimensional model on the screen and input a few variables—type of material, thickness and tensile strength, among others. It then solves the equations and supplies a contour map that indicates in red areas of highest stress.

Yale alumni have already become finite element proselytizers, and New York City's Metropolitan Opera is one beneficiary. Geoff Webb was snapped up by the opera house as a technical designer after completing a master's thesis in 1988 at Yale on finite element analysis. In working on a set for Wagner's *Flying Dutchman*, Webb had to make adjustments to the specifications for a 50-foot staircase of steel tubing that descended from the Dutchman's floating ship. The software showed that the steps might have flexed up and down slightly underfoot when the singer made his entrance on it.

Webb is looking at still other software from the design engineer's toolbox. So computerized methods for designing concealed aircraft may get quite visible use by those working hidden behind the curtain.—*Gary Stix*

FINITE ELEMENT ANALYSIS helped design the Metropolitan Opera's set for the* Flying Dutchman. *Photo: Winnie Klotz.

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

Buckeyballs

Fullerenes open new vistas in chemistry

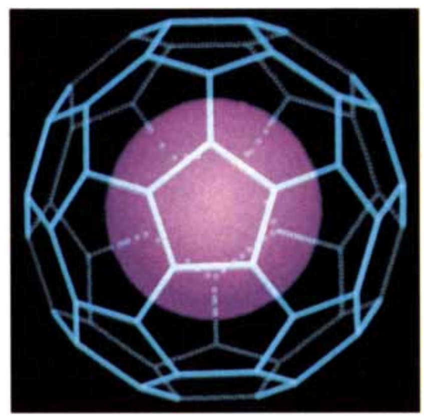
Researchers in Arizona and Germany have given new luster to soot. They have isolated from it quantities of a 60-atom carbon cluster called buckminsterfullerene, in honor of the designer of the geodesic dome. The isolate, a glossy, brown crystal, constitutes the third form of carbon (the others are diamond and graphite), one that could open up new vistas in synthetic chemistry. "The mind races over the possibilities of use," says Richard E. Smalley of Rice University, who helped to discover the cluster, posited its structure and gave it its name. "It's a new starting material for organic compounds, similar to the discovery of benzene in 1825." And, he adds, "it looks like it's going to be dirt cheap."

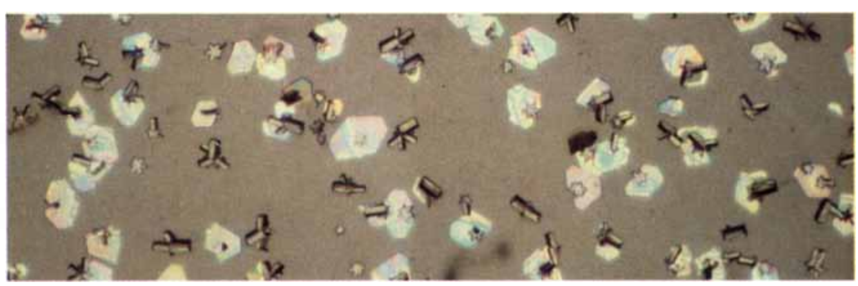
The production of gram quantities of buckminsterfullerene—buckeyball, for short—has two immediate consequences. First, it gives chemists their first look at a bulk form of a carbon species that was discovered, in molecules, only five years ago. Second, techniques of bulk chemistry, such as crystallography, have proved beyond doubt that C-60's constituents indeed assume the soccerball shape—resembling Fuller's domes—conjectured in 1985 by Smalley, his colleagues from Rice and Harold W. Kroto of the University of Sussex.

The mass-production technique was discovered by accident, while researchers were looking at still smaller clusters. "We were making milligram quantities in 1983 but didn't know what it was, and it was only after the work at Rice that it occurred to us we might be producing carbon 60," says Donald R. Huffman of the University of Arizona in Tucson, who worked with Wolfgang Krätschmer of the Max Planck Institute for Nuclear Physics in Heidelberg.

Huffman and Krätschmer achieved their synthesis by creating an arc between graphite electrodes in helium held at one seventh of atmospheric pressure. They collected the soot that formed in the vessel, dissolved it in benzene to remove unwanted species of carbon and were left with a mixture of nearly pure C-60 and C-70, an egg-shaped cousin. The solution is reddish and precipitates into brown crystals.

If the C-70 impurity is removed, the crystals show a mustard color, a fact that greatly satisfies Smalley and his collaborators at Rice and Kroto. Theory predicted that C-60 would be yellow and what they called their "search for the yellow vial" had been under way for five years, ever since buckeyball





BUCKMINSTERFULLERENE'S soccerball shape was originally modeled in paper (top left) *by Richard E. Smalley of Rice University. He now wants to "shrink wrap" ions inside the carbon 60 cage* (top right). *Crystals of the material were recently isolated* (bottom) *by Donald R. Huffman of the University of Arizona and Wolfgang Krätschmer of the Max Planck Institute for Nuclear Physics. Krätschmer provided the micrograph.*

appeared in the annals of chemistry as a strange spike in a mass spectrograph. The researchers figured that the C-60 spike meant that that cluster size was particularly stable. They thought about the kind of structures that might account for this stability, considering both open, sheetlike models and the closed, cagelike alternatives.

After crude modeling attempts using toothpicks, gummy bears and frustrating computer graphics programs, Smalley says he found the crucial analytical method. He taped together paper polygons in his kitchen during the small hours of an August morning in 1985. When hexagons alone failed to close his cage, Smalley paused to drink a beer. After he resumed his work, he found that 12 pentagons and 20 hexagons could form a ball if the pentagons were not allowed to share a side and if their vertices were made to point at one another (to preserve symmetry). He had stumbled on the truncated icosahedron, so called because it can be constructed by lopping off the corners of the icosahedron, the last of Plato's five "perfect" polyhedra.

The shape is indeed perfect. "Sixty is the largest number of proper rotations in the icosahedral group," Smalley says. "That, in turn, is the largest point group—the largest group where symmetry operations, rotations, reflections, etc. leave one point unmoved. This makes C-60 the most symmetric possible molecule."

This roundest of round molecules could have a wide range of commercially useful properties. It could serve as the kernel of a wonderful lubricant, Smalley predicts, provided that its 60 dangling bonds are first tied up, perhaps by adding 60 fluorine atoms. (Each carbon atom has four bonding sites, so that it can either form a double bond with one of its three neighbors or hook onto an outside reactant.) Another aspect of buckeyball's symmetry is the freedom it gives electrons to wander throughout its whole network. This works in the manner of a geodesic dome: the strain imposed on the straight linkages is spread throughout the cage. In a building, this makes for a strong, light roof; in a carbon cluster, it makes for easy adding and subtracting of electrons, a reversible reaction that Smalley thinks may create a whole new class of rechargeable batteries.

Most exciting of all, Smalley says, is the prospect of "shrink wrapping" metal ions inside the carbon cage's central cavity. That would permit reactive or radioactive materials to be encapsulated for specialized storage. Stuffed buckeyballs may also serve as tiny handles. "We don't have an easy way to hold a single atom—ion traps require that it be an ion," Huffman says. "But the smaller atoms should fit inside a buckeyball even without being ionized."

The technique would also allow fine-tuning of the buckeyballs' electrochemical and optical properties. "In principle, you can put 92 different elements inside and get 92 different properties," says Smalley, who has dubbed the new compounds metallofulleronium.

Indeed, buckeyball jargon is running well ahead of commercial applications. What do you call compounds made of positively charged buckeyballs? Buckides. What's a buckeyball called when it's wearing a Medusa's head of organic molecules, attached by their hydrogen-bond heads? A fuzzyball. Larger cousins to the buckeyball appear at even numbers following 60 carbon atoms. What are they called? Giant fullerenes. In theory, it should be possible to put quite large molecules into the larger fullerenes and perhaps to nest them one inside the other. What do you call nested fullerenes? Russian eggs, Smalley quips.

Get used to the new terminology: it is sure to figure in tomorrow's textbooks. And in the not too far-off future, fullerenes may also be showing up in products ranging from consumer electronics to the family medicine cabinet. —*Philip E. Ross*

Another Small Step?

Space veterans plan a low-budget moon mission

A group of volunteers is trying to put together a private mission that will revive the heady spirit of the early U.S. space program. This makeshift team wants to send an unmanned spacecraft to the moon perhaps as early as 1992—the first U.S. return there since the Apollo program ended in the mid-1970s. Before its tiny craft can set forth on the journey of more than 200,000 miles, the group needs to raise $12 million, build three control centers spotted around the globe and talk the Soviets into doing the launch.

The goal of the 25-member team—a third of them employees of the National Aeronautics and Space Administration who have been devoting spare time to the project—is to build a lunar satellite that would weigh no more than 660 pounds when fueled and would gather data that NASA has talked about since the 1960s—namely, a chemical, magnetic and gravitational map of the entire moon. "I've wanted this data for 20 years," says Alan B. Binder, who worked at Lockheed on experiments for future NASA trips to the moon and Mars before moving on to head up the project.

The private moon mission began as a collaboration between Lunar Exploration, Inc., organized by a group of Houston space enthusiasts, and the Space Studies Institute of Princeton in New Jersey. That institute, under the leadership of Princeton University Professor Emeritus Gerard K. O'Neill, advocates using solar power, lunar minerals and other space resources to construct extraterrestrial human habitats. Binder was asked to manage the project because of his experience as a scientist for the camera team on the Viking mission to Mars in 1976—a $1-billion venture. (Because Lockheed has agreed to keep him on the payroll, he is the only paid member of the group.)

By orbiting around the poles, the so-called *Lunar Prospector* would be able to make a complete record of the lunar surface that was impossible before because of the Apollo program's equatorial orbits. The spacecraft—to be designed and built by Omni Systems, an El Segundo, Calif., start-up—would spin like a top instead of having to be constantly reoriented by firing its jets.

The experiments would be carried out in a 100-kilometer-high polar orbit using relatively simple instruments that can detect radiation either facing toward or away from the moon. A major goal of the year-long mission would be to answer a question that has intrigued scientists since the Apollo days: Is there frozen water at the lunar poles? If water is found, it could be used as oxygen and hydrogen fuel to support the type of lunar colony proposed by NASA and President George Bush as part of the Space Exploration Initiative to send missions to both the moon and Mars starting in the next decade.

LUNAR PROSPECTOR ***chief Alan B. Binder stands beside a model of the spacecraft. Photo: Space Studies Institute.***

In the meantime NASA has provided encouragement and has donated an instrument left over from the Apollo program—a gamma-ray spectrometer—which makes up the bulk of what the group says are "assets" of $2 million. "I think philosophically people in NASA support it," says William J. Huffstetler, director of NASA's New Initiatives Office, a project planning group at the Johnson Space Center in Houston. "This represents the kind of activities and spirit that could be found during the Apollo and Gemini days."

So Space Marketing Concepts, an Atlanta-based marketing firm that organizes promotions around the theme of space, is trying to line up a corporate sponsor or two. It would like to entice a fast-food chain, a soft-drink company or another large corporate entity. One idea that it hopes might net a spare million or two from a corporate sugar daddy is a contest whose winners—a child from the U.S. and one from the Soviet Union—would each push a button to launch the Soviet rocket carrying the spacecraft.

Members of the *Lunar Prospector* team were scheduled to make their second trip to the Soviet Union in December to discuss terms of a launch with NPO Energia, a Soviet aerospace organization. Binder says he has a "letter of intent" stating that the Soviets will do the launch for a nominal fee: "They've wanted to get into the commercial space business, they've wanted to fly a full-scale lunar polar-orbiting mission and there's also the whole *glasnost* thing. They want to be our buddies."

Whether it is possible to privatize a scientific mission that has no direct payoff is anybody's guess. But if the team can pull it off, NASA might have to rethink the agency's approach to costly space endeavors. "We're trying to remind people that we can do simple missions the way they used to be done," Binder says. —*Gary Stix*

THE ANALYTICAL ECONOMIST

Second-class jobs

Small business is part of the American dream. The term calls to mind the mom-and-pop grocery store, as well as the high-technology start-up that makes millionaires of its first dozen employees. Conservative policymakers throughout the past decade have averred that small businesses are the primary engine of job creation in the U.S., pointing for evidence to Silicon Valley and the "Massachusetts Miracle."

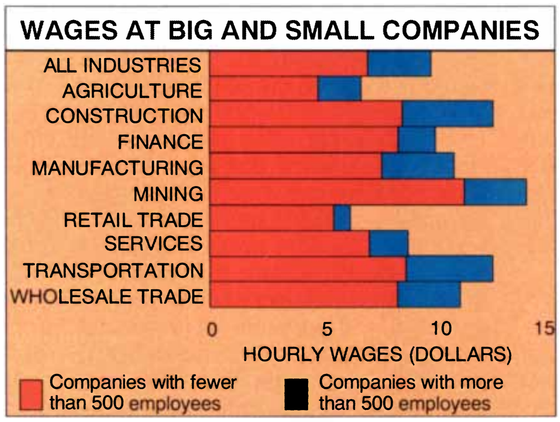
Not for economist James Medoff of Harvard University. "That's just garbage," he says. "It's crazy." He contends that if small businesses do create more jobs than large ones, it is only because those jobs are worth less. In general, he says, small businesses offer lower pay for the same work, fewer benefits and less protection from workplace hazards and unlawful discrimination. In *Employers Large and Small,* published last spring, Medoff, Charles Brown of the University of Michigan at Ann Arbor and James Hamilton of Duke University make the case that the roughly 54 percent of U.S. workers employed by companies of fewer than 500 people are second-class citizens compared with their big-company brethren.

Fewer than 500? That's the definition the Commerce Department uses to compile statistics on small business. Although stereotypical companies with a handful or two of workers account for the majority of small firms, it is these "big" small businesses that account for the majority of small-business employment. Investment banking house Kohlberg Kravis Roberts & Company, which engineered the takeovers of RJR Nabisco, Safeway and Beatrice, is a small business, for example.

If small businesses aren't necessarily small, neither are they generally high-tech entrepreneurial start-ups. High-technology manufacturing by small companies is growing significantly faster than is manufacturing as a whole, but it is still only a small fraction of the sector's employment. According to the Small Business Administration, the economic sectors dominated by small businesses are agriculture and forestry, wholesale trade and retail trade. (The rapidly growing service sector, which encompasses everything from fast food to corporate law, is evenly split between large and small firms.)

Myths aside, just what role do small businesses play in creating jobs? Between 1976 and 1986 the U.S. economy grew by about 22.3 million jobs. Small businesses accounted for 57 percent of that growth, just slightly more than might be expected from their share of total employment. Assertions that small firms create 80 percent or more of new jobs ignore the fact that more than 60 percent of small firms close within six years of starting, taking their employment with them.

The jobs small companies do create fall short on several counts. Numbers from the Bureau of Labor Statistics show that average pay is about 30 percent less than that in big companies. Even after taking into account the generally lower skill levels of small-business workers, Brown says, the differential for comparable work is about 10 percent. Small employers are also significantly less likely to offer fringe benefits such as health insurance or paid vacation time.



SOURCE: CURRENT POPULATION SURVEY, 1983

The supposed heady entrepreneurial atmosphere of a small company does not seem to compensate workers for lower wages and benefits—they quit far more often than their big-company counterparts. Moreover, a small firm in any given industry is more likely to lay off workers than is a large one.

Paradoxically, Medoff says, contrasts between working conditions at small and large firms bring to mind a similar contrast in Japan—a country whose industrial organization is generally considered quite different from that of the U.S. There the lifetime employment of workers in the enormous conglomerates is buffered by a second tier of subcontractors, whose job security and wages are substantially inferior. Although small U.S. companies are not so closely tied to large ones, the effect on employees appears much the same, he notes.

All these disadvantages raise the question of why small companies—other than those on their way to becoming big companies—continue to thrive and why people work for them. Medoff contends that small businesses enjoy undue protection from local and national authorities. Small firms are explicitly exempted from certain regulations governing pollution, financial disclosure, discrimination and workplace safety. This boon naturally lowers their cost of doing business, Medoff says: "If the I.R.S. promised you that you'd never be audited, wouldn't you cheat?"

In addition, small firms may also receive de facto regulatory exemption because the government does not police them. A 1987 study by David Weil of Harvard found that factories employing fewer than 100 people were roughly one sixth as likely to be inspected by the Occupational Safety and Health Administration as those that employ more than 500.

Medoff acknowledges that regulations may disproportionately burden small firms, but he argues that issues such as worker safety should take precedence. If compliance costs small companies so much that they can no longer make a profit, he says, "maybe they should be out of operation."

In their book, Medoff, Brown and Hamilton suggest that small companies have achieved their relative freedom from regulation by selling the image of small business as job engine to the public and to politicians. Political action committees that represent the small-business view, they point out, make up about a quarter of all PAC spending in national elections, behind large corporations and just ahead of organized labor. Furthermore, small businesses are small only on a national scale—a company that employs several hundred people is generally a major force in its hometown.

That may help to explain why people work for small businesses: in many towns, there are no large employers. Brown also points out that large companies have longer waiting lists for applicants, and so those who need jobs in a hurry may have to settle for a small company. But should those small companies continue to hire and employ their workers under a different set of rules than their larger counterparts?

—*Paul Wallich and Elizabeth Corcoran*

MATHEMATICAL RECREATIONS

Tools for computer graphics make an invisible world seem less alien

by A. K. Dewdney

On the first day of the earth year 2991, the *Armstrong* interstellar spacecraft touched down on the fourth planet orbiting the star Tau Ceti. The *Armstrong's* crew detected movement from the northeast and focused the ship's camera on a distant rocky cliff. There on a ledge was a nest made of rock crystals and an egg that resembled a fried pastry, a French cruller to be exact. The egg began to dissolve, and from it emerged a snakelike creature composed of two intertwined rings. The mission biologist quickly dubbed it a "gorgonoid." As the probe moved closer to get a better look at the gorgonoid, the creature stiffened in fright and bounced off the cliff into an acetylene river.

To be sure, the world of the gorgonoid is science fiction, but its image resides in a computer at the IBM Thomas J. Watson Research Center. Clifford A. Pickover, a graphics wizard at IBM, created the alien I call the gorgonoid to demonstrate powerful, new tools for computer graphics. He has developed the techniques as part of his mission to help other scientists visualize the intricate shapes produced by physical phenomena or derived from theories. Pickover, whose microscopic biomorphs appeared in these pages in July 1989, describes his creations as "graphics from an unseen world" [*see illustration on opposite page*].

Although the gorgonoid egg looks like an alien life-form, it is actually a model based on physical principles discovered in terrestrial laboratories. If one could peel away the "shell" of the gorgonoid egg, one would find a frame composed of two "wires." One is bent into a circle; the other winds around the circle in a spiral that rejoins itself. If the wires were charged with a certain voltage, they would generate an electric force that would be stronger at points close to the wire frame than at points farther away. Pickover's computer program finds all the points representing a given strength of the force and then plots them to form the shell of the gorgonoid egg. Pickover calls this imaging technique voltage sculpture.

Pickover engages in the art of voltage sculpture to depict a variety of atomic structures from single molecules to the complex spiral of DNA. Because the voltage sculpture displays the electric forces surrounding the molecules, investigators may be able to deduce how some molecules produced by living cells can fit certain receptor sites in other cells.

The young gorgonoid is not a voltage sculpture but what might be called a worm necklace. Like the egg, the gorgonoid is based on two wire loops, one winding around the other. To make the body of the gorgonoid, Pickover adorns the wires with spherical beads: large ones for the circular wire, small ones for the spiral one. The beads are spaced evenly along the wires, and consecutive beads overlap.

The mature gorgonoid is a worm necklace made of three wires: the first wraps around the second, which in turn curls around the third. The mature gorgonoid also has an eye made from three nearly concentric spheres that intersect to form an iris from one sphere and a pupil from another.

A mature gorgonoid can spot a predator a mile away through an ammonia haze—an important survival strategy when it is being stalked by a pacmantis. This cup-shaped creature spends half its time basking in the rays of Tau Ceti. But when the pacmantis gets hungry, it rolls on the ground, opening and closing its mouth like the computer sprite known as Pac-man.

The anatomy of the pacmantis is no more complicated than the morphology of the gorgonoid. To bring the pacmantis to life, Pickover creates a computer pendulum. He simulates a ball that is tied to one end of a rigid wire; the other end is connected to a pivot that allows the wire and ball to swing freely in all directions.

Initially, the pendulum is pushed sideways with a certain velocity and swings down under the influence of gravity. After it swings back and forth, it arrives at a point that is a certain distance away from its starting point. In the course of its subsequent swings, the ball covers most of the available space within the sphere of possible positions.

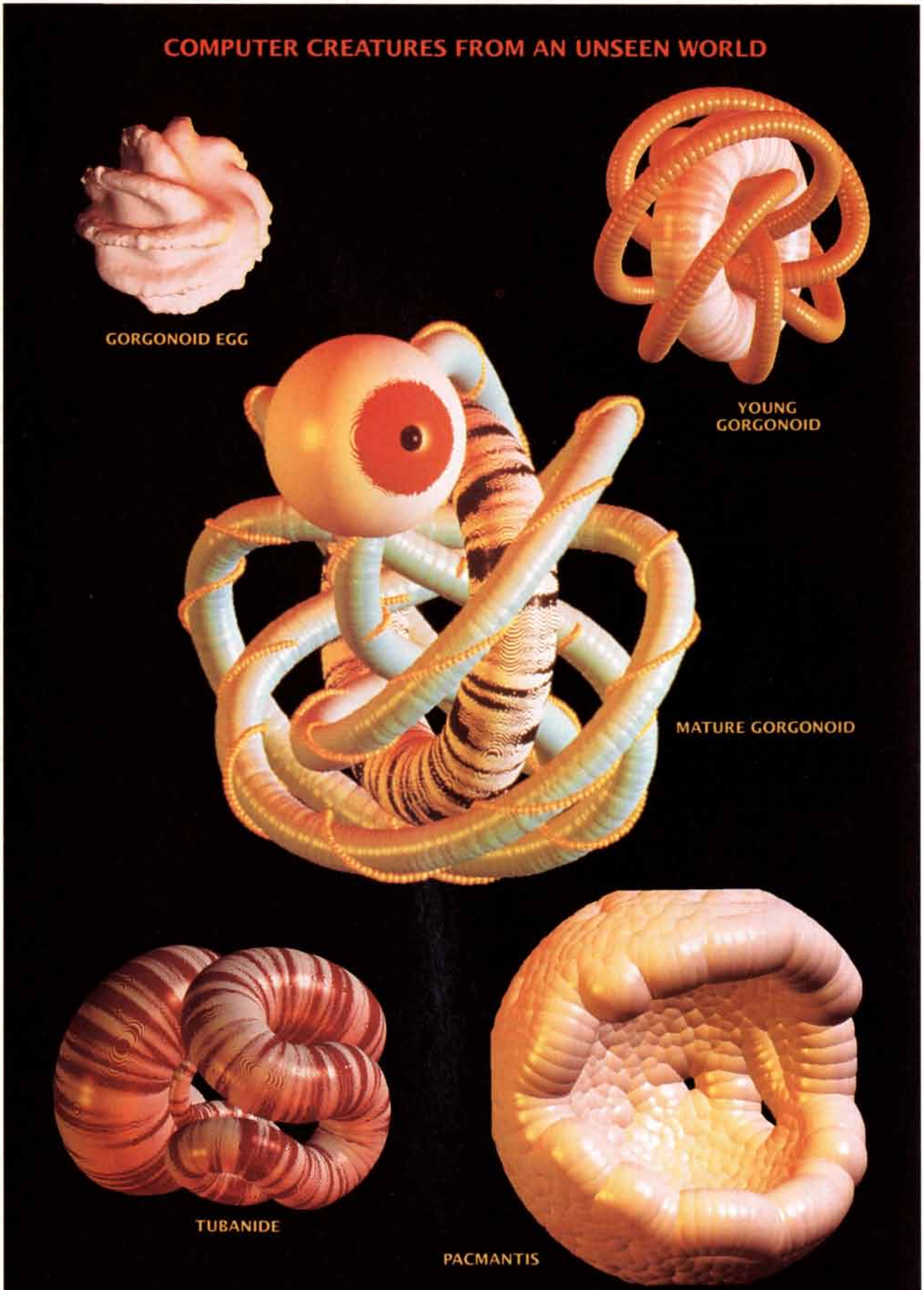
As the pendulum swings, Pickover's computer periodically takes snapshots of the ball. When many images of the ball are displayed simultaneously, they form a shell. By rotating the shell 90 degrees, one sees the exoskeleton of a pacmantis in its proper orientation.

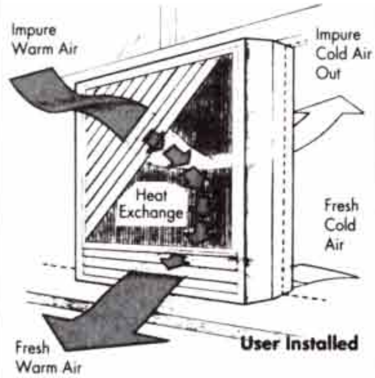
Although the pacmantis will occasionally munch on a gorgonoid, it prefers to dine on the tubanides that live in ammonia oceans. These succulent shellfish resemble certain ammonoids that flourished on the earth during the Mesozoic era. The tubanide has an attractively striped shell, which begins as a straightforward open spiral but subsequently curves back on itself—like the product of a demented tuba maker. As a result of its twisted shell, the tubanide tumbles as it swims, making it easy prey for the pacmantis.

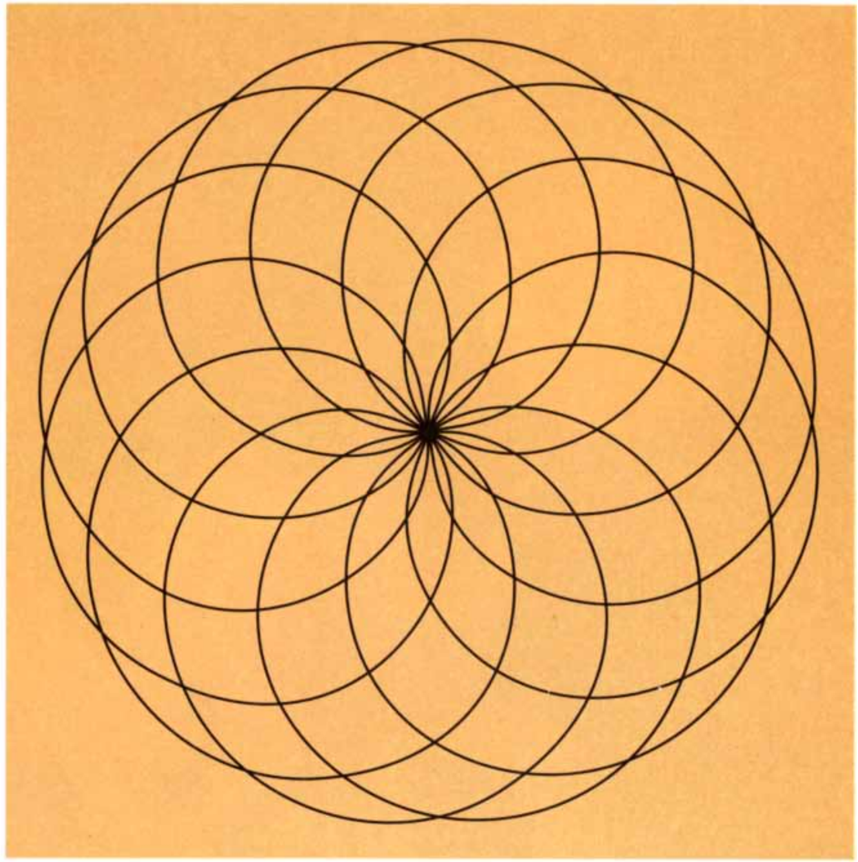
Tubanides were spawned from Pickover's collaboration with Australian conchologist Chris Illert. Pickover and Illert studied a bizarre ammonoid called *Nipponites mirabilis.* Most ammonoids, like the modern *Nautilus,* have regular, logarithmically spiraling shells that allow the animal to move smoothly through the water. During the early stages of growth, the shell of *N. mirabilis* grows much like that of other ammonoids, but later it twists and turns

ANSWERS TO NOVEMBER COLUMN

To exercise mental muscles, I challenged readers last November to solve three little posers. First, if someone shuffles an ace, king and queen, places them face down, removes one at random, then draws one of the remaining cards, the probability that it will be an ace is one third, the same as if none of the cards had been removed. Second, the proof that no one works does not itself work. It subtracts apples, eight-hour work days, from oranges, 24-hour work days. Finally, if a bottle and a cork together cost $1.10 and the bottle costs $1 more than the cork, then the cork must cost five cents.


COMPUTER CREATURES FROM AN UNSEEN WORLD
GORGONOID EGG
YOUNG GORGONOID
MATURE GORGONOID
TUBANIDE
PACMANTIS





A spherical Lissajous figure

in all directions. Illert hoped to investigate such unusual growth patterns by searching for a mathematical description of the irregular spiral.

He came up with a formula that has a simple interpretation. The orientation of the opening of the shell determines the direction of the shell's growth. In ordinary spiral growth, the orientation of the opening would remain fixed in relation to the adjacent rings of the shell. But the growth of *N. mirabilis* can be nicely simulated if the shell opening is rotated according to an exponential rule: as the shell grows, the opening rotates more and more. This hypothesis produces a proper ammonoidal appearance for young *N. mirabilis* and yields a twisting, irregular-looking spiral for the older animal.

Pickover and Illert demonstrated that the tubanide is a good model for the adult *N. mirabilis.* To render tubanides in living 3-D, Pickover used the worm-necklace technique. He colored the tubanide by alternating between crimson and white spheres.

The concepts behind worm necklaces and voltage sculptures are as simple as advertised, but I am tempted to add the performer's warning: "Do not try this at home!" After all, Pickover has access to computers specifically designed for graphics. His computer system can automatically shade and hide surfaces; it can show light from several sources reflecting off a surface; and it can produce, in an instant, a view of any three-dimensional object from any angle. The skin of Pickover's creatures is therefore only a few keystrokes away.

Although such facilities are not available in home computers, Pickover would not discourage amateur programmers from creating some exquisite alien graphics called spherical Lissajous figures [*see illustration above*]. In 1857 the French mathematician Antoine Lissajous first described these sinusoidal figures that today parade on the screens of oscilloscopes. A single Lissajous curve is traced out on the screen as a bright dot moves up and down and side to side any number of times and eventually returns to its starting point.

Spherical Lissajous figures have the same properties as their two-dimensional relatives, except that they lie on the surface of a sphere. To represent this three-dimensional curve in three dimensions, one needs three separate equations, each involving a single variable, t, which one can think of as time.

$$x = R\sin(At)\cos(Bt)$$
$$y = R\sin(At)\sin(Bt)$$
$$z = R\cos(At)$$

R, A and B are constants. For each value of t, the three formulas collectively specify a single point in three-dimensional space. As the value of t is incremented (as time passes), the formula produces a succession of points that generate the spherical Lissajous curve.

By setting values for R, A and B, the rates at which the curves oscillate, one may generate fascinating figures. The curve will close back on itself unless the ratio of A to B is an irrational number, not a likely event in a computer.

Readers can devise a simple computer program to view a spherical Lissajous curve on a two-dimensional screen. The program should first ask for the values of R, A and B. The program should then enter a loop where the value of t is increased from, say, 1 to 1,000. For each value of t, the program should calculate x and y according to the formulas. The x coordinate, for example, will multiply R by the sine of A times t, then by the cosine of B times t. Finally, the program should plot the point (x, y).

Some caveats accompany this algorithm. First, the numbers x and y may have to be specially modified so that the point being plotted will appear on the screen. If necessary, add a suitable constant. Second, the values of t may have to change more gradually to produce a solid-looking curve rather than a string of widely spaced points.

Pickover is delighted at the potential of his tools and similar techniques to help not only scientists but also artists. As examples of artists who have already exploited such possibilities, he cites William Latham of the IBM U.K. Scientific Center, John Lewis of the New York Institute of Technology and Donna J. Cox of the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign. As electronic devices shrink and get faster still, even computers like Pickover's sophisticated system will find their way into smaller and more affordable boxes. The implications for science and art will be equally great as more and more graphics emerge from unseen worlds.

FURTHER READING

COMPUTERS, PATTERN, CHAOS, AND BEAUTY. Clifford A. Pickover. St. Martin's Press, 1990.

BOOKS

Earth as a peppercorn, the greening of Gondwana, travel in the tropics

by Philip Morrison

THE FLOWERING OF GONDWANA, by Mary E. White. Photography of fossils by Jim Frazier. Princeton University Press, 1990 ($49.50).

Like the trademark marsupials, the modern Australian flora are strangely distinctive. The unique eucalypts, a protean genus of plants as diverse as stunted shrubs and towering trees, easily dominate a wide rim around nearly the whole landmass. "If a continent can be... characterized by an aroma," it is the scent of the oils of eucalypt leaves and twigs that symbolizes Australia. The arid plateau surrounded by that rim of dominant eucalyptus bears a scrubby vegetation, one dominated almost as thoroughly by another genus, the acacias, which are tropical legumes, their thorny kin found in all hot lands. Fine photographs present a magnificent white ghost gum of the rimland; showy pink eucalypt blossoms, their pollen conspicuously yellow and thrust well forward, flowers that have drawn not a bee visitor—here bees are few—but a honey eater bird; and the golden brushblossoms of a wattle, a thornless acacia. Gondwana flowered forth at last.

FOUND ONLY IN AUSTRALIA, **Petrophile linearis** *is a member of an ancient Gondwanan family of plants, the Proteaceae.*

Gondwana is the name given to the great southern supercontinent wherein long ago South America, Africa, Antarctica, Australia and most of South Asia nestled cheek by jowl. It was named for an ancient people of the Central Indian highlands, "the land of the Gonds." (Affixing an extra "-land" is redundant, dropped officially in 1980.)

This enticing book offers in spectacular support of its clear text about 400 large and colorful photographs of plants, fossil and alive, a comprehensive sequence of apposite maps, and many diagrams and charts. Its Sydney-based author believes she was born a botanist. The case is strong. Her parents before her "compiled a *Flora of Southern Africa,*" and now she spreads before us the best book on paleobotany this reviewer has seen, a book based on her own professional studies and fossil collecting over 35 years. She has as well enlisted splendid help from talented specialist friends.

In Australian tablelands—it is the flattest of continents as well as the smallest—are some of the very oldest rocks on the earth; not far off, we see the earliest signs of life, in fossil stromatolites 3.5 billion years old, their counterparts alive yet on the tidal flats. After a stirring watery prelude, the author exhibits stage by stage the history of Australian plants on land. Gondwana's is an old, old story; the narrative covers about 600 million years. In those early days the seas and tidewater teemed with organisms, yet "still the land is lifeless." Came a greening that lined every water margin; soon the filigree of green caught all the land and the "bloom of green... becomes a pelt."

The method behind the entire argument is an instance of the spiral growth of science. When Joseph Hooker, Darwin's botanical friend and mentor, botanized in Australia, he believed that related forms found widely separated in the past or the present clamored for diffusion of species across wide oceanic sea barriers. The Kauri pine is one of the distinctive southern conifers that thrive today in the refuge of the tropical Australian coastal rain forest. One striking spread celebrates a living branch of that pine, two tightbound white tree cones among oval green leaves. The same form grows wild today in the Argentine, in the Falkland Islands, near the Cape of Good Hope, on the lonely rock of Kerguelen halfway to Australia, in New Zealand, but nowhere else.

The old botanists took this as evidence that some kind of land connection must once have bridged the roaring ocean; pines cannot swim (Darwin tested the survival of seeds in seawater). Continents must have been linked; just how, the early botanists could not know. Still stronger is the claim of the paleobotanist. Here in sharp focus a fossil twig of that very Kauri pine imprints rock that once was an ancient lake bottom not so far from Sydney. Some show among fallen twigs and leaves the prints of immature little cones that the species still sheds each season. The whole conifer and cycad rain forest assemblage was abundant over all Australia 175 million years ago. But far off in South Africa, rocks from that same epoch hold in abundance pollen which is unmistakably that of the southern conifers.

Like the 19th-century students of plant geography, our author does not invoke seismic epicenters or magnetic records or any other physical evidences of continental drift. Such evidence is welcome, quantitative, profoundly confirming, yet in a way not essential. How else could Occam's razor shave the story of five southern landmasses all greened with the same early unflowering forests, now oceans apart? The only other argument needed is reliable correlation in time, the harvest of two brilliant centuries of paleontology. The botanists were first with the story, correct if not compelling. Now botanist White can confidently cite dates and climates and positions for the opening and severing of overland routes, firm geophysical input to her intricate and revealing botanical logic based on fossils visible to eye or microscope.

The first evidence of flowering comes about 140 million years ago, when the South Atlantic rift parted the shoulder of Brazil from the Gulf of Guinea. The earliest flowers, magnolialike and laurellike, left pollen traces there. Earlier still, flowering plants made their way to north and south. Later they reached Australia by way of South Africa, where 50 or 60 million years ago the vegetation was like that of Australia and New Zealand at the same time. The last link was with icy Antarctica.

Consider an exercise for the reader who may be a paleobotanical novice. How can blue gum and thornless wattle, found nowhere else, now dominate Down Under? Why were they not immigrants from afar? The answer is clear in rock and map: for at least the 30 million years since the first eucalypt pollen appears in Australian rock, all land contact has been severed. Old Gondwana was disassembled, and Australia cast adrift to compound fragrances of its own.

Ever since then Australia has been more than an ark of survival; it is a lonely Eden. It would in time become home to new forms, passengers born during the voyage. Grasses entered just in time to catch the marsupial ark. Eventually grasslands would allow the "short hopping jumps of... early Marsupials" to strengthen into the longer rhythms of the kangaroo. Unique flowering plants as strange as the stumpy black desert grass-trees and the red-stemmed kangaroo paw decorate the close of the chronicle. The triangle of India has long since shoved ashore to fold and wrinkle broad Asia, but the docking of Australia is still under way, although only yesterday the nimblest of mammals jumped with their dogs across the narrowing gap.

The Thousand Yard Model or, The Earth as a Peppercorn, by Guy Ottewell (paperbound, $5). **Astronomical Calendar 1991**, by Guy Ottewell (paperbound, $15). Both from Astronomical Workshop, Furman University, Greenville, SC 29613.

3-D Star Maps, by Richard Monkhouse and John Cox. Harper & Row, Publishers, 1989 ($15.95).

The sky is as three-dimensional as your house. In the crisp title of his 16-page brochure, celestial geometer Guy Ottewell describes an explicit and imaginative solar system model for impromptu realization outdoors. No atlas page and no orrery can do an honest job of mapping the solar system proportionately. But walk out 1,000 yards from a bowling-ball sun, far past pinhead moon and peppercorn Earth, to a distant coffee-bean Neptune and beyond, and you gain a real grasp of the scale of the solar system.

A poet's concreteness enriches the entire account. How much better to code the planets, not as nine spheres differing only in diameter but as objects with roughly the right sizes. That all planets are near-spheres is easily learned but that Jupiter is chestnut size to Earth's peppercorn is memorable without numbers. Nor is the author insensitive to practicality. A bowling ball is heavy and expensive. Better to use any eight-inch inflatable ball and streetwise to send a kid back the 40 yards from Mars to fetch it before you all enter the long asteroid gap; some passerby might take it while you are blocks away. This booklet is well worth the price for readers who will never do the 1,000 yards, an eye-opener for informal teachers, a piece of virtuoso pedagogy simple and right from first to last.

Around midday on this coming July 11, the heavens will spectacularly proclaim the unseen depth of the sky vault in the most visible of ways. The new moon will cross centrally in front of the sun and hide it for nearly seven minutes. Every year Ottewell's celestial calendar offers monthly sky maps for planets and stars, a geometric treasury of the year's events in diagram and table, the moon in its phases, showers of meteors, comets and more. Solar eclipses are well plotted; this year's, pretty neatly centered in space and time, will last about as long as any eclipse can. A novel diagram shows the bulk of the moon's shadow as projected on the plane of its earth orbit for each new and full moon of 1991; the symmetry of this particular eclipse is made evident. The path of totality crosses Baja California under desert skies, an easy tourist's drive for the mobile and eager public of San Diego and Los Angeles, so that this may be the most sought-out eclipse ever. (If you plan to go, look further for practical detail on when and where to stand.)

Beyond the sun's family the abyss of the sky can be plumbed only by the full powers of modern astronomy. In a brilliant work of computer graphics and modern typography, two London author-designers have given us a magical view of stars and galaxies at depth. They provide about two dozen stereo pages, red and green dotting a grayish ground, that induce a striking perception of depth. The trick is the use of cardboard glasses, right eye green and left eye red, two pairs of which are thoughtfully provided.

Every stereo page has a carefully annotated key map in white on black. One set of maps shows the bright stars of the whole sky; one all the stars nearby, most of them telescopically faint; and a third set, galaxies and their clusters far out into the cosmos, the distances known by redshift. The depth perceived is to scale, appearing here nearer, there beyond, the apparent plane of its page. The maps are a little trickier: distances have been represented not linearly but logarithmically to allow representation of the vast range of cosmic scales.

Professor Monkhouse teaches software at Imperial College, London, and his colleague is an artist and programmer. Their compact book is rich both graphically and intellectually, with a lot to say about astronomy in little text with no long tables. Coordinate grids are plotted, bright stars are mapped by color classes, the methods of distance estimating are summarized and so on. There is plenty of novelty here to attract casual stargazers and more still for all who would enjoy a look at Barnard's Star or Centaurus A along the third cosmic dimension.

A NEOTROPICAL COMPANION: AN INTRODUCTION TO THE ANIMALS, PLANTS, AND ECOSYSTEMS OF THE NEW WORLD TROPICS, by John C. Kricher. Princeton University Press, 1989 ($47.50; paperbound, $16.95).

RAINFORESTS: A GUIDE TO RESEARCH AND TOURIST FACILITIES IN SELECTED TROPICAL FOREST SITES IN CENTRAL AND SOUTH AMERICA, by James L. Castner. Feline Press, 1990 (P.O. Box 7219, Gainesville, FL 32605) (paperbound, $16.95).

Every day now big jets rise off wintry runways for the neotropics a few hours away. Almost everyone who has found pleasure along the bookshelves of science has dreamt of the green depths so eloquently reported from the days of Alexander von Humboldt, past William Beebe and P. W. Richards, to the moderns who documented the changes that now feed our anxieties. These books, compact, brimful and personal, may just persuade you to seize the day while the petroleum still flows.

Kricher is compact and richly substantive. A biologist of New England woodlands who has spent much time in the rain forests, he has written an intense and lively field guide, his keen eye on what unity can be found among the gorgeous excesses of diversity. Why is tropical nature so diverse? Our theories are also diverse and less than compelling. Perhaps biodiversity is itself explosive by mutual evolution within a rich tapestry of interacting organisms: once the density of species exceeds a few times what is common in temperate zones, it may simply run away through its own opportunism.

Here are toucan and quetzal, the superb kapok, tree sacred to the Maya, the oilbirds whose echoing calls fill the dark Trinidad caves, the moss-backed sloth, the clamorously self-advertised golden and crimson poison frogs. Prepare for the feeding flocks that suddenly animate the silent forest with a universal bustle, to leave within minutes, the narrow path as still as before. For an episode of memorable horror, praise the noisome giant cockroaches under the flashlight beam, creatures so much at home burrowed into the droppings of bats and humans. In that world of just-in-time, the swift turnover of nutrient keeps most natural inventories as frugal as Toyota's.

There are no highways in the upper Amazon; you must get there by boat or by air. James Castner is a young entomologist from Florida who has gone both thriftily and well. His flair is not only for the practical but also for the feel and texture of places and people. That makes his small book a credible and convincing firsthand guide to a selected three dozen lodges, inns, camps, field labs and nature preserves in seven countries of the neotropics, from Trinidad to Amazonian Peru. It offers help equally to the short-term naturalist tourist and to the student or researcher who plans an entire season's campaign (even to lists of sources of grant support). He tells where you might cross paths with old Beebe himself, how with good luck you can expect to see the cloud forest quetzal or the wary large mammals of the forest and how to visit by invitation and chutzpa among the working biologists. There are many lists, among them books, biologists, travel agencies and Spanish biological nouns. Site photographs, names, telephone numbers are all here, as well as how to get good maps and to rent cars that work: neither is always easy.

Don't go alone into the forest, but whenever you go take a book or two, wait patiently beside some fruit-laden tree to see just who else comes by, and drop a postcard in report to one of these enthralled authors.

THE CITIES OF ANCIENT MEXICO: RECONSTRUCTING A LOST WORLD, by Jeremy A. Sabloff. Thames and Hudson, 1989 ($19.95).

Author Sabloff is among those critical archaeologists who have for a decade or so sought to test the vision—the visions?—of the past generation or two of scholarly Mayanists, who founded so much on their subjective inferences. It is an agreeable surprise to find his compact volume built, city by city, around eight two-page vignettes, each a little essay putting the reader into the life of some unnamed person of the past. What could be more like the donnish speculative reconstructions by Professor Sir Eric Thompson decades back, all of them good reading as well?

Among Sabloff's major points is the depth of unity across time and space among the Mesoamerican cultures. His vignettes begin with the Olmec, 2,500 years earlier than the Aztec pyramid of Tenochtitlan. Nor is the empathy of the essayist limited to the elite, the priests and rulers about whom we learn directly from the monuments and the glyphs. The cast is much larger today; we can be sure there were weavers for the ornate costumes, as the poet Brecht reminded us long ago. Nowadays there is careful digging around local peasants' houses to assess the practice of agriculture, hardly to be found by excavations at temples alone. A major section is titled "How Do We Know?" and each vignette is supported there by an account of the basis of its analogies, some of course still conjectural.

The colossal basalt portrait heads that are the famous mark of Olmec times were certainly not hauled by oxen. "Hundreds of men" spent immense effort. But they were skillful enough not to damage the cotton-padded 15-ton carving they maneuvered slowly up the earthen ramp from the raft that had brought it nearby. The husband tells his story at day's end to his weaver wife, center of the Olmec vignette. His work must for a while be at the ropes, while she weeds their upland field alone, laboring to sever taproots with a stone ax. Can she manage time for loomwork on the cloak commissioned of her by the chief's son? All too soon the leader's image will be uncovered in high ceremony in the hilltop plaza of the small city near their village, one hot early summer day on the Gulf coast about 1000 B.C.

Astronomer and priest at Monte Alban, where on two sacred days a year the zenith sun casts no shadow at midday; aristocratic ball player staking fame and even life on the game at the great court at Uxmal; architect of the monument to Palenque's ruler preparing the day's funeral 15 years after work had begun on the temple tomb; and a flower farmer who toils in the muck of the lake gardens at the southern edge of the Valley of Mexico—these are the people who come clear in the shadow of the past. The intricate fabric of this chronicle is shown us not as by a confident merchant but rather as though we have visited a weaver of ideas, still at his loom.

The illustrations in black-and-white are handsome and helpful, in particular the maps and plans, and a final gazetteer lists some 50 cities, with a paragraph for each that adds to some practicalities a crisp comment on the site's meaning.

ESSAY

Technology and competitiveness

by B. R. Inman
and Daniel F. Burton

The once commanding U.S. advantage in technology has slipped away. The culprit? International competition. The consequence? U.S. government and business relationships have been fundamentally altered, with major implications for U.S. science and technology policy. Although the repercussions are still being sorted out, the basic message is clear: when it comes to technology, U.S. public policy can no longer afford to be preoccupied with basic research and military issues; economic security and industrial competitiveness are also vital considerations.

Five technology stories illustrate the changing public policy debate: consumer electronics, semiconductors, superconductors, the FSX fighter airplane and high-definition television. Taken together, they hold important lessons for U.S policymakers.

The decline of U.S. consumer electronics was the first shot across industry's bow. American inventors pioneered this field with such innovations as the phonograph, radio broadcasting, television receivers, the transistor, color television, portable radios and videocassette recorders. Until 1970, American industry dominated the consumer electronics business. Since then, it has lost virtually the entire market. The industry learned too late that technical breakthroughs must be relentlessly followed up with manufacturing systems capable of high-quality, low-cost, volume production. The federal government contributed to the demise of the industry by allowing imports to be dumped in the U.S. market while foreign markets remained essentially closed to American producers.

If the collapse of consumer electronics could be dismissed in some circles as a low-wage, commodity industry without real importance to the U.S. economy, semiconductors could not. Semiconductors personified Silicon Valley high tech to many Americans, and the trials of this industry captured the public imagination. Following on the heels of consumer electronics, the sequence of events seemed all too familiar. Invented in the U.S., semiconductors spawned a vital American industry that captured a worldwide market. In 1976, however, Japan launched the Very Large Scale Integration (VLSI) program. Ten years later the Japanese had captured 65 percent of the world market for computer memory products. The lesson learned was a hard one: even high-tech U.S. industry could be devastated in short order by foreign competition. The decline of the U.S. semiconductor industry reaffirmed the fact that leading-edge product technology by itself is not enough; process technology, world-class manufacturing capability and government-industry cooperation are also essential.

Close on the tracks of the semiconductor fiasco came the superconductor breakthrough. The series of events that followed highlighted two key issues. First, these events demonstrated just how quickly technical knowledge is diffused around the world. The finding at the IBM laboratory in Switzerland set off an international chain reaction of intense verification efforts and new research advances throughout Japan, the U.S., Europe and China. Second, superconductors riveted U.S. government attention on one overriding issue: how to translate basic research into products and processes that yield the sought-after military and economic benefits. The demise of the U.S. consumer electronics and semiconductor industries was not lost on public officials.

In the spring of 1989 another technological issue made headlines—the FSX fighter. The debate revolved around a proposed joint venture between the U.S. and Japan to produce a new tactical fighter based on General Dynamics' F-16 aircraft. Critics feared it would be a giveaway of American technology and would help Japan achieve its long-standing goal of creating a civilian aerospace industry. Proponents argued that it would actually keep Japan from building its own plane and that the U.S. would gain access to Japanese technology. Although a modified version of the agreement was eventually signed, the debate brought out a deeper policy question: Should American economic interests be given as much weight as traditional security concerns in the formulation of U.S. foreign policy?

For the past two years, Washington has been embroiled in a policy dispute over high-definition television. Backers of HDTV hail it as the most significant development since color television and think it will drive several related technologies. Critics have downplayed it as a technical advance that is limited to a narrow market. For many, the controversy has come to symbolize the government's difficulty in developing a coherent technology policy. Technical feasibility, standards, market demand, military implications and the role of foreign industry are all hotly debated issues. In addition, the HDTV controversy has drawn attention to the fact that a solid manufacturing base is essential to technological leadership. It has also emphasized the close linkages among several related technologies. As one government official stated, "The next generation of televisions is the next generation of computers."

Taken together, these developments have prompted recognition of the need for a workable U.S. public policy framework for technology. Although the debate is still unfolding, there is growing awareness that action is essential in four areas: strengthening the U.S. manufacturing base, upgrading the policymaking machinery for technology, rebuilding the technological infrastructure and widening federal R&D efforts to include more commercially relevant technology. In each area, the need is clear; the means are less apparent.

The debate about technology and competitiveness has enormous implications for U.S. national security and economic growth. As such, it promises to have a far-reaching impact not only on the strategic objectives the nation pursues but also on the institutions that shape our policies. In the past, U.S. policies have tended to impede rather than to assist efforts of the private sector to bring new technology rapidly to market, because the U.S. government has not addressed the commercialization of technology as a public policy issue. We can no longer afford to ignore this issue. Arcane topics, such as technical standards, patent law, R&D funding and intellectual property, that once were viewed as the responsibility of obscure bureaucrats will increasingly engage public officials at the highest levels. And government agencies will have to be refocused to address the new priority of technology and competitiveness.

The broad contours of a responsive technology policy are beginning to take shape. The challenge ahead is to implement specific policies and programs that produce concrete results.

B. R. INMAN is president of Inman Associates. DANIEL F. BURTON is executive vice president of the Council on Competitiveness.

SCIENCE/SCOPE®

A lightweight laser-illumination warning system may soon aid combat crews in avoiding laser-supported weaponry. The system, developed by Hughes Aircraft Company, is designed to provide tactical aircraft, combat vehicles and ships with data on a threat laser's bearing, pulse rate, width and intensity. The sensor provides a 190-degree azimuth and 110-degree elevation field of view. The warning unit weighs less than two pounds and has been successfully flight tested aboard an F-4 and A-7D aircraft.

Military aircraft may stand a better chance of avoiding detection as a result of technology being developed by Hughes for the U.S. Navy. By combining an infrared sensor and a carbon dioxide laser rangefinder through a single aperture, Navy aircraft may be able to accurately locate targets without emitting signals or energy that could trigger enemy detection and countermeasures. The infrared sensors will locate targets passively, by detecting heat difference between objects and their backgrounds, while the laser will provide accurate range information over great distances and reduce the sensor's false alarm rate.

A new onboard signal processing chip can increase the efficiency of communications satellites. The Hughes-designed very large scale integrated (VLSI) circuit will allow satellites to sort and arrange simultaneously received signals and retransmit them in a single "downlink." Normally, a satellite returns signals to Earth in multiple downlinks, in the same configuration as the signals are received. This splits the satellite's power among the various transmissions and requires a large ground station. A single downlink enables the use of small, simplified ground stations. Without these new VLSI circuits, the electronics to perform these functions would be the size of a filing cabinet.

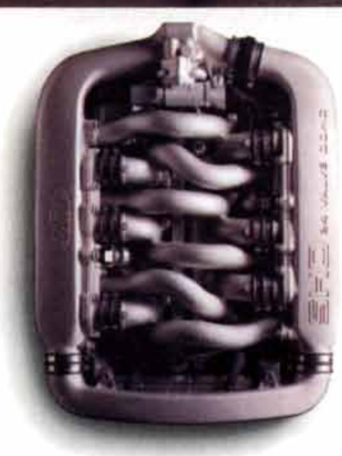
New applications of gallium arsenide technology will improve the performance of satellite communications receivers. Hughes is developing a new technology, called high electron mobility transistors (HEMT), for the next generation of advanced space communication equipment. HEMT devices are built on an indium phosphide substrate with alternating layers of aluminum indium arsenide and gallium indium arsenide. Laboratory tests indicate a factor of 15 improvement in the sensitivity of receivers using this new technology. Improved sensitivity will reduce the size of receiving antennas required by communications satellites, lowering their weight and their manufacturing and launch costs.

Three-quarters of a billion dollars are on the way. And in 1990, Hughes Aircraft Company's Ground Systems Group is anticipating many more new contracts, in Air Traffic Control (ATC), Air Defense, and the commerical sector. These new programs have created excellent career opportunities for experienced and motivated individuals in systems engineering, software engineering, and test engineering. Appropriate background is preferred in some positions. For immediate consideration, please send your resume to Hughes Aircraft Company, Ground Systems Group, Dept. 1343-T, PO Box 4275, Fullerton, CA 91634. Equal Opportunity Employer. Proof of U.S. citizenship may be required.

For more information write to: P.O. Box 45068, Los Angeles, CA 90045-0068

HUGHES

Subsidiary of GM Hughes Electronics



24-VALVE
220 HP
FORD
TAURUS
SHO

Executive conference room.

When your career's on the fast track, you need a car that belongs there. The Taurus SHO. Its distinctive appointments include fully-articulated sport seats (leather trim optional) and performance instrumentation. The Taurus SHO is perfect for a closed conference, even if your only meeting is with the available JBL stereo and your favorite compact disc.

The business end.

But your executive image doesn't have to be conservative. Nothing sets you apart like Taurus SHO's hidden asset; a 24-valve, 220 horsepower

Super High Output (SHO) V-6, precisely controlled by a newly refined, smoother-shifting five-speed manual transmission.

Performing assets.

Rounding out Taurus SHO's portfolio are a special sports suspension, 4-wheel disc anti-lock brakes, and standard driver's side air bag supplemental restraint system to be used with your safety belt. And with new 16″ cast aluminum wheels, the SHO will quickly improve your outlook.

The Taurus SHO. Because making work a pleasure is the essence of good business.

Buckle up—together we can save lives.

Ford Taurus SHO

Have you driven a Ford...lately?


Gate to gate service.
Standing under the torii gates of Kyoto, you're a long way from where you began this journey. But United made it easy.
With nonstops to Japan from five U.S. gateways, United gives you the convenience you need, and a level of service that makes passing through a pleasure.
Come fly the friendly skies.
UNITED
AIRLINES
TOKYO · OSAKA · HONG KONG · SEOUL · TAIPEI · SYDNEY · MELBOURNE · BEIJING · SHANGHAI · AUCKLAND · SINGAPORE · MANILA · BANGKOK
株式会社
三栄組
昭和五十九年五月吉日建之